Molecular Evolution of Trehalose-6-Phosphate Synthase (TPS) Gene Family in *Populus*, *Arabidopsis* and Rice

Hai-Ling Yang¹*, Yan-Jing Liu², Cai-Ling Wang¹, Qing-Yin Zeng²*

1 College of Life Sciences and Biotechnology, Beijing Forestry University, Beijing, China, 2 State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China

Abstract

Trehalose-6-phosphate synthase (TPS) plays important roles in trehalose metabolism and signaling. Plant TPS proteins contain both a TPS and a trehalose-6-phosphate phosphatase (TPP) domain, which are coded by a multi-gene family. The plant TPS gene family has been divided into class I and class II. A previous study showed that the *Populus*, *Arabidopsis*, and rice genomes have seven class I and 27 class II TPS genes. In this study, we found that all class I TPS genes had 16 introns within the protein-coding region, whereas class II TPS genes had two introns. A significant sequence difference between the two classes of TPS proteins was observed by pairwise sequence comparisons of the 34 TPS proteins. A phylogenetic analysis revealed that at least seven TPS genes were present in the monocot–dicot common ancestor. Segmental duplications contributed significantly to the expansion of this gene family. At least five and three TPS genes were created by segmental duplication events in the *Populus* and rice genomes, respectively. Both the TPS and TPP domains of 34 TPS genes have evolved under purifying selection, but the selective constraint on the TPP domain was more relaxed than that on the TPS domain. Among 34 TPS genes from *Populus*, *Arabidopsis*, and rice, four class I TPS genes (*AtTPS1*, *OsTPS1*, *PtTPS1*, and *PtTPS2*) were under stronger purifying selection, whereas three Arabidopsis class I TPS genes (*AtTPS2*, 3, and 4) apparently evolved under relaxed selective constraint. Additionally, a reverse transcription polymerase chain reaction analysis showed the expression divergence of the TPS gene family in *Populus*, *Arabidopsis*, and rice under normal growth conditions and in response to stressors. Our findings provide new insights into the mechanisms of gene family expansion and functional evolution.

Citation: Yang H-L, Liu Y-J, Wang C-L, Zeng Q-Y (2012) Molecular Evolution of Trehalose-6-Phosphate Synthase (TPS) Gene Family in *Populus*, *Arabidopsis* and Rice. PLoS ONE 7(8): e42438. doi:10.1371/journal.pone.0042438

Editor: Natarajan Kannan, University of Georgia, United States of America

Received March 14, 2012; Accepted July 9, 2012; Published August 8, 2012

Copyright: © 2012 Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from the National Basic Research Program of China (2009CB119104), the Natural Science Foundation of China (NSFC 30800873), and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-J-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: qingyin.zeng@ibcas.ac.cn (Q-YZ); yhailing@yahoo.com.cn (H-LY)

Introduction

Trehalose (α-D-glucopyranosyl α-D-glucopyranoside) is a non-reducing disaccharide in which two glucose units are linked in an α,α-1,1-glycosidic linkage. Trehalose plays important roles in protecting plants from heat, cold, and osmotic and dehydration stress [1,2,3,4]. The biosynthesis of plant trehalose consists of two enzymatic steps. In the first, trehalose-6-phosphate synthase (TPS) catalyses the transfer of glucose from UDP-glucose to glucose 6-phosphate (G6P) to produce trehalose-6-phosphate (T6P). Subsequently, T6P is dephosphorylated into trehalose by trehalose-6-phosphate phosphatase (TPP). The plant TPS proteins contain the TPS and TPP domains, whereas TPP proteins contain only TPP domains. Plant TPP proteins have TPP activities [3]. However, although plant TPS proteins contain TPP domains, many studies have not detected TPP activity [6,7]. The TPP domains in plant TPS proteins appear to have lost enzymatic activity during evolution. The evolutionary basis of this loss is not yet understood.

Each of the *Arabidopsis* and rice genomes has 11 TPS genes, and *Populus* contains 12 TPS genes [6,7,8]. These TPS genes have been divided into class I and class II [6,7,8]. Many studies have shown functional divergence among members of this gene family. Among the 11 rice TPS genes, functional complement assays performed in yeast tps1 and tps2 mutants revealed that only OsTPS1 encodes an active TPS enzyme and that no OsTPS protein possesses TPP activity [7]. Among four *Arabidopsis* class I TPS genes, only *AtTPS1* has TPP enzymatic activity, and no protein has significant TPP activity [6]. The *AtTPS1* gene plays important roles in the control of the stress response, cell and embryonic development, glucose sensing, and starch synthesis [9,10,11]. None of the seven *Arabidopsis* class II TPS genes (*AtTPS3–11*) apparently evolved under relaxed selective constraint. Additionally, a reverse transcription polymerase chain reaction analysis showed the expression divergence of the TPS gene family in *Populus*, *Arabidopsis*, and rice under normal growth conditions and in response to stressors. Our findings provide new insights into the mechanisms of gene family expansion and functional evolution.
to contribute to the expansion of this gene family? (2) Do the TPS and TPP domains of TPS genes undergo similar selection pressure? (3) Which factors drive the functional divergence of class I and class II TPS genes? In order to address these questions, in this study, we examined the evolutionary characterisation of the TPS gene family in Populus, Arabidopsis, and rice. Arabidopsis thaliana is an important model for flowering plants (particularly eudicots), and rice (monocotyledon) is one of the most important food crops in the world. Perennial Populus is the most important model tree system of plant genomics currently available. These three angiosperm plants have completely sequenced and well-annotated genomes and have undergone at least one round of genome-wide duplication. Through a comprehensive analysis of gene sequences, gene structures, molecular evolution, and gene expression patterns, we provide a useful framework for further functional characterisation of the plant TPS gene family.

Results

Phylogenetic Analyses and Gene Structures of the TPS Gene Family in Populus, Arabidopsis, and Rice

Previous studies identified 12 TPS genes in the Populus genome and 11 in each of the Arabidopsis and rice genomes (Table S1); these TPS genes are divided into class I and class II subfamilies [6,7,8]. Populus, Arabidopsis, and rice contain two (PtTPS1 and PtTPS2), four (AtTPS1, 2, 3, and 4), and one (OsTPS1) class I TPS genes, respectively, and the remaining 27 TPS genes are class II TPS genes [6,7,8]. In this study, we constructed a phylogenetic tree of the 34 TPS genes from Populus, Arabidopsis, and rice (Figure 1A). The phylogenetic tree showed that the 34 TPS genes were divided into distinct two clades (clades A and B) with 100% bootstrap support. Clades A and B contained 27 class II and seven class I TPS genes, respectively. The seven class I TPS genes were further divided into two subclades (clades B1 and B2) with high bootstrap support. Clade B1 contained four class I TPS genes (AtTPS1, OsTPS1, PtTPS1, and PtTPS2) from Populus, Arabidopsis, and rice, whereas clade B2 contained only three Arabidopsis class I TPS genes (AtTPS2, 3, and 4).

We identified the nodes that lead to rice-specific and Populus–Arabidopsis–specific subclades (red circles in Figure 1A). These nodes represented the most recent common ancestral genes before the monocot and dicot split. The subclades defined by such nodes were designated as orthologous groups. Interestingly, subclade A1 only contained Populus and Arabidopsis class II TPS genes, indicating that genes may have been lost from the rice genome. Subclade B2 only contained three Arabidopsis class I TPS genes. We predicted that the TPS genes in subclade B2 might exist in the monocot–dicot common ancestor and were subsequently lost in Populus and rice. The number of subclades indicated that there were at least seven ancestral TPS genes in the monocot–dicot common ancestor. To further validate this conclusion, we performed a joint phylogenetic analysis with the TPS genes from Populus, Arabidopsis, rice, and other basal angiosperm species. To date, no complete genome sequence is available for basal angiosperm species. The most comprehensive genomic resources are the Amborella Genome Database [http://amborella.luck.psu.edu/] and Ancestral Angiosperm Genome Project (http://ancangio.uga.edu/). From these two databases, 15 TPS genes from Amborella trichopoda, Persea americana, Liriodendron tulipifera, Aristolochia fimbriata, and Nymphoides advena were identified using TBLASTN searches. A phylogenetic tree with 49 TPS genes from Populus, Arabidopsis, rice, and five basal angiosperm species supported that there were at least seven ancestral TPS genes in the monocot–dicot common ancestor (red circles in Figure S1).

TPS genomic sequences were retrieved from the Populus, Arabidopsis, and rice genomes based on available information to investigate the intron/exon structures of the TPS genes. We found that all class I TPS genes had 16 introns within the protein-coding region by sequence comparisons between cDNAs and genomic DNA, whereas class II TPS genes had two introns (Figure 1B). The contrast in gene structures between class I and class II TPS members suggested their evolutionary divergence.

Structural Features of TPS Proteins

Pairwise comparisons of the 34 TPS full-length protein sequences revealed some notable features (Figure 2). Seven class I TPS proteins showed 46.7–78.0% pairwise sequence identity. Twenty-seven class II TPS proteins had 43.1–94.4% pairwise sequence identity. But the protein sequences between the two TPS classes were significantly different (independent-sample t-test, P<0.0001) and only had 25.3–30.5% pairwise sequence identity.

The crystal structures of the Eschscholzia coli TPS protein (OtsA) have been resolved [15], which facilitated our understanding of the structural features of TPS family proteins. In this study, we modelled the three-dimensional (3D) structures of PtTPS1 and PtTPS5, representing two different classes of TPS proteins, based on the known 3D structure of the OtsA protein. TPS is a UDP-glucose (UDP-Glc)-dependent glycosyltransferase. It catalyses the synthesis of T6P using UDP-Glc as the donor and G6P as the acceptor. The X-ray structure of OtsA showed six key catalytic residues that could interact with the UDP-Glc substrate (Figure 3) [15]. Sequence comparisons showed that five sites among the six key catalytic residues were absolutely conserved in the seven class I TPS proteins, whereas only one site was absolutely conserved in the 27 class II TPS proteins (Figure S2). The predicted 3D structure of PtTPS1 showed that six residue sites could interact with the UDP-Glc substrate (Figure 3). However, only three sites in the PtTPS5 protein could interact with the UDP-Glc substrate (Figure 3). Thus, low pairwise sequence identity and a structural difference between the two classes of TPS proteins suggested their functional divergence.

Chromosomal Locations and Gene Duplications in the TPS Gene Family of Populus, Arabidopsis, and Rice

The physical locations of 10 among 12 TPS genes were assigned to nine Populus chromosomes, whereas the other two were assigned to two unattributed scaffold fragments. All 12 TPS genes were dispersed in Populus chromosomes (Figure S3). A phylogenetic analysis of the TPS gene family from Populus, Arabidopsis, and rice showed that Populus contained five recently duplicated gene pairs (PtTPS3/5, PtTPS4/6, PtTPS7/8, PtTPS9/10, and PtTPS11/12) (Figure 1A). A previous analysis of the Populus genome identified paralogous segments created by the whole-genome duplication event in the Salicaceae (salicoid duplication), which occurred 60–65 million years ago [16]. In this study, we found four duplicate pairs (PtTPS4/6, PtTPS7/8, PtTPS9/10, and PtTPS11/12) that were each located in a pair of paralogous blocks (Figure 4A), which indicated that the four duplicated gene pairs were created by a whole-genome duplication event. Additionally, comparisons of the 60-kb flanking genomic regions of the paralogous gene pair PtTPS3/3 using Mauve software showed that this duplicate gene pair was formed by a segmental duplication event (Figure 4C).

In rice, all 11 TPS genes were dispersed in chromosomes 1, 2, 3, 5, 8, and 9 (Figure S3). The maximum-likelihood (ML) tree in Figure 1A shows that the rice genome contained four rice-specific duplicated gene pairs (OsTPS2/6, OsTPS4/10, OsTPS7/11, and OsTPS8/9). Segmental duplicated blocks have been identified in the rice genome (http://rice.plantbiology.msu.edu/). We found
that three duplicated gene pairs (OsTPS2/6, OsTPS7/11, and OsTPS8/9) were each located in a pair of paralogous blocks (Figure 4B), indicating that these three duplicated gene pairs were formed by a segmental duplication event.

In the Arabidopsis genome, except for AtTPS2 and AtTPS3, which were tandemly arrayed in chromosome 1, the other nine TPS genes were dispersed in chromosomes 1, 2, and 4 (Figure S3). The ML tree in Figure 1A shows that the Arabidopsis genome contained two Arabidopsis-specific duplicated gene pairs (AtTPS2/3 and AtTPS7/8). AtTPS7 and AtTPS8 were created by a recent Arabidopsis whole-genome duplication event [17]. A previous co-linearity analysis suggested that the AtTPS2 and AtTPS3 genes arose from segmental duplication of the AtTPS1 gene region, followed by a tandem duplication giving rise to the AtTPS2 and AtTPS3 gene pair [8].

**Molecular Evolution Analyses**

The ratio (ω) of the synonymous substitution rate (dS) versus the non-synonymous substitution rate (dN) provides a sensitive measure of selective pressure acting on a protein-coding gene. Homologous genes with ω ratios of 1, <1, or >1 are usually assumed to be evolving under neutral evolution, purifying selection, or positive selection, respectively. Plant TPS proteins consist of an N-terminal TPS domain and a C-terminal TPP domain [3]. To test for deviations in the substitution pattern of the two domains, we partitioned the 34 TPS sequences into TPS domain and TPP domain regions. The ω values were calculated across all pairwise comparisons within each of the 34 TPS genes using the YN00 program in the PAML software package [18]. A plot of dS/dN for the TPS versus TPP domains is shown in Figure 5. The results suggested that both domains evolved under purifying selection but that the selective
constraint on the TPP domain was more relaxed than that on the TPS domain (t-tests, $P<0.0001$).

The TPS gene family in Populus, Arabidopsis, and rice was divided into two subfamilies (class I and class II) (Figure 1A). The TPS genes within the same class showed conserved gene structures and high protein sequence similarity, but gene structural variation and low protein sequence similarity were observed between the two classes of TPS genes. To test whether there were changes in selective pressure between the two classes of TPS genes, we conducted two branch-specific models (one-ratio model and two-ratio model) using the CODEML program in the PAML software package. The one-ratio model assumed a single $\omega$ ratio for TPS genes in the tree. This was compared using a likelihood ratio test (LRT) with the two-ratio model, which assumed different $\omega$ ratios in the two TPS genes clades. The test was conducted independently for the full-length gene, TPS, and TPP domain regions on the four unrooted trees (Figure 6). The comparison of the one-ratio model to the two-ratio model using LRTs showed that the two-ratio model was a significantly better fit than the one-ratio model for the full-length gene, TPS, and TPP domains ($P<0.002$) (tree 1 in Table 1), suggesting significant differences in selective pressures between the two classes of TPS genes. The mean $\omega$ value in the class II TPS genes was lower than that in the class I TPS genes (Table 1), indicating that the class II TPS genes (clade A in tree 1, Figure 6) were under stronger purifying selection than the class I TPS genes (clade B in tree 1, Figure 6). Class I TPS genes were further divided into two subclasses: clade B1 and B2 (tree 1 in Figure 6). We further tested whether there were changes in selective pressures between the class II TPS genes...
Figure 5. \( \frac{d_N}{d_S} \) plot for the TPP domain versus the TPS domain of each pair of 34 TPS genes. These TPS genes were from Populus, Arabidopsis, and rice. doi:10.1371/journal.pone.0042438.g005

and class I TPS genes in clade B1 or clade B2. The PAML analysis showed that the class I TPS genes in clade B1 were under stronger purifying selection than the class II TPS genes in clade A for the full-length TPS genes and the TPS domain regions (\( P<0.001 \)) (tree 2 in Table 1). However, the selective pressure for the TPP domain was not significantly different between the two clades. This indicated that the TPS domain contributed to the selective pressure change between the TPS genes in clades A and B1 (\( P<0.001 \)) (tree 2 in Figure 6). Clade B2 only contained three Arabidopsis class I TPS genes (AtTPS2, 3, and 4). The PAML analysis showed that three Arabidopsis class I TPS genes in clade B2 were under more relaxed selection pressure than the class II TPS genes in clade A in the full-length gene, the TPS domain, and the TPP domain regions (\( P<0.001 \)) (tree 3 in Table 1). Additionally, the PAML analysis also showed that the three Arabidopsis class I TPS genes in clade B2 are under more relaxed selection pressure than the four class I TPS genes in clade B1 (\( P<0.001 \)) (tree 4 in Table 1). Taken together, these results reveal that four class I TPS genes (AtTPS1, OsTPS1, PtTPS1, and PtTPS2) were under stronger purifying selection among 34 TPS genes from Populus, Arabidopsis, and rice, whereas three Arabidopsis class I TPS genes (AtTPS2, 3, and 4) apparently evolved under relaxed selective constraint.

Expression Patterns of TPS Gene Family in Populus, Arabidopsis, and Rice

The tissue-specific expression patterns of the TPS gene family in Populus and Arabidopsis were examined by reverse transcription polymerase chain reaction (RT-PCR) analysis under normal growth conditions and in response to stress treatments (\( \text{H}_2\text{O}_2 \), NaCl, salicylic acid, and drought). In Populus, ten TPS genes (PtTPS1, 3, 4, 5, 6, 7, 8, 9, 10, and 12) were expressed in all tissues under all growth conditions, whereas two TPS genes (PtTPS2 and PtTPS11) were neither expressed in any tissue nor responded to any treatment applied in this study (Figure 7A). We did not find the PtTPS2 EST sequence in the Populus trichocarpa expressed sequence tag (EST) database, but five PtTPS17 ESTs (expressed in bud, floral bud, and leaf tissues) were identified (Table S2). Eight Arabidopsis TPS genes (AtTPS1, 5, 6, 7, 8, 9, 10, and 11) were expressed in all the tissues examined, and two TPS genes (AtTPS2 and 3) were not expressed in any tissues examined (Figure 7B). AtTPS4 was only expressed in root and flower bud tissues. In the Arabidopsis thaliana EST database, only a single AtTPS2 EST (expressed in seeds) was identified, but we did not find the AtTPS3 EST sequence (Table S2). In rice, we only analysed tissue-specific expression patterns of the TPS gene family under normal growth conditions. Eight rice TPS genes (OsTPS1, 2, 3, 4, 5, 6, 10, and 11) were expressed in all tissues, whereas OsTPS2 was not expressed in the tissues examined. OsTPS6 and OsTPS7 were selectively expressed in some specific tissues (Figure 7C). An EST search showed that OsTPS9 was expressed in the panicle and pistil (Table S2). In this study, both the RT-PCR and EST search did not identify expression of the PtTPS2 and AtTPS3 genes. These two genes might be expressed at sub-detectable levels, or they are only induced in response to treatments and/or in tissues not examined in our study, or they are pseudogenes.

Discussion

The plant TPS genes play essential roles in the regulation of sugar metabolism, embryonic development, and response to abiotic stress. Plant TPS genes fall into two distinct classes. Many studies have shown distinct functional divergence between the two classes of TPS genes. Our molecular evolution analysis revealed that the TPS domains of four class I TPS genes (AtTPS1, OsTPS1, PtTPS1, and PtTPS2) among 34 TPS genes from Populus, Arabidopsis, and rice, were under stronger purifying selection, indicating functional conservation of the TPS domain among the four TPS genes. Sequence comparisons and predicted 3D structures also indicated that class I TPS proteins had TPS activity. Functional complement assays performed in yeast tps1 and tps2 mutants revealed that the OsTPS1 and AtTPS1 had TPS activities but no TPP activity [6,7]. Thus, the results of the molecular evolution analysis were well supported by these experimental results. Sequence and structural analyses showed that class II TPS proteins might not interact with the UDP-Glc substrate, suggesting that class II TPS proteins are not directly involved in trehalose metabolism. However, many class II TPS genes were expressed in different tissues in Populus, Arabidopsis, and rice, suggesting that class II TPS genes might have new functions. For example, Arabidopsis AtTPS5 plays a role in thermotolerance, possibly through its interaction with the transcriptional co-activator MFB1c [19]. Interestingly, three Arabidopsis class I TPS genes (AtTPS2, 3, and 4) apparently evolved under relaxed selective constraint, indicating functional divergence. Functional complement assays performed in yeast tps1 and tps2 mutants revealed that all AtTPS2, AtTPS3, and AtTPS4 had no TPS or TPP activity [6]. The possible molecular causes for the loss of AtTPS2 and AtTPS4 activity were considered to accumulate deleterious mutations and alter protein conformations [6]. Additionally, the expression patterns of AtTPS2, AtTPS3, and AtTPS4 had significant differences compared with AtTPS1. AtTPS1 was expressed in roots, stems, rosette leaves, flower buds, and ripening silique [20]. However, the expressions of AtTPS2 and AtTPS3 were not detected by RT-PCR in Arabidopsis. AtTPS4 was only expressed in root and flower bud tissues. Other studies found that AtTPS2 is expressed in seeds [21]. AtTPS2-specific GUS expression was detected only in the chalazal endosperm [6]. Thus, the molecular evolution analysis, gene expression patterns, and functional complement assays indicated functional divergence between the three Arabidopsis class I TPS genes and AtTPS1.

Gene duplication and subsequent functional divergence of the duplicate genes has been recognised as an important source of evolutionary novelty. Several possible evolutionary fates of duplicate genes have been proposed [22,23,24]: (1) non-
functionalisation, in which one duplicate gene accumulates deleterious mutations as a pseudogene, whereas another duplicate gene maintains its original function; (2) neo-functionalisation, in which one duplicate copy accumulates beneficial mutations and acquires a new function, whereas another duplicate copy retains the original function; (3) sub-functionalisation, in which one duplicate copy accumulates beneficial mutations and acquires a new function, whereas another duplicate copy retains the original function.
segmental duplication events, and one duplicate pair, AtTPS2/3, was formed by tandem duplication. The gene expression-pattern analyses showed that two duplicate copies shared similar expression patterns in five duplicate gene pairs (PtTPS3/5, PtTPS7/8, PtTPS9/10, and AtTPS8/9), suggesting that two duplicate genes had redundant functions or that functional divergence not identified in our study had occurred, such as protein function. For the duplicate pair AtTPS2/3, AtTPS2 was expressed in seeds, whereas AtTPS3 was not expressed in all tissues examined, suggesting that AtTPS3 may have become a pseudogene or evolved a new function not identified in our study. One copy of each of four duplicate gene

Table 1. Summary statistics for detection of selection using branch specific models of PAML.

| Tree  | Model       | Estimates of parameters | In L   | 2 \( .dl \) | P     |
|-------|-------------|-------------------------|--------|-------------|-------|
| Tree 1| Full length gene | One ratio \( \omega = 0.09382 \) for all branches | -47300.58064 |             |       |
|       |             | Two ratios \( \omega_A = 0.08476 \) for clade A, \( \omega_B = 0.13127 \) for clade B | -47274.88473 | 51.39182 | <0.0001 |
|       | TPS domain  | One ratio \( \omega = 0.07925 \) for all branches | -28400.95963 |             |       |
|       |             | Two ratios \( \omega_A = 0.07222 \) for clade A, \( \omega_B = 0.10836 \) for clade B | -28388.21796 | 25.48336 | <0.0001 |
|       | TPP domain  | One ratio \( \omega = 0.10915 \) for all branches | -14778.40868 |             |       |
|       |             | Two ratios \( \omega_A = 0.10133 \) for clade A, \( \omega_B = 0.14354 \) for clade B | -14773.41535 | 9.98666 | 0.0017 |
| Tree 2| Full length gene | One ratio \( \omega = 0.08217 \) for all branches | -45848.25709 |             |       |
|       |             | Two ratios \( \omega_A = 0.08778 \) for clade A, \( \omega_B = 0.05162 \) for clade B1 | -45829.27108 | 37.97203 | <0.0001 |
|       | TPS domain  | One ratio \( \omega = 0.06743 \) for all branches | -27558.71280 |             |       |
|       |             | Two ratios \( \omega_A = 0.07500 \) for clade A, \( \omega_B = 0.03016 \) for clade B1 | -27529.46938 | 58.48685 | <0.0001 |
|       | TPP domain  | One ratio \( \omega = 0.09627 \) for all branches | -13788.87778 |             |       |
|       |             | Two ratios \( \omega_A = 0.09945 \) for clade A, \( \omega_B = 0.07508 \) for clade B1 | -13787.24660 | 3.26236 |       |
| Tree 3| Full length gene | One ratio \( \omega = 0.09565 \) for all branches | -42556.07775 |             |       |
|       |             | Two ratios \( \omega_A = 0.08393 \) for clade A, \( \omega_B = 0.35356 \) for clade B2 | -42442.09292 | 227.96967 | <0.0001 |
|       | TPS domain  | One ratio \( \omega = 0.08229 \) for all branches | -25567.87248 |             |       |
|       |             | Two ratios \( \omega_A = 0.07111 \) for clade A, \( \omega_B = 0.33343 \) for clade B2 | -25487.99989 | 159.74518 | <0.0001 |
|       | TPP domain  | One ratio \( \omega = 0.10881 \) for all branches | -13198.32846 |             |       |
|       |             | Two ratios \( \omega_A = 0.09997 \) for clade A, \( \omega_B = 0.30455 \) for clade B2 | -13179.70794 | 37.24105 | <0.0001 |
| Tree 4| Full length gene | One ratio \( \omega = 0.12320 \) for all branches | -10909.79129 |             |       |
|       |             | Two ratios \( \omega_A = 0.06510 \) for clade B1, \( \omega_B = 0.35030 \) for clade B2 | -10798.89073 | 221.80112 | <0.0001 |
|       | TPS domain  | One ratio \( \omega = 0.09574 \) for all branches | -6148.55613 |             |       |
|       |             | Two ratios \( \omega_A = 0.04285 \) for clade B1, \( \omega_B = 0.31121 \) for clade B2 | -6063.71598 | 169.68030 | <0.0001 |
|       | TPP domain  | One ratio \( \omega = 0.13771 \) for all branches | -3405.43807 |             |       |
|       |             | Two ratios \( \omega_A = 0.08930 \) for clade B1, \( \omega_B = 0.31125 \) for clade B2 | -3387.03778 | 36.80058 | <0.0001 |

NOTE: All trees were shown in figure 6.
doi:10.1371/journal.pone.0042438.t001

molecular evolution, in which each descendant copy adopts some of the tasks of the ancestral gene. In this study, nine TPS duplicate pairs (AtTPS8/9, PtTPS3/5, 4/6, 7/8, 9/10, and 11/12, OsTPS2/6, 7/11, and 8/9) were created by segmental duplication events, and one duplicate pair, AtTPS2/3, was formed by tandem duplication. The gene expression-pattern analyses showed that two duplicate copies shared similar expression patterns in five duplicate gene pairs (PtTPS3/5, PtTPS4/6, PtTPS7/8, PtTPS9/10, and AtTPS8/9), suggesting that two duplicate genes had redundant functions or that functional divergence not identified in our study had occurred, such as protein function. For the duplicate pair AtTPS2/3, AtTPS2 was expressed in seeds, whereas AtTPS3 was not expressed in all tissues examined, suggesting that AtTPS3 may have become a pseudogene or evolved a new function not identified in our study. One copy of each of four duplicate gene
pairs (PtTPS11/12, OsTPS2/6, OsTPS7/11, and OsTPS8/9) was expressed in all tissues examined, whereas the other was expressed only in a specific tissue. This sub-functionalisation may be the evolutionary fate of the four duplicate gene pairs.

Methods

Phylogenetic Analyses

A previous study identified 12 TPS genes in the Populus genome (Build version 1.2) [8]. Eleven TPS genes were identified in each of the Arabidopsis and rice genomes [7,25]. These TPS protein sequences were aligned using MUSCLE software [26] and further adjusted manually using BioEdit [27]. The phylogenetic relationships of the Populus, Arabidopsis, and rice TPS gene family were reconstructed based on full-length protein sequences using a ML procedure implemented in PHYML [28] with the JTT (Jones, Taylor & Thornton) amino acid substitution model. The Escherichia coli TPS protein (OtsA) was used as an out group in the phylogenetic analysis. Sequence comparisons between cDNA and genomic DNA were performed using BioEdit software and further adjusted manually using BioEdit [27].

Homology Modelling

The crystal structures of the Escherichia coli TPS protein (Protein Data Bank code no.: 1GZ5) was used as a template for constructing structural models of PtTPS1 and PtTPS5. The sequences were aligned using the Align 2D structure alignment program (Homology Module in InSightII software; Accelerx, San Diego, CA, USA). Structures were automatically built using the modeller module of InSightII. All structures were verified by the profile-3D program in InSightII. The models were selected according to the model evaluation score calculated by profile-3D.

Molecular Evolution Analyses

To evaluate variation in selective pressure in a phylogeny, the branch-specific models of CODEML in the software package PAML [18] were used to estimate $\theta$ under different assumptions. Analyses were conducted under two $a$ priori assumptions: a one-ratio model in which one $\theta$ value was assumed for the entire tree, and a two-ratio model in which $\theta$ values were allowed to vary between the two different clades. To verify which of the models best fit the data, LRTs were performed by comparing twice the difference in log-likelihood values between pairs of the models using a $\chi^2$ distribution, with the degrees of freedom equal to the differences in the number of parameters between the models [18].

Expression of TPS Genes in Populus, Arabidopsis, and Rice

Seedlings of P. trichocarpa (Torr. & Gray) were cultivated in potting soil for 2 months. Populus seedlings were separately irrigated and sprayed with 150 mM NaCl, 0.5% H$_2$O$_2$, or 3.5 mM salicylic acid solution. Each treatment was conducted with five replicate plants for 12 hr. Drought stress was evaluated by withholding water for 2 weeks. After the stress treatments, total RNA was isolated from leaf, shoot, bud, phloem, and root tissues of each seedling. The 4-week-old Arabidopsis plants (Arabidopsis thaliana ecotype Columbia-0) were transferred into a 250 mM NaCl solution for a 5 hr salt treatment, 1 mM H$_2$O$_2$ solution for 5 hr for the oxidative stress treatment, and Whatman filter paper for 2 hr for the drought-stress treatment, respectively. Arabidopsis plants were irrigated with 5 mM salicylic acid solution for 24 hours for the salicylic acid treatment. After the stress treatments, total RNA was isolated from root, stem, rosette leaf, cauline leaf, and flower bud tissues of each Arabidopsis plant. The seedlings of Oryza sativa L. ssp. japonica were grown individually in 10x10 cm pots for 3 months. Total RNA was isolated from mature leaf, immature leaf, leaf sheath, stem, and root tissues of each rice seedling. Total RNA was isolated using an Aurum Total RNA Kit (Bio-Rad Laboratories, Hercules, CA, USA). Then, total RNA was treated with RNase-free DNase I (Promega, Madison, WI, USA) and reverse transcribed into cDNA using an RNA PCR Kit (AMV) version 3.0 (TaKaRa Bio, Shiga, Japan). Thirty-four specific primer pairs were designed based on the multiple sequence alignment of the TPS sequences (Table S3). In all RT-PCR analyses, the actin gene was used as an internal control. PCR conditions were optimised to consist of an initial denaturation of 3 min at 95°C, followed by 35 cycles of 30 s at 94°C, 30 s at 60°C, and 30 s at 72°C, with a final extension of 5 min at 72°C. After RT-PCR, the PCR products from each sample were analysed on 1% agarose gels and validated by DNA sequencing. All RT-PCR results were obtained from three independent experiments. Expression patterns of the TPS genes in this study were also inferred from a BLAST analysis of the EST database at the National Centre for Biotechnology Information (NCBI) EST database. To date, the NCBI EST database has 1,520,700 Arabidopsis thaliana ESTs, 987,318 Oryza sativa japonica ESTs, and 89,943 Populus trichocarpa ESTs. A minimum cut-off E value ($\Delta e \leq -20$) was applied to select significant matches. A threshold of at least 97% sequence identity was employed.

Supporting Information

Figure S1  Phylogenetic tree of TPS genes from Populus, Arabidopsis, rice, and five basal angiosperm species. Red circles indicate the most recent common ancestral TPS genes among Populus, Arabidopsis, and rice. (TIF)

Figure S2  Sequence alignment of TPS proteins from Populus, Arabidopsis, and rice. Conserved residues in TPS proteins are shaded in black and gray. The key catalytic residues interacting with UDP-glucose in the crystal structure of E. coli OtsA are indicated by red arrows. Conserved key catalytic residues interacting with UDP-glucose are shaded in green. (TIF)

Figure S3  Genomic localisation of the TPS genes in Populus (A), Arabidopsis (B), and rice (C). (TIF)

Table S1  The TPS genes used to reconstruct phylogenetic trees. (DOC)

Table S2  The numbers of TPS ESTs identified from rice, Arabidopsis and Populus EST databases in NCBI. (DOC)

Table S3  Primers used to detect the expression of TPS genes. (DOC)
Acknowledgments

We thank the three anonymous reviewers for valuable comments and suggestions that helped to improve the manuscript. We thank Zhi-Ling Yang for protein structure homology modeling.

References

1. Garg AK, Kim JK, Owens TG, Ramvola AP, Choi YD, et al. (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proc Natl Acad Sci USA 99: 15988–15993.
2. Elbein AD, Pan YT, Pastuszak I, Carroll D (2003) New insights on trehalose: a multifunctional molecule. Glycobiology 13: 171–178.
3. Goddijn OJM, van Dun K (1999) Trehalose metabolism in plants. Trends Plant Sci 4: 315–319.
4. Jang JC, Oh SJ, Seo JS, Choi WB, Song SI, et al. (2003) Expression of a bifunctional fusion of the Escherichia coli genes for trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase in transgenic rice plants increases trehalose accumulation and abiotic stress tolerance without stunting growth. Plant Physiol 131: 516–524.
5. Shima S, Matsui H, Tahara S, Imai R (2007) Biochemical characterization of rice trehalose-6-phosphate phosphatases supports distinctive functions of these plant enzymes. FEBS J 274: 1192–1201.
6. Vandesteene L, Ramon M, Le Roy K, Van Dijck P, Rolland F (2010) A single active trehalose-6-P synthase (TPS) and a family of putative regulatory TPS-like proteins in Arabidopsis. Mol Plant 3: 406–419.
7. Zang B, Li H, Li W, Deng XW, Wang X (2011) Analysis of trehalose-6-phosphate synthase (TPS) gene family suggests the formation of TPS complexes in rice. Plant Mol Biol 76: 507–522.
8. Lunn JE (2007) Gene families and evolution of trehalose metabolism in plants. Funct Plant Biol 34: 540–563.
9. Gomez LD, Gilday A, Feil R, Lunn JE, Graham IA (2010) ArTPS-mediated trehalose 6-phosphate synthesis is essential for embryonic and vegetative growth and responsiveness to ABA in germinating seeds and stomatal guard cells. Plant J 64: 1–13.
10. Gomez LD, Basu S, Gilday A, Li Y, Graham IA (2006) Delayed embryo development in the Arabidopsis trehalose-6-phosphate synthase 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. Plant J 46: 68–84.
11. van Dijken AJ, Schuerppmann H, Sneekens SC (2004) Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. Plant Physiol 135: 969–977.
12. Ramon M, De Smet I, Vandesteene L, Naudts M, Leyman B, et al. (2009) Extensive expression regulation and lack of heterologous enzymatic activity of the Class II trehalose metabolism proteins from Arabidopsis thaliana. Plant Cell Environ 32: 1013–1132.
13. Chary SN, Hicks GR, Choi YG, Carter D, Raikhel NV (2008) Trehalose-6-phosphate synthase/phosphatase regulates cell shape and plant architecture in Arabidopsis. Plant Physiol 146: 97–107.
14. Glinski M, Weckwerth W (2005) Differential multivariate phosphorylation of the trehalose-6-phosphate synthase gene family in Arabidopsis thaliana: A mass spectrometry-based process for multiparallel peptide library phosphorylation analysis. Mol Cell Proteomics 4: 1614–1623.
15. Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R, Davies GJ (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. Chem Biol 9: 1337–1346.
16. Tukan AG, Difazio S, Janssen S, Bohllmann J, Grigoriev I, et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 315: 1596–1604.
17. Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res 13: 174–184.
18. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586–1591.
19. Suzuki N, Bajad S, Shuman J, Shulaev V, Mittler R (2008) The transcriptional co-activator MBF1c is a key regulator of thermotolerance in Arabidopsis thaliana. J Biol Chem 283: 9269–9275.
20. van Dijken AJ, Schuerppmann H, Sneekens SC (2004) Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. Plant Physiol 135: 969–977.
21. White JA, Todd J, Newman T, Focks N, Girke T, et al. (2000) A new set of genes for trehalose-6-phosphate synthase/phosphatase regulates cell shape and plant architecture in Arabidopsis thaliana. Genome Res 10: 1531–1545.
22. Ohno S (1970) Evolution by gene duplication. Heidelberg: Springer-Verlag.
23. Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc Biol Sci 256: 119–124.
24. Force A, Lynch M, Pickett FB, Amores A, Yan YL, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–1545.
25. Leyman B, Van Dijck P, Thevelein JM (2001) An unexpected plethora of trehalose biosynthesis genes in Arabidopsis thaliana. Trends Plant Sci 6: 510–513.
26. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.
27. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98.
28. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
29. Darling AE, Mau B, Perna NT (2010) ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5: e11147.

Author Contributions

Conceived and designed the experiments: HLY QYZ. Performed the experiments: YJL CLW. Analyzed the data: HLY QYZ. Wrote the paper: HLY QYZ.