Plant Omics Journal POJ

POJ 6(3):215-223 (2013) ISSN:1836-3644

Comprehensive and evolutionary analysis of protein tyrosine phosphatases (PTP) in the green plants

Lujun Yu†, Dongru Feng†, Wenyan Li, Hongbin Wang, Jinfa Wang and Bing Liu*

Guangdong Key Laboratory of Plant Resources and State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, 510275, Guangzhou, P.R. China

***Corresponding author: liubing@mail.sysu.edu.cn.**
†These authors contributed equally to this work.

Abstract

Protein tyrosine phosphatases (PTP) exhibit specific functions in regulation of organism growth and development, cooperating with protein tyrosine kinase (PTK) to balance tyrosine phosphorylation. In the current study, we investigated a comprehensive analysis of PTP complements (PTPome) in five representative major plant lineages, including green algae, bryophytes, pteridophytes, monocots and eudicots, and classified them into ten subfamilies according to the catalysis manner, sequence phylogenetic and domain organization analysis. Furthermore, we proposed a PTP subfamily stepwise shift model, which was consistent with plant species complexity. An original PTP kit containing seven subfamilies was detected in *Chlamydomomas reinhardtii*, with one acquisition and one reduction step followed by species radiation. Lipid phosphatases were only detected in land plants, but not aquatic plants. Additionally, comparisons of green plant PTPome with human PTPome revealed significant differences in gene number, subfamilies and domain combinations. Comprehensive analysis provides insight into PTP evolution in plants and the basis for functional genomics.

Keywords: *Chlamydomonas reinhardtii*; gene family; HMMER; lipid phosphatases; PTP loop; PTP kit.
Abbreviations: AAA- ATPases associated with a variety of cellular activities; aDSP- atypical dual specificity phosphatase; AFK- Actin-Fragmin Kinase; *At- Arabidopsis thaliana*; CaMB- calmodulin binding; CDC25- cell division cycle phosphatase 25; Cdk- cyclin-dependent kinase; *Cr-* Chlamydomonas reinhardtii; DSP- dual specificity phosphatase; E_set- E or "early" set of sugar utilizing enzymes (including AMP-activated protein kinase beta subunit glycogen binding domain); ED- Eya domain; EYA- Eyes absent phosphatase; FHA- Forkhead associated domain; GH- Gelsolin homology domain; GRAM- glucosyltransferases, Rab-like GTPase activators and myotubularins motif; GP- Green plant; GPPTP- Green plant protein tyrosine phosphatase; HMM- hidden Markov model; HTH_ARSR- helix_turn_helix, Arsenical Resistance Operon Repressor; LMW-PTP- low molecular weight protein tyrosine phosphatase; LP- Lipid phosphatases; MKP- Mitogen-activated protein kinase phosphatase; mRNA capping- mRNA capping enzyme; mRNA_cap_C-, mRNA capping enzyme, C-terminal; MTM- myotubular myopathy; MTMR- myotubular myopathy related; *Ol- Ostreococcus lucimarinus*; *Os- Oryza sativa*; *Ot- Ostreococcus tauri*; PDZ- Postsynaptic Density-95/Discs large/ZO1; PI3P- phosphatidylinositol 3-phosphate; PI35P2- phosphatidylinositol 3,5-bisphosphate phosphatases; PIP3- phosphatidylinositol 3,4,5-trisphosphate; PK- protein kinase; *Pp- Physcomitrella patens*; PP- protein phosphatase; PPDK- Pyruvate phosphate dikinase; PRL- phosphatase of regenerating liver; *Pt- Populus trichocarpa*; PTEN-C2- C2 domain of PTEN tumour-suppressor protein; PTP- protein tyrosine phosphatase; PTPL- protein tyrosine phosphatase-like; PTEN- phosphatase and tensin homolog deleted on chromosome 10; *Sm- Selaginella moellendorfii*; STP- Serine/Threonine phosphatase; *Vc- Volvox carteri*; *Zm- Zea mays*; ZnF_C_2H_2- Zinc finger C_2H_2-type.

Introduction

Reversible protein phosphorylation is a ubiquitous chemical modification mechanism, which regulates organism growth and development (Luan, 2003; Ghelis et al., 2008). Protein phosphatases (PPs) dephosphorylate a specific substrate, and counteract the effects of protein kinases (PKs), equilibrating protein phosphorylation, which have been thoroughly studied, revealing the evolutionary conservation and functional diversity of the enzymes in many eukaryotes (Manning et al., 2002; Ghelis 2011). In addition, PPs can be divided into serine/threonine phosphatases (STP) and protein tyrosine phosphatases (PTP) based on substrate specificity (Luan 2003). Compared with most STPs, PTPs serve more specific roles in protein phosphorylation regulation (Tonks and Neel 2001; Luan, 2003).

PTP is a large and complex superfamily, which is traditionally grouped into Cys-based PTP and Asp-based PTP, according to the specific catalytic signatures (Alonso et al., 2004; Kim and Ryu 2012). Cys-based PTPs are divided into class I, class II (low molecular weight phosphatases, LMW-PTP) and class III (CDC25) due to their independent evolutionary origins (Alonso et al., 2004). Class I PTP is the largest PTP family, which is comprised of a classic PTP; dual-specific protein phosphatases (DSP) and lipid phosphatases (LPs) in kinetoplasts (Brenchley et al., 2007); and classic PTPs and DSPs identical to humans (Alonso et al., 2004). There is also a kind of PTP-like (PTPL) set of Class IV genes, similar to Class I, which conserved cysteine residue is replaced, resulting in their primary structures different from other PTPs (Bellec et al., 2002).

Relative to $C\times_5R$ in Cys-based PTPs, which use “C” as the nucleophile and “R” for phosphate binding (Tabernero et al., 2008), the Asp-based PTPs share a D×D×T/V motif, which

use first "D" as the nucleophile and second "D" for metal ion binding (Meinhart et al., 2005). The Asp-based PTPs are mainly made up of Eya, in which the inherent tyrosine phosphatase activity has been experimentally demonstrated (Rayapureddi et al., 2005); and RNA polymerase II C-terminal domain phosphatases (Kerk et al., 2008), which require more accurate sequence based and experimental characterization (Alonso et al., 2004).

Recently, PTP functional roles in cellular processes have been elucidated in several model species, facilitating a more global approach to PTP study, including *Homo sapiens* (*Hs*) (Alonso et al., 2004; Kim and Ryu, 2012), sea urchins (Byrum et al., 2006), the human malaria parasite *Plasmodium* (Wilkes and Doerig 2008), and a single lineage of three kinetoplastid species (Brenchley et al., 2007). These reports have addressed animal and protozoan PTP complements (PTPome), but only two plant PTPome has been well examined in *Arabidopsis thaliana* (Kerk et al., 2002) and *Oryza sativa* (Singh et al., 2010). To date, the genomes of many different plant lineages have been sequenced, making it possible to analyze PTPome in green plants. Although PTPs in four green plant species have been previously resolved (Kerk et al., 2008), but classification schemes were still not complete. Since conserved catalytic and accessory domains were used to characterize and delimit different subfamilies in PTPs (Brenchley et al., 2007; Andreeva and Kutuzov 2008; Andersen et al., 2001; Alonso et al., 2004); here, we used comparative genomic approaches to evaluate the Cys-based PTPome evolution from five plant lineages. First, we identified all candidate PTP homologs from ten species, representing five green plant lineages. Second, we evaluated PTP catalysis signatures, catalytic domain conservation and domain organization for GPPTP classification to assess green plant PTPome (GPPTPome) evolution form lower to higher green plants. Finally, we performed a comprehensive comparison between GPPTPome and *Homo sapiens* PTPome (HsPTPome) in terms of gene number, subfamilies and domain combinations.

Results and discussion

GPPTPome identification and classification

There are totally 281 PTP genes detected by the HMMER (Soding, 2005) in five representative major plant lineages as follows: (Tables 1 and S1): *Ostreococcus tauri* (*Ot*), *Ostreococcus lucimarinus* (*Ol*), *Chlamydomonas reinhardtii* (*Cr*), and *Volvox carteri* (*Vc*) in green algae; *Physcomitrella patens* (*Pp*) in bryophytes; *Selaginella moellendorfii* (*Sm*) in pteridophytes; *Arabidopsis thaliana* (*At*) and *Populus trichocarpa* (*Pt*) in eudicots; and *Oryza sativa* (*Os*) and *Zea mays* (*Zm*) in monocots.

All plant PTPs were classified into four subfamilies, based on PTP loop sequence (Andersen et al., 2001), Weblogo (Crooks et al., 2004) analysis and domain organizations combined with mode of catalysis, including Subs. I (PF00102, termed PTP; PF00782 or PF03162, termed DSP), II (PF01451), III (PF00581) and IV (PF04387) (Fig. 1A). The DSP genes were further divided into six subfamilies by phylogeny reconstruction, as follows: Subs. I-B (phosphatase and tensin homolog deleted on chromosome 10, PTEN), I-C (myotubular myopathy related, MTMR), I-D (plant and fungi atypical-DSPs, PFA-DSP), I-E (mRNA), I-F (atypical DSP, aDSP) and I-G (CDC14-like) (Fig. 2). Domain organization analysis showed 70 GPPTP genes (25%) exhibiting multi-domain organization, with diverse accessory domains (ADs), suggesting diverse multi-functions of plant PTP genes (Fig. 1 and Table S2).

Detail analysis of PTP subfamilies

Sub. I-A (classic PTP)

In humans, 38 Sub. I-A PTP genes (Table S3) represent about one-third of the total PTPs, and are comprised of 17 intracellular non-receptor PTPs (NRPTPs) and 21 transmembrane receptor-like enzymes (RPTPs), which are responsible for separate intra- and intercellular responses (Brenchley et al., 2007). Sixteen Sub. I-A PTP genes were found in the GPs except *Ot* and *Ol* (Fig. 1C), with single PTP domain NRPTPs (Fig. 1B type 1), similar with TriTryp PTPome (Brenchley et al., 2007). Considering that same proportion of phosphotyrosines between *At* and *Hs* (Sugiyama et al., 2008), there was only 1 to 5 classic PTP in GP, indicating that the *Hs* Sub. I-A PTP gene function was replaced by DSP in GPs. Phylogenetic analysis showed that all GP genes grouped together and distinct from the *Hs* clade (Fig. S1). The same lineage separately clustered together, which was consistent with plant phylogenetic history.

Lipid Phosphatases (LP)

Sub. I-B (PTEN)

In humans, five PTEN genes (Table S3) use phosphoinositide as the substrate and play important roles in apoptosis during tumor suppression (Maehama et al., 2001). Seventeen PTEN were identified in the land plants (Fig. 1C), with DSP domain and PTEN-C2 motif organization (Fig. 1B type 2). Phylogenetic analysis showed that DSP and PTEN-C2 domain with the similar tree topology (Figs. 3A and S2), which suggested co-evolution between the DSP and PTEN-C2 domains. Plant PTEN had the characteristic PTP loop signature (Fig. 1A), suggesting that they could dephosphorylate PIP3 and participate in the phosphatidylinositol 3-kinase (PI3K) signaling pathway (Gupta et al., 2002).

Sub. I-C (MTMR)

In humans, MTMR genes regulate phosphoinositide phosphatidylinositol 3-phosphate (PI3P) and phosphatidylinositol 3, 5-bisphosphate phosphatases (PI35P2) in development (Maehama et al., 2001; Kerk and Moorhead 2010). Less than 16 genes in humans (Table S3), only eight MTMR genes were identified in land plants (Fig. 1C), with the GRAM, MTMR and DSP domains organization (Fig. 1B types 3-4). Phylogenetic analysis showed that DSP and MTMR domain with the similar tree topology (Figs. 3B and S3), which suggested co-evolution between the DSP and MTMR domains, the similar as PTEN.

Sub. I-D (PFA-DSP)

PFA-DSP, was not identified in human, but detected in *Saccharomyces cerevisiae* (*Siw14*) (Table S3), serving in actin cytoskeleton organization and endocytosis (Roma-Mateo et al., 2011). Twenty-five PFA-DSP genes were detected only in the land plants (Fig. 1C), with only DSP domain (Fig. 1B type 5). Phylogenetic analysis showed differentiated plant genes clustered into two clades; one was comprised of three *Pp* genes, and the other remained 22 land genes and *Siw14* (Fig. 3C). Our lab identified that PFA-DSP regulated the drought and pathogen response in rice (He et al., 2012; Liu et al., 2012a).

Table 1. PTP gene identification in the selected green plant species.

Lineage	Organism	Genome Size * (Mb)	No. Gene	No. PTP	Nomenclature
Green Algae	*Ostreococcus tauri*	12.56	7725	10	*OtPTP*
	Ostreococcus lucimarinus	13.2	7651	13	*OlPTP*
	Chlamydomonas reinhardtii	120	15256	28	*CrPTP*
	Volvox carteri	138	15544	24	*VcPTP*
Bryophytes	*Physcomitrella patens*	480	35938	32	*PpPTP*
Pteridophytes	*Selaginella moellendorfii*	212.5	22285	39	*SmPTP*
Eudicots	*Arabidopsis thaliana*	140	27500	28	*AtPTP*
	Populus trichocarpa	480	45555	38	*PtPTP*
Monocots	*Oryza sativa*	420	41000	37	*OsPTP*
	Zea mays	2500	50000	32	*ZmPTP*
Total				281	

*, The genome information of each species was attained from GenBank release version 193.0.

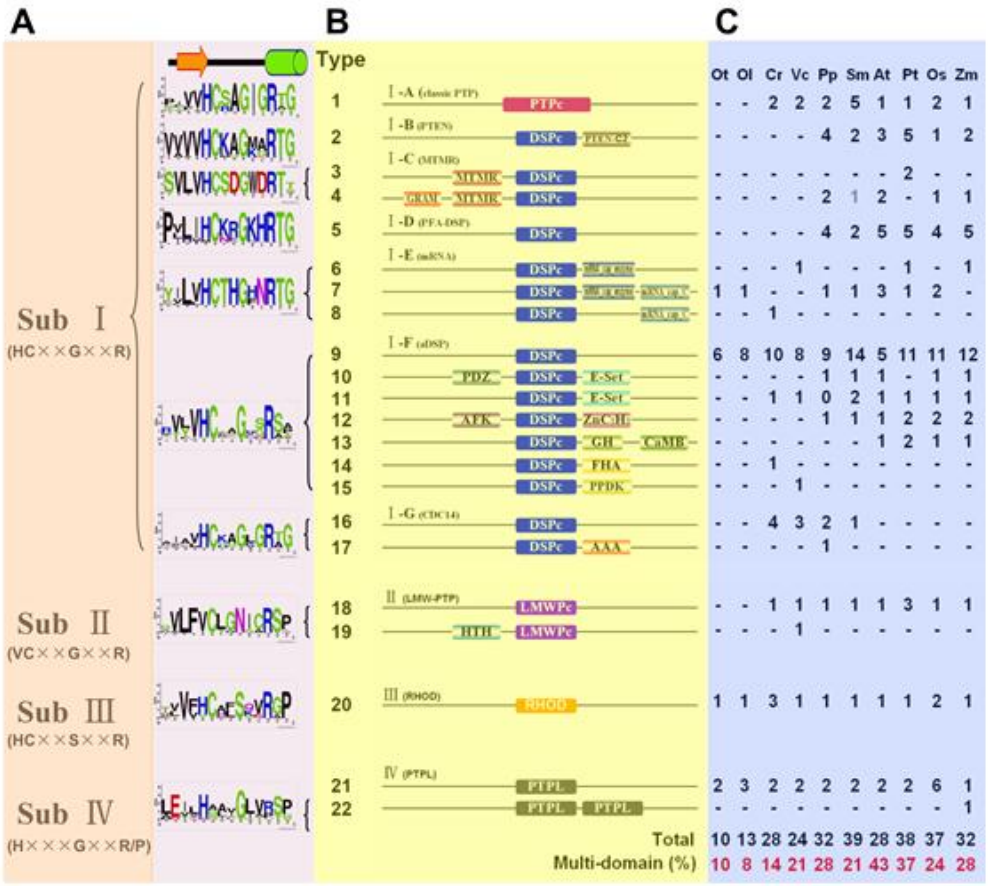

Fig 1. PTPome overview in green plants.
(a) PTP loop sequence characterization of each PTP subfamily. Weblogo analysis of the PTP loop sequences of GPPTPs. The overall stack height indicates the sequence conservation at that position, while the symbol height within the stack indicates the relative frequency of each amino acid at that position. PTP loop of Subs. I, II, III and IV PTP has the "HC××G××R", "VC××G××R", "HC××S××R" and "H×××G×R/P" signature sequences, respectively. (b) GPPTP domain organizations. Domains are colored according to the domain and classed within the ten PTP subfamilies. PTP domains: PTPc, protein-tyrosine phosphatase (PF00102); DSPc, dual specificity phosphatase (PF00782 and PF03162); LMWPc, low molecular weight phosphotyrosine protein phosphatase (PF01451); RHOD, Rhodanese-like domain (PF00581); PTPL, protein tyrosine phosphatase-like protein (PF04387). Accessory domains: PTEN-C2 tensin-type, protein kinase C conserved region 2 tensin-type (PF10409); MTMR, myotubular myopathy related (PF06602); GRAM, glucosyltransferases, rab-like GTPase activators and myotubularins motif (PF02893); mRNA capping, mRNA capping enzyme (PF01331); mRNA_cap_C, mRNA capping enzyme, C-terminal (PF03919); PDZ, postsynaptic density-95/discs large/ZO1 (PF00595); E_set, E or "early" set of sugar utilizing enzymes (including AMP-activated protein kinase beta subunit glycogen binding domain) (PF00686); AFK, actin-fragmin kinase (PF09192); ZnF_C2H2, zinc finger C2H2-type (PF00096); GH, gelsolin homology domain (PF00626); CaMB, CaM-binding ((Yamakawa et al. 2004)); FHA, forkhead associated domain (PF00498); PPDK, pyruvate phosphate dikinase (PF01326); AAA, ATPases associated with a variety of cellular activities (PF00004); HTH_ARSR, helix_turn_helix, arsenical resistance operon repressor(PF01022). (c) GPPTPs distribution in the type of domain organization. "-" indicated absence of this type of domain organization.

Three LP subfamilies were found only in land but not aquatic plants, suggesting that LPs might be integral in the transition from aquatic to terrestrial environments. Three LP subfamilies demonstrated different phosphoinositide preferences, indicated by variation in PTP loop characteristics, which might be due to charge-charge attractions present in PTEN and PFA-DSP, and charge-charge repulsion in the MTMR (Kim and Ryu, 2012) (Fig. 1A).

Sub. I-E (mRNA)

Human *RNGTT* (Table S3) is an mRNA capping enzyme, participates the mRNA capping process by transferring a GMP cap onto the end of mRNA. *Cel-1* (Table S3) shows both RNA triphosphatase (RTPase) and guanylyltransferase (GTase) activity (Takagi et al., 2003). Fourteen genes were identified in GPs sampled (Fig. 1C), with three types domain organization (Fig. 1B types 6-8). Phylogenetic analysis showed that DSP domain and mRNA capping domain with the similar topology (Figs. S4 and S5): genes from the same lineage clustered together, which suggested co-evolution between DSP and mRNA capping domains in the plants.

Sub. I-F (aDSP)

In humans, 19 Sub. I-F genes (Table S3) are smaller and poorly characterized enzymes (Alonso et al., 2004). Totally, 140 aDSP genes were identified in the plants, with diverse domain organization (Fig. 1B types 9-15) and clustered six main groups by the DSP domain phylogenetic analysis (Fig. 2). I-F-1 was comprised of four aquatic PTP genes, grouped closer to the I-E (Fig. 2), with DSP domain (Fig. 1B type 9 and 15). I-F-2 was comprised of 29 genes identified in all GPs (Fig. 2) with three domain organizations (Fig. 1B types 9-11). The E-set domain in the type 10 and 11, responsible for starch or glycogen metabolism (Niittyla et al., 2006), exhibited a reverse domain array, compared with its human *Laforin* (Table S3). I-F-3, I-F-4 and I-F-6 were comprised of 12, 17 and 6 genes in plants (Fig. 2), with only DSP domain (Fig. 1B type 9). I-F-5 was comprised of 72 genes identified in plants (Fig. 2), with four domain organizations (Fig. 1B types 9 and 12-14). The aquatic plants genes mainly contained only single DSP domain (Fig. 1B types 9, 14). Land plants genes contained AFK-DSP-ZnF_C_2H_2 domain organization (Fig. 1B type 12), mediating actin phosphorylation; and DSP-GH-CaMB domain organization (Fig. 1B type 13), joining Ca signaling and tyrosine phosphorylation. Weblogo analysis indicated that "AYLM" amino acid residues behind the C×$_5$R active site were conserved in I-F-5 (Fig. S6), the same as the HsMKPs, suggesting that those genes were GP MAPK phosphatases (GPMKP) and played pivotal roles in stress and hormone responses (Katou et al., 2007; Ulm et al., 2002; Ulm et al., 2001; Rainaldi et al., 2007; Strader et al., 2008b; Strader et al., 2008a; Lee and Ellis 2007; Xu et al., 1998; Yoo et al., 2004; Naoi and Hashimoto 2004; Quettier et al., 2006; Lee et al., 2009; Monroe-Augustus et al., 2003; Bartels et al., 2009). But the GPMKPs contained the different domain organizations (Fig. 1B types 9 and 12-14), compared with the HsMKPs (Alonso et al., 2004), which was combined of the inactive RHOD and DSP domains (Table S3), suggesting marked MKP divergence between plants and humans.

Sub. I-G (CDC14)

In humans, four CDC14 (cell division cycle 14, CDC14) proteins (Table S3) serve as a dephosphorylation antagonist of cyclin-dependent kinase (CDK) activity (Trautmann and McCollum, 2002). Ten CDC14 genes were identified in the green algae, bryophyte and pteridophyte (Figs. 1C and S7), with one DSP domain (Fig. 1B type 16), except *PpPTP-07* containing AAA domain (Fig. 1B type 17). The fact that CDC14 was absent in angiosperms, might be related to selection for the plant cell cycle pathway.

Sub. II (LMW-PTP)

In humans, *ACP1* (Table S3) is integral in cell regulation (Bottini et al., 2002). Eleven LMW-PTP were identified in plants without *Ot/Ol* (Fig. 1C), with LMW-PTP domain (Fig. 1B types 18-19). Phylogenetic analysis showed that there were two clades (Fig. 4A). The first showed that the same lineage genes were clustered together, respectively. The second clade only contained *VcPTP-24* and *PtPTP-37*, suggesting gene expansion and differentiation in the two plants.

Sub. III (RHOD)

In humans, three RHOD genes (Table S3) play function in CDK activation (Tabernero et al., 2008). Thirteen RHOD genes were identified in plants (Fig. 1C), with one RHOD domain (Fig. 1B type 20), similar as sulfur-transferases and Arc2 reductases (Bordo and Bork, 2002). Phylogenetic analysis showed that same lineage genes were respectively clustered together, which was also supported by their gene length and the special long N-terminal extension (Fig. 4B). It was argued that whether CDC25 works in the higher GPs (Boudolf et al., 2006), because of losing the N-terminal extensions from the green alga (*Cr/Vc*) to eudicots and monocots (Fig. 4B). In contrast, RHOD genes might function as Acr2 reductases (Duan et al., 2007; Liu et al., 2012b), suggesting that RHOD genes participated in varied signaling pathways (e.g. arsenic detoxification pathway) and elicited diverse functions between plants and humans.

Sub. IV (PTPL)

In humans, four PTPL genes (Table S3) play roles as PTP competitive inhibitor (Moorhead et al., 2009). Twenty-five PTPL genes were identified in plants (Fig. 1C), with one PTPL domains (Fig. 1B type 21), except two PTPL in *ZmPTP-23* (Fig. 1B type 22). Weblogo analysis revealed the characteristic H×××G××R/P signature sequence, similar to HC××G××R, but with a key conserved residue substitution (Fig. 1A), suggesting binding function without dephos-phorylation (Moorhead et al., 2009). Phylogenetic analysis of the PTPL domain showed two clades with the same pattern (Fig. 4C), which was consistent with PTHR11035-SF1 and PTHR11035-SF2 classifications in the PANTHER database.

GPPTPome evolution

GP species within the same lineage were much more similar in PTP number, subfamilies, and domain organization, but notable differences across lineages (Figs. 1 and 5A). The multi-domain PTP showed a GP conserved sequence array, including types 2-4, 6-7 and 10-13 (Fig. 1B), which suggested there was a common ancestor. The number and proportion of PTP multi-domains (Fig. 1B) in each species increased across each GP lineage, which was in consistent with their taxonomic hierachy.

Above all, an "original PTP kit" containing seven subfamilies was identified in *Cr/Vc* (Fig. 5B): I-A and II were

specific to the P-Tyr protein substrate (Alonso et al., 2004);

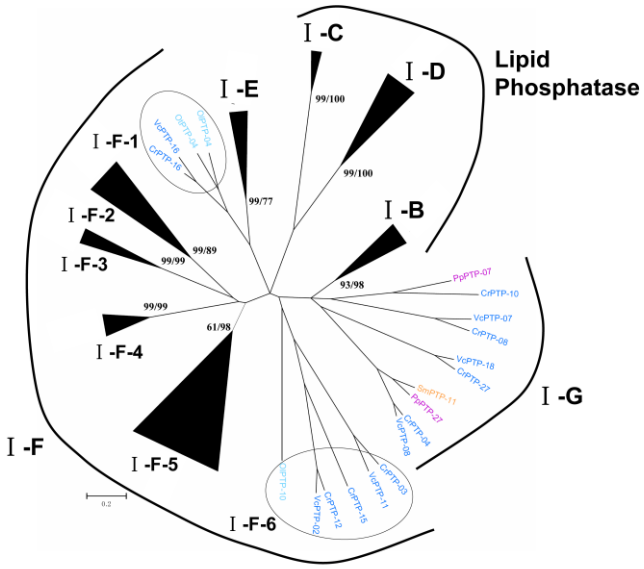

Fig 2. Phylogenetic relationships of DSPs in green plants. The unrooted phylogenetic tree of 215 Cys-based PTP genes with DSP (Subs. I-B to I-G) domains was constructed by the neighbor-joining (NJ) and maximum likelihood (ML) methods. Corresponding bootstrap values are shown and split by "/", respectively. Sequences are color coded by organism: blue for *Ot/Ol* and *Cr/Vc*; purple for *Pp*; brown for *Sm*.

I-E participated in the mRNA capping process, characteristic of eukaryotic organisms; I-F exhibited diverse PTPs, diversifying and leading to the origin of additional PTP subfamilies. The I-G was integral in cell cycle pathway selection (Clifford et al., 2008); III participated the cell cycle of ancestral GPs; and IV might function as a competitive inhibitor of PTP activities (Bellec et al., 2002).

Furthermore, there might be two selective forces during green plant radiation. I-A, I-G and II PTP in *Ot/Ol* exhibited dephosphorylation specificity and functional effectiveness, reflecting optimization of environmental adaptation. The other branch of green alga radiated into the angiosperms. As GPs adapted to land environments, LPs (I-B, I-C and I-D) were favored in terrestrial plants, and I-G was lost in the angiosperms (Fig. 5B). Three lipid phosphate subfamilies arose providing land plants the capacity to dephosphorylate phosphoinositide, favored in hydrophobic environments, which was concomitant with their movement onto land. The loss of I-G in eudicots and monocots might be due to selection for the cell cycle pathway in angiosperms, which possess unique B-type CDKs (Boudolf et al., 2006).

GPPTPome and HsPTPome comparisons

PTP evolution was further illustrated by performing a comprehensive comparison between GPPTPome and HsPTPome in gene numbers, subfamilies and accessory domains. There were comprised of 108 PTP genes in HsPTPome, notably greater than GPPTPome (Fig. 6; Table 1). GPPTP showed marked differences with HsPTP in the proportion of substrate specificity (Fig. 6). There was a lower proportion of P-Tyr substrate PTP (0-15%), and a higher proportion P-Tyr or P-Ser/Thr as substrates PTP (39-78%), which were distinctly different from HsPTP (36% and 39%), with comparable lipid substrate PTP (13-36%) in land plants and humans (20%). Relative to 12 PTP subfamilies found in HsPTPome, there were 4-11 subfamilies in green plants (Fig. 6), with three human exclusive PTP subfamilies (slingshots (SSH), regenerating liver phosphatases (PRL) and mitogen-activated protein kinase phosphatase (MKP) subfamilies), and one land plant exclusive PTP subfamily (I-D subfamily), suggesting different evolutionary pressures and LP preferences between plants and human. In addition, 24 and 15 different kinds of accessory domains were respectively detected in HsPTPome and GPPTPome, with special two seriate PTP domain in HsPTPome. Moreover, GPPTPome didn't have any extracellular domains, which was detected in HsPTP genes, including carbonic anhydrase-like (CA), fibronectin-like (FN), immunoglobulin-like (Ig) and meprin (Alonso et al., 2004), suggesting different regulation mechanism response to the extracellular signals.

Materials and methods

PTP data retrieval

To identify all candidate PTP homologs in green plants, the following procedures were performed: (a) Green plant genome downloads. Total 10 green plant genomes, representing five major lineages, were downloaded from public databases as follows: green algae (*Ot*, *Ol*, *Cr*, and *Vc*, JGI, http://genome.jgi-psf.org/); bryophyte (*Pp*, JGI, http://genome.jgi-psf.org/); pteridoplyte (*Sm*, JGI, http://genome.jgi-psf.org/); eudicots (*At*, TAIR, http://www.arabidopsis.org/; *Pt*, JGI, http://genome.jgi-psf.org/); monocots (*Os*, RGAP, http://rice.plantbiology.msu.edu/; *Zm*, http://www.maize sequence.org/).
(b) Selection of Hidden Markov Models (HMM) for PTP domains. PTP diversity was represented by six distinct Pfam profiles defining catalytic domains as follows: PF00102 (Y-phosphatase), PF01451 (low molecular weight, LMWP), PF00782 (DSPc), PF03162 (Y-phosphatase2), PF00581 (Rhodanese-like domain), and PF04387 (PTP-like, PTPL). The corresponding PTP profiles were downloaded from Pfam (Finn et al., 2006) and used to identify PTP candidate sequences in the ten green plant genomes.
(c) Retrieval of PTP sequences. PTP candidate catalytic domains were identified by use of the HMMER v2.0 search (Soding 2005) with an E-value cutoff of $<10^{-3}$ parameter. The candidate PTP domains selected from the HMMER search were compared at both the protein and DNA levels to further delete redundant PTPs using the following criteria: (1) sequences exceeding 95% identity at the DNA sequence level were considered the same gene (Zhang et al., 2001); (2) if the sequences were isoforms of one gene locus, the longest functional PTP was chosen; and (3) all potential PTP multiple sequence alignments were performed to determine if critical conserved and catalytic residues existed. All candidate PTP sequences were submitted to Pfam (Finn et al., 2006) and InterProScan (Hunter et al., 2009) to obtain information regarding PTP domains. (d) Repetitive database searches for PTP data. The above green plant PTP candidates were used to construct HMM profiles and HMMER v2.0 was used to search new GPPTP genes until no new PTP genes were identified. Finally, we identified complementary and non-redundant GPPTP data sets in the selected species, termed GPPTPome.

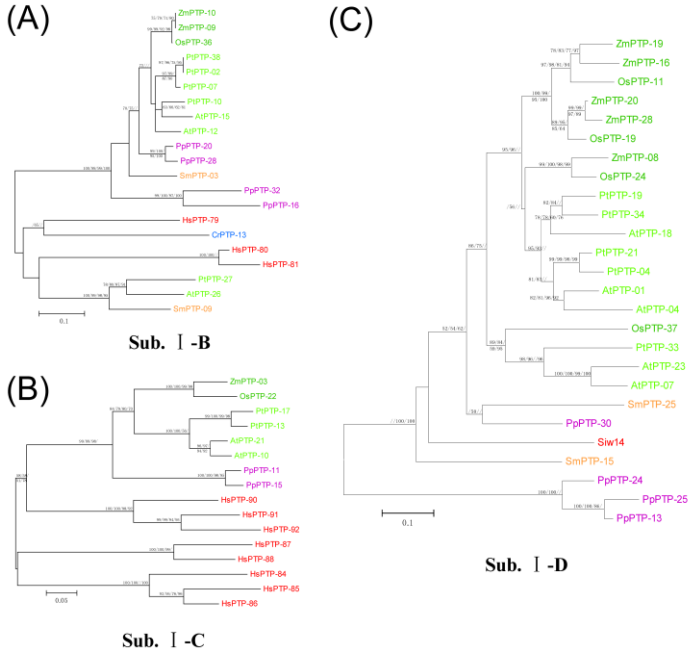

Fig 3. Phylogram of lipid phosphatases. (a) Phylogram of Sub. I-B (PTEN) GPPTPs. The phylogram of Sub. I-B DSP domain (PF00782) includes GPPTP I-B genes, CrPTP-13 and homologs of the Hs genes (Alonso et al. 2004) (Table S3) as comparative genes. (b) Phylogram of Sub. I-C (MTMR) GPPTPs. The phylogram of I-C DSP domain (PF00782) includes GPPTP I-C genes and homologs of the Hs genes (Alonso et al. 2004) (Table S3) as comparative genes. (c) Phylogram of Sub. I-D (PFA-DSP) GPPTPs. The phylogram of I-D DSP domain (PF03162) includes GPPTP I-D genes and homolog of the S. cerevisiae genes (Table S3) as comparative genes. Sequences are color-coded by organism: blue for Ot/Ol and Cr/Vc; purple for Pp; brown for Sm; green for At/Pt and Os/Zm; and red for Hs or Siw14. Four phylogenetic trees depict results of neighbor-joining (NJ), minimal evolution (ME), maximum parsimony (MP) and maximum likelihood (ML) methods with corresponding bootstrap values (> 50) shown and split by "/", respectively.

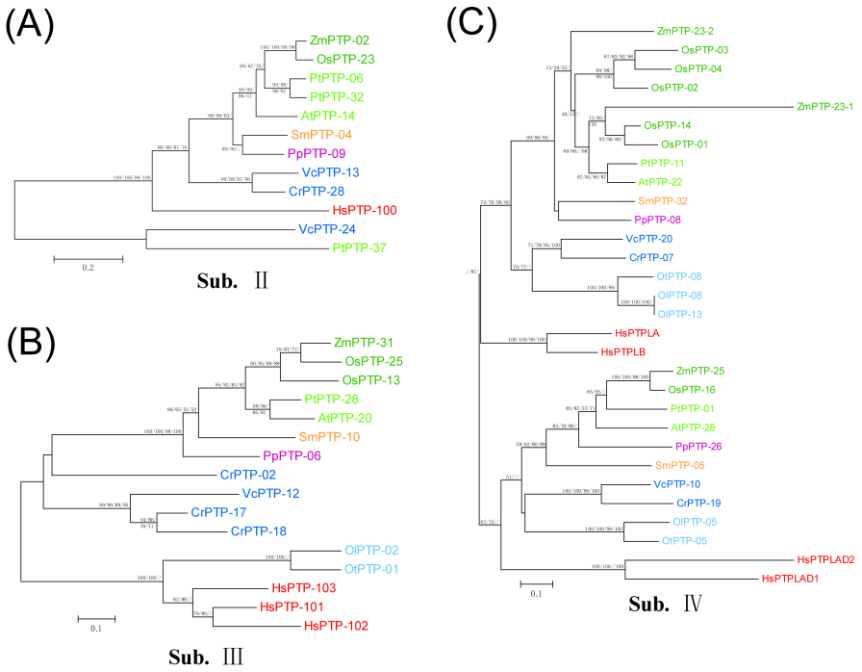

Fig 4. Phylogram of Subs. II, III and IV GPPTPs. (a) Phylogram of Sub. II (LMW-PTP) GPPTPs. The phylogram of II LMWPc domain (PF01451) includes GPPTP II genes and homologs of the *Hs* genes (Table S3) as comparative genes. (b) Phylogram of Sub. III (RHOD) GPPTPs. The phylogram of III RHOD domain (PF00581) includes GPPTP III genes and homologs of the *Hs* genes (Alonso et al. 2004) (Table S3) as comparative genes. (c) Phylogram of Sub. IV (PTPL) GPPTPs. The phylogram of IV PTPL domain (PF04387) includes GPPTP IV genes and homologs of the *Hs* genes (Alonso et al. 2004) (Table S3) as comparative genes.

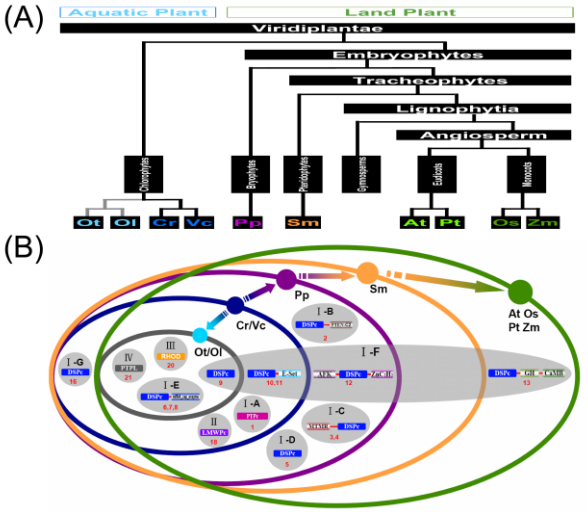

Fig 5. The model of GPPTPome evolution. (a) Phylogenetic relationships of the five plant lineages. *Chlamydomonas reinhardtii* (Cr) represents the ancestral species, and following plant radiation, two pathways can be inferred: one branch follows Cr evolving to *Volvox carteri* (Vc), the bryophytes represented by Physcomitrella patens (Pp), pteridoplytes represented by *Selaginella moellendorfii* (Sm), eudicots represented by *Arabidopsis thaliana* (At) and Populus trichocarpa (Pt), and the monocots represented by *Oryza sativa* (Os) and *Zea mays* (Zm) from ancestral to derived taxa; the other lineage branches from Cr but does not diversify and maintains an ancestral algal lineage represented by *Ostreococcus tauri* (Ot) and *Ostreococcus lucimarinus* (Ol). (b) GPPTP subfamilies shift in evolution model. The Venn diagram indicates a PTP subfamily stepwise shift model according to green plant taxonomy location from the left to the right. The Cr/Vc (dark blue) contained the "original PTP kit" comprised of seven PTP subfamilies: I-A (classic PTP), I-E (mRNA), I-F (atypical DSP, aDSP), I-G (CDC14-like), II (LMW-PTP), III (RHOD) and IV (PTPL). One branch of plant radiation indicated a decrease in PTP subfamilies (I-A, I-G and II) existed in Ot/Ol (sky blue). On the other branch leading to the angiosperms, the addition of I-B (PTEN), I-C (MTMR) and I-D (PFA-DSP) subfamilies occurred in the land plants (Pp [purple], Sm [brown] and the angiosperms [green]); and the I-G subfamily was lost in the angiosperms. Domain organization types are shown by pattern and corresponding red Arabic numerals in Fig. 1.

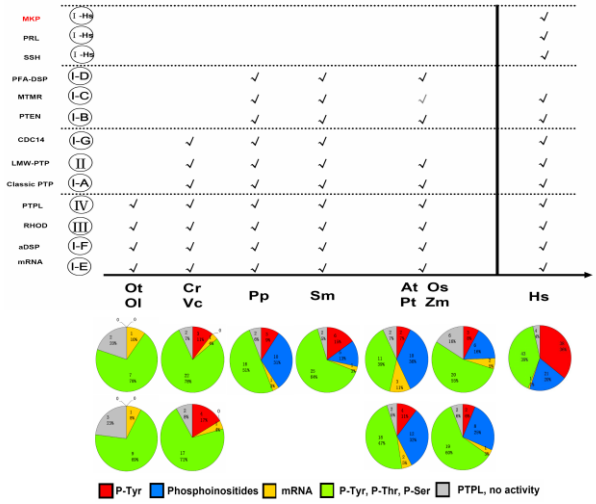

Fig 6. Comparison of GPPTPome and HsPTPome. The upper section shows the PTP subfamilies distribution in each PTPome. The "√" represents the presence of the PTP subfamily in each species. The dashed lines show the stepwise shift of the PTP subfamilies. HsMKPs (red font, because HsMKPs have a special sequence characterization different from GPMKPs), PRL and SSH only exist in the HsPTPome (Alonso et al. 2004). The lower pie charts depict the number and proportion of PTP substrate specificity in each PTPome. P-Tyr represents strict tyrosine specificity, including the PTP subfamilies I-A (classic PTP), II (LMW-PTP) and PRL; Phosphoinostides represent the I-B (PTEN), I-C (MTMR) and I-D (PFA-DSP) PTP subfamilies, which can dephosphorylate lipids; mRNA indicates mRNA substrate, including the PTP subfamily I-E (mRNA); P-Tyr, P-Thr and P-Ser indicates tyrosine and serine/threonine substrates, including the PTP subfamilies I-F (atypical DSP, aDSP), I-G (CDC14-like), III (RHOD), SSH, and MKP. PTPL represents the PTP subfamily IV (PTPL), which shows no catalytic activity.

Domain combination analysis

Following GPPTP data retrieval, we investigated GPPTP gene domain combinations. Pfam (Finn et al., 2006) and InterProScan (Hunter et al., 2009) were employed to characterize every PTP nonphosphatase domain (GPPTP accessory domain) accessory domain. Accessory domains identified in the data base analysis corresponding to the Pfam profiles were used in the HMMER v2.0 search (Soding 2005) in the selected plants with an E-value cutoff of $<10^{-3}$ parameter.

Multiple sequence alignment and phylogenetic analysis

PTP catalytic domain alignments were generated with Clustal X v2.0 (Larkin et al., 2007) using the following parameters: BLOSUM 30 for the protein weight matrix, and gap extension penalty as the default and manual edit applying BioEdit (Hall and Cantley, 1999). Detailed sequence analyses were performed for each GPPTP subfamily with the catalytic PTP domains and accessory domains. The evolutionary relationships among GPPTP proteins (GPPTP subfamilies) and a GPPTP protein phylogenetic tree was reconstructed using MEGA v4.0 (Tamura et al., 2007) with the following options: neighbor-joining (NJ), minimal evolution (ME), and maximum parsimony (MP) analyses. NJ and ME methods included parameters of the p-distance model. Pairwise deletions of gaps/missing data were used in the Poisson correction data set for the distance model. For both methods, topologies were generated as the consensus of 1,000 bootstrap alignment replicates. The maximum likelihood (ML) trees were built using PhyML v3.0 (http://www.atgc-montpellier.fr/phyml) (Guindon and Gascuel 2003) with the following options: Blosum62 amino acid substitution model and the default BIONJ distance-based tree as a starting tree with 100 bootstrap values. Finally, the phylogenetic trees were visualized using MEGA v4.0.

Conclusion

Our study has identified 281 PTP genes in green plants and classified them into ten subfamilies, based on the mode of

PTP catalysis, phylogenetic analysis and domain organization analysis. Based on species and subfamily levels, we found that same lineage species showed similar PTPome, but marked differences among lineages plants, suggesting a PTP subfamily stepwise shift model during green plants evolution, with lipid phosphatases only detected in land plants. Our comprehensive and evolutionary analysis of PTP genes contributed to the understanding of plant PTP distribution and evolution, and comparisons between plant and animal PTPome elucidated differences in PTPs between divergent biological organisms.

Acknowledgements

This research was supported by grants from the National Natural Science Foundation of China (No. 30970237 and No. 31100199), the Natural Science Foundation of Guangdong Province, P. R. China (No. 8151027501000016) and the Fundamental Research Funds for the Central Universities 12lgpy35.

References

Alonso A, Sasin J, Bottini N, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T (2004) Protein tyrosine phosphatases in the human genome. Cell. 117(6):699-711.

Andersen JN, Mortensen OH, Peters GH, Drake PG, Iversen LF, Olsen OH, Jansen PG, Andersen HS, Tonks NK, Moller NP (2001) Structural and evolutionary relationships among protein tyrosine phosphatase domains. Mol Cell Biol. 21(21):7117-7136.

Andreeva AV, Kutuzov MA (2008) Protozoan protein tyrosine phosphatases. Int J Parasitol. 38(11):1279-1295.

Bartels S, Anderson JC, Besteiro MA, Carreri A, Hirt H, Buchala A, Metraux JP, Peck SC, Ulm R (2009) MAP kinase phosphatase1 and protein tyrosine phosphatase1 are repressors of salicylic acid synthesis and SNC1-mediated responses in Arabidopsis. Plant Cell. 21(9):2884-2897.

Bellec Y, Harrar Y, Butaeye C, Darnet S, Bellini C, Faure JD (2002) Pasticcino2 is a protein tyrosine phosphatase-like involved in cell proliferation and differentiation in Arabidopsis. Plant J. 32(5):713-722.

Bordo D, Bork P (2002) The rhodanese/Cdc25 phosphatase superfamily. Sequence-structure-function relations. EMBO Rep. 3(8):741-746.

Bottini N, Bottini E, Gloria-Bottini F, Mustelin T (2002) Low-molecular-weight protein tyrosine phosphatase and human disease: in search of biochemical mechanisms. Arch Immunol Ther Exp (Warsz). 50(2):95-104.

Boudolf V, Inze D, De Veylder L (2006) What if higher plants lack a CDC25 phosphatase? Trends Plant Sci. 11(10):474-479.

Brenchley R, Tariq H, McElhinney H, Szoor B, Huxley-Jones J, Stevens R, Matthews K, Tabernero L (2007) The TriTryp phosphatome: analysis of the protein phosphatase catalytic domains. BMC Genomics. 8:434.

Byrum CA, Walton KD, Robertson AJ, Carbonneau S, Thomason RT, Coffman JA, McClay DR (2006) Protein tyrosine and serine-threonine phosphatases in the sea urchin, Strongylocentrotus purpuratus: identification and potential functions. Dev Biol. 300(1):194-218.

Clifford DM, Chen CT, Roberts RH, Feoktistova A, Wolfe BA, Chen JS, McCollum D, Gould KL (2008) The role of Cdc14 phosphatases in the control of cell division. Biochem Soc Trans. 36(Pt 3):436-438.

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res. 14(6):1188-1190.

Duan GL, Zhou Y, Tong YP, Mukhopadhyay R, Rosen BP, Zhu YG (2007) A CDC25 homologue from rice functions as an arsenate reductase. New Phytol. 174(2):311-321.

Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. Nucleic Acids Res. 34:D247-251.

Ghelis T (2011) Signal processing by protein tyrosine phosphorylation in plants. Plant Signaling Behavior. 6(7):942-951.

Ghelis T, Bolbach G, Clodic G, Habricot Y, Miginiac E, Sotta B, Jeannette E (2008) Protein Tyrosine Kinases and Protein Tyrosine Phosphatases Are Involved in Abscisic Acid-Dependent Processes in Arabidopsis Seeds and Suspension Cells. Plant Physiology. 148(3):1668-1680.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52(5):696-704.

Gupta R, Ting JT, Sokolov LN, Johnson SA, Luan S (2002) A tumor suppressor homolog, AtPTEN1, is essential for pollen development in Arabidopsis. Plant Cell. 14(10):2495-2507.

Hall TA, Cantley LC (1999) BioEdit: a user-friendly biological sequence alignme editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 41:95-98.

He H, Su J, Shu S, Zhang Y, Ao Y, Liu B, Feng D, Wang J, Wang H (2012) Two homologous putative protein tyrosine phosphatases, OsPFA-DSP2 and AtPFA-DSP4, negatively regulate the pathogen response in transgenic plants. Plos One. 7(4):e34995.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2009) InterPro: the integrative protein signature database. Nucleic Acids Res. 37:D211-215.

Katou S, Kuroda K, Seo S, Yanagawa Y, Tsuge T, Yamazaki M, Miyao A, Hirochika H, Ohashi Y (2007) A calmodulin-binding mitogen-activated protein kinase phosphatase is induced by wounding and regulates the activities of stress-related mitogen-activated protein kinases in rice. Plant Cell Physiol. 48(2):332-344.

Kerk D, Bulgrien J, Smith DW, Barsam B, Veretnik S, Gribskov M (2002) The complement of protein hosphatase catalytic subunits encoded in the genome of Arabidopsis. Plant Physiol. 129(2):908-925.

Kerk D, Moorhead GB (2010) A phylogenetic survey of myotubularin genes of eukaryotes: distribution, protein structure, evolution, and gene expression. BMC Evol Biol. 10:196.

Kerk D, Templeton G, Moorhead GB (2008) Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. Plant Physiol. 146(2):351-367.

Kim SJ, Ryu SE (2012) Structure and catalytic mechanism of human protein tyrosine phosphatome. BMB reports. 45(12):693-699.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics. 23(21):2947-2948.

Lee JS, Ellis BE (2007) Arabidopsis MAPK phosphatase 2 (MKP2) positively regulates oxidative stress tolerance and inactivates the MPK3 and MPK6 MAPKs. J Biol Chem. 282(34):25020-25029.

Lee JS, Wang S, Sritubtim S, Chen JG, Ellis BE (2009) Arabidopsis mitogen-activated protein kinase MPK12 interacts with the MAPK phosphatase IBR5 and regulates auxin signaling. Plant J. 57(6):975-985.

Liu B, Fan J, Zhang Y, Mu P, Wang P, Su J, Lai H, Li S, Feng D, Wang J, Wang H (2012a) OsPFA-DSP1, a rice protein tyrosine phosphatase, negatively regulates drought stress responses in transgenic tobacco and rice plants. Plant Cell Rep. 31(6):1021-1032.

Liu W, Schat H, Bliek M, Chen Y, McGrath SP, George G, Salt DE, Zhao F-J (2012b) Knocking out *ACR2* does not affect arsenic redox status in *Arabidopsis thaliana*: Implications for As detoxification and accumulation in plants. PLOS ONE. 7(8):e42408.

Luan S (2003) Protein phosphatases in plants. Annu Rev Plant Biol. 54:63-92.

Maehama T, Taylor GS, Dixon JE (2001) PTEN and myotubularin: novel phosphoinositide phosphatases. Annu Rev Biochem. 70:247-279.

Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci. 27(10):514-520.

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. Genes Dev 19 (12):1401-1415.

Monroe-Augustus M, Zolman BK, Bartel B (2003) IBR5, a dual-specificity phosphatase-like protein modulating auxin and abscisic acid responsiveness in Arabidopsis. Plant Cell. 15(12):2979-2991.

Moorhead GB, De Wever V, Templeton G, Kerk D (2009) Evolution of protein phosphatases in plants and animals. Biochem J. 417(2):401-409.

Naoi K, Hashimoto T (2004) A semidominant mutation in an Arabidopsis mitogen-activated protein kinase phosphatase-like gene compromises cortical microtubule organization. Plant Cell. 16(7):1841-1853.

Niittyla T, Comparot-Moss S, Lue WL, Messerli G, Trevisan M, Seymour MD, Gatehouse JA, Villadsen D, Smith SM, Chen J, Zeeman SC, Smith AM (2006) Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. J Biol Chem. 281(17):11815-11818.

Quettier AL, Bertrand C, Habricot Y, Miginiac E, Agnes C, Jeannette E, Maldiney R (2006) The phs1-3 mutation in a putative dual-specificity protein tyrosine phosphatase gene provokes hypersensitive responses to abscisic acid in Arabidopsis thaliana. Plant J. 47(5):711-719.

Rainaldi M, Yamniuk AP, Murase T, Vogel HJ (2007) Calcium-dependent and -independent binding of soybean calmodulin isoforms to the calmodulin binding domain of tobacco MAPK phosphatase-1. J Biol Chem. 282(9):6031-6042.

Rayapureddi JP, Kattamuri C, Chan FH, Hegde RS (2005) Characterization of a plant, tyrosine-specific phosphatase of the aspartyl class. Biochemistry. 44(2):751-758.

Roma-Mateo C, Sacristan-Reviriego A, Beresford NJ, Caparros-Martin JA, Culianez-Macia FA, Martin H, Molina M, Tabernero L, Pulido R (2011) Phylogenetic and genetic linkage between novel atypical dual-specificity phosphatases from non-metazoan organisms. Mol Genet Genomics. 285(4):341-354.

Singh A, Giri J, Kapoor S, Tyagi A, Pandey G (2010) Protein phosphatase complement in rice: genome-wide identification and transcriptional analysis under abiotic stress conditions and reproductive development. BMC Genomics. 11(1):435.

Soding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics. 21(7):951-960.

Strader LC, Monroe-Augustus M, Bartel B (2008a) The IBR5 phosphatase promotes Arabidopsis auxin responses through a novel mechanism distinct from TIR1-mediated repressor degradation. BMC Plant Biol. 8:41.

Strader LC, Monroe-Augustus M, Rogers KC, Lin GL, Bartel B (2008b) Arabidopsis iba response5 suppressors separate responses to various hormones. Genetics. 180(4): 2019-2031.

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Mol Syst Biol. 4:193.

Tabernero L, Aricescu AR, Jones EY, Szedlacsek SE (2008) Protein tyrosine phosphatases: structure-function relationships. FEBS J. 275(5):867-882.

Takagi T, Walker AK, Sawa C, Diehn F, Takase Y, Blackwell TK, Buratowski S (2003) The Caenorhabditis elegans mRNA 5'-capping enzyme. In vitro and in vivo characterization. J Biol Chem. 278(16):14174-14184.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 24 (8):1596-1599.

Tonks NK, Neel BG (2001) Combinatorial control of the specificity of protein tyrosine phosphatases. Curr Opin Cell Biol. 13(2):182-195.

Trautmann S, McCollum D (2002) Cell cycle: new functions for Cdc14 family phosphatases. Curr Biol. 12(21):R733-735.

Ulm R, Ichimura K, Mizoguchi T, Peck SC, Zhu T, Wang X, Shinozaki K, Paszkowski J (2002) Distinct regulation of salinity and genotoxic stress responses by Arabidopsis MAP kinase phosphatase 1. Embo J. 21(23):6483-6493.

Ulm R, Revenkova E, di Sansebastiano GP, Bechtold N, Paszkowski J (2001) Mitogen-activated protein kinase phosphatase is required for genotoxic stress relief in Arabidopsis. Genes Dev. 15(6):699-709.

Wilkes JM, Doerig C (2008) The protein-phosphatome of the human malaria parasite Plasmodium falciparum. BMC Genomics. 9(1):412.

Xu Q, Fu HH, Gupta R, Luan S (1998) Molecular characterization of a tyrosine-specific protein phosphatase encoded by a stress-responsive gene in Arabidopsis. Plant Cell. 10(5):849-857.

Yamakawa H, Katou S, Seo S, Mitsuhara I, Kamada H, Ohashi Y (2004) Plant MAPK phosphatase interacts with calmodulins. J Biol Chem. 279(2):928-936.

Yoo JH, Cheong MS, Park CY, Moon BC, Kim MC, Kang YH, Park HC, Choi MS, Lee JH, Jung WY, Yoon HW, Chung WS, Lim CO, Lee SY, Cho MJ (2004) Regulation of the dual specificity protein phosphatase, DsPTP1, through interactions with calmodulin. J Biol Chem. 279(2):848-858.

Zhang, L., Pond, S.K., and Gaut, B.S. (2001). A survey of the molecular evolutionary dynamics of twenty-five multigene families from four grass taxa. J Mol Evol. 52:144-156.