Original article

dbPPT: a comprehensive database of protein phosphorylation in plants

Han Cheng[1,†], Wankun Deng[1,†], Yongbo Wang[1], Jian Ren[2], Zexian Liu[1,]* and Yu Xue[1,]*

[1]Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China and [2]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

*Corresponding author Tel/Fax: +86 27 87793172; Email: lzx@hust.edu.cn

Correspondence may also be addressed to Yu Xue, Tel: +86 27 87793903; Fax: +86 27 87793172; Email: xueyu@hust.edu.cn

[†]These authors contributed equally to this work.

Abstract

As one of the most important protein post-translational modifications, the reversible phosphorylation is critical for plants in regulating a variety of biological processes such as cellular metabolism, signal transduction and responses to environmental stress. Numerous efforts especially large-scale phosphoproteome profiling studies have been contributed to dissect the phosphorylation signaling in various plants, while a large number of phosphorylation events were identified. To provide an integrated data resource for further investigations, here we present a comprehensive database of dbPPT (database of Phosphorylation site in PlanTs, at http://dbppt.biocuckoo.org), which contains experimentally identified phosphorylation sites in proteins from plants. The phosphorylation sites in dbPPT were manually curated from the literatures, whereas datasets in other public databases were also integrated. In total, there were 82 175 phosphorylation sites in 31 012 proteins from 20 plant organisms in dbPPT, presenting a larger quantity of phosphorylation sites and a higher coverage of plant species in comparison with other databases. The proportions of residue types including serine, threonine and tyrosine were 77.99, 17.81 and 4.20%, respectively. All the phosphoproteins and phosphorylation sites in the database were critically annotated. Since the phosphorylation signaling in plants attracted great attention recently, such a comprehensive resource of plant protein phosphorylation can be useful for the research community.

Database URL: http://dbppt.biocuckoo.org

Introduction

As one of the most ubiquitous post-translational modifications (PTMs), protein phosphorylation is critical for almost all types of cellular processes such as cell cycle and signaling transduction (1, 2). Through transferring a phosphate group from ATP to the hydroxyl group of a residue, in most cases, serine, threonine and tyrosine (3), phosphorylation could alter the structural and/or functional state of the protein (4). Phosphorylation was reversibly catalyzed by kinases and phosphatase (3), and then contributed to the dynamic cellular signaling in organisms. In plants, phosphorylation plays important roles in a variety of critical processes such as regulation of photosystem (5), metabolism (6). For example, early in 1980s, Bennett *et al*. (7) found that a membrane-bound protein kinase could regulate excitation energy transfer through phosphorylation of surface-exposed segments of the light-harvesting pigment–protein complex in the photosystem. Bachmann *et al*. (8) identified S543 as the major phosphorylation site which could regulate nitrate metabolism through control the activity of leaf nitrate reductase in spinach. Recently, phosphorylation was found to be heavily involved in the plant circadian system through various aspects including circadian oscillator function, circadian entrainment, circadian output pathways and circadian regulations of cellular processes such as ABA signaling (9). Accumulated studies also showed that phosphorylation was highly implicated in the response of various stresses (10). Taken together, understanding of phosphorylation signaling is critical to reveal the physiological and pathological processes in plants.

Recently, rapid progress in phosphopeptide enrichment methods and mass spectrum (MS)-based identification techniques greatly advanced large-scale detection of phosphorylation events in plants (11). Early in 2004, Heintz *et al*. (12) published a protocol for identification of phosphorylation in moss *Physcomitrella patens*, with a detection of 253 distinct phosphopeptides. However, although the phosphoproteomics techniques in yeast and mammals developed quickly, phosphoproteome analysis in plants is still challenging (13). In 2008, Sugiyama *et al*. (14) identified over 2000 of phosphotyrosine sites in *Arabidopsis* with the extent of phosphotyrosines, demonstrating the great advancement of phosphoproteomics in plants (15). Quantitative techniques were also introduced (16, 17), while recently a number of studies employed quantitative phosphoproteomic to study the phosphorylation signaling dynamics in plants (18–20). With the accumulation of studies on phosphorylation in plants, a large number of phosphoproteins and sites were identified, while data maintenance and sharing became increasingly challenging.

To deposit the precious phosphorylation information for easy retrieving, a handful of studies have been contributed. Although databases such as PhosphoSitePlus (21), Phospho.ELM (22), dbPTM (23) and SysPTM (24) provided comprehensive information of phosphorylation sites, these databases were mainly focused on yeast and animals, while data for plant is limited. Meanwhile, a dozen of databases focused on plants were developed, such as PlantsP (25), PhosPhAt (26) and P^{3}DB (27). However, there were limited phosphorylation data or lower coverage of species in most of these databases. For example, PlantsP contains only 4969 phosphoproteins (25), while PhosPhAt and MPPD are exclusively developed for *Arabidopsis* and *Medicago*, respectively (26, 28). Thus, a more comprehensive database is needed for further studies of phosphorylation in plants.

In this study, we provided a large collection of 82 175 phosphorylation sites in 31 012 phosphoproteins that were manually curated from literatures. The proportions for phosphorylated serines, threonines and tyrosines were 77.99, 17.81 and 4.20%, respectively. The phosphoproteins were from 20 plant species, whereas 29 058 phosphorylation sites in 9003 phosphoproteins were from *Arabidopsis*. Most phosphorylation sites were mapped to the sequences from the UniProt database. All the phosphoproteins were well annotated with various descriptions and the source references for phosphorylation identifications were also provided. We further constructed the database of dbPPT (database of Phospho-sites in PlanTs) to provide user-friendly online services for data retrieval and access. The keyword-based search and blast search services were implemented. The browse service was also provided to query information for a specific organism of interest. As an application of the dbPPT database, the most significant sequence motifs for phosphorylation in *Arabidopsis* were analyzed. Taken together, dbPPT serves as the most comprehensive data resource for phosphorylation in plants, and would be helpful for the research community

Database construction and content

To construct the most comprehensive database for phosphorylation in plants, we integrated data from multiple different sources including PubMed, P^{3}DB and PhosPhAt. The literature database of PubMed (http://www.ncbi.nlm.nih.gov/pubmed) was searched to retrieve published plant phosphorylation-related studies through various keywords combinations such as 'plant(s) phosphoproteomic', 'plant(s) phosphoproteome' and more than 20 plant species names combined with 'phosphoproteomic' or 'phosphoproteome' like '*Arabidopsis thaliana* phosphoproteomic', '*Arabidopsis thaliana* phosphoproteome'. To avoid missing data,

additional keywords like 'plant(s) large-scale phosphoryl-ation', 'plant(s) MS phosphorylation' and other related key-words were also employed to search PubMed. The retrieved literatures were manually checked to collect the experimen-tally identified phosphorylation substrates and sites in plants. The collected sites were cross checked with other databases such as P³DB and PhosPhAt.

Since there might be differences among protein sequences from different versions or sources, we employed the UniProt (Release 2014_06) database (29) as the source of bench-mark sequences. For most species, their sequences could be found in UniProt. However, the sequences for a number of species such as *Brassica napus* could not be fully mapped to UniProt entries. Therefore, sequence information from other public databases such as Ensembl (Release 75) (30) and NCBI Protein (31) were employed as the benchmark to map the collected phosphorylation sites. A full list of the data-bases used in this study was provided in Supplementary Table S1. To provide further information for the curated phosphoproteins, the annotations were also retrieved from the UniProt and other public databases. To ensure the high quality of dbPPT and allow the user to query further infor-mation about the identified phosphorylation sites, the pri-mary references were also provided.

In total, 82 175 identified phosphorylation sites in 31 012 phosphoproteins were provided in dbPPT, while the numbers (and proportions) of phosphoserine, phosphothreonine and phosphotyrosine were 64 089 (77.99%), 14 632 (17.81%) and 3454 (4.20%), respectively (Supplementary Table S2). Obviously, most of the identified phosphorylation sites were phosphoserines, while the number of phosphotyrosine was limited. However, with the advancement of phosphoproteo-mics techniques, more tyrosine phosphorylation will be iden-tified (32). As the model organism for plants, *Arabidopsis* attracted special attentions from the research community as 29 058 phosphorylation sites in 9003 proteins were identi-fied. Other plants especially the crops also have considerable number of identified phosphorylation events. For *Zea mays*, *Medicago truncatula* and *Oryza sativa* subsp. *Japonica*, there were 13 580, 14 836 and 8603 identified phosphorylation sites, respectively. The heatmaps for the data summarization of the numbers and proportions of phosphorylation sites were presented in Figure 1, while the details were shown in Supplementary Table S2. All the phosphorylation sites and sequences of proteins in the dbPPT database can be down-loaded at http://dbppt.biocuckoo.org/download.php.

Database usage

To provide convenient services for the research commu-nity, browse and search functions with various options are developed in a user-friendly manner for the dbPPT

database. The browse function is developed for organism-based querying among the 20 plant species. Using the im-portant crop wheat as an example, the workflow for browsing is shown in Figure 2. The plant species are pre-sented in an evolutionary tree on the 'BROWSE' page with images and organism names, while user could click the image or organism name for one species to browse the phosphoproteins (Figure 2A). If '*Triticum aestivum*' is se-lected, then all the identified phosphoproteins of wheat are listed (Figure 2B). The dbPPT entry, sequence ID (the ID in the database from which the sequence of the protein is retrieved, for example, UniProt Accesion) and the protein names (including gene name, gene synonyms, protein name and protein synonyms) are shown for browsing. User could click the dbPPT entry for the protein of interest to view the details of the phosphoprotein (Figure 2B). General infor-mation for the protein such as protein ID and gene ID are shown in this page (Figure 2C). The phosphorylation sites are presented with position number, peptide with 7 amino acids upstream and downstream of the modified residue, the data source of the site and the link to the primary refer-ence (Figure 2C). The 'dbPPT' in the 'Source' column means the site is collected in this study, while other such as 'P³DB' represents that the site is cross checked with the P³DB database. Annotations such as gene ontology and Interpro are also provided for further understanding of the phosphoprotein.

To provide convenient querying services, four search op-tions are implemented in dbPPT. In the home page of dbPPT, a simple search option is provided, which allows user to perform keyword-based search in specific area (Figure 3A). An example with SNF1-type serine–threonine protein kinase SnRK2.4 is shown by searching the keyword 'snrk2' in the 'Protein Name' area. After submitting the query, the results will be shown in a tabular format with dbPPT ID, Sequence ID and protein/gene names/synonyms (Figure 3B). Furthermore, three additional search options including advanced search, multiple search, BLAST search are also implemented and available in the 'SEARCH' page of dbPPT. In advanced search option, three keywords can be used to perform a complex search in three areas through combination with three operators of 'and', 'or' and 'exclude' (Figure 3C), while the results are shown in the format as the simple search option (Figure 3D). In multiple search option, several keywords such as a list of gene names can be used to search one area (Figure 3E), all the proteins hits any one of these keywords will be presented in the results (Figure 3F). In addition, the BLAST search option is designed to search phosphoproteins in dbPPT with a protein sequence in the FASTA format. User could submit a protein sequence to run a homologous search through NCBI BLAST packages (33) with a specified *E*-Value threshold (Figure 3G).
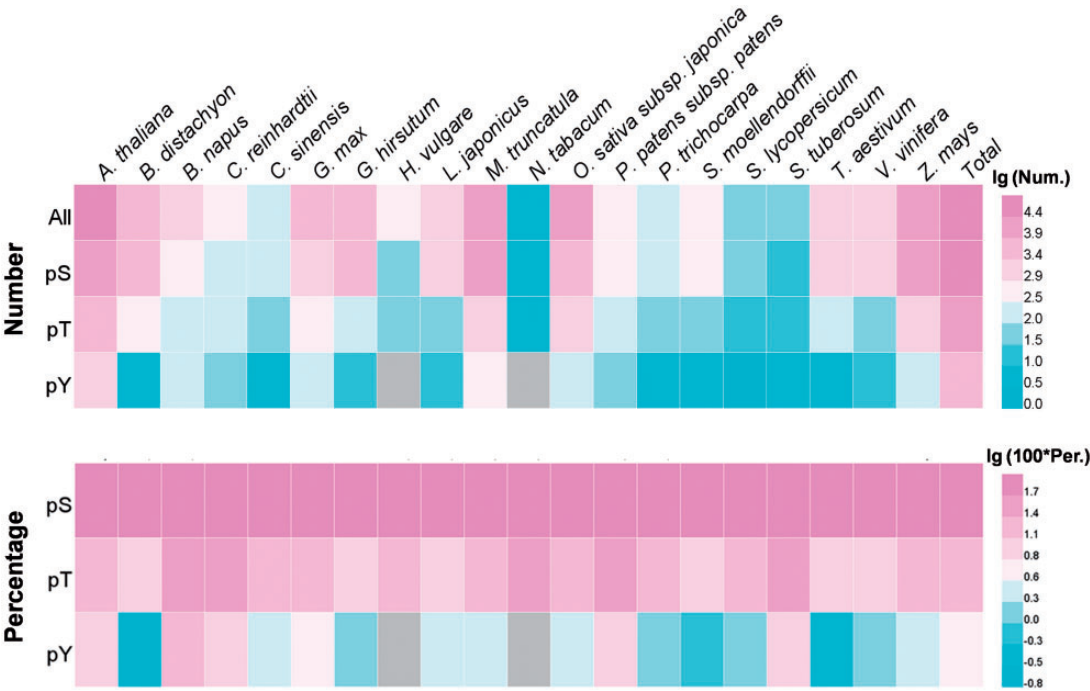
Figure 1. The heatmap for the distribution of site numbers and percentages for phosphoserine (pS), phosphothreonine (pT) and phosphotyrosine (pY) in different species.
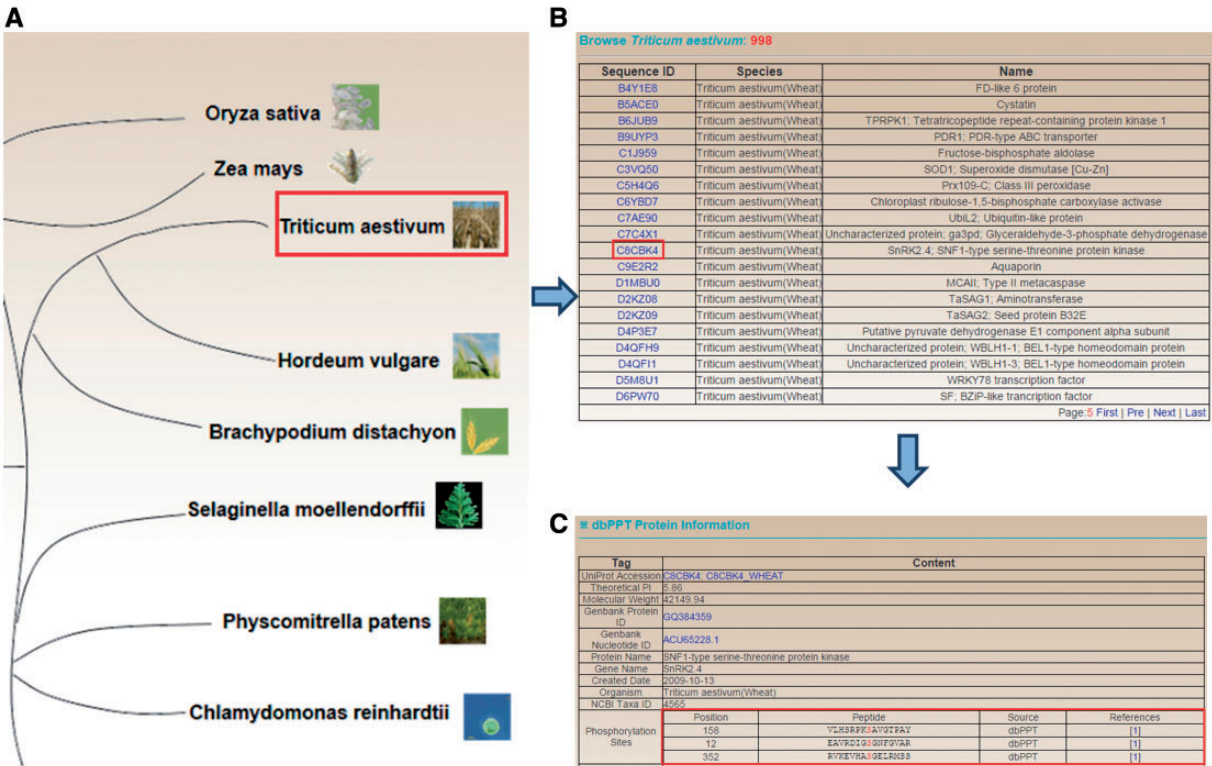

Figure 2. The browse option developed for browsing by plant species in the dbPPT database. (A) The 20 plant species which contain identified phosphorylation sites. (B) The list of identified phosphoprotein in *T. aestivum*. (C) The detailed information of *T. aestivum* SNF1-type serine–threonine protein kinase.
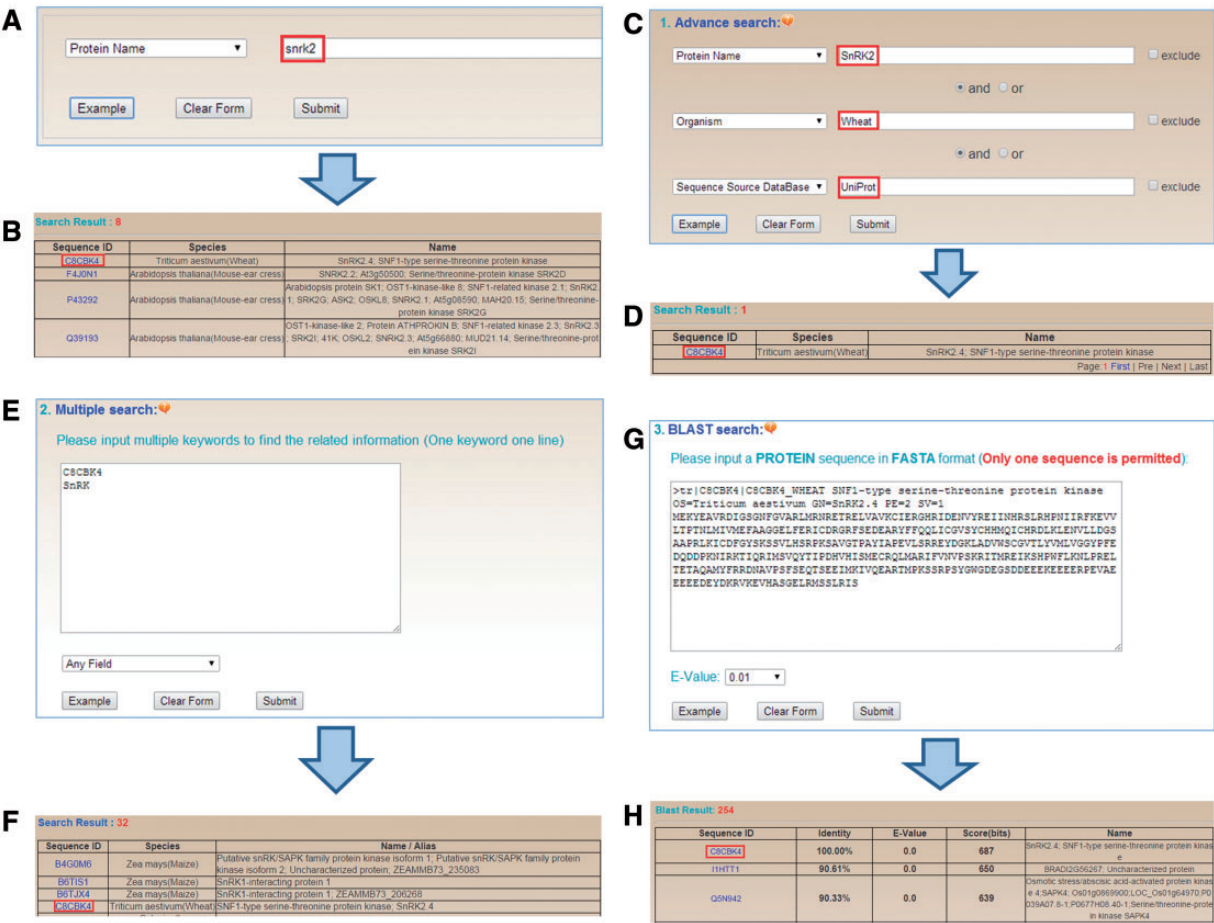
Figure 3. The search options in the dbPPT database. (A) dbPPT can be queried with a simple keyword. (B) The results of simple search for 'snrk2' in the area 'Protein Name'. (C) Advanced search allows users to perform a complex search using multiple keywords in several areas and three operators including 'and', 'or' and 'exclude'. (D) Advanced search result of 'SnRK2' in wheat while its full-length sequence was retrieved from UniProt. (E) Multiple search allows user to search multiple keywords in one query event. (F) The results of searching 'C8CBK4' and 'SnRK' in any field. (G) BLAST search was implemented for search similar phosphoproteins with FASTA format protein sequence. (H) BLAST search for SNF1-type serine–threonine protein kinase in wheat.

Search results of BLAST search is presented in a tabular format with dbPPT ID, sequence similarity evaluations including identity, E-value and score, sequence ID and protein/gene name/synonyms (Figure 3H). Furthermore, all the phosphoproteins in the dbPPT database could be accessed directly through the link of 'http://dbppt.biocuckoo.org/getTable.php?seqID=accession', while the 'accession' is the sequence accession such as UniProt accession.

Conclusion and future direction

As one of the most important PTMs, protein phosphorylation was involved in almost all aspects of biology processes (1, 2). Phosphorylation in yeast and animals was extensively studied in recent years while tens of thousands of phosphorylation sites were identified (21–24). Although the advancement for understanding of phosphorylation in plants is limited by comparison with yeast and animals,

recently more and more studies were contributed to this area to dissect the important roles of phosphorylation in the physiological and pathological processes in plants (11). Besides the requirement for maintaining the phosphorylation data which was in explosive increase, a comprehensive data resource was the foundation of systematic analyses of phosphorylation signaling. Although a number of studies have been contributed (25–28), more efforts are needed.

In this study, we constructed the dbPPT database with 82 175 phosphorylation sites in 31 012 proteins, which could be employed to perform systematic studies. Since *Arabidopsis* was the well-studied organism in plants, we analyzed the sequence motifs for phosphorylation. The motifs for phosphoserine, phosphothreonine and phosphotyrosine were calculated with Motif-x (Supplemental Tables S3–S5) (34), while the top four motifs with the highest motif scores were presented in
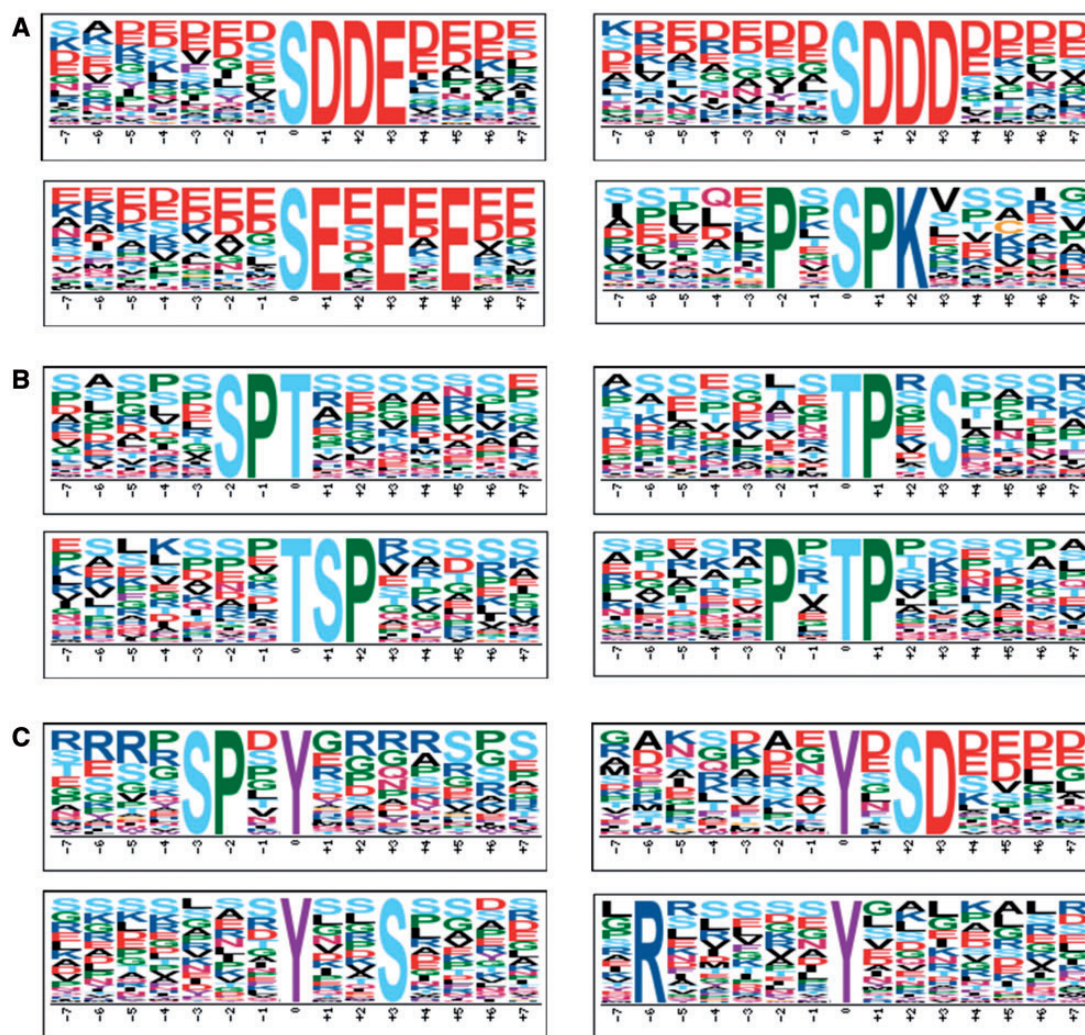
Figure 4. The top four most representative motifs discovered by Motif-x for phosphoserine, phosphothreonine and phosphotyrosine in *Arabidopsis*. (A) Motifs for phosphoserine. (B) Motifs for phosphothreonine. (C) Motifs for phosphotyrosine. The position 0 represents the modified residues.

Figure 4. For phosphoserine, the top four most representative motifs were pS-D-D-E, pS-D-D-D, pS-E-*X*-E-*X*-E and P-*X*-pS-P-K (pS, pT and pY represent the phosphoserine, phosphothreonine and phosphotyrosine, respectively. *X* represents any amino acid) (Figure 4A). The top four motifs with the highest scores for phosphothreonine were S-P-pT, pT-P-*X*-S, pT-S-P and P-*X*-pT-P (Figure 4B). Previously, it was proposed that [ST]-P motif might be recognized by MAPKs and CDKs (22), which indicated that MAPKs and CDKs might be the primary kinases for the identified phosphoserines and phosphothreonines. Acidic motifs such as pS-D-*X*-[DE], pS-E-*X*-E and pS-[DE] were potential substrates of CKII (35, 36), while other motifs such as R-S-*X*-pS, L-*X*-R-*X*-*X*-pS and R-*X*-*X*-pS were associated with CPKs and CaMKs activity (36, 37). The top four most representative motifs for phosphotyrosine were calculated as S-P-*X*-pY, pY-*X*-S-D, pY-*X*-*X*-S and R-*X*-*X*-*X*-*X*-*X*-pY, while the potential kinases for

these motifs needed to be dissected. In general, most of the motifs such as [ST]-P, pS-D-*X*-[DE], R-*X*-*X*-pS were conserved between plants and mammals, which indicated that their corresponding kinases played crucial roles in both mammals and plants.

Besides the results of motif analyses, which might be helpful for prediction of phosphorylation sites in plants, actually, PhosphAt database provided prediction service for *Arabidopsis* (26). To evaluate the coverage of positive hits on experimentally identified phosphorylation sites for the prediction function of PhosphAt, the 94 284 predicted phosphorylation sites from 25 840 substrates in PhosphAt were downloaded for computation. To avoid redundancy in the two datasets, the phosphopeptides with six amino acids upstream and six amino acids downstream around the modified residue were employed for comparison, which resulted in 75 343 and 25 640 non-redundant tridecapeptides for PhosphAt and dbPPT, respectively.

From comparison, it was observed that there were 3966 overlapping phosphopeptides between the two datasets, which indicated 15.47% of experimentally identified phosphorylation sites could be predicted as positive hits.

With the development of phosphoproteome techniques, more efforts will be contributed to this area. The systematic methods to dissect phosphorylation signaling in yeast and animals such as network-based analyses could be employed in plants (38), while the kinase and phosphatase information in plants could also be integrated to reveal the reversible regulation of phosphorylation (39). Furthermore, other PTMs such as acetylation and ubiquitination could also be employed to dissect the PTMs signaling in plants (40, 41). Taken together, here we built the dbPPT database with the manually curated phosphorylation events, which will be a useful resource for both computational and experimental scientists. The dbPPT database will be routinely updated to follow research progresses of protein phosphorylation studies in plants.

Supplementary Data

Supplementary data are available at *Database* Online.

References

1. Pawson,T. and Scott,J.D. (2005) Protein phosphorylation in signaling—50 years and counting. *Trends Biochem. Sci.*, **30**, 286–290.
2. Hunter,T. (2009) Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.*, **21**, 140–146.
3. Adams,J.A. (2001) Kinetic and catalytic mechanisms of protein kinases. *Chem Rev.*, **101**, 2271–2290.
4. Johnson,L.N. and Barford,D. (1993) The effects of phosphorylation on the structure and function of proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **22**, 199–232.
5. Bennett,J. (1991) Protein–phosphorylation in green plant chloroplasts. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **42**, 281–311.
6. Huber,S.C. (2007) Exploring the role of protein phosphorylation in plants: from signalling to metabolism. *Biochem. Soc. Trans.*, **35**, 28–32.
7. Bennett,J., Steinback,K.E. and Arntzen,C.J. (1980) Chloroplast phosphoproteins: regulation of excitation energy transfer by phosphorylation of thylakoid membrane polypeptides. *Proc. Natl Acad. Sci. USA*, **77**, 5253–5257.
8. Bachmann,M., Shiraishi,N., Campbell,W.H. *et al.* (1996) Identification of Ser-543 as the major regulatory phosphorylation site in spinach leaf nitrate reductase. *Plant Cell*, **8**, 505–517.
9. Kusakina,J. and Dodd,A.N. (2012) Phosphorylation in the plant circadian system. *Trends Plant Sci.*, **17**, 575–583.
10. Rampitsch,C. and Bykova,N.V. (2012) The beginnings of crop phosphoproteomics: exploring early warning systems of stress. *Front. Plant Sci.*, **3**, 144.
11. de la Fuente van Bentem,S. and Hirt,H. (2007) Using phosphoproteomics to reveal signalling dynamics in plants. *Trends Plant. Sci.*, **12**, 404–411.
12. Heintz,D., Wurtz,V., High,A.A. *et al.* (2004) An efficient protocol for the identification of protein phosphorylation in a seedless plant, sensitive enough to detect members of signalling cascades. *Electrophoresis*, **25**, 1149–1159.
13. Kersten,B., Agrawal,G.K., Iwahashi,H., *et al.* (2006) Plant phosphoproteomics: a long road ahead. *Proteomics*, **6**, 5517–5528.
14. Sugiyama,N., Nakagami,H., Mochida,K. *et al.* (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. *Mol. Syst. Biol.*, **4**, 193.
15. Kersten,B., Agrawal,G.K., Durek,P. *et al.* (2009) Plant phosphoproteomics: an update. *Proteomics*, **9**, 964–988.
16. Stulemeijer,I.J., Joosten,M.H., and Jensen,O.N. (2009) Quantitative phosphoproteomics of tomato mounting a hypersensitive response reveals a swift suppression of photosynthetic activity and a differential role for hsp90 isoforms. *J. Proteome Res.*, **8**, 1168–1182.
17. Kline-Jonakin,K.G., Barrett-Wilt,G.A., and Sussman,M.R. (2011) Quantitative plant phosphoproteomics. *Curr. Opin. Plant Biol.*, **14**, 507–511.
18. Umezawa,T., Sugiyama,N., Takahashi,F. *et al.* (2013) Genetics and phosphoproteomics reveal a protein phosphorylation network in the abscisic acid signaling pathway in *Arabidopsis thaliana*. *Sci. Signal*, **6**, rs8.
19. Yang,Z., Guo,G., Zhang,M. *et al.* (2013) Stable isotope metabolic labeling-based quantitative phosphoproteomic analysis of Arabidopsis mutants reveals ethylene-regulated time-dependent phosphoproteins and putative substrates of constitutive triple response 1 kinase. *Mol. Cell Proteomics*, **12**, 3559–3582
20. Xue,L., Wang,P., Wang,L. *et al.* (2013) Quantitative measurement of phosphoproteome response to osmotic stress in arabidopsis based on Library-Assisted eXtracted Ion Chromatogram (LAXIC). *Mol. Cell Proteomics*, **12**, 2354–2369.
21. Hornbeck,P.V., Kornhauser,J.M., Tkachev,S. *et al.* (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.*, **40**, D261–D270.

22. Dinkel,H., Chica,C., Via,A. *et al.* (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.*, **39**, D261–D267.

23. Lu,C.T., Huang,K.Y., Su,M.G. *et al.* (2013) DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. *Nucleic Acids Res.*, **41**, D295–D305.

24. Li,J., Jia,J., Li,H. *et al.* (2014) SysPTM 2.0: an updated systematic resource for post-translational modification. *Database*, **2014**, bau025.

25. Gribskov,M., Fana,F., Harper,J. *et al.* (2001) PlantsP: a functional genomics database for plant phosphorylation. *Nucleic Acids Res.*, **29**, 111–113.

26. Zulawski,M., Braginets,R., and Schulze,W.X. (2013) PhosPhAt goes kinases—searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. *Nucleic Acids Res.*, **41**, D1176–D1184.

27. Yao,Q., Ge,H., Wu,S. *et al.* (2014) P(3)DB 3.0: From plant phosphorylation sites to protein networks. *Nucleic Acids Res.*, **42**, D1206–D1213.

28. Rose,C.M., Venkateshwaran,M., Grimsrud,P.A. *et al.* (2012) Medicago PhosphoProtein Database: a repository for *Medicago truncatula* phosphoprotein data. *Front. Plant Sci.*, **3**, 122.

29. UniProt,C. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.

30. Flicek,P., Amode,M.R., Barrell,D. *et al.* (2014) Ensembl 2014. *Nucleic Acids Res.*, **42**, D749–D755.

31. Coordinators,N.R. (2014) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **42**, D7–D17.

32. de la Fuente van Bentem,S. and Hirt,H. (2009) Protein tyrosine phosphorylation in plants: more abundant than expected? *Trends Plant Sci.*, **14**, 71–76.

33. Boratyn,G.M., Camacho,C., Cooper,P.S. *et al.* (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, **41**, W29–W33.

34. Schwartz,D. and Gygi,S.P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.*, **23**, 1391–1398.

35. Meggio,F. and Pinna,L.A. (2003) One-thousand-and-one substrates of protein kinase CK2?. *FASEB J.*, **17**, 349–368.

36. Wang,X., Bian,Y., Cheng,K. *et al.* (2013) A large-scale protein phosphorylation analysis reveals novel phosphorylation motifs and phosphoregulatory networks in Arabidopsis. *J. Proteomics*, **78**, 486–498.

37. White,R.R., Kwon,Y.G., Taing,M. *et al.* (1998) Definition of optimal substrate recognition motifs of Ca2+-calmodulin-dependent protein kinases IV and II reveals shared and distinctive features. *J. Biol. Chem.*, **273**, 3166–3172.

38. Liu,Z., Wang,Y., and Xue,Y. (2013) Phosphoproteomics-based network medicine. *FEBS J.*, **280**, 5696–5704.

39. Wang,Y., Liu,Z., Cheng,H. *et al.* (2014) EKPD: a hierarchical database of eukaryotic protein kinases and protein phosphatases. *Nucleic Acids Res.*, **42**, D496–D502.

40. Gao,T., Liu,Z., Wang,Y. *et al.* (2013) UUCD: a family-based database of ubiquitin and ubiquitin-like conjugation. *Nucleic Acids Res.*, **41**, D445–D451.

41. Liu,Z., Wang,Y., Gao,T. *et al.* (2014) CPLM: a database of protein lysine modifications. *Nucleic Acids Res.*, **42**, D531–D536.