The Plant PTM Viewer, a central resource for exploring plant protein modifications

Patrick Willems[1,2,3,4], Alison Horne[5], Thomas Van Parys[5], Sofie Goormachtig[1,2], Ive De Smet[1,2], Alexander Botzki[5], Frank Van Breusegem[1,2,*] and Kris Gevaert[3,4,*]

[1]*Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium;* [2]*VIB Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium;* [3]*Department of Biomolecular Medicine, Ghent University, 9000 Ghent, Belgium;* [4]*VIB Center for Medical Biotechnology, VIB, 9000 Ghent, Belgium;* [5]*VIB Bioinformatics Core, VIB, 9052 Ghent, Belgium*

[*] For correspondence (e-mail kris.gevaert@vib-ugent.be and frank.vanbreusegem@psb.ugent.be)

Keywords: Protein modification, database, proteomics, mass spectrometry, plant, proteoform, PTM code, ubiquitin, phosphorylation, acetylation

Significance Statement

Present-day proteomics studies in plants outline extensive and diverse posttranslational modification (PTM) protein landscapes that are often unnoticed due to the lack of a plant-specific protein database that allows a user-friendly exploration of the available PTM information. To equip the plant community with a central PTM resource, the Plant PTM Viewer integrates approximately 370,000 protein sites of 19 modification types in five plant species and offers innovative tools to hypothesize their interplay and impact on protein functions.

SUMMARY

Posttranslational modifications (PTMs) of proteins are central in any kind of cellular signaling. Modern mass spectrometry technologies enable comprehensive identification and quantification of various PTMs. Given the increased numbers and types of mapped protein modifications, a database is necessary that simultaneously integrates and compares site-specific information for different PTMs, especially in plants for which the available PTM data are poorly catalogued. Here, we present the Plant PTM Viewer (http://www.psb.ugent.be/PlantPTMViewer), an integrative PTM resource that comprises approximately 370,000 PTM sites for 19 types of protein modifications in plant proteins from five different species. The Plant PTM Viewer provides the user with a protein sequence overview in which the experimentally evidenced PTMs are highlighted together with an estimate of the confidence by which the modified peptides and, if possible, the actual modification sites were identified and with functional protein domains or active site residues. The PTM sequence search tool can query PTM combinations in specific protein sequences, whereas the PTM BLAST tool searches for modified protein sequences to detect conserved PTMs in homologous sequences. Taken together, these tools help to assume the role and potential interplay of PTMs in specific proteins or within a broader systems biology context. The Plant PTM Viewer is an open repository that allows the submission of mass spectrometry-based PTM data to remain at pace with future PTM plant studies.

INTRODUCTION

Protein modifications (PTMs) are largely responsible for a part of all chemically different forms of proteins or so-called proteoforms (Smith *et al.,* 2013). Various proteoforms may be engaged in different protein complexes (Zhang *et al.,* 2016) or have altered stabilities (Nelson and Millar, 2015), structural conformations (Cho *et al.,* 2011) and activities (Finkemeier *et al.*, 2011, Huang *et al.,* 2018a). PTMs can also occur in a coordinated manner, resulting in a complex and interconnected landscape of protein regulation (Hunter, 2007, Minguez *et al.,* 2012). Although over 400 PTMs have been reported, only a subset has been comprehensively characterized on a proteome-wide level by means of mass spectrometry (MS) due to the lack of tools needed for their enrichment. On the contrary, thanks to technological advances in high-precision MS, up to thousands of PTM sites can now be identified in a single proteomics experiment (Choudhary and Mann 2010).

Thus far, protein phosphorylation is the best-characterized PTM in plants. For example, for *Arabidopsis thaliana* (Arabidopsis), approximately 57,000 and 48,000 phosphorylation sites are available in the PhosPhAt 4.0 (Heazlewood *et al.,* 2008) and the Plant Protein Phosphorylation database (Yao *et al.*, 2014), respectively. Another well characterized plant PTM is lysine acetylation, which, besides other lysine modifications, is available for Arabidopsis and other species in the Protein Lysine Modification (PLM) database (Xu *et al.*, 2017). Several studies highlighted the functional importance of lysine acetylation in Arabidopsis (Finkemeier *et al.*, 2011, Wu *et al.*, 2011, Hartl *et al.*, 2017), but also, for example, in rice (*Oryza sativa*) seed development (Wang *et al.*, 2017). Whereas lysine acetylation is reversible, acetylation of the α-amino group at protein N-termini is not and happens either cotranslationally or posttranslationally after protein cleavage events (Bienvenut *et al.*, 2012, Drazic *et al.*, 2016). For instance, in plants such a posttranslational N-terminal acetylation often occurs on plastid protein N-termini exposed after transit peptide cleavage (Bienvenut *et al.*, 2012, Köhler *et al.,* 2015, Rowland *et al.,* 2015). Besides phosphorylation and acetylation, PTMs reported of late include N-terminal myristoylation (Majeran *et al.,* 2018), *O*-linked N-acetylglucosaminylation (*O*-GlcNAcylation) (Xu *et al.,* 2017) in Arabidopsis, lysine modifications, such as 2-hydroxyisobuturylation (Meng *et al.,* 2017) and malonylation (Mujahid *et al.,* 2018) in rice (*O. sativa* spp. *japonica*), and SUMOylation in Arabidopsis (Rytz *et al.*, 2018). In addition to the identification of PTMs, differences in overall PTM levels have been assessed, for instance under stress conditions (Jacques *et al.,* 2015, Vu *et al.,* 2018) or in developmental processes (Wang *et al.,* 2017). However, no integrative plant databases exist that contain such recently characterized modifications or their dynamics. Other plant protein modifications that have been detected by MS, but rarely catalogued, comprise cysteine oxidation and proteolytic processing. Oxidized cysteine sites have been discovered in Arabidopsis and *Chlamydomonas reinhardtii*, after differential labeling of oxidized and reduced thiols (Liu *et al.,* 2014). Proteolytically processed sites have been found in Arabidopsis by enrichment of newly exposed protein N-termini via positional proteomics (Kleifeld *et al.,* 2010, Venne *et al.,* 2015), identifying substrates of the protease METACASPASE9 (Tsiatsiani *et al.,* 2013), of the chloroplast protein import-associated processing (Köhler *et al.,* 2015, Rowland *et al.,* 2015) and N-end rule degradation pathway(Zhang *et al.,* 2015*,* 2018).

Interestingly, the increasing numbers of different PTMs open up exciting possibilities to discover PTM interplays. Here, we refer to a PTM interplay as an initial PTM that can act as a signal and influence the occurrence of another PTM (Hunter, 2007). Given the high quantity of identified PTMs, (species-specific) databases are needed that compile PTM data for further exploration. Examples of such databases are the PLM database (Xu *et al.*, 2017), dbPTM (Lee *et al.,* 2006), PTMcode (Minguez *et al.,* 2015) and iPTMnet (Huang *et al.,* 2018b). As such, databases often integrate various PTM types, facilitating the exploration of PTM interplays and leading to novel hypotheses and new experiments. For instance, incentivized by the overlap of

protease cleavage and phosphorylation sites, human caspases were shown to cleave after phosphorylated serines (Seaman *et al.,* 2016). In plants, PTM interplays have been described as well. Based on observations on plant nitrite reductase and pyruvate dehydrogenase, nearby phosphorylation is inhibited by methionine oxidation that can act as a phosphorylation-regulating redox switch (Hardin *et al.,* 2009, Miernyk *et al.,* 2009). In addition, one or several phosphorylated residues have been shown to induce ubiquitination and subsequent protein degradation in a *cis*-regulatory manner (Hardtke *et al.,* 2000).

Given the vast amount of characterized PTMs on plant proteins, the challenge is now to typify their functions. Recently, a community effort has been proposed that would collect and assemble all PTM data for plant proteomes with the aim to determine systematically the possible functional implications for plant proteins and the biological processes in which they are involved (Friso and van Wijk, 2015). Although some PTM databases (Minguez *et al.,* 2015, Huang *et al.,* 2016*,* 2018b) contain plant protein modifications, PTM data are still largely unclassified and hardly explored in an integrative manner. To address this issue, we developed the Plant PTM Viewer (http://www.psb.ugent.be/PlantPTMViewer). In addition to the simple 'viewing' of PTMs that provides a probability measure of the correctness of the identification of the modified peptide, we present several bioinformatics tools that help discover the hypothetical roles of plant PTMs in a protein-centric or systematic manner. Moreover, a visualization of the quantitative differences in PTM levels between conditions has been attempted. Hence, the Plant PTM Viewer offers the necessary framework to decode further plant PTM interplays to steer future plant protein studies.

RESULTS

A central resource for the exploration of PTMs of plant proteins

Plant PTM data were collected from 105 publications, often from supplementary tables. Modified peptide sequences gave rise to 371,742 PTM sites in 98,227 proteins (including protein isoforms resulting from mRNA splicing) for five plant species, outcompeting the amount stored in other PTM databases (Figure 1; Table S1). Predominant is the number of phosphorylation sites. Although more phosphopeptides are identified in Arabidopsis than in maize (*Zea mays*), i.e., 19,409 versus 14,627 unique phosphopeptide sequences, a high number of maize protein phosphosites are present due to the increased ambiguity in peptide-to-protein matching (143,208 sites; Figure 1). For instance, the 14,627 phosphopeptide sequences in maize match on average 11 proteins (isoform-level), whereas the 19,409 Arabidopsis phosphopeptides match 2.25 proteins on average. Noteworthy, some basic import criteria were considered to ensure a minimal level of PTM reliability (see Experimental Procedures). Of the 19 PTM types reported in plants (Table S2), several had not been compiled for plants in any other PTM database. Additional information and statistics per PTM type are available via the 'PTM Info' tab on the website. Besides the option to browse for PTMs per protein ('Protein Search' tab), one of the aims of the Plant PTM Viewer is to find the hypothetical role of a given protein/proteoform. To this end, we introduced several tools that enable the functional dissemination of PTMs, including the possibility to search for user-defined PTM motifs ('PTM Search' tab) and to retrieve conserved PTMs in homologous proteins or protein regions ('PTM BLAST' tab). For each of the tools, a tutorial is available on the database website (http://www.psb.ugent.be/PlantPTMViewer/).

Examination of the complex PTM landscape for a protein of interest

With the 'Protein Search' tab, a protein of interest can be searched, either via a protein identifier, a protein description, or a protein-matching amino acid sequence. The Plant PTM Viewer returns a protein-centric overview of all available PTMs in a tabular format, *i.e.* the 'PTM table', and a protein sequence overview in which all PTM sites are highlighted.

As a test case, we present data for ubiquitination, a highly conserved protein-based modification in which the small protein ubiquitin is covalently attached to substrate proteins often targeting them for degradation. Intriguingly, ubiquitin itself can be subjected to PTMs. For instance, in human ubiquitin, besides the polyubiquitin chain formation, substrate-bound ubiquitin can be acetylated, phosphorylated and modified by other ubiquitin-like molecules, referred to as the ubiquitin code (Swatek and Komander, 2016). To view the PTMs on the plant ubiquitin and to explore possible parallels to the human ubiquitin code, we examined the highly conserved 76-amino-acid ubiquitin protein sequence in the 'Protein Search' tab. Several ubiquitin-encoding plant genes were found from which the protein sequence overview of the 76-amino-acid-long ubiquitin protein domain was extracted (Figure 2A). In the provided ubiquitin protein sequences, the different experimentally evidenced ubiquitin modifications are shown, in which lysine ubiquitination is a prominent modification. For the Arabidopsis ubiquitin, six out of seven lysines were ubiquitinated (Maor *et al.*, 2007, Sarraco *et al.*, 2009, Walton *et al.*, 2016). In addition, the modification sites are detailed in the PTM table (Figure 2A, bottom), in which every modified amino acid site and the respectively identified peptide sequence(s) can be consulted per individual study/experiment. For example, in Arabidopsis, Lys48 ubiquitination was reported in three independent studies (Maor *et al.*, 2007, Sarraco *et al.*, 2009, Walton *et al.*, 2016), giving an idea on the general prevalence of this modification. Besides polyubiquitination, various PTMs were identified on plant ubiquitins. Lys6 and Lys48 were found to be acetylated in Arabidopsis ubiquitin (Hartl *et al.*, 2017, Liu *et al.*, 2018). Also in ubiquitin of wheat (*Triticum* sp) and rice, lysines were targeted by acetylation (Wang *et al.,* 2017) and by PTMs metabolically related to acetylation, such as 2-hydroxyisobutyrylation (Meng *et al.*, 2017), succinylation (He *et al.*, 2016) and malonylation (Mujahid *et al.*, 2018). In rice, all ubiquitin lysines, except Lys29, are modified by acetylation, malonylation, butyrylation and/or succinylation (Figure 2A). Notably, the same six lysines had been found to be acetylated in human (Swatek and Komander, 2016). Moreover, acetylation at Lys6 or Lys48 inhibited polyubiquitination in human, thereby shaping the ubiquitin branches (Ohtake *et al.,* 2015). Thus, we hypothesize that ubiquitin lysine acetylation, or metabolically related PTMs, might also influence the branching of polyubiquitin chains in plant proteins.

Besides information on peptides and PTMs, the 'PTM table' provides additional peptide-level metadata that are reported in MS studies (Figure 2A, bottom). For approximately 80% of the collected PTM data either a peptide score, a posterior error probability (PEP), a modification site localization probability and/or a precursor mass error deviation could be gathered. Based on the extracted peptide score, we labeled PTMs with a low, medium or high confidence (see Experimental Procedures or online tutorial). The attributed confidence can be intuitively viewed with a 'traffic-light' color scheme in the 'PTM table' (Figure 2A) that can be extended to provide the confidence metadata, thus allowing users to intuitively assess the reliability of PTMs, a particularly interesting feature when downstream experiments are triggered by certain PTMs. As an example, the phosphorylation at Ser65 had a MASCOT ion score of 27 and a modification site probability of 99.7% (Roitinger *et al.,* 2015), but given this relatively low MASCOT ion score, a low confidence is attributed to this modification. In contrast, Lys48 ubiquination identified by two peptides and three studies is considered with high confidence (Figure 2A). In favor of Ser65 phosphorylation is its identification in rice, albeit also with a low confidence (MASCOT ion score of 32). In human cells, this serine was shown to be phosphorylated by PINK1, leading to Parkin activation (Gladkova *et al.,* 2018). As Ser65 phosphorylation generates a structurally and functionally altered ubiquitin that can act as an independent signal (Wauer *et al.,* 2015), its occurrence is intriguing in the plant ubiquitin protein. However, careful inspection of the provided confidence metadata warrants a cautious interpretation of this event. Hence, confidence assessment of PTMs is crucial, because false positives can arise in MS results. Therefore, circumspect consideration is advised of the

experimental details available in the Plant PTM Viewer, such as the false discovery rate (FDR) and search settings used. For further validation of PTMs, additional experiments are recommended before starting large follow-up studies that very often focus on one single modification event in a single protein. For instance, for quite some PTMs, peptides can be synthesized and analyzed by the same LC-MS/MS setup as used for the discovery experiment. When the MS/MS spectral profiles of the synthetic peptide and the natural peptide match, an extra level of certainty can be added to the identified, modified peptide.

The Plant PTM Viewer can store differential peptide abundancies between conditions, when the basic statistics are available (*e.g.*, log2-fold changes and *P* or *Q* values). In such cases, a comparison is considered a different experiment, enabling visualization of PTM variabilities in a heat map-like display (Figure 2A) and swift evaluation of PTM intensities between conditions. For instance, differential lysine acetylation was assessed during early seed development in rice (Wang *et al.,* 2017). Here, Lys6 ubiquitin acetylation significantly increased after 3 and 7 days of seed development (Figure 2A, experiment ID 14a and 14b, respectively). Interestingly, several ubiquitin-modifying enzymes have been implicated in seed size regulation of Arabidopsis, rice and wheat (Li and Li, 2014). Hence, differential ubiquitin acetylation might contribute to the seed development regulation. However, PTM abundancy should preferably be normalized to protein abundancy for a correct interpretation. Although this normalization is applied in some studies (Vu *et al.*, 2016), it is not standard practice and it is thus advisable to consult the description of the differential analysis in the respective publications. In summary, by combining PTM data from diverse plant proteomics studies, a putative ubiquitin code in plants could easily be sketched, analogously to the human ubiquitin code. Nevertheless, the potential of the Plant PTM Viewer goes beyond ubiquitin, because the modification landscape of any plant protein of interest can be rapidly explored.

Another tool that facilitates hypothesis generation is the visualization of PTMs in a protein sequence view alongside their InterPro protein domains (Finn *et al.,* 2017) and sites that are annotated in the UniProt Knowledgebase (UniProtKB) (UniProt Consortium, 2018) as binding or catalytic sites. As an example, the Arabidopsis PEROXISOMAL 3-KETOACYL-COA THIOLASE3 (KAT2, AT2G33150) is given. KAT2 is one of three KAT enzymes in Arabidopsis that catalyzes peroxisomal β-oxidation and is strongly produced in plants (Wiszniewski *et al.*, 2012). KAT2 contains a N-terminal peroxisomal targeting sequence of 34 residues and its posttranslational cleavage has experimentally been observed becauses Leu35 has been identified as a N-terminally exposed residue in three independent experiments (Venne *et al.,* 2015, Zhang *et al.,* 2015, 2018). In the N-terminal thiolase protein domain (InterPro: IPR020616), a multitude of other cleavage sites is evident, but also lysine acetylation, ubiquitination and S-nitrosylation of Cys138 (Figure 2B). Interestingly, this Cys is crucial for the enzymatic activity, because it forms the acyl-thioester intermediate. Thus, the observed S-nitrosylation reported by two independent experiments (Fares *et al.*, 2011, Hu *et al.,* 2015) hints at a putative redox control of KAT2. In fact, Cys138 of KAT2 forms a disulfide with Cys192 and is inactivated when oxidized (Pye *et al.*, 2010). In the Plant PTM Viewer, peptides are flagged that ambiguously match proteins encoded by different genes. Such an example is the peptide 'QCSSGLQAVADVAAAIK' with the reported S-nitrosylated Cys138, because it corresponds to the active sites of both KAT2 and KAT1 (AT1G04710) and it is a duplicated gene with reduced expression (Wiszniewski *et al.*, 2012). Taken together, the Plant PTM Viewer steers the functional interpretation of PTMs in relation to annotated regulatory regions of a protein, thereby paving the road from PTM 'site-viewing' to protein function.

Assessment of the PTM interplay with the 'PTM search' tool

The 'PTM search' tool allows the search for ambiguous or specific amino acid motifs that are modified by one or more PTMs (for details, see help section or tutorial). This ambiguity search

can be fine-tuned to address specific or more system-wide research questions. Searches can be restricted to the retrieval of PTMs residing in known enzyme recognition sites (such as kinases or proteases). For instance, the substrate specificity was profiled for the Arabidopsis MITOGEN-ACTIVATED KINASE 3 (MPK3) and MPK6 by means of a synthetic peptide library, delivering a Leu/Pro-X-pSer-Pro-Arg/Lys consensus motif for MPK6 (Sörensson *et al.,* 2012). Examination of this consensus motif in Arabidopsis proteins by '[LP]XS(ph)P[RK]' with the 'PTM search' tool returned 177 phosphosites uniquely matching 159 (representative) proteins (Figure 3, left panel), including the AT1G80180.1 protein that had been experimentally verified to be a MPK3 and a MPK6 substrate (Sörensson *et al.,* 2012). In addition, combinations of PTMs can be looked for to discover potential PTM interplays. Here, the possible interplay between phosphorylation and ubiquitination was considered, in which phosphorylated residues could inhibit or promote ubiquitination and act as so-called phosphodegrons (Filipčik *et al.,* 2017). To this end, we looked for 'X(Ph)X{0,4}K(Ub)', thus seeking phosphorylation of any residue ('X(Ph)'), followed by lysine ubiquitination ('K(Ub)'), spaced by none or maximal four amino acids ('X{0,4}'). We also queried the reversed scenario, namely 'K(Ub)X{0,4}X(Ph)', in which a ubiquitinated residue is N-terminal to a phosphorylated one. Together, these two queries returned 182 hits in 123 proteins (Figure 3, middle panel). Thus, such protein regions might hold putative phosphodegrons that promote or inhibit subsequent ubiquitination. As such, the 'PTM search' tool can propose specific candidates and sites that might be of interest for future experimental validation. Moreover, the search results can optionally be restricted to match specified InterPro domains or UniProtKB-annotated protein sites.

The co-occurrence of different PTMs on the same amino acid is the most obvious form of PTM interplay, hinting at competition between modifying enzymes for a given amino acid. Such a co-occurrence of PTMs can be explored with the term 'X(xx&xx)', *i.e*. any amino acid ('X') targeted by at least two different modifications ('xx&xx'). As an example, we focused again on phosphorylation. Phosphosites overlapped with 29 *O*-GlcNAc sites in 28 proteins and, surprisingly, also with 144 N-terminal residues exposed after proteolytic processing in 119 proteins (Figure 3, right panel). Further inspection of these phosphorylated sites revealed that 47 of these phosphosites resided at Thr2 or Ser2, thus, at protein N-termini undergoing N-terminal methionine excision. Such a phosphorylation at protein N-terminal residues had been reported to occur in photosynthetic membranes (Vener *et al.,* 2001). In some cases, the N-terminal modification status was even more complex. Ser2 of the NUCLEOSOME ASSEMBLY PROTEIN1 (NAP1, AT2G19480.1) could be uncovered by N-terminal methionine excision and be N-terminally acetylated (Linster *et al.*, 2015, Zhang *et al.,* 2015, 2018), phosphorylated (Roitinger *et al.,* 2015) and even N-terminally ubiquitinated (Walton *et al.,* 2016). Although this N-terminal ubiquitination was reported for only 10 sites, novel techniques, such as UbiSite, are bound to detect additional sites in the future (Akimov *et al.*, 2018). Hence, the N-terminal modification landscape of some proteins might be more complex and dynamic than anticipated.

Discovery of conserved PTMs in proteins within or across species with 'PTM BLAST'

Whereas the 'PTM sequence search' tool returns specific PTM patterns in proteins, the 'PTM BLAST' tool can identify similar protein regions with conserved PTMs within or across species. 'PTM BLAST' can use a (modified) protein sequence as query for a default protein BLAST to all the proteins in the Plant PTM Viewer (Figure 4A). In a next step, the obtained and aligned protein sequences are cross-checked for alignment of differing ('align') or identical ('match') PTM types. The top results are sorted by the numbers of PTM alignments (Figure 4A), allowing the swift identification of conserved PTMs within or across plant species. BLAST alignments can be visualized and their respective PTMs are indicated. Ambiguous peptide sequences matching both aligned proteins can be omitted. Such peptides can appear in

protein alignments within species due to ambiguous peptide-to-protein assignment. 'PTM BLAST' can be launched from each PTM Viewer protein overview, querying the entire protein sequence or a region thereof, thus restricting the BLAST search to a region of interest, such as a protein domain.

To evaluate the evolutionary conservation of Arabidopsis PTMs, we ran the 'PTM BLAST' for all canonical Arabidopsis proteins, keeping track of the number of Arabidopsis PTMs that aligned to other types of PTMs per species. For instance, for phosphorylation that is the most abundant PTM type represented in all species (Figure 1), 1,619 phosphosites in Arabidopsis aligned exactly to phosphosites of any other plant species (Figure 4B). An example is OPEN STOMATA 1 (OST1, AT4G33950), in which both Ser171 and Ser175 are conserved phosphosites that reside in the OST1 activation loop. Noteworthy, in Arabidopsis, both serines were phosphorylated after hyperosmotic or abscisic acid treatment and Ser175 was crucial for the OST1 phosphorylation activity (Belin *et al.,* 2006). As an illustration of a 'PTM BLAST' alignment, we present a protein region alignment between the Arabidopsis OST1 and its maize ortholog SRK2E (UniProt accession B4FQ40) that, besides Ser171 and Ser175, has four supplementary conserved phosphosites (Figure 4C).

Besides querying a Plant PTM Viewer protein, any custom protein sequence, with or without specified PTMs included in the Plant PTM Viewer (Table S2), can be searched via the 'PTM BLAST'. As a case, we selected the human glyceraldehyde-3-phosphate dehydrogenase (GAPDH), because it is an essential housekeeping gene for which a rich PTM diversity is evident, as substantiated by the 25 sites of five PTM types reported in UniProtKB and included in the PTM Viewer (UniProtKB P04406) (Figure S1A). Therefore, the complex PTM landscape of the human GAPDH was used as input for the 'PTM BLAST', resulting in exact PTM alignments for the GAPDH orthologs of all the five plant species. For GAPDH C2 (GAPC2) of Arabidopsis, six conserved PTMs were retrieved, including four phosphorylation sites, a Lys acetylation site and a S-nitrosylation site (Figure S1B). Thus, 'PTM BLAST' can return aligned PTMs for a given, non-plant input sequence as illustrated here for the human GAPDH.

DISCUSSION AND PERSPECTIVES

The Plant PTM Viewer enables researchers to consult the PTM landscape for plant proteins in a user-friendly manner. The current collection of plant PTM data contains approximately 370,000 sites of 19 PTM types of five plant species. The Plant PTM Viewer offers various tools that facilitate the functional interpretation and hypothesis formulation of PTMs. Of note, it also reports on the confidence level at which a given modified peptide and, if possible, the actual modified site had been correctly identified, thereby providing the user some sort of additional priority ranking scheme. In contrast to other PTM databases, the 'PTM search' tool allows the search for combinations of PTMs or of single PTMs in user-defined amino acid motifs. Thanks to an easy fine-tuning, the biological research question at hand can be addressed and putative protein PTM candidates can be hypothesized. In addition, the 'PTM BLAST' tool can reveal aligned PTMs across or within species, thereby enhancing the confidence and relevance of these PTMs. In the future, these tools could be extended with, for instance, assessment and visualization of overrepresented gene sets and sequence motifs in the case of a PTM search, or multiple sequence alignments of homologous protein regions delivered by 'PTM BLAST'. Like the database information, all the results of the Plant PTM Viewer can straightforwardly be exported, facilitating the plant PTM research by directing both future *in silico* systematic analyses and wet-lab experiments.

The Plant PTM Viewer strives to be and remain a central resource for plant researchers. To this end, we strongly encourage the plant science community to actively contribute to the Plant PTM Viewer by submitting peer-reviewed PTM data. The solid Plant PTM Viewer framework for protein visualization and tools can easily be expanded to deal with additional

PTM sites, including those from newly characterized PTM types and additional plant species. Despite the phosphorylation dominance in the PTM research, we expect extra PTM types to be better characterized in the future. However, due to the increasing wealth of reported PTMs in proteins, an important future challenge will be to determine whether PTMs are functionally important for protein functions or not. In this respect, acquisition of quantitative information will become increasingly vital to determine abundant PTMs or those that differ between conditions. Although differential PTM data are still in their infancy, we feel that an initial step has been taken toward a differential PTM browser, similar to GENEVESTIGATOR for differential gene expression (Hruz *et al.,* 2008).

EXPERIMENTAL PROCEDURES

PTM data and metadata collection and processing

From 105 peer-reviewed MS studies on plant proteomes, PTMs with their respective peptide sequences were collected from supplemental files available from publications deposited in the PRIDE repository (Vizcaíno *et al.,* 2016). In addition, Arabidopsis phosphopeptide data were extracted from the PhosPhAt 4.0 database (Heazlewood *et al.,* 2008). Peptide metadata, such as peptide score, peptide precursor mass deviation, modification localization probability and posterior error probability (PEP), were assembled when available. PEP can be considered as a "local FDR" for single spectrum hits and is more stringent than the FDR used to filter sets of spectral matches. Most PEP values reported here were obtained with a software, such as MaxQuant (Cox and Mann 2008), Proteome Discoverer (Thermo Scientific) or with post-processing algorithms, such as Percolator (Käll *et al.*, 2007). Besides PEP, peptide scores, such as the MASCOT ion score (Perkins *et al.*, 1999), Andromeda score (Cox et al., 2011) and the SEQUEST cross correlation (XCorr) score (Eng *et al.*, 1994), were stored. Certain basic import requirements were used. For these three peptide scores, separate minimal score thresholds were used: the ion scores had to be at least 20 for MASCOT, 40 for Andromeda and 2 for XCorr (for score distributions, see Figure 5A). The peptide score was used to categorize PTMs with a low, medium or high confidence in an experiment (Figure 5B). When multiple spectra for a PTM were identified in one single experiment, confidence metadata were extracted from the best scoring spectrum. For the remaining 41,136 PTM entries (19.8%) without peptide score or derived from other search engines, no confidence was assigned (Figure 5). Nevertheless, these confidence assignments serve uniquely as an intuitive first impression of the PTM reliability.

In addition to peptide-level confidence measurements, the probability in modification site localization within a peptide was measured by algorithms, such as PhosphoRS (Taus *et al.,* 2011) or the PTM Score implemented in MaxQuant (Olsen *et al.,* 2006). Here, a site probability of at least 0.75 was mandatory as import requirement. By application of these filtering criteria when possible, 127,914 PTM sites within non-redundant peptides were assembled that provided PTMs from five species: Arabidopsis, maize, common wheat, rice and *Chlamydomonas reinhardtii* (Figure 1).

PTM-carrying peptide sequences were reallocated to all proteins (*i.e*., including splice forms) from the latest reference proteomes of these five species (Table S3), resulting in approximately 370,000 PTM sites in proteins. Ambiguous protein interference frequently arises in plants due to extensive duplication events (Vanneste *et al.*, 2014). Hence, in the Plant PTM Viewer, approximately 20% of the PTM sites matched proteins encoded by different genes. To address this issue, ambiguous peptides that corresponded with multiple proteins were flagged in the PTM Viewer. When differential abundance statistics between conditions were reported, the log2-fold change and significance values (*e.g., P* value and FDR) were stored as well, a current occurrence for approximately 4% of the non-redundant list of PTM peptides. To offer a user-friendly indication of the PTM confidence, we designed a color scheme labeling PTMs with high, medium or low confidence (green to grey, Figure 1).

The PTM data were accompanied with various protein and experimental metadata. Protein identifiers were cross-referenced with UniProtKB (UniProt Consortium, 2018) and the plant comparative genomics platform PLAZA (Van Bel *et al.,* 2018) for identical protein sequences. For the original MS studies, the respective PubMed identifier and the PRIDE (Vizcaíno *et al.,* 2016) accession data of the raw MS data, if deposited, were incorporated. In addition, relevant experimental information, such as plant tissue and genotype, stress condition and PTM enrichment methodology were retained as well. As the PTM reliability is highly dependent on the used MS instrumentation and subsequent *in silico* analysis, we appended experimental details, such as MS instrument, search algorithm, protein database, FDR strategy and threshold, precursor mass tolerance and specified (variable/fixed) modifications.

System configuration

The Plant PTM Viewer is accessible via http://www.psb.ugent.be/PlantPTMViewer. A web interface for browsing, searching and displaying of data was implemented in PHP. To enable the search for PTMs, sequence-based queries were translated to specific regular expressions by means of various wildcards and operators. For instance, 'X' can be used to denote any amino acid and '[ST]' can specify either serine or threonine. Hence, complex patterns can be retrieved, whereby amino acids and PTM types can be specified in a flexible manner. Screen resolution is auto-adaptable for user-friendly browsing from wide screens to smart phones. The Plant PTM Viewer can be accessed programmatically. The Plant PTM Viewer data are maintained in a MySQL relational database, of which an overview is given in Figure S2. For enhancement of the functional interpretation, the Plant PTM Viewer is connected to the InterPro domain database (Finn *et al.,* 2017), as well as to UniProtKB (UniProt Consortium, 2018).

ACKNOWLEDGMENTS

The authors thank Lam Dai Vu, Tingting Zhu, Natalia Nikonorova, Silke Jacques, Alan Walton, Lukas Braem, and the PTM-characterizing plant research community for providing the necessary input, Michiel Van Bel and Frederik De Laere for IT support, Simon Stael and Kai Xun Chan for critical reading and comments, and Martine De Cock for help with the manuscript preparation. This work was supported by grants from the Research Foundation-Flanders (grant number G.0038.09N [to K.G. and F.V.B.] and Excellence of Science project no. 30829584 [to F.V.B.]), the Ghent University Multidisciplinary Research Partnership ["Ghent BioEconomy" (Project 01MRB 510W) to F.V.B.] and the Ghent University Special Research Fund (BOF 01J11311 [to F.V.B.]).

REFERENCES

Akimov, V., Barrio-Hernandez, I., Hansen, S.V.F., Hallenborg, P., Pedersen, A.-K., Bekker-Jensen, D.B., Puglia, M., Christensen, S.D.K., Vanselow, J.T., Nielsen, M.M., Kratchmarova, I., Kelstrup, C.D., Olsen, J.V. and Blagoev, B. (2018) UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nat. Struct. Mol. Biol.* **25**, 631-640.

Belin, C., de Franco, P.-O., Bourbousse, C., Chaignepain, S., Schmitter, J.-M., Vavasseur, A., Giraudat, J., Barbier-Brygoo, H. and Thomine, S. (2006) Identification of features regulating OST1 kinase activity and OST1 function in guard cells. *Plant Physiol*. **141**, 1316-1327.

Bienvenut, W.V., Sumpton, D., Martinez, A., Lilla, S., Espagne, C., Meinnel, T. and Giglione, C. (2012) Comparative large scale characterization of plant *versus* mammal proteins reveals similar and idiosyncratic *N*-α-acetylation features. *Mol. Cell. Proteomics* **11**, M111 015131.

Cho, E.J., Yuen, C.Y.L., Kang, B.-H., Ondzighi, C.A., Staehelin, L.A. and Christopher, D.A. (2011) Protein disulfide isomerase-2 of *Arabidopsis* mediates protein folding and localizes to both the secretory pathway and nucleus, where it interacts with maternal effect embryo arrest factor. *Mol. Cells* **32**, 459-475.

Choudhary, C. and Mann, M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol*. **11**, 427-439.

Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol*. **26**, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794-1805.

Drazic, A., Myklebust, L.M., Ree, R. and Arnesen, T. (2016) The world of protein acetylation. *Biochim. Biophys. Acta* **1864**, 1372-1401.

Eng, J.K., McCormack, A.L. and Yates J.R. III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989.

Fares, A., Rossignol, M. and Peltier, J.-B. (2011) Proteomics investigation of endogenous *S*-nitrosylation in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* **416**, 331-336.

Filipčik, P., Curry, J.R. and Mace, P.D. (2017) When worlds collide---mechanisms at the interface between phosphorylation and ubiquitination. *J. Mol. Biol*. 429, 1097-1113.

Finkemeier, I., Laxa, M., Miguet, L., Howden, A.J.M. and Sweetlove, L.J. (2011) Proteins of diverse function and subcellular location are lysine acetylated in Arabidopsis. *Plant Physiol.* **155**, 1779-1790.

Finn, R.D., Attwood, T.K., Babbitt, P.C. *et al.* (2017) InterPro in 2017---beyond protein family and domain annotations. *Nucleic Acids Res*. **45**, D190-D199.

Friso, G. and van Wijk, K.J. (2015) Posttranslational protein modifications in plant metabolism. *Plant Physiol*. **169**, 1469-1487.

Gladkova, C., Maslen, S.L., Skehel, J.M. and Komander, D. (2018) Mechanism of parkin activation by PINK1. *Nature* **559**, 410-414.

Hardin, S.C., Larue, C.T., Oh, M.-H., Jain, V. and Huber, S.C. (2009) Coupling oxidative signals to protein phosphorylation via methionine oxidation in *Arabidopsis*. *Biochem. J*. **422**, 305-312.

Hardtke, C.S., Gohda, K., Osterlund, M.T., Oyama, T., Okada, K. and Deng, X.W. (2000) HY5 stability and activity in *Arabidopsis* is regulated by phosphorylation in its COP1 binding domain. *EMBO J*. **19**, 4997-5006.

Hartl, M., Füßl, M., Boersema, P.J., Jost, J.-O., Kramer, K., Bakirbas, A., Sindlinger, J., Plöchinger, M., Leister, D., Uhrig, G., Moorhead, G.B.G., Cox, J., Salvucci, M.E., Schwarzer, D., Mann, M. and Finkemeier, I. (2017) Lysine acetylome profiling uncovers novel histone deacetylase substrate proteins in *Arabidopsis*. *Mol. Syst. Biol.* **13**, 949.

He, D., Wang, Q., Li, M., Damaris, R.N., Yi, X., Cheng, Z. and Yang, P. (2016) Global proteome analyses of lysine acetylation and succinylation reveal the widespread involvement of both modification in metabolism in the embryo of germinating rice seed. *J. Proteome Res*. **15**, 879-890.

Heazlewood, J.L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D. and Schulze, W.X. (2008) *PhosPhAt*: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res*. **36**, D1015-D1021.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P. (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinform.* **2008**, 420747.

Hu, J., Huang, X., Chen, L., Sun, X., Lu, C., Zhang, L., Wang, Y. and Zuo, J. (2015) Site-specific nitrosoproteomic identification of endogenously *S*-nitrosylated proteins in Arabidopsis. *Plant Physiol.* **167**, 1731-1746.

Huang, H., Arighi, C.N., Ross, K.E., Ren, J., Li, G., Chen, S.-C., Wang, Q., Cowart, J., Vijay-Shanker, K. and Wu, C.H. (2018b) iPTMnet: an integrated resource for protein post- translational modification network discovery. *Nucleic Acids Res.* **46**, D542-D550.

Huang, J., Willems, P., Van Breusegem, F. and Messens, J. (2018a) Pathways crossing mammalian and plant sulfenomic landscapes. *Free Radic. Biol. Med.*, in press (doi: 10.1016/j.freeradbiomed.2018.02.012).

Huang, K.-Y., Su, M.-G., Kao, H.-J., Hsieh, Y.-C., Jhong, J.-H., Cheng, K.-H., Huang, H.-D. and Lee, T.-Y. (2016) dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. *Nucleic Acids Res.* **44**, D435-D446.

Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol. Cell* **28**, 730-738.

Jacques, S., Ghesquière, B., De Bock, P.-J., Demol, H., Wahni, K., Willems, P., Messens, J., Van Breusegem, F. and Gevaert, K. (2015) Protein methionine sulfoxide dynamics in *Arabidopsis thaliana* under oxidative stress. *Mol. Cell. Proteomics* **14**, 1217-1229.

Käll, L., Canterbury, J.D., Weston, J., Noble, W.S. and MacCoss, M.J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**, 923-925.

Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R.K., Starr, A.E., Foster, L.J., Kizhakkedathu, J.N. and Overall, C.M. (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28**, 281-288.

Köhler, D., Montandon, C., Hause, G., Majovsky, P., Kessler, F., Baginsky, S. and Agne, B. (2015) Characterization of chloroplast protein import without Tic56, a component of the 1-megadalton translocon at the inner envelope membrane of chloroplasts. *Plant Physiol.* **167**, 972-990.

Li, N. and Li, Y. (2014) Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* **5**, 332.

Linster, E., Stephan, I., Bienvenut, W.V., Maple-Grødem, J., Myklebust, L.M., Huber, M., Reichelt, M., Sticht, C., Møller, S.G., Meinnel, T., Arnesen, T., Giglione, C., Hell, R. and Wirtz, M. (2015) Downregulation of N-terminal acetylation triggers ABA-mediated drought responses in *Arabidopsis*. *Nat. Commun.* **6**, 7640.

Liu, P., Zhang, H., Wang, H. and Xia, Y. (2014) Identification of redox-sensitive cysteines in the Arabidopsis proteome using OxiTRAQ, a quantitative redox proteomics method. *Proteomics* **14**, 750-762.

Liu, S., Yu, F., Yang, Z., Wang, T., Xiong, H., Chang, C., Yu, W. and Li, N. (2018) Establishment of dimethyl labeling-based quantitative acetylproteomics in *Arabidopsis*. *Mol. Cell. Proteomics* **17**, 1010-1027.

Maor, R., Jones, A., Nühse, T.S., Studholme, D.J., Peck, S.C. and Shirasu, K. (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Mol. Cell. Proteomics* **6**, 601-610.

Majeran, W., Le Caer, J.-P., Ponnala, L., Meinnel, T. and Giglione, C. (2018) Targeted profiling of *Arabidopsis thaliana* subproteomes illuminates co- and posttranslationally N-terminal myristoylated proteins. *Plant Cell* **30**, 543-562.

Meng, X., Xing, S., Perez, L.M., Peng, X., Zhao, Q., Redoña, E.D., Wang, C. and Peng, Z. (2017) Proteome-wide analysis of lysine 2-hydroxyisobutyrylation in developing rice (*Oryza sativa*) seeds. *Sci. Rep*. **7**, 17486.

Miernyk, J.A., Johnston, M.L., Huber, S.C., Tovar-Méndez, A., Hoyos, E., and. Randall, D.D. (2009) Oxidation of an adjacent methionine residue inhibits regulatory seryl-phosphorylation of pyruvate dehydrogenase. *Proteomics Insights* **2**, 15-22.

Minguez, P., Parca, L., Diella, F., Mende, D.R., Kumar, R., Helmer-Citterich, M., Gavin, A.-C., van Noort, V. and Bork, P. (2012) Deciphering a global network of functionally associated post-translational modifications. *Mol. Syst. Biol*. **8**, 599.

Minguez, P., Letunic, I., Parca, L., Garcia-Alonso, L., Dopazo, J., Huerta-Cepas, J. and Bork, P. (2015) PTMcode v2: a resource for functional associations of post-translational modifications within and between proteins. *Nucleic Acids Res*. **43**, D494-D502.

Mujahid, H., Meng, X., Xing, S., Peng, X., Wang, C. and Peng, Z. (2018) Malonylome analysis in developing rice (*Oryza sativa*) seeds suggesting that protein lysine malonylation is well-conserved and overlaps with acetylation and succinylation substantially. *J. Proteomics* **170**, 88-98.

Nelson, C.J. and Millar, A.H. (2015) Protein turnover in plant biology. *Nat. Plants* **1**, 15017.

Ohtake, F., Saeki, Y., Sakamoto, K., Ohtake, K., Nishikawa, H., Tsuchiya, H., Ohta, T., Tanaka, K. and Kanno, J. (2015) Ubiquitin acetylation inhibits polyubiquitin chain elongation. *EMBO Rep*. **16**, 192-201.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567.

Pye, V.E., Christensen, C.E., Dyer, J.H., Arent, S. and Henriksen, A. (2010) Peroxisomal plant 3-ketoacyl-CoA thiolase structure and activity are regulated by a sensitive redox switch. *J. Biol. Chem.* **285**, 24078-24088.

Roitinger, E., Hofer, M., Köcher, T., Pichler, P., Novatchkova, M., Yang, J., Schlögelhofer, P. and Mechtler, K. (2015) Quantitative phosphoproteomics of the ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-mutated and Rad3-related (ATR) dependent DNA damage response in *Arabidopsis thaliana*. *Mol. Cell. Proteomics* **14**, 556-571.

Rowland, E., Kim, J., Bhuiyan, N.H. and van Wijk, K.J. (2015) The Arabidopsis chloroplast stromal N-terminome: complexities of amino-terminal protein maturation and stability. *Plant Physiol*. **169**, 1881-1896.

Rytz, T.C., Miller, M.J., McLoughlin, F., Augustine, R.C., Marshall, R.S., Juan, Y.-t., Charng, Y.-y., Scalf, M., Smith, L.M. and Vierstra, R.D. (2018) SUMOylome profiling reveals a diverse array of nuclear targets modified by the SUMO ligase SIZ1 during heat stress. *Plant Cell* **30**, 1077-1099.

Saracco, S.A., Hansson, M., Scalf, M., Walker, J.M., Smith, L.M. and Vierstra, R.D. (2009) Tandem affinity purification and mass spectrometric analysis of ubiquitylated proteins in Arabidopsis. *Plant J.* **59**, 344-358.

Seaman, J.E., Julien, O., Lee, P.S., Rettenmaier, T.J., Thomsen, N.D. and Wells, J.A. (2016) Cacidases: caspases can cleave after aspartate, glutamate and phosphoserine residues. *Cell Death Differ*. **23**, 1717-1726.

Smith, L.M., Kelleher, N.L., Linial, M. *et al.* (2013) Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186-187.

Sörensson, C., Lenman, M., Veide-Vilg, J., Schopper, S., Ljungdahl, T., Grøtli, M., Tamás, M.J., Peck, S.C. and Andreasson, E. (2012) Determination of primary sequence

specificity of *Arabidopsis* MAPKs MPK3 and MPK6 leads to identification of new substrates. *Biochem J*. **446**, 271-278.

Swatek, K.N. and Komander, D. (2016) Ubiquitin modifications. *Cell Res*. **26**, 399-422.

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C. and Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res*. **10**, 5354-5362.

Tsiatsiani, L., Timmerman, E., De Bock, P.-J., Vercammen, D., Stael, S., van de Cotte, B., Staes, A., Goethals, M., Beunens, T., Van Damme, P., Gevaert, K. and Van Breusegem F. (2013) The *Arabidopsis* METACASPASE9 degradome. *Plant Cell* **25**, 2831-2847.

UniProt Consortium (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. **45**, D158-D169 [Erratum *Nucleic Acids Res*. **46**, 2699].

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F. and Vandepoele, K. (2018) PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**, D1190-D1196.

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014) Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res.* **24**, 1334-1347.

Vener, A.V., Harms, A., Sussman, M.R. and Vierstra, R.D. (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. *J. Biol. Chem*. **276**, 6959-6966.

Venne, A.S., Solari, F.A., Faden, F., Paretti, T., Dissmeyer, N. and Zahedi, R.P. (2015) An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatography (ChaFRADIC) to study proteolytic events in *Arabidopsis thaliana*. *Proteomics* **15**, 2458-2469.

Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R. and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res*. **44**, D447-D456 [Erratum *Nucleic Acids Res*. **44**, 1033].

Vu, L.D., Stes, E., Van Bel, M., Nelissen, H., Maddelein, D., Inzé, D., Coppens, F., Martens, L., Gevaert, K. and De Smet, I. (2016) Up-to-date workflow for plant (phospho)proteomics identifies differential drought-responsive phosphorylation events in maize leaves. *J. Proteome Res.* **15**, 4304-4317.

Vu, L.D., Zhu, T., Verstraeten, I., van de Cotte, B., The International Wheat Genome Sequencing Consortium, Gevaert, K. and De Smet, I. (2018) Temperature-induced changes in the wheat phosphoproteome reveal temperature-regulated interconversion of phosphoforms. *J. Exp. Bot*. **69**, 4609-4624.

Walton, A., Stes, E., Cybulski, N., Van Bel, M., Iñigo, S., Nagels Durand, A., Timmerman, E., Heyman, J., Pauwels, L., De Veylder, L., Goossens, A., De Smet, I., Coppens F., Goormachtig, S. and Gevaert, K. (2016) It's time for some "site"-seeing: novel tools to monitor the ubiquitin landscape in *Arabidopsis thaliana*. *Plant Cell* **28**, 6-16.

Wang, Y., Hou, Y., Qiu, J., Li, Z., Zhao, J., Tong, X. and Zhang, J. (2017) A quantitative acetylomic analysis of early seed development in rice (*Oryza sativa* L.). *Int. J. Mol. Sci*. **18**, 1376.

Wauer, T., Swatek, K.N., Wagstaff, J.L., Gladkova, C., Pruneda, J.N., Michel, M.A., Gersch, M., Johnson, C.M., Freund, S.M.V. and Komander, D. (2015) Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *EMBO J*. **34**, 307-325.

Wiszniewski, A.A.G., Smith, S.M. and Bussell, J.D. (2012) Conservation of two lineages of peroxisomal (Type I) 3-ketoacyl-CoA thiolases in land plants, specialization of the genes in Brassicaceae, and characterization of their expression in *Arabidopsis thaliana*. *J. Exp. Bot.* **63**, 6093-6103.

Wu, X., Oh, M.-H., Schwarz, E.M., Larue, C.T., Sivaguru, M., Imai, B.S., Yau, P.M., Ort, D.R. and Huber, S.C. (2011) Lysine acetylation is a widespread protein modification for diverse proteins in Arabidopsis. *Plant Physiol.* **155**, 1769-1778.

Xu, S.-L., Chalkley, R.J., Maynard, J.C., Wang, W., Ni, W., Jiang, X., Shin, K., Cheng, L., Savage, D., Hühmer, A..F.R., Burlingame, A.L. and Wang, Z.-Y. (2017) Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **114**, E1536-E1543.

Xu, H., Zhou, J., Lin, S., Deng, W., Zhang, Y. and Xue, Y. (2017) PLMD: An updated data resource of protein lysine modifications. *J. Genet. Genomics* **44**, 243-250.

Yao, Q., Ge, H., Wu, S., Zhang, N., Chen, W., Xu, C., Gao, J., Thelen, J.J. and Xu, D. (2014) P^3DB 3.0: From plant phosphorylation sites to protein networks. *Nucleic Acids Res.* **42**, D1206-D1213.

Zhang, H., Deery, M.J., Gannon, L., Powers, S.J., Lilley, K.S. and Theodoulou, F.L. (2015) Quantitative proteomics analysis of the Arg/N-end rule pathway of targeted degradation in Arabidopsis roots. *Proteomics* **15**, 2447-2457.

Zhang, H., Gannon, L., Hassall, K.L., Deery, M.J.., Gibbs, D.J., Holdsworth, M.J., van der Hoorn, R.A.L., Lilley, K.S. and Theodoulou, F.L. (2018) N-terminomics reveals control of Arabidopsis seed storage proteins and proteases by the Arg/N-end rule pathway. *New Phytol*. **218**, 1106-1126.

Zhang, T., Chen, S. and Harmon, A.C. (2016) Protein--protein interactions in plant mitogen-activated protein kinase cascades. *J. Exp. Bot.* **67**, 607-618.

FIGURES

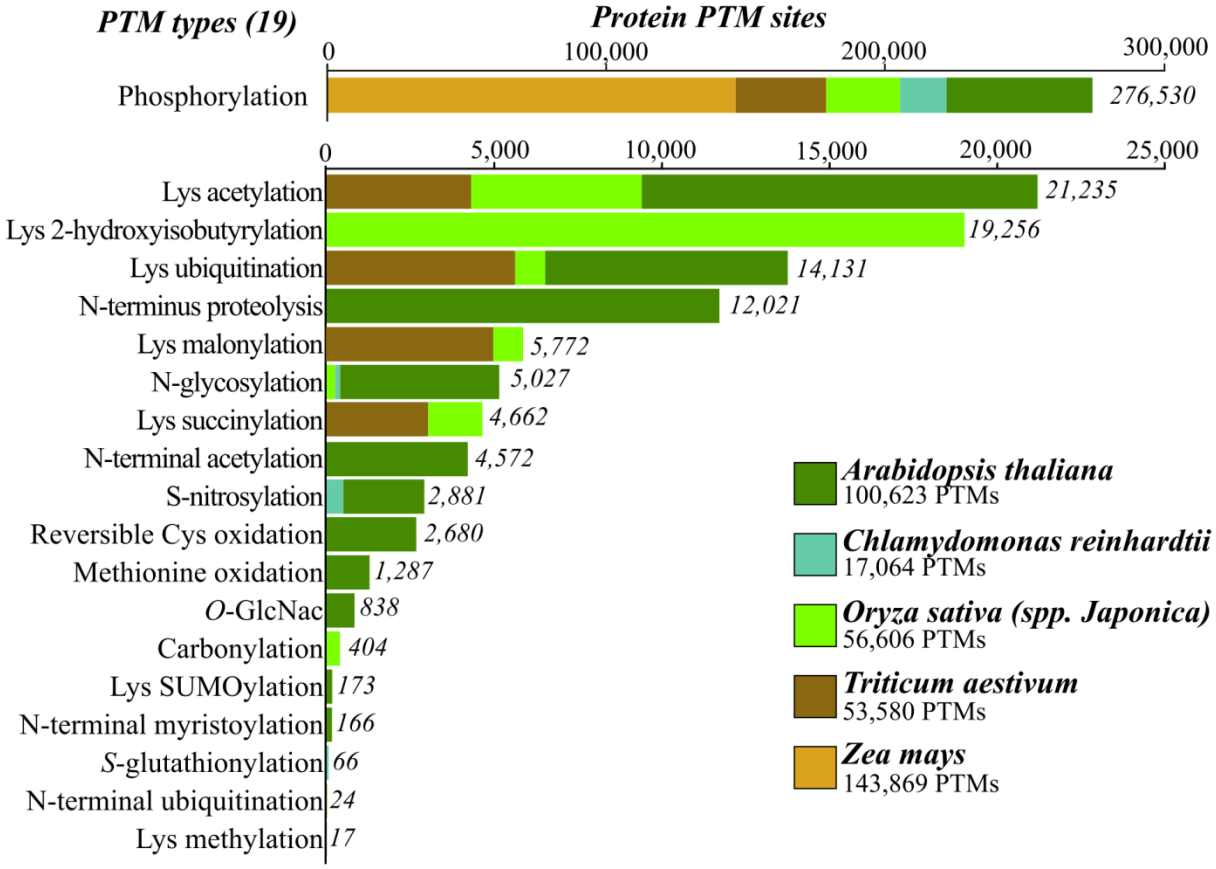


Figure 1. Plant PTM Viewer datasets. PTM sites identified by mass spectrometry for protein (isoforms) from five plant species. All distinct protein PTM sites per species were displayed (x-axis) according to the PTM type (19 types) (y-axis). For phosphorylation sites, the scale was more than 10-fold greater than that for other PTM types. Bars were colored according to species. For more information regarding the PTM types, the reader is referred to the 'PTM Info' tab online.

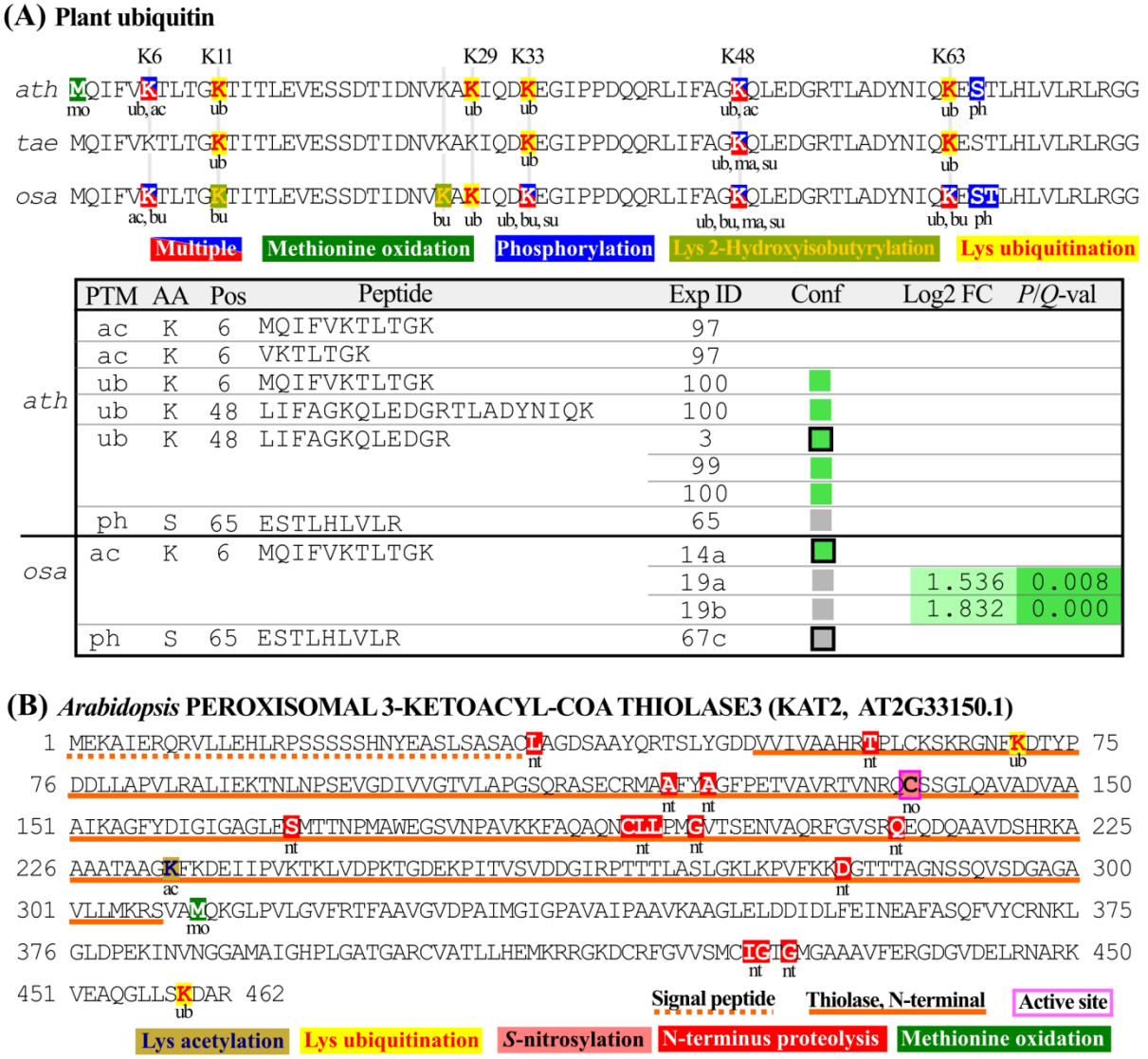


Figure 2. PTM protein overviews.
(A) The PTM protein sequence overview (top) and PTM table (bottom) of the ubiquitin protein domain encoded by Arabidopsis (*ath*) polyubiquitin 10 (AT4G05320), wheat (*tae*) TRIAE_CS42_6BS_TGACv1_515033_AA1666790, or rice (*osa*) LOC_Os06g46770. Modified residues are color coded and the respective two-letter code abbreviations of these modifications (Table S2) are indicated below the protein sequence. Ubiquitin lysines were specified with their respective position (K6, K11, K27, K29, K33, K48, and K63). A few modification sites described in the text for Arabidopsis and rice ubiquitins are shown in the PTM table (bottom). For detailed information regarding the PTM table, the reader is referred to the tutorial and the respective protein overview pages online. (B) PTM protein sequence overview of Arabidopsis PEROXISOMAL 3-KETOACYL-COA THIOLASE3 (KAT2, AT2G33150.1). The active site residue is indicated in a purple rectangle and the N-terminal thiolase protein domain (IPR020616 and the peroxisomal signal peptide are underlined by an orange solid line and dotted line, respectively. Abbreviations: Ac, acetylation; Bu, 2-hydroxyisobutyrylation; Ma, malonylation; Mo, methionine oxidation; No, S-nitrosylation; Nt, N-terminal; Ph, phosphorylation; Ro, reversible cysteine oxidation; Su, succinylation; Ub, ubiquitination.

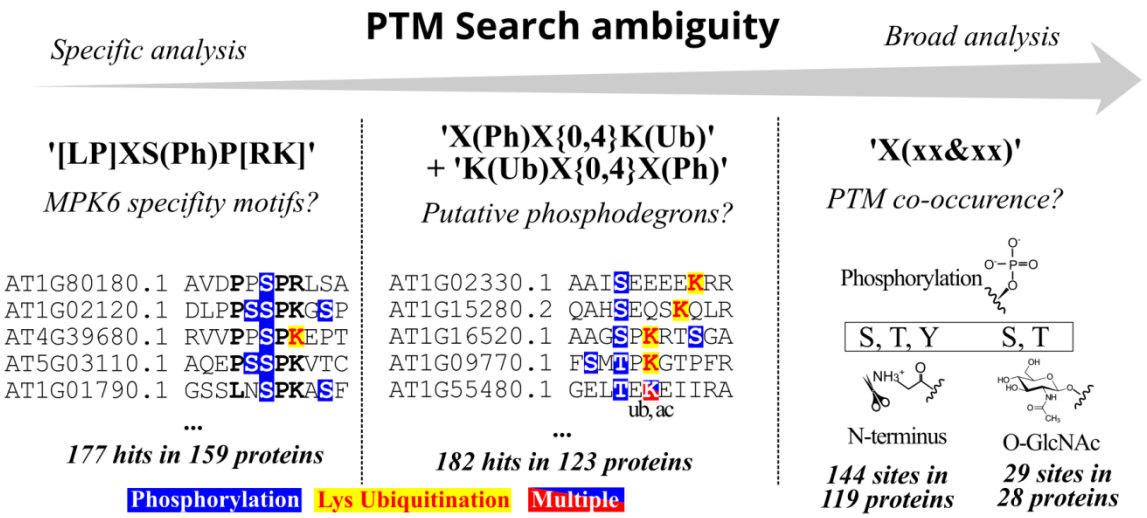


Figure 3. PTM Search with three queries and increased sequence and/or PTM ambiguity. For the MPK6 specificity and phosphodegron searches (left and middle panel), PTMs were highlighted in the returned hits, in which the consensus motif for '[LP]XS(Ph)P[RK]' is in bold. For the co-occurrence 'X(xx&xx)' search (right panel), the co-occurring phosphorylation and N-terminal residues exposed after proteolytic processing and *O*-GlcNAc are shown. These PTM searches can be reproduced online by selecting *Arabidopsis thaliana* as species and unique peptides as advanced options. Additional information regarding the 'PTM search' tool is provided in the help section and the tutorial.

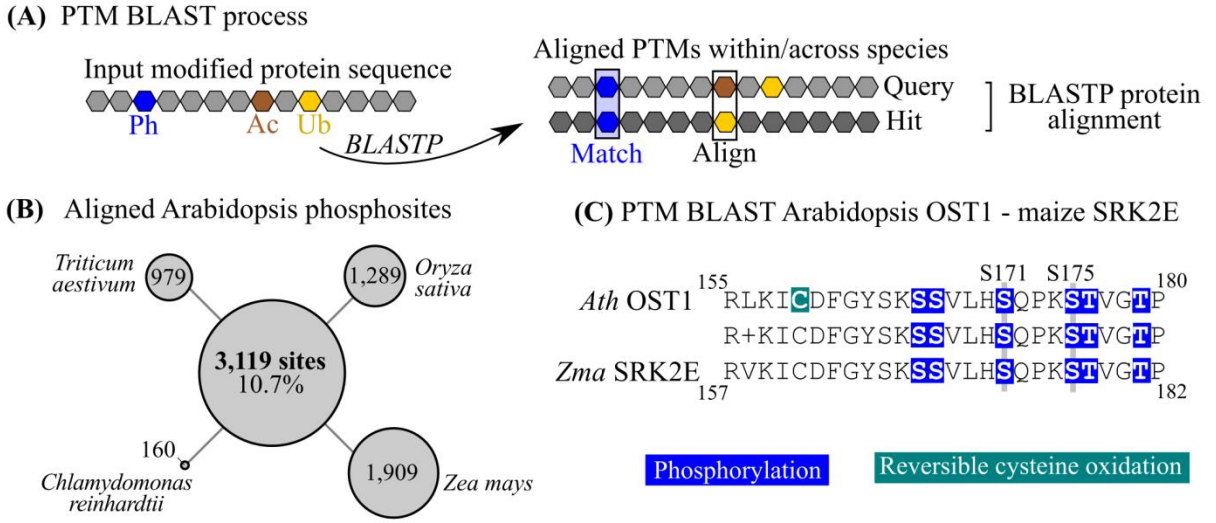


Figure 4. 'PTM BLAST' and sequence alignment.
(A) A default protein BLAST (BLASTP) carried out by 'PTM BLAST' for a given modified protein sequence, *e.g.*, phosphorylation (Ph, blue), acetylation (Ac, brown) and ubiquitination (Ub, yellow). Protein alignments of significant BLASTP hits can be visualized with highlighted PTMs. Two PTM alignment cases are counted: the alignment of modified residues regardless of the PTM type ('align' type; *e.g.* Ac and Ub) and of the same PTM type ('match' type). (B) Alignment of Arabidopsis phosphorylation sites in orthologous proteins. The number of Arabidopsis phosphorylation sites that aligned exactly to at least one phosphosite in an orthologous protein (BLASTP E-value < 0.5, alignment length > 80% longest sequence and amino acid identity > 60%) was counted for other included plant species. (C) 'PTM BLAST' alignment of conserved phosphosites, including Ser171 (S171) and Ser175 (S175), between Arabidopsis OPEN STOMATA1 (OST1, AT4G33950.1) and its maize ortholog SRK2E (Zm00001d033339_P002).

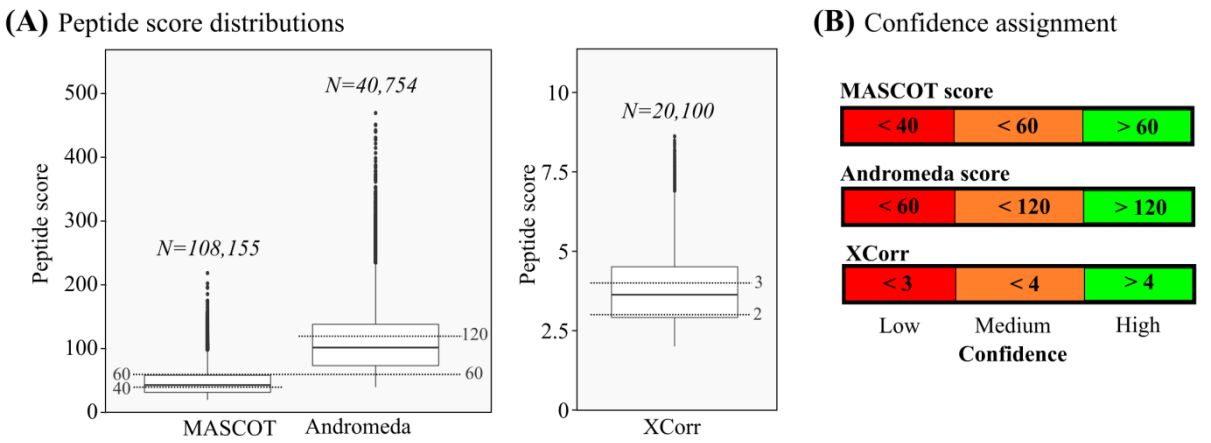


Figure 5. Confidence assignment scheme for PTM data entries (non-redundant modified peptide sites from publications).
(A) Peptide score distributions for PTMs identified by MASCOT (MASCOT ion score, 108,155 peptides), MaxQuant (Andromeda score, 40,754 peptides) and search algorithms, such as SEQUEST measuring a cross-correlation score (XCorr, 20,100 peptides). Thresholds used to assign confidence labels are indicated with a dotted line. (B) Assignment of a high (green), medium (orange) or low (red) confidence based on the respective PTM peptide scores. When no peptide score had been reported in the respective publication, PTMs were attributed a low confidence.

SUPPLEMENTAL INFORMATION

Supplemental Table S1. Comparison of the Plant PTM Viewer unique PTM peptides per species *versus* other PTM sites stored in different PTM databases. For the Plant PTM Viewer, the numbers of unique PTM peptides sequences are shown. Statistics from dbPTM (http://dbptm.mbc.nctu.edu.tw/) were obtained by browsing data per plant species. Experimental PTMCode v2 sites were extracted from Minguez *et al*. (2015). Sites of iPTMnet were derived from https://research.bioinformatics.udel.edu/iptmnet/stat.

Species	**Database**			
	Plant PTM Viewer	PTMCode v2	iPTMnet	dbPTM
Arabidopsis thaliana	56,409	5,443	32,685	27,583
*Oryza sativa (*spp. *japonica)*	25,092	0	4,376	0
*Oryza sativa (*spp.*indica)*	0	0	569	0
Zea mays	16,432	0	138	10,728
Chlamydomonas reinhardtii	21,291	0	0	0
Triticum aestivum	25,473	0	0	0
Medicago truncatula	0	0	0	12,693

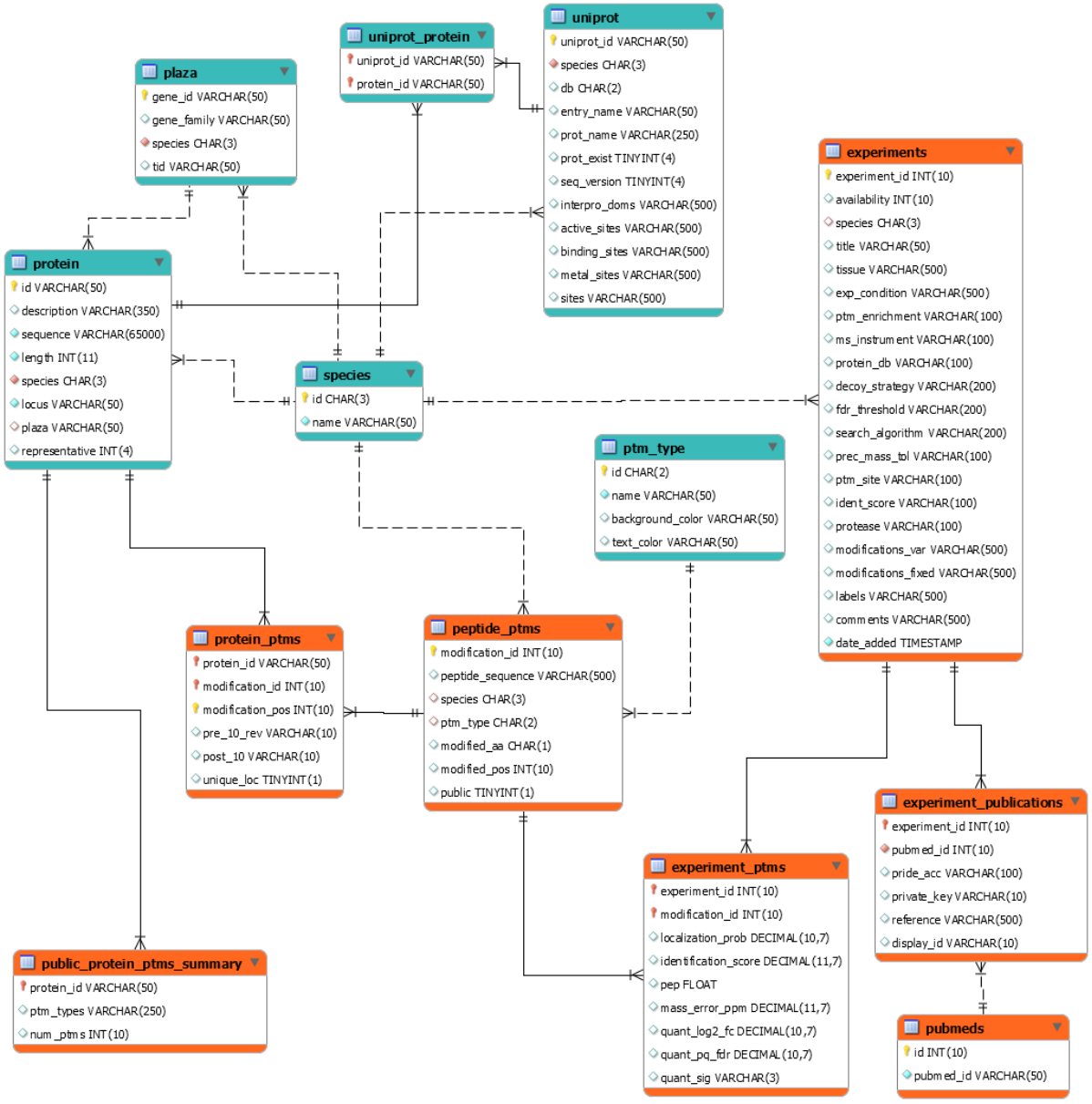
Supplemental Table S2. Nineteen plant PTMs included in the Plant PTM Viewer. The two-letter code abbreviation and respective amino acid residue occurrence and PTM viewer experiments are displayed per PTM type. Additional information and statistics per PTM can be consulted on the online 'PTM info' tab.

PTM	Abbreviation	Amino acid	Experiments
Acetylation	Ac	N-term, K	32
2-Hydroxyisobuturylation	Bu	K	1
Carbonylation	Ca	R, K, P, T	2
Malonylation	Ma	K	2
Methylation	Me	K	1
Methionine oxidation	Mo	M	2
Myristoylation	My	G (Nt)	1
N-glycosylation	Ng	S, T	4
S-nitrosylation	No	C	4
N-terminus	Nt	N-term	13
O-GlcNAcylation	Og	S, T	1
Phosphorylation	Ph	S, T, Y, N-term (rare)	79
Reversible cysteine oxidation	Ro	C	7
S-glutathionylation	Sg	C	1
Sumoylation	Sm	K	1
Succinylation	Su	K	2
Ubiquitination	Ub	K, N-term (rare)	3

Supplemental Table S3. Species reference annotations.

Species	Proteins	Annotation
Arabidopsis thaliana	48,359	Araport11
Zea mays	131,496	AGPv4
Triticum aestivum	154,140	TGACv1
*Oryza sativa (*spp. *japonica)*	49,061	JGIv7
Chlamydomonas reinhardtii	19,526	JGIv5.5

Sources: Araport11, http://araport.org/; AGPv4 and TGACv1, https://plants.ensembl.org/; JGIv7 and JGIv5.5, https://phytozome.jgi.doe.gov/.



Supplemental Figure S1. Plant PTM Viewer MySQL database scheme. Variable tables are indicated in orange and are automatically updated during novel PTM data import. Blue tables are fixed and include species, protein, PTM type annotation and cross-references to UniProtKB and PLAZA.

(A) Human GAPDH PTM sequence input

>sp|P04406|G3P_HUMAN Glyceraldehyde-3-phosphate dehydrogenase
MGKVKVGVNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLN**Y(Ph)**MVY**M(Ox)**FQYDSTHGKFHGTV
KAENGKLVINGNPI**T(Ph)**IFQERDP**S(Ph)**KIKWGDAGAEYVVESTGVFTTMEKAGAHLQGGAKRVII
S(Ph)APSADAPMFVMGVNHEKYDNSLKII**S(Ph)**NA**S(Ph)C(No)T(Ph)**TNCLAPLAKVIHDNFGIVEGL
MT**T(Ph)**VHAI**T(Ph)**A**T(Ph)**PQKTVDGPSGK(**Ac**|**Ma**)LWRDGRGALQNIIPAS**T(Ph)**GAA**K(Ma)**AVG
K(Ac)VIPELNG**K(Ac)**L**T(Ph)**GMAFRVP**T(Ph)**ANV**S(Ph)**VVDLT**C(No)**RLEKPA**K(Ac)**YDDIKKVVKQAS
EGPLKGILGYTEHQVVSSDFNSDTHSSTFDAGAGIALNDHFVKLI**S(Ph)**WYDNEFGYSNRVVDLMA
HMA**S(Ph)**KE

(B) PTM Alignment Human GAPDH - Arabidopsis GAPC2

```
GAPDH    3 KVKVGVNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVYYFQYDSTHGKF-HGTVK
           K+++G+NGFGRIGRLV R        V++VA+NDPFI   YM YMF+YDS HG++ H  +K
GAPC2    5 KIRIGINGFGRIGRLVARVVLQRDDVELVAVNDPFITTEYMTYMFKYDSVHGQWKHHELK

GAPDH   62 AENGKLVINGN-PITIFQERDPSKIKWGDAGAEYVVESTGVFTTMEKAGAHLQGGAKRVI
            ++ K ++ G  P+T+F  R+P  I WG+AGA++VVESTGVFT  +KA AHL+GGAK+V+
GAPC2   65 VKDDKTLLFGEKPVTVFGIRNPEDIPWGEAGADFVVESTGVFTDKDKAAAHLKGGAKKVV

GAPDH  121 ISAPSADAPMFVMGVNHEKYDNSLKIISNASCTTNCLAPLAKVIHDNFGIVEGLMTTVHA
           ISAPS DAPMFV+GVN  +Y + L I+SNASCTTNCLAPLAKVI+D FGIVEGLMTTVH+
GAPC2  125 ISAPSKDAPMFVVGVNEHEYKSDLDIVSNASCTTNCLAPLAKVINDRFGIVEGLMTTVHS

GAPDH  181 ITATPQKTVDGPSGKLWRDGRGALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTAN
           ITAT QKTVDGPS K WR GR A  NIIP+STGAAKAVGKV+P LNGKLTGM+FRVPT +
GAPC2  185 ITAT-QKTVDGPSMKDWRGGRAASFNIIPSSTGAAKAVGKVLPSLNGKLTGMSFRVPTVD

GAPDH  241 VSVVDLTCRLEKPAKYDDIKKVVKQASEGPLKGILGYTEHQVVSSDFNSDTHSSTFDAGA
           VSVVDLT RLEK A YD+IKK +K+ SEG +KGILGYTE  VVS+DF  D  SS FDA A
GAPC2  244 VSVVDLTVRLEKAATYDEIKKAIKEESEGKMKGILGYTEDDVVSTDFVGDNRSSIFDAKA

GAPDH  301 GIALNDHFVKLISWYDNEFGYSNRVVDLMAHMA
           GIAL+D FVKL+SWYDNE+GYS+RVVDL+ HM+
GAPC2  304 GIALSDKFVKLVSWYDNEWGYSSRVVDLIVHMS
```

Supplemental Figure S2. 'PTM BLAST' of human glyceraldehyde-3-phosphate dehydrogenase (GAPDH). (A) 'PTM BLAST' input derived from the modification sites for the UniProtKB record of human GAPDH (P04406). (B) Aligned PTM view of human GAPDH with *Arabidopsis* GAPC2 (AT1G13440.1), of which the top PTM BLAST match with six aligned PTMs of identical type. The 'PTM BLAST' results can easily be recreated by searching the modified human GAPDH sequence provided in (A).