Large-Scale Comparative Phosphoproteomics Identifies Conserved Phosphorylation Sites in Plants\[W\][OA]

Hirofumi Nakagami\(^2\), Naoyuki Sugiyama\(^2\), Keiichi Mochida, Arsalan Daudi, Yuko Yoshida, Tetsuro Toyoda, Masaru Tomita, Yasushi Ishihama\(^3\), and Ken Shirasu*\(^3\)

RIKEN Plant Science Center (H.N., K.M., A.D., K.S.) and RIKEN Bioinformatics and Systems Engineering Division (Y.Y., T.T.), Tsurumi-ku, Yokohama 230–0045, Japan; Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997–0017, Japan (N.S., M.T., Y.I.); and PRESTO, Japan Science and Technology Agency, Chiyoda-ku, Tokyo 102–0075, Japan (Y.I.)

Knowledge of phosphorylation events and their regulation is crucial to understand the functional biology of plants. Here, we report a large-scale phosphoproteome analysis in the model monocot rice (Oryza sativa japonica ‘Nipponbare’), an economically important crop. Using unfractionated whole-cell lysates of rice cells, we identified 6,919 phosphopeptides from 3,393 proteins. To investigate the conservation of phosphoproteomes between plant species, we developed a novel phosphorylation-site evaluation method and performed a comparative analysis of rice and Arabidopsis (Arabidopsis thaliana). The ratio of tyrosine phosphorylation in the phosphoresidues of rice was equivalent to those in Arabidopsis and human. Furthermore, despite the phylogenetic distance and the use of different cell types, more than 50% of the phosphoproteins identified in rice and Arabidopsis, which possessed ortholog(s), had an orthologous phosphoprotein in the other species. Moreover, nearly half of the phosphorylated orthologous pairs were phosphorylated at equivalent sites. Further comparative analyses against the Medicago phosphoproteome also showed similar results. These data provide direct evidence for conserved regulatory mechanisms based on phosphorylation in plants. We also assessed the phosphorylation sites on nucleotide-binding leucine-rich repeat proteins and identified novel conserved phosphorylation sites that may regulate this class of proteins.

Model systems have provided excellent bases for the understanding of a wide range of biological processes. For Arabidopsis (Arabidopsis thaliana), the premier model system in plant science, large sets of genetic resources, and analytical tools are now available to assist studies on the functions of plant genes (Somerville and Koornneef, 2002). However, transferring information from Arabidopsis to other plant species still remains a considerable challenge. In particular, whether information on protein modifications from model dicots is readily applicable to evolutionarily divergent monocot species remains unclear because of limited evidence about whether conserved residues are modified in the same manner.

Phosphorylation events govern a wide range of biological processes in plants and other organisms (Sugiyama et al., 2008). Therefore, a characterization of conserved phosphoproteome features in plants will promote the understanding of core regulatory systems and eventually may lead to the improvement of agronomically important plant species. Recent progress in phosphoproteomics technology paved the way for the identification of a few thousand phosphorylation sites from unfractionated plant cells by simple one-step phosphopeptide enrichment methods, and we have found more than 2,000 phosphorylation sites in Arabidopsis (Sugiyama et al., 2008). This partial Arabidopsis phosphoproteome revealed an unexpected proportion of Tyr phosphorylation (Sugiyama et al., 2008; de la Fuente van Bentem and Hirt, 2009; Kersten et al., 2009; Reiland et al., 2009). Comparable phosphoproteome data from other plant species were not available until Medicago phosphoproteome studies were recently reported (Kersten et al., 2009; Grimsrud et al., 2010).

Rice (Oryza sativa) has become an excellent model since its genome sequence was determined (Goff et al., 2002; Kikuchi et al., 2003; Rensink and Buell, 2004). Importantly, rice is a monocot that is taxonomically distinct from the dicot Arabidopsis. Thus, the rice phosphoproteome is a useful reference with which to compare Arabidopsis phosphoproteome features.
Since rice is the major staple food for a very significant proportion of the global population and also serves as a reference plant for biofuel production, the understanding of its phosphoproteome is an agriculturally important issue. Here, we present a large-scale identification of phosphorylation sites in rice. Furthermore, we updated the Arabidopsis phosphoproteome by adding new Arabidopsis phosphorylation sites. Using the data sets, we identified a large number of conserved phosphorylation sites in rice and Arabidopsis, providing, to our knowledge, the first overview of phosphoproteome conservation between dicots and monocots. Using this information, we identified conserved phosphorylation sites in nucleotide-binding leucine-rich repeat (NB-LRR) proteins, most of which are located to important regions for conferring resistance against pathogens. Furthermore, comparative analyses against the recently published Medicago phosphoproteome revealed highly conserved phosphorylation sites in three distinct plant species.

RESULTS
Identification of Rice and Arabidopsis Phosphorylation Sites

In order to perform comparative analyses of rice and Arabidopsis phosphoproteomes, a large-scale data set of rice phosphorylation sites was collected from non-stimulated suspension-cultured rice cells using three different phosphopeptide enrichment methods, lactic acid-modified titania and β-hydroxypropanoic acid-modified zirconia-utilized hydroxy acid-modified metal oxide chromatography (Ti-HAMMOC and Zr-HAMMOC) and Fe(III)-immobilized metal-ion affinity chromatography (Fe-IMAC), as detailed in our previous Arabidopsis phosphoproteome study (Sugiyama et al., 2008). In addition, we used strong cation-exchange (SCX) prefractionation before Ti-HAMMOC. Four replicates of the analysis were performed with each method, and a total of 32 liquid chromatography-mass spectrometry (LC-MS) runs were performed for each sample. Our approach identified 5,523 unique phosphorylation sites on 3,393 proteins from unfractonated rice cell lysates (Table I; Supplemental Table S1). Interestingly, SCX prefractionation prior to Ti-HAMMOC did not lead to the detection of more phosphopeptides than Ti-HAMMOC without prefractionation (3,755 phosphorylation sites by Ti-HAMMOC and 2,687 sites by SCX and Ti-HAMMOC). Probably, Ti-HAMMOC alone was sufficiently selective to enrich the phosphopeptides. In this case, SCX prefractionation resulted in a reduced recovery because of the additional step.

Furthermore, we identified 1,776 novel phosphorylation sites in the nonstimulated Arabidopsis cells that we used for our Arabidopsis phosphoproteome analysis (Sugiyama et al., 2008). Combined with our previously reported data set (Sugiyama et al., 2008), we identified a total of 3,948 unique phosphorylation sites on 2,244 proteins in Arabidopsis (Table I; Supplemental Table S2). The combined data set was used for the following analyses. A randomized decoy database estimated a 2.2% false-positive rate for identified peptides.

Features of Rice and Arabidopsis Phosphoproteomes

With our conventional method of phosphorylation site localization in identified phosphopeptides (Sugiyama et al., 2008), the contributions of phospho-Ser (pS), phospho-Thr (pT), and phospho-Tyr (pY) sites were estimated to be 84.8%, 12.3%, and 2.9% in rice and 82.7%, 13.1%, and 4.2% in Arabidopsis. To verify our results, we applied the PTM score method, which was developed by Olsen et al. (2006), to our data and calculated probabilities for phosphorylation site localization (Supplemental Table S2; Olsen et al., 2006). Phosphorylation sites with $P > 0.75$ are ranked as class I sites (Olsen et al., 2006). Since the kinase motif that is used for the definition of class II sites (Olsen et al., 2006) is not well known in plants, we defined sites with $0.75 \geq P > 0.50$ as class II sites (Trost et al., 2009). We further divided class II into two subclasses, $0.75 \geq P > 0.666$ as class II-a and $0.666 \geq P > 0.50$ as class II-b. As shown in Table II, the proportions of pS, pT, and pY sites within class I were estimated to be 89.5%, 8.9%, and 1.6% in rice and 87.7%, 9.9%, and 2.4% in Arabidopsis. The proportion of pY among the class I phosphosides in epidermal growth factor-stimulated human HeLa cells has been reported to be 1.8% (Olsen et al., 2006). Applications of equivalent data-processing methods clearly indicate that the proportion of Tyr phosphorylation in plants is equivalent to the proportion in humans.

Tandem mass spectrometry (MS/MS) spectra occasionally contain fragment ions from phosphopeptides.

Table I. Number of identified phosphopeptides, phosphoproteins, and phosphorylation sites from rice

| Variable | Rice | Arabidopsis |
|----------|------|-------------|
| No. of unique phosphopeptides$^b$ | 6,919 | 5,143 |
| No. of phosphoproteins | 3,393 | 2,244 |
| No. of unique phosphorylation sites | 5,523 | 3,948 |
| Phosphorylated residues (Ser:Thr:Tyr) | 4,685:680:158 (84.8%, 12.3%, 2.9%) | 3,267:517:164 (82.7%, 13.1%, 4.2%) |

$^a$The number of Arabidopsis data used in this study, including new as well as published data (Sugiyama et al., 2008). $^b$The number of phosphopeptides is based on unique sequences containing missed cleavage products, oxidation of Met, and phosphorylation of different sites.
with different phosphorylation sites (Supplemental Table S3). In such cases, the PTM score and other existing scoring methods are not able to properly evaluate phosphorylation site localization. Therefore, we developed a novel site-determining ion combination (SIDIC) method (Supplemental Document S1) that can assess the mixed MS/MS spectra and can be used complementary to the existing methods. As shown in Table II, application of the SIDIC method estimated a similar pY ratio as estimated by the PTM score application.

To obtain an overview of phosphorylation events in rice and Arabidopsis, cellular localization, molecular function, and biological processes relating to the phosphoproteins identified were analyzed and compared with those of experimentally characterized proteins from the cell lysates used for phosphopeptide enrichment in this study (Fig. 1). Comparison of the rice and Arabidopsis phosphoproteomes showed that distribution patterns of different types of phosphoproteins are generally similar in these species (Fig. 1). Nuclear and plasma membrane proteins were found to be phosphorylated to an overproportional extent in rice and Arabidopsis (Fig. 1A). In contrast, plastid and ribosome proteins were less frequent targets of phosphorylation (Fig. 1A). Frequent phosphorylation of proteins associated with "kinase activity," "signal transduction," and "protein modification process" indicated that kinases themselves are often regulated by phosphorylation (Fig. 1, B and C). Proteins related to "transcription factor activity," "transcription regulator activity," and "transcription" were also found to be targeted frequently by phosphorylation, indicating that transcriptional events are controlled by phosphorylation (Fig. 1, B and C).

Overlap between Rice and Arabidopsis Phosphoproteomes

To study the conservation of plant phosphoproteomes, we investigated whether orthologous proteins in rice and Arabidopsis are phosphorylated in the same manner. For this purpose, we generated alignments of orthologous protein sequences and mapped the identified phosphopeptides onto the alignments (Fig. 2). Orthologous protein groups were constructed by cluster analysis using the OrthoMCL algorithm (Li et al., 2003; Chen et al., 2007), and orthologous protein sequences were aligned using ClustalW (Thompson et al., 1994). Of the phosphoproteins identified, 865 rice proteins and 403 Arabidopsis proteins appeared to be unique to their species (Fig. 3; Table III). Among 2,528 rice and 1,841 Arabidopsis phosphoproteins that possessed ortholog(s), more than half of the corresponding ortholog(s) were found to be phosphorylated as well (Fig. 3; Table III). Strikingly, nearly one-quarter of the ortholog(s) were phosphorylated at equivalent sites or neighboring sites (Tables III and IV). We identified a number of conserved actual phosphorylation sites in the two species that were located on important factors involved in various physiological processes as listed in Table IV. This information will help to unravel conserved phosphorylation-regulated molecular mechanisms in plants. It should be noted that many of the orthologs of the identified phosphoproteins whose phosphorylation at equivalent sites was not detected in this study contain conserved phosphorylatable residues at the equivalent sites. For instance, orthologs of 345 rice and 277 Arabidopsis phosphoproteins categorized to "Distant sites are phosphorylated in orthologs" in Table III contain conserved putative phosphorylation sites. Therefore, further in-depth phosphoproteome analyses are likely to detect additional overlaps of the rice and Arabidopsis phosphoproteomes.

Recently, a large-scale phosphoproteome analysis of *Medicago truncatula* root was reported (Grimsrud et al., 2010). Therefore, we further investigated overlaps among *Medicago*, rice, and Arabidopsis phosphoproteomes. Of the reported 829 *Medicago* phosphoproteins, 756 proteins possessed orthologs either or both in rice and Arabidopsis (Supplemental Fig. S1). Significantly, among the 756 proteins, nearly 70% (526) of corresponding orthologs in rice or Arabidopsis were identified as phosphoprotein in our analysis (Supplemental Fig. S1; Supplemental Table S4). Moreover, approximately 29% (215) of the orthologs were phosphorylated at equivalent sites (Supplemental Tables S4 and S5). These results indicate that observation of the overlap between rice and Arabidopsis phosphoproteomes is not specifically derived from our experimental conditions and that phosphoproteomes are conserved in a similar manner in various plant species.

### Table II. Distribution of phosphorylated residues

| Method                        | Rice          | Arabidopsis  |
|-------------------------------|---------------|--------------|
|                               | pS:\pT:\pY    | pS:\pT:\pY   |
| Conventional method           | 4,685:680:158 (84.8%, 12.3%, 2.9%) | 3,267:517:164 (82.7%, 13.1%, 4.2%) |
| PTM score-derived             | 4,685:680:158 (84.8%, 12.3%, 2.9%) | 3,267:517:164 (82.7%, 13.1%, 4.2%) |
| localization probability      | 3,333:331:59 (89.5%, 8.9%, 1.6%) | 2,281:258:63 (87.7%, 9.9%, 2.4%) |
| Class I (P > 0.75)            | 3,333:331:59 (89.5%, 8.9%, 1.6%) | 2,281:258:63 (87.7%, 9.9%, 2.4%) |
| Class I+II-a (P > 0.666)     | 3,482:386:78 (88.2%, 9.8%, 2.0%) | 2,408:299:96 (85.9%, 10.7%, 3.4%) |
| Class I+II (P > 0.50)        | 3,951:521:138 (85.7%, 11.3%, 3.0%) | 2,823:444:152 (82.6%, 13.0%, 4.4%) |
| Extended conventional method  | 3,694:390:63 (89.1%, 9.4%, 1.5%) | 2,612:285:57 (88.4%, 9.7%, 1.9%) |
| (SIDIC)                       | 4,022:473:95 (87.6%, 10.3%, 2.1%) | 2,690:306:75 (87.6%, 10.0%, 2.4%) |

*Note: pS, pT, and pY represent phosphorylation sites at serine, threonine, and tyrosine residues, respectively.*
Figure 1. GO analysis of the identified proteins and phosphoproteins. The identified rice proteins, rice phosphoproteins, Arabidopsis proteins, and Arabidopsis phosphoproteins are represented by blue, red, green, and purple bars, respectively. A, Cellular localization. B, Molecular function. C, Biological process.
Features of Preferentially Phosphorylated Protein Groups in Plants

To further characterize patterns of phosphorylation targets in plants, we investigated which categories of proteins were overrepresented among the proteins whose ortholog(s) were phosphorylated in both rice and Arabidopsis (phospho-targeted protein group; Table III) or among the proteins whose phosphorylation sites were conserved between rice and Arabidopsis (conserved phospho-sites protein group; Table III). We used the “weight” method of the topGO tool (Alexa et al., 2006) to compare the Gene Ontology (GO) “molecular function” distribution of all rice phosphoproteins with that of the rice phosphoproteins whose Arabidopsis orthologs are phosphorylated at corresponding and/or noncorresponding sites. Proteins of the GO categories “nucleotide binding,” “RNA binding,” “hydrolase activity,” “kinase activity,” “enzyme regulator activity,” “signal transducer activity,” and “transporter activity” were overrepresented in the phospho-targeted protein group (red bars in Fig. 4). Interestingly, proteins annotated as having “kinase activity” or “signal transducer activity” were more strongly overrepresented in the conserved phospho-sites protein group (blue bars in Fig. 4) than in the phospho-targeted protein group (red bars). Conversely, proteins annotated as being “RNA binding” or possessing “transporter activity” were underrepresented in the conserved phospho-sites protein group (Fig. 4). These results indicate that proteins related to the GO categories presented in Figure 4 might be preferentially targeted by phosphorylation in plants and that proteins related to “kinase activity” and “signal transducer activity” frequently carry highly conserved phosphorylation sites. Proteins related to “RNA binding” and “transporter activity” are likely to be preferential targets for phosphorylation in plants; however, the targeted phosphorylation sites are diversified among different plant species.

Assessment of Phosphorylation Sites on NB Proteins

Disease resistance (R) proteins play a key role to detect pathogen effectors and thereby trigger effector-triggered immunity, which often results in hypersensitive response (HR)-type cell death (Chisholm et al., 2006; Jones and Dangl, 2006). The majority of R proteins belong to the NB-LRR protein family, which is one of the most variable protein families in plants (Monosi et al., 2004; Chisholm et al., 2006; Jones and Dangl, 2006; Shirasu, 2009) The rice genome encodes coiled-coil (CC)-class NB-LRR genes but not Toll/Interleukin-1 receptor (TIR)-class NB-LRR genes, which exist in dicot species (Monosi et al., 2004). Despite extensive studies on R proteins, little is known about the conserved molecular mechanisms that reg-
ulate their functions. In particular, R protein regulation by its phosphorylation has not been reported so far.

Our phosphoproteome analysis identified phosphorylation sites on NB-LRR proteins that may function as R protein (Fig. 5). We found eight phosphopeptides on nine CC-NB-LRR proteins from rice and five phosphopeptides on five CC-NB-LRR proteins, five phosphopeptides on four TIR-NB-LRR proteins, and two phosphopeptides on two TIR-NB proteins, which are highly homologous to TIR-NB-LRR proteins but lack the LRR domain, from Arabidopsis (Fig. 5). Strikingly, many of the phosphorylation sites are found to be located on conserved regions, which are important for R protein functions (Fig. 5). For example, the N-terminal 13 amino acids of potato (*Solanum tuberosum*) R x were shown to be required for triggering HR (Rairdan et al., 2008), and we have identified two rice proteins that were phosphorylated at the corresponding region (Fig. 5B). Interestingly, Arabidopsis RPM1 at the same region was recently reported to be phosphorylated in PhosPhAt, the Arabidopsis protein phosphorylation site database (underlined with green in Fig. 5B; Durek et al., 2010). These data indicate that the N-terminal border region of the CC domain is targeted by phosphorylation and may regulate its protein functions. Similarly, the N-terminal border region of the TIR domain was found to be phosphorylated, although we could not locate the exact phosphorylation site. PhosPhAt data also indicate that a conserved Tyr residue in the corresponding region, which is also found in tobacco (*Nicotiana tabacum*) N and flax (*Linum usitatissimum*) L6, is phosphorylated (Fig. 5C). Importantly, substitution of the Tyr residue in tobacco N to a nonphosphorylatable residue Phe alters its HR-inducing activity (Dinesh-Kumar et al., 2000). Furthermore, substitution to the phosphorylatable residue Ser also altered the activity but to a lesser extent compared with the Phe substitution (Dinesh-Kumar et al., 2000). These observations may indicate that the corresponding Tyr kinase can also phosphorylate Ser but less efficiently. As shown in Figure 5D, another well-conserved Tyr residue in the TIR domain was found to be phosphorylated. The Pfam database shows that this Tyr residue is highly conserved in plant species but not in other organisms.

The NB domain, which has ATPase activity, is well conserved in plant and animal proteins. Structure analysis of the NB domain of mammalian Apoptotic Protease-Activating Factor1 (APAF1) indicated that several conserved motifs, including hhGRExE (Fig. 5E), P-loop (Fig. 5F), RNBS-B (Fig. 5G), and MHD (Fig. 5H), are important for the ATPase activity (Takken et al., 2006). Mutation analyses of R proteins revealed that these motifs are important for their functions (Dinesh-Kumar et al., 2000; Bendahmane et al., 2002; Tornero et al., 2002; Howles et al., 2005). We found phosphorylation sites within or close to these motifs (Fig. 5, E–H). Although some of the phosphorylation sites are not well conserved among the R proteins, the phosphorylated Thr residue in the P-loop (Fig. 5F) and the phosphorylated Ser residue in RNBS-B (Fig. 5G) are conserved among the R proteins from six different plant species. Taken together, our data indicate that NB-LRR type R proteins are likely to be regulated by phosphorylation. Moreover, the existence of highly conserved phosphorylation sites among different plant species indicates that a conserved phosphorylation-dependent regulation system may control the diversified R proteins. This information may help to understand how R protein functions are regulated.

Table III. Conserved phosphoproteins in rice and Arabidopsis

| Type of Phosphoprotein                                      | Rice   | Arabidopsis |
|------------------------------------------------------------|--------|-------------|
| Proteins unique to plant species                           | 865    | 403         |
| Proteins with orthologs in another plant species           | 2,528 (100%) | 1,841 (100%) |
| Proteins whose orthologs are not phosphorylated            | 1,255 (49.6%) | 806 (43.8%) |
| Proteins whose orthologs are phosphorylated                | 1,273 (50.4%) | 1,035 (56.2%) |
| Equivalent sites are phosphorylated in orthologs           | 458 (18.1%) | 464 (25.2%) |
| Neighboring sites are phosphorylated in orthologs          | 138 (5.5%) | 127 (6.9%) |
| Distant sites are phosphorylated in orthologs              | 677 (26.8%) | 444 (24.1%) |
| Total                                                      | 3,393  | 2,244       |

*a*Ratios against the identified phosphoproteins with orthologs in another plant species. *b*Phosphopeptides from two plant species are partially aligned but phosphorylation sites are not aligned.

Figure 3. Overlap between the rice and Arabidopsis phosphoproteomes. The numbers of rice and Arabidopsis phosphoproteins are shown in red and blue, respectively. The hatched area in the center represents orthologous phosphoproteins that were phosphorylated in both species.
## Table IV. Selected conserved phosphorylated sites within the orthologous proteins

| Ortholog No. | Rice (TIGR Identifier) | Arabidopsis (TAIR Identifier) | Alignment of Identified Phosphopeptides |
|--------------|------------------------|-------------------------------|----------------------------------------|
| 1,155        | LOC_Os05g45420.1       | AT3G01090.1: AKIN10           | DGHFLKTSCGSPNYAAPVEISGK                |
|              |                        |                               | ------TSCGSPNYAAPVEISGK                  |
| 10,444       | LOC_Os04g31340.1       | AT3G48530.1: KING1            | VEDLWEQKQPQLSPENEK                     |
|              |                        |                               | VEDLWEAEPQLSPESEK                      |
| 11,443       | LOC_Os06g6090.1        | AT2G43790.1: MPK6             | VTSESDFMTEDVYVTR                       |
| 1,109        | LOC_Os03g49640.1       | AT4G08500.1: MEKK1            | SLEFPEP---TSFR                         |
| 348          | LOC_Os07g38120.1       | AT5G19450.1: CDCPK1           | EGSLQLEGAN                             |
| 222          | LOC_Os02g14130.1       | AT4G18710.1: BIN2             | GEANISYICSR                            |
|              |                        | AT1G06390.1: GSK1             | GEANISYICSR                            |
| 752          | LOC_Os09g12240.1       | AT1G53610.1: BRL1             | GSAELQTQLYKL                           |
|              |                        | AT3G13380.1: BRL3             | GSAELQTQLYKL                           |
| 2,507        | LOC_Os02g03060.1       | AT3G48750.1: CDKA;1           | TPTHEVVTLYWR                           |
|              |                        | AT5G10270.1: CDKC;1           | TSYSDHHTGNYTNR                         |
| 2,323        | LOC_Os01g72970.1       | AT5G04510.1: PDK1             | LASIDSFPSDR---                        |
| 991          | LOC_Os06g38470.1       | AT4G303080.1: HDA6            | IWSGTAT                                  |
|              | LOC_Os03g44500.1       | AT5G04510.1: PDK1             | LASSIDSFPSDR---                        |
|              | LOC_Os05g05240.1       | AT4G303080.1: HDA6            | IWSGTAT                                  |
|              | LOC_Os03g44500.1       | AT5G04510.1: PDK1             | LASSIDSFPSDR---                        |
|              | AT1G08420.1            | AT1G08420.1: HDA6             | LASSIDSFPSDR---                        |
| 13,068       | LOC_Os12g06520.1       | AT2G27210.1: BSU1             | LNQHTQDQ---TQDLHK                       |
| 621          | LOC_Os02g52780.1       | AT1G08420.1: HDA6             | QGSLTPLR                                |
|              | LOC_Os03g44500.1       | AT2G27210.1: BSU1             | QGSLTPLR                                |
| 11,287       | LOC_Os09g07300.1       | AT1G43700.1: VP1              | SNSMDGNS---=PFEGAESLSSG--LPDYAK        |
| 7,908        | LOC_Os05g05240.1       | AT3G143700.1: VP1             | SNSMDGNS---=PFEGAESLSSG--LPDYAK        |
| 2,347        | LOC_Os06g38470.1       | AT5G05g05240.1: HDA6          | IWSGTAT                                  |
| 313          | LOC_Os02g54070.1       | AT5G05g05240.1: HDA6          | IWSGTAT                                  |
| 1,355        | LOC_Os01g08560.1       | AT4G38130.1: HD1              | HADSDVEMDDVPLDSSR                      |
|              | LOC_Os05g08401.1       | AT4G38130.1: HD1              | WDPDSMDMVD-----DD---RKIPSR             |
| 587          | LOC_Os03g5730.1        | AT5G05g05240.1: HDA6          | FYMEPETSDSGMASGAHTGR                    |
| 1,140        | LOC_Os03g5730.1        | AT5G05g05240.1: HDA6          | FYMEPETSDSGMASGAHTGR                    |
|              | AT3G09840.1: CDC48     | AT5G05g05240.1: HDA6          | FYMEPETSDSGMASGAHTGR                    |

(Table continues on following page.)
In order to investigate the conservation of phosphoproteomes in plants, large-scale data sets of phosphorylation sites are required from species whose genome sequences are available. In the case of the well-characterized model plant Arabidopsis, 1,172 phosphopeptides were identified from the plasma membrane of elicitor-stimulated and untreated cultured cells (Benschop et al., 2007) and 3,029 phosphopeptides were identified from the plasma membrane of elicitor-stimulated and untreated cultured cells. This information is continued in the subsequent tables.
were identified from nonstimulated seedlings (Reiland et al., 2009). In contrast, there was only one report of a few hundred phosphopeptides identified in rice embryos (Zhang et al., 2007). When we initiated our study, this report was the only phosphoproteome information obtained by shotgun-type analysis available from plants other than Arabidopsis, and the amount of information was too small to allow for meaningful comparison with the Arabidopsis data sets. Therefore, we performed large-scale phosphoproteomics in rice to obtain sufficient information for a comparative analysis.

The false-positive rate obtained in this study appears to be relatively high (2.2%) compared with other published works (Olsen et al., 2006; Swaney et al., 2009). However, this value was calculated after removing redundancy of identified phosphopeptides in the final list, whereas other works calculated the rate for redundant peptide lists generated from SCX-fractionated LC-MS runs (Olsen et al., 2006; Swaney et al., 2009). Generally, the overlap of identified phosphopeptides between SCX fractions is higher than the overlap of false-positive peptides. Therefore, the false-positive rate from redundant peptide lists would increase to 2% to 5% if the peptide redundancy were removed. In fact, when we applied the calculation method used in these published works, the false-positive rate in this study was estimated to be 1.1%, a similar level to those in the representative published works (Olsen et al., 2006; Swaney et al., 2009). On the other hand, regarding the false-positive identification for each phosphopeptide in our final list, each confidence level in identification should be evaluated. Therefore, we decided to classify these phosphopeptides into three categories based on Mascot probability score, such as class A (99.9% confidence), class B (99.0% confidence), and class C (95.0% confidence). The class category is described in Supplemental Table S1. For the false-positive rates, we obtained 0.12% and 1.06% for class A and class B phosphopeptides, respectively.

In our previous study, we had applied our own criteria for the identification of phosphorylation sites based on the presence of site-determining ions; it was assumed that fragment ions indicated phosphorylation sites unambiguously (Sugiyama et al., 2008). We found that the proportion of pY in plants was equivalent to that in humans (Sugiyama et al., 2008). Recently, de la Fuente van Bentem and Hirt (2009) challenged this view by pointing out that the phosphorylation site assignment in our study might not have been unambiguous without visual inspection of all phosphopeptide spectra and without applying scoring methods as previously reported and that, therefore, the pY ratio in plants could be much lower than in animals. To assess this possibility, we applied the PTM score-based method to the Arabidopsis and rice phosphoproteome data sets to compare them precisely with a human data set that had been evaluated by identical methods (Olsen et al., 2006). The results we obtained confirmed that the pY ratio in plants is equivalent to that in humans (Table II).

Phosphopeptides with different phosphorylation sites often exhibit similar retention times in LC-MS analysis, and MS/MS spectra occasionally contain fragment ions from overlapping phosphopeptides. We synthesized 90 phosphopeptide pairs having the same amino acid sequences with different phosphorylation location sites and found that approximately 30% of pS-pY pairs and 60% of pS-pS pairs have overlapped peaks under the LC-MS gradient condition used in this study (Supplemental Table S3). In such cases, none of the existing scoring methods provide reliable data. Therefore, we extended the approach used in our previous Arabidopsis study and characterized the reliability of possible phosphorylation site localizations by counting the observed as well as the observable SIDICs (Supplemental Document S1). In contrast to the PTM score methodology, this approach is not probability based and is closer to manual inspections for the presence of y- and b-ions. In this approach, ambiguous phosphorylation sites are divided into indistinguishable and undetermined sites. Indistinguishable sites are defined by the pres-
ence of site-determining ions in cases in which the spectra also contain site-determining ions for other possible phosphorylation sites. Note that we consider only unique y- and b-ions as being site-determining (there often are cases where some given y- and b-ions have similar mass-to-charge ratio \([m/z]\) values as other y- and b-ions, and these ions are not considered to be site determining). Based on the numbers of "unambiguous" or "unambiguous and indistinguishable" sites by this extended method, we calculated pY ratios in rice and Arabidopsis (Supplemental Table S2) and confirmed that the obtained values were similar to those by the PTM score-derived localization probability method (Table II). The complementary use of different approaches might help to reduce the occurrence of false-positive and false-negative assignments of phosphorylation site localizations.

Recently, Reiland et al. (2009) reported a phosphoproteome analysis of Arabidopsis seedlings that identified a similar number of phosphorylation sites as our previous study (Sugiyama et al., 2008). In contrast to our results, Reiland et al. (2009) reported a proportion of pY in Arabidopsis of only 0.38%. This discrepancy may be due to three reasons. First, the pY ratio may differ between cell cultures and seedlings, as discussed by Reiland et al. (2009). Second, Tyr-phosphorylated proteins occur at low abundance compared with Ser- and Thr-phosphorylated proteins; therefore, the phosphopeptide detection sensitivity of LC-MS setups as well as the efficiency of phosphopeptide enrichment techniques could affect the recovery rate for Tyr-phosphorylated peptides. Third, the counting methods for phosphorylation sites differed between the two studies. For example, Reiland et al. (2009) did not take the redundancy of phosphorylation sites into account and counted each site on proteins multiple times, resulting in overestimation of pS and underestimation of the pY ratio. When we apply their method...
to recount our Arabidopsis class I phosphorylation sites, the numbers of pS, pT, and pY are 3,387 (98.8%), 320 (8.5%), and 65 (1.7%), respectively, representing an underestimation of the pY ratio.

For phosphorylation-site localization, utilization of a recently developed electron transfer dissociation (ETD) method to obtain MS/MS spectra of phosphopeptide is expected to be quite useful (Syka et al., 2004). Recent Medicago phosphoproteome analysis utilized the ETD technology and proved that ETD is superior in phosphorylation-site localization compared with the currently most used collision-induced dissociation method (Grimsrud et al., 2010). By utilizing the ETD method, Grimsrud et al. (2010) concluded that the proportion of pY in Medicago is 1.3%, similar to our findings.

To examine the conservation of phosphorylation sites, we mapped phosphopeptides onto aligned sequences of orthologs as shown in Figure 2. Compared with PhosphoBlast (Wang and Klemke, 2008) analysis, this method avoids the risk of matching peptides from unrelated nonhomologous proteins. In addition, the method allows matching particular peptides, such as peptides from the border of conserved and nonconserved regions of a protein. For instance, phosphopeptides from Arabidopsis HSP70 and its rice orthologs in Table IV may not be matched by the PhosphoBlast tool.

We found that a number of physiologically important orthologous protein pairs from rice and Arabidopsis are phosphorylated at equivalent sites (Table IV). For instance, Arabidopsis MEKK1, a mitogen-activated protein (MAP) triple kinase that regulates plant immunity responses and redox homeostasis (Ichimura et al., 2006; Nakagami et al., 2006; Suarez-Rodriguez et al., 2007), and its rice homolog were found to be phosphorylated at the N-terminal putative regulatory domain. Since nothing is known about regulatory mechanisms of MEKK1 in plants so far, our phosphorylation data may provide a clue to understand signaling cascades of MEKK1. Another good example is Arabidopsis VirE2-Interacting Protein1 (VIP1), which is a host plant factor utilized by the plant pathogen Agrobacterium for its delivery of transfer DNA into the host cells (Tzfira et al., 2001). VIP1 is regulated through phosphorylation by MAP kinases, and Agrobacterium appears to exploit the phosphorylation-dependent machinery of its hosts (Djamei et al., 2007). Interestingly, we found that VIP1 and its rice homolog are phosphorylated at equivalent sites, which are distinct from the MAP kinase-targeting sites (Djamei et al., 2007). Thus, VIP1 may be regulated by types of kinases other than MAP kinases.

Despite a lack of unambiguous experimental evidence, phosphorylation-dependent regulation of protein function observed in one plant species was often assumed to be conserved in orthologous proteins in other plant species (Nuhse et al., 2004). Using our data sets consisting of several thousand phosphopeptides, we demonstrated that orthologous proteins from different species actually are frequently phosphorylated in a similar manner. It should be noted that the rice and Arabidopsis cell cultures used in this study were generated from different plant tissues and were maintained under different light conditions. Since these differences appear to alter the phosphoproteome status, the similarity between the phosphoproteomes of different species can be expected to be much higher than reported here, if identical cell types grown under similar conditions are tested. Significantly, the Medicago phosphoproteome reported by other groups highly overlapped with the rice and Arabidopsis phosphoproteomes (Supplemental Fig. S1), despite the fact that the Medicago study utilized different types of plant material and a distinct analytical platform. These data may indicate that plant phosphoproteomes are far more conserved than we would have expected. Nevertheless, our results demonstrate the value of comparative characterizations of phosphoproteomes.

In order to reveal common features of the phosphoproteomes in different plants, we have identified mostly novel phosphorylation sites from rice than those published to date. All phosphorylation site data for rice and Arabidopsis, including mass spectral profiles (Supplemental Figs. S2 and S3) and the phosphorylation profiles of orthologs, will be freely available on a Web-based RIKEN database called Plant Phosphoproteome Database (http://phosphoproteome.psc.database.riken.jp), which is connected to the Keio University PepBase-P phosphopeptide database (http://pepdbase.iab.keio.ac.jp) with annotated MS/MS spectra. As investigated for phosphorylation sites on NB-LRR proteins, these database resources will facilitate research on phosphorylation events in plants. The large data set will also be a useful resource for statistical and bioinformatic analyses.

**MATERIALS AND METHODS**

**Plant Material**

Suspension-cultured rice cells (Oryza sativa japonica ‘Nipponbare’) derived from seed scutella were maintained using modified N-6 medium as described previously (Desaki et al., 2006). Rice cells were incubated on a rotary shaker at 25°C in the dark, and a 2-ml aliquot of loosely packed cells was transferred to 100 ml of fresh medium every week. Every 2 weeks, cell clusters were filtered through a 20-mesh screen to make fine aggregates. An Arabidopsis (Arabi-
dopsis thaliana) cell suspension line (ecotype Landsberg erecta) derived from stem explants (Maor et al., 2007) was grown in Murashige and Skoog medium (pH 5.7) containing 3% Suc, 10 ml L⁻¹ MES, 100 mg L⁻¹ thiamine-HCl, 1 mg L⁻¹ pyridoxine-HCl, 1 mg L⁻¹ nicotinic acid, 0.5 mg L⁻¹ L-Asparagine, and 0.05 mg L⁻¹ 6-benzylaminopurine under a 16-h-light/8-h-dark cycle at 22°C. Seven-day-old rice and Arabidopsis suspension cultures were harvested by vacuum filtration, frozen immediately in liquid nitrogen, and kept at −80°C until analysis.

**Sample Preparation for LC-MS**

Arabidopsis or rice cells (0.2 g, wet) were frozen in liquid nitrogen and then disrupted with a Multi-beads shocker (MB400U; Yasui Kikai). The disrupted cells were suspended with 1 ml of 0.1 M Tris-HCl (pH 8.0) and then treated with protein phosphatase inhibitor cocktail 1 and 2 (Sigma) and protease inhibitors (Sigma) according to the manufacturer’s protocol. The homogenate was centrifuged at 1,500g for 10 min, and the supernatant was added with urea at a final concentration of 8 M. The protein amount in the solution was measured with a bicinchoninic acid protein assay kit (Thermo Scientific). The solution was reduced with 10 mM dithiothreitol for 30 min at room temper-
Nakagami et al.

ature, alkylated with 50 mU iodoacetamide for 30 min at room temperature in the dark, and digested with Lys-C (1:100, w/w) for 3 h at room temperature, followed by dilution 4-fold with 50 mU ammonium bicarbonate and digestion with trypsin (1:100, w/v) overnight at room temperature (Saito et al., 2006). These digested samples were acidified with the addition of trifluoroacetic acid (TFA) and were desalted using StageTips with C18 Empore disc membranes (3M; Rappsilber et al., 2003) as described below. The desalted peptide mixtures before phosphopeptide enrichment were analyzed by LC-MS/MS, and proteins in the cell lysates were identified. Phosphopeptide enrichment based on HAMMOC was performed using lactic acid-modified titania (Ti-HAMMOC) and β-hydroxypropanoic acid-modified zirconia (Zr-HAMMOC), as described below. Fe-IMAC using Phos-Select (Sigma) was also conducted to enrich the phosphopeptides as described below. The eluates from HAMMOC and IMAC were concentrated in a vacuum evaporator for nano-LC-MS/MS analysis. SCX prefractionation before Ti-HAMMOC was performed as detailed previously (Ishihama et al., 2006). Four fractions were prepared for the subsequent Ti-HAMMOC phosphopeptide enrichment as described above.

Desalting with C18 Stage Tip

For desalting after tryptic digestion, a disc cut out from the membrane with a 10-gauge syringe needle was inserted into a pipette tip (D-100; Gilson). The pipette tip was conditioned with 100 µL of 0.1% TFA, 80% acetonitrile, and then equilibrated with 100 µL of 0.1% TFA, 5% acetonitrile by centrifugation at 1,000g for 1 min. The tryptic digest corresponding to 100 µg of protein was loaded onto the tip by centrifugation at 1,000g for 5 min. The tip was washed with 100 µL of 0.1% TFA, 5% acetonitrile by centrifugation at 1,000g for 1 min.

The peptides were eluted with 100 µL of 0.1% TFA, 80% acetonitrile by centrifugation at 1,000g for 1 min.

For desalting after enrichment of phosphopeptides, a disc cut out from the membrane with a 16-gauge syringe needle was inserted into a pipette tip (D-200; Gilson). Desalting was performed in the same manner as described above except the volume of solvent and the loading amount of sample were changed. In each step, 20 µL of solvents was used and whole sample eluted from one MOC tip or IMAC tip was loaded to one desalting tip.

Enrichment of Phosphopeptides with HAMMOC

Custom-made HAMMOC tips were prepared as follows. A disc cut out from C8 Empore disc membranes (3M) with a 20-gauge syringe needle was inserted into a 0.1- to 10-µL pipette tip (Eppendorf) as a frit. Then, slurry of 0.5 mg of bulk titania (particle size, 10 µm; GL Sciences) or zirconia (particle size, 10 µm; Zircrom Separations) beads in 10 µL of methanol was packed into the tip by centrifugation at 1,000g for 1 min. Prior to loading samples, the HAMMOC tips were equilibrated with 100 µL of 0.1% TFA, 80% acetonitrile with hydroxy acids as selectivity enhancers (solution A) by centrifugation at 2,000g for 1 min. As the enhancer, lactic acid was used at a concentration of 300 mg mL−1 for the titania HAMMOC tip and β-hydroxypropanoic acid at 100 mg mL−1 for zirconia. The desalted tryptic digest from a total 100 µg of Arabidopsis or rice proteins was diluted with 100 µL of solution A, and a 50-µL aliquot was loaded to the HAMMOC tip four times by centrifugation at 1,000g for 5 min. After successive washing with solution A and 0.1% TFA, 80% acetonitrile by centrifugation at 2,000g for 1 min, the peptide was eluted with 50 µL of 0.5% ammonium hydroxide or 1% disodium hydrogenphosphate by centrifugation at 1,000g for 5 min. The eluted fraction was acidified with TFA and desalted using C18 StageTips as described above. The desalted sample was concentrated in a vacuum evaporator, followed by dissolution with 10 µL of solution A for subsequent nano-LC-MS/MS analysis.

Enrichment of Phosphopeptides with IMAC

A disc cut out from C8 Empore disc membranes (3M) with a 16-gauge syringe needle was inserted into a pipette tip (D-200; Gilson) as a frit. Then, 30 µL of slurry of Phos-Select (Sigma) was packed into the tip by centrifugation at 1,000g for 1 min. The tip was equilibrated with 200 µL of 0.3% TFA, 0.1% TFA, 50% acetonitrile by centrifugation at 1,200g for 3 min. The desalted tryptic digest from a total 100 µg of Arabidopsis or rice proteins was diluted with 100 µL of water and loaded into the IMAC tip. The tip was washed with 50 µL of 0.3% TFA, 50% acetonitrile by centrifugation at 1,200g for 3 min. The peptides were 100 µL of 0.5% ammonium hydroxide or 1% disodium hydrogenphosphate by centrifugation at 1,200g for 3 min. The eluted fraction was acidified, desalted, and analyzed with nano-LC-MS/MS analysis as described above.

Nano-LC-MS System

An LTQ-Orbitrap XL (Thermo Fisher Scientific) coupled with a Dionex Ultimate3000 pump and an HTC-PAL autosampler (CTC Analytics) was used for nano-LC-MS/MS analyses. A self-pulled needle (150 mm length × 100 µm i.d., 6-µm opening) packed with ReproSil C18 materials (3 µm; Dr. Maisch GmbH) was used as an analytical column with “stone-arch” frit (Ishihama et al., 2002). A spray voltage of 2,400 V was applied. The injection volume was 5 µL, and the flow rate was 500 nL min−1. The mobile phases consisted of 0.5% acetic acid (A) and 0.5% acetic acid and 80% acetonitrile (B). A three-step linear gradient of 5% to 10% B in 5 min, 10% to 40% B in 60 min, 40% to 100% B in 5 min, and 100% B for 10 min was employed. The MS scan range was m/z 300 to 1,500. The top-10 precursor ions were selected in the MS scan by Orbitrap with resolution R=10,000 and for subsequent MS/MS scan, the top trap in the automated gain control mode, where automated gain control values of 5.00e + 05 and 1.00e + 04 were set for full MS and MS/MS, respectively. The normalized collision-induced dissociation was set to 35.0. A lock mass function was used for the LTQ-Orbitrap XL to obtain constant mass accuracy during gradient analysis (Olsen et al., 2005). Note that we did not use multistage activation or neutral loss-triggered MS3.

Database Searching

Mass Navigator version 1.2 (Mitsui Knowledge Industry) with the default parameters for LTQ-Orbitrap XL was used to create peak lists on the basis of the recorded fragmentation spectra. The m/z values of the isotope peaks were converted to the corresponding monoisotopic peaks when the isotope peaks were selected as the precursor ions. In order to improve the quality of MS/MS spectra, Mass Navigator discarded all peaks of less than 10 absolute intensity and with less than 0.1% of the most intense peak in MS/MS spectra (Supplemental Fig. S4; Ravichandran et al., 2009). Peptides and proteins were identified by means of automated database searching using Mascot version 2.2 (Matrix Science) in The Institute for Genome Research genome annotation project database (all.pep, ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_5.0/all.chrs; chrC.pep, ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/chloroplast.dir; chrM.pep, ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/mitochondrion.dir) and The Arabidopsis Information Resource database (TAIR7_pep_20070425, ftp://ftp.arabidopsis.org/home/tair/Sequences blast_databases/TAIR7_blastdatabases/) with a precursor mass tolerance of 3 ppm, a fragment ion mass tolerance of 0.8 Da, and strict trypsin specificity (Olsen et al., 2004), allowing for up to two missed cleavages. Carbamidomethylation of Cys was set as a fixed modification, and oxidation and Met and phosphorylation of Ser, Thr, and Tyr were allowed as variable modifications. Peptides were considered identified if the Mascot score was over the 95% confidence limit based on the “identity” score of each peptide and if at least three successive y- or b-ions with a further two or more y-, b-, and/or precursor-origin neutral loss ions were observed, based on the error-tolerant peptide sequence tag concept (Mann and Wilm, 1994). A randomized decoy database created by a Mascot Perl program was used to estimate the false-positive rate for identified peptides with these criteria. Note that most sulfated peptides could be discriminated from phosphopeptides because of the ultrahigh accuracy of the Orbitrap instrument that we used. The number of unique phosphopeptides was counted based on the amino acid sequence and the modification sites.

The identified peptide sequences were searched in the protein databases to examine whether any single peptide matched multiple proteins. When a peptide matched multiple proteins other than splicing variants, all matched proteins were treated as identified phosphoproteins. For unique phosphorylation site counting, phosphopeptides that overlapped on a protein were merged at first, in order to remove redundancy, and the number of phosphorylated sites in unique merged phosphopeptides was counted according to the strategy reported elsewhere (Olsen et al., 2006).

Synthetic Phosphopeptide Analysis

All synthetic phosphopeptides were obtained from Sigma Genosys. Synthetic phosphopeptides were measured by the nano-LC-MS system using
Phosphorylation Site Localization

Phosphorylation site localization was evaluated using three approaches. A conventional approach used in our previous work was based on the presence of site-determining ions, either y- or b-ions in the peak lists of the fragment ions, which supported the phosphorylation sites unambiguously (Sugiyama et al., 2008). This approach was only applied to the Mascot top-hit peptides to confirm whether Mascot phosphorylation site assignment was correct or not. The other was based on PTM scores to assign probabilities for each of the possible sites based on their site-determining ions. Then, potential phosphorylation sites were grouped into four categories depending on their PTM localization probabilities (Olsen et al., 2006). In this study, PhosCalc version 1.2 (Maclean et al., 2008), downloaded from http://www.ayeaye.tsl.ac.uk/PhosCalc/, was used to calculate PTM scores and PTM localization probabilities, and the four categories were defined as class I (localization probability, \( P > 0.75 \)), class II-a (0.75 \( \geq P > 0.666 \)), class II-b (0.666 \( \geq P > 0.5 \)), and class III \( (P < 0.5). \)

The SIDIC was based on the presence of site-determining y- or b-ions, similar to the conventional approach. But this approach was applied to each of the possible sites as the PTM score-based approach. In addition, in order to characterize the reliability of the candidate site localization, we evaluated the number of the observed SIDICs as well as the observable SIDICs. When the observed site-determining ions for one candidate were consistently observed in each MS/MS spectrum, the phosphorylation site was defined as an unambiguous site. The existence of indistinguishable sites was inferred when MS/MS spectra showed two or more candidate-supporting site-determining ions. Spectra without site-determining ions resulted in the identification of phosphopeptides with undetermined sites. More details of this approach are described in Supplemental Document S1.

Bioinformatics

Orthologous protein groups were constructed by cluster analysis using the OrthoMCL algorithm with default parameter settings (Li et al., 2003; Chen et al., 2007). Splicing variants were removed from the rice and Arabidopsis databases that were used for searches. For multiple sequence alignments of orthologous proteins, ClustalW (Thompson et al., 1994) was employed with default parameter settings. The identified phosphopeptides were mapped onto the alignments using an in-house software written in Perl. GO information for rice and Arabidopsis proteins was obtained from the Michigan State University rice genome annotation project database (http://rice.plantbiology.msu.edu/index.shtml) and The Arabidopsis Information Resource (http://www.arabidopsis.org/). The map2slim script (http://search.cpan.org/~cmungall/go-perl/scripts/map2slim) was used to count GO slim terms.

Retention times of synthetic phosphopeptides.

List of phosphopeptides that match Arabidopsis orthologs.

Details of the SIDIC.

Acknowledgments

We thank Dr. Naoto Shibuya (Meiji University) for kindly providing the suspension-cultured rice cells and Sumiko Ohnuma (Keio University) for technical assistance.

Received April 6, 2010; accepted May 11, 2010; published May 13, 2010.

Literature Cited

Alexa A, Rahnenfuhrer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22: 1600–1607

Bendahmane A, Farnham G, Moffit P, Baulcombe DC (2002) Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato. Plant J 32: 195–204

Benschop JJ, Mohammed S, O’Flaherty M, Heck AJ, Sliijper M, Menke FL (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol Cell Proteomics 6: 1198–1214

Chen F, Mackey AJ, Vermunt JK, Roos DS (2007) Assessing performance of orthology detection strategies applied to eukaryotic genomes. PLoS One 2: e2683

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124: 803–814

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

Desaki Y, Miya A, Venkatesh B, Tsuyumu S, Yamane H, Kaku H, Minami E, Shibuya N (2006) Bacterial lipopolysaccharides induce defense responses associated with programmed cell death in rice cells. Plant Cell Physiol 47: 1530–1540

Dinesh-Kumar SP, Tham WH, Baker BJ (2000) Structure-function analysis of the tobacco mosaic virus resistance gene N. Proc Natl Acad Sci USA 97: 14789–14794

Djamei A, Pitzschke A, Nakagami H, Rajh I, Hirt H (2007) Trojan horse strategy in Agrobacterium transformation: abusing MAPK defense signaling. Science 318: 453–456

Durek P, Schmidt R, Hazlewood JL, Jones A, Maclean D, Nagel A, Kersten B, Schulze WX (2010) PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. Nucleic Acids Res 38: D628–D634

Goff SA, Rice D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J,Sessions A, Oeller P, Varma H, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296: 92–100

Grimsrud PA, den Os D, Wenger CD, Swaney DL, Schwartz D, Sussman MR, Ane JM, Coon JJ (2010) Large-scale phosphoprotein analysis in Medicago truncatula roots provides insight into in vivo kinase activity in legumes. Plant Physiol 152: 19–28

Howles P, Lawrence G, Finnegan J, McFadden H, Ayliffe M, Dodds P, Ellis J (2005) Autoactive alleles of the flax L6 rust resistance gene induce non-race-specific rust resistance associated with the hypersensitive response. Mol Plant Microbe Interact 18: 570–582

Ichimura K, Casais C, Peck SC, Shinozaki K, Shirasu K (2006) MEKK1 is
required for MPK4 activation and regulates tissue-specific and temperature-dependent cell death in Arabidopsis. J Biol Chem 281: 36969–36976.

Ishihama Y, Rappsilber J, Andersen JS, Mann M (2002) Microcolumns with self-assembled peptide for proteomics. J Chromatogr A 979: 233–239.

Ishihama Y, Rappsilber J, Mann M (2006) Modular stop and go extraction tips with stacked disks for parallel and multidimensional peptide fractionation in proteomics. J Proteome Res 5: 988–994.

Jones JD, Dangl JL (2006) The plant immune system. Nature 444: 323–329.

Kersten B, Agrawal GK, Durek P, Neigenfind J, Schulze W, Walther D, Rakwal R (2009) Plant phosphoproteomics: an update. Proteomics 9: 964–988.

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al. (2003) Collection, mapping, and annotation of over 20,000 cDNA clones from japonica rice. Science 301: 376–379.

Li L, Stocekct CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178–2189.

Maclean D, Burrell MA, Studholme DJ, Jones AM (2008) PhosCalc: a tool for evaluating the sites of peptide phosphorylation from mass spectrometer data. BMC Res Notes 1: 30.

Mann M, Wilm M (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem 66: 4390–4399.

Monosi B, Wissel RJ, Pennill L, Hulbert SH (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem 66: 4390–4399.

Maor R, Jones A, Nuhse TS, Studholme DJ, Peck SC, Shirasu K (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics 6: 601–610.

Nakagami H, Soukupova H, Schikora A, Zarsky V, Hirt H (2006) A mitogen-activated protein kinase kinase mediates reactive oxygen species homeostasis in Arabidopsis. J Biol Chem 281: 38697–38704.

Nuhse TS, Stensballe A, Jensen AM, Peck SC (2006) LC/MS sample pretreatment in proteomics. Anal Chem 78: 759–775.

Reiland S, Messerli G, Baerenfeller K, Gerrits B, Endler A, Grossmann J, Grussme W, Ragnisky S (2009) Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. Plant Physiol 150: 889–903.

Rensink WA, Buell CR (2004) Arabidopsis to rice: applying knowledge from a weed to enhance our understanding of a crop species. Plant Physiol 135: 622–629.

Saito H, Oda Y, Sato T, Kurimoto S, Ishihama Y (2006) Multiplexed two-dimensional liquid chromatography for MALDI and nanoelectrospray ionization mass spectrometry in proteomics. J Proteome Res 5: 1803–1807.

Shirasu K (2009) The HSP90-SGT1 chaperone complex for NLR immune sensors. Annu Rev Plant Biol 60: 139–164.

Somerville C, Koornneef M (2002) A fortunate choice: the history of Arabidopsis as a model plant. Nat Rev Genet 3: 883–889.

Suarez-Rodriguez MC, Adams-Philpills L, Liu Y, Wang H, Su SH, Lester PJ, Zhang S, Bent AF, Krysan PJ (2007) MEKK1 is required for flag1-induced MPK4 activation in Arabidopsis plants. Plant Physiol 143: 661–669.

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

Swayne DL, Wenger CD, Thomson JA, Coon JJ (2009) Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. Proc Natl Acad Sci USA 106: 995–1000.

Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci USA 101: 9528–9533.

Takken FL, Albrecht M, Tameling WI (2006) Resistance proteins: molecular switches of plant defence. Curr Opin Plant Biol 9: 383–390.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

Tornero P, Chao RA, Luthin WN, Goff SA, Dangl JL (2002) Large-scale structure-function analysis of the Arabidopsis RPM1 disease resistance protein. Plant Cell 14: 433–450.

Trost M, English L, Lemieux S, Courcelles M, Desjardins M, Thibault P (2009) The phagosomal proteome in interferon-gamma-activated macrophages. Immunity 30: 143–154.

Tzfira T, Vaidya M, Citovsky V (2001) VIP1, an Arabidopsis protein that interacts with Agrobacterium VirE2, is involved in VirE2 nuclear import. Proc Natl Acad Sci USA 98: 139–164.

Wang Y, Klemke RL (2008) PhosphoBlast, a computational tool for comparing phosphoprotein signatures among large datasets. Mol Cell Proteomics 7: 145–162.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2: a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189–1191.

Zhang X, Ye J, Jensen ON, Roepstorff P (2007) Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment. Mol Cell Proteomics 6: 2032–2042.