

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/297403333>

PlantPReS: A database for plant proteome response to stress

Article in *Journal of Proteomics* · March 2016

DOI: 10.1016/j.jprot.2016.03.009

CITATIONS

3

READS

125

10 authors, including:



Mohammad Reza Ghaffari

Agricultural Biotechnology Research Institute o...

44 PUBLICATIONS 269 CITATIONS

SEE PROFILE



Mehdi Mirzaei

Macquarie University

61 PUBLICATIONS 731 CITATIONS

SEE PROFILE



Mehdi Alikhani

Royan Institute

20 PUBLICATIONS 103 CITATIONS

SEE PROFILE



Paul A Haynes

Macquarie University

197 PUBLICATIONS 5,676 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Gastric cancer biomarkers [View project](#)



Transdifferentiation [View project](#)

All content following this page was uploaded by [Seyed Ahmad Mousavi](#) on 26 June 2016.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

Accepted Manuscript

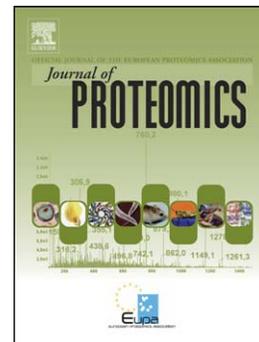
PlantPRoS: A database for plant proteome response to stress

Seyed Ahmad Mousavi, Farhad Movahedi Pouya, Mohammad Reza Ghaffari, Mehdi Mirzaei, Akram Ghaffari, Mehdi Alikhani, Mohammad Ghareyazie, Setsuko Komatsu, Paul A. Haynes, Ghasem Hosseini Salekdeh

PII: S1874-3919(16)30063-X
DOI: doi: [10.1016/j.jprot.2016.03.009](https://doi.org/10.1016/j.jprot.2016.03.009)
Reference: JPROT 2448

To appear in: *Journal of Proteomics*

Received date: 16 January 2016
Revised date: 23 February 2016
Accepted date: 2 March 2016



Please cite this article as: Mousavi Seyed Ahmad, Pouya Farhad Movahedi, Ghaffari Mohammad Reza, Mirzaei Mehdi, Ghaffari Akram, Alikhani Mehdi, Ghareyazie Mohammad, Komatsu Setsuko, Haynes Paul A., Salekdeh Ghasem Hosseini, PlantPRoS: A database for plant proteome response to stress, *Journal of Proteomics* (2016), doi: [10.1016/j.jprot.2016.03.009](https://doi.org/10.1016/j.jprot.2016.03.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

PlantPreS: a database for plant proteome response to stress

Seyed Ahmad Mousavi^{1, 4}, Farhad Movahedi Pouya¹, Mohammad Reza Ghaffari¹, Mehdi Mirzaei², Akram Ghaffari¹, Mehdi Alikhani⁴, Mohammad Ghareyazie³, Setsuko Komatsu⁵, Paul A. Haynes², Ghasem Hosseini Salekdeh^{1, 4*}

¹Agricultural Biotechnology Research Institute of Iran, Karaj, Tehran, Iran

² Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW, Australia

³ Computer Science Department, University of California, Davis, Davis, CA

⁴Department of Molecular Systems Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran

⁵National Institute of Crop Science, Tsukuba, Ibaraki 305-8518, Japan

Corresponding author:

Ghasem Hosseini Salekdeh, Agricultural Biotechnology Research Institute of Iran, P.O. Box 31535-1897, Karaj, Iran; Tel.: +98-26-32700845; Fax: +98-26-32704539; Email: h_salekdeh@abrii.ac.ir

Abstract

About 75% of plant yield potential has been estimated to be lost to environmental stresses, even in developed agricultures. To facilitate the biotechnological improvement of crop productivity, genes and proteins that control crop adaptation to a wide range of environments will need to be identified. Due to the challenges faced in text/data mining, there is a large gap between the data available to researchers and the hundreds of published plant stress proteomics articles. Plant stress proteome database (PlantPReS; www.proteome.ir) is an open online proteomic database, which currently (as of October 2015) comprises > 20413 entries from 456 manually curated articles, and contains >10600 unique stress responsive proteins. Since every aspect of the experiments, including protein name, accession number, plant type, tissue, stress types, organelles, and developmental stage has been digitized, experimental data can be rapidly accessed and integrated. Furthermore, PlantPReS enables researchers to perform multiple analyses on the database using the filtration mode, and the results of each query indicate a series of proteins for which a set of selected criteria are met. The query results can be displayed in either text or graphical format.

Key words: Database; Plant Stress; Proteomics

Introduction

It is estimated that only 10% of the world's arable land may be categorized as free from stress [1]. The impact of environmental stress on crop productivity varies with the source; 26%, 20% and 15% of arable lands are under drought, mineral and freezing stresses, respectively. Even in developed agricultures, about 75% of yield potential has been calculated to be lost to environmental stresses [2]. By 2030, we will need to double world food production in order to feed eight billion mouths [3]. This will not be achieved unless new varieties of food crops are bred that will improve yields both on marginal lands and those currently abandoned. To meet these challenges, proteins that regulate crop adaptation to a wide range of environments will need to be identified to facilitate the biotechnological improvement of crop productivity. Proteomics has proved to be a powerful approach for discovery of proteins and pathways that may eventually be used to improve crop yield under stress conditions. It is well suited to hypothesis-limited research because it offers an entry point for identifying possibly significant changes in plant protein levels against a background of unresponsive proteins. The application of proteomics in crop breeding is usually initiated by detection of stress responsive proteins after comparison between stressed and control plants. Identification of these differentially expressed candidate proteins may then reveal that some have functions clearly consistent with the stress tolerance trait. So far, hundreds of plant stress proteomics papers have been published which provide a wealth of data on plant responses to stress at the proteome level [4-7].

Data mining is an essential step in drilling down to fundamentals of research, allowing the data to be easily and fully understandable to plant biologists, and to be integrated in breeding programs. However, the promise of text and data mining to facilitate and enhance research fundamentally has not yet been achieved, mainly because great numbers of stress-associated

proteins are not deposited in databases. Although there are several databases containing plant stress information such as RiceSRTFDB [8], QlicRice [9], PRGdb [10], STIFDB2 [11], PSRGC [12], MCDRP [13], and PSPDB [14], these databases do not provide sufficient information about stress responsive proteins available in literature. The only plant stress protein database, PSPDB, retrieved proteins from UniProt [15] through gene ontology term search for the various plant stress-related terms. As of last update, PSPDB hosted 2,064 manually curated plant stress proteins [14]. However, this database does not provide comprehensive information on stress responsive information available in literature, and lacks important information such as the expression pattern of proteins under various environmental conditions. As a complement to these databases, we have developed Plant Stress Proteome Database (PlantPreS; www.proteome.ir). PlantPreS is a manually curated database and comprises > 20413 entries from 456 manually curated articles which contain >10600 unique proteins (as of October 2015). This database has a user-friendly interface with a number of useful features, including a search engine, analysis tools, gene ontology, and a function for cross-referencing useful external databases. The expression pattern of proteins have been manually collected and curated from proteomics papers or their supplementary data. Since plant stress responsive proteins have been published with a wide range of accession numbers during the last two decades, PlantPreS converts entries into sequences followed by BLAST search against the whole database. This also allows extending our search to orthologs of our proteins of interest. PlantPreS is a valuable database for the vast majority of researchers working in proteomics and plant stress areas because it helps to make sense of the overwhelming amount of information available in literature.

Data collection and content

We searched NCBI PubMed and Google scholar with keywords including plant, crop, stress, abiotic, biotic, proteome and proteomics. We collected 456 relevant literature articles manually and read through the full text, tables, and supplementary data of each article to identify stress responsive proteins. Once a non-redundant list of proteins was extracted from the literature further information regarding the protein was derived and integrated into protein information. For data extraction, we defined a reference table which include common features from each article such as protein accession identifier, spot number, plant organism, protein match organism, tissue, sub-tissue, organelle, stress, stress details, genotype, expression pattern and sampling stage. Furthermore, we converted all protein accessions into the UniProtKB accession to achieve a uniform accession type. The corresponding sequence of each accession number was also included in reference table. The PlantPreS website was implemented using ASP.NET by Visual Studio 2010 (.NET Framework 3.5) on Parallel Plesk (version 11) server. The programming language used was C# and SQL Server 2008 R2 was used as the database management system, database schematic structure is shown in Fig. 1.

An overview of PlantPreS statistics is presented in Fig 2. This includes > 20413 entries from 456 manually curated articles which contain >10600 unique proteins. All entries were classified based on plant species, stresses imposed and tissues and organelles collected from each study. From collected data, we listed 95 different stresses (53 biotic and 42 abiotic), 112 plant species, 26 tissues and 12 organelles. The most frequent plant entries were belong to *Oryza sativa*, *Triticum aestivum*, *Arabidopsis thaliana*, *Glycine max* and *Zea mays* (Fig. 2A). In addition, drought, salinity, cold, and oxidative along with fungus were major classes of abiotic and biotic stress proteomics studied respectively (Fig. 2B and 2C).

Data access

The data in PlantPreS can be easily accessed by different means. In ‘General Search’, users can query the database by ‘protein name’ or ‘protein accession’, ‘plant’, and ‘stress’ types, which results in display of protein-centered information in the same page (Fig. 3A). The protein search result provides the following information: protein ID (UniProt), protein name, spot number (if available), organism, tissue, expression pattern, and also link to references that validate the response of the protein to stress. The outputs can be illustrated as pie charts based on plant, tissue, stress and organelle types (Fig. 3B). Furthermore, clicking on each protein ID in the result page leads to protein information including protein name, gene name, organism, stress, expression pattern, sequence, general ontologies, and link to other databases such as UniProtKB, NCBI gene [16], RefSeq [17], KEGG [18], Pfam [19], PRIDE [20] and Protein Model Portal [21]. In ‘protein accession’ search, the accession number (UniProtKB accession, NCBI GI numbers, RefSeq accession numbers, and Entrez Gene IDs) of a query is converted to sequence before being searched against the sequence database. This approach allows the identification of homologous proteins for any given query. The percent identity of query with PlantPreS proteins will also be displayed. If a similar protein is not found, it is highly recommended to search the sequence of query in PlantPreS database using ‘sequence search’ tool. A customized BLAST (NCBI version 2.2.28+) [22] tool has been made available that searches a user-defined query against the sequences available in the database. It may be useful in characterization of stress responsive sequences and retrieval of homologous sequences from the database, based on sequence similarity.

A filter can be applied to restrict the output based on the similarity of query to the entries in the PlantPreS database. Additionally, the sequence search tool is capable of searching up to 20 different protein identifiers simultaneously. The ‘article search’ makes it possible to search by article title, author name or PubMed ID; this option allows user to have a glimpse of the article details.

For each search, up to 500 proteins (entries) are generated as a list view in the browser. It is also possible to download the complete output file in a tab-delimited format by clicking on the corresponding link (Fig. 3C). For ‘batch search’, an excel file containing one sheet per protein will be generated. PlantPreS provides another way to view and retrieve all stress responsive proteins in which all the entries are accessible to the researchers for a multifactorial analysis. In the “database analysis” section, one may analyze the database for a specific plant species, stress, tissue, organelle and expression pattern, or a combination of these five factors. The results can be presented as a list in the browser (for less than 500 entries) or as a link to the output file.

Discussion and future perspectives

Although over the years a large number of plant databases have been published, yet, to our knowledge, there is no resource available that provides detailed information about proteins associated with stress. Our database is the first manually curated database for plant proteomics articles. Although manually extracting data from articles might be a time consuming and tedious task for administrators, it provides a wealth of information at the fingertips of plant researchers. PlantPreS is the first expandable database which provides the expression pattern of stress responsive proteins extracted from more than 400 publications. The researchers can search the database in different ways such as general, sequence and article search. One of the major

challenges in manual data curation and annotation of plant proteome data was differences among protein accession numbers used in different articles for similar proteins. To deal with this issue, we converted all protein accessions into the UniProtKB accession to achieve a uniform accession type. This also provides a link to additional information for each protein such as gene ontology and protein modifications. Furthermore, a customized BLAST tool has been made available that is very useful in characterization of stress responsive sequences and retrieval of homologous sequences from the database, based on sequence similarity. This is particularly helpful when no result is retrieved by 'general search'. We intend to increase the number of protein entries in the future and expand the database as new studies are published. We believe that our database will be useful to many researchers in the field, and have an impact in plant biology, especially in cereal crops.

Acknowledgment

This project has been supported by Agricultural Biotechnology Research Institute of Iran and Iranian National Science Foundation.

References

- [1] Dudal R. Inventory of the major soils of the world with special reference to mineral stress hazards. Plant adaptation to mineral stress in problem soils Proceedings of a workshop held at the National Agricultural Library, Beltsville, Maryland, November 22-23, 1976: Cornell Univ. Agricultural Experiment Station; 1976. p. 3-13.
- [2] Boyer JS. Plant productivity and environment. *Science*. 1982;218:443-8.
- [3] Lindskog P. How to cut world hunger in half. *Science (New York, NY)*. 2005;310:1768.
- [4] Salekdeh GH, Komatsu S. Crop proteomics: aim at sustainable agriculture of tomorrow. *Proteomics*. 2007;7:2976-96.
- [5] Kosová K, Vítámvás P, Prášil IT, Renaut J. Plant proteome changes under abiotic stress—contribution of proteomics studies to understanding plant stress response. *Journal of proteomics*. 2011;74:1301-22.
- [6] Barkla BJ, Castellanos-Cervantes T, de Leon JL, Matros A, Mock HP, Perez-Alfocea F, et al. Elucidation of salt stress defense and tolerance mechanisms of crop plants using proteomics--current achievements and perspectives. *Proteomics*. 2013;13:1885-900.
- [7] Chakraborty S, Salekdeh GH, Yang P, Woo SH, Chin CF, Gehring C, et al. Proteomics of Important Food Crops in the Asia Oceania Region: Current Status and Future Perspectives. *Journal of proteome research*. 2015;14:2723-44.
- [8] Priya P, Jain M. RiceSRTFDB: a database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. *Database (Oxford)*. 2013;2013:bat027.
- [9] Smita S, Lenka SK, Katiyar A, Jaiswal P, Preece J, Bansal KC. QlicRice: a web interface for abiotic stress responsive QTL and loci interaction channels in rice. *Database (Oxford)*. 2011;2011:bar037.
- [10] Sanseverino W, Hermoso A, D'Alessandro R, Vlasova A, Andolfo G, Frusciante L, et al. PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucleic Acids Res*. 2013;41:D1167-71.
- [11] Naika M, Shameer K, Mathew OK, Gowda R, Sowdhamini R. STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. *Plant Cell Physiol*. 2013;54:e8.

- [12] [Wanchana S, Thongjuea S, Ulat VJ, Anacleto M, Mauleon R, Conte M, et al. The Generation Challenge Programme comparative plant stress-responsive gene catalogue. Nucleic Acids Res. 2008;36:D943-6.](#)
- [13] [Gour P, Garg P, Jain R, Joseph SV, Tyagi AK, Raghuvanshi S. Manually curated database of rice proteins. Nucleic acids research. 2014;42:D1214-D21.](#)
- [14] [Anil Kumar S, Hima Kumari P, Sundararajan V, Suravajhala P, Kanagasabai R, Kavi Kishor PB. PSPDB: Plant Stress Protein Database. Plant Molecular Biology Reporter. 2014;32:940-2.](#)
- [15] [Activities at the Universal Protein Resource \(UniProt\). Nucleic Acids Res. 2014;42:D191-8.](#)
- [16] [Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2011;39:D52-7.](#)
- [17] [Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences \(RefSeq\): current status, new features and genome annotation policy. Nucleic Acids Res. 2012;40:D130-5.](#)
- [18] [Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42:D199-205.](#)
- [19] [Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222-30.](#)
- [20] [Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDentifications \(PRIDE\) database and associated tools: status in 2013. Nucleic Acids Res. 2013;41:D1063-9.](#)
- [21] [Haas J, Roth S, Arnold K, Kiefer F, Schmidt T, Bordoli L, et al. The Protein Model Portal--a comprehensive resource for protein structure and model information. Database \(Oxford\). 2013;2013:bat031.](#)
- [22] [Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.](#)

Figure Legends

Figure 1. Schematic representation illustrating the PlantPreS data structure, curation and navigation processes

Figure 2. Overall distributions of the plants, Abiotic and biotic and stresses in PlantPreS. **(A)** Distribution of the 10 most frequent plants in the PlantPreS. **(B)** Distributing of the 10 most frequent abiotic stresses and biotic stresses **(C)** in the PlantPreS.

Figure 3. Screen shot example of PlantPreS showing different search tools and outputs. **(A)** Search panel and home page of database **(B)** Searching results page with several pie charts for plant analysis, tissue analysis, organelle analysis, stress analysis. **(C)** Generating records in a tab-delimited file format.

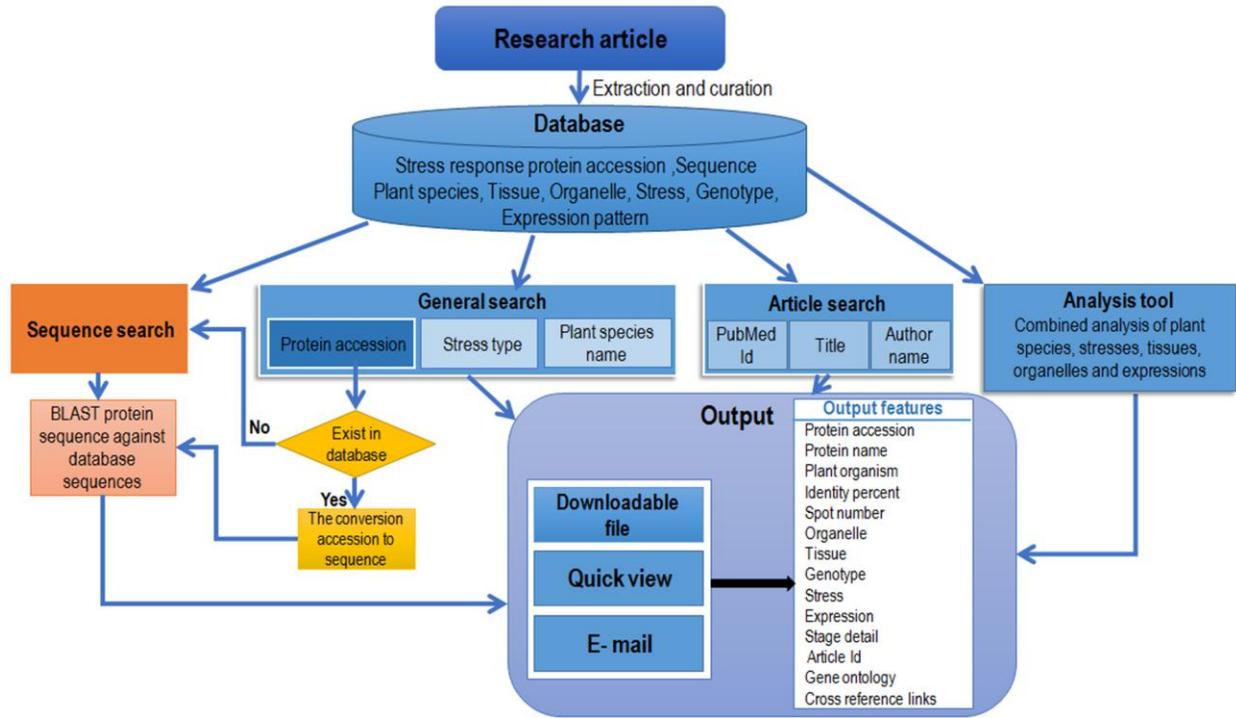


Figure 1

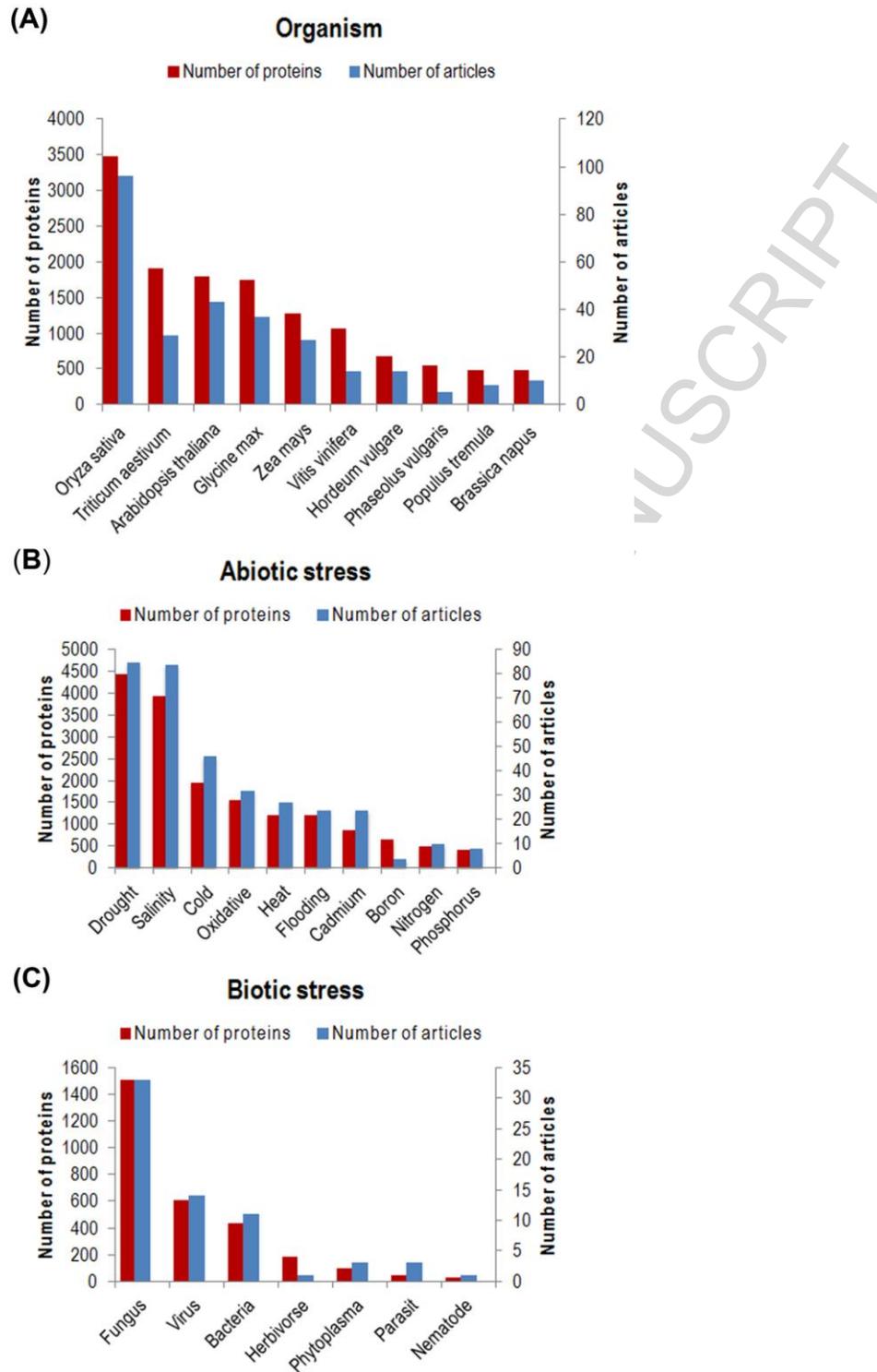


Figure 2

(A) **PlantPreS**

Plant Proteome Response to Stress

Home | General search | Sequence search | Article search | Database analysis | About us | Contact us | Help

Queries information

Please choose one of these options: Protein, Stress or Plant.
Acceptable queries:

- > **Protein:** UniProt accession, GI number, Protein Refseq, Gene Id, Protein name or Gene name
- > **Stress:** Type of stress i.e. "Cold" or "Drought"
- > **Plant:** Plant name i.e. "Rice" or "Arabidopsis thaliana"

> **Batch protein:** Only UniProt accession or GI number are acceptable. The delimiter is ',' comma.

> Result with more than 500 records will not be shown in the browser and can only be downloaded.

> **Database statistics**

> Last update: Oct-09-2015

> Number of articles: 456

> Number of entries: 20413

General search

Q24182

Email for sending result (optional)

Job name (optional)

Example: [Protein](#) [Batch protein](#) [Stress](#) [Plant](#)

Protein (search using blast) Stress Plant

Search by Enter any type of Uniprotid or GI No. or Protein name or Gene name or NCBI GeneID or Refseq Id.

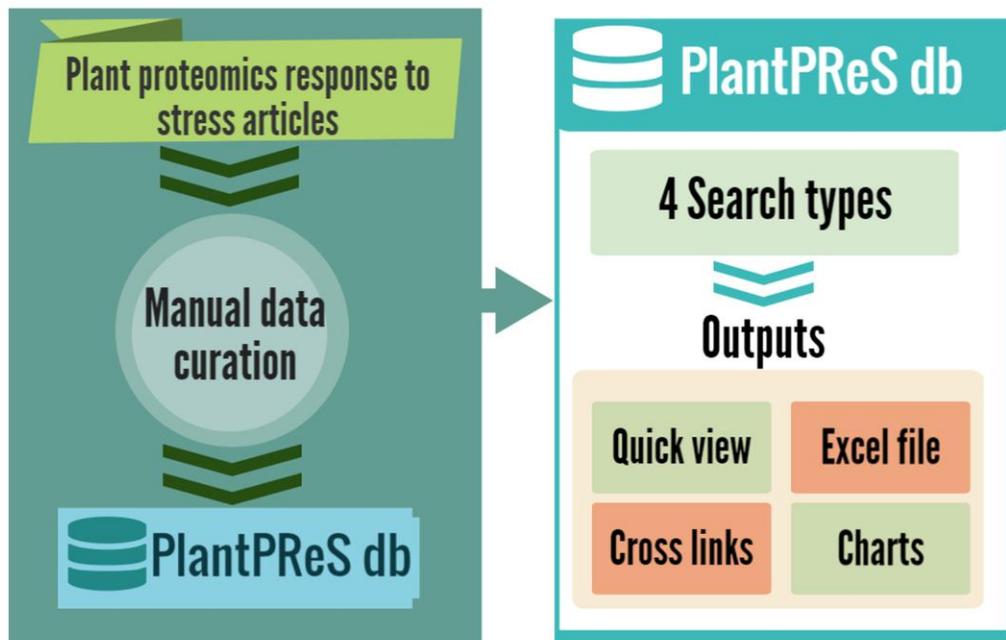
Identity percent (Just for protein) All result Identity >=90% Identity >=80% Identity >=70%

[Plant analysis](#)
[Tissue analysis](#)
[Organelle analysis](#)
[Stress analysis \(B\)](#)

417 number of founded [Download full result as .txt Tab delimited file format \(You can open it with spreadsheet program\)](#) (C)

Protein ID	Protein name	Percent identity	Spot Number	Protein source organism	Tissue	Organelle	Stress	Expression	Article Id
Q24182	Endosperm luminal binding protein	100%	35	Rice (Oryza sativa)	Leaf	Cell	Heat	Up regulated	19819927
Q24182	Endosperm luminal binding protein	100%	17	Rice (Oryza sativa)	Leaf	Cell	Rhizoctonia solani	Up and Down regulated	20507456
Q24182	Endosperm luminal binding protein	100%	70	Rice (Oryza sativa)	Leaf	Cell	Cadmium	Up regulated	21117708
Q24182	Endosperm luminal binding protein	100%	114	Rice (Oryza sativa)	Leaf	Cell	Oxidative	Down regulated	18407957
Q24182	Endosperm luminal binding protein	100%	18	Rice (Oryza sativa)	Leaf	Cell	Oxidative	Down regulated	21172955

Figure 3



Graphical Abstract

Significance

The promise of text and data mining to facilitate and enhance research fundamentally has not yet been achieved, mainly because great numbers of stress-associated proteins are not deposited in databases. PlantPReS is a valuable database for the vast majority of researchers working in proteomics and plant stress areas. It has a user-friendly interface with a number of useful features, including a search engine, analysis tools, gene ontology, a function for cross-referencing useful external databases, and the expression pattern of stress associated proteins.

ACCEPTED MANUSCRIPT

Highlights

- Plant stress proteome database is an open online proteomic database
- It comprises > 20413 entries from 456 manually curated articles
- It contains >10600 unique stress responsive proteins
- PlantPreS enables researchers to perform multiple analyses on the database

ACCEPTED MANUSCRIPT