

DBPR: DataBase of Plant Research

Shahid Ullah (✉ dridechina@gmail.com)

The Seventh Affiliated Hospital Sun Yat-sen University <https://orcid.org/0000-0001-6694-5590>

Wajeeha Rahman

S-Khan Lab Mardan KPK PAKISTAN

Farhan Ullah

S-Khan Lab Mardan KPK Pakistan

Gulzar Ahmad

S-Khan Lab Mardan KPK Pakistan

Muhammad Ijaz

S-Khan Lab Mardan KPK Pakistan

Tianshun Gao

The Seventh Affiliated Hospital Sun Yat-sen University

Research Article

Keywords: habdsk, DBPR, Plant database, Protein, DNA

Posted Date: February 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-153167/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Databases of plant have been an integral part of modern biology. Enormous quantities of data are produced from plants and improving with time, many databases have been published with large amounts of data in various fields, including proteomics, transcriptomics, genomics, and metabolomics. To make it easier, a comprehensive database is needed to sort and organized all these data to one platform and give an easy and friendly finding way to plant research community, because, databases are a standard strategy for analyzing, storing, and processing such big data. Therefore, we have compiled 225 plant databases on a single platform and grouped them into 5 categories on the basis of various typescripts, such as the Protein, DNA, RNA, Pathway and Expression database, which can be indexed by clicking on the name of the category or by clicking on the picture expression or directly searching in the specified search bar in the database. DBPR is a comprehensive plant resource database that is available on <https://www.habdisk.org/dbpr.php> . Computer platform such as PHP, HTML, CSS, and JavaScript has been used to build and will be updated timely.

Introduction

The fastest growing level of plant data sets over the past few decades is likely due to the application of next-generation sequencing as well as mass-spectrometry technologies to experimental model and plants research[1]. The change in the volume and complexity of Omics (proteomics, transcriptomics, genomics, metabolomics) data creates challenges, mostly related to the collection, storage, and data sharing within the scientific community[2]. The increase in data collection is limiting the existing computational connectivity and formal bioinformatics algorithms, which have become insufficient to accommodate gigantic-data inputs and analyzes. Several major challenges, such as the collection, analysis, visualizing, and storage of data sets, need to be tackled by modifying existing structures and creating new tools and databases for fast and efficient processing of big data. Many articles have been published in well know journals [3–6] (Tab: 1), which have collected the databases of different organisms and research area, e.g. “Biological databases for human research[7]” have collected 74 human databases which have been published in Genomics Proteomics & Bioinformatics (GPB), “Online Databases for Taxonomy and Identification of Pathogenic Fungi and Proposal for a Cloud-Based Dynamic Data Network Platform[8]” have collected 24 fungi databases and published in the journal of clinical microbiology, so that a well comprehensive plant database is also needed for the plant research community to sort and save all the plant data for future researchers, cause database of plants has been an integral part of modern biology. Enormous quantities of data are produced from plants[9] [10] [11] [12], such as protein functions in particularly sequences, MPIM database [13], P3DB [14], plant RNA database and website incorporates knowledge from numerous independent computer-assisted reaches and databases such as “PsnoRNA database” [15], “PceRBase database” [16] and “CSRDB” [17], The Pathway database is a database of biochemical pathways for metabolic, signaling, reaction and control[18, 19]. eg, “MetaCrop”[12], “PLaMoMdb” [20] while plant DNA database have genomic details of different plants, such as PLAZA [21], “Planteome” [22], “AtGDB” [23], According to such huge research[24], we have collected and

integrated 225 far more popular and accessible plant databases from the most recent published lectures and created a well-known plant database (DBPR) that will be a convenient and friendly forum for the scientific community. Further, we have divided the plant database into five categories: DNA, RNA, Protein, Expression, and Pathway databases, which have three forms of searching option, browse by name or image expression, or search by name in the search bar. According to the published databases research work, many databases have been noticed in the different research area[6], which have been provided the latest database in the form of table[3–6]. Other hands, to make it's easier and clearer to plant researchers we have provided the plant databases table as well as database, which is highlighted in the (Tab: 1) and will be updated over time.

Result And Discussion

Construction and content of the DBPR

In this work, we have used many keywords, such as plant database, a database of plant, biological databases, etc. in a variety of search engines, like PubMed, (<https://pubmed.ncbi.nlm.nih.gov/>) Google, (<https://www.google.com/>) Google Scholar (<https://scholar.google.com/>) and have manually collected the plant database from published research work and journals such as NAR and database journals. Computational platforms PHP, HTML, MySQL, CSS, and JavaScript have been used to delete all broken (dead) links, making a database and organize the data. Finally, have provided a comprehensive collection of plant databases to one platform named “DBPR” which can be friendly operate, have updated data and will update data in a timely manner (Fig: 1).

Classification of Plant databases

The numbers of databases are growing at a very high speed due to the ongoing research and technology in all the research fields[24], some of the omics databases are available on the “National Center for Biotechnology Information (NCBI) [9] in different research area. for more easy access we have collected Plant databases and divided them into many classes, which are given below.

Plant DNA database

Database of plant DNA is a database in which genomic details of different plants are available in various databases, such as “PLAZA” [21], “Planteome” [22], “AtGDB” [23], etc. It's tough to get the knowledge you need regarding plant genomes because reported research is scattered through a large range of publications and the majority are undisclosed, Database brings together information from smaller databases and literature [25]. And we've compiled the strongest information on plant genomes to make the work simpler for the researcher.

Plant RNA database

The Plant RNA database and web-site incorporate knowledge from numerous independent computer-assisted searches and databases such as “PsnRNA database” [15], “PceRBase database” [16] and

“CSRDB” [17], etc. The RNA database was used to classify different plant species, and such sequences are used in the database as alignments. The database ultimately establishes a unifying nomenclature for all other RNAs i.e Small nucleolar RNAs, processing and modification of other RNAs, such as ribosomal and small nuclear spliceosomal RNAs, SnoRNAs are a large family of relatively well characterized non-coding RNAs (ncRNAs)[26].

Plant protein database

Plant protein databases have become an important part of contemporary biology. For plant protein structures, functions, and especially sequences, enormous amounts of data are generated. In the study of a new protein, database queries are always the first step. Comparison between proteins or between groups of proteins offers knowledge about the interaction between proteins inside or through genomes or across species, and therefore provides much greater knowledge than can be gained from researching a single protein alone[27]. In a fact, there are also commonly accessible secondary sources originating from sample datasets. Such repositories restructure and annotate the data or offer predictions. Using several repositories can also help researchers consider a protein's structure and work, although certain plant protein databases are well recognized such as MPIM database[13], P3DB[14], etc. they are far from being completely used in the field of protein sciences. our database gives readers a starting point for discovering the value of online plant protein repositories.

Plant expression database

Plant expression database is a publicly accessible co-expressed gene sets database that will be a powerful resource for a broad spectrum of experimental projects, including targeting genes for functional detection of regulatory work e.g, “SoyNet database”[28], “BarleyBase”[29]. Here we report the construction of the Arabidopsis thaliana trans-factor and cis-element prediction database (ATTED-II) that provides co-regulated, co-expressed gene relationships, On co-expressed genes resulting from microarray data and the cis elements expected[30].

Plant pathway database

The Plant Pathway database or Reactome database is a database of biochemical pathways for metabolic, signaling, reaction and regulation[19]. Such as “MetaCrop” [31], “PathoPlant” [32], “PLaMoMdb” [20]. These data bases also provide bioinformatics tools for the researchers to anticipate and analyze biological and biochemical pathways. Examples of biological pathways in Reactome include classical intermediary metabolism, signaling, transcriptional regulation, apoptosis and disease. The plant pathways databases accommodate the distinct type of reactions that is available for computational analysis.

Statistics of the DBPR

With the rapid growth of plant databases, we tried to gather all online plant databases to one database and give easy access to all plant researchers, therefore, we have provided access to 225 redundant databases (Tab-S1). We have collected all category-wide and year-wise database and have classified into

5 categories which shown in the Fig-2A. further, the category-wise growth of the plant databases shows the big difference of each and every category from the last decade of the plant research Fig-2B. while the year-wise growth gives clear figures of the improvement of plant database research with the passage of time Fig: 2C.

Usages of the DBPR Database

To provide useful and updated plant research the DBPR is developed in an easy and friendly searching way. For fast search our database gives 3 options to accesses the data, user can direct click on the categories name or by the image expression icon (Fig. 3A), which will give all the category databases on single click (Fig. 3B), on further clicking user can get their needed information. For searching single database user can type the name of the required database in the navigation bar, which is on the top of the database (Fig. 3C), we have highlighted MMP database as an example.

Conclusion

Plant Databases have been a standard way of sorting and analyzing vast volumes of knowledge in various fields of plant research, on government and private sectors. Many plant databases have been published in different research areas which have their own scope and value, according to that published work we have collected all plant databases to one platform and provided an easy way to the plant science community. All the plant databases are divided into 5 categories according to their external and internal function, each and every class has its own databases, such as DNA databases, RNA database, Protein database, pathway database, and Expression database. To avoid wasting time we have provided a list as well as a database of the databases named DBPR, which can be searched in 3 ways, first users can search by clicking the name of the category, image expression, or can directly type the name of the needed database in the search bar which is on the top of the navigation bar. Computer programs, PHP, HTML, MySQL JavaScript, and CSS have been used to build DBPR and will be updated timely.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

These data will be available under the journal rule and regulation

Competing interests

The anthers don't have any compete of interests.

Funding

This project is supported by Shenzhen's introduction of talents and research start-up (392020)

Authors' contributions

Dr. Shahid Ullah and Dr. Tianshun Gao supervised the project. Wajeeha Rahman, Mr. Gulzar Ahmad, Mr. Farhan Ullah, Mr Muhammad Ijaz collected and verified the data carefully, Dr shahid Ullah have wrote the manuscript, all authors reviewed the manuscript and agreed to submit.

Acknowledgement

To avoid future conflict and plagiarism issue, DBPR database is uploaded on (<http://habdsk.org/dbpr.php>) so that we have provided some content in this article.

References

1. Rhee SY, Crosby B, *Biological databases for plant research*. 2005, Am Soc Plant Biol.
2. Ullah S, et al. EDBCO-19: Emergency Data Base of COVID-19. *J Clin Med Res*. 2020;2(4):1–4.
3. Xu D, Xu Y. *Protein databases on the internet*. Current protocols in molecular biology, 2004. 68(1): p. 19.4. 1-19.4. 15.
4. Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. *Nucleic acids research*. 2006;34(suppl_1):D504–6.
5. Harper R. Access to DNA and protein databases on the Internet. *Curr Opin Biotechnol*. 1994;5(1):4–18.
6. Rigden DJ, Fernández XM. The 27th annual Nucleic Acids Research database issue and molecular biology database collection. *Nucleic Acids Res*. 2020;48(D1):D1–8.
7. Zou D, et al. Biological databases for human research. *Genom Proteom Bioinform*. 2015;13(1):55–63.
8. Prakash PY, et al. Online databases for taxonomy and identification of pathogenic fungi and proposal for a cloud-based dynamic data network platform. *J Clin Microbiol*. 2017;55(4):1011–24.
9. Coordinators NR. Database resources of the national center for biotechnology information. *Nucleic acids research*. 2018;46(Database issue):D8.
10. Consortium U. UniProt: the universal protein knowledgebase. *Nucleic acids research*. 2018;46(5):2699.
11. Spannagl M, et al. PGSB PlantsDB: updates to the database framework for comparative plant genome research. *Nucleic acids research*. 2016;44(D1):D1141–7.

12. Caspi R, et al. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic acids research*. 2007;36(suppl_1):D623–31.
13. Orsini M, et al., *Amyotrophic lateral sclerosis: new perspectives and update*. *Neurology international*, 2015. 7(2).
14. Gao J, et al. P3DB: a plant protein phosphorylation database. *Nucleic acids research*. 2009;37(suppl_1):D960–2.
15. Brown JW, et al. Plant snoRNA database. *Nucleic acids research*. 2003;31(1):432–5.
16. Yuan C, et al. PceRBase: a database of plant competing endogenous RNA. *Nucleic acids research*. 2017;45(D1):D1009–14.
17. Johnson C, et al. CSRDB: a small RNA integrated database and browser resource for cereals. *Nucleic Acids Res*. 2007;35(suppl_1):D829–33.
18. Wheeler DL, et al., *Database resources of the National Center for Biotechnology Information*. *Nucleic Acids Res*. 2005.
19. Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic acids research*. 2018;46(D1):D649–55.
20. Guan D, et al., *PlaMoM: a comprehensive database compiles plant mobile macromolecules*. *Nucleic acids research*, 2016: p. gkw988.
21. Vandepoele K, et al. pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environ Microbiol*. 2013;15(8):2147–53.
22. Cooper L, et al. The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic acids research*. 2018;46(D1):D1168–80.
23. Dong Q, Schlueter SD, Brendel V. PlantGDB, plant genome database and analysis tools. *Nucleic acids research*. 2004;32(suppl_1):D354–9.
24. KhanSA JM, et al. Tremendous Contribution of Dr. Shahid Ullah to Scientific Community during COVID-19 Pandemic in the Form of Scientific Research. *J Clin Med Res*. 2020;2(5):1–7.
25. Garcia S, et al. Recent updates and developments to plant genome size databases. *Nucleic acids research*. 2014;42(D1):D1159–66.
26. Leader DJ, et al. Splicing-independent processing of plant box C/D and box H/ACA small nucleolar RNAs. *Plant molecular biology*. 1999;39(6):1091–100.
27. Xu D. *Protein databases on the internet*. *Current protocols in protein science*, 2012. 70(1): p. 2.6. 1-2.6. 17.
28. Kim E, Hwang S, Lee I. SoyNet: a database of co-functional networks for soybean *Glycine max*. *Nucleic acids research*. 2017;45(D1):D1082–9.
29. Shen L, et al. BarleyBase—an expression profiling database for plant genomics. *Nucleic Acids Res*. 2005;33(suppl_1):D614–8.
30. Obayashi T, et al. ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. *Plant Cell Physiol*. 2018;59(1):e3–3.

31. Grafahrend-Belau E, et al. MetaCrop: a detailed database of crop plant metabolism. Nucleic acids research. 2007;36(suppl_1):D954–8.
32. Bülow L, et al., *PathoPlant®: a database on plant-pathogen interactions*. In silico biology, 2004. 4(4): p. 529–536.

Tables

Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

Figures

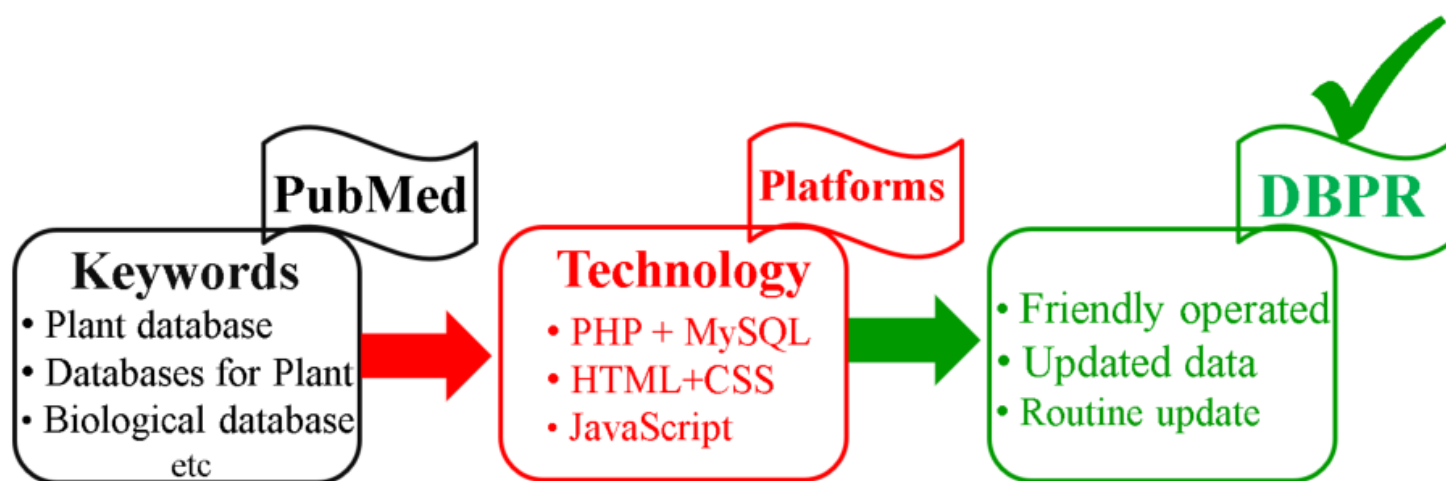


Figure 1

Procedure for the collection of plant data, construction of the DBPR and technologies used.

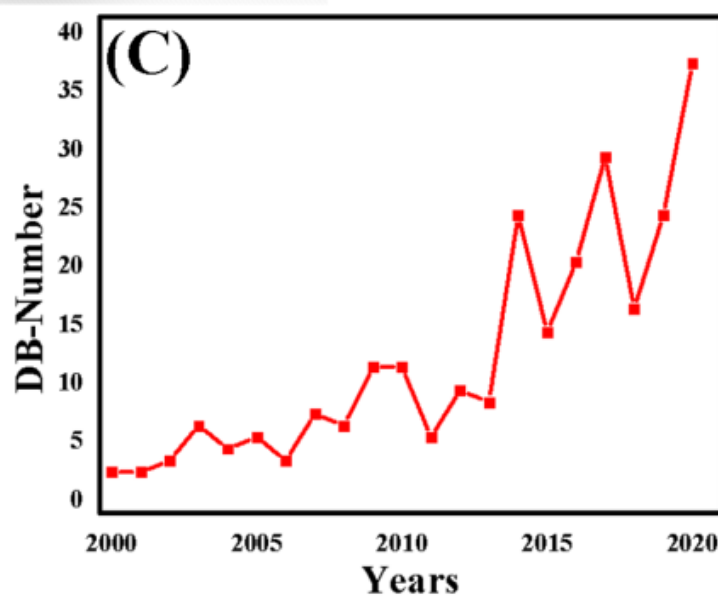
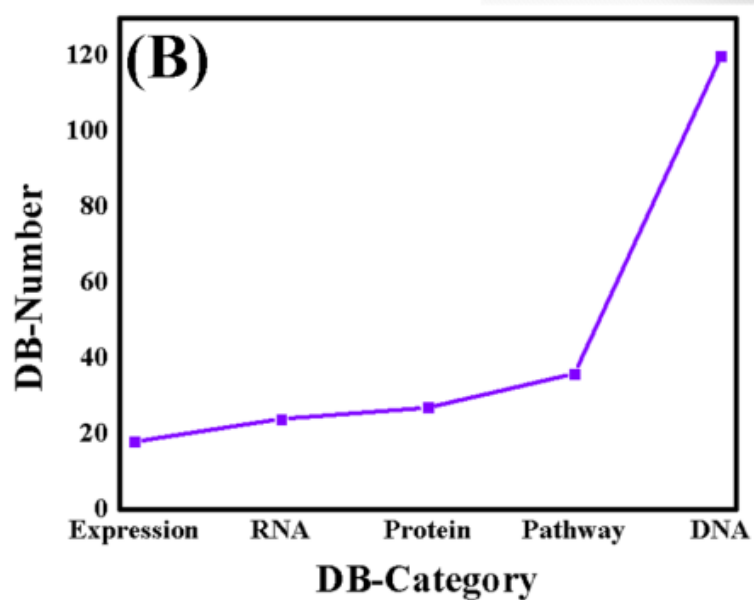
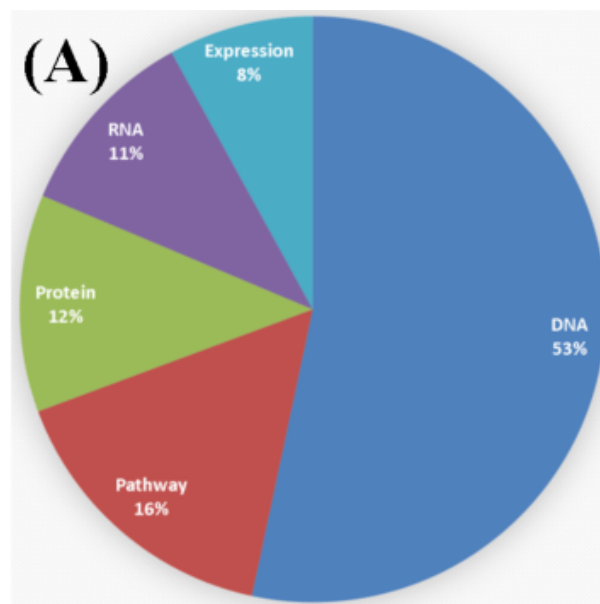


Figure 2

The statistics data of the DBPR. (A) Distribution of the category, (B) Chronological order of the DBPR, (C) Category-wise improvement of the DBPR.

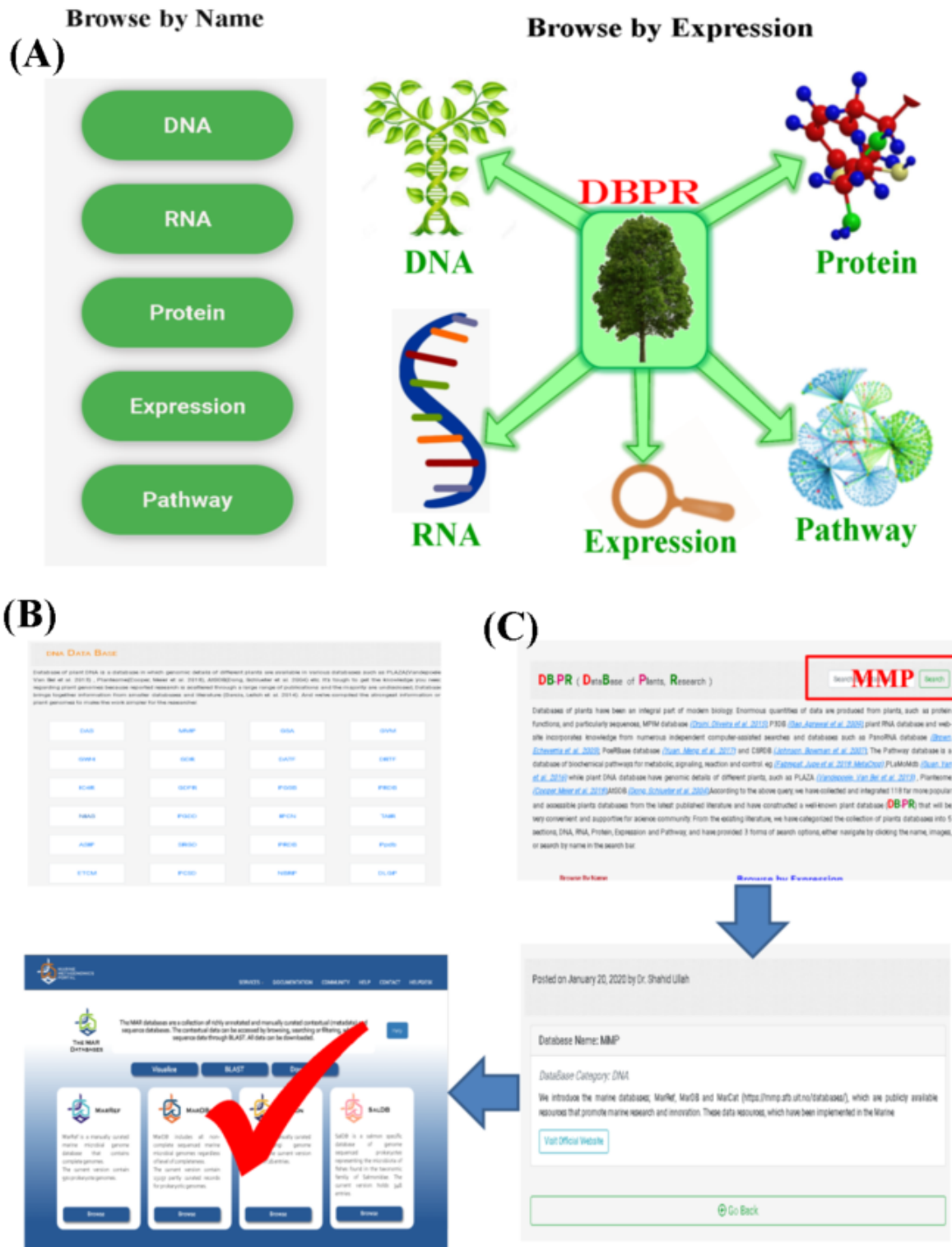


Figure 3

The searching procedure of the DBPR. (A) User can search by either clicking the name or image expression of the category, (B) Have shown the table of the clicked category, (C) Have shown the search bar with an example.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.jpg](#)
- [TableS1.xlsx](#)