Original Article

MED-PDB: An online database of medicinal plants

Bhakti Sargia, Bharat Singh, Nisha Gupta, Lokesh Kumar Gahlot, Trisha Gulati, Yasha Hasija

Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi-110042

Correspondence: Yasha Hasija, Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi-110042, E_mail: yashahasija@dce.edu

ABSTRACT

Medicinal plants have been anticipated to be one of the most valuable resources in therapeutic practices for human diseases. A range of plants in the form of herbal medications have been publicized to be therapeutically significant in a large number of diseases, counting cancers, diabetes, autoimmune disorders, epidermal infections, dermatological disorders, etc. WHO also claims that medicinal plants are extremely important for the population of developing countries.

Plant extracts have been used in raw, crude as well as processed form. Despite the clear evidence of the medicinal usage of plants, there is no such central repository that houses all the medicinal plants and their usage. To fill this void, the present study aims to compile and curate the medicinal plants with their medicinal values from the published literature.

The present database host information on 147 plants species, 53 plant families & subfamilies, 435 types of diseases, 369 types of active compound and covers worldwide geographical distribution.

The database has been made freely available online at the URL http://genomeinformatics.dtu.ac.in/medicinalplant/.

We believe that the present database may help in pharmacological and clinical exploration of plant species revealing the subsequent role of active compounds in various human diseases.

Keywords: Medicinal plants, active compounds, disease, database

Introduction

Several research model revolutions including both experimental and theoretical have made it plausible to fathom the modes of functioning of biological processes at molecular level. Computational paradigms of systems with extensive properties have also been proved to be the foundation for prognosis of biological behavior, giving rise to new discoveries and investigations[1]. In order to improve the efficiency of such biological transformation it is necessary to aid the processing, amalgamation, elucidation of the immense heap of biological data provided by various research communities. For many years, Databases have been proved to be a usual way of dealing and handling vast oodles of information in miscellaneous fields, including industry, academic restraints, and government subdivisions.

The practice of database technologies has garnered the attention of a division of the biological community, but its operation has been sparse to a significant part of the community though these assets are followed by numerous people of the research community. This can circumscribe not only the usages of these data to its utmost volume but also guide to misuse of the data. Adding to it, many experimental biologists are figuring data on a massive scale and are in need of establishing and organizing their own databases.

The motivation of establishing a consolidated database is to explicate the process by which major database resources pertinent for plant research obtain, analyze, and beget their data accessible, to identify the current constraints and the future endeavors of these resources to promote application of databases to research problems and respective goals.

The notion of plant specific databases is matter to adjustment as researchers are flaring their room of research. Consider the availability of various gene sequences of many organisms from various data sources that has enabled clinicians and researchers to easily and rapidly access and compare the sequences of interest[2]. Adding to biological database based transformation, various single species databases are available which only deals with taxonomically related species for e.g. Databases for grains, cereals, and night shades. Other examples of unrestricted databases include the ones which are based on particular domains of data like metabolism, genome annotation, orthologous relationships etc.

Access this article online	
Website: www.japer.in	**E-ISSN:** 2249-3379

How to cite this article: B. Sargia, B. Singh, N. Gupta, LK Gahlot, T. Gulati, Y. Hasija. MED-PDB: An online database of medicinal plants. J Adv Pharm Edu Res 2018;7(4):204-207.
Source of Support: Nil, **Conflict of Interest:** None declared.

Considering the storage and ongoing availability of such huge information there is a need to develop fairly independent database with good quality documentation and proper design to facilitate barrier free data exchange. A problem in building a database is the paucity of acknowledgment of this work as a real scientific effort. Many of the databases are public endeavors constructed with the help of software developers under the guidance of a biologist, sharing their experience via conferences may help in improving the problem. Widely held database based papers describe only the content and offer slight material on the design and operation of the software with no schema available.

Another crucial difficulty in this field is the inadequate capability to access and practically assimilate data from these multiple databases in an allied manner. Several types of databases along with plenty of software applications make it difficult for the researcher to extract the exact information in time and the representation of data in all the accessible form through these data sources puts an added load on researchers who wants to utilize the resource information.

Currently, we wish to deploy computer-readable data model of plant metabolomics based database MED-PDB, which incorporates a database schema for the formation of globally available plant metabolite information relating the active component to the effective metabolic target involved in various diseases, and a user interface to browse the certain defined attributes in context of disease. It will provide an enhanced and automatous understanding about ongoing biological research transition and will help in better implementation of novel methods and technologies[3,4].

Dataset and Feature Selection

For robust data prognostication, it is advantageous to incorporate several biological data sources and it has been realized that both plants and their medicinal uses are substantial for analyzing whether unknown active compound is used in a disease or not[5-7]. Consequently, for our model constructing motive we incorporateddisease, botanical name, family, common name, Geographical Distribution, morphology of plant, type of extract, active compound, administration, biological target and PMID[1]. Medicinal uses of all the plants were the findings from NCBI. The parameters of Medicinal use of plants were defined according to type of extract and their correlation with disease cure etc.

Methodology

The primary data in the MED-PDB depicts affiliation of various plant extracts along with their active compounds to various diseases. All the complicated medical conditions that are presumed to be rife have been deliberated for the purpose of this work. The information on the relationship of plant active compound with disease was obtained from the articles published in subsections of NCBI.

The plant and disease associations for MED-PDB were manually curated from the germane articles published in PubMed extracted deploying the keywords such as "Disease AND Plant Metabolites" or "Disease name AND Active Compounds of plants". The information from the current databases was sensibly evaluated and amended with apropos to the original articles. Then we created a data model deploying the free open-source version of the MySQL Workbench to minimize tautology[8]. Each ingress in MED-PDB contains the information on Disease, Botanical name, Family and Common Name, Geographical Distribution, Morphology of Plant, Type of Extract, Active Compound, Administration, Biological Target and PubMed ID (PMID).

User Interface

MED-PDB provides a comprehensible interface to query complete information on disease and plant association. Users can query the database through disease, botanical name, common name or family[5]. Further, MED-PDB interface permits the selection of the attributes, such that only the anticipated information on Plant-Disease association can be observed without encumbering the screen with data of least significance for the user[9].

Structure of MED-PDB Database

The MED-PDB database holds the data in the form of differentially categorized flat file format sub-locations. The user can access it through computer or mobile phone via the internet. The simple design of the database includes a front end search window and a back end database repository file.

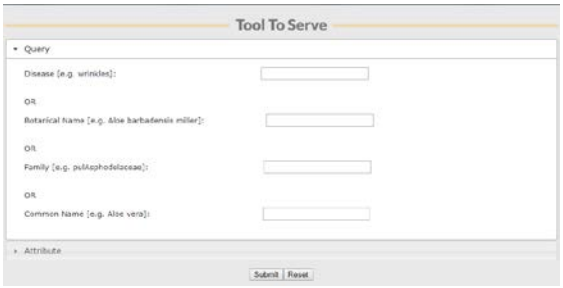

Figure 1. The front end search window having four basic search section comprising Disease, Botanical name, Family and Common name that allow the user to make an input of interest.

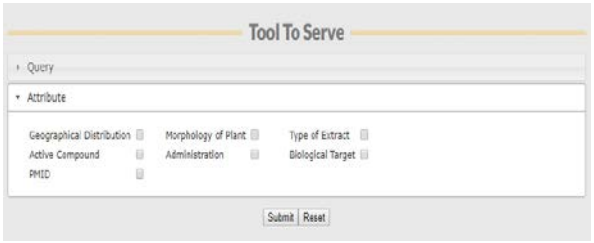

Figure 2. A window which allows the user to select certain attributes in context of the search query

These attributes include Geographical Distribution, Morphology of Plant, Type of Extract, Active Compound, Administration, Biological Targets and PubMed ID (PMID)[6,7].

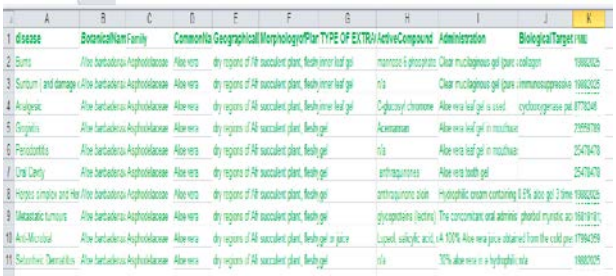

Figure 3. Back end flat data file with predefined sections containing the related information regarding the query.

Description of Attributes

1. **Disease: -** An abnormal condition with specific sign and symptoms. In our database we have provided information about 435 types of medical conditions.

2. **Botanical name: -**International code for nomenclature of distinct plant species. It defines the species and genus information. We have provided information about 147 plant species collected from the manually peered data.

3. **Family:** - It defines the homologous sharing group which contains various plant species of different genera's. MED-PDB holds description about 53 plant families and subfamilies.

4. **Common Name:** - A common epithet of a plant. It is very useful when a researcher wants to get the information about the role of a traditional plant in any disease.

5. **Geographical Location:** - Denizen of the medicinal plant. MED-PDB covers the world-wide distribution of various plants.

6. **Morphology of Plant:** - MED-PDB provides the information about the external appearance of the plant.

7. **Type of Extract:** - For medicinal purposes plant extracts are prepared in various different ways by taking different plant parts like root, stem, leaf etc. MED-PDB also contains the information about the 369 type of extracts used for the experimental purpose.

8. **Administration:** - During clinical procedure the way used to administer the drug determines a crucial role[6]. MED-PDB search output is able to give the information to the user about the same.

9. **Active Compounds:** - Use of various plant parts for the medicinal purpose depends on the type of target being used. It is necessary to know the nature of active compound in order to determine the proper functioning or interaction of target with the target compound. Plants are said to be the abundant source of various medically important Active compounds for e.g. Alkaloids, Terpenoids, Glycosides, Natural Phenol, Phenazines, Biphenyls etc. MED-PDB provides information about the type of active compound involved in disease metabolism.

10. **Biological Target:** - Functioning of any target compound depends on the nature of target it is going to interact with. MED-PDB provides the information about the Biologically Active site interacting with active compound.

11. **PMID:** - It is the unique identification number assigned for each and every record which is stored in PubMed. The section of MED-PDB can provide the information about the literature for the whole work in just one click.

Web Implementation

For the expedition of data retrieval, user-friendly web interface was evolved. XHTML and CSS were deployed for forming presentation layer of MED-PDB, the application server deployed was Apache. MySQL was used for backend database and PHP was used as a programming language[8].

Database Accessibility

The basic unit of MED-PDB is the various medicinal properties of plants, which is depicted in the online delivery model of database as a Variant Report. Synopsis is provided for maximum plant species, including the various disease, botanical name, family, common name, Geographical Distribution, morphology of plant, type of extract, active compound, administration, biological target and PMID. Synopsis is ensued by the list of PMID curetted for the variant. Each study divides into a set of observations, with each observation comprising five core fields of data.

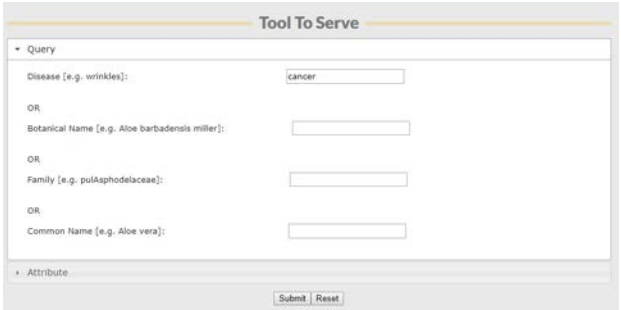

Figure 4. Introducing Cancer as the query input in front end search window at MED-PDB.

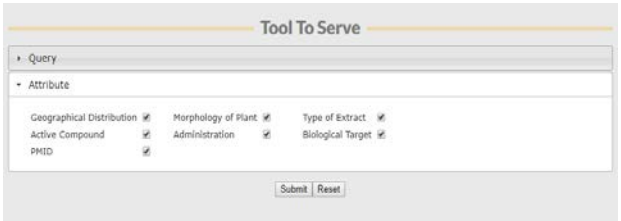

Figure 5. Selection of attributes in the front end search window at MED-PDB

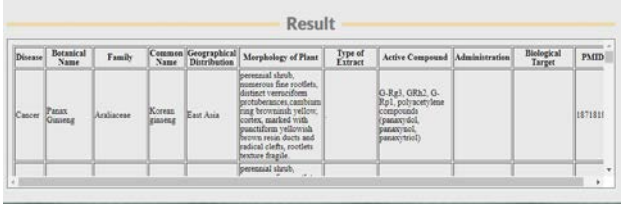

Figure 6. Search result for the desired input (Cancer) showing description about selected attributes generated by Back End Flat file database.

The MED-PDB is a cachet that has collated the literature on medicinal plants with the diseases. By permitting a user to swiftly overlay the earlier observed correlations, we have made it plausible to give meaning to active compound of plant extract in a clinical context, helping usher both clinical and potential treatment of possibly severe disease on an individual basis.

Conclusion

This report has delineated the concept of MED-PDB, and conferred some of its salient aspects. The MED-PDB has given a new dimension to researchers who are looking for various aspects of plants, medicinal properties and their active compounds which are used in therapeutics of disease. MED-PDB will foster sharpened and quick availability of the information rather than setting up an individual data centre for each plant or its various properties as done earlier.

The above mentioned information is the legitimate result of manually curetting and studying various research papers reviewed at available database at NCBI. MED-PDB database saves a lot of money and time of researchers as the data from different sources is present in a single database. In the absence of MED-PDB database, researchers might have to invest a huge amount of money and time for getting information about various plants and related aspects mentioned in the database.

MED-PDB database compiles a good number of attributes of medicinal plants which can serve as a basis for research prospects. MED-PDB has given the latitude to access the information from anywhere without any circumscription of purchasing policy etc. which makes it a powerful tool.

Acknowledgement

Authors thank Aashiyan Singh, Aroma Kaul and Sanchit Varma for their contribution in annotation.

References

1. National Research Council. Catalyzing inquiry at the interface of computing and biology. National Academies Press, 2006.
2. Mehmood MA, Sehar U, Ahmad N "Use of Bioinformatics Tools in Different Spheres of Life Sciences" Journal of Data Mining in Genomics Proteomics 5:158, 2014
3. R. Prakash. "Traditional Uses of Medicinal Plants in Uttarakhand Himalayan Region". Scholars Academic Journal of Biosciences (SAJB), pp 345-353, 2014.
4. W. A. Shehri. "Cloud Database Database as a service". International Journal of Database Management Systems (IJDMS)Vol.5, No.2, 2013.
5. Hassan, Bassam, Medicinal Plants (Importance and Uses) Pharmaceutica Analytica Acta, 2012
6. Nwachukwu, C. U., et al. "Identification and traditional uses of some common medicinal plants in Ezinihitte Mbaise LGA, of Imo State, Nigeria." Report and Opinion 2.6 (2010): 1.
7. A. Maroyi, "Traditional use of medicinal plants in south-central Zimbabwe: review and perspectives". Journal of Ethnobiology and Ethno medicine, pp 9-31, 2013.
8. A. Kaplun, J.D. Hogan, F. Schacherer, A.P. Peter, S. Krishna, B.R. Braun, R. Nambudiry, M.G. Nitu, R. Mallelwar, A. Albayrak. "A comprehensive manually curated pharmacogenomics Database" The Pharmacogenomics Journal, 2015.
9. Isha Srivastava, Lokesh Kumar Gahlot, Pooja Khurana, Yasha Hasija, dbAARD & AGP: A computational pipeline for the prediction of genes associated with age related disorders, In Journal of Biomedical Informatics, Vol. 60, pp 153-161, 2016.

Author Biography

Bharat Singh is currently pursuing his PhD from MNNIT Allahabad in Medical Biotechnology. He has done M.Tech in Biomedical Engineering from Delhi Technological University and B.Tech from Sharda University He was a part of various projects during his master's program which include developing NutriGenDB: nutrigenetic variants database, developing medicinal plant database MED-PDB and Disease module study.
Email: bharat30.1990@gmail.com

Bhakti Sargia has done her B.Tech in Biotechnology Engineering from Sir Padampat Singhania University Udaipur and M.Tech in Biomedical Engineering from Delhi Technological University. She has worked on projects such as laser and its interaction with human body skin, report on computational intelligence: a thriving field in skin cancer etc.
Email: sbhakti101@gmail.com

Nisha Gupta has done her B.Tech in Biotechnology and M.Tech in Biomedical Engineering.
Email: nisha.biotech2012@gmail.com

Lokesh Kumar Gahlot has done B.Tech in Computer Engineering (2008-12), M.Tech in Biomedical Engineering (2013-15). He is currently working as Software Engineer II at Thermo Fisher Scientific. He is the reviewer for the journal Soft Computing. He has 2 international publications and a book chapter to his credit.
Email: lok.gahlot@gmail.com

Trisha Gulati is pursuing her B.Tech in Biotechnology Engineering from Delhi Technological University. She has worked on projects like Biochemical Analysis (AIIMS), Analysis of pure cultures obtained from environmental samples (DTU) and Preclinical Models in Behavioral Neuropharmacology (DRDO).
Email: trish.gulati@gmail.com

Yasha Hasija (B.Tech, M.Tech, and Ph.D.) completed her Ph.D. from CSIR-IGIB (Delhi) and University of Pune. She is currently working as Assistant Professor at Department of Biotechnology, Delhi Technological University. She has published several papers in national and international journals of high repute, and has been awarded several prestigious awards. She is the Project Investigator of several sponsored research

projects from Govt. agencies including SERB, CSIR-OSDD and DBT. Her broad areas of research include genome informatics, genome annotation, microbial informatics, integration of genome-scale data for systems biology and personalized genomics.