

Development and Deployment of an Ethnobotanical and Phytochemical Knowledge Database of Malaysia

Shakina Mohd Talkah^a, Ilyia Zulkiflee^a, Mohd Shahir Shamsir^{a*}

^aDepartment of Biological and Health Sciences, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

*Corresponding author: shahir@fbb.utm.my

Article history

Received :29 July 2013
Received in revised form :
23 September 2013
Accepted :29 September 2013

Graphical abstract



Abstract

Currently, all the information regarding ethnobotanical, phytochemical and pharmaceutical information of South East Asia are scattered over many different publications, depositories and databases using various digital and analogue formats. Although there are taxonomic databases of medicinal plants, they are not linked to phytochemical and pharmaceutical information which are often resides in scientific literature. We present Phyknome; an ethnobotanical and phytochemical database with more than 22,000 species of ethnoflora of Asia. The creation of this database will enable a biotechnology researcher to seek and identify ethnobotanical information based on a species' scientific name, description and phytochemical information. It is constructed using a digitization pipeline that allow high throughput digitization of archival data, an automated dataminer to mine for pharmaceutical compounds information and an online database to integrated these information. The main functions include an automated taxonomy, bibliography and API interface with primary databases such as Global Biodiversity Information Facility (GBIF). We believe that Phyknome will contribute to the digital knowledge ecosystem to elevate access and provide tools for ethnobotanical research and contributes to the management, assessment and stewardship of biodiversity. The database is available at <http://mapping.fbb.utm.my/phyknome/>.

Keywords: Biodiversity; database; ethnobotany; herbs; medicinal plants

Abstrak

Sehingga kini, kesemua maklumat berkaitan etnobotani, kimia fitologi and farmasi tumbuhan Asia Tenggara berselerakan didalam pelbagai hasil penerbitan, pangkalan maklumat didalam pelbagai bentuk samada analog atau digital. Walaupun ada pangkalan maklumat tumbuhan yang mempunyai kegunaan perubatan, biasanya ia tidak akan dipaut dengan maklumat fitokimia dan perubatannya yang selalunya hanya dijumpai didalam penerbitan saintifik dan jurnal. Kami mempersembahkan Phyknome; satu pangkalan maklumat etnobotani dan fitokimia yang mempunyai lebih dari 22,000 spesis etnoflora dari Asia. Kewujudan pangkalan data ini akan membolehkan penyelidik bidang bioteknologi untuk mencari dan mendapatkan maklumat berdasarkan nama saintifik sesuatu sepis beserta ringkasan maklumat dan maklumat bahan kimia aktif. Ia dibangunkan menggunakan proses pendigitalan yang membolehkan maklumat analog di lombong secara pantas dan pada skala yang besar. Fungsi utama Phyknome adalah seperti rangkaian taksonomi, pautan otomatik kepada pangkalan maklumat lain seperti Global Biodiversity Information Facility (GBIF). Kami percaya bahawa penwujudan Phyknome akan menyumbang kepada ekosistem maklumat kepelbagaian biologi dan membantu pengawasan, urustadbir kepelbagaian biologi yang mapan. Pangkalan data ini boleh didapati di <http://mapping.fbb.utm.my/phyknome/>.

Kata kunci: Etnobotani; fitokimia; Malaysia; biodiversiti

© 2013 Penerbit UTM Press. All rights reserved.

1.0 INTRODUCTION

Biological diversity or biodiversity is the variety of life found in an ecosystem and the variation in their genetic makeup. In recent years, we have seen the sharp increase of demand and

subsequent availability of online information on biodiversity. The increase in the number of biodiversity's conservation initiatives makes it important to elevate the sharing and accessibility of biodiversity information, especially the ethnobotanical and phytochemical information as these have

been the source of majority of active compounds that become the basis of drugs in the treatment of ailments and diseases. The term ethnobotany is described by Harshberger as the study of the complex interaction between plants and human [9]. The interaction is defined as the plants that are involved in human survival, living surroundings and the world's ecosystem. 'Phyto' is a Greek word for plants. Literally, phytochemicals means chemicals produced by plants. The term generally used for plants chemicals that have beneficiary effects on human health [7]. The current evolution of ethnobotany and phytochemicals as interdisciplinary area now has incorporated various fields such as anthropology, economics, linguistics, indigenous law, pharmacology and most importantly the field of medicine [8]. The value and importance of ethnobotany studies and the need for protection and documentation of species that might be on the verge of extinction makes it crucial to preserve this indigenous knowledge for sustainable use for the benefit of the next generation. Although there are an abundance of data and information, there are only a few specific databases that cater for ethnobotany. Malaysia has the potential to be one of the leading countries that provide the best biodiversity data because it has a complex tropical rain forest ecosystem containing potential plants sources that have evolved for over 100 million years [1]. The fact that Malaysia is one of 17 designated 'megadiversity' shows that Malaysia shoulders a responsibility to conserve its natural resources and the ethnobiological information within it. However, over the last 30 years, Malaysia shows rapid socio-economic growth that leads to destruction of natural resources, plants and animal biodiversity through ecosystem destruction and degeneration. The increasing recognition of the role of biodiversity in economic, scientific and educational aspects of life have increased its importance and are considered as key issues in sustainable development policies [2]. An establishment of proper plant biodiversity data management will assist towards bioresource loss prevention and sustainable development.

The increase in demand and amount of biodiversity data has created a global need for structured information domain by various institutional and national databases for biodiversity and conservation research. Traditionally, respective communities in Malaysia have practiced and established its ethnobotany legacy and culture. The Chinese community relies on the Traditional Chinese Medicine and the Indians with their Ayurvedic traditions. The Malays have similar use plants for dietary, medicinal and ritual applications, usually within the boundary of the Islamic practise. However, there is a conspicuous scarcity of Malaysian biodiversity databases that focuses on ethnobiology even with the existing cultural and historical legacy with only a few robust and notable examples. Among the leading examples is the GlobinMed (<http://www.globinmed.com/>), a resource on traditional and complementary medicine created by the Institute of Medical Research (IMR), Flora of Peninsular Malaysia Online (FPM: <http://www.tfbc.frim.gov.my>), a searchable flora that includes ancillary data such as habit, habitat, taxonomic synonymy, distribution within Peninsular Malaysia and Singapore. There are agriculturally focused databases such as AgrobIS (<http://agrobis.mardi.gov.my/>), an information system developed by MARDI to provide the public direct access to data of more than 40,000 accessions of Plant Genetic Resource for Food and Agriculture (PGRFA) which includes fruits, rice, vegetables, and medicinal plants. The system also consists of information on 2,500 isolates of microbial genetic resources and about 30,000 specimens of arthropods. The Department of Agriculture is also developing Agriculture Information System Geodatabase Portal (AgrIS) <http://www.agris.doa.gov.my>, a Web-based system integrates agricultural data in all the States of

Malaysia into a single unified system where one can access up-to-date information about soil, rainfall, pests, land use etc. There are smaller institutional, state and research group databases such as The Langkawi Mangrove Biodiversity Database <http://www.mangrove.my> and the the Johor Biodiversity Database by the Johor Biodiversity and Biotechnology Corporation at <http://www.jbiotech.gov.my/jbiodi/>, However, such commendable effort is disproportionate to the existing untapped biodiversity data and the present research community in Malaysia. Global Biodiversity Information Facility (GBIF: www.gbif.org) is a sophisticated international organization based on Among the 14 contributing institutions in the, there are no databases originating from Malaysia.

The conspicuous absence of an ethnobotanical database has motivated us to develop a comprehensive database for Malaysia. The database would possess features and datasets digitized from recognized local traditional knowledge, scientific literature and books on herbs or plants within Malaysia as well as sourcing from established online database with similar themes.

2.0 CONSTRUCTION AND CONTENT

To create this portal, we would conduct research on the best data model, data types and data compatibilities of biodiversity data generated by Malaysian researchers. As one of 12 megadiversity countries and a mixed heritage of three major cultures found in Asia, Malaysia offers a substantial amount of traditional knowledge regarding plants with medicinal value. Adding the assimilation of different cultures in Malaysia such as indigenous peoples, Sabahan and Sarawakian natives increases the richness of the ethnobotanical information. Since 1966, an English botanist, Isaac Henry Burkill in his extensive research compiled 1,300 plant species with their medicinal value in Peninsula Malaysia [1]. The construction of the datanase follows the principles proposed by Ningthoujam and coworkers [2] (Figure 1).

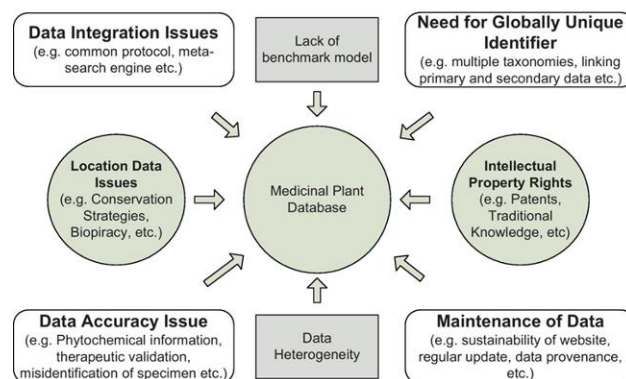


Figure 1 Challenges facing the utilization of databases for medicinal plant research [2]

The system was built using Rapid Prototyping methodology which is a variant of the Rapid Application Development (RAD) method. Books and journals containing information about plants with medicinal value were searched and collected from the library and the World Wide Web. Online leading databases that have been sources for taxonomic data collection are Integrated Taxonomic Information System – ITIS (<http://www.itis.gov/>), Encyclopedia for Life - EOL (<http://eol.org/>), Biodiversity Heritage Library (<http://www.biodiversitylibrary.org/>) and Natural Resources

Conservation Service by United States Department of Agriculture (<http://plants.usda.gov/java/>).

We scanned and manually collected literature from UTM and UKM libraries, pages that contained the desired information were scanned and converted into pdf form using Optical Character Recognition (OCR) software. The data collection and user requirement analysis process highlighted the inherently complex nature of biodiversity data that is associated with the plant species. Another complexity is socially generated through potentially competing groups, government agencies and varied interest. Addition of political and commercial interests within these groups would sometimes spark volatility into the debate. Not only that there are multiple, disparate data providers with specific user communities, but also the biodiversity data is diverse, and includes taxonomic names, museum collections, publications, genomic and phenotypic data [4]. The content of the database includes information on more than 22,000 species with varying degree of information completeness. The content types can be summarised in Table 1.

2.1 Survey on Existing Database

Despite the fact that there are very limited organizations that implemented Drupal as their platform for biodiversity database management system, Scratchpad showed a pretty good reputation due to the fact that a lot of researchers adapt it as a platform to store their research collections. However, an ethnobotanical database based on Drupal platform is yet to be found. The research will apply the usage of Drupal as new platform for ethnobotanical database management system due to its flexibility towards implementing system's requirement.

Several medicinal plant databases were reviewed and data content was browsed to survey on what type of data each database offered. The databases were gained from journal citations and keywords searching. Table below project the results from the review:

Table 1 Survey on existing database

Online Databases	Database Summary	Database Content
Prelude Medicinal Plants Database (http://www.metafro.be/prelude)	Provide data collections of Royal Museum for Central Africa (RMCA) on zoological, geological, historical and anthropological.	<ul style="list-style-type: none"> • Scientific name • family name • vernacular name • plant images • symptoms of illness • article references
Customary Medicinal Knowledgebase - CMKb (http://biolinfo.org/cmkb)	Provide information on customary medicinal plant knowledge that scattered in different literature resources possessed by Australian Aboriginal people.	<ul style="list-style-type: none"> • Species information such as kingdom, scientific name • Synonym • common name • native language name • Habit • Habitat • author citation <p>This database also offers medicinal information, chemical information, biological activity information and biogeography information.</p>
FRLHT Encyclopedia of Indian Medicinal Plants (http://envis.frlht.org/indian-medicinal-plants-database.php)	Developed by Foundation for Revitalization of Local Health Traditions (FRLHT) – provide information on Indian plant with medical value heritage based on the research from Institute of Ayurveda and Integrative Medicine.	<ul style="list-style-type: none"> • Scientific c name • family name • vernacular name • references which they used in (Ayurveda, Unani and Sidha).
Traditional Knowledge Digital Library (http://www.tkdl.res.in/tkdl/langdefault/commmon/)	Provide information on formulations of medicinal plants based on Ayurveda, Unani and Siddha books.	<ul style="list-style-type: none"> • Scientific name • common name based on Ayurveda, Unani and Siddha and references in those traditional books.
Yuthog Foundation for Tibetan Medicine (http://www.yuthog.org/?Home)	Developed by Yuthog Foundation for preservation and recognition of Tibetan traditional medicine based on natural plant.	<ul style="list-style-type: none"> • Scientific name • family name • Tibetan common name • detailed location • medical usage • part used • brief description of the species.

Online Databases	Database Summary	Database Content
TradiMed (http://www.tradimed.com/)	This website offers information on Chinese and Korean medical classics which combine traditional knowledge and modern scientific regarding plants with medicinal value. However, the data only accessible when users purchase the coupon offers by the website.	<ul style="list-style-type: none"> • Scientific name • Chinese name • common name • botanical resources • symptom illnesses • active chemical constituents • processing method.
Ethnobotany of the Peruvian Amazon (http://www.biopark.org/peru/plants-amazon.html)	The database offers information on plants that commonly used by Peruvian Amazon.	<ul style="list-style-type: none"> • Scientific name • family name • common name by Amazonian • brief comments about its medicinal use.
SEPASAL (www.rbgekew.org.uk/ceb/sepasal/)	The database offers enquiry service about useful “wild” and semi-domesticated plants of tropical and subtropical drylands, developed and maintained at the Royal Botanical Gardens, Kew.	<ul style="list-style-type: none"> • Scientific name • common name • synonyms • conservation status • habitat • detailed taxonomic information (from class to genus) • geographical distribution • description • illustration • uses • cultivation
Plants for a Future (http://www.pfaf.org/user/default.aspx)	The database content is a compilation data of Ken and Addy Fern work their experimental site in Cornwall. They have been experimenting 1500 species of edible plants on the field.	<ul style="list-style-type: none"> • Scientific name • common name • family name • synonyms • known hazards • habitats • images • medical uses • edible uses • cultivation details • references
NAPRALERT (http://www.napralert.org/Default.aspx)	Relational database of natural products, including ethnomedical and pharmacological information from all organisms including marine organisms. Require payment to view full results.	<ul style="list-style-type: none"> • Scientific name • active compound information • pharmacological activity
EthnoMedicinals (http://www.ethnomedicinals.com/index.htm)	General database on taxonomy of medicinal plants.	<ul style="list-style-type: none"> • Provide links to other related plant databases.
Dr. Duke's Phytochemical and Ethnobotanical Databases (http://www.ars-grin.gov/duke/)	A compilation of research data by Dr James A. Duke regarding ethnobotanical uses on certain plant species.	<ul style="list-style-type: none"> • Queries for chemicals and activities in particular plants • ethnobotanical uses • specific chemical searches for chosen plants are available in this website.

2.2 Content Types and Functionalities

We examined the entire available online biodiversity database generated to benchmark the functionalities of our database (Figure 2). We also included picture, automatic links and API to third party content, bulk import function using spreadsheet formats, automatic generation of taxonomic information based on GBIF auto-generated taxon data as well as an installer version that is packaged with the well-known XAMPP for Windows Operating System (OS) to facilitate biologist. We added Featured Plants function that randomly display species available in the system (Figure 3), Featured Researchers function that randomly display researchers that involved in

Phyknome and a species counter that display species numbers that are currently uploaded in the system.

Table 2 Contents types and functionalities of Phyknome

Content Types	Functionality
Species Account	<p>Main species page for each species in Phyknome. Fields in this content type are:</p> <ol style="list-style-type: none"> 1. Scientific Name (which is unique id for each species). 2. Synonyms (unique for each species). 3. Vernacular Name. 4. Chinese Name. 5. Taxonomy Information (Phylum, Class, Family, Genus). 6. Description. 7. Image. 8. Medicinal Uses. 9. Parts Use. 10. Precaution. 11. Other Uses. 12. Chemical Constituents
Researcher	Provide profile pages for researchers who developed Phyknome.
Gallery	Contributed module that provide gallery view of uploaded images
Biblio	With presence from Biblio module, all literatures that involved in 'Phyknome' data mining were integrated with direct linking of scholars search engine which are Google Scholar and PubMed. Users can also import the literature source to their EndNote and BibTex.

Piper nigrum

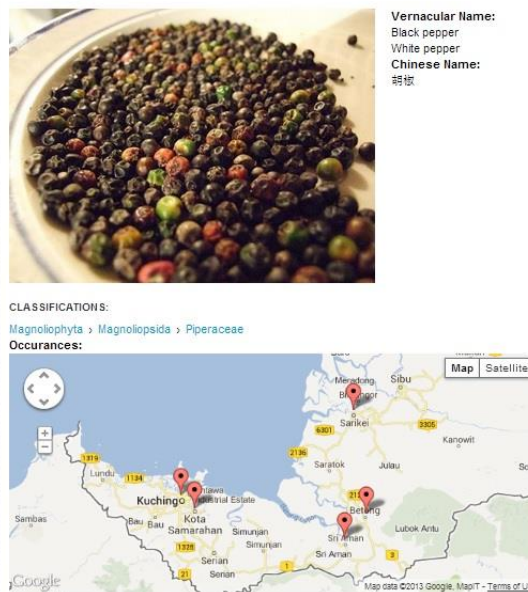


Figure 4 Exemplar of GIS data

We added GIS functionalities (refer Figure 4) that will create a unique reference base (geographic location and maps), i.e. natural vegetation, soil, land use, topography, hydrology that enables biodiversity data to be combined and analysed to produce novel associations between environmental features and relationships between different species. Using GIS will facilitate visualizing the relationship between species and geographical locations, inter and intra species locations and their relationships and examining distribution characteristics of species [3]. The temporal context of biodiversity data is important as it reflect trends in range, numbers, distribution, genetics, and proportions over time. In Malaysia, although GIS and its various mapping derivatives have been used in biodiversity research, none of the database described earlier provides for dynamic queryable mapping solutions and at best, maps are provided only as images and pdfs. We used retrospective georeference where the textual descriptions of places where data and specimens were collected (locality descriptions such as “Sungai Tiram, Kampung Teluk Sengat, etc”) were converted into their corresponding geographic coordinates—especially from legacy or archival data that are descriptive in nature This will create quantitative map coordinates and generate new usability of the legacy data by making it amenable to spatial and numerical analysis. To retrospectively georeferencing is a daunting task given the biodiversity data itself are legacy in nature and almost all biodiversity data records in Malaysia are in printed form. An automated workflow that harvests, simplifies and improves the efficiency is desirable given the large amount of data available in Malaysia.

3.0 CONCLUSION

The construction of Phyknome which specializes on ethnobotanical information offers comprehensive database with more than 22,000 species of plants with nutritional value based on ethnobotanical data primarily from local culture. The content contains taxonomy classification of the specific plants, medicinal usage and active compound that is readily accessible



Figure 2 Screenshot of Phyknome (<http://mapping.fbb.utm.my/phyknome/>)



Figure 3 Featured plants showing exemplar species pages

to public and researchers on the website. We hope that the database will preserve Malaysian ethnobotanical heritage and catalyse new plant-based drug discovery in the future.

Acknowledgement

We would like to thank Universiti Teknologi Malaysia for funding this project (GUP 01H25).

References

- [1] Napis, S., et al. 2001. *Biodiversity Databases For Malaysian Flora and Fauna: an Update*. In Proceedings of IWS2001 (Internet Workshop 2001).
- [2] Zeti, A. M., et al. 2009. *Bioinformatics in Malaysia: Hope, Initiative, Effort, Reality, and Challenges*. PLoS Computational Biology. 5(8): e1000457.
- [3] Burkill, I. H. et al. 1935. *A Dictionary of the Economic Products of the Malay Peninsula*. London: Published on Behalf of the Governments of the Straits Settlements and Federated Malay States by the Crown Agents for the Colonies.
- [4] Ningthoujam, S. S. et al. 2012. Challenges in Developing Medicinal Plant Databases for Sharing Ethnopharmacological Knowledge. *Journal of Ethnopharmacology*.
- [5] Page, R. D. M. 2008. Biodiversity Informatics: The Challenge of Linking Data and the Role of Shared Identifiers. *Briefings in Bioinformatics*. 9(5): 345–354.
- [6] Oberlies, N. H. et al. 2009. Mapping of Sample Collection Data: GIS Tools for the Natural Product Researcher. *Phytochem Lett*. 2(1): 1–9.
- [7] Saroya, A. S. 2011. Herbalism Phytochemistry and Ethanopharmacology. *Science Publishers*. 36
- [8] Alcorn, J. B. 1995. *The Scope and Aims of Ethnobotany in a Developing World*. Ethnobotany.
- [9] Harshberger, J. W. 1896. *The Purpose of Ethnobotany*. Botanical Gazette.