Paithumbase - Biometrical Traits based Query System for Studying Mungbean [*Vigna radiata* (L.) Wilczek] Phenomics

*N.Senthil[1], J.Murukarthick[1] ,M.Pandiyan[2], A.Karthikeyan[1], M.Sudha[1],M.Raveendran[1]
P.Jayamani[3], S.Kalaiselvi[4], P.Nagarajan[1]

[1]Genomics and Proteomics Laboratory, Centre for Plant Molecular Biology & Biotechnology, Tamil Nadu Agricultural University, Coimbatore Tamilnadu, India
[2]National Pulses Research Centre, Vamban, Tamil Nadu Agricultural University, Pudhukottai, Tamilnadu,India
[3]Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore. Tamilnadu, India
[4]Avinashilingam University for Home Science and Higher Education for Women, Coimbatore, Tamilnadu, India

*Corresponding author

ABSTRACT

Paithumbase is an established database for identifying the phenotypic variations of the mungbean germplasm accessions preserved in the National Pulses Research Centre, Tamil Nadu Agricultural University. This database provides information on both quantitative and qualitative traits of the mungbean germplasm. The database architecture was built with a GUI interface, developed in PHP and HTML as front end and MySQL as the backend language. It is promising that this database will provide valuable biological and bioinformatics resources for plant physiologists and geneticists to dissect the major phenotypic traits to develop highly adapted plant material to sustain future breeding programs. The plant breeders can easily decide the choice of the parental lines involved in the varietal improvement programs by using this database. Paithumbase is available for public access at http://www.tnaugenomics.com/paithumbase/.

Keywords: Mungbean germplasm, MySQL, Paithumbase, PHP, TNAU Genomics

1. INTRODUCTION

Mungbean [*Vigna radiata* (L.) Wilczek] (2n=22) belongs to the family *Fabaceae* and subfamily *papillonaceae*, an excellent source of easily digestible proteins with low flatulence which complements the staple diet in Asia. Crop germplasm diversity contributes to developing improved crop cultivars aimed at increasing crop productivity. The miniaturization of crop collections with almost full representation of genetic diversity in the form of a database (~1% of the entire collection) approach is an effective methodology to enrich and enhance crop improvement programs [1]. A detailed phenotypic characterization of the experimental material is a critical requirement for a meaningful analysis of any crop plant. Understanding the different phenotypic characters of the germplasm will provide sufficient knowledge on developing strategies for enhancing breeding program. The correct choice of parents to be used in crossing will depend largely on the aims and objectives of the breeder and is generally based on phenotypic and genotypic expression of the characters they display. The two main criteria are as follows: (a) the characteristics of the parents such as agronomic and morphological traits, disease reaction and flowering behavior and (b) the breeding performance of the parents as revealed by progeny tests of previous crosses [2]. The major goal of this research is to develop a unique web portal containing detailed characterization of morphological traits of mungbean (*Vigna radiata*). For that 646 mungbean germplasm accessions were assessed in National Pulses Research Centre (NPRC), Tamil Nadu Agricultural University, in which all the accessions have 55 morphological traits (18 quantitative and 37 qualitative traits) with its values stored in the database. The front page of the Paithumbase was developed using Dreamweaver, usually used to design standard front pages. PHP and HTML are used to design dynamic and static web pages and the data are stored in MySQL database which is connected with web server. Hence, the present database is undertaken, to retrieve the genetic diversity among the different germplasm lines through morphological traits, to access or retrieve the information along with the map for searching the location through a Keyhole Markup Language (KML) embedded viewer. KML is an XML language focusing on geographic visualization, including annotation of maps and images which provides a user friendly interface to bring out the design and detail of the accessions in a particular location. This can be achieved by querying the database which is publicly accessible via the World Wide Web. This database can be accessed through the link, http://www.tnaugenomics.com/paithumbase/.

2. DATABASE ARCHITECTURE- GUI

Paithumbase has been designed in Relational Database Management System (RDBMS) using MySQL as the backend language. The Architecture of this database is designed aesthetically with a graphical user interface (GUI). The front end language PHP which is a scripting language was used in webpage development which is embedded into HTML [5]. The Database Administrator is provided with the privilege to access the backend to add or delete data objects, and edit the existing data objects after approval. The MySQL tables are stored in the tnaugenomics web server [6] [7] (www.tnaugenomics.com/database/admin).

The user interface is built on standard web technologies including HTML, JavaScript, PHP and AJAX [8]. One best

workaround is to use Asynchronous JavaScript Technology and XML (AJAX), which makes asynchronous calls to the Web server to retrieve XML documents to dynamically manipulate Web pages from JavaScript [3]. The relational data must first be converted to XML documents and be stored on the Web server, allowing the curator to update fields without a refresh of the page. A Web-enabled database can provide an effective and cost-efficient means to collect and distribute data and make that data available at anytime, anywhere. A KML embedded files in Google Earth has exposed geographic information technology virtually to everyone and geo-browsers are now used by a growing number of scientists as they enable data with a spatial component to be overlain on top of remotely sensed imagery to identify the places where a particular germplasm is grown [4]. The architecture and the capability of Paithumbase may be of interest in terms of knowledge sharing to the scientific and public community. In addition detailed database schema was described in Fig.1.
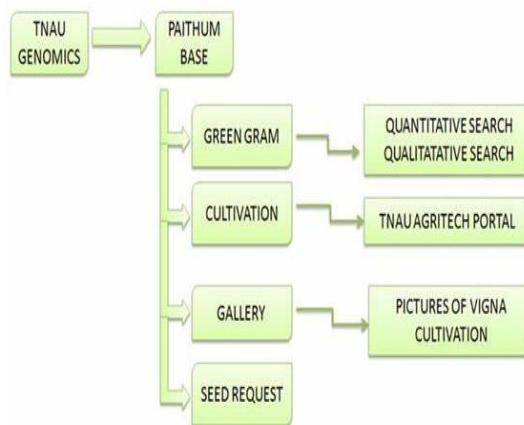

Fig.1.Paithumbase Database schema

2.1.DATA CURATION

The database is flooded with 648 accessions of Mungbean, with an extent of 55 phenotypic characters including 38 qualitative and 17 quantitative traits. The Qualitative and Quantitative data distribution have been plotted (Fig. 1). The data available here have been curated manually by experimental results and resources. The phenotypic traits are indicators of plant interactions with their environment and the resource and hence, they are to be studied extensively [9]. The traits include plant height (cm), the number of branches per plant, length of the branches (cm), the number of clusters per plant, the number of clusters per branch, the length of the pod (cm),the number of pods per plant, the number of seeds per pod, dry matter production (g), days to initial maturity, days to 50 per cent flowering, days to initial flowering, protein content (%), the breadth of the seed (cm), the length of the seed (cm), the single plant yield (g), hundred seed weight (g), days to full maturity and days to emergence, growth habit, growth pattern, hypocotyl colour (recorded after 10 days of emergence), stem colour, primary leaf shape, terminal leaflet shape, terminal leaflet length (recorded for the leaf at fourth node), leaf pubescence, leaf colour (recorded at 50 per cent flowering), petiole colour, petiole length (recorded for the leaf

at fourth node), leaf senescence (recorded when 50 per cent pods have dried), peduncle length, raceme position (recorded when the first pod changes colour), calyx colour, corolla colour (wings and standard), pod colour at immature pod, colour of ventral suture of immature pod, pod colour at mature stage, attachment of mature pod to peduncle, pod pubescence (recorded when the first pod changes colour), pod peak shape, cross – section of the mature pods, seed colour, cluster on seed surface, seed shape, twinning tendency, branching pattern, leafiness (at 50 per cent flowering), flowering period, shattering in the field, reaction to Cercospora leaf spot, reaction to powdery mildew, reaction to bacterial leaf spot, reaction to mungbean yellow mosaic virus, reaction to bruchids infestation.
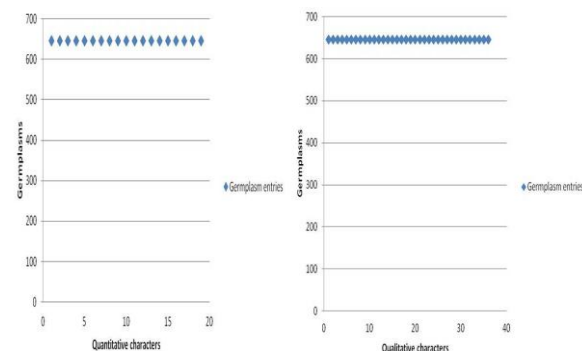

Fig.1 Scatter diagram depicts the Qualitative and Quantitative Characters Distribution in the Database.

3. WEB RETRIEVAL SYSTEM

The Vigna database was designed efficiently to provide search results and frequently update the data obtained from the NPRC, TNAU. However, its normalized table structure represents necessarily suitable for the retrieval system, which joins the relevant tables. The web search systems are connected to the web retrieval database. In this way, the application programs of the search systems are not directly affected by the modification of schema. The web retrieval database has been constructed by joining many tables. Paithumbase provides flexible user search by providing detailed search options, including the location of the germplasms (origin), accession and range. The retrieval system provides with a KML-formatted file to view the accessions originating from specific locations providing the Earth map and the map location, to be clearer with the location of the germplasm (Fig. 2). This database is provided with a user friendly query-able platform for the qualitative characters. The accessions along with the particulars are displayed by querying the database. This web retrieval system is expected to contribute to a greater understanding of the association of genetic characteristics with phenotype and as a unique resource for genetic diversity research [10].
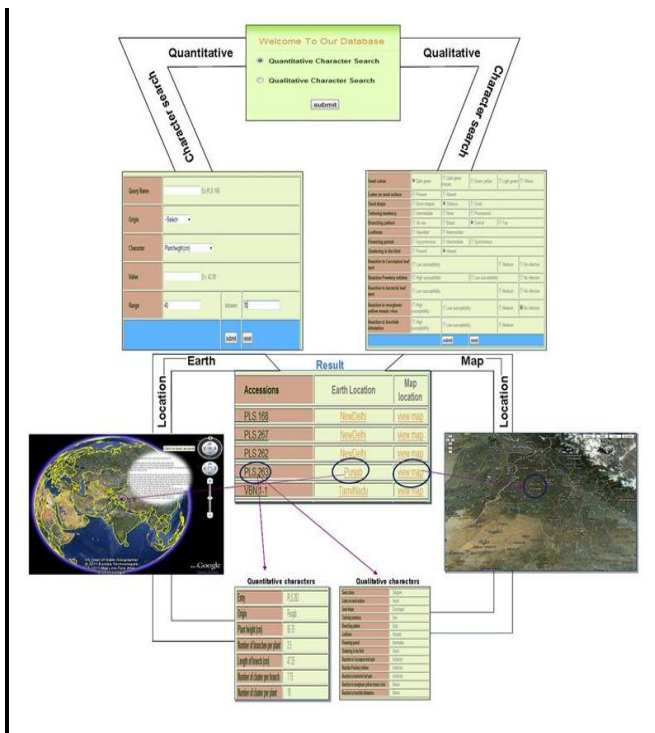
Fig 2. Paithumbase web retrieval system

4. EXTERNAL LINKS

In order to complement to Tamil Nadu Agricultural University, and provide additional information to the breeders and the farmers, Paithumbase is provided with links to TNAU Agri portal website, containing and providing information about resources, areas of cultivation, schemes and services and other link and forms. An accessible link to TNAU Genomics web portal and a link for the breeders, plant biologists and farmers for seed request are also incorporated. Thus, it works as a central hub of the mungbean by preserving, and distributing the knowledge available at TNAU.

5. RESULTS AND DISCUSSION

Germplasm collection was the engine of the crop improvement programmes [11]. In the past, much of the important information generated from this collection was accessible only locally, in field books or researchers' files. Major international initiatives for germplasm collection, conservation and much collected material are still not used because it is difficult to access. However, more and more institutions are generating experimental data on a large scale and are in need of developing and managing databases of their own. Effective use of this information may strongly promote the biological studies, and may lead to many important findings [12]. The primary goal of Paithumbase is to develop a data-integrative approach for *Vigna radiata*. Phenotype description ranges from quantitative traits like plant height to qualitative traits like reactions to bruchids infestation including all the experimental results from the field. Major objective is provide a versatile database structure that can be simple, require little maintenance, may be run on a desktop computer and yet, has the potential to be scaled to a large, well-curated database running on a server. The detailed and extensive description of the data should stay within the realm

of the organism-specific source databases. They have also discussed in more detail the associated difficulties of integrating phenotypic terminology that differs significantly between each organism-specific research community. We adopted a practical approach clearly intended to allow for high-level data integration and easy integration of new upcoming data. The data content will be updated frequently when there is a need. The *Vigna* database now has to prove that the method of integration applied here can add value to the scientific exploitation of phenome data [13]. The National Plant Genetic Resources Laboratory, the Asian Vegetable Research and Development Centre and the National Bureau of Plant Genetic Resources are involved in maintaining the germplasm collections, by preserving the seeds which are less useful for plant breeders. The Tropical African germplasm collections (303 accessions) are maintained in the National Gene Bank of Kenya and the Crop Plant Genetic Resources Centre, which are collections of field germplasms. The "ECPGR DATABASE", is maintained by the Austrian Agency for Health and Food Safety (AGES) in Linz, Austria. The database contains passport data representing the *Vigna* collections maintained in the European Gene Banks. It provides the search criteria for the crop name, institute code, sub taxa, species, origin, accession number, name, site etc., [14]. This provides less information for the plant breeders and biologists working on *Vigna* species. But the *Vigna* database provides more features with quantitative search like accession number, origin, a KML embedded viewer showing the location of *Vigna* species on an earth map and map location, character search, values and range search. Qualitative search allows the user to search through a series of 38 characters along with its descriptors to have a quick and clear idea for the geneticists and plant breeders working on this crop. An additional feature of link to TNAU Agritech portal webpage, TNAU genomics webpage, and a seed request mail form.The design of *Vigna* database provides a generic database framework from which the interfaces ranging from simple to complex may be used as a gateway to the data. In future we will have plan to update the genotype informations also. How ever. This design further provides less redundant data of the genetic diversity of species for the people interested in the diversity research.

6. REFERENCES

[1] Upadhyaya HD, Yadav D, Dronavalli N, Gowda CLL and Singh S. 2010. "*Mini core germplasm collections for infusing genetic diversity in plant breeding programs*". Electronic Journal of Plant Breeding, 1(4):1294-1309.

[2] Badaloo G H, Domaingue R and Ramdoyal K. 1998. "*A critical review of parental choice and cross prediction techniques in the msiri sugar cane breeding programme AMAS,*" Food and Agricultural Research Council, Reduit, Mauritius.

[3] Conlin D C. 2006. "*Introduction to Databases on The Web*",http://www.access.qbuilt.com/html/dbdriven_web_sites.html.

[4] Henry A. 2009. "*Using Google Earth for Internet GIS [MSc Dissertation] Geographical Information Science,*" Institute of Geography, School of Geosciences.

[5] Cosentino C. 2000. "*Essential PHP for Web Professionals*", Prentice Hall Computer Books, 1 edition. pp.205.

[6] Murukarthick J, Sreedevi G S, Senthil, N, Raveendran, M, *et al*. 2011. *A web accessible resource for investigating cassava phenomics and genomics information*. **Bioinformation** 6(10): 391-392.

[7] Murukarthick J, Senthil N, Raveendran M, Prabhakaran P, *et al.* 2011. *Biogen Base – An Interactive Maize Database for Phenomics Platform*. **International Journal of Computer Application.** 2: 56–61.

[8] Yamazaki Y and Jaiswal P. 2005. *Biological ontologies in rice databases. An introduction to the activities in Gramene and Oryzabase*. **Plant Cell Physiol** 46: 63-8. [Pmid: 15659431]

[9] Okuno K. 1998. "*Proceedings of the international symposium on rice germplasm evaluation and enhancement*", p.101.

[10] Heilmann C. 2007. *Beginning JavaScript with DOM Scripting and Ajax*, APress.

[11] Fabrice Grassein, Irène Till-Bottraud, Sandra Lavorel. "*Plant resource-use strategies: the importance of phenotypic plasticity in response to a productivity gradient for two subalpine species*," **Annals of Botany**. June 21, 2010. pp.154

[12] Masaru Takeya, Fukuhiro Yamasaki and Shihomi Uzuhashi. "*NIASGBdb: NIAS Genebank databases for genetic resources and plant disease information*," **Nucleic Acids Research.** 39(14): October 2010, doi:10.1093/nar/gkq916. 2011.

[13] The ECPGR Vigna Database. Austrian Agency for Health and Food Safety (AGES) Wieningerstrasse 8. A-4020 Linz, Austria http://www.genbank.at/en/ecpgr-vigna.html.

[14] Abdullah Kahraman, Andrey Avramov and Lyubomir G. Nashev. "*PhenomicDB: a multi-species genotype/phenotype database for comparative phenomics*," **Bioinformatics,** 21(3): 418-420. doi: 10.1093/bioinformatics/bti010. 2005.