UNDER PEER REVIEW

Research paper

TNAU PDB - Tamil Nadu Agricultural University Proteome DataBase - Black Gram Proteome

ABSTRACT

Tamil Nadu Agricultural University Proteome DataBase (TNAU PDB) – Black gram proteome is an open accessible database that focuses on proteome of Black gram (*Vigna mungo*). Currently the database contains reference maps based on Two-Dimensional Polyacrylamide Gel Electrophoresis (2D-PAGE) of proteins obtained from artificially aged black gram seeds which is compared with the fresh seeds. It provides information about experimentally identified properties such as molecular weight, pI value, differentially expressed protein due to accelerated ageing and protein sequences obtained using MALDI-TOF mass spectrometry. This database runs on WAMP server with HTML as the front end and MySQL as the backend using PHP as interface and it is hosted in TNAU genomics domain. The basic intention of this database is to provide the detailed information about proteome of black gram. This will help us to have improve understanding of adaptive and general protective mechanism related to seed aging and effect on germination .

Keywords: Accelerated ageing, Black gram, Database, Proteomics, Proteome database, TNAU PDB.

1. INTRODUCTION

Black gram (*Vigna mungo* L) is a protein rich food, containing about 26 per cent protein, which is almost three times that of cereals and ranks fourth among the major pulses cultivated in India. Black gram supplies a major share of protein requirement of vegetarian population of the country. It is consumed in the form of split pulse as well as whole pulse, which is an essential supplement of cereal based diet. In India, black gram occupies 12.7 per cent of total area under pulses and contribute 8.4 per cent of total pulses production. However, area and production of black gram has declined from 3.01 million ha and 1.30 million tonnes in 2000-01 to 2.97 million ha and 1.23 million tonnes, respectively in 2009-10 [1]. In the current scenario of flourishing omics segment that provides deep insight into the mechanism of crop physiology and assist in further crop improvement, proteomics of blackgram is not yet completely exploited as that of cereals like paddy [2]. Since analysis of the proteome provides a direct link of genome sequence with biological activity, data on proteomes are slowly mounting. Till date, there is no separate database concentrating on proteomes of black gram. Thus we had developed our database to integrate the essential data derived in our lab which would help the researchers to understand the biological process and thereby assists in further development. The main objective of this database is to promote quantitative and qualitative proteome research in black gram that helps in understanding the mechanism of various biotic and abiotic stress related

UNDER PEER REVIEW

proteins which in turn helps in annotating their function. Currently detailed proteome information about the blackgram pods are made available in the database.

2. METHODOLOGY

2.1. Database content and source

The black gram proteome database consists of proteome information about pods, leaves, stem and root. Currently, only proteome data of pods alone made available. Possible changes in proteins due to artificial ageing on dry seeds of blackgram were analyzed and the quantitative and qualitative proteome changes were recorded. Artificial ageing was done in an incubator maintained at 40°C and 95 per cent relative humidity for 6 days. Proteins were isolated by taking fresh seeds as control and 6 days artificially aged seeds as test. In the first dimension, 150µg of protein was loaded on a 17 cm IPG strip with a linear gradient of pH 4-7. In the second dimension, 12% SDS –PAGE gels were used with molecular weight standards. Proteins were visualized by silver staining as suggested by Blum *et al* [3]. Their properties such as molecular weight, pI values and their expression were calculated using image analyzer software. Significant spots were cut, digested in trypsin and analysed in MALDI-TOF for sequencing [4]. The generated data were uploaded in the database.

2.2. Languages and Software's Used

The front end of the web application is developed on HTML 5.0 (Hyper Text Markup Language) and the validations are done using javascript. The server side scripting was done on PHP 5.3.0 (Hypertext Pre Processor) and the application was connected to the database using MySQL 5.0.8. Web application was created by using wamp server 2.2 and each spot in the gel image was linked to the corresponding protein information with the help of Macromedia dreamweaver.

2.3. Database Architecture

The application is built on a three tier architecture model consisting of presentation layer, logic layer and the database layer. The presentation layer is the front end of the application created using HTML with which the user interacts. The middle layer is the application server or logical layer created using PHP, which serves the application with data and accepts the requests from the user. The data layer contains the relational database which contains the data to be fetched by the application.

3. RESULT AND DISCUSSION

UNDER PEER REVIEW

3.1. Flow of Data

TNAU genomics is the home page under the link http://www.tnaugenomics.com/, from link TNAU PDB under the tab databases, proteomics data of blackgram can be accessed as shown in the Fig (3.1.)

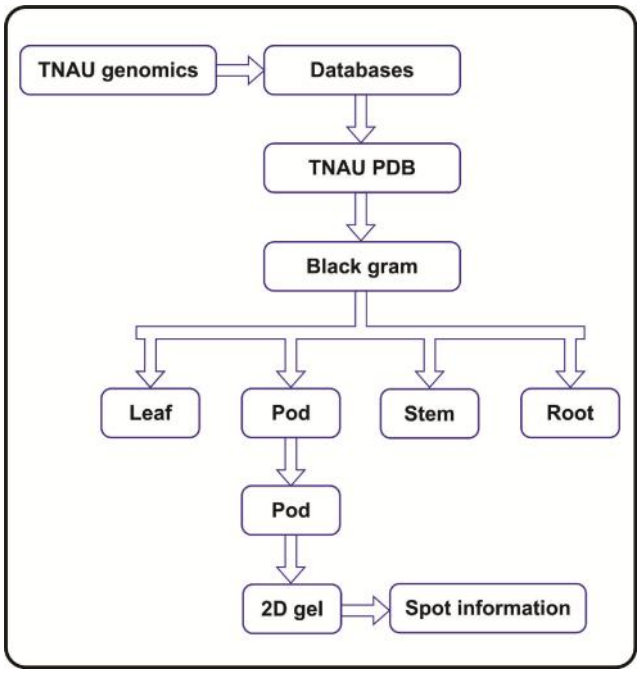

Fig. 3.1. Flow of data

3.2. Database schema and its features

The data is classified based on the tissue such as pods, leaf, stem, root, etc. from which it is identified. Proteomes obtained using the 2D PAGE and MALDI-TOF methods are displayed using image analyzer software. The description of each spot is stored in the database and they are mapped to the corresponding spots. Based upon the request from the user the application queries the database and fetches the spot information.

Currently, database contains the proteome information obtained from pods of black gram.The reference 2D-PAGE gel shows the position of each identified protein in that crop. The entire list of protein can also be obtained by selecting the crop name displayed inside the table. By selecting the spot in the gel image the entire information about that protein can be obtained. The results are displayed in such a manner that they are also compared with their control (Fig. 3.2). All the reference maps are also displayed under 2D gel section for a quick reference. The experiment protocols are listed under the protocol section. Some of the major proteomics tools like includes Mascot, Compute pI/Mw tool in ExPASy and ExPASy Proteomics tools are displayed under proteome tools section. Similar database for maize was also published for our laboratory [5].

UNDER PEER REVIEW

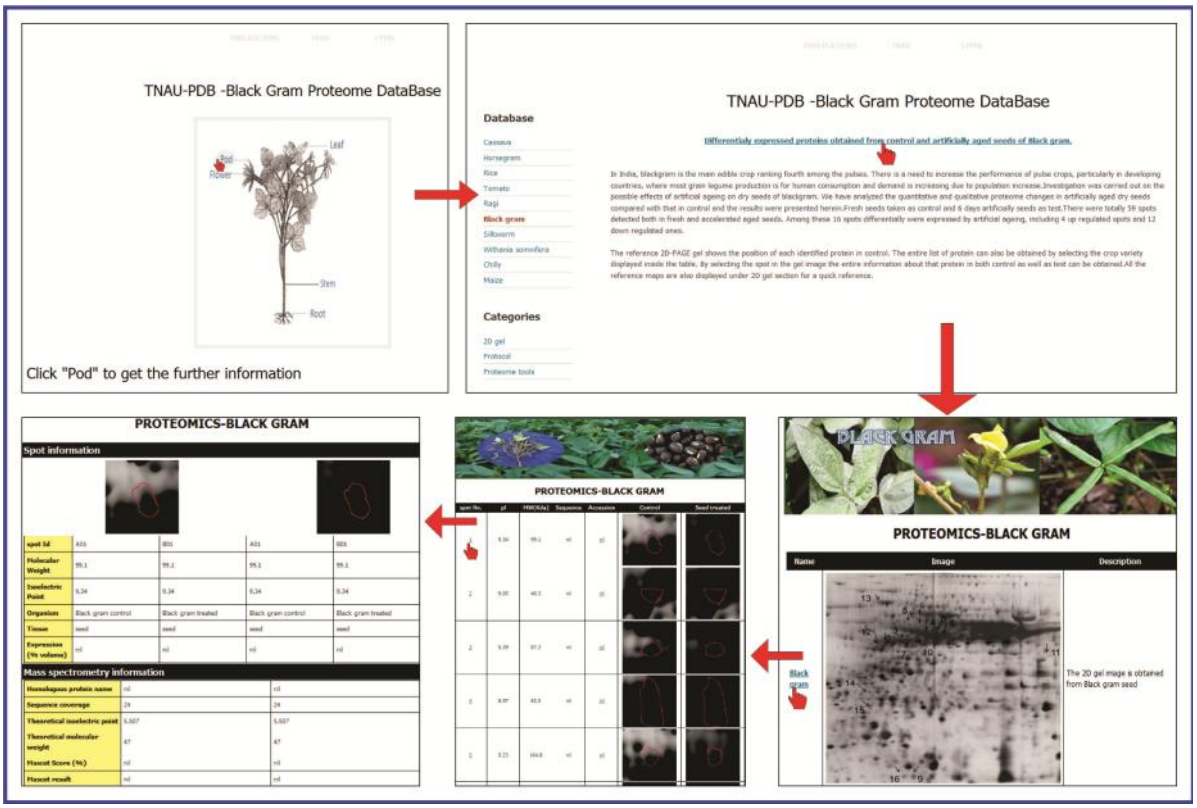

Fig. 3.2. Procedure to access differentially expressed proteins due to artificial ageing in blackgram seeds

3.3. Utility

The primary users of this database will be plant breeders, seed technologist, students and other researchers concentrating on blackgram improvement. It shows the proteome changes that could take place at the dry state of aged seeds. Determination of the expression patterns in response to stress, and an understanding of their functions in stress adaptation will provide the researchers with the basis for effective genetic engineering strategies for improving the tolerance of crops to various stresses.

3.4. Future Developments

The database content will be updated routinely. In the near future proteome data for whole plant parts such as root nodules, stem, leaves and expression of protein under different environmental stress conditions will be made available. With the availability of ample proteome data, architecture of the web page might be modified to suit to the excess flow of information.

AUTHORS' CONTRIBUTIONS

This work was carried out by the first author SS in collaboration between all authors. Guided by the expertise advice of the second author NS and last author MR, the corresponding author NS designed

UNDER PEER REVIEW

the study, first author SS and third author VVK designed the web page and wrote the first draft of the manuscript. Fourth author NJS and fifth author SS did the proteomic analysis. All the authors managed the analyses of the study. However, all authors read and approved the final manuscript.

REFERENCE

1. ASSOCHAM. Study paper on “Emerging pulses scenerio in 2015 - productivity, awareness and affordability to drive pulses economy in India”. The Associated Chambers of Commerce and Industry of India, New Delhi, India. 2012; 1 – 18.
2. Setsuko Komatsu, Keiichi Kojima, Kouji Suzuki, Kazuo Ozaki, Kenichi Higo. Rice Proteome Database based on two-dimensional polyacrylamide gel electrophoresis: its status in 2003. Nucleic Acids Research, 2004; 32 (Database issue): D388 - D392.
3. Blum H, Beier H, Gross HJ. Improved silver staining of plant proteins,RNA and DNA in polyacrylamide gels. Electrophoresis 1987; 8: 93-99.
4. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. 1996; 68: 850-858.
5. Murukarthick, J., Senthil N, Raveendran M, Prabhakaran P, Sreedevi G, Sumanth Kumar M, Shobhana VG, Dhanya S, Khushboo R, Arumugasamy S, Ravikesavan R, Jagadish SVK. 2011. Biogen Base – An Interactive Maize Database for Phenomics Platform. International Journal of Computer Application. 2: 56–61.