A web accessible resource for investigating cassava phenomics and genomics information: BIOGEN BASE

Murukarthick Jayakodi[1], Sreedevi Ghokhilamani selvan[1], Senthil Natesan[1]*, Raveendran Muthurajan[1], Raghu Duraisamy[1], Jana Jeevan Ramineni[1], Sakthi Ambothi Rathinasamy[1], Nageswari Karuppusamy[2], Pugalenthi Lakshmanan[3], Mohan Chokkappan[4]

[1]Genomics and Proteomics Laboratory, Department of Plant Molecular Biology and Biotechnology, Centre for Plant Molecular Biology, Tamil Nadu Agricultural University, Coimbatore - 641 003; [2]Tapioca and Castor Research Station, Yethapur, Tamil Nadu Agricultural University, Coimbatore - 641 003; [3]Horticultural College and Research Institute, Tamil Nadu Agricultural University, Coimbatore – 641 003; [4]Central Tuber Crops Research Institute, Trivandrum, India; Natesan Senthil - Email: senthil_natesan@tnau.ac.in; *Corresponding author

Abstract:
The goal of our research is to establish a unique portal to bring out the potential outcome of the research in the Casssava crop. The Biogen base for cassava clearly brings out the variations of different traits of the germplasms, maintained at the Tapioca and Castor Research Station, Tamil Nadu Agricultural University. Phenotypic and genotypic variations of the accessions are clearly depicted, for the users to browse and interpret the variations using the microsatellite markers. Database (BIOGEN BASE - CASSAVA) is designed using PHP and MySQL and is equipped with extensive search options. It is more user-friendly and made publicly available, to improve the research and development of cassava by making a wealth of genetics and genomics data available through open, common, and worldwide forum for all individuals interested in the field.

Availability: http://www.tnaugenomics.com/biogenbase/casava.php.

Keywords: TNAU Genomics, cassava database, MySQL, PHP, and Bioinformatics.

Background:
Cassava, *Manihot esculenta* Crantz, is one of the important staple food crops in the world which is being grown under varying environmental conditions. Cassava grown in Tamil Nadu, Southern part of India is analyzed for the varietal variability and is studied extensively. There are a lot of databases operating on cassava, having cDNA, mRNA, ESTs, an online archive in the Cassava Database Community, helping the researchers. The current database provides accessions which are analyzed phenotypically and genotypically for understanding their variations at the allelic level. The Tapioca and Castor Research Station, maintains the field gene banks which provide variation both qualitatively and quantitatively. The accessions are analyzed for their morphological traits and marker details and stored in the database. The variations in the accessions can be clearly studied by querying the search options. The phenotype search, genotype search, variation search through images and the frequency distribution by graphs are retrieved by querying the database which is publicly accessible via the World Wide Web. The information obtained from SSR markers is also useful for making predictions about crossing and selection aimed at increasing the efficiency of parental selection and varietal development. The linkage groups and the references under which the SSR Markers are mapped are also identified and stored in the database.

Methodology:
The entries of BIOGEN BASE - CASSAVA database are generated from the text mining of the enormous work going in the genomics and proteomics laboratory of Tamil Nadu Agricultural University. BIOGEN BASE - CASSAVA is designed using PHP and MySQL [1]. The accessions cover both qualitative and quantitative traits. A GUI interface is developed in PHP to access the germplasm and genotype data via web (http://www.tnaugenomics.com/biogenbase/casava.php). The database administrators are given the backend access to add, modify and remove data from the database. The frontend PHP webpage connects to MySQL database hosted on tnaugenomics web server to retrieve the data [2]. Information such as Germplasm, Phenotype, Genotype and Variations are stored in the database. The assessed data of germplasm are captured in germplasm table. The phenotype [3] table consists of shape of the plant, color of the bud, emergence

of eye (bud), alignment of internodes, color of the young leaf, shape of the leaf lobes, color of the petiole, color of the upper surface of lamina, colour of the under surface of lamina, colour of the lamina nerve, colour of the young stem, shape of the tuber, colour of the outer skin of the tuber, colour of the inner skin of the tuber, colour of pulp, tuber attachment on parent cutting, incidence of the cassava mosaic disease, starch content (%), height of the plant (m), number of leaf lobes (nos.), length of the petiole (cm), number of branches (nos.), number of tubers per plant (nos.), length of the tuber (cm), girth of the tuber (cm) and yield per plant (kg). The genotype table contains information about the amplicon size, gel image, linkage group, forward primer, reverse primer, allelic size of different varieties including SSR marker. The variation table consists of images for all the qualitative traits and graphs for all the quantitative traits. The database is supported with a user manual to have a better understanding on the database.
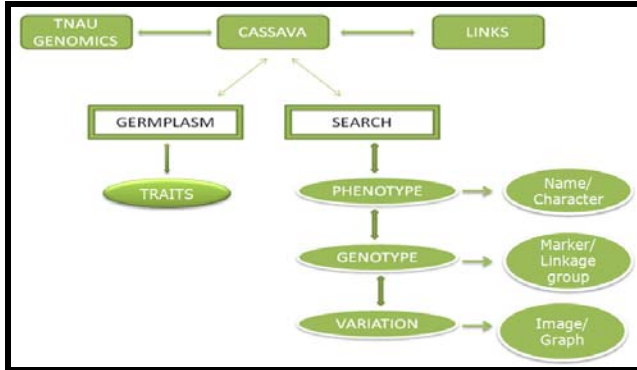

Figure 1: Database Schema

Utility:

BIOGEN BASE - CASSAVA was created to combine data obtained by genotyping activity of the CASSAVA research group with the available molecular and phenotypic data in order to create a unique site and tool for Cassava. The database as it is can have a broader platform of users like public researchers and breeders, from private, regional and national regions. Its maintenance and further improvement is ensured, and it is our interest to keep it updated and usable in the long term. In fact, breeders use information and make crosses and selection, also marker-assisted, with the use of this database. To date, some other databases such as AppleBreed, Panzea, Germinate and PlantDB, Maize GDB, RIKEN Cassava Full-Length cDNA Database **[4]**, Cassava Online Archive **[5]**, have been designed to provide a way to store and correlate genotypic to phenotypic information on plant genetic collections, including genetic stocks and commercial varieties. With respect to the existing databases, the unique way in which the BIOGEN BASE database is designed, enables an intensive cross-talk between all the integrated data sources. In fact, BIOGEN BASE provides an integrated System of data sources already existing, to join the data produced by the genomics and proteomics laboratory, to facilitate breeders in better approaching the large resources available for molecular breeding. It represents a first attempt of simplification from different sources with the mainly applied purpose of giving a tool for marker-assisted breeding. The BIOGEN BASE database is, as far as the authors know, the first database that enables querying both molecular and phenotypic data specific for Cassava (**Figure 1**). It supports breeders and geneticists in their exploration of germplasm, available in various collections. The information stored in this database helps the breeders not only to understand the genetics of economically important traits and to identify marker-trait associations, but also enables them to choose the most interesting genotypes that should be crossed to obtain a desired progeny. The variation and frequency distribution graph may help the users of this database to be more flexible.

References:
- **[1]** http://www.php.net/
- **[2]** Murukarthick J *et al*. *International Journal of Computer Applications* 2011 2: 56
- **[3]** Raghu D *et al*. *International Journal of Integrative Biology* 2007 1(2):141-148.
- **[4]** http://amber.gsc.riken.jp/cassava/
- **[5]** http://cassava.psc.riken.jp/index.pl

Edited by P Kangueane
Citation: **Murukarthick** *et al*. Bioinformation 6(10): 391-392 (2011)