Bioinformation by Biomedical Informatics Publishing Group

_____

A database for coconut crop improvement

Velamoor Rajagopal[*], **Ramaswamy Manimekalai, Krishnamurthy Devakumar, Rajesh, Anitha Karun, Vittal Niral, Murali Gopal, Shamina Aziz, Marimuthu Gunasekaran, Mundappurathe Ramesh Kumar and Arumugam Chandrasekar.**

Central Plantation Crops Research Institute, (ICAR), Kudlu P.O, Kasaragod – 671124, Kerala, India;
Velamoor Rajagopal* - Email: bioinfcpcri@yahoo.co.uk; * Corresponding author

Abstract:

Coconut crop improvement requires a number of biotechnology and bioinformatics tools. A database containing information on CG (coconut germplasm), CCI (coconut cultivar identification), CD (coconut disease), MIFSPC (microbial information systems in plantation crops) and VO (vegetable oils) is described. The database was developed using MySQL and PostgreSQL running in Linux operating system. The database interface is developed in PHP, HTML and JAVA.

Availability: http://www.bioinfcpcri.org

Keywords: coconut; crop improvement; database; oils; germplasm; plant disease; cultivar

Background:

Coconut (*Cocos nucifera* L.) is an important multipurpose perennial oil crop of humid tropics. It is grown in nearly two million hectares in India. It supports millions of small and marginal farm families. The crop also serves many people in the fragile ecosystem of coastal and island areas. Therefore, it is important to develop strategies to improve coconut crop productivity. Coconut crop improvement involves systematic monitoring of parameters essential for crop development. This could be achieved by developing databases containing information similar to many other plant databases. [1,2,3] Here, we describe the development of a database containing information for coconut crop improvement.

Methodology:

Database Implementation:

Data is stored in MySQL and PostgreSQL running in a Linux server. The database architecture is shown in Figure 1A.

Database Interface:

The database interface is implemented using PHP, HTML and Java. [4] The sample interface is given in Figure 1B.

Database description:

The Coconut Cultivar Identification Database (CCID) was created by representing CCID information using binary data (1 or 0) of six isozymes (PPO - Polyphenol oxidase, EST-Esterase, PER- Peroxidase, MDH - Malate dehydrogenase, ADH - Alcohol dehydrogenase, GOT- Glutamate oxaloacetate transaminase) and microsatellite markers (Figure 1C). [5] The database can be queried for the above mentioned isozymes and microsatellite alleles. [2] The query

is matched with the database content and the possible match for each coconut cultivar is obtained. For each cultivar, accession details, morphological yeild, genetics related and biochemical characteristics are provided in the database.

CPCRI (Central Plantation Crops Research Institute) holds the largest coconut germplasm collection comprising of 353 accessions including 132 exotic (from 22 countries) and 221 indigenous accessions. The coconut germplasm database has the passport data, morphological, genetic characteristic data, accession number and the place of collection. Same interface also provides the details of passport data with sub field of accession and collection data. Characterization data includes sub fields of germination, plant morphology, inflorescence, fruit, flower, copra yield, oil content and pictures of fruit/flower for selected cultivar (Figure 1D).

The coconut disease database consists of fungal pathogens of coconut. The database contains 303 entries of nucleotide sequences for *Colletotrichum gloeosporioides*, 3 entries of nucleotide sequences for *Exerohilum rostratum*, 64 entries of protein sequences for *Colletotrichum gloeosporioides* and 1 entry of sequence for *Exerohilum rostratum*. Nucleotide and protein sequences for *Ganoderma lucidum* and *Ganoderma applanatum* pathogens were collected from NCBI, EMBL and stored in PostgreSQL database. [6,7] Provision was also made to locate the publications on these topics and can be queried using KEYWORD. Options were also provided to save the sequence file in a *.txt format and FASTA format (Figure 1E).

Figure 1: (A) The relational database management architecture is shown. **MIFSPC** = Microbial information system of plantation crops; **CGD** = Coconut Germplasm Database; **VOD** = Vegetable Oil Database; **CCID** = Coconut Cultivar Identification Database; **CDD** = Coconut Disease Database; **INFOSYS =** Information System; (B) Search query interface for coconut cultivar identification based on Isozyme markers; (C) Search query interface for coconut cultivar identification based on SSR markers; (D) Search query interface for coconut germplasm database; (E) Search query results for coconut disease database; (F) Overall database structure for the vegetable oils; (G) Search query interface for the vegetable oil database; (H) Search query interface for CPCRI nutrient study; (I) Search query interface for Microbial Information System on Plantation Crops

The vegetable oil database provides information for nutritional, health, value addition and quality control using the interface. Information is available on coconut, corn, olive, palm, palm kernel, peanut, safflower, seasame, soybean and sunflower oil (Figure 1F, G & H). The variation in the fatty acid profile between different hybrids of coconut can be obtained from the vegetable oil database.

The Microbial Information Systems provides the role of microbes in plantation crops and research related data on bio-inoculants, bio-control agents, decomposers, edible mushrooms and photo-galary. Information on microorganisms obtained from PHYLOBASE **[8]** and other published sources were also stored in the database. This gives detailed information on nitrogen fixers, phosphate solubilizers, plant growth promoting *Rhizobacteria's*, vesicular arbuscular mycorrhizae, insect pathogens and antagonists. The DECOMPOSERS suite provides information on cellulose and lignin degraders. Details of edible mushrooms like *Pleurotus*, *Calocybe* and *Agaricus* are also available with corresponding nucleotide/protein sequences **[9,10]** and protein 3D structure **[11]** for every species in the database (Figure 1I).

Utility:
The database finds utility in coconut crop improvement. The database is freely available in public domain and login information is available from the authors.

Future developments:
We plan to include datasets of other crops like cocoa and arecanut. An advanced graphical tool will be added to facilitate display and visualization of the underlying data.

Bioinformation
by Biomedical Informatics Publishing Group

open access

www.bioinformation.net

Web Database

_____

References:
[1] http://www.ncbi.nlm.nih.gov/Genbank/
[2] J.. Macas, *et al., Bioinformatics,* 18:28 (2002) [PMID: 11836208]
[3] N. Srinivasan, *et al., Ann.Appl.Biol.*, 129:433 (1996)
[4] S. Rudd, *et al., Nucl. Acids Res.*, 33:D628 (2005) [PMID: 15608276]
[5] A. J. Robinson, *et al., Bioinformatics,* 20:1475 (2004) [PMID: 14962913]
[6] T. Kulikova, *et al., Nucleic Acids Res.,* 32**:**D27 (2004) [PMID: 14681351]
[7] D. L. Wheeler, *et al., Nucl. Acids Res.*, 33:D39 (2005) [PMID: 15608222]
[8] http://62.245.228.5/
[9] http://www.ebi.ac.uk/
[10] http://www.ncbi.nlm.nih.gov/unigene/
[11] http://www.pdb.bnl.gov/

Edited by P. Kangueane
Citation: Rajagopal *et al.,* Bioinformation 1(2): 75-77 (2005)