Short scientific report *Journal of Plantation Crops, 2008, 36 (1): 61-64*

COCOS-SATDB: A Microsatellite Database for Coconut (*Cocos nucifera* L.)

(Manuscript Received: 14-08-07, Revised: 15-10-07, Accepted: 07-02-08)

Keywords: coconut, database, microsatellites

The last decade has witnessed the rapid buildup of sequencing data from many organisms. This has necessitated the development and utilization of new tools, which may allow simple access and analysis of this vast information. Design and development of databases, which may cater to the particular sequence type of any one organism, may provide efficient tools for retrieval and analysis of useful data.

Microsatellites or simple sequence repeats (SSRs) are short tandemly repeated sequence motifs of approximately 1-8 bp in length, which are scattered throughout the genome of both prokaryotes and eukaryotes (Dong *et al.,* 2005). The origin of such polymorphism appears to be most likely due to slippage events during DNA replication (Schlotterer and Tautz, 1992). They are present in both coding and non-coding regions and are characterized by a high degree of length polymorphism (Andrew *et al.,* 2004). They have numerous applications in plant breeding: phylogenetic studies, gene tagging, genome mapping, variety protection, QTL analysis etc.

SSRs are becoming increasingly attractive markers in coconut genetic diversity studies and QTL mapping (Perera *et al*., 1999; Rivera *et al*., 1999; Baudouin *et al*., 2006; Devakumar *et al*., 2006). In the present study, the development of a coconut microsatellite database is described which would provide centralized access to all the publicly available coconut microsatellites.

For database implementation, data was stored in MySQL running on a Linux server. The three-level scheme of Prasad *et al.* (2005) was followed in the development of database (Fig.1). The database interface

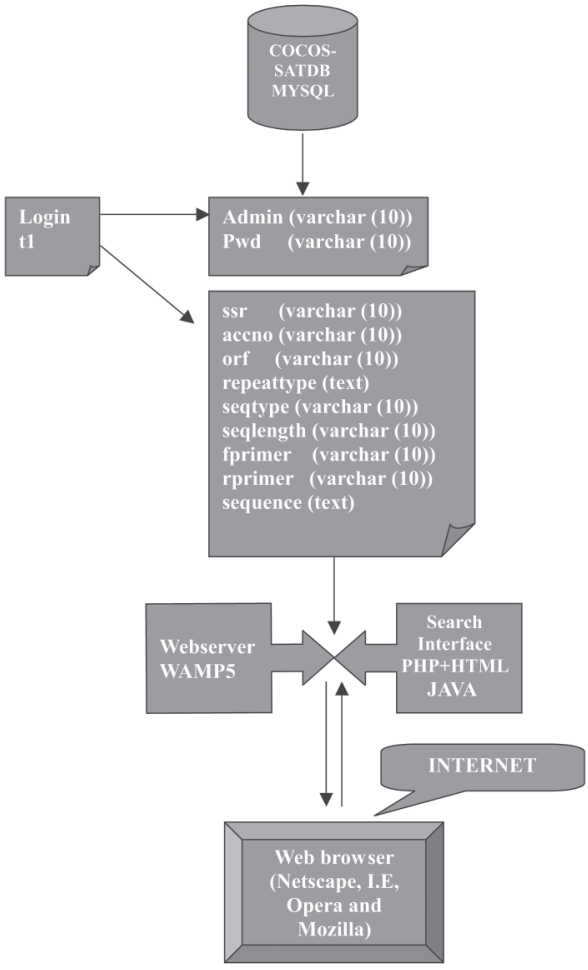

Fig. 1. Workflow schematic diagram illustrating the multiple forms of data interrogation

was implemented using PHP, HTML and JAVA. The sample interface is given in Fig. 2.

Each entry in the Coconut Microsatellite Database includes an identifier for the 267 SSRs and its

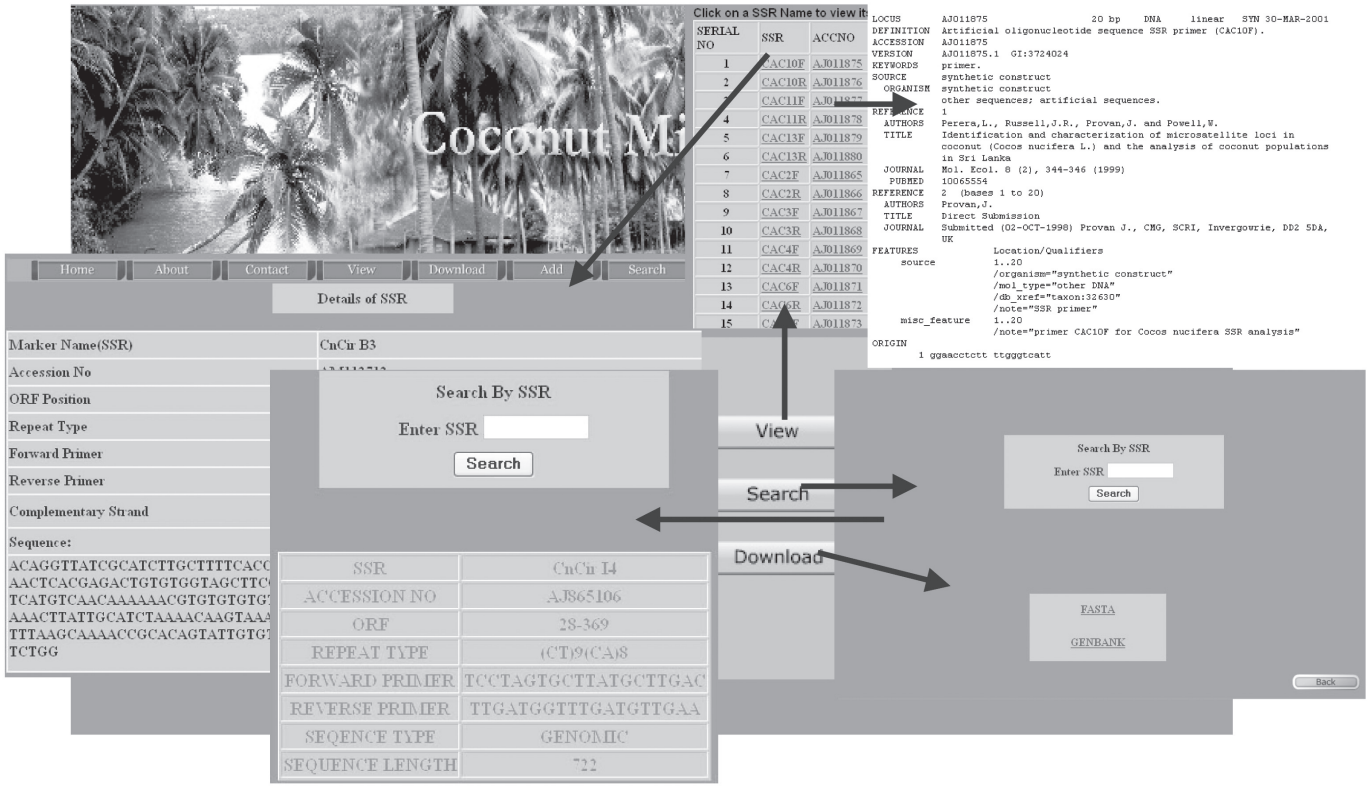

Fig. 2. The overall database search interface

corresponding Accession Number. The Accession Number is an alphanumeric number, which identifies the complete detail of a particular clone (SSR). The database contains the ORF (Open Reading Frame) for each of the SSRs specified and the sequence of each marker.

The user-friendly interface for the database has been developed using PHP, a server side scripting language with MySQL as the backend. This acts as a comprehensive and integrated resource for retrieval of the information from Coconut Microsatellite Database.

The query results are displayed in tabular format showing the SSR name, the Accession Number, the ORF, the repeat type, sequence length, sequence type and the FASTA format of the corresponding SSR. Links are provided for the SSR, Accession Number and the display format (FASTA) in the table. Linked SSR name takes the user to the respective page where the details of SSR including Forward Primer, Reverse Primer and sequence of that particular SSR are displayed. Accession Number link takes the user to the GENBANK details (Dennis *et al.,* 2000) and by clicking FASTA, the complete sequence of the particular SSR can be viewed.

The individual project pages contain access to all public data currently available for coconut microsatellites. The standardized project information includes Accession Numbers, sequences in GENBANK and FASTA format and the ORF details. Standardized panel screened data are available for download directly from the project page. Currently, this database houses information on 267 annotated coconut microsatellites, which can be viewed. A microstaellite information page displays the sequences along with the repeat type and primers (both forward and reverse primers).

Database search interface: The initial SSR search result page displays SSR identifiers. The individual SSR entry links to a page where details of the SSR are displayed (Fig. 3) with links to the corresponding pages. Markers can be searched by marker name. With the accession number, GENBANK details can be retrieved.

The web pages are organized such that users can easily access the data of interest regardless of the navigation starting point. For example, the home page has links to view the details of all the SSR from which marker details, sequence file in FASTA format and GENBANK details regarding each Accession Number can be viewed. A general tool bar is also included in each page for the ease of navigation through the site. This is a new resource freely accessible at CPCRI Bioinformatics Centre Website (http://

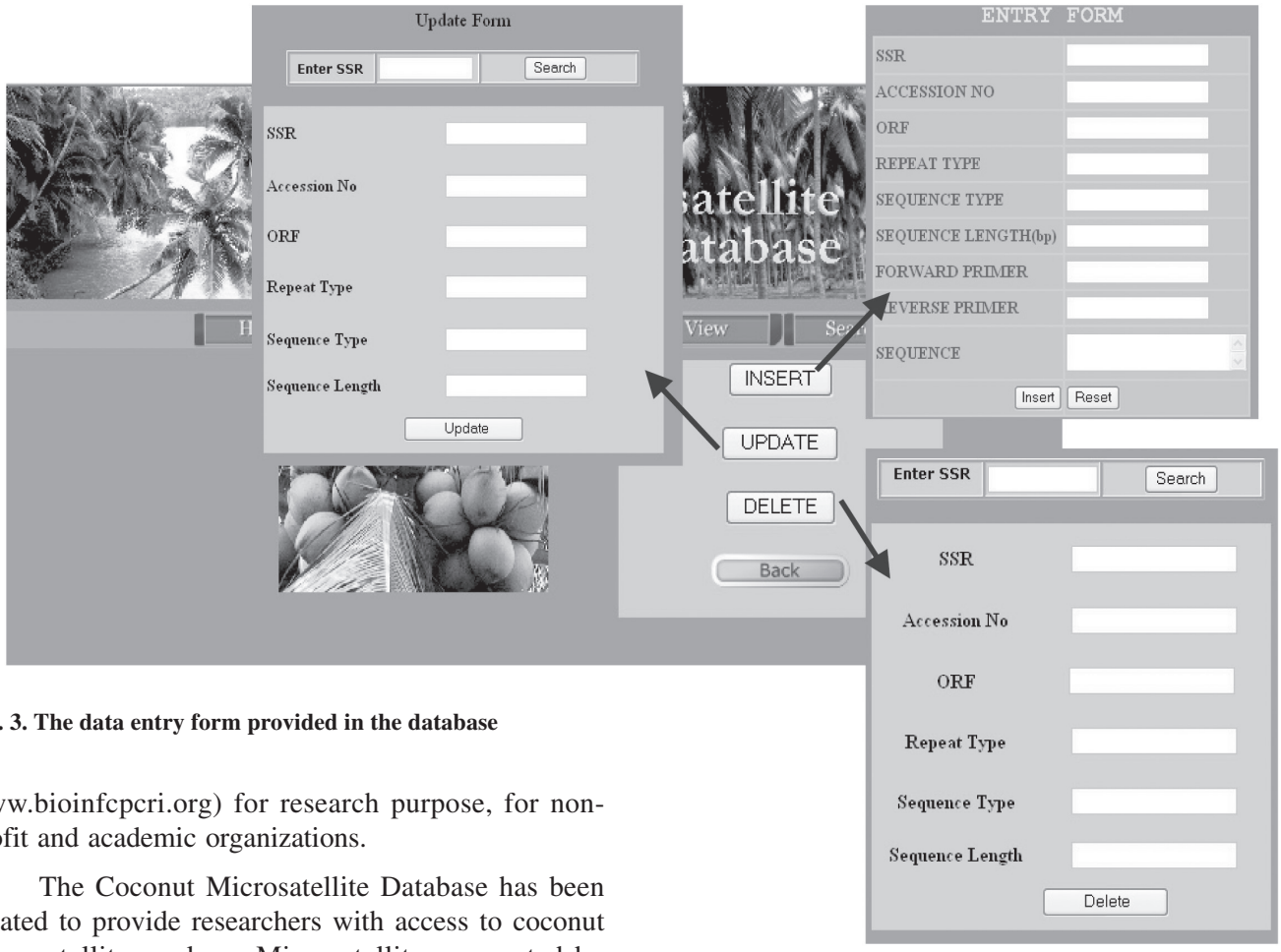

Fig. 3. The data entry form provided in the database

www.bioinfcpcri.org) for research purpose, for non-profit and academic organizations.

The Coconut Microsatellite Database has been created to provide researchers with access to coconut microsatellite markers. Microsatellites generated by different research groups have been included in the database. Access to this database is provided through integrated web tools, which allow users to access individual or combined project data via search interfaces. It also provides download and view of microsatellites, their primers, Open Reading Frames and their subsequent repeat type. Overall this database will be a good resource for coconut researchers involved in fingerprinting/ character tagging using microsatellites.

Future development will focus on incorporation of additional microsatellites to the existing data as and when they are developed and validated for improving the tools and functionality of the web interface. Annotation of SSRs with known homology will also be carried out.

Acknowledgement

We thank the web resource providers for data collection and development. This work was supported by a grant from the Govt. of India, Department of Biotechnology (BTISnet), New Delhi - 110003.

References

Andrew J. R., Christopher G., Jacqueline, B., Gary Barker and David, E. 2004. Simple sequence repeat marker loci discovery using SSR primer. *Bioinformatics* **20**: 1475-1476.

Baudouin L., Lebrun, P., Konan, J.L., Ritter, E., Berger, A. and Billotte, N. 2006. QTL analysis of fruit components in the progeny of a Rennell Island Tall coconut (*Cocos nucifera* L.) individual. *Theor. Appl. Genet*. **112**: 258-268.

Dennis, A. B., Karsch-Mizrachi, I., David J. L., James O. and David L.W.2000. Genbank. *Nucl. Acid Res*. **28**: (1): 15-18.

Devakumar, K., Jayadev, K., Rajesh, M.K., Chandrasekar, A., Manimekalai, R., Kumaran, P.M. and Parthasarathy, V.A. 2006. Assessment of genetic diversity of Indian coconut accessions and their relationship to other cultivars using microsatellite markers. *Plant Genetic Resources Newsletter* **145**: 38-45.

Dong, Q., Shannon, D., Schlueter and Brendel, V. 2005. Plant GDB, plant genome database and analysis tools, *Nucl. Acid Res*. **32**: 354-359.

Perera, L., Russell, J.R., Provan.J. and Powell, W. 1999. Identification and characterization of microsatellite loci in coconut (*Cocos nucifera* L.) and the analysis of coconut populations in Sri Lanka, *Molecular Ecology* **8**: 335-346.

Prasad M. D., Muthulakshmi, M., Arunkumar, K. P., Madhu, M., Sreenu, V. B., Pavithra, V., Bose, B., Nagarajaram, H.A., Mita, K., Shimada, T. and Nagaraju, J. 2005. SilkSatDb: a microsatellite database of the silkworm, *Bombyx mori, Nucl. Acid Res.* **33**: 403-406.

Rivera, R., Edwards, K.J., Barker, J.H.A, Arnold, G.M., Ayad, G., Hodgkin, T. and Karp. 1999. Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. *Genome* **42**: 668-675.

Schlotterer, C. and Tautz, D. 1992. Slippage synthesis of simple sequence DNA, *Nucl. Acid Res*. **20**: 211–215.

Bioinformatics Centre,
Central Plantation Crop Research Institute (ICAR),
Kasaragod - 671124, Kerala, India.

Praseeda, B,
Chandrasekar, A,
Rajesh, M.K,
Anitha Karun,
George V. Thomas