Construction of EST-SSR Marker Database In Coconut

Preethi P.[1], Divyalakshmi B.[1], Naganeeswaran S.[1], Hemalatha N. [2,*] and Rajesh M.K.[1]

ICAR-Central Plantation Crops Research Institute, Kasaragod, Kerala, India[1]

St. Aloysius Institute of Management & Information Technology, Mangalore, Karnataka, India[2]

ABSTRACT: The 'Coconut EST-SSR Marker Database' is a curated and integrated web-based relation database providing access to coconut microsatellites. Using this database, users can access information of EST-SSR markers in coconut, designed from transcriptome data, and gene sequences used to generate these markers. Users are provided with provision to upload EST-SSR data in FASTA formats. Results are also available in downloadable formats. This readily accessible web-enabled database provides a more efficient utilization of molecular marker resources in coconut and will help accelerate basic and applied research in coconut molecular breeding.

KEYWORDS*:* coconut; molecular marker; EST-SSR; database

I. INTRODUCTION

Coconut (*Cocos nucifera*) belongs to the monocotyledonous family Arecaceae. It is perennial oil yielding economic tree crop of tropics. There are two main types of coconut: 'tall', the naturally cross pollinating group with more economic value, and 'dwarf', the naturally self-pollinating group with reduced size and growth habit. It is believed that the 'dwarf' originated from earliest 'tall' coconut palms and maintained most of its original genome because of its autogamous behavior [1-3]. Evaluation of genetic diversity is an important step towards effective utility in coconut breeding programmers. The diversity analysis of a coconut germplasm has traditionally been based on morphological and biochemical features. Earlier attempts made however, have shown a few inflorescence descriptors and fruit components as useful descriptors in assessing coconut germplasm [2]. Efforts made to reveal coconut genetic diversity by isozyme analysis too had met with meager success, mainly due to technical limitations [4, 5]. The conventional method for studying coconut diversity has so many other problems like long growth period, large area required for planting, inherent cross pollinated nature of the crop, and the laborious work involved. All these limitations can be overcome by molecular markers.

Various molecular markers have been used for coconut genetic studies viz., RAPD [6, 7], RFLP [2], AFLP [8] and SSR [9-13]. Simple Sequence Repeats (SSRs) or microsatellites are becoming the most important molecular markers in both animals and plants [14]. SSR markers or microsatellite are tandem repeats (mono, di, tri, tetra and above nucleotide units) interspersed throughout the genome and can be amplified using primers that flank these regions. These markers are an ideal tool for genetic diversity and relationship studies due to their high information content, reproducible, co-dominant and multi-allelic nature and high discriminating power [15, 16] and also SSRs cover the whole genome and have become markers of choice for diversity analysis and genome analysis [17- 21]. SSRs are highly abundant and exhibit broad levels of polymorphism in eukaryotic [22, 23] and prokaryotic genomes [24, 25]. Their variability in length is caused by slip-strand mutations and that may affect the local structure of DNA molecule or encoded protein [26]**.** The identification and production of new SSRs from genomic DNA libraries are laborious and need more time which discounted the advantages of the molecular marker [27]. SSRs can be developed at cheaper costs through the mining of transcriptome databases using various software's. Plant tissues can be used to generate cDNAs from mRNA and sequenced to generate ESTs that are assembled into a non-redundant set of sequences (contigs and singletons).

Molecular markers developed from ESTs are known as EST-SSR markers which are easily obtained by electronic search of EST databases and are useful in developing linkage maps and in marker assisted breeding programmers [28, 29]. The SSRs marker developed from expressed sequence tag (ESTs) sequence can be used as an

efficient functional marker in genic region [30]. Only two earlier studies, viz., Xiao *et al.* [31] and Xia *et al.* [32] have reported development of gene-based SSR markers (EST-SSRs) from coconut transcriptome data. Publically available gene based SSR markers are few which limits the scope for molecular breeding in coconut palms. Also, there is no EST-SSR marker database in case of coconut palm. Construction of a EST-SSR marker database would definitely help coconut researchers and formed the objective of this study.

II. DATABASE DESIGN AND IMPLEMENTATION

Database design

The design of the 'Coconut EST-SSR Marker Database' followed three schema architecture (Fig. 1). The database was developed in Windows environment using Adobe Dreamweaver as IDE (Integrated Development Environment). Dreamweaver was used because it is a more fully featured HTML web and programming editor. The whole database interfaces were coded using PHP (Preprocessor Hyper Text) version 5.3.4 and HTML (Hyper Text Markup Language). PHP is a widely used open source general purpose scripting language especially suited for web development and can be embedded in to HTML. Other than user interfaces, every database should need some database management systems. We have used custom MySQL database version 5.1.53. It is the most popular open source relational SQL (Structured Query Language) management system. MySQL acts as a container that holds all the types of data such as text, audio, images etc. Another most important part of database is "Apache", the webserver within the Windows. By running a local apache server on windows, the webpage's can be viewed in a browser. All the animations and validations in the database were done using JQUERY version 1.11.3 i.e., fast, small, feature rich JavaScript library. Database also includes Cascading Style Sheets (CSS) that deals with looks and formatting of the user interfaces.

Database Implementation: Database was developed in Windows environment, and hosted in Linux platform.
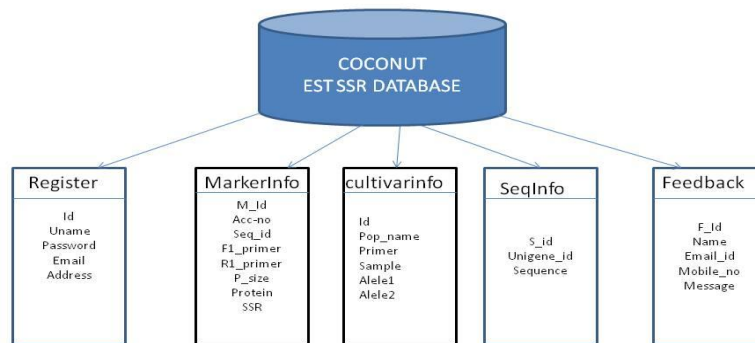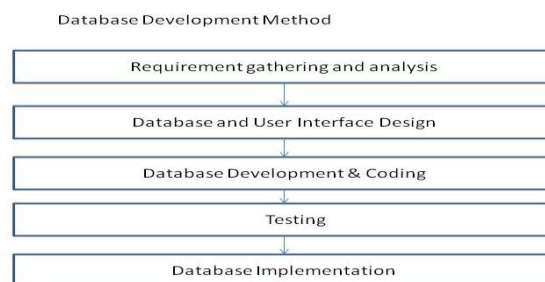

Fig. 1. Overall schema of coconut EST-SSR marker database

III. RESULTS

Database description: The 'Coconut EST SSR Marker Database' is a multiuser system. Only registered users can access the database. Once the user has been registered, the user can access the database using created username and password. For that, users have login option in the home page (Fig. 2). Some options are only limited to the 'Administrator' (admin) of the database. Separate pages are created for user and admin (Fig. 3). The database is composed of five tables, which stores the all the data of the database including 'Accession Number', 'Primer name', 'Forward primer sequence', 'Reverse primer sequence', 'Product size', the 'Protein coded' etc. The database also contains details of the gene sequences that were used to generate the EST-SSR markers.

Fig.1. Screen shot of the 'Home Page'

Fig.2. Screen shot of the 'User Page'

        The query results are displayed in tabular format showing the 'Marker name' and 'Unigene id'. The 'Marker name' link takes the user to the new page giving 'Marker information', which displays the sequences along with the repeat type and primers (both forward and reverse primers). By giving the 'Unigene id', the complete sequence of the particular EST-SSR marker can be viewed. Currently, this database stores information on 143 coconut EST SSR markers that can be viewable and downloadable easily.

The main interfaces in the database are upload interface, search interface, download interface.

Upload Interface: Upload interface helps user to upload EST-SSR data and gene sequences in to the database (Fig. 4)

Fig.4. Screen shot of the 'Upload Interface'

Search Interface: The EST-SSR search result page displays 'Marker name' and 'Unigene id'. The individual marker entry links to a page where details of the markers are displayed (Fig. 5 and Fig. 6). Markers can be searched by marker name. It is also possible to search sequences based on 'Unigene Id' (Fig .7)
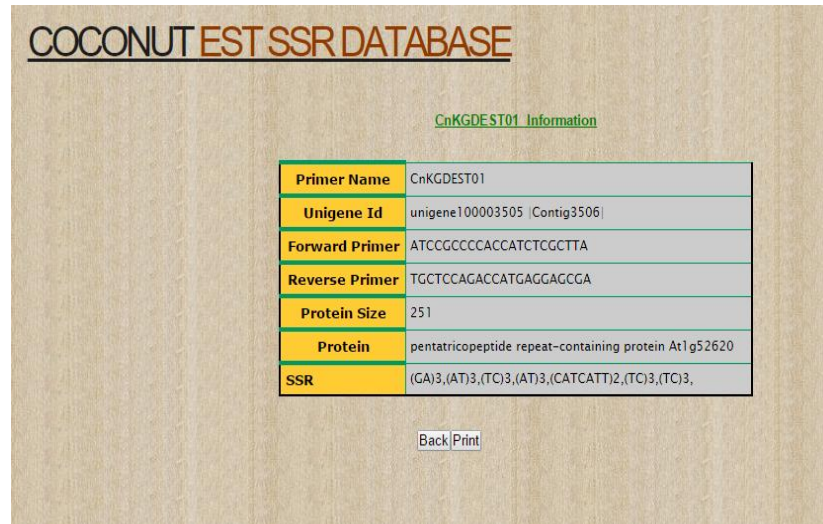

Fig.5. Screen shot of search interface of coconut EST-SSR marker database

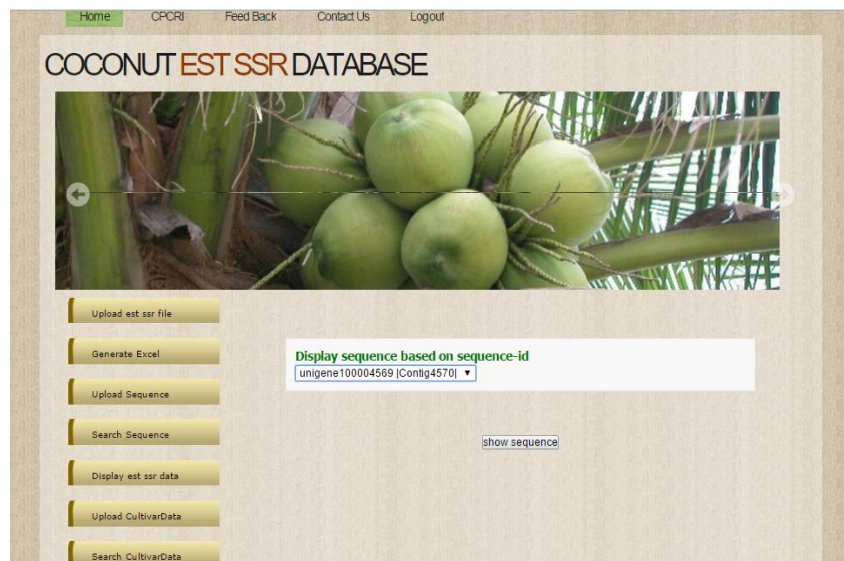**Fig.6. Displaying the result of search (EST-SSR Marker Information)**

Fig.7. Searching Gene sequence based on 'Unigene Id'

Download Interface: Using 'Download interface', user can download the whole database contents as excel sheet.

III. CONCLUSION

The Coconut EST-SSR Marker Database' has been created to provide researchers with access to coconut EST SSR data. Access to this database is provided through integrated web tools which allow users to access data via search interfaces. It also provides facility to download and view the Markers, their subsequent repeat type. 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 EST-SSR from which marker details, sequences can be viewed. A general tool bar is also included in each page for the ease of navigation through the site. Overall, we hope that this database would immensely benefit coconut researchers.

REFERENCES

1. Ashburner, G.R., Thompson, W.K., Halloran, G.M. and Foale, M.A., 'Fruit component analysis of South Pacific coconut palm populations', GRACE, vol. 44, pp. 327-335, 1997.
2. Lebrun, P., N'cho, Y.P., Seguin, M., Grivet, L. and Baudouin, L., 'Genetic diversity in coconut (*Cocos nucifera* L.) revealed by RFLP markers', Euphytica, 101:103-108, 1998.
3. Perera, L., Russel, J.R., Provan, J. and Powell, W., 'Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos nucifera* L.)', Genome, 43:15-21, 2000.
4. Carpio, C.B., 'Biochemical studies of *Cocos nucifera* L'. Kalikasan Philipp. J. Biol., 11: 319-338, 1982.
5. Fernando, W.M.U., Abeyasekera, J. D. J. and Jayatileke, R., 'Patterns of isozyme variation in *Cocos nucifera* L.', Proc. of Annual Sessions of the Sri Lanka Association for the Advancement of Science, 51:84- 86, 1995.
6. Upadhyay, A., Jayadev, K., Manimekalai, R. and Parthasarathy, V. A., 'Genetic relationship and diversity in Indian coconut accessions based on RAPD markers', Sci. Hort., 99:353-362, 2004.
7. Rajesh, M. K., Jerard, B. A., Preethi, P., Regi Jacob Thomas and Anitha Karun., 'Application of RAPD markers in hybrid verification in coconut', Crop Breed. Appld. Biotech., 14: 36-41, 2014.
8. Perera, L., Russell, J.R., Provan, J., McNicol, J.W. and Powell, W., 'Evaluating genetic relationship between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling', Theor.Appl. Genet., 96: 545-550, 1998.
9. Rivera, R., Edwards, K.J., Barker, J.H., Arnold, G.M., Ayad, G., Hodqkin, T. and Karp. A., 'Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L.', Genome, 42:668-675, 1999.
10. Perera, L., Russel, J.R., Provan, J. and Powell, W., 'Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos nucifera* L.)', Genome, 43:15-21, 2000.
11. Rajesh, M.K., Arunachalam, V., Nagarajan, P., Lebrun, P., Samsudeen, K. and Thampan, C., 'Genetic survey of ten Indian coconut landraces by Simple Sequence Repeats (SSR)', Sci. Hort., 118:282-287, 2008.
12. Rajesh, M.K., Nagarajan, P., Jerard, B.A., Arunachalam, V. and Dhanapal, R., 'Microsatellite Variability of coconut accessions (*Cocos nucifera* L.) from Andaman and Nicobar Islands', Curr. Sci., 94:1627-1631, 2008.
13. Rajesh, M. K., Samsudeen, K., Rejusha, P., Manjula, C., Shafeeq Rahman and Anitha Karun., 'Characterization of Annur and Bedakam ecotypes of coconut from Kerala State, India, using microsatellite markers', Int. J. Biodiver., 7 pages, 2014.
14. Tautz, D. and Renz, M., 'Simple sequences are ubiquitous repetitive components of eukaryotic genomes', Nucleic Acids Res., 12:4127-4138, 1984.
15. Morgante, M. and Olivieri, A.M., 'PCR- amplified microsatellites as markers in plant genetics', Plant J., 3: 175-182, 1993.
16. Powell, W., Machray, G.C. and Provan, J., 'Polymorphism revealed by simple sequence repeats', Trends Plant Sci., 1: 215-222, 1996.
17. Morand, M.E., Brachet, S., Rossignol, P., Dufour, J. and Frascaria-Lacoste, N., 'Generalized heterozygote deficiency assessed with microsatellites in French common ash populations', Mol. Ecol., 11: 377–385, 2002.
18. Zeid, M., Schon, C., Link, W., 'Genetic diversity in recent elite faba bean lines using AFLP markers', Theor. Appl. Genet., 107: 1304–1314, 2003.
19. Yi, G., Lee, J.M., Lee, S., Choi, D. and Kim, B.D., 'Exploitation of pepper EST-SSRs and an SSR-based linkage map', Theor. Appl. Genet., 114: 113–130, 2006.
20. Wijarat, P., Keeratinijakal, V., Toojinda, T., Vanavichit, A. and Tragoonrung, S., 'Genetic evaluation of *Andrographis paniculata* (Burm. f.) Nees revealed by SSR, AFLP and RAPD markers', J. Med. Plants Res., 6: 2777–2788, 2012.
21. Jiao, Y., Jia, H.M., Li, X.W., Chai, M.L., Jia, H.J., Chen, Z., Wang, G.Y., Chai, C.Y., Weg, E.V. and Gao, Z.S., 'Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*)', BMC Genomics, 13: 201–216, 2012.
22. Weber, J.L., 'Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms', Genomics, 7: 524-530.1990.
23. Katti, M.V., Ranjekar, P.K. and Gupta,V.S, 'Differential distribution of simple sequence repeats in eukaryotic genome sequences', Mol. Biol. Evol., 18: 1161-1167, 2001.
24. Field, D. and Wills, C., 'Long polymorphic microsatellites in simple organisms', Proc. Biol. Sci., 263: 209-215, 1996.
25. Gur-Arie, R., Cohen, C.J., Eitan, Y., Shelef, L., Hallerman, E.M. and Kashi, Y., 'Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism', Genome Res., 10: 62-71, 2000.
26. Mrazek, J., Guo, X. and Shah, A., 'Simple sequence repeats in prokaryotic genomes', Proc. Natl. Acad. Sci, 104, 8472– 8477, 2007.
27. Wang, H., Huan, P., Lu, X. and Liu, B., 'Mining of EST-SSR markers in clam Meretrix meretrix larvae from 454 shotgun transcriptome', Genes Genet. Syst., 86, p. 197–205, 2011.
28. Varshney, R.K., Sigmund, R., Bo¨rner, A., Korzun, V., Stein, N., Sorrells, M.E., Langridge, P. and Graner, A., 'Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice', Plant Sci., 168, pp. 195–202, 2005.
29. Yasodha, R., Sumathi, R., Chezhian, P., Kavitha, S and Ghosh, M., 'Eucalyptus microsatellites mined in silico: survey and evaluation', J. Genet., 87, 21–25, 2008.
30. Li, Y.C., Korol, A.B., Fahima, T. and Nevo, E., 'Microsatellites within genes: structure, function, and evolution', Mol. Biol. Evol., 21:991–1007, 2004.
31. Xiao, Y., Luo, Y., Yang, Y., Fan, H., Xia, W., Mason, A.S., Zhao, S., Sager, R. and Qiao, F., 'Development of microsatellite markers in *Cocos nucifera* and their application in evaluating the level of genetic diversity of *Cocos nucifera'*, POJ, 6:193–200, 2013.
32. Xia, W., Xiao, Y., Liu, Z., Luo, Y., Mason, A.S., Fan, H., Yang, Y., Zhao, S. and Peng, M., 'Development of gene-based simple sequence repeat markers for association analysis in *Cocos nucifera'*, Mol. Breed., 34:525–535, 2014.