CMSD: A Database of Microsatellite Markers Genotyped in Indigenous Coconut Accessions

Bobby Paul*, MK Rajesh, VP Jayakrishnan, MP Ramesh Kumar and George V Thomas
Bioinformatics Centre, Central Plantation Crops Research Institute, Kasaragod-671124, Kerala

The increased use of molecular markers in varietal identification has led to the need for an appropriate storage place and this has paved way for the development of a marker database containing allelic data of coconut accessions as revealed by microsatellite markers. The main features of the information system are a SSR database, SSR analysis tool, variety identification tool, data submission form and brief information about microsatellites. The database contains details of 11200 allelic data of 400 palms comprising 62 indigenous coconut accessions analyzed using 14 microsatellite primer pairs. Variety identification tool contains a display of 14 pairs of allelic entry corresponding to the 14 microsatellite primers. The result pages display the information about the query sample matches to a reference sample in the database and the score values they attain. Based upon the maximum score value, it can be concluded that the unknown sample is genetically related to the particular reference sample. The SSR allelic size data have been used for creating a DNA barcode system too, providing a more convenient handling of numerical data. This identification system with data warehouse can be accessed from the home page of Bioinformatics Centre, Central Plantation Crops Research Institute (ICAR), Kasaragod (http://www.bioinfcpcri.org/cmsd/home.php).

Key Words: Microsatellites, Coconut, Database

Introduction

Morphological and physiological markers are frequently used to determine distinctness, uniformity and stability (DUS) for registration and granting of Plant Breeder's Rights to new varieties of crops. New varieties have to be distinct from all existing varieties by at least one character. Additionally, they have to meet established standards with respect to uniformity and stability of the characteristics used to demonstrate distinctness. Morphological characters used for registration purposes can also be used for varietal identification. However, many of the morphological descriptors have limitations- they are multi-genic, quantitative or continuous characters and their expression can be altered by environmental factors. Furthermore, the number of registered varieties increases over time because of which, it is impossible for any testing authority to check efficiently each newly submitted variety against all existing varieties in common knowledge. Hence, to circumvent these problems, it has been suggested that modern methods should be evaluated for variety identification and related uses (Cooke, 1999).

Polymorphisms existing in DNA molecule of plant genomes can be exploited for genotyping of cultivars. DNA-based markers have many advantages for plant variety identification over the more traditionally used morphological characters because of their independence from environmental influences, their generally high level of polymorphism and their almost unlimited availability. There are several approaches to DNA profiling, of which microsatellites have proven to be efficient tools for characterizing coconut cultivars. Microsatellites or simple sequence repeat (SSR) loci are short tandem repeats of 1-6 bp, which are ubiquitous in both prokaryotes and eukaryotes both in protein-coding and non-coding regions (Toth *et al*., 2000). Presently, they serve as the most widely used markers in plants. They have also been used widely in coconut in a variety of applications including genetic distance measurements, phylogeny reconstructions, population genetics, genetic mapping and elucidating evolutionary history (Rivera *et al*., 1999; Perera *et al*., 1999; Merrow *et al*., 2003; Devakumar *et al*., 2006; Rajesh *et al*., 2008). Databases of DNA profiles of major coconut accessions are essential to the application of DNA marker-based variety identification.

The main objective of this present study was to build a database containing allelic data of indigenous coconut accessions as revealed by microsatellite markers.

Materials and Methods

Four hundred palms representing 62 coconut accessions (3-4 palms per accession) native to India were subjected to diversity analysis. DNA was extracted from spindle leaves of the palms using the standardized protocol (Upadhyay *et al.,* 1999) with slight modifications. A

* *Author for Correspondence E-mail: bobbypaul29@gm*

total of 14 highly polymorphic SSR primer pairs from the coconut microsatellite kit (Baudouin and Lebrun, 2002) were used in the present study. PCR reaction was conducted in volumes of 20 µl containing 35 ng genomic DNA, 0.2 µM each of forward and reverse primers, 50 µM of each dNTPs (M/s Bangalore Genei Pvt. Ltd., Bangalore), 1X buffer (10 mM Tris-Hcl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$) and 0.3 Unit of *Taq* DNA polymerase (M/s Bangalore Genei Pvt. Ltd., Bangalore). PCR amplifications were performed on an Eppendorf gradient thermal cycler with a PCR profile of 94ºC for 5 min followed by 30 cycles of 1 min at 94ºC, 2 min at the different annealing temperatures standardized for the individual SSR locus, and 2 min at 72ºC with a final extension for 5 min at 72ºC.After amplification, a volume of 8 µl of loading buffer (98 per cent formamide, 10 mM EDTA, 0.005 per cent each of xylene cyanol and bromophenol blue as tracking dyes) was added to each of the amplified product and denatured at 94ºC for 5 min, snap cooled using ice and separated on 5 per cent denaturing polyacrylamide gels containing 7 M urea at a constant power of 100 W. The patterns of amplified products across the samples were resolved by silver staining. The alleles were scored individually based on comparison with the molecular ladder and also the control samples.

The user friendly web interface for querying the database was done in PHP, a server side scripting language. The back end of the database was developed with MySQL (www.mysql.com). 11,200 allelic data were stored in a single table and there are 31 fields.

Results and Discussion

The Coconut Microsatellite Database (CMSD) is committed to provide the scientific community with comprehensive information on microsatellites occurring in indigenous coconut accessions. Three level schema architecture and flow of data has been shown in Figure 1. The screenshot of the CMSD homepage has been shown in the Figure 2. Currently CMSD contains 11200 allelic information of 400 coconut palms comprising 62 accessions extracted using 14 microsatellite markers. There is provision for updating any number of newly derived microsatellite data and the updation process is fully automated.

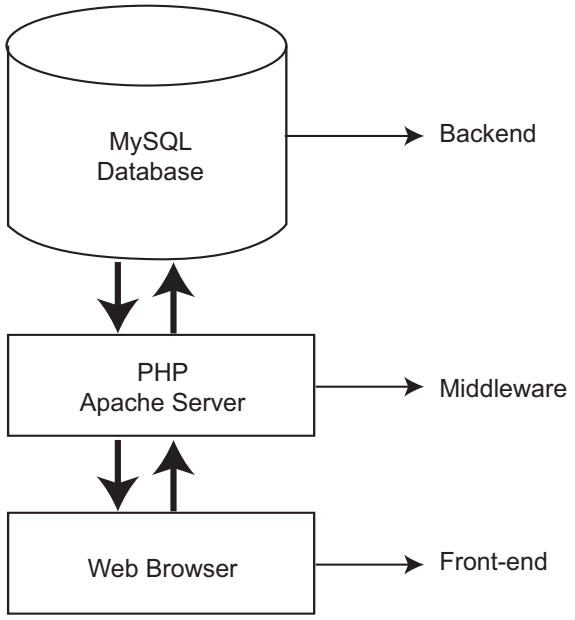

Fig. 1: An illustration of three level architecture used in the development of CMSD

The CMSD interface provides mainly search and variety identification tool and SSR analyzer tool. In 'search tool', users can obtain the details of microsatellite markers by selecting either the microsatellite primer or accession-wise option (Fig. 3). If the primer option is

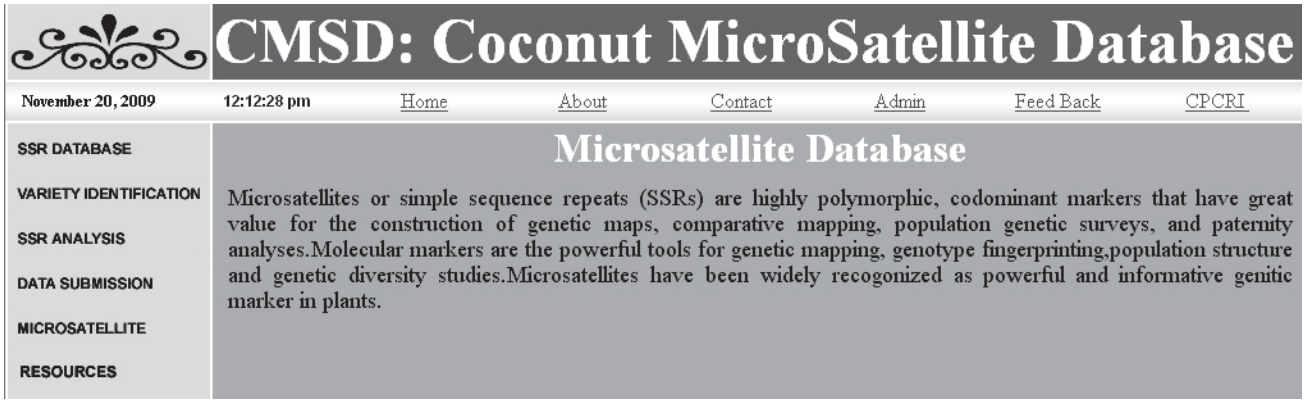

Fig. 2: A screen shot of CMSD home page

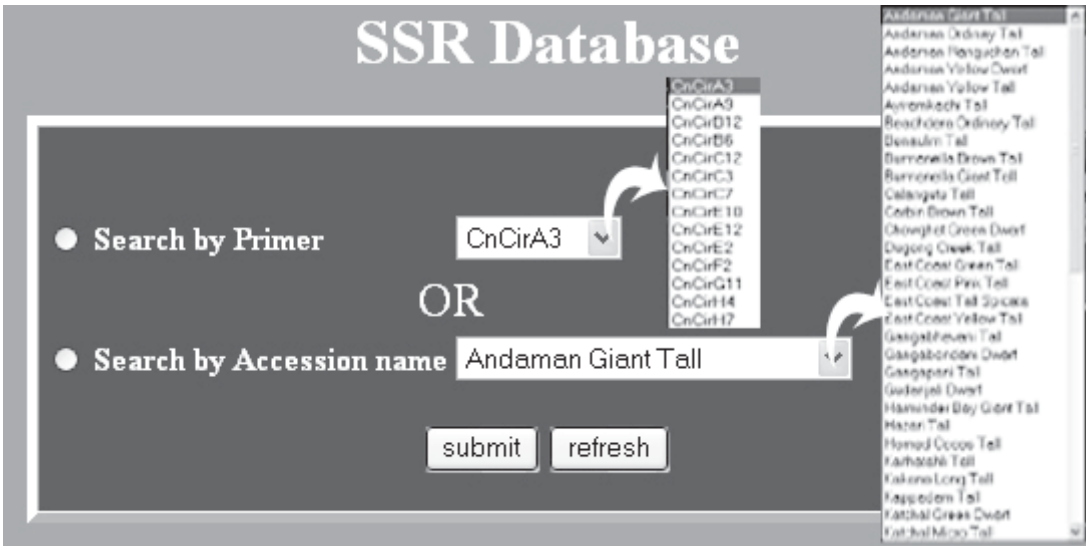

Fig. 3: Search tool with options for selecting either by the microsatellite primer or by accession

chosen, the allelic information of the 400 samples with primer sequence, annealing temperature, number of alleles, product sizes etc. are displayed. The accession-wise selection displays allelic size of the 14 microsatellite primers of the accession selected by the user. The 'variety identification tool' displays the highly identical sample population to a query with their obtained score, percent identity, percent similarity and matched alleles (Fig. 4). This tool is highly beneficial to the researchers aiming to identify an unknown coconut population from a mixed

Sl.No.	Population	Sample	Score	Similarity	Identity	Match Primers
1	West Coast Tall	RT1	28	100	100	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir A9 Allele 1, CnCir A9 Allele 2, CnCir B6 Allele 1, CnCir B6 Allele 2, CnCir B12 Allele 1, CnCir B12 Allele 2, CnCir C3 Allele 1, CnCir C3 Allele 2, CnCir C7 Allele 1, CnCir C7 Allele 2, CnCir Allele C12 1, CnCir C12 Allele 2, CnCir E2 Allele 1, CnCir E2 Allele 2, CnCir E10 Allele 1, CnCir E10 Allele 2, CnCir E12 Allele 1, CnCir E12 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir G11 Allele 1, CnCir G11 Allele 2, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 1, CnCir H7 Allele 2
2	West Coast Tall	RT6	18	64.29	64.29	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir A9 Allele 1, CnCir A9 Allele 2, CnCir B6 Allele 1, CnCir B6 Allele 2, CnCir C3 Allele 1, CnCir Allele C12 1, CnCir C12 Allele 2, CnCir E10 Allele 1, CnCir E10 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir G11 Allele 1, CnCir G11 Allele 2, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 2
3	West Coast Tall	WCT8	17	60.71	60.71	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir A9 Allele 1, CnCir A9 Allele 2, CnCir B6 Allele 1, CnCir B6 Allele 2, CnCir C7 Allele 1, CnCir C7 Allele 2, CnCir E10 Allele 1, CnCir E10 Allele 2, CnCir E12 Allele 1, CnCir E12 Allele 2, CnCir F2 Allele 1, CnCir G11 Allele 2, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 2
4	Ayiramkachi Tall	AYRT8	15	53.57	53.57	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir B6 Allele 2, CnCir C3 Allele 1, CnCir C7 Allele 2, CnCir E10 Allele 1, CnCir E10 Allele 2, CnCir E12 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir G11 Allele 1, CnCir G11 Allele 2, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 2
5	Ayiramkachi Tall	AYRT6	15	53.57	53.57	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir A9 Allele 1, CnCir A9 Allele 2, CnCir B6 Allele 1, CnCir B6 Allele 2, CnCir C7 Allele 2, CnCir E10 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir G11 Allele 1, CnCir G11 Allele 2, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 2
6	Katchal Micro Tall	KM4	15	53.57	53.57	CnCir A3 Allele 1, CnCir B6 Allele 1, CnCir B6 Allele 2, CnCir C7 Allele 1, CnCir C7 Allele 2, CnCir E10 Allele 2, CnCir E12 Allele 1, CnCir E12 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir G11 Allele 1, CnCir G11 Allele 2, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 2
7	Benaulim Tall	BENT7	15	53.57	53.57	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir B6 Allele 2, CnCir C3 Allele 1, CnCir C7 Allele 1, CnCir C7 Allele 2, CnCir E10 Allele 2, CnCir E12 Allele 1, CnCir E12 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir G11 Allele 1, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 2
8	Tiptur Tall	TPT2	15	53.57	53.57	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir B6 Allele 2, CnCir B12 Allele 1, CnCir C3 Allele 1, CnCir C3 Allele 2, CnCir C7 Allele 1, CnCir E10 Allele 2, CnCir E12 Allele 1, CnCir E12 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir H4 Allele 1, CnCir H4 Allele 2, CnCir H7 Allele 2
9	Nadora Tall	NAD89	15	53.57	53.57	CnCir A3 Allele 1, CnCir A3 Allele 2, CnCir A9 Allele 1, CnCir A9 Allele 2, CnCir B6 Allele 1, CnCir B6 Allele 2, CnCir B12 Allele 2, CnCir Allele C12 1, CnCir C12 Allele 2, CnCir E10 Allele 2, CnCir E12 Allele 1, CnCir E12 Allele 2, CnCir F2 Allele 1, CnCir F2 Allele 2, CnCir H7 Allele 2

Fig. 4: The 'variety identification tool' displaying the highly identical sample population to a query with their obtained scores, percent identity, percent similarity and matched alleles

population. CMSD also provides an interface to a 'SSR analyzer tool', a programme designed to predict mono- to deca- microsatellite repeats from any given nucleotide sequence. The data submission form helps in easy submission of newly derived data from anywhere in the world. This facility is very useful to maintain the database with current information which is not publicly accessible. The 'resources page' provides basic information about microsatellites and protocols for microsatellite marker genotyping. The individual SSR allelic size values were applied for creating barcodes making it possible to easily identify the cultivars at DNA level at first sight and thus providing a more convenient handling of numerical data. User comments and feedback are always welcome through the 'feedback page' provided. CMSD is hosted by the Bioinformatics Centre, CPCRI and is accessible via the World Wide Web interface at http://www.bioinfcpcri.org/cmsd/home.php.

We have demonstrated that the selection of microsatellite markers and construction of a consensus database as carried out in this work constitute a suitable system for use in network activities in which a common, centrally held database is continually fed with data from different laboratories. The existence of such DNA databases would facilitate the testing of new varieties of crops against all existing ones, reducing the need for individual laboratories or testing centres to maintain their own large reference collections (Donini *et al*., 2000).

The ability of identifying unknown samples with such a database was examined by genotyping 62 populations and their samples subsequent comparison to known reference varieties. As well as being useful in its own right and serving as a tool for variety identification, the constructed database of more than might be of interest for breeding companies and variety registration agencies for additional purposes such as description of the gene pool represented in indigenous coconut accessions. To maximise the value of the current database, however, it would be necessary to include additional characters, such as pest and disease resistance information. Furthermore, such a database only retains its value if it is permanently updated with new data. There are some important issues that arise from this methodology, such as the need for a careful selection of markers (rejecting any that cause difficulties in any laboratory, for instance), duplication of analyses in at least two laboratories, and having information about the possible heterogeneity of varieties for identification and related purposes. Last but not least, to be of any practical value, there must be information on a sufficiently large number of accessions. Given attention to these factors, it is clearly possible to construct consensus databases that have a wide range of practical applications in the variety and seeds area and for genetic resource analysis generally.

Acknowledgement

This work was supported by a grant from the Govt. of India, Department of Biotechnology (BTISnet), New Delhi-110003.

References

Baudouin L and P Lebrun (2002) The development of a microsatellite kit and dedicated software for use with coconuts. *Burotrop. Bull*. **17**:16-20.

Cooke RJ (1999) Modern methods for cultivar verification and the transgenic plant challenge. *Seed Sci. Technol.* **27**: 669-680.

Devakumar K, K Jayadev, MK Rajesh, A Chandrasekhar, R Manimekalai, PM Kumaran and VA Parthasarathy (2006) Assessment of genetic diversity of Indian coconut accessions and their relationship to other cultivars using microsatellite markers. *Plant Genet. Resour. Newsl.* **145:** 38-45.

Donini P, RJ Cooke and JC Reeves (2000) Molecular markers in variety and seed testing. In: AD Arencibia (eds) Plant Genetic Engineering: Towards the Third Millennium. *Elsevier Science B.V.* pp 27-34.

Merrow AW, RJ Wisser, JS Brown, DN Kuhn, RJ Schnell and TK Broschat (2003) Analysis of genetic diversity and population structure within Florida coconut (*Cocos nucifera* L.) using microsatellite DNA, with special emphasis on the Fiji Dwarf cultivar. *Theor. Appl. Genet*. **106:** 715-726.

Perera L, JR Russell, J Provan and W Powell (1999) Identification and characterization of microsatellites in coconut (*Cocos nucifera* L.) and the analysis of coconut population in Sri Lanka. *Mol. Ecol***. 8:** 344-346.

Rajesh MK, V Arunachalam, P Nagarajan, P Lebrun, K Samsudeen and C Thamban (2008) Genetic survey of ten Indian coconut landraces by simple sequence repeats (SSRs). *Scientia Hort.* **118**: 282-297.

Rivera R, KJ Edwards, JHA Barker, GM Arnold, G Ayad, T Hodgkin and A Karp (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. *Genome*, **42**: 668-675.

Toth Gabor Zoltan Gaspari and Jerzy Jurka (2000) Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Res.* **10**: 967-981.

Upadhyay A, VA Parthasarathy, G Seema and A Karun (1999) An efficient method of DNA extraction from coconut. *Agrotropica* **11:** 35-38.