

Microsatellites mining in date palm (*Phoenix dactylifera* L.) and their cross transferability across *Arecaceae* family

Kalathil Palliyarakkal Manju¹, Ramaswamy Manimekalai^{*2}, Sudalaimuthu Asari Naganeeswaran¹, Vadivel Arunachalam³, Anitha Karun¹

¹Biotechnology Section, Central Plantation Crops Research Institute (ICAR), Kasaragod, Kerala, India

²Biotechnology Section, Sugarcane Breeding Institute, Coimbatore, Tamil nadu, India

³ICAR Research Complex for Goa, Goa, India

*Corresponding author: rmanimekalaiicar@gmail.com

Abstract

Phoenix dactylifera L. (date palm) is the first publicly available nuclear genome sequence under *Arecaceae* family and indeed under the entire order Arecales. These genomic sequences were mined for microsatellites which could be used for marker assisted selection for important traits. We have developed an integrated, configurable and time effective microsatellite mining and annotation pipeline for this purpose. Microsatellite survey of date palm whole genome shotgun sequences using the developed pipeline detected a total of 166,760 perfect repeats with an average of one SSR per 2.2kb. Putative functions of these SSR-containing sequences within the proximity of genic regions were predicted. Primers targeting the functionally important SSR regions were designed and a set of eight SSR primers synthesized and were validated *in vitro* across five members of the *Arecaceae* family (coconut, arecanut, oilpalm, date palm and palmyrah palm). We have found that 28.6% of the SSRs are common among the members of arecaceae family. Those SSRs could be used in molecular marker analysis of less studied palm family members like arecanut and palmyrah palm.

Keywords: Arecanut; coconut; date palm; SSRs.

Abbreviations: e value_Expected value; kb_Kilo Base; Mbp_Million Base Pairs; MISA_MicroSATellite identification tool; mM_Milli Molar; PVPP_Poly Vinyl Poly Pyrrolidone; SDS_Sodium Dodecyl Sulphate; SSR_Simple Sequence Repeats; TIGR_The Institute for Genomic Research.

Introduction

Date palm (*Phoenix dactylifera* L.) is one of the most economically important dioecious monocotyledonous palms endemic to North Africa and Middle East. It is an evergreen, perennial, highly heterozygous, diploid (2n=36) plant with long generation time (Nixon, 1951). It is believed to have originated in the lands around the Persian Gulf and in ancient times was especially abundant between the Nile and Euphrates rivers (Kwaasi, 2003). The whole genome of date palm consists of ~658Mbp and Al-Dous et al. (2011) analyzed the ~380Mb partial sequences (~56% of the whole genome) as it is a good candidate for improving agricultural yields in arid environments. Simple Sequence Repeats (SSRs) or microsatellites are short DNA repeats, evenly dispersed across the genome (Hearne et al. 1992), more frequently in non-coding regions of the genome, of both eukaryotes and prokaryotes (Tautz 1989; Morgante and Olivieri, 1993). The general characteristics of plant and animal genomes are that plants are rich in AT repeats whereas AC repeat is the common type of repeat in animals (Powell et al. 1996; Rajwant et al. 2011).

The microsatellite density profiles in the *Arecaceae* family showed the predominant occurrence of dinucleotide repeats in the expressed genes of palm members (Manju et al. 2011). Several reports on the mining of SSRs across plant genomes are reported over the years. There is a non-random distribution of microsatellite repeats in coding and non-

coding regions in chloroplast genome of *Brassicaceae* family (Sumit et al. 2010). Comparative analysis of microsatellites in different plants showed that SSRs occupied 0.84% of genome in *Arabidopsis thaliana* (L), 0.97% in *Pinus taeda*, 0.26% in *Mesostigma viride*, 2.41% in *Chlamydomonas reinhardtii*, 2.67% in *Syntrichia ruralis*, 3.46% in *Physcomitrella patens*, 3.48% in *Gnetum gnemon*, 3.57% in *Oryza sativa*, 4.33% in *Marchantia polymorpha*, 4.64% in *Adiantum capillus-veneris*, and 4.66% in *Selaginella spp.* (Victoria et al. 2011). In human genome, 3% encompasses SSR motifs (Subramanian et al. 2003). The presence of SSR motifs near the gene sequences are found important for the development of functional markers (Grover and Sharma, 2007). One of the most important implications of *in silico* mining of microsatellites led to the understanding of cross-transferability among closely related species. The cross-amplification of soybean SSR loci within and among legume genera shared 65% of SSR cross transferability (Peakall et al. 1998). Cross-genera transferability among cassava, rubber tree and physic nut, reported the highest transferability (59.18%) in cassava (Whankaew et al. 2011). Cross-amplification of *Medicago truncatula* and *Trifolium repens* across 12 species of forage legumes (Amaresh, 2008) are reported. Akkak et al. (2009) evaluated the cross transferability of SSR markers from date palm itself across 14 other *Phoenix* species.

In the present study, we reported the mining of SSRs in the date palm whole genome shotgun sequences, the annotation and characterization of SSR-containing sequences and the cross transferability of a few of the microsatellite markers across five palm species. This is essential for genetic analysis in less studied palms such as arecanut, palmyrah and coconut.

Results and Discussion

Pipeline Implementation

We generated analysis pipeline which reduced the usage of different analysis tools for mining and annotation. The work flow of the analysis pipeline is shown in Fig. 1. The pipeline was made available in downloadable form from the URL http://nagu.inbios.org/ssr_front_end/. The usage of this tool can be extended to the SSR detection and annotation in genome/EST sequences of other crops and projects.

SSR distribution in date palm genome

We analyzed the distribution of perfect microsatellites in date palm (~366 Mbp, i.e. ~56% of the nuclear genome (~658Mbp)) non-redundant shotgun genomic sequences. A total of 166,760 SSRs (Table 1) with perfect repeats were detected with an average of 455 SSRs/Mb (i.e. one SSR in every 2.2kb of sequence). The number of SSR containing sequences is 62,639 and the repeat regions were ranged from 1 to 38 in single SSR-containing sequences. Also the SSR-containing sequence length ranges from 100bp to 93,730bp. The AT (96.6%) and GC (3.4%) content in the SSR-containing sequences were also tabulated and 72.4% of SSR-containing sequences comprise greater than 1000bp of nucleotides.

The frequency of SSRs in date palm showed similar trend with coffee (1 SSR/1.5kb), cotton (1 SSR/20kb) (Sithichoke et al. 2009), soybean genome (1 SSR/7.4kb) (Cardle et al. 2000), mungbean (1 SSR/6.7kb), cucumber (1 SSR/1.8kb) (Cavagnaro 2010), grapevine (1 SSR/1.97kb) and poplar (1 SSR/1.19kb) (Cavagnaro, 2010).

Mononucleotide repeats are the most abundant repeat types in the date palm whole genome, representing nearly 62% (Table 2) of all the repeats, followed by di- (29%) and trinucleotides (7%) (Fig. 2). Tetra-, penta- and hexa repeat types contributed only ~2% of the entire repeat types. Among the mononucleotide repeats, A/T repeats contributed 97% of repeats. The density of mono/di repeats spanned across the genome is given in Fig. 3. The A and T repeat together gives density of 270.23 SSR/Mbp. The other mon/di repeats together represents density of 130.80 SSR/Mbp. The AT repeats motif alone contributed 18% among the entire dinucleotides. Lagercrantz et al. (1993) reported the frequent occurrence of (AT)_n motif among dinucleotides as the peculiar feature of most plant genomes. This is in accordance with the present study and other studies such as microsatellite detection in cpDNA of 12 *Brassicaceae* family members (Sumit et al. 2010) and whole genome sequences of *Sorghum bicolor* (Yonemaru et al. 2009) and pine genomic sequences (Chagne et al. 2004).

It was reported that the microsatellite frequency is inversely related to the number of repeat units (Cavagnaro et al. 2010). However, mononucleotide showed a deviation in the present study, a variation was observed at higher repeat numbers. The minimum size of mono repeat units to be considered as perfect repeats was five, with an average of 24.5% of mono repeats. The percentage of SSRs among the

mono repeats decreased gradually as the number of repeat units increased upto 20 at an SSR percentage of 0.39 to total genome. But, unit size of greater than 20 showed a higher percentage (1.74%) of mono repeats (Fig. 4).

Functional annotation of SSR containing sequences

The functional annotation was done for 62,639 SSR-containing sequences. Arabidopsis, rice and maize were taken for comparison separately. As the function of genes that contain SSRs and the role of SSR motif in the function are poorly understood (Varshney et al. 2005), the sequences having SSRs around the 1kb adjacent regions were filtered. A total of 5299 sequences were annotated using Arabidopsis as background database.

After the first round annotation, the pipeline generated a total of 525,214 redundant annotated regions in 62,639 date palm sequences against Arabidopsis data-sets. From which 5229 annotated regions located within <1 kb neighboring functional units from 3233 sequences were filtered out. A trinucleotide repeat of the type (GGC)₅ occurred in a particular gene located within the ATPase protein-coding region with 0 e value. The frequent occurrence of trinucleotide repeats in the exonic regions that encode amino acid runs was reported (Sharma et al. 2011). Also the amino acid glycine finds the most frequent amino acid runs in *A. thaliana* as well as in *O. sativa* (Sharma et al. 2011).

Date palm SSR primers

For the annotated SSR regions, a set of 102 primers flanking functionally important annotated SSR regions were designed (supplementary table 1) and validated *in silico* and eight primer pairs were synthesized and validated in the laboratory. We focused on the designed primers to obtain selective annotated portions by narrowing down to disease-resistant gene related proteins, kinase, F-box family protein, etc (Table 3).

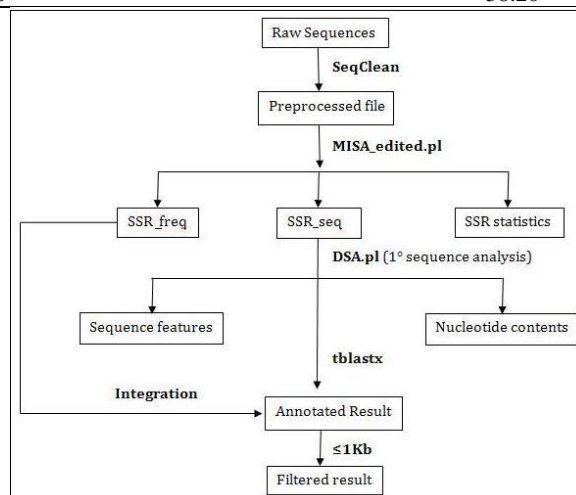
Cross-species amplification

The cross-species amplification of microsatellite loci is an economical method for the detection of SSR markers in other species and it can be treated as a source of developing microsatellites in lesser studied species (Whankaew et al. 2011). Here we have validated the cross-transferability of eight date palm SSR loci in economically important palms such as coconut, arecanut, oil palm, palmyrah and date palm. Out of the eight primers, amplification in the date palm reference DNA sample at the expected size was obtained for seven primers (Table 4). Oil palm DNA produced 42.8% of amplification with date palm primers. Coconut, arecanut and palmyrah showed 28.6%, 14.35% and 14.3% of amplification, respectively (Table 5). The primer pd4AT targeting (ATAA)₅ repeat motif amplified in all the five members (Fig. 5). And the primer set pd4GA produced amplification only in arecanut samples. The four primers, pd2CA, pd4AA, pd5GA and pd5TT did not give amplification across the members.

Many other studies reported low percentage of cross-amplification outside the genus. But, in our survey we detected slightly higher percentage of cross amplification between oil palm and date palm and between coconut and date palm. High rates of SSR loci across species (>50%) within a *Glycine* genus were reported (Peakall et al. 1998). Only 3%–13% of amplifications are detected outside the

Table 1. Sequence and microsatellite density information in date palm genome.

Microsatellite data types	Density
Total number of date palm contigs examined	141,832
Total size of examined sequences (bp)	366,035,356
Total number of identified SSRs	166,760
Number of SSR containing sequences	62,639
Number of sequences containing more than 1 SSR	32,628
Number of SSRs present in compound formation	19,538
Total AT%	61.79
Total GC%	38.20

**Fig 1.** Work flow of SSR mining annotation pipeline. The flow chart of SSR mining tool shows the pipeline initiation with raw date palm contigs retrieved from the public domain processed using different configured third party tools and the SSR sequences undergone annotation using tblastx program. The annotated results are sorted manually to filter out the $\leq 1\text{kb}$ proximity SSR sequences.**Table 2.** Distribution of perfect microsatellites with ≥ 5 repeats.

SSR Type	No. of SSRs	Percentage	Density (SSR/Mb)
Mononucleotide	103102	61.83	280.9
Dinucleotide	48005	28.78	130.8
Trinucleotide	11696	7.01	31.9
Tetranucleotide	3312	2.0	9.0
Pentanucleotide	469	0.28	1.3
Hexanucleotide	176	0.10	0.5
Total	166760	100	454.4

The distribution of different repeat types, percentage and the density calculated by number of SSRs/366Mbp were given in the Table.

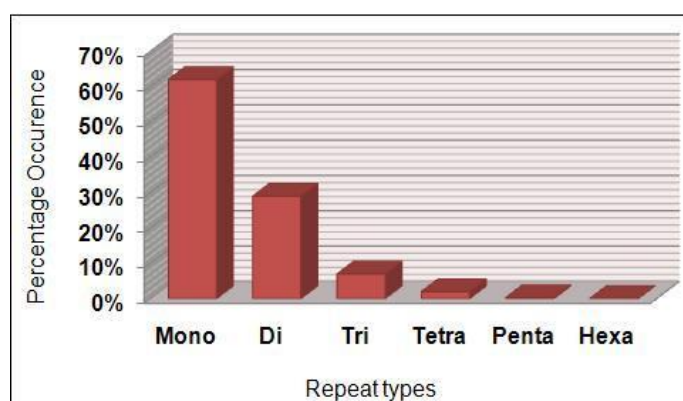
**Fig 2.** The percentage distribution of repeat types across the date palm genome. The distribution of different repeat types mined from date palm whole genome shotgun sequences. Mono nucleotides occurred abundantly among the total 166,760 SSRs.

Table 3. Details of date palm SSR primers used for cross amplification analysis and its annotated proteins.

Sl. No:	Primer Name	SSR Types	Primer sequence (5'-3')	Annotation
1	pd2CA	(CA)10	F: GGCACCTATTACGCTCTA R: GGTATCCCTCGCATTG	F-box family protein [<i>Populus trichocarpa</i>]
2	pd2CT	(CT)10	F: CTTGGTAAGCTCCAGTA R: GATGCGTCATGAGATAGG	Seven transmembrane domain protein [<i>Zea mays</i>]
3	pd2GA	(GA)10	F: AGCACGTATGAAACAGT R: TTGAGGGAAAGTGAGGT	F-box family protein-like [<i>Oryza sativa japonica</i> Group]
4	pd4AT	(ATAA)5	F: ATTCGGGTCATGGATGGCA R: CCCAGTCCGGAGTGACAG	PERK1-like protein kinase [<i>Nicotiana tabacum</i>]
5	pd4AA	(AAAT)5	F: TGTCCCTCAGCATAAACCAT R: ATGGAGATTGATCGGCGA	Disease resistance gene [<i>Pinus sylvestris</i>]
6	pd4GA	(GAAG)6	F: CAGCAGAGAAAGGTCGCA R: GGAAAGCGTGGATAACCC	Similar to mitogen-activated protein kinase10 [<i>Rattus norvegicus</i>]
7	pd5GA	(GAAAA)5	F: GCAACAAGGCAGGTCCAA R: CATTTCGGAAGGAGCCGTTT	4-coumarate-CoA ligase [<i>Juglans nigra</i>]
8	pd5TT	(TTTTG)5	F: TTTATCCCCTTCGTTGCTGC R: GGACTCACCAATGGCAGA	Hypothetical protein OsI_01213 [<i>Oryza sativa indica</i> Group]

Date palm primers listed with primer names, SSR types, primer sequences and its annotated proteins.

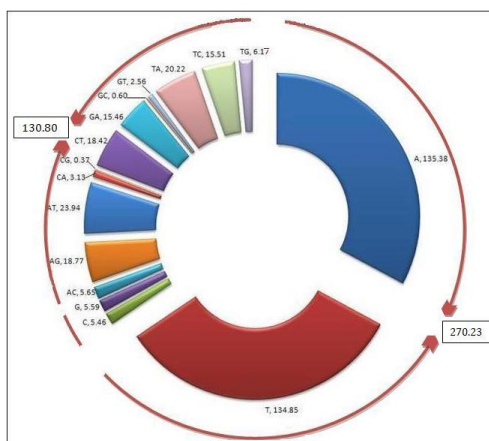


Fig 3. Percentage occurrence of different mono and di nucleotide repeats in the date palm genome. Each section of the graph describes the density of mono/di repeats calculated by SSR/366Mbp size genome of date palm. The A and T repeat together gives density of 270.23 SSR/Mbp.

genera (Peakall et al. 1998). Thus it is showed that there is a certain level of common SSR loci between members of *Arecaceae*. The tagging of important gene regions in other palms can also be possible.

Materials and Methods

Source of sequences

A total of 142,304 whole genome shotgun sequences of date palm were retrieved from the NCBI ftp site (<http://www.ncbi.nlm.nih.gov/Ftp/>). Complete non-redundant protein sequences of *Arabidopsis thaliana* L. (33410) (<https://www.arabidopsis.org/>), *Oryza sativa* (66249) (<http://rice.plantbiology.msu.edu/>) and *Zea mays* (10,446) (<http://maize.jcvi.org/>) were collected from The Institute of Genome Research, TIGR gene indices.

Mining of microsatellites

A total of 142,304 raw date palm contigs taken for the analysis via the Seqclean tool (<http://compbio.dfci.harvard.edu/tgi/software/>) integrated with Blast and Univec (<http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html>)

database. After trashing the erroneous regions, the program produces cleaned filtered sequence file which would be used as input for the succeeding processes. The perl script of the MicroSatellite Identification tool (MISA) (Thiel et al. 2003) was configured to obtain a sequence file containing SSR-containing sequences only. This was locally achieved for easy processing and analysis of repeat sequences. Furthermore, perl scripts were adopted for the primary analysis such as the calculation of the number of SSRs, AT and GC content and length of the repeat containing contigs.

Annotation of SSR containing sequences

A total of 62,639 SSR-containing sequences underwent sequence annotation. Similarity searches were performed with the standalone version 2.2.21 of BLAST against the non-redundant TIGR gene model sequences. Three main plant reference models (*Arabidopsis*, rice and maize) were used for comparison. Since the SSR region covers a very small percentage within the huge contigs sequences, the functional prediction of SSR-containing sequences was done with high stringency along with filtering the blast hits. When the SSR region comes within a space of 1kb (1000bp) sequences to the functional regions, those were counted as sequences within the proximity of genic regions (<1kb). Such

Table 4. Validation of primers synthesized in date palm DNA.

Sl No:	Primer Name	Amplified (Yes/No)	Expected amplicon Size
1	pd2CA	Yes	248
2	pd2CT	No	510
3	pd2GA	Yes	556
4	pd4AT	Yes	247
5	pd4AA	Yes	1117
6	pd4GA	Yes	277
7	pd5GA	Yes	577
8	pd5TT	Yes	208

The screened date palm SSR primers with amplification result and expected amplicon size.

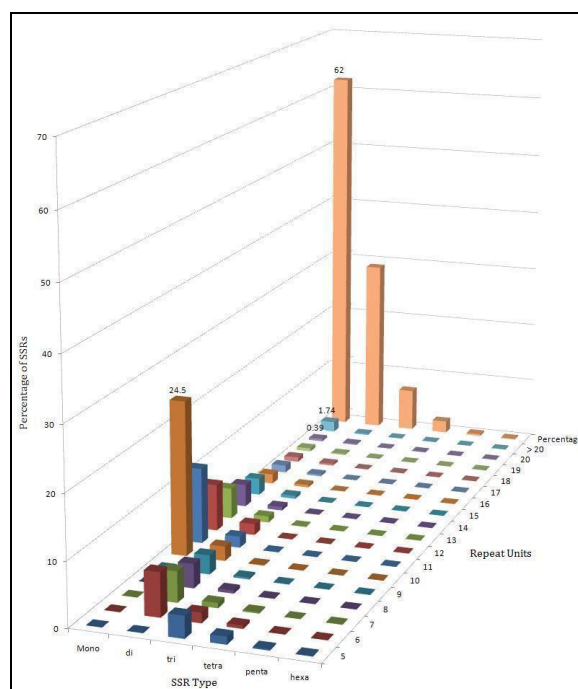


Fig 4. Graphical representation of the number of repeats under each SSR type and percentage of each SSR type in date palm genome. The graph was based on a total of 166,760 SSRs detected in 366 Mbp of date palm sequences. The percentages of all the repeat types against repeat units are shown in the figure. The mono repeat type starts with a repeat unit of 10 whereas the others were considered from five repeat units.

hits with SSR region within <1kb of the targeted proteins and e values $< e^{-15}$ were selected. The integration of MISA result and the annotation output was performed using locally developed perl script. Furthermore, redundancy was removed manually by scrutinizing the final annotation results.

Primer designing

Primers were designed flanking to the repeat regions using FastPCR (<http://primerdigital.com/fastpcr>) Primer3 software (<http://primer3.sourceforge.net/>) and the quality of the designed primers was evaluated for the formation of primer dimers/self-annealing and loopings using stand alone Oligoanalyzer tool.

Plant material

Coconut (*Cocos nucifera* L.), arecanut (*Areca catechu*), oil palm (*Elaeis*), date palm (*Phoenix dactylifera* L.) and palmyrah (*Borassus flabellifer*) are the five different palms collected from the family of *Areaceae*. Immature leaves were used for DNA extraction.

DNA extraction

DNA was extracted from the immature leaves of all the palms using SDS method (extraction buffer- 100mM Tris, pH 8, 1.4 M NaCl, 50 mM EDTA). The leaf samples was frozen in liquid nitrogen and ground into fine powder. To that 0.5 g of poly vinyl (poly) pyrrolidone (PVPP) and 1ml of 10% SDS was added, mixed well and transferred to a centrifuge tube containing pre-heated (65°C) extraction buffer with β -mercaptoethanol. The mixture was incubated at 65°C for 1 hour with intermittent mixing. After incubation, chloroform: isoamyl alcohol mixture (24:1 V/V) was added and homogenized by gentle inversion for 15 min and centrifuged at 8,000 rpm for 20 min at 4°C. The clear aqueous phase was then transferred into a fresh centrifuge tube and 2/3rd volume of ice-cold isopropanol was added. The DNA pellets were collected in 1.5 ml micro centrifuge tubes, washed thrice with 70% alcohol, air dried and the pellets were dissolved in TE buffer. The DNA was purified by RNase treatment and quantified using spectrophotometer. The quality was checked in 0.8% agarose gel and the samples were diluted to 20 ng/ μ l for further use.

Table 5. Amplifications using date palm SSR primers in different palms.

Sl No:	Name	Size	Coconut	Arecanut	Oil palm	Palmyrah palm	Date palm
1	pd2CA	248	-	-	-	-	+
2	pd2GA	556	-	-	+	++	+
3	pd4AA	1117	-	-	-	-	+
4	pd4AT	247	+	+	+	++	+
5	pd4GA	277	+	-	+	+	+
6	pd5GA	577	-	-	-	-	+
7	pd5TT	208	-	-	-	-	+
Crop wise amplification			28.6	14.3	42.8	14.3	

⁺ indicates bands at the expected size, ++ indicates multiple bands presents including the one at expected size, - indicates absence of band.

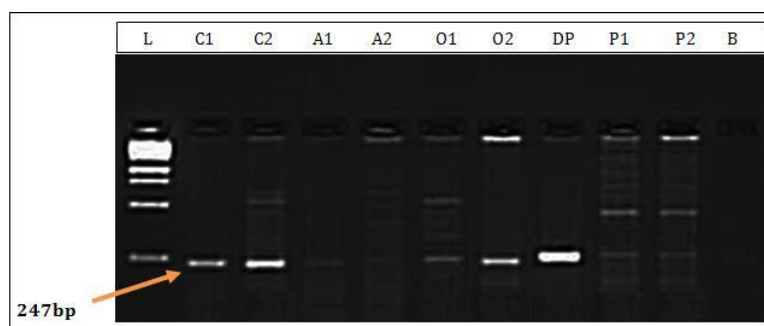


Fig 5. Gel picture of the cross amplification of palms DNA using pd4AT primer. Figure shows the amplification in the five members of Arecaceae family with the primer pd4AT at an expected size of 247bp. Lane L- 1KB Ladder, C1, C2- Coconut, A1, A2- arecanut, O1, O2- Oil palm, DP- Date palm, P1, P2- Palmyrah palm, B- Blank.

PCR and gel electrophoresis

The amplification of the microsatellite (using eight primers, Table 3) was carried out in 30 µl reaction mixture containing 20 ng of template DNA, 1X assay buffer (MgCl₂), 200µM of dNTP mix, 20 picomoles primer mix and 0.5units of Taq DNA polymerase. The amplification was carried out in thermal cycler (BIORAD Gradient Thermo Cycler). The PCR consisted of a total of 35 cycles with an initial template denaturation at 94°C at 5 min for 1 cycle then at 94°C for 1 min followed by 30 sec annealing at different date palm primer annealing temperature and then 35 cycles of primer extension at 72°C for 1min and a final extension at 72°C for 10 min. The amplified fragments were separated by electrophoresis in 3% agarose gel stained with ethidium bromide (Sambrook et al. 1989) and were visualized in a gel documentation system (Syngene Bioimaging System). Data were scored according to the presence or absence of clear and reproducible bands. Bright and unambiguous bands at the expected size were read as positive for cross-amplification.

Conclusion

The developed pipeline is useful for sequence processing, SSR mining, analysis and annotation of genomic sequences. Numerous SSR genetic markers are detected in date palm. Based on the transferable nature of SSR markers across species, these SSRs can be used in other palm members. This information can be used for the genetic improvement of the crop in future.

Acknowledgements

This work was supported by grants from Department of Information Technology, India. Our sincere thanks to Dr. George.V.Thomas, Director, Central Plantation Crops

Research Institute, Kasaragod, for his broad guidance and support.

References

- Akkak A , Scariot V, Marinoni DT, Boccacci P, Beltramo C, Botta R (2009) Development and evaluation of microsatellite markers in *Phoenix dactylifera* L. and their transferability to other *Phoenix* species. *Biologia Plantarum*. 53:164-166.
- Al-Dous EK, Binu G, Maryam EM, Moneera YJ, Hao W, Yasmeen MS, Eman KA, Srinivasa C, Ana CP, Jeremy DB, Vincent A, John O, Imad JS, Khaled MS, Jeffrey LB, Robert RK, Joel AM (2011) De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat Biotechnol*. 29:521-527.
- Amaresh C (2008) Transferability of SSR markers across twelve species of forage legumes for germplasm characterization and evaluation. *The Ind J Genet Plant Breed*. 68:189-194.
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*. 156:847-854.
- Cavagnaro PF, Douglas AS, Luming Y, Philipp WS, Timothy TH, Chinnappa DK, Sanwen H, Yiqun W (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics*. 11:569.
- Chagne D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JM, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor Appl Genet*. 109: 1204-1214.
- Grover A, Sharma PC (2007) Microsatellite motifs with moderate GC content are clustered around genes on Arabidopsis thaliana chromosome 2. *In Silico Biol*. 7: 201-13.

- Hearne CM, Ghosh S, Todd JA (1992) Microsatellites for linkage analysis of genetic traits. *Trends Genet.* 8: 287–294.
- Kwaasi AAA (2003) Date palm and sandstorm-borne allergens. *Clinical Experimental Allergy.* 33:419-426.
- Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* 21:1111–1115.
- Manju KP, Manimekalai R, Arunachalam V (2011) Microsatellites in palm (*Arecaceae*) sequences. *Bioinformation.* 7: 347- 351.
- Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J.* 3:175-182.
- Nixon RW (1951) The date palm: "Tree of life" in subtropical deserts. *Economic Botany.* 5: 274-301.
- Peakall R, Gilmore S, Keys W, Morgante M, Rafalski (1998) Cross-Species Amplification of Soybean (*Glycine max*) Simple Sequence Repeats (SSRs) Within the Genus and Other Legume Genera: Implications for the Transferability of SSRs in Plants. *Mol Biol Evol.* 15:1275–1287.
- Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1:215–222.
- Rajwant K, Manoj KR, Sanjay K, Rohtas S, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica.* 177:309–334.
- Sambrook J, Fritsch EF, Manitis T (1989) *Molecular cloning: a laboratory manual.* 2nd ed. N. Y., Cold spring Harbor Laboratory Press. pp. 1659.
- Sharma A, Chauhan RS (2011) Repertoire of SSRs in the castor bean genome and their utilization in genetic diversity analysis in *Jatropha curcas*. *Comp Funct Genomics.* 1-9.
- Sithichoke T, Prakrit S, Pichahpuk U, Juntima C, Duangjai S, Worapa S, Warunee S, Somvong T, Peerasak S (2009) Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). *BMC Plant Biology.* 9: 137.
- Subramanian S, Mishra RK, Singh L (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biology.* 4: R13.
- Sumit GG, Praveen A, Yashbir SB (2010) Analysis of SSR dynamics in chloroplast genomes of Brassicaceae family. *Bioinformation.* 5:16–20.
- Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17:6463–6471.
- Thiel T, Michalek V, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet.* 106:411–422.
- Varshney RK, Andreas G, Mark ES (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23:48-55.
- Victoria FC, Da Maia LC, De Oliveira AC (2011) *In silico* comparative analysis of SSR markers in plants. *BMC Plant Biol.* 11:15.
- Whankaew S, Supanath K, Chalernpol P, Duncan RS, Jarunya N, Kanokporn T (2011) Cross-genera transferability of (simple sequence repeat) SSR markers among cassava (*Manihot esculenta* Crantz), rubber tree (*Hevea brasiliensis* Muell. Arg.) and physic nut (*Jatropha curcas* L.). *Afr J Biotech.* 10:1768-1776.
- Yonemaru JI, Ando T, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). *DNA Res.* 16:187–193.