

Analysis of Simple Sequence Repeats Information from Floral Expressed Sequence Tags Resources of Papaya (*Carica papaya* L.)

Priyanka Priyanka¹, Dileep Kumar¹, Anurag Yadav², Kusum Yadav^{1*}, U. N. Dwivedi¹

¹Department of Biochemistry, University of Lucknow, Lucknow, India

²College of Basic Sciences & Humanities, Sardarkrushinagar Agricultural University Dantiwada, Banaskantha, India

Email: *anukusum@gmail.com

How to cite this paper: Priyanka, P., Kumar, D., Yadav, A., Yadav, K. and Dwivedi, U.N. (2017) Analysis of Simple Sequence Repeats Information from Floral Expressed Sequence Tags Resources of Papaya (*Carica papaya* L.). *American Journal of Plant Sciences*, 8, 2315-2331.

<https://doi.org/10.4236/ajps.2017.89155>

Received: July 27, 2017

Accepted: August 21, 2017

Published: August 24, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Papaya (*Carica papaya* L.) is one of the most economically, medicinally and nutritionally important tropical fruit crops. Expressed sequence tags (ESTs) derived simple sequence repeat (SSR) markers are more valuable as they are derived from conserved genic portion. Development of EST-SSRs markers through *in silico* approach is cheaper, less time consuming and labour-intensive. In this study, we aimed to mine SSRs and developed EST-SSR primers from papaya floral ESTs. A total of 75,846 papaya floral ESTs were downloaded from public database National Centre for Biotechnology Information (NCBI). A total of 26,039 floral unigenes (7961 contigs and 18,078 singletons) were generated after assembly of these ESTs. From these floral unigenes, 433,782 perfect SSRs, 204,968 compound SSRs and 6061 imperfect SSRs were mined, respectively. In perfect SSRs, mononucleotide repeats were most abundant (94.7%) followed by tri- (3.1%) and di-nucleotide repeats (1.7%). The frequencies of tetra-, hexa- and penta-nucleotide repeats accounted for only (0.17%), (0.04%) and (0.03%), respectively. In mononucleotide repeats, the most abundant motif was A/T (69.3%) and in di- and tri-nucleotide repeats were AG/CT (61%) and AAG/CTT (31%), respectively. In imperfect SSRs, mononucleotide repeats (56.5%) were most abundant. 176 different types of motifs were identified. A total of 3807 primer pairs for floral papaya ESTs were successfully designed. These developed EST-SSR primers are being used for the genetic improvement of papaya such as study of cross-transferability across genera/species, evaluation of genetic diversity, and identification of sex-specific markers. These EST derived SSRs can also be used in filling gaps in existing linkage maps in papaya.

Keywords

Papaya (*Carica papaya* L.), *In Silico*, Simple Sequence Repeats, Expressed Sequence Tags (ESTs), SSR Mining, EST-SSR, SSR Motifs, Primer Pairs

1. Introduction

Papaya (*Carica papaya* L.) is an edible fruit crop of the family Caricaceae, originally native to Central and South America and distributed in tropical and subtropical regions worldwide. It is a diploid species ($2n = 18$) and a dicotyledonous plant. It has a small genome size of 372 Mbp [1] [2] [3]. It is short-lived, semi-woody, herb-like and perennial tropical plant. The fruit production starts after nine to ten months from germination period [4] [6].

According to the percentage of US recommended daily allowances, papaya fruit ranked first among 35 most commonly used fruits. It is highly nutritious and contains antioxidant vitamins (A, C and E), thiamine, folate, riboflavin, niacin, potassium, iron, calcium and fibre. It contains no starch and is low in calories [5]. A proteolytic enzyme, papain (EC: 3.4.22.2) is extracted from the latex of unripe fruit which is commonly used in food processing such as in tenderization of meat, to clarify beer and juice and in industry for making soap, shampoo, lotions, skin care products and toothpastes [6]. It can also be used in several medical applications such as for digestion improvement and in treatment of fever, ulcers, muscular dystrophy and osteoporosis [7].

It is a trioecious species with three types of sex: female, hermaphrodite and male. The hermaphrodite plants are widely grown as every plant of hermaphrodite produces fruits. Female plants are commercially important for papain production, while male plants have no use except for pollination [8]. Female plants need 6% - 10% male plants in the field for the purpose of fruit production [9]. Since the use of seeds produces seedlings of unknown sex, farmers have to plant seedlings in large amounts and thin out the female or hermaphrodite plants after 3 to 4 months when it is possible to identify the sex of the seedlings from their floral buds [10]. If the sex of papaya is identified before their transplantation to the field at the seedling stage, then a desired ratio of male and female plants (5% males: 95% females) would be achieved for cultivation and resources like planting space, fertilizers and water could be devoted to female and hermaphrodite plants. Papaya is considered as a fruit model crop for genetic, genomics and molecular studies owing to its several features such as short generation time, small genome size, primitive sex chromosomes and efficient breeding system [11].

Microsatellites or simple sequence repeats (SSRs) consist of one to six (bp) tandem repeats (mono-, di-, tri-, tetra- and penta-, hexa-nucleotides), and are found in all genomes including prokaryotes and eukaryotes [12] [13]. They are also termed as simple sequence length polymorphisms [14], microsatellite [15], short tandem repeats [16]. They are located in both coding and non-coding regions of the genome [17]. SSRs are most important over other PCR-based mo-

lecular markers like random amplified polymorphic DNA (RAPD), inter simple sequence repeats (ISSR) and amplified fragment length polymorphism (AFLP) due to their sequence-specificity, multi-allelic nature, co-dominant inheritance, high distribution in the genome, easy detection by PCR, high rate of transferability, hyper-variability and high reproducibility [13] [18] [19] [20]. The polymorphic nature of SSR was observed by Litt and Luty (1989) which is generated due to variation in repeats number. The origin and evolution of microsatellites occur due to slippage of DNA strand which creates mispairing [21] and repetitive errors generated during replication of DNA [22], or unequal recombination between sister chromatids during meiosis [23]. The principle of polymorphism detection involves the designing of primers from flanking sequences near the portion of microsatellite repeat motif. Amplification of genomic DNA with specific primers flanking the SSR motifs is performed using PCR and running agarose or denaturing polyacrylamide gel for visualization of variations in alleles. There are two types of SSRs on the basis of their location: 1) SSRs that are distributed throughout the genome are called genomic-SSRs, 2) SSRs that are found in genic or expressed portion of the genome is called as genic-SSRs or Expressed Sequence Tags-SSRs (EST-SSRs). Genic-SSRs act as functional molecular markers because “putative function” can be determined by publically available databases via computational approaches.

There are two traditional methods for the development of genomic SSR markers, 1) SSR-enriched genomic library and 2) nonenriched genomic library construction. Both the methods involve construction of genomic DNA library, following the hybridization with tandemly repeated oligonucleotides probes, cloning and sequencing of candidate clones [24], which makes these methods of development very tedious, time consuming, expensive and labor-intensive [25]. On the other hand, with the advancement of modern genomics, genic or EST-SSRs are comparatively easier to develop as large numbers of ESTs of various organisms are available in various data banks. Availability of these large amounts of freely accessible data makes possible to develop EST-based SSR markers through database mining. The development of EST-SSRs or genic-SSRs through *in silico* approach is a fast, efficient, requires less cost, time and labor as compared to the development of genomic-SSRs [26] [27].

ESTs are the short (200 - 800 bases), and single pass random sequence reads of cDNAs derived from cDNA libraries. EST-SSRs are more advantageous than the genomic SSRs due to less time consuming, easily available, cheapest to develop, detect variations in expressed portion of the genome and sequence-specificity. Moreover, EST-SSRs show high rate of transferability, which means EST-SSR markers isolated from one species, can be transferred to other related species/ genera or within the same family due to conserved genic regions [27] [28]. Therefore, EST-SSRs have been utilized in several plants for various applications such as to study genetic diversity [29], cross-transferability [30], comparative analysis [31] and in linkage map construction [32]. In papaya, several microsatellite markers have been developed for the study of genetic diversity [33] [34] and mark-

er-assisted selection (MAS) [35], but most of these SSRs are genomic in nature.

Complete papaya genome has been sequenced by Ming *et al.* [5], which generated enormous amount of ESTs and other DNA sequences which are freely accessible at NCBI (<http://www.ncbi.nlm.nih.gov>) and the availability of several SSR mining tools like MISA [13], TROLL [36], SciRoKo [37], Msat commander [38], etc., makes it possible to utilize available ESTs for the development of genic SSRs which could be applied for papaya crop improvements. Only few studies of microsatellite analysis from genomic sequences [39] and from ESTs [40] have been performed in papaya. Moreover, only limited genic or EST-SSR markers, which emerge from transcribed portion of the genome, therefore becomes more important, are available in *C. papaya*. Therefore, the present study was undertaken to develop genic SSRs by utilizing the available EST database of *C. papaya*. The study has following two objectives: 1) *In-silico* approach to mine SSRs from the available papaya ESTs from the NCBI database and, 2) to develop EST-SSR primers. These developed primers could be used for estimation of genetic diversity, cross-transferability across species and genera, in comparative-genomics study and in identification of sex specific markers in papaya.

2. Materials and Methods

The methodology of *in silico* mining and development of EST-SSR primers from papaya floral ESTs are shown in (Figure 1).

2.1. Retrieval of Floral Papaya EST Sequences

EST sequences of *C. papaya* are available at NCBI (www.ncbi.nlm.nih.gov/nucest/). A total of 75,846 papaya floral EST sequences (male, female and hermaphrodite flower) before meiosis and after meiosis stage were retrieved from EST database (dbEST) of NCBI in FASTA format. These EST sequences were submitted by Ming *et al.* [5].

2.2. EST Sequences Processing

ESTs are single pass DNA sequences so, they are more error prone. EST sequences may contain vector/adaptor contaminations, low complexity sequences and poly-A/T tails. Therefore, EST sequences were initially screened using DDBJ VecScreen tool (<http://ddbj.nig.ac.jp/vecscreen/>) for identification of vector contamination. It detects vectors, adaptors and other suspect contaminations by NCBI's UniVec core vector/adaptor library. EST sequences were then processed using SeqTrim NEXT [41] with its default parameters. The program takes a FASTA format sequence file as an input. It removes vector/adaptor contamination, low complexity regions and trimming of poly-A, poly-T tails from the EST sequences according to the given parameters.

2.3. Assembly of Floral Papaya EST Sequences

All the processed floral EST sequences were assembled using SeqMan DNA-STAR Lasergene ver. 9.0 program with its default parameters (minimum matching

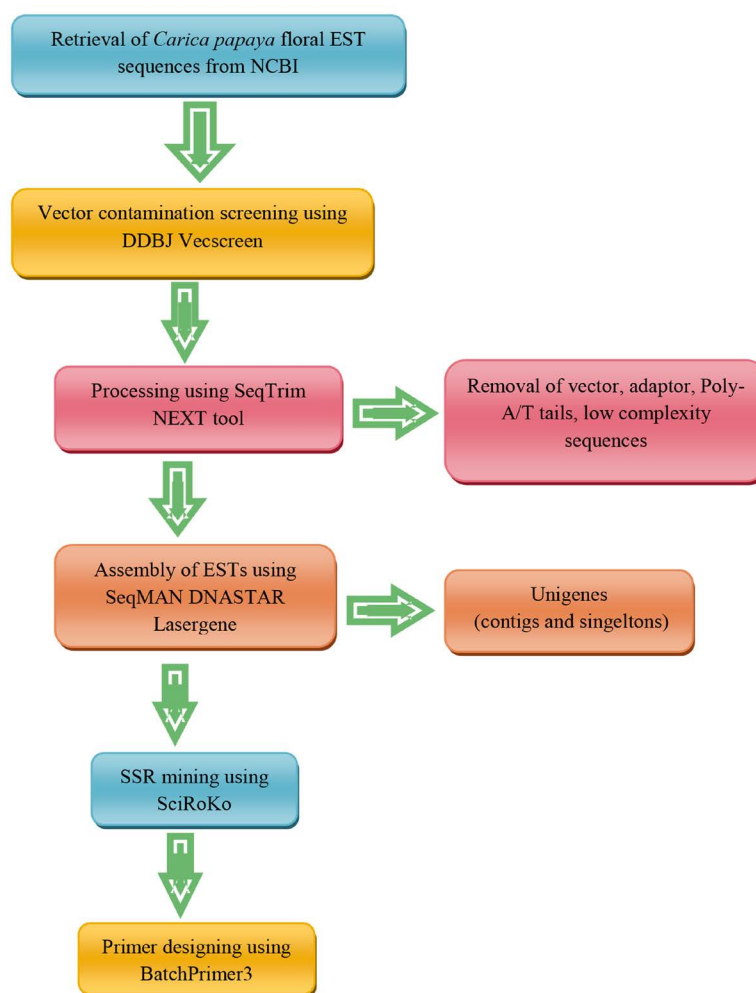


Figure 1. Flowchart showing methodology of *in silico* mining and development of EST-SSR primers from papaya floral ESTs.

percent = 80%). This software provides contig, singletons and statistical information. The sequences which cannot be grouped due to their low similarity to other ESTs results in singletons. Contigs and singletons constitute non-redundant dataset therefore were used for SSRs identification.

2.4. Detection of Genic Microsatellite

The potential SSRs were detected in the assembled floral ESTs by submitting the sequences to a SSR mining tool, SciRoKo 2.1 version. The minimum repeat unit was defined as 4 for mono- and di-nucleotide, 3 for tri-, tetra-, penta- and hexanucleotides, respectively [42] (the numbers here indicating repeat unit *i.e.* minimum number of times the motif was repeated). Imperfect SSR analysis was done under the mismatched and fixed penalty search mode of SciRoKo tool. This program takes a FASTA formatted sequence file as an input and produces an output file with sequence name, counts of SSR, SSR type, SSR motif, repeat number, the length of the sequence and GC content. SciRoKo is freely available on internet which can be downloaded and installed in the PC.

2.5. Primer Designing

Microsatellites containing floral EST sequences were used to design flanking forward and reverse EST-SSR primer pairs using online software BatchPrimer3 v1.0 with default parameters

(<http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>). BatchPrimer3 is a primer design tool based on Primer3 [43] that can accept in input up to 500 sequences at a time. The major criteria for primer designing were as follows: primer length (18 - 23 bp, with optimum value 20 bp); T_m (57°C - 63°C, with optimum value 60°C); GC content (40% - 60%, with the optimum value 50%); maximum T_m difference between forward and reverse primer 1.5°C and product size range (100 - 300 bp with optimum value 150 bp). Twenty eight primer pairs were custom synthesized from these designed primers by Eurofins Genomics, Bangalore, India.

3. Results

3.1. Retrieval, Processing and Assembly of Papaya Floral ESTs

A total of 75,846 papaya floral ESTs were downloaded from NCBI in FASTA format. All EST sequences were screened by DDBJ VecScreen for identification of vector, adaptor contaminations, low complexity sequences and poly-A/T tails. EST sequences were processed using SeqTrim NEXT for the removal of these contaminations. A total of 59,522 floral EST sequences were obtained after processing (Table 1). Processed floral EST sequences were assembled using SeqMan DNASTAR Lasergene ver. 9.0 program with its default parameters. A total of 26,039 floral unigenes (7960 contigs and 18,079 singletons), were generated after assembly of papaya floral EST sequences (Table 1). These assembled floral unigenes were further utilized for mining of SSRs.

Table 1. Summary of *in silico* mining of EST-SSRs from papaya floral EST database.

Parameters	Value
Total number of raw EST sequence	75,846
ESTs remaining after processing (vector and PolyA/T removal)	59,522
Number of contigs	7960
Numbers of singletons	18,079
Unigenes (contigs + singletons)	26,039
Total number of perfect SSRs	433,782 SSRs
Total number of compound SSRs	204,968 SSRs
Total number of Imperfect SSRs	6061 SSRs
Average density of Perfect SSR	3610.84 SSR/Mb
Average imperfect SSR count	5118.52 SSR/Mb
Average compound SSR count	50.45 SSR/Mb

3.2. Frequency, Distribution and Characterization of SSR Repeat Types

SSRs in the floral ESTs were mined using the SSR mining tool, SciRoKo. The mined EST-SSRs were classified into three types on the basis of repeat sequences; perfect SSRs, containing single motif; imperfect SSRs, with a pair of bases are present within the repeat motif that does not match the motif sequence; and compound SSRs, containing more than two adjacent different motifs [35]. Mining from floral unigenes resulted in 433,782 perfect SSRs with average density of 3610.84 SSR/Mb, 204,968 compound SSRs with average density of 50.45 SSR/Mb and 6061 imperfect SSRs with average density of 5118.52 SSR/Mb, respectively (Figure 2). The frequency distribution of mined perfect SSR repeat types is presented in Figure 3 and Figure 4. It was observed that mononucleotide repeats (411,156; 94.7%) were the most abundant perfect repeat type, followed by trinucleotide repeats (13,792; 3.1%), and dinucleotide repeats (7697; 1.7%) in floral unigenes, respectively. The frequencies of tetra-, hexa- and penta-nucleotide repeat types accounted for only (772; 0.17%), (203; 0.04%) and (162; 0.03%) in floral unigenes, respectively (Figure 3, Table 2).

3.3. Frequency, Distribution and Characterization of SSR Repeat Motifs

During standardization, the reverse complements of microsatellite motifs were considered, and similar microsatellite motifs are grouped together, for example,

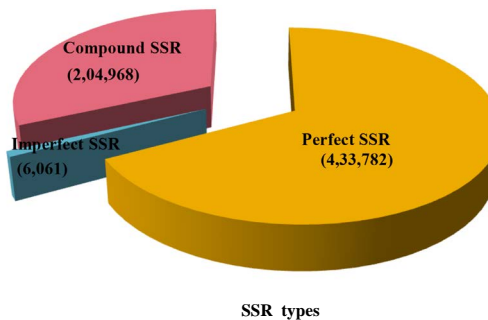


Figure 2. Abundance of different SSR types in papaya floral unigenes.

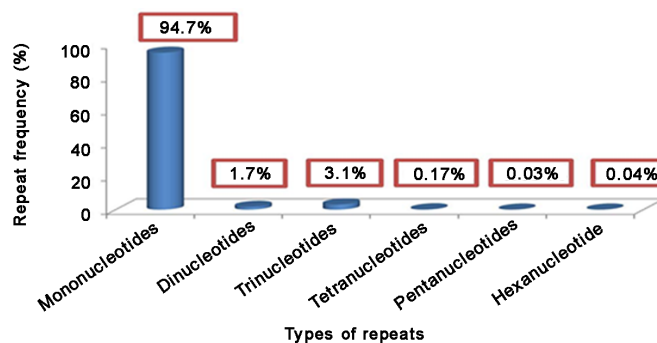


Figure 3. Frequency distribution of perfect SSR repeat types in papaya floral unigenes.

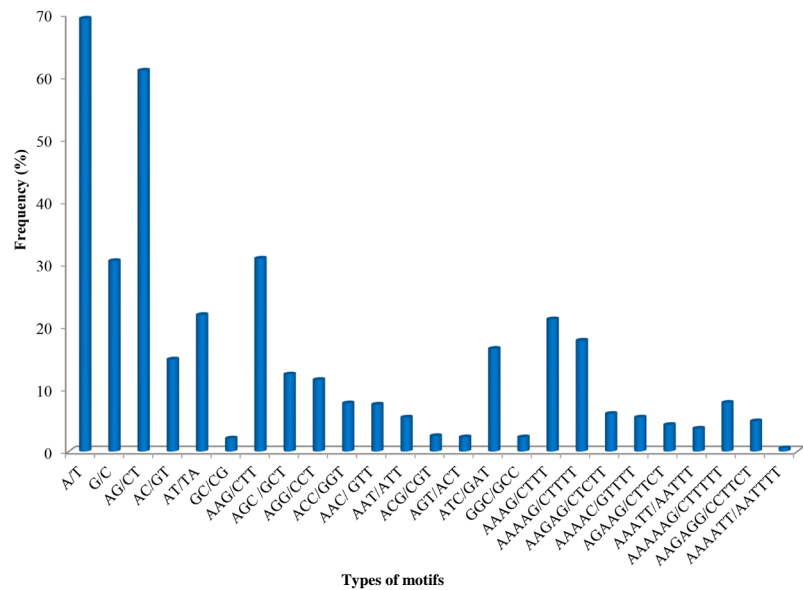


Figure 4. Frequency distribution of different repeat motifs in papaya floral unigenes.

a poly-A repeat is equivalent to a poly-T repeat on a complementary strand, AC is equivalent to CA in different reading frames and to TG, GT on a complementary strand. Similarly in trinucleotide an AGC motif is equivalent to CGA and GCA in different reading frames and to GCT, TCG and TGC on complementary strand. Thus, there are two possible combinations for mononucleotide motifs, four possible dinucleotide motifs, ten possible trinucleotides, 33 possible tetranucleotides, 102 for pentanucleotides, and 350 for hexanucleotide motifs (**Table 2**). Frequency distribution of abundant mono-, di-, tri-, tetra-, penta- and hexanucleotides perfect SSR motifs is presented in **Figure 5**. In mononucleotide repeats, the most abundant SSR motif was A/T (69.3%), while G/C was accounted only (30.6%) in floral unigenes. Among dinucleotide repeats, most frequent motif was AG/CT (61%), while GC/CG (2.1%) was least frequent in floral unigenes. In trinucleotide repeats, most abundant motif was AAG/CTT (31%), while GGC/GCC (2.3%) was least frequent trinucleotide motif. Among tetranucleotides repeats, AAAG/CTTT (21.3%) was most frequent motif. Among pentanucleotide repeats, AAAAG/CTTTT (17.9%) was most abundant motif. In hexanucleotide repeats, AAAAAG/CTTTT (7.8%) was most frequent in floral unigenes.

A total of 176 different types of motifs of imperfect SSRs were identified in floral unigenes. In imperfect SSRs, mononucleotide repeats 3426 (56.5%) was most abundant (**Table 3**).

3.4. Primer Designing

In this study, a total of 3807 primer pairs for floral papaya ESTs (except mononucleotide repeats) were successfully designed using BatchPrimer3 v1.0. Twenty eight primer pairs were custom synthesized from these designed primers. The details of EST-SSR primers along with their T_m , product size, GC%, corresponding SSR motifs are listed in (<https://goo.gl/sTJUdn>).

Table 2. The occurrence of different perfect SSR motif types in papaya floral unigenes.

Motif types	Occurrence	Frequency (%)
Mononucleotides		
A/T	285,074	69.3%
G/C	126,082	30.6%
Subtotal	411,156	
Dinucleotides		
AG/CT	4695	61%
AC/GT	1147	14.9%
AT/TA	1693	22%
GC/CG	162	2.1%
Subtotal	7697	
Trinucleotides		
AAG/CTT	4280	31%
AGC/GCT	1725	12.5%
AGG/CCT	1606	11.6%
ACC/GGT	1082	7.8%
AAC/GTT	1053	7.6%
AAT/ATT	758	5.5%
ACG/CGT	351	2.5%
AGT/ACT	318	2.3%
ATC/GAT	2298	16.6%
GGC/GCC	321	2.3%
Subtotal	13,792	
Tetranucleotides		
AAAG/CTTT	165	21.3%
Others	607	
Subtotal	772	
Pentanucleotides		
AAAAG/CTTTT	29	17.9%
AAGAG/CTCTT	10	6.1%
AAAAC/GTTTT	9	5.5%
AGAAG/CTTCT	7	4.3%
AAATT/AATTT	6	3.7%
Others	101	
Subtotal	162	
Hexanucleotides		
AAAAAG/CTTTTT	16	7.9%
AAGAGG/CCTTCT	10	4.9%
AAAATT/AATTTT	1	0.5%
Others	176	
Subtotal	203	
Overall total	433,782	

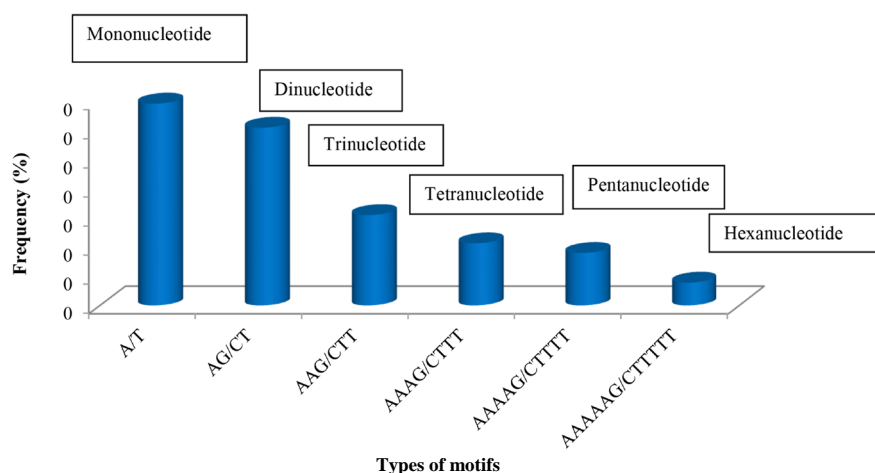


Figure 5. Frequency distribution of abundant mono-, di-, tri-, tetra-, penta- and hexanucleotides perfect SSR motifs in papaya floral unigenes.

Table 3. Summary of the imperfect SSR identified in papaya floral unigenes.

SSR motif types	Count	Types of motifs	Average mismatches	Amount of SSR /Mbp in papaya floral ESTs	GC Content
Mononucleotide	3426	2	0.19	171.11	0.27
Dinucleotide	753	4	0.55	37.61	0.38
Trinucleotide	1237	10	0.62	61.78	0.42
Tetranucleotide	257	25	0.37	12.84	0.31
Pentanucleotide	184	43	0.38	9.19	0.34
Hexanucleotide	204	92	0.56	10.19	0.44
Total	6061	176	0.44	50.45	0.36

4. Discussion

4.1. Identification and Characterization of Papaya Floral EST-SSRs

In this study, a total of available 75,846 *C. papaya* floral ESTs were downloaded from NCBI, USA. Assembly generated a total of 26,039 floral unigenes. All three types namely perfect, imperfect and compound SSRs were identified using the SciRoko program. The perfect SSR types (433,782) were most abundant, followed by compound (204,968) and imperfect SSRs (6061) in papaya floral unigenes. The amount of perfect EST-SSRs (433,782) mined in the present study is higher as compared to the earlier report of papaya in which 10,688 SSRs were identified [39]. In the present study, the average frequency or density of perfect, imperfect and compound SSR were identified as 3610.84 SSR/Mb, 50.45 SSR/Mb and 5118.52 SSR/Mb respectively which is higher than the previous studies of papaya in which density of perfect SSRs reported were 1340 SSR/Mb [39], 746 SSR/Mb [44] and 656 SSR/Mb [45]. Such variations in the density of identified microsatellites are usual among different reports, mainly due to differences in the algorithms, parameter settings, minimal repeat length, the SSR search crite-

ria, the size of the dataset for analysis, and the database-mining tools [46].

Characterization of SSR analysis revealed that, the mononucleotide repeats were most abundant 411,156 (94.7%) in papaya floral unigenes which is similar to previous reports on several plant species namely pea [47], olive [48], *Camellia sinensis* [49], tobacco [50] and *Taxodium zhong shansa* [51], but in contrast to those from pineapple [52], forage legume [53] in which di- and tri-nucleotides were identified as the most abundant repeats. The frequency of mononucleotide repeats was found to be highest (94.7%) as compared to other repeat types in our study which is similar to previous report of papaya in which mononucleotides contributed maximum (69.1%) [40]. The abundance of mononucleotide repeats in assembled ESTs suggests that they are present within the expressed regions and not at the end of the mRNA sequences [54]. Mononucleotide repeats have been used to study the population genetic in chloroplast genomes [55]. Trinucleotides (13,792; 3.1%) were the second most abundant repeats followed by dinucleotides (7697; 1.7%). Rest of the SSR types including tetra-, penta- and hexa-nucleotides were found in low frequency only 772 (0.17%), 203 (0.04%), 162 (0.03%) respectively.

In mononucleotides repeats, the most abundant motif was A/T (69.3%), which is similar with previous reports on *Allium sativum* [54] and *Humulus lupulus* [56]. In dinucleotides, AG/CT motif (61%) was most abundant in this study. The similar trend of AG/CT was also found in several plant species such as coffee [57], Madagascar periwinkle [58] and European hazelnut [59]. AG/CT is a common dinucleotide motif among plant genome [12]. One possible explanation is that AG/CT motifs frequently occur in 5' UTRs and involved in gene regulation [46]. GC/CG motif (2.1%) was least frequent in floral unigenes, indicating selective pressure against this class of repeats. In trinucleotides, AAG/CTT motif (31%) was most abundant and AGT and GGC were the least frequent which is in agreement with earlier reports on *Pisum sativum* [60] and *Salix, Eucalyptus* [61] but in contrast to finger millet [62], in which CGG motif was most abundant indicating that the abundance of EST-SSRs usually varies between different plant species. According to Morgante *et al.*, [63], AAG/CTT is most common trinucleotide motif among dicotyledonous plants, while CCG/CGG is a specific feature of monocot genome [12].

In this study, imperfect SSRs were also mined in which mononucleotides repeats 3426 (56.5%) were most abundant. The variations in frequency, distribution and abundance of SSRs among different plant species depend on various factors such as, the SSR search criteria, the size of the dataset for analysis, and the database-mining tools [46].

4.2. Development of Floral EST-SSR Primers

Microsatellites are usually characterized by the presence of conserved flanking sequences. In this study, a total of 3807 primer pairs for floral papaya ESTs were successfully designed. The remaining SSR containing sequences either fail to

generate primer-pair due to any or all of the following reasons, 1) flanking sequences are too short, 2) due to unavailability of flanking site for primer designing, or 3) it did not match the primer designing criteria of BatchPrimer3 v 1.0 software [46].

5. Conclusion

In papaya, a large amount of EST database has facilitated the identification of genic SSRs. The present study examined the frequency, type and distribution of microsatellites in floral ESTs of papaya and highlights the development of EST-SSR primers in papaya. The development of EST derived SSRs via *in silico* saves both costs and time and less labor-intensive approach. A total of 26,039 floral unigenes were generated after assembly from papaya floral EST sequences. 433,782 perfect SSRs, 204,968 compound SSRs and 6061 imperfect SSRs were identified in floral unigenes. 3807 primer pairs for floral papaya ESTs were designed and 28 primer pairs were custom synthesized from these designed primers. The floral EST-derived SSR primers reported in this study are being used in genetic diversity analysis, its cross transferability analysis among *Carica* and its related genera, and identification of sex specific markers among female, hermaphrodite and male plants of various papaya varieties. These primers could also be useful in comparative mapping to study the order of genes among closely related *C. papaya* species, and for markers-assisted selection of desirable traits (disease resistance) in papaya.

Acknowledgements

The financial assistance in the form of research projects sanctioned by Department of Science and Technology (DST), Government of India, New Delhi and Council of Science and Technology (CST), Uttar Pradesh, India and Junior Research Fellowships (JRFs) to Priyanka and Dileep Kumar by Department of Biotechnology (DBT), Government of India, New Delhi and University Grants Commission (UGC), Government of India, New Delhi, respectively, are gratefully acknowledged.

References

- [1] Araujo, F.S., Carvalho, C.R. and Clarindo, W.R. (2010) Genome Size, Base Composition and Karyotype of *Carica papaya* L. *Nucleus*, **53**, 25-31.
<https://doi.org/10.1007/s13237-010-0007-8>
- [2] Arumuganathan, K. and Earle, E.D. (1991) Nuclear DNA Content of Some Important Plant Species. *Plant Molecular Biology Reporter*, **93**, 208-219.
<https://doi.org/10.1007/BF02672069>
- [3] Damasceno-Junior, P.C., Costa, F.R., Pereira, T.N.S., Freitas Neto, M. and Pereira, M.G. (2009) Karyotype Determination in Three *Caricaceae* Species Emphasizing the Cultivated Form (*C. papaya* L.). *Caryologia*, **62**, 10-15.
<https://doi.org/10.1080/00087114.2004.10589660>
- [4] Organization for Economic Co-Operation and Development (OECD) (2005) Con-

- sensus Document on the Biology of Papaya (*Carica papaya*), OECD Environment, Health and Safety Publications, Series on Harmonization of Regulatory Oversight in Biotechnology No. 33, France.
- [5] Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V. and Lewis, K.L. (2008) The Draft Genome of the Transgenic Tropical Fruit Tree Papaya (*Carica papaya* Linnaeus). *Nature*, **452**, 991-996.
<https://doi.org/10.1038/nature06856>
- [6] Morton, J.F. and Miami, F.L. (1987) Papaya. In: Morton, J., Ed., *Fruits of Warm Climates*, Miami, 336-346.
- [7] Aravind, G., Bhowmik, D., Duraivel, S. and Harish, G. (2013) Traditional and Medicinal Uses of *Carica papaya*. *Journal of Medicinal Plants Studies*, **1**, 7-15.
- [8] Urasaki, N., Tokumoto, M., Ban, Y., Kayano, T., Tanaka, H., Oku, H., Chinen, I. and Terauchi, R. (2002) A Male and Hermaphrodite Specific RAPD Marker for Papaya (*Carica papaya* L.). *Theoretical and Applied Genetics*, **104**, 281-285.
<https://doi.org/10.1007/s001220100693>
- [9] Eustice, M., Yu, Q., Lai, C.W., Hou, S., Thimmapuram, J., Liu, L., Alam, M., Moore, P.H., Presting, G.G. and Ming, R. (2008) Development and Application of Microsatellite Markers for Genomic Analysis of Papaya. *Tree Genetics and Genomes*, **4**, 333-341. <https://doi.org/10.1007/s11295-007-0112-2>
- [10] Ma, H., Moore, P.H., Liu, Z., Kim, M.S., Yu, Q., Fitch, M.M., Sekiota, T., Paterson, A.H. and Ming, R. (2004) High Density Linkage Mapping Revealed Suppression of Recombination at the Sex Determination Locus in Papaya. *Genetics*, **166**, 419-436.
<https://doi.org/10.1534/genetics.166.1.419>
- [11] Bedoya, G.C. and Nunez, V. (2007) A SCAR Marker for the Sex Types Determination in Colombian Genotypes of *Carica papaya*. *Euphytica*, **153**, 215-220.
<https://doi.org/10.1007/s10681-006-9256-7>
- [12] Li, Y.C., Korol, A.B., Fahima, T. and Nevo, E. (2004) Microsatellites within Genes: Structure, Function and Evolution. *Molecular Biology and Evolution*, **21**, 991-1007.
<https://doi.org/10.1093/molbev/msh073>
- [13] Thiel, T., Michalek, W., Varshney, R.K. and Graner, A. (2003) Exploiting EST Databases for the Development and Characterization of Gene-Derived SSR-Markers in Barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, **106**, 411-422.
<https://doi.org/10.1007/s00122-002-1031-0>
- [14] Tautz, D. (1989) Hypervariability of Simple Sequences as General Source for Polymorphic DNA Markers. *Nucleic Acids Research*, **17**, 6463-6472.
<https://doi.org/10.1093/nar/17.16.6463>
- [15] Litt, M. and Luty, J.A. (1989) A Hypervariable Microsatellite Revealed by *in Vitro* Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. *American Journal of Human Genetics*, **44**, 397-401.
- [16] Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991) DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. *American Journal of Human Genetics*, **49**, 746-756.
- [17] Toth, G., Gaspari, Z. and Jurka, J. (2000) Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Research*, **10**, 967-981.
<https://doi.org/10.1101/gr.10.7.967>
- [18] Kalia, R.K., Rai, M.K., Kalia, S., Singh, R. and Dhawan, A.K. (2011) Microsatellite Markers: An Overview of the Recent Progress in Plants. *Euphytica*, **177**, 309-334.
<https://doi.org/10.1007/s10681-010-0286-9>

- [19] Powell, W., Machray, G.C. and Provan, J. (1996) Polymorphism Revealed by Simple Sequence Repeats. *Trends in Plant Science*, **1**, 215-222.
- [20] Zane, L., Bargelloni, L. and Patarnello, T. (2002) Strategies for Microsatellite Isolation: A Review. *Molecular Ecology*, **11**, 1-16. <https://doi.org/10.1046/j.0962-1083.2001.01418.x>
- [21] Levinson, G. and Gutman, G.A. (1987) Slipped-Strand Mispairing: A Major Mechanism for DNA Sequence Evolution. *Molecular Biology and Evolution*, **4**, 203-221.
- [22] Katti, M.V., Ranjekar, P.K. and Gupta, V.S. (2001) Differential Distribution of Simple Sequence Repeats in Eukaryotic Genome Sequences. *Molecular Biology and Evolution*, **18**, 1161-1167. <https://doi.org/10.1093/oxfordjournals.molbev.a003903>
- [23] Innan, H., Terauchi, R. and Miyashita, N.T. (1997) Microsatellite Polymorphism in Natural Populations of the Wild Plant *Arabidopsis thaliana*. *Genetics*, **146**, 1441-1452.
- [24] Senan, S., Kizhakayil, D., Sasikumar, B. and Sheeja, T.E. (2014) Methods for Development of Microsatellite Markers: An Overview. *Notulae Scientia Biologicae*, **6**, 1-13.
- [25] Nakatsuji, R., Hashida, T., Matsumoto, N., Tsuru, M., Kubo, N. and Hirai, M. (2011) Development of Genomic and EST-SSR Markers in Radish (*Raphanus sativus* L.). *Breeding Science*, **61**, 413-419. <https://doi.org/10.1270/jsbbs.61.413>
- [26] Bhattacharyya, U., Pandey, S.K. and Dasgupta, T. (2014) Identification of EST-SSRs and FDM in Sesame (*Sesamum indicum* L.) through Data Mining. *Scholarly Journal of Agricultural Science*, **4**, 60-69.
- [27] Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N. and Balyan, H.S. (2003) Transferable EST-SSR Markers for the Study of Polymorphism and Genetic Diversity in Bread Wheat. *Molecular Genetics Genomics*, **270**, 315-323. <https://doi.org/10.1007/s00438-003-0921-4>
- [28] Varshney, R.K., Thiel, T., Stein, N., Langridge and Graner, A. (2002) *In Silico* Analysis on Frequency and Distribution of Microsatellites in ESTs of Some Cereal Species. *Cell Molecular Biology Letter*, **7**, 537-546.
- [29] Anjali, N., Dharan, S., Nadiya, F. and Sabu, K.K. (2015) Development of EST-SSR Markers to Assess Genetic Diversity in *Elettaria cardamomum* Maton. *International Journal of Applied Science Biotechnology*, **3**, 188-192.
- [30] Haq, S.U., Jain, R., Sharma, M., Kachhwaha, S. and Kothari, S.L. (2014) Identification and Characterization of Microsatellites in Expressed Sequence Tags and Their Cross Transferability in Different Plants. *International Journal of Genomics*, **2014**, Article ID: 863948. <https://doi.org/10.1155/2014/863948>
- [31] Bhati, J., Sonah, H., Jhang, T., Singh, N.K. and Sharma, T.R. (2010) Comparative Analysis and EST Mining reveals High Degree of Conservation among Five Brassicaceae Species. *Comparative and Functional Genomics*, **2010**, Article ID: 520238.
- [32] Studer, B., Kolliker, R., Muylle, H., Asp, T., Frei, U., Roldan-Ruiz, I., Barre, P., Tomaszewski, C., Meally, H., Barth, S., Skot, L., Armstead, I.P., Dolstra, O. and Lubberstedt, T. (2010) EST-Derived SSR Markers Used as Anchor Loci for the Construction of A Consensus Linkage Map in Ryegrass (*Lolium* spp.). *BMC Plant Biology*, **10**, 177. <https://doi.org/10.1186/1471-2229-10-177>
- [33] Jesus, O.N.D., Freitas, J.P.X.D., Dantas, J.L.L. and Oliveira, E.J.D. (2013) Use of Morpho-Agronomic Traits and DNA Profiling for Classification of Genetic Diversity in Papaya. *Genetics and Molecular Research*, **12**, 6646-6663. <https://doi.org/10.4238/2013.July.11.8>
- [34] Sengupta, S., Das, B., Prasad, M., Acharyya, P. and Ghose, T.K. (2013) A Compar-

- tive Survey of Genetic Diversity among a Set of *Caricaceae* Accessions Using Microsatellite Markers. *Springer Plus*, **2**, 345. <https://doi.org/10.1186/2193-1801-2-345>
- [35] Oliveira, E.J.D., Silva, A.D.S., Carvalho, F.M.D., Santos, L.F.D., Costa, J.L., Amorim, V.B.D.O. and Dantas, J.L.L. (2010) Polymorphic Microsatellite Marker Set for *Carica papaya* L. and Its Use in Molecular-Assisted Selection. *Euphytica*, **173**, 279-287. <https://doi.org/10.1007/s10681-010-0150-y>
- [36] Castelo, A.T., Martins, W.S. and Gao, G.R. (2002) Tandem Repeat Occurrence Locator. *Bioinformatics*, **18**, 634-636. <https://doi.org/10.1093/bioinformatics/18.4.634>
- [37] Kofler, R., Schlotterer, C. and Lelley, T. (2007) SciRoKo: A New Tool for Whole Genome Microsatellite Search and Investigation. *Bioinformatics*, **23**, 1683-1685. <https://doi.org/10.1093/bioinformatics/btm157>
- [38] Faircloth, B.C. (2008) Msatcommander: Detection of Microsatellite Repeat Arrays and Automated, Locus-Specific Primer Design. *Molecular Ecology Resources*, **8**, 92-94. <https://doi.org/10.1111/j.1471-8286.2007.01884.x>
- [39] Wang, J., Chen, C., Na, J.K., Yu, Q., Hou, S., Paull, R.E., Moore, P.H., Alam, M. and Ming, R. (2008) Genome-Wide Comparative Analysis of Microsatellite in Papaya. *Tropical Plant Biology*, **1**, 278-292. <https://doi.org/10.1007/s12042-008-9024-z>
- [40] Zeng, F., Yu, Q., Hou, S., Moore, P.H., Alam, M. and Ming, R. (2014) Features of Transcriptome in Trioecious Papaya Revealed by a Large-Scale Sequencing of ESTs and Comparative Analysis in Higher Plants. *Plant Omics Journal*, **7**, 450-460.
- [41] Falgueras, J., Lara, A.J., Fernandez-Pozo, N., Canton, F.R., Perez-Trabado, G. and Claros, M.G. (2010) SeqTrim: A High-Throughput Pipeline for Pre-Processing Any Type of Sequence Read. *BMC Bioinformatics*, **11**, 38. <https://doi.org/10.1186/1471-2105-11-38>
- [42] Fluch, S., Burg, A., Kopecky, D., Homolka, A., Spiess, N. and Vendramin, G.G. (2011) Characterization of Variable EST SSR Markers for Norway Spruce (*Piceaabies* L.). *BMC Research Notes*, **4**, 401. <https://doi.org/10.1186/1756-0500-4-401>
- [43] Rozen, S. and Skaletsky, H.J. (2000) Primer3 on the WWW for General Users and for Biologist Programmers. In: Krawetz, S. and Misener, S., Eds., *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Humana Press, Totowa, 365-386.
- [44] Shi, J., Huang, S., Fu, D., Yu, J., Wang, X., Hua, W., Liu, S., Liu, G. and Wang, H. (2013) Evolutionary Dynamics of Microsatellite Distribution in Plants: Insight from the Comparison of Sequenced Brassica, Arabidopsis and Other Angiosperm Species. *PLoS ONE*, **8**, e59988. <https://doi.org/10.1371/journal.pone.0059988>
- [45] Vidal, N.M., Grazziotin, A.L., Ramos, H.C.C., Pereira, M.G. and Venancio, T.M. (2014) Development of a Gene-Centered SSR Atlas as a Resource for Papaya (*Carica papaya*) Marker-Assisted Selection and Population Genetic Studies. *PLoS ONE*, **9**, e112654. <https://doi.org/10.1371/journal.pone.0112654>
- [46] Varshney, R.K., Graner, A. and Sorrells, M.E. (2005) Genic Microsatellite Markers in Plants: Features and Applications. *Trends Biotechnology*, **23**, 48-55.
- [47] Mishra, R.K., Gangadhar, B.H., Nookaraju, A., Kumar, S. and Park, S.W. (2012) Development of EST-Derived SSR Markers in Pea (*Pisum sativum*) and Their Potential Utility for Genetic Mapping and Transferability. *Plant Breeding*, **131**, 118-124. <https://doi.org/10.1111/j.1439-0523.2011.01926.x>
- [48] Adawy, S.S., Mokhtar, M.M., Alsamman, A.M. and Sakr, M.M. (2013) Development of Annotated EST-SSR Database in Olive (*Olea europaea*). *International Journal of Science and Research*, **4**, 1063-1073.
- [49] Sahu, J., Sarmah, R., Dehury, B., Sarma, K., Sahoo, S., Sahu, M., Barooah, M., Modi,

- M.K. and Sen, P. (2012) Mining for SSRs and FDMs from Expressed Sequence Tags of *Camellia sinensis*. *Bioinformation*, **8**, 260-266. <https://doi.org/10.6026/97320630008260>
- [50] Cai, C., Yang, Y., Cheng, L., Tong, C. and Feng, J. (2015) Development and Assessment of EST SSR Marker for the Genetic Diversity among Tobaccos (*Nicotiana tabacum* L.). *Russian Journal of Genetics*, **51**, 591-600. <https://doi.org/10.1134/S1022795415020064>
- [51] Cheng, Y., Yang, Y., Wang, Z., Qi, B., Yin, Y. and Li, H. (2015) Development and Characterization of EST-SSR Markers in *Taxodium "zhongshansa"*. *Plant Molecular Biology Reporter*, **33**, 1804-1814. <https://doi.org/10.1007/s11105-015-0875-9>
- [52] Wohrmann, T. and Weising, K. (2011) *In Silico* Mining for Simple Sequence Repeat Loci in a Pineapple Expressed Sequence Tag Database and Cross-Species Amplification of EST-SSR Markers Across Bromeliaceae. *Theoretical and Applied Genetics*, **123**, 635-647. <https://doi.org/10.1007/s00122-011-1613-9>
- [53] Ding, X., Jia, Q., Luo, X., Zhang, L., Cong, H., Liu, G. and Bai, C. (2015) Development and Characterization of Expressed Sequence Tag-Derived Simple Sequence Repeat Markers in Tropical Forage Legume *Stylosanthes guianensis* (Aubl.) Sw. *Molecular Breeding*, **35**, 202. <https://doi.org/10.1007/s11032-015-0370-x>
- [54] Chand, S.K., Nanda, S., Rout, E. and Joshi, R.K. (2015) Mining, Characterization and Validation of EST Derived Microsatellites from the Transcriptome Database of *Allium sativum* L. *Bioinformation*, **11**, 145-150. <https://doi.org/10.6026/97320630011145>
- [55] Tripathi, K.P., Roy, S., Maheshwari, N., Khan, F., Meena, A. and Sharma, A. (2009) SSR Polymorphism in *Artemisia Annuua*: Recognition of Hotspots for Dynamics Mutation. *Plant Omics Journal*, **2**, 228-237.
- [56] Singh, S., Gupta, S., Mani, A. and Chaturvedi, A. (2012) Mining and Gene Ontology Based Annotation of SSR Markers from Expressed Sequence Tags of *Humulus lupulus*. *Bioinformation*, **8**, 114-122. <https://doi.org/10.6026/97320630008114>
- [57] Aggarwal, R.K., Hendre, P.S., Varshney, R.K., Bhat, P.R., Krishnakumar, V. and Singh, L. (2007) Identification, Characterization and Utilization of EST-Derived Genic Microsatellite Markers for Genome Analyses of Coffee and Related Species. *Theoretical and Applied Genetics*, **114**, 359-372. <https://doi.org/10.1007/s00122-006-0440-x>
- [58] Mishra, R.K., Gangadhar, B.H., Yu, J.W., Kim, D.H. and Park, S.W. (2011) Development and Characterization of EST Based SSR Markers in Madagascar Periwinkle (*Catharanthus roseus*) and Their Transferability in Other Medicinal Plants. *Plant Omics Journal*, **4**, 154-162.
- [59] Boccacci, P., Beltramo, C., Prando, M.A.S., Lembo, A., Sartor, C., Mehlenbacher, S.A., Botta, R. and TorelloMarinoni, D. (2015) *In Silico* Mining, Characterization and Cross-Species Transferability of EST-SSR Markers for European Hazelnut (*Corylus avellana* L.). *Molecular Breeding*, **35**, 21. <https://doi.org/10.1007/s11032-015-0195-7>
- [60] Teshome, A., Bryngelsson, T., Dagne, K. and Geleta, M. (2015) Assessment of Genetic Diversity in Ethiopian Field Pea (*Pisum sativum* L.) Accessions with Newly Developed EST-SSR Markers. *BMC Genetics*, **16**, 102. <https://doi.org/10.1186/s12863-015-0261-5>
- [61] He, X., Zheng, J., Zhou, J., He, K., Shi, S. and Wang, B. (2015) Characterization and Comparison of EST-SSRs in *Salix*, *Populus*, and *Eucalyptus*. *Tree Genetics and Genomes*, **11**, 820. <https://doi.org/10.1007/s11295-014-0820-3>

- [62] Babu, B.K., Pandey, D., Agrawal, P.K., Sood, S. and Kumar, A. (2014) *In Silico* Mining, Type and Frequency Analysis of Genic Microsatellites of Finger Millet (*Eleusine coracana* (L.) Gaertn.): A Comparative Genomic Analysis of NBS-LRR Regions of Finger Millet with Rice. *Molecular Biology Reports*, **41**, 3081-3090. <https://doi.org/10.1007/s11033-014-3168-8>
- [63] Morgante, M., Hanafey, M. and Powell, W. (2002) Microsatellites are Preferentially Associated with Nonrepetitive DNA in Plant Genomes. *Nature Genetics*, **30**, 194-200. <https://doi.org/10.1038/ng822>



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ajps@scirp.org