

Development of Microsatellite Markers in Mango (*Mangifera indica* L.) using 5' anchored PCR

Kittipat Ukoskit

Department of Biotechnology, Faculty of Science and Technology,
Thammasat University (Rangsit Campus) Khong Luang Pathum Thani, 12121 Thailand

Abstract

A 5' anchored PCR was utilized to develop microsatellite markers in mango. Di-repeat primers containing 3 or 7 degenerate bases at the 5' end of the primers were used to amplify genomic DNA flanked by simple sequence repeats. The PCR products were cloned to generate enriched microsatellite libraries. Clones containing microsatellite fragments were randomly selected and subsequently sequenced. A specific primer was designed and used in combination with a 5' anchored primer to produce a locus-specific microsatellite marker. The results indicated that 7 degenerated bases of 5' anchored primers reduced the efficiency of 5' anchored PCR. Sixteen microsatellite markers were developed and were able to reveal 46 alleles in 8 mango varieties.

Keywords: Mango, Microsatellite Markers, 5' anchored PCR

1. Introduction

Total world mango production reached over 26 million tons in 2002. The annual mango production ranked fifth in the world fruit production, following bananas, oranges, grapes, and apples. A few countries (India, China, Mexico, Pakistan and Indonesia) including Thailand, account for over 75% of the world production [1].

Mango cultivars are commonly divided into two groups, monoembryonic and polyembryonic. Monoembryonic mangos, which are mostly subtropical (Indian types), produce a single zygotic embryo from seeds, whereas polyembryonic mangos, which are mostly tropical (southeast Asian types), produce at least one embryo, and usually one of them is zygotic, and the remaining are derived from the nucellus. Most of the mango cultivars in Thailand belong to the Southeast Asian types. Currently, there are over 200 different mango cultivar names in Thailand. Many cultivar names are thought to be synonyms for the same cultivar, which in turn causes confusion among breeders and farmers, because many clonally propagated mango cultivars have unique local and regional names. Cultivar identification based on phenotypic traits is inaccurate due to the influence of the

environment and the limiting number of discriminating traits. Recently, molecular identification has been carried out with different molecular systems in many fruit tree species [2].

Microsatellites, or simple-sequence repeats (SSRs), have proven to be particularly valuable in plant breeding and fingerprinting because they are multi-allelic, co-dominantly inherited, widely dispersed across the genome, easily scored, and their analysis can be automated [3,4,5,6]. In species such as mango and most fruit trees, no or few sequences are available in public sequence databases or in related species. The conventional method for developing microsatellite markers involves library construction and screening, DNA sequencing, PCR primer design and PCR optimization [5]. The development of microsatellite markers, therefore, has been time-consuming and expensive. In addition, the recovery rate of useful microsatellites is generally low due to a variety of causes including non-polymorphic loci, primer pairs failing to give a good amplification, and clones positive in the screening, which nevertheless do not contain microsatellites, or lack of sufficient flanking sequence suitable for primer design. As a consequence, large-scale microsatellite markers

have only been developed for commercially importance crop species such as cereals, maize, rice, wheat, and soybeans [7,8,9].

Several different approaches have been described to improve the efficiency of microsatellite marker development. These strategies have attempted to target specific regions of the genome containing microsatellites [7,10] and to reduce the cost of marker development by using techniques that reduce the number of primers required [11,12]. In 5' anchored PCR [11,13], a primer was designed to anneal at the 5' end of the repeated sequence anchored by the additional non-repeat bases. The 5' anchored PCR product contained a microsatellite at both ends. A specific primer can be designed and used in combination with a 5' anchored primer to generate a locus-specific microsatellite marker.

The work presented in this report describes the isolation of microsatellite markers from mango using 5' anchored PCR. The effect of the number of anchored degenerate bases at the 5' end of the anchored primer on amplification of the original repeat length products in enriched microsatellite libraries was examined.

2. Materials and methods

2.1 Plant materials and DNA isolation

The eight commercial mango varieties used to observe for polymorphic microsatellite markers were Num Dok Mai-Ta Why, Num Dok Mai-Si Thong, Bun Ban Dan, Choke Anun, Kiao Sway, Fha Lan, Mae Look Duok and Mon Dueun Kao. All varieties represent commercial varieties for Thailand mango production. Total DNA was extracted from 1-5 g of young leaves from plants grown in the field according to the procedure described by Gawel and Jarret [14]. Num Dok Mai-Ta Why and Num Dok Mai-Si Thong represent a close genetic relationship.

2.2 Microsatellite libraries construction by 5' anchored PCR

Microsatellite libraries were constructed using the method of Fisher et al. (1996) [11]. Briefly, degenerate primers designed to anchor at the 5' end of microsatellite sequences were used to amplify genomic DNA isolated from Num Dok Mai - Si Thong. Di-repeated primers containing 3 and 7 degenerate bases at the 5' end

of repeats were comprised of BDB(CA)₇ and CHY(GA)₇ and KKVRVRV(CT)₆, KKVHVHV(GT)₆, KKKYRYYY(AC)₆, KKRVRVR(TC)₆ and KKHVHVH(TG)₆, respectively (R=A,G; Y=C,T; K=G,T; H=A,C,T; B=C,G,T; V=A,C,G; D=A,G,T). PCRS were carried out in 20 µl volumes containing 1X PCR buffer (1.5 mM MgCl₂, 10 mM Tris HCl pH 8.3, 50 mM KCl), 3 U *Taq* polymerase, 50 pmol 5' anchored primer, 30 ng genomic DNA, 200 µM each of dNTPs. Reactions were performed in a PCR system (2400 Perkin-Elmer) using the following amplification profile; initial denaturation at 94 °C for 3 min; 5 cycles of denaturation at 93 °C for 30 s, annealing at 59 °C (or $T_m + 2$ °C, where T_m is the melting temperature of a primer) and extension at 72 °C for 30 s; 35 cycles of denaturation at 93 °C for 30 s, annealing at 57 °C (or T_m) and extension at 72 °C for 30 s; final extension at 72 °C for 2 min. Products were resolved on 6% denaturing polyacrylamide gels containing 1X TBE buffer and 5 M urea at 80 W constant power for 2-4 h and visualized using a silver staining procedure. The microsatellite libraries were constructed from PCR products using TOPO TA Cloning kit (Quigen) according to the manufacturer's instructions.

2.3 Screening, Primer design and PCR analysis

Every clone in the microsatellite libraries generated by this method contains two terminal microsatellites. Instead of screening clones containing microsatellites using repeated sequence probed hybridization, a number of clones greater than the number of observed bands were randomly selected for sequencing. The selected clones were sequenced using M13 universal primers and an ABI 310 automated sequencer according to the manufacturer's instructions. The redundancy of clones containing microsatellites was determined and a unique sequence was selected. Specific primers were designed using Primer 3 software [15]. For analysis of polymorphism of a microsatellite locus, the anchored primer was used in combination with the specific primer. Touch down PCRs were performed in a PCR system (2400 Perkin Elmer) using a denaturing step of 30 s at 92 °C and an extension step of 30 s at 72 °C. The initial annealing was done for 60 s at 10 °C above the T_m of the specific primer and subsequently lowered by 1 °C each cycle until

the T_m was reached, which was maintained for the remaining 35 cycles of amplification. The PCR was performed in a volume of 20 μ l containing 50 ng of genomic DNA, 20 pmol of the 5' anchored primer, 4 pmol of the specific primer, 200 μ M each of dNTPs, 1X PCR buffer, and 1 U of *Taq* polymerase. The PCR products were separated on 6% denaturing polyacrylamide gels and visualized using a silver staining procedure. The degree of polymorphism was quantified using the polymorphism information content (PIC) [16].

3. Results and discussion

An anchored primer was designed to anneal at the 5' end of the repeated sequence at

different genomic locations anchored by degenerate bases. The degenerate bases were designed to avoid priming from within a repeated sequence. The anchored primers were expected to amplify complete microsatellite regions. Therefore every PCR product should contain repeated sequences at both ends. A 5' anchored PCR generated multiple bands varying in size from approximately 200 bp to 1000 bp (Fig 1.). Two types of polymorphisms, codominance and dominance, may occur. Codominant polymorphisms resulting from the difference in number of repeats were indicated by ladders of bands of similar but slightly different sizes between the genotypes (lane 1 and 2 in Fig. 1d as indicated by the arrow).

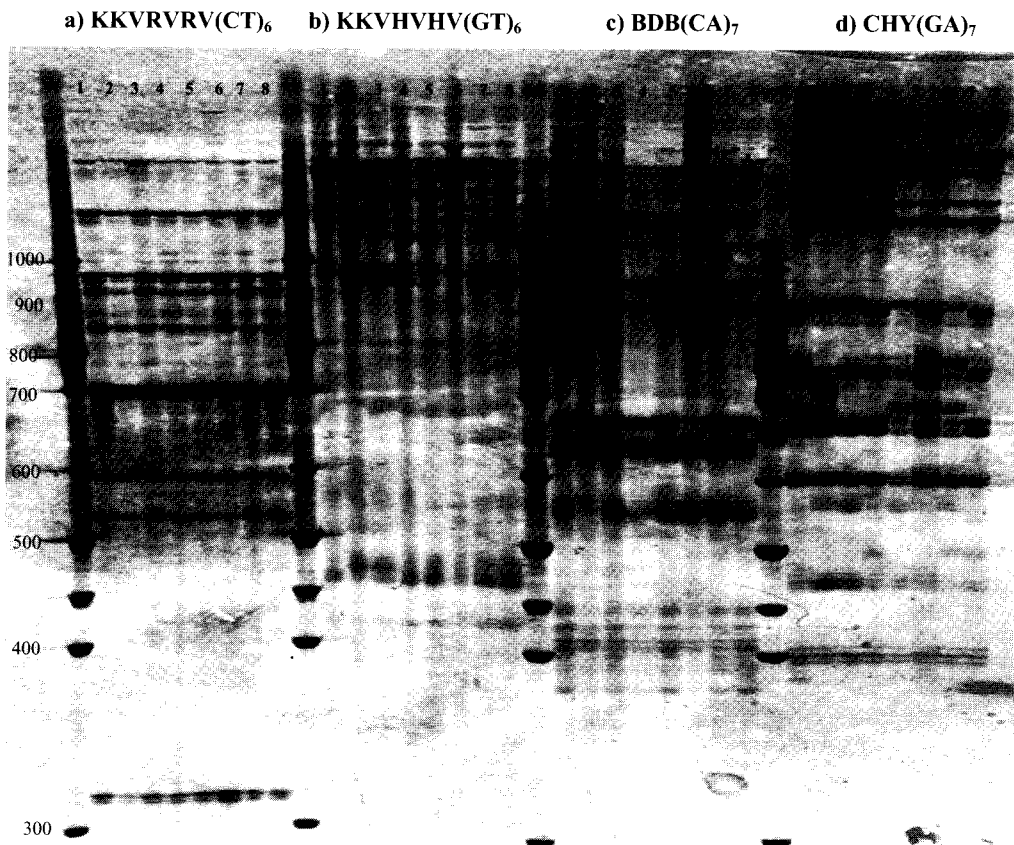


Fig. 1. Examples of polymorphism among 8 mango varieties detected using four 5' anchored primers. The arrow in 1b indicates polymorphic products generated by the 5' anchored primer working in a manner similar to RAPD. The arrow in 1d indicates the repeated length polymorphism. #Lanes M) 100 bp ladder, 1) Num Dok Mai – Si Thong 2) Choke Anun, 3) Mae Look Duok, 4) Bun Ban Dan, 5) Fha Lan, 6) Num Dok Mai – Ta Why, 7) Mon Dueun kao and 8) Kiao Sway

A 5' anchored PCR works in a manner similar to Random Amplified Polymorphic DNA (RAPD), amplifying disperse regions in the genome and not specifically a microsatellite as indicated by the presence or absence of the bands of the same size [17]. The example illustrated in Fig 1b is indicated by the arrow. The band was absent in 2 mango varieties, Num Dok Mai-Ta Why (1) and Num Dok Mai-Si Thong (6), but present in the other 6 varieties. The pattern similarity seen in these 2 varieties inferred close genetic relationships as presented in the previous work (unpublished data).

3.1 Library screening and sequence characterization

The enrichment method involves amplifying regions of the genome where two adjacent and oppositely oriented microsatellites

provide primer binding sites. A total of 64 clones were randomly selected from all microsatellite libraries and subsequently sequenced. Every clone contained microsatellites at both ends with at least as many repeats as in the 5' anchored primers. In addition the presence of additional internal microsatellites was observed in some clones, implying that there may be clustering of microsatellites in some genomic regions. Even though the enrichment procedure used in this experiment yielded a very high frequency of clones containing microsatellites, there was evidence for increased redundancy. The result showed that of 64 clones, 39 clones had unique sequences (Table 1). The types and number of repeats of 39 unique clones are presented in Table 2.

Table 1. Number of 5' anchored degenerated bases, sequence of 5' anchored primers, number of selected clones from all microsatellite libraries, number of unique clones analyzed from the total number of selected clones, and number of clones containing more repeats than those present in the 5' anchored primer.

5' anchored bases	Primers	Number of selected clones	Number of unique clones	Number of clones containing more repeats than those in the primers
3	BDB (CA) ₇	10	4	4
	CHY(GA) ₇	7	5	5
	Total	17	9	9
7	KKVRVRV (CT) ₆	14	3	0
	KKVHVHV(GT) ₆	4	1	1
	KKRYRYY(AC) ₆	9	9	2
	KKRVRVR(TC) ₆	9	9	1
	KKHVHVH(TG) ₆	11	8	5
	Total	47	30	9

In this experiment 5' anchored primers were designed in such a way that in addition to the microsatellite sequence, 3 or 7 degenerated bases were added immediately 5' to the microsatellite. The 5' degenerated bases are expected to anchor the primer to the 5' end of a microsatellite. In practice, this approach was difficult to optimize. These difficulties arise because the 5' anchored primers fail to anchor at the 5' end of microsatellites, instead slipping to the 3' of the microsatellites during PCR [13]. Each amplified fragment therefore contains exactly the number of repeats as in the primers and any repeat length polymorphism is lost. In

another case of RAPD, the 5' anchored primers were used one at a time during PCR, amplifying genomic location flanked by inversely oriented, closely spaced repeated sequences. Sequence analysis of 39 unique clones revealed that 18 microsatellite markers contained more terminal repeats than found in the 5' anchored primer (Table 2). All 9 unique clones produced by the 5' anchored primers containing of 3 anchored bases had terminal microsatellites of more than 7 repeats, ranging from 8 to 15 repeats, confirming that anchoring had been successful. Whereas 21 out of 30 unique clones produced by the 5' anchored primers containing 7

anchored bases, had an equal number of repeats as the primer. The results indicated that low specificity of primers to the microsatellites due to a large number of 5' anchored bases permitted

slippage of the primer to the 3' end of the target microsatellite loci, resulting in loss of repeat length variation. These 21 clones therefore were not selected for designing specific primers.

Table 2 Microsatellites isolated with di-repeated primers containing 3 (clone number 1-9) and 7 (clone number 10-39) degenerated bases at the 5' end of the repeats

No.	5' terminal repeat ^a	Internal repeat	3' terminal repeat ^a
1	CGT (CA) ₇	(AT) ₅ (TA) ₄ (TA) ₅	(TG) ₁₄ ATG
2	CGG (CA) ₁₀	-	(TG) ₁₄ ACC
3	TGT (CA) ₈	(TG) ₄	(TG) ₁₄ ACC
4	CAT (CA) ₁₄	(TA) ₅ (TA) ₄	(TG) ₁₄ ACG
5	CCT (GA) ₁₁	-	(TC) ₇ GAG
6	CAC (GA) ₇	-	(TC) ₈ ATG
7	CAC (GA) ₈	-	(TC) ₁₂ GGG
8	CTC (GA) ₈	(CT) ₆	(TC) ₁₀ GGG
9	CAT (GA) ₁₅	-	(TC) ₇ GGG
10	GTGAGGG (CT) ₆	-	(AG) ₆ CTGCTCA
11	GTCACGG (CT) ₆	-	(AG) ₆ CCTTCCA
12	TGGAAGG (CT) ₆	-	(AG) ₆ CCGTGAC
13	TGGGACC (GT) ₇	-	(AC) ₆ TCGCACA
14	GTATGCT (AC) ₆	-	(GT) ₆ AATGTCA
15	TGGCACT (AC) ₆	-	(GT) ₆ GACGCCC
16	GTACGTT (AC) ₆	-	(GT) ₆ GGCGTCA
17	GTGTGCC (AC) ₈	-	(GT) ₆ AATGCCA
18	GGATGCT (AC) ₆	-	(GT) ₆ GACATAC
19	GGGCGTT (AC) ₆	-	(GT) ₆ GGCGTAC
20	TGGCGCT (AC) ₆	-	(GT) ₁₇ GGCGTCC
21	GGACGCT (AC) ₆	(TA) ₄	(GT) ₆ GATGCCA
22	GGGTGCC (AC) ₆	(TA) ₄	(GT) ₆ GATGCCA
23	TGGGGGG (TC) ₆	-	(GA) ₆ CGCGTCA
24	GGGAGGG (TC) ₆	-	(GA) ₆ TTCCCCC
25	TTGGAGG (TC) ₇	-	(GA) ₆ TTCCCCA
26	GGAAGGG (TC) ₆	-	(GA) ₆ TTCGTCA
27	TGGGACA (TC) ₆	-	(GA) ₆ CTTCCCC
28	TGACGGG (TC) ₆	(GA) ₅	(GA) ₆ CGCTCCC
29	GGAAGGG (TC) ₆	-	(GA) ₆ CTCGCAC
30	TGGCGCG (TC) ₆	-	(GA) ₆ CCTGTCC
31	GGAGGGG (TC) ₆	-	(GA) ₆ TGCCCCA
32	GGTGTCT (TG) ₆	-	(CA) ₆ GTGGTAC
33	GGCGTCC (TG) ₆	-	(CA) ₆ ACGTCA
34	TGTCTGC (TG) ₇	-	(CA) ₆ AGAGGCA
35	TGTGTGT (TG) ₆	(TA) ₆	(CA) ₇ GCATTCC
36	TGTGTGT (TG) ₈	(AT) ₇	(CA) ₆ TCGGACC
37	TGAGTGT (TG) ₆	(AT) ₇	(CA) ₆ GCGCAAC
38	GGTCAGC (TG) ₇	(AG) ₆ (AG) ₈ (GC) ₄	(CA) ₆ ACGGTCC
39	GTAAGT (TG) ₆	(TA) ₅ (TC) ₅ (GT) ₉ (AG) ₅	(CA) ₉ TTAGGCC

^aThe previous and later specific bases of the 5' and 3' terminal repeats, respectively, were sequenced from the selected clones generated by amplification with a 5' anchored primer containing 5' degenerated bases.

3.2 Marker evaluation

In order to develop locus-specific microsatellite markers, one additional specific primer was required to amplify the terminal

microsatellites. The internal microsatellites were amplified by two specific primers as normal. From the total of 18 clones (Table 1) having more number of terminal repeats than that found

in the 5' anchored primers, 5 clones could not be used for designing specific primers due to the clones lacking a sufficient sequence suitable for primer design. Locus-specific primers were designed from 13 unique clones producing 16 microsatellite markers. A specific primer was used in combination with an anchored primer to determine if they amplified PCR products in the expected size range and if they were

polymorphic across 8 mango varieties. The total of 16 microsatellite markers revealed 46 alleles among the fragments amplified from the 8 mango varieties (Table 3). Fourteen polymorphic microsatellite markers produced an average of 2.8 alleles ranging from 2 to 6 alleles per locus with the average of 0.475 polymorphism information content (PIC) varied from 0.186 to 0.637.

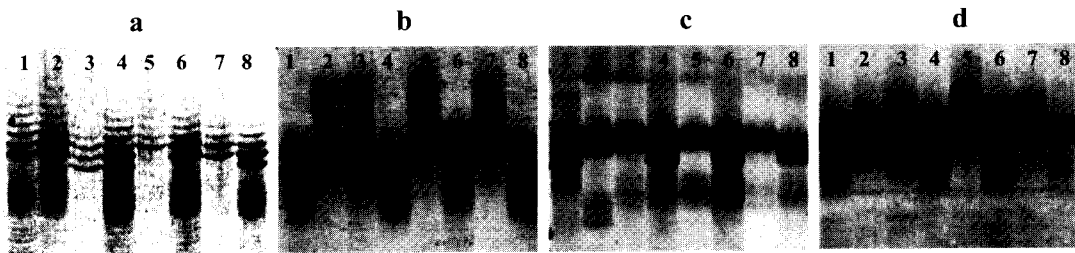


Fig. 2. Examples of polymorphisms among 8 mango varieties genotyped with the 5' anchored primer in combination with the specific primer; a) BDB(CA)₇ and AAGGTCATTGGGTTTCATTCG b) KKVHVHV(GT)₆ and AGAAATGGTTGGTGGTGGTC c) KKRYRYY(AC)₆ and TGGTATTCAAGCATGGTCCTC and d) CHY(GA)₇ and TGGTATTCAAGCATGGTCCTC. Lanes 1) Num Dok Mai – Si Thong Num, 2) Choke Anun, 3) Mae Look Duok, 4) Bun Ban Dan, 5) Fha Lan, 6) Dok Mai – Ta Why, 7) Mon Dueun kao and 8) Kiao Sway.

Table 3 Sixteen locus-specific microsatellite markers developed in this experiment, sequences of primer pairs, the expected product size, the number of alleles of each marker and polymorphism information content (PIC)

Name of Marker	Sequences (5'-3')		Product size (bp)	Number of Alleles	PIC
	Forward	Reverse			
MITG175	TGCGTCTTGTGTGTGTGTGT	GGAATGCTGTGTGTGTGTG	175	3	0.422
MITG436-1	AGAAATGGTTGGTGGTGGTC	ACCGTGTGTGTGTGTGTGC	167	3	0.632
MITG436-2	GGTCAGCTGTGTGTGTGTG	CAATTCAATGCTTTGGATGCT	158	6	0.602
MITG962	TGTTTCGATTGCAAACCTTTT	GGCCTAATGTGTGTGTGTG	299	2	0.422
MICA231-1	TGGAAGGACCATGCTTGAAT	GGTCACACACACACACACA	161	2	0.536
MICA231-2	CGGCACACACACACACA	AAGGTCATTGGGTTTCATTCG	150	3	0.558
MICA235	TGTCACACACACACACACA	AATGGAAGGACCATGCTTGA	163	2	0.408
MIGA179	CCTGAGAGAGAGAGAGAGA	GAGAGAGAGAGAGAGGTGG	176	3	0.497
MIGA203	TGAAGGATAGGTGTGGTG	CATGAGAGAGAGAGAGAGA	158	2	0.186
MIGA224	CACGAGAGAGAGAGAGAGA	GGGTCTCAGAGGGAGGATT	187	2	0.437
MIGA253	CATGAGAGAGAGAGAGAGA	AAAGGAAAGGCAGGGAAATG	153	3	0.422
MIGA326	GACAGACAAAGCCAGCAGAA	CCCGAGAGAGAGAGAGAGA	297	2	0.422
MIAC251-1	CCTTGGGTTCATTTCGCTAAA	GGACGCCACACACACACAC	165	3	0.637
MIAC251-2	TGGCGCTACACACACACAC	CACACACACACACACACG	229	3	0.552
MIAC326	TGGTATTCAAGCATGGTCCTC	TGGCATCACACACACACAC	244	5	0.615
MITC138	TCTCCCTTCATCGATTGTCC	GGAGCGTCTCTCTCTCCA	122	2	0.258

In conclusion, 5' anchored PCR was applicable for development of microsatellite markers in mango. This methodology amplifies genomic fragments containing microsatellite sequences. However anchoring of the 5'

anchored primer may not be exclusive at the 5' end. Most microsatellite sequences depend on the number of 5' degenerated bases. In this study, 7 degenerated bases of 5' anchored primers reduced the efficiency of 5' anchored

PCR while 3 degenerate bases of the 5' anchored primer showed superior results. Microsatellite markers developed from this experiment offer potential use of the markers for varietal identification and genome mapping.

4. Acknowledgments

Financial support from the Faculty of Science and Technology, Thammasat University is gratefully acknowledged. I also thank Ganlayarat Posudsavang, Kanchana Punyawaew, Pornrung Phupa and Suchonma Udomlertpreecha for their technical assistance.

5. References

- [1] FAOSTAT, FAO Statistics Database on the World Wide Web. <http://apps.fao.org>, 2004.
- [2] Wünsch A. and Hormaza J.I., Cultivar Eight Identification and Genetic Fingerprinting of Temperate Fruit Tree Species Using DNA Markers, *Euphytica*, Vol. 125, pp. 59–67, 2002.
- [3] Morgante M., Olivieri A.M., PCR Amplified Microsatellites as Markers in Plant Genetics, *Plant J.*, Vol. 3, pp. 175–182, 1993.
- [4] Queller D.C., Strassmann J.E., Hughes C.R., Microsatellites and Kinship, *Trends Ecol. Evol.*, Vol. 8, pp. 285–288, 1993.
- [5] Powell W., Morgante M., McDevitt R., Vendramin G.G., Rafalski J.A., Polymorphic Simple Sequence Repeat Regions in Chloroplast Genomes: Applications to the Population Genetics of Pines, *Proc. Natl. Acad. Sci. USA*, Vol. 92, pp. 7759–7763, 1995.
- [6] Kashi Y., King D., Soller M., Simple-Sequence Repeats as a Source of Quantitative Variation, *Trends Genet.*, Vol. 13, pp. 74–78, 1997.
- [7] Cregan P.B., Jarvik T., Bush A.L., Shoemaker R.C., Lark K.G., Kahler A.L., Kaya N., Van T.T., Lohnes D.G., Chung J. and Specht J.E., An Integrated Genetic Linkage Map of the Soybean, *Crop Sci.*, Vol. 39, pp. 1464–1490, 1999.
- [8] Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal M.W., A Microsatellite Map of Wheat, *Genetics*, Vol. 141, pp. 2007–2023, 1998.
- [9] Senior M.L., Chin E.C.L., Lee M., Smith J.S.C. and Stuber C.W., Simple Sequence Repeat Markers Developed from Maize Sequences Found in the GenBank Database: Map Construction, *Crop Sci.*, Vol. 36, pp. 1676–1683, 1996.
- [10] Hayden MJ and Sharp P.J., Targeted Development of Informative Microsatellite (SSR) Markers, *Nucleic Acids Res.*, Vol. 29, e44, 2001.
- [11] Fisher P.J., Gardner R.C. and Richardson T.E., Single Locus Microsatellites Isolated using 5' Anchored PCR, *Nucleic Acids Res.*, Vol. 24, pp. 4369–4371, 1996.
- [12] Hayden M.J. and Sharp P.J., SequenceTagged Microsatellite Profiling (STMP): A Rapid Technique for Developing SSR Markers, *Nucleic Acids Res.*, Vol. 29, pp. e43, 2001.
- [13] Zietkiewicz E., Rafalski A. and Labuda D., Genome Fingerprinting by Simple Sequence Repeat (SSR)-Anchored Polymerase Chain Reaction Amplification, *Genomes*, Vol. 20, pp. 176–183, 1994.
- [14] Gawel N.L. and Jarret R.L., A Modified CTAB DNA Extraction Procedure of *Musa* and *Ipomoea*, *Plant Mol. Bio. Rpt.*, Vol. 9, pp. 262–266, 1991.
- [15] Rozen S. and Skaletsky H.J., Primer3 Available from http://www-genome.wi.mit.edu/genome_software/other/primer3.html, 1998.
- [16] Botstein D., White R.L., Skolnick M.H. and Davies R.W., Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms, *American Journal of Human Genetics*, Vol. 32, pp. 314–31, 1980.
- [17] Caldeira R.L., Carvalho O.S., Lage R.C., Cardoso P.C.M. and Oliveira G.C., Sequencing of Simple Sequence Repeat Anchored Polymerase Chain Reaction Amplification Products of *Biomphalaria Glabrata*, *Mem. Inst. Oswaldo Cruz*, Vol. 97, No. 1, pp. 23–26, 2002.