Arctoa (2014) **23**: *145-149*
doi: 10.15298/arctoa.23.12

SIMPLE SEQUENCE REPEATS MINING USING COMPUTATIONAL APPROACH IN CHLOROPLAST GENOME OF *MARCHANTIA POLYMORPHA*

ПОИСК ПРОСТЫХ НУКЛЕОТИДНЫХ ПОВТОРОВ В ХЛОРОПЛАСТНОМ ГЕНОМЕ *MARCHANTIA POLYMORPHA* С ПОМОЩЬЮ КОМПЬЮТЕРНОЙ ПРОГРАММЫ

ASHEESH SHANKER[1]

АШЕЕШ ШАНКЕР[1]

Abstract

Simple sequence repeats (SSRs) or microsatellites are found in DNA sequences and consist of short repeat motifs of 1-6 bp. These repeats play important role in the development of molecular markers, phylogenetics, population genetics and evolutionary biology. The present analysis was conducted to detect chloroplastic SSRs (cpSSRs) in *Marchantia polymorpha*. The chloroplast genome sequence of *M. polymorpha* was downloaded from the National Center for Biotechnology Information (NCBI) and mined with the help of a Perl script named MISA. A total of 66 cpSSRs were detected in 121.024 kb sequence mined showing density of 1 SSR/1.83 kb. Depending on the repeat units, the length of SSRs ranged from 12 to 17 bp for mono, 12 to 64 bp for di, 12 to 21 bp for tri, 12 to 24 bp for tetra, 15 bp for penta and 18 bp for hexa nucleotide repeats. Mononucleotide repeats were the most frequent repeat type (42.42%) followed by dinucleotide (25.76%) and tetranucleotide (21.21%) repeats. PCR primers were successfully designed for 45 (68.18%) cpSSRs of *M. polymorpha*.

Резюме

Простые нуклеотидные повторы (SSRs), или микросателлиты, встречающиеся в последовательностях ДНК, состоят из коротких повторов из 1-6 нуклеотидов. Они важны для разработки молекулярных маркеров для филогенетических, популяционно-генетических и эволюционных исследований. Представленный анализ проведен с целью определения хлоропластных SSR *Marchantia polymorpha*. Использован хлоропластный геном *M. polymorpha* из ГенБанка и в нем проведен поиск с помощью оригинального скрипта MISA на языке Perl. Выявлено 66 SSR маркеров в последовательности длиной 121024 bp; таким образом, имеется 1 SSR повтор на 1.83 kbp. Длина SSR повторов варьирует от 12 до 17 bp для моно-, 12 до 64 bp для ди-, 12 до 21 bp для три-, 12 до 24 bp для тетра-, 15 bp для пента- и 18 bp для гекса-нуклеотидных повторов. Мононуклеотидные повторы наиболее часты (42.42%), реже встречаются ди- (25.76%) и тетра- нуклеотидные повторы (21.21%). Праймеры успешно разработаны для 45 (68.18%) SSR маркеров.

KEYWORDS: Bryophytes, Marchantia, Chloroplast, Microsatellites, Simple Sequence Repeats

INTRODUCTION

Bryophytes are broadly classified into liverworts, mosses and hornworts, and are considered as the earliest land plants. Chloroplast and mitochondrial genome sequences based phylogenetic analysis showed liverworts as the earliest diverging lineage and hornworts as sister group to vascular plants (Shanker, 2013a,b,c).

Among land plants the first complete chloroplast genome sequence determined was that of *Marchantia polymorpha* (Ohyama *et al.,* 1986; Ohyama *et al.,* 1988). Since then several studies to elucidate the structure and organization of chloroplast genomes have been conducted. However, a small number of organelle genome sequences of bryophytes are available (Shanker, 2012a,b).

Simple sequence repeats (SSRs) or microsatellites are found in DNA sequences and consist of short repeat motifs of 1-6 bp (Shanker *et al.,* 2007b). These repeats are present in both coding and non-coding regions of DNA sequences, and have been widely used as molecular markers in plant genomes (Gupta *et al.,* 2003; Jakobsson *et al.,* 2007; Blair & Hurtado, 2013). Studies were conducted to mine SSRs in chloroplast genomes of bryophytes (Shanker, 2013d,e). However, mining of other available organelle genome sequences of bryophytes will help to know the occurrence of SSRs in them.

Biotechnological methods of SSR identification are costly and consume time. Contrary to this bioinformatic approaches offer rapid and economical SSR extraction using sequences deposited in public databases (Shanker *et al.,* 2007a). Therefore the present analysis was conducted to identify cpSSRs in *M. polymorpha*. Moreover, the distribution of these repeats in coding and non-coding regions of chloroplast genome was analyzed.

[1] – Department of Bioscience and Biotechnology, Banasthali University, Banasthali-304022, Rajasthan, India; e-mail: ashomics@gmail.com

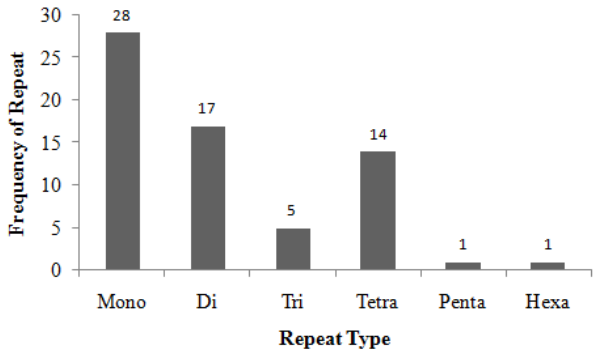

Fig. 1. Frequency distribution of mono-hexa repeats.

MATERIALS AND METHODS

Chloroplast genome sequence of M. polymorpha. The complete chloroplast genome sequence of *M. polymorpha* (NC_001319, 121024 bp; Ohyama *et al.,* 1986) was downloaded from NCBI (www.ncbi.nlm.nih.gov) in FASTA and GenBank format.

Identification of chloroplastic simple sequence repeats. A Perl script named MISA (available at http://pgrc.ipkgatersleben.de/misa/misa) was used to identify SSRs in chloroplast genome sequence of *M. polymorpha*. MISA takes FASTA formatted DNA sequence file as an input and generates information of perfect and compound SSRs, if found. A perfect SSR consist of same repeating motif without interruptions, e.g., $(GTG)_8$. In compound SSR, two or more SSRs are found adjacent to one another, e.g., $(GTG)_8(AT)_{16}$ (Bachmann & Bare, 2004). The length of SSRs in this study was defined as ≥12 for mono, di, tri and tetranucleotide, ≥15 for pentanucleotide and ≥18 for hexanucleotide repeats. Maximum difference between two compound SSRs was taken as 0. Since the information of coding and non-coding regions of chloroplast genome has been available in GenBank file, therefore, the mined cpSSRs were classified as coding, non-coding and coding-non-coding (occurrence of few bases of SSR in coding as well as in non-coding regions or vice-versa) SSRs based on the presence of these repeats in respective regions.

Primer designing for identified SSRs. PCR primers were designed for identified SSRs considering 200 base pair of SSR flanking regions using Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/) with default parameters of GC content, melting temperature, primer and PCR product size.

RESULTS AND DISCUSSION

In this study SSRs were identified in chloroplast genome sequence of *M. polymorpha* considering a minimum length of 12 bp. A total of 66 perfect SSRs were detected with length variation from 12 to 64 bp. Compound SSRs were found to be totally absent in chloroplast genome sequence of *M. polymorpha*. The frequency of identified SSRs (mono-hexa) is presented in Fig. 1. Detail information of mined SSRs motif, their length, start-end position and the region in which they lie is presented in Table 1. It is evident from this table that out of total cpSSRs detected, 14 (21.21%) found in coding, 48 (72.73%) in non-coding and only 4 (6.06%) in coding-non-coding regions. The occurrence of these repeats in various regions attributed to the evolutionary processes that fine tune distribution of SSR repeat types in genome (Lin & Kussell, 2012).

Mononucleotides were found to be most frequent repeat (28, 42.42%) followed by dinucleotide (17, 25.76%), tetranucleotide (14, 21.21%) and trinucleotide (5, 7.58%) repeats. Penta and hexanucleotide repeats were found with equal frequencies (1, 1.52%). An attempt was made to design PCR primers for all mined SSRs, however, these were successfully designed for 45 (68.18%) cpSSRs of *M. polymorpha.* A list of designed PCR primers, their length, product size etc. is presented in Table 2.

The mined SSRs represent a density of 1 SSR/1.83 kb in 121.024 kb sequence mined. The density of cpSSRs in *M. polymorpha* found to be higher than the density of cpSSRs in *Anthoceros formosae* (1 SSR/2.4 kb; Shanker, 2013d), rice (1 SSR/6.5 kb; Rajendrakumar *et al.,* 2007), EST-SSRs in barley, maize, wheat, rye, sorghum and rice (1 SSR/6.0 kb; Varshney *et al.,* 2002), cotton and poplar (1 SSR/20 kb and 1 SSR/14 kb respectively; Cardle *et al.,*, 2000), Unigenes sequences of *Citrus* (1 SSR/12.9 kb; Shanker *et al.,* 2007a). However, the density of cpSSRs in *M. polymorpha* found to be lower than the cpSSRs density in family Solanaceae (1 SSR/1.26kb; Tambarussi *et al.,* 2009). The selection of SSR detection tools, parameters taken (e.g. minimum length of SSRs) and amount of data analyzed might be the cause of variations in SSR density.

The higher occurrence of mononucleotides in this study shows consistency with earlier studies of cpSSRs in *Anthoceros formosae* (Shanker, 2013d), rice (Rajendrakumar *et al.,* 2007), Solanaceae species (Tambarussi *et al.,* 2009), *Saccharum* spp. (Melotto-Passarin *et al.,* 2011) and *Olea* species (Filiz & Koc, 2012). Moreover the abundance of cpSSRs in non-coding regions is also in agreement with previous studies (Hancock, 1995; Shanker, 2013d).

As a concluding remark, *in silico* mining of complete chloroplast genome sequence of *M. polymorpha* saves time, cost and provides sufficient number of SSRs for this liverwort. The identified SSRs can be used to develop SSR markers and for various other purposes.

LITERATIRE CITED

BACHMANN, L. & P.T.J. BARE 2004. Allelic variation, fragment length analyses and population genetic model: A case study on *Drosophila* microsatellites. – *J. Zool. Syst. Evol. Research* **42**: *215-222*.

BLAIR, M.W. & N. HURTADO 2013. EST-SSR markers from five sequenced cDNA libraries of common bean (*Phaseolus vulgaris* L.) comparing three bioinformatic algorithms. – *Mol. Ecol. Resour*. **13**: *688-695*.

CARDLE, L., L. RAMSAY, D. MILBOURNE, M. MACAULAY, D. MARSHALL & R. WAUGH 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. – *Genetics* **156**: *847-854*.

No.	MOTIF	LENGTH	START	END	REGION
1	(ATAAA)3	15	450	464	Non coding
2	(AT)7	14	2290	2303	Non coding
3	(TAA)4	12	2510	2521	Non coding
4	(TA)32	64	3589	3652	Non coding
5	(AAC)7	21	4559	4579	Non coding
6	(AT)6	12	5318	5329	Non coding
7	(TA)17	34	5379	5412	Non coding
8	(T)12	12	5491	5502	Non coding
9	(A)14	14	7640	7653	Coding
10	(T)14	14	9981	9994	Non coding
11	(A)15	15	13493	13507	Coding
12	(A)12	12	14664	14675	Coding
13	(A)13	13	15824	15836	Coding
14	(T)12	12	22002	22013	Non coding
15	(AT)24	48	22693	22740	Non coding
16	(AAT)6	18	23325	23342	Non coding
17	(A)13	13	23923	23935	Non coding
18	(T)13	13	26530	26542	Non coding
19	(AAAC)3	12	26632	26643	Non coding
20	(AT)7	14	26664	26677	Non coding
21	(T)12	12	26745	26756	Non coding
22	(A)15	15	30115	30129	Coding
23	(A)13	13	33266	33278	Coding
24	(ATTA)3	12	36685	36696	Non coding
25	(TA)8	16	38168	38183	Non coding
26	(A)13	13	38339	38351	Non coding
27	(TTAA)6	24	41278	41301	Coding-Non-Coding
28	(A)12	12	41323	41334	Non coding
29	(TA)13	26	41988	42013	Non coding
30	(A)16	16	50459	50474	Non coding
31	(TAA)4	12	50760	50771	Non coding
32	(AT)6	12	53692	53703	Non coding
33	(ATTA)3	12	53937	53948	Non coding
34	(TA)8	16	56040	56055	Non coding
35	(AT)16	32	56076	56107	Non coding
36	(A)12	12	58041	58052	Non coding
37	(TTA)4	12	59082	59093	Non coding
38	(A)13	13	60479	60491	Coding
39	(TTTG)3	12	63125	63136	Coding
40	(T)12	12	63943	63954	Non coding
41	(TA)6	12	64048	64059	Non coding
42	(A)16	16	64355	64370	Coding-Non-Coding
43	(TTATAT)3	18	64657	64674	Non coding
44	(AT)6	12	64681	64692	Non coding
45	(T)13	13	64817	64829	Non coding
46	(TA)6	12	64853	64864	Non coding
47	(TA)7	14	67473	67486	Non coding
48	(TATT)4	16	68702	68717	Non coding
49	(TAAA)3	12	70571	70582	Non coding
50	(T)13	13	72980	72992	Non coding
51	(T)12	12	75959	75970	Coding
52	(TTTA)3	12	76223	76234	Non coding
53	(T)15	15	76621	76635	Coding-Non-Coding
54	(T)13	13	77537	77549	Non coding
55	(TTTA)3	12	79756	79767	Non coding
56	(T)13	13	79811	79823	Non coding
57	(T)17	17	80642	80658	Coding
58	(AGGT)3	12	87572	87583	Coding
59	(TTTA)3	12	95400	95411	Non coding
60	(AATA)3	12	96864	96875	Coding
61	(AT)6	12	98203	98214	Non coding
62	(AT)7	14	98579	98592	Non coding
63	(A)16	16	99738	99753	Non coding
64	(AATT)3	12	99773	99784	Coding-Non-Coding
65	(T)14	14	106439	106452	Coding
66	(CTAC)3	12	114535	114546	Coding

Table 1: Information of mined SSRs in chloroplast genome sequence of *M. polymorpha.*

FILIZ, E & I. KOC 2012. *In silico* chloroplast SSRs mining of *Olea* species. – *Biodiversitas* **13**: *114-117*.

GUPTA, P.K., S. RUSTGI, S. SHARMA, R. SINGH, N. KUMAR & H.S. BALYAN 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. – *Mol. Genet. Genomics* **270**: *315-323*.

HANCOCK, J.M. 1995. The contribution of slippage-like processes to genome evolution. – *J. Mol. Evol*. **41**: *1038-1047*.

JAKOBSSON, M., T. SALL, C. LIND-HALLDEN & C. HALLDEN 2007. Evolution of chloroplast mononucleotide microsatellites in *Arabidopsis thaliana*. – *Theor. Appl. Genet.* **114**: *223-235*.

LIN, W.H. & E. KUSSELL 2012. Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. – *Nucleic Acids Res.* **40**: *2399-2413*.

MELOTTO-PASSARIN, D.M., E.V. TAMBARUSSI, K. DRESSANO, V.F. DE MARTIN & H. CARRER 2011. Characterization of chloroplast DNA microsatellites from *Saccharum* spp. and related species. – *Genet. Mol. Res*. **10**: *2024-2033*.

OHYAMA, K., H. FUKUZAWA, T. KOHCHI, H. SHIRAI, T. SANO, S. SANO, K. UMESONO, Y. SHIKI, M. TAKEUCHI, Z. CHANG, S. AOTA, H. INOKUCHI & H. OZEKI 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. – *Nature* **322**: *572-574*.

OHYAMA, K., H. FUKUZAWA, T. KOHCHI, T. SANO, S. SANO, H. SHIRAI, K. UMESONO, Y. SHIKI, M. TAKEUCHI, Z. CHANG, H. INOKUCHI & H. OZEKI 1988. Structure and organization of *Marchantia polymorpha* chloroplast genome. I. Cloning and gene identification. – *J. Mol. Biol*. **203**: *281-298*.

RAJENDRAKUMAR, P., A.K. BISWAL, S.M. BALACHANDRAN, K. SRINIVASARAO & R.M. SUNDARAM 2007. Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. – *Bioinformatics* **23**: *1-4*.

SHANKER, A. 2012a. Chloroplast genomes of bryophytes: a review. – *Archive for Bryology* **143**: *1-5*.

SHANKER, A. 2012b. Sequenced mitochondrial genomes of bryophytes. – *Archive for Bryology* **146**: *1-6*.

SHANKER, A. 2013a. Paraphyly of bryophytes inferred using chloroplast sequences. – *Archive for Bryology* **163**: *1-5*.

SHANKER, A. 2013b. Inference of bryophytes paraphyly using mitochondrial genomes. – *Archive for Bryology* **165**: *1-5*.

SHANKER, A. 2013c. Combined data from chloroplast and mitochondrial genome sequences showed paraphyly of bryophytes. – *Archive for Bryology* **171**: *1-9*.

SHANKER, A. 2013d. Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. – *Archive for Bryology* **191**: *1-6*.

SHANKER, A. 2013e. Mining of simple sequence repeats in chloroplast genome of a parasitic liverwort: *Aneura mirabilis*. – *Archive for Bryology* **196:** *1-4*.

SHANKER, A., A. BHARGAV, R. BAJPAI, S. SINGH, S. SRIVASTAVA & V. SHARMA 2007a. Bioinformatically mined simple sequence repeats in UniGene of *Citrus sinensis*. – *Sci. Hort.* **113**: *353-361*.

SHANKER, A., A. SINGH & V. SHARMA 2007b. *In silico* mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites. – *Microbiol. Res.* **162***: 250-256*.

TAMBARUSSI, E.V., D.M. MELOTTO-PASSARIN, S.G. GONZALEZ, J.B. BRIGATI, F.A. DE JESUS, A.L. BARBOSA, K. DRESSANO & H. CARRER 2009. *In silico* analysis of simple sequence repeats from chloroplast genomes of Solanaceae species. – *Crop Breed. Appl. Biotech.* **9**: *344-352*.

VARSHNEY, R.K., T. THIEL, N. STEIN, P. LANGRIDGE & A. GRANER 2002. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. – *Cell & Mol. Biol. Lett*. **7**: *537-546*.

Table 2: List of primers designed for identified SSRs along with primer length, melting temperature and product size.

No.	MOTIF	START	END	LEFT / RIGHT PRIMER	PRIMER LENGTH	Tm	GC%	PRODUCT SIZE (bp)
1	(AT)7	2290	2303	ATGAAGGAGTGCGATTCGTT	20	57.613	45	152
				AAAAATTAATTGGTGAAATTTTGCGTT	27	57.205	22.222	
2	(TAA)4	2510	2521	AACGCAAAATTTCACCAATTAATTTTT	27	57.205	22.222	261
				TCGAAGAATCCTTCTATTTCAATGGT	26	58.661	34.615	
3	(TA)17	5379	5412	GCCCAGGCTATATTAATAATATCCATT	27	57.186	33.333	300
				AGACAACCCATCATAAGATTAAAACA	26	57.122	30.769	
4	(A)14	7640	7653	ACTGTCTCATCGCATGGAGG	20	59.537	55	196
				AGCAGCTCCGTCTGCTAAAA	20	59.678	50	
5	(T)14	9981	9994	TCGAAATAGAGCAAAAACCTAAAGGA	26	59.062	34.615	250
				CGTCATCCGGCTCAAATAGT	20	57.775	50	
6	(AT)24	22693	22740	TGACCTCAATTTCAGATAGTCAAATT	26	57.131	30.769	393
				CACTCGGCCATCTCTCCAAA	20	59.75	55	
7	(AAT)6	23325	23342	TGCAATTGATTACTACTTTGCCA	23	57.026	34.783	220
				GCCTTTGTATGGCAAGCTTCA	21	59.454	47.619	
8	(A)13	23923	23935	CGAATCCTTCCGTCCCAGAC	20	60.179	60	231
				GCAGGACCAGCATACATCCA	20	59.819	55	
9	(T)13	26530	26542	TGTGTCTAGAAATAACCAGCATGC	24	59.364	41.667	192
				TGTTTGTTTGTTTTTCAAAAGTAGGCA	27	59.773	29.63	
10	(AAAC)3	26632	26643	AGAAATAACCAGCATGCTGATTAAT	25	57.407	32	250
				TGCAATTCCATAAAAATTTTCCTGAA	26	57.085	26.923	
11	(AT)7	26664	26677	AAAGTTTTGCCTACTTTTGAAAAACA	26	57.615	26.923	100
				TGCAATTCCATAAAAATTTTCCTGAA	26	57.085	26.923	
12	(ATTA)3	36685	36696	CCAGCGGATTTACAGTCCGT	20	60.108	55	163
				TCTCTTTCAAGGAGGCGACG	20	59.756	55	
13	(TA)8	38168	38183	TCCAATAGGTTATCAATCTTTCCGA	25	57.868	36	387
				CACTGAGTTAAAAGGGCAATAAGA	24	57.182	37.5	
14	(TTAA)6	41278	41301	CTGCTGCAGCTGGTTTTGAA	20	59.614	50	223
				TCGAGGGTTCAAATCCCTCTC	21	59.168	52.381	
15	(A)12	41323	41334	CTGCTGCAGCTGGTTTTGAA	20	59.614	50	223
				TCGAGGGTTCAAATCCCTCTC	21	59.168	52.381	
16	(A)16	50459	50474	GCATTACAAGTGCGACGCTC	20	60.247	55	249
				CCAATTTCGCCATACCCCCT	20	60.106	55	
17	(TAA)4	50760	50771	ACAAAAGAGGTGCAGAGACTCA	22	59.564	45.455	248
				CTGGGGGTAGAGGGACTTGA	20	59.957	60	
18	(AT)6	53692	53703	TACCGCTGAGCTATAGCCCT	20	59.887	55	216
				CTCTCGCCGTATGAAAGCGA	20	60.248	55	
19	(ATTA)3	53937	53948	TCGCTTTCATACGGCGAGAG	20	60.248	55	176
				AGAGCTAAAGCAAGATTAGAAGCA	24	58.261	37.5	
20	(A)12	58041	58052	TGGGTTTGTAGCTCAGTGGA	20	58.57	50	247
				AGCGCCAATTAATCCACCAA	20	58.154	45	
21	(TTA)4	59082	59093	TGAATTATATAACGCTGCTCCTTGT	25	58.712	36	292
				GCAGGAAAAATTAATCCAACTAAAGGA	27	58.689	33.333	
22	(A)13	60479	60491	TCAAAACGTTCTTGGCAGTCT	21	58.361	42.857	235
				AGCTCTAATCCAAGCTAATTTTCTGT	26	58.885	34.615	
23	(TTTG)3	63125	63136	ACCCCAATAGAGGCTAGTACGA	22	59.824	50	175
				ACCGTAAGATGGTTAGCCGT	20	58.811	50	
24	(TA)6	64048	64059	TTACAGATTGGTCTATCCCGAAA	23	57.057	39.13	345
				ACGCTAATGTTAAAGCACCGA	21	58.309	42.857	
25	(A)16	64355	64370	TCGGTGCTTTAACATTAGCGT	21	58.309	42.857	209
				TCCAGACAACAAAGCTTCAACC	22	59.312	45.455	
26	(TTATAT)3	64657	64674	TGGAGACCTACGTTCTACCGA	21	59.721	52.381	183
				TGGTAGCGCGTTTGTTTTGG	20	59.97	50	
27	(AT)6	64681	64692	TGGAGACCTACGTTCTACCGA	21	59.721	52.381	183
				TGGTAGCGCGTTTGTTTTGG	20	59.97	50	
28	(TA)6	64853	64864	CCAAACTGCGCTACATCCCT	20	60.393	55	297
				ACAGGTGCAGTAGAAAGATATGT	23	57.316	39.13	
29	(TATT)4	68702	68717	ACTTTCGGAACACCAATAGGCA	22	60.225	45.455	233
				ACAATGGAGAGATTGGTCCCA	21	58.722	47.619	

Table 2 (cont.): List of primers designed for identified SSRs along with primer length, melting temperature and product size.

No.	MOTIF	START	END	LEFT / RIGHT PRIMER	PRIMER LENGTH	Tm	GC%	PRODUCT SIZE (bp)
30	(TAAA)3	70571	70582	GCTCAAGTGGAATTTGGAGCG	21	60.136	52.381	232
				TGATTCCTAAAGTACCTACCAACA	24	57.061	37.5	
31	(T)13	72980	72992	TGAATAGTAAAGACTTCGTTAAATCCA	27	57.068	29.63	280
				TCATCCGGCTCCAACATTAATGA	23	60.118	43.478	
32	(T)12	75959	75970	CCCATTCCACCTAAAACTTTTGGA	24	59.409	41.667	241
				AACAGTTCAAGTACCTGCTACT	22	57.042	40.909	
33	(TTTA)3	76223	76234	TCGCAATTGTATCATTCCCCA	21	57.711	42.857	192
				AGGAACTCGAGTTTTTGGTCCA	22	59.828	45.455	
34	(T)15	76621	76635	TGGCACTGCTTCTTTAACAACA	22	58.98	40.909	195
				GGCCCTCCCTAATCCATCCA	20	60.772	60	
35	(T)13	77537	77549	TGCGTTTTAATACTATAACAAAAACCC	27	57.377	29.63	199
				GCTATGCTTAGTGTGTGACTCG	22	59.145	50	
36	(T)13	79811	79823	TCATTTAGCATCAGAAAGGATGACT	25	58.343	36	277
				AGGAAAAATAGGCAAGTGAAAAAGGT	26	59.9	34.615	
37	(T)17	80642	80658	ACAGTATATCCTGTTGTCGTACCT	24	58.801	41.667	212
				TTATGAATCAAGTTAAGTACCCAGTAC	27	57.12	33.333	
38	(AGGT)3	87572	87583	CTTCCAGCCAATGTCCGAGT	20	60.036	55	153
				TTATCTCGCGCCCCTAGGTA	20	59.889	55	
39	(TTTA)3	95400	95411	AGCAGTGCTAAGGCTTCTCG	20	60.108	55	194
				ACGCTCTAAGGTTATAAATGGCA	23	57.969	39.13	
40	(AATA)3	96864	96875	TTTCTCGTGGTCCAGCATCC	20	60.036	55	227
				GCTTTGCCAGGTATGAGTGG	20	58.902	55	
41	(AT)6	98203	98214	ACGGGATTACTAAACCTGCAGA	22	59.164	45.455	226
				TGAGACTACTCGTAGTATGGGTCT	24	59.592	45.833	
42	(A)16	99738	99753	TCTGGTAATTTCATAAAGCTCACAA	25	57.006	32	165
				CCTAAAGGCAAAATAGAAGGGCA	23	58.977	43.478	
43	(AATT)3	99773	99784	TCTGGTAATTTCATAAAGCTCACAA	25	57.006	32	165
				CCTAAAGGCAAAATAGAAGGGCA	23	58.977	43.478	
44	(T)14	106439	106452	TGGAAAATCCCATCTGTTTGCT	22	58.491	40.909	248
				AGGATCTATGCGTGCTCGAC	20	59.688	55	
45	(CTAC)3	114535	114546	TTATCTCGCGCCCCTAGGTA	20	59.889	55	153
				CTTCCAGCCAATGTCCGAGT	20	60.036	55	