International Journal of Medicine and Pharmaceutical Sciences (IJMPS)
ISSN(P): 2250-0049; ISSN(E): 2321-0095
Vol. 5, Issue 3, Jun 2015, 25-32

TRANS STELLAR
• Journal Publications • Research Consultancy

IN SILICO ANNOTATION AND MINING OF SSR MARKERS FROM EXPRESSED SEQUENCE TAGS OF *WITHANIA SOMNIFERA* DUNAL

CHAURASIA PRATIBHA[1], DETROJA RAJESH[2] & BORA MADHUMATI[3]

[1,3]Department of Biotechnology, Genetics and Bioinformatics, N. V. Patel college of Pure and Applied Sciences, S. P. University, V.V. Nagar, Gujarat, India

[2]Department of Bioinformatics, RGITBT, Bharati Vidyapeeth, Katraj, Pune, Maharashtra, India

ABSTRACT

Withania somnifera commonly known as Ashwagandha is a member of the family *Solanaceae*. Microsatellites or simple sequence repeats (SSRs) present in Expressed Sequence Tags (ESTs) provide an opportunity for low cost SSR marker development. SSRs located in open reading frames (ORF) were analysed for functional annotation using Gene Ontology. Seven hundred and forty one EST sequences were mined, examined and assembled to get full-length sequences. Maximum frequency distribution was shown by mononucleotide SSR motifs i.e. 88.37% whereas, minimum frequency was observed for dinucleotide SSR (4.65%) where, AT/TA (55.55%) was the most frequent repeat. Maximum trinucleotide motifs code for lysine and leucine (40%). Flanking primer pairs were designed *insilico* for the SSR containing sequences. Functional annotation of SSRs in the sequences was characterized under headings like biological process, cellular component and molecular function. Thus *insilico* approaches provide an attractive and alternative way to conventional laboratory methods for rapid and economic development of SSR markers by utilizing freely available genomic sequences in public databases.

KEYWORDS: Medicinal Plant, *Withania somnifera*, Primers, Expressed Sequence Tag (EST), Molecular Markers, Simple Sequence Repeats (SSR)

INTRODUCTION

BACKGROUND

Exploring medicinal values of plant is the way to successful applications in molecular farming, health food, functional food, and plant resistance [1]. Ashwagandha (*Withania somnifera*) is a medicinal plant having major profitable use in ayurvedic medicine. Ashwagandha possesses antioxidant, antitumor, antistress, anti-inflammatory, immunomodulatory, hematopoetic, anti-ageing, anxiolytic, anti-depressive rejuvenating properties and also influences various neurotransmitter receptors in the central nervous system [2]. Roots of *Withania somnifera* are used for the treatment of asthma, bronchitis, edema, leucoderma, anorexia, consumption, asthenia, anaemia, exhaustion, aging, insomnia, neurasthenia, infertility, impotence, repeated miscarriage, paralysis, memory loss, multiple sclerosis, immune- dysfunction, carcinoma, rheumatism and arthritis [3-6] etc. The main chemical constituents of ashwagandha are alkaloids and steroidal lactones. These include isopelletierine, anaferine, and withanolides, withaferins respectively. Saponins containing an additional acyl group (sitoindoside VII and VIII), and withanolides with a glucose at carbon 27 (sitoindoside IX and X) are also reported. These metabolites are rich in medicinal properties like anti-inflammatory, immunomodulator, anti-tumour, nervine, mild sedative, and analgesic, reproductive tonic, aphrodisiac and anti anaemic.

Simple sequence repeat (SSR) or microsatellite markers detect differences in the length of mono to hexa-nucleotide repeat sequences [7]. SSR markers constitute a useful tool for genetic diversity analysis in that they enable multi-allele detection which are highly transferable across species and are flexible enough so that they can be used with various laboratory systems [8]. This method is very efficient and cost effective and provides higher polymorphism as compared to other marker system. The major frequencies of length polymorphism linked with microsatellites provide the foundation for development of a marker system that has widespread application in genetic research including studies of genetic variation, linkage mapping, gene tagging and evolution [9].SSRs are used widely as molecular markers because of their multiallelic nature, co-dominant inheritance and relative abundance. The major drawback of using SSRs as markers has been their time consuming process. However, with fast-paced boost of nucleic acid in recent years, it became realistic to screen for microsatellites in database for numerous plant species. Variations in SSR regions originate mostly from errors during the replication process and frequent DNA polymerase slippage. These errors create base pair insertions or deletions respectively in larger or smaller regions [10]. The use of EST or cDNA-based SSRs has been reported for several species including grape [11], sugarcane [12], durum wheat [13], rye [14], medicinal plant like basil [15] and periwinkle [16]. There are diverse SSR identification software's such as MISA [17], SSR Finder, SSRIT, TRF, TROLL and sputnik.

The current study is designed to investigate the genetic diversity among reported species of Withania populations and to explore the possibility of using EST-SSR markers for fingerprinting of cultivars. Different types of SSRs and their percentage distributions were scrutinized. The forward and reverse primer pairs were designed from the flanking ends of SSRs. The functional annotation of these SSR containing sequence was done. The annotation analyzes the possible function of ESTs and also detects functional domain markers linked to SSR-ESTs enabling investigation on gene ontology. The aim of present study is based on identification of i) frequency and distribution of SSRs in EST ii) functional annotation and prediction of amino acids from SSR loci and iii) development and validation of polymorphism of EST-SSR marker in ashwagandha.

MATERIALS AND METHODS

Sequence Data Source: Seven hundred and forty one EST sequences of *Withania somnifera* were identified in EST database at NCBI. These sequences were isolated from different plant tissues like leaves, stem, root, etc. There are redundant sequences in the EST, therefore we used CAP3 (Contigs Assembly Program) assembler for sequence assembly. (http://www.genome.clemson.edu/resources/online_tools/cap3).

Microsatellite Identification: After pre-processing, SSRs were detected using MIcroSAtellite identification tool (MISA) (http://pgrc.ipk-gatersleben.de/misa/misa.html) written in the Perl scripting language [18]. EST derived SSRs were considered to contain repeat motifs ranging in length from 1 to 6 bp. The minimum numbers of repeats were 10 for mononucleotides, 6 for dinucleotides and 5 for trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides. The analysis of SSRs was carried out on the basis of their types (mono to hexanucleotides), number of repeats, percentage frequency of occurrences of each SSR motif and their distribution in the sequence.

Gene Ontology Classifications: Classification of SSR-ESTs was performed using BLAST2GOsoftware [19]. Blast2GO allows automatic and high throughput sequence annotation and integrate functional information for annotation-based data mining. The ontology classification was performed to analyse biological process, molecular function and cellular component. These characterizations were based on scrutinized SSR repeats.

Marker Development: For the development of microsatellite markers, we designed primer pairs for all the identified microsatellites. The microsatellites (excluding monomers) were used for designing primers pairs. The primers were designed from the flanking sequences having microsatellite repeats using Primer3 software [20] (Whitehead Institute, USA). Forward and reverse primer pairs were designed for marker development. The optimum and maximum primer sizes were set to 18 and 24 nucleotides, respectively. The GC % was set to 50.0 to 60.0 and the Tm value between 55- 60^0C.

RESULTS AND DISCUSSIONS

ESTs are often represented by redundant cDNA sequences making them difficult to analyse effectively for SSRs. To eliminate the redundancy in sequences CAP3 program has been used. The resulting sequences were Contigs (62 sequence) and Singlets (562 sequence). MISA tool used for microsatellite scrutinizing in both contig (Table 1) and singlet sequences (Table 2). The study of occurrence of different types of SSR repeats revealed that percentage allocation of mononucleotide SSRs was 88.37%, dinucleotide SSRs 4.65% and trinucleotide 6.97% in Contigs (Figure 1). Similarly the occurrence of different types of SSR repeats in Singlets showed that percentage allocation of mononucleotide SSRs was 91.60%, dinucleotide SSRs 3.14 % and trinucleotide 5.24% (Figure 2).

Amino Acid Distribution: The triplet codon forms an open reading frame (ORF) translated to proteins. The trinucleotide SSRs are triplet codon that code for a particular amino acid. It was observed that all triplet codons of Contig sequences contain leucine, serine, glycine, threonine, isoleucine and asparagine. While in singleton sequences, asparagine and valine followed by leucine, serine, and threonine followed by methionine and lysine followed by glycine have been observed. Comparative analyses of different properties of observed amino acids were identified in both contigs and singlets (Figure 3).

Gene Ontology Classification: The widest use of ontologies within biology is for conceptual annotation- a representation of stored knowledge is computationally more amenable than natural language. Gene ontology based functional annotation of SSR- ESTs was performed through Blast2GO (http://www.blast2go.com). BLAST best hit were retained meeting the following criteria: E-value < 1e-5, and similarity >=70%. The most significant matches for the SSR-ESTs with unique SSR motif were considered. Based on Blast2GO analysis, putative functions could be assigned to 328 SSR loci. The GO software has developed three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. There are three separate aspects to this effort: first, the development and maintenance of the ontologies themselves; second, the annotation of gene products, which entails making associations between the ontologies and the genes and gene products in the collaborating databases; and third, the development of tools that facilitate the creation, maintenance and use of ontologies. (http://geneontology.org/)

The biological process is a collection of molecular events with a defined beginning and end. Mutant phenotypes often reflect disruptions in biological processes (http://geneontology.org/). These genes regulate all biological functions related to photosynthesis, cell signalling, stress, etc. In biological process corresponding to SSR-ESTs, the most frequent observed process in present studies was oxidation reduction followed by responses to cadmium ion, cold, metabolic process and red light etc (Figure 4).

The cellular component ontology is the parts of a cell or its extracellular environment. It describes locations, at the levels of sub cellular structures and macromolecular complexes. Examples of cellular components include 'nuclear inner

membrane' etc. Generally, a gene product is located in or is a subcomponent of a particular cellular component. The cellular component ontology includes multi-subunit enzymes and other protein complexes, but not individual proteins or nucleic acids.

(http://geneontology.org/). In the present context the most frequent is chloroplast envelope, apoplast, plasma membrane, chloroplast stroma, chloroplast thylakoids, cytosol etc (Figure 5).

Molecular function is the elemental activities of a gene product at the molecular level, such as binding or catalysis may include transporting things around, binding to things, holding things together and changing one thing into another. Here the most frequent is ATP binding, metal ion binding, protein binding, monooxygenate activity, copper ion binding and zinc ion binding etc. (Figure 6).

Primer Designing: A total of 215 (29%) EST sequences allowed for primer designing from 741 SSR containing sequences. The remaining SSR primers were unsuitable for primer construction. Around, 70 primer pairs (supplementary data 1) were successfully developed on the basis of following standard parameters by using Primer3software [21] (a) the target amplicon size of 100–500 bp, (b) the optimum annealing temperature 55-60^{o}C, (c) average GC content 50-60% and (d) the primer length 18-24 bp (Table 3). Putative functions of SSR loci were assigned by comparison with the non-redundant sequence database at NCBI using the BLASTX2.2.17 software [22][23].

CONCLUSIONS

Withania somnifera (L.) Dunal has attracted a big attention worldwide, because of its potential as a medicinal plant. However, very little genomic information has been known about this plant. Furthermore, Genomic SSRs have neither gene function nor close linkage to transcriptional regions, while EST-SSRs are potentially linked with functional genes that perhaps control certain important genetic characters [24]. NGS (Next Generation Sequencing) is increasingly being used for genomic and transcriptomic profiling of medicinal plants and may largely replace traditional fingerprinting techniques in future. NGS is proving an efficient tool in identification of SSR marker [25]. Among different classes of molecular markers, microsatellite or simple sequence repeat (SSR) markers are the most favoured for a variety of applications in plant genetics and breeding because of their multi-allelic nature, reproducibility, codominant inheritance, high abundance and extensive genome coverage [26].

In silico approaches have been used here to mine ever increasing EST sequences in public databases. The publicly available collections of 741 ESTs from *Withania somnifera* have been assembled and clustered using CAP3 assembly program. Assembly of EST sequences resulted in 624 non-redundant EST sequences which were reported to have 329 EST-SSRs. Among all the percentage frequencies, mono-nucleotide SSRs were maximum and dinucleotides minimum. Functional annotation of 328 SSR-EST was performed and 170 have significant matches. Finally SSR-ESTs were used in primer designing for *Withania somnifera* that can be applied in studies of genetic variation, linkage mapping and comparative genomics. The functional annotation of the SSR-ESTs showed that most of them are associated with expressed proteins and therefore, trait linked genes [27]. This study demonstrates the utility of computational approaches for mining SSRs from ever increasing repertoire of publicly available plant EST sequences present in different databases.

REFERENCES

1. Katare D. P.,Aeri V.and Bora, M (2009) Secondary metabolites and metabolic engineering J. Cell Tiss. Res. Vol. 9(3) 2027-2036.
2. Pattipati S, Amanpreet S, and Shrinivas K (2003). Effect of *Withania somnifera* root extract on Haloperidol induced Orofacial Dyskinesia: Possible mechanism of action. J. Med. Food. 6(2), 107-114.
3. Nadkarni K.M., & Nadkarni A.K. (1976). Dr. K.M. Nadkarni's Indian materia medica: With Ayurvedic, Unani-tibbi, Siddha, allopathic, homeopathic, naturopathic & home remedies, appendicces & indexes. Bombay: Popular Prakashan.
4. Frawley, David and Vasant Lad (1986). The yoga of Herbs: An Ayurvedic Guide to Herbal Medicine. Santa Fe:Lotus Press.
5. Dash and Bhagwan (1991). Materia Medica of Ayurveda. New Delhi: B.Jain Publishers.
6. Kirtikar K.R and Basu B.D (1993). Indian Medicinal plants. 2nd ed. Vol.1-4. Reprint. Delhi:Periodical Experts.
7. Kenta Shirasawa et al. (2013). Development of *Capsicum* EST–SSR markers for species identification and *in silico* mapping onto the tomato genome sequence. Mol Breed. Jan; 31(1): 101–110.
8. 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.117:309–334.
9. Kantety RV, La Rota M, Matthews DE, and Sorrells ME (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol. 48(5-6):501-10.
10. Victoria C F, Maia L and Oliveira A (2011). *In silico* comparative analysis of SSR markers in plants, BMC Plant Biology, 11:15.
11. Scott K.D, Eggler P. Seaton G, Rossetto M, Ablett E.M, Lee L.S and Henry R.J (2000). Analysis of SSRs derived from grape ESTs. Theor Appl Genet. 100:723–726.
12. Cordeiro G.M, Casu R, McIntyre C.L, Manners J.M, and Henry R.J (2001). Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to *Erianthus* and sorghum. Plant Sci. 160(6):1115-1123.
13. Eujayl I, Sorrells ME, Baum M, Wolters P and Powell W (2002). Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet.104 (2-3):399-407.
14. Hackbut B and Wehling P (2002). Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed. 121, 17-25.
15. Gupta S, Shukla R, Roy S, Sen N and Sharma A (2010). *In silico* SSR and FDM analysis through EST sequences in *Ocimum basilicum* POJ 3(4):121-128.
16. Joshi R K, Kar B and Nayak S (2011). Exploiting EST databases for the mining and characterization of short sequence repeat (SSR) markers in *Catharanthus roseus L*. Bioinformation. 5(9):378-381.
17. 17.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.)Theor Appl Genet., 106:411–422.

18. Thiel T (2002). http://pgrc.ipk-gatersleben.de/misa/misa.html.
19. 19.Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A,Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G (2000).Gene Ontology: tool for the unification of biology. Nat. Genet. 25(1):25-9.
20. Koressaar T,and Remm M (2007). Enhancements and modifications of primer design program Primer3 Bioinformatics 23(10):1289-91.
21. Rozen,S. and Skaletsky,H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol., 132:365–386.
22. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang ZH, Miller W, and Lipman DJ (1997). Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.
23. Gong, Y. M., S. C. Xu, W. H. Mao, Q. Z. Hu, G. W. Zhang, J. Ding, and Y. D. Li (2010). Developing new SSR markers from ESTs of pea (*Pisum sativum* L.). J. Zhejiang Univ. Sci. B 11, 702—707.
24. Shaohua Zeng, Gong Xiao, Juan Guo, Zhangjun Fei, Yanqin Xu, Bruce A Roe and Ying Wang (2010).Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. ET Zucc.) Maxim, *BMC* Genomics, 11:94.
25. 25.Kaur, S, N. O. Cogan, L. W. Pembleton, M. Shinozuka ,K.W. Savin, M. Materne ,and j. W. Forster (2011). Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genomics12:265-275.
26. Gupta PK and Varshney RK. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185.
27. Neeraj Jain, Ganesh B. Patil, Poonam Bhargava, Rajani and S. Nadgauda (2014). *In Silico* Mining of EST-SSRs in *Jatropha curcas* L. towards Assessing Genetic Polymorphism and Marker Development for Selection of High Oil Yielding Clones. Amer. J. Plant Sci., 5: 1521-1541.

APPENDICES

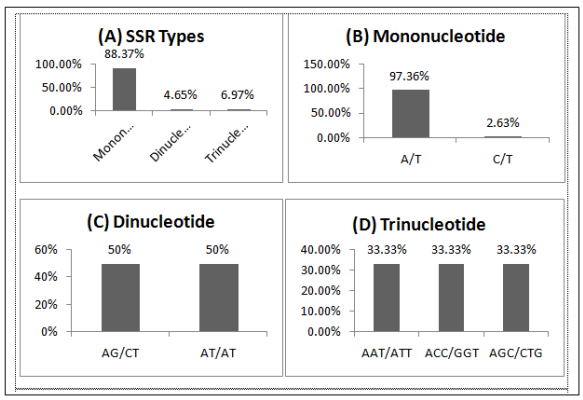

Figure 1: Contig Sequences (A) Distribution of Different SSRs; (B) Distribution of Mononucleotide SSRs; (C) Distribution of Dinucleotide SSRs; (D) Distribution of Trinucleotide SSRs

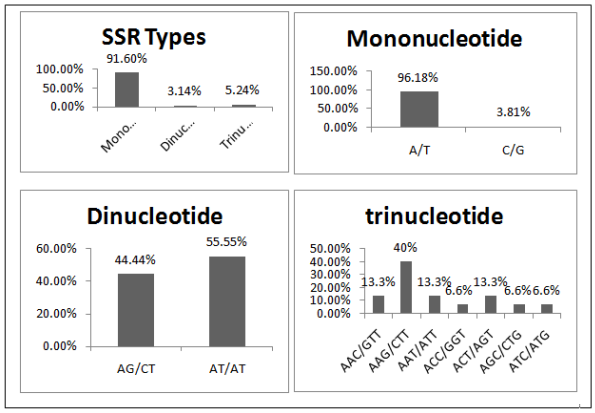

Figure 2: Singlet Sequences (A) Distribution of Different SSRs; (B) Distribution of Mononucleotide SSRs; (C) Distribution of Dinucleotide SSRs; (D) Distribution of Trinucleotide SSRs

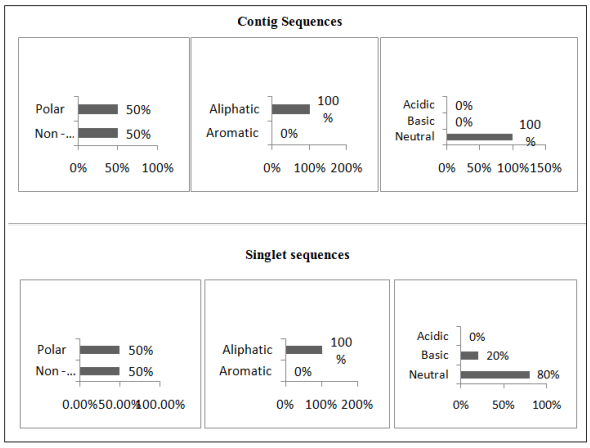

Figure 3: Percentage Frequency of Aminoacids Based on their Properties in Both Contig and Singlet Sequences

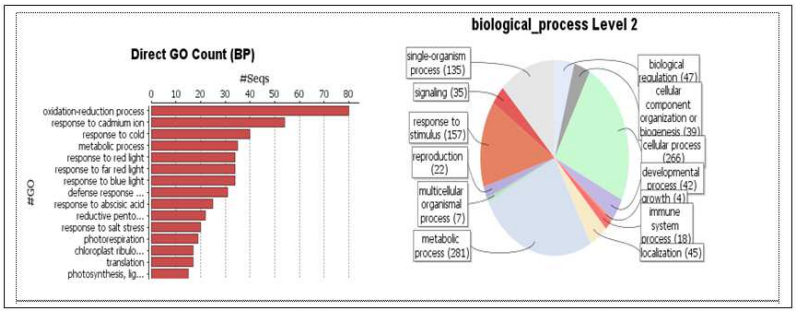

Figure 4: Gene Ontology Classification of EST Sequences Containing SSR's Based on Biological Process

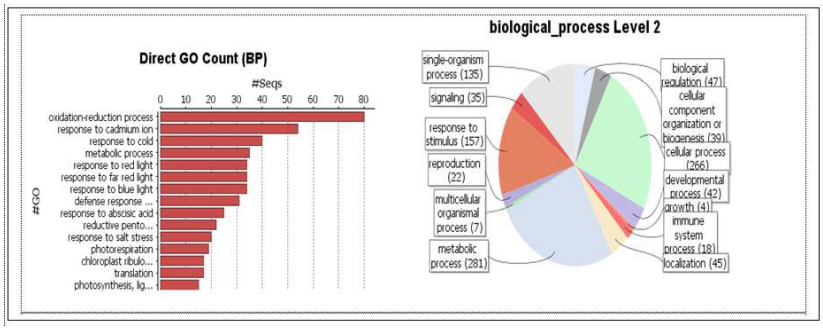

Figure 5: Gene Ontology Classification of EST Sequences Containing Ssrs Based on Cellular Component

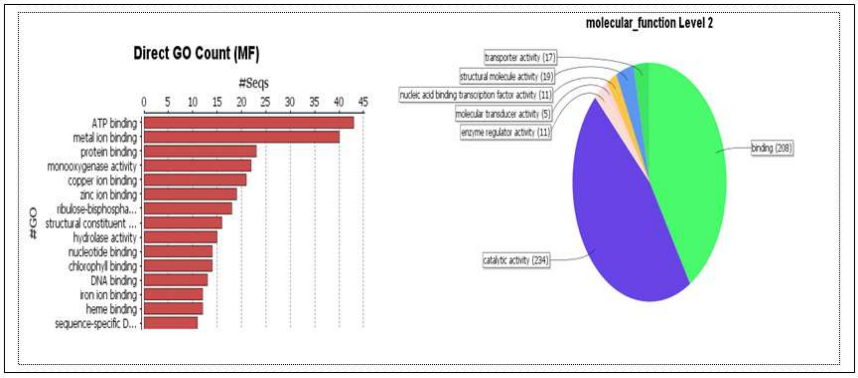

Figure 6: Gene Ontology Classification of EST Sequences Containing SSRs Based on Molecular Function

Table 1: Results of Microsatellite Search (for Contigs)

Total Number of Sequences Examined	**62**
Total size of examined sequences (bp)	47517
Total number of identified SSRs	43
Number of SSR containing sequences	34
Number of sequences containing more than 1 SSR	9
Number of SSRs present in compound formation	2

Table 2: Results of Microsatellite Search (for Singlets)

Total number of sequences examined	562
Total size of examined sequences (bp)	346576
Total number of identified SSRs	286
Number of SSR containing sequences	244
Number of sequences containing more than 1 SSR	35
Number of SSRs present in compound formation	27

Table 3: Primers Designed for EST Sequences Containing SSR

Sr.No.	ID	SSR Type	SSR Motif	TM(C%)	GC%	Forward Primer	Reverse Primer	Expected Product Size(bp)
WS17	gi|366844230	p2	(CT)4	59.98	45	TCCCAAAACCACACAGAACA	CCTCAACACCATTTTCAGCA	200
WS20	gi|366844164	p2	(CT)3	60.14	45	GGAATGGCAATCCATGAAAC	GGGCAGTAGGGACACTTGAA	213
WS07	gi|366844558	p2	(CG)3	59.93	50	CCTCTGGTCGATGGAACAAT	CGAAGGAACCGTTCTCAGTC	201
WS08	gi|366844550	p2	(TG)4	59.99	45	CTGAGCAAAATGGTGCAGAA	CCCAAATTGCGAAGACTTTC	181
WS09	gi|366844546	p2	(CA)3	60.01	60	GAAGCAGCCACTCCCTGTAG	TTTGCTGCTCCGAATTTTCT	207
WS10	gi|366844544	p2	(TG)3	59.87	50	AGAACATCAGTGGGCAGCTT	TGCTGCTTCAATTTCAGTGG	156
WS22	gi|366843902	p2	(GT)3	60.02	50	GCAAAAGTAGCCAAGCAAGG	TGGAGACGTCATGAAACCAA	150
WS19	gi|366844166	p2	(CA)3	59.96	50	CTTCGTGGTCCCATAAGCAT	TCCCCTTGTACGCAGGTAAC	196
WS01	gi|366878006	p2	(CT)3	60.12	55	GATGTGGGGCTGTTGTCTCT	CAGCGATTTCAGCAATTGAA	242
WS02	gi|366878000	p2	(AC)3	60.1	45	TAATAAGCCGTCATGCCACA	TCCAGTGGGAACATTCAACA	205
WS03	gi|366849188	p3	(ATC)3	60.05	55	TGTGCCAGTAGTGGAAGCAG	GGAAAAGGTGGATGTGGAGA	215
WS04	gi|366849186	p3	(AGC)3	60.05	40	TCAATGCTGGGAACATTCAA	CACCTTTTGCTGGTGCAGTA	235
WS06	gi|366844584	p3	(GAT)3	59.92	50	GGGACGTGCTATATCCGAAA	TGCTTCAATGGCTCGATATG	157
WS11	gi|366844538	p3	(AGC)3	59.98	55	CATCAGTGCCAGAGGACTCA	CGAATCTGAACGCATCTTGA	210
WS12	gi|366844536	p3	(CAT)3	59.99	55	ACAGGTCTTCAGGAGCTGGA	ATGGTTGCCCTGTTCTGTTC	183
WS13	gi|366844498	p3	(TGA)3	59.99	50	GCGGCATCACTTTCTTTAGC	TTTTGAAGCCCTCAACCATC	190
WS18	gi|366844228	p3	(AAG)6	60.28	50	TGGTGCTAACGTGGATGCTA	GTCGTCCTCCTTGTGTGCTT	154
WS21	gi|366844162	p3	(TTG)3	60	50	AAGTCCGGTGATGTTTGTG	GGAAGTTGCCAAGAAAGCTG	250
WS05	gi|366849110	p4	(GAAA)3	60.12	45	GTCCACTTTGCGTCGAATTT	GGGGCGTAGTTCATCCATAA	162
WS14	gi|366844556	p4	(AGAA)3	59.76	50	CACCATCATGGATCACAAGG	AGAGAAGGCTTCCCCTTCAG	173
WS35	gi|366849142	p4	(ATTT)3	59.12	50	CTGCCGCATTACCAACT	GGGTGGTGGCAATGAATGTT	236
WS36	gi|366844522	p4	(ATGA)3	58.03	45	GCCAAGTGAAAACAAGCCTA	AAACCAATCGAAGTAAAAGATCAAA	179
WS37	gi|366844543	p4	(TGTT)3	59.85	40.91	TCGATGAGTGAAGAAGGTCAAA	TCTCAAACACGCACACAAGA	161
WS38	gi|366844509	p4	(CTAT)3	60.13	50	TCCACTTTCCAGTGGTCACA	GGAGCACAATTCCGGTAAAA	174
WS39	gi|366843909	p4	(GATG)3	60.31	50	AGTAAACCGATTCGGGAAGG	TGCTTTCAACTGGAGCATTG	187
WS40	gi|366843909	p4	(CTAG)3	60.31	50	AGTAAACCGATTCGGGAAGG	TGCTTTCAACTGGAGCATTG	187
WS41	gi|366843825	p4	(GATT)3	59.75	55	GCACGAGGAGCAACTTCTCT	CCATGCCTTGTTCTCCTCAT	236
WS42	gi|285804433	p4	(TTGA)3	60.45	40	AAAGGCGGTTGATTGATTGA	AACACCGTACCGGAAACAAA	152
WS30	gi|366844556	p5	(AAAAG)3	59.85	55	GGTACCTCAACGGGTTCGTA	TCTTGCAATGAGCCTCACAC	153
WS31	gi|366844502	p5	(ATTTT)3	60.03	50	GAGAGCAACAACCAGCAACA	AATTTTCATGACCGGAGCTG	204
WS32	gi|366843906	p5	(TCTTT)4	59.83	50	TGATGTTTCCTCGTCGTCTG	GCCAAACTGGTCATAAATGGA	150
WS33	gi|366843547	p5	(TGGGG)3	60.07	50	TAGCCTCATTGGGGAAGTTG	ACTGGAAACATGCCTCCTTG	179
WS34	gi|366843529	p5	(TATTT)3	60.19	50	GAGCATTGGATGATCCTGCT	TATTTTCCAAGGCGTTGACC	210
WS24	gi|366849154	p6	(CAGGGT)3	59.99	45	CAATTGCTTCTGCCACTTCA	TCATCACCACGGTCACTGTT	241
WS25	gi|366844574	p6	(GGCTTT)3	59.96	50	TCCGGTGGACTTTAGGTTTG	GGGACGGTCAGTTCAATCTC	192
WS26	gi|366844561	p6	(TCTCCA)3	60.15	55	CCTCTCCATCTCCATCTCCA	CACCAAAGCAGCAGCAATAA	188
WS27	gi|366844149	p6	(TAAAAC)3	60.09	40	TTTTGGAAGCATTGGACACA	CAGTGACTGGAGCAGGAACA	222
WS28	gi|366844147	p6	(CATCAC)3	60.03	45	CAGTTTTGCAGTGCTTTGGA	AGCGGAGCTAACCAAGTTGA	195
WS29	gi|285804426	p6	(CATAGG)3	59.19	33.33	TCCAAATCACAAAAAGTGCAA	GTTTACTCCCATCCGTTGGA	203
WS23	gi|366843485	p8	(CTCGTCCT)3	60.09	40	TCCATTGCAAAGGTTGTGAA	ATGATGATGACGACGACGAA	210