Plant Omics Journal POJ

POJ 4(3):154-162 (2011) ISSN:1836-3644

Development and characterization of EST based SSR markers in Madagascar periwinkle (*Catharanthus roseus*) and their transferability in other medicinal plants

Raghvendra Kumar Mishra, Baniekal Hiremath Gangadhar, Jae Woong Yu, Doo Hwan Kim*, Se Won Park*

Department of Molecular Biotechnology, College of Life and Environmental Sciences, Konkuk University, Seoul, South Korea 143-701

***Corresponding authors: Doo Hwan Kim, dhkim@konkuk.ac.kr; Se Won Park, sewpark@konkuk.ac.kr**

Abstract

Simple sequence repeat (SSR) markers derived from expressed sequence tags (ESTs) are important resources for gene discovery and mapping. The objectives of the work presented here are to develop EST based SSR markers, their polymorphism in *Catharanthus roseus* and to study Cross-species transferability in other medicinal plants. *C. roseus* is a plant of great medicinal value and contains more than 130 alkaloids namely vincristine, vinblastine, ajamalcine, serpentine, etc. In this study, we analyzed 19899 ESTs and identified 2034 SSRs from 1763 SSR-containing unique ESTs. Of the 2034 EST-SSRs, 42.8% were dinucleotide repeats, 41.9% tri, 6.8% tetra, 4.7% penta and 3.6% were hexanucleotide repeats. From these SSR-containing ESTs, a total of 170 primer pairs were designed among them 20 were selected randomly for the validation of amplification and polymorphism in *C. roseus*. Out of these 20 primer pairs, 16 (80%) primer pairs could amplify and among them 11 (55%) exhibited polymorphic profiles in *C. roseus* genotypes. Furthermore, transferable analysis of these primer pairs revealed high Cross-species transferability (31-57%) in three medicinally important plants *Croton macrostachyus*, *Salvia schimperi* and *Ageratum conizoides*. The high reproducibility, scorable bands per marker (3.2), transferability and polymorphism demonstrate the potential utility of these EST-SSR markers in molecular mapping, quantitative trait loci (QTL) analysis and comparative mapping in *C. roseus* and other medicinal plants.

Keywords: *Catharanthus roseus,* Expressed Sequence Tag, Simple Sequence Repeats, Transferability, Terpenoid Indole Alkaloids.
Abbreviations: EST- Expressed Sequence Tag; SSR- Simple Sequence Repeats, NCBI- National Center for Biotechnology Information; QTL- Quantitative trait loci, MISA-MIcroSAtellite; BLAST- Basic Local Alignment Search Tool; TIA- Terpenoid Indole Alkaloids.

Introduction

Catharanthus roseus is a diploid (2n = 2x =16) species of *Catharanthus* genus. It is one of the important medicinal plant containing more than 130 mono-terpenoid indole alkaloids (TIAs) in different plant parts like root, shoot and stem (van der Heijden et al., 2004; Shukla et al., 2006). No other single plant species is reported to produce such a wide range of alkaloids that are useful against cancer and other human diseases (van der Heijden et al., 2004). The biosynthesis of alkaloids in *C. roseus* is governed by a complex terpenoid indole alkaloid (TIAs) pathway involving several genes (St-Pierre et al., 1999). The alkaloid profiles vary between the aerial and underground organs of the plant at different developmental stages and this difference in profiles are regulated by differential expression of genes in TIA pathway. Attempts have been made to study the differential expression of TIA biosynthetic genes and around 155 expressed sequence tags (ESTs) were developed (Shukla et al., 2006). Expressed sequence tag (EST) is presently the most widely accepted nucleotide sequence for marker based study as it represents transcribed part of the genome. A large amount of EST database present on public domain (NCBI) can be exploited for development of SSR markers and functional annotation by using software based *In silico* analysis. The development of SSR markers by this method is more preferable today (Scott et al., 2000; Temnykh et al., 2000) as they are cost effective, need less time to development and more informative in comparison to conventional method (Zane et al., 2002). The EST-SSR markers not only help in molecular mapping but also provide an opportunity for gene discovery when they show linkage with a trait of interest (Thiel et al., 2003). EST-SSR markers have more conserved region in comparison to markers that generated from genomic sequences, therefore they show more transferability between species (Portis et al., 2007; Varshney et al., 2005). All these properties make EST-SSRs more popular among the existing markers for development of molecular maps or QTL analysis. Also they are being used widely in different plant species like rice (Kantety et al., 2002), bread wheat (Kantety et al., 2002; Gupta et al., 2003), *Capsicum* (Minamiyama et al., 2006; Portis et al., 2007), sugarcane (Cordeiro et al., 2001) and cotton (Park et al., 2005) for molecular mapping, genetic diversity, transferability etc. However in *C. roseus*, development of EST-SSR markers is not reported till date. In the present study, we examined 19899 *C. roseus* EST sequences available on NCBI database based on (1) frequency and

Table 1. Summary of the *In silico* search for SSRs in *C. roseus*

Parameters used in screening	Data generated by MISA
Total number of sequences examined	19899
Total size of examined sequences (bp)	10339774
Total number of identified SSRs	2034
Number of SSR containing sequences	1763
Number of sequences containing more than 1 SSR	227
Number of SSRs present in compound formation	139
Dinucleotide	872
Trinucleotide	853
Tetranucleotide	139
Pentanucleotide	96
Hexanucleotide	74

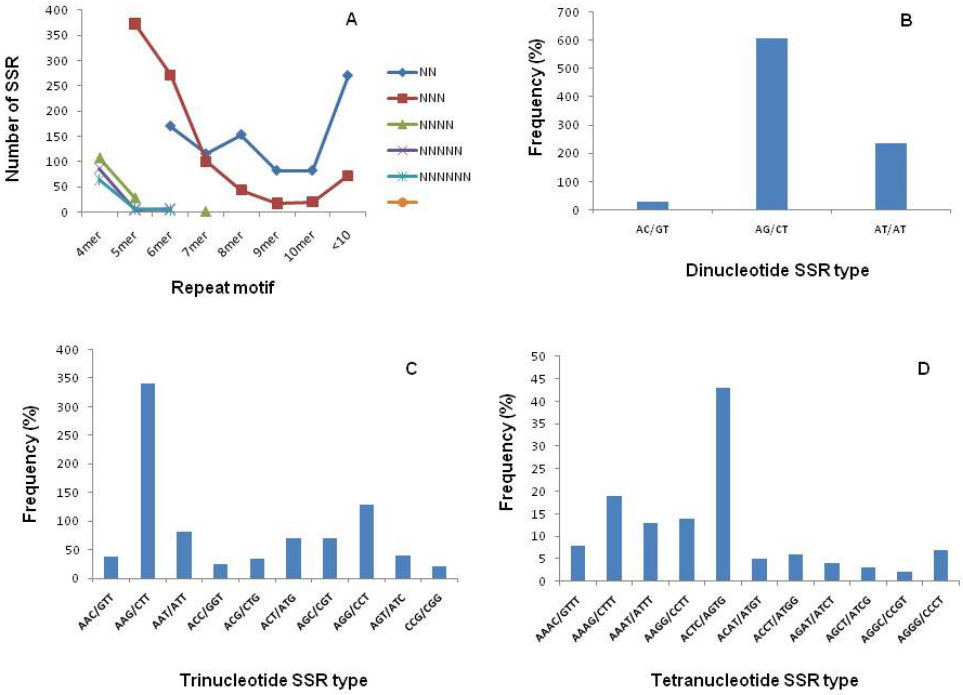

Fig 1. Frequency distribution of different nucleotide repeats in identified SSR sequences. (A) Different nucleotide repeats in *C. roseus*, (B) Dinucleotide, (C) Trinucleotide and (D) Tetranucleotide.

distribution of SSRs in EST (2) functional annotation and prediction of amino acid from SSR loci and (3) development and validation of polymorphism of EST-SSR markers in *C. roseus* cultivar. We tested their applicability for studying transferability in other medicinal plants.

Materials and methods

Detection of SSR-containing sequences

A total of 19899 *C. roseus* EST sequences (equivalent to 10.3 Mbp) were retrieved from NCBI database. For EST-SSR development, we analyzed all these sequences by tandem repeat finder and then sequences were assembled by CAP3 programme (Huang and Madan, 1999). After pre-processing, SSR containing sequences were identified by a perl script based program, MISA software (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/).

Primer designing

A total of 803 (45.5%) EST sequences allowed for primer designing from 1763 SSR containing sequences. The remaining 960 (54.45%) SSR primers were inappropriate for primer construction. Out of 803, 170 primer pairs (supplementary data 1) were successfully developed on the basis of following standard parameters by using primer3 software (Rozen et al., 2000) (a) the target amplicon size of 100–500 bp, (b) the optimal annealing temperature to 55-60^{o}C, (c) average GC content 50-60% and (d) the primer length at 18-24 bp. Putative functions of SSR loci were assigned by comparison with the non-redundant sequence database at NCBI using the BLASTX2.2.17 software (Altschul et al., 1997). Finally, 20 primer pairs were randomly selected for validation of amplification and assessment of the polymorphism in *C. roseus* (Gong et al., 2010).

Table 2. Characteristics of 20 EST-derived SSRs for *Catharanthus roseus*.

S.No.	ID	SSR type	SSR motif	Tm (^{o}C)	GC%	Forward Primer	Reverse Primer	Expected Product size (bp)
RG01	gil164561865	p3	(CTT)7	59.86	50	GGAACCAAGGATGTTAGAGTGG	CTGCAACGGTTACTAGAGAGTAGAGC	157
RG07	gil164556449	p3	(GCA)5	60.00	57	GAGGAGGTGTCTCTCATGCTG	CGACCTCAACAGAAGGTTCG	175
RG08	gil164555839	p3	(GCA)6	60.30	55	AGAAGGAAGTGGTGGTGCTG	GTTTACAGGGGGAGGAGGAG	159
RG09	gil171846161	p2	(TA)7	60.59	61	CTGGGGCTCTTGGGTTTC	GCTGCTTCTTCTGCCGTCTC	282
RG10	gil164561674	p2	(AT)9	60.79	63	CGTCAGCTCAGCAGTGGAC	CTTCCTCCTTACCCCTAGCC	241
RG11	gil164561349	p2	(AT)11	62.09	60	GGCACGAGGCATCCTTACTC	CCACAGCTCTGGTAGCTCCT	115
RG15	gil164556819	p2	(AG)8	59.58	60	GAGAGAGAGAGAGCGGCAAG	GTGGGTCTCCCACAATAGCC	263
RG18	gil164561715	p4	(AAGA)4	60.64	50	CATTCTCTTCTCGAGGCTTCTG	ACCCCCATGACAGTCAAGATAG	280
RG20	gil164556358	p3	(AGC)8	59.75	55	AGCTGAACAAGGAGCTGAGG	GCGCCGGTTCTTTCTTCT	104
RG21	gil164559832	p5	(TGAAA)4	60.24	50	CCCTTCCTGAGAGACTCAAATG	CCAAGCACTCTTCATCTCAGG	273
RG22	gil164553982	p3	(ATC)6	60.16	60	CGAGGAGGTGTTAGCAGAGC	GGAGAGATGATGGGCTGATG	278
RG23	gil164553785	p3	(TCT)8	59.83	60	GAGCCATAGCACCAGAGGTC	GGGTCAGCCCAGTAGCTGTA	118
RG24	gil164551888	p2	(TC)7	62.63	60	CACGAGGGGGTGGATTAGTG	GTAACCAGGAGGGGGTGTCT	203
RG25	gil164551881	p2	(TA)11	60.09	61	CACCCTTTCCTGGCACAC	GTAGTGGCGGGAGGTGATAG	148
RG26	gil116658717	p2	(TC)14	59.82	61	GGACAGCGAGCTTTGAGC	GGAGGATGCCACACCAAG	126
RG27	gil164556759	p5	(TGAAA)4	60.89	50	GTTATGACGTGTGAAGCAGCAG	CCCAAGCACTCTTCATCTCAG	202
RG28	gil164555337	p5	(TCCTT)4	61.53	50	CTCTTGGTGATGCTAACAGAGACC	CCTTGATTCCTTCCTTCCTCCT	172
RG29	gil164554532	p3	(GCA)6	59.02	55	GTGAAAGCCTCCGAAGAGTC	CTTCGGTAGCAAAGGGACTG	186
RG30	gil164554493	p3	(GAT)5	60.07	55	GGACCTCTGCCTGTTATGGA	CGCTCCCTTGTCCAGAGTAA	205
RG42	gil164558595	p6	(AGTTGG)6	61.25	50	GGCTCTAATACGTTGCAGTACACC	GAAAGCTCCTGTCTCATCAACC	278

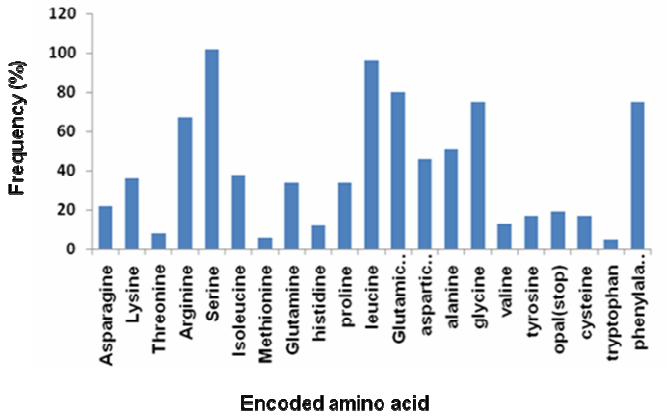

Fig 2. Distribution of EST-SSR encoded amino acids in *C. roseus*.

Plant materials

In this study, we used two *C. roseus* cultivars namely Prabal (Lucknow, India) (Dwivedi et al., 2000) and Pink (Rajdhani nursery, New Delhi, India) and wild plants of three other medicinally important plants namely *Ageratum conizoides*, *Salvia schimperi* and *Croton macrostachyus* (Geyid et al., 2005) were collected from Afar, Ethiopia. The *C. roseus* (*Apocynaceae*) and *S. schimperi* (*Lamiaceae*) belongs to euasterids I and *A. conizoides* (*Asteraceae*) belong to euasterids II. Euasterids I and II are belongs to same clad asterids. Whereas *C. macrostachyus* (*Euphorbiaceae*) comes in separate clad rosids; both clad rosids and asterids comes under eudicots (The Angiosperm Phylogeny Group). All the plants were grown at green house facility of Konkuk University in 2010. Genomic DNA was isolated from leaves of plants using CTAB method (Khanuja et al., 1999).

PCR amplification for EST-SSR polymorphism and transferability

The polymorphic and transferability study of developed 20 primer pairs, were carried out by previously described method (Gupta et al., 2003; Shokeen et al., 2007; Gong et al., 2010; Hu et al., 2010). The PCR reactions were performed using the M-Cycler (Bio-Rad, USA). Each 20 ml PCR reaction mix contained 50 ng of DNA, 2 µl of 10X PCR buffer, 2 mM $MgCl_2$, 0.2 mM of each dNTPs, 1 µM of each primer and 1U of Taq DNA polymerase (Invitrogen, USA). A touchdown PCR programme consisting of an initial denaturation at 95 ^{0}C for 2 min followed by 15 cycles of denaturation at 95 ^{0}C for 20 s, annealing at 65 ^{0}C for 50 s and extension at 72 ^{0}C for 50 s was used. The annealing temperature decreased by 1^{0}C after every cycle till it reached an annealing temperature of 55 ^{0}C. The next 25 cycles consisted of denaturation for 20 s at 95 ^{0}C, annealing for 50 s at 55 ^{0}C and extension for 50 s at 72 ^{0}C followed by a final extension of 7 min at 72 0C. The PCR products were resolved on 3% metaphore agarose gel in Tris-borate EDTA buffer, gels were stained with ethidium bromide (0.5 µg/ml) and analysed using a gel documentation system (Alpha Imager 2200, Alpha Innotech Corp., USA).

Results

Frequency and distribution of EST-SSRs

A total of 19899 *C. roseus* ESTs retrieved from the NCBI data base (http://www.ncbi.nlm.nih.gov/dbest/) were used to evaluate the presence of SSR motifs. After pre-processing with CPA3, MISA software was used to identify SSRs from ESTs. The following search criteria were used for identification of SSRs: 6 repeat units for dinucleotide repeats, 5 repeat units for trinucleotide repeats, 4 repeat units for tetranucleotide repeats, and 3 repeat units for penta and hexa nucleotides repeats. From the total of 19899 ESTs, we have detected 1763 EST sequences carrying 2034 (10.22%) SSRs. Among these SSRs, 872 (42.8%) had dinucleotide repeats and 853 (41.9%) trinucleotide repeats (Table 1). Among the dinucleotide repeats, motif AG/CT (69.6%) was the most common followed by AT/AT (27.2%). Among the trinucleotide repeats, AAG/CTT (39.9%) was most common followed by AGG/CTT (15.1%) and AAT/ATT (9.5%). The most common tetranucleotide motif was ACTC/AGTG (29.5%) followed by AAAG/CTTT (14.7%). Among the penta and hexa nucleotide repeats, motif AAATG/ACTTT (32.3%) and AAAGCC/CGGTTT (45.9%), were found respectively (Fig. 1).

Distribution and nature of encoded amino acids

Each trinucleotide motif present in SSR loci codes a specific amino acid which plays an important role in biological, cellular and metabolic process in plants. In our study, we have found 853 trinucleotides repeats motif representing 21 amino acids (Fig. 2) including stop codons. The percentage of trinucleotide motifs coding for serine was highest (11.95%) closely followed by leucin (11.25%), while the tryptophan were least present (0.6%). These encoded amino acids were further divided into four categories such as (a) hybrophobic and hydrophilic, (b) aromatic and aliphatic, (c) polar and non polar and (d) acidic and basic. The identified SSRs of *C. roseus* coded for all the four types of amino acids but at varied frequencies (Fig. 3). The most common essential amino acid coded in *C. roseus* was leucin (11.25%). Other essential amino acids present were phenylalanine (8.8%), isoleucine (4.4%), lysine (4.2%) and histidines (1.4%).

Development of EST-SSR markers and their polymorphism

From the 19899 *C. roseus* ESTs, a total of 1763 SSR containing sequences were used for primer development. However, only 803 (45.5%) SSR containing sequences could be used for EST-SSR primer development and rest were found not suitable due to short flanking sequences. The primers developed in this study comprised of 202 (25.1%) dinucleotide repeats, 374 (46.6%) trinulceotide repeats, 108 (13.4%) tetranucleotide repeats, 61 (7.6%) pentanucleotide repeats and 58 (7.2%) hexanucleotide repeat SSRs. Out of 803 SSR containing sequences 170 primer pairs were successfully designed. The amplification and polymorphism of these primer pairs were tested by a sub set of randomly selected 20 primer pairs in *C. roseus* cvs. Prabal and Pink (Table 2). Out of twenty, sixteen (80%) primer pairs showed amplification in *C. roseus*. Five primers amplified single alleles while eleven primer pairs amplified multiple alleles. These eleven primer pairs were used to study polymorphism in *C. roseus* cvs. Prabal and Pink var. All these eleven primer

Table 3. Polymorphism analysis of 11 EST-SSR primers generated in present study

Primer name	Gene Bank accession no.	Allele number and band size range (bp)	*Catharanthus* "Prabal" (bp)	*Catharanthus* "Pink" (bp)	E-value	Putative function
RG01	FD4250061	2(200-270)	250,270	200,240	3.00E-42	Hypothetical protein[*Oryza sativa* Indica group]
RG04	FD660772.1	3 (140-260)	140,180,230	230,260	1.00E-07	Pathogenesis-related transcriptional factor and ERF [*Medicago truncatula*]
RG05	FD424897.1	3 (170-220)	170, 190,220	190,220	2.00E-12	hypothetical protein [*Oryza sativa* Indica Group]
RG07	FD420596.1	2 (180-260)	180,260	180,240	2.00E-67	S18.A ribosomal protein [*Arabidopsis thaliana*]
RG08	FD425272.1	3 (150-260)	150,200	150, 220,260	3.00E-56	major intrinsic protein 1 [*Solanum tuberosum*]
RG11	FD417203.1	3 (200-300)	200,245,300	210, 245	4.00E-30	carbonic anhydrase [*Solanum lycopersicum*]
RG12	FD417407.1	4 (130-280)	130, 155, 200,280	130,155,280	2.00E-102	chlorophyll A/B binding protein, putative [*Ricinus communis*]
RG13	FD415410.1	3 (200-260)	200,235	200, 210, 260	4.00E-21	extensin-like protein [*Cucumis sativus*]
RG15	EG559729.1	3 (180-280)	170, 225,290	225,280	6.00E-30	Protein PIR, putative [*Ricinus communis*]
RG17	FD418705.1	4 (170-390)	170, 225,290	225,290,300, 390	1.00E-08	chloroplast ribulose-1,5-bisphosphate carboxylase [*Solenostemon scutellarioides*]
RG18	FD418321.1	5 (160-275)	160,190,200, 250, 275	190,200,250	3.00E-49	Thylakoid membrane phosphoprotein 14 kDa, chloroplast precursor [*Ricinus communis*]

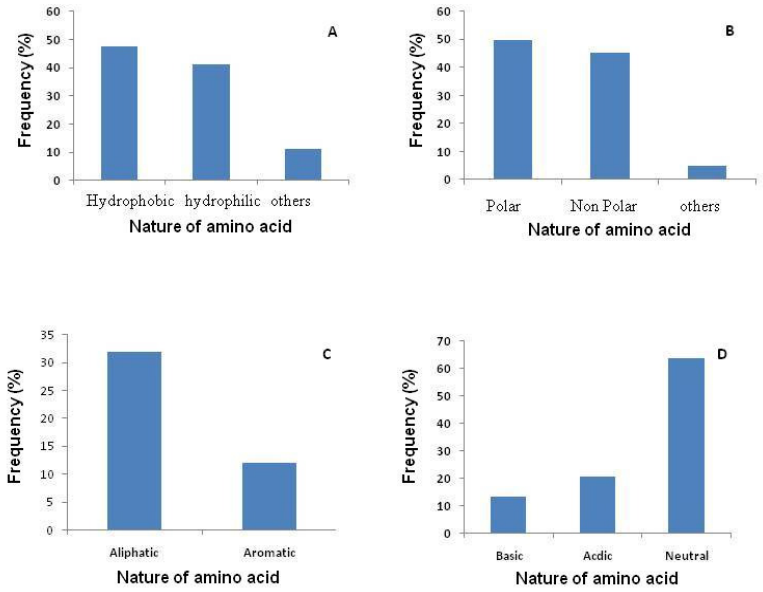

Fig 3. Frequency and distribution of amino acid on the basis of their behavior. (A) Hydrophobic and hydrophilic, (B) Polar and non polar, (C) Aromatic and aliphatic and (D) Acidic and basic amino acids.

pairs successfully amplified the DNA from both plants and produced 35 alleles with an average of 3.2 alleles per locus and sizes range from 100 to 500 bp (Table 3).

Transferability analysis of EST-SSRs

Transferability of the developed EST-SSR primers was studied by Cross-species amplification of three medicinally important plants *A. conizoides*, *S. schimperi* and *C. macrostachyus*. Out of 20 primer pairs tested, 11 (55%) primers showed amplification in *A. conizoides*, 9 (45%) in *S. schimperi* and 6 (30%) primers showed amplification in *C. macrostachyus*. Four primer pairs namely RG01, RG05, RG07 and RG11 showed amplification in all the plants including *C. roseus* (Table 4) showing the Cross-species transferability of the *In silico* developed EST-SSRs of *C. roseus* in *A. conizoides*, *S. schimperi* and *C. macrostachyus*.

BLASTX analysis

Based on BLASTX analysis, putative functions could be assigned to 225 SSR loci. On the basis of their significant match, all SSR loci were divided into three groups; (a) Biological function: These genes regulate all biological functions related to photosynthesis, cell signaling, stress, etc. In this study with the help of BLASTX analysis, we have identified 80 (35.5%) SSR loci were shown to have similarity with the active biological putative proteins, 34% with photosynthetic putative proteins, 15% with cell signaling, 12.5% with different stress (oxidative, cold and ethylene stress) related proteins and 11% with protease inhibitors (Fig, 4A). (b) Cellular function: A total of 125 (55.5%) SSR loci were shown to have similarity with putative cellular proteins like DNA binding transcriptional factors, common structural proteins, ribosomal proteins, cell wall proteins, etc. Among them 22.4% showed similarity with common structural proteins, 21.6% with ribosomal proteins, 17.6% with auxin related hormone dormancy and 8.8% with cell wall putative proteins (Fig. 4B). (c) Metabolic function: *C. roseus* is an important medicinal plant containing more than 130 primary and secondary metabolites that are useful for treating cancer and other diseases in humans. Based on BLASTX analysis, 20 SSR loci were shown to have putative function in primary and secondary metabolism and involved in sucrose synthesis, glycolysis, aspartyl and serine protease, nitric oxide synthesis etc. Some SSR loci shown to have homology with enzyme serine hydroxymethyltransferase/Glycine hydroxymethyl-transferase of primary metabolism and enzymes O-methyltransferase, flavonoid O methyltransferase and strictosidine β D glucosidase of secondary metabolism (Fig. 4C). Strictosidine β D glucosidase play an important role in terpenoid indole alkaloid (TIAs) pathway.

Discussion

Catheranthus roseus, commonly known as "periwinkle (evergreen)" is highly studied plant because of its medicinal properties. The secondary metabolites produced by this plant are antileukemic (vincristine and vinblastine) and antihypertensive (ajmalicine and serpentine) (van der Heijden et al., 2004). However, extremely low yields hamper the extensive use of these alkaloids for medicinal purposes. Molecular marker based techniques were applied for screening of high yielding varieties and their mapping. So far, 42 genomic SSRs were developed and successfully used them in genotype screening of *C. roseus* (Shokeen et al., 2007). However, these genomic SSR are time consuming, labour intensive and costly. Now days, EST-SSRs are evolved as an important class of molecular markers owing to their easy availability, hyper variability, and suitability for high throughput analysis, high polymorphism and transferability in comparison to other available markers. Presently EST-derived SSR markers are being used in marker assisted selection (MAS), development of high yielding varieties, molecular mapping and quantitative trait loci (QTL) analysis (Varshney et al., 2005). These markers are more useful for the genetic improvement and diversity analysis in comparison to other marker types. *In silico* analysis provides an opportunity to mine SSRs from EST database at low cost, in less time and may provide abundant information (Gupta et al., 2003). Apart from *C. roseus,* the other plant species used in our study namely *A. conizoides*, *S. schimperi* and *C. macrostachyus* are important medicinal plants being used for curing various diseases in humans (Geyid et al., 2005). Except for *C. roseus*, no prior reports available regarding development and characterization of molecular markers in these plants (Shokeen et al., 2007). In the present study we have developed EST-derived SSR markers from 19899 ESTs available on public database by *In silico* analysis. The frequency of EST-SSRs (10.22%) identified in present study is much higher compared to earlier studies in arabidopsis (2.4%), almond (4.1%), peach (4.1%) and rosa (4.8%) (Jung et al., 2004). The difference in the frequencies of EST-SSRs could be attributed to the "search criteria" used, type of SSR motif, size of sequence data analysis and the mining tool used (Portis et al., 2007; Gupta and Prasad, 2009). The most abundant repeat motifs found in the present study were AG/CT (69.6%), AAG/CTT (39.9%) and ACTC/AGTG (30.2%) for dinucleotide, trinucleotide and tertranucleotide repeats, respectively, which was quite similar to earlier report in rubber tree (Feng et al., 2009). Although the development of EST-SSR markers was reported in other plant species wheat (Gupta et al., 2003), cucumber (Hu et al., 2010) medicago (Gupta and Prasad, 2009), no EST based SSR markers of *C. roseus* have been reported and their Cross-species transferability was studied. In this study, 170 primer pairs were developed and their efficacy of amplification and polymorphism were evaluated by random selection method as reported in pea, medicago and cucumber (Gong et al., 2010; Gupta and Prasad, 2009; Hu et al., 2010). Randomly selected 20 primer pairs were used to check the application capability of all 170 primer pairs. Out of these 20 primer pairs, 16 (80%) primers produced amplification in *C. roseus*. The amplification percentages of these primers are much higher than previous studies in rubber tree and pea (Feng et al., 2009; Gong et al., 2010). Out of sixteen, eleven primer pairs (55%) showed polymorphism and generated 35 alleles with an average of 3.2 alleles per locus, which was close agreement with earlier studies (Shokeen et al., 2007; Gupta and Gopalakrishna, 2010). These result proved that primer pairs developed in the present study are highly polymorphic. Out of 20 primer pairs were used for Cross-species transferability study, 55% primers showed amplification in *A. conizoides*, 44% in *S. schimperi* and 30% showed amplification in *C. macrostachyus*. Transfer success rate is relatively low in *C. macrostachyus* because it is distantly related to *C. roseus*. These results are in close agreement with previous studies in soybean plant where the transfer success rates decrease with the increasing genetic distance (Peakall et al., 1998). In case of *A. conizoides* and *S. schimperi,* the transferability rate is medium because EST-SSR decreased slightly in the more closely related species (Gao et al., 2005). The transferable rates generated by these primers are satisfactory and provide a new tool for further studying of

Table 4. Transferability study of four EST-SSR markers in three medicinal importance plants along with allele number and band size

Primer pair	Number of alleles	Size of alleles (bp)	Size of alleles in each plants (bp)		
			A. conizoides	*C. macrostachyus*	*S. schimperi*
RG01	3	170-240	200,210,240	170,220	220
RG05	3	200-290	200,210	220	200, 290
RG07	3	190-230	190,230	200, 230	220
RG11	2	200-225	225	200	200

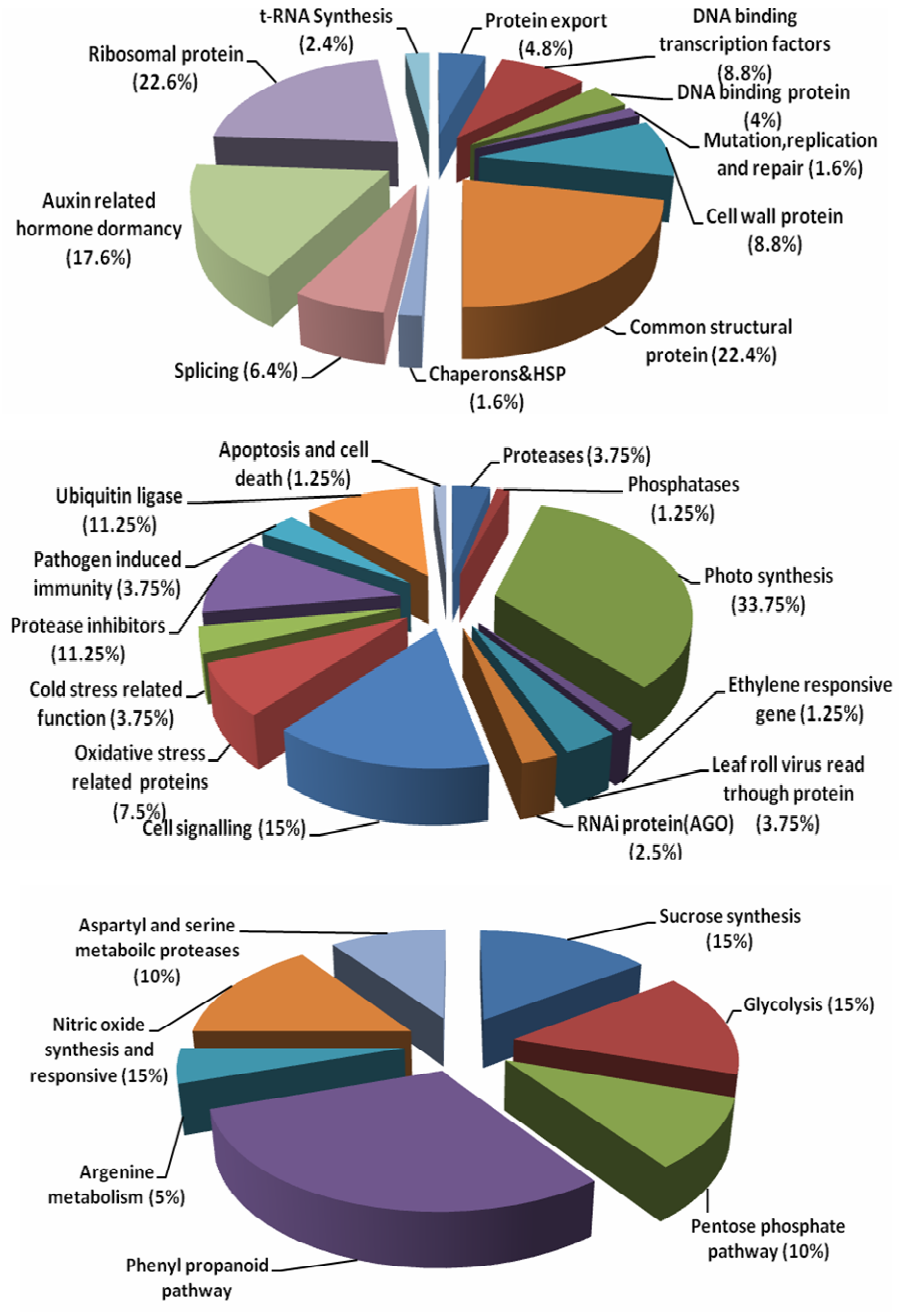

Fig 4. Classification of EST-SSRs on the basis of significant match with putative proteins (A) Cellular putative proteins, (B) Biological putative proteins and (C) Metabolism putative proteins.

these plants. Our results further confirm the reports on transferability of EST-SSRs (Gupta et al., 2003; Bandopadhyay et al., 2004; Saha et al., 2004; Varshney et al., 2005, Gupta and Gopalakrishna, 2010; Gupta and Prasad, 2009). Each trinucleotide motif represents an amino acid which has several putative functions. In *C. roseus* 853 trinucleotide motifs encoded 21 amino acids. The amino acids serin (Ser), leucine (leu) and glutamic acid (glu) were dominant in *C. roseus* similar to previous study (Victoria et al., 2011). The highest percentage of serine (11.95%) in *C. rosus* was close agreement with fern species and Gnetum (Victoria et al., 2011). The percentage of tryptophan (0.6%) was lower in *C. roseus*, however in arabidopsis, pinus and oryza tyrosine, aspartic acid were lowest (Victoria et al., 2011). The putative function of ESTs was assigned by BLASTX analysis and classified them into three groups based on homology: biological, cellular and metabolic process. The EST-SSR primer pairs were developed for Serine hydroxymethyltransferase / glycine hydroxylmethyl transferase, O-methyltransferase, flavonoid O-methyltrans-ferase and strictosidine β-D-gluco sidase (sgd). Strictosidine β-D-glucosidase (sgd) is a key enzyme of the TIA biosynthetic pathway in *Catharanthus*, which determines the fate of the post strictosidine pathway (Shukla et al., 2006; Geerlings et al., 2000). The significant value of this work is, the SSR loci have shown homology with putative function, indicate a good approach for using these SSR loci as a molecular marker to saturate primary and secondary metabolic pathways in plants (Victoria et al., 2011). The polymorphism and transferability analysis of EST-SSR markers indicated the value of developed markers. The applicability of Cross-species amplification of developed EST-SSR provides a good opportunity for studying of unknown medicinal plants. The high amplification tendency, polymorphism and transferability of these primer pairs suggest that markers developed in this study are highly useful in markers assisted selection (MAS), genetic diversity studies, linkage mapping, comparative mapping and QTL analysis. The work presented here provides supplement to ongoing efforts for development of well matured *C. roseus* molecular map. The newly developed EST-SSRs are informative for phylogenetic and transferability studies and represent an important valuable source of gene based SSR markers.

Acknowledgements

This paper was supported by Konkuk University in 2010. We are grateful to Mr. Turasafawo Jarso, Samara University, Afar, Ethiopia for providing plant material.

Conflicts of interest statement

The authors declare that they have no conflicts of interest.

Supplementary data 1

EST-SSRs developed in this study are available as supplementary data 1.

References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang ZH, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST:a newgeneration of protein database search programs. Nucleic Acids Res 25:3389–3402

Bandopadhyay R, Sharma S, Rustgi S, Singh R, Kumar A, Balyan HS, Gupta PK (2004) DNA polymorphism among 18 species of *Triticum–Aegilops* complex using wheat EST-SSRs. Plant Sci 166:349-356

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to Erianthus and Sorghum. Plant Sci 160:1115-1123

Dwivedi S, Singh M, Singh A, Sharma S, Uniyal GC, Kumar S (2000) Assessment of genetic divergence for its purposeful exploitation in periwinkle *Catharanthus roseus.* J Genet Breed 54:95-99

Feng SP, Li WG, Huang HS, Wang JY, Wu YT (2009) Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol Breed 23:85-97

Gao LZ, Zhang CH, Jia JZ (2005) Cross-species transferability of rice microsatellites in its wild relatives and the potential for conservation genetic studies. Genet Resour Crop Evol 52:931–940

Geerlings A, Ibanez MML, Memelink J, van der Heijden R, Verpoorte R (2000) Molecular cloning and analysis of strictosidine b-D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. J of Bio Chem 275:3051-3056

Geyid A, Abebe D, Debella A, Makonnen Z, Aberra F, Teka F, Kebede T, Urga K, Yersaw K, Biza T, Mariam BH, Guta M (2005) Screening of some medicinal plants of Ethiopia for their anti-microbial properties and chemical profiles. J of Ethno 97:421-427

Gong YM, Xu SC, Mao WH, Hu QZ, Zhang GW, Ding J, Li YD (2010) Developing new SSR markers from ESTs of pea (*Pisum sativum* L.) J. of Zhejiang Univ. 11**:** 702-707

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan, HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genom 270:315-323

Gupta SK, Gopalakrishna T (2010) Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other *Vigna* species. Genome 53:508-523

Gupta S, Prasad M (2009) Development and characterization of genic SSR markers in *Medicago truncatula* and their transferability in leguminous and non-leguminous species. Genome 52:761-771

Hu JB, Zhoub XY, Li JW (2010) Development of novel EST-SSR markers for cucumber (*Cucumis sativus*) and their transferability to related species. Scientia Horticul 125:534-538

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868-877

Jung S, Jesudurai C, Staton M, Du Z, Ficklin S, Cho I, Abbott A, Tomkins J, Main D (2004) GDR (Genome Database for Rosaceae): integrated web resources for rosaceae genomics and genetics research. BMC Bioinform 5:130

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barely, maize, rice, sorghum and wheat. Plant Mol Biol 48:501-510

Khanuja SPS, Shasany AK, Darokar MP, Kumar S (1999) Rapid isolation of DNA from dry and fresh samples of plant producing large amounts of secondary metabolites and essential oils. Plant Mol Biol Rep 17:1-7

Minamiyama Y, Tsuro M, Hirai M (2006) An SSR-based linkage map of *Capsicum annuum*. Mol Breed 18:157-169

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genom 274:428-441

Peakall R, Gilmore S, Keys W (1998) Cross-species amplification of soybean (Glycine max) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275-1287

Portis E, Nagy I, Sasva Z, Stagelri A, Barchi L, Lanteri S (2007) The design of *Capsicum* spp. SSR assays via analysis of *In silico* DNA sequence, and their potential utility for genetic mapping. Plant Sci 172:640-648

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers, in: S. Krawetz, S. Misener (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, 365-386

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang LJ, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783-791

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henery RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723-726

Shokeen B, Sethy NK, Kumar S, Bhatia S (2007) Isolation and characterization of microsatellite markers for analysis of molecular variation in the medicinal plant Madagascar periwinkle (*Catharanthus roseus* (L.) G. Don). Plant Sci 172: 441-451

Shukla AK, Shasany AK, Gupta MM, Khanuja SPS (2006) Transcriptome analysis in *Catharanthus roseus* leaves and roots for comparative terpenoid indole alkaloid profiles. J of Exp Bot 57:3921-3932

St-Pierre B, Vazquez Flota FA, De Luca V (1999) Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. The Plant Cell 11:887-900

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697-712

The Angiosperm Phylogeny Group (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Botani J of the Linne Soci 141:399-436

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411-422

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537-546

vander Heijden R, Jacobs DI, Snoeijer W, Hallard D, Verpoorte R (2004) The *Catharanthus* alkaloids: pharmacognosy and biotechnology. Curr Medi Chem 11:607-62

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers: their characteristics, development and application to plant breeding and genetics. Trends Biotech 23:48-55

Victoria1 FC, da Maia LC, de Oliveira1 AC (2011) *In silico* comparative analysis of SSR markers in plants. BMC Plant Biol 2011: 11-15

Zane L, B argelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1-16