



## Computational Recognition and Characterization of EST-SSR in *Vinca rosea*

Manisha Ray<sup>1</sup>, Bhabani Shankar Sahoo<sup>2</sup>, Souvagyalaxmi Sahoo<sup>3\*</sup>

<sup>1,2</sup> Tectona Biotech Resource Center, Bhubaneswar, Odisha, India

<sup>3</sup> Senior Scientist, Tectona Biotech Resource Centre, Bhubaneswar, Odisha, India

### Abstract

*Vinca rosea* (*Catharanthus roseus*), normally known as Sadabahar; is a decorative plant in addition to having effective medicinal values, belongs to the family Apocynaceae. The expressed sequence tags (EST) are regarded as potential source for the development of genic microsatellite markers, gene discovery, comparative genomics and genomic studies. Thus this study conveyed to analyze the EST sequences from dbEST of NCBI for *Vinca rosea*, which had been assembled in CAP3 program after which analyzed in SSR detection tool GRAMENE and outcomes 20 numbers of SSR-EST sequences were subjected to functional annotation through InterProScan. The SSR-ESTs successfully designed 48 numbers of primer pairs by using Primer3. These findings probably beneficial to investigate molecular markers which have useful importance and have to additionally facilitate the evaluation of genetic variety in apocynaceae family especially in *Vinca rosea* species.

**Keywords:** *Vinca rosea*, EST sequences, microsatellites, FDM analysis, primer design

### 1. Introduction

*Catharanthus roseus*, the Madagascar periwinkle, rose periwinkle, is a species of flowering plant in the family Apocynaceae. It was formerly included in the genus *Vinca* as *Vinca rosea*, also *Sadabahar* in local; is a fast growing ornamental plant which is grown in gardens but has immense medicinal value like anti-tumor, anti-diabetic, anti-microbial, anti-oxidant and anti-mutagenic effects, however, many life-saving drugs are developed from this plant to treat fatal diseases like vinblastine and vincristine fight against cancer [1].

It is local and endemic to India Ocean Island of Madagascar, but it has been cultivated as a decorative and medicinal plant throughout the tropical and subtropical areas worldwide inclusive of the Southern and United States [1]. Also it has ended up naturalized in Asia, Africa, America, southern Europe, Brazil, British West Indies and Oceania [2]. Europe associated species had been used for the proprietary suppression of the flow of milk. In the British West Indies it has been used to deal with diabetic ulcer and in the Philippines has been suggested as being an effective oral hypoglycemic agent [1]. In China, it is used as a traditional folk remedy to remedy dysmenorrheal, diabetes and malaria [3]. Although it is miles a treasured medicine, it is also referred as “flower of loss of life” in Italy, because of its excessive toxicity. Therefore a hit management and improvement of this plant resource requires information of the genetic variety and structure of herbal population to maximize the conservation and usage of germplasms of *Vinca rosea*. Therefore successful management and development of this plant resource requires understanding of the genetic diversity and structure of natural population to maximize the conservation and utilization of germplasms of *Vinca rosea* [4]. Hence, a study become undertaken to mine SSRs from ESTs of *Vinca rosea* and develop primers, which would be further examined and used in plant breeding technique with the aid of measure its genetic diversity [5].

Since the 1980s, molecular markers have become regularly

utilized in plant genetic research and breeding applications. Several types of markers are regularly used for cultivar fingerprinting, linkage map construction, mapping alleles for desirable traits, marker-assisted selection, and assessment of population structure; microsatellites among them [6]. The Expressed Sequence Tags (ESTs) are easiest and cheapest source for SSR improvement. SSR based molecular makers are regularly used in plant genetics because of their excessive reproducibility, co-dominant inheritance, multi allelic nature, abundance in the genome, excessive reproducibility, hyper polymorphism, excessive charge of transferability across genera and species and excessive information content material. EST-SSRs offer various compensation including ease of access, presence in gene-wealthy areas, and excessive transferability across species and genera, which allow them to serve as anchor markers for comparative mapping and evolutionary studies [7]. But because of cost effective the use of microsatellites in plant species has been reduced, therefore, here the use of *in silico* approach has been appreciated due to less cost effective and less time consuming to identify the EST-SSRs, to investigate their functionality and to screened the applicable SSRs for primer design.

### 2. Materials and Methods

#### 2.1 Retrieval of EST Sequences

The Expressed sequence tag (EST) sequences of *V. rosea* were retrieved from EST database (dbEST) (<https://www.ncbi.nlm.nih.gov/nucest/?term>) of National Centre for Biotechnology Information (NCBI) web server (<https://www.ncbi.nlm.nih.gov/>).

#### 2.2 Sequence Assembly Analysis

The collected EST sequences of *V. rosea* were taken for assembly analysis by using CAP3 (<http://doua.prabi.fr/software/cap3>) sequence assembly program, which allows to assemble a set of contiguous or contigs sequence as well as the singleton sequences.

### 2.3 Detection of SSR containing EST sequences

The resulted contigs and singleton sequences were subjected to further analysis to find out those sequences which contained the single sequence repeats (SSR) sequences through GRAMENE database (<http://www.gramene.org/>) containing SSR marker search tool i.e. SSRIT (<http://archive.gramene.org/db/markers/ssrtool>), which predicts the SSR marker contains EST sequences.

### 2.4 Functional Annotation of EST-SSRs

The functional annotation of EST-SSRs was carried out through Inter Pro Scan (<https://www.ebi.ac.uk/interpro/search/sequence-search>), as it is a resource that provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and important sites.

### 2.5 Primer Designing

The functionally annotated SSR containing EST sequences were again computed in Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) to collect the appropriate primer sequences or the forward and reverse primer from the given nucleotide sequences to design primer.

## 3. Results

### 3.1 Sequence Retrieval

There were overall 20,181 numbers of EST sequences of *Vinca rosea* had been found in EST database of NCBI, out of which most effective the sequences which were with a 1000 or greater  $\geq 1000$  in base pair length were selected for better effects. So, on the basis of this above bp length best 50 numbers of EST sequences out of 20,181 sequences were collected from EST database of NCBI and subjected to

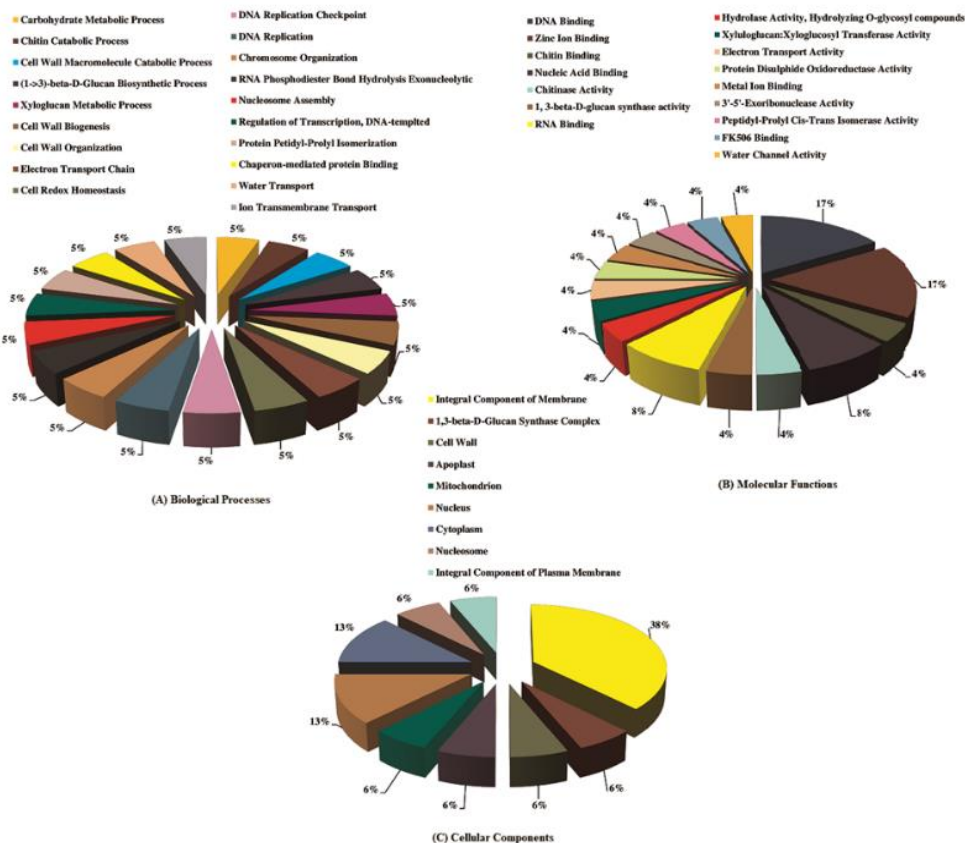
further analysis.

### 3.2 Sequence Assembly Analysis and Identification of SSR

After the retrieval  $\geq 1000$  bp sized 50 numbers of EST sequences of *Vinca rosea* the sequences were analyzed for assembly analysis through the program CAP3 which assemble the reads of EST sequences and predicts only 1 number of contigs and 47 numbers of singleton sequences from 50 numbers of EST sequences. The predicted contigs and singleton sequences had been subjected for identity of SSR containing sequences through GRAMENE database. Simple Sequence Repeat Identification Tool (SSRIT) is one of the tools of GRAMENE database for SSR identification, which results 48 numbers of SSR sequences i.e. out of 48 numbers of contigs and singleton sequences 20 numbers of sequences were contained the SSRs and termed as SSR-ESTs, based on the parameters like at maximum motif length was tetramer and minimum number of repeats was 4.

### 3.3 Functional Domain Marker analysis

All the SSR containing EST sequences i.e. SSR-ESTs were functionally annotated through Inter Pro Scan analysis. Using Inter Pro Scan, functional domains for 20 SSR-EST sequences, were identified from provided databases in Inter Pro Scan such as pattern scan, Signal PHMM, TMHMM, HMM Panther, and F Print Scan. The SSR-FDM provided information regarding the putative functions like Biological process, Molecular functions and Cellular components of transcribed genetic markers. Thus the functional analysis was examined 19 numbers of Biological processes (Fig. 1A), 16 numbers of Molecular Functions (Fig. 1B) and 9 numbers of Cellular components (Fig. 1C).



**Fig 1:** The Graphical Representation of Functional Domain Analysis of selected SSR-ESTs through InterProScan; ((A) Biological Processes, (B) Molecular Functions, (C) Cellular Components)

### 3.4 Primer Development

The resulted SSR-ESTs were analyzed for designing primer through Primer3 tool by followed these major parameters for primer pair development such as; minimum and maximum product sizes of 103–250 bp (optimal: 150 bp); primer length of 18–25 bases (optimal: 21 bases); GC content of 57.45% –

61.76% (optimal: 50%); annealing temperatures of 31.82 °C –60 °C (optimal: 56 °C); and default values for the other parameters. So, on the basis of these above criteria all the 20 numbers of EST-SSR sequences were gave 20 numbers of appropriate forward and reverse primer pairs through Primer3 (Table 1).

**Table 1:** List of Primers obtained from EST-SSRs through Primer3

Sequence ID	Forward Primer(left)	Tm (°C)	GC (%)	Reverse Primer(right)	Tm (°C)	GC (%)	Product Size(BP)
FD424855.1	TATTGCCCTGGT GATCCTTC	59.89	50.00	TGCCAAACCCT GTACAAACA	60.00	45.00	152
FD424718.1	TTGCAAGTGCTCT GGATGTC	59.99	50.00	ACCGATGCATTG CACAGTAA	60.14	45.00	201
FD424961.1	TTGCAAGTGCTCT GGATGTC	59.99	50.00	ACCGATGCATTG CACAGTAA	60.14	45.00	201
FD424473.1	GGGGCTATTGCCA TATCAGA	59.88	50.00	TGTGGAGTCATC CAAAACCA	59.94	45.00	135
FD424457.1	GCTTGGAGAGGT CACTGAGG	59.99	60.00	GCCTCAAGTCA AAGACACC	60.00	55.00	243
FD423945.1	TTGCCAAATCAACA GGCAGAG	59.99	45.00	GATGCCAAAA GGATGAGAA	60.01	45.00	153
FD423104.1	TGCCCTTGTGTG TGATTTA	59.96	45.00	ATTCCCCAAAGC TCGTTTTT	59.95	40.00	246
FD422546.1	TGGGGATGATCAT TTTCGTT	60.13	40.00	GCCTGGGGAGA AATTAGAGG	60.03	55.00	244
FD422543.1	TAAGTTTGGACAG CCAAGG	60.10	50.00	ATATATGGGGCA GGGAAAGG	60.00	50.00	216
FD422181.1	GATGGGCTCACGT TACGAGT	60.14	55.00	GTCAACGTACAA GGC GGATT	60.00	50.00	226
FD422057.1	AAGGGAACAAAG AGCAGCAA	59.99	45.00	GGAATTCCCATT CCTTCGTT	60.13	45.00	151
FD422231.1	CCAAGAATCAAA CCGAAGGA	60.04	45.00	ATGGGATCATGG AAAACCAA	59.99	40.00	178
FD421419.1	TGAATGGGGGAG CTATTTCA	60.41	45.00	GTCTCGCTTGTTT CCAACAT	60.12	50.00	150
FD421568.1	GATGGGCTCACGT TACGAGT	60.14	55.00	GTCAACGTACAA GGC GGATT	60.00	50.00	226
FD418491.1	TATACGGACGGCA AATGTCA	59.95	45.00	GATGGCATCTTG GGAGGTAA	59.89	50.00	175
FD419085.1	AGACTGGCCGAA GCAAAGTA	60.02	50.00	CTCGTAACAAG GGAGATGC	59.98	55.00	231
FD418645.1	AGATGGTGTGGT GGGAAAGC	59.93	50.00	AGGGCCCTCAATT GATTTTT	59.78	50.00	235
FD417519.1	CGACTGAAGCTGC ATCACAT	60.02	50.00	ACTGGGATCAGG CCTAAGGT	59.96	55.00	230
FD416626.1	CGTCAGCAGCTAC CAAATGA	60.01	50.00	CAACAGGACCCT TTCCTTCA	60.08	50.00	230
FD416624.1	CTTCCAATTGGGT TTGCTGT	59.97	45.00	AAGGATCCCCAA AGCCCTTA	59.98	45.00	231

### 4. Discussion

Based at the preceding reviews or studies records on the genetic and evolutionary diversity of decorative as well as medicinal plant *Vinca rosea*, a radical look at on the populace degree is required to assess the quantity of ultimate genetic resources and to inform management plans. So, the use of simple sequence repeat (SSR) markers derived from expressed sequence tags (ESTs) is beneficial as those are vital resources for gene discovery and mapping<sup>[8]</sup> and additionally for plant breeding. Thus as consistent with the located consequences there had been a total of 50 numbers of EST sequences out of 20,181 sequences from dbEST<sup>[9]</sup> and have been considered in this study based on the bp size ( $\geq 1000$ ) and subjected to assembly program through CAP3 assembler, as the program has a capability to clip 5' and 3' low-quality regions of reads, uses base quality values in computation of overlaps between reads, construction of multiple sequence alignments of reads, and generation of consensus sequences. This program also uses forward–reverse constraints to correct assembly errors and link contigs<sup>[10]</sup>. So, on the premise of all of the functions of CAP3 program, resulted 48 numbers of contigs and singletone sequences, which had been similarly analyzed beneath SSRIT tool of GRAMENE database as it gives critical precious information approximately all forms of markers<sup>[11]</sup> to screened out the SSR containing EST sequences, i.e. those sequences contains single sequence repeats (SSR) or microsatellites, are widely-used marker system in plant genetics and forensics and useful for primer design<sup>[12]</sup>; so, 20 numbers of EST-SSR sequences were identified based on the set motif length then analyzed for functional domain analysis through various databases in InterProScan<sup>[13]</sup>, which resulted biological process (all biological processes involved in equal proportion (Fig. 1A)), molecular functions (most frequently analyzed molecular

functions were DNA binding and Zinc ion binding (Fig. 1B)), and cellular components (mostly occurred cellular component was integral component of membrane (Fig. 1C)). Then this work was finished at the primer development via Primer3 by using the above defined parameters, which have the competencies to enhance the technique of primer layout in splendid quantity, because Primer3 software has been broadly used for primer layout, regularly in high-throughput genomics programs<sup>[14]</sup>. Finally the advanced or designed forward and reverse primer sets from 20 EST-SSRs believed to involve within the utility of PCR-primarily based technologies in gene expression and genetic range evaluation<sup>[15]</sup>. The predicted facts about the distribution of EST-SSRs within the species *Vinca rosea* will assist researchers to estimate genetic variety and populace structure within this species and may be beneficial for the development of a breeding method for *Vinca rosea*.

### 5. Conclusion

Development and application of molecular markers is of monstrous significance within the examination of the genetic composition, inter-species variability, and evolutionary relationships of *Vinca rosea*. This study demonstrates an approach to develop computationally mined 20 numbers of SSRs from ESTs, which may involve in extra studies to research the genetic courting or linkage mapping between the widely dispensed species of *Vinca rosea*. The functional annotation of SSR-ESTs might be used to provide best functional marker to study about the functionally variability between the plant species and at last the designed primers from SSR-ESTs is probably beneficial for PCR amplification for DNA polymorphism of related plant sorts in destiny factors.

## 6. Acknowledgement

We really thanks to Tectona Biotech Resource Centre (TBRC), Odisha, for providing necessary facilities, as well as we present our heartily gratitude to Dr. Shovan Kumar Mishra, Director of TBRC, for his endless support throughout this work.

## 7. References

1. Mishra JN, Verma NK. A brief study on *Catharanthus roseus*: A review. International Journal of Research in Pharmacy and Pharmaceutical Sciences. 2017; 2:20-23.
2. Rojas-Sandoval J. *Catharanthus roseus* (Madagascar periwinkle). Department of Botany-Smithsonian, NMNH, Washington DC, USA, 2015. <https://www.cabi.org/isc/datasheet/16884>
3. Renjini KR, Gopakumar G, Latha MS. The Medicinal Properties of Phytochemicals in *Catharanthus roseus* -A Review. European Journal of Pharmaceutical and Medical Research. 2017; 4(11):545-551.
4. Guo Q, Wang JX, Su LZ, Lv W, Sun YH, Li Y. Development and Evaluation of a Novel Set of EST-SSR Markers Based on Transcriptome Sequences of Black Locust (*Robinia pseudoacacia* L.). Genes. 2017; 8(7):177.
5. Radhika V, Aswath C, Reddy DCL, Shweta Bhardwaj A. In silico microsatellite development in arum lily (*Zantedeschia aethiopica*). Journal of Horticultural Science. 2011; 6(1):37-40.
6. Simko I. Development of EST-SSR Markers for the Study of Population Structure in Lettuce (*Lactuca sativa* L.). Journal of Heredity, 2008; 100(2):256-262.
7. Kumar B, Kumar U, Yadav HK. Identification of EST-SSRs and molecular diversity analysis in *Mentha piperita*. The Crop Journal. 2015; 3(4):335-342.
8. Mishra RK, Gangadhar BH, Yu JW, Kim DH, Park SW. Development and characterization of EST based SSR markers in Madagascar periwinkle (*Catharanthus roseus*) and their transferability in other medicinal plants. Plants Omics Journal. 2011; 4(3):154-16.
9. Parmar EK, Fougat RS, Patel CB, Zala HN, Patel MA, Patel SK *et al.* Validation of dbEST-SSRs and transferability of some other solanaceous species SSR in ashwagandha [*Withania Somnifera* (L.) Dunal]. 3 Biotech. 2015; 5(6):933-938.
10. Huang X, Madan A. CAP3 A DNA Sequence Assembly Program. Genome Research. 1999; 9(9):868-877.
11. Naithani S, Geniza M, Jaiswal P. Variant Effect Prediction Analysis Using Resources Available at Gramene Database. Methods in Molecular Biology Plant Genomics Databases, 2016, 279-297.
12. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017; 33(16):2583-2585.
13. Bhattacharyya U, Pandey SK, Dasgupta T. Identification of EST-SSRs and FDM in sesame (*Sesamum indicum* L.) through data mining. Scholarly Journal of Agricultural Science. 2014; 4(2):60-69.
14. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M *et al.* Primer3-new capabilities and interfaces. Nucleic Acids Research. 2012; 40(15):e115.
15. Yang X, Scheffler BE, Weston LA. Recent Developments in Primer Design for DNA Polymorphism and mRNA profiling in higher plants. BioMed Central. 2006; 2:4.