

RESEARCH ARTICLE

Development of EST-SSR markers based on transcriptome and its validation in ginger (*Zingiber officinale* Rosc.)

Venugopal Vidya¹, Duraisamy Prasath^{1*}, Mohandas Snigdha¹, Ramasamy Gobu¹, Charles Sona¹, Chandan Suravi Maiti²

1 ICAR-Indian Institute of Spices Research, Kozhikode, Kerala, India, **2** School of Agricultural Sciences and Rural Development (SASARD), Nagaland University, Nagaland, India

* dprasath@gmail.com



OPEN ACCESS

Citation: Vidya V, Prasath D, Snigdha M, Gobu R, Sona C, Maiti CS (2021) Development of EST-SSR markers based on transcriptome and its validation in ginger (*Zingiber officinale* Rosc.). PLoS ONE 16(10): e0259146. <https://doi.org/10.1371/journal.pone.0259146>

Editor: Sanghamitra Nayak, Siksha O Anusandhan University, INDIA

Received: July 8, 2021

Accepted: October 13, 2021

Published: October 27, 2021

Copyright: © 2021 Vidya et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This work was supported by grants from the Department of Biotechnology, Government of India (BT/PR24650/NER/95/794/2017). The funders had no role in the experimental design, data analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare that no competing interests exist.

Abstract

Ginger (*Zingiber officinale* Rosc.) is an economically important and valuable spice crop around the world. It is used as food, spice, condiment, and medicine. A considerable extent of genetic diversity in ginger occurs in the Western Ghats and North-Eastern India. However, genetic diversity studies at the molecular level in ginger is limited due to limited availability of genetic and genomic information. In the present study, for the first time, we have identified and validated expressed sequence tag (EST)-simple sequence repeat (SSR) markers from ginger. We obtained 16,790 EST-SSR loci from 78987 unigenes, and 4597 SSR loci in the predicted 76929 coding sequences from RNA-Seq assembled contigs of ginger through Illumina paired-end sequencing. Gene ontology results indicate that the unigenes with SSR loci participate in various biological processes such as metabolism, growth, and development in ginger. One hundred and twenty-five primer pairs were designed from unigenes and coding sequences. These primers were tested for PCR optimization, characterization, and amplification and identified 12 novel EST-SSR markers. Twelve flanking polymorphic EST-SSR primers were validated using 48 ginger genotypes representing North-Eastern India and different eco-geographical adaptations by PCR amplification and allele sizing through capillary electrophoresis. Twelve EST-SSR primers generated a total of 111 alleles with an average of 9.25 alleles per locus and allele sizes ranging between 115-189bp. This study implies that the SSR markers designed from transcriptome sequences provides ample EST-SSR resources, which are helpful for genetic diversity analysis of Zingiberaceae species and molecular verification of ginger genotypes.

Introduction

Ginger (*Zingiber officinale* Rosc.), an important spice and medicinal plant is known since antiquity. It belongs to the family Zingiberaceae and has various medicinal properties such as antioxidant, antimicrobial, anti-diabetic, and anti-inflammatory effects [1]. It is believed to be originated in South East Asia [2]. A wide array of diversity occurs in the Western Ghats, India [3, 4], and maximum variation within cultivated ginger occurs in North-Eastern India.

However, less genomic information and limited molecular markers availability in ginger hinders breeding and genetic studies.

To conserve the genetic resources and achieve improved crop productivity, it is essential to understand genetic diversity [5]. Due to the quick advancement in molecular biology, progress in DNA based markers was growing rapidly [6]. Microsatellites, also known as simple sequence repeats, are co-dominant DNA markers of short tandem repeats of 2 to 6 bp of nucleotides that are extensively present in whole genomes of various species. Microsatellites are of two types viz, genomic SSR and expressed-sequence-tag-based/ genic SSR (EST-SSR). Being co-dominant in nature, SSR markers have an edge over other markers, such as, high level of polymorphism, high specificity and repeatability [7]. Hence, SSR markers are widely used to study germplasm characterization, QTL mapping, linkage mapping and marker-assisted selection (MAS) [8].

With minimal molecular information available for ginger, molecular characterization was attempted earlier by using ISSR [9, 10], AFLP [11], RAPD [12] and SSR [13, 14] molecular markers. Since the EST-SSRs tightly linked to the genes controlling different traits can reveal local adaptation and environmental heterogeneity compared to other neutral markers [15, 16]. Besides that, it can be very efficiently transferred to other related species even if they are designed for a specific crop because of their location in conserved regions [17–19]. With improvements in next-generation sequencing (NGS) technology, such as low cost *de novo* transcriptome sequencing, has facilitated the identification of microsatellite loci [20]. The *de novo* assembly of transcriptomes is requisite for analysing functional genomics or marker development in non-model organisms, especially when there is a lack of a sequenced genome [20, 21].

Previously, we sequenced the transcriptome of ginger [22] using Illumina sequencing. In this study, we assembled 78987 unigenes, identified EST-SSR loci, designed primer pairs based on these data. The primary goal of this study was to develop EST-SSR markers from transcriptome sequence for the genetic diversity study among ginger genotypes and to preserve their diversity. In addition, these markers may be helpful in the ginger taxonomic, and evolutionary history research.

Materials and methods

De novo assembly and functional annotation of the transcriptome

The ginger transcriptome was sequenced earlier using Illumina sequencing [22]. The raw reads of this transcriptome sequence (PRJNA311170) were initially processed to ensure the accuracy of *de novo* assembly and subsequent analyses of SSR. Fast-QC (v0.11.7) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to control raw reads by filtering the empty reads, the adaptor sequences, and the sequences with low quality. Further filtering and quality control of the reads was done with Trimmomatic v.0.32 with default settings [23]. Finally, *de novo* assembly of the transcripts was completed using Trinity Assembler (<https://github.com/trinityrnaseq/trinityrnaseq/releases>) with default k-mer parameter (K = 25), and contigs were obtained by assembling overlap information between the sequences. After the assembly, unigene sequences were developed by CD-HIT (<http://weizhongli-lab.org/cd-hit/>) and using the unigene data coding regions were extracted using TransDecoder (<https://github.com/TransDecoder/TransDecoder/releases>).

The functional annotation of the non-redundant unigenes were done by Blastx against NCBI non redundant protein database and GO annotation and then plotted with functional classification using the BLAST2GO module of OmicsBox programme (www.biobam.com/omicsbox).

Development of EST-SSR markers

EST-SSR from the unigenes and CDS were mined using MicroSatellite identification tool (MISA, <http://pgrc.ipk-gatersleben.de/misa/misa.html>). The repeat sequence motifs included mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides with a minimum repeat number of 10, 6, 5, 5, 5, and 5, respectively. Batch Primer 3, an online web tool, was used to design SSR primer pairs from the flanking sequences of the identified microsatellite motifs [24]. Primers were designed considering the following conditions (a) primer length of 18 to 25 bp with 20 bp as the optimum; (b) PCR product size ranging from 100 to 300 bp; (c) melting temperature (T_m) between 55°C and 65°C with a difference of no greater than 5°C between the T_m values of the forward and reverse primers; and (d) GC content of 40% to 70% with an optimum of 50%. The designed primer quality was verified using NetPrimer (<http://www.premierbiosoft.com/netprimer/index.html>). For this study, a total of 125 primers were selected randomly and synthesized by Integrated DNA Technologies (IDT, Coralville, IA).

Plant material and DNA isolation

A total of 48 ginger genotypes, which includes 27 landraces of the North-Eastern States of India, 19 cultivated genotypes, one exotic accession, and *Zingiber zerumbet* (S1 Table) were collected from the National Active Germplasm Site of ginger at ICAR-IISR, Experimental Farm, Kozhikode, Kerala, India. The collected rhizomes were planted, and young disease-free leaves were collected from the 48 genotypes. Genomic DNA was isolated from the fresh leaf samples using Doyle and Doyle [25] with modifications such as double centrifugation with chloroform isoamyl alcohol (24:1) to get a clear supernatant free from polyphenols. The DNA was dissolved in 50 μ L of nuclease-free water. The quantity and quality of DNA samples were determined using the DeNovix DS-11 spectrophotometer (DeNovix, Wilmington, DE, USA) and by electrophoresis on 0.8% agarose gel, then diluted to 50 ng/ μ L and stored at -20°C until the PCR analysis.

EST-SSR markers amplification and validation

A set of 125 EST-SSR primers designed randomly to screen 48 ginger genotypes for validating SSR locus using PCR amplification (Agilent Surecycler 8800, USA). A total of 20 μ L PCR reaction mix contained 10 μ L of PCR master mix (Emerald Amp GT PCR, Takara), 0.5 μ L (10 μ M) each forward and reverse primers, 1 μ L template genomic DNA (50ng/ μ L), and 8 μ L nuclease-free water. PCR amplification was done under the following conditions: initial denaturation of template DNA at 94°C for 5 min followed by 35 cycles of 94°C for 45 s, 50 to 65°C (depending on the melting temperature of the primer pair used) for 45 s and 72°C for 1 min followed by a final extension at 72°C for 10 min. PCR products were separated on 3.5% agarose gel, stained with ethidium bromide, and gels were visualized using a gel documentation system (Syngene Gel Doc; Syngene, Synoptics Ltd, UK). After analyzing the agarose gel, 12 primers that showed polymorphism were further selected for capillary electrophoresis. The amplified PCR products of the selected primers were loaded into a QIAxcel capillary gel electrophoresis system (QIAGEN, Germany) to confirm SSRs in amplified genomic DNA fragments. The allelic sizes of each sample were measured in the form of virtual gel images and electropherogram peaks using QIAxcel Screengel Software (QIAGEN, v1.5).

Data processing and genetic analysis

Reproducible and consistent SSR bands were scored as present (1) or absent (0) into a binary matrix separately for the 48 ginger genotypes. The polymorphic information content (PIC) of

Table 1. Summary of assembly statistics of the ginger transcriptome data.

Total contigs	79049
Total transcripts	193666
GC (%)	44.7
N50	1236
Median contig length (bp)	629
Average contig length (bp)	912.83
Total assembled bases	176783582

<https://doi.org/10.1371/journal.pone.0259146.t001>

selected SSR primers was calculated by the formula: $PIC = 1 - \sum(P_i)^2$, where P_i is the frequency for the i^{th} microsatellite allele. The genetic similarity (GS) among the genotypes was calculated by Jaccard's similarity coefficients. The dendrogram was constructed through SAHN clustering method, and similarity coefficients generated using the unweighted pair group method with arithmetic mean (UPGMA) in NTSYS-pc version 2.01 [26]. Principal coordinate analysis (PCoA) based on the simple matching coefficient using eigen vector matrices was performed to confirm the grouping in the software GenAlex version 6.5.

Results

Identification of frequency and distribution of EST-SSRs

RNA sequences of the ginger reported in a previous study [22] were assembled to form contigs. A total number of 78987 contigs were generated from 145942 transcripts with an average contig length of 840.72 bp. The N50 value was 1,099 and the GC content was 44.79%. The median contig length was 574 bp. The predicted 78987 unigene contains both coding and non-coding sequences (Tables 1 and 2). TransDecoder identified long open reading frames (ORFs) within transcripts and scores them according to their sequence composition. The predicted coding regions were 76929.

A total of 16790 SSRs were identified using MISA software in 145942 sequences, with 1742 unigenes containing more than one SSR locus. Additionally, 710 SSRs were involved in compound formation. The SSR repeats units ranged from one to six, and the number of SSRs with each repeat motif varied widely. Among the SSRs, mono-nucleotide repeat motifs were found

Table 2. Frequency distribution of the 12 most frequent EST-SSR repeat motifs in ginger.

Repeats	5	6	7	8	9	10	11	12	13	14	15	>16	Total
A/T	-	-	-	-	-	4027	1749	890	610	395	266	527	8464
AG/CT	-	800	473	297	194	154	103	90	23	39	43	70	2286
AGG/CCT	599	283	167	92	20	22	1	5		2		0	1191
AAG/CTT	623	230	127	66	18	29	7	7	1	3		1	1112
CCG/CGG	457	196	74	31	21	6						0	785
AT/AT	-	228	107	73	50	25	23	10	4	8	1	1	530
AGC/CTG	304	128	52	32	1	1	5					0	523
AC/GT	-	145	66	44	21	23	13	10	8	2	2	8	342
ATC/ATG	170	36	13	25	3	12	1					0	260
AAC/GTT	106	69	22	11	1	6	1					0	216
ACG/CGT	100	38	33	16	8							0	195
ACC/GGT	97	51	15	5	2							0	170
AAT/ATT	74	32	18	18	4	1		2				0	149
C/G	-	-	-	-	-	46	19	8	8	5	2	2	90

<https://doi.org/10.1371/journal.pone.0259146.t002>

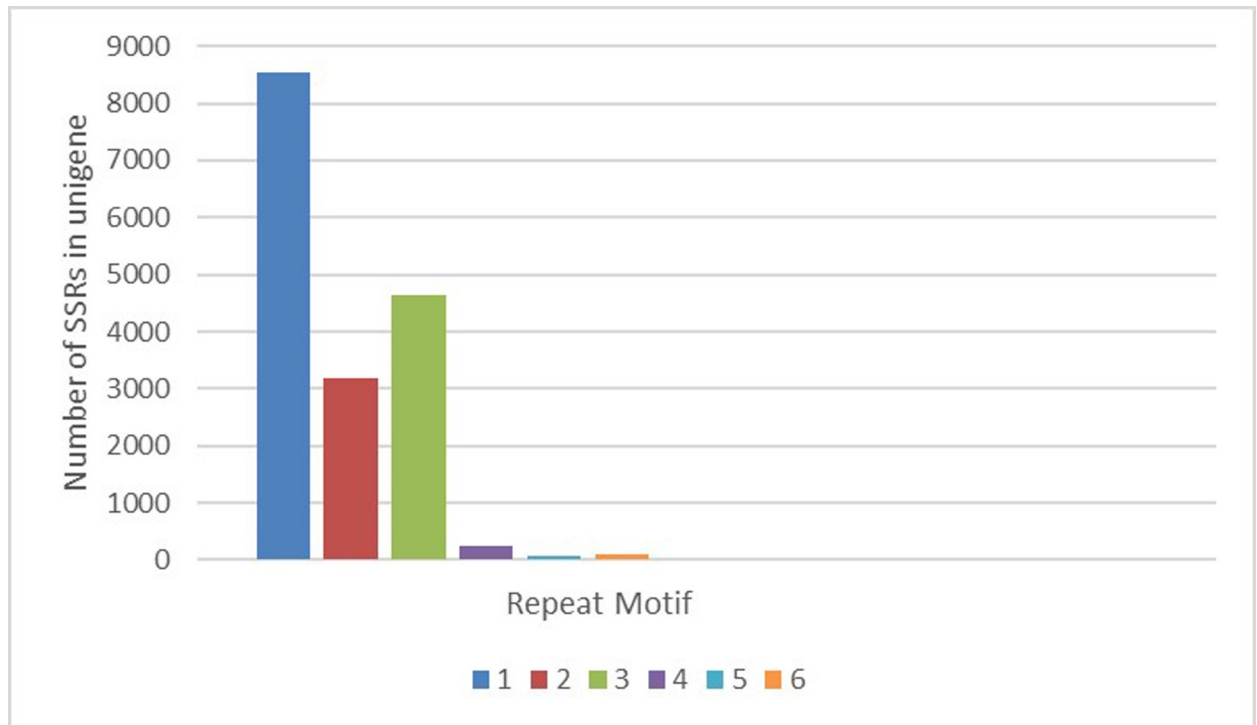


Fig 1. Distribution to different repeat type classes in unigenes.

<https://doi.org/10.1371/journal.pone.0259146.g001>

most abundant (8554; 50.94%), followed by tri- (4639; 27.62%), di- (3176; 18.91%), tetra- (252; 1.50%), hexa (101; 0.61%), and penta- (68; 0.40%) repeat motifs. Among the two types of molecular mononucleotide repeats (A/T) n was the most abundant when compared to (C/G) (Fig 1). While 4597 SSRs were identified in the predicted 76929 coding sequences, 334 sequences containing more than one SSR locus were identified. Additionally, 191 SSRs involved in compound formation. SSRs with trinucleotide repeat motifs were most abundant (3020; 65.6%), followed by mono (809; 17.6%), di- (666; 14.48%), hexa- (48; 1.04%), tetra (41; 0.89%), and penta- (13; 0.28%) repeat motifs (Fig 2).

Functional annotation and classification of unigenes. The unigenes were functionally annotated using BLAST2GO module of OmicsBox. 18222 unigenes were mapped to GO database. The sequence length of the annotated unigenes ranges from 301 to 14178 nucleotides. 10269 unigenes were mapped to known enzymes with EC numbers. 15965 unigenes had similarity measures >70%. The E-value distribution ranged from 0 to 9.85E-11. The annotated unigenes were grouped into three clusters: molecular function, biological process, and cellular component (Fig 3 and S2 Table). The cluster biological process consisted of maximum number of GO terms (1951) whereas cellular component consisted of least (506). Molecular Function component included 1407 terms. Majority of the unigenes in biological process were distributed in sphingosine and ceramide biosynthetic pathways. In molecular function, it was chiefly distributed in ATP binding, mRNA binding and zinc ion binding. A high proportion of unigenes were in the cellular component of membrane proteins.

Development and validation of the SSR markers

Only the SSRs with di-, tri-, tetra-, penta- and hexa-nucleotide repeats were considered as potential candidates for EST-SSR marker development. 125 primer pairs were designed

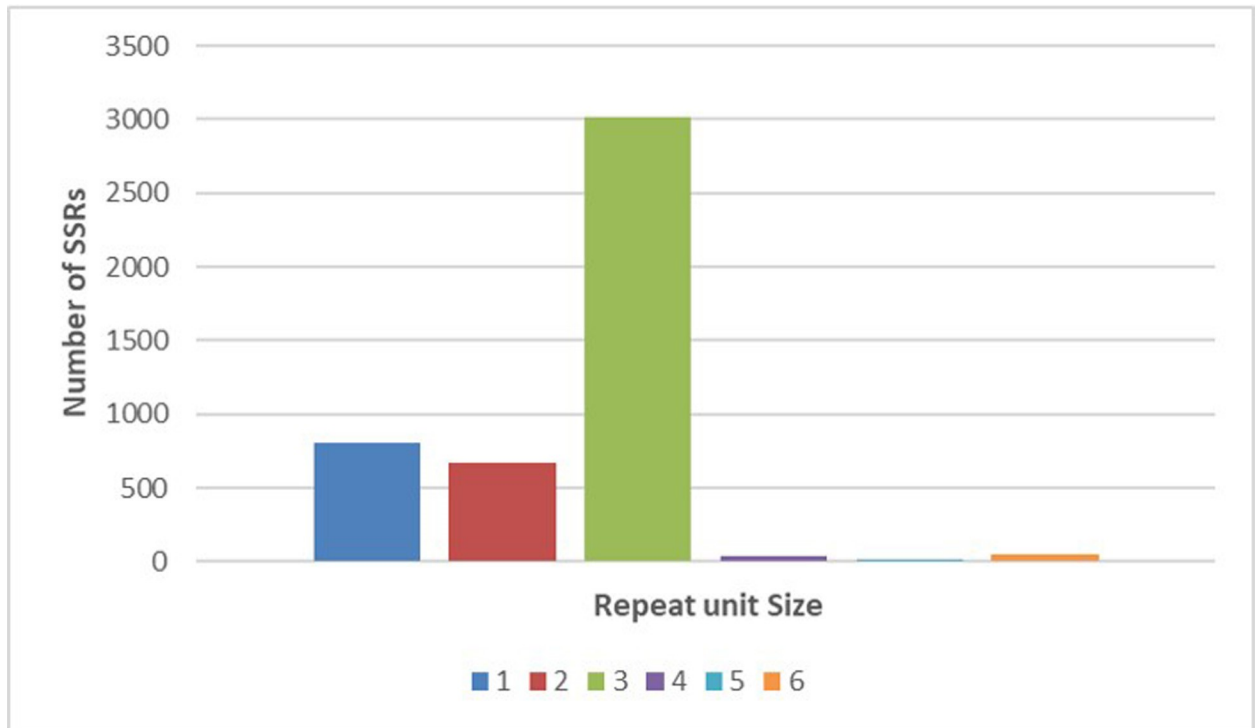


Fig 2. Distribution to different repeat type classes in CDS.

<https://doi.org/10.1371/journal.pone.0259146.g002>

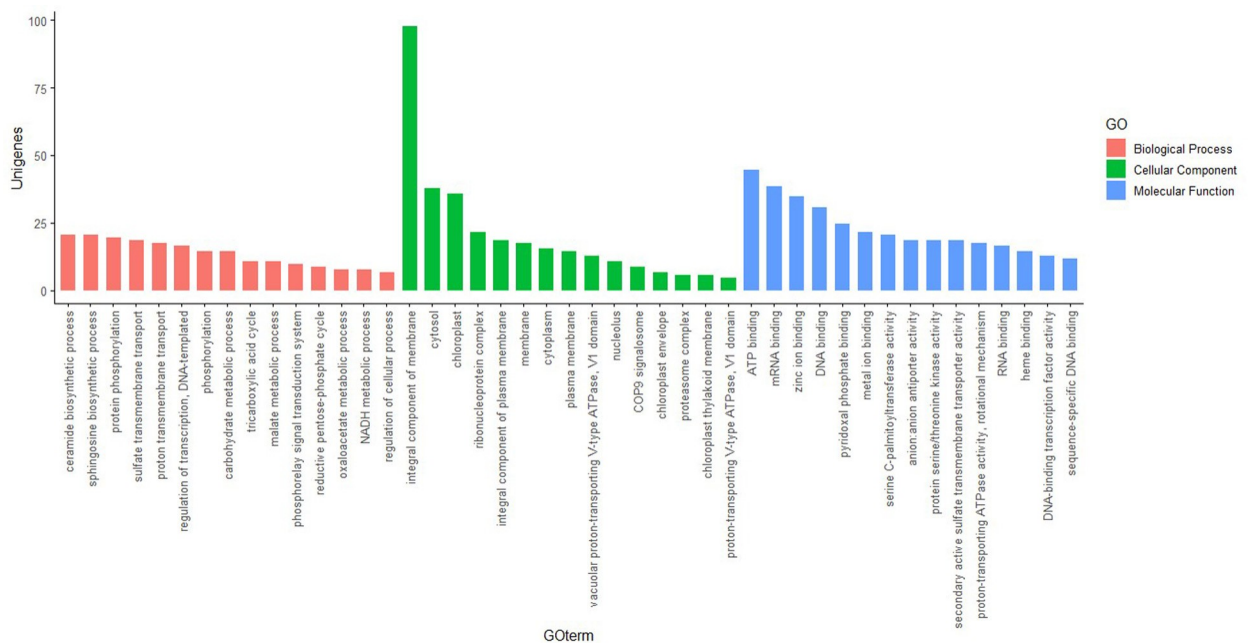


Fig 3. Gene Ontology classification of unigenes. Unigenes were assigned to three categories: cellular component, molecular function, and biological process.

<https://doi.org/10.1371/journal.pone.0259146.g003>

randomly from the unigenes, and CDS, and was tested for PCR optimization, characterization, and amplification. The validation of the designed EST-SSR primers was undertaken by PCR amplification and separation of the PCR products on a 3.5% agarose gel electrophoresis. Out of 125 primers, 92 primers produced amplicons of the expected size. However, the resolution of the EST-SSR primers was poor as the size of the bands observed on the agarose gel was uniform for most of the genotypes. To understand the allelic variation among the genotypes, we have selected 12 prominent EST SSR primers, which showed polymorphism between north-eastern genotypes, red ginger type genotypes and *Z. zerumbet* for capillary electrophoresis to study the allelic difference among the 48 genotypes. The characteristics associated with 12 polymorphic SSR markers are listed in Table 3. In total, 111 alleles were identified with an average of 9.25 alleles per locus varying from 5–15, which generated 1610 amplicons in total.

Table 3. Characteristics of 12 selected primers designed for analysing genetic diversity of ginger and PIC value.

Primers	Repeat motif	Sequence	Total alleles	Monomorphic amplicons	Polymorphic amplicons	Total number of amplicons	Allele size range (bp)	PIC value
ZO SSR2	(TTC)7	F-TGATTCGGTATCAACTCCAT R-CAAGGAAGACTTCAACTCCAT	10	0	10	138	117–154	0.81
ZO SSR25	(AAG)5	F-CTGAGTCCAGTGTGTATAGG R-GTTCTTGCTCGCTAGATCAC	10	0	10	66	143–178	0.81
ZO SSR16	(GTGATG)7	F-ATCAAGGAAAAGACCTCAAAG R-CATTATCATCGTCTTCTCTG	8	0	8	75	115–163	0.67
ZO SSR36	(GAC)8	F-GAGGACTACTCCGATATGAGC R-GGAGTTAGGGTTAGGGATTG	8	0	8	202	162–189	0.82
ZO SSR21	(CCT)6	F-TCCTCCTCTTTCTCCTCTC R-GCTAGAAGCGAGGGATTT	6	0	6	123	164–182	0.75
ZO SSR64	(GGC)6	F-TCCAGAGGCTCTCAGCTT R-ACTGCGACGACTCAGGTC	9	0	9	74	122–149	0.75
ZO SSR111	(TCT)10	F-CTAAGGGGCTCCTCTTTC R-CAGCTGGAAGCAGCTATG	11	0	11	123	124–160	0.89
ZO SSR108	(TTTA)5	F-GATCTCCTGCTTGTATCTCTC R-TGTTCTAGGTGTTGTGGAAG	8	0	8	164	117–145	0.8
ZO SSR35	(GAG)7	F-GGTCCAAGGTCTTTAAGCAT R-ACGAAGACAACGATATCAGC	12	0	12	186	129–195	0.83
ZO SSR38	(GAAGGC)5	F-GAAGGAGGCTCTCGAAGT R-GCACCTGCTTACAGTTACAAT	5	0	5	64	158–188	0.51
ZO SSR73	(CCT)5	F-GCTCTCCCTTCGAAAAC R-GCGTAGGTGCAGAAGTAGTTA	9	0	9	144	154–178	0.84
ZO SSR91	(GCA)7	F-CTCCATCCTATCAACTGTCCAC R-ACATTCTGAAGCTCTTGCAAT	15	0	15	251	126–181	0.9
TOTAL			111			1610		0.939
Average			9.25			134.1667		0.78

<https://doi.org/10.1371/journal.pone.0259146.t003>

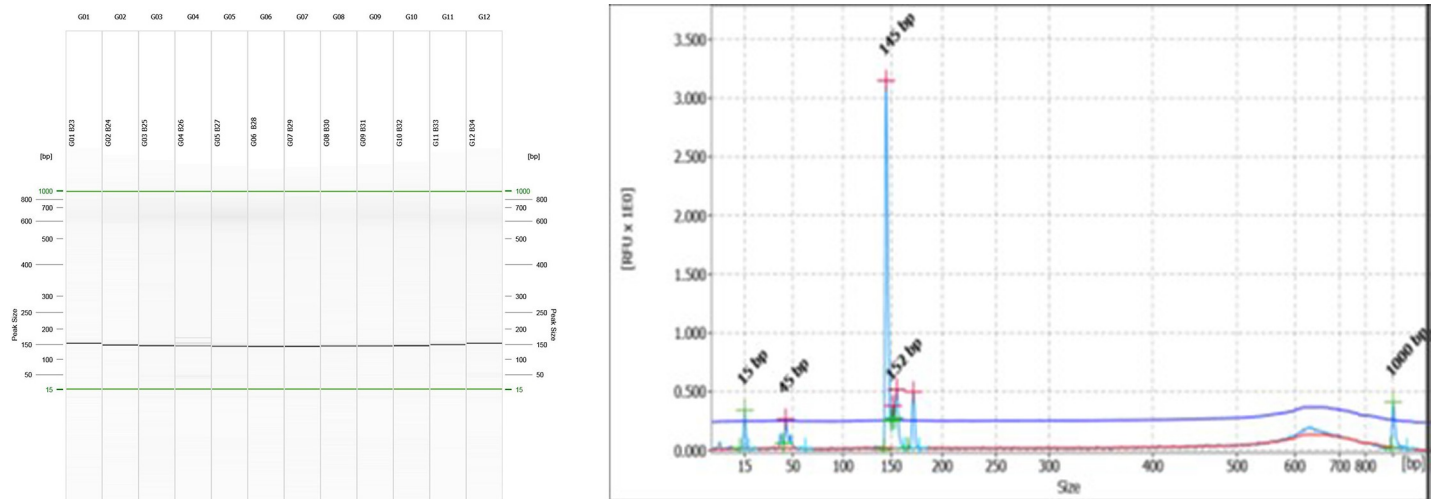


Fig 4. Validation of SSR markers in germplasm accessions of ginger. A. Gel image of PCR amplification of SSR marker ZOSSR25 on 48 germplasm accessions of ginger as captured on QIAxcel ScreenGel software. Numbers refer to accession numbers as indicated in S1 Table. The last lane is DNA molecular weight standard 50-1000bp v2.0 Qx DNA size marker. B. A representative electropherogram showing different allele sizes (sample G04 B26, variety Athira).

<https://doi.org/10.1371/journal.pone.0259146.g004>

In this study, the PIC values of the selected 12 EST-SSRs ranged from 0.51 to 0.90, with an average of 0.78. The lowest PIC value was observed in ZOSSR38 primer with 5 alleles, while the highest PIC value was obtained for ZOSSR91 with 15 alleles (Table 3 and S1 Fig). The ginger variety, Athira, showed a distinct polymorphic band with ZOSSR25, visible from agarose gel and confirmed with capillary electrophoresis (Fig 4).

Genetic diversity analysis

The Jaccard's similarity coefficients ranged between 0.07 and 0.84 (S3 Table). The least similarity coefficient of 0.07 was observed between 9079–9019 and *Z. zerumbet*–Sourabh genotypes, while the highest similarity coefficient 0.84 was observed between 9080–9042.

The UPGMA-based dendrogram placed all 48 genotypes into three main clusters at 15% similarity (Fig 5). Cluster I was the larger cluster and subdivided into four sub-clusters, cluster IA comprised of 10 genotypes (9015, 9062, Chitra, Nadia, 9061, Aswathy, Rio-de, 9019, 9058, Rejatha). Cluster IB comprised of 9021, Bhaise, 9070, Mahim, 9068, 9066, Gorubadhane, 9043, 9081, Exotic Red ginger, 9045, Mahima and Maran. Cluster IC includes 9030, Chandra, 9041, 9042, 9080, 9077, 9071, 9044, Varada, Himachal local, Karthika, Suruchi, Mohini, Suravi and Sourabh and cluster ID has Athira. Cluster II involves all the red ginger type genotypes of the North Eastern Region, India (9040, 9046, 9063, 9078, 9073, 9076, 9079). Cluster III contain *Z. zerumbet*. The genetic relationship among the genotypes was further studied using principal coordinate analysis (PCoA), according to the cluster analysis (Fig 6, S4 Table).

Discussion

Characterization of EST-SSRs in ginger

Sequence analysis in non-model organisms has been widely studied by *de novo* assembly of short reads of the transcriptome without a reference genome [27, 28]. Because of the high polymorphism, reproducing capacity, and codominance, microsatellite markers are in great demand and widely used in DNA fingerprinting, genetic diversity, population structure analysis, and marker-assisted crop breeding [29, 30]. However, the traditional way of SSR

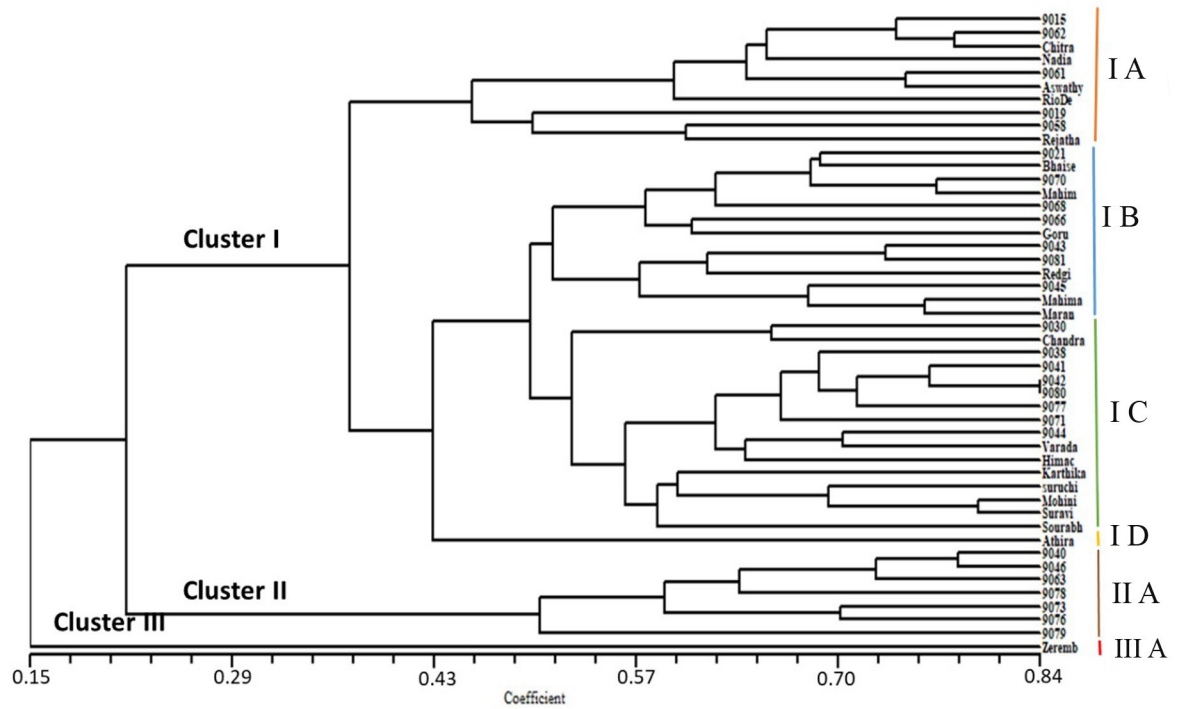


Fig 5. Genetic relationship of ginger accessions as revealed by SSR markers. Dendrogram of 48 ginger genotypes revealed by cluster analysis of genetic similarity estimates generated by Jaccard coefficient based on 12 EST-SSR markers.

<https://doi.org/10.1371/journal.pone.0259146.g005>

development is a very costly and time-consuming process and hence is not very efficient [31]. In this present study, we have used already available transcriptome data of ginger generated in our lab through Illumina paired-end RNA-seq technology [22]. N50 and an average length of

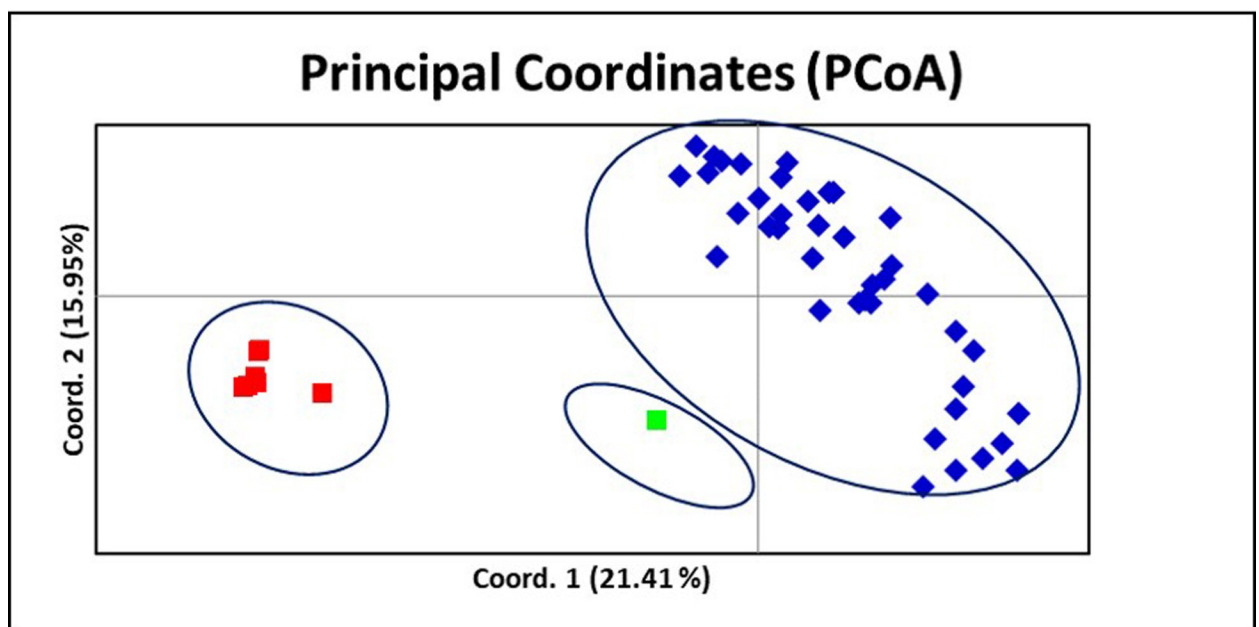


Fig 6. Principle component analysis (PCA) of ginger genotypes based on 12 polymorphic EST-SSR markers.

<https://doi.org/10.1371/journal.pone.0259146.g006>

all unigenes was 1,099 bp (Table 2), which is comparable with N50 reported from the *de novo* transcriptome assembly of *Curcuma longa* (N50 = 1,515 bp) [32]; *Zanthoxylum bungeanum* (N50 = 846 bp) [33], *Zantedeschia rehmannii* Engl. (N50 = 1,476 bp) [34], *Cicer arietinum* L. (N50 = 1,192 bp) [35] and *Ipomoea batatas* (N50 = 765 bp) [36]. The functional prediction was performed using Nr and GO protein databases. The unigenes were classified into 45 subcategories, including 15 functions each in molecular functions, biological process, and cellular component aspects (Fig 3). This result is similar with earlier studies on *Triticum aestivum*, *Curcuma alismatifolia* and *Curcuma longa* as the unigenes are classified into 47, 51 and 25 subcategories with GO database respectively [32, 37, 38]. We identified 16,790 EST-SSR loci from 145,942 unigenes and 4,597 SSR loci identified in the predicted 76,929 coding sequences. Trinucleotide repeats were the most frequent SSR repeats when excluding mono repeats followed by di and tetra repeats. Annadurai et al., (2013) [32] in turmeric and Wei et al., (2016) [34] in coloured calla lily reported di and tri-nucleotide motifs as most frequently occurred. In UTRs, a higher percentage of SSRs were mapped than the CDS region mostly due to the long-term natural selection which altered microsatellites that could easily lead to phenotype changes, and SSRs in the non-coding sequences could ensure the stability of germplasm resources [39]. Moreover, in general, EST-SSRs with tri-repeats remained most common among the monocot and dicots because open reading frames do not disturb the triplet codon with insertions and deletions within translated regions [29, 40, 41]. Among mononucleotide repeats, as in most plants, A/T repeats were far more abundant than G/C repeats [42]. Among the di-nucleotide repeats, AG/CT showed prevalence over other di repeats in contrast to Awasthi et al. (2017) [14] where maximum frequencies of TA and GA repeats and low frequencies of other di-nucleotide repeats were reported. The most abundant trinucleotide repeat motif in ginger was AGG/CCT, AAG/CTT and closely followed by CCG/CGG (12.94%); similar results were reported in *Curcuma alismatifolia* [38] and calla lily [34]. Besides, we noticed that, GC rich tri-repeats were more abundant than AT-rich repeats, supporting the fact that GC richness and consequent codon usage bias can be considered specific features of monocot genomes [38].

Genetic diversity in ginger

A total of 125 primer pairs were designed based on the ginger transcriptome sequences, and 92 primer pairs were successfully amplified. However, the remaining 33 primers did not produce amplicon in the ginger genotypes, indicating the primer either developed from an erroneously assembled transcript or the primer sequence exon-exon junctions [43]. In this study, 12 EST-SSR markers (Table 3) were obtained and verified, offering an informative and applicable approach for evaluating genetic relationships within and among the ginger genotypes. Besides, the PIC value is normally used to quantify polymorphism for a marker locus and is determined by both the number of alleles and their frequency distribution within the population [43]. All the ginger SSRs validated here showed a moderate level of informativeness (PIC > 0.5), with an average PIC value of 0.78.

The present study demonstrated the utility of the 12 newly developed polymorphic EST-SSR markers to evaluate genetic diversity among ginger genotypes. From the analysis results, the 12 markers divided the 48 genotypes into three main groups using the UPGMA cluster analysis. The dendrogram revealed that the North Eastern collections are highly diverse and falls in a different cluster along with other cultivated ginger varieties; also, red ginger type north-eastern collection fell into a single cluster and which is not clustered along with exotic Red ginger indicating the diverse nature of North Eastern red ginger collections as a separate cluster. Previous reports on EST-SSR in ginger focused on cross-species transferability [14, 44]; however, we report 12 novel EST-SSR which can be used for the diversity analysis of wide range of ginger germplasm and landraces.

Conclusion

The present work constitutes a considerable progress in identifying enormous number of informative SSR loci in ginger from transcriptome sequences. The expressed sequence tag (EST)–SSR markers will be helpful in the future investigation of genetic diversity, population structure, and phylogeography of the species. Furthermore, all the markers were successfully cross amplified in *Z. zerumbet* species, suggesting that they may also be used to study other related species in the Zingiberaceae family. Moreover, this study will aid in the development of conservation and management of existing ginger germplasm and endemic landraces unique for a particular geographical location.

Supporting information

S1 Fig. Gel images of SSRs with highest and lowest alleles. Gel image of PCR amplification of SSR marker ZOSSR38 with least alleles on 48 germplasm accessions of ginger as captured on QIAxcel ScreenGel software.

(TIF)

S2 Fig. Gel images of SSRs with highest and lowest alleles. Gel image of PCR amplification of SSR marker ZOSSR91 with highest alleles on 48 germplasm accessions of ginger as captured on QIAxcel ScreenGel software.

(TIF)

S1 Table. Ginger genotypes used for validation of EST-SSR markers and diversity analysis.

(XLSX)

S2 Table. Gene ontology annotation of unigenes.

(XLSX)

S3 Table. Jaccards similarity coefficient.

(XLSX)

S4 Table. PCoA analysis and Eigen values using GenALEx software.

(XLSX)

Author Contributions

Conceptualization: Venugopal Vidya, Duraisamy Prasath.

Data curation: Venugopal Vidya, Mohandas Snigdha, Ramasamy Gobu, Charles Sona.

Formal analysis: Venugopal Vidya.

Funding acquisition: Duraisamy Prasath.

Investigation: Venugopal Vidya, Ramasamy Gobu.

Project administration: Duraisamy Prasath.

Resources: Duraisamy Prasath.

Software: Charles Sona.

Supervision: Duraisamy Prasath.

Validation: Chandan Suravi Maiti.

Writing – original draft: Venugopal Vidya.

Writing – review & editing: Duraisamy Prasath, Mohandas Snigdha.

References

1. Ravindran P.N, Nirmal Babu K. "Ginger: The genus Zingiber". 2005. CRC press.
2. Burkill IH. A dictionary of the economic products of the Malay Peninsula. A Dictionary of the Economic Products of the Malay Peninsula. 1966;2.
3. Muralidharan VK, Velayudhan KC. A note on the occurrence of wild ginger in Western Ghats. South Indian Hort. 1983; 31: 259–260.
4. Sasikumar B, George KJ, Zachariah TJ. A note on a ginger (*Zingiber officinale* Rosc.) type collected from Western Ghats, South India. 1995.
5. Fu Y-B. Understanding crop genetic diversity under modern plant breeding. Theoretical and Applied Genetics. 2015; 128: 2131–2142. <https://doi.org/10.1007/s00122-015-2585-y> PMID: 26246331
6. Kumar B, Kumar U, Yadav HK. Identification of EST–SSRs and molecular diversity analysis in *Mentha piperita*. The Crop Journal. 2015; 3: 335–342.
7. Morgante M, Olivieri AM. PCR-amplified microsatellites as markers in plant genetics. The plant journal. 1993; 3: 175–182. PMID: 8401603
8. Tuler AC, Carrijo TT, N6ia LR, Ferreira A, Peixoto AL, da Silva Ferreira MF. SSR markers: a tool for species identification in Psidium (Myrtaceae). Molecular Biology Reports. 2015; 42: 1501–1513. <https://doi.org/10.1007/s11033-015-3927-1> PMID: 26476530
9. Kizhakkayil J, Sasikumar B. Genetic diversity analysis of ginger (*Zingiber officinale* Rosc.) germplasm based on RAPD and ISSR markers. Scientia horticulturae. 2010; 125: 73–76.
10. Baruah J, Pandey SK, Begum T, Sarma N, Paw M, Lal M. Molecular diversity assessed amongst high dry rhizome recovery Ginger germplasm (*Zingiber officinale* Roscoe) from NE-India using RAPD and ISSR markers. Industrial Crops and Products. 2019; 129: 463–471.
11. Kavitha PG, Kiran AG, Raj RD, Sabu M, Thomas G. Amplified fragment length polymorphism analyses unravel a striking difference in the intraspecific genetic diversity of four species of genus *Zingiber Boehm.* from the Western Ghats, South India. Current Science. 2010; 242–247.
12. Palai SK, Rout GR. Identification and genetic variation among eight varieties of ginger by using random amplified polymorphic DNA markers. Plant Biotechnology. 2007; 24: 417–420.
13. LEE S-Y, Fai WK, Zakaria M, Ibrahim H, Othman RY, GWAG J-G, et al. Characterization of polymorphic microsatellite markers, isolated from ginger (*Zingiber officinale* Rosc.). Molecular Ecology Notes. 2007; 7: 1009–1011.
14. Awasthi P, Singh A, Sheikh G, Mahajan V, Gupta AP, Gupta S, et al. Mining and characterization of EST-SSR markers for *Zingiber officinale* Roscoe with transferability to other species of Zingiberaceae. Physiology and Molecular Biology of Plants. 2017; 23: 925–931. <https://doi.org/10.1007/s12298-017-0472-5> PMID: 29158639
15. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant science. 2001; 160: 1115–1123. [https://doi.org/10.1016/s0168-9452\(01\)00365-x](https://doi.org/10.1016/s0168-9452(01)00365-x) PMID: 11337068
16. Varshney RK, Sigmund R, B6rner A, Korzun V, Stein N, Sorrells ME, et al. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Science. 2005; 168: 195–202.
17. Varshney RK, Graner A, Sorrells ME. Genomics-assisted breeding for crop improvement. Trends in plant science. 2005; 10: 621–630. <https://doi.org/10.1016/j.tplants.2005.10.004> PMID: 16290213
18. Fan L, Zhang M-Y, Liu Q-Z, Li L-T, Song Y, Wang L-F, et al. Transferability of newly developed pear SSR markers to other Rosaceae species. Plant Molecular Biology Reporter. 2013; 31: 1271–1282. <https://doi.org/10.1007/s11105-013-0586-z> PMID: 24415844
19. Guo R, Mao Y-R, Cai J-R, Wang J-Y, Wu J, Qiu Y-X. Characterization and cross-species transferability of EST–SSR markers developed from the transcriptome of *Dysosma versipellis* (Berberidaceae) and their application to population genetic studies. Molecular breeding. 2014; 34: 1733–1746.
20. Faure D, Joly D. Next-generation sequencing as a powerful motor for advances in the biological and environmental sciences. Genetica. 2015; 143: 129–132. <https://doi.org/10.1007/s10709-015-9831-8> PMID: 25736916
21. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nature biotechnology. 2012; 30: 434–439. <https://doi.org/10.1038/nbt.2198> PMID: 22522955
22. Prasath D, Karthika R, Habeeba NT, Suraby EJ, Rosana OB, Shaji A, et al. Comparison of the transcriptomes of ginger (*Zingiber officinale* Rosc.) and mango ginger (*Curcuma amada* Roxb.) in response to the bacterial wilt infection. PLoS One. 2014; 9: e99731. <https://doi.org/10.1371/journal.pone.0099731> PMID: 24940878

23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
24. You FM, Huo N, Gu YQ, Luo M, Ma Y, Hane D, et al. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC bioinformatics*. 2008; 9: 1–13. <https://doi.org/10.1186/1471-2105-9-1> PMID: 18173834
25. Doyle J, Doyle JL. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull*. 1987; 19: 11–15.
26. Rohlf FJ. NTSYS-PC, numerical taxonomy system for the PC ExeterSoftware, Version 2.1. Applied Biostatistics Inc Setauket, USA. 2000.
27. Russell J, Hackett C, Hedley P, Liu H, Milne L, Bayer M, et al. The use of genotyping by sequencing in blackcurrant (*Ribes nigrum*): developing high-resolution linkage maps in species without reference genome sequences. *Molecular breeding*. 2014; 33: 835–849.
28. Wei L, Li S, Liu S, He A, Wang D, Wang J, et al. Transcriptome analysis of *Houttuynia cordata* Thunb. by Illumina paired-end RNA sequencing and SSR marker discovery. *PLoS one*. 2014; 9: e84105. <https://doi.org/10.1371/journal.pone.0084105> PMID: 24392108
29. Wang H, Lei Y, Yan L, Wan L, Cai Y, Yang Z, et al. Development and validation of simple sequence repeat markers from *Arachis hypogaea* transcript sequences. *The Crop Journal*. 2018; 6: 172–180.
30. Shi X, Wu H, Li W, Guo W, Zheng Y, Yu S, et al. Development of polymorphic EST-SSR markers in *Itea chinensis* (Iteaceae) and cross-amplification in related species. *Applications in plant sciences*. 2018; 6: e1013. <https://doi.org/10.1002/aps3.1013> PMID: 29732244
31. Tang D-Q, Lu J-J, Fang W, Zhang S, Zhou M-B. Development, characterization and utilization of GenBank microsatellite markers in *Phyllostachys pubescens* and related species. *Molecular Breeding*. 2010; 25: 299–311.
32. Annadurai RS, Neethiraj R, Jayakumar V, Damodaran AC, Rao SN, Katta MA, et al. *De novo* transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PLoS one*. 2013; 8: e56217. <https://doi.org/10.1371/journal.pone.0056217> PMID: 23468859
33. Feng S, Zhao L, Liu Z, Liu Y, Yang T, Wei A. *De novo* transcriptome assembly of *Zanthoxylum bungeanum* using Illumina sequencing for evolutionary analysis and simple sequence repeat marker development. *Scientific reports*. 2017; 7: 1–11. <https://doi.org/10.1038/s41598-016-0028-x> PMID: 28127051
34. Wei Z, Sun Z, Cui B, Zhang Q, Xiong M, Wang X, et al. Transcriptome analysis of colored calla lily (*Zantedeschia rehmannii* Engl.) by Illumina sequencing: *de novo* assembly, annotation and EST-SSR marker development. *PeerJ*. 2016; 4: e2378. <https://doi.org/10.7717/peerj.2378> PMID: 27635342
35. Garg R, Patel RK, Tyagi AK, Jain M. *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA research*. 2011; 18: 53–63. <https://doi.org/10.1093/dnares/dsq028> PMID: 21217129
36. Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, et al. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC genomics*. 2010; 11: 1–14. <https://doi.org/10.1186/1471-2164-11-1> PMID: 20044946
37. Yang ZJ, Peng ZS, Yang H. Identification of novel and useful EST-SSR markers from *de novo* transcriptome sequence of wheat (*Triticum aestivum* L.). *Genetics and Molecular Research*. 2016; 15: 15017509.
38. Taheri S, Abdullah TL, Rafii MY, Harikrishna JA, Werbrouck SP, Teo CH, et al. *De novo* assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. *Scientific reports*. 2019; 9: 1–14. <https://doi.org/10.1038/s41598-018-37186-2> PMID: 30626917
39. Bhargava A, Fuentes FF. Mutational dynamics of microsatellites. *Molecular biotechnology*. 2010; 44: 250–266. <https://doi.org/10.1007/s12033-009-9230-4> PMID: 20012711
40. Kumpatla SP, Mukhopadhyay S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome*. 2005; 48: 985–998. <https://doi.org/10.1139/g05-060> PMID: 16391668
41. Sharifova S, Mehdiyeva S, Shahmuradov I. Mining and Survey of Simple Sequence Repeats in Expressed Sequence Tags of Tomato Species. *Albanian Journal of Agricultural Sciences*. 2016; 15: 65.
42. Gao Z, Wu J, Liu Z, Wang L, Ren H, Shu Q. Rapid microsatellite development for tree peony and its implications. *BMC genomics*. 2013; 14: 1–11. <https://doi.org/10.1186/1471-2164-14-1> PMID: 23323973
43. Guo X, Elston R. Linkage information content of polymorphic genetic markers. *Human heredity*. 1999; 49: 112–118. <https://doi.org/10.1159/000022855> PMID: 10077733
44. Sakthipriya M, Sabu KK. Development and cross-genera transferability of ginger EST-SSR markers for cardamom. *Current Bioinformatics*. 2018; 13: 95–99.