



OPEN ACCESS

EDITED BY

Mulatu Geleta,
Swedish University of Agricultural Sciences,
Sweden

REVIEWED BY

Rajwant Kalia,
Central Arid Zone Research Institute (ICAR),
India
Romit Seth,
North Carolina State University,
United States

*CORRESPONDENCE

Ambika Baldev Gaikwad
✉ ambika.gaikwad@icar.gov.in;
✉ ambikabg@gmail.com

RECEIVED 08 February 2023

ACCEPTED 05 April 2023

PUBLISHED 10 May 2023

CITATION

Gaikwad AB, Kumari R, Yadav S, Rangan P,
Wankhede DP and Bhat KV (2023) Small
cardamom genome: development and
utilization of microsatellite markers from a
draft genome sequence of *Elettaria
cardamomum* Maton.
Front. Plant Sci. 14:1161499.
doi: 10.3389/fpls.2023.1161499

COPYRIGHT

© 2023 Gaikwad, Kumari, Yadav, Rangan,
Wankhede and Bhat. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Small cardamom genome: development and utilization of microsatellite markers from a draft genome sequence of *Elettaria cardamomum* Maton

Ambika Baldev Gaikwad*, Ratna Kumari, Sheel Yadav,
Parimalan Rangan, Dhammaprakash Pandhari Wankhede
and KV Bhat

Division of Genomic Resources, Indian Council of Agriculture Research-National Bureau of Plant
Genetic Resources, New Delhi, India

Small cardamom (*Elettaria cardamomum* Maton), the queen of spices, is the third most expensive spice in the world after saffron and vanilla, valued highly for its aroma and taste. This perennial herbaceous plant is a native of coastal parts of Southern India and displays a significant amount of morphological diversity. Its genetic potential has not been exploited due to lack of genomic resources limiting our understanding of the genome and important metabolic pathways which give it the economic advantage in the spice industry. Here, we report upon the *de novo* assembled, draft whole genome sequence of cardamom variety, Njallani Green Gold. We used a hybrid assembly strategy using the reads from the Oxford Nanopore, Illumina and 10x Genomics GemCode sequencing chemistries. The assembled genome length was 1.06 Gb (gigabases) which is close to the estimated genome size of cardamom. More than 75% of the genome was captured in 8000 scaffolds with a N50 of 0.15 Mb. The genome appears to have a high repeat content and 68055 gene models were predicted. The genome is close to *Musa* species and displays an expansion and contraction in different gene families. The draft assembly was used for *in silico* mining of simple sequence repeats (SSRs). A total of 2,50,571 SSRs were identified of which 2,18,270 were perfect SSRs and 32,301 were compound SSRs. Among the perfect SSRs, trinucleotides were most abundant (1,25,329) and hexanucleotide repeats appear least (2,380). From the 2,50,571 SSRs mined, 2,27,808 primer pairs were designed based on flanking sequence information. Wet lab validation was performed for 246 SSR loci and based on their amplification profiles, 60 SSR markers were used for diversity analysis of a set of 60 diverse cardamom accessions. The average number of alleles detected per locus were 14.57 with a minimum of 4 and maximum of 30 alleles. Population structure analysis revealed the presence of high degree of admixtures which could primarily be due to cross-pollination prevalent in this species. The SSR markers identified would help in the development of gene or trait-linked markers which can be

subsequently used for marker-assisted breeding for crop improvement in cardamom. The information on utilization of the SSR loci for generation of markers has been developed into a public database, 'cardamomSSRdb' that is freely available for use by the cardamom community.

KEYWORDS

small cardamom, whole genome sequence (WGS), simple sequence repeats (SSRs), allele, genetic diversity, microsatellite marker database

Introduction

Small cardamom (*Elettaria cardamomum* Maton) is a monocot species belonging to the family Zingiberaceae with $2n=4x=48$. This species is believed to have originated in the rainforests of the Western Ghats of South India (Kuriakose et al., 2009; Hrideek et al., 2011; Nadukeri et al., 2020a, b). As it is a shade loving plant, it is mainly grown under canopies which provides suitable environment for plant growth. In India, it is majorly grown in the three southern states of Kerala, Karnataka and Tamil Nadu, with Kerala alone accounting for more than 58% of the annual production (Horticultural Statistics at a glance 2018). Owing to its pleasant taste and aroma, cardamom is commonly referred to as the 'Queen of spices'. The seeds and fruits of cardamom are the economically important parts and have found wide applicability across the culinary and traditional medicine spectrum (Verma et al., 2009; Kaliyaperumal et al., 2020). The medicinal properties of cardamom are attributed to the essential oil which is extracted from its fruits. This oil possesses antibacterial, anti-inflammatory and antispasmodic activities (al-Zuhair et al., 1996; Sobral et al., 2014; Alam et al., 2021).

The Zingiberaceae family comprises of more than 1500 species (Li et al., 2021). Most species of this family are valued as natural sources of spices, perfumes, herbal medicines, dyes, etc (Kumar et al., 2013). The molecular studies for species delineation within the family have so far relied on the use of the nuclear internal transcribed spacer (ITS) based loci and traditional chloroplast markers (Kress et al., 2002; Ngamriabsakul et al., 2000). The availability of chloroplast genomes of eighteen different species belonging to four different genera of Zingiberaceae is a valuable resource in order to understand the evolutionary dynamics within the family (Li et al., 2021). However, the lack of whole genome sequences (WGS) limits marker identification and utilization. So far, the WGS information is available for only two Zingiberaceae species i.e., *Curcuma longa* (Chakraborty et al., 2021; Yin et al., 2022) and *Curcuma alismatifolia* (Dong et al., 2022). The scarcity of genomic resources in this family creates a bottleneck for carrying out future studies which are aimed at crop improvement. Therefore, the generation of comprehensive genomic resources is of paramount importance for deriving insights into species characterization and phylogenetic diversity. In small cardamom, RAPD (Random Amplified Polymorphic DNA) markers have been used previously for molecular characterization of genotypes (Radhakrishnan and Mohanan, 2005). ISSR (Inter Simple Sequence

Repeat) markers have been also been used to assess the genetic variability among accessions which included wild collections, landraces, feral and released varieties (Jose et al., 2014; Anjali et al., 2016). The development of other marker systems which are sequence based will expand the scope and reliability for conducting genetic diversity analysis. Simple sequence repeats (SSR) or microsatellite markers, are one of the most widely used markers for genetic diversity analysis (Cavagnaro et al., 2010; Wang et al., 2010; Feng et al., 2016; Ali et al., 2019; Kumari et al., 2019; Kapoor et al., 2020; Li et al., 2022). Their relative abundance, multi-allelic nature, co-dominance, high levels of reproducibility and ease with which these can be scored, are some of the reasons for their widespread utility across a wide range a species. Furthermore, they are present across both the coding and non-coding regions of the genome and can be used for cross-species amplification in related species and genera. In cardamom, previously, SSR markers have been developed using selective hybridization enrichment method and these were subsequently used for diversity and cross transferability studies (Cyriac et al., 2016). However, these were very few in number (140). Also, a few EST-SSRs (200 in number) have been developed using the EST sequences from *Curcuma longa* and used for diversity analysis in cardamom (Anjali et al., 2015). It is important to identify a comprehensive set of SSR markers which would provide genome-wide coverage. This would facilitate comparative and functional genomics studies for cardamom in the future.

Here, we report the WGS of one of the most popular, high-yielding varieties of cardamom, Njallani Green Gold. The genome has been assembled using long and short reads and an exhaustive number of SSRs were identified. A select set was then used to genotype a collection of accessions for the purpose of validation and subsequently a database was developed for ease of access and utilization. This will serve as a rich source of genomic resources in small cardamom.

Materials and methods

Genome sequencing and assembly

Total genomic DNA was isolated from the leaves of the popular farmer's variety 'Njallani Green Gold' (Figure 1) using the CTAB extraction method (Doyle and Doyle, 1987). The purified DNA was checked on 0.8% agarose gel and quantified on NanoDrop (DS-11 spectrophotometer, DeNovix, Wilmington, Delaware). Illumina

(paired-end and 10X chromium) and Oxford Nanopore genomic libraries were prepared as per the manufacture's protocol and sequenced on Illumina HiSeq X Ten sequencer and Oxford Nanopore P24 PromethION sequencing platforms, respectively. The genome size was evaluated by k-mer (k=16) distribution analysis with Jellyfish using the Illumina paired-end reads (105.11 Gb). The Bioinformatics pipeline was as follows. The Nanopore data correction was carried out using Canu (version 1.6). The corrected Nanopore data was further used for *de novo* assembly using Wtdbg2 (<https://github.com/ruanjue/wtdbg2>). This assembly was further polished using Pilon (Walker et al., 2014). BWA (version 0.7.17) Mem algorithm was used to map the Illumina data on assembly obtained using Wtdbg2. The 10X chromium data along with Nanopore assembly was used by ARCS (<https://github.com/bcgsc/arcs>) for further scaffolding.

Repeat masking

Tandem repeat finder was utilized to identify tandem repeats in polished genome. In brief, *de novo* repeat library was generated

using RepeatModeler version 2.0.1 (<http://www.repeatmasker.org/RepeatModeler/>) to predict transposable elements in the unannotated genome assemblies. Three, *de novo* repeat-finding programs, RECON version 1.08 (Bao and Eddy, 2002), RepeatScout version 1.0.5 and LtrHarvest (genometools-1.5.9)/Ltr_retriever (version 2.9.0) were employed to identify the boundaries of repetitive elements and to build consensus models of interspersed repeats. RepeatMasker version 4.1.0 (<https://www.repeatmasker.org/>) was used to annotate DNA sequences for interspersed repeats and identification of low complexity DNA sequences.

BUSCO genome assessment

We selected Benchmarking Universal Single-Copy Orthologs (BUSCOs; version 3, http://busco.ezlab.org/embryophyta_odb10) to assess the completeness of the assembled genome against *Arabidopsis thaliana* species with 1375 Plant BUSCOs.



FIGURE 1
The flowers and capsules of small cardamom (*Elettaria cardamomum* Maton).

Gene prediction

Seqping (<https://sourceforge.net/projects/seqping/>, version 0.1.45.1) was used for gene prediction of the masked sequence. It is an automated pipeline that performs gene prediction using self trained HMM (hidden Markov model) models and transcriptomic data. The program processes the genome and transcriptome sequences of a target species through Glimmer HMM, SNAP, and AUGUSTUS training pipeline that ends with the program MAKER2 combining the predictions from the three models in association with the transcriptomic evidence. The pipeline generated species-specific HMMs and is able to predict genes that are not biased to other model organisms. Genes were clustered using CDHIT v4.6 (<http://cd-hit.org>) software at a sequence similarity of 90%.

Gene annotation

Genes were mapped against NR database (updated May 2020) using blastx (NCBI-2.2.29+) and Plant Transcription factor database (<http://plantfdb.cbi.pku.edu.cn/>, version 5.0). Gene annotation was done using Uniprot database and KEGG (Kyoto Encyclopedia of Genes and Genomes) database (<https://www.genome.jp/kegg/pathway.html>). GO visualisation was done using WEGO version 2.0.

Phylogenetic analysis

BUSCO was used for phylogenomics. BUSCO was done for all the 14 species to predict conserved genes. *Amborella trichopoda*, *Ananas comosus*, *Arabidopsis thaliana*, *Cocos nucifera*, *Elettaria cardamomum*, *Glycine max*, *Musa acuminata*, *Musa balbisiana*, *Oryza sativa*, *Phoenix dactylifera*, *Populus deltoides*, *Solanum lycopersicum*, *Sorghum bicolor* and *Vitis vinifera*. Complete BUSCO genes were extracted from all the genomes. Alignment of the peptide sequences of genes was done using MUSCLE (version 3.8.31) program and analysed using MEGAX (<https://www.megasoftware.net/>). UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method was used for constructing phylogenetic tree.

In silico mining of microsatellite repeats

The assembled genome sequence was searched for presence of different microsatellite (di to hexa nucleotide) simple repeats as well as complex repeats following the default parameter of MISA -MISatellite identification tool (<http://pgrc.ipk-gatersleben.de/misa/>). The SSRs were identified from the draft genome using MISA perl scripts (Beier et al., 2017). The search criteria included minimum of six repeats of dinucleotides, minimum five repeats for trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides.

Development and validation of SSR primers

The SSR primers were designed from the flanking sequences of SSRs identified in the genomic sequences of cardamom by using software Primer 3 (Rozen and Skaletsky, 2000; Untergasser et al., 2012). Primers were designed for 2,27,808 of the 2,50,571 SSRs identified. Out of the 2,27,808 SSR primers, 246 primers were synthesized for wet lab validation (Supplementary Table 1). PCR was carried out in a total volume of 20 μ l comprising of 1X PCR buffer, 2.5mM MgCl₂, 1 μ M primer, 0.2mM of each dNTPs, 1U Taq DNA polymerase (NEB) and 15 ng template DNA in thermal cycler (Eppendorf). The amplification conditions were as follows: Initial denaturation at 95°C for 5 min followed by 35 cycles of denaturation at 95°C for 1 min, annealing at 44-52°C for 1 min and extension at 72°C for 1 min followed by final extension at 72°C for 10 mins.

Genetic diversity analysis

The plant material included 60 accessions (Table 1) of *Elettaria cardamomum* collected from Karnataka and Kerala (Nadukeri et al., 2020b). DNA was isolated from the leaves using the CTAB extraction method and quantified on the Nanodrop spectrophotometer (DS-11 spectrophotometer, DeNovix, Wilmington, Delaware). Genetic diversity was analysed by amplification at 60 SSR loci (Table 2). The amplified products were resolved on QIAxcel multi-capillary system using QIAxcel High Resolution Kit 1200 (QIAGEN, No. 929002, New Delhi, Qiagen India Pvt. Ltd.), 50-800 bp v2.0 Qx DNA size marker (QIAGEN, No. 929561) and 15bp/1000 bp QX alignment marker (QIAGEN, No. 929521). PCR products were separated with high resolution run method OM700 with a sample injection time of 10 seconds. The allelic sizes of each sample were resolved and calculated in the form of gel profiles and peaks using QIAxcel ScreenGel software (QIAGEN, v1.5). The SSR amplification products were scored across the lanes according to their molecular weight. The polymorphism information content (PIC) for each pair of SSR primers was calculated using the formula:

$$PIC_j = 1 - \sum_{i=1}^n P_i^2$$

where i is i -th allele of the j -th marker, n is the number of the j -th marker's alleles, P is the allele frequency (Botstein et al., 1980). The data was analyzed using Polysat package of R (Clark and Jasieniuk, 2011) choosing Bruvo distance (Bruvo et al., 2004) for generating distance matrix and dendrogram was generated using hierarchical clustering.

Population structure

The SSR allelic data on 60 accessions was used to run the software STRUCTURE version 2.3.4 (Pritchard et al., 2000). In order to determine the correct K value, the run parameters were set

TABLE 1 List of cardamom accessions used for genetic diversity analysis (as per Nadukeri et al., 2020b).

S.No.	Genotype	Origin of genotype	Morphological attributes
1	CL692	ZAHRS Mudigere	Prostrate panicle, oval/oblong in shape, bold, good yield
2	CL688	ZAHRS Mudigere	Prostrate panicle, bold capsules
3	Wayanad	Wayanad, Kerala	Thick stem from the base
4	CL671	ZAHRS Mudigere	Prostrate panicle
5	CL622	ZAHRS Mudigere	Medium bold capsules with pale green colour
6	D751	Mudigere, Karnataka	Medium bold capsules with pale green colour
7	PV4	Pampadumpara	Prostrate panicle, forms a part of the Walayar collection
8	D168	Mudigere, Karnataka	Semi-erect panicle
9	PDP4	Pampadumpara	Erect panicle and medium plant height
10	PDP14	Pampadumpara	Erect panicle and medium plant height
11	Pink Pseudostem	Myladumpara	Prostrate panicle and pink coloured pseudo stem
12	CL73	ZAHRS, Mudigere	Prostrate panicle, short span of flowering
13	M1	ZAHRS, Mudigere(Local check)	Prostrate panicle, tolerant to thrips and shoot borer, pubescent leaves
14	M2	ZAHRS, Mudigere (Local check)	Prostrate panicle, Clonal selection from CL-683
15	Dharmada	Sri Lanka	Erect panicle, green colour capsule
16	Ceylone-5	Periyar Tiger reserve, Kerala	Erect panicle and good plant height
17	P10	Pampadumpara	Round capsules with medium plant height
18	PDP12	Pampadumpara	Semi-erect panicle, late and long flowering span
19	K5	Kalmane, Mudigere, Karnataka	Semi-erect panicle, late flowering
20	D516	Mudigere, Karnataka	Semi-erect panicle, round oblong capsules
21	KMRD2	Kammardi, KoppaTaluk, Karnataka	Semi-erect panicle, green colour capsules with good plant height
22	D11	Mudigere, Karnataka	Semi-erect panicle, green capsules
23	D140	Mudigere, Karnataka	Semi-erect, green capsules
24	KMRD10	Kammardi, KoppaTaluk, Karnataka	Semi-erect, green colour capsules with good plant height
25	P1	Pampadumpara	Short panicle, early maturing variety with slightly ribbed light green capsules.
26	CL679	ZAHRS Mudigere	Bold capsules, panicle length is good
27	GreenGold	Vandanmedu, Idukki	Semi-erect panicle, dark green coloured bold capsules, robust in nature, perform well under intensive cultivation
28	HS1	Handi selection, Mudigere	Prostrate panicle, drought tolerant and good yield, round capsules
29	MCC61	Myladumpara	Erect panicle, bold capsules and produce more tillers
30	MCC12	Myladumpara	Erect panicle, terminal flowering
31	MCC21	Myladumpara	Erect panicle, bold capsules and produce more tillers
32	CSS-800	IISR, Appangala	Semi-erect panicle, bold, dark deep green capsule
33	CSS-872	IISR, Appangala	High yielding and dark green capsules
34	ICRI-2	Myladumpara	Erect panicle, clonal selection from MCC-61
35	SKP21	ICRI, Sakleshpur	Prostrate panicle, early and short span of flowering

(Continued)

TABLE 1 Continued

S.No.	Genotype	Origin of genotype	Morphological attributes
36	SKP72	ICRI, Sakleshpur	Prostrate panicle, early and short span of flowering, good plant height
37	CL-698	ZAHRS, Mudigere	Bold capsules, panicle length is good.
38	MCC18	Myladumpara	Bold capsules, number of bearing suckers are more
39	CL-668	ZAHRS Mudigere	Medium bold capsules with green in colour
40	CCS800	IISR, Appangala	High yielding and dark green capsules
41	SEL-98	ZAHRS, Mudigere	Bold capsules, panicle length is good.
42	10-6-D10	ZAHRS, Mudigere	Superior for number of bearing suckers and high yielding
43	2-5-D11	ZAHRS, Mudigere	Superior for number of bearing suckers, high yield and medium bold capsules
44	12-7-D11	ZAHRS, Mudigere	Good plant height with more number of bearing suckers
45	26-16-D11	ZAHRS, Mudigere	Medium oval shape capsules
46	29-9-D11	ZAHRS, Mudigere	Medium bold oval shape capsules
47	S-1	Pampadumpara	Semi-erect panicle
48	MHC-73	Myladumpara	Medium plant height, with pale yellow capsules
49	MHC-26	Myladumpara	Erect panicle, oval shape capsules
50	MCC-309	Myladumpara	Erect to semi-erect panicle, oval shape capsules
51	APG-284	IISR, Appangala	Good yield, bold capsule and dark green colour leaves
52	APG-293	IISR, Appangala	Good yield and bold capsule
53	MHC-10	Myladumpara	Prostrate panicle, medium plant height with pale yellow capsules
54	SKP-170	ICRI, Sakleshpur	Semi-erect panicle, with round to oblong capsules
55	MHC-18	Myladumpara	Erect panicle, a hybrid of the germplasm collections, deep green, angular bold capsules, performance is fairly good even under rainfed conditions
56	MHC-13	Myladumpara	Prostrate panicle, medium plant height, with green to pale yellow capsules
57	SKP-165	ICRI, Sakleshpur	Good plant height with more number of bearing suckers
58	PS-44	Pampadumpara	Bold capsules, good panicle length.
59	MCC-200	Myladumpara	Early flowering type
60	CCS-1	IISR, Appangala	Maximum bearing and bold capsules

as Length of burn-in 10,000, Number of MCMC after burn-ins at 50000 and number of iterations at 20. The result data was used to estimate the delta (k) values and these were plotted against the K values to obtain a biplot to determine the correct K as 5 following Evanno method (Earl and vonHoldt, 2012). Following inference of correct K as 5, the analysis was done to determine the allelic affinities of each putative panicle type groups using the parameters, Length of burn-in 500,000, Number of MCMC after burn-ins at 7,50,000 and number of iterations at 20. The results presented are based on this final analysis. Based on panicle type, cardamom can be classified into the Malabar (prostrate), Mysore (erect) and Vazhukka (semi-erect or intermediate) types. Unambiguous data on panicle type was available for only 39 of the 60 accessions used in this study (Table 1). Accordingly, there were 15, 10 and 14 accessions under the three panicle type groups, Malabar (prostrate), Mysore (erect) and Vazhukka (semi-erect),

respectively and hence only these 39 accessions were used for the analyses of population sub-structure.

Development of microsatellite marker database

Elettaria cardamomum microsatellite database (CardamomSSRdb) is an interactive and relational online database that contains comprehensive information on small cardamom genomic SSRs that were identified using Misa script and primers designed using Primer 3 software. It is based on “three tier level schema architecture and organization” with client, server and database. PHP has been used to design the dynamic, interactive and user-friendly interface of the database and My SQL server for storing the genomic SSR data in tabulated form. Other than

TABLE 2 List of 60 SSR primer pairs used for diversity analysis in 60 accessions of cardamom.

S. No.	Primer ID	Repeat type	Forward primer sequence	Reverse primer sequence	No. of alleles	Allele size (bp)
1	EC1	(TG)20	CTTTGGTTGTATAGGAAATG	TATGCTGCTAACTCTTTCTC	5	338
2	EC4	(TG)18	ATAAATTAGAATTCGCTGTG	TATGGCCTTGTAATAGATGT	4	225
3	EC5	(TG)17	ACTTGTCCGTGATACTTATG	AGAAGCAAAGGAAGAATATC	7	315
4	EC7	(TG)16	CTTGGATGTAGTTTACAAT	CTCTCAATGCTAATTCCTAA	4	300
5	EC8	(TG)15	ATACAAACCATTAAAACCAA	AACAATATTGGCACATAGTC	4	123
6	EC12	(TG)14	TGTAATATCCTTTTCTGGA	TGTC AAGGATACTTTCTCAC	7	241
7	EC14	(TG)14	ATTCATTTGAACATAAGCAC	AGCATGATATTACACCAAAC	8	269
8	EC24	(TC)18	TAAGTTTGTATTTTGAGGG	CTCTACATCGTCTATGCAAT	11	304
9	EC25	(TC)17	GGAGTGATAAAACACTACCA	CTCGTATTGTTTCAGAAAAG	7	240
10	EC28	(TC)17	GTTCTCGTAATTTCAATGTC	CCAATCAAACCTTCATAAAAC	9	273
11	EC29	(TC)16	ATACAGAGTAGGCAATTGAA	GCTTGAGAATACTACCAAAA	6	349
12	EC33	(TC)15	ATATTAATGTTTGCTAACCG	AAATCACAAAGCAAATAAG	8	340
13	EC34	(TC)15	ATTGACTATTTTCATGTTTGC	TAAGAAACCAAAATAGATGG	13	289
14	EC35	(TC)14	ATTTATACTATCGCTGTTGC	GAAACGAGAAAGAGGAAG	7	206
15	EC37	(TC)14	TGACAGCGTATGTTTATTTA	AAATCATATTCATTCAAACC	14	313
16	EC41	(TC)14	CAAGTTACTTCTTCGGTTTA	CCAGCTTTTGAGATTAGATA	8	274
17	EC43	(TC)14	TTTCTAGTTTCATTATCTCCG	AGTTCACTAGCTATCTCCCT	17	308
18	EC46	(TC)14	TATGGTTAATTGCATTCTCT	TGTTCTAACATCTTGAATCC	30	298
19	EC47	(TC)14	TAGATGAAGACTGAGAATGC	CGACAAAAGAAAGTACAGAC	18	271
20	EC48	(TC)14	CCAAACTTTTTATTATCCAC	ACATTGAAGTGTGTGAGAT	8	176
21	EC49	(TC)14	TGAACTTCAATTTCTTGTTT	TTTGTTTATCGGACTTCTTA	13	333
22	EC83	(GT)17	GTACATATGGGTTGATTGAC	ACAAGCTAATGATCTTGGTA	11	306
23	EC86	(GT)13	GCATGATACCACTAACAAAT	AGAGTGTGCAAAAATAAAG	16	208
24	EC89	(GA)18	GAAAATCAGTGGATTGATAA	AGATTTTGGAAATCTCTCTT	26	342
25	EC90	(GA)18	GTGCTACGCTGTACTTTTAT	TTAAGGAATAGCTGAAATTG	19	341
26	EC92	(GA)17	TCTAATGATAGCTTGATCGT	AGTAGCTGATACAGAATTGC	26	219
27	EC93	(GA)16	AAATGCTTCCATATAACTCA	TAATTGCAAAGAGGAATAAG	16	296
28	EC95	(GA)16	TCTTCTTTCTCCCTAGAAT	TGTTTCAATCAAATACATCA	13	253
29	EC96	(GA)16	AAGCCCATCTAATTAAGTT	CTGTCCCTTGTTACTCTTG	12	336
30	EC102	(GA)15	GATATGATGATCCTTTTGTG	TGTTCTACTATTTAGCTGCC	9	138
31	EC104	(GA)14	CTACTTGGCTACCCTGTT	CTGCCGTAGAAGTAAATAAG	12	258
32	EC106	(GA)14	AACTTCTTGTGATGAAAAA	CTGTTAGAAGTATTCGCATT	19	260
33	EC124	(CT)17	GTGCAGAAAAGAACAATAAT	AATACCAAAGCTCATCTACA	24	272
34	EC130	(CT)15	ATTATAGTAATTACCGCTGC	CTTTATCGATGACAAGGTC	15	314
35	EC132	(CT)15	GTGGTGTTTTCGAGAGTAA	AAGAAGAAGAAGGACAAAAT	20	311
36	EC140	(CT)14	CATTTAGGATAGTCTTTTGC	CTGACATTAACGAATCAAT	15	120
37	EC141	(CT)14	GTAAATGTCAATTGGAAAAG	CTTCTAACGTGACAAAAACT	20	173
38	EC142	(CT)14	AACAAAATGAGTCAAACAAC	AACATGAGAAATGACAGAAG	20	199

(Continued)

TABLE 2 Continued

S. No.	Primer ID	Repeat type	Forward primer sequence	Reverse primer sequence	No. of alleles	Allele size (bp)
39	EC145	(CT)14	TTTCTCACTGAGGACTAGC	CTGGTAATAATGGGTAATTG	15	306
40	EC168	(AG)18	ACTAAACGAGTGATACCTGA	CTCTCCCTCAATTTCTTAAT	17	190
41	EC169	(AG)17	CCACGACTTTAATCTACTTG	AGAAACCAAGAAAAGAAGCTC	26	299
42	EC172	(AG)17	ATCAAAGCATACTCTCTCAA	AGAGATTTTCTTCTGTTTT	15	207
43	EC181	(AG)15	ATGAATGGTGTTTAGAGTTG	TCAATGATGCATAACAATTA	26	327
44	EC182	(AG)15	ATATACTAGTGAACCACGGA	CAAAATCTCTCCTTTTCTCT	16	313
45	EC183	(AG)15	TAATGTAATCATCAATGCAA	TGAGTTTTACTGTCCAGAAC	22	158
46	EC187	(AG)14	GAATCTGTATGATGCTTGT	TGTATCCGTTCAAATAACTT	24	298
47	EC192	(AG)14	TGTTCTAGAACATAACAT	TTGCTTCCTAAAATCTATTG	22	322
48	EC193	(AG)14	CAAAAACGACCTAGTTAAAG	GACGATCAATTAACCTCTA	17	222
49	EC195	(AG)14	ATCAAATAAATGGAATGAAA	AAATTCAGGTCTTACCG	12	339
50	EC196	(AG)14	AAATATTA AAAACACGTTTCG	TTCAGACTTTGAACTCAACT	11	186
51	EC197	(AG)14	TTAATTCTCCATTGTCTAA	AATTACATTATGGCTTGAG	15	280
52	EC212	(AC)18	GAATAACCATCTGTCAGAAA	ACAAAACGATAGAAGAAATTG	9	347
53	EC232	(TCT)14	ACCAATTGAAATGTGATTAG	CGAAAATACITATTGTCCAC	11	244
54	EC234	(GAA)14	GAGAATTGCCITTTGTAATA	CTTTCCACATAGTTCAGTGT	18	195
55	EC235	(GAA)13	GCTATTGTACGGAAGATAAAA	AGATCAAACGCTGTAATTT	18	289
56	EC236	(CTT)15	AGATTGAACACTAATTGGAA	GAAGAAGATGATGATGAGAA	15	305
57	EC240	(AGA)13	CCAATTCACITGGTTATATT	ATTAATGAAGATGAAGCAGA	17	269
58	EC242	(AAT)11	CCTAAGTGAGAATAGCTTCA	AATTATCTCTCTCCAGAT	17	240
59	EC243	(AAG)13	AACAAAAGGGAGAAAGTTAT	TGATCTACCAGTATTTCCAC	17	257
60	EC245	(AAG)11	TACCTTCGTA CTCCAATAG	GTAGATGAAATCCAGTACA	13	282

genomic SSR data and statistics, the database also contains step by step pictorial tutorial to facilitate hassle free usage by the user.

Results and discussion

Genome sequencing and assembly

The cardamom genome was sequenced using a hybrid sequencing approach which included the long read chemistry

from the Oxford Nanopore, Illumina short read data and 10x Genomics GemCode linked read data (Table 3). Similar strategy involving the use of more than one sequencing chemistries has been employed for whole genome sequencing of *Curcuma longa* and *Curcuma alismatifolia*, both belonging to the Zingiberaceae family (Chakraborty et al., 2021; Dong et al., 2022; Yin et al., 2022) and also in other species (Hu et al., 2019; Wang et al., 2019). The Nanopore sequencing generated a total data of 141.90 Gb. The average read length obtained was 28 Kb. After data correction, the reads were

TABLE 3 Metrics of Nanopore, Illumina and 10 X GemCode Sequencing data.

Nanopore		Illumina		10 X GemCode	
Number of reads	13964122	Number of reads	696152122	Number of reads	612562452
Total data	141.90 Gb	Total data	105.11 Gb	Total data	92.4 Gb
Mean Read Length	28 Kb	Read Length	151 bp	Read Length	151 bp
Largest Read Length	969 Kb	Number of contigs	13145	Number of scaffolds	8000
Number of contigs	30638	Minimum contig length	12736 bp		
Minimum contig length	3419 bp				

assembled *de novo* and an assembled sequence of 1.06 Gb was generated. Based on K-mer depth distribution analysis the total length of the genome was estimated to be ~ 1.3 Gb (Supplementary Table 2). A genome sequencing depth of 108.46 x with more than 75% of the genome was captured in the assembled sequence. The assembled genome was polished using the Illumina paired-end (PE) data. The Illumina sequencing chemistry generated a total of 105.11 Gb data. The assembly of the reads resulted in generation of 30,000 contigs. The minimum contig length increased more than four times from 3419 bp to 12736 bp after polishing. The 10 X GemCode sequencing data along with the Nanopore *de novo* assembly was used for scaffolding. The final assembly consisted of 8000 scaffolds with a scaffold N50 of 0.15 Mb (Table 4). The assembled genome length was 1.06 Gb with 39.36% GC content. Previously, the genome size for cardamom has been estimated to be 1.4 Gb based on flow-cytometry analysis with a 2C nuclear DNA of 2.87 pg (Anjali et al., 2016). Our assembled genome length of 1.06 Gb reveals that we were able to sequence > 75% of the predicted genome length. The draft genome assembled in the present study can provide a base to build on for any future refinements in the genome assembly in terms of reducing the gaps present. The assembly quality in terms of completeness was estimated using BUSCO (Benchmarking Universal Single-Copy Orthologs) and more than 72% of the genes were found to be complete (C) with more than 99% single-copy genes, on analysis with 1,375 total BUSCO groups for *Arabidopsis thaliana* (Supplementary Table 3).

Genome annotation

Genome annotation led to identification of 68,055 gene models. The non-protein coding genes included 478 tRNA genes and 1852 other small RNA genes. The repetitive sequences constituted 71% of the assembled genome (Table 5), with retrotransposons contributing to as high as 47% of the repeats. Amongst the retrotransposons the LTR retrotransposons accounted for 46% of the assembled sequence with ~ 36% of Copia and ~ 9% of Gypsy elements. The tandem repeats in the form of satellites (satDNA)

TABLE 4 Cardamom genome assembly and annotation.

Assembly features	
Length of genome	1.06 Gb
Number of scaffolds	8000
N50	0.15 Mb
Longest scaffold	3.9 Mb
Smallest scaffold	33.3 Kb
Mean length	0.13 Mb
GC content (%)	39.36
Number of genes	68,055
Number of tRNA genes	478
Number of sRNA genes	1852

TABLE 5 Organization of repetitive sequences in the cardamom genome.

Repeats	Length	In genome (%)	Number of elements
Total interspersed repeats	759115910	71.15	–
Retrotransposons	504982486	47.33	324440
DNA transposons	10371139	0.97	11829

constituted 0.11% of the genome with 1223 satellite elements identified.

A total of 38,400 genes were annotated (Supplementary Table 4) and 14,731 genes were mapped to the Plant Transcription factor database (Supplementary Table 5). The enrichment analysis of the genes identified in the cardamom genome was performed through the Gene Ontology (GO) and KEGG enrichment analysis. The GO analysis revealed that the genes were enriched in the GO terms of cellular component, molecular function and biological process. Amongst the GO category of cellular component, the GO term membrane (GO: 0016020) was most abundant. Similarly for the GO category of molecular function, the GO term, GO: 16390, binding was most abundant. For biological process GO category, the GO terms cellular process (GO: 0009987) and metabolic process (GO: 0008152) were most represented (Supplementary Table 6 and Supplementary Figure 1). The KEGG enrichment analysis revealed an enrichment of genes involved in metabolic pathways and biosynthesis of secondary metabolites (Supplementary Table 7 and Supplementary Figure 2). The species belonging to the Zingiberaceae family are known to be rich sources of secondary metabolites which have found immense utility in field of traditional medicine, identification of genes involved in biosynthesis of these compounds would help in understanding the genetic basis of secondary metabolite production in cardamom (Vairappan et al., 2013; Ghasemzadeh et al., 2016).

Comparative genomic analysis

In order to gain an insight into cardamom evolution, we compared the genome of cardamom with other species. Highest similarities were observed with the wild banana species, *Musa acuminata* with 7,265 orthologous genes identified, followed by the banana species, *Musa balbisiana* with 4,996 orthologous genes (Figure 2A and Supplementary Table 8). A total of 3,708 gene families were identified in cardamom. The conservation of these gene families was examined amongst four different species that included *Musa acuminata* (Family: Musaceae), *Oryza sativa* (Family: Poaceae), *Phoenix dactylifera* (Family: Arecaceae) and *Ananas comosus* (Family: Bromeliaceae). It was observed that highest number of gene families, 31% (1179) were conserved between cardamom and *Musa acuminata*. This is in agreement with the taxonomic classification, as both *Musa acuminata* and cardamom belong to the same order of Zingiberales (D'Hont et al., 2012). A total of 732 gene families were conserved amongst all the five species with 2,466 gene families specific to cardamom (Figure 2B).

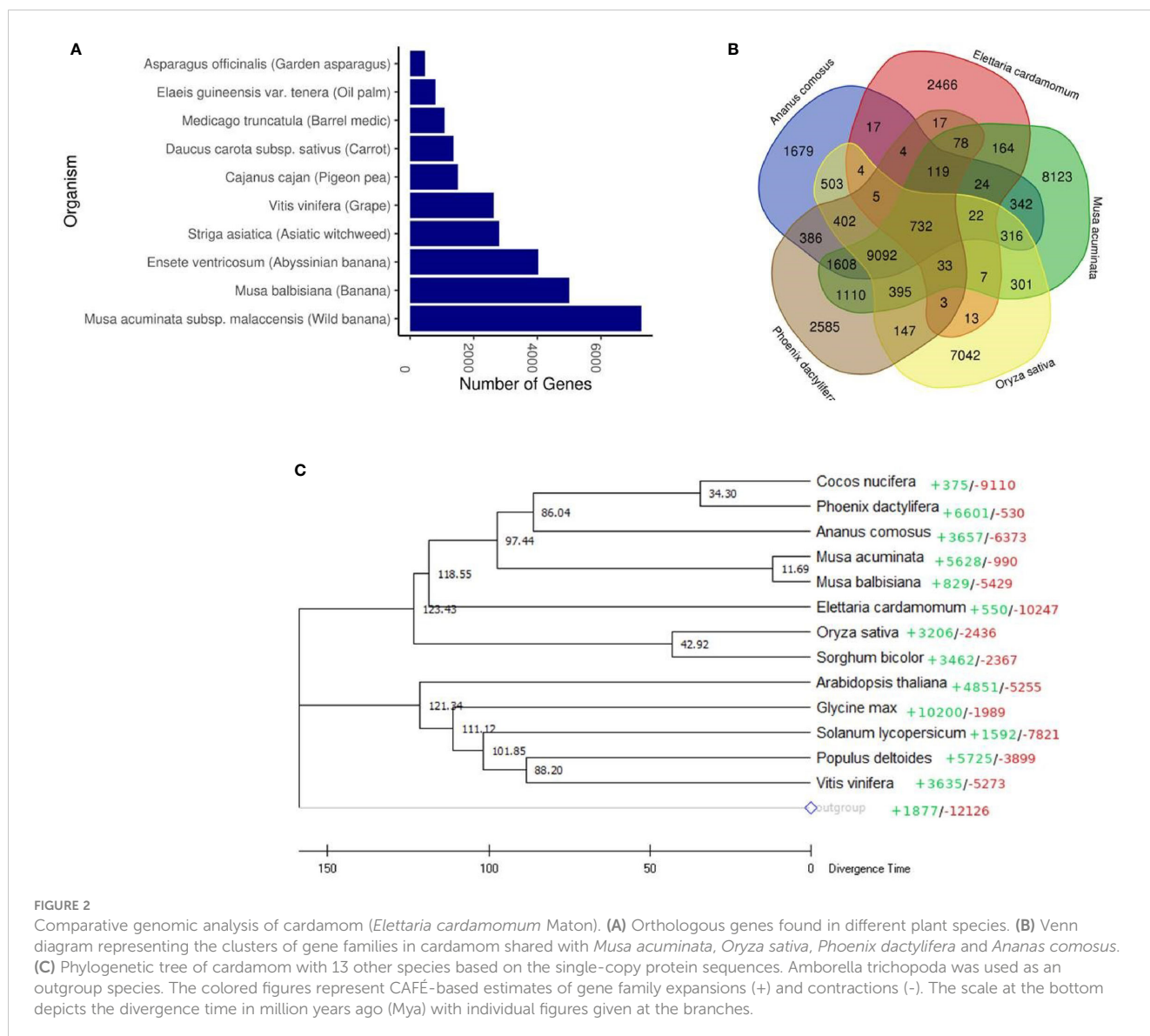


FIGURE 2 Comparative genomic analysis of cardamom (*Elettaria cardamomum* Maton). **(A)** Orthologous genes found in different plant species. **(B)** Venn diagram representing the clusters of gene families in cardamom shared with *Musa acuminata*, *Oryza sativa*, *Phoenix dactylifera* and *Ananas comosus*. **(C)** Phylogenetic tree of cardamom with 13 other species based on the single-copy protein sequences. *Amborella trichopoda* was used as an outgroup species. The colored figures represent CAFE-based estimates of gene family expansions (+) and contractions (-). The scale at the bottom depicts the divergence time in million years ago (Mya) with individual figures given at the branches.

A phylogenetic tree was generated based on the single-copy protein sequences identified in cardamom and thirteen other species. These species were selected based on their taxonomic classification with representation from diverse taxonomic groups. Amongst the monocotyledons (Commelinids), we selected *Sorghum bicolor*, *Oryza sativa*, *Ananas comosus*, *Elettaria cardamomum*, *Musa acuminata*, *Musa balbisiana*, *Phoenix dactylifera*, *Cocos nucifera* and amongst the eudicots, *Arabidopsis thaliana*, *Glycine max*, *Solanum lycopersicum*, *Populus deltoides* and *Vitis vinifera* were selected (Figure 2C). The basal angiosperm *Amborella trichopoda* was used as an outgroup species. The phylogenetic analysis clearly separated the Commelinids from the eudicots. It was observed that the genome of cardamom was most similar to the genomes of *Musa acuminata* and *Musa balbisiana*. This is in corroboration with the previous studies where it has been observed that the Zingiberaceae species like *C. longa*, *Z. officinale* and *C. alismatifolia* are phylogenetically closer to the Musaceae family species, *M. acuminata* and *Musa balbisiana* (D’Hont et al.,

2012; Dong et al., 2022). The time of divergence of Musaceae from the Zingiberaceae (*Elettaria cardamomum*) was estimated to be ~80 Mya (million years ago). Additionally, we observed that the Arecales (*Phoenix dactylifera*, *Cocos nucifera*) are more closely related to Zingiberales (including *Musa*) than to Poales, as has been reported previously (D’Hont et al., 2012; Yin et al., 2022). However, amongst the poales, *Ananas comosus* was observed to be placed closer to Arecales than Poales. The earlier divergence of *Ananas comosus* from the Poales might be a reason for this observation (Qian and Jin 2016). Another plausible reason for this could be the genome assembly of pineapple used in the reported analysis. The assembly used is the MD2 v1 assembly (Ming et al., 2015) which is not at the chromosome level and has lesser number of genes predicted and annotated, than the newer version of the MD2 assembly (Yow et al., 2021). The CAFE based analysis of gene family expansion and contraction revealed that 550 gene families show expansion in cardamom and 10247 gene families were contracted.

Identification and characterization of SSRs from the cardamom genome

The SSRs (dinucleotides to hexanucleotides) were identified from the assembled contigs. A total 2,50,571 SSRs were identified. Among these, 2,18,270 (87%) were perfect and 32,301 (12%) were compound SSRs. Among the perfect SSRs, trinucleotides were most abundant (1,25,329), followed by dinucleotides (68,831), tetranucleotides (16,794) and pentanucleotides (4936). The hexanucleotide repeats were least in number (2,380) (Figure 3). The repeat motif (AT/TA) was highest in number among the dinucleotides (Figure 4A) and the repeat motifs (AAG/AGA/GAA/CTT/TTC/TCT) were the most abundant among trinucleotides (Figure 4B). Our results are in agreement with the previous studies where dinucleotide repeats have been reported as the most abundant repeat type in crops like *Oryza sativa* (McCouch et al., 2002), mung bean (Tangphatsornruang et al., 2009), cranberry (Zhu et al., 2012), pigeonpea (Dutta et al., 2011), black alder (Lepais and Bacles, 2011), maqui (Bastías et al., 2016) and black pepper (Kumari et al., 2019). In some crops trinucleotides were the most abundant type of SSRs like in *Glycine max* (Xin et al., 2012), *Brachypodium* (Sonah et al., 2011), foxtail millet (Zhang et al., 2014) and watermelon (Zhu et al., 2016). In cotton, however hexanucleotides were the most abundant SSRs in *Gossypium hirsutum* and pentanucleotides in *G. raimondii* genome (Wang et al., 2015). The differences may also be due to use of different software programs and also the input parameters used for identification of SSRs and the completeness of the genome assemblies used for prediction.

Development and validation of SSR markers

SSR markers play an important role in genetic diversity analysis, gene tagging, mapping and marker assisted selection. In earlier studies, the SSR markers from related genera like *Amomum subulatum*, *Zingiber officinale* and *Curcuma longa* were tested for transferability in cardamom (Anjali et al., 2015; Cyriac et al., 2015;

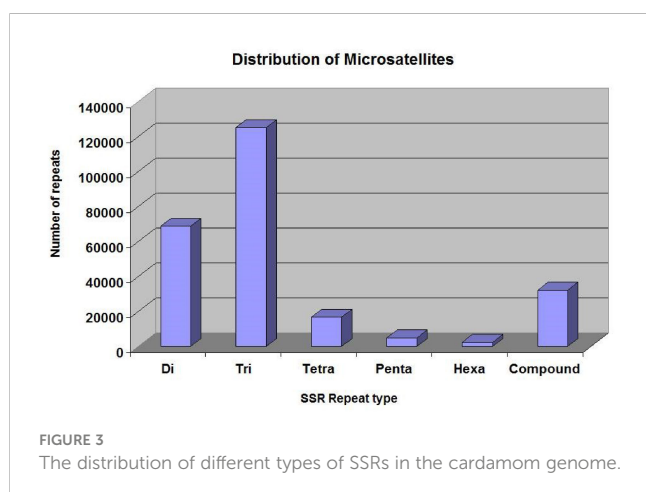
Sakthipriya and Sabu, 2018). The present report of genomic SSR markers in small cardamom paves the way for further downstream applications like genetic diversity analysis, QTL mapping, marker assisted selection etc. The flanking sequences of SSRs identified from whole genome sequences of cardamom were used for designing primers. A total of 2,27,808 primers were designed from 2,50,571 SSRs mined. For wet lab validation, 246 primers were synthesized and tested for amplification in three cardamom genotypes. Out of 246 primer pairs, 136 produced amplification product of expected size of which 60 were further used for diversity analysis in a set of 60 cardamom accessions. A representative amplification profile of 24 cardamom accessions with SSR primer EC 86 as resolved on QIAxcel multi-capillary system is shown in Figure 5.

Genetic diversity analysis

From 60 primers, a total of 874 alleles were detected with an average of 14.57 alleles per locus. The number of alleles ranged from a minimum of 4 to a maximum of 30. The marker EC 49 had the highest value for polymorphism information content (PIC) i.e., 0.80 while the marker EC 12 had the least PIC value of 0.08. We identified a set of SSR primers with high PIC (> 0.6) i.e., EC 1, EC 33, EC 47, EC 89, EC 90, EC 106, EC 183, EC 193, EC 196 and EC 212. Based on diversity analysis, the 60 accessions were clustered into two different groups which were each further divided into two smaller groups (Figure 6). The average genetic distance between the accessions was estimated to be 0.61 with highest distance of 0.75 between the two accessions 29-9-D11 and MCC12 and least distance of 0.17 between the cardamom accessions, Wayanad and CL-688. Compared to the previously reported ISSR (Anjali et al., 2016) and SSR (Cyriac et al., 2016) markers in cardamom, the average number of polymorphic bands generated by the SSR markers identified in the present study was much higher. The number of alleles detected by the use of ISSRs and SSRs ranged from 2 to 7 in the previous studies. Here, we obtained substantially higher number of alleles with a comparable number of cardamom accessions used. This implies that the SSR markers identified and used in the present study have significantly higher discriminatory powers and can be efficiently utilized for germplasm characterization. The availability of highly polymorphic SSR markers would also allow identification and mapping of genomic loci governing superior agronomic traits in cardamom.

Population structure

The Structure analysis revealed presence of five groups among the 60 accessions of cardamom (Figures 7A, B). For the 39 accessions, which were categorized as the Malabar (prostrate panicle, group 1), Mysore (erect panicle, group 2) and Vazhukka (intermediate or semi-erect panicle, group 3) types, the proportion of membership of each panicle type groups in each of the five clusters is indicated in Table 6. The inferred ancestry of these accessions in the five clusters is given in Supplementary Figure 3.



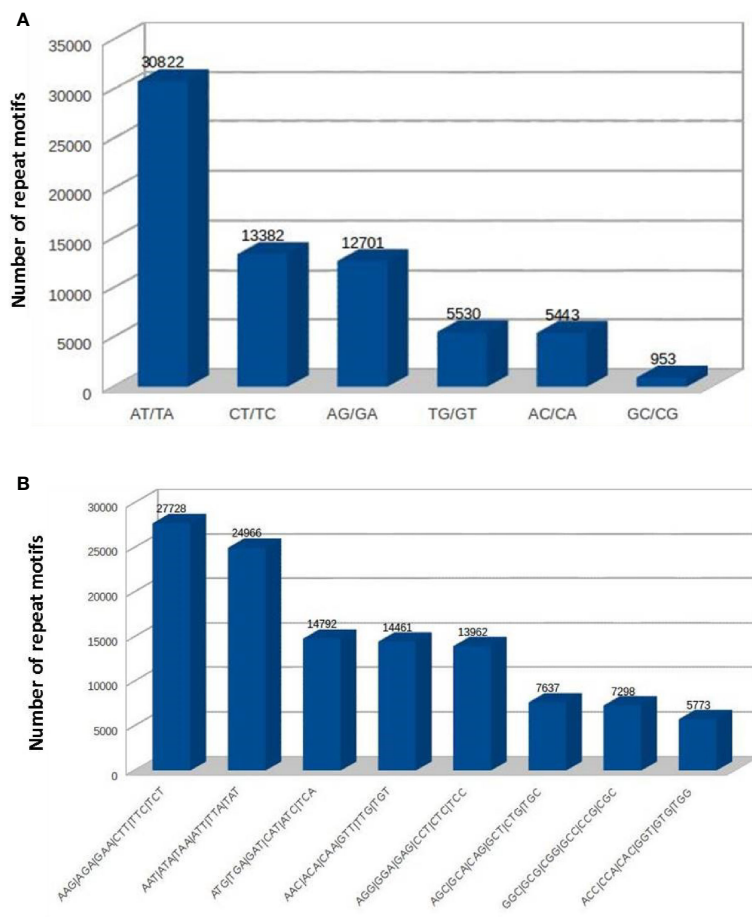


FIGURE 4 The frequencies of SSRs with different repeat sequence motifs in cardamom genome (A) dinucleotides. (B) trinucleotides.

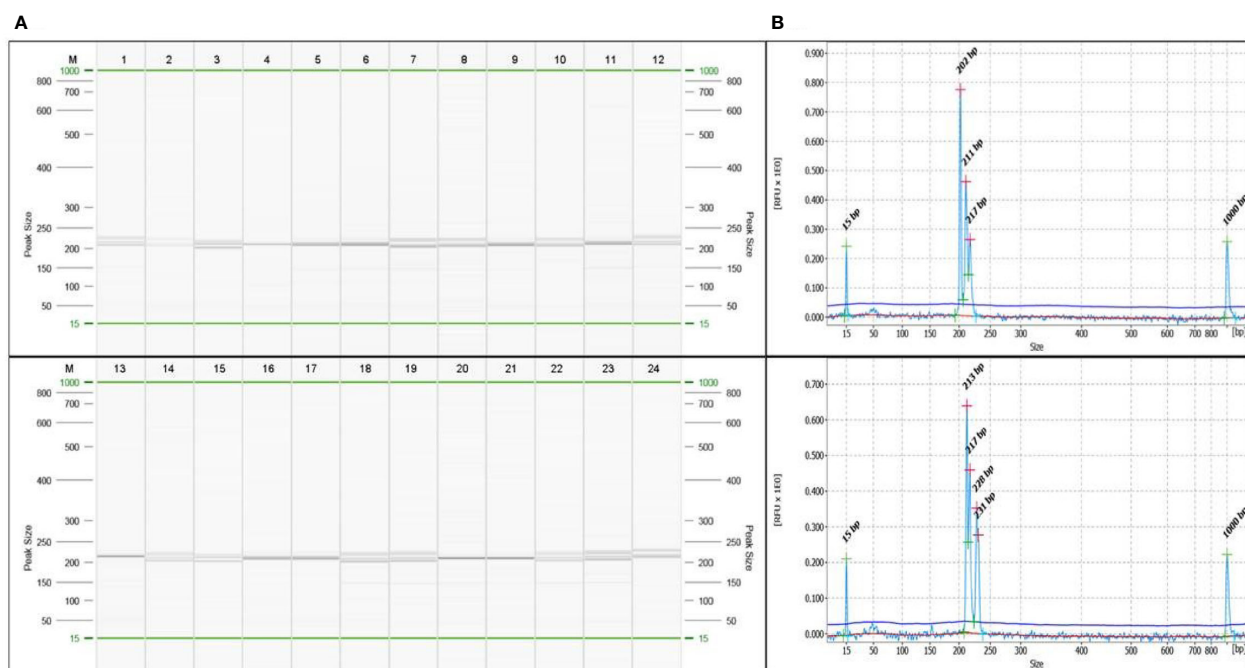


FIGURE 5 (A) A representative profile of amplification at locus EC86 in 24 accessions of cardamom captured on QIAxcel ScreenGel software. The lane marked 'M' is DNA molecular weight standard 50-800 bp v2.0 Qx DNA size marker. (B) A representative electropherogram showing different allele sizes in sample 3 and sample 24 for the marker EC86.

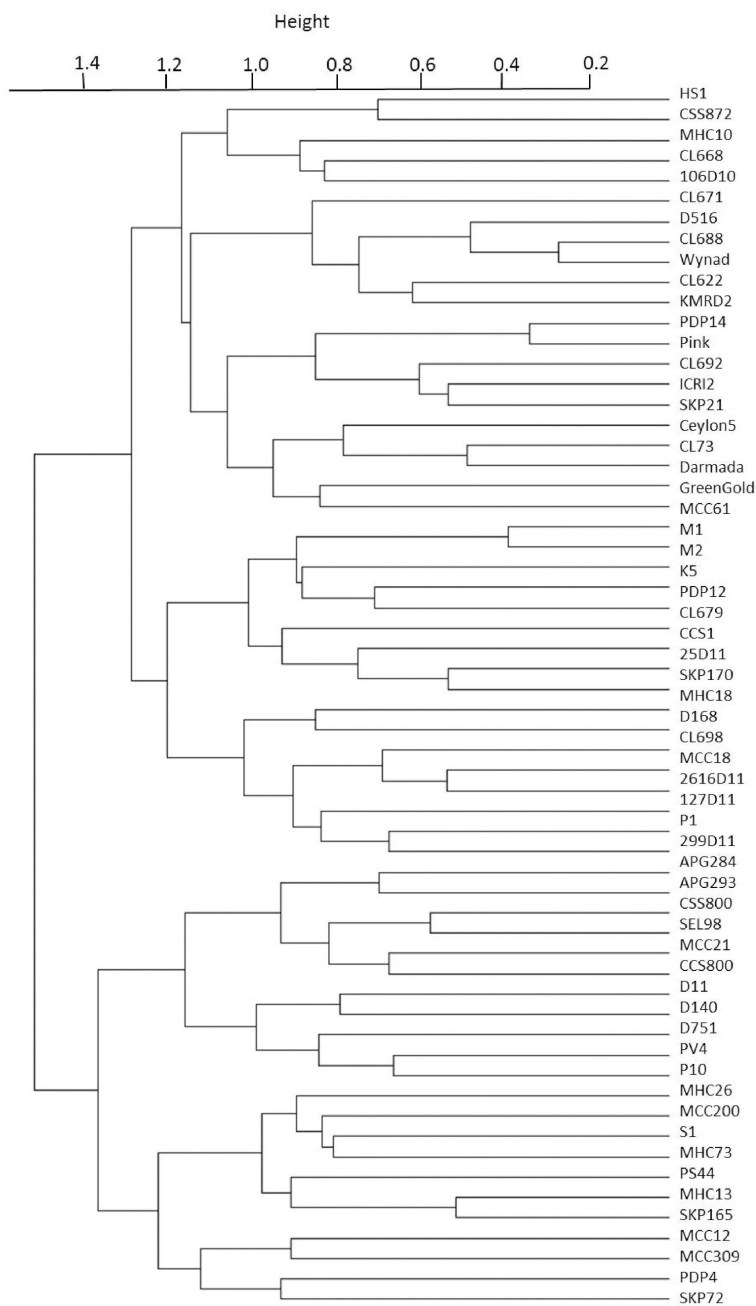


FIGURE 6
Hierarchical clustering of 60 cardamom accessions at 60 SSR loci based on Bruvo distance.

The group 1 cultivars were distributed in all five clusters although the clusters 2 and 5 had major proportion. However, cultivars from group 2 were mainly placed in cluster 4 with sizeable number in cluster 3 and 5. The pattern of grouping for group 3 cultivars was mostly comparable to that for group 1 cultivars, with greater representations in clusters 2, 3, 4 and 5 which is explainable since these are cultivars with intermediate type panicles. This was also supported by the mean values of *Fst* (pairwise fixation statistic) for the five groups indicating further the lack of genetic differentiation of cultivars from different panicle type groups (Supplementary Table 5). Similar observations have been made previously for the

cardamom accessions with different panicle types (Prasath and Venugopal, 2004). No well-defined clustering was observed for the Malabar, Mysore and Vazhukka panicle type accessions. The possibility of existence of a common ancestral origin might be a reason for the lack of genetic differentiation observed. These inferences were further substantiated by the bar diagrams of genotypic constitutions of the cultivars analysed (Figure 7C). The cultivar-wise bar plots indicated presence of high degree of genetic admixtures in all panicle type groups. The bar plot indicated presence of 13 cultivars with least or no admixture of alleles from other groups (1-CL692, 2-CL688, 7-M1, 8-M2, 9-HS1, 10-SKP21,

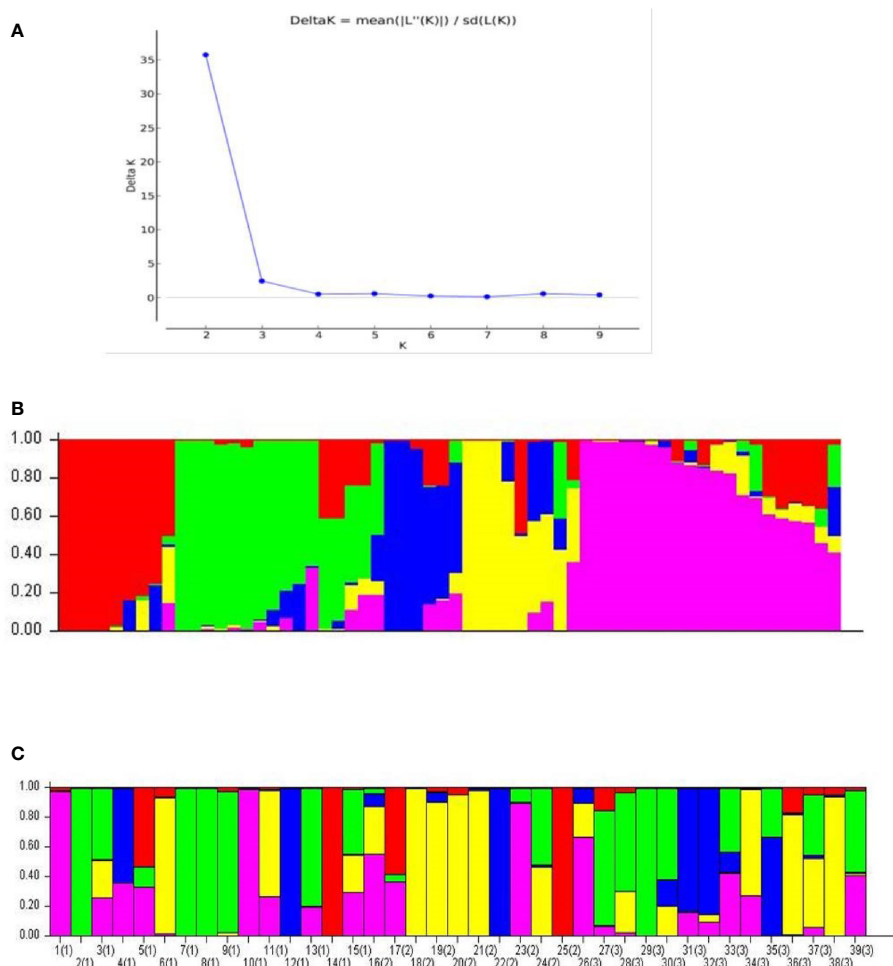


FIGURE 7 Population structure analysis for the 60 cardamom accessions. **(A)** Delta K (ΔK) plot from Structure Harvester for estimation of different numbers of subpopulations. **(B)** Population structure of 60 cardamom accessions with $K = 5$. **(C)** Population structure for the 39 different accessions of cardamom belonging to different panicle types i.e., Malabar (prostrate), Mysore (erect) and Vazhukka (intermediate or semi-erect) types.

12-CCS800, 14-SKP165, 18-Darmada, 21-MCC-12, 22-MCC21, 25-MHC13, 29-D516 and 38-MCC309). The presence of large-scale admixtures due to predominant cross-pollinations prevailing in this species appears to have resulted in lack of population differentiation among the panicle type groups of cardamom. However, lack of admixture in the 13 cultivars listed above needs to be related to their significance to agronomic performance, yield, adaptability, propagation procedures and their pedigree.

Green cardamom db

The green cardamom db is a freely available public database which can be utilized to search for microsatellite repeats and their corresponding primers for PCR based amplification of SSR markers at the 2,27,808 loci in cardamom. The database provides a SSR search option under the Microsatellite tab. Genomic SSRs can be searched based on the length and the sequence of the core motif.

TABLE 6 Proportion of membership of each of the pre-defined populations in the five clusters.

Given pop	Inferred Clusters					Number of individuals
	1	2	3	4	5	
1 Malabar (prostrate)	0.111	0.388	0.111	0.143	0.248	15
2 Mysore (erect)	0.165	0.072	0.118	0.462	0.183	10
3 Vazhukka (semi-erect)	0.034	0.342	0.201	0.265	0.158	14

This database would serve as an important repository for information regarding SSR markers developed for cardamom and facilitate comparative and functional genomics studies for cardamom in the future (<http://www.nbpgr.ernet.in:9092/>).

Conclusion

The WGS of cardamom was assembled *de novo* and the sequence length of the assembled genome was 1.06 Gb. More than 75% of the genome was sequenced through a hybrid sequencing strategy which involved three different sequencing chemistries i.e., the Oxford Nanopore, Illumina and 10x Genomics GemCode. The genome appears to be very rich in repeat regions and 68,055 gene models were predicted. At its current assembly levels, it appears very close to the members of the *Musa* family. The sequence information generated was utilized for mining and characterization of SSRs. A total of 2,50,571 SSRs were identified across the genome of cardamom and about sixty SSR markers were used for diversity analysis for a set of sixty cardamom accessions collected on the basis of their morphology and distribution in the native growing areas. These markers were highly polymorphic with an average of 14.57 alleles amplified across the cardamom accessions. The population structure analysis revealed the presence of high degree of genetic admixtures due to prevalent cross-pollinations in the species. The availability of SSR markers providing genome-wide coverage would find immense utility in future studies aimed at crop improvement through marker – assisted breeding in cardamom. Through this study, we have developed novel genomic resources in cardamom which were hitherto not reported.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

Conceptualization and funding acquisition (AG and KB); Investigation and data analysis (AG, PR, RK, KB); Manuscript

References

- Alam, A., Rehman, N. U., Ansari, M. N., and Palla, A. H. (2021). Effects of essential oils of *Elettaria cardamomum* grown in India and Guatemala on gram-negative bacteria and gastrointestinal disorders. *Molecules* 26 (9), 2546. doi: 10.3390/molecules26092546
- Ali, A., Pan, Y. B., Wang, Q. N., Wang, J. D., Chen, J. L., and Gao, S. J. (2019). Genetic diversity and population structure analysis of *Saccharum* and *Erianthus* genera using microsatellite (SSR) markers. *Sci. Rep.* 9, 395. doi: 10.1038/s41598-018-36630-7
- al-Zuhair, H., el-Sayeh, B., Ameen, H. A., and al-Shoora, H. (1996). Pharmacological studies of cardamom oil in animals. *Pharmacol. Res.* 34, 79–82. doi: 10.1006/phrs.1996.0067

preparation (SY, AG, DW). All authors contributed to the article and approved the submitted version.

Funding

This work is funded by the grant received under the project Indian Council of Agricultural Research-Consortium Research Platform on Genomics, NBFGR, Lucknow, India (project number 1007341).

Acknowledgments

The authors acknowledge the funding from ICAR-Consortium Research Platform on Genomics and the facilities provided by ICAR-National Bureau of Plant Genetics Resources, New Delhi. The authors acknowledge Dr. Sadashiv Nadugeri for providing the leaf material of accessions used for diversity analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1161499/full#supplementary-material>

- Anjali, N., Ganga, K. M., Nadiya, F., Shefeek, S., and Sabu, K. K. (2016). Intraspecific variations in cardamom (*Elettaria cardamomum* maton): assessment of genomic diversity by flow cytometry, cytological studies and ISSR analysis. *Springerplus* 5, 1560. doi: 10.1186/s40064-016-3226-x

- Anjali, N., Sowmya, D. S., Nadiya, F., and Sabu, K. K. (2015). Development of EST-SSR markers to assess genetic diversity in *Elettaria cardamomum* maton. *Int. J. Appl. Sci. Biotechnol.* 3 (2), 188–192. doi: 10.3126/ijasbt.v3i2.12380

- Bao, Z., and Eddy, S. R. (2002). Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* 12, 1269–1276. doi: 10.1101/gr.88502

- Bastias, A., Correa, F., Rojas, P., Almada, R., Muñoz, C., and Sagredo, B. (2016). Identification and characterization of microsatellite loci in maqui (*Aristotelia chilensis* [Molina] stunz) using next-generation sequencing (NGS). *PLoS One* 11, e0159825. doi: 10.1371/journal.pone.0159825
- Beier, S., Thiel, T., Muñich, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33 (16), 2583–2585. doi: 10.1093/bioinformatics/btx198
- Botstein, D., White, R. L., Skalnick, M. H., and Davies, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* 32, 314–331.
- Bruvo, R., Michiels, N. K., D'Sousa, T. G., and Schulenberg, H. (2004). A simple method for calculation of microsatellite genotypes irrespective of ploidy level. *Mol. Ecol.* 13, 2101–2106. doi: 10.1111/j.1365-294X.2004.02209.x
- Cavagnaro, P. F., Senalik, D. A., Yang, L., Simon, P. W., Harkins, T. T., Kodira, C. D., et al. (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics* 11, 569. doi: 10.1186/1471-2164-11-569
- Chakraborty, A., Mahajan, S., Jaiswal, S. K., and Sharma, V. K. (2021). Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *Commun. Biol.* 4, 1–12. doi: 10.1038/s42003-021-02720-y
- Clark, L., and Jasieniuk, M. (2011). POLYSAT: an R package for polyploid microsatellite analysis. *Mol. Ecol. Resour.* 11, 562–566. doi: 10.1111/j.1755-0998.2011.02985.x
- Cyriac, A., Paul, R., Anupama, K., Senthil, R. K., Sheeja, T. E., Nirmal, K. B., et al. (2016). Isolation and characterization of genomic microsatellite markers for small cardamom (*Elettaria cardamomum* maton) for utility in genetic diversity analysis. *Physiol. Mol. Biol. Plants* 22 (2), 219–229. doi: 10.1007/s12298-016-0355-1
- Cyriac, A., Paul, R., Prasath, D., Deepesh, P. V., Babu, N. K., and Parthasarathy, V. A. (2015). Transferability of ginger, turmeric and large cardamom SSR primers to small cardamom (*Elettaria cardamomum* maton). *J. Trop. Agric.* 53 (2), 107–115.
- D'Hont, A., Denoed, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., Noel, B., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241
- Dong, Q., Zou, Q. C., Mao, L. H., Tian, D. Q., Hu, W., Cao, X. R., et al. (2022). The chromosome-scale assembly of the *Curcuma alismatifolia* genome provides insight into anthocyanin and terpenoid biosynthesis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.899588
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Dutta, S., Kumawat, G., Singh, B. P., Gupta, D. K., Singh, S., Dogra, V., et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) millspaugh]. *BMC Plant Biol.* 11, 17. doi: 10.1186/1471-2229-11-17
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Feng, S., He, R., Lu, J., Jiang, M., Shen, X., Jiang, Y., et al. (2016). Development of SSR markers and assessment of genetic diversity in medicinal *Chrysanthemum morifolium* cultivars. *Front. Genet.* 7. doi: 10.3389/fgene.2016.00113
- Ghasemzadeh, A., Jaafar, H. Z., Ashkani, S., Rahmat, A., Juraimi, A. S., Puteh, A., et al. (2016). Variation in secondary metabolite production as well as antioxidant and antibacterial activities of *Zingiber zerumbet* (L.) at different stages of growth. *BMC Complement Altern. Med.* 22, 16. doi: 10.1186/s12906-016-1072-6
- Horticultural statistics at a glance. (2018). Ministry of Agriculture and Farmer's Welfare, GOI. *Nat. Commun.* Available at: <https://nhb.gov.in>.
- Hrideek, T. K., Radhakrishnan, V., Mohanan, K. V., Kuruvilla, K., Madhusoodanan, K., and Thomas, J. (2011). Genetic divergence in some elite landraces of small cardamom (*Elettaria cardamomum* Maton). *J. Plantation Crops* 39, 201–202.
- Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., et al. (2019). The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat. Commun.* 10, 4702. doi: 10.1038/s41467-019-12607-6
- Jose, S., Anisha, C. S., Mathew, M. K., Kumar, A. O., and Rao, Y. S. (2014). ISSR markers as a tool for assessing genetic diversity in small cardamom (*Elettaria cardamomum* maton). *Int. J. Trop. Agric.* 32 (3/4), 465–467.
- Kaliyaperumal, A. K., Murugan, M., Dhanya, M. K., and Warkentin, D. T. (2020). Botany, traditional uses, phytochemistry and biological activities of cardamom [*Elettaria cardamomum* (L.) maton] – a critical review. *J. Ethnopharmacol.* 246, 112244. doi: 10.1016/j.jep.2019.112244
- Kapoor, M., Mawal, P., Sharma, V., and Gupta, R. C. (2020). Analysis of genetic diversity and population structure in *Asparagus* species using SSR markers. *J. Genet. Eng. Biotechnol.* 18, 50. doi: 10.1186/s43141-020-00065-3
- Kress, W. J., Prince, L. M., and Williams, K. J. (2002). The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *Am. J. Bot.* 89, 1682–1696. doi: 10.3732/ajb.89.10.1682
- Kumar, K. M., Asish, G. R., Sabu, M., and Balachandran, I. (2013). Significance of gingers (Zingiberaceae) in Indian system of medicine - ayurveda: an overview. *AncSci Life.* 32 (4), 253–261. doi: 10.4103/0257-7941.131989
- Kumari, R., Wankhede, D. P., Bajpai, A., Maurya, A., Prasad, K., Gautam, D., et al. (2019). Genome wide identification and characterization of microsatellite markers in black pepper (*Piper nigrum*): a valuable resource for boosting genomics applications. *PLoS One* 14 (12), e0226002. doi: 10.1371/journal.pone.0226002
- Kuriakose, G., Sinu, P. A., and Shivanna, K. R. (2009). Domestication of cardamom (*Elettaria cardamomum*) in Western ghats, India: divergence in productive traits and a shift in major pollinators. *Ann. Bot.* 103 (5), 727–733. doi: 10.1093/aob/mcn262
- Lepais, O., and Bacles, C. F. E. (2011). *De novo* Discovery and multiplexed amplification of microsatellite markers for black alder (*Alnus glutinosa*) and related species using SSR-enriched shotgun pyrosequencing. *J. Hered.* 102 (5), 627–632. doi: 10.1093/jhered/esr062
- Li, D. M., Li, J., Wang, D. R., Xu, Y. C., and Zhu, G. F. (2021). Molecular evolution of chloroplast genomes in subfamily zingiberoideae (Zingiberaceae). *BMC Plant Biol.* 21, 558. doi: 10.1186/s12870-021-03315-9
- Li, X., Qiao, L., Chen, B., Zheng, Y., Zhi, C., Zhang, S., et al. (2022). SSR markers development and their application in genetic diversity evaluation of garlic (*Allium sativum*) germplasm. *Plant Divers.* 44 (5), 481–491. doi: 10.1016/j.pld.2021.08.001
- McCouch, S. R., Teytelman, L., Xu, Y. B., Lobos, K. B., Clare, K., Walton, M., et al. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res.* 9, 199–207. doi: 10.1093/dnares/9.6.199
- Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47, 1435–1442. doi: 10.1038/ng.3435
- Nadukeri, S., Umesha, K., Naik, M. K., Srinivasa, V., Ankegowda, S. J., Hanumantharaya, L., et al. (2020a). Variability studies in small cardamom (*Elettaria cardamomum*) genotypes for growth characters. *J. Pharmacognosy Phytochem.* 9 (6), 1346–1349.
- Nadukeri, S., Umesha, K., Naik, M. K., Srinivasa, V., Ankegowda, H. L., and Shankarappa, K. S. (2020b). Evaluation of working collections of cardamom (*Elettaria cardamomum* maton) for morphological attributes under hill zone of karnataka. *Int. J. Chem. Stud.* 8 (6), 2215–2219. doi: 10.22271/chemi.2020.v8.i6af.11102
- Ngamriabsakul, C., Newman, M., and Cronk, Q. (2000). Phylogeny and disjunction in roscoea (Zingiberaceae). *Edinburgh J. Bot.* 57 (1), 39–61. doi: 10.1017/S0960428600000032
- Prasath, D., and Venugopal, V. A. (2004). Genetic diversity and conservation of cardamom (*Elettaria cardamomum* maton) in India. *Plant Genet. Resour. Newsl.* 138, 55–60.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945
- Qian, H., and Jin, Y. (2016). An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J. Plant Ecol.* 9, 233–239. doi: 10.1093/jpe/rtv047
- Radhakrishnan, V. V., and Mohanan, K. V. (2005). Molecular characterization of some elite genotypes of cardamom (*Elettaria cardamomum* maton). *Indian J. Genet. Plant Breed.* 65 (3), 227–228.
- Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365–386. doi: 10.1385/1-59259-192-2:365
- Sakthipriya, M., and Sabu, K. K. (2018). Development and cross-genera transferability of ginger EST-SSR markers for cardamom. *Curr. Bioinf.* 13, 95–99. doi: 10.2174/1574893611666161128123827
- Sobral, M. V., Xavier, A. L., Lima, T. C., and Sousa, D. P. (2014). Antitumor activity of monoterpenes found in essential oils. *Sci. World J.* 2014, 953451. doi: 10.1155/2014/953451
- Sonah, H., Deshmukh, R. K., Sharma, A., Singh, V. P., Gupta, D. K., Gacche, R. N., et al. (2011). Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in. *PLoS One* 6 (6), e21298. doi: 10.1371/journal.pone.0021298
- Tangphatsornruang, S., Somta, P., Uthapaisanwong, P., Chanprasert, J., Sangsrakru, D., Seehalak, W., et al. (2009). Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) wilczek). *BMC Plant Biol.* 9, 137. doi: 10.1186/1471-2229-9-137
- Untergasser, A., Cutcutache, I., Korosaaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40 (15), e115. doi: 10.1093/nar/gks596
- Vairappan, C., Elias, U. M., Tamachandram, T. R., and Kamada, T. (2013). Secondary metabolites from rhizome of *Curcuma caesia* roxb. (Zingiberaceae). *Biochem. Systematics Ecol.* 48, 107–110. doi: 10.1016/j.bse.2012.11.008
- Verma, S. K., Jain, V., and Katewa, S. S. (2009). Blood pressure lowering, fibrinolysis enhancing and antioxidant activities of cardamom (*Elettaria cardamomum*). *Indian J. Biochem. Biophysics* 46, 503–506. doi: 10.1007/s00122-011-1689-2
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9 (11), e112963. doi: 10.1371/journal.pone.0112963
- Wang, Z., Miao, H., Liu, J., Xu, B., Yao, X., Xu, C., et al. (2019). *Musa balbisiana* Genome reveals subgenome evolution and functional divergence. *Nat. Plants* 5, 810–821. doi: 10.1038/s41477-019-0452-6

- Wang, C., Ulloa, M., Shi, X., Yuan, X., Sasaki, C., Yu, J. Z., et al. (2015). Sequence composition of BAC clones and SSR markers mapped to upland cotton chromosomes 11 and 21 targeting resistance to soil-borne pathogens. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00791
- Wang, Jm., Yang, Jm., Zhu, Jh., Jia, Qj., and Tao, Yz. (2010). Assessment of genetic diversity by simple sequence repeat markers among forty elite varieties in the germplasm for malting barley breeding. *J. Zhejiang UnivSci B.* 11 (10), 792–800. doi: 10.1631/jzus.B0900414
- Xin, D., Sun, J., Wang, J., Jiang, H., Hu, G., Liu, C., et al. (2012). Identification and characterization of SSRs from soybean (*Glycine max*) ESTs. *Mol. Biol. Rep.* 39, 9047–9057. doi: 10.1007/s11033-012-1776-8
- Yin, Y., Xie, X., Zhou, L., Yin, X., Guo, S., Zhou, X., et al. (2022). A chromosome-scale genome assembly of turmeric provides insights into curcumin biosynthesis and tuber formation mechanism. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1003835
- Yow, A. G., Bostan, H., Castanera, R., Ruggieri, V., Mengist, M. F., Curaba, J., et al. (2021). Improved high-quality genome assembly and annotation of pineapple (*Ananas comosus*) cultivar MD2 revealed extensive haplotype diversity and diversified FRS/FRF gene family. *Genes* 13, 52. doi: 10.3390/genes13010052
- Zhang, S., Tang, C., Zhao, Q., Li, J., Yang, L., Qie, L., et al. (2014). Development of highly polymorphic simple sequence repeat markers using genome-wide microsatellite variant analysis in foxtail millet [*Setaria italica* (L.) p. beauv]. *BMC Genomics* 15, 78. doi: 10.1186/1471-2164-15-78
- Zhu, H., Senalik, D., McCown, B. H., Zeldin, E. L., Speers, J., Hyman, J., et al. (2012). Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* ait.). *Theor. Appl. Genet.* 124 (1), 87–96. doi: 10.1007/s00122-011-1689-2
- Zhu, H., Song, P., Koo, D. H., Guo, L., Li, Y., Sun, S., et al. (2016). Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. *BMC Genomics* 17, 557. doi: 10.1186/s12864-016-2870-4