Industrial Crops & Products 111 (2018) 158–164

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Research paper

Development of genomic simple sequence repeat (gSSR) markers in cumin and their application in diversity analyses and cross-transferability

Ruchika Bharti, Sushil Kumar*, Mithil J. Parekh

Department of Agricultural Biotechnology, Anand Agricultural University, Anand-388 110, India

ARTICLE INFO

Keywords:
Cumin
Diversity
DNA marker
SSR

ABSTRACT

Cumin (*Cuminum cyminum* L.), an industrially important seed spices, has diverse usage in food and pharma industries. Therefore, documentation and characterization of cumin germplasm is essential for its genetic improvement. However, reliable co-dominant marker system especially simple sequence repeat (SSR) is not available in cumin. To develop SSR markers in cumin, DNA survey sequencing was performed on Illumina Miseq next-generation sequencing (NGS) platform. A total of 8086 sequences were detected having microsatellite repeat sequences. A set of 6421 primer pairs (including 1278 mononucleotide) were generated from genomic sequences of cumin, in which di- and tri-nucleotides repeat were more common. For marker validation, 25 primer pairs were randomly selected and amplified in 30 cumin genotypes and five other species of family Apiaceae namely, dill seed, parsley, celery, fennel and ajwain. From 23 markers, a total of 79 alleles were identified with the mean of 3.43 alleles per marker with polymorphism information content (PIC) ranged from 0.03 to 0.70. The average gene diversity and heterozygosity was 0.44 and 0.53, respectively. With a mean dissimilarity coefficients of 0.398, five clusters were identified through neighbour-joining analysis. The highest genetic distance was observed between GC-1/GP-39 and GC-3/GP-39, while the lowest level was found between GP-47 and CE-2. Of 23 primers, 21 (91.30%) exhibited cross-amplification in at least one of the species of Apiaceae.

1. Introduction

Cumin (*Cuminum cyminum* L., 2n = 14) is an industrially important seed spices crop of family Apiaceae. This aromatic herbaceous crop is cultivated for food, drugs and essential oil due to presence of important metabolites (Heidari and Sadeghi, 2014). The key producer of cumin is Asia, Middle East and North Africa with India and Iran as the largest producers and exporters (Lim, 2013). It is extensively used in foods, beverages, liquors, medicines, toiletries and perfumery (Kumar et al., 2015). Cumin seeds contain volatile oil (2.5–5%) composed primarily of aldehydes and hydrocarbons. Due to diuretic, emmanogogic, carminative and antispasmodic properties of seeds, cumin has role in traditional medicine to cure diseases like toothache, whooping cough, dyspepsia, diarrhea, epilepsy and jaundice (Parashar et al., 2014).

Cumin is a drought tolerant crop of tropic and semi-tropic regions. But narrow genetic base due to monotypism makes this crop vulnerable to biotic and abiotic stresses consequently low seed and volatile oil yield. The scope for genetic improvement of cumin is limited as information on genetic diversity and intraspecific relatedness is inadequate (Parashar and Malik, 2014). Though, traditional phenotypic based variability assessment has reported in cumin but it is time and energy consuming and costly, and exhibit genotype × environment interactions (Rukhsar et al., 2017). Therefore, dominant markers like RAPD, ISSR and AFLP have been used in cumin for genetic analysis of germplasm as well as genetic diversity investigation. However, cumbersome nature, poor consistency and low reproducibility limit the effectiveness of dominant markers.

As compared to dominant markers, simple sequence repeats (SSRs) or microsatellites are marker of choice for crop breeding as it is more informative due to co-dominant inheritance, multi-allelism, abundance, and reproducibility. Till date, no SSR marker has been developed in cumin though work on cross-transferability has been carried out (Kumar et al., 2014). This limited work on cumin using microsatellites promoted to develop simple sequence repeat markers. Previously, the SSR markers development through genomic library was a time consuming, laborious, and expensive task but next-generation sequencing (NGS) makes the SSR development faster and cheaper (Li et al., 2016). In present study, NGS technology has been used in cumin to develop novel SSR markers first time. The efficacy of developed SSRs was

* Corresponding author.
E-mail address: sushil254386@yahoo.com (S. Kumar).

http://dx.doi.org/10.1016/j.indcrop.2017.10.018
Received 25 July 2017; Received in revised form 7 October 2017; Accepted 9 October 2017

Table 1
Statistics of SSRs identified in cumin genome.

Features	Values
Total number of sequences examined	68,050
Total size of examined sequences (bp)	46,833,856
Total number of identified SSRs	10,759
Number of SSR containing sequences	8086 (11.9%)
Number of sequences containing more than one SSR	1905
Number of SSRs present in compound formation	1733

subsequently evaluated through assessing genetic diversity in cumin genotypes and by studying transferability in other seed spices of family apiaceae.

2. Materials and methods

2.1. DNA sequencing and primer amplification

Healthy and fresh leaves of two week older seedlings were collected followed by DNA extraction using CTAB protocol (Doyle and Doyle, 1987). Cumin variety GC-3 was used for DNA sequencing and SSR identification. For next generation DNA sequencing, sequencing library was prepared using the Illumina TruSeq (LT) DNAPCR-Free sample Prep Kits (FC-121–3001). After measuring DNA on Qubit 2.0 fluorometer (Life Technologies), genomic DNA (1.5 ug) was fragmented ultrasonically using Covaris S2 focused ultrasonicator. Sheared DNA was cleaned up purification beads followed by end repairing, A-tailing and ligation using adapters. The library was cleaned on magnetic stand-96 using purification beads. The final library was analysed on Agilent 2100 Bioanalyzer HS-DNA chip to verify fragment size distribution. This constructed library was sequenced using MiSeq (2 × 300 bp).

To validate the developed SSR, a set of 30 genotypes of cumin were procured from National Research Centre on Seed Spices, (ICAR-NRCSS), Tabiji, Ajmer, India. In current study, five other seed spices of family Apiaceae namely, dill seed, parsley, celery, fennel and ajwain were collected for cross-amplification of developed SSR. The extracted DNA was quantified using Nanodrop N.D.1000 (V.3.3.0, Thermo Scientific, USA). The quality and integrity of genomic DNA was evaluated on ethidium bromide stained 0.8% agarose gel by visual comparison with λ DNA standard (Invitrogen, USA). DNA was diluted to 20 ng/μl for SSR marker amplification.

2.2. SSR detection and primer design

After removing low quality reads with PHRED < 20, short sequences ≤100 bp and exact duplicates, high quality sequences were utilized for de novo assembly using MIRA (Mimicking Intelligent Read Assembly) computational tool. PrimerPro v1.0 (http://webdocs.cs.ualberta.ca/~yifeng/primerpro/), was used for SSR motifs mining, and optimal primer screening. Motif size search ranged from mono- to deca-nucleotides, with minimum repeat units set as follows: mono-nucleotides ≥10; di-, tri-, tetra-, penta-, and hexanucleotides ≥5; the remaining repeat motifs ≥5

Table 3
Frequencies of different repeat motifs of di- and tri-nucleotide repeats in cumin SSRs.

Repeat motif	Repeats number								
	5	6	7	8	9	10	> 10	Total	%
AC/GT	–	200	127	101	53	42	101	624	8.05
AG/CT	–	402	227	159	94	59	260	1201	15.50
AT/AT	–	1287	1044	992	701	524	1357	5905	76.25
CG/CG	–	13	1	–	–	–	–	14	0.18
Total	–	1902	1399	1252	848	625	1718	7744	-
AAC/GTT	77	31	8	4	3	3	1	127	11.26
AAG/CTT	96	24	11	2	2	3	10	148	13.13
AAT/ATT	203	87	40	32	24	25	114	525	46.58
ACC/GGT	56	12	–	2	–	–	1	71	6.29
ACG/CTG	8	1	–	–	–	–	–	9	0.79
ACT/ATG	53	14	5	3	3	3	7	88	7.80
AGC/CGT	3	–	–	–	–	–	–	3	0.26
AGG/CCT	22	7	1	–	–	1	–	31	2.75
AGT/ATC	69	21	5	8	7	3	7	120	10.64
CCG/CGG	5	–	–	–	–	–	–	5	0.44
Total	592	197	70	51	39	38	140	1127	–

2.3. Polymerase chain reactions (PCR) parameters and gel analysis

PCR reaction for SSR amplification was performed in SensoQuest thermocycler with 15 μL reaction mixture containing 1.5 μL template DNA (30 ng), 7.5 μL PCR master mix (2x, EmeraldAMP, Takara Biotech Inc), 0.6 μL $MgCl_2$ (50 mM), 1.0 μL diluted primer and 4.4 μL nuclease free water. The PCR reaction conditions were an initial denaturation of 94 °C from 5 min, followed by 35 cycles of 94 °C for 45 s, ΔT°C (primer specific) for 45 s, 72 °C for 45 s, and a final extension at 72 °C for 7 min. The amplified products were separated on non-denaturing polyacrylamide gels (6% PAGE) and were silver-stained. The silver-stained PAGE gels were scanned using scanner (Bio-6000, Microtek, Taiwan) at 600 dpi.

2.4. Data analysis

For data analysis, genotype wise SSR amplicon sizing (base pair) was carried out manually. The number of alleles per locus, major allele frequency, gene diversity, heterozygosity and polymorphism information content (PIC) were calculated by Power marker v.3.25 (Liu and Muse, 2005). The inter-genotype genetic distance was investigated through DARwin 6.0 with Neighbor-Joining (NJ)

Table 2
Distribution and frequencies of SSR repeat types with repeat numbers in cumin.

Motif length	Repeats number										
	5	6	7	8	9	10	11	12	> 12	Total	%
Dinucleotide	–	1902	1399	1252	848	625	380	296	1042	7744	85.49
Trinucleotide	592	197	70	51	39	38	21	20	99	1127	12.44
Tetranucleotide	–	97	27	8	5	4	1	1	4	147	1.62
Pentanucleotide	–	17	5	–	–	–	–	–	–	22	0.24
Hexanucleotide	–	11	5	1	–	–	–	–	1	18	0.20
Total	592	2224	1506	1312	892	667	402	317	1146	9058	–
%	6.53	24.55	16.62	14.48	9.84	7.36	4.43	3.49	12.65	–	–