



Research Article

Development and characterization of novel set of polymorphic SSR markers for Dill (*Anethum graveolens* L.)

Geetika Jethra¹, Sharda Choudhary² and Vinay Sharma*³¹Banasthali Vidyapith, P.O. Banasthali Vidyapith, Rajasthan-304022, India.²ICAR-National Research Centre on Seed Spices, Ajmer, Rajasthan-305 206, India.³Amity University Rajasthan, Jaipur 303002, India

Received: 21-05-2018; Accepted: 30-05-2018

Abstract: Dill (*Anethum graveolens* L.), belonging to the Apiaceae family, is considered one of the most important and economically important seed spice grown and cultivated all over India. But, due to minimal genetic and genomic resources, most of the past studies were based on physiological and biochemical analysis rather than molecular breeding. In the present study an attempt was made to develop and design *in-silico* and computationally derived highly informative microsatellite markers. Current investigation was the first attempt to develop and design microsatellite markers from Dill EST and genomic sequences. A total of 1556 SSRs were found with a density of 1SSR/0.38 kb in 589.390 kb sequence searched. Of the total, 1265, 81.30% were perfect and 291, 18.70% were identified as compound SSRs in nature. Mononucleotides (1089, 69.98%) were most frequent followed by trinucleotides (254, 16.32%), dinucleotides (183, 11.76%), tetranucleotides (22, 1.41%), hexanucleotides (6, 0.38%) and the least frequent were pentanucleotides (2, 0.13%).

Key words: *Anethum graveolens* L., Apiaceae, Microsatellite, Simple sequence repeats, Seed Spices

Introduction

Dill (*Anethum graveolens* L.) belonging to the parsley family (Umbelliferae) also known as Apiaceae family, is often grown as annual herb both in our country and around the world for its aromatic nature and valuable biochemical compounds (Karklelienė *et al.*, 2014). It is indigenous to South Europe and is cultivated in England, Germany, Rumania and the Mediterranean region. It is also cultivated throughout India chiefly in Punjab, Uttar-Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Assam and West Bengal (Khan *et al.*, 1993; Kokate *et al.*, 1998). The name “Dill” originated from an Old English word “dile”, meaning to soothe or lull. Dill seeds are much esteemed by Indians, and used as a condiment and medicine. Dill is believed to be one of the most crucial seed spices having the carminative property of relieving gas (Vokk *et al.*, 2011). It is considered to have very good anti-gas properties, and hence is used as mukhwas, or an after-meal digestive. It is also traditionally given to mothers immediately after childbirth (Jana and Shekhawat, 2010). Dill seeds and its essential oil is not only used for food flavouring but also for bactericidal, fungicidal, and medicinal applications like colic pains, vomiting, hiccups, ulcers, abdominal pains, eye diseases, uterine pains etc. and in fragrances (Bakkali *et al.*, 2008). Dill is the most valuable seed spice with various nutritional benefits. Mineral matter, ascorbic acid and other biologically active compounds supply human food allowances with the stuff of natural origin, therefore

investigations in this case are very important and useful (Singh, 2012).

Microsatellite markers, also known as simple sequence repeats (SSRs), are short tandem repeat motifs, 1–6 nucleotides in length. They occur both in protein coding and non-coding regions of DNA sequences (Bassil *et al.*, 2005). SSRs illustrate comparatively high level of length polymorphism owing to mutations of one or more repeats. SSRs are the genetic markers of choice for many applications and are favourable due to their extensive genome coverage, relatively high abundance, high reproducibility, and high level of polymorphism with multiple alleles, co-dominant inheritance, inter-specific and inter-generic transferability, and amenability to automated high-throughput genotyping, and ease of sharing among laboratories (Wright and Bentzen, 1994). DNA markers are not affected by growth stage or environment, and hence, can be employed at any plant growth stage, and used for indirect selection of alleles at loci tightly linked to them. SSRs remain unaltered by the phenotypic characteristics of the plant (Collard *et al.*, 2005) and are transferable across all the genotypes of Dill and related species of apiaceae. SSR markers have been widely used in plant genetic studies and have practical applications in plant breeding and the management of plant collections. Applications include cultivar identification, identification of duplicates in collections, parentage and genetic diversity analyses,

*Corresponding Author:

Prof. Vinay Sharma, FNASc

Dean -Academics,

Amity University Rajasthan, Jaipur 303002, INDIA

Phone: 1436 405695 (Direct)/ 01436 405678- 3355 Cell: +91 9352671355; 9887550949

Email: vinaysharma30@yahoo.co.uk; vsharma4@jpr.amity.edu



quantitative trait locus (QTL) analysis marker-assisted selection (MAS), and linkage mapping (Ellegren, 2004; Hearne *et al.*, 1992; Parida *et al.*, 2009). Moreover, SSRs have been used in comparisons of genome structure and homology between related species, including diploids and polyploids, functional, evolutionary and comparative genomic studies (Sathuvalli and Mehlenbacher, 2013). The computational approach for developing SSR markers from ESTs provides a better platform than the conventional approach (Boccacci *et al.*, 2005).

In spite of huge potential of SSR markers in modern genetic studies, microsatellite development and research on Dill is limited due to no or very less prior sequence information high expense and time consumption (Fu *et al.*, 2014). Due to these limitations, only a few microsatellite markers have been reported till date in Dill, which is the main factor for under study of this economically and medicinally significant crop.

In the present study, SSRs were developed and identified from Dill EST and genomic sequences with SSR length suitable for PCR amplification using *in-silico* approach (Jena *et al.*, 2012). Microsatellite markers developed using *in-silico* techniques were additionally verified and confirmed using different genotypes of Dill. Due to this approach a large number of amplified products were obtained, which were consecutively cloned and sequenced to develop highly specific homologous primer for Dill crop.

Materials and Methods

Retrieval of genome/ EST sequences

Available Genome and EST sequences for Dill (*Anethum graveolens*) were downloaded in GenBank and FASTA format from National Centre for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>).

Refinement of genome/ EST sequences

TRIMEST program from EMBOSS (<http://ngfnblast.gbf.de/uk/emboss.html>) has been used for eliminating the poly T / A ends of the EST and gene sequences prior to executing the sequence sets. The following settings were set to furnish most favourable results based on trial and experimental runs on numerous sequences: a) Minimum length of a poly-A tail (integer): 4, b) Number of contiguous mismatches allowed in a tail : 4, c) Write the reverse complement when poly-T is removed? Yes, and d) Remove poly-T tails at the 5' end of the sequence? Yes

Mining of simple sequence repeats

Microsatellite markers were identified and developed for Dill by *in silico* analysis. *In silico* identification and analysis of SSRs were

accomplished by two different perl scripts i.e. MICroSATellite MISA (<http://pgrc.ipkgatersleben.de/misa>) and SSRF (developed locally). The two scripts were competent to locate and categorize both single (where each repeat follows the next without interruptions) and compound (where two or more repeat units are adjacent to each other) microsatellites. In the present study, microsatellites containing motifs ranging from one to six nucleotides in size were considered. Therefore, the minimum repeat length were defined as ≥ 10 -mono, ≥ 6 -di, ≥ 4 -tri, ≥ 3 -tetra, penta and hexa nucleotide, respectively (Jethra *et al.*, 2017). The maximum difference between two SSRs was kept 0. Subsequently, data generated by SSR mining was also investigated to derive SSRs distribution and frequency.

Analysis of mined SSRs

Data generated after SSR mining was analyzed with the help of MS Excel which includes calculations of frequency and distribution of SSRs, and the percentages of different types of repeats. Analysis of SSRs distributed in coding and non-coding regions was done with the help of generated MISA file from where given SSRs start and end positions were taken and searched into GenBank file of respective organisms.

Functional annotation and Gene Ontology

Sequences of coding SSRs for all the organisms under present study were grouped in a single .txt for carrying out the annotation analysis. SSR sequences were compared against SwissProt database and the sequences with significant matches to protein entries of database were then classified according to corresponding protein gene ontology (GO) descriptors: 1) biological process 2) molecular function and, 3) cellular localization using BLAST2GO (<http://www.blast2go.de>).

Functional domain marker analysis

The annotated sequences were then subjected into InterProScan (<http://www.ebi.ac.uk/Tools/InterProScan/>) for functional domain markers (FDM) analysis. It is used for searching functionally important molecular markers.

Primer Designing

Primer pairs were designed for selectively important SSRs for Dill (*Anethum graveolens*) with 200 nucleotide long SSR flanking regions using the web-based software Primer3 (<http://bioinfo.ut.ee/primer3.0.4.0/primer3/input.htm>). The standard parameters including GC content, melting temperature, primer size and size of the PCR product were used as default.

In-silico PCR

The selected primers were subjected to *in-silico* PCR amplification and gel simulation using software SPCR (Cao *et al.* 2005) with default parameters.

Based on its results and genotype availability primers were subjected to wet lab validation.

Validation of Microsatellite Markers

We obtained a total of 123 microsatellite markers and tested them for amplification on Dill accessions. PCR amplifications were conducted in C1000TM Thermal cycler (BIO-RAD) PCR reaction for SSR was carried out in 20 µl of reaction mix containing 25ng genomic DNA (Gürcan *et al.*, 2010). PCR amplification was carried out in programmable thermal cycler from Bio-Rad Laboratories, Inc. Berkeley, California, under following conditions: an initial denaturation at 94°C for 4 min followed by 35 cycles at 94°C for 30 sec, respective primer annealing temperature ranging from 55°C to 65°C for 1 min and 72°C for 1 min. Final elongation step of 10 mins was carried out at 72°C.

Results and Discussions

Distribution and Frequency of Microsatellites

The investigation for microsatellites in 14,651 Dill EST and genomic sequences was carried out for the present study. In *Anethum graveolens* a total of 1556 SSRs were found with a density of 1SSR/0.38 kb in 589.390 kb sequence searched (Table 1). Out of 1556 SSR identified for Dill 1265, 81.30% were perfect and 291, 18.70% were identified as compound in nature (Fig 1). Among the factors that cause the generation of repeat sequences in the genome, replication slippage is often considered as the major mechanism. Mononucleotides (1089, 69.98%) were most frequent followed by trinucleotides (254, 16.32%), dinucleotides (183, 11.76%), tetranucleotides (22, 1.41%), hexanucleotides (6, 0.38%) and the least frequent were pentanucleotides (2, 0.13%) (Fig 2). Repeat motif A (505, 46.37%) was found to be the most abundant among mononucleotides closely followed by T (487, 44.71%). Mononucleotides G and C were also present with the frequencies of 54, 4.95% and 43, 3.94%, respectively.

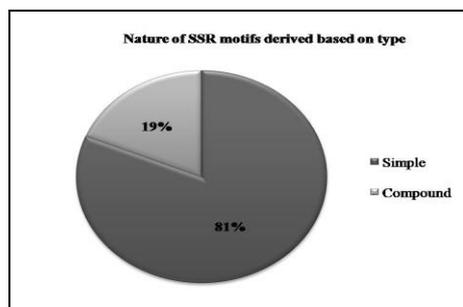


Figure 1. Nature of SSR motifs derived in Dill based on type

Among dinucleotides, repeat motif AT (58, 31.69%) was followed by TA (57, 31.14%), GA (25, 13.67%), CT/TC (14, 7.65% each) and AG (10, 5.46%). The least available dinucleotides were AC (3, 1.63%) and

TG (2, 1.09%). Trinucleotide repeat motifs CTT (23, 9.05%) were the most frequent followed by AAG/TCT (20, 7.87% each), TAA/TTA (18, 7.08% each), TTC (16, 6.29%), ATA (14, 5.51%), AGA/ATT (11, 4.33%) and TTG (10, 3.93%). Other than above mentioned trinucleotides, 22 different trinucleotides AAT/CAA (8, 3.14%), TGA (7, 2.75%), AAC/GAA/GCT/ GGT (6, 2.36%), CTG/TAT/TCA (5, 1.96%), TGC (4, 1.57%), AGC/ATG/CAC/ CAG/GCA/GGA (3, 1.18%), CTA/GTT/TAG/TCG (2, 0.78%) and TCC (1, 0.39%) were also identified with the frequencies ranging below 10 and comprising a total of 36.61%.

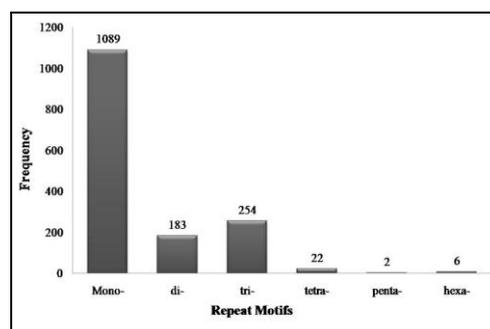


Figure 2. Distribution and frequency of SSRs in Dill based on length

Among tetranucleotides, repeat motifs AGGT (5, 22.72%) had the highest frequency, followed by CAAT/CTAC/TTGA (3, 13.63%) and AAAT/ATCA/ATTT/TCCT (2, 9.10%). Only one pentanucleotide TTTTA (2, 100%) was found for Dill. Hexanucleotide repeat motifs ATGGGA/CATTCC (3, 50% each) were also present with the same frequency. The results in this study for Dill also justified that, the frequency of SSR motifs decreased with increase in the length of motifs. The frequencies of all nucleotide repeat type with repeat numbers in Dill genome are presented in table 1. The lengths of the SSRs in the different groups varied greatly with the largest variation (12–64 bp) being present in dinucleotide repeats and the smallest variation (15–20 bp) in pentanucleotide repeats.

Table 1. Statistics of SSRs identified in Dill genome

Features	Values
Total number of sequences examined	225
Total size of examined sequences (bp)	589390
Total number of identified SSRs	1556
Number of SSR containing sequences	110
Number of sequences containing more than one SSR	74
Number of SSRs present in compound formation	291

Functional annotation of SSR containing sequences

Gene Ontology (GO) and high-level functional category of SSRs were assigned based on Blast2Go function annotation system. For Dill, out of 1013

SSR containing sequences identified, 505 (49.85%) were classified into cellular components by the GO terms (Fig 3) followed by biological processes (324, 31.98%) and molecular functions (184, 18.16%). In the category of molecular function, the largest sets of genes were involved in the binding activity (64, 34.78%) followed by the genes involved in catalytic activity (48, 26.09%), structural molecule activity (40, 21.74%), electron carrier activity (20, 10.87%) with least number of genes involved in transporter activity (12, 6.52 %). Genes identified for Dill under the category of cellular component were involved mainly in cell, cell parts (111, 21.98%), macromolecular complex (95, 18.81%), membrane (62, 12.28%), membrane part (56, 11.09%), organelle part (52, 10.30%) and organelle with frequency of 18 (3.56%). Additionally, genes involved in biological process were further sub-categorized into metabolic process (142, 43.83 %) followed by cellular process (125, 38.58%) and single organism process (57, 17.59%).

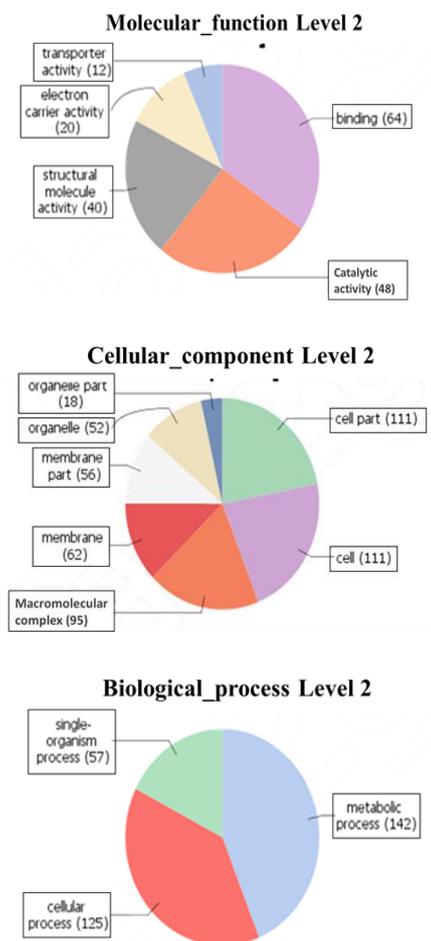


Figure 3. Biological process, molecular function, and cellular component of coding SSRs identified in Dill.

Confirmation of functionally putative SSRs

The results were merged to get the confirmed number of annotations for Dill (Fig 4). These

confirmed annotations determined the SSRs with both the GO id and corresponding assigned protein families as functionally putative SSRs.

In-silico PCR

SSR marker of gene *ycf4*, *petA* and *matK* (Fig 5) showed maximum transferability of 100% for Dill and all the seed species of family Apiaceae.

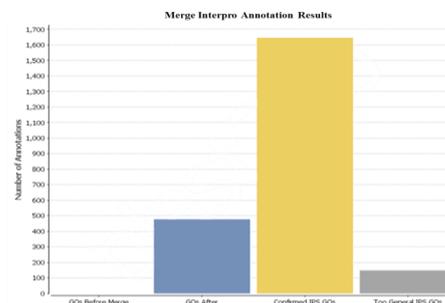


Figure 4. Confirmed functional annotation of SSRs in Dill.

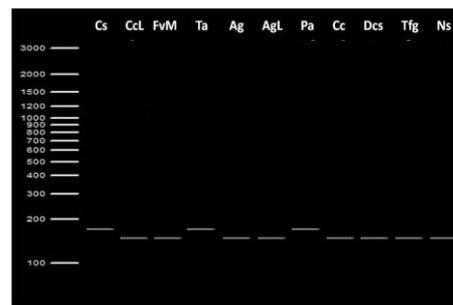


Figure 5. Transferability of coding SSR markers *matK* of *Anethum graveolens*

In *Apium graveolens* L., only the primers illustrating clear and predominant banding pattern were considered eliminating all the primers with faint or no banding pattern, reducing the number of good quality markers to 39 (82.97%). Among these high quality markers, 13 (33.33%) were monomorphic and 26 (66.66%) were polymorphic. It was observed that the number of loci per SSR ranged from two to six. These primers can also be used as specific markers for Apiaceae family.

Conclusion

Since last few years research and developmental activities based on Dill production and quality has started in India. Rigorous attempts have been made for the improvement of Dill yield with enhanced quality but recently microsatellite markers have proved very significant and one of the key technique. *In-silico* mining of SSR's has facilitated the existing tedious and time consuming methods by automating the development of SSR markers and classification for Dill. The present study is the first attempt to develop SSRs and design microsatellite markers from Dill EST and genomic SSR sequences. A total of 1556 SSRs were found with a density of 1SSR/0.38 kb in 589.390 kb sequence

searched. Mononucleotides (1089, 69.98%) were most frequent followed by trinucleotides (254, 16.32%), dinucleotides (183, 11.76%), tetranucleotides (22, 1.41%), hexanucleotides (6, 0.38%) and the least frequent were pentanucleotides (2, 0.13%).

Acknowledgement

The authors gracefully acknowledge the Director, National Research Centre on Seed Spices, Tabiji, Ajmer for financial and infrastructural support.

References

- Bakkali F, S Averbeck, D Averbeck, and M Idaomar. "Biological effects of essential oils – a review." *Food and Chemical Toxicology* 46 (2008): 446–475. Print
- Bassil NV, R Botta, and SA Mehlenbacher. "Microsatellite markers in hazelnut: isolation, characterization and cross-species amplification." *Journal of the American Society for Horticultural Science* 130 (2005): 543–549. Print
- Boccacci P, A Akkak, NV Bassil, SA Mehlenbacher, and R Botta. "Characterization and evaluation of microsatellite loci in European hazelnut (*Corylus avellana* L.) and their transferability to other *Corylus* species." *Molecular Ecology Notes* 5 (2005): 934–937. Print
- Cao Y, L Wang, K Xu, C Kou, Y Zhang, G Wei, J He, Y Wang, and L Zhao. "Information theory-based algorithm for in silico prediction of PCR products with whole genomic sequences as template." *BMC Bioinformatics* 6 (2005): 190. Print
- Collard BCY, MZZ Jahufer, JB Brouwer, and ECK Pang. "An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts." *Euphytica* 142 (2005): 169–196. Print
- Ellegren H. "Microsatellites: simple sequences with complex evolution." *Nature Reviews Genetics* 5 (2004): 435–445. PMID:15153996
- Fu N, PY Wang, XD Liu, and HL Shen. "Use of EST-SSR markers for evaluating genetic diversity and fingerprinting celery (*Apium graveolens* L.) cultivars." *Molecules* 19.2 (2014): 1939–55. Print
- Gürçan K, and SA Mehlenbacher. "Transferability of microsatellite markers in the Betulaceae." *Journal of the American Society for Horticultural Science* 135 (2010): 159–173. Print
- Hearne CM, S Ghosh, and JA Todd. "Microsatellites for linkage analysis of genetic traits." *Trends in Genetics* 8 (1992): 288–294. PMID:1509520
- Jana S, and GS Shekhawat. "Anethum graveolens: An Indian traditional medicinal herb and spice." *Pharmacognosy Review* 4.8 (2010): 179–184. Print
- Jena SN, A Srivastava, KM Rai, A Ranjan, SK Singh, and T Nisar. "Development and characterization of genomic and expressed SSRs for levant cotton (*Gossypium herbaceum* L.)." *Theoretical and Applied Genetics* 124.3 (2010): 565–76. Print
- Jethra G, S Choudhary, and V Sharma. "Development of New Set of EST-SSR markers for Evaluating Genetic Diversity and Fingerprinting Local Celery (*Apium graveolens* L.) Cultivars." *Current Biomarkers* 7 (2017): 52–63. Print
- Karklelienė R, E Dambrauskienė, D Juškevičienė, A Radzevičius, M Rubinskienė, and P Viškelis. "Productivity and nutritional value of dill and parsley." *Horticultural Science (Prague)* 41.3 (2014): 131–137. Print
- Khan MMA, SH Samiullah, MM Afaq, RK Afridi, and FA Khas. "Yield and Quality of Dill (*Anethum sowa* L.) in relation of basal and foliar application of Nitrogen and Phosphorus." *Glimpses in Plant Research, Medicinal Plants: New Vistas of Research (Part 1)*, Eds. Govil, J. N., Singh V.K. and Shamima Hashmi, Today and Tomorrow's Printers and Publishers, New Delhi, India (1993): 275–282.
- Kokate CK, AP Purohit, and Gokhale. "Pharmacognosy." *Nirali Prakashan, Pune, India* (1998): 283. Print
- Parida SK, SK Kalia, K Sunita, V Dalal, Hemaprabha G, and Selvi A. "Informative genomic microsatellite markers for efficient genotyping application in sugarcane." *Theoretical and Applied Genetics* 118 (2009): 327–338. PMID:18946655
- Sathuvalli VR, and SA Mehlenbacher. "De novo sequencing of hazelnut bacterial artificial chromosomes (BACs) using multiplex Illumina sequencing and targeted marker development for eastern filbert blight resistance." *Tree Genetics & Genomes* 9 (2013): 1109–1118. Print
- Singh S. "Chemical Constituents of Essential Oil from Anethum Sowa Kurz Seed." *Journal of Chemical and Pharmaceutical Research*. 4.9 (2012): 4156–4160. Print
- Vokk R, T Lõugas, K Mets, and M Kravets. "Dill (*Anethum graveolens* L.) and Parsley (*Petroselinum crispum* (Mill.) Fuss) from Estonia: Seasonal Differences in Essential Oil Composition." *Agronomy Research*. 9.2 (2011): 515–520. Print
- Wright JM, and P Bentzen. "Microsatellites: Genetic markers for the future." *Reviews in Fish Biology and Fisheries* 4 (1994): 384–388. Print

Cite this article as:

Geetika Jethra, Sharda Choudhary and Vinay Sharma. Development and Characterization of Novel Set of Polymorphic SSR Markers for Dill (*Anethum graveolens* L.). *Annals of Plant Sciences* 7.6 (2018) pp. 2347-2351.

 <http://dx.doi.org/10.21746/aps.2018.7.6.8>

Source of support: Director, National Research Centre on Seed Spices, Tabiji, India.

Conflict of interest: Nil