International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2019; 7(2): 1070-1076

Received: 12-01-2019
Accepted: 16-02-2019

Shiwani
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

MK Mandavia
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

RS Tomar
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Correspondence
Shiwani
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Development and validation of simple sequence repeat markers from genome of dill (*Anethum graveolens* L.)

Shiwani, MK Mandavia and RS Tomar

Abstract
Dill (*Anethum graveolens* L.) is one of the important spice crops. Dill is an important medicinal and industrial plant. Considering the importance of this herbal spice plant various biochemical studies has been carried out and which are still going on including traditional and modern approaches. In present study, Ion Torrent Genome Sequencing (Ion S5) technology was used to generate dill draft genome using genotype Gujarat Dill 1. Gujarat Dill-1 genotype was sequenced using next generation sequencing platform Ion S5 which yielded 10.44 Gb of raw data. After quality check and its control total of 3.2 Gb of data was remained in the data set. In order to maintain the quality of assembled data contigs were filtered, after filter 32 contigs were remained with 5000 bp of minimum, 42,775 bp of maximum and 9,568 bp of average contig length. After Blast2Go analysis approximately 542 putative functions were found from 32 input sequences, which were contributing for 498 GO terms (biological processes (206), cellular components (157) and molecular processes (135). Out of 32 sequences, all most all sequences showed positive InterPro result while 5 sequences did not showed any InterPro results and 15 sequences were scanned with GOs. 47 IPS IDs were identified, among them 30 were IPS domains, 14 were IPS family and 4 were IPS active sites. SSR primers from assembled data were designed *in silico* using batchprimer3 v1.0 online program. Total 80 primers were identified from 32 contig data, in which most of them were tetranucleotide repeats (40%) followed by dinucleotide repeats (22.5%), pentanucleotide repeats (20%) and trinucleotide repeats (7.5%). For validation of SSR primers, among identified 80 primers only 25 primers were used which were having melting temperature more than 55 ºC. Primers were validated using five Dill genotype (Ajmer Dill1, Ajmer Dill2, Gujarat Dill1, Gujarat Dill2, Gujarat Dill3). Among 25 primers, 19 primers amplified the DNA. From 19 primers, three primers were found polymorphic among five genotypes of Dill. These primer were related to those sequence which were having putative function for cytochrome c oxidase. The polymorphic primer can be used in genetic diversity analysis of Dill and its relatives.

Keywords: *de novo*, IPS, SSR, putative

Introduction
Dill (*Anethum graveolens* L.) is an annual or biennial herb. The genus name *Anethum* is derived from Greek word aneeson or aneeton, which means strong smelling. The *A. graveolens* are diploid 2n = 22 (Ma *et al*., 1984) [8] and a typical out crossing species (Snell and Aarssen, 2005) [14]. The plant is a native species in Southwest Asia and is cultivated in Europe, India and the United Statesill (*Anethum graveolens* L.) is an annual or biennial herb. The genus name *Anethum* is derived from Greek w (Tucker, 2008). The chief components of dill herb oil are phellandrene (35%) and 3, 9-epoxy-p-menth-1-ene (25%), of dill fruit oil, limonene (up to 70%) and carvone (up to 60%) (limonene and carvone are closely related; together they represent about 95% of the oil). Carvone can be applied as germination suppressor, e.g. in potatoes. Dill has been reported to possess anti hyperlipidaemic and anti hypercholesterolaemic activity (Yazdanparast and Alavi, 2001) [19].

Genome sequencing refers to methods for determining the order of the nucleotides bases adenine, guanine, cytosine and thymine of an organism's genome at a single time. Numerous technologies have been developed to analyze and quantify the sequencing. Any high-throughput sequencing technology can be used for genome sequencing such as the Illumina Genome Analyzer, Applied Biosystems SOLiD ™, Roche 454 Life Sciences system, Ion Torrent. Genome sequencing work will be helpful to decipher information from coding and noncoding part of Dill. It also provides facts of genome construct its constituent and genic variation. SSR developed from genome data will be helpful to develop a breeding program for

mapping and marker assisted selection. The main genetic tools used for the identification and breeding of cultivars of domesticated species are morphological and molecular marker (Tanksley *et al.,* 1993) [15]. SSRs have been the most widely employed class of molecular markers used in genetic studies with applications in many fields of genetics including genetic resources conservation, population genetics, molecular breeding and paternity testing (Ellegren, 2004) [3]. This range of applications is due to the fact that SSR markers are codominant, multiallelic, and highly reproducible, have high resolution, are amenable to high throughput and are based on polymerase chain reaction (PCR) (Oliveira *et al*., 2006) [10]. As a convention, SSRs are regions in the genome where a group of bases (1–8 bp long) are repeated in tandem (Richard *et al*., 2008) [13]. These regions can be isolated either by data mining of existing sequences or by generating SSR-enriched libraries. Microsatellites or simple sequence repeats (SSRs) are found throughout the eukaryotic genomes and occur in both coding and noncoding regions. Microsatellites are stretches of DNA consisting of tandemly arranged units in 1-6 bp (Gupta *et al.,* 1996; Thiel *et al.,* 2003) [13, 17], characterized by their co-dominant inheritance, wide genomic distribution, hyper variable and multiallelic nature (Powell *et al.,* 1996; Parida *et al.,* 2009) [12, 11]. They are also termed as simple sequences (Tautz, 1989) [16], Short Tandem Repeats (STRs) (Edwards *et al.,* 1991) [2] and Simple Sequence Repeats (SSRs) (Jacob *et al.,* 1991) [6]. The existence of microsatellites in a wide range of evolutionarily diverse eukaryotic genome (from yeasts to humans) was first documented by (Hamada *et al.,* 1982). Microsatellites are ubiquitous in the coding and noncoding regions with a higher density of simple sequence motifs in the noncoding regions of eukaryotes. In plants, SSRs are much more abundant and preferentially associated within untranslated regions (UTRs) of the transcribed regions. With the importance of dill as a medicinal and spices crop, this research has been challenge to discover the genes responsible for different cellular, biological and molecular aspect through genome sequencing. Genome sequencing work will be helpful to decipher information from coding and noncoding part of an organism. It also provides facts of genome construct its constituent and genic variation. SSR developed from genome data will be helpful to develop a breeding program for mapping and marker assisted selection. Polymorphic SSR markers development would be useful for the population genetic studies and germplasm management of dill. Sequencing of *Anethum graveolens* genome will provide a model for characterization of metabolic pathway, involved in synthesis of bioactive compound, comparative evolutionary studied among various Apiaceae family members & help annotate their genome.

Material and Method

Sample collection and DNA isolation

For this study five genotypes of dill were used namely, Ajmer Dill 1, Ajmer dill 2, Gujarat Dill 1, Gujarat Dill 2, Gujarat Dill 3. The seeds of Dill genotypes Gujarat Dill 1, Gujarat Dill 2, Gujarat Dill 3 were obtained from Seed Spices Research Station, Sardarkrushinagar Dantiwada Agricultural University, Jagudan and seeds of Ajmer Dill 1, Ajmer dill 2 collected from National Research Centre on Seed Spices, Ajmer. Genomic DNA isolated from fresh seedling tissue by following modified CTAB method as described by Doyle and Doyle (1990) [1]. After DNA isolation the concentration was determined by using Picodrop PET01 with software v2.08 (Picodrop Ltd., Cambridge U.K). The ratio at A_{260}/A_{280} was 1.85. For the good quality and purity of genomic DNA, plant DNA extraction kit (Nucleopore DNA extraction kit) also was used in the study.

Sequencing of sample by ION Torrent S5

gDNA fragment library were prepared using Ion Shear™plus enzyme mix II (as per the Ion S5 library preparation protocol), then fragments were purified by Agencourt® AMPure® XP reagent (1.8X sample volume) followed by adapter ligation and nick repair. Adapter ligated fragments were purified. Size selection of library performed on 2% E-Gel. Stop the run when the 500-bp ladder band is at the top edge of the collection well. Fragments of the desired size are enriched by amplifying library. Emulsion PCR was carried out to generate multiple copies of fragments on the Ion Sphere (IPS) beads using the Ion torrent OT2 machine. Sequencing of each IPS was carried out in Ion Torrent S5 sequencer system using Ion 530/540 chip.

Data Analysis

Raw data quality analysis using CLC Version9.5.4. Quality control check of raw sequence data coming from high throughput sequencing pipelines provides a modular set of analyses which you can use to give a quick impression of whether your data has any problems of which you should be aware before doing any further analysis. These tools analyze FASTA format file to calculate different sequence statistics and calculates the average quality score for each read and overall average quality score for all the reads. *de novo* genome assembly was also carried by using CLC Version 9.5.4 *de novo* Assembler (Merida *et al.,* 2013) [9]. Functional annotation was carried out by using Blast2Go software to align the consensus sequences from the assembled contigs and the singleton sequences (Liu *et al.,* 2013) [7]. Identification of putative genes was carried out using CLC software Genomics Workbench (CLC Genomics Workbench v9.5.4 or according to the manufacturer's instructions (User manual according respective company) by company (Zhou *et al.,* 2012) [20].

Identification of genomic SSR marker, primer designing and validation

Genomic SSR markers were designed from those contigs which were involved in biological, molecular and cellular function. SSR markers were designed using a high through put web tool Batch Primer3 v1.0 (Guimaraes *et al.,* 2012) [4]. For validation of designed SSR markers in five genotype of dill, the mastermix was prepared in a microfuge tube in which the buffer was added first followed by sterile water, Primer, dNTPs mix followed by Taq DNA polymerase (Table 1). At the last DNA was added in each tube separately. The reagents were mixed gently by tapping against the tube. The tubes were then placed in the Thermal Cycler for amplification. The PCR condition for thermal cycler is given in table 2. PCR products were subjected to electrophoresis with marker DNA of known molecular weight in 1.8% agarose gel. After electrophoresis, the gel was carefully taken out of the casting tray and photographed in GeneSys gel documentation system.

Result and Discussion

Genome sequencing and *de novo* assembly of the raw data

The study was initiated with the objective of genome analysis of Gujarat Dill1. A total of three run were performed on Ion Torrent S5 next generation system. In first run of sample, total data generated was 4.34 Gb. ISPs loading was 88% on the Ion 530™ chip and 13,378,296 total numbers of reads were

generated after removing of polyclonal (29%) and low quality data (41%), and adaptor dimer with mean length of 325 bp. The loading of second run sample's ISPs was 73% on the chip and total number of reads of 5,771,965 was obtained after removing of polyclonal (19%) and low quality data (73%) and adapter dimer, with mean read length of 300 bp. Total data generated in second run was 1.73 Gb. Second run was carried out by using bar codes for different samples, out of which dill was barcoded with IonXpress_002 barcode name. For dill 492,151,055 base pairs were generated and total reads were 220,205 and mean length of reads were 286 bp. Third run was carried out in Ion 540 ™ chip and total of 5.6 Gb data was generated and there were 27,361,042 numbers of reads with 197 bp mean read length. The data of all three run is shown in Table 3.

Total of 10.44 Gb raw data was generated through sequencing. Raw sequence data were processed for quality assessment using CLC workbench v9.5.4, in which the adapter sequences were removed and trimmed on quality bases. Total sequences and nucleotides in data sets were 16,179,010 and 3,378,117,867 respectively. In order to get quality reads the data was trimmed. *de novo* assembly of 15,940,402 reads had yield data of 321,805,497 base pairs (321 Mb). Total numbers of contigs were 555,413 in number with 200 bp of minimum, 10,247 bp of maximum contig length. N25, N50 and N75 measurement of contigs were found 1279 bp, 661 bp, and 421 bp respectively via *de novo* assembly. In order to maintain the quality of assembled data contigs were filtered, in which only those contigs who have sequence above and equal to 5000 bp were selected for the annotation. This whole process was carried out in CLC Genome Workbench. After filter 32 contigs were remain with 5000 bp of minimum, 42,775 bp of maximum and 9,568 bp of average contig length.

Blast2go analysis of genome sequencing data

Sequence from dill (Gujarat Dill 1) cultivar was assembled using CLC workbench. Validation and functional annotation of these sequences were done using Blast2Go tool. Data analysis progress of functional annotation is given in (Figure 2). Total 32 sequences were functionally annotated out of which 31.84 (99.5%) were showed positive interpro, while 30 (93.75%) were got Blast hits. 87.5% (28) and 65.62% (21) sequences were mapped and annotated respectively.

Gene ontology (GO) IDS and sequence distribution

Gene ontology mainly divided in to three groups, biological process, cellular components and molecular function. Collectively 498 numbers of GO IDs were found which were grouped into biological process, cellular components and molecular function, respectively responsible to cellular process and metabolic process. During gene ontology direct Go count were generated. In molecular function maximum sequences from total sequences respond to ubiquinone activity and protein binding showing maximum Go count followed by Magnesium ion binding, iron-sulfur cluster binding. In biological processes maximum sequences were showing maximum Go count to ATP synthesis followed by protein – chromophore linkage and aerobic respiration. In cellular component maximum sequences having maximum Go count which respond to integral component of the membrane followed by integral component of the plasma membrane.

Biological process

Dill plant genome 32 sequences contributing for total 206 biological process. Out of total sequences 21 sequences were responsible for cellular metabolic process followed by 20 sequences for organic substance metabolic process, 19 sequences for primary metabolic process and 18 sequences for nitrogen compound metabolic process (Figure 3).

Molecular function

Dill genome 32 sequences contributing for binding, catalytic, enzyme regulator, catalytic activity and transporter type of 135 molecular functions (Figure 4). Out of total sequences maximum 9 sequences contributing for oxidoreductase activity and 8 sequences for cation binding followed by 7 sequences for tetrapyrrole binding and 6 sequences for nucleotide binding.

Cellular component

In the cellular component cell, cell part, organelle, envelope, extracellular region, extracellular region part, macromolecular complex, organelle part such type of process were recognised. Total 157 cellular processes were identified in the total sequences. In the result (Figure 5) maximum dill sequences were contributing for intracellular part and organelle followed by intrinsic component of membrane and respiratory chain.

Putative/hypothetical genes identification

Putative genes identification is a requirement for gene investigation in the era of genomics. The strategy for discovery of potential ORFs at a large-scale in dill genomes described here will contribute to their annotation and identifies new potential regulators of diverse biological processes in plants and that improve understanding of plant biology. Here analysis revealed that remarkably, 542 putative genes were found in dill genotype (Gujarat Dill 1) from total 32 contigs. Out of 542 total putative genes, maximum number of genes (155) found in Dill_D_(single)_trimmed_contig_32 may have putative function of DNA repair and minimum number of gene (1) found in Dill_D_(single)_trimmed_contig_19. Out of total 32 sequences, 26 sequences were there which revealed GO terms. Total of 498 GO terms were identified, among all 44 GO terms were found in Dill_D_(single)_trimmed_contig_4 which contribute for biological processes, metabolic function and cellular components. Detailed analysis of putative functions identified in 32 contigs represent in supplementary table 1.

SSR marker identification and primer design using BatchPrimer3, V 1.0

Genomic SSRs of dill were identified and developed from the draft genome of dill genotype, Gujarat Dill-1 by using online tool BatchPrimer3 v1.0. For the identification of SSRs FASTA files of 32 contigs were used. Using the filtered assembly report, 80 SSR markers were developed, among 80 SSR primers 18 were dinucleotide, 6 were trinucleotide, 40 were tetranucleotide and 16 were pentanucleotide. Total selected 80 primers met the following parameters: 100–300 final product length (optimal 200 bp), primer size from 19 to 27 bp (optimal 23 bp) and GC content 40%–70% (optimal 50%); the annealing temperature was set at 50–60 °C (optimal 55 °C). Tm and GC % was found between 54ºC-57 ºC and 40%-50% respectively. Length of SSR primers was between 20-23 bp with 137-187 bp product size.

Validation of SSR markers

In order to validate the SSR primers in 5 different genotypes of dill (Ajmer Dill-1, Ajmer Dill-2, Gujarat Dill-1, Gujarat Dill-2, Gujarat Dill-3), 25 SSRs were selected manually out of 80 primers. Criteria used to select the SSR primers for validation purpose was based on the melting temperature (Tm) and GC content (GC%) of the primers (Supplementary table 2). Those primers were selected which were having melting temperature above 54ºC and GC% above 40% and primers were chosen from those contigs which were involved in putative function. Firstly, all the 25 primers were screened on different annealing temperature and total of 19 primers amplified the DNA.

Polymorphism pattern of SSR

All the 19 SSRs primers were amplified a total of 23 bands. Among 19 amplified primers, only three primers showed polymorphism (Figure 6), rest were monomorphic. The SSR primer Dill 8, Dill 15, Dill 20 and Dill 23 produced maximum number of 2 bands, while others produced only 1 band. Out of 23 bands, 5 bands were polymorphic and 18 bands were monomorphic. Among the 5 polymorphic bands, 4 bands were shared polymorphic within two or more varieties, while one band was unique-polymorphic. The amplified fragments ranged from 86-342 bp. The largest amplicone of 342 bp was amplified by SSR primer Dill 23 and smallest fragment of 86 bp was found with SSR primer Dill 22. The percent polymorphism obtained for SSR primers were ranged from 0 % to 100%. The polymorphic information content (PIC) was calculated for each primers and it was 0.48 and 0.49 for Dill15 and Dill23 respectively. The SSR primer index (SPI) was 0.96 (Dill 15) and 0.98 (Dill 23).

Conclusion

The whole study provides important information about the biological functions, molecular functions and cellular components occurring in dill. This study can be of great importance in the field of research to understand metabolic pathways and genetic basis of all the physiological processes occurring in dill plant. It also provides data through which we can identify gene and functional elements of genome and give basis for annotation of complete plant genome. The designed primers having putative function can be used in genetic diversity studies.

Table 1: Preparation of reaction mixture for SSR

Sr. No.	Reagent	Quantity
1	PCR buffer (10X)	2 µl
2	Taq polymerase (3 U.µl^{-1})	0.3 µl
3	dNTPs mix (2.5 mM each)	0.06 µl
4	Primer-F (25 pmoles.µl^{-1})	1 µl
5	Primer-R (25 pmoles.µl^{-1})	1 µl
6	Template DNA (50 ng.µl^{-1})	1 µl
7	Millipore sterile distilled water	14.74 µl
Total		20 µl

Table 2: PCR conditions for SSR

Sr. No	Steps	Temperature (ºC)	Duration
1	Initial Denaturation	94	3.0 min
2	Denaturation	94	30 sec
3	Annealing	57-59	1 min
4	Extension	72	1 min
Repeat the steps 2 to 4 for 40 times			
5	Final extension	72	5.0 min
6	Hold	4	--

Table 3: Raw data of run on Ion S5

Sr. no.	Parameter	1st run	2nd run	3rd run
1	Total number of reads	13,378,296	5,771,965	99,125,389
2	Total bases (Gb)	4.34	0.5	5.6
3	Mean length (bp)	325	286	197
4.	ISP Loading (%)	88	73	90

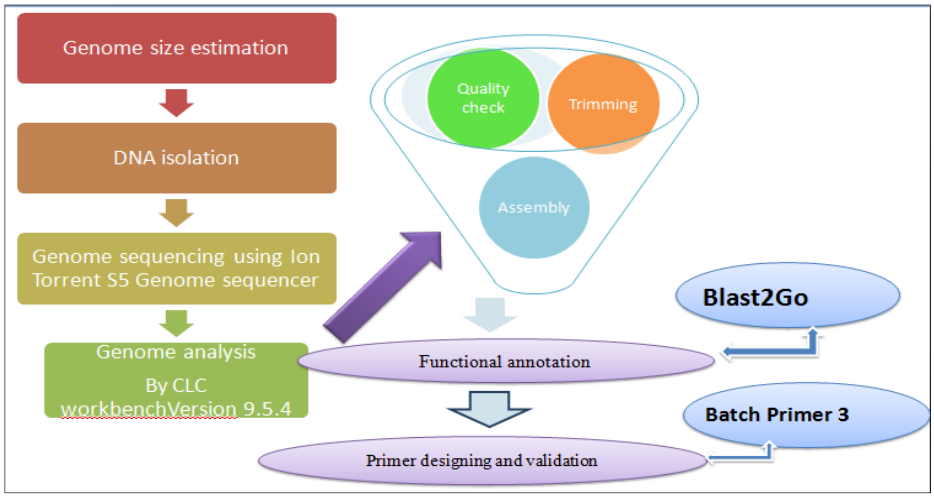

Fig 1: Overview of the workflow implemented in genome sequencing and primer designing

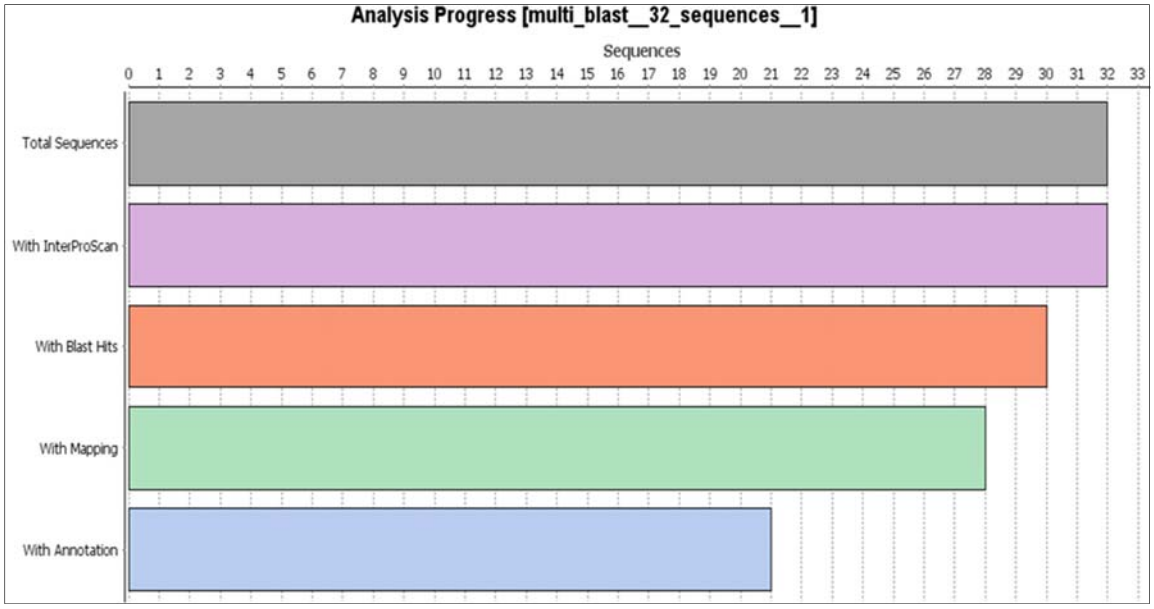

Fig 2: Analysis progress of Blast2Go result

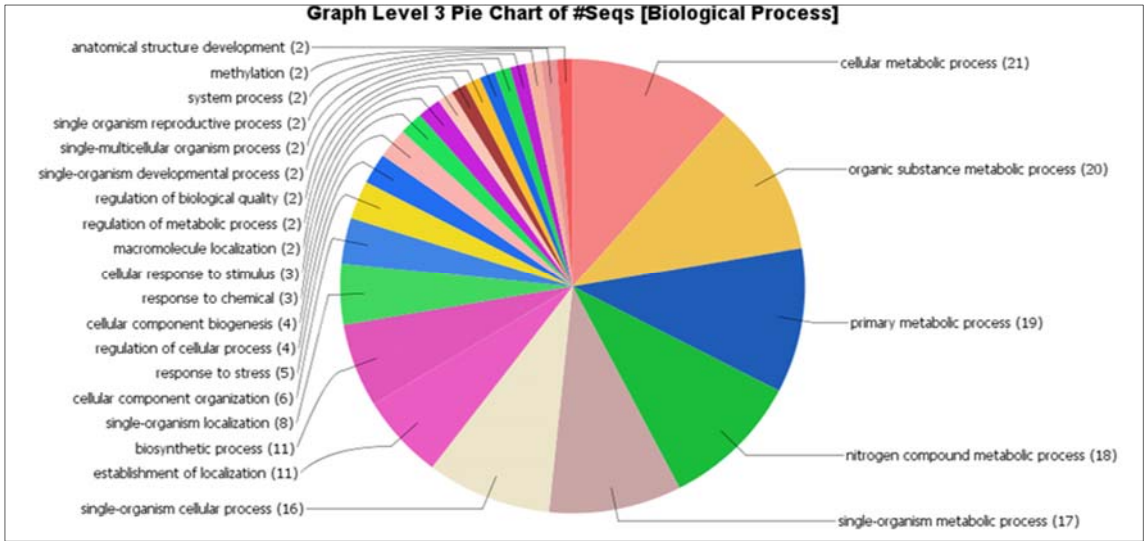

Fig 3: Pie chart at graph level 3 of biological process.

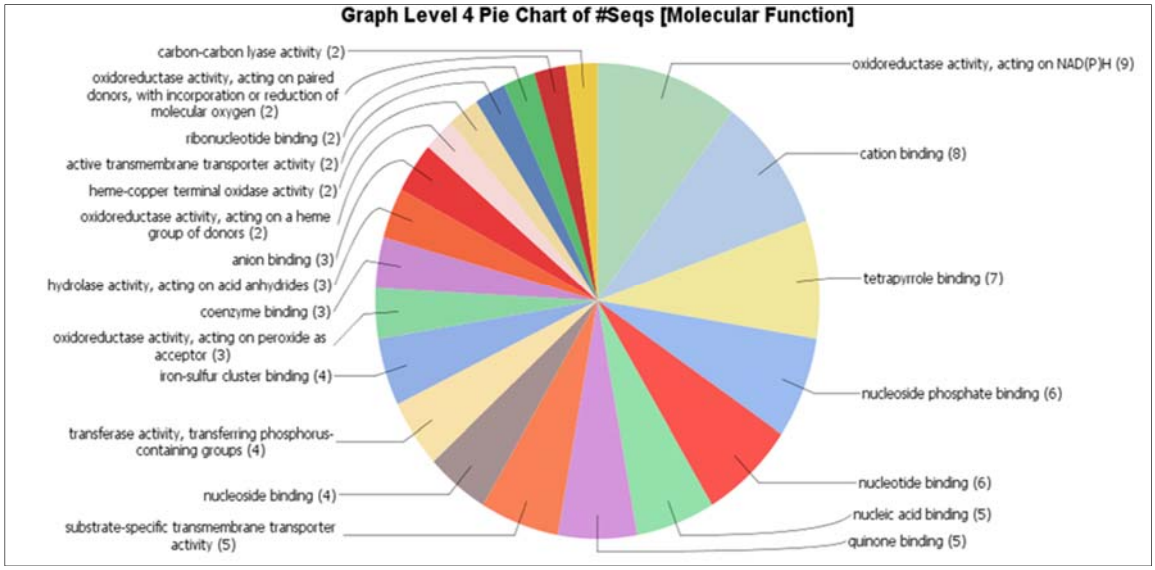

Fig 4: Pie chart at graph level 4 of molecular function.

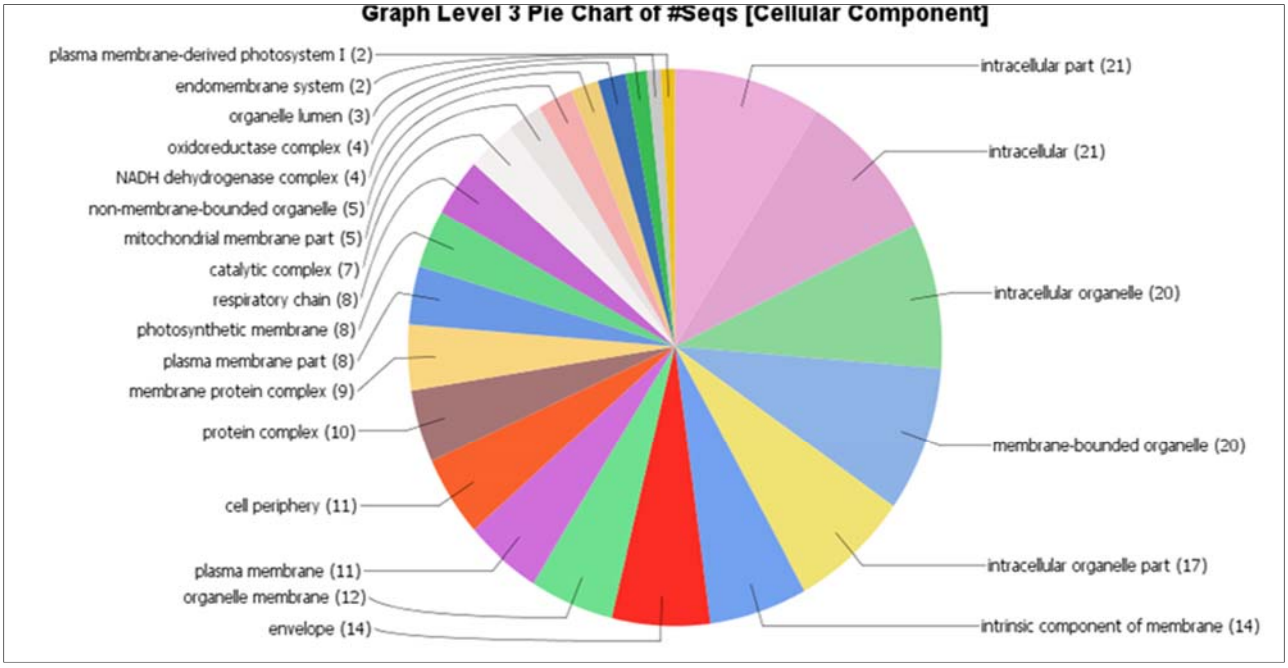

Fig 5: Pie chart at graph level 3 of cellular components.\

Fig 6: Primer amplification (1: Gujarat Dill-1, 2: Gujarat Dill-2, 3: Gujarat Dill-3, 4: Ajmer Dill-1, 5: Ajmer Dill-2)

Supplementary Information

Table 1: List of Putative/hypothetical gene's function identified

S. No.	Sequence ID	Putative function	Len	Identity (%)	#GO
1	Dill_D_(single)_trimmed_contig_17	NADH-ubiquinone oxidoreductase chain 5,NADH dehydrogenase subunit 5	6354	61.85	36
2	Dill_D_(single)_trimmed_contig_24	Ribosomal S1	6445	52.21	12
3	Dill_D_(single)_trimmed_contig_22	DNA-directed RNA polymerase subunit beta, RNA polymerase subunit beta	6464	55.69	19
4	Dill_D_(single)_trimmed_contig_31	Uncharacterized mitochondrial g01280 (ORF291)	6493	56	11
5	Dill_D_(single)_trimmed_contig_23	Photosystem I P700 chlorophyll a apo A1	10349	79.81	21
6	Dill_D_(single)_trimmed_contig_16	ATP synthase subunit a (ATPase 6)	5000	82.3	9
7	Dill_D_(single)_trimmed_contig_5	Holliday junction ATP-dependent DNA helicase	5290	48	11
8	Dill_D_(single)_trimmed_contig_28	Uncharacterized protein ORF104	5293	46	0
9	Dill_D_(single)_trimmed_contig_15	----	5380	0	17
10	Dill_D_(single)_trimmed_contig_19	Neuronal acetylcholine receptor subunit alpha-3 Flags: Precursor	5582	42	0
11	Dill_D_(single)_trimmed_contig_29	tRNA dimethylallyltransferase, Dimethylallyl diphosphate	6054	48.5	0
12	Dill_D_(single)_trimmed_contig_25	NADH dehydrogenase [ubiquinone] iron-sulfur 2 (mitochondrion)	6516	75.84	2
13	Dill_D_(single)_trimmed_contig_21	NADH-ubiquinone oxidoreductase chain 2	7233	55.84	0
14	Dill_D_(single)_trimmed_contig_11	TAR1_KLULA (TAR1)	7302	69.33	13
15	Dill_D_(single)_trimmed_contig_27	Root phototropism 2 (BTB POZ domain-containing RPT2)	7833	42	22
16	Dill_D_(single)_trimmed_contig_1	----	8081	0	13
17	Dill_D_(single)_trimmed_contig_8	ubiquinone oxidoreductase chain 6	11407	62.83	6
18	Dill_D_(single)_trimmed_contig_9	Photosystem II CP43 reaction center	8338	96.43	16
19	Dill_D_(single)_trimmed_contig_10	NADH-ubiquinone oxidoreductase chain 5	9025	61.42	23
20	Dill_D_(single)_trimmed_contig_20	Probable cytochrome c biosynthesis	9968	56.74	12
21	Dill_D_(single)_trimmed_contig_14	ATP synthase subunit a	5446	79.6	19
22	Dill_D_(single)_trimmed_contig_13	NADH-ubiquinone oxidoreductase chain 4 (NADH dehydrogenase subunit 4)	10462	58.07	23
23	Dill_D_(single)_trimmed_contig_4	YCF2_DAUCA (Ycf2)	17831	91.62	44
24	Dill_D_(single)_trimmed_contig_32	DNA repair RAD16 (ATP-dependent helicase RAD16)	5257	47.23	19

25	Dill_D_(single)_trimmed_contig_18	COX1_SOYBN (Cytochrome c oxidase subunit, Cytochrome c oxidase polypeptide I)	11400	75.39	11
26	Dill_D_(single)_trimmed_contig_26	DNA polymerase (S-1 DNA ORF 3)	7149	43.29	7
27	Dill_D_(single)_trimmed_contig_7	DNA-directed RNA polymerase subunit beta	10561	88.48	40
28	Dill_D_(single)_trimmed_contig_30	POLX_TOBAC (Retrovirus-related Pol poly from transposon TNT 1-94)	5505	48.83	19
29	Dill_D_(single)_trimmed_contig_2	ATP synthase subunit beta	42776	97.63	12
30	Dill_D_(single)_trimmed_contig_6	ATP synthase subunit alpha	12027	86.08	23
31	Dill_D_(single)_trimmed_contig_12	COX2_DAUCA (Cytochrome c oxidase subunit 2, Cytochrome c oxidase polypeptide II)	11075	75.39	6
32	Dill_D_(single)_trimmed_contig_3	TI214_TOBAC (TIC 214 Translocon at the inner envelope membrane of chloroplasts 214)	22288	88.33	32

References

1. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990; 12:13-15.
2. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. American Journal of Human Genetics. 1991; 49:746-756.
3. Ellegren H. Microsatellites: simple sequences with complex evolution. Nature Review Genetics. 2004; 5:435-445.
4. Guimaraes PM, Brasileiro ACM, Morgante CV, Martins ACQ, Pappas G *et al*. Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. *BMC Genomics.* 2012; 13:387.
5. Gupta PK, Balyan H, Sharma PC, Ramesh B. Microsatellites in plants: a new class of molecular markers. Current Science. 1996; 70:45-54.
6. Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK. Genetic mapping of a gene causing hypertensive rat. Cell. 1991; 67:213-224.
7. Liu T, Zhu S, Tang Q, Chen P, Yu Y, Tang S. De novo assembly and characterization of transcriptome using Illumina paired-end sequencing and identification of CesA gene in ramie (*Boehmerianivea* L. Gaud).BMC Genomics. 2013; 14:125.
8. Ma XH, Qin RL, Xing WB. Chromosome observations of some medical plants in Xinjiang. Acta Phytotaxonomica Sinica. 1984; 22:243-249.
9. Merida AM, Plaza JJG, Canada A, Blanco AM, Lopez MDCG, Rodrguez JM *et al*. *De Novo a*ssembly and functional annotation of the olive (*Oleaeuropaea*) transcriptome. DNA Research. 2013; 36:1-16.
10. Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira ML. Origin, evolution and genome distribution of microsatellites. Gentetic and molecular Biology. 2006; 29:294–307.
11. Parida SK, Kalia SK, Sunita K, Dala, V, Hemaprabha G, *et al*. Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. Theoretical and Applied Genetic. 2009; 118:327-338.
12. Powell W, Machray G, Provan J. Polymorphism revealed by simple sequence repeats. Trends in Plant Science. 1996; 1:215-222.
13. Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiology and Molecular Biology Review. 2008; 72:686-727.
14. Snell R,Aarssen LW. Life history traits in selfing versus outcrossing annuals: exploring the 'time-limitation' hypothesis for the fitness benefit of self-pollination. BioMed Central Ecology. 2005; 5:2.
15. Tanksley SD, Ahn S, Anderson JA, Sorrells ME. Homoeologous relationships of rice, wheat and maize chromosomes. Molecular and General Genetics. 1993; 241:483-490.
16. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nuclics Acids Research. 1989; 17:6463-6471.
17. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoritical applied Genetics. 2003; 106:411-422.
18. Tucker AO. 'Dill Grolier multimedia encyclopedia. Grolier onlinehttp://gme.grolier.com/cgi-bin/article?asset id=0085590-0.
19. Yazdanparast R, Alavi M. Anti hyperlipidaemic and anti hypercholesterolaemic effects of *Anethum graveolens* leaves after the removal of furocoumarins. *Cytobios.* 2001; 105(410):185-191.
20. Zhou Y, Gao F, Liu R, Jinchao F, Hongjie L. *de novo* sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*. BMC Genomics. 2012; 13:266-279.