International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2019; 7(2): 1034-1042

Received: 17-01-2019
Accepted: 21-02-2019

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

A Draft Genome Sequencing of Dill (*Anethum graveolens* L.) Using Ion Torrent Sequencing Technology

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. For the measurement of the genome size, flow cytometer (Accuri C6) was used, and genome size of the dill haploid cell was found 2.6 Gb approximately. Gujarat Dill-1 genotype was sequenced using next generation sequencing platform Ion S5 which yielded 10.44 Gb of raw data. The data was checked for quality. After quality check and its control total of 3.2 Gb of data was remained in the data set. This 3.2 Gb data was de novo assembled. *de novo* assembly of 15,940,402 reads had yield data of 321,805,497 base pairs (321 Mb). 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. In Blast2go total 32 sequence (32 contigs) 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. Quality filter of assembly yielded data of 306 Kb (306,184 bp). This data was selected for functional annotation. Through functional annotation of assembled data, total of 542 putative genes were identified which were involved in 498 putative function. It included biological processes, molecular function and cellular components. Sequence similarity was also tested based on protein domain conserved region through InterProScan. Out of 32 sequences, all most all sequences showed positive InterPro result. 47 IPS IDs were identified, among them 30 were IPS domain, 14 were IPS family and 4 were IPS active sites. Assembled sequences of dill (Gujarat dill-1) were divided in to four main classes which covered 18 sequences. Among them 11 sequences were grouped into class oxidoreductases followed by class transferases (3 sequences), class hydrolases (2 sequences) and class lyases (2 sequences).

Keywords: dill, *de novo*, assembly, contig, gene ontology

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) [4] and a typical out crossing species (Snell and Aarssen, 2005). The plant is a native species in Southwest Asia and is cultivated in Europe, India and the United States (Tucker, 2008) [9]. It is also successfully cultivated in Taiwan. Dill is also known as Sowa in India. In India dill is native to northern India. A variant called east Indian Dill or *Sowa* (*Anethum graveolens* var *sowa* Roxb. ex, Flem.) occurs in India and is cultivated for its foliage as a cold weather crop throughout the Indian sub-continent, Malaysian archipelago and Japan. *Anethum graveolens* L. is the sole species of the genus *Anethum*, though classified by some botanists in the related genus *Peucedanum* as *Peucedanum graveolens* L. 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) [10].
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, IonS5. 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. The broadest application of the new sequencing approaches to plant species may be whole genome sequencing (WGS) to reveal the full sequence and genetic structure of genomes. In WGS projects such as strawberry (Shulaev *et al*., 2011) [7] and wheat (Brenchley *et al*., 2012) [1] whole genomic DNA content was first randomly cut into fragments of different sizes. Then sequencing was carried out and the obtained reads were assembled using powerful bioinformatics tools. The WGS approach can be accomplished not only for resequencing, but also for *de novo* projects.

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. 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. This study provides complete map of genome, gene expression and metabolic network as well as its interplay which will be helpful to understand the biology from growth to seed setting in plant and be supportive base to build hypothesis and theory for future study. The Dill cultivar Gujarat Dill 1 used for this purpose.

Material and Methods

Sample collection and DNA isolation

The seeds of Dill genotypes Gujarat Dill 1 was obtained from Seed Spices Research Station, Sardarkrushinagar Dantiwada Agricultural University, Jagudan. Genomic DNA isolated from fresh seedling tissue by following modified CTAB method as described by Doyle and Doyle (1990) [2]. 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.

Genome size estimation by flow cytometer

For the measurement of the genome size, flow cytometer (Accuri C6) was used. The young leaves of dill were used in flow cytometry study, in which many intercalating dyes were used to such type study. But here (PI) propidium iodide dye was used (Nath *et al.*, 2014) [6]. After the following all procedure flow cytometer give a one pick position. Using the following equation, calculate of C-value and genome size. Dill samples and reference material were analyzed on an EPICS Elite ESP cytometer (Beckman- Coulter, Hialeah, FL, USA) with an air-cooled argon laser at 488 nm using 20 mW. *Catharanthus roseus* (1.51 pg /2C) were used as internal biological reference material.

$$\text{Sample 2C value(DNA pg or Mbp)} = \text{Reference 2C value} \times \frac{\text{sample 2C mean peak position}}{\text{reference 2C mean peak position}}$$

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® AM Pure® 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 Version 9.5.4. 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) [5]. 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) [3]. 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) [11].

Results and Discussion

Determination of genome size by flow cytometry

After following the procedure flow cytometer gave a one pick position. Using the following equation, calculate of C-value and genome size.

$$\text{Sample 2C value(DNA pg or Mbp)} = \text{Reference 2C value} \times \frac{\text{sample 2C mean peak position}}{\text{reference 2C mean peak position}}$$

Reference pick position= 389,565 pg (Figure 1)
Sample pick position = 1,361,465 pg (Figure 2)
Reference 2C value = 1.51 pg (*Catharanthus roseus*)
Sample 2C Value = 5.277 pg
Sample C value = 2.638 pg

After calculation sample C value = 2.638 pg, here 2C mention cause dill is a diploid plant, but genome size is estimated as haploid one. So sample C value should multiply with 978 to get the data in Mb and that value used fixing for the all samples or plants.

Sample C value = 2.638 pg

Sample genome size = 2.638 X 978 Mb = 2579.96 Mb approx

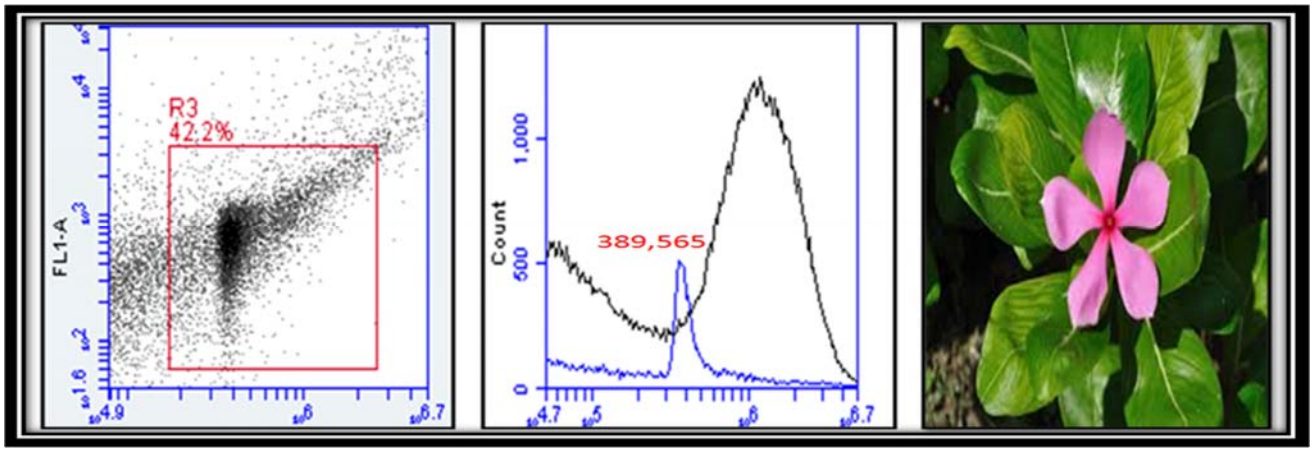

Fig 1: Scatter plot and pick position of *Catharanthus roseus* generated by accuri C6 flow cytometer (as a reference plant).

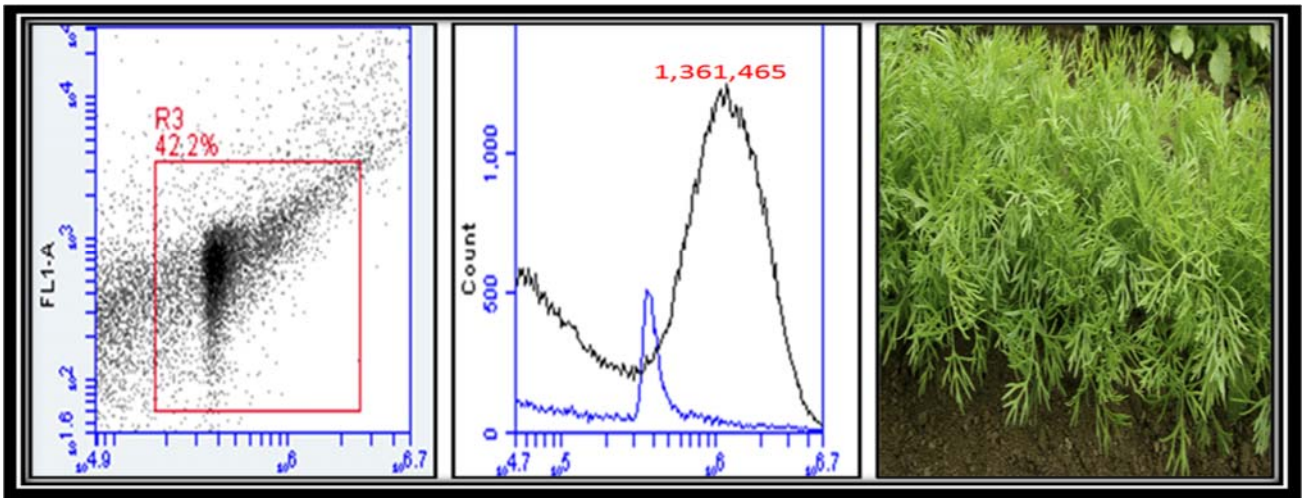

Fig 2: Scatter plot and pick position of dill (*Anethum graveolens*) generated by accuri C6 flow cytometer.

2.6 Gb

Genome sequencing of dill (*Anethum graveolens* L.)

Genome sequencing of Gujarat Dill-1 genotype of dill was done using next generation sequencing platform Ion torrent (Ion S5). For this pure genomic DNA was extracted by using kit, and then proceeded for the different steps to prepare the sample for the sequencing using protocol suggested by the instrument manual. Sequencing was followed in three replications. Genomic DNA concentration was measured using Qubit® Fluorometer using ds DNA assay kit. Concentration of DNA was 13.2, 11.9 and 12.7 ng/µl respectively for three different runs. Next step was DNA library construction, in this step gDNA libraries were constructed for which Ion Shear™ Plus 10X Reaction Buffer was used to fragment the genomic DNA in many fragments followed by adaptor ligation. Fragmented libraries were size selected using E-gel at 2% agarose. 400 bp library was eluted from the gel and then library was amplified, again final amplified library concentration was checked using Qubit® Fluorometer. The final library concentration was 1.14, 1.25, and 1.26 ng/µl in respect to total run. The final library was measured for their concentration and the concentration of DNA library was calculated in pmol. The concentration of the all samples was adjusted to 100 pmol by dilution with nuclease-free water and it was 4318, 4734 and 9545 pmol respectively for all three runs. Clonal based amplification is done using Ion one touch PCR (OT2) for every library followed by library enrichment. ISPs (sample) were loaded on each new Ion 530 chip for sequencing.

Total three runs were performed of Ion S5. 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.

Data analysis

Total of 10.44 Gb raw data was generated through sequencing. Raw sequence data were processed for quality assessment and other processing steps 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. It was found that per base sequence data was not good, some reads were having length more than 420 bp. The detail of reads and read length given in table 1. In order to get quality reads the data was trimmed. Total sequences in data sets before trimming were 16,179,010 with average read length 208.8 bp. After quality filter (trimming) total sequences in data sets remain 15,940,402 with average read length 203.9 bp.

***de novo* assembly of the raw data**

After reads filter, total sequence in data sets remains 15,940,402 which yield 3.2 Gb of data. *de novo* assembly (draft) of dill genome was carried out by using CLC genome workbench v9.5.4 *de novo* Assembler. *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 and 579 bp of average contig length. N25, N50 and N75 measurement of contigs were found 1279 bp, 661 bp, and 421 bp respectively via *de novo* assembly (Table 2). 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 (Table 3).

Table 1: Before trimming /after trimming report of reads

Sr. No.	Performed run	Before Trimming		After Trimming		Percentage of sequence remain after trimmed
		Number of reads	Average read length (bp)	Number of reads	Average read length (bp)	
1.	Dill	16,179,010	208.8	15,940,402	203.9	98.53%

Table 2: CLC genome work bench (v9.5.4) *de novo* assembler generated assembly data from genome sequencing of Gujarat Dill-1 genotype of dill

Sr. No.	Total sequences	Number of Contigs	Minimum contig length	Maximum contig length	Average length	N25	N50	N75
Contig measurements								
1.	321,805,497	555,413	200	10,247	579	1279	661	421

Table 3: CLC genome work bench (v9.5.4) Trimmed Assembly report

Sr. No.	Total sequences	Number of Contigs	Minimum contig length	Maximum contig length	Average length	N25	N50	N75
Contig measurements								
1.	306,184	32	5000	42,775	9,568	17,831	10,349	6,516

Blast2go analysis of genome sequencing data

Blast 2GO is all in one tool for functional annotation of (novel) sequences and the analysis of annotation data. Total 32 sequence (32 contigs) 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. In data distribution total of 32 sequences were analysed, in which sequences analysed with different parameter like blast, mapping and annotaion. Species distribution of assembled sample had shown highest similarity with *ubiuinone* with 10 blast hit (17.37%) followed by *Marchantia polymorpha* with 4 blast hit (6.9%) and *Nicotiana* species with 3 blast hit (5.21%) (Figure 3). A maximum blast hit was approximately 55 found in others species. Top hit species distribution shows the species distribution of our sample information, where our sample data shows Top-Hits with majorly with *Nicotiana sylvestra* belongs to solanaceae family which is followed by *Ostreid herpesvirus*, this virus belongs to Malacoherpesviridae and its host are molluscs. Percent similarity distribution of assembled sequences, out of total 32 sequences, 36% sequences did not show any similarity with available sequences in database while 30% sequences show n more than 75% similarity and 35% sequences shown less than 75% similarity with available sequences in database**.** Mapping database distribution of dill genome showed highest similarity with UniprotKB followed by Saccharomyces Genome Database and The Arabidopsis Information Resources (TAIR) (Figure 4). Annotation distribution of Go of dill plant genome varies from 0-38. Among total sequences maximum were those which were zero Go score. Some sequences were having 6-26 Go score. Data has been shown in (Figure 5).

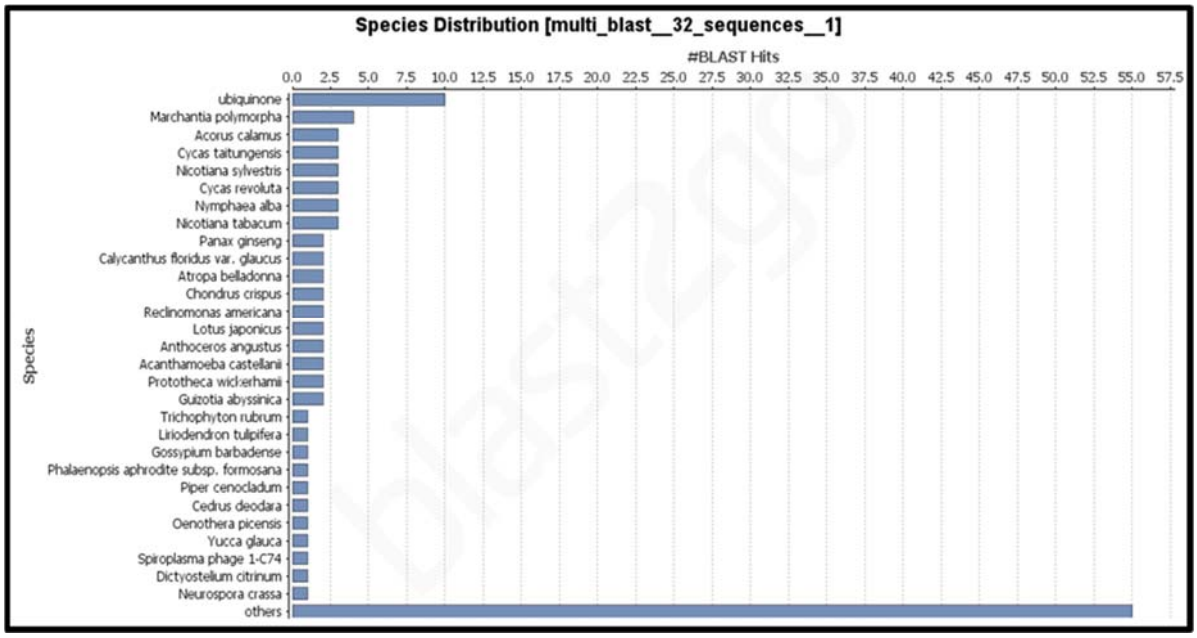

Fig 3: Species distribution of Blast2Go result.

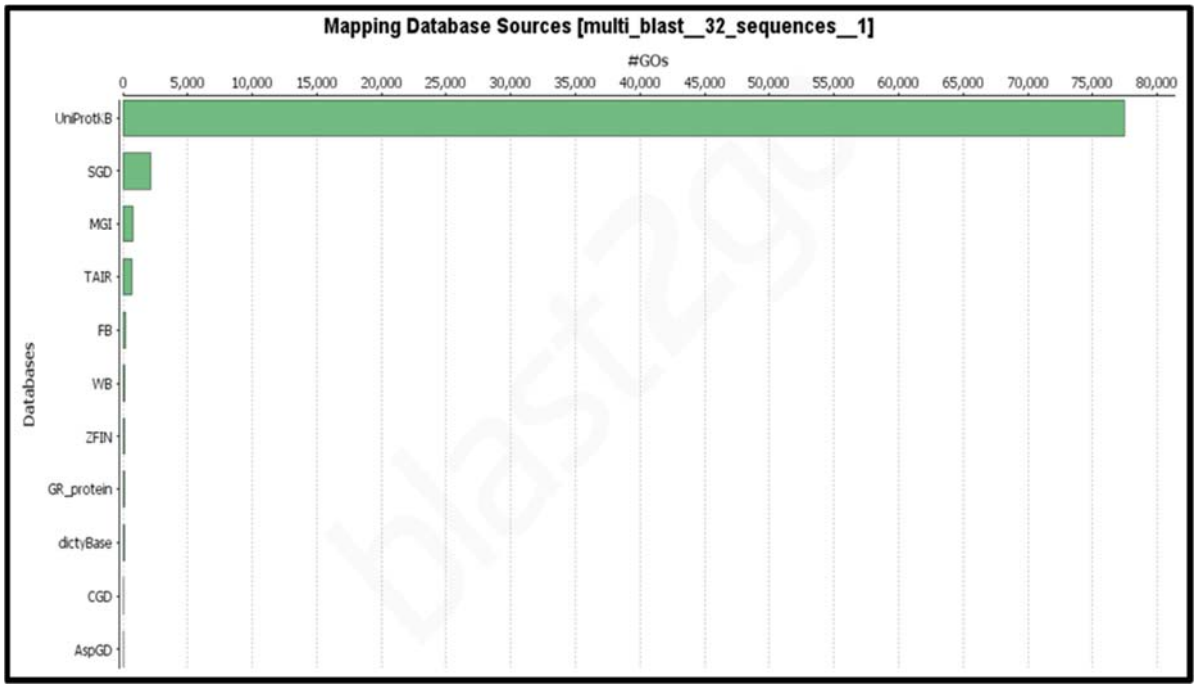

Fig 4: Mapping database distribution of input sequences in Blast2go analysis

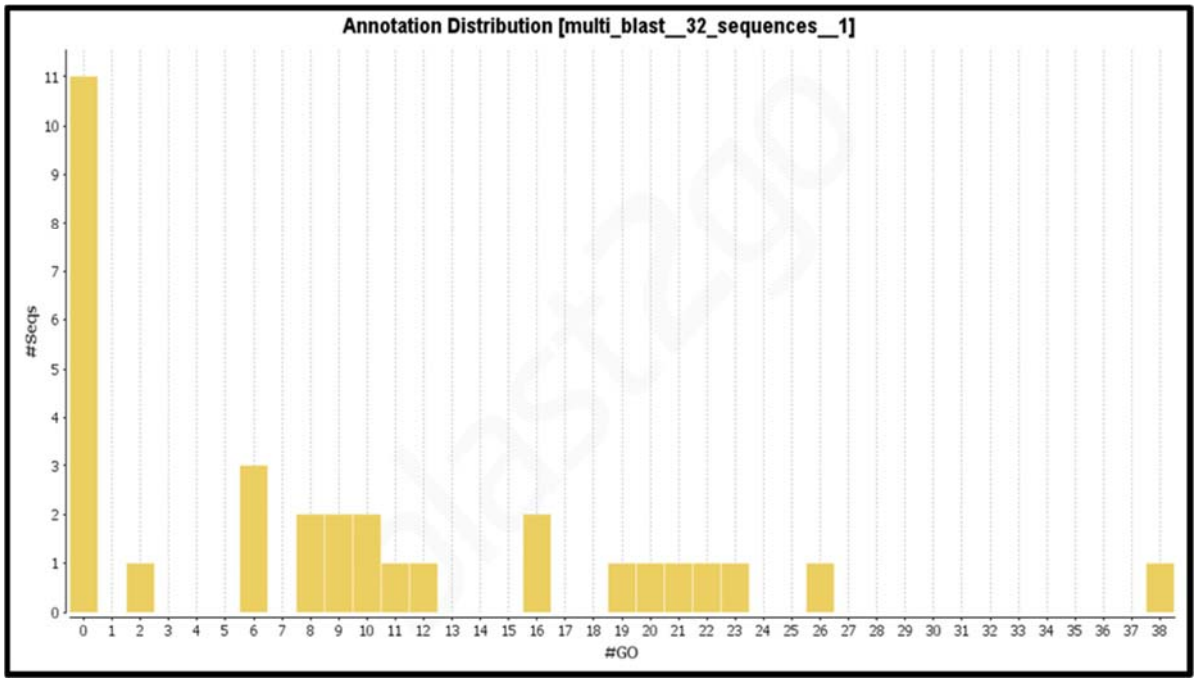

Fig 5: Annotation distribution of Blast2GO result.

Gene ontology (go) ids and sequence distribution

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) (Figure 6). 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 aerbic respiration. In cellular component maximum sequences having maximum Go count which respond to integral component of the membrane followed by integral component component of the plasma membrane.

Sequence similarity was also tested base on protein domain conserved region through InterProScan. Out of 28 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 from the sequences which were positive interproscan, among them 30 IPS domains were identified, 14 IPS family and 3 IPS active sites were identified respectively. Details of the IPS IDs were given in supplementary Table 1(A, B, C).

Table 1A: List of Interproscan Domains identified

S. No.	IPS Domain	#Seq	Sequence
1	(IPR027417) P-loop containing nucleoside triphosphate hydrolase	4	Dill_D_(single)_trimmed_contig_2, Dill_D_(single)_trimmed_contig_4, Dill_D_(single)_trimmed_contig_6, Dill_D_(single)_trimmed_contig_32
2	(IPR025452) Domain of unknown function DUF4218	3	Dill_D_(single)_trimmed_contig_28, Dill_D_(single)_trimmed_contig_27, Dill_D_(single)_trimmed_contig_5
3	(IPR004100) ATPase, F1/V1/A1 complex, alpha/beta subunit, N-terminal domain	2	Dill_D_(single)_trimmed_contig_2, Dill_D_(single)_trimmed_contig_6
4	(IPR003593) AAA+ ATPase domain	2	Dill_D_(single)_trimmed_contig_2, Dill_D_(single)_trimmed_contig_4
5	(IPR000194) ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain	2	Dill_D_(single)_trimmed_contig_2, Dill_D_(single)_trimmed_contig_6
6	(IPR033732) ATP synthase, F1 complex, alpha subunit nucleotide-binding domain	1	Dill_D_(single)_trimmed_contig_6
7	(IPR025312) Domain of unknown function DUF4216	1	Dill_D_(single)_trimmed_contig_15
8	(IPR024937) Domain X	1	Dill_D_(single)_trimmed_contig_10
9	(IPR007120) DNA-directed RNA polymerase, subunit 2, domain 6	1	Dill_D_(single)_trimmed_contig_7
10	(IPR003959) ATPase, AAA-type, core	1	Dill_D_(single)_trimmed_contig_4
11	(IPR033944) Cytochrome c oxidase subunit I domain	1	Dill_D_(single)_trimmed_contig_18
12	(IPR008972) Cupredoxin	1	Dill_D_(single)_trimmed_contig_12
13	(IPR007641) RNA polymerase Rpb2, domain 7	1	Dill_D_(single)_trimmed_contig_7
14	(IPR029014) NiFe hydrogenase-like	1	Dill_D_(single)_trimmed_contig_25
15	(IPR023616) Cytochrome c oxidase-like,	1	Dill_D_(single)_trimmed_contig_18

	subunit I domain		
16	(IPR007645) RNA polymerase Rpb2, domain 3	1	Dill_D_(single)_trimmed_contig_7
17	(IPR024034) ATPase, F1/V1 complex, beta subunit, C-terminal	1	Dill_D_(single)_trimmed_contig_2
18	(IPR025724) GAG-pre-integrase domain	1	Dill_D_(single)_trimmed_contig_30
19	(IPR023366) ATP synthase subunit alpha-like domain	1	Dill_D_(single)_trimmed_contig_6
20	(IPR000330) SNF2-related, N-terminal domain	1	Dill_D_(single)_trimmed_contig_32
21	(IPR007642) RNA polymerase Rpb2, domain 2	1	Dill_D_(single)_trimmed_contig_7
22	(IPR014001) Helicase superfamily 1/2, ATP-binding domain	1	Dill_D_(single)_trimmed_contig_32
23	(IPR011759) Cytochrome C oxidase subunit II, transmembrane domain	1	Dill_D_(single)_trimmed_contig_12
24	(IPR012337) Ribonuclease H-like domain	1	Dill_D_(single)_trimmed_contig_30
25	(IPR000793) ATP synthase, alpha subunit, C-terminal	1	Dill_D_(single)_trimmed_contig_6
26	(IPR004332) Transposase, MuDR, plant	1	Dill_D_(single)_trimmed_contig_29
27	(IPR014724) RNA polymerase Rpb2, OB-fold	1	Dill_D_(single)_trimmed_contig_7
28	(IPR001584) Integrase, catalytic core	1	Dill_D_(single)_trimmed_contig_30
29	(IPR002541) Cytochrome c assembly protein	1	Dill_D_(single)_trimmed_contig_20
30	(IPR001750) NADH: quinone oxidoreductase/Mrp antiporter, membrane subunit	1	Dill_D_(single)_trimmed_contig_21

Table 1B: List of Interproscan Family identified

S. No.	IPS Family	#Seqs	Sequence
1	(IPR004242) Transposon, En/Spm-like	2	Dill_D_(single)_trimmed_contig_28, Dill_D_(single)_trimmed_contig_27
2	(IPR003918) NADH: ubiquinone oxidoreductase	1	Dill_D_(single)_trimmed_contig_13
3	(IPR015712) DNA-directed RNA polymerase, subunit 2	1	Dill_D_(single)_trimmed_contig_7
4	(IPR005722) ATP synthase, F1 complex, beta subunit	1	Dill_D_(single)_trimmed_contig_2
5	(IPR003569) Probable cytochrome c biosynthesis protein, plants	1	Dill_D_(single)_trimmed_contig_20
6	(IPR008543) Uncharacterised protein family Ycf2	1	Dill_D_(single)_trimmed_contig_4
7	(IPR003567) Cytochrome c-type biogenesis protein	1	Dill_D_(single)_trimmed_contig_20
8	(IPR005294) ATP synthase, F1 complex, alpha subunit	1	Dill_D_(single)_trimmed_contig_6
9	(IPR000883) Cytochrome c oxidase subunit I	1	Dill_D_(single)_trimmed_contig_18
10	(IPR000568) ATP synthase, F0 complex, subunit A	1	Dill_D_(single)_trimmed_contig_14
11	(IPR000932) Photosystem antenna protein-like	1	Dill_D_(single)_trimmed_contig_9
12	(IPR008896) Protein TIC214	1	Dill_D_(single)_trimmed_contig_3
13	(IPR005869) Photosystem II CP43 reaction centre protein	1	Dill_D_(single)_trimmed_contig_9
14	(IPR002092) DNA-directed RNA polymerase, phage-type	1	Dill_D_(single)_trimmed_contig_26

Table 1C: List of Interproscan Functional sites identified

S. No.	IPS Site	#Seqs	Sequence
1	(IPR020003) ATPase, alpha/beta subunit, nucleotide-binding domain, active site	2	Dill_D_(single)_trimmed_contig_2, Dill_D_(single)_trimmed_contig_6
2	(IPR007121) RNA polymerase, beta subunit, conserved site	1	Dill_D_(single)_trimmed_contig_7
3	(IPR023011) ATP synthase, F0 complex, subunit A, active site	1	Dill_D_(single)_trimmed_contig_14

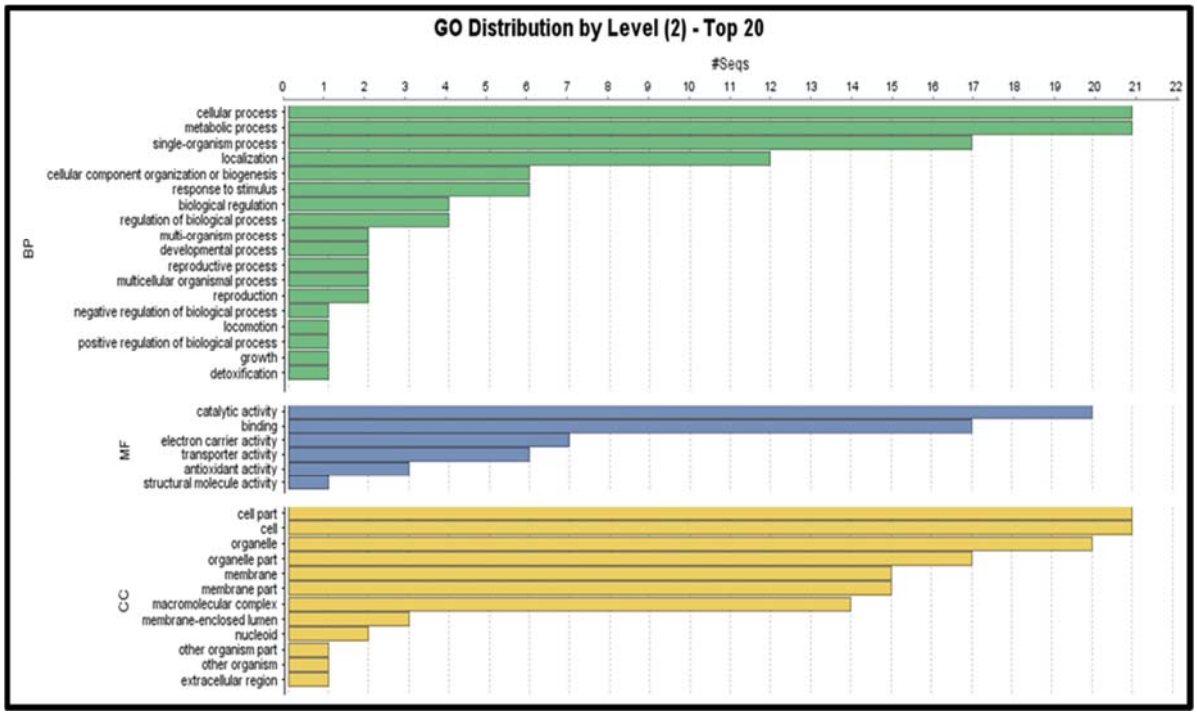

Fig 6: GO distribution by level (2) of top-20 for dill sequences

Enzyme class distribution

Assembled sequences of dill (Gujarat dill-1) were divided in to four main classes which covered 18 sequences. Among them 11 sequences were grouped into class oxidoreductases followed by class transferases (3 sequences), class hydrolases (2 sequences) and class lyases (2 sequences) (Figure 7).

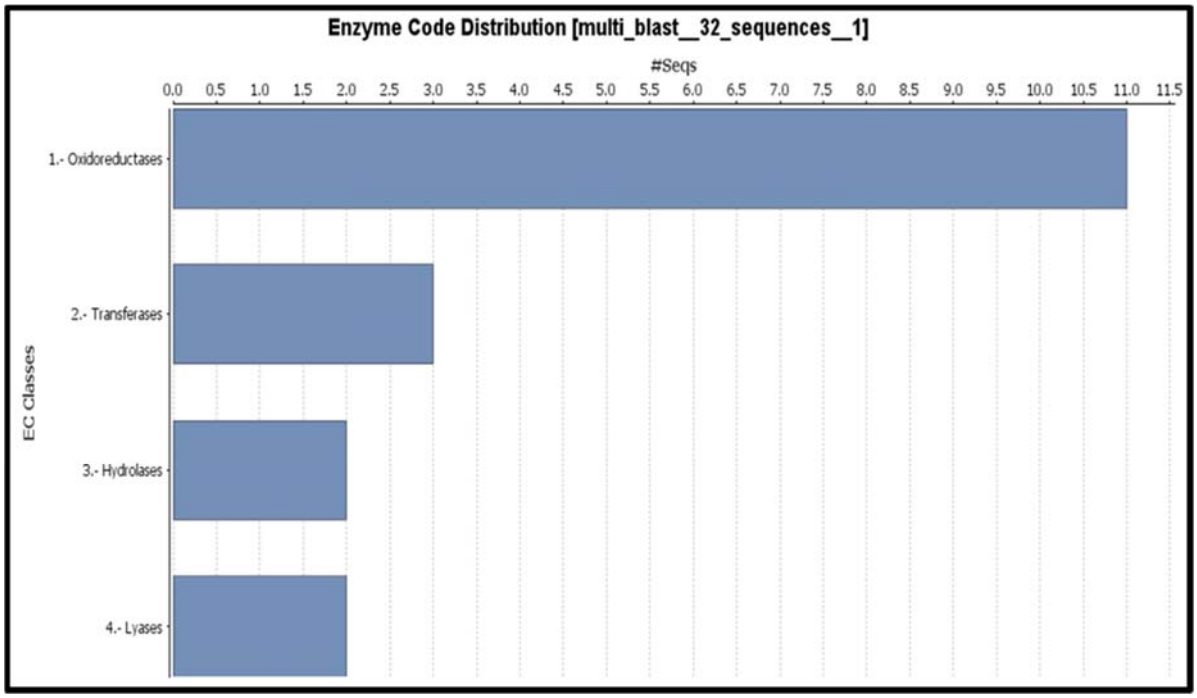

Fig 7: Number of sequences and enzymes in each class.

Conclusion

This information constitutes an important resource for genetic and evolutionary studies. Whole genome has a great importance in the field of research to understand metabolic pathways, genetic basis of disease and its relation with the risk of developing susceptibility against any adverse condition. Genomic co-dominant markers like SSR can be developed from genome sequencing which become useful for the QTL mapping programme. Whole genome also give an idea about evolution of genetic structure and function as well as study of molecular phylogeny of individual species with the genome of evolutionary proximate species. It also provides data through which we can identify gene and functional elements of genome and give basis for annotation of complete plant genome. Data and information generated

after analysis of whole genome can be used for transcriptome study.

Reference

1. Brenchley R, Spannag M, Pfeifer M, Barker GL, Amore R, Allen AM *et al*. Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature. 2012; 491:705-710.
2. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990; 12:13-15.
3. 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.
4. Ma XH, Qin RL, Xing WB. Chromosome observations of some medical plants in Xinjiang. Acta Phytotaxonomica Sinica. 1984; 22:243-249.
5. 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.
6. Nath S, Sanjaya KM, Sumita J. An improved method of genome size estimation by flow cytometry in five mucilaginous species of hyacinthaceae. International Scoiety for Advancement of Cytometry. 2014; 85A:833-840.
7. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL *et al*. The genome of woodland strawberry (*Fragariavesca*). Nature genetics. 2011; 43:109-116.
8. 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.
9. Tucker AO. ''Dill'' Grolier multimedia encyclopedia. Grolier online, 2008. http://gme.grolier.com/cgi-bin/article?asset id=0085590-0.
10. 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.
11. 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.