CHAPTER-V

SUMMARY AND CONCLUSION

Celery is one of the most ancient and important seed spice crop. Celery is commonly known as *ajmod* or *karnaulli* and has a diploid chromosome 2n=2x=22. Celery belongs to the Apiaceae family. Celery has been widely used as both food and medication since middle ages. It has many health promoting properties. Very limited efforts have been made towards its genetic improvement and phytochemical study. To initiate any genetic improvement or breeding program the knowledge of the amount of genetic variability existing in the material, type of gene action governing the yield and its components and a clear-cut understanding of other important genetic parameters are essential. Therefore, this research has been conducted, so that one can get the deep knowledge about its genome and this can be useful for successful exploitation of the genetic parameters and formulation of efficient breeding program. The present investigation was undertaken to sequence the genome and to develop SSRs from genomic libraries and thereafter to validate SSR.

5.1 Determination of genome size

Ajmer Celery-1 was used for the sequencing process. The experimental material was procured from the National Research Centre on Seed Spices, Ajmer (Rajasthan), India. Genome size estimation was carried out using flow cytometer (Accuri C6). Young leaves were used for the sample preparation using Galbraith's cell lysis buffer. Genome size was estimated as "C" which means for haploid cell, so the genome size of dill (*Apium graveolens* L.) measured approximately 3.94 Gb.

5.2 Genome sequencing of celery (*Apium graveolens* L.) Ajmer Celery-1

Genome sequencing of Ajmer Celery-1 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. For *in vitro* amplification of the single stranded library of quantitated DNA fragments, the emulsion (emPCR) was performed, further, to generate a library of clonally amplified, bead-immobilized, single-stranded DNA fragments representative of the entire DNA by using OT2 emPCR Kit. Total volume was subjected for

enrichment and after enrichment to the emPCR for clonally amplified bead ligated DNA fragments. After the completion of enrichment, ISPs were loaded on each new Ion 530 and Ion 540 chip for sequencing. Two runs were performed of Ion S5. 6.01 Gb data was generated during the first run of sample. ISPs loading were 92% on the Ion 530™ chip and 19,533,024 bp were the total reads generated after the removal of polyclonal (25%) and low quality data (23%), with mean length 308 bp. The percentage of usable reads was 58%. Second run was carried out in Ion 540 ™ chip and 19.6 Gb data was generated in total. This run was preceded by using barcodes for different samples, out which celery was barcoded with IonXpress_001 barcode name. ISPs loading were 90% on the chip. The total number of reads generated was 99,125,389 bp after removing polyclonal (18%) and low quality data (7%), with mean length 198 bp. For celery genome, total of 3.3 Gb data was generated and there were 16,179,010 numbers of reads with 209 bp mean read length.

Two runs were carried out in this research study. The data produced after the first run was 6.01 Gb. The second run was preceded by using barcodes for specific samples and that generated 3.3 Gb data. Raw data (reads) generated from sequencing (NGS) were assessed with the help of CLC Genomics Workbench v9.5.4, in which per-sequence coverage analysis was carried out. The total sequences were 27,361,042 bp (combined 2 runs) and total nucleotides were 5,412,056,746 bp. Later, trimming process generated a trim report describing the number of reads to be 27,218,172 bp with average read length of 194.4 bp.

5.3 EXPERIMENTAL ANALYSIS

After reads filtering, 4.4 Gb of high-quality data was generated. *De novo* assembly of celery was carried out with the help of CLC genomics workbench (Version 9.5.4) *de novo* Assembler. The *de novo* assembly of 27,218,172 reads yielded assembled reads of 441,399,737 from celery sample with 827,971 numbers of contigs. This comprised of 32.9% adenine, 17.2% cytosine, 17.1% guanine and 32.9% thymine. The minimum, maximum and average length of contigs were 200, 15978 and 533 respectively. N25, N50 and N75 of contigs were found to be 1014 bp, 591 bp, and 395 bp respectively. N50 contig size is a value such that at least half of the genome is contained in contigs of that size larger. N25 and N75 are the equivalent values at which 25% and 75% of the genome is covered by contigs of that size larger.

Blast2GO is an all in one tool for functional annotation of (novel) sequences and the analysis of annotation data. It is suitable platform for functional genomics research in non-model species. In *de* novo assembly, 827,971 numbers of contigs were generated. Somehow, it is a tedious job to perform annotation 827,971 contigs. Therefore annotation for only selected contigs was carried out. Contigs with size more than 5000 bp were filtered out and used for the annotation (375 sequences). Total 375 sequences were functionally annotated out of which 374 showed positive InterProScan and 355 got Blast hits. 351 and 290 sequences were mapped and annotated respectively.

The assembled samples showed highest similarity with UDP-forming with 74 Blast hits followed by Protein-PII with 73 Blast hits, while result of Top-Hit species distribution showed equal similarity with ADP-ribose, Cauliflower mosaic virus, ADP-forming, Phalaenopsis aphrodite subsp. Formosana, Manihot esculenta, marchantia polymorpha, UDP-forming, Panax ginseng and Ostreid herpesvius 1 with one Blast hit. E-value describes the number of hits expected by chance when searching a database of particular size. At 1E-01 the highest hits obtained is 3900. Negligible hits were seen after 1E-100. High scoring segment pair (HSP) distribution indicated that 35,900 hits distribution was found between hits 0 to 2. Mapping database distribution of celery genome showed highest similarity with UniprotKB (375,000 GOs) followed by Saccharomyces Genome Database (SGD) (85,000 GOs) and The Arabidopsis Information Resources (TAIR) (75,000 GOs). Annotation distribution of GO graph ranged from 0 to 90. 0-50 GOs were obtained for various sequences. Very fewer sequences were seen with GOs after 100 score.

Sequence similarity was also tested based on the protein domain conserved region through InterProScan. 320 sequences showed positive InterPro result while 50 sequences did not show any InterPro results and 130 sequences was scanned with GOs.

The assembled sequences of celery were divided into six main classes of enzymes. Among all the sequences, 22 sequences were grouped into Oxidoreductases class followed by Transferases class (66 sequences), Hydrolases class (95 sequences), Lyases class (17 sequences) Isomerases class (11 sequences) and Ligases class (07 sequences).

KEGG is a multispecies, integrated resource consisting of genomic, chemical and network information with cross-references to numerous outside databases and containing a complete set of building blocks (genes and molecules) and wiring diagrams (biological pathways) to represent cellular functions. In present study KEGG analysis gave 76 pathways.

5.4 OVERVIEW OF SSRs MARKERS DEVELOPMENT

In this study, genomic SSRs of celery were identified and developed by draft genome. This novel polymorphism SSR markers were effectively applied to genotype and genetic diversity analyses. The SSR primers were designed using BatchPrimer3 (version 1.0) online software. For the identification of SSRs, FASTA file of 375 contigs were used. Total 506 primers met the following parameters: 100–300 final product length (optimal 200 bp), primer size from 19 to 23 bp (optimal 21 bp) and GC content 30%–70% (optimal 50%); the annealing temperature was set at 50–60 °C (optimal 55 °C). In order to validate the SSR primers in 3 different genotypes of celery (Ajmer Celery-1, Amy vishnagar and Tall Utah), 25 SSRs were selected manually. Criteria used to select the SSR primers for validation purpose was based on the melting temperature (Tm) and GC content (GC%) of the primers. Those primers were selected which were having melting temperature above 55ºC and GC% above 30%.

5.5 VALIDATION OF SSR MARKERS

Out of 100 primers, 68 SSRs primers showed amplification and gave a total of 101 bands. The SSR primer FEN5003 and FEN6003 produced maximum number of 3 bands. Out of 101 bands, 43 bands were polymorphic with an average of 0.63 bands per primer and 58 bands were monomorphic. Among the 43 polymorphic bands, 29 bands were sheared polymorphic while 14 bands were unique-polymorphic (Table 4.14). The amplified fragments ranged from 90-1272 bp. The largest amplicone of 1272 bp was amplified by SSR primer AJ22 and smallest fragment of 90 bp was found with SSR primer CELL18 and DILL7.

CONCLUSION

In the end whole result is concluded as; genome size of celery (*Apium graveolens* L.) haploid cell that is C value was approximately 3.94 Gb. Ajmer Celery-1 genotype was sequenced using next generation sequencing platform Ion S5 which yielded 4.4 Gb of raw data. This data was *de novo* assembled. The *de novo* assembly of 27,218,172 reads yielded assembled reads of 441,399,737 (441 Mb) from celery sample with 827,971 number of contigs. In order to maintain the quality of assembled data, contigs were filtered, in which only those contigs which had sequence above and equal to 5000 bp were selected for the annotation. Sequence similarity was tested based on protein domain conserved region through InterProScan. Out of 375 sequences, 320 sequences showed positive InterPro result. The assembled sequences of celery were divided into six main classes of enzymes. Among all the sequences, 22 sequences were grouped into Oxidoreductases class followed by Transferases class (66 sequences), Hydrolases class (95 sequences), Lyases class (17 sequences) Isomerases class (11 sequences) and Ligases class (07 sequences). In order to validate the SSR primers in 3 different genotypes of celery (Ajmer Celery-1, Amy Vishnagar and Tall Utah), 25 SSRs were selected. Total 100 primers were used, 25 each from Celery, Ajwain, Fennel and Dill. Ajwain, Fennel and Dill are the crops from Apiaceae family. So cross-species primers were used for the validation process. Among 100 primer pairs, 68 primers were amplified and gave a total of 101 bands. Out of 101 bands, 43 bands were polymorphic with an average of 0.63 bands per primer and 58 bands were monomorphic. Among the 43 polymorphic bands, 29 bands were sheared polymorphic, while 14 bands were unique-polymorphic.