



E-ISSN: 2278-4136  
P-ISSN: 2349-8234  
JPP 2019; 8(3): 1779-1794  
Received: 14-03-2019  
Accepted: 18-04-2019

**Shubha R**  
Department of Genetics and  
Plant Breeding, Junagadh  
Agricultural University,  
Junagadh, Gujarat, India

**Rukam S Tomar**  
Department of Genetics and  
Plant Breeding, Junagadh  
Agricultural University,  
Junagadh, Gujarat, India

## Transcriptome sequencing, *De novo* assembly from three developmental growth stages in fenugreek (*Trigonella foenum-graecum* L.)

**Shubha R and Rukam S Tomar**

### Abstract

Fenugreek (*Trigonella foenum-graecum* L.) popularly known as “Methi” belongs to family fabaceae and subfamily papilionaceae. Gujarat Methi Variety 2 was selected for transcriptome study from the three developmental stages that were vegetative stage (20-30 DAS), reproductive stage (50-60 DAS) and maturity stage (80-90 DAS) which were sequenced using Ion S5 genome sequencing machine. The *de novo* assembly yielded assembled reads from vegetative stage, total reads were 59,306, and the largest contig was 5773 bp and total number of base pairs were 2,562,931,387, with a consensus length of 39 bp. Similarly, assembly yielded assembled reads of 30,648,677 from reproductive stage with a consensus length of 40 bp, total number of contigs generated in reproductive stage, were 57,415 and the largest contig was 7,167 bp, total number of base pairs were 3,143,890,944. In maturity stage assembled reads of 16,924,096 with a consensus length of 81 bp were generated. Total number of contigs generated in maturity stage were, 46,721 and largest contig was 4,434 bp and the total number of base pairs were 3,205,640,970.

**Keywords:** Fenugreek, developmental stages, transcriptome sequencing, *de novo* assembly

### Introduction

Fenugreek (*Trigonella foenum-graecum* L.) popularly known as “Methi” belongs to family fabaceae and subfamily papilionaceae. It is an important spice crop largely grown in the northern India during rabi season. Fenugreek is used as an herb (dried or fresh leaves), spice (seeds), and vegetable (fresh leaves, sprouts, and microgreens). Cuboid-shaped, yellow- to amber-colored fenugreek seeds are frequently encountered in the cuisines of the Indian subcontinent, used both whole and powdered in the preparation of pickles, vegetable dishes, dal, and spice mixes such as panch phoron and sambar powder. They are often roasted to reduce bitterness and enhance flavour (BBC – 2017) <sup>[1]</sup>.

Fruit is a curved seed-pod, with ten to twenty flat and hard, yellowish-brown seeds. They are angular- rhomboid, oblong or even cubic, and have a deep furrow dividing them into two unequal lobes. It contains three important chemical constituents with medicinal value; *i.e.*, 1) steroidal sapogenins, 2) galactomannans and 3) isoleucine. These constituents have placed fenugreek among the most commonly recognized “nutraceutical” or health food products. It has been reported that fenugreek contains 81 phytonutrients and diosgenin, a steroid saponin found in fenugreek seeds, is the most bioactive component. Diosgenin is often used as a raw precursor for the production of steroidal drugs and hormones such as testosterone, glucocorticoids and progesterone. Studies reveal that a maximum level of diosgenin [(25R)-5-spirosten- 3h-ol] is found to be in young leaves (20mg/g dry weight) and in mature seeds with the percentage range from 0.28% - 0.92%. Reported that steroidal sapogenins were effective agents for the treatment of hypocholesterolemia, a disorder often associated with diabetes (Raju *et al.*, 2004) <sup>[2]</sup>.

In current study, Gujarat Methi Variety 2 was selected for transcriptome study from the three developmental stages that were vegetative stage (20-30 DAS), reproductive stage (50-60 DAS) and maturity stage (80-90 DAS) (Figure 1) which were sequenced using Ion S5 genome sequencing machine.

### Transcriptome sequencing and *De novo* assembly

The transcriptome is the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition. The key aims of transcriptomics are: to catalogue all species of transcript, including mRNAs, non-coding RNAs and small RNAs; to determine the transcriptional structure of genes, in terms of their start sites, 5' and 3' ends, splicing patterns and other post-transcriptional modifications; and to quantify the changing

### Correspondence

**Shubha R**  
Department of Genetics and  
Plant Breeding, Junagadh  
Agricultural University,  
Junagadh, Gujarat, India

expression levels of each transcript during development and under different conditions. *De novo* genome sequencing is used to sequence uncharacterized genomes where there is no reference sequence available, or *De novo* transcriptome assembly is the method of creating a transcriptome without the aid of a reference genome. *De novo* assembly is a detailed genetic analysis of any organism where no reference is available and *de novo* sequencing has been performed.

## 2. Materials and Methods

Present investigation on “Transcriptome sequencing, *de novo* assembly from three developmental growth stages in fenugreek (*Trigonella foenum-graecum* L.)” was undertaken at the Junagadh Agricultural University (JAU), Junagadh during 2015-16 and 2016-17. Laboratory studies on various aspects were carried out at Biotech Cell, Department of Biotechnology, College of Agriculture, Junagadh Agriculture University, Junagadh, Gujarat. Fenugreek seeds of GMV-2 (Gujarat Methi Variety-2) were sown in plot, under the natural environmental condition, for the collection of the samples from three developmental stages *i.e.*, vegetative stage (20-30 DAS), reproductive stage (50-60 DAS) and maturity stage (70-80 DAS). All the tissues were kept in RNA later at -20°C till they were used for isolation of RNA.

### Protocol for RNA isolation by TRIZOL method

Total RNA was isolated from GMV-2 variety of fenugreek at three developmental stages that are vegetative stage, reproductive stage and maturity stages at an interval of 30 days. Total RNA was isolated from the leaves of fenugreek plant using “TRIZOL RNA isolation protocol” and mRNA was purified with Magnetic Oligo (dT) beads.

### A. Homogenization of Tissues

To obtain high yield of RNA, grinding is a critical point in the extraction of genetic material. So, grinding should be fine as much as possible and temperature should be low as much as possible to avoid degradation. Took 50 mg of sample and grinded it with the help of liquid nitrogen in mortar and pestle that was sterilized and baked, then add 1 ml of Trizol/Tri reagent/Tri extract per 50 - 100 mg of tissue in 2 ml microcentrifuge tube. Sample volume should be less than 100 µl. Mixed the sample by vigorously shaking by briefly vortexing until the sample is thoroughly resuspended.

### B. Phase Separation

In this step, incubated the samples for 5 minutes at room temperature. After that added 0.2 ml of chloroform to each tube and shaken vigorously by hand for 15 seconds, then incubated the samples for 5 minutes at room temperature. After the incubation centrifuge for 15 minutes at 12,000 ×g (RCF) at 4 °C or room temperature.

### C. RNA Precipitation

Transferred the upper aqueous phase to a fresh tube. Then added 0.5 ml of isopropyl alcohol to precipitate RNA (If this RNA will be used for RT-PCR, first add 50 µl isopropyl alcohol to precipitate RNA, mixed it and incubate samples at room temperature for 5 min and centrifuge at 12,000 ×g (RCF) for 10 minutes at 4 °C or room temperature), Mix it gently, incubate the samples at room temperature for 10 minutes and centrifuge at 12,000 ×g (RCF) for 10 minutes at 4°C or room temperature. The RNA will form a white minute pellet on side or bottom of the tube.

### D. RNA Washing

Discard the supernatant gently and wash the pellet with 1 ml of chilled 75% ethanol (Freshly prepared in DEPC water). Mix samples by vortexing for 15 secs, then centrifuge at 12,000 ×g (RCF) for 5 minutes at 4°C or room temperature. Repeat this step twice.

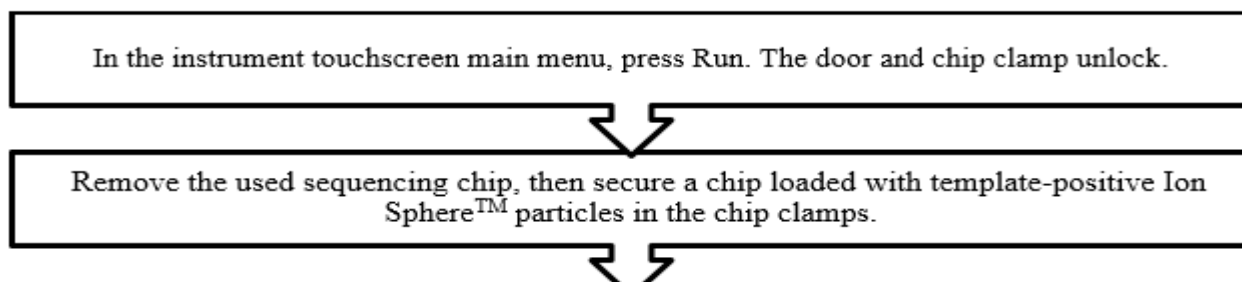
### E. Dissolving and storage of RNA

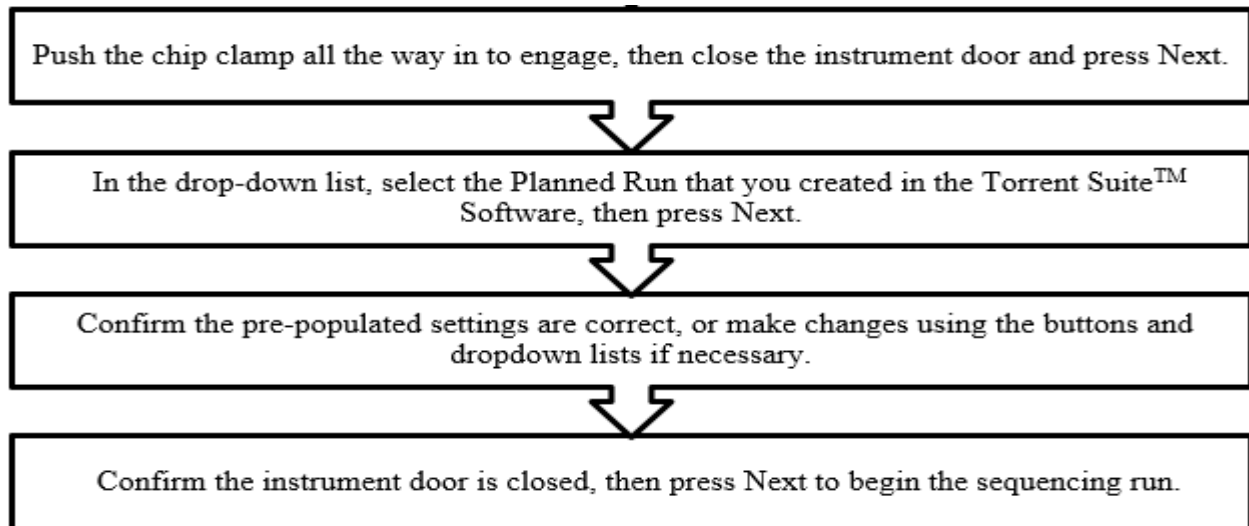
Removed the supernatant gently, then air dry the pellet not more than 5 minutes. Dissolved the pellet in 60 µl RNase free water (commercial) or autoclaved DEPC treated water. For the short time storage we can put it in -20°C and for long time storage put it in -80°C. mRNA was isolated by Dynabeads® mRNA DIRECT™ Micro Kit of life technology. Whole transcriptome library was prepared by Ion Total RNA-Seq Kit v2 of life technology. Temtube preparation was prepared by Ion S5™ Temtube OT2 Kit of life technology. Sequencing reagents used from the Ion S5™ Sequencing 340 Kit of life technology. mRNA was purified with Magnetic Oligo (dT) beads in accordance with the manufacturer’s instructions by Life technologies.

### Chip loading for Ion S5™ Sequencing Polymerase

1. Inject 100 µl of the Flushing solution into the chip loading port two times. After each injection, discard the solution that is expelled from the opposite port.
2. Inject 100 µl of 50% Annealing Buffer into the chip loading port three times. Do not introduce air bubbles. After each injection, remove the expelled liquid from the opposite port.
3. Combine 6 µl of Ion S5™ Sequencing Polymerase with 60 µl of 50% Annealing buffer.
4. Inject 65 µl of the polymerase solution into the chip loading port and remove the expelled liquid from the exit port. Be careful to avoid introducing air bubbles.
5. Allow the chip to incubate for 5 minutes, then immediately proceed to for run.

### Performing of the run





### Bioinformatics analysis

#### Reads and reads quality (QC)

FastQC (v1.0) aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. These tools analyze FASTA format file to calculate different sequence statistics. AvgQuality.pl requires equal file (quality in FASTA format) and calculates the average quality score for each read and overall average quality score for all the reads. N50Stat.pl takes FASTA sequence file to calculate various statistics, such as total number of reads/sequences in the file, total and individual (A, T, C, G and N) number of bases, G+C and A+T counts, and minimum, maximum, average, median, N25, N50, N75 length. These tools will facilitate non-experts to assess various sequence statistics

#### De novo assembly

*De novo* assembly was done by using CLC genomic workbench (v10.0.1) from QIAGEN Bioinformatics tools, which produced assembled contigs from genotype GMV- 2 of three different developmental stages that were vegetative stage, reproductive stage and maturity stage. After assembly we got the N50 value which is an important statistics as this value assess the quality of the sequence assembly and the higher the value of N50 is, the better is the assembly.

### 3. Results and Discussion

#### Isolation of total RNA and mRNA

Concentration of total RNA and mRNA was measured in Qubit® Fluorometer using RNA assay kit. The total RNA concentration was 10 µg, 15 µg and 2 µg from vegetative, reproductive and maturity stage respectively (Figure 2). Total mRNA concentration was 6.1 µg, 7.63 µg and 1.52 µg from vegetative, reproductive and maturity stage respectively.

#### Construction of whole transcriptome library

In present study fragment library was constructed for Ion S5 sequencing platform size selection and purification of fragmented. The library was constructed using enzymatic fragmentation. The enzymatic time was for 3 min to make a library with average read of 200 bp. Concentration of amplified cDNA was measured by HS assay kit using Qubit® Fluorometer. The final library concentration of vegetative stage, reproductive stage and maturity stage was 25.4µg, 38.7µg and 5.35µg respectively (Figure 3).

#### Transcriptome sequencing using Ion S5 sequencing machine

Transcriptome sequencing was done by using Ion S5 sequencing machine (Table 1). After the enrichment, samples were loaded on each new Ion 540v1 chip for sequencing. After run the total data generated was 19.19 Gb, total reads were 97,919,901, usable reads 72% with a mean read length of 96 bp, after removing the polyclonal (19.7%), low quality reads of 8.6% and adapter dimer 00.1%, final library ISPs was 71.8%. Chip type used was 540v1 and instrument Ion S5 and chip barcode was DABG00671. Total number of reads for vegetative stage were 23,936,197 with a mean length of 110bp, for reproductive stage total number of reads were 32,004,586 with a mean length of 101bp and for maturity number of reads were generated was 39,376,696 with a mean length of 85bp. Total reads obtained after quality filtering (trimming) were 23,046,642; 30,913,376 and 37,139,025 from vegetative, reproductive and maturity stages respectively (Figure 4).

#### Quality check of sequencing data

The first step of data analysis was to check the quality of receiving sequencing data by running the FastQC (v1.0) program which generates an output file of raw sequencing data with quality scores across all bases.

#### Distribution of sequence lengths.

In case of untrimmed Illumina or SOLiD reads it will just contain a single peak. On x axis and y axis, sequence length in base-pairs is shown and number of sequences featuring a particular length normalized to the total number of sequences shown, respectively. Sequence length distribution was in between 8-358bp, maximum reads length lies between 179-200bp and minimum number of reads length lies between 8-350bp. In graph maximum sequence length was 110 bp in vegetative stage (Figure 5). Sequence length distribution was in between 8-363bp, maximum reads length lies between 60-130bp and minimum number of reads length lies between 9-229bp. In graph maximum sequence length was 90bp in reproductive stage (Figure 6). Sequence length distribution was in between 8-365bp, maximum reads length lies between 40-130bp and minimum number of reads length lies between 9-190bp. In graph maximum sequence length was 100 bp in maturity stage (Figure 7).

Per sequence GC- content:- The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases). Average GC content was 44%

for vegetative stage (Figure 8), 43% for reproductive stage (Figure 9) and 46% for maturity stage (Figure 10).

In case of untrimmed data, the coverage of per base position was 100% at level of 299 bp read length. In the nucleotide contribution the four DNA nucleotides with ambiguous bases. The four (A/T/G/C) nucleotides observed between 20-29% at level of 299 bp base position for vegetative stage (Figure 11). For reproductive stage the four (A/T/G/C) nucleotides observed between 20-30% at level of 299 bp base position (Figure 12). For maturity stage the four (A/T/G/C) nucleotides observed between 20-30% at level of 299 bp base position (Figure 13).

The central red line is the median value. The yellow box represents the inter-quartile range (25-75%). The upper and lower whiskers represent the 10% and 90% points. The blue line represents the mean quality.

The y-axis on the graph shows the quality scores. The higher the score the better the base call. The background of the graph divides the Y-axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The quality of calls on most platforms will degrade as the run progresses, so it is common to see base calls falling into the orange area towards the end of a read.

Example of good quality of an average read in the sample. In this picture X-axis represents the position of each base in a read and the Y-axis represents the quality score. A score above 20 is considered as good quality. In the present study, quality score for vegetative stage is 24 (Figure 14), 25 for reproductive stage (Figure 15) and 25 for maturity stage (Figure 16).

A total of 23,936,197; 32,004,586 and 39,376,696 raw sequence reads were obtained from vegetative, reproductive and maturity stages respectively, before filtering. Total reads obtained after quality filtering (trimming) were 23,046,642; 30,913,376 and 37,139,025 from vegetative, reproductive and maturity stages respectively.

The data generated have been submitted in National Centre for Biotechnology Information (NCBI) with three IDs, Bioproject, Biosample and SRA (Sequence Read Archive). For Bioproject the ID is PRJNA383660 for all the three developmental stages of fenugreek genotype (GMV-2). For Biosample the IDs were SAMN06806434, SAMN06806435 and SAMN06806436, for vegetative, reproductive and maturity stage respectively. For SRA, SRR and SRP IDs were SRR5482216 and SRP105251 for vegetative stage, SRR5482215 and SRP105251 for reproductive stage and for maturity stage IDs were SRR5482214 and SRP105251.

#### Quality filtering of transcriptome sequencing data

In order to filter the genome data CLC genomic workbench (v10.0.1) was used, in which different trim settings were followed. Removal of low quality sequence with limit = 0.05 and removal of ambiguous nucleotides, maximum 2 nucleotides were allowed. Removal of terminal nucleotides that is one nucleotide from 5' end and one nucleotide from the 3' end. Removal of sequences on length minimum length 25 nucleotides and maximum length 280 nucleotides.

Total sequences in vegetative stage data sets before trimming were 23,936,197 with average read length 110.2 bp. After quality filter (trimming) total sequences in data sets remain 23,046,642 with average read length 111.2 bp (Figure 17). In

reproductive stage data sets before trimming were 32,004,586 with average read length 101.2 bp and after quality filter (trimming) were 30,913,376 with average read length 101.7 (Figure 18) and in maturity stage data sets of raw reads were 39,376,696 with 84.2bp and after trimming raw reads were 37,139,025 with average read length of 86.3bp (Figure 19) (Table 2). Total number of nucleotides before trimming were 2,638,563,660 bp; 3,240,242,092 bp and 3,334,341,888 bp in vegetative stage (Figure 20), reproductive stage (Figure 21) and maturity stage (Figure 22) respectively. After trimming of data the nucleotides were 2,562,931,387 bp; 3,143,890,944 bp and 3,205,640,970 bp in vegetative stage, reproductive and maturity stage respectively (Table 3).

#### Statistics of the *De novo* assembly

After reads filtering, total sequence in data sets remain 91,099,043 which yield 8.8 Gb of data. *De novo* assembly (draft) of fenugreek transcriptome was carried out by using CLC genomic workbench (v10.0.1) *de novo* assembler. *De novo* assembly of vegetative stage produced 59,306 contigs yielded data of 29,146,721 reads. Out of which, 29.5% were adenine, 19.0% were cytosine, 22.5% were guanine and 29.0% were thymine. Total number of contigs were 59,306 in number with 39 bp of minimum, 5773 bp of maximum and 491 bp of average contig length. *De novo* assembly of reproductive stage yielded 57,415 contigs had yield data of 30,648,677 reads. Out of which 29.6 % were adenine, 18.8% were cytosine, 22.2% were guanine and 29.4% were thymine. Total number of contigs were 57,415 in number with 40 bp of minimum, 7,167 bp of maximum and 534 bp of average contig length. *De novo* assembly of maturity stage produced 46,721 reads yielded data of 16,924,096. Out of which 29.1 % were adenine, 19.4% were cytosine, 23.1% were guanine and 28.5% were thymine. Total number of contigs were 46,721 in number with 81 bp of minimum, 4,434 bp of maximum and 362 bp of average contig length (Table 4). The N25 contig set is calculated by summarizing the lengths of the biggest contigs until reach 25 % of the total contig length. The N50 contig set is calculated by summarizing the lengths of the biggest contigs until reach 50 % of the total contig length. The N75 contig set is calculated by summarizing the lengths of the biggest contigs until reach 75 % of the total contig length. The minimum contig length in this set is the number that is usually used to report the N25, N50 and N75 value of a *de novo* assembly (Table 5).

However, Garg *et al.* (2011) [3] got assembly of ~107 million high-quality trimmed reads using Velvet followed by Oases with optimal parameters into a non-redundant set of 53,409 transcripts ( $\geq 100$  bp). The average length of transcripts was 523 bp and N50 length of 900 bp. In present study, total number of trimmed reads were 76,719,494 with N50 length of 556; 639 and 368 bp. Like wise, Vaidya and Ghosh (2012) got a total of 42 million high quality reads.

Patel *et al.* (2014) [4] they got a total of 627,117 million single reads were generated with N50 of 470 bp and sequence assembly contained total 7,256 contigs. In present study, the average N50 value of present study from all the developmental study was 521 bp, revealed that the N50 value of present study is better than the Patel *et al.* (2014) [4]. More the N50 value better is the assembly.

**Table 1:** Statistical data of run on PGM

Index	Parameter	Vegetative stage	Reproductive stage	Maturity stage
1	Total number of reads	23,936,197	32,004,586	39,376,696
2	Total bases (Gb)	5.42	6.72	7.05
2	Mean length (bp)	110	101	85
4	ISP Loading (%)	93	93	93

**Table 2:** Statistical data showing the transcriptome reads before and after trimming of data

Index	Developmental stages	Raw reads before trimming	Raw reads after trimming	Average length before trimming(bp)	Average length after trimming(bp)
1	Vegetative stage	23,936,197	23,046,642	110.2	111.2
2	Reproductive stage	32,004,586	30,913,376	101.2	101.7
3	Maturity stage	39,376,696	37,139,025	84.2	86.3

**Table 3:** List of nucleotides in data set before and after trimming

Index	Developmental stages	Before trimming	After trimming
		Total nucleotides in data set(bp)	Total nucleotides in data set(bp)
1	Vegetative stage	2,638,563,660	2,562,931,387
2	Reproductive stage	3,240,242,092	3,143,890,944
3	Maturity stage	3,334,341,888	3,205,640,970

**Table 4:** Nucleotide distribution level chart of all the three developmental stages of fenugreek

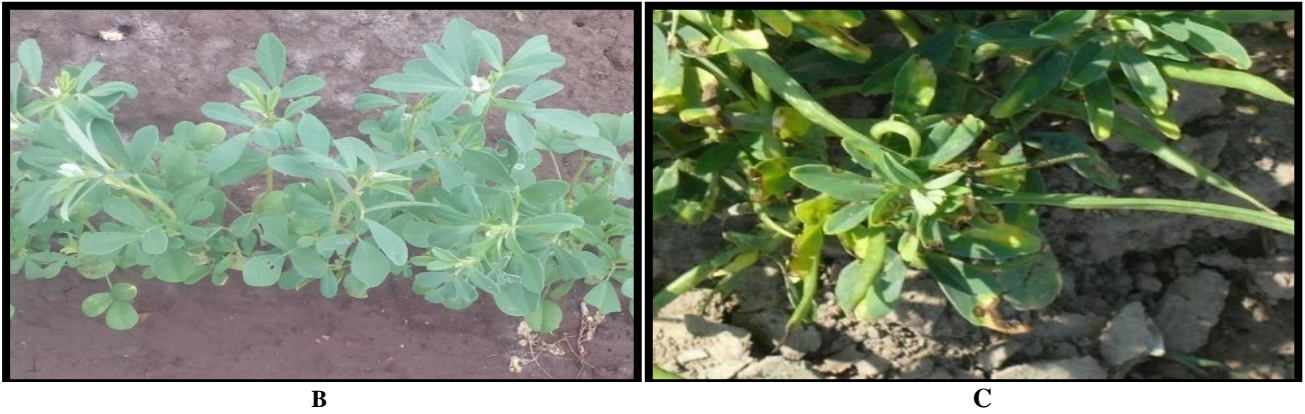
Index	Nucleotide	Developmental stage					
		Vegetative stage		Reproductive stage		Maturity stage	
		Count	Frequency	Count	Frequency	Count	Frequency
1	Adenine	8,596,476	29.5%	9,061,992	29.6%	4,920,110	29.1%
2	Cytosine	5,537,860	19.0%	5,762,432	18.8%	3,282,550	19.4%
3	Guanine	6,545,389	22.5%	6,816,526	22.2%	3,905,019	23.1%
4	Thymine	8,466,996	29.0%	9,007,727	29.4%	4,816,417	28.5%

**Table 5:** CLC genomic workbench (v10.0.1) *de novo* assembler generated data of *De novo* assembly contig measurements of all the three developmental stages of fenugreek

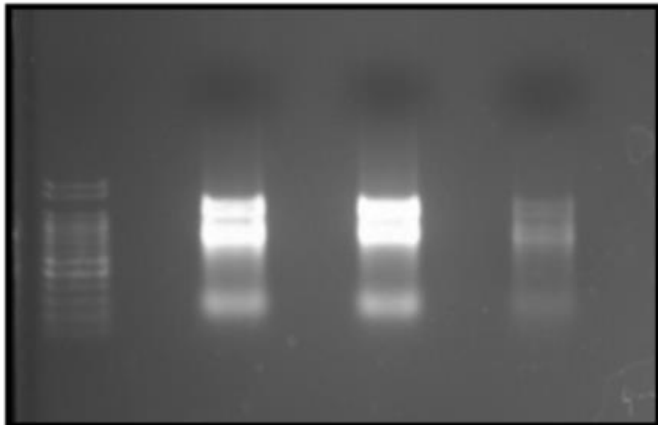
Index	Contig measurements	Vegetative stage	Reproductive stage	Maturity stage
1	N25	984	1,174	567
2	N50	556	639	368
3	N75	347	368	266
4	Minimum(bp)	39	40	81
5	Maximum(bp)	5773	7,167	4,434
6	Average length (bp)	491	534	362
7	Count	59,306	57,415	46,721
8	Total reads	29,146,721	30,648,677	16,924,096
9	Total bases (bp)	2,562,931,387	3,143,890,944	3,205,640,970



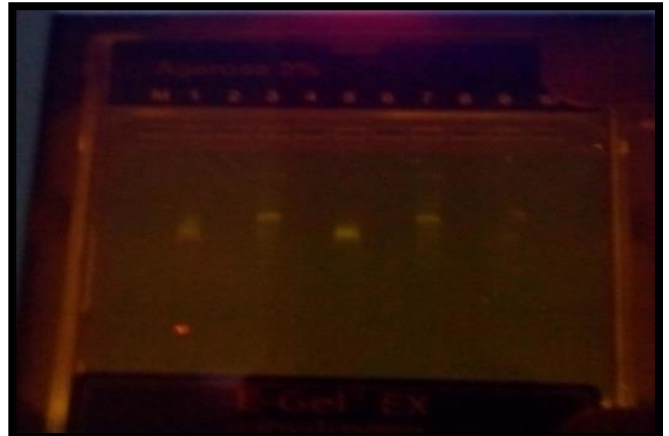
A



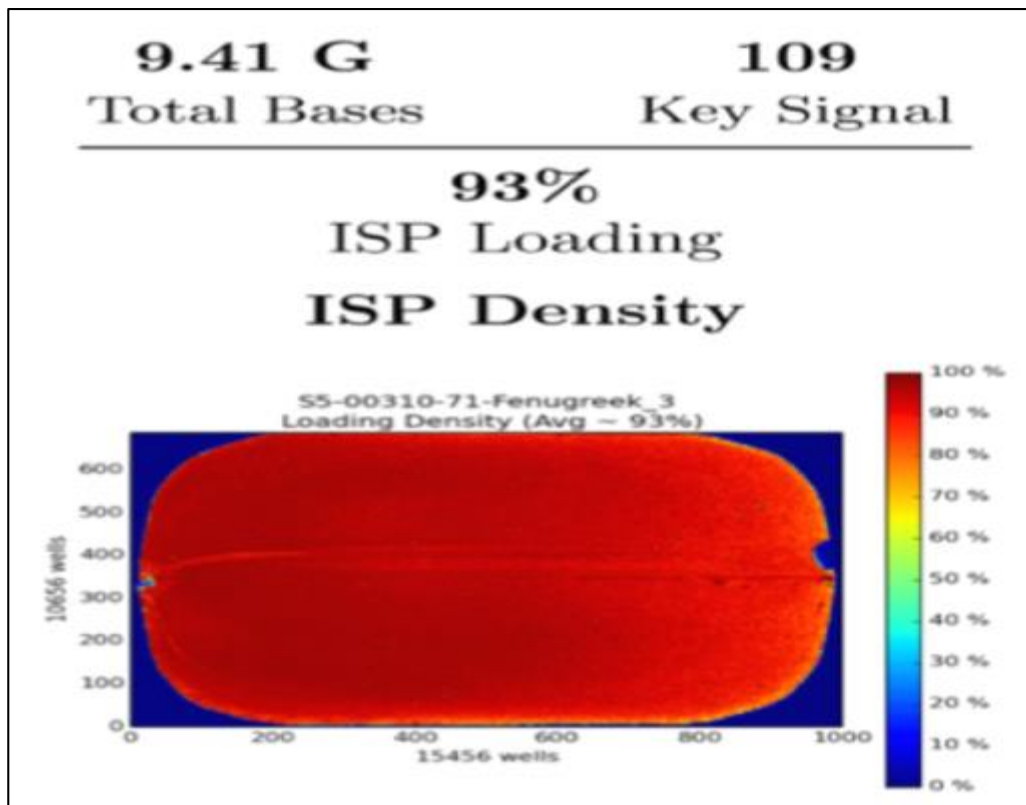
**Fig 1:** Developmental stages of fenugreek (GMV-2) (a) Vegetative stage, (b) reproductive stage and (c) maturity stage

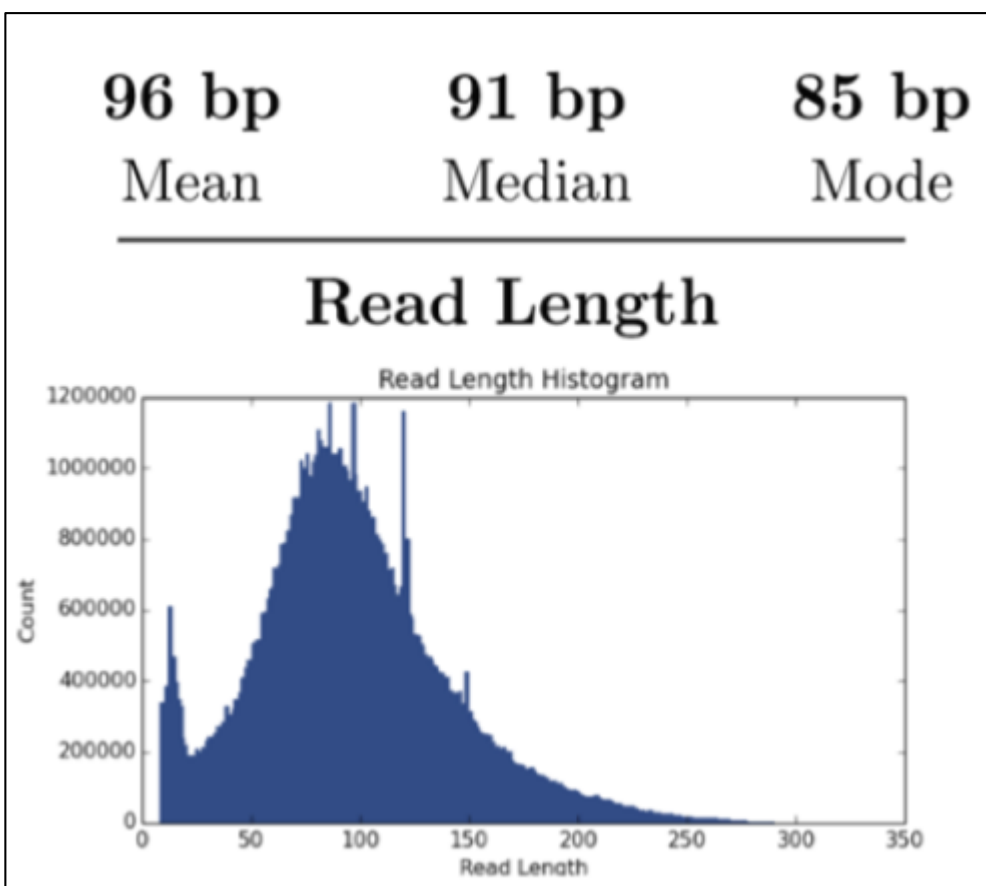
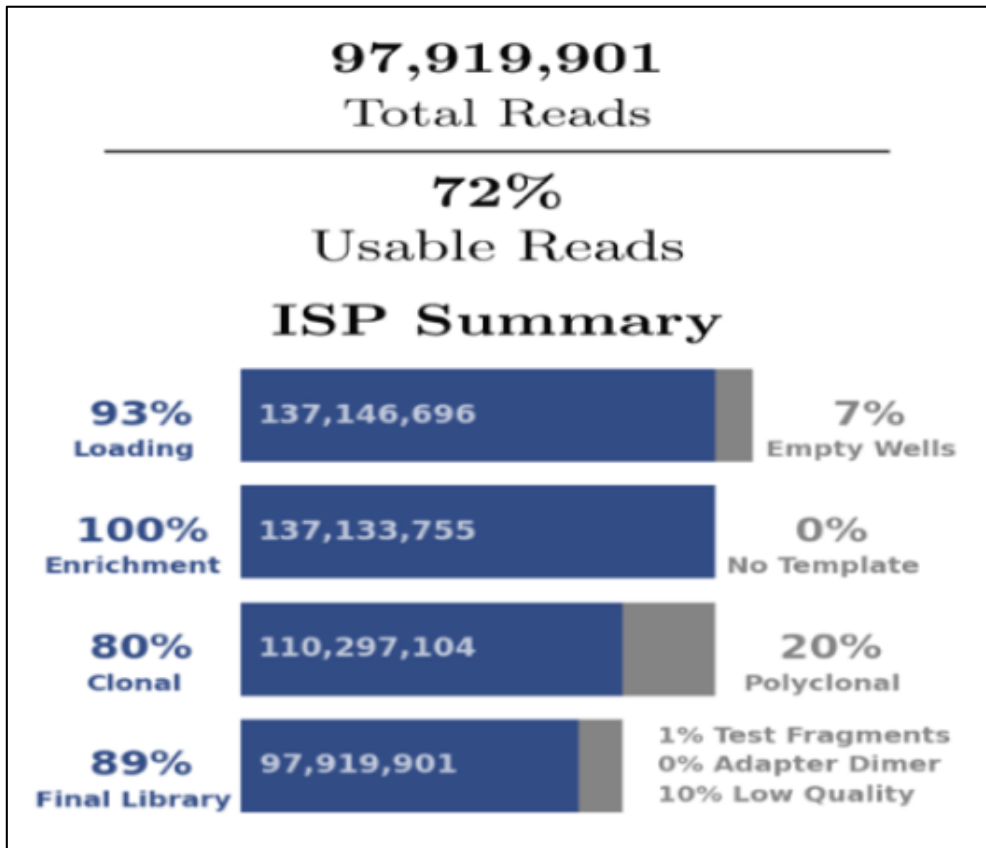


**Fig 2:** RNA bands of different developmental stages of fenugreek (GMV-2) genotype, that are vegetative stage (V), reproductive stage (R) and maturity stage (M).



**Fig 3:** E-gel image of cDNA library construction





**Fig 4:** ISP loading summary of all the three developmental stages of fenugreek variety GMV-2

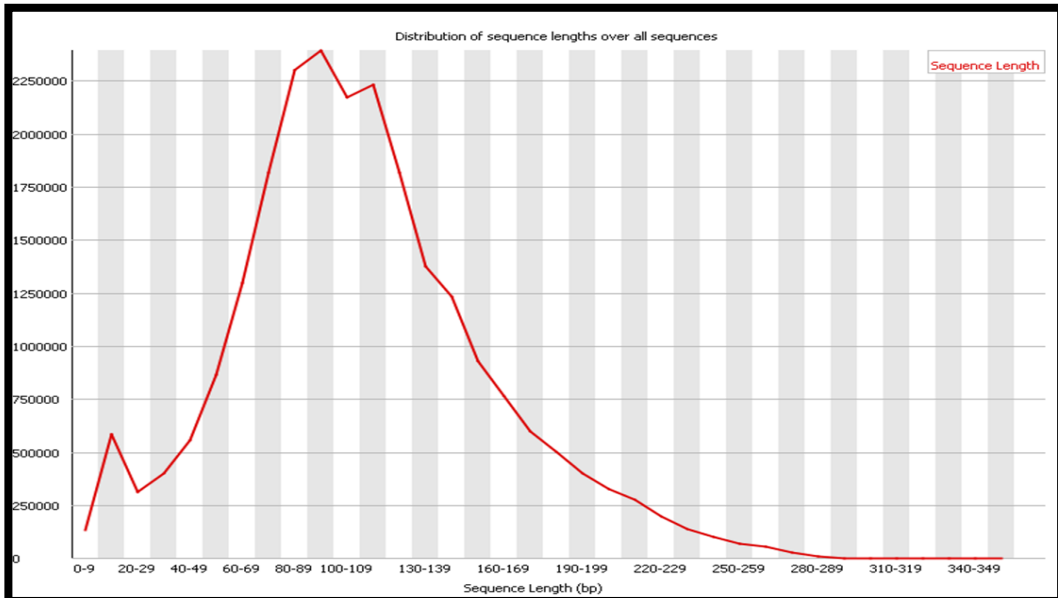


Fig 5: Distribution of sequence length chart of vegetative stage

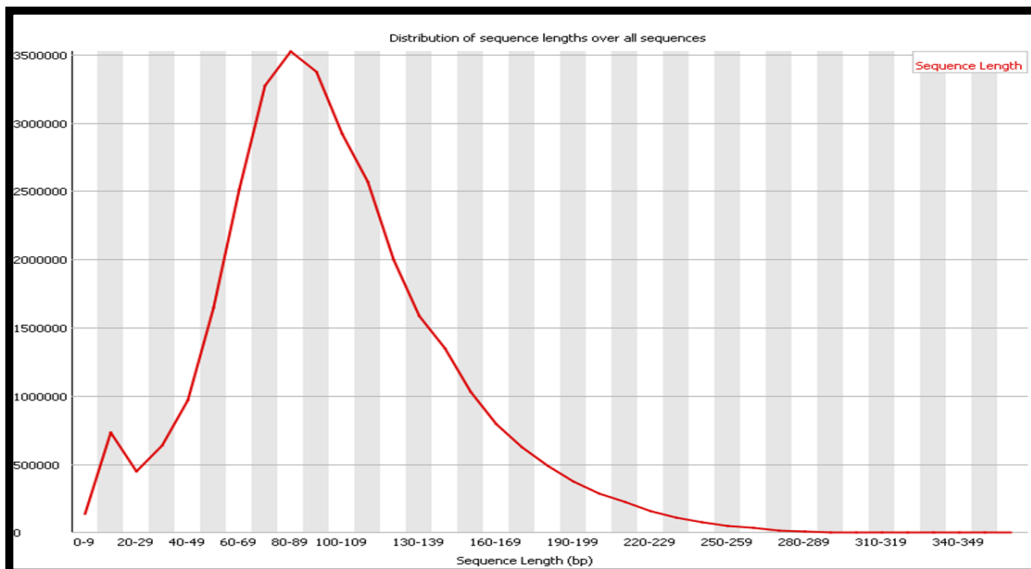


Fig 6: Distribution of sequence length chart of reproductive stage

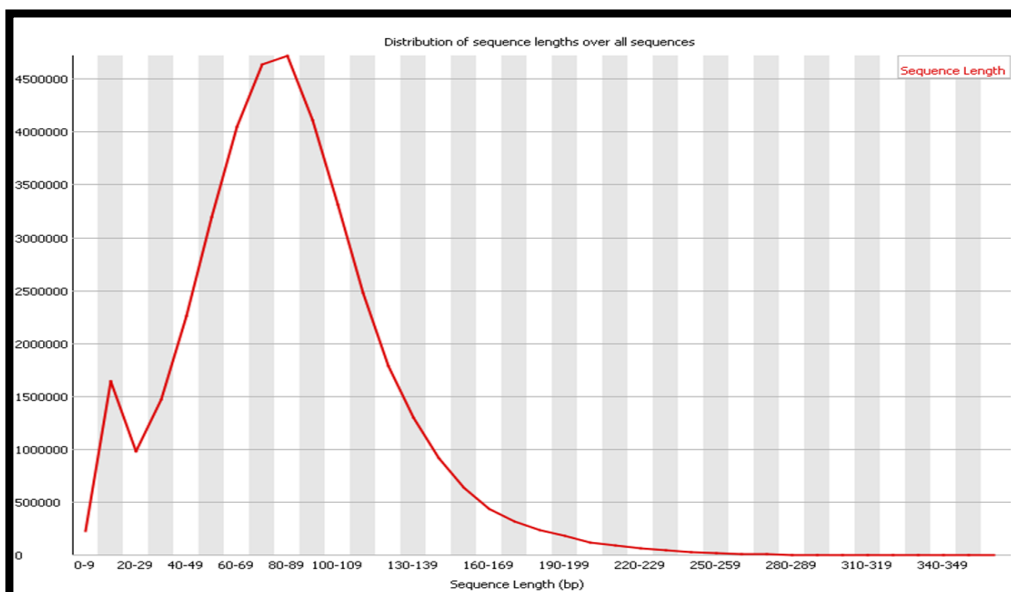
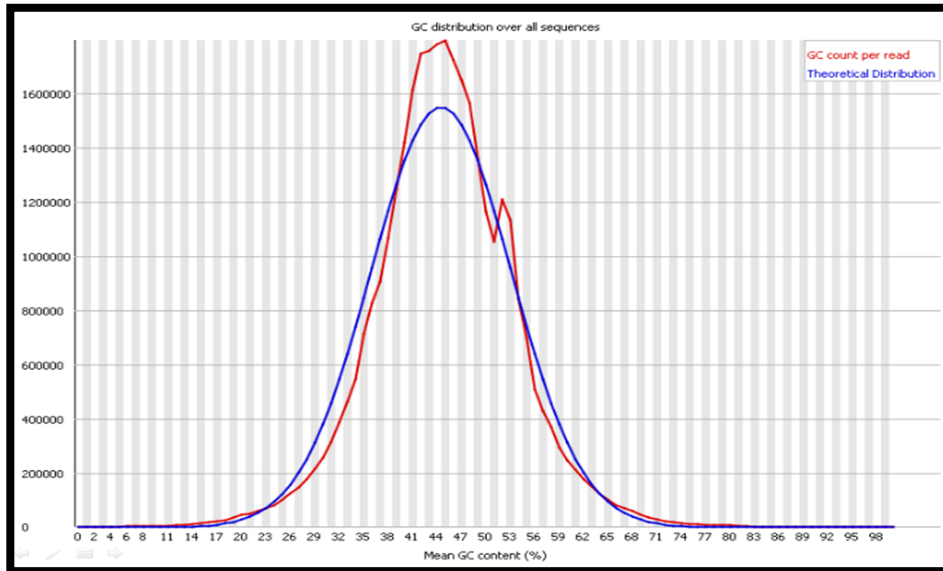
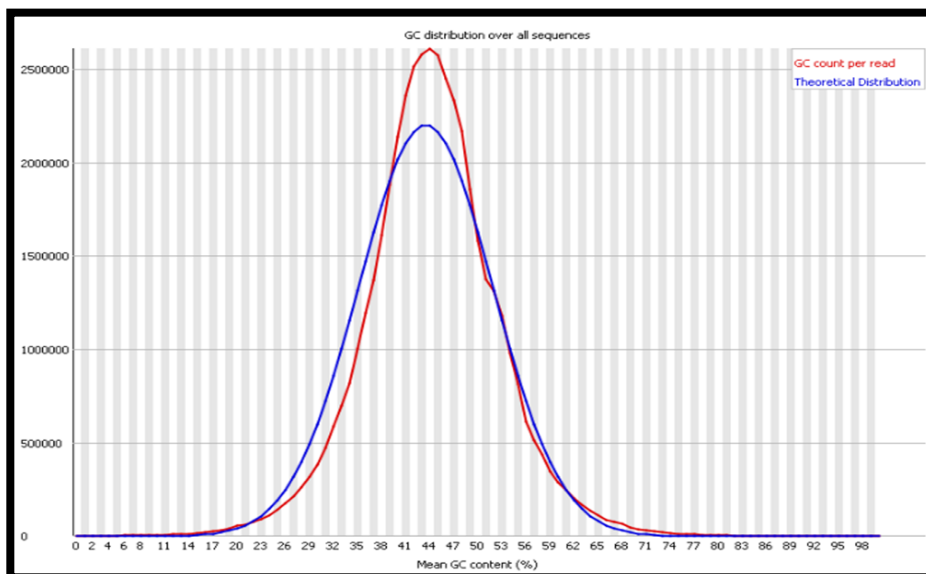


Fig 7: Distribution of sequence length chart of maturity stage

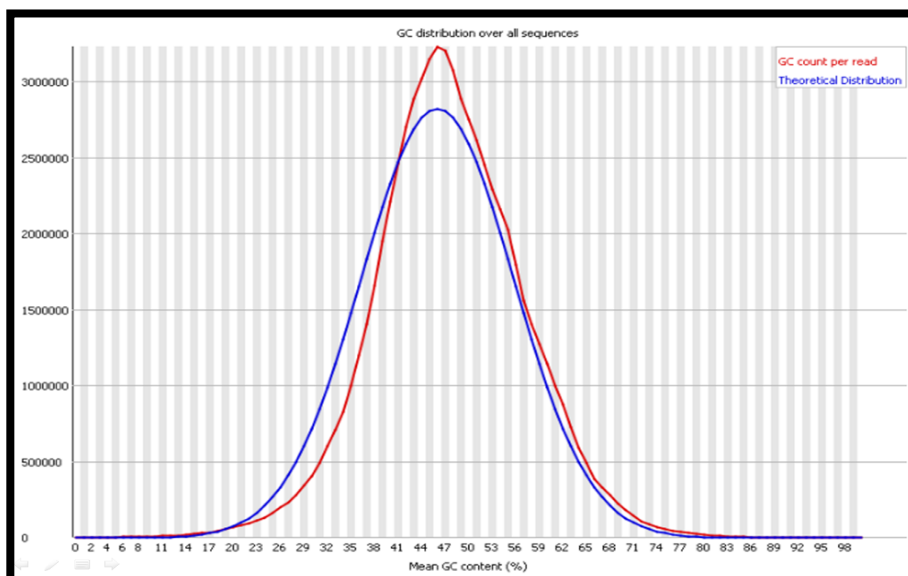




**Fig 8:** Per sequence GC content of quality check graph of vegetative stage



**Fig 9:** Per sequence GC content of quality check graph of reproductive stage



**Fig 10:** Per sequence GC content of quality check graph of maturity stage

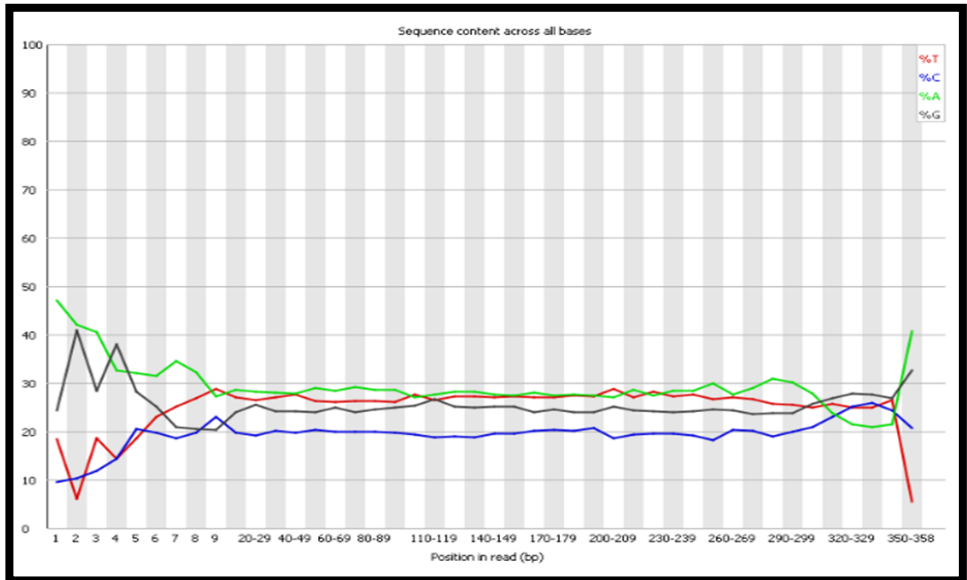


Fig 11: Per base sequence content of quality check graph of vegetative stage

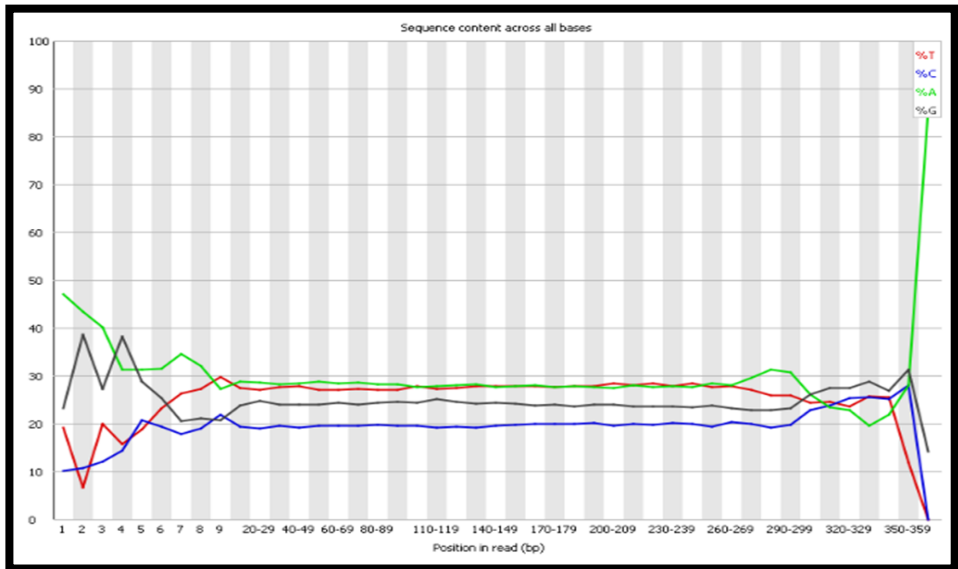


Fig 12: Per base sequence content of quality check graph of reproductive stage

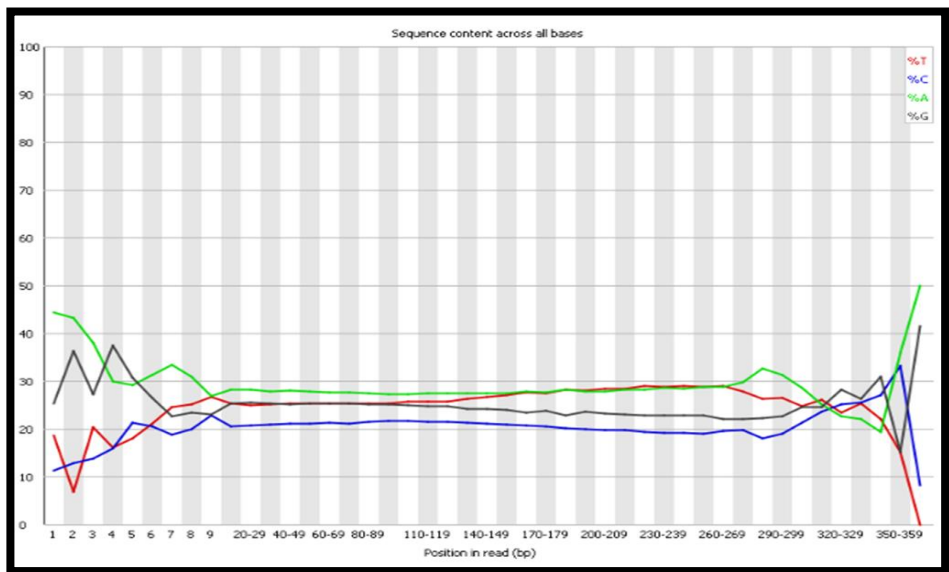


Fig 13: Per base sequence content of quality check graph of maturity stage

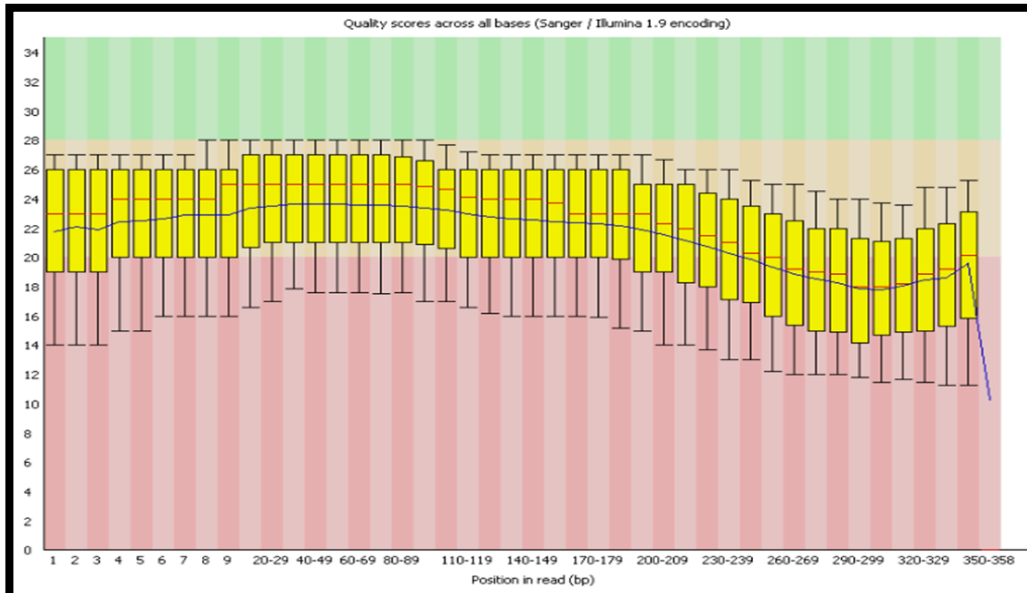


Fig 14: Per base sequence quality of vegetative stage

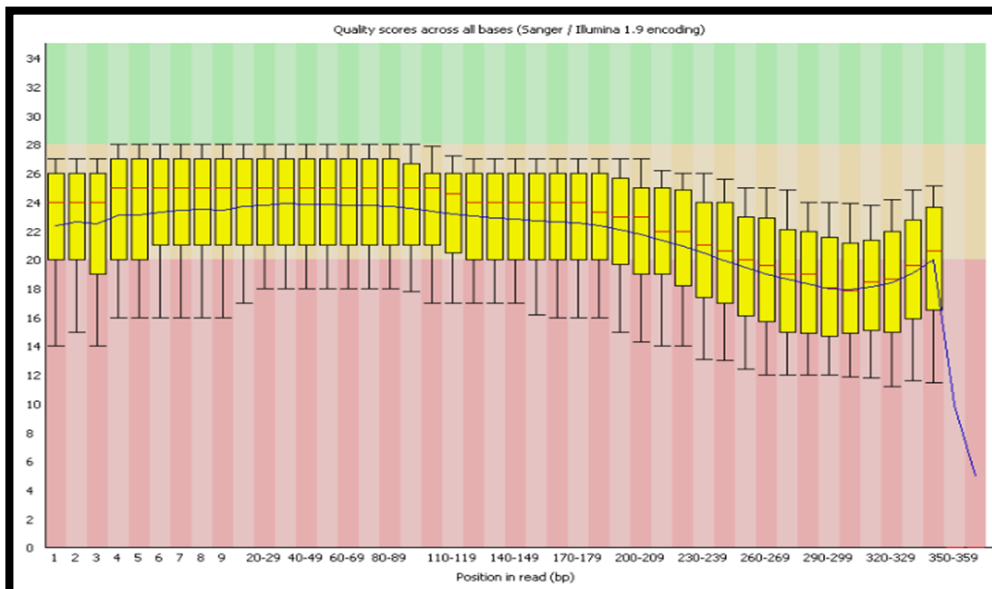


Fig 15: Per base sequence quality of reproductive stage

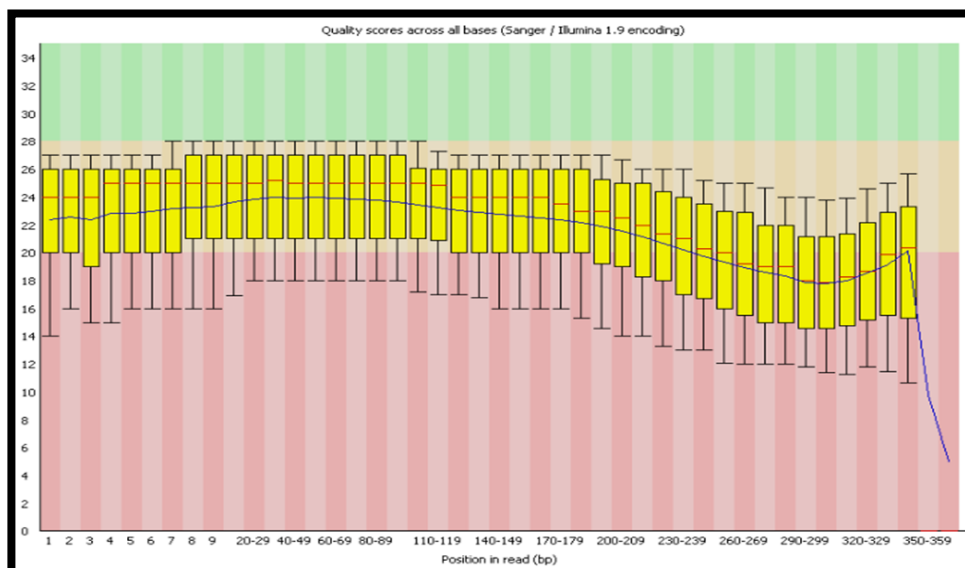
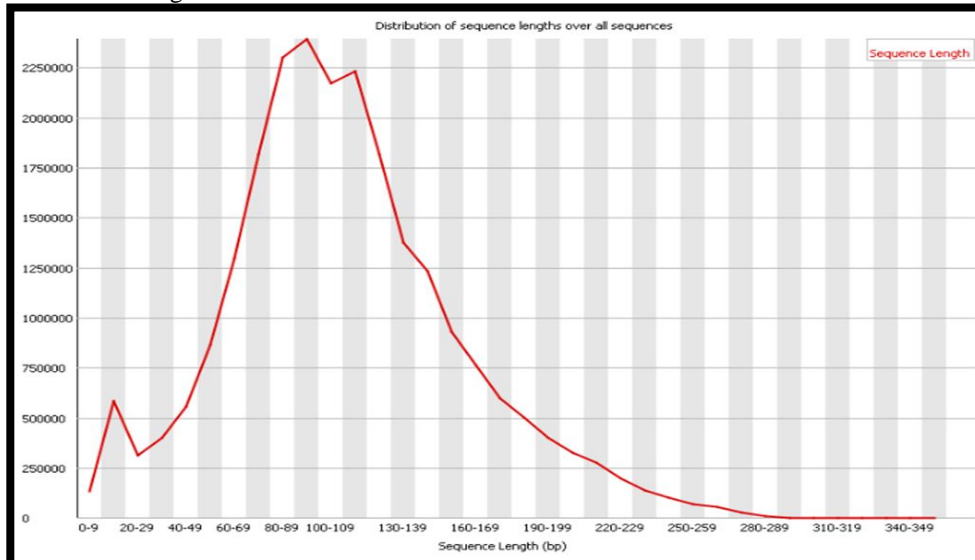


Fig 16: Per base sequence quality of maturity stage

Before trimming of data



After trimming of data

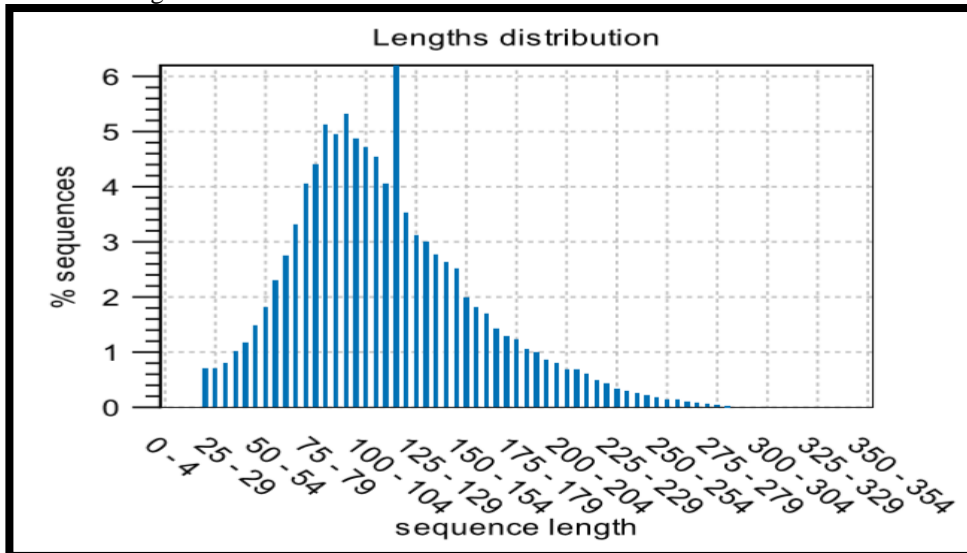
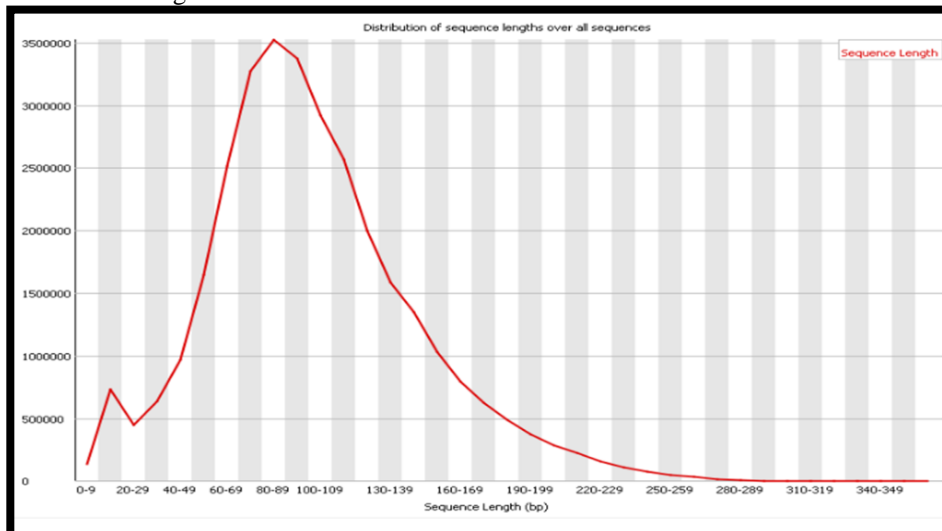


Fig 17: Before and after trimming of data of sequence length distribution graph of vegetative stage

Before trimming of data



After trimming of data

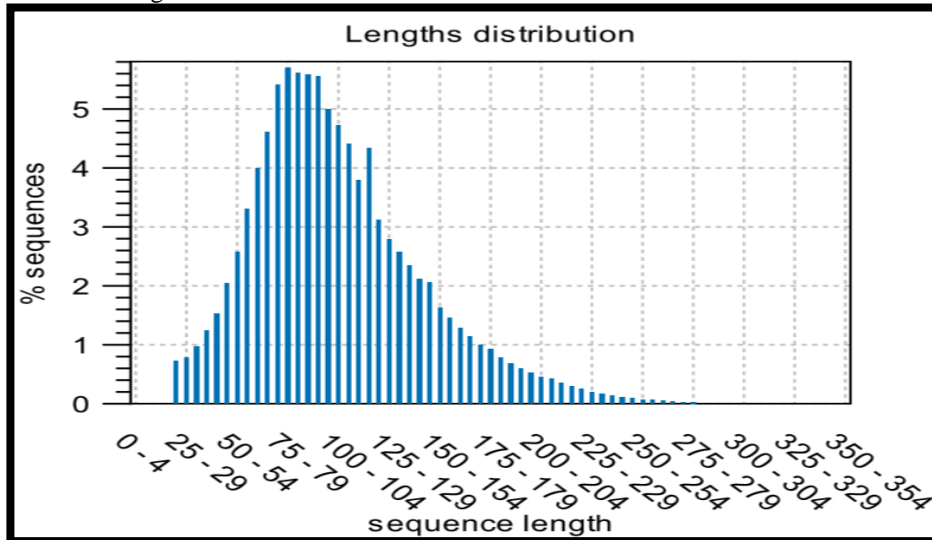
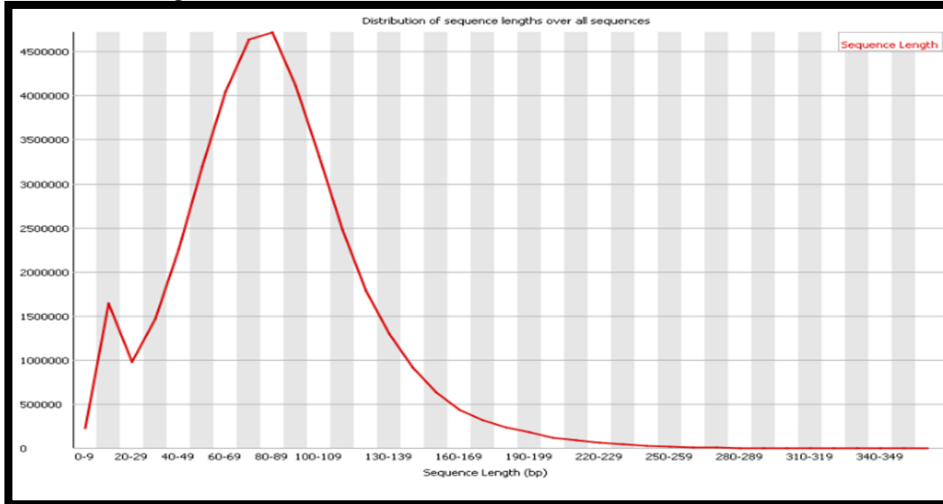


Fig 18: Before and after trimming of data of sequence length distribution graph of reproductive stage

Before trimming of data



After trimming of data

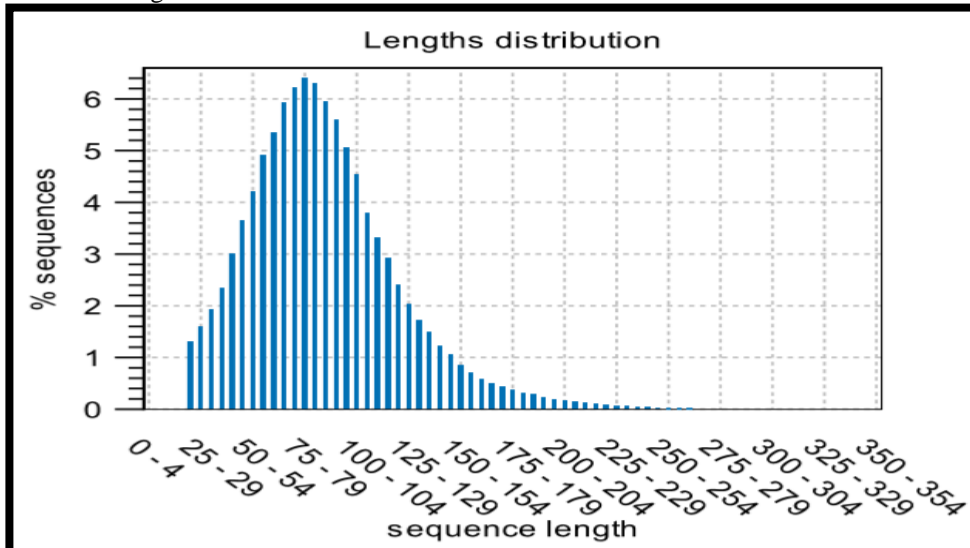
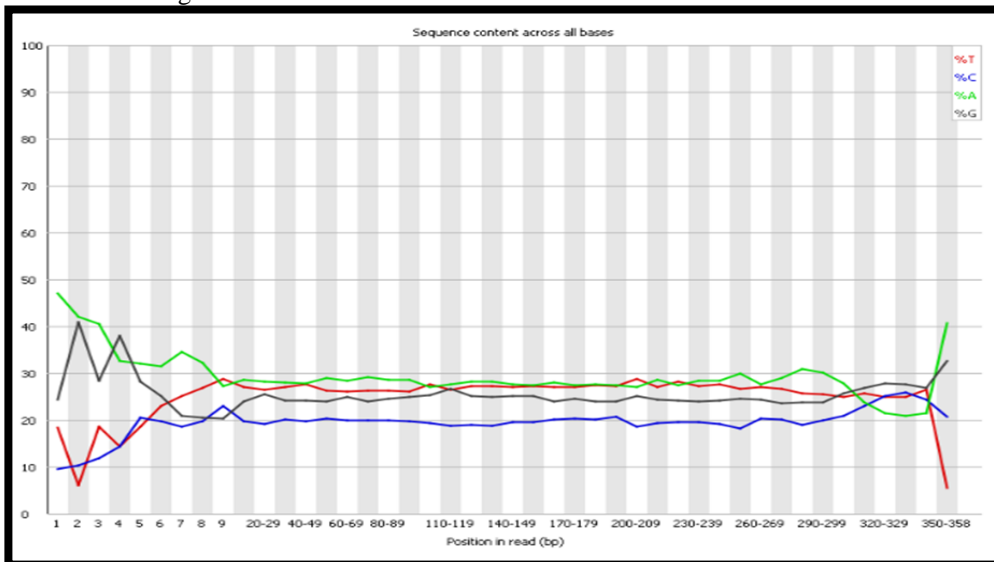


Fig 19: Before and after trimming of data of sequence length distribution graph of maturity stage

Before trimming of data



After trimming of data

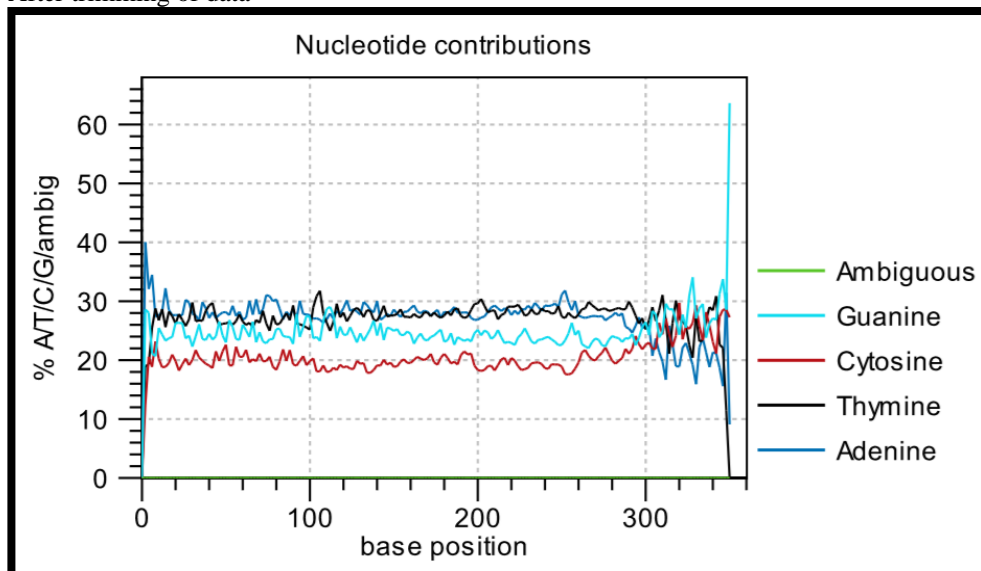
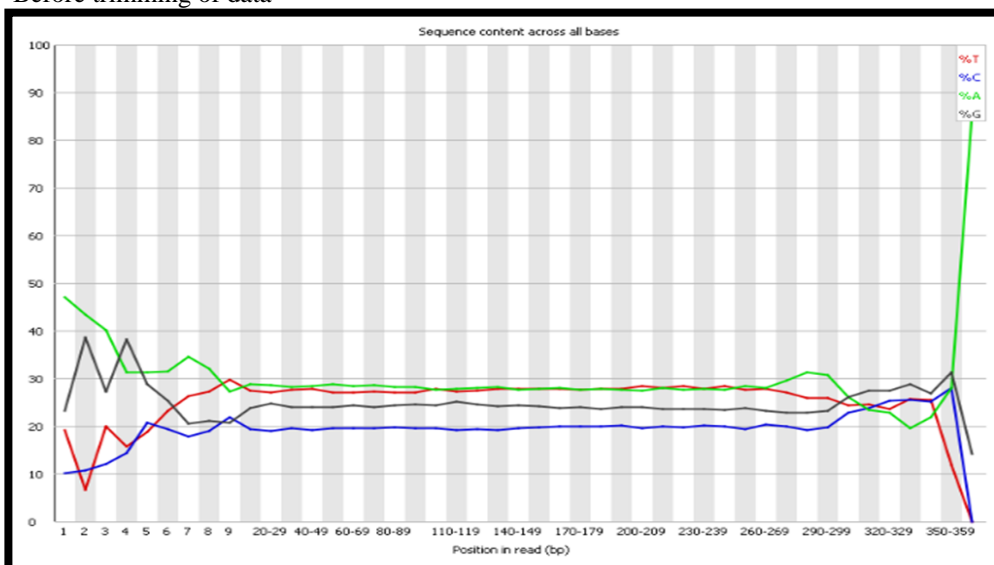


Fig 20: Per base nucleotide distribution trimmed chart of vegetative stage

Before trimming of data



After trimming of data

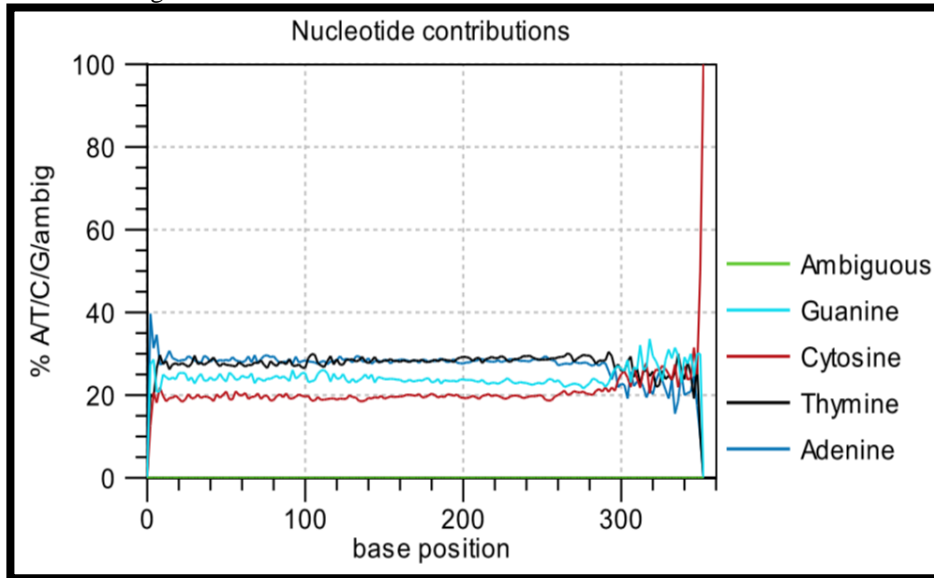
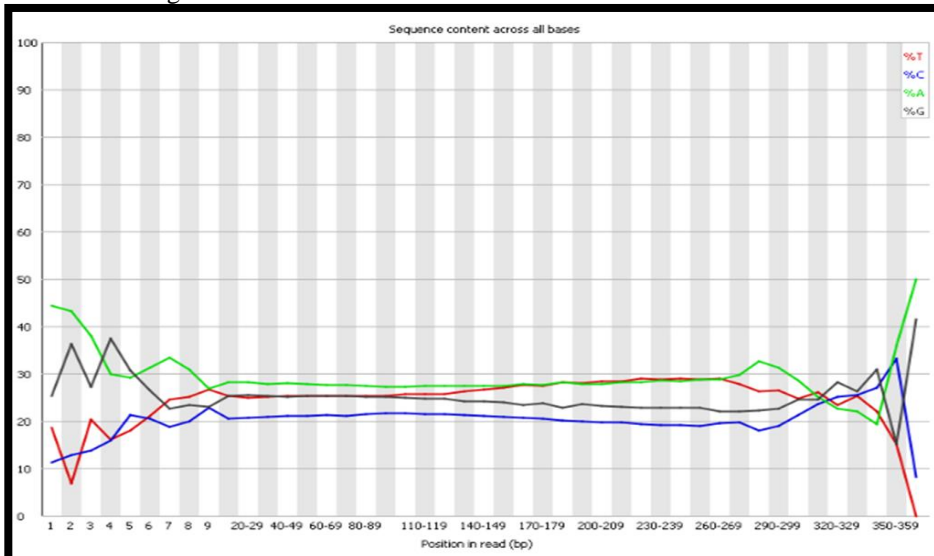


Fig 21: Per base nucleotide distribution trimmed chart of reproductive stage

Before trimming of data



After trimming of data

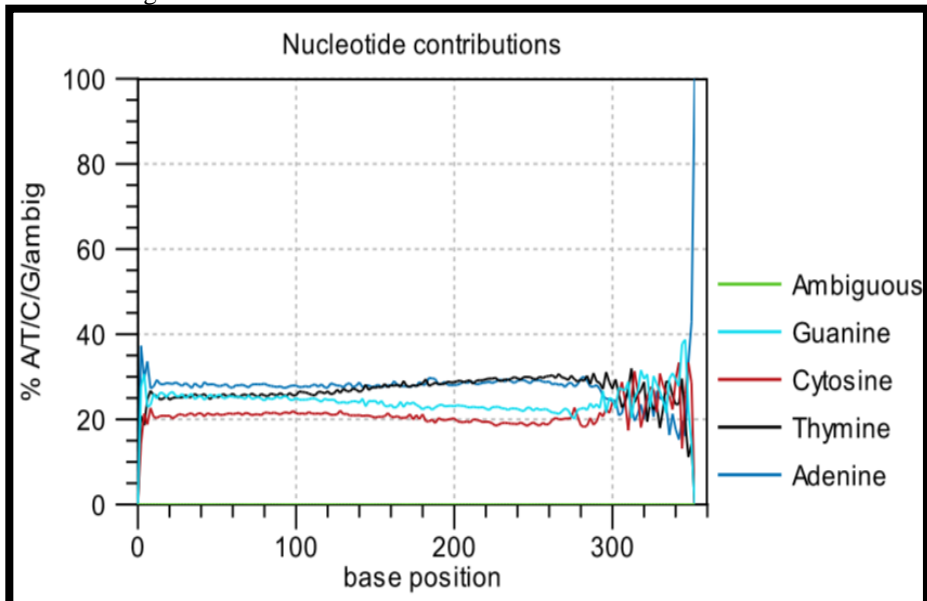


Fig 22: Per base nucleotide distribution trimmed chart of maturity stage

#### 4. Conclusion

Fenugreek (*Trigonella foenum-graecum* L.) is extensively used as a spice crop in India and the Mediterranean region and is known to possess a number of medicinal properties. Transcriptome sequence thus constitutes a meaningful resource to develop a large number of popular molecular markers such as microsatellite marker (SSRs). The *de novo* transcriptome analysis of this very important Indian medicinal ayurvedic herb brings out for the first time novel transcripts related to saponin, diosgenin biosynthesis from fenugreek, which has anticancer and anti-oxidant properties. *Trigonella foenum-graecum* L. the study demonstrates the feasibility of generating a large scale of sequencing information by Ion torrent and efficient assembling. Results provide a valuable resource for fenugreek research. The developed molecular markers are foundation for further genetic linkage analysis and gene localization and they will be essential to accelerate the process of breeding. It is a viable alternative source for different metabolites like, isoleucine, saponins and galactomannans production. Transcripts generated in the present study will be useful in the understanding different metabolites processes and formation.

#### 5. References

1. BBC - Food - Fenugreek recipes. 2017.
2. Raju S Sajad, Pradyuman K. Fenugreek: A review on its nutraceutical properties and utilization in various food product. Journal of the Saudi Society of Agricultural Sciences. 2004; 10:23-26.
3. Garg R, Patel RK, Tyagi AK, Jain M. *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Research. 2011; 18(1):53-63.
4. Patel SS, Dipti B, Hetalkumar J, Panchal. *De novo* RNA Sequencing Assembly and Annotation of *Trigonella foenum-graecum* L. (SRR066197). Legume Genomics and Genetics. 2014; 5:5-7.
5. Vaidya K, Ghosh A. *De novo* transcriptome sequencing in *Trigonella foenum-graecum* L. to identify genes involved in the biosynthesis of diosgenin. Crop Science Society of America. 2012; 6:211-213.