www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; SP-12(12): 1683-1688
© 2023 TPI
www.thepharmajournal.com
Received: 16-10-2023
Accepted: 20-11-2023

Divyesh Vasava
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Corresponding Author:
Divyesh Vasava
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Comparative transcriptome analysis reveals sex-biased expression of sex differentiation in spine gourd (*Momordica dioica*)

Divyesh Vasava

Abstract
Dioecious plant Spine Gourd (*Momordica dioica*), a commercial and nutritional vegetable crop, serves as a model for studying the mechanisms of sex determination and differentiation in plants. However, this mechanism is still unclear. Herein, based on using next-generation transcriptome sequencing (NGS) and comparative analyses were used to identify differentially expressed genes (DEGs) in female and male flower buds. A total of 71.7 million reads were generated. We found a total of 2567, 2059, 3020, 1447 DEGs up-regulated in the female Fe_Fl1_vs_Ml_Fl1; Fe_Fl2_vs_Ml_Fl2; Fe_Fl3_vs_Ml_Fl3; Fe_Fl4_vs_Ml_Fl4 while 4396, 4136, 2408, and 4038 DEGs were respectively down-regulated in female in the same groups. A further 2462, 293, 552 DEGs up-regulated in the female Pre_Fe_vs_Pre_Ml; Dev_Fe_vs_Dev_Ml; while 1864, 246, 826 DEGs were respectively down-regulated in female in the same groups. Moreover, the gene expression pattern of 5 unigenes differentially expressed between male and female flowers revealed by RNA-Seq was confirmed by real-time quantitative PCR (qRT-PCR). This comparative transcriptome analysis provides an invaluable resource for gene expression, genomics, and functional genomics studies in *M. dioica* and its related species. This study also represents a first step toward the investigation of genes involved in spine gourd sex determination.

Keywords: Sex determination, DEGs, transcriptome sequencing

Introduction

The vegetable plant, *M. dioica* (2n = 28), belonging to cucurbitaceae family, is known varyingly as spine gourd, akakara, bodakakara, kakor, teasle gourd, kantola, kakrol, parora, kheksa, kankoda, golbandra, dharkarela, and batkarila (Bawara, Dixit, Chauhan, Dixit, & Saraf, 2010). It is predominant in distribution worldwide, mainly in India, Sri Lanka, Myanmar, and Bangladesh. In India, it is disseminated in all the states except in north-east (Bharathi *et al.*, 2011) [3], but generally confined in the states like Orissa, Bihar, Uttar Pradesh and West Bengal. *M. dioica* is raised effectively in hot and humid areas and almost seems to be day-neutral. The nature of this vegetable is perennial, rhizomatous and climbs up to 3-10 m height with tapering root, which gets bulged and elongated with subsequent years. Nutritionally and medicinally, it is a highly valuable crop, and the tender green fruits are mainly used for cooking purposes (Salvi, 2015) [14]. Differentially expressed several genes are presence at various growth stages allows sex determination a dynamic phenomenon in higher plants (Charlesworth & Mank, 2010) [4]. In these conditions, the discovery and characterization at a specific stage of a few sex-linked genes may not be capable of accounting for the whole process of sex determination in the species concerned (Renner, 2014) [11]. In other terms, to grasp the dynamics of plant sex determination, it is important to characterize various genes from separate developmental stages of dioecious plants. Large-scale studies of gene expression such as mRNA differential show, SSH, PCR reverse transcription and microarray have previously been used in a wide variety of plants to determine the critical stages of sex determination (Cho *et al.*, 2005) [5]. However, they were of limited application in understanding sex determination due to the poor sensitivity of the methods, inconvenience in cross-hybridization and non availability of the total genome sequences. RNA sequencing has become widely popular in recent years for deciphering the genetic networks that govern floral sex determination and growth in many species.

RNA sequencing has become widely popular in recent years for deciphering the genetic networks that govern floral sex determination and growth in many species. Using 454 pyro sequencing technologies, (Rocheta *et al.*, 2014) [10] analyzed the transcriptome of male and

female flowers in monoecious *Quercus suber* and identified some genes involved in pollen development and ovule formation. (Xu, Huang, Yang, & Yao, 2016) [18] studied transcriptome of flower buds at various developmental stages in *Jatropha curcas* and they were obtained 15 sex-related genes that lead to stamen differentiation and embryo sac growth. Other studies have also been conducted to detect the candidate genes underlying sex determination in a variety of plant species, such as wild grapevine (Ramos *et al*., 2017; Ramos *et al*., 2014) [9, 10], bitter gourd (Shukla *et al*., 2015), and castor bean (Tan *et al*., 2016) [17]. Although many genes involved in regulating floral sex have been identified in different species, the gene expression information in florescence buds and the factors regulating floral sex in spine gourd remain poorly understood. The workflow of RNA-sequencing experiment was shown in Figure 1.

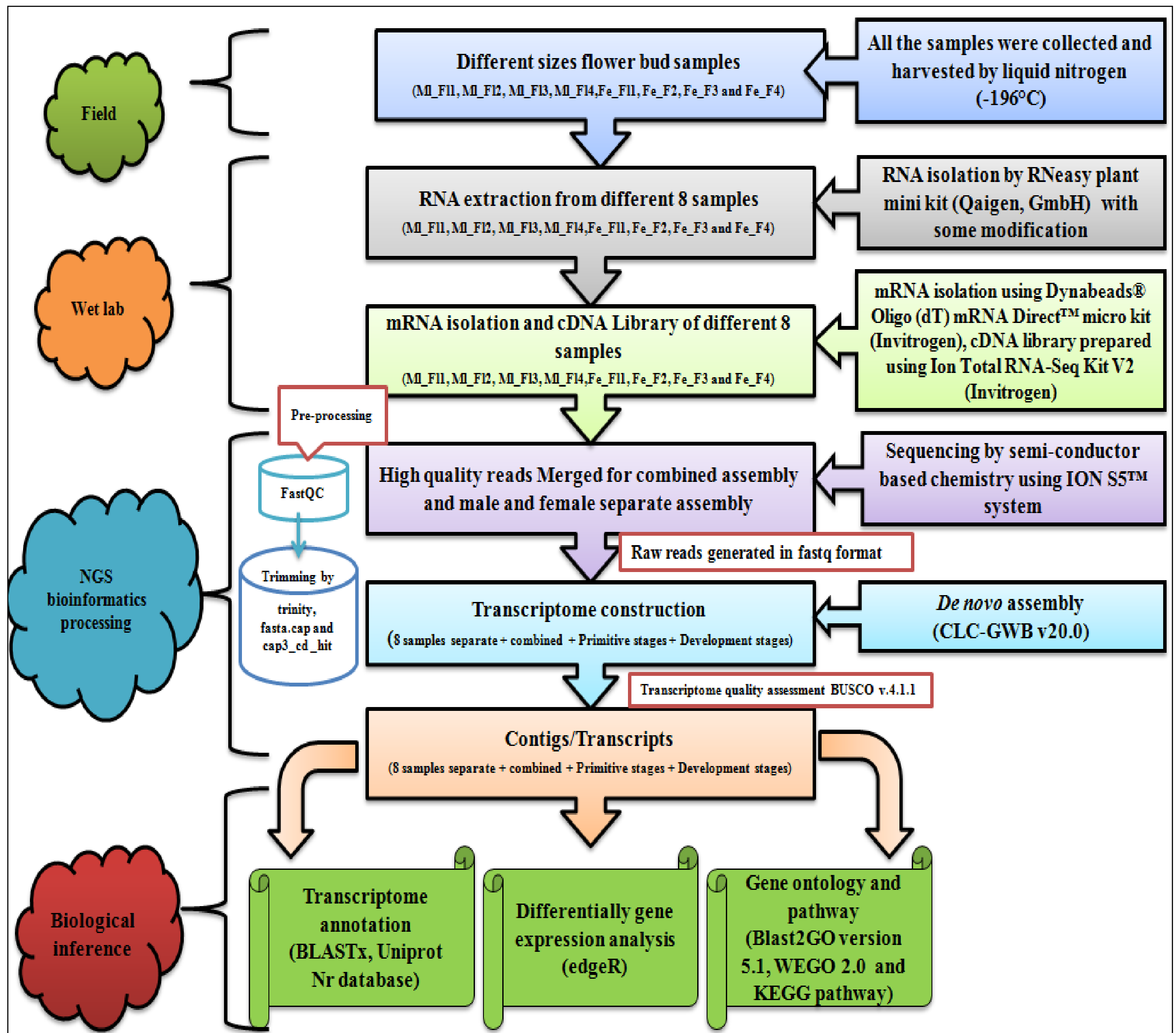

Fig 1: The workflow of RNA-sequencing experiment

2. Materials and Methods

2.1 Plant Materials

The female plant and the male plant were derived from local germplasm, both of which were cultivated at the Department of Biotechnology, Junagadh Agricultural University, Junagadh. Different floral buds from three female and male individuals (the same genotype) were sampled and pooled together. All samples were immediately frozen separately in liquid nitrogen and stored at -80°C prior to RNA extraction.

2.2 RNA isolation, cDNA library construction and sequencing

Total RNAs were isolated from young floral buds using RNeasy plant mini kit (Qaigen, GmbH). mRNA was purified with the Dynabeads® Oligo (dT) mRNA DIRECT™ MICRO Kit according to the manufacturer's instructions. The purified RNA was subsequently fragmented into 150-200 bp pieces. The quantity of mRNA at different developmental bud formation stages were measured by using Qubit 2.0 flurometer. Fragmentation of mRNA was performed on 10-12 µl total RNA using RNase III enzymes. The concentration of the fragmented library was determined by using Qubit® 2.0 fluorometer. Next generation transcriptome sequencing was performed using ION S5 sequencer machine.

2.3 De novo transcriptome assembly of the raw data
Trinity (Grabherr *et al*., 2011) [7] r2013-08-14 was employed as assemblers described in. Additionally, different settings of word size and bubble size set default in CLC-GWB to study the effect of these settings on the assembly quality statistics including contigs number, N50, contig average length; mapping and comparative metrics. Two packages CD-HIT and CAP3 were employed in redundancy reduction of the assembly. The parameters of "-s 0.95 -c 0.95 -n 10" were used in CD-HIT to cluster those contigs of 95% identity and of at least 95% length of the longest representative contigs in the cluster. The parameters "-merge yes -cov_cutoff 1 -edgeFractionCutoff 0.01 -min_trans_lgth 300" were used for contig merging, error correction and length filtering. The overlap percent identity cutoff "-p 95" and other default parameters were used in CAP3 to assemble contigs of defined identity into longer sequences.

2.4 Transcriptome quality assessment
A combination of different metrics was used in assessing the quality of the transcriptome assemblies, including N50 length statistics, rate of reads mapping back, recovery of widely conserved and expressed orthologs, full-length count and coding potentials. The RSEM scores obtained were employed to assess the quality of the assemblies. The transcriptome assembly completeness was assessed by BUSCO v.4.1.1 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) [16].

2.5 Differential Gene Expression (DGE) analysis
Differentially expressed genes (DEGs) were detected using edgeR (Robinson, McCarthy, & Smyth, 2010) [12] and DESeq (Anders & Huber, 2010) [1] in spine gourd between the various stages of bud development in male and female flowers. We used a stringent value of P value ≤ 0.05 and an absolute value of fold change ≥ 2 as the thresholds for identifying significant differences in gene expression between two stages of bud formation. There were taken the following five comparisons: female_1 vs male_1; female_2 vs male_2; female_3 vs male_3; female_4 vs male_4 and all_female1 vs all_male.

2.6 Gene ontology and pathway assignments
The Differentially expressed genes (DEGs) were analyzed for GO classifications using Blast2GO (Conesa & Götz, 2008) [6] version 5.1, GO terms with significant gene number differences was analyzed by WEGO 2.0 and their KEGG pathway annotation was analyzed by BLAST2GO/Omics Box with Biobam bioinformatics software. [WEGO 2.0 (https://wego.genomics.cn/)];[Biobam bioinformatics solution: (https://www.biobam.com/omicsbox/?cn-reloaded=1)]

2.7 Data analysis
Basic assembly statistics were determined by QUAST (Quality Assessment for Genome Assemblies). Venn diagrams were created by the online tool VIB-UGENT (http://bioinformatics.psb.ugent.be/webtools/Venn/) to compare male and female flower buds DEGs. All analyses in a Linux environment were conducted at the high performance computer clusters at Department of Biotechnology, J.A.U., Junagadh.

3. Result and Discussion
3.1 Transcriptome sequencing (ION S5) run report
Sequencing through ION S5 sequencer generated a total data approximately 11.09 Gb for all the eight samples with a total reads of 71.7 million and total bases of 11.09 billion. The individual reads for eight samples Ml_Fl1, Ml_Fl2, Ml_Fl3, Ml_Fl4, Fe_Fl1, Fe_Fl2, Fe_Fl3 and Fe_Fl4 were 65,35,531; 59,56,656; 67,66,785; 1,44,99,619; 87,06,271; 71,12,467; 1,26,67,401 and 94,59,928. The average read length for all the samples varied from 123 bp to 171 bp with the largest read length in sample Fe_Fl3 and smallest read length in sample Ml_Fl4. Run summary of transcriptome sequencing of *Momordica dioica in* Ion S5™ sequencer shown in Figure 2.

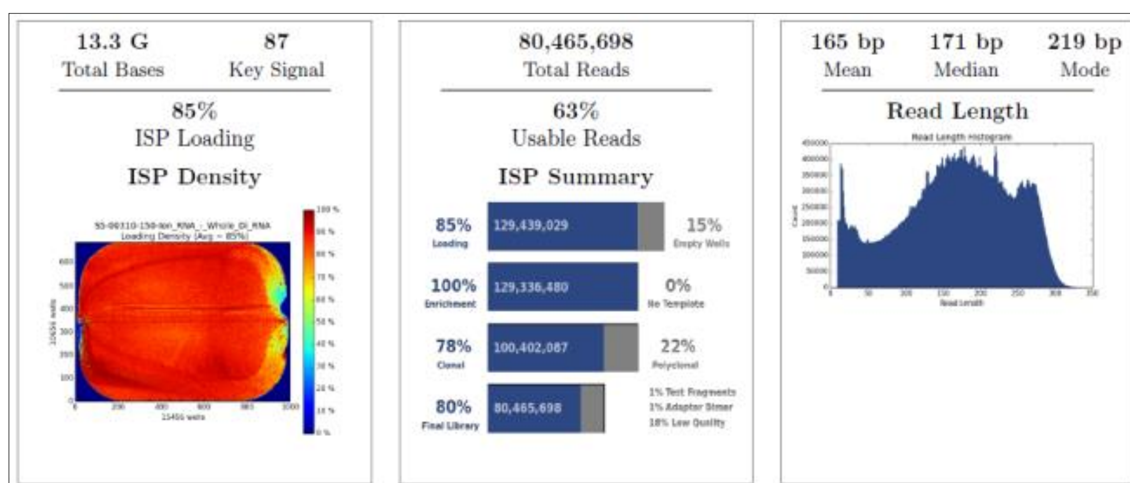

Fig 2: Run summary of transcriptome sequencing of Momordica dioica in Ion S5™ sequencer

3.2 Transcriptome assembly and annotation
A total of 71,704,658 raw reads from the eight libraries were generated. After removing adaptor sequences, ambiguous reads and low-quality reads, a total of 59,309,564 clean reads with an average length of 200 bp were obtained. Among these, 27,769,523 were from female flowers and 31,540,041 were from male flowers (Table 1). All clean reads are available from the NCBI Short Read Archive (SRA) database under accession number SRR13874708 to SRR13874715. Basic statistics report of male samples generated by FASTQC was shown in Table 1 & 2. Average GC-content of all different sizes female sequences in data sets were 45%, 44%, 46% and 47%. Average GC-content of all different sizes male sequences in data sets were 46%, 44%, 45% and 48%.

Table 1: Summary of sequence analysis of male flower buds

Sr. No.	Sample ID	Number of Reads	Avg. length	Number of reads after trim	Percentage trimmed (%)	Avg. length after trim
1	ML_Fl1	87,06,271	169.34	73,28,759	84.18	153.14
2	ML_Fl2	71,12,467	157.86	59,47,987	83.63	140.96
3	ML_Fl3	126,67,401	171.23	105,24,518	83.08	153.66
4	ML_Fl4	94,59,928	166.32	77,38,777	81.81	145.8
Average			166.1875		83.175	148.39

Table 2: Summary of sequence analysis of female flower buds

Sr. No.	Sample ID	Number of Reads	Avg. length	Number of reads after trim	Percentage trimmed (%)	Avg. length after trim
5	Fe_Fl1	65,35,531	154.04	49,26,371	75.38	129.74
6	Fe_Fl2	59,56,656	143.04	43,41,226	72.88	114.48
7	Fe_Fl3	67,66,785	163.29	52,57,632	77.7	137.24
8	Fe_Fl4	144,99,619	123.74	132,44,294	91.34	117.36
Average			146.0275		79.325	124.705

3.3 Comparison of female and male flower transcriptome:

Comparison between male and female flower buds was carried out using EdgeR software (Robinson *et al*., 2010) [12] in which R script was used to count the fold change from count matrix generated from RSEM (RSEM was used for transcript abundance estimation). EdgeR was used to identify the differentially expressed transcripts for all the pair-wise comparisons between the eight samples. Differentially expressed transcripts at a minimum fold change of >2 with *P*-values at most *P*<0.05 were extracted and GO enrichment analysis was performed.

The results showed that comparative differentially expressed genes analyzed in Male vs Female with different developmental stages of flower bud formation like Ml1 vs Fe1, Ml2 vs Fe2, Ml3 vs Fe3, Ml4 vs Fe4. For better understanding the differential expressed genes were took two different comparisons that was primitive stage vs developmental stage and all female vs all male. Primitive stage means combination of Ml1+Ml2 vs Fe1+Fe2 and developmental stages means combination of Ml3+ Ml4 vs Fe3+ Fe4. All female vs all male means Fe1+Fe2+Fe3+Fe4 vs Ml1+Ml2+Ml3+Ml4. The comparative gene expression analysis was performed in female relative to male samples; male samples were taken as a control and female as a treatment in every comparison. So every negative value showed down regulated in female and up regulated in male and positive value showed up regulated in female and down regulated in male.

3.4 Sexuality-specific genes between female and male of *M. Dioica*

We found a total of 2567, 2059, 3020, 1447 DEGs up-regulated in the female Fe_Fl1_vs_Ml_Fl1; Fe_Fl2_vs_Ml_Fl2; Fe_Fl3_vs_Ml_Fl3; Fe_Fl4_vs_Ml_Fl4 while 4396, 4136, 2408, and 4038 DEGs were respectively down-regulated in female in the same groups. A further 2462, 293, 552 DEGs up-regulated in the female Pre_Fe_vs_Pre_Ml; Dev_Fe_vs_Dev_Ml; while 1864, 246, 826 DEGs were respectively down-regulated in female in the same groups. We functionally annotated the *M. dioica* female and male transcriptomes by performing a GO analysis on DEGs. Our results (Fig. 3) showed that the most significantly enriched terms were associated with the biological process like 20% Obsolete oxidation reduction process, 11% Protein phosphorylation, 10% Regulation of transcription DNA templated; cellular component like 29% Integral component of membrane, 17% Nucleous, 10% Cytoplasm and molecular function like 23% ATP binding, 13% Metal ion binding, 7% DNA binding in female *M. dioica* flower buds. GO enrichment in Spine gourd male flower bud, there were various components found in biological process like: 18% Obsolete oxidation reduction process, 12% Protein phosphorylation, 10% Regulation of transcription DNA templated; cellular components found like: 30% Integral component of membrane, 17% Nucleous, 9% Cytoplasm; molecular function like: 24% ATP binding, 12% Metal ion binding, 7% DNA binding. KEGG (Kyoto Encyclopaedia of Genes and Genomes) analysis was performed on the up regulated and down regulated differentially expressed genes in order to map them with various biological pathways**.** There were different significant pathways of differentially expressed genes was identified, pathway network analysis was conducted using KEGG in male, female and combined of male and female flower buds. According to KEGG analysis of Spine gourd assembly, total 147 pathways were found from male flower bud assembly, 146 pathways were found from female flower bud assembly and 145 pathways were found from combined of male and female assembly. Among them 22 pathways were extracted from up and down regulated DEGs. Moreover, they were involved in the pollen regulation, stamen regulation and flower development. A large no of genes involved in purine metabolism, thiamine metabolism, folate biosynthesis, oxidative phosphorylation, cysteine and methionine metabolism were detected in our results. KEGG pathways of sex-biased genes involved in male and female flower bud regulation were shown in Table 3.

Table 3: KEGG pathways of sex-biased genes involved in male and female flower bud regulation

Sr. No.	Gene Name	Enzyme	Pathway	Pathway ID
1	contig6628.p1	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730
2	contig416.p4	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730
3	contig21480.p2	ec:4.4.1.9 - synthase	Cyanoamino acid metabolism	map00460
4	contig2107.p1	ec:1.14.11.15 - 3beta-dioxygenase	Diterpenoid biosynthesis	map00904
5	contig10318.p1	ec:2.7.7.23 - diphosphorylase	Amino sugar and nucleotide sugar metabolism, O-Antigen nucleotide sugar biosynthesis	map00520, map00541
6	contig10318.p1	ec:2.7.7.44 - uridylyltransferase	Amino sugar and nucleotide sugar metabolism, Ascorbate and aldarate metabolism, Pentose and glucuronate interconversions	map00520, map00053, map00040
7	contig10318.p1	ec:2.7.7.10 - uridylyltransferase	Amino sugar and nucleotide sugar metabolism, Galactose metabolism	map00520, map00052
8	contig1920.p1	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730
9	contig20357.p1	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730
10	contig1950.p2	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730
11	contig5883.p1	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730
12	contig1920.p3	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730
13	contig2682.p1	ec:5.1.2.3 - epimerase	Fatty acid degradation, Butanoate metabolism	map00071, map00650
14	contig2682.p1	ec:1.1.1.35 - dehydrogenase	Tryptophan metabolism, Benzoate degradation, Fatty acid degradation, Lysine degradation, Caprolactam degradation, Toluene degradation, Butanoate metabolism, Valine, leucine and isoleucine degradation, Fatty acid elongation, Geraniol degradation, Carbon fixation pathways in prokaryotes	map00380, map00362, map00071, map00310, map00930, map00623, map00650, map00280, map00062, map00281, map00720
15	contig17423.p1	ec:2.5.1.47 - synthase	Cysteine and methionine metabolism, Sulfur metabolism	map00270, map00920
16	contig5459.p2	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730

3.5 Validation of DGEs by quantitave Real Time PCR (qRT PCR)

(Mohanty, Nayak, Jha, & Joshi, 2017) [8] analyzed to validate sex specific expression genes in transcriptome of male and female flower buds in Ivy gourd by qRT-PCR. They were taken 10 DEGs among them 6 were up regulated and 4 were down regulated in male flower buds. In similar to, among 5 DEGs: 1 was DEG up regulated and 4 DEGs were down regulated in female flower buds at development stage of Spine gourd. Comparison of transcript expression levels between transcriptome data and qRT-PCR depicted positive correlation although the value for fold change did not exactly match but remained consistent in up down regulated expression.

4. Conclusion

Dioecious species have been long considered unique tools to study the developmental programs involved in the formation of separate male and female flowers. Here, a broad flowering transcriptome composed of eight independent libraries was obtained for early and late developmental stages of male and female flowers of *M. dioica*, dioecious plant. In the future, to further enhance our knowledge on sex-specific genetic networks, individual EST libraries could be obtained for each phonological stage to fine map male and female flower specific regulators. Comparative studies revealed a subset of transcripts that were differentially expressed in the different libraries, many of which have a known role in flower and/or plant development. Transcriptome analysis also revealed a group of genes expressed exclusively in each type of flower gender that may have a functional role in male and female flower organ development or in sex specification. Some of the genes that showed differential expression have not been previously characterized in other species and others have not, to our knowledge, been implicated in flower development. Thus, it would be very interesting to perform functional studies using the above mentioned genes to identify its roles in plant reproduction or flower development in *M. dioica* and other flowering species.

5. Acknowledgments

Authors would like to thank Junagadh Agricultural University for providing laboratory and data analysis facility during experiment.

6. References

1. Anders S, Huber W. Differential expression analysis for sequence count data. Nature Precedings; c2010. p. 1-1.
2. Bawara B, Dixit M, Chauhan N, Dixit V, Saraf D. Phyto-pharmacology of *Momordica dioica* Roxb. ex. Willd: a review. Int J Phytomed. 2010, 2(1).
3. Bharathi L, Munshi A, VINOD, Chandrashekaran S, Behera T, Das A, *et al*. Cytotaxonomical analysis of *Momordica* L. (Cucurbitaceae) species of Indian occurrence. J Genet. 2011;90:21-30.
4. Charlesworth D, Mank JE. The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics. 2010;186(1):9-31.
5. Cho J, Koo DH, Nam YW, Han CT, Lim HT, Bang JW,

et al. Isolation and characterization of cDNA clones expressed under male sex expression conditions in a monoecious cucumber plant (*Cucumis sativus* L. cv. Winter Long). Euphytica. 2005;146:271-281.
6. Conesa A, Götz S. Blast 2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics; c2008.
7. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-652.
8. Mohanty JN, Nayak S, Jha S, Joshi RK. Transcriptome profiling of the floral buds and discovery of genes related to sex-differentiation in the dioecious cucurbit *Coccinia grandis* (L.) Voigt. Gene. 2017;626:395-406.
9. Ramos MJN, Coito JL, Fino J, Cunha J, Silva H, Almeida DPG, *et al*. Deep analysis of wild Vitis flower transcriptome reveals unexplored genome regions associated with sex specification. Plant Mol Biol. 2017;93:151-170.
10. Ramos MJN, Coito JL, Silva HG, Cunha J, Costa MMR, Rocheta M, *et al*. Flower development and sex specification in wild grapevine. BMC Genomics. 2014;15:1-21.
11. Renner SS. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot. 2014;101(10):1588-1596.
12. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-140.
13. Rocheta M, Sobral R, Magalhães J, Amorim MI, Ribeiro T, Pinheiro M, *et al*. Comparative transcriptomic analysis of male and female flowers of monoecious Quercus suber. Front Plant Sci. 2014;5:599.
14. Salvi J. Nutritional Composition of Momordica dioica fruits: As a wild vegetable. J Food Pharm Sci. 2015;3(2):18-22.
15. Shukla A, Singh V, Bharadwaj D, Kumar R, Rai A, Rai A, *et al*. De novo assembly of bitter gourd transcriptomes: gene expression and sequence variations in gynoecious and monoecious lines. PLoS One. 2015;10(6):e0128331.
16. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210-3212.
17. Tan M, Xue J, Wang L, Huang J, Fu C, Yan X, *et al*. Transcriptomic analysis for different sex types of *Ricinus communis* L. during development from apical buds to inflorescences by digital gene expression profiling. Front Plant Sci. 2016;6:1208.
18. Xu G, Huang J, Yang Y, Yao YA. Transcriptome analysis of flower sex differentiation in *Jatropha curcas* L. using RNA sequencing. PLoS One. 2016;11(2):e0145613.