Transcriptome analysis and validation of sex-specific differential expressed transcripts in Sea buckthorn

By

Mehul Salaria 131573
Tashil Sharma 133806

Report of the project completed under the supervision of

Dr. Anil Kant

June 2017

Submitted in the partial fulfillment of degree

DEPARTMENT OF BIOTECHNOLOGY JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, SOLAN

TABLE OF CONTENTS

CERTIFICATE OF ORIGINALITY

This is to certify that the thesis entitled **" Transcriptome analysis and validation of sex-specific differential expressed transcripts in Sea buckthorn "** which is being submitted by **Mehul Salaria** and **Tashil Sharma** in partial fulfillment of the requirements for the award of the degree of Bachelors of Technology in Biotechnology by the Jaypee University of Information Technology, Waknaghat has been carried out under the supervision of **Dr. Anil Kant**. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma**.**

Signature of Supervisor
Dr. Anil Kant
Assistant Professor
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology
Waknaghat, Solan, India- 173234

Date:

ACKNOWLEDGEMENT

We feel privileged to express our sincere gratitude towards our mentor, Dr. Anil Kant, for his relentless support and unsurpassed guidance, and most of all for his unmatched patience with which he chose to nurture young minds like us at every step. His efforts have been a stimulus which we will always cherish. We have no words to express our gratitude for everything he has contributed. We thank him from the bottom of our hearts.

We are thankful to Prof. Vinod Kumar, Vice chancellor; Dr. Samir Dev Gupta, Director (Admin); Brig (Retd). K.K. Marwah, Director (Academic affairs) for his support and encouragement. The research facility at JUIT is also gratefully acknowledged.

Our special thanks to our HOD Dr. R. S. Chauhan for his invaluable suggestions, scientific attitude, continuous support, guidance, cooperation and encouragement. We gratefully acknowledge the help and support rendered by him.

We feel indebted to Mr. Ankush Bansal and Mr. Raman (PhD Scholars) for their valuable time and patience throughout.

We are also thankful to all the technical and non-technical staff of our labs including Mr. Baleshwar, Mrs. Mamta, Mrs. Somlata and Mr. Kamlesh for their assistance during the project work.

Signature and name of the students:

Mehul Salaria (131573) Tashil Sharma (133806)

Date: Date:

LIST OF FIGURES

7 B	This figure illustrates the Co-expression networks of 8 genes which are found to be involved in flowering pathways in Sea buckthorn male transcriptome dataset. It is an undirected graph where each gene is represented on a node and a set of co-expressed genes are connected to the node with an edge.	39

LIST OF TABLES

ABSTRACT

Sea buckthorn is a dioecious plant found at high altitude with various medicinal properties such as antioxidants, folic acid, vitamins (B1, B2, K, C, A, E), minerals, phenolics, flavonoids and carotenes. As a result it is beneficial in curing or providing benefits against diseases like blood pressure, cardiovascular diseases, skin, acne problems, hyperlipidemia, asthma, etc. The plant has both male and female reproductive organs in separate individuals. The female plants produce orange berries, soft, juicy, and rich in oils. In this article, whole transcriptome *de novo* assemblies of male and female flower bud samples were carried out using Illumina NextSeq 500 platform to determine the role of the genes involved in sex determination. Moreover, genes with differential expression in male and female transcriptome were identified to understand the underlying sex determination mechanism. Current study showed 63,904 and 62,272 CDS in female and male transcriptome datasets respectively. Out of 16831annotated CDS, 625 were upregulated and 491 were found to be downregulated. To understand the potential regulatory roles of differentially expressed genes in metabolic networks and biosynthetic pathways; KEGG mapping, Gene Ontology and co-expression network analysis was performed. Comparison with FLOR-ID database resulted 8 differentially expressed genes viz. PKL, FYPP, TPL, SFR6, LDL1, PRP8A,SUS4, UBP12; known to be broadly involved in flowering, photoperiodism, embryo development and cold response pathways.

CHAPTER 1

INTRODUCTION

Sea buckthorn (*Hippophae rhamnoides* L.) is a medicinal plant widely distributed in temperate zones of Asia and Europe and subtropical regions of Asia at high altitudes. Sea buckthorn berries, along with bark and leaves, have been used for medicinal and nutritional purposes in Russia and China for hundreds of years [1]. Its berries are a potential source of many bioactive compounds including vitamins, organic acids, fatty acids, flavonoids and antioxidants, thus provides various health related benefits [2]. The oil in pulp and seeds of berries exhibits cardio protective, anti-platelet and anti-ulcer activities, thus making this plant important for commercial as well as research interests [2]. Sea buckthorn grows in harsh environmental of cold desert, thus is tolerant to extreme temperatures, drought and soil salinity prevailing in such areas. The root system of Sea buckthorn is highly advanced hence, acts as an excellent biotic choice for prevention of soil erosion on deep unreliable slopes of high altitude. The soils of high mountains are generally poor in nutrients, especially in organic matter and nitrogen. Sea buckthorn can grow on such soils because of its ability to fix nitrogen through its association with *Frankia*, a nitrogen-fixing actinomycete [3]. Approximately about 180 kg of nitrogen per hectare per annum can be fastened in soil in a Sea buckthorn forest [4]. Thus, Sea buckthorn is a good plantation shrub for high altitude area as it offers people opportunity to maintain a more sustainable lively hood as well as prevent soil erosion in fragile topography.

The demand of Sea buckthorn berries has jumped increasingly in the last few years due to their rapid use in commercial products like juice, oil, alcoholic beverages, biscuits, ice-cream, tea, jam, candies, etc [5]. The huge demand also permits its intensive cultivation, instead of their collection from wild resources and genetic improvement in term of productivity and quality. Improvement of dioecious Sea buckthorn through breeding projects aimed at producing both females and male cultivars. Since berries occur on female plants, therefore, production of female cultivars was preferred over male cultivars in breeding programs [5]. For the breeding program to be successful in dioecious plants, the early identification of progeny's gender is necessary. But unfortunately, the gender of Sea buckthorn seedlings cannot be known until flowering, which usually takes place after 3-4 years in the field [6]. This represents a serious problem for plant breeders who have to retain a large number of a superfluous male for several years. To reduce efforts and cost, a large proportion of the males could be discarded at an early stage in the evaluation process.

In dioecious plants, gender determination is regulated at the genetic level by X/Y chromosome system as well as by loci located on autosomal chromosomes[7]. Moreover various molecular marker-based studies such as RAPD, SSR, ISSR, SCAR etc. were being performed from past

several years for gender identification in Sea buckthorn [8]–[11]. Unfortunately, none of the marker-based studies in Sea buckthorn were able to link a marker with genetic locus responsible for sex determination. Thus, mechanism governing the sex determination in Sea buckthorn remains unknown. The genetic control of sex determination is well understood in several model plant systems like *Silene latifolia* [12]–[14], *Cucumis sativus* [15]–[17], *Salix* [18], [19] etc. Different spatial and temporal developmental stages of the flower have been used to decipher the mRNA transcripts involved in sex determination in dioecious plants like *S. latifolia*, *Rumexacetosa*, *Actinidia chinensis*, etc. Flowering genes like *APETALA 2*, *CLAVATA 1* and *SEPTALA 3* showed differential expression among male and female flowers of plants like *Z. mays*, *S. latifolia*, *A. officinalis* indicating their role in sex determination in the above-mentioned plants.

During past three decades, Expressed Sequence Tags (ESTs) have played an informative role in gene discovery as well as their function analysis, especially for non-model organisms [20]. The speed and efficiency of gene discovery have improved significantly due to the emergence of next generation sequencing technologies which have potential to generate millions of reads and expose the complete transcriptome profiling of any organism. RNA-Seq based on next-generation sequencing has become widely functional to obtain mass sequence data for molecular marker development, transcriptional analysis and gene discovery [21].

Till date, there is no report on transcriptome analysis of female plants directed to decipher genes involved in sex determination of Sea buckthorn [22]. Therefore, this study was aimed to elucidate the molecular basis of sex determination through high throughput next generation sequencing technology (RNA-seq) of male and female flower buds in Sea buckthorn. Further, to assess the functional linkages among Gene Involved in Sex Determination (GISD), co-expression networks have been constructed to obtain large gene expression networks between multiple genes to deduce gene association and relevant role in pathways.

CHAPTER 2

LITERATURE SURVEY

Since, early sex-determination in the case of dioecious plants has commercial utilities as various can be saved if undesired female or male plants can be eliminated during early stage of commercial planting or research purpose. The commercial plantation of Sea buckthorn needs only 10% males for appropriate pollination [23]. The most early detection of male and female Sea buckthorn is only possible after flowering which is a major problem for plant breeders who have to retain large numbers of superfluous males for many years and hence this leads to wastage of time and funds in case of commercial plantations. There is more complication in Sea buckthorn which multiplies vegetatively in the field through suckers which leads to excessive dependence of female and males upon their initial proportion. This problem was noticed in the Kelong area of Lahul and Spiti, India, where Sea buckthorn plantation was done a few years back by the Forest department. Many of the plants turned out to be male and the entire forest is at present conquered by unproductive males. Molecular marker based studies such as RAPD, SSR, ISSR, SCAR were conducted since last few years for gender identification in *Hippophae rhamnoides* [24-27]. For the full potential utilization of the Sea buckthorn flora found in India, gender differentiating markers are necessary for all the three species present in India. Thus, a study shows the development of sex specific markers for *H. rhamnoides* which were tested on the collected populations of male and female plants of *H. salicifolia* and *H. tibetana*. Female specific *HrX1* SCAR marker was capable of differentiating female plants from male plants throughout the three species of Sea buckthorn i.e. *H. rhamnoides, H. salicifolia* and *H.tibetana.*

For the detection of differences between male and female plants in reproductive organs; differential growth, repression or abortion of sex organs in unisexual flowers are the major concerns [28,29]. The genes play role in the flower development like meristem identity genes, organ identity genes and flowering time genes which can act as the probable candidates for sex determination in dioecious plants. Apart from floral regulatory genes, sex-based determination is also reliable on the regulatory networks which alter sex expression depending upon environmental cues like photoperiodism and temperature.

The genetic control of sex determination can be analyzed well from many model plant systems like *Silene latifolia* [30-32], *Cucumissativus* [33-35], *Salix* [36, 37], etc. Molecular and genetic studies show that the mechanisms controlling flower development are conserved in distantly related dicotyledonous plant varieties [38]. The genomic resources which are generated from these model plants could be beneficial for identification of the potential GISD in Sea buckthorn. Different spatial and temporal development stages of flower were used to interpret the mRNA transcripts responsible for sex determination in dioecious plants like *S. latifolia*, *Rumexacetosa*, *Actinidiachinensis*, etc. [39-41]. Therefore, the identification of potential candidates for sex determination in Sea buckthorn differential expression of known flowering genes along with the transcription factors was analyzed with the help of quantitative Real Time PCR (qRT-PCR) in three temporal Floral Development Stages (FDS) of both female and male Sea buckthorn flowers.

As per the studies, the first floral development stage of Sea buckthorn recorded higher female specific expression for *HrAP1*, *HrCRY2*, *HrNEF1* and *HrAG*. Whereas in male flowers; *HrAP2*, *HrLFY*, *HrFRI* and *HrGI* recorded higher expression levels. At the second floral development stage the expression level of almost all the studied genes was observed higher in male flowers as compared to female flowers except *HrCRY2* and *HrLFY*. In the third floral development stage, higher female specific expression of *HrAP1*, *HrCRY2*, *HrEF1* and *HrFIL* was noted, whereas *HrCRY1*, *HrCO* and *HrPHYB* had male specific expression. *HrCO* showed consistent higher expression in all male floral development stages. On the other hand, *HrCRY2* noted elevated expression levels in all the female floral development stages.

The morphological analysis of the female sex organs i.e. ovary at FST 3 indicated that the ovary might have developed in stages between FST 2 and FST 3. The expression level of the genes *HrAP1*, *HrCLV1*, *HrFIL*, *HrCRY2*, *HrGI*, *HrEF1* and *HrETR1* increased in FST 3 compared to FST 2. However the expression level of the gene *HrLFY* decreased at FST 3 with respect to FST 2. In case of male floral buds the distinct male floral organs: anthers were observed at MST 2 which matured through MST 3. Thus, the development of the stamens started in between MST 1 and MST 2 while stamens matured through MST 2 and 3. The expression of the genes *HrAP2*, *HrCLV1*, *HrAG*, *HrSEP3*, *HrYAB5*, *HrCRY1*, *HrPHYB*, *HrCO*, *HrCOLK*, *HrFRI*, *HrFRILK*, *HrEF1*, *HrERS1*, *HrETR1*, *HrX1* and *HrNEF1* inclined in MST 2 compared MST 1. But the expression of *HrLFY* declined in MST 2 compared MST 1. During the maturation of male floral bud from MST 2 to MST 3, the expression of only *HrCO* increased while the rest of the above mentioned genes had reduced expression levels in MST 3 with respect to MST 2.

Female specific SCAR marker ***HrX1*** is able to differentiate female plants from male plants in three species of Sea buckthorn, namely, *H. rhamnoides*, *H. salicifolia* and *H. tibetana.* Applicability of this single marker (***HrX1***) in all the three species has circumvented the need for de-novo development of sex linked markers in *H. salicifolia* and *H. tibetana*, thus saving both the time and resources. It is the first report in Sea buckthorn that sequence of sex linked marker has shown homology with known plant gene, which needs further investigation for its potential role in sex determination. Thus as per the stated hypothsis the female specific SCAR marker *HrX1* developed for *H. rhamnoides* were able differentiate gender in other two species of Sea buckthorn i.e. *H. salicifolia* and *H. tibetana*.

HrCO has shown consistent higher expression in male floral buds only. While H***rCRY2*** was expressed throughout the development of female floral buds while its expression was very low in all the development stages of male flowers. The expression level of ***HrAP1, HrFIL, HrCRY2*** and ***HrGI*** increased only in female flowers during the development of female floral organs while the level of expression of ***HrAP2, HrAG, HrSEP3, HrYAB5, HrCRY1, HrPHYB, HrCO, HrCOLK, HrFRI, HrFRILK, HrERS1, HrX1*** and ***HrNEF1*** increased along with the development of male floral organs in male floral buds. In case of both male and female flowers the expression of ***HrLFY*** gene decreased as the sex organ development started.

Efforts have been made to analyse different marker systems used to access genetic diversity and phylogenetic relationship in seabuckthorn [42-44]. Till now, very limited information is accessible on microsatellite markers in seabuckthorn [45]. It has been known for long that microsatellite variation sometimes leads to modification in gene expression which leads to phenotypic variation. As a result, microsatellite variations could be implicated as a solution in the process of adaptation and evolution [46] Studies have been carried out to comprehend distribution of microsatellites in seabuckthorn transcriptome and review their functional significance in increasing Unigene Specific Microsatellite Markers.

Microsatellites have a vital role in seabuckthorn transcriptome. Out of total putative unigenes, few of them showed that microsatellite repeats are present. Some of the unigenes possessed more than one microsatellite repeat. Moreover, compound formation also involved microsatellite repeats. The alterations in prevalence observed earlier in the studies by different groups could be due to the different tools used to screen microsatellites[47,48] and dissimilarity in the size of sequence dataset.

It was observed that the primer pairs intended to amplify microsatellite sequence loci successfully resulted in amplification of target loci in all the accessions of H. rhamnoides. Some alleles were detected for different loci. Although, the number of polymorphic markers obtained in small marker set is quite high as compared to previous study [49] where some polymorphic marker were obtained. It is possible that there could be biased selection of class I microsatellites that is carrying sequences from the whole data set for marker development. It was found that Class I microsatellites were more polymorphic in comparison to class II microsatellites in earlier studies as well. Proportion of polymorphic markers was later improved by analysing selected sequences for their 'Var'score. In this study, out of total amplified putative markers, few microsatellites markers showed to be. In some markers polymorphism could not be seen and the var score could not be calculated as they were involved in compound formation. Therefore, selecting sequences with class I microsatellites and suitable Var score, can permit superior return of polymorphic markers. Greater number of alleles observed in a smaller population in theis study may not be credited alone to structural features of microsatellites, mainly when markers are designed from genic regions. Leaving aside the choice of the technique for the evaluation (in terms of accuracy and resolution) of amplicons, this level of polymorphism can be recognized as the nature of the genes, which needs more investigation.

As microsatellite markers that are specific for unigene are transferable from one taxon to another, microsatellites acknowledged in this study provide a functional resource for marker based applications in seabuckthorn. These sites can be further exploited for recording allelic variance, and for the development of functional markers based on genes. Further investigation and study of microsatellite positive genes involved in a variety of pathways may help in considering the regulatory role(s) of different microsatellite loci in plants.

CHAPTER 3

MATERIALS AND METHODS

Plant material

The different stages of flower buds designated as MST I, MST II, MST III for males and FST I, FST II, FST III for females were collected from Defence Institute of High Altitude Research (DIHAR), Jammu & Kashmir, India (Geographic Coordinates—34°08' 236" N, 77° 34' 345" E). The different stages of male and female flower buds were pooled separately; frozen using liquid nitrogen and kept at -80°C for generation of NGS transcriptomes.

Total RNA isolation, Illumina NExtSeq PE library preparation and quality check

Total RNA was isolated from the male and female flower bud samples using ZR plant RNA Miniprep^{TM} (ZYMO Research) according to the manufacturer's guidelines. The quality and quantity of the isolated RNA was analyzed on 1% RNA Agarose gel and NanoDrop respectively. The RNA-Seq paired-end sequencing libraries were constructed from the Quality Check (QC) passed RNA samples using illumine TruSeq stranded mRNA sample preparation kit.

Cluster Generation and Sequencing

Paired-End (PE) sequencing enables the template fragments to be sequenced in both the forward and reverse directions on NextSeq 500. The adapters were intended to permit re-synthesis of the reverse strands followed by cleavage of the forward strand during sequencing. The reverse strand copied will further be used to sequence from the opposite end of the fragment. The PE libraries were prepared from total RNA using TruSeq stranded mRNA Library Prep Kit. The mean of the library fragment size distribution was 452 bp to 453 bp for male and female flower buds respectively. The libraries were sequenced on NextSeq 500 using 2 x 75 bp chemistry.

High Quality reads generation

The sequenced raw data was organized to achieve high-quality clean reads using Trimmomatic v0.35 to remove adapter sequences, ambiguous reads (reads with unknown nucleotides "N"

larger than 5%), and low-quality sequences (reads with more than 10% quality threshold (QV <20 phred score). A minimum length of 50 nt (nucleotide) after trimming was applied. After removing the adapter and low-quality sequences from the raw data, 51,301,600 (2 x 75 bp) and 42,922,794 (2 x 75 bp) high-quality reads were retained for Sea buckthorn female and male flower bud samples respectively. For *de novo* assembly of both the samples these high quality (QV>20), paired-end reads were used. For filtration; the following parameters were considered: (a) Adapter trimming (b) Sliding window: Conduct a sliding window trimming of 20 bp, cutting once the average quality within the window falls below a threshold of 20; (c) Leading: If threshold quality is below 20, bases are cut off at the starting of a read (d) Trailing: If threshold quality reaches below 20, bases are cut off at the end of a read; (e) Minilength: If the read was below 50 bp length then it is then dropped off.

De-novo transcriptome assembly and Validation

The filtered high-quality reads of both the samples of Sea buckthorn female and male were assembled into transcripts using velvet v1.2.10 and oases v0.2.09 on optimized K-mer 31[77], [78].During the assembly large amounts of misassembled, erroneous and poorly supported transcripts can be encountered. Thereafter, using BWA v0.7.12; high-quality reads were mapped back to their respective assembled transcripts for validation [79]. The complete workflow for transcriptome analysis is depicted in (Figure 1).

Figure 1. Complete framework showing the steps conducted to perform male and female transcriptome based analyses like functional annotation using BlastX, pathway analysis using KASS, differential gene expression using DEseq and gene ontology using Blast2GO.

Coding Sequence (CDS) Prediction

The TransDecoder was used to predict coding sequences from transcripts. It identifies candidate coding regions within transcript sequences based on the following parameters: (1) A minimum length open reading frame (ORF) is found (2) A log-likelihood score similar to what is computed by the GeneID software is > 0 (3) The above coding score is greatest when the ORF is scored in the first reading frame as compared to scores in the other five reading frames (4) In case, candidate ORF seems to be enclosed by the other candidate ORF where the longer ORF is considered. However, multiple ORFs can be reported by a single transcript (applicable for chimeras, operons, etc) [53].

Gene Ontology Analysis

Gene Ontology (GO) annotations of the coding sequence (CDS) were identified using the Blast2GO program [54]. GO terms were employed for categorizing the functions of predicted CDS using three main domains which are: (a) Biological Processes (BP), (b) Molecular Functions (MF) (c) Cellular Components (CC). To retrieve GO terms for annotated CDS; GO mapping uses following criteria: 1) BLASTX result accession IDs are used to identify gene names or symbols, retrieved gene names or symbols are then searched in the species-specific entries GO database; 2) BLASTX results are used to retrieve Uniprot IDs making use of Protein Information Resource (PIR) which includes Protein Sequence Database (PSD), Universal Protein Resource (UniProt), SwissProt, TrEMBL, RefSeq, GenPept and Protein Data Bank (PDB) databases; 3) Accession IDs are searched directly in the dbxref table of GO database; 4) BLASTX result accession IDs are searched directly in the gene product table of GO database.

KEGG Pathway Analysis

To identify the potential involvement of the predicted CDS of Sea buckthorn female and male samples in biological pathways, CDS were formulated into reference canonical pathways in KEGG [55]. All the CDS were categorized under five processes namely: Metabolism, Cellular processes, Genetic information processing, Environmental information processing and Organism Systems. The output of KEGG analysis includes KEGG Orthology (KO) assignments and corresponding Enzyme commission (EC) numbers and metabolic pathways of predicted CDS using KEGG automated annotation server KASS (http://www.genome.jp/kaas-bin/kaas_main).

Differentially expressed gene Analysis

The mapping of the high-quality reads to their respective set of CDS of each sample was done using BWA aligner for read count calculation [56]. The hit accessions based on BLAST against non-redundant (nr) database were identified for differential gene expression analysis. The analysis to study differentially expressed genes was performed using a negative binomial distribution model i.e; DeSeq v1.8.1 package (http://www-huber.embl.de/users/anders/DESeq) [57]. Further these genes were classified as up and down-regulated based on their log fold change (FC) values calculated by FC = Log2 (Sea buckthorn male / Sea buckthorn female) formula. FC value greater than zero were considered as up-regulated and less than zero as down-regulated. P-value threshold was used to filter statistically significant results.

An overall linkage hierarchical clustering was carried out using Multiple Experiment Viewer (MEV v4.8.1) on the top most 100 differentially expressed genes [58]. Heat map (cluster) shows the expression of genes/ Level of expression is represented by the $\log_2$ ratio of gene abundance between Sea buckthorn female vs. male. Differentially expressed gene identified in Sea buckthorn female and male was examined using hierarchical clustering. A heat map was created using the log-transformed and normalized value of genes as per Pearson un-centered correlation distance as well as based on complete linkage method.

Comparison with FLOR-ID Database

Flowering Interactive (FLOR-ID) database contains fully defined genes and pathways that are involved in flowering-time. To identify these flowering genes, in-house database of FLOR-ID nucleotide sequences has been constructed and further used for comparison against Sea buckthorn female and male transcriptome by using standalone BLASTN program.

Network Reconstruction

Genes shortlisted after comparison with FLOR-ID database were further considered for co-expression network reconstruction. Construction of co-expression networks were derived via non parametric Pearson Correlation Coefficient (r) which was calculated using in-house perl script.

$$r = \frac{\left[M^{-1}\sum_{i=1}^{M} j_i k_i\right] - \left[M^{-1}\sum_{i=1}^{M} \frac{1}{2}(j_i + k_i)^2\right]}{[M^{-1}\sum_{i=1}^{M} \frac{1}{2}({j_i}^2 + {k_i}^2)] - [M^{-1}\sum_{i=1}^{M} \frac{1}{2}(j_i + k_i)^2]} \quad (1)$$

Where *ji*, *ki* are the degrees of targets at both the ends of the *ith* connection and *M* represents the total connections in the network.

CHAPTER 4

RESULTS AND DISCUSSION

Paired end sequencing of cDNA Library and De novo Assembly of Transcriptome

Whole transcriptome sequencing was performed of Sea buckthorn female and male flower bud samples were performed by Illumina NextSeq500 platform using 2 x 75 bp chemistry. After trimming low quality reads and adapter sequence; the number of high quality reads which were observed in female floral bud sample was 51,301,600 and in the male was 42,922,794. We obtained 69,457 and 69,390 validated transcripts for Sea buckthorn female and male respectively (Table 1).

Table 1. Distribution of no. of transcript according to their length on the basis of gene expression

Range of transcript	Sea buckthorn Female	Sea buckthorn Male
200 ≤ transcript < 500	12596	13486
500 ≤ transcript < 1000	15076	14737
1000 ≤ transcript < 2000	25824	24717
2000 ≤ transcript < 3000	11188	11066
3000 ≤ transcript < 4000	3345	3531
4000 ≤ transcript < 5000	909	1122
Transcript ≥ 5000	519	731

CDS Prediction

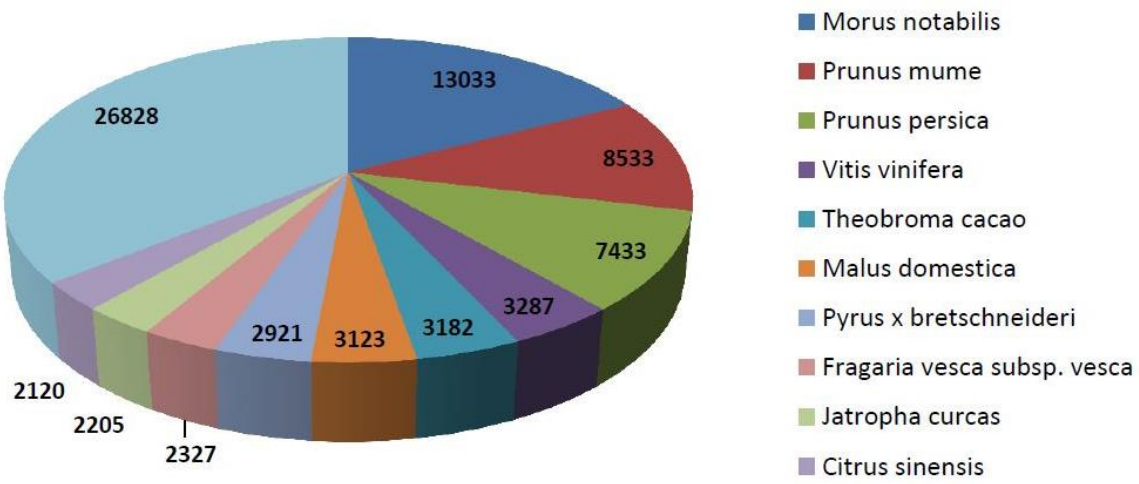
The CDS prediction was carried out on assembled transcripts. The number of Coding DNA Sequences (CDS) was obtained to be 63,904 and 62,272 from female and male floral bud samples respectively. This accounted a total CDS length of 65,128,017 bases having the mean value 1,019 in the females and 66,059,358 bases with 1,060 as the mean value in the males. The maximum length of CDS was obtained to be 7,050 in the female and 12,972 in the male samples; whereas the minimum length was found to be 297 bases in both the samples (Table 2).

Table 2. Distribution of No. of CDS after differential expression analysis according to their length

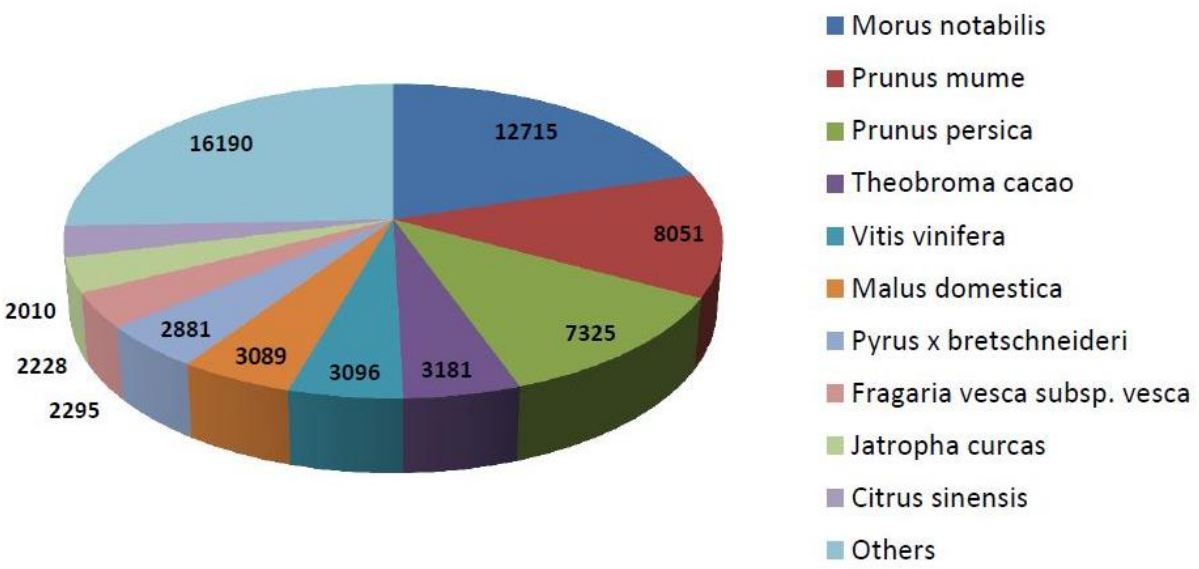
Range of CDS	Sea buckthorn Female	Sea buckthorn Male
200 ≤ CDS < 500	15346	14821
500 ≤ CDS < 1000	22703	21514
1000 ≤ CDS < 2000	20298	19727
2000 ≤ CDS < 3000	4400	4546
3000 ≤ CDS < 4000	961	1157
4000 ≤ CDS < 5000	150	350
CDS ≥ 5000	46	157

Functional annotation

The predicted CDS were searched against NCBI Nr protein database using Basic local alignment search tool (BLASTX; E-value-1e-05). 63,904 CDS for Sea buckthorn female and 62,272 for male flower bud samples were finally annotated. Out of the above CDS, 1472 and 1510 CDS from each sample had no significant BLAST hits whereas the majority of hits were found to be against the *Morus notabilis* followed by *Prunus mume* (Figure 2).



(A)



(B)

Figure 2. Top BLAST results (A) Seabuckthorn Female (B) Seabuckthorn Male

Gene Ontology

From the Gene Ontology (GO) analysis, 26,506 and 26,071 CDS were annotated with Gene ontology terms (Figure 3 A-B). The data obtained based on the gene ontology distribution showed that in female sample 18,231 CDS were involved in biological processes, 21,037 in molecular functions and 14,668 in cellular components. Whereas, in the case of male floral bud sample 17,947 CDS were involved in biological processes, 20,841 in molecular functions and 14,207 in cellular components.

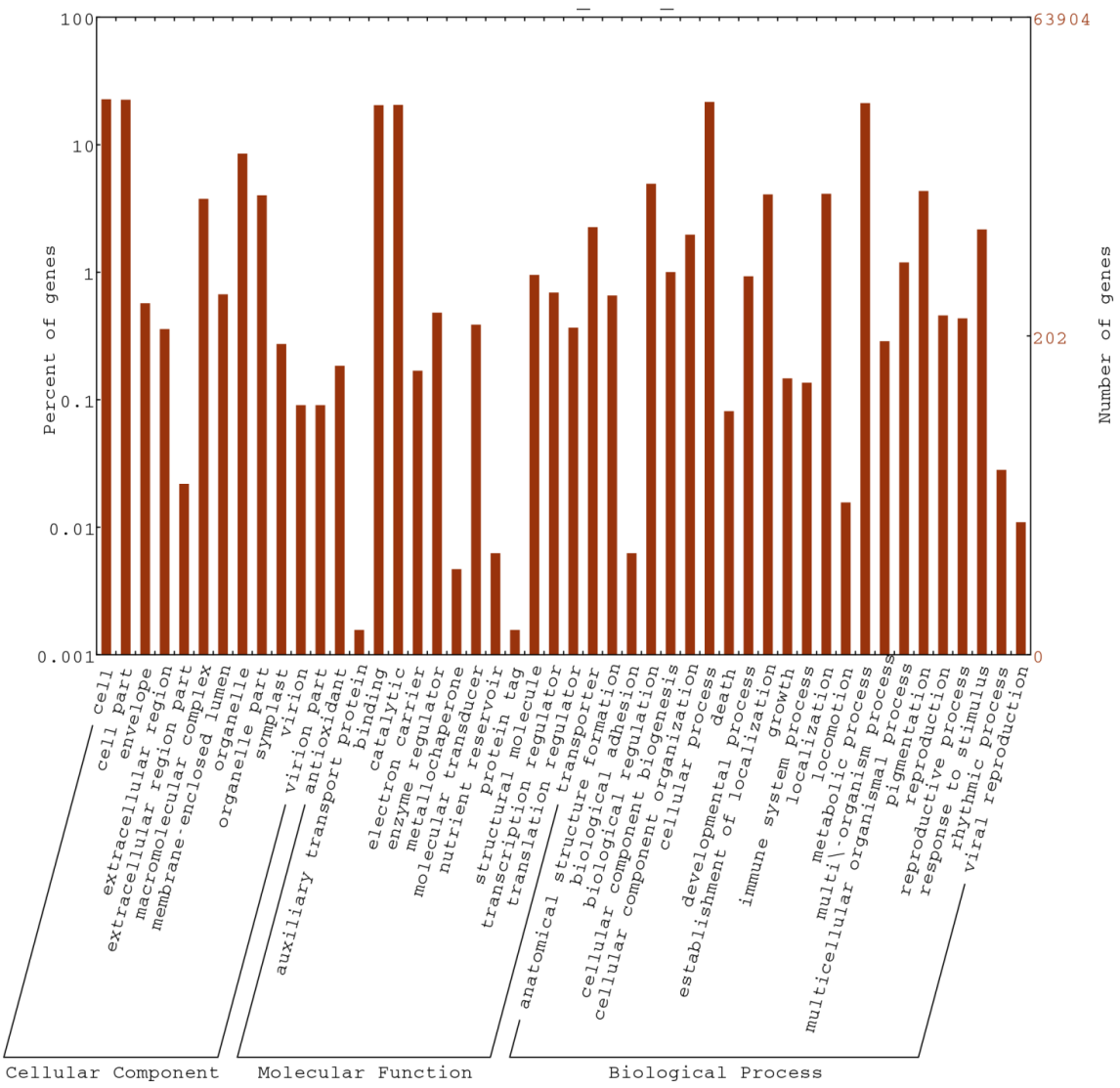


Figure 3 A. Gene Ontology: Cellular Component, Molecular Function and Biological Processes for (A) Sea buckthorn female. The figure shows a set of genes which are classified on the basis of the (1) cellular component describing locations at the levels of sub cellular structures and macromolecular complexes (2) Molecular function describing the functions of gene products and the abilities they possess (3) Biological processes gives the insight about the collection of molecular events with a defined beginning and end.

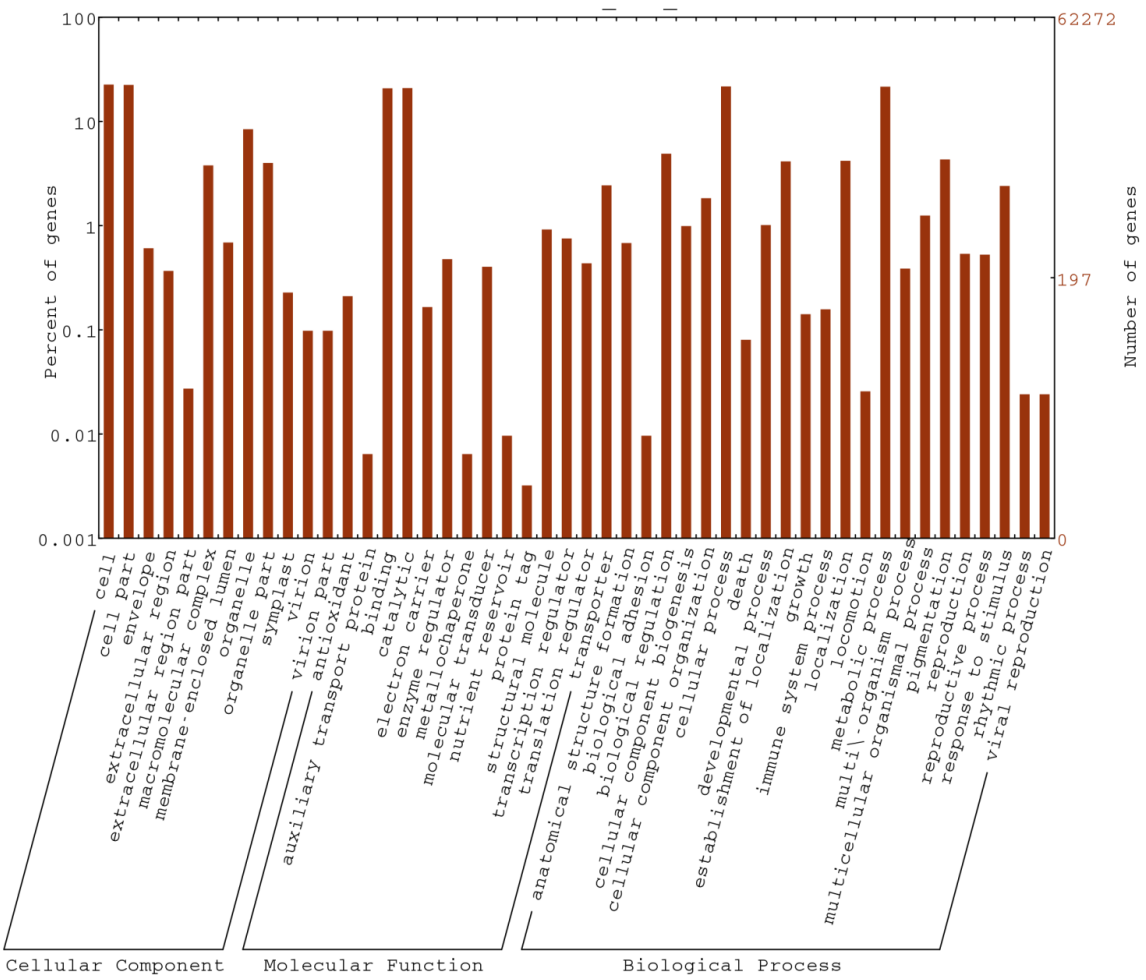


Figure 3 B. Gene Ontology: Cellular Component, Molecular Function and Biological Processes for (B) Sea buckthorn male. The figure shows a set of genes which are classified on the basis of the (1) cellular component describing locations at the levels of sub cellular structures and macromolecular complexes (2) Molecular function describing the functions of gene products and the abilities they possess (3) Biological processes gives the insight about the collection of molecular events with a defined beginning and end.

The molecular level differences on the basis of GO terms in Sea buckthorn female and male flower buds were represented as GO terms (Female, Male).

In biological processes, regulation of cellular processes (2680, 2630) was observed followed by establishment of localization (2909, 2882), single organism metabolic process (4530, 4465), biosynthesis process (4618, 4454), single organism cellular process (5017, 4886), nitrogen compound metabolic process (5092, 4864), organic substance metabolic process (11210, 10984), primary metabolic process (10492, 10354) and cellular metabolic process (10146, 10048).

In molecular functions, ion-binding (3997, 4014) was observed followed by hydrolase activity (4447,4482), carbohydrate derivative binding (4746, 4729), transferase activity (5394, 5362), organic cyclic compound binding (9013, 8901), heterocyclic compound binding (9010, 8899) and small molecule binding (5872, 5830).

In cellular components, protein complex (1679, 1674) was observed followed by intracellular organelle part (2503, 2430), membrane bounded organelle (4532, 4417), intracellular organelle (5426, 5234), intrinsic component of membrane (6599, 6476), intracellular (7708, 7368) and intracellular part (6998, 6735).

Above mentioned biological processes and cellular components doesn't show much difference in number but there is a slight increase in GO terms in female CDS compared to the male. While in molecular functions there was increase in GO terms in male compared to female especially in ion-binding and hydrolase activity. Ageez et al., 2005 discussed the role of hydrolase activity in male fertility signaling cascade [59].

KEGG Annotation

KASS pathway analysis for both Sea buckthorn Female and Sea buckthorn Male sample was carried out. Categorization of CDS was done in 25 various functional KASS pathway classification. The majority of CDS were annotated in signal transduction pathway, Transport and catabolism pathway, Translation and carbohydrate metabolism in Sea buckthorn Female and Sea buckthorn Male samples respectively (Table 3).

Table 3. KEGG Database categorises Pathway classification of Sea buckthorn Male and Sea buckthorn Female Sample CDS

Category	Pathways	Seabuckthorn Male	Seabuckthorn Female
Metabolism	Carbohydrate Metabolism	1312	1075
	Energy Metabolism	753	686
	Lipid Metabolism	638	617
	Nucleotide Metabolism	370	343
	Amino acid Metabolism	835	710
	Metabolism of other amino acids	319	306
	Glycan biosynthsis and metabolism	255	234
	Metabolism of cofactors and vitamins	510	521
	Metabolism of terpenoids and polyketides	306	261
	Biosynthsis of other secondary metabolites	296	232
	Xenobiotics biodegradation and metabolism	110	86
	Genetic Information Processing		
	Transcription	711	661
	Translation	1530	1372
	Folding, sorting and degradation	1283	1178
	Replication and repair	297	286
Environmental Information Processing	Membrane transport	66	62
	Signal transduction	1979	1684
	Signalling molecules and interaction	1	1
Cellular Processes	Transport and catabolism	1020	863
	Cell motility	133	104
	Cell growth and death	564	517
	Cellular community – eukaryotes	192	163
	Cellular community – prokaryotes	138	120
Organismal System	Environmental adaptation	496	393

Differential Gene Expression

Differential gene expression between Sea buckthorn female and male flower bud samples was carried out using scatter plot where each genes was represented by a dot. Expression level of each gene in the Sea buckthorn Male condition is represented by the vertical position whereas horizontal position represents its Sea buckthorn Female strength (Figure 4 A). As a result, genes falling above the diagonal are over-expressed and genes falling below the diagonal are under-expressed in comparison to their median expression level in experimental grouping of the experiment.

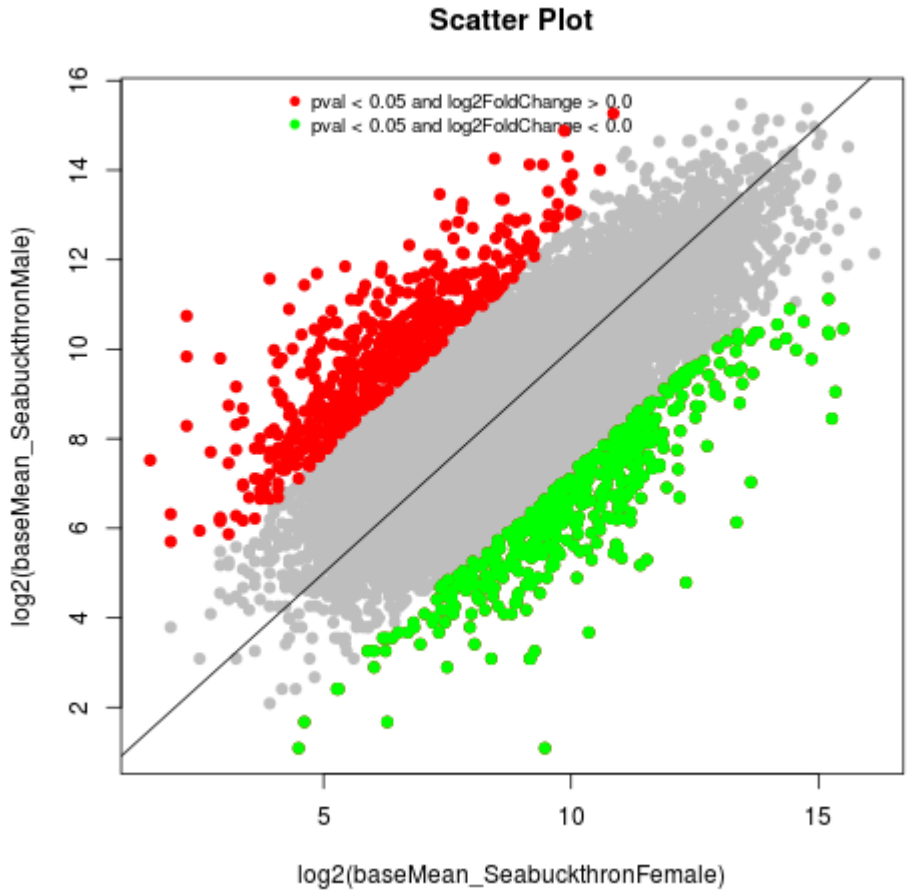


Figure 4 A. Scatter-plot showing the expression of Sea buckthorn male vs Sea buckthorn female; where X-axis represent the log2(baseMean _SeabuchthornFemale) and log2(baseMean _SeabuchthornMale) is represented on the Y-axis. The green dots represented the data having pval < 0.05 and log2 fold change < 0, whereas the red dots represented the pval < 0.05 but log2 fold change > 0.

The R script was used to represent the graphical information and distribution of differentially expressed genes which were found in Sea buckthorn female and male flower bud samples. The

volcano plot was used to cross-check the differential expressed genes where X-axis represents the log fold change in Sea buckthorn male flower bud sample to female sample where red block on the right side of zero represents the upregulated genes and green block on the left side of zero represents significant downregulated genes. Whereas Y-axis represents the negative log of p-value (value <=0.05) of the performed statistical analysis where data points with low p-value (highly significant) are appearing at the top of the plot. Grey block shows the non-differentially expressed genes (Figure 4 B).

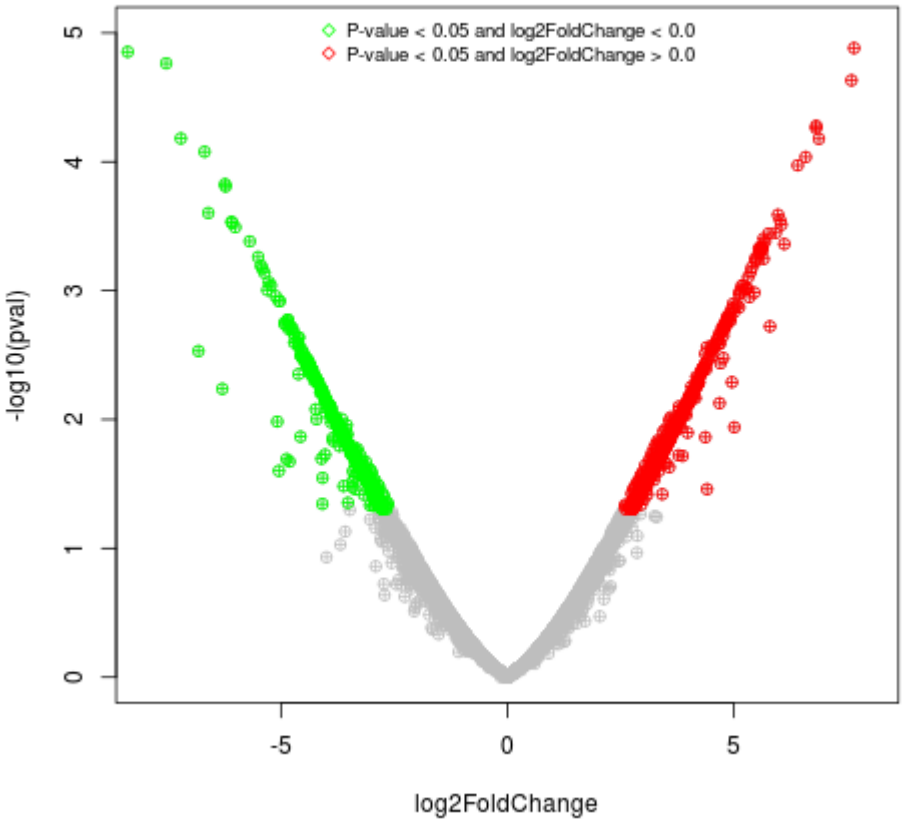


Figure 4 B. Volcano-plot showing the upregulated and downregulated expression of Sea buckthorn male vs female. In this, log2 Fold change was plotted on the X-axis and –log10 pval on the Y-axis; similarly here also green points represented the data having pval < 0.05 and log2 fold change < 0, whereas the red points represented the pval < 0.05 but log2 fold change > 0.

After performing scatter and volcano plots, it was identified that a total of 16831 genes were commonly expressed, out of which 625 genes were upregulated and 491 were found to be

downregulated. From this data, top 50 upregulated and downregulated genes were considered for hierarchal clustering (Figure 5).

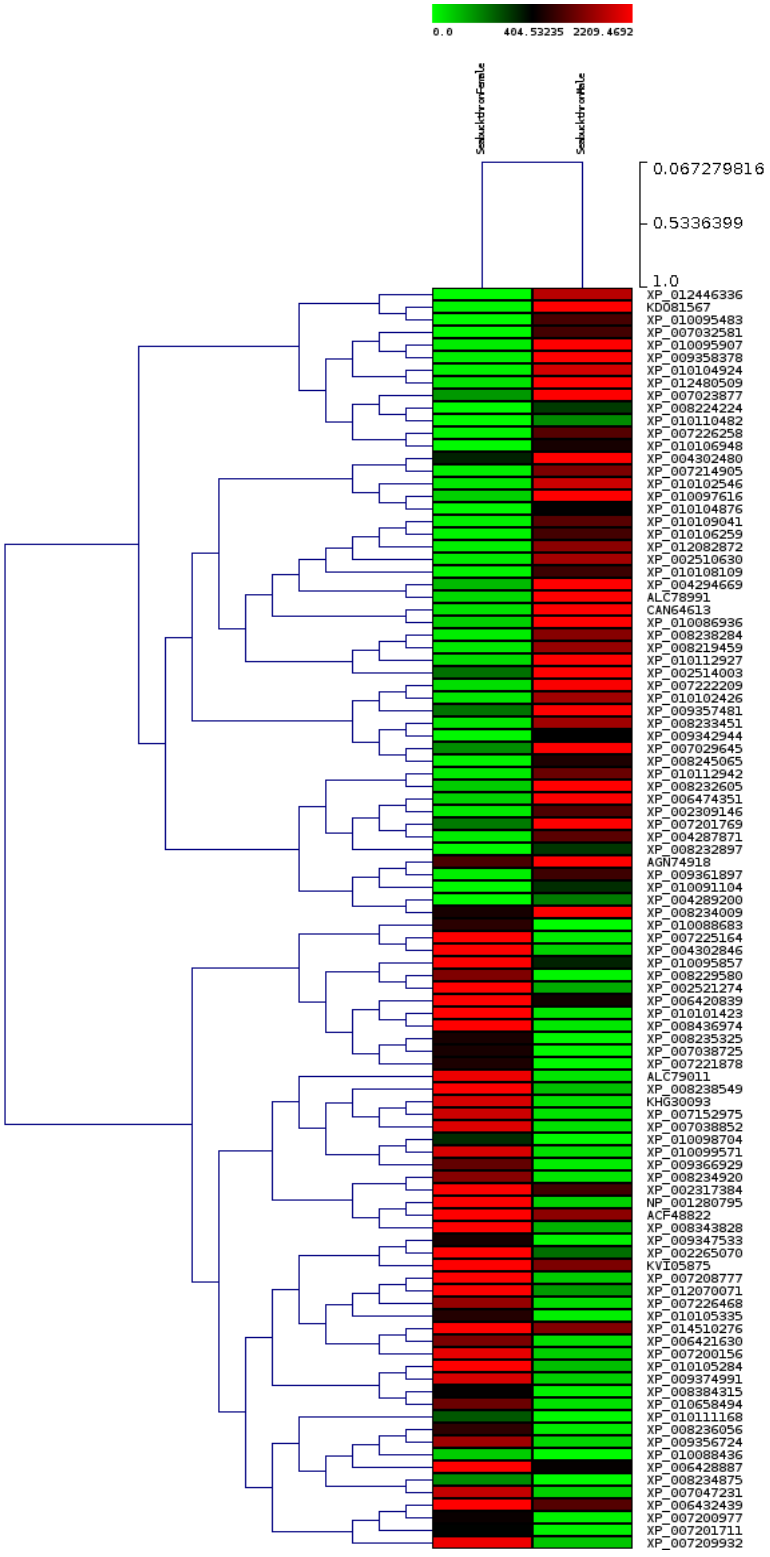


Figure 5. The figure depicts the Hierarchal clustering of Seabuckthorn Male and Female transcriptome. This explains the gene expression data with proper upregulation and downregulation patterns labelled with Accession IDs.

Differentially Expressed Genes Association with Pathways

Number of genes showing differential expression across male and female transcriptome dataset was found to be associated with various primary metabolism pathways (Table 7). In order to investigate the metabolic connections among various differentially expressed pathways we examined glycolysis, amino acid metabolism, citric acid cycle, oxidative phosphorylation, pentose phosphate pathway, carotenoid biosynthesis pathway, flavonoid pathway, purine metabolism, pyrimidine metabolism, vitamin B6, photosynthesis and terpenoid biosynthesis pathway. On the basis of expression (FPKM values), top contributing upregulated and downregulated genes were considered from male and female transcriptomes (Figure 6).

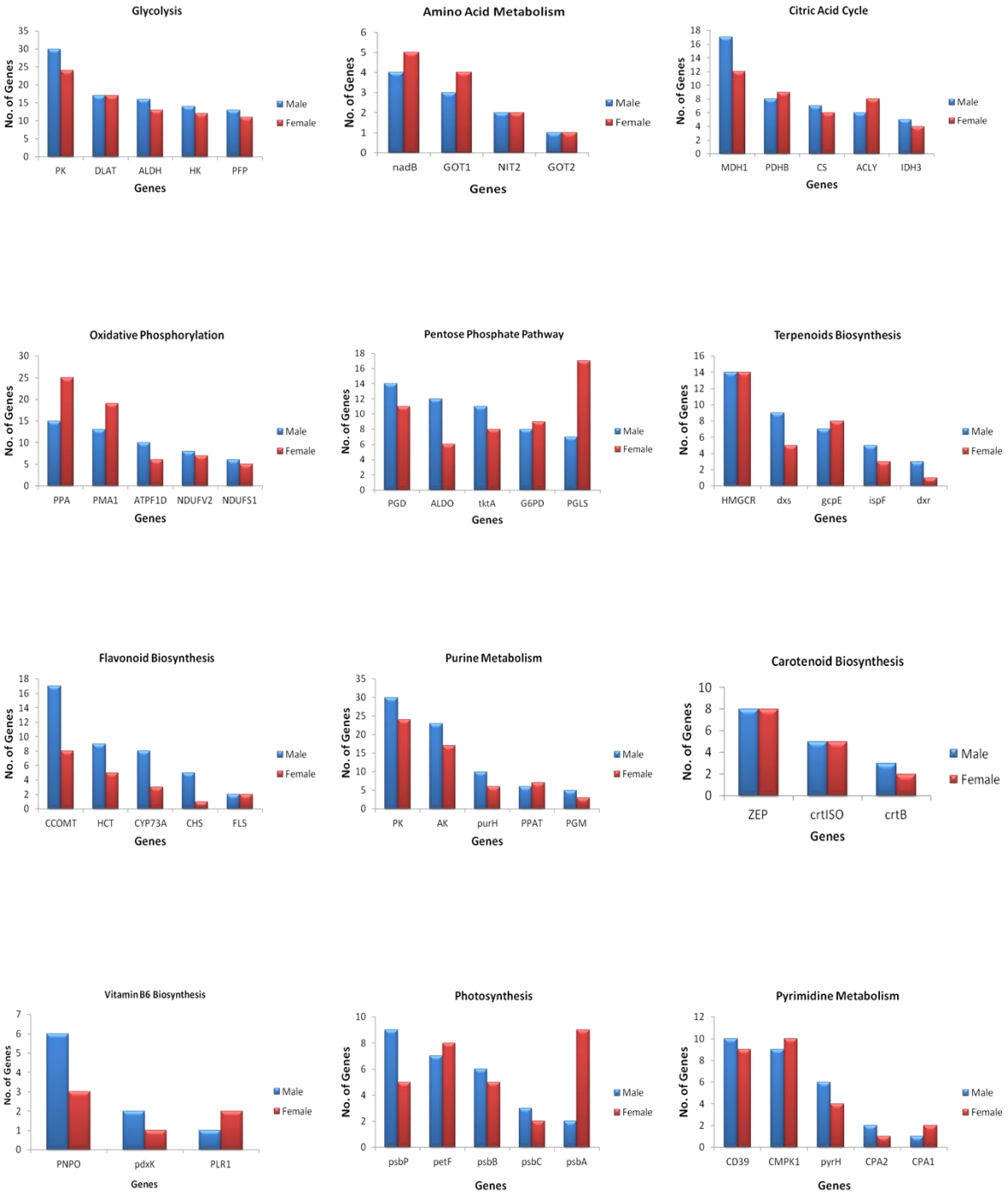


Figure 6. Pathway analysis of a few selected pathways was conducted and analyzed to form the mentioned results. The figure shows a set of graphs for various pathways showing the number and the name of different genes involved in particular pathways.

Transcriptome comparison with FLOR-ID database and their co-expression analysis

Upregulated and downregulated genes in primary metabolism pathways provide a global view of the regulatory interactions but current study is focused on identification of the genes involved in

sex determination. Henceforth, comparison of male and female transcriptome was performed against FLOR-ID (Table 4 A-B).

Table 4 A-B. Differentially expressed genes and their function in Sea buckthorn Female and Male respectively

Gene	Accession ID	Female_FPKM	Function	References
SUS4	XP_010088975	447.892179	proteosome assembling	[79]
SUS4	XP_010105926	455.4197786	RNA polymerase II transcription cofactor activity	[79]
SUS4	XP_012434276	443.1874292	Sequence specific DNA binding	[80]
SUS4	XP_010663135	469.534028	transcription regulatory region DNA binding	[81]
SUS4	XP_010105404	447.892179	Zinc ion binding	[79]
UBP12	XP_012067472	1143.254	ATP binding	[82]
UBP12	XP_010106080	1157.368	iron ion binding	[79]
PRP8A	KDP26737	4861.888	iron ion binding	[82]
PRP8A	XP_007225164	5123.473	serine-type peptidase activity	[83]
LDL1	CAN66971	1431.185	GTP Binding	[84]
LDL1	XP_002284183	1450.004	intracellular protein transfer	[81]
LDL1	XP_012076821	1506.461	metal ion binding	[82]
SFR6	XP_010112723	38.57895	auxin response factor	[79]
SFR6	XP_010090781	39.5199	zinc ion binding	[79]
SFR6	XP_010091609	59.27985	metal binding	[79]
SFR6	XP_010090105	64.92555	ATP binding	[79]
TPL	XP_012070482	462.0064	ATP binding	[82]
TPL	XP_010105926	455.4198	RNA polymerase II transcription cofactor activity	[79]
TPL	XP_010105267	470.475	transmembrane tranporter activity	[79]
TPL	XP_010105404	447.8922	Zinc ion binding	[79]
FYPP3	XP_00722182	319.923	ATP binding	[83]

	7			
FYPP3	XP_010097967	325.5687	glycosyltransferase activity	[79]
FYPP3	XP_010097707	306.7497	metabolic processes	[79]
FYPP3	XP_010096006	322.7458	transmembrane tranportation	[79]
PKL	EYU30458	210.7728	photo-lyase activity	[85]
PKL	XP_010113025	220.1823	Srna processing	[79]
PKL	XP_010096533	226.7689	Zinc ion binding	[79]

Gene	Accession ID	Male_FPKM	Function	References
SUS4	XP_010106948	572.8254	Auxin responsive	[79]
SUS4	XP_007012865	568.5743	Integral for cell membrane	[86]
SUS4	XP_010104831	573.8881	oxidoreductase activity	[79]
UBP12	XP_002324381	1133.96	sarcosine oxidase activity	[87]
PRP8A	XP_010091605	8729.476	Cellulose biosynthesis	[79]
PRP8A	CAP39915	9702.96	endonuclease activity	[87]
LDL15	XP_010086879	1363.516	ATP Binding	[79]
LDL17	XP_007201181	1374.143	chromatin remodelling	[83]
LDL21	XP_012082872	1392.21	Lipid metabolism	[84]
SFR9	XP_010096709	71.20464	ATP binding	[79]

SFR 14	XP_002282789	77.58117	DNA repair	[81]
SFR 12	KJB27319	74.3929	intracellular protein transport	[80]
TPL	XP_007220571	1794.995	flavin adenine dinucleotide binding	[83]
TPL	XP_010101004	1753.547	Lipolytic acyl hydrolase	[79]
FYP P3	XP_002511751	1826.877	ATP binding	[88]
FYP P3	XP_010102546	1853.446	ATP binding	[79]
FYP P3	XP_008348502	1818.375	ethylene resposive transcription	[89]
PKL	XP_010106948	572.8254	Auxin response factor	[79]
PKL	XP_007012865	568.5743	cell adhesion	[90]

Surprisingly, analysis resulted in only 8 flowering genes which were further considered for coexpression analysis (Figure 7 A-B).

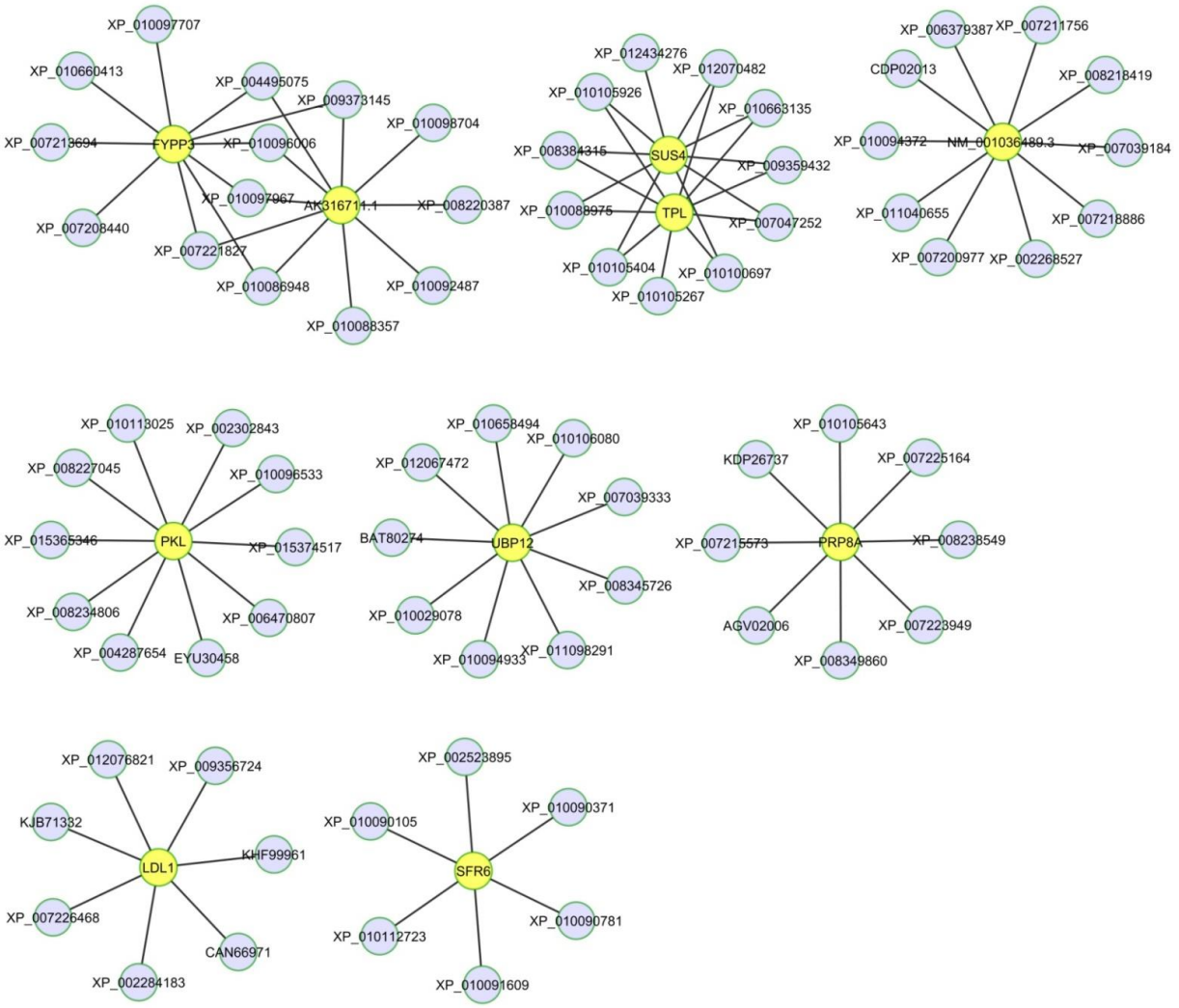


Figure 7 A. This figure illustrates the Co-expression networks of 8 genes which are found to be involved in flowering pathways in Sea buckthorn female transcriptome dataset. It is an undirected graph where each gene is represented on a node and a set of co-expressed genes are connected to the node with an edge.

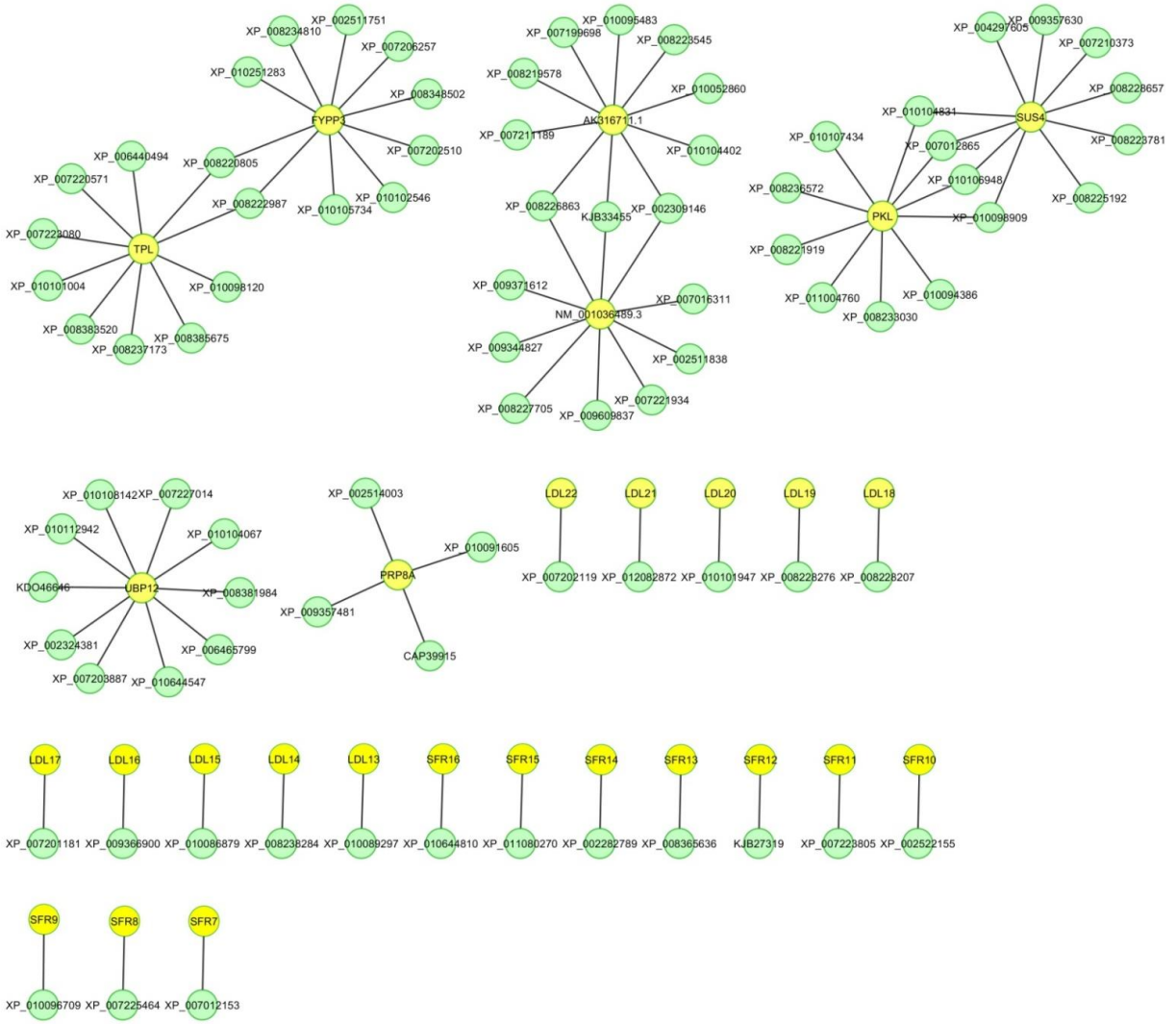


Figure 7 B. This figure illustrates the Co-expression networks of 8 genes which are found to be involved in flowering pathways in Sea buckthorn male transcriptome dataset. It is an undirected graph where each gene is represented on a node and a set of co-expressed genes are connected to the node with an edge.

PKL is a chromatin-remodelling factor and is known to inhibit the embryonic genes expression; hence plays the key role in the post germination growth [60]. Its function is also seen during carpel differentiation [61]. As per our analysis, PKL along with a bunch of other genes (L484_024032, L484_008274, L484_008224, L484_006547, TCM_037681) showing similar expressions contribute together in the gibberllin-signal transduction pathway. Gibberllin signal mechanism is well connected to the root development by the PKL gene. This gene mutate and shut down the mechanism between the embryonic and adult pathway mechanisms and even affects the development of shoot, along with the help of its sister genes [48].

Flowering plants need a photoreceptor protein in order to sense the length of days/nights and other seasonal changes, so as to signalize flowering [49]. In our analysis FYPP gene was seen to control photoperiodism in the plant by dephosphorylating the phytochromes and hence modulating their signals to control flowering time [64]. FYPP and its co-expressed genes (RCOM_1612610, PRUPE_ppa010815mg, PRUPE_ppa015809mg, L484_014223) work for the photoperiodism pathway by regulating their expressions. The overexpression of these genes postponed the time of flowering and their suppression led to the increase in the rate of flowering [51].

TPL is known to code for WD-40 sequence repeats and along with its other protein factors and found to restrict the promotion of root enhancing genes in the upper embryo division at the time of transitional embryogenesis [52]. This process takes place in the restraining temperature conditions. TPL along with a couple of a few more genes (CICLE_v10021651mg, L484_026254, L484_013182, PRUPE_ppa000037mg) incurs mutations which lead to the proper functioning of the Auxin-signalling pathway [67]. The transcriptional mechanism of Auxin (AUX) is essential for the progress and development of roots and vessels at the time of embryogenesis. TPL gene is known to be a transcriptional co-repressor and helps in modulation of hormones like AUX during the plant developmental stages [52].

Mutations in the SFR gene help the plant in dealing with the freezing temperature[54]. The SFR gene along with a few more genes (PRUPE_ppa010657mg, L484_025827, B456_004G2909001, VIT_04s0008g05690, PRUPE_ppa000230mg) co-ordinate to provide resistance against such harsh conditions. The absence of SFR6 gene mutant often leads to its unsuccessful response towards its defense mechanism which triggers the expression of the Cold-On Regulated gene via C-box binding factor [55].

For any plant to reproduce effectively, it is very important that the change from transition state into flowering state undergoes efficiently. Failure in proper functioning of the LDL genes hampers the DNA methylation on a silenced floral repressor (FWA), but the repression of the FLOWERING LOCUS C (FLC) is not affected by the DNA methylation[56]. Thus, various gene members of the LDL group (PRUPE_ppa003322mg, L484_001327, L484_008195, L484_021828, PRUPE_ppa006923mg) were observed in the silencing mechanisms [57].

Mutation in the PRP8A gene can lead to defect in the cell-division of embryonic suspensor [72],[73]. A few more genes (hbn1, RCOM_1038560, L484_026453, CICLE_v10004979mg) were analyzed together in the degradation of mRNAs which contain the premature translation termination codons i.e, PTCs. These steps of targeted degradation come under non-sense

mediated mRNA decay (NMD) pathway. Apart from degradation of mRNAs, NMD pathway also works to maintain the expression of some wild-type genes [74].

SUS4 gene functions in the cleavage of sucrose for the production of energy in the form of glucose and fructose for different metabolism pathways [75]. Sucrose is the important substrate which is required for the initial steps (especially in the non-photosynthetic tissues) of the carbohydrate biosynthesis [61]. SUS4 gene along with some other genes (TCM_037681, L484_006547, L484_008224, PRUPE_ppa001250mg) helps sucrose in formation of chromoplasts from chloroplasts for the accumulation of carotenoids [62]. The biosynthesis of carotenoids is further controlled by a set of different other pathways which combine to work in a systematic way to produce good results.

Ubiquitin carboxyl-terminal hydrolase enhance the fusion of mitochondria and even modulates the photoperiodic flowering pathway [78]. UBP12 along with many other genes (L484_023231, CISIN_1g027463mg, POPTR_0018s03430g, L484_014252, PRUPE_ppa001185mg) work to coordinate in the degradation of ubiquitin regulated proteins leading to the stabilized functioning of photoperiodic flowering pathway. Mutations in UBP genes often result in early flowering and even in over expression of FLOWERING LOCUS T. LATE ELONGATED HYPOCOTYL, CIRCADIAN CLOCK ASSOCIATED1 and TIMING OF CAB EXPRESSION1 (Cui et al., 2013).

Current study provides a holistic view and insight of male and female transcriptomic analysis but does not provide an exact gene map for the differentiation of male and female organs. Predicted CDS showing high differential expression were found unannotated, proving to be a hurdle in deciphering the underlying molecular mechanism. Although, our analysis revealed various pathways and expression based categorization of predicted genes but this could not pave a defined path to sync differentially expressed genes with already known flowering and sex determination mechanisms.

CHAPTER 5

CONCLUSION

The outcomes of this study shed light on *de novo* sequencing and characterization of transcriptomes of Sea buckthorn female and male floral buds. Transcriptome study provides a window to look into differential gene expression and gain a deep insight into regulatory mechanism at flowering and sex differentiation stages. Co-expression network analysis of selected transcripts and pathway mapping might show the diverse role of differentially expressed genes in various other studies which are related to yield content. Current study will prove an add-on to the previous transcriptomic studies and can aid a helping hand in further marker selection studies for sex based analysis on Sea buckthorn as no genome source is available till date.

REFERENCES

[1] V. Singh, *Seabuckthorn (Hippophae L.): a multipurpose wonder plant*. Daya Books, 2008.

[2] T. Fatima *et al.*, "Fatty Acid Composition of Developing Sea Buckthorn (Hippophae rhamnoides L.) Berry and the Transcriptome of the Mature Seed," *PLOS ONE*, vol. 7, no. 4, p. e34099, Apr. 2012.

[3] A. D. L. Akkermans, W. Roelofsen, J. Blom, K. Huss-Danell, and R. Harkink, "Utilization of carbon and nitrogen compounds by Frankia in synthetic media and in root nodules of Alnus glutinosa, Hippophaë rhamnoides, and Datisca cannabina," *Can. J. Bot.*, vol. 61, no. 11, pp. 2793–2800, Nov. 1983.

[4] L. Rongsen, *Seabuckthorn: a multipurpose plant species for fragile mountains*. Internat. Centre for Integrated Mountain Development, 1992.

[5] R. K. Kalia, R. Singh, M. K. Rai, G. P. Mishra, S. R. Singh, and A. K. Dhawan, "Biotechnological interventions in sea buckthorn (Hippophae L.): current status and future prospects," *Trees*, vol. 25, no. 4, pp. 559–575, Aug. 2011.

[6] S. M. Gupta, A. Grover, P. Pandey, and Z. Ahmed, "Female plants of Hippophae salicifolia D. Don are more responsive to cold stress than male plants," *Physiol. Mol. Biol. Plants Int. J. Funct. Plant Biol.*, vol. 18, no. 4, pp. 377–380, Oct. 2012.

[7] D. Charlesworth, "Plant sex determination and sex chromosomes," *Heredity*, vol. 88, no. 2, pp. 94–101, Feb. 2002.

[8] A. Chawla, A. Kant, T. Stobdan, R. B. Srivastava, and R. S. Chauhan, "Cross-species application of sex linked markers in H. salicifolia and H. tibetana," *Sci. Hortic.*, vol. 170, pp. 281–283, May 2014.

[9] A. Jain, R. Ghangal, A. Grover, S. Raghuvanshi, and P. C. Sharma, "Development of EST-based new SSR markers in seabuckthorn," *Physiol. Mol. Biol. Plants Int. J. Funct. Plant Biol.*, vol. 16, no. 4, pp. 375–378, Dec. 2010.

[10] G. Korekar *et al.*, "Identification and validation of sex-linked SCAR markers in dioecious Hippophae rhamnoides L. (Elaeagnaceae)," *Biotechnol. Lett.*, vol. 34, no. 5, pp. 973–978, May 2012.

[11] A. Sharma, G. Zinta, S. Rana, and P. Shirko, "Molecular identification of sex in Hippophae rhamnoides L. using isozyme and RAPD markers," *For. Stud. China*, vol. 12, no. 2, pp. 62–66, Jun. 2010.

[12] L. F. Delph, A. M. Arntz, C. Scotti-Saintagne, and I. Scotti, “The genomic architecture of sexual dimorphism in the dioecious plant Silene latifolia,” *Evol. Int. J. Org. Evol.*, vol. 64, no. 10, pp. 2873–2886, Oct. 2010.

[13] S. Lebel-Hardenack, E. Hauser, T. F. Law, J. Schmid, and S. R. Grant, “Mapping of sex determination loci on the white campion (Silene latifolia) Y chromosome using amplified fragment length polymorphism,” *Genetics*, vol. 160, no. 2, pp. 717–725, Feb. 2002.

[14] S. Matsunaga, S. Kawano, H. Takano, H. Uchida, A. Sakai, and T. Kuroiwa, “Isolation and developmental expression of male reproductive organ-specific genes in a dioecious campion, Melandrium album (Silene latifolia),” *Plant J. Cell Mol. Biol.*, vol. 10, no. 4, pp. 679–689, Oct. 1996.

[15] S. Adhikari, T. K. Bandyopadhyay, and P. Ghosh, “Hormonal control of sex expression of cucumber (Cucumis sativus L.) with the identification of sex linked molecular marker,” *The Nucleus*, vol. 55, no. 2, pp. 115–122, Aug. 2012.

[16] C. Foucart, A. Boualem, B. Lasseur, J. Eleblu, I. Fahraj, and A. Bendahmane, “[Sex determination in cucurbits],” *Biol. Aujourdhui*, vol. 206, no. 1, pp. 57–62, 2012.

[17] T. Wu, Z. Qin, X. Zhou, Z. Feng, and Y. Du, “Transcriptome profile analysis of floral sex determination in cucumber,” *J. Plant Physiol.*, vol. 167, no. 11, pp. 905–913, Jul. 2010.

[18] J. Liu *et al.*, “Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (Salix suchowensis),” *PloS One*, vol. 8, no. 4, p. e60181, 2013.

[19] V. Semerikov, U. Lagercrantz, V. Tsarouhas, A. Rönnberg-Wästljung, C. Alström-Rapaport, and M. Lascoux, “Genetic mapping of sex-linked markers in Salix viminalis L,” *Heredity*, vol. 91, no. 3, pp. 293–299, Sep. 2003.

[20] D. Ma *et al.*, “Transcriptome Analysis for Identification of Genes Related to Gonad Differentiation, Growth, Immune Response and Marker Discovery in The Turbot (Scophthalmus maximus),” *PLoS ONE*, vol. 11, no. 2, Feb. 2016.

[21] O. Simsek, D. Donmez, and Y. A. Kacar, “RNA-Seq Analysis in Fruit Science: A Review,” *Am. J. Plant Biol.*, vol. 2, no. 5, p. 1, Feb. 2017.

[22] A. Chawla, T. Stobdan, R. B. Srivastava, V. Jaiswal, R. S. Chauhan, and A. Kant, “Sex-Biased Temporal Gene Expression in Male and Female Floral Buds of Seabuckthorn (Hippophae rhamnoides),” *PLOS ONE*, vol. 10, no. 4, p. e0124890, Apr. 2015.

[23] Dwivedi S. and Ahmed Z., "Sea buckthorn (*Hippophae sp.*)-a Potential Underutilized Fruit Plant for Cold Arid India." *In: XXVII International Horticultural Congress-IHC2006: International Symposium on Asian Plants with Unique Horticultural 769.* pp 297-302, 2006.

[24] Jain A., Ghangal R., Grover A. *et al.*, "Development of EST-based new SSR markers in Sea buckthorn", *Physiology and Molecular Biology of Plants,* vol. 16, pp. 375-378,2010.

[25] Korekar G., Sharma R., Kumar R. *et al.*, "Identification and validation of sex-linked SCAR markers in dioecious *Hippophae rhamnoides* L. (Elaeagnaceae)", *Biotechnol Lett,* vol. 34, pp. 973-978, 2012.

[26] Persson H.A. and Nybom H., "Genetic sex determination and RAPD marker segregation in the dioecious species sea buckthorn (*Hippophae rhamnoides* L.)", *Hereditas,* vol. 129, pp. 45-51, 1998.

[27] Sharma A., Zinta G., Rana S. *et al.*, "Molecular identification of sex in *Hippophae rhamnoides* L. using isozyme and RAPD markers", *Forestry Studies in China,* vol. 12, pp. 62-66, 2010.

[28] Chuck G., "Molecular mechanisms of sex determination in monoecious and dioecious plants", *Advances in botanical research,* vol. 54, pp. 53-83, 2010.

[29] Matsunaga S., "Sex chromosome-linked genes in plants", *Genes and Genetic Systems,* vol. 81, pp. 219, 2006.

[30] Delph L.F., Arntz A.M., Scotti- Saintagne C. *et al.*, "The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*", *Evolution,* vol. 64, pp. 2873-2886, 2010.

[31] Lebel-Hardenack S., Hauser E., Law T.F. *et al.*, "Mapping of sex determination loci on the white campion (*Silene latifolia*) Y chromosome using amplified fragment length polymorphism", *Genetics,* vol. 160, pp. 717-725, 2002.

[32] Matsunaga S., Kawano S., Takano H. *et al.*, "Isolation and developmental expression of male reproductive organ-specific genes in a dioecious campion, Melandrium album (Silene latifolia)", *The Plant Journal,* vol. 10, pp. 679-689, 1996.

[33] Adhikari S., Bandyopadhyay T.K. and Ghosh P., "Hormonal control of sex expression of cucumber (*Cucumis sativus* L.) with the identification of sex linked molecular marker", *The Nucleus,* vol. 55, pp. 115-122, 2012.

[34] Foucart C., Boualem A., Lasseur B. *et al.*, "Sex determination in cucurbits", *Biologie aujourd'hui,* vol. 206, pp. 57, 2012.

[35] Wu T., Qin Z., Zhou X. *et al.*, "Transcriptome profile analysis of floral sex determination in cucumber", *Journal of plant physiology,* vol. 167, pp. 905-913, 2010.

[36] Liu J., Yin T., Ye N. *et al.*, "Transcriptome Analysis of the Differentially Expressed Genes in the Male and Female Shrub Willows (Salix suchowensis)", *PloS one,* vol. 8, pp. e60181, 2013.

[37] Semerikov V., Lagercrantz U., Tsarouhas V. *et al.*, "Genetic mapping of sex-linked markers in *Salix viminalis* L", *Heredity,* vol. 91, pp. 293-299, 2003.

[38] Yanofsky M.F., "Floral meristems to floral organs: genes controlling early events in

Arabidopsis flower development", *Annual review of plant biology*, vol. 46, pp. 167-188, 1995.

[39] Hardenack S., Ye D., Saedler H. *et al.*, "Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion", *The Plant Cell Online*, vol. 6, pp. 1775-1787, 1994.

[40] Ainsworth C., Crossley S., Buchanan-Wollaston V. *et al.*, "Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression", *The Plant Cell Online*, vol. 7, pp. 1583-1598, 1995.

[41] Kim H.B., Jun S.-S., Choe S. *et al.*, "Identification of differentially expressed genes from male and female flowers of kiwifruit", *African Journal of Biotechnology*, vol. 9, pp. 6684-6694, 2010.

[42] H. Knight, S. G. Mugford, B. Ulker, D. Gao, G. Thorlby, and M. R. Knight, "Identification of SFR6, a key component in cold acclimation acting post-translationally on CBF function," *Plant J. Cell Mol. Biol.*, vol. 58, no. 1, pp. 97–108, Apr. 2009.

[43] M. Zhao, S. Yang, X. Liu, and K. Wu, "Arabidopsis histone demethylases LDL1 and LDL2 control primary seed dormancy by regulating DELAY OF GERMINATION 1 and ABA signaling-related genes," *Front. Plant Sci.*, vol. 6, Mar. 2015.

[44] G. G. Simpson, "The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time," *Curr. Opin. Plant Biol.*, vol. 7, no. 5, pp. 570–574, Oct. 2004.

[45] B. W. Schwartz, E. C. Yeung, and D. W. Meinke, "Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of Arabidopsis," *Development*, vol. 120, no. 11, pp. 3235–3245, Nov. 1994.

[46] C. Koncz, F. deJong, N. Villacorta, D. Szakonyi, and Z. Koncz, "The Spliceosome-Activating Complex: Molecular Mechanisms Underlying the Function of a Pleiotropic Regulator," *Front. Plant Sci.*, vol. 3, 2012.

[47] A. H. Benkovics, T. Nyikó, Z. Mérai, D. Silhavy, and G. D. Bisztray, "Functional analysis of the grapevine paralogs of the SMG7 NMD factor using a heterolog VIGS-based gene depletion-complementation system," *Plant Mol. Biol.*, vol. 75, no. 3, pp. 277–290, Feb. 2011.

[48] M. Salanoubat *et al.*, "Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana," *Nature*, vol. 408, no. 6814, pp. 820–822, Dec. 2000.

[49] H. Fu, S. Y. Kim, and W. D. Park, “A potato Sus3 sucrose synthase gene contains a context-dependent 3’ element and a leader intron with both positive and negative tissue-specific effects.,” *Plant Cell*, vol. 7, no. 9, pp. 1395–1403, Sep. 1995.

[50] H. Li and R. Durbin, “Fast and accurate short read alignment with Burrows-Wheeler transform,” *Bioinforma. Oxf. Engl.*, vol. 25, no. 14, pp. 1754–1760, Jul. 2009.

[51] B. J. Haas *et al.*, “De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis,” *Nat. Protoc.*, vol. 8, no. 8, pp. 1494–1512, Aug. 2013.

[52] A. Conesa and S. Götz, “Blast2GO: A comprehensive suite for functional analysis in plant genomics,” *Int. J. Plant Genomics*, vol. 2008, p. 619832, 2008.

[53] M. Tanabe and M. Kanehisa, “Using the KEGG database resource,” *Curr. Protoc. Bioinforma.*, vol. Chapter 1, p. Unit1.12, Jun. 2012.

[54] S. Anders and W. Huber, “Differential expression analysis for sequence count data,” *Genome Biol.*, vol. 11, no. 10, p. R106, 2010.

[55] E. A. Howe, R. Sinha, D. Schlauch, and J. Quackenbush, “RNA-Seq analysis in MeV,” *Bioinforma. Oxf. Engl.*, vol. 27, no. 22, pp. 3209–3210, Nov. 2011.

[56] R. Ekblom and J. B. W. Wolf, “A field guide to whole-genome sequencing, assembly and annotation,” *Evol. Appl.*, vol. 7, no. 9, pp. 1026–1042, Nov. 2014.

[57] S. Huang, H. Zeng, J. Zhang, S. Wei, and L. Huang, “Interconversions of different forms of vitamin B6 in tobacco plants,” *Phytochemistry*, vol. 72, no. 17, pp. 2124–2129, Dec. 2011.

[58] S. Johansson, S. Lindstedt, and H. G. Tiselius, “Metabolic interconversions of different forms of vitamin B6,” *J. Biol. Chem.*, vol. 249, no. 19, pp. 6040–6046, Oct. 1974.

[59] A. Ageez, Y. Kazama, R. Sugiyama, and S. Kawano, “Male-fertility genes expressed in male flower buds of Silene latifolia include homologs of anther-specific genes,” *Genes Genet. Syst.*, vol. 80, no. 6, pp. 403–413, Dec. 2005.

[60] J. Ogas, S. Kaufmann, J. Henderson, and C. Somerville, “PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 24, pp. 13839–13844, Nov. 1999.

[61] Y. Eshed, S. F. Baum, and J. L. Bowman, “Distinct mechanisms promote polarity establishment in carpels of Arabidopsis,” *Cell*, vol. 99, no. 2, pp. 199–209, Oct. 1999.

[62] C. B. Taylor, “GA Signaling: Genes and GTPases,” *Plant Cell*, vol. 10, no. 2, pp. 131–134, Feb. 1998.

[63] K. C. Hamner and J. Bonner, “Photoperiodism in Relation to Hormones as Factors in Floral Initiation and Development,” *Bot. Gaz.*, vol. 100, no. 2, pp. 388–431, Dec. 1938.

[64] D.-H. Kim, J.-G. Kang, S.-S. Yang, K.-S. Chung, P.-S. Song, and C.-M. Park, “A phytochrome-associated protein phosphatase 2A modulates light signals in flowering time control in Arabidopsis,” *Plant Cell*, vol. 14, no. 12, pp. 3043–3056, Dec. 2002.

[65] M. Ni, “Integration of light signaling with photoperiodic flowering and circadian rhythm,” *Cell Res.*, vol. 15, no. 8, pp. 559–566, Aug. 2005.

[66] H. Szemenyei, M. Hannon, and J. A. Long, “TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis,” *Science*, vol. 319, no. 5868, pp. 1384–1386, Mar. 2008.

[67] K. S. Osmont and C. S. Hardtke, “The topless plant developmental phenotype explained!,” *Genome Biol.*, vol. 9, no. 4, p. 219, 2008.

[68] H. Knight, E. L. Veale, G. J. Warren, and M. R. Knight, “The sfr6 mutation in Arabidopsis suppresses low-temperature induction of genes dependent on the CRT/DRE sequence motif,” *Plant Cell*, vol. 11, no. 5, pp. 875–886, May 1999.

[69] R. Ghangal, S. Raghuvanshi, and P. C. Sharma, “Expressed sequence tag based identification and expression analysis of some cold inducible elements in seabuckthorn (Hippophae rhamnoides L.),” *Plant Physiol. Biochem.*, vol. 51, pp. 123–128, Feb. 2012.

[70] R. K. Kalia, R. Singh, M. K. Rai, G. P. Mishra, S. R. Singh, and A. K. Dhawan, “Biotechnological interventions in sea buckthorn (Hippophae L.): current status and future prospects,” *Trees*, vol. 25, no. 4, pp. 559–575, Aug. 2011.

[71] C.-J. Ruan, K. Rumpunen, and H. Nybom, “Advances in improvement of quality and resistance in a multipurpose crop: sea buckthorn,” *Crit. Rev. Biotechnol.*, vol. 33, no. 2, pp. 126–144, Jun. 2013.

[72] A. Wang, Q. Zhang, D. Wan, YanzhuoYang, and J. Liu, “Nine microsatellite DNA primers for Hippophae rhamnoides ssp. sinensis (Elaeagnaceae),” *Conserv. Genet.*, vol. 9, no. 4, pp. 969–971, Aug. 2008.

[73] A. Grover, B. Ramesh, and P. C. Sharma, "Development of microsatellite markers in potato and their transferability in some members of Solanaceae," *Physiol. Mol. Biol. Plants*, vol. 15, no. 4, pp. 343–358, Oct. 2009.

[74] R. K. Varshney, A. Graner, and M. E. Sorrells, "Genic microsatellite markers in plants: features and applications," *Trends Biotechnol.*, vol. 23, no. 1, pp. 48–55, Jan. 2005.

[75] P. C. Sharma, A. Grover, and G. Kahl, "Mining microsatellites in eukaryotic genomes," *Trends Biotechnol.*, vol. 25, no. 11, pp. 490–498, Nov. 2007.

[76] A. Jain, R. Ghangal, A. Grover, S. Raghuvanshi, and P. C. Sharma, "Development of EST-based new SSR markers in seabuckthorn," *Physiol. Mol. Biol. Plants Int. J. Funct. Plant Biol.*, vol. 16, no. 4, pp. 375–378, Dec. 2010.

[77] X. Zhou *et al.*, "Regulatory control of high levels of carotenoid accumulation in potato tubers," *Plant Cell Environ.*, vol. 34, no. 6, pp. 1020–1030, Jun. 2011.

[78] X. Cui *et al.*, "Ubiquitin-specific proteases UBP12 and UBP13 act in circadian clock and photoperiodic flowering regulation in Arabidopsis," *Plant Physiol.*, vol. 162, no. 2, pp. 897–906, Jun. 2013.

[79] N. He *et al.*, "Draft genome sequence of the mulberry tree Morus notabilis," *Nat. Commun.*, vol. 4, p. 2445, Sep. 2013.

[80] A. H. Paterson *et al.*, "Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres," *Nature*, vol. 492, no. 7429, pp. 423–427, Dec. 2012.

[81] O. Jaillon *et al.*, "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla," *Nature*, vol. 449, no. 7161, pp. 463–467, Sep. 2007.

[82] L. Zhang *et al.*, "Global Analysis of Gene Expression Profiles in Physic Nut (Jatropha curcas L.) Seedlings Exposed to Salt Stress," *PLOS ONE*, vol. 9, no. 5, p. e97878, May 2014.

[83] The International Peach Genome Initiative *et al.*, "The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution," *Nat. Genet.*, vol. 45, no. 5, pp. 487–494, May 2013.

[84] R. Velasco *et al.*, "The first genome sequence of an elite grapevine cultivar (Pinot noir; Vitis vinifera L.): coping with a highly heterozygous genome," presented at the COST 858: Grape genome bioinformatics, 2006, p. 1.

[85] U. Hellsten *et al.*, “Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 48, pp. 19478–19482, Nov. 2013.

[86] J. C. Motamayor *et al.*, “The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color,” *Genome Biol.*, vol. 14, no. 6, p. r53, Jun. 2013.

[87] G. A. Tuskan *et al.*, “The genome of black cottonwood, Populus trichocarpa (Torr. & Gray),” *Science*, vol. 313, no. 5793, pp. 1596–1604, Sep. 2006.

[88] J. Matousek, L. Orctová, J. Skopek, K. Pesina, and G. Steger, “Elimination of hop latent viroid upon developmental activation of pollen nucleases,” *Biol. Chem.*, vol. 389, no. 7, pp. 905–918, Jul. 2008.

[89] A. P. Chan *et al.*, “Draft genome sequence of the oilseed species Ricinus communis,” *Nat. Biotechnol.*, vol. 28, no. 9, pp. 951–956, Sep. 2010.

[90] E. Tacken *et al.*, “The role of ethylene and cold temperature in the regulation of the apple POLYGALACTURONASE1 gene and fruit softening,” *Plant Physiol.*, vol. 153, no. 1, pp. 294–305, May 2010.