3 Biotech (2018) 8:96
https://doi.org/10.1007/s13205-018-1122-5

ORIGINAL ARTICLE

Comparative de novo transcriptome analysis of male and female Sea buckthorn

Ankush Bansal[1] · Mehul Salaria[1] · Tashil Sharma[1] · Tsering Stobdan[2] · Anil Kant[1]

Received: 1 July 2017 / Accepted: 16 January 2018 / Published online: 25 January 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Sea buckthorn is a dioecious medicinal plant found at high altitude. The plant has both male and female reproductive organs in separate individuals. 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 transcriptomes were identified to understand the underlying sex determination mechanism. The current study showed 63,904 and 62,272 coding sequences (CDS) in female and male transcriptome data sets, respectively. 16,831 common CDS were screened out from both transcriptomes, out of which 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 were performed. Comparison with Flowering Interactive Database (FLOR-ID) resulted in eight differentially expressed genes viz. CHD3-type chromatin-remodeling factor PICKLE (*PKL*), phytochrome-associated serine/threonine–protein phosphatase (*FYPP*), protein TOPLESS (*TPL*), sensitive to freezing 6 (*SFR6*), lysine-specific histone demethylase 1 homolog 1 (*LDL1*), pre-mRNA-processing-splicing factor 8A (*PRP8A*), sucrose synthase 4 (*SUS4*), ubiquitin carboxyl-terminal hydrolase 12 (*UBP12*), known to be broadly involved in flowering, photoperiodism, embryo development, and cold response pathways. Male and female flower bud transcriptome data of Sea buckthorn may provide comprehensive information at genomic level for the identification of genetic regulation involved in sex determination.

Keywords Sea buckthorn · Transcriptome · Network reconstruction · Differential expression · Regulatory networks · Computational biology · Co-expression

Introduction

Sea buckthorn (*Hippophae rhamnoides* L.) is 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 since hundreds of years (Singh 2008). Berries of the Sea buckthorn plant are a potential source of many bioactive compounds including vitamins, organic acids, fatty acids, flavonoids, and antioxidants, thus providing various health-related benefits (Fatima et al. 2012). The oil in pulp and seeds of berries exhibits cardioprotective, antiplatelet, and anti-ulcer activities, thus making this plant important for commercial as well as research interests (Fatima et al. 2012). Sea buckthorn grows in the harsh environment of cold deserts, thus showing tolerance towards extreme temperatures, drought, and soil salinity prevailing in such areas. The root system of Sea

Ankush Bansal, Mehul Salaria, and Tashil Sharma contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-018-1122-5) contains supplementary material, which is available to authorized users.

✉ Anil Kant
anil_kantv@yahoo.com

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh 173234, India

[2] Defence Institute of High Altitude Research, Defence R&D Organisation, Leh, Jammu and Kashmir, India

buckthorn is highly advanced and acts as an excellent biotic choice for prevention of soil erosion on deeply unreliable slopes of high altitude. Mountain soil deprive in nutrients, especially in organic matter and nitrogen. Sea buckthorn providing habitable environment due to its ability to fix nitrogen through its association with *Frankia*, a nitrogen-fixing actinomycete (Akkermans et al. 1983). Approximately 180 kg of nitrogen per hectare per annum can be fastened in the soil in a Sea buckthorn forest (Rongsen 1992). Thus, Sea buckthorn is a good plantation shrub for high altitude areas as it offers people with the opportunity to maintain a more sustainable lively hood as well as prevent soil erosion in fragile topography.

The demand of Sea buckthorn berries has increased in the last few years due to high consumption of commercial products like juice, oil, alcoholic beverages, biscuits, ice-cream, tea, jam, candies, etc. (Kalia et al. 2011). The huge demand also permits its intensive cultivation instead of their collection from wild resources and also their genetic improvement in terms 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 (Kalia et al. 2011). For the breeding program to be successful in dioecious plants, the early identification of progeny's gender is necessary. However, unfortunately, the gender of Sea buckthorn seedlings cannot be known until flowering, which usually takes place after 3–4 years in the field (Gupta et al. 2012). 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 (Charlesworth 2002). Moreover, various molecular studies such as random amplification of polymorphic DNA (RAPD), simple sequence repeat (SSR), inter-simple sequence repeat (ISSR), sequence characterized amplified region (SCAR), etc. were being performed from past several years for gender identification in Sea buckthorn (Sharma et al. 2010; Jain et al. 2010; Korekar et al. 2012; Chawla et al. 2014). Unfortunately, none of the 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* (Matsunaga et al. 1996; Lebel-Hardenack et al. 2002; Delph et al. 2010), *Cucumis sativus* (Wu et al. 2010; Foucart et al. 2012; Adhikari et al. 2012), *Salix* (Semerikov et al. 2003; Liu et al. 2013), 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*, *Rumex acetosa*, *Actinidia chinensis*, etc. (Di Stilio et al. 2005; Jeong et al. 2010). 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*, and *A. Officinalis* (Adam et al. 2011; Uragami et al. 2016) indicating their role in sex determination in the above-mentioned plants.

During the 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 organism. The speed and efficiency of gene discovery have improved significantly due to the emergence of the next-generation sequencing technologies which have potential to generate millions of reads and expose complete transcriptome profiling of any organism (Kumar et al. 2017; Bansal et al. 2017; Bansal and Srivastava 2018). RNA-Seq based on the next-generation sequencing has become widely functional to obtain mass sequence data for molecular marker development, transcriptional analysis, and gene discovery (Ghangal et al. 2013).

Until date, there is no report on transcriptome analysis of female plants directed to decipher genes involved in sex determination of Sea buckthorn. 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. Furthermore, 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.

Materials and methods

Plant material

Three different samples of floral buds for the current study were collected on the basis of phenological observations on flowering of Sea buckthorn from the Defence Institute of High Altitude Research (DIHAR), Jammu and Kashmir, India (geographic coordinates—34°08′236″N, 77°34′345″E) (Chawla et al. 2015). Flower buds start developing from the month April and flowers open in the start of May to mid-May. The flower bud samples were collected in the month of April at 10 day interval, starting from dormant winter bud to when buds are about to open. This is period when female and male reproductive tissues are formed in the flower buds. Male Stage I, Male Stage II, and Male Stage III were pooled together. In similar fashion, Female Stage I, Female Stage II,

and Female Stage III were pooled together for transcriptomic analysis to get understanding of overall transcript expression during flowering time. Flower buds were immediately frozen in liquid nitrogen and were stored at − 80 °C generation of NGS transcriptomes.

Total RNA isolation, Illumina NextSeq 500 PE library preparation, and quality check

Total RNA was isolated from the male and female flower bud samples using ZR plant RNA Miniprep™ (ZYMO Research) according to the manufacturer's guidelines. The quality and quantity of the isolated RNA were 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. Transcriptome sequencing of collected male and female samples were outsourced from Eurofins Scientific laboratories. The adapters were intended to permit re-synthesis of the reverse strands followed by cleavage of the forward strand during sequencing. The reverse strand copies were further used to sequence 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 452–453 bp for male and female flower buds, respectively. The libraries were sequenced on NextSeq 500 using 2 × 75 bp chemistry.

High-quality read generation

The sequenced raw data were assembled 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) (Bolger et al. 2014). 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 × 75 bp) and 42,922,794 (2 × 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 this high quality (QV > 20), paired-end reads were used. The following parameters were considered for filtration of high-quality reads: (1) adapter trimming; (2) sliding window: conduct a sliding window trimming of 20 bp, cutting once the average quality within the window falls below a threshold of 20; (3) leading: if threshold quality is below 20, bases are cutoff at the starting of a read; (4) trailing: if threshold quality reaches below 20, bases are cut-off at the end of a read; and (5) Mini_length: if the read was below 50 bp length, 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 (Li and Durbin 2009; Haas et al. 2013). During the assembly, large amounts of misassembled, erroneous, and poorly supported transcripts can be encountered. Thereafter, high-quality reads were mapped back to their respective assembled transcripts for validation using BWA v0.7.12 (Conesa and Götz 2008). The complete workflow for transcriptome analysis is depicted in Fig. 1.

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 greater when the ORF is scored in the first reading frame as compared to scores in the other five reading frames; and (4) in case, candidate ORF seems to enclosed by the other candidate ORF, where the longer ORF is considered. However, multiple ORFs can be reported by a single transcript (Tanabe and Kanehisa 2012).

Gene ontology analysis

Gene ontology (GO) annotations of the coding sequence (CDS) were identified using the Blast2GO program (Anders and Huber 2010). GO terms were employed for categorizing the functions of predicted CDS using three main domains which are: (1) biological processes (BP); (2) molecular functions (MF); and (3) cellular components (CC). To retrieve GO terms, the following criteria was used: (1) BLASTX result's accession IDs were used to identify gene symbols. Retrieved gene symbols were then searched in the species-specific entries GO database; (2) BLASTX results were further 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 were searched directly in the dbxref table of GO database; and (4) BLASTX result accession IDs were searched directly in the gene product table of GO database.

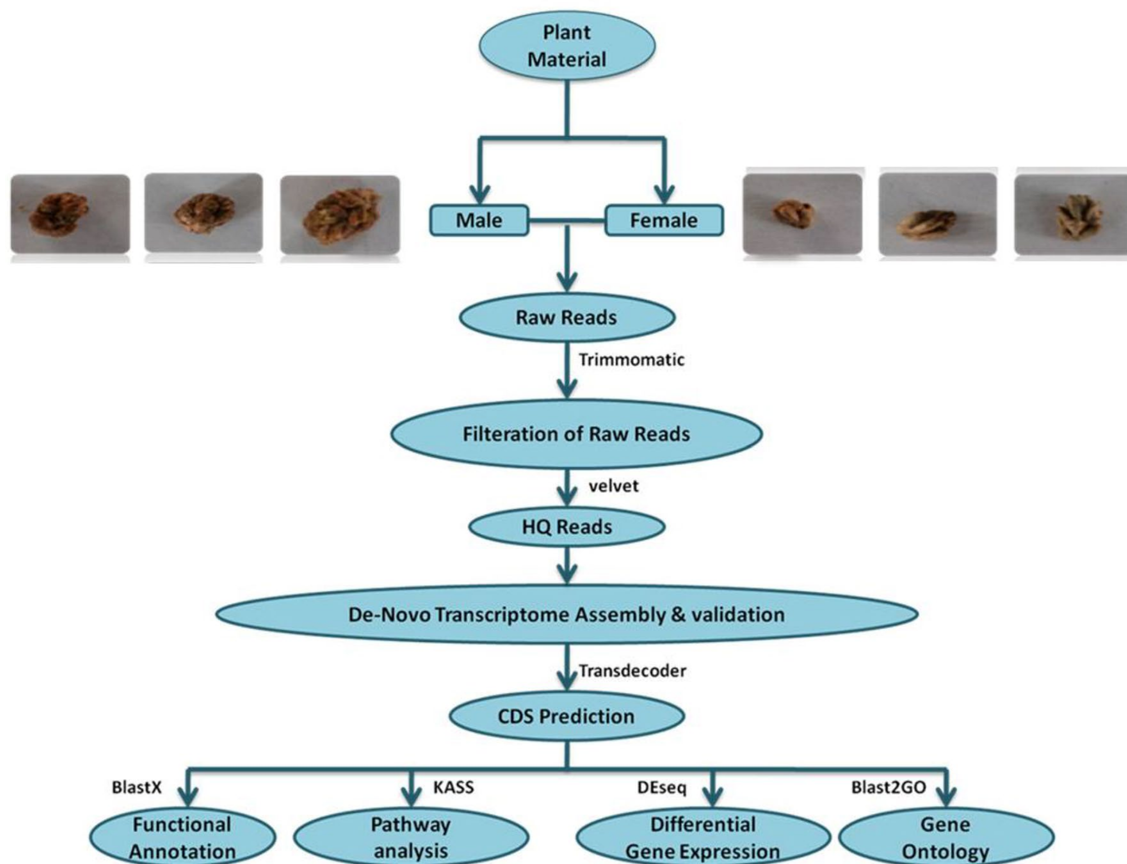

Fig. 1 Complete framework showing the steps conducted to perform male and female transcriptome-based analyses like functional annotation using BlastX, pathway analysis using KASS, differential gene expression using DEseq, and gene ontology using Blast2GO

KEGG pathway analysis

To identify the potential involvement of the predicted CDS of Sea buckthorn female and male flower bud samples in biological pathways, CDS were formulated into reference canonical pathways in KEGG (Howe et al. 2011). 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 along with metabolic pathways of predicted CDS can be obtained 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 respective set of CDS was done using BWA aligner (Ekblom and Wolf 2014). The differentially expressed gene analysis was performed using a negative binomial distribution model, i.e., DeSeq v1.8.1 package (http://www-huber.embl.de/users/anders/DESeq) (Huang et al. 2011). Furthermore, these genes were classified as upregulated and downregulated based on their log fold change (FC) values, where FC = $\log_2$ (Sea buckthorn male/Sea buckthorn female). FC value greater than zero was considered as upregulated and less than zero as downregulated. *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 topmost 100 differentially expressed genes (Johansson et al. 1974). Heat map (cluster) shows the expression of genes/level of expression and is represented by the $\log_2$ ratio of gene abundance between Sea buckthorn female vs. male. Differentially expressed genes identified in Sea buckthorn female and male was examined using hierarchical clustering. A heatmap 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, an in-house database of FLOR-ID nucleotide sequences has been constructed and used for comparison against Sea buckthorn female and male transcriptome 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 was derived via non-parametric Pearson correlation coefficient (r) which was calculated using in-house Perl script:

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

where j_i and k_i are the degrees of targets at both the ends of the ith connection and M represents the total connections in the network.

Table 1 On the basis of gene expression distribution of no. of transcript according to their length

Range of transcript	Sea buckthorn female	Sea buckthorn male
200 ≤ transcript < 500	12,596	13,486
500 ≤ transcript < 1000	15,076	14,737
1000 ≤ transcript < 2000	25,824	24,717
2000 ≤ transcript < 3000	11,188	11,066
3000 ≤ transcript < 4000	3345	3531
4000 ≤ transcript < 5000	909	1122
Transcript ≥ 5000	519	731

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	15,346	14,821
500 ≤ CDS < 1000	22,703	21,514
1000 ≤ CDS < 2000	20,298	19,727
2000 ≤ CDS < 3000	4400	4546
3000 ≤ CDS < 4000	961	1157
4000 ≤ CDS < 5000	150	350
CDS ≥ 5000	46	157

Results and discussion

Paired-end sequencing of cDNA library and de novo assembly of transcriptome

Whole transcriptome sequencing was performed on Sea buckthorn female and male flower bud samples by Illumina NextSeq 500 platform using 2 × 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).

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 1019 in the females and 66,059,358 bases with 1060 as the mean value in the males. The maximum length of CDS was obtained to be 7050 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).

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* (Fig. 2).

Gene ontology

From the gene ontology (GO) analysis, 26,506 and 26,071 CDS was annotated with gene ontology terms (Fig. 3a, 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

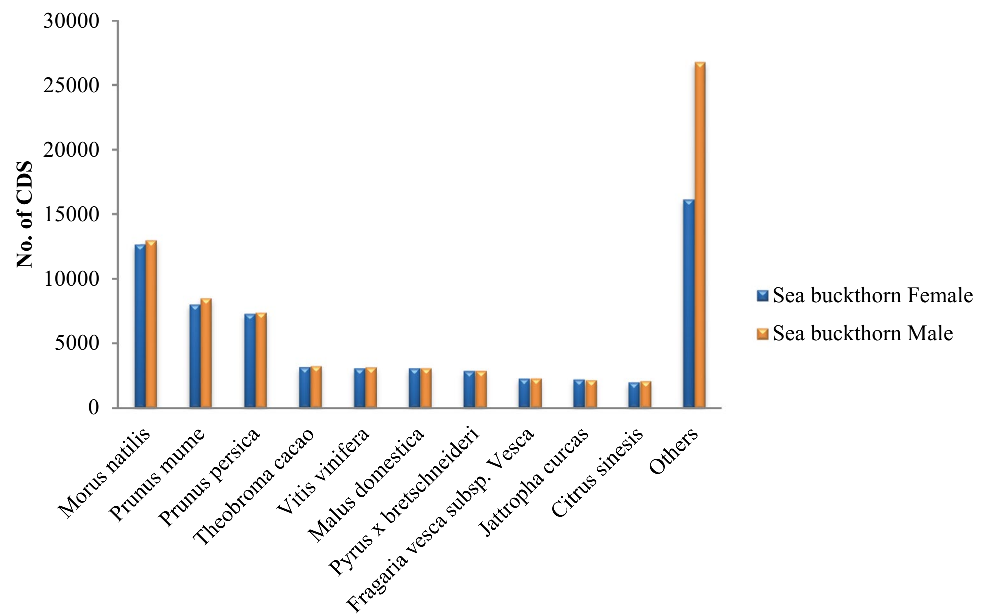

Fig. 2 Top BLAST results for sea buckthorn female and sea buckthorn male

biological processes, 20,841 in molecular functions, and 14,207 in cellular components.

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 (11,210, 10,984), primary metabolic process (10,492, 10,354), and cellular metabolic process (10,146, 10,048).

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, a 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).

Biological processes and cellular components mentioned above display a slight increase in GO terms in female CDS compared to the male. While in molecular functions, there was an 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.

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 classifications. 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).

Differential gene expression

Differential gene expression between Sea buckthorn female and male flower bud samples was carried out using scatter plot, where each gene was represented by a dot. The expression level of each gene in the Sea buckthorn male condition is represented by the vertical position and the horizontal position represents its Sea buckthorn female strength (Supplementary Fig. 1A). As a result, genes falling above the diagonal are over-expressed and genes falling below the diagonal are under-expressed in comparison with their median expression level in experimental grouping of the experiment.

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-to-female flower

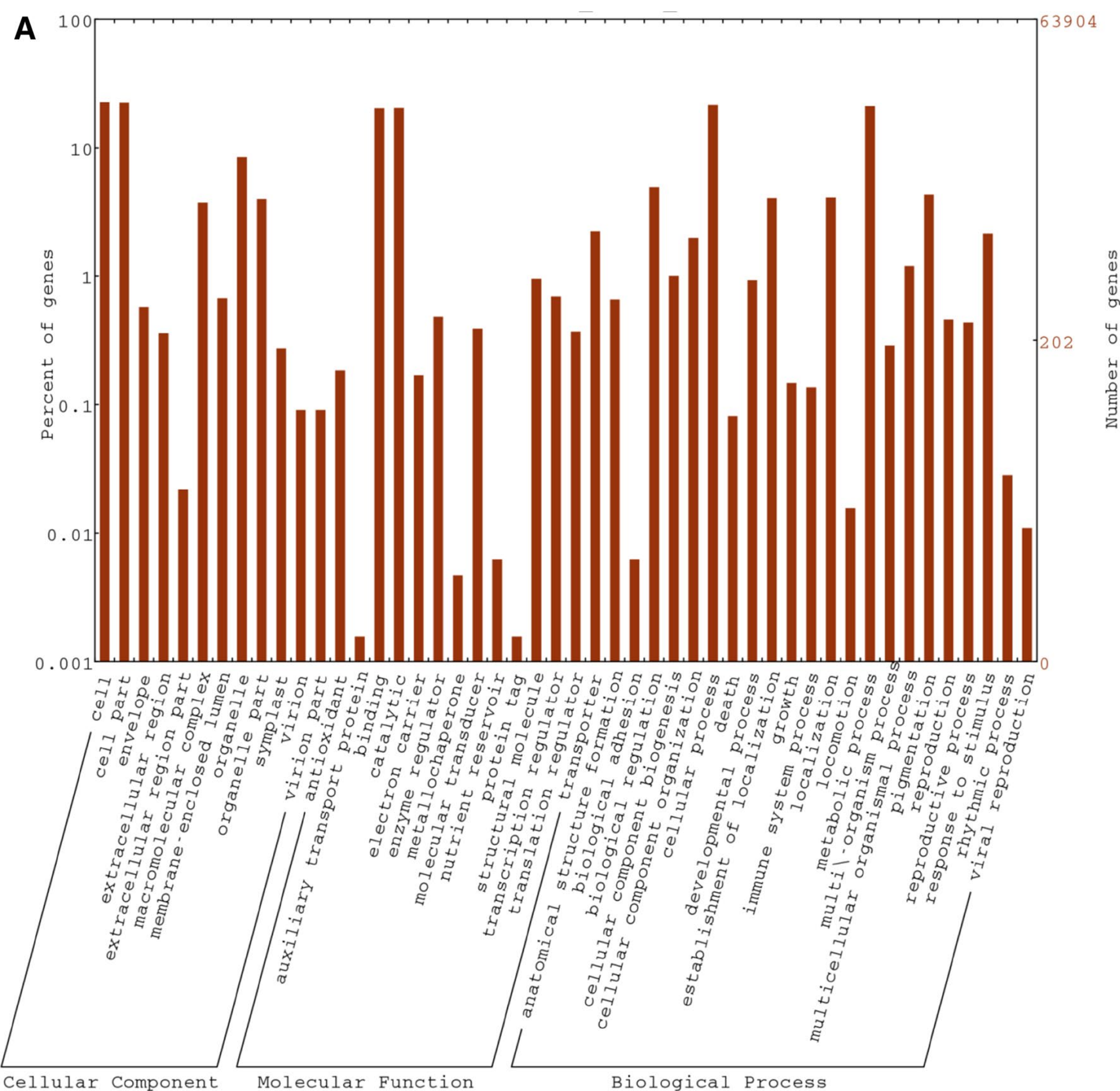

Fig. 3 **a** Gene ontology: cellular component, molecular function and biological processes for **a** sea buckthorn female. The figure shows a set of genes which are classified on the basis of (1) cellular component describing locations at the levels of subcellular structures and macromolecular complexes, (2) molecular function describing the functions of gene products and the abilities they possess, (3) biological processes give the insight about the collection of molecular events with a defined beginning and end. **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 (1) cellular component describing locations at the levels of subcellular structures and macromolecular complexes, (2) molecular function describing the functions of gene products and the abilities they possess, and (3) biological processes gives the insight about the collection of molecular events with a defined beginning and end

bud sample. Red block on the right side of zero represents the upregulated genes and green block on the left side of zero represents significantly downregulated genes. *Y*-axis represents the negative log of *p* value (value ≤ 0.05) of the performed statistical analysis, where data points with a low *p* value (highly significant) were appearing at the top of the plot. Grey block shows the non-differentially expressed genes (Supplementary Fig. 1B).

After performing scatter and volcano plots, it was identified that a total of 16,831 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 (Fig. 4).

Differentially expressed gene association with pathways

A number of genes showing differential expression across male and female transcriptome data sets were found to be associated with various primary metabolism pathways (Supplementary Table 1A–B). To investigate the metabolic

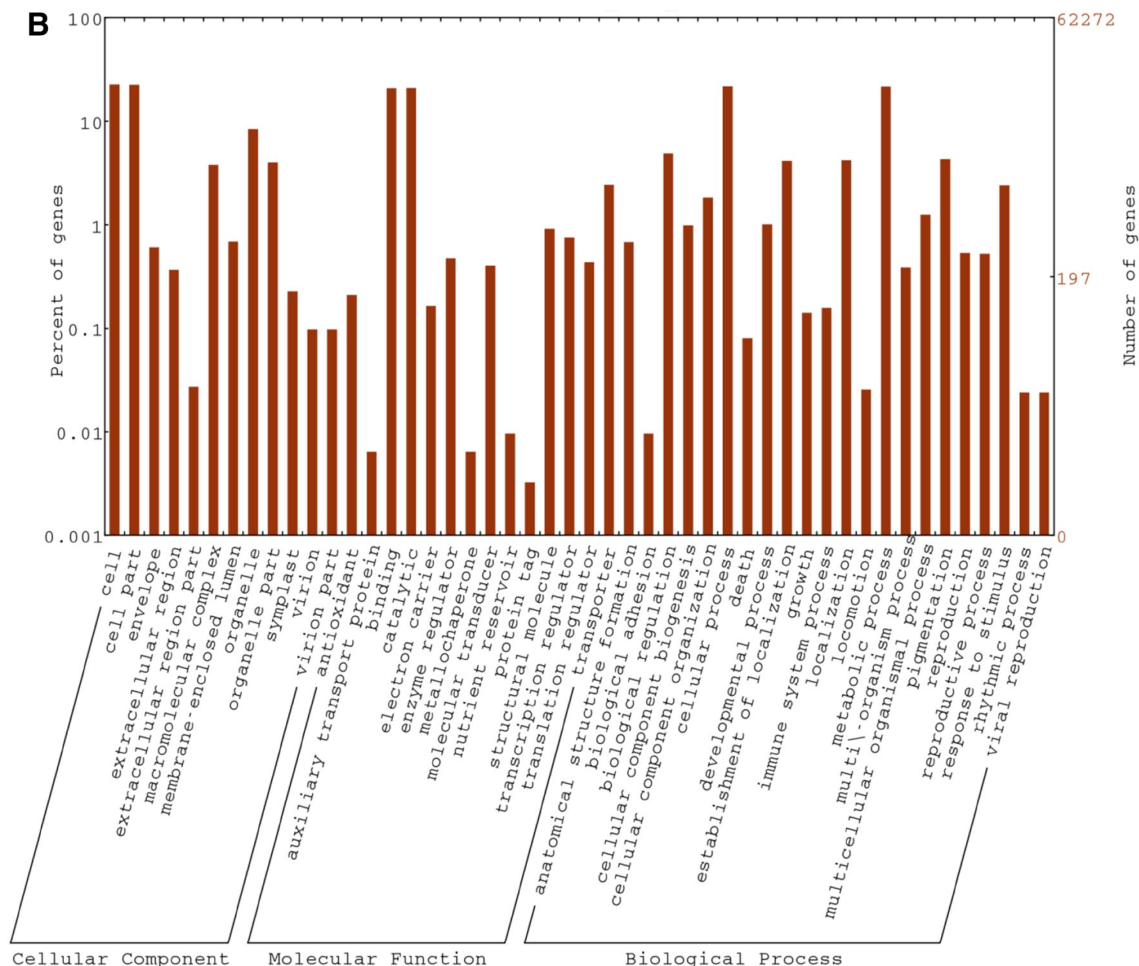

Fig. 3 (continued)

connections among various differentially expressed pathways, 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 were examined. On the basis of expression (FPKM values), top contributing upregulated and downregulated genes were considered from male and female transcriptomes (Fig. 5).

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

Upregulated and downregulated genes in primary metabolism pathways provide a global view of the regulatory interactions, but the current study is focused on identification of the genes involved in sex determination. Henceforth, comparison of male and female transcriptomes was performed against FLOR-ID (Supplementary Table 1A–B). Surprisingly, only eight flowering genes provided successful outcomes which were further considered for co-expression analysis (Supplementary Table 2, Fig. 6a, b).

PKL is a chromatin-remodelling factor and is known to inhibit the embryonic genes expression and hence plays the key role in the post-germination growth (Ogas et al. 1999). Its function is also seen during carpel differentiation (Eshed et al. 1999). As per analysis, *PKL* along with a bunch of other genes (L484_024032, L484_008274, L484_008224, L484_006547, and TCM_037681) showing similar expressions contributes together in the gibberellin-signal transduction pathway. Gibberellin-signal mechanism is well connected to the root development by the *PKL* gene. This gene mutates and shuts 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 (Taylor 1998).

Flowering plants need a photoreceptor protein to sense the length of days/nights and other seasonal changes, so as to signalize flowering (Hamner and Bonner 1938). *FYPP* gene was seen to control photoperiodism in the plant by

Table 3 KEGG database categorizes pathway classification of Sea buckthorn male and Sea buckthorn female sample CDS

Category	Pathways	Sea buckthorn male	Sea buckthorn 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 biosynthesis and metabolism	255	234
	Metabolism of cofactors and vitamins	510	521
	Metabolism of terpenoids and polyketides	306	261
	Biosynthesis 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
	Signaling 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

dephosphorylating the phytochromes and hence modulating their signals to control flowering time (Kim et al. 2002). *FYPP* and its co-expressed genes (RCOM_1612610, PRUPE_ppa010815 mg, PRUPE_ppa015809 mg, and 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 (Ni 2005).

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 (Szemenyei et al. 2008). This process takes place in the restraining temperature conditions. *TPL* along with a couple of a few more genes (CICLE_v10021651 mg, L484_026254, L484_013182, and PRUPE_ppa000037 mg) incurs mutations which leads to the proper functioning of the auxin-signalling pathway (Osmont and Hardtke 2008). 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 the modulation of hormones like AUX during the plant developmental stages (Szemenyei et al. 2008).

Mutations in the SFR gene help the plant in dealing with the freezing temperature (Knight et al. 1999). The *SFR* gene along with a few more genes (PRUPE_ppa010657 mg, L484_025827, B456_004G2909001, VIT_04s0008g05690, and PRUPE_ppa000230 mg) coordinates 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 (Knight et al. 2009).

For any plant to reproduce effectively, it is very important that the change from transition state into flowering state undergoes efficiently. Failure in the 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 (Zhao et al. 2015). Thus, various gene members of the *LDL* group (PRUPE_ppa003322 mg, L484_001327, L484_008195, L484_021828, and PRUPE_ppa006923 mg) were observed in the silencing mechanisms (Simpson 2004).

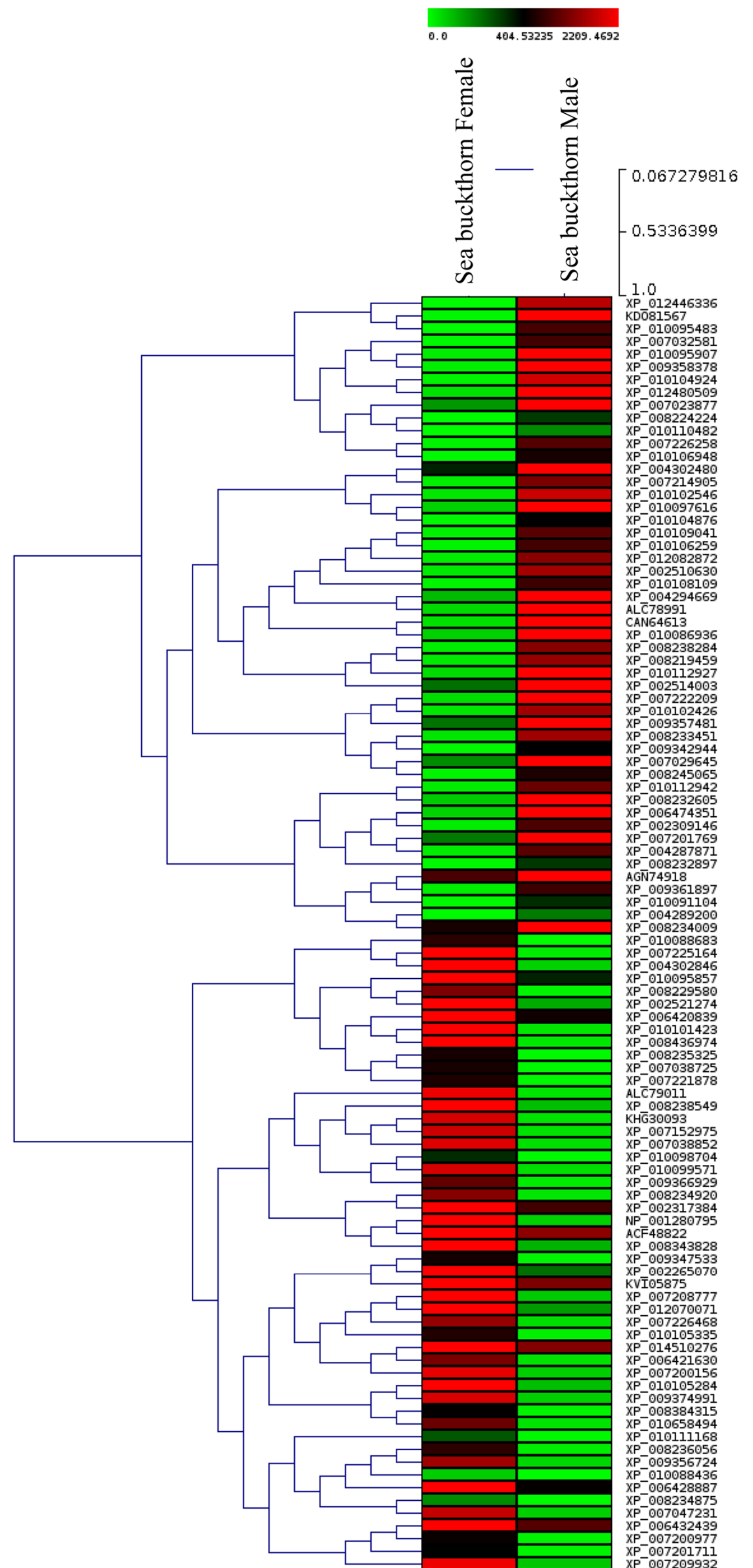

Fig. 4 Hierarchal clustering of Sea buckthorn Male and Female transcriptome. This explains the gene expression data with proper upregulation and down-regulation patterns labeled with Accession IDs

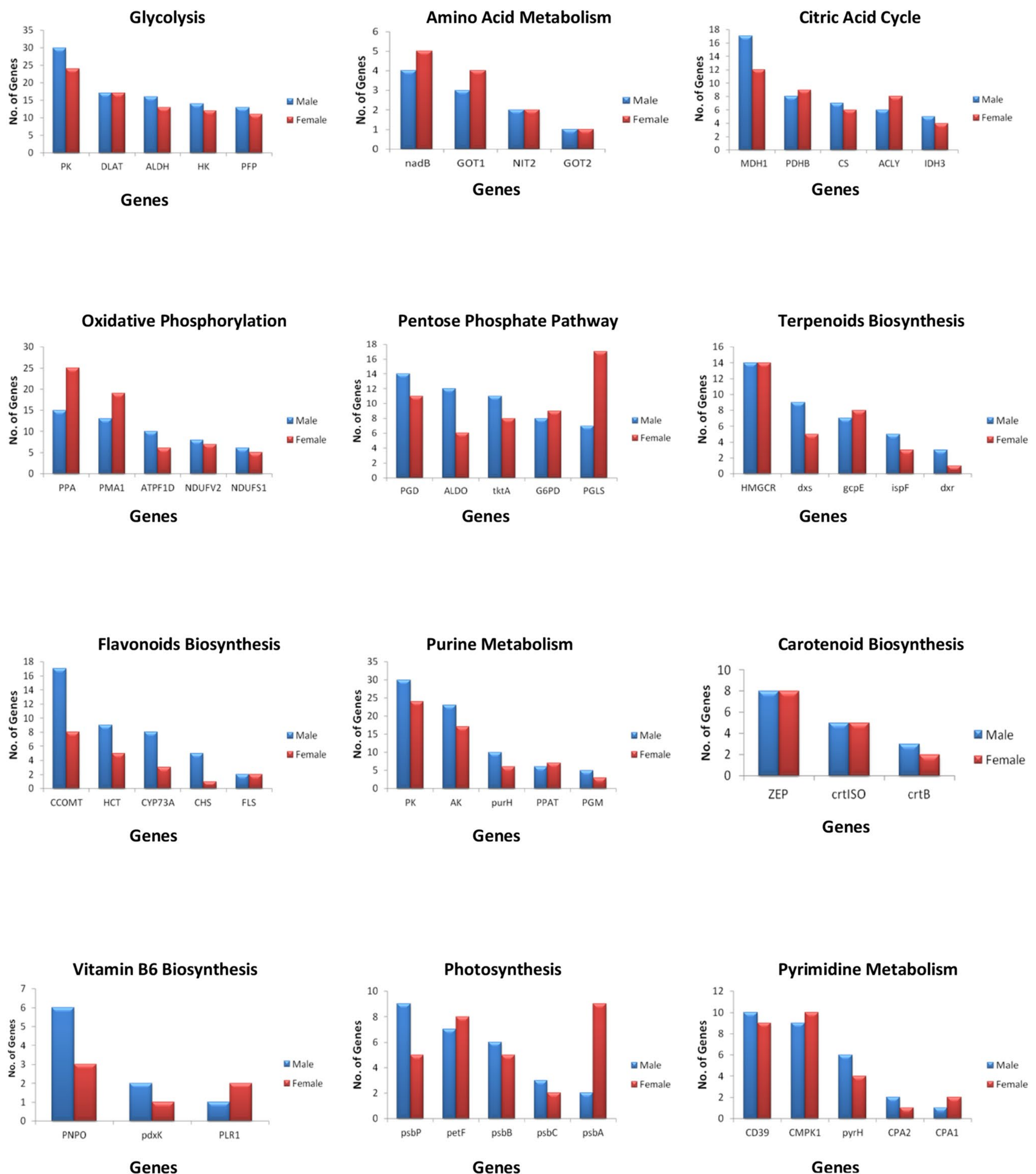

Fig. 5 Pathway analysis of a few selected pathways was conducted and analyzed to form the mentioned results. The figure shows a set of graphs for various pathways showing the number and the name of different genes involved in particular pathways

Mutation in the *PRP8A* gene can lead to a defect in the cell-division of embryonic suspensor (Schwartz et al. 1994; Koncz et al. 2012). A few more genes (hbn1, RCOM_1038560, L484_026453, CICLE_v10004979 mg) 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

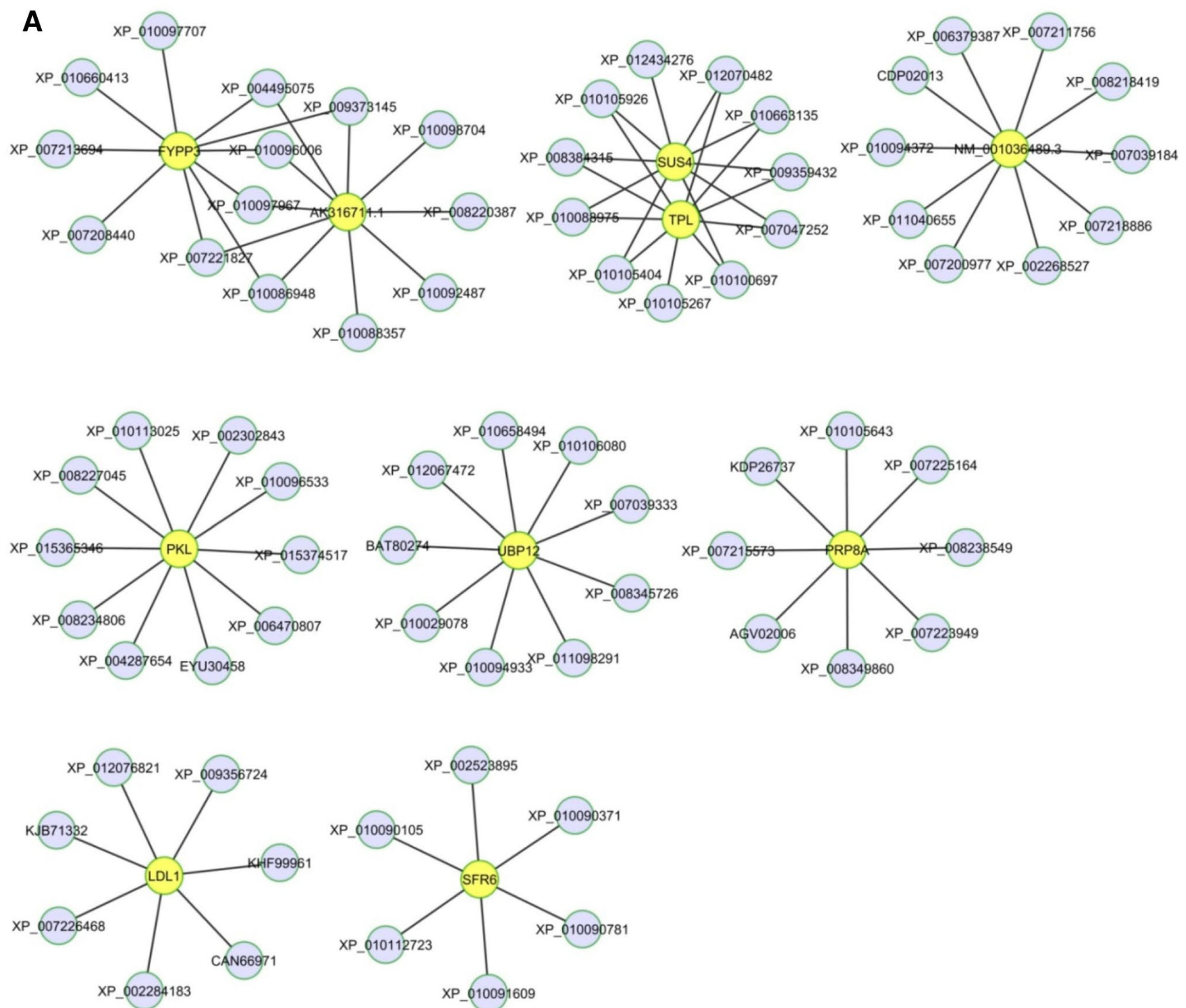

Fig. 6 **a** Co-expression networks of eight genes which are found to be involved in flowering pathways in Sea buckthorn female transcriptome data set. It is an undirected graph, where each gene is represented by a node and a set of co-expressed genes are connected to the node with an edge. **b** Co-expression networks of eight genes which are found to be involved in flowering pathways in Sea buckthorn male transcriptome data set. It is an undirected graph, where each gene is represented by a node and a set of co-expressed genes are connected to the node with an edge

degradation of mRNAs, NMD pathway also works to maintain the expression of some wild-type genes (Benkovics et al. 2011).

SUS4 gene functions in the cleavage of sucrose for the production of energy in the form of glucose and fructose for different metabolism pathways (Salanoubat et al. 2000). Sucrose is the important substrate which is required for the initial steps (especially in the non-photosynthetic tissues) of the carbohydrate biosynthesis (Fu et al. 1995). *SUS4* gene along with some other genes (TCM_037681, L484_006547, L484_008224, and PRUPE_ppa001250 mg) helps sucrose in the formation of chromoplasts from chloroplasts for the accumulation of carotenoids (Zhou et al. 2011). 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 (Baranski and Cazzonelli 2016).

Ubiquitin carboxyl-terminal hydrolase enhances the fusion of mitochondria and even modulates the photoperiodic flowering pathway (Cui et al. 2013). *UBP12* along with many other genes (L484_023231, CISIN_1g027463 mg, POPTR_0018s03430 g, L484_014252, and PRUPE_ppa001185 mg) work to coordinate in the degradation of ubiquitin-regulated proteins leading to the stabilized functioning of the photoperiodic

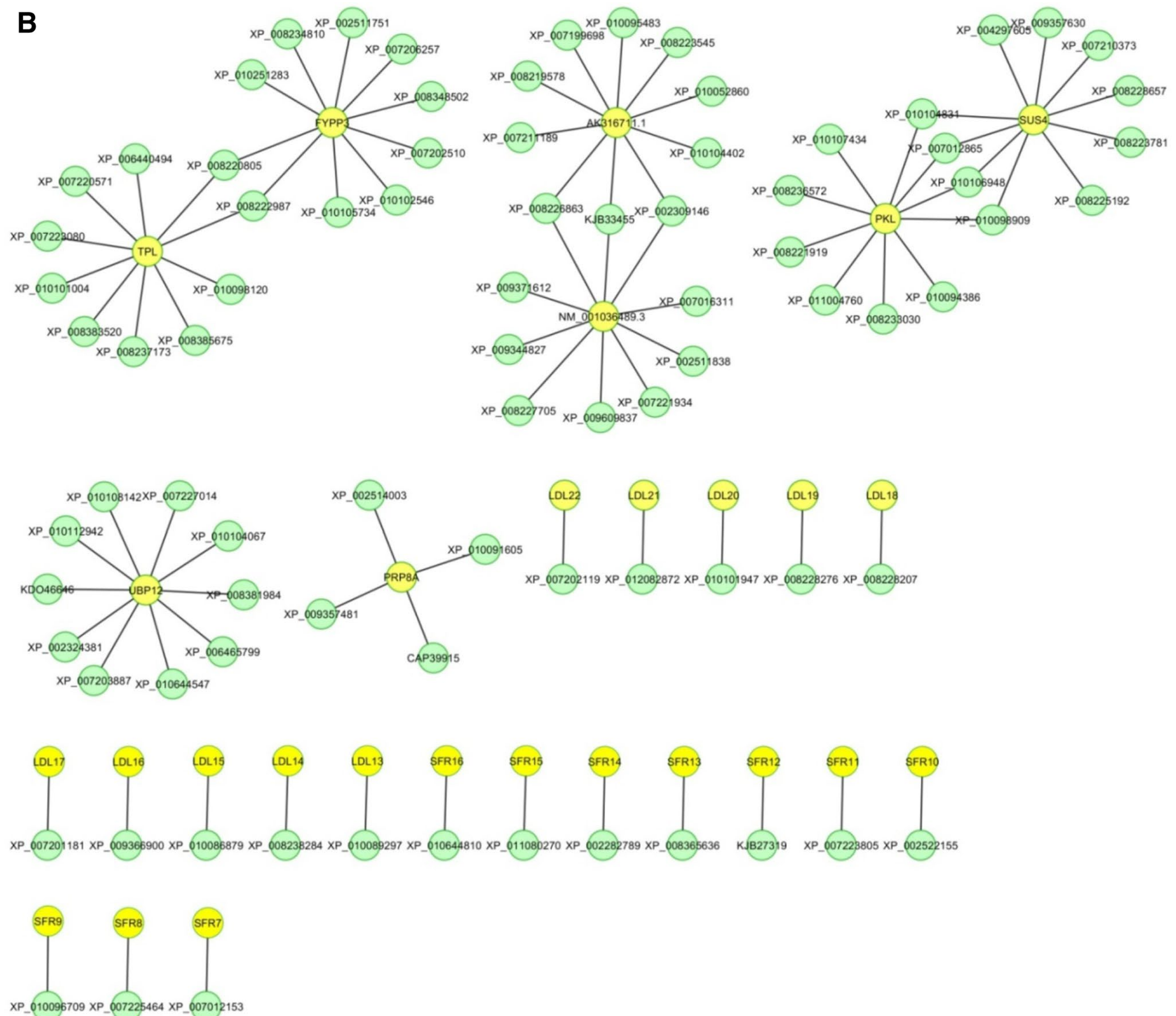

Fig. 6 (continued)

flowering pathway. Mutations in *UBP* genes often result in early flowering and even in overexpression of FLOWERING LOCUS T. LATE ELONGATED HYPOCOTYL, CIRCADIAN CLOCK ASSOCIATED1 and TIMING OF CAB EXPRESSION1 (Cui et al. 2013).

The 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.

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 shows the diverse role of differentially expressed genes in various other studies

which are related to yield content. The current study will prove an add-on to the previous transcriptomic studies and can aid a helping hand in further molecular marker selection studies for sex-based and other genotyping based differential analysis on Sea buckthorn as no genome source is available till date.

Data availability Transcriptome data for female and male flower buds are freely accessible at http://www.bioinfoindia.org/sbt.

Acknowledgements Authors thank the Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology for providing facilities to carry out experiments. All authors sincerely thank Defence Research and Development Organisation (DRDO), Ministry of Defence, Govt. of India for funding the project on Seabuckthorn in the form of financial support to AK.

Author contributions AK and TSS conceived the ideas; AK and AB designed the framework. AB, TAS, and MS carried out experiment and analyzed the transcriptome data. All authors read and improvised the manuscript.

Compliance with ethical standards

Conflict of interest Authors declare no conflict of interest.

References

Adam H, Collin M, Richaud F et al (2011) Environmental regulation of sex determination in oil palm: current knowledge and insights from other species. Ann Bot 108:1529–1537. https://doi.org/10.1093/aob/mcr151

Adhikari S, Bandyopadhyay TK, Ghosh P (2012) Hormonal control of sex expression of cucumber (*Cucumis sativus* L.) with the identification of sex linked molecular marker. Nucleus 55:115–122. https://doi.org/10.1007/s13237-012-0061-5

Ageez A, Kazama Y, Sugiyama R, Kawano S (2005) Male-fertility genes expressed in male flower buds of *Silene latifolia* include homologs of anther-specific genes. Genes Genet Syst 80:403–413. https://doi.org/10.1266/ggs.80.403

Akkermans ADL, Roelofsen W, Blom J et al (1983) 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 61:2793–2800. https://doi.org/10.1139/b83-309

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106

Bansal A, Srivastava PA (2018) Transcriptomics to metabolomics: a network perspective for big data. IGI Glob. https://doi.org/10.4018/978-1-5225-2607-0.ch008

Bansal A, Singh TR, Chauhan RS (2017) A novel miRNA analysis framework to analyze differential biological networks. Sci Rep 7:14604. https://doi.org/10.1038/s41598-017-14973-x

Baranski R, Cazzonelli CI (2016) Carotenoid biosynthesis and regulation in plants. In: Kaczor A, Baranska M (eds) Carotenoids. John Wiley & Sons Ltd, Hoboken, pp 159–189. https://doi.org/10.1002/9781118622223.ch10

Benkovics AH, Nyikó T, Mérai Z et al (2011) Functional analysis of the grapevine paralogs of the SMG7 NMD factor using a heterolog VIGS-based gene depletion–complementation system. Plant Mol Biol 75:277–290. https://doi.org/10.1007/s11103-010-9726-0

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Charlesworth D (2002) Plant sex determination and sex chromosomes. Heredity 88:94–101. https://doi.org/10.1038/sj.hdy.6800016

Chawla A, Kant A, Stobdan T et al (2014) Cross-species application of sex linked markers in *H. salicifolia* and *H. tibetana*. Sci Hortic 170:281–283. https://doi.org/10.1016/j.scienta.2014.03.013

Chawla A, Stobdan T, Srivastava RB, Jaiswal V, Chauhan RS, Kant A, Albertini E (2015) Sex-Biased Temporal Gene Expression in Male and Female Floral Buds of Seabuckthorn (Hippophae rhamnoides). PLOS ONE 10(4):e0124890

Conesa A, Götz S (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:619832. https://doi.org/10.1155/2008/619832

Cui X, Lu F, Li Y et al (2013) Ubiquitin-specific proteases UBP12 and UBP13 act in circadian clock and photoperiodic flowering regulation in *Arabidopsis*. Plant Physiol 162:897–906. https://doi.org/10.1104/pp.112.213009

Delph LF, Arntz AM, Scotti-Saintagne C, Scotti I (2010) The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*. Evol Int J Org Evol 64:2873–2886. https://doi.org/10.1111/j.1558-5646.2010.01048.x

Di Stilio VS, Kramer EM, Baum DA (2005) Floral MADS box genes and homeotic gender dimorphism in *Thalictrum dioicum* (Ranunculaceae)—a new model for the study of dioecy. Plant J 41:755–766. https://doi.org/10.1111/j.1365-313X.2005.02336.x

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. Evol Appl 7:1026–1042. https://doi.org/10.1111/eva.12178

Eshed Y, Baum SF, Bowman JL (1999) Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. Cell 99:199–209. https://doi.org/10.1016/S0092-8674(00)81651-7

Fatima T, Snyder CL, Schroeder WR et al (2012) Fatty acid composition of developing sea buckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed. PLoS ONE 7:e34099. https://doi.org/10.1371/journal.pone.0034099

Foucart C, Boualem A, Lasseur B et al (2012) Sex determination in cucurbits. Biol Aujourdhui 206:57–62. https://doi.org/10.1051/jbio/2012005

Fu H, Kim SY, Park WD (1995) 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 7:1395–1403

Ghangal R, Chaudhary S, Jain M et al (2013) Optimization of de novo short read assembly of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. PLoS ONE 8:e72516. https://doi.org/10.1371/journal.pone.0072516

Gupta SM, Grover A, Pandey P, Ahmed Z (2012) 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 18:377–380. https://doi.org/10.1007/s12298-012-0133-7

Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084

Hamner KC, Bonner J (1938) Photoperiodism in relation to hormones as factors in floral initiation and development. Bot Gaz 100:388–431. https://doi.org/10.1086/334793

Howe EA, Sinha R, Schlauch D, Quackenbush J (2011) RNA-Seq analysis in MeV. Bioinformatics (Oxf Engl) 27:3209–3210. https://doi.org/10.1093/bioinformatics/btr490

Huang S, Zeng H, Zhang J et al (2011) Interconversions of different forms of vitamin B6 in tobacco plants. Phytochemistry 72:2124–2129. https://doi.org/10.1016/j.phytochem.2011.07.019

Jain A, Ghangal R, Grover A et al (2010) Development of EST-based new SSR markers in seabuckthorn. Physiol Mol Biol Plants Int J Funct Plant Biol 16:375–378. https://doi.org/10.1007/s12298-010-0037-3

Jeong JH, Lee JW, Kim KS et al (2010) Antioxidant and antimicrobial activities of extracts from a medicinal plant, sea buckthorn. J Korean Soc Appl Biol Chem 53:33–38. https://doi.org/10.3839/jksabc.2010.006

Johansson S, Lindstedt S, Tiselius HG (1974) Metabolic interconversions of different forms of vitamin B6. J Biol Chem 249:6040–6046

Kalia RK, Singh R, Rai MK et al (2011) Biotechnological interventions in sea buckthorn (*Hippophae* L.): current status and future prospects. Trees 25:559–575. https://doi.org/10.1007/s00468-011-0543-0

Kim D-H, Kang J-G, Yang S-S et al (2002) A phytochrome-associated protein phosphatase 2A modulates light signals in flowering time control in *Arabidopsis*. Plant Cell 14:3043–3056. https://doi.org/10.1105/tpc.005306

Knight H, Veale EL, Warren GJ, Knight MR (1999) The sfr6 mutation in Arabidopsis suppresses low-temperature induction of genes dependent on the CRT/DRE sequence motif. Plant Cell 11:875–886. https://doi.org/10.1105/tpc.11.5.875

Knight H, Mugford SG, Ulker B et al (2009) Identification of SFR6, a key component in cold acclimation acting post-translationally on CBF function. Plant J Cell Mol Biol 58:97–108. https://doi.org/10.1111/j.1365-313X.2008.03763.x

Koncz C, deJong F, Villacorta N et al (2012) The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. Front Plant Sci. https://doi.org/10.3389/fpls.2012.00009

Korekar G, Sharma RK, Kumar R et al (2012) Identification and validation of sex-linked SCAR markers in dioecious *Hippophae rhamnoides* L. (Elaeagnaceae). Biotechnol Lett 34:973–978. https://doi.org/10.1007/s10529-012-0852-4

Kumar V, Bansal A, Chauhan RS (2017) Modular design of picroside-II biosynthesis deciphered through NGS transcriptomes and metabolic intermediates analysis in naturally variant chemotypes of a medicinal herb, *Picrorhiza kurroa*. Front Plant Sci. https://doi.org/10.3389/fpls.2017.00564

Lebel-Hardenack S, Hauser E, Law TF et al (2002) Mapping of sex determination loci on the white campion (Silene latifolia) Y chromosome using amplified fragment length polymorphism. Genetics 160:717–725

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics (Oxf Engl) 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liu J, Yin T, Ye N et al (2013) Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). PLoS ONE 8:e60181. https://doi.org/10.1371/journal.pone.0060181

Matsunaga S, Kawano S, Takano H et al (1996) Isolation and developmental expression of male reproductive organ-specific genes in a dioecious campion, *Melandrium album* (Silene latifolia). Plant J Cell Mol Biol 10:679–689. https://doi.org/10.1046/j.1365-313X.1996.10040679.x

Ni M (2005) Integration of light signaling with photoperiodic flowering and circadian rhythm. Cell Res 15:559–566. https://doi.org/10.1038/sj.cr.7290325

Ogas J, Kaufmann S, Henderson J, Somerville C (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. Proc Natl Acad Sci USA 96:13839–13844. https://doi.org/10.1073/pnas.96.24.13839

Osmont KS, Hardtke CS (2008) The topless plant developmental phenotype explained! Genome Biol 9:219. https://doi.org/10.1186/gb-2008-9-4-219

Rongsen L (1992) Seabuckthorn: a multipurpose plant species for fragile mountains. Centre for Integrated Mountain Development. http://agris.fao.org/agris-search/search.do?recordID=XF2015037660

Salanoubat M, Lemcke K, Rieger M et al (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. Nature 408:820–822. https://doi.org/10.1038/35048706

Schwartz BW, Yeung EC, Meinke DW (1994) Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of *Arabidopsis*. Development 120:3235–3245

Semerikov V, Lagercrantz U, Tsarouhas V et al (2003) Genetic mapping of sex-linked markers in *Salix viminalis* L. Heredity 91:293–299. https://doi.org/10.1038/sj.hdy.6800327

Sharma A, Zinta G, Rana S, Shirko P (2010) Molecular identification of sex in *Hippophae rhamnoides* L. using isozyme and RAPD markers. For Stud China 12:62–66. https://doi.org/10.1007/s11632-010-0012-7

Simpson GG (2004) The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. Curr Opin Plant Biol 7:570–574. https://doi.org/10.1016/j.pbi.2004.07.002

Singh V (2008) Seabuckthorn (Hippophae L.): a multipurpose wonder plant. Daya Books, Delhi

Szemenyei H, Hannon M, Long JA (2008) TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. Science 319:1384–1386. https://doi.org/10.1126/science.1151461

Tanabe M, Kanehisa M (2012) Using the KEGG database resource. In: Curr Protoc Bioinforma. https://doi.org/10.1002/0471250953.bi0112s38 **(chapter 1, unit 1.12)**

Taylor CB (1998) GA Signaling: genes and GTPases. Plant Cell 10:131–134. https://doi.org/10.1105/tpc.10.2.131

Uragami A, Ueno R, Yamasaki A et al (2016) Productive differences between male and female plants in white asparagus production using the rootstock-planting forcing culture technique. Hortic J 85:322–330. https://doi.org/10.2503/hortj.MI-115

Wu T, Qin Z, Zhou X et al (2010) Transcriptome profile analysis of floral sex determination in cucumber. J Plant Physiol 167:905–913. https://doi.org/10.1016/j.jplph.2010.02.004

Zhao M, Yang S, Liu X, Wu K (2015) Arabidopsis histone demethylases LDL1 and LDL2 control primary seed dormancy by regulating DELAY OF GERMINATION 1 and ABA signaling-related genes. Front Plant Sci. https://doi.org/10.3389/fpls.2015.00159

Zhou X, Mcquinn R, Fei Z et al (2011) Regulatory control of high levels of carotenoid accumulation in potato tubers. Plant Cell Environ 34:1020–1030. https://doi.org/10.1111/j.1365-3040.2011.02301.x