Transcriptome Sequencing and Analysis of Seabuckthorn (*Hippophae* Sp.)

11

Saurabh Chaudhary
and Prakash C. Sharma

Abstract

Seabuckthorn (*Hippophae* Sp.) is known for its immense medicinal, nutritional, and agronomical properties and has been utilized for centuries as traditional medicine and food supplements in various countries of Asia and Europe. The revolutionary development in the field of sequencing from first to the third generation opens a whole new horizon to explore the transcriptome of non-crop but important plant species. Many next-generation sequencing (NGS)-based transcriptome studies exploring different prospects such as therapeutic uses, stress management, and sex determination have been conducted for seabuckthorn in the last decade. Currently, a complete transcriptome profile from leaf and root tissues along with other transcriptomes profiled under different environmental conditions is available in the public domain for seabuckthorn. In the current chapter, we have highlighted the recent developments and techniques used for NGS-based transcriptome profiling for seabuckthorn. We have also discussed the challenges and prospects of NGS-based transcriptome studies in seabuckthorn. The transcriptome profiles generated for seabuckthorn could further be exploited to isolate molecular markers, microRNAs, other small and long non-coding RNAs, and gene families specific to this plant. Finally, a strategy for the utilization of third (next to next) generation sequencing to explore the seabuckthorn transcriptome is also discussed. The current chapter could serve as a good reference for transcriptome studies in seabuckthorn and other related plant species.

Keywords

Seabuckthorn · Transcriptome · Next-generation sequencing · Third-generation sequencing · Molecular markers · MicroRNA

11.1 Introduction

Seabuckthorn (*Hippophae* Sp.), a member of the family Elaeagnaceae, is a multipurpose plant with significant medicinal, nutritional, and ecological importance. Seabuckthorn is a hardy, deciduous shrub, widely distributed in various regions of Europe, Asia, and North America covering a vast area of India Himalayas, China, and Russia (Bartish et al. 2002; Yang et al.

S. Chaudhary (✉)
Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK
e-mail: ChaudharyS6@cardiff.ac.uk; s.chaudhary0525@gmail.com

P. C. Sharma
University School of Biotechnology, Guru Gobind Singh Indraprastha University, Dwarka 16C, New Delhi 110078, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

P. C. Sharma (ed.), *The Seabuckthorn Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-031-11276-8_11

2005). Since ancient times, the various parts of seabuckthorn including berries, leaves, and young stems have been utilized in traditional Indian, Chinese, and Tibetan medicines, to treat cough, cold, influenza, bowel irregularities, gastric ulcer, skin infections, and wounds (Sharma and Kalkal 2018). In addition to medicinal and nutritional values, seabuckthorn holds up many important ecological properties. The extensive root system developed by seabuckthorn helps in preventing soil erosion and maintaining soil fertility and land reclamation. Moreover, the presence of root nodules in seabuckthorn helps in fixing atmospheric nitrogen using symbiotic association with an actinomycete, Frankia (Stewart and Pearson 1967; Wei et al. 2007). Seabuckthorn also can survive in adverse climatic conditions such as drought, salinity stress, and extreme temperature ranging from −40 °C to 40 °C (Chaudhary and Sharma 2015; Sharma and Deswal 2020). Overall, the immense medicinal, nutritional, and ecological importance attracts the attention of many researchers worldwide toward seabuckthorn and makes it suitable as a reference plant for developing high-yield, nutritious, and stress-tolerant crop varieties. However, to be a reference, complete knowledge of the seabuckthorn genome is essential.

Over the last decade, whole-genome transcriptome profiling using next-generation sequencing (NGS) techniques has become the most favorable tool to understand the several transcriptional dynamics in many plant species, in different conditions and/or developmental stages. The emergence of NGS techniques allows the production of snapshots of the genome using whole transcriptome sequencing of many non-crop but important plant species. Advancement in the sequencing technique from the era of first-generation (Sanger Sequencing) to the current third-generation sequencing (Illumina, Pacific Biosciences, and Oxford Nanopore Technology) makes the goal of seabuckthorn genome sequencing achievable soon. In the recent past, several transcriptome studies have been conducted for seabuckthorn in tissue and condition-dependent manner, to elucidate the major pathways and genes behind making it a multi-purpose wonder plant. In the current chapter, we first give a brief explanation of the basic steps involved in transcriptome sequencing and analysis using the NGS technique and later tried to cover the transcriptome-based studies conducted in seabuckthorn till now. We also discussed how transcriptome data generated for seabuckthorn is already utilized and could be beneficial in the future as well, for the discovery of molecular markers, potential microRNA (miRNA), and their targeted genes, small and non-coding RNAs.

11.2 Transcriptome Sequencing and Bioinformatics Analysis

Whole-genome transcriptome profiling can be utilized for several applications including differential gene expression (DGE), gene discovery, pathway analysis, alternative splicing, and understanding many other transcriptional dynamics. Among various transcriptome profiling, RNA-Seq (Emrich et al. 2014; Mortazavi et al. 2008; Nagalakshmi et al. 2008) technique and Illumina platform are the most widely utilized worldwide. Different NGS platforms use different kits for sequencing library preparation. However, the basic protocol for any transcriptome sequencing and analysis remains the same and flexible to modifications according to the experimental requirements. Here, we briefly explain the generalized protocol used in the RNA-Seq experiment using Illumina sequencing platforms. The basic protocols (Fig. 11.1a) include the following steps:

1. **Extraction of total RNA and quality control**: Several conventional methods including TRIzol Reagent (Invitrogen) and commercial kits such as Qiagen RNeasy Plant Mini Kit are available for extraction of good quality RNA from plant tissues. Once the total RNA is extracted, it is further processed for quality check with Agilent 2100 Bioanalyzer system. The RNA integrity number (RIN) value above 9 is the good quality RNA for NGS.

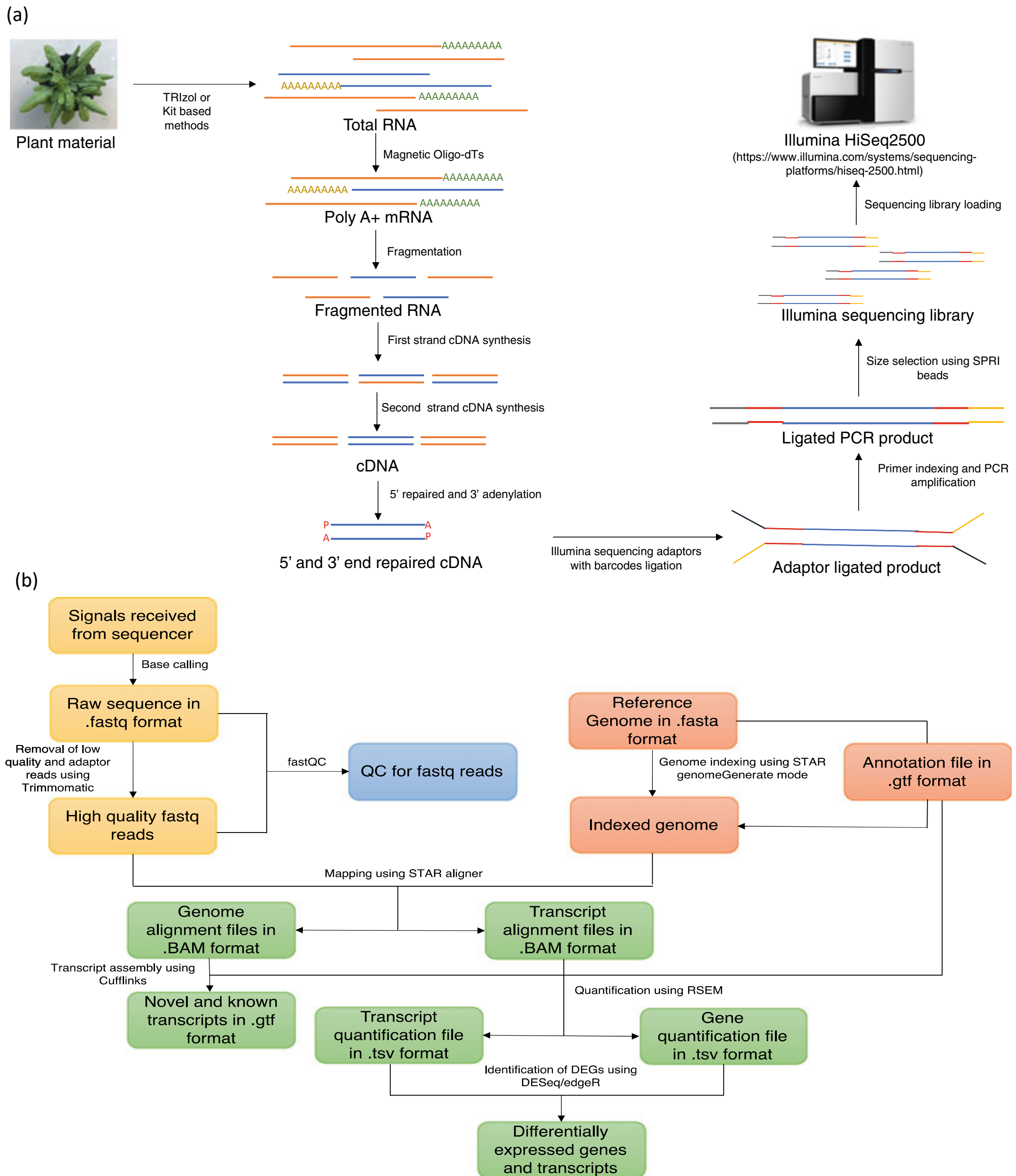


Fig. 11.1 Transcriptome sequencing and bioinformatics analysis using RNA-Seq technique and Illumina platform. **a** Steps involved in preparation of next-generation sequencing (NGS) library preparation and sequencing on Illumina HiSeq2500 platform. **b** Flowchart representing stepwise bioinformatics pipeline used to analysis the sequences received from Illumina HiSeq2500 platform

2. **Isolation of mRNA from total RNA**: To isolate the poly-A-containing mRNA, oligo (dT) beads are used.
3. **Library construction and sequencing**: Depending upon the NGS platform used in the study, each library from the isolated

mRNA can be constructed according to the commercial kit protocol. For example, for Illumina sequencing platforms, TruSeq RNA Library Prep Kit v2, from Illumina San Diego, CA, is available. The major steps in the sequencing library preparations include (i) mRNA fragmentation, (ii) first- and second-strand cDNA synthesis, (iii) adaptor ligation, (iv) polymerase chain reaction (PCR) amplification, and (v) size selection and quality control (Fig. 11.1a). Once the sequencing library of the desired size is prepared, it is ready for sequencing.

During sequencing, the NGS platforms generate signals which simultaneously converted to nucleotides with quality control of each nucleotide in fastaq (fastq) format using in-built base calling programs. High-throughput sequencing platforms such as Illumina generate millions of short reads up to 150 bp in fastq format from each run which requires a bioinformatics pipeline and various computational tools to assign those reads a biological meaning. Like sequencing platforms, one can choose a bioinformatics pipeline and tool from many for their experiment needs. For the plant species with no reference, genome availability requires de novo assembly of short reads before downstream analysis of identification of DEGs and annotation. An example of such a case is seabuckthorn which is discussed in detail later in this chapter. Here, we present the generalized bioinformatics pipeline for RNA-Seq data analysis (Fig. 11.1b) that could be applied for plant species whose reference genome is available and well-annotated. The general steps in RNA-Seq bioinformatics analysis (Fig. 11.1b) include the following steps:

1. **Quality control of sequencing reads**: The raw reads received after sequencing in fastq format contain low-quality and adaptor read contamination. To access the quality of the short reads before and after quality control, FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) can be used. After the quality check, Trimmomatic (Bolger et al. 2014) a flexible read trimming tool is applied to remove the low-quality, too short and too long, and adaptor reads from the raw fastq sequences. The high-quality (HQ) fastq reads generated after trimming could also be accessed for quality check with FastQC, again.
2. **Mapping of HQ short reads to reference genome**: The HQ reads are then mapped using STAR aligner (Dobin et al. 2013) to the index reference genome to generate transcriptome and genome alignment files in BAM format. An index genome reference could be generated using the '-generateGenome' parameter of STAR aligner, and passing reference genome in FASTA format, and the gene and transcript annotation file in Gene Transfer Format (gtf). Once the gene and transcript alignment to the reference genome has been completed, the gene and transcript aligned files in BAM format could be utilized in downstream analysis.
3. **Transcriptome reconstruction/assembly**: The gene alignment file in BAM format can be used to reconstruct using Cufflinks (Trapnell 2012; Pertea et al. 2015). The assembled transcripts are used for the discovery of novel and known transcripts and reannotation.
4. **Gene and transcript quantification**: Any tool that identifies the DEGs or differentially expressed transcripts requires either Transcript Per Million (TPM) or Fragments Per Kilobase of Transcript Per Million (FPKM) of aligned reads. Therefore, the next step includes the generation of quantification files to assign TPM or FPKM value to every mapped read in alignment files. This can be achieved using RNA-Seq by Expectation-Maximization (RSEM) quantification tool (Djebali et al. 2017). RSEM takes the transcript alignment file in BAM format and used the gene annotation file in gtf format as a reference to generate gene and transcript quantification files in tab-separated values (tsv) files.
5. **Identification of DEGs**: Once FPKM value is assigned to each gene and transcript, the tsv files are required to pass through DESeq (https://github.com/getopt/DESEQ) or edgeR (https://github.com/StoreyLab/edge) programs to identify the DEGs.

11.3 Complete Transcriptome Profiling of Seabuckthorn: Challenges and Breakthrough

Despite the medicinal, nutritional, and ecological importance of seabuckthorn, its genomic and transcriptomic data was scarce till the first decade of the twenty-first century. By the period, research was mainly focused on the biochemical characterization and therapeutic uses of seabuckthorn extracts. It was the second decade of the twenty-first century when the revolutionary advancement in sequencing technology and bioinformatics efficient tools developed and transcriptome data on seabuckthorn starts accumulating in the public domain. The advent of NGS makes it possible to perform complete transcriptome profiling in tissue and the condition-dependent manner in seabuckthorn (Fatima et al. 2012a; Ghangal et al. 2013; Bansal et al. 2018; Ye et al. 2018). However, despite the advancement in sequencing techniques and bioinformatics efficient tools, there were challenges in generating and analyzing seabuckthorn transcriptome data. For instance, high-quality RNA is the major requirement of any NGS platform to perform transcriptome sequencing, whereas getting a high-quality RNA from seabuckthorn tissues is a challenging task because of the presence of high content of secondary metabolites and bioactive compounds (Ghangal et al. 2009). Moreover, the unavailability of the complete and annotated genome, reference-based assembly, mapping, and quantification of transcriptome were difficult to perform with high accuracy in seabuckthorn. Nevertheless, in the last one decade tremendous efforts were made in generating and analyzing transcriptome data to explore the molecular mechanisms responsible to make a seabuckthorn multipurpose wonder plant. A list of NGS-based transcriptome studies with potential purposes is summarized in Table 11.1.

11.3.1 Transcriptome Analysis for Fatty Acid Composition in Seabuckthorn Mature Seeds

The first transcriptomic study in seabuckthorn using NGS technology was conducted to identify the genes for the fatty acid biosynthesis in

Table 11.1 Next-generation sequencing-based transcriptome analysis in seabuckthorn

Seabuckthorn species	Tissues	NGS platform	Objective of the study	References
Hippophae rhamnoides (ssp. *mongolica*)	Mature seed	454-GS-FLX	Identification of genes for fatty acid biosynthesis	Fatima et al. (2012a)
Hippophae rhamnoides	Leaf and root	Illumina HiSeq2000	De novo assembly and annotation of seabuckthorn transcriptome	Ghangal et al. (2013)
Hippophae rhamnoides	Leaf	Illumina HiSeq2000	Identification of genes for cold and freeze tolerance	Chaudhary and Sharma (2015)
Hippophae rhamnoides	Flower buds	Illumina NextSeq500	Identification of underlying sex determination mechanism	Bansal et al. (2018)
Hippophae rhamnoides subsp. *sinensis*	Leaf	Illumina HiSeq2000	Identification of genes for drought tolerance	Ye et al. (2018)
Hippophae rhamnoides	Leaf, root, and stem	Illumina HiSeq2000	Molecular response under elevated CO_2	Zhang et al. (2018)
Hippophae rhamnoides	Berry pulp	Illumina HiSeq2500	Molecular mechanism behind high level of palmitoleic acid in seabuckthorn berry pulp oil	Ding et al. (2019)

seabuckthorn mature seeds (Fatima et al. 2012a). This was the first genomic resource generated for seabuckthorn using NGS technology (Table 11.1). The transcriptome sequencing was performed in the mature seeds of the Canadian cultivar *Hippophae rhamnoides* (ssp. *mongolica*) using the 454-GS-FLX sequencing platform. The authors also performed the gas chromatography-mass spectrometry (GC-MS) profiling of fatty acids in mature berry pulp and seeds. Since seabuckthorn is known for its nutraceutical values, high content of essential fatty acids, linoleic and α-linolenic acids, and monounsaturated palmitoleic acid was reported in its seed and berry pulp oils, respectively (Yang and Kallio 2001; Tiitinen et al. 2006). To identify the genes for fatty acid biosynthesis in seabuckthorn, the total RNA was extracted from seabuckthorn seeds to make a cDNA library for 454-GS-FLX sequencing. The benefit of using 454-GS-FLX was the longer reads than generated in Illumina sequencing, which further enhances the transcriptome mapping efficiency. In total, 500,392 reads were generated in the study which assembled into 89,141 and 97,392 unigenes using Fiesta 2 and GenomeQuest assembler, respectively. Among all, approximately 49% of unigene sequences were shown homology with the Arabidopsis and other plant databases. The sequence homology search showed maximum sequence similarity between seabuckthorn and *Vitis vinifera*. Interestingly, 10,011 seabuckthorn seed transcripts were found to be involved in various metabolic pathways under the category of 'biosynthesis.' The amino acid followed by a fatty acid, lipids, secondary metabolites, and carbohydrate biosynthesis pathways was dominating in the seabuckthorn seed transcriptome. The GO functional enrichment analysis suggested that seabuckthorn seed transcriptomes are enriched in several cellular and metabolic processes. Further in metabolic processes, the sub-terms including lipid biosynthesis and fatty acid biosynthesis were highly represented. Many genes encoded for enzymes involved in lipid biosynthesis were also reported in the study.

The data generated in the study is available for the research community at NCBI short-read archive (SRA) under the accession number SRS290823.1. Collectively, the transcriptome analysis of mature seeds provides the first genomic resource for seabuckthorn using NGS technology and proved to be a valuable resource to predict many genes behind the nutraceutical values of seabuckthorn. The presence of high contents of fatty acids and related genes helps in providing the hardiness to the plant to make it tolerable to extreme environmental conditions as well. Moreover, the correlation between GC-MS and transcriptome data made the study a reference platform for further metabolic and transcriptomic studies in seabuckthorn and other important plant species.

11.3.2 De Novo Assembly of Seabuckthorn Transcriptome

The first comprehensive transcriptome study in seabuckthorn (*Hippophae rhamnoides* L.) was carried out using NGS technology from leaf and root tissues (Ghangal et al. 2013) (Table 11.1). The overall experimental design, bioinformatics pipeline, and workflow used for de novo short-read assembly of seabuckthorn transcriptome are presented in Fig. 11.2. In the study, the authors generated approximately 94 million paired-end (PE) sequencing reads from leaf and root tissues of seabuckthorn using the Illumina HiSeq2000 platform. The raw Illumina short reads generated in the study were made available for the research community and can be fetched from NCBI-SRA under the study accession number SRP011938. Additionally, an efficient de novo assembly of seabuckthorn transcriptome was also optimized using six assembly tools including Velvet (Zerbino and Birney 2008), Oases (Schulz et al. 2012), ABySS (Simpson et al. 2009), SOAPdenovo (Luo et al. 2012), CLC Genomics Workbench, and Trinity (Henschel et al. 2012).

In the study, two parallel approaches were used. In the first approach, short-read assemblers

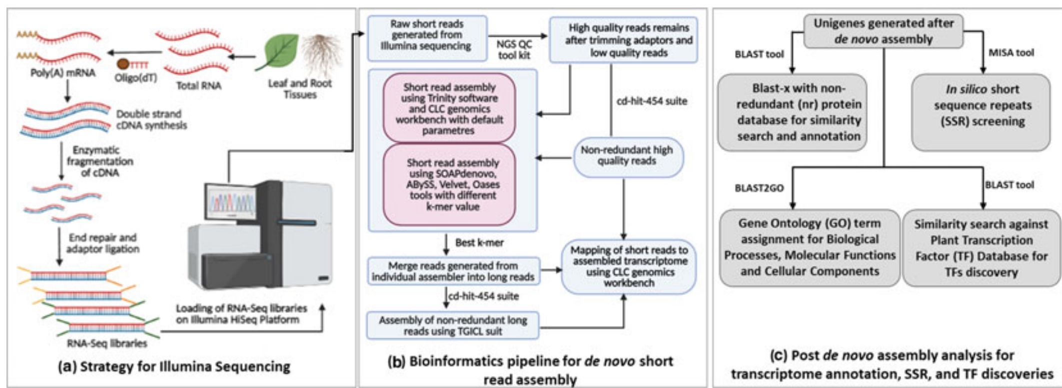


Fig. 11.2 Flowchart (created with BioRender.com) showing **a** experimental design, **b** bioinformatics pipeline, and **c** annotation analysis used for de novo assembly of seabuckthorn transcriptome. **a** Total RNA was extracted from the leaf and root tissues (Ghangal et al. 2009) followed by extraction of poly(A+) mRNA using oligo (dT) beads. A complementary strand of poly(A+) mRNA was synthesized (cDNA synthesis) using random hexamer primers. End repair, sequencing adaptor ligation, and PCR amplification were performed using the Illumina TrueSeq sequencing kit. Finally, the sequencing libraries were loaded on the Illumina HiSeq2000 platform. **b** The raw short reads generated from Illumina sequencing were quality controlled (QC) by NGS QC Toolkit (Patel and Jain 2012) to remove sequencing adaptors and low-quality reads. Redundancy from high-quality (HQ) reads was removed using a cd-hit-45 suit (Fu et al. 2012). Both non-redundant and redundant HQ reads were assembled using two approaches (i) best k-mer length and (ii) additive k-mer followed by TGICL suit (Pertea et al. 2003). Six assembly tools including Velvet (Zerbino and Birney 2008), Oases (Schulz et al. 2012), ABySS (Simpson et al. 2009), SOAPdenovo (Luo et al. 2012), CLC genomics workbench, and Trinity (Henschel et al. 2012) were used for optimizing de novo short-read assembly (Ghangal et al. 2013). **c** Unigenes generated from de novo assembly were processed for similarity search using Blast-x parameter of BLAST tool included in BLAST2GO (Conesa and Götz 2008). The gene ontology (G) term assigned to unigenes and their contribution in biological processes (BP), molecular functions (MF), and cellular components were also analyzed. MISA tool (Thiel 2003) and BLAST tool were used for simple sequence repeat (SSR) marker and transcription factor (TF) discoveries in the seabuckthorn transcriptome, respectively. Plant transcription factor database (Guo et al. 2008) was used as a reference for the identification of TFs in seabuckthorn transcriptome

were optimized based on k-mer length, whereas in the second approach, reads assembled based on best k-mer length were further re-assembled with additive k-mer strategy using long-read assembler TGICL suit (Pertea et al. 2003). Among all, ABySS with the additive k-mer approach followed by TGICL assembly was found to be the most efficient transcriptome assembly. Around 88 million high-quality (HQ) sequencing reads were assembled into 88,297 unique transcripts which represent 53 Mb of seabuckthorn transcriptome. In the final assembly, the average read length of the transcripts was 610 bp and N50 read length 1198 bp, observed. All the HQ short reads were further remapped to the assembled transcriptome, and approximately 91% of reads were uniquely mapped. The sequence homology results suggested a high similarity between *Vitis vinifera* (grapes) and the seabuckthorn transcriptome. Further, gene annotation was performed by assigning gene ontology (GO) terms to the assembled transcripts using the BLAST2GO tool (Conesa and Götz 2008), in which 38,830 (43.9%) seabuckthorn assigned with one GO term. Seabuckthorn is highly enriched in secondary metabolites and bioactive compounds, and GO analysis also revealed the significant enrichment of its transcripts in primary metabolic process, cellular metabolic process, and biosynthetic process. Further, GO analysis showed the dominance of seabuckthorn transcripts in response to biotic stimulus, response to endogenous stimulus, response to abiotic stimulus, cellular response to stimulus, and response to external stimulus, which suggests that many transcripts in seabuckthorn respond to various environmental stresses. In the end, the

Table 11.2 Statistics of short-read de novo assembly in seabuckthorn transcriptome (Ghangal et al. 2013)

Number of raw reads	94,013,936
Total number of bases in raw reads	8,461,254,240
Total number of high-quality (HQ) reads	86,253,874
Total number of bases in HQ reads	7,762,848,660
Total number of unigenes generated in accepted transcriptome assembly	88,297
Number of transcripts with significant hit with non-redundant proteins	41,340
Number of transcripts with at least one gene ontology (GO) term	38,830
Number of transcripts with short sequence repeats (SSRs)	10,980
Number of SSRs identified in the transcriptome	13,299
Number of transcription factors (TFs) identified in the transcriptome	7421
Number of TF families identified	80

seabuckthorn transcriptome was screened for the simple sequence repeat (SSR) markers (Jain et al. 2014) and transcription factors (TFs). The authors identified 7421 TFs representing 8.4% of seabuckthorn transcriptome distributed among 80 TF families. The most abundant TFs were belonging to C3H, MADS, bHLH, NAC, and FAR1 families. The complete statistics of the seabuckthorn transcriptome are provided in Table 11.2.

The de novo short-read assembly and annotation of seabuckthorn transcriptome later open a new horizon of knowledge for the research community working on molecular aspects of seabuckthorn and/or other related plant species. This is further supported by several transcriptome studies using the NGS technique in a tissue and condition-dependent manner in seabuckthorn (Table 11.1). The seabuckthorn transcriptome data generated in the study is proven to be a reference for many other transcriptomic and genomic studies in seabuckthorn and other unexplored but important plant species.

11.4 Seabuckthorn Transcriptome Analysis for Different Prospects

The initial seabuckthorn transcriptome studies using NGS technology and optimization of efficient de novo assembly open a new horizon of genomic resources to the research community to explore the plant for several prospects. Consequently, afterward, several studies were reported in a tissue and condition-dependent manner (Table 11.1) to explore the expression and identification of key genes responsible for medicinal, nutraceutical, and agronomical properties of seabuckthorn.

11.4.1 DeepSAGE-Based Transcriptome Analysis in Seabuckthorn Under Cold and Freeze Stress

In addition to medicinal and nutraceutical values, seabuckthorn can survive in varying and extreme environmental conditions including temperature ranging from −40 °C to 40 °C (Chaudhary and Sharma 2015; Sharma and Deswal 2020). Therefore, the seabuckthorn genome is anticipated to harbor several genes for abiotic stress tolerance. In 2015, Chaudhary and Sharma generated transcriptomic data in seabuckthorn using the deep Serial Analysis of Gene Expression (deepSAGE) technique. DeepSAGE is a modified version of classical SAGE (Velculescu et al. 1995) with longer tags that can be sequenced using NGS technology. The advancement in sequencing technologies provides NGS-based gene expression profiling techniques a boost to present the global picture of thousands of

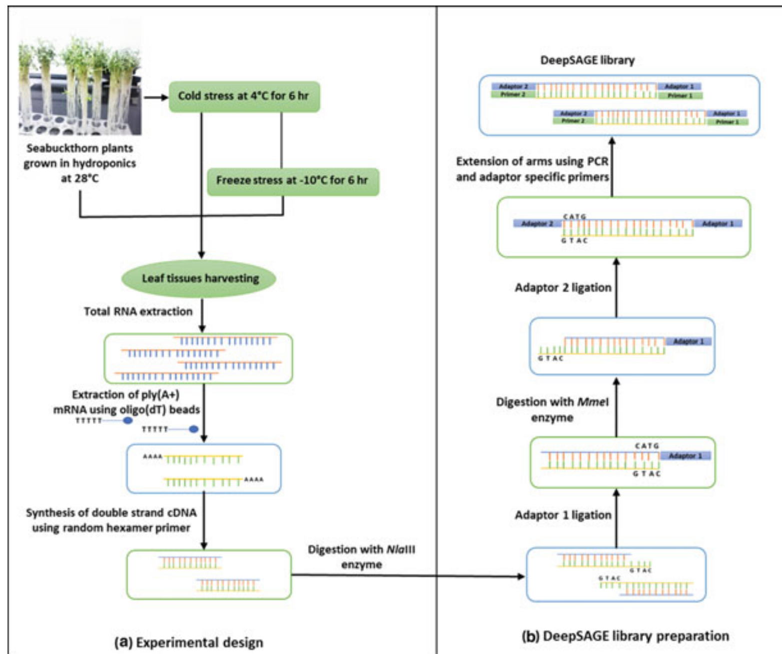


Fig. 11.3 Flowchart representing experimental design and deep Serial Analysis of Gene Expression (deepSAGE) library preparation used for identification of genes for cold and freeze tolerance in seabuckthorn (Chaudhary and Sharma 2015). **a** The experimental design showing seabuckthorn plants growing in hydroponics at laboratory conditions. At the six- to eight-leaf stage, the plant leaves were harvested from normal temperature (28 °C) as control, and plants were treated with cold stress (4 °C) for 6 h. For freeze stress, plant leaves were harvested after shifting plants from cold to freezing temperature (−10 °C) for 6 h. Harvested leaf tissues from all three conditions were processed for total RNA isolation using the modified CTAB method (Ghangal et al. 2009), followed by extraction of poly(A+) mRNA using oligo(dT) beads. The cDNA synthesis was performed by using poly(A+) mRNA as template strands and random hexamer primers. **b** Showing preparation of deepSAGE library. The double-stranded cDNA was further digested with restriction enzyme *Nla*III having CATG recognition site leaving sticky end where Illumina adaptor 1 was ligated. In the second enzymatic digestion *Mme*I, an endonuclease cut at 17 bp downstream of the CATG site, where second Illumina adaptor 2 ligated. The adaptor-containing tags were further amplified using Illumina adaptor-specific primers with polymerase chain reaction (PCR). Finally, the amplified tags (deepSAGE library) were loaded on the Illumina HiSeq2000 platform for sequencing

differentially expressed genes (DEGs) in a tissue and condition-dependent manner. In the study (Chaudhary and Sharma 2015), the authors employed the deepSAGE technique to identify genes for cold and freeze tolerance in seabuckthorn. The flowchart representing the experimental design and deepSAGE library preparation used in the study is provided in Fig. 11.3. The deepSAGE technique proved to be efficient for seabuckthorn as a large number of tags generated in the study further fetched many DEGs under cold and freeze stress (Table 11.3).

In total, approximately 36 million raw tags were generated from the three deepSAGE libraries including control (28 °C), cold (4 °C), and freeze (−1 °C) stress in the study. After filtering low-quality tags and adaptor trimming, 43 bp long 35.5 million clean tags were left. The cleans tags were further mapped to the reference genome using bowtie (Langmead et al. 2009), whereas the BLAST tool of BLAST2GO (Conesa and Götz 2008) was used to align tags with the seabuckthorn transcriptome (Ghangal et al. 2013). The differential gene expression analysis was performed based on the variations among the number of tags using the algorithm developed by Audic and Claverie (Audic and Claverie 1997). In total, 11,922 DEGs were identified in the

Table 11.3 Statistics of deepSAGE-based transcriptome analysis in seabuckthorn under cold and freeze stress (Chaudhary and Sharma 2015)

Description	Number
Total number of raw tags in control library (28 °C)	11,803,601
Total number of clean tags in control library (28 °C)	11,566,379
Total number of raw tags in cold stress library (4 °C)	12,265,029
Total number of clean tags in cold stress library (4 °C)	12,010,122
Total number of raw tags in freeze stress library (−10 °C)	12,113,951
Total number of clean tags in freeze stress library (−10 °C)	11,908,614
Tags mapped to genes in control library (28 °C)	43,179
Tags mapped to genes in control library (4 °C)	44,719
Tags mapped to genes in control library (−10 °C)	47,384
Number of DEGs in control versus cold stress	3872
Number of DEGs in control versus freeze stress	4052
Number of DEGs in cold versus freeze stress	3998

study in response to cold and freezing stress. Among all DEGs, 6,529 were upregulated, whereas 5383 were downregulated in different experimental conditions (control, cold, and freeze; Table 11.3). Later, the GO terms were assigned to the DEGs to identify their dominance in all three categories, viz. molecular functions, cellular components, and biological functions. The GO terms suggested that DEGs under cold and freeze stress are highly enriched in several metabolic processes including cellular, primary, and macromolecule metabolic processes in seabuckthorn. Also, approximately 28 and 15% DEGs were enriched in the category of response to external stimulus and stress. Kyoto encyclopedia of genes and genome based (KEGG) pathway (Du et al. 2014) analysis suggested that the DEGs involved in photosynthetic, plant hormone signal transduction, carotenoid biosynthesis, and vitamin B6 metabolism pathways were found to be significantly enriched in response to cold and freezing stress. The author concluded the study with the validation of 22 randomly selected DEGs with quantitative real-time polymerase chain reaction (qRT-PCR).

In conclusion, deepSAGE-based identification of cold and freeze stress-responsive genes is proved to be an efficient and powerful approach for seabuckthorn. The large amount of raw data generated in the study is available in the public domain and can be accessed with NCBI-SRA accession number SRP049042, whereas the processed data is available at NCBI-Gene Expression Omnibus (GEO) with accession number GSE62489. Many DEGs identified in the study provide useful resource and reference for seabuckthorn and other related plant species. The cold and freeze-responsive genes unique to seabuckthorn could further be characterized and helpful in developing stress tolerance crop varieties using different genetic engineering tools in the future.

11.4.2 De Novo Transcriptome Analysis of Male and Female Seabuckthorn

Considering the medicinal, nutraceutical, and agronomical values (Sharma and Kalkal 2018), seabuckthorn attracts the attention for cultivated farming worldwide. Since seabuckthorn is dioecious, both male and female cultivars are required in various breeding programs for yield and biomass improvement. However, as only female plants bear fruit berries, which are the most commercially consumable part of seabuckthorn, they are preferred over male cultivars in breeding programs (Kalia et al. 2011). Recognizing males and females efficiently at an early stage is therefore essential for successful

breeding programs in seabuckthorn. Toward this goal, an NGS-based comparative transcriptome study between male and female flower buds was conducted in seabuckthorn (Table 11.1) (Bansal et al. 2018). In the study, author employed high-throughput RNA-Seq technique to understand the underlying molecular mechanism and gene involved in seabuckthorn sex determination. The basic strategy for de novo transcriptome analysis for sex determination in seabuckthorn includes the collection of male and female flower buds for total RNA extraction followed by NGS library preparation. The NGS library prepared from male and female flower buds was sequenced from both the ends (PE) on the Illumina Next-Seq500 platform. The raw reads generated in the study were quality checked by Trimmomatic (Bolger et al. 2014) to fetch 51,301,600 and 42,922,794 HQ reads in female and male floral buds, respectively. The HQ reads were assembled into 69,457 transcripts in female buds and 69,390 transcripts in male buds, using Velvet (Zerbino and Birney 2008) assembler. The transcripts were further used for coding DNA sequence (CDS) prediction using TransDecoder (https://github.com/TransDecoder/TransDecoder/wiki) which resulted in the identification of 63,904 and 62,272 CDS in female and male flower buds, respectively. Differential gene expression analysis performed by DESeq pipeline (Anders and Huber 2010) further suggests that among 16,831 common CDS sequences between female and male flower buds, 625 were upregulated and 491 were downregulated. The DEGs were also analyzed for functional annotation using BlastX, GO terms (Conesa and Götz 2008), KEGG pathway (Du et al. 2014), and co-expression network analysis. The comparative analysis between DEGs and Flowering Interactive Database (FLOR-ID) (Bouché et al. 2016) showed significant expression changes in eight genes including (i) phytochrome-associated serine/threonine–protein phosphatase (FYPP), (ii) CHD3-type chromatin-remodeling factor PICKLE (PKL), (iii) protein TOPLESS (TPL), (iv) lysine-specific histone demethylase 1 homolog 1 (LDL1), (v) sensitive to freezing 6 (SFR6), (vi) sucrose synthase 4 (SUS4), (vii) pre-mRNA-processing-splicing factor 8A (PRP8A), and (viii) ubiquitin carboxyl-terminal hydrolase 12 (UBP12) between female and male flower buds.

The transcriptomic data generated from the female and male flower buds in the study by Bansal et al. (2018) is publicly available at http://www.bioinfoindia.org/sbt. The DEGs identified in the study provide new insight into the molecular mechanism involved in flowering and sex determination stages in seabuckthorn. Moreover, in the future, the transcriptomic data could also be utilized in mining molecular markers including SSRs for sex determination in seabuckthorn and cross-amplification in other related pant species.

11.4.3 Seabuckthorn Transcriptome Analysis Under Drought Stress

In addition to cold, drought is another abiotic stress which affects agriculture productivity drastically worldwide. Fortunately, seabuckthorn is one of those plants which can stand in extreme environmental conditions including drought and could serve as a model plant for drought stress tolerance. Therefore, recently RNA-Seq-based transcriptome profiling was performed for seabuckthorn (*Hippophae rhamnoides* subsp. *sinensis*) under drought conditions (Table 11.1) (Ye et al. 2018). The basic strategy used in the study is like other Illumina-based sequencing methodologies. Briefly, five-month-old seabuckthorn plants were divided into two groups: one with normal water supply (control) and the other with two weeks of drought stress (DS). The leaf tissues were harvested for extraction of total RNA in two biological replicates. The total RNA was processed for Illumina library preparation followed by sequencing on the Illumina HiSeq platform.

Approximately 33 million HQ reads generated were assembled using Trinity software (Henschel et al. 2012) into 122,803 unigenes. In addition to many DEGs among control and DS, the author identified a gene which involved in

wax biosynthesis and was upregulated in DS as compared to control. The author suggests that wax prevents water loss through nonstomatal transpiration and plays a crucial role in DS tolerance in seabuckthorn. Another important discovery of the study includes differential expression in 18 genes involved in ABA biosynthesis. In plants under abiotic stress including drought stress, the ABA biosynthesis increases (Vishwakarma et al. 2017). Similarly, in seabuckthorn author suggests that during DS the ABA biosynthesis potentially occurs in roots and is translocated to other parts of the plant to trigger the closing of stomata and prevent water loss. Overall, involvement in the variation in expression of many ABA biosynthesis-related genes confers the drought-tolerant ability of seabuckthorn. Additionally, out of 4438 TFs identified in the study, only 98 showed differential expression, suggesting most TFs required by one or other gene families (Samad et al. 2017) and are continuously expressed in most of the conditions in seabuckthorn. Collectively, the authors identified many genes which show differential expression in response to DS in seabuckthorn. These include genes coding for small-molecule protein such as late embryogenesis abundant (LEA), reactive oxygen species (ROS)-related genes, and gene code for aquaporins (AQP). In concluding remarks, author suggests that the DEGs identified in seabuckthorn under DS could serve as a valuable resource for engineering the drought tolerance in other plant species. The data generated in the study is accessible at NCBI-SRA under the bioproject number PRJNA449450 (SUB3883118).

11.4.4 Transcriptome Analysis in Response to Carbon Dioxide (CO_2) Elevation in Seabuckthorn

Atmospheric carbon dioxide (CO_2) is the major carbon (C) source for land plants; however, its ever-increasing concentration has become a major concern worldwide. Many studies suggest that an increasing level of CO_2 affects agricultural productivity and plant fitness. Therefore, understanding plant response to elevated CO_2 at the genomic level is crucial. In 2018, a comparative transcriptome analysis was performed for two seabuckthorn cultivars of *Hippophae rhamnoides* species under elevated CO_2 (Table 11.1) (Zhang et al. 2018). The two cultivars of seabuckthorn 'zhongguo' and 'xiangyang' were grown in laboratory conditions and provided different levels of CO_2 for three months. The Illumina library was prepared from the pooled RNA extracted from leaf, root, and stem tissues of both the cultivars at all three concentrations of CO_2. Each library was sequenced on Illumina HiSeq2000 platform. The clean reads generated in the study were assembled into 198,785 transcripts and 95,204 unigenes. The authors identified a total of 4740 DEGs in both seabuckthorn cultivars, in response to elevated CO_2 levels. The GO results suggest the different responses of both cultivars under elevated CO_2 levels. For instance, in 'xiangyang' seabuckthorn the photosynthesis, photosystem I, and chloroplast were dominant as compared to 'zhongguo' seabuckthorn where only photosynthesis is a highly enriched term under elevated CO_2 levels. Differential expressions in TFs such as WRKY, MYB, and NAC were also observed with the changing CO_2 concentrations, suggesting their role in elevated CO_2. Overall, this is the first comprehensive transcriptome study under elevated CO_2 conditions in seabuckthorn, which would be useful in understanding the molecular aspects of increasing CO_2 levels on plant fitness. The data generated in the study is available for the research community at NCBI-SRA under the accession number SRP067785.

11.4.5 RNA-Seq-Based Fatty Acids Analysis in Seabuckthorn Berry Pulp

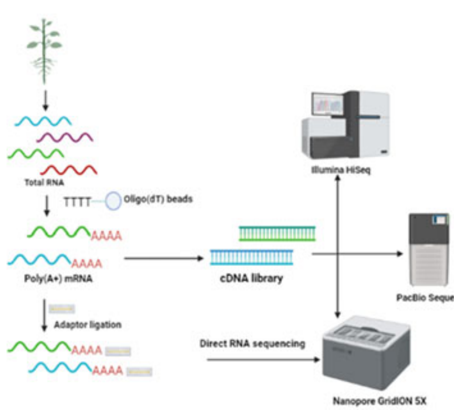
Seabuckthorn berry pulp contains many bioactive compounds and bioactive oils including a high level of uncommon fatty acid (FA) palmitoleic acid (Dulf 2012; Fatima et al. 2012a, b).

To understand the molecular mechanism behind the high level of palmitoleic acid in growing seabuckthorn berry pulp, transcriptomic data was generated from the berry pulp of two seabuckthorn lines 'Za56' and 'TF2–36' (Table 11.1) (Ding et al. 2019). The authors identified many unigenes including 1737 involved in lipid biosynthesis similar to previously reported in the mature seeds of Canadian seabuckthorn cultivar (Fatima et al. 2012a). Further differential gene expression analysis suggests 139 DEGs between the two lines. Among 139 DEGs, 15 genes involved in fatty acid and triacylglycerol (TAG) biosynthesis were validated by qRT-PCR. The transcripts related to these biosynthesis pathways showed higher expression 'Za56' than 'TF2–36' seabuckthorn lines. Overall, DEG analysis from the berry pulp of two lines of seabuckthorn suggested that expression of the gene code for delta9-ACP-desaturase (ACP-Δ 9D) correlates with the high accumulation of palmitoleic acid in 'Za56' line of seabuckthorn. The study provided a basic understanding at the molecular level of the high accumulation of palmitoleic acid and essential oils in seabuckthorn berry pulp.

11.5 Perspectives of Second- and Third-Generation Sequencing Technology and Seabuckthorn

The progress in sequencing technologies and their role in plant science research in the last decade has been astonishing. The genomic/transcriptomic studies of many non-model but important plant species have been driven by the advent of NGS technology where high-throughput sequencing data can be generated in a robust, less labor, and cost-effective manner. In addition to genome sequencing, NGS facilitates genome resequencing, transcriptome, epigenome, and epi-transcriptome sequencing to elucidate the molecular complexities in many plant species. Moreover, the third-generation sequencing platforms developed by Pacific Biosciences (PacBio IsoSeq) and Oxford Nanopore Technologies (ONT), because of their long-read length, facilitate plant science research and potentially uncover the various complexity of plant genome and transcriptome. However, the use of third-generation sequencing technology is still limited in many non-models but important plant species such as seabuckthorn. As previously discussed in this chapter, the second-generation sequencing (NGS) techniques particularly Illumina sequencing have been utilized the most in the exploration of seabuckthorn transcriptome. In this section, we introduce the readers to the third-generation (next-to-next generation) sequencing platforms and how these long-read sequencing platforms could facilitate the seabuckthorn genomic/transcriptomic research. The basic technology difference between Illumina, PacBio, and ONT platforms and their potentials is presented in Fig. 11.4.

The most common platform for second-generation sequencing used is Illumina HiSeq which has the potential to generate hundreds of millions of short reads per run. Illumina sequencing is widely used for several transcriptomes and genome-wide studies in different organisms. In seabuckthorn, most of the NGS-based transcriptome studies are conducted on the Illumina HiSeq platform (Ghangal et al. 2013; Chaudhary and Sharma 2015; Bansal et al. 2018; Ye et al. 2018; Zhang et al. 2018; Ding et al. 2019). The Illumina HiSeq has many advantages over other sequencing platforms including the ability to generate a large number of short reads which helps in the accurate quantification of genome-wide steady-state RNAs. Its potential to sequence poly(A+) mRNA significantly facilitates the understanding of the transcriptional dynamic such as alternative splicing, polyadenylation site analysis, and discovery of small and long non-coding RNAs (Chaudhary et al. 2019; Reddy et al. 2020; Zhu et al. 2020). However, reconstruction of full-length alternative transcripts and RNA modification predictions are the major limitations of Illumina sequencing (Reddy et al. 2020). Third-generation sequencing platforms such as PacBio Sequel and ONT GridION 5X have the capability to sequence at single-base resolution with unexceptionally long reads. The utilization of third-generation sequencing



Applications	Limitations
• Quantification of steady state RNAs • Transcriptome profiling • Differential gene expression • Gene discoveries • Alternative splicing analysis • Small and long non-coding RNA discoveries • Ribosomal profiling • Epigenomics	• Reconstructing full length alternative transcripts • Detection of RNA modification • Not suitable for epi-transcriptomic • Requirement of high efficient bioinformatics tools
• Differential gene expression • Alternative splicing analysis • Alternative polyadenylation analysis • Epi-transcriptomics • Gene discoveries • Isoform discoveries	• High cost
• Differential gene expression • Alternative splicing analysis • Alternative polyadenylation analysis • Epi-transcriptomics • Gene discoveries • Isoform discoveries	• High cost • High error rate in base calling in direct RNA sequencing

Fig. 11.4 Representation of strategy used in second- and third-generation sequencing platforms with potential applications and limitations. The second-generation (Illumina HiSeq) platform and third-generation Pacific Biosciences (PacBio Sequel) required cDNA libraries for sequencing, whereas Nanopore GridION of Oxford Nanopore Technology (ONT) can sequence direct RNA (DRS sequencing) as well cDNA library. This figure is created with BioRender.com

techniques is being increased rapidly in many models and crop plants such as Arabidopsis, rice, sorghum, and maize (Reddy et al. 2020). On one hand, where PacBio allows full-length transcript variants from cDNA libraries, ONT has the same capability by sequencing direct poly(A+) mRNA as well as cDNA library (Reddy et al. 2020). The third-generation sequencing platforms can uncover many aspects which are difficult in second-generation sequencing technologies. However, with the high cost and a large amount of starting material required as in the case of ONT, these high-performance sequence techniques are still away from their application in non-model plant species such as seabuckthorn.

11.6 Conclusion

Seabuckthorn is known for its nutraceutical, medicinal, and ecological potentials from ancient history. In the last decade, the advancement in NGS technology makes it feasible to understand the molecular aspects of this multipurpose plant by transcriptome sequencing. In the current chapter, we discussed the recent developments in seabuckthorn transcriptome studies and emphasize NGS-based transcriptome studies conducted so far in seabuckthorn. The data generated in various seabuckthorn transcriptome studies using NGS could serve as a valuable resource for microsatellite markers, microRNA, small and non-coding RNA, and gene discoveries. The data generated in the transcriptome studies conducted under various stress such as cold, drought, and CO_2 elevation could also prove to be a valuable resource for engineering stress-tolerant crop plants. We also emphasize the ability of second- and third-generation sequencing technologies and their potential in exploring seabuckthorn genome/transcriptome for various purposes.

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: r106
Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 7(10):986–995
Bansal A, Salaria M, Sharma T, Stobdan T, Kant A (2018) Comparative de novo transcriptome analysis of male and female Sea buckthorn. 3 Biotech 8(2):1–15
Bartish IV, Jeppsson N, Nybom H, Swenson U (2002) Phylogeny of Hippophae (Elaeagnaceae) inferred from parsimony analysis of chloroplast DNA and morphology. Syst Bot 27(1):41–54

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120

Bouché F, Lobet G, Tocquin P, Périlleux C (2016) FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. Nucleic Acids Res 44(D1):D1167-1171

Chaudhary S, Khokhar W, Jabre I, Reddy ASN, Byrne LJ, Wilson CM, Syed NH (2019) Alternative splicing and protein diversity: plants versus animals. Front Plant Sci 10:708

Chaudhary S, Sharma PC (2015) DeepSAGE based differential gene expression analysis under cold and freeze stress in seabuckthorn (*Hppophae rhamnoides* L.). PLoS One 10(3):e0121982

Conesa A, Götz S (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:1–12

Ding J, Ruan C, Du W, Guan Y (2019) RNA-seq data reveals a coordinated regulation mechanism of multi-genes involved in the high accumulation of palmitoleic acid and oil in sea buckthorn berry pulp. BMC Plant Biol 19(1):1–18

Djebali S, Wucher V, Foissac S, Hitte C, Corre E, Derrien T (2017) Bioinformatics pipeline for transcriptome sequencing analysis. Methods Mol Biol 1468:201–219

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21

Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y (2014) KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. Mol Biosyst 10(9):2441–2447

Dulf FV (2012) Fatty acids in berry lipids of six sea buckthorn (*Hippophae rhamnoides* L., subspecies *carpatica*) cultivars grown in Romania. Chem Cent J 6(1):1–12

Emrich SJ, Brad Barbazuk W, Li L, Schnable PS (2014) Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res 17(1):69–73

Fatima T, Snyder CL, Weselake RJ, Krishna P (2012b) Transcriptome and metabolite analysis of polyunsaturated fatty acidrich seabuckthorn (*Hippophae rhamnoides*) seed. Pharm Biol 50(5):667

Fatima T, Snyder CL, Schroeder WR, Cram D, Datla R, Wishart D, Weselake RJ, Krishna P (2012a) Fatty acid composition of developing seabuckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed. PLoS ONE 7(4):e34099

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150–3152

Ghangal R, Raghuvanshi S, Chand Sharma P (2009) Isolation of good quality RNA from a medicinal plant seabuckthorn, rich in secondary metabolites. Plant Physiol Biochem 47(11–12):1113–1115

Ghangal R, Chaudhary S, Jain M, Purty RS, Chand Sharma P (2013) Optimization of *de novo*short read assembly of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. PLoS ONE 8(8):e72516

Guo AY, Chen X, Gao G, Zhang H, Zhu QH, Liu XC, Zhong YF, Gu X, He K, Luo J (2008) PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res 36:D966-969

Henschel R, Nista PM, Lieber M, Haas BJ, Wu LS, Leduc RD (2012) Trinity RNA-Seq assembler performance optimization. In: ACM International conference proceeding series, pp 1–8

Jain A, Chaudhary S, Sharma PC (2014) Mining of microsatellites using next generation sequencing of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. Physiol Mol Biol Plants 20(1):115–123

Kalia RK, Singh R, Rai MK, Mishra GP, Singh SR, Dhawan AK (2011) Biotechnological interventions in seabuckthorn (*Hippophae* L.): current status and future prospects. Trees Struct Funct 25(4):559–575

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):1–10

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1(1):2047–2217

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(1):621–628

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science (80-)320:1344–1349

Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7(2):e30619

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 5 (2003):651–652

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33(3):290–295

Reddy ASN, Huang J, Syed NH, Ben-Hur A, Dong S, Gu L (2020) Decoding co-/post-transcriptional complexities of plant transcriptomes and epitranscriptome using next-generation sequencing technologies. Biochem Soc Trans 48:2399–2414

Samad AFA, Sajad M, Nazaruddin N, Fauzi IA, Murad AMA, Zainal Z, Ismail I (2017) MicroRNA and transcription factor: key players in plant regulatory network. Front Plant Sci 8:565

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 2 8(8):1086–1092

Sharma PC, Kalkal M (2018) Nutraceutical and Medicinal Importance of seabuckthorn (*Hippophae* sp.). In: Grumezescue AM, Holban AM (eds) Therapeutic, probiotic, and unconventional foods, Elsevier, pp 227–253

Sharma B, Deswal R (2020) Ecophysiolomic analysis of stress tolerant Himalayan shrub *Hipppophae rhamnoides* shows multifactorial acclimation strategies induced by diverse environmental conditions. Physiol Plant 168(1):58–76

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19 (6):1117–1123

Stewart WDP, Pearson MC (1967) Nodulation and nitrogen-fixation by *Hippophaë rhamnoides* L. in the field. Plant Soil 26(2):348–360

Thiel T (2003) MISA—microsatellite identification tool. Bioinformatics 33(16):2583–2585

Tiitinen KM, Yang B, Haraldsson GG, Jonsdottir S, Kallio HP (2006) Fast analysis of sugars, fruit acids, and vitamin C in seabuckthorn (*Hippophaë rhamnoides* L.) varieties. J Agric Food Chem 54(7):2508–2513

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Prot 7(3):562–578

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270 (5235):484–487

Vishwakarma K, Upadhyay N, Kumar N, Yadav G, Singh J, Mishra RK, Kumar V, Verma R, Upadhyay RG, Pandey M, Sharma S (2017) Abscisic acid signaling and abiotic stress tolerance in plants: a review on current knowledge and future prospects. Front Plant Sci 8:161

Wei W, Chen L, Fu B, Huang Z, Wu D, Gui L (2007) The effect of land uses and rainfall regimes on runoff and soil erosion in the semi-arid loess hilly area, China. J Hydrol 335(3–4):247–258

Yang Y, Yao Y, Xu G, Li C (2005) Growth and physiological responses to drought and elevated ultraviolet-B in two contrasting populations of *Hippophae rhamnoides*. Physiol Plant 124(4):431–440

Yang B, Kallio HP (2001) Fatty acid composition of lipids in seabuckthorn (*Hippophaë rhamnoides* L.) berries of different origins. J Agric Food Chem 49 (4):1939–1947

Ye G, Ma Y, Feng Z, Zhang X (2018) Transcriptomic analysis of drought stress responses of seabuckthorn (*Hippophae rhamnoides* subsp. *sinensis*) by RNA-Seq. PLoS ONE 13(8):e0202213

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):821–829

Zhang G, Zhang T, Liu J, Zhang J, He C (2018) Comprehensive analysis of differentially expressed genes reveals the molecular response to elevated CO_2 levels in two seabuckthorn cultivars. Gene 660:120–127

Zhu S, Ye W, Ye L, Fu H, Ye C et al (2020) PlantAPAdb: a comprehensive database for alternative polyadenylation sites in plants. Plant Physiol 182(1):228–242