RNA-Seq Analysis in Fruit Science: A Review

Ozhan Simsek¹, *, Dicle Donmez², Yildiz Aka Kacar¹, ²

¹Faculty of Agriculture, Department of Horticulture, Cukurova University, Adana, Turkey
²Institute of Natural and Applied Sciences, Department of Biotechnology, Cukurova University, Adana, Turkey

Email address:
ozhan12@gmail.com (O. Simsek), dicledonmez4@gmail.com (D. Donmez), ykacar@cu.edu.tr (Y. A. Kacar)

*Corresponding author

To cite this article:

Received: February 11, 2017; Accepted: February 14, 2017; Published: February 28, 2017

Abstract: Fruit breeding is an ancient technology with dynamic current techniques and an exciting future. There are a number of restraints to conventional fruit breeding which are especially limiting in tree fruits with their long juvenile period, large plant size, and which are represented by unique, highly-selected heterozygous genotypes. Biotechnology offers to minimize disadvantages of classical breeding techniques. In this sense, fruit breeding refers to the purposeful genetic improvement of fruit crops through various techniques including selection, hybridization, mutation induction, and molecular techniques. Among molecular techniques, sequencing technology have been used for many years and recently a new concept titled “RNA-Seq” have been started to performed to understand molecular mechanisms in fruits. RNA-Seq analysis is an effective tool to understand which genes involved and expressed in different mechanisms and organs/cells of a plant. Recently, many articles have been published using RNA-Seq in fruits. In the present review, we illustrated how to apply different RNA-Seq platforms in fruits with examples.

Keywords: Sequencing, Illumina, Gene, Breeding, Fruit

1. Introduction

1.1. Importance of Fruit Science

Among the horticultural plants; fruits are the primary source of essential minerals and vitamins. Fruits will continue to expand due to demand of better tasting, more varied and nutritious. Reproductive behavior in plant species is a crucial part of ecosystem functioning. Therefore, patterns of production of fruit have received much attention from a wide range of ecologists [1, 2]. How do we keep fruit production on par with the burgeoning population? Although conventional plant breeding techniques have made considerable progress in the development of improved varieties, they have not been able to keep pace with the increasing demand for vegetables and fruits in the developing countries. Therefore an immediate need is felt to integrate biotechnology to speed up the crop improvement programs. Especially modern biotechnological methods have been taken part in breeding strategies of fruit crops in addition to classical plant breeding strategies. Several biotechnological methods can be applied to plant to have better ones in the process of fruit breeding [3].

RNA-Seq analysis is a magic method in order to understand which genes involved and expressed in different mechanisms and organs/cells of a plant. Recently, many article have been published related to RNA-Seq in fruits. In the present review, we illustrated how to apply RNA-Seq in fruits with examples.

1.2. RNA-Seq Technology

High-throughput sequencing of cDNA fragment populations, commonly known as RNA-Seq. RNA-Seq is a powerful tool for transcriptome analysis and uses deep-sequencing technologies to produce millions of short cDNA reads. The resulting reads are either aligned to a reference genome or reference transcripts, or assembled de novo (without the genomic sequence) to produce a genome-scale transcription map that consists of both the transcript structure and level of expression for each gene at any particular developmental stage [4, 5, 6]. RNA-Seq has been applied successfully in transcriptome profiling of species without genome sequencing data [7, 8].

The recent RNA-Seq based on NGS (next-generation...
sequencing) enables studies to be carried out on species without corresponding sequenced genome information as a reference [9]. It has become widely applied to model as well as non-model organisms to obtain mass sequence data for molecular marker development, gene discovery and transcriptional analysis [9, 10]. Compared with traditional laboratory methods, RNA-Seq is a high throughput technology, overcoming the weakness of microarrays in exploring unknown genes. Furthermore, it has great advantages in examining transcriptome fine structure, such as detection of allele-specific expression and splice junction variation [11]. This powerful technology, available since only a couple of years, is already making substantial contributions towards the understanding of genome expression and regulation in living organisms [7, 12, 13]. This technology allows us to survey multiple levels of natural variation at unprecedented resolution [14] and can be very useful in the assessment of alternative splicing and the detection of novel gene structures [15]. Werner [16] indicated the great potential contribution of this technology to functional genomics with a special focus on gene regulation by transcription factor binding rates [17].

In principle, RNA-Seq allows analysis of all expressed transcripts, with three key goals: (i) annotating the structures of all transcribed genes including their 5’ and 3’ ends and all splice junctions [18, 19], (ii) quantifying expression of each transcript [20, 21] and (iii) measuring the extent of alternative splicing [5, 22, 23]. Many methods have been recently developed for strand-specific RNA-Seq, and they fall into two main classes. One class relies on attaching different adaptors in a known orientation relative to the 5’ and 3’ ends of the RNA transcript. These protocols generate a cDNA library flanked by two distinct adaptor sequences, marking the 5’ end and the 3’ end of the original mRNA. A second class of methods relies on marking one strand by chemical modification, either on the RNA itself by bisulfite treatment or during second-strand cDNA synthesis followed by degradation of the unmarked strand [23]. For RNA-Seq four high-throughput DNA sequencing technologies (Roche 454, Illumina, Helicos BioSciences and Life Technologies) can be used, and new technologies are in development by others. Among these methods, the most widely used system to generate RNA-Seq data is the Illumina system mainly due to the cheaper cost per base sequenced. Given the importance of sequencing capabilities, such as throughput, read length, error rate and ability to perform paired reads, for RNA-Seq as well as genomic studies, NGS companies are constantly improving their platforms to provide the best sequencing performance at the lowest cost [24, 25]. The GS instrument was introduced in 2005, developed by 454 Life Sciences, as the first next-generation system on the market. The method has recently increased the achieved reading length to the 400–500 base range, with paired-end reads, and as such is being applied to genome (bacterial, animal, human) sequencing. One spectacular application of the system was the identification of the culprit in the recent honey-bee disease epidemics. A relatively high cost of operation and generally lower reading accuracy in homopolar stretches of identical bases are mentioned presently as the few drawbacks of the method. The next upgrade 454 FLX Titanium will quintuple the data output from 100 Mb to about 500 Mb, and the new picotiter plate in the device uses smaller beads about 1 mm diameter. The device, schema of operation, its further developments and list of publications with applications can be found at http://www.454.com/index.asp [26, 27].

The Illumina Genome Analyser produces over 100 million short reads (35–76 bases, depending on the sequencing chemistry) leading to 3–6 gigabases of sequencing data in one run. A megabase costs about USS4. We refer to this technology as a short read technology. It is based on solid-phase bridge PCR and uses a ‘sequencing by synthesis’ approach, with fluorescent dye-labelled reversible terminator nucleotides. It uses fragmented double-stranded DNA as template. Fragments of up to 10 kb can be used for the construction of paired end sequencing libraries. The technology is also referred to as Solexa sequencing. The accuracy of the produced sequence data is greater 98.5% (http://www.solexa.com) [24, 28].

The Helicos Genetic Analysis System (http://www.helicosbio.com) was the first commercially available single-molecule sequencing system (SMS) on the market [29]. Helicos method produces average read lengths of 35 bp across 600 million to 1 billion reads, totaling 21 – 35 Gb per run at a rate of > 1 Gb/h. The Helicos platform lends itself well to multiplexing with up to 96 samples per channel or 4800 samples per run (http://www.helicosbio.com) [30].

The Applied Biosystems SOLiD system is based on emulsion PCR in combination with sequencing by ligation with dye-labelled oligonucleotides [31]. It produces up to one billion short reads (up to 50 bases) per run, leading to a total sequence output of up to 30 gigabases per single read run. As templates it uses fragmented double-stranded DNA. Fragment sizes for the construction of paired end sequencing libraries can be up to 10 kb. The sequences produced are 99.94% accurate (http://www3.appliedbiosystems.com/AB_Home) [24, 28].

RNA-Seq technology represents the latest and most powerful tool for characterizing transcriptomes [32]. RNA-Seq is its ability to provide information on transcripts that are expressed at very low levels, limited only by the total number of reads that are generated [33]. Because RNA-Seq is performed using tagged libraries of short cDNAs, prepared from fragmented or unfragmented RNA, it does not require prior knowledge of the sequences to be profiled [17].

2. Instances of RNA-Seq Analysis in Fruit Crops

Many RNA-Seq articles have been recently published in fruits to investigate genes expressed and involved in different mechanisms. RNA-Seq is also good tool to understand expression level of genes. Different RNA-Seq platforms have
been released for several years. These platforms were used for RNA sequencing in fruit species.

A comparative transcriptomic analysis between HW (Hot Water) treated and non-treated tomato fruit before and after cold storage was carried out. In the study massive sequencing were perform on a 5500 SOLID System (with Exact Call Chemistrymodule). RNA-Seq analysis detected a large number of differentially expressed genes that ranged from 2235 (heat shock) to 5433 (cold storage). Three clusters of genes were identified after 2 weeks of cold storage: the chilling-response included the down-regulation of genes involved in photo-synthesis, metabolism of cell wall, lipid and ethylene, as well as the up-regulation of genes for trehalose synthesis and transcription factors (DOF and MYB); the chilling-susceptibility was associated with the down-regulation of genes involved in carotenoid biosynthesis, which correlates with the main CI symptom of uneven ripening; meanwhile, the chilling-tolerance was related to the up-regulation of genes for heat stress (heat shock proteins and heat shock transcription factors) and detoxification (glutathione S-transferases) [34].

In another study, in order to understand the systemic effect of the fungal symbiosis on the tomato fruit, Zouari et al. [35] used RNA-Seq to perform global transcriptome profiling on Moneymaker tomato fruits at the turning ripening stage. In this study fruits were collected at 55 days after flowering, from plants colonized with *Funnelliformis mosseae* and from control plants, which were fertilized to avoid responses related to nutrient deficiency. RNA-Seq analysis were performed using the Illumina Genome Analyzer (Solexa). Transcriptome analysis identified 712 genes that are differentially expressed in fruits from mycorrhizal and control plants. Gene Ontology (GO) enrichment analysis of these genes showed 81 overrepresented functional GO classes. Up-regulated GO classes include photosynthesis, stress response, transport, amino acid synthesis and carbohydrate metabolism functions, suggesting a general impact of fungal symbiosis on primary metabolisms and, particularly, on mineral nutrition. Down-regulated GO classes include cell wall, metabolism and ethylene response pathways.

*De novo* RNA-Seq of ripe blackberries grown under field conditions was performed using Illumina Hi-Seq 2000. Almost 9 billion nucleotide bases were sequenced in total. Following assembly, 42,062 consensus sequences were detected. For functional annotation, 33,040 (NR), 32,762 (NT), 21,932 (Swiss-Prot), 20,134 (KEGG), 13,676 (COG), 24,168 (GO) consensus sequences were annotated using different databases; in total 34,552 annotated sequences were identified [6]. Gupta et al. [36] performed high-throughput transcriptome sequencing (RNA-Seq) and differential gene expression analysis of five stages of berry development and ripening. RNA-Seq expression profiling showed that blueberry growth, maturation, and ripening involve dynamic gene expression changes, including coordinated up- and down-regulation of metabolic pathway enzymes and transcriptional regulators. Analysis of RNA-Seq alignments identified developmentally regulated alternative splicing, promoter use, and 3’ end formation. Kim et al. [37] optimized a procedure for high-quality RNA isolation from vegetative and reproductive tissues of climaergic and non-climaergic plum cultivars and conducted high-throughput transcriptomics. Researchers were identified 20 candidate reference genes from significantly non-differentially expressed transcripts of RNA-Seq data and verified their expression stability using qRT-PCR on a total of 141 plum samples which included flesh, peel, and leaf tissues of several cultivars collected from three locations over a 3-year period.

Fruit skin color is one of the most important traits for both the commercial and esthetic value of strawberry. Anthocyanins are the most prominent pigments in strawberry that bring red, pink, white, and yellow hues to the fruits in which they accumulate. Zhang et al. [8] conducted a *de novo* assembly of the fruit transcriptome of woodland strawberry and compared the gene expression profiles with yellow (Yellow Wonder, YW) and red (Ruegen, RG) fruits. The cDNA library was sequenced on the Illumina sequencing platform (Hi-Seq™ 2500). *De novo* assembly yielded 75,426 unigenes, 21.3% of which were longer than 1,000 bp. Among the high-quality unique sequences, 45,387 (60.2%) had at least one significant match to an existing gene model. A total of 595 genes, representing 0.79% of total unigenes, were differentially expressed in YW and RG. Among them, 224 genes were up-regulated and 371 genes were down-regulated in the fruit of YW.

Wang et al. [38] investigated to facilitate isolation of genes controlling important horticultural traits of peach, transcriptome sequencing was conducted. The cDNA libraries were sequenced using Illumina Hi-Seq2000 sequencer. A total of 133 million pair-end RNA-Seq reads were generated from leaf, flower, and fruit, and 90% of reads were mapped to the peach draft genome. Sequence assembly revealed 1,162 transcription factors and 2,140 novel transcribed regions (NTRs). Of these 2,140 NTRs, 723 contain an open reading frame, while the rest 1,417 are non-coding RNAs.

Liu et al. [39] were reported that Solexa sequencing has been used to discover small RNA populations and compare miRNAs on genome-wide scale in watermelon grafting system. A total of 11,458,476, 11,614,094 and 9,339,089 raw reads representing 2,957,751, 2,880,328 and 2,964,990 unique sequences were obtained from the scions of self-grafted watermelon and watermelon grafted on-to bottle gourd and squash at two true-leaf stage, respectively. 39 known miRNAs belonging to 30 miRNA families and 80 novel miRNAs were identified in our small RNA dataset. Compared with self-grafted watermelon, 20 (5 known miRNA families and 15 novel miRNAs) and 47 (17 known miRNA families and 30 novel miRNAs) miRNAs were expressed significantly different in watermelon grafted on to bottle gourd and squash, respectively. miRNAs expressed differentially when watermelon was grafted onto different rootstocks, suggesting that miRNAs might play an important
role in diverse biological and metabolic processes in watermelon and grafting may possibly by changing miRNAs expressions to regulate plant growth and development as well as adaptation to stresses. Feng et al. [40] were designed to obtain transcript sequence data and examine gene expression in bayberry developing fruit based on RNA-Seq and bioinformatic analysis, to provide a foundation for understanding the molecular mechanisms controlling fruit quality changes during ripening. Four RNA samples containing various tissues and fruit of different development and ripening stages were sequenced using the latest Illumina deep sequencing technique. RNA-Seq generated 1.92 G raw data, which was then de novo assembled into 41,239 UniGenes with a mean length of 531 bp. Approximately 80% of the UniGenes (32,805) were annotated against public protein databases, and coding sequences (CDS) of 31,665 UniGenes were determined. Over 3,600 UniGenes were differentially expressed during fruit ripening, with 826 up-regulated and 1,407 down-regulated. GO comparisons between the UniGenes of these two types and interactive pathways (Ipath) analysis found that energy-related metabolism was enhanced, and catalytic activity was increased. All genes involved in anthocyanin biosynthesis were up-regulated during the fruit ripening processes, concurrent with color change.

Zhang et al. [41] investigated to identify novel as well as conserved miRNAs in citrus, deep sequencing of small RNA library combined with microarray was performed in precocious trifoliate orange (an early flowering mutant of trifoliate orange, Poncirus trifoliata L. Raf.). Illumina technology were used in sequencing analysis. A total of 114 conserved miRNAs belonging to 38 families and 155 novel miRNAs were determined. The miRNA star sequences of 39 conserved miRNAs and 27 novel miRNAs were also discovered among newly identified miRNAs, providing additional evidence for the existence of miRNAs.

RNA-Seq analysis could be also performed for molecular marker developments. There are some articles in fruits in order to develop molecular markers (Pomegranate [42], (Pomelo) [43], (Banana) [44].

Some recent examples by using RNA-Seq technologies, and mechanisms for fruits were presented Table 1.

<table>
<thead>
<tr>
<th>Species</th>
<th>Mechanism</th>
<th>RNA-Seq Technology</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweet Orange (Citrus sinensis)</td>
<td>Fruit ripening</td>
<td>Illumina Hi-Seq 2000</td>
<td>[45]</td>
</tr>
<tr>
<td>Sweet Orange (Citrus sinensis)</td>
<td>Heterozygous</td>
<td>Solid Bioscope</td>
<td>[46]</td>
</tr>
<tr>
<td>Mango (Mangifera indica L.)</td>
<td>Characterization of transcriptome and plastid genome</td>
<td>454-GS-FLX Titanium</td>
<td>[47]</td>
</tr>
<tr>
<td>Pomegranate (Punica granatum L.)</td>
<td>Molecular Marker Development</td>
<td>Illumina</td>
<td>[48]</td>
</tr>
<tr>
<td>Mango (Mangifera indica Linn)</td>
<td>Fruit development and ripening</td>
<td>Illumina Hi-Seq 2000</td>
<td>[49]</td>
</tr>
<tr>
<td>Apple (Malus communis)</td>
<td>Transcriptome sequencing</td>
<td>Illumina Hi-Seq 2000</td>
<td>[50]</td>
</tr>
<tr>
<td>Trifoliate orange Poncirus trifoliata (L.)</td>
<td>Phosphorus deficiency</td>
<td>Illumina Hi-Seq 2000 and 454-GS-FLX Titanium</td>
<td>[51]</td>
</tr>
<tr>
<td>Apple (Malus xiaojinensis)</td>
<td>Iron tolerance</td>
<td>Illumina Hi-Seq 2000</td>
<td>[52]</td>
</tr>
<tr>
<td>Cheery (Prunus avium L.)</td>
<td>Fruit development and exocarp</td>
<td>Illumina Hi-Seq 2000</td>
<td>[53]</td>
</tr>
<tr>
<td>Sweet Orange (Citrus sinensis)</td>
<td>Fruit ripening</td>
<td>Illumina Hi-Seq 2000</td>
<td>[54]</td>
</tr>
<tr>
<td>Sweet Orange (Citrus sinensis)</td>
<td>Cold tolerance under storage conditions</td>
<td>Illumina Hi-Seq 2000</td>
<td>[55]</td>
</tr>
<tr>
<td>Nectarine (Prunus persica)</td>
<td>Self-incompatibility</td>
<td>Illumina Hi-Seq 2000</td>
<td>[56]</td>
</tr>
<tr>
<td>Xianghui lemon (Citrus limon (L.) Burm. F)</td>
<td>Biosynthesis of anthocyanin</td>
<td>Illumina/Solexa</td>
<td>[57]</td>
</tr>
<tr>
<td>Cherry (Prunus avium L.)</td>
<td>Chitosan and salicylic acid applications</td>
<td>Illumina/Solexa</td>
<td>[58]</td>
</tr>
<tr>
<td>Sweet Orange (Citrus sinensis)</td>
<td>Color of fruits</td>
<td>Illumina Hi-Seq 2000</td>
<td>[59]</td>
</tr>
<tr>
<td>Apple (Malus sieversii v. niedzwetzkyana X Malus domestica)</td>
<td>Pnicarp colors and contents</td>
<td>Illumina</td>
<td>[60]</td>
</tr>
<tr>
<td>Pummelo (Citrus grandis)</td>
<td>Biosynthesis of anthocyanin</td>
<td>Illumina Hi-Seq 2000</td>
<td>[61]</td>
</tr>
<tr>
<td>Apple (Malus sieversii v. niedzwetzkyana)</td>
<td>Development of Molecular Markers</td>
<td>Illumina</td>
<td>[62]</td>
</tr>
<tr>
<td>Pummelo (Citrus grandis)</td>
<td>Iron deficiency</td>
<td>Illumina</td>
<td>[63]</td>
</tr>
<tr>
<td>Almond (Prunus mongolica maxim)</td>
<td>Drought stress</td>
<td>Illumina Hi-Seq 2000</td>
<td>[64]</td>
</tr>
<tr>
<td>Cherry (Prunus pseudocerasus)</td>
<td>Gene expression in seeds</td>
<td>Illumina</td>
<td>[65]</td>
</tr>
<tr>
<td>Trifoliate orange (Poncirus trifoliata (L.) Raf.)</td>
<td>Cold stress</td>
<td>Illumina</td>
<td>[66]</td>
</tr>
<tr>
<td>Apple (Malus x domestica)</td>
<td>Russeting in exocarp</td>
<td>Illumina</td>
<td>[67]</td>
</tr>
<tr>
<td>Apple (Malus domestica)</td>
<td>Sorbitol synthesis</td>
<td>Illumina Hi-Seq 2000</td>
<td>[68]</td>
</tr>
<tr>
<td>Banana (Musa balbisiana)</td>
<td>Transcriptome sequencing, molecular marker development</td>
<td>Ion Torrent</td>
<td>[69]</td>
</tr>
<tr>
<td>Blackberry (Rubus sp. Var. Lochness)</td>
<td>Transcriptome sequencing</td>
<td>Illumina Hi-Seq 2000</td>
<td>[70]</td>
</tr>
<tr>
<td>Strawberry (Fragaria vesca)</td>
<td>Post-harvest auxin abscisic acid response</td>
<td>Illumina Hi-Seq 2000/2500</td>
<td>[71]</td>
</tr>
<tr>
<td>Grapevine (Vitis vinifera X Vitis labrusca)</td>
<td>Transcriptome sequencing</td>
<td>Illumina Hi-Seq 2500</td>
<td>[72]</td>
</tr>
<tr>
<td>Grape (Vitis vinifera)</td>
<td>Bud dormancy</td>
<td>Illumina Hi-Seq 2000</td>
<td>[73]</td>
</tr>
</tbody>
</table>

3. Conclusion

Many fruit crops are generated by hybridization and selection. However, crop hybridization breeding has limitations that are difficult to overcome [3]. This requires enormous amounts of labor and land resources, although fast track breeding techniques and molecular technologies may accelerate breeding and selection processes. Finally advances in biotechnology must be attached to breeding efforts.
Recently different RNA-Seq platforms were released for researchers. It seems that RNA-Seq platforms and bioinformatics for data analysis will continue to have more significance for fruits.

References


