Chapter 14
RNA-Seq Data Analysis for Studying Abiotic Stress in Horticultural Plants

V.V. Mironova, C. Weinholdt, and I. Grosse

Abstract Initiating the project on sequencing the *Arabidopsis thaliana L.* genome at the end of the twentieth century, researchers one day wished to expand the accumulated knowledge on *Arabidopsis* genetics to horticultural plants. The future arrived with the appearance of high-throughput sequencing technologies that allowed the investigation of transcriptomes of non-model plants at an unprecedented pace. RNA-seq experiments provide a unique opportunity of studying in depth the molecular-genetic basis for plant response to environmental cues. Here we substantiate the potential of RNA-seq experiments in applications to horticultural plants. The basic steps in RNA-seq data analysis and available software packages are presented in the first section. Examples of RNA-seq data analyses, including studies of gene expression changes under various stresses in horticultural plants, and transcriptome analyses of the tolerance to abiotic stresses in horticultural plants are given in the second section.

Keywords Genomics • Horticultural plants • RNA-seq • Stress response • Transcriptomics

V.V. Mironova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Novosibirsk State University, Novosibirsk, Russia
e-mail: victoria.v.mironova@gmail.com

C. Weinholdt
Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany

I. Grosse (✉)
Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany

German Center of Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
e-mail: grosse@informatik.uni-halle.de

Y. Kanayama, A. Kochetov (eds.), *Abiotic Stress Biology in Horticultural Plants*, DOI 10.1007/978-4-431-55251-2_14

14.1 Introduction

The transcriptome is the full set of RNA molecules generated by a cell or a population of cells. The transcriptome includes mRNAs, tRNAs, miRNAs, and other noncoding RNAs. In contrast to the genome, the transcriptome varies significantly with cell type, developmental stage, biotic and abiotic conditions. There are two sources of transcriptome variability: variability of the transcribed sequences and variability of expression levels (Wang et al. 2009). Several technologies have been developed for studying transcriptomes, but only the application of next-generation-sequencing methods to transcriptomes (RNA-seq) has provided the opportunity of studying both sources of variability simultaneously and with high resolution (McGettigan 2013).

RNA-seq experiments have shown that both the transcribed sequences and their expression levels vary significantly under different biotic and abiotic stresses. Alternative splicing, alternative transcription start sites, alternative transcription stop sites, and RNA editing are widely present in transcriptomes of well-studied species (Zavolan et al. 2003; Trapnell et al. 2010). Moreover, a great variety of previously uncharacterized noncoding RNAs has been found by RNA-seq experiments, and sequencing of small RNAs (miRNA-seq) is nowadays as popular as standard sequencing of mRNAs (Motameny et al. 2010; Lee et al. 2010).

RNA-seq experiments provide additional benefits for species without sequenced genomes or with poor genome annotations. For horticultural plants, the availability of information about transcribed sequences has a fundamental impact on many areas of plant biology such as plant phylogenetics, reverse genetics, DNA fingerprinting, and marker-assisted selection. Here, we review applications of RNA-seq techniques to study abiotic stresses in horticultural plants. The review consists of two sections. The first section describes technical aspects of handling RNA-seq data. The second section describes examples of applications of RNA-seq techniques to horticultural plants subjected to various abiotic stresses such as drought, salinity, flood, cold, and mineral deficiency.

14.2 Analysis of RNA-Seq Data

Different next-generation sequencing technologies exist that convert input RNA material into millions of short reads. Modern sequencing platforms are based on the sequencing-by-synthesis technology with either a DNA polymerase (e.g., Roche 454, Illumina, Helicos, Pacific Biosciences) or a ligase (e.g., Life Technologies SOLiD, Complete Genomics) as a key component. The sequencing platforms can be further categorized as either *single-molecule-based* such as Helicos and Pacific Biosciences or *ensemble-based* such as Illumina and SOLiD (Metzker 2010). So far Illumina and SOLiD have the smallest error rates, less than 1 % per base, which is of special importance for the analysis of miRNAs. Both platforms are widely used because of their high sequencing capacity, which makes it possible to measure low-abundance transcripts (Metzker 2010).

RNA-seq data analyses consist of a sequence of steps that must be adapted and optimized depending on the goal of the experiment, the RNA material, and the species. The analysis can be divided in two main parts: (i) general steps that must be performed in each RNA-seq analysis (Sect. 14.2) and (ii) specific steps that vary from analysis to analysis for long RNAs (Sect. 14.2.2) and small RNAs (Sect. 14.2.3).

14.2.1 General Steps of RNA-Seq Data Analysis

General steps of RNA-seq data analysis involve quality control and filtering of reads (Sect. 14.2.1.1) and assembling the filtered reads and connecting them with a reference genome or transcriptome by mapping or *de novo* assembly (Sect. 14.2.1.2).

14.2.1.1 Quality Control and Filtering

Reads obtained from a sequencing platform must be quality controlled and filtered for possible sequencing errors, artifacts, or contaminations. Typically, the following three types of reads or sub-reads are eliminated: (i) low-quality reads or sub-reads including a high percentage of low-quality bases or a high percentage of uncalled bases; (ii) reads or sub-reads including sequencing artifacts such as duplicate reads, adapter sequences, barcodes, or a strong bias in the GC content; and (iii) reads or sub-reads including DNA contamination or RNA contamination from other species.

The tool *fastQC* (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) can be used for visualizing the read qualities, the distribution of read-lengths, which is important for the analysis of RNA-seq data of small RNAs, or the GC content. An extensive comparison of methods addressing errors of Illumina data, published by Del Fabbro et al. (2013), shows that quality control and filtering reduce the need for disk space and computation time for subsequent mapping and de novo assembly steps.

Table 14.1 provides an overview of different tools for quality control and filtering. Column 2 of Table 14.1 provides information about the underlying algorithmic approach such as the *running sum* approach utilized by Cutadapt (Martin 2011) or different *window-based* approaches utilized by other tools such as *FASTX quality trimmer* (hannonlab.cshl.edu/fastx_toolkit), *PRINSEQ* (Schmieder and Edwards 2011), *Trimmomatic* (Lohse et al. 2012), and *sickle* (github.com/najoshi/sickle). Column 3 provides information about the capability of handling paired-end reads. Here, tools such as *sickle* or *Trimmomatic* can process paired-end reads simultaneously, whereas *Cutadapt*, *FASTX quality trimmer*, or *PRINSEQ* process the two read sets independently. Columns 4 and 5 provide information about the capability of handling color-space reads relevant for processing SOLiD data and about the capability of adapter removal. Clipping of sequencing adapters by tools such as *Cutadapt* or *Trimmomatic* is indispensable for the analysis of RNA-seq data of small RNAs, because the typical read length of 50–200 bp is greater than the typical length of small RNAs so that adapters or barcodes are also sequenced.

Table 14.1 Selection of tools for quality control and filtering grouped by their algorithmic approach and other useful features

Tool	Algorithmic approach	Processing of paired-end reads	Processing of SOLiD data	Adapter removal
Cutadapt	Running sum	No	Yes	Yes
FASTX quality trimmer	Window based	No	Yes (for fixed-length trimming)	Yes (by FASTX Clipper)
PRINSEQ	Window based	No	No	No
Trimmomatic	Window based	Yes	No	Yes
Sickle	Window based	Yes	No	No

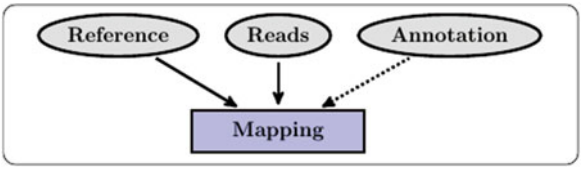

Fig. 14.1 Read mapping for case (i) where a reference genome or transcriptome and an annotation are available. Reads can be mapped to the reference genome or transcriptome using the annotation

14.2.1.2 Assembly of Reads and Connecting them to a Reference Genome or Transcriptome

Three situations can occur when analyzing RNA-seq data of horticultural plants: (i) an annotated reference genome or transcriptome is available, (ii) a reference genome or transcriptome is available, but it is not annotated, and (iii) no reference genome or transcriptome is available. We address typical approaches for these three cases in the sections "Mapping Assembly," "Transcriptome Assembly," and "De Novo Transcriptome Assembly."

Mapping Assembly

In case (i), in which both a reference genome or transcriptome and an annotation are available, mapping of quality-filtered and trimmed reads can be performed directly without additional preprocessing steps (Fig. 14.1). Mapping is the process of determining the position of each read in the reference genome or transcriptome. Hence, mapping is a fundamental step of each RNA-seq data analysis. To date more than 80 mapping tools are available, which differ strongly in their algorithmic strategies and their mapping behavior (Fonseca et al. 2012). Mapping tools can be grouped by several criteria (Li and Homer 2010; Alamancos et al. 2014), and we concentrate on the following two criteria in this review.

Table 14.2 Selection of mapping tools from Li and Homer (2010) and Alamancos et al. (2014)

Tools	Algorithmic approach	Alignments reported	SOLiD data	Splice-aware mapper	Use of annotation
MAQ	Hash table	B,R	Yes	No	–
SHRiMP	Hash table	B, U, S	Yes	No	–
SOAP	Hash table	B, R, S	No	No	–
Bowtie	Suffix/prefix tree	A, B, R, S	Yes	No	–
Bowtie 2	Suffix/prefix tree	A, B, R, S	No	No	–
BWA	Suffix/prefix tree	R, S	Yes	No	–
PALMapper	Suffix/prefix tree	A, B	Yes	Yes	No
segemehl	Suffix/prefix tree	A, B	No	Yes	No
STAR	Suffix/prefix tree	A, B, S	No	Yes	Yes
TopHat 2	Suffix/prefix tree	B, S	Yes	Yes	Yes

Mapping tools are grouped by their algorithmic approach (column 2), by their treatment of reads that map to multiple locations (column 3), and by their capability of handling color-space reads from SOLiD (column 4), handling split reads (column 5), and using the annotation (column 6) in case of splice-aware mapping tools. If a read maps to multiple locations (column 3), all alignments can be reported (A), only the best alignment can be reported (B), a randomly selected alignment can be reported (R), only unique alignments can be reported (U), or a user-defined number of alignments can be reported (S).

First, mapping tools can be grouped according to their algorithmic approach. They can be based on hash tables, such as *MAQ* (Li et al. 2008a), *SOAP* (Li et al. 2008b), and *SHRiMP* (Rumble et al. 2009), or on suffix or prefix tries, such as *Bowtie* (Langmead et al. 2009), *Bowtie 2* (Langmead and Salzberg 2012), *BWA* (Li and Durbin 2009), *segemehl* (Hoffmann et al. 2014), and *TopHat 2* (Trapnell et al. 2009) (Table 14.2).

Second, mapping tools can be grouped by their capability of handling reads that span exon–exon junctions. Such reads, which are called split reads, can be neglected by *unspliced mappers* such as *PALMapper* (Jean et al. 2010), segemehl, *STAR* (Dobin et al. 2013), or *TopHat 2* (Table 14.2). *Splice-aware* mappers make it possible to map reads to splicing junctions and thereby increase the number of mappable reads and provide helpful information for detecting novel isoforms. Mappers such as *TopHat 2* and *STAR* (Table 14.2) can optionally use the annotation for increasing the accuracy of mapping to known splice junctions.

Transcriptome Assembly

In case (ii), in which the reference genome or transcriptome is available but an annotation is missing, read mapping can be performed as in section "Mapping Assembly", but mapping tools cannot benefit from the optional use of the annotation. However, an annotation can be generated by one of the following two approaches (Fig. 14.2).

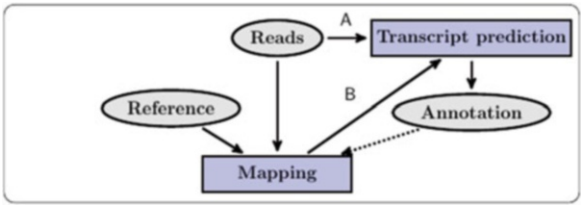

Fig. 14.2 Missing annotation can be substituted by computational transcript prediction directly from the reads (A) or using the mapped reads for a genome-guided assembly of the genes (B)

Evidence-based transcript prediction tools such as *mGene* (Schweikert et al. 2009) or *Augustus* (Keller et al. 2011) can be used for obtaining an annotation of transcripts or exon-intron structures. *Augustus* can integrate evidence from expressed sequence tags (EST), tandem mass spectrometry (MS/MS) data, protein alignments, and genomic alignments (Stanke et al. 2008) and is capable of identifying coding exons even from transcripts with very low abundance (Steijger et al. 2013).

Alternatively, *genome-guided* transcriptome assembly tools such as *Cufflinks* (Trapnell et al. 2010), *Scripture* (Guttman et al. 2010), and *GRIT* (Boley et al. 2014) use reads mapped by splice-aware mappers to a reference genome or transcriptome for identifying transcript models. One advantage of using genome-guided transcriptome assembly tools is that these tools make it possible to discover novel transcripts and to assemble low-abundance transcripts (Martin and Wang 2011).

De Novo Transcriptome Assembly

In case (iii), in which a reference genome or transcriptome is not available, a reference transcriptome must first be generated by a de novo assembly tool (Fig. 14.3). This step is often necessary when working with non-model species, and in this case it is advisable to have a sufficiently high sequencing depth and to perform paired-end sequencing.

Historically, *de novo* assembly tools such as *ABySS* (Birol et al. 2009), *SOAPdenovo* (Luo et al. 2012), and *Velvet* (Zerbino and Birney 2008) were developed for assembling genomes, but specialized tools for performing de novo assembly of transcriptomes have also been developed in the meantime. Examples of de novo transcriptome assembly tools are *OASES* (Schulz et al. 2012), *Rnnotator* (Martin et al. 2010), and *SOAPdenovo-trans* (Xie et al. 2014), *Trans-ABySS* (Simpson et al. 2009), and *Trinity* (Haas et al. 2013).

Table 14.3 shows several features of these five *de novo* transcriptome assembly tools. Specifically, column 2 shows the algorithmic approach, column 3 the capability of detecting alternative isoforms, and column 4 the capability of quantifying isoform-specific expression. From column 2 of Table 14.3 we see that *Trinity* has a k-mer size fixed to k = 25 and that the k-mer sizes for the tools *Oases*, *Rnnotator*, *SOAPdenovo-trans*, and *Trans-AbySS* are variable. We see from column 3 that

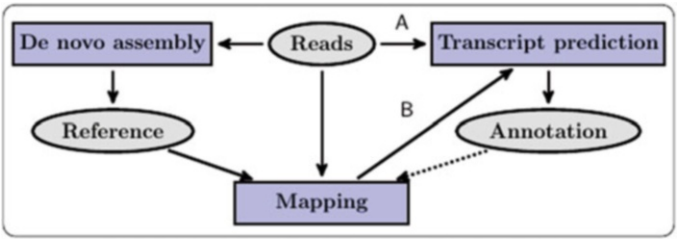

Fig. 14.3 The missing reference genome or transcriptome can be substituted by de novo transcriptome assembly. Subsequent mapping to the *de novo* assembled reference transcriptome can be performed and the annotation can be predicted as described by (Fig. 14.2)

Table 14.3 Selection of de novo transcriptome assembly tools

Tool	Algorithmic approach	Detection of alternative isoforms	Quantification of isoforms
OASES[a]	Variable k-mer	No	No
Rnnotator[a]	Variable k-mer	No	No
SOAPdenovo-trans	Variable k-mer	Yes	No
Trans-Abyss	Variable k-mer	Yes	Isoform read coverage
Trinity	Single k-mer	Yes	Yes[b]

De novo transcriptome assembly tools are grouped by their algorithmic approach (column 2) and by their capability of detecting alternative isoforms (column 3) and of quantifying isoform-specific expression (column 4)

[a]Uses *Velvet* for de novo transcriptome assembly

[b]Uses *RSEM* for calculating *RKPM* values (see Sect. 14.2.2.1)

SOAPdenovo-trans, *Trans-AbySS*, and *Trinity* are capable of detecting alternative isoforms. De novo assembly tools can consume hundreds of gigabytes of RAM and can run for weeks even on a high-performance-computing cluster (Martin and Wang 2011). Rapaport et al. (2013) provides further details on and a comprehensive review of de novo transcriptome assemblers.

Additionally, the tool *scaffold_builder* (Silva et al. 2013) has been developed for scaffolding preassembled contigs against a genome from an evolutionarily related species with a sufficiently high degree of sequence similarity.

After de novo transcriptome assembly has been performed and a resulting reference transcriptome is available, reads can be mapped to the de novo assembled reference transcriptome as described in section "Mapping Assembly" and illustrated by Figs. 14.1 and 14.2. Additionally, *evidence-based* transcript prediction tools or genome-guided transcriptome assembly tools can be applied for annotating exons in the reference transcriptome (Fig. 14.2).

14.2.2 Specific Steps of RNA-Seq Data Analysis of Long RNAs

Mapped reads can be used for addressing numerous tasks such as detecting differentially expressed transcripts (*Sect. 14.2.2.1*) or calling single-nucleotide polymorphisms (*Sect. 14.2.2.2*).

14.2.2.1 Detecting Differentially Expressed Transcripts

In this section we describe the detection of differentially expressed transcripts, which consists of three steps called quantification, normalization, and statistical testing (Fig. 14.4).

Quantification

As starting point for the detection of differentially expressed transcripts, RNA abundances corresponding to each transcript must be quantified: this can be accomplished by (i) *count-based* approaches or by (ii) *model-based* approaches.

Count-based approaches simply quantify the relative RNA abundance per transcript by counting the number of mapped reads per annotated transcript. Count-based approaches can be divided in two groups. Only uniquely mapped reads are counted by tools of group (a), whereas all mapped reads are counted by tools of group (b). Examples for tools of group (a) are *HTSeq-count* (Anders et al. 2014) and *GenomicRanges* (Lawrence et al. 2013). Examples for tools of group (b) are *IRanges* (Lawrence et al. 2013) and *BEDTools* (Quinlan and Hall 2010).

Several tools such as *featureCounts* (Liao et al. 2014) belong to both groups as they allow counting both uniquely mapped and all mapped reads. Approaches for quantifying the relative RNA abundance based on all mapped reads often lead to a biased quantification from cross-mapping of reads from close homologs (Anders et al. 2013). whereas, approaches based on counting only uniquely mapped reads

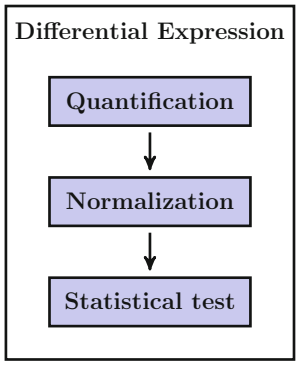

Fig. 14.4 Detection of differentially expressed transcripts. Current approaches for detecting differentially expressed transcripts consist of three steps quantification, normalization, and statistical testing

typically lead to a less biased quantification of the relative RNA abundance and thus are preferable for detecting differentially expressed transcripts.

Model-based approaches such as *Cufflinks* (Trapnell et al. 2010) and *RSEM* (Li and Dewey 2011) combine the quantification step with the subsequent normalization step. This combination leads to normalized values for relative RNA abundances per transcript. Examples are the RPKM normalization for single-end reads, where RPKM stands for *reads per kilobase of exon per million mapped reads* (Pang et al. 2013), or the FPKM normalization for paired-end reads, where FPKM stands for *fragments per kilobase of exon per million mapped reads* (Trapnell et al. 2010).

Normalization

Subsequent to the quantification step, a normalization step is required for obtaining normalized values of relative RNA abundances that are comparable across different samples and different libraries. Raw counts obtained from count-based quantification approaches should not be compared to each other without normalization for reasons of different library sizes, different technical biases of library preparation, and different nucleotide compositions (Kvam et al. 2012; Rapaport et al. 2013).

A comparison of several normalization approaches published by Dillies et al. (2013) shows that RPKM normalization does typically not improve the results of count-based quantification approaches and should thus be replaced by *upper quartile normalization*, *median* normalization (Bolstad et al. 2003), *DESeq* normalization (Anders et al. 2012), or *TMM* normalization (Robinson et al. 2010). According to Dillies et al. (2013), the normalization tools *DESeq* and *TMM* yield the most robust results with respect to different library sizes and different library compositions.

Statistical Testing

Pipelines for the detection of differentially expressed transcripts require normalized values of relative RNA abundances in at least two groups of samples as input, then compute some test statistics for each of the transcripts from these normalized input data, and finally rank the transcripts by their computed test statistics.

Popular software packages for the detection of differentially expressed transcripts are *baySeq* (Hardcastle and Kelly 2010), *DESeq* (Anders and Huber 2010), *edgeR* (Robinson et al. 2010), and *PoissonSeq* (Li et al. 2012) (Table 14.4). Each of these software packages includes several normalization methods for which the statistical test is optimized. For example, the Bioconductor (Gentleman et al. 2004) packages *DESeq* and *edgeR* use a variation of Fisher's exact test adapted to the negative binomial (NB) distribution for calculating the significance of the change of the normalized relative RNA abundances between two conditions, whereas the software package *PoissonSeq* uses a statistical test based on the Poisson distribution. *DEXseq* (Anders et al. 2012) is a special software package devoted to the detection of differentially expressed exons, and a detailed description of *DESeq* and *edegR* is published by Anders et al. (2013).

Table 14.4 Selection of software packages for detection of differentially expressed transcripts

Tools	Input	Normalization	Model distribution	Test statistic	Differential expression
DESeq	Gene counts	DESeq normalization (normalization factor by median of scaled counts)	NB	Adapted Fisher's exact test for NB	Genes
edgeR	Gene counts	TMM (weighted trimmed mean of log expression ratios)	NB	Adapted two-sided Fisher's exact (binomial test)	Genes
baySeq	Gene counts	TMM, Quantile	NB	Empirical Bayes approach	Genes
PoissonSeq	Gene counts	Quantile, TMM	Poisson	Score statistic on the basis of a Poisson log-linear model	Genes
EBSeq	Isoform/ gene counts or RSEM-EBSeq pipeline	Median, Quantile	NB	Bayesian method: estimate posterior probability	Genes and isoform
DEXseq	Exon counts	DESeq	NB	NbinomTest	Exons
Cuffdiff 2	FPKM	FPKM	NB	*t* test-like statistics for FPKM	Genes and isoforms

The software packages are grouped by the input data from the quantification step (column 2), by the supported normalizations (column 3), by the statistical model for the normalized values of the relative RNA abundance (column 4), by the statistical test (column 5), and by the type of transcript that can be handled such as genes, exons, or isoforms

Soneson and Delorenzi (2013) compare 11 different software packages for detecting differentially expressed transcripts. They find that different software packages are optimal for different situations and that the resulting sets of differentially expressed genes can differ strongly between the software packages. As a result, the authors recommend using more than one software package for detecting differentially expressed transcripts.

An alternative approach for detecting differentially expressed transcripts is based on model-based methods for quantification and normalization. One popular example is the *Tuxedo suite* pipeline, which is composed of the mapping tool *TopHat 2*, the quantification and normalizing tool *Cufflinks*, and the tool *Cuffdiff 2* for the detection of differentially expressed genes or isoforms (Trapnell et al. 2013). A detailed description of the TopHat-Cufflinks pipeline has been published by Trapnell et al. (2010).

Detecting differentially expressed transcripts in cases where no reference genome or transcriptome is available, but where a de novo transcriptome assembly is performed for generating a reference transcriptome, is an important task for the analysis of RNA-seq data of horticultural plants and other non-model species. In this case, the software package *RSEM* is often used, which provides the RSEM-EBSeq pipeline and the tool *EBSeq* for the detection of differentially expressed transcripts (Leng et al. 2013) (Table 14.4).

14.2.2.2 SNP Calling

Calling single-nucleotide polymorphisms (SNPs) is a very important task in many RNA-seq experiments including studies on plants. SNP calling can be performed directly on the sorted output of mapped reads (Sect. 14.2.1). One limitation of calling and analyzing SNPs from RNA-seq data is that only SNPs in exonic regions can be detected. Three popular SNP callers are *Freebayes* (https://github.com/ekg/freebayes), *GATK* (DePristo et al. 2011), and *Samtool* (Li et al. 2009). An extended list of SNP callers can be found at http://seqanswers.com/wiki/SNP_discovery. The standard output format is the variant calling format (vcf), so most SNP callers write their output in vcf files. vcf files can then be filtered for significant SNPs by *vcftools* (Danecek et al. 2011), and their content can be visualized, for example, by the *Integrative Genomics Viewer* (IGV) (Thorvaldsdóttir et al. 2013).

14.2.3 Specific Steps of RNAseq Data Analysis of Small RNAs

Mapped reads from a pool of small RNAs can be used for addressing numerous tasks such as detecting differentially expressed small RNAs, predicting novel miRNAs, or predicting miRNA targets, and we address these tasks in Sects. 14.2.3.1–14.2.3.3. Detailed reviews of RNA-seq data analyses of small RNAs have been published by Motameny et al. (2010) and Gomes et al. (2013).

14.2.3.1 Detecting Differentially Expressed Small RNAs

One main task of the analysis of small RNA-seq data is the detection of differentially expressed miRNAs, siRNAs, snoRNAs, tRNA, or rRNAs. This task can be accomplished by the same steps as described in Sect. 14.2.2.1 with the only addition that the reference genome or transcriptome or the annotation should be related to the class of small RNAs to be analyzed. In case of miRNAs, such information is available from the database *miRBase* (Kozomara and Griffiths-Jones 2011), which allows using the miRNA hairpin structure as reference for mapping or using genome-related miRNA annotation files for quantification.

14.2.3.2 Predicting Novel Plant miRNAs by *plantDARIO*

The tool *plantDARIO* (http://plantdario.bioinf.uni-leipzig.de/), an extension of the tool *DARIO* (Fasold et al. 2011), can be used for the prediction of novel miRNAs, tRNAs, C/D-box snoRNAs, and H/ACA-box snoRNAs in plants. *PlantDario* is specifically tailored to plants and uses the tool *NOVOMIR* (Teune and Steger 2010) for the prediction of novel miRNAs and the tool *SnoReport* (Hertel et al. 2008) for predicting novel snoRNAs. *PlantDARIO* provides basic features including quality control, quantification, and normalization in addition to predicting novel small RNAs. Currently, *plantDARIO* allows analyses of small RNAs in *Arabidopsis thaliana*, *Beta vulgaris*, and *Solanum lycopersicum*.

14.2.3.3 Predicting miRNA Targets

The prediction of miRNA targets in mammals is reviewed by Witkos et al. (2011). For plants, however, there are only a few tools available for the prediction of miRNA targets. Two noteworthy examples are *psRNATarget* (Witkos et al. 2011) and *TAPIR* (Bonnet et al. 2010). Additionally, software packages for the analysis of miRNA such as *miRDeep-P* (Yang and Li 2011), *miREvo* (Wen et al. 2012), *MirTools 2.0* (Wu et al. 2013), *seqBuster* (Pantano et al. 2010), or *miRanalyzer* (Hackenberg et al. 2011) can be used for the prediction of plant miRNA targets. These software packages can contain additional tools for predicting differentially expressed miRNAs (Table 14.5, column 3), novel miRNAs (Table 14.5, column 6), isoforms (isomiRs) (Table 14.5, column 5), and point mutations (Motameny et al. 2010; Git et al. 2010).

Table 14.5 Selection of software packages for the analysis of miRNAs

Tool	Quantification	Differential expression	Prediction of miRNA targets	Prediction of miRNA isoforms	Prediction of novel miRNAs	Organism
miRanalyzer	Yes	Yes	Yes	Yes	Yes	A, P
MirTools 2.0	Yes	Yes	Yes	No	Yes	A, P
psRNATarget	Yes	No	Yes	No	No	P
TAPIR	Yes	No	Yes	No	No	P
miREvo	Yes	No	Yes	No	Yes	A, P
miRDeep-P	Yes	No	No	No	Yes	P
plantDARIO	Yes	No	No	No	Yes	P
SeqBuster	Yes	No	No	Yes	Yes	A, P

Several of these software packages are not tailored to plants. The software packages are grouped by their capability of quantifying relative miRNA abundance (column 2), of detecting differentially expressed miRNAs (column 3), of predicting miRNA targets (column 4), miRNA isoforms (column 5), and novel miRNAs (column 6), and of analyzing animal (A) and/or plant (P) miRNAs (column 7)

14.3 RNA-Seq Data on Abiotic Stresses of Horticultural Plants

Although abiotic stresses in model plants have been studied mainly by microarray techniques, nowadays we observe a dramatic growth of transcriptome profiling by RNA-seq experiments in non-model plants. Here we review recently published results on RNA-seq derived transcriptome data analyses for plants under abiotic stresses such as cold, drought, or salinity. One can classify the experiments in two groups: (i) the study of transcriptional response in horticultural plants and (ii) the study of transcriptomes of plants that are well adapted to abiotic stresses such as endemics or wild cultivars. Several studies have been also performed to investigate the molecular basis of adaptation under selection processes, for example, in wheat (Jia et al. 2013) and in tomato (Koenig et al. 2013). In some studies (Massa et al. 2013; O'Rourke et al. 2013), cross-species comparisons of stress-induced transcriptomes have uncovered differentially expressed orthologs and defined evolutionary conserved genes. Using the example of cold-responsive transcriptome studies, we show which bioinformatics methods researchers have used for transcriptome assembly and annotation and for the detection of differentially expressed genes (Table 14.6). For other abiotic stresses studies, we only give an overview of the RNA-seq data analysis tasks.

14.3.1 Cold

Recently, a number of studies have been published on the analysis of cold-responsive transcriptomes based on RNA-seq experiments. Two different questions that have been studied are (i) plant cold resistance and acclimation and (ii) the harmful effect of low temperatures. In the first case, the transcriptomes of plants known for their adaptability to cold have been studied. Examples are *Ammopiptanthus mongolicus*, an evergreen broadleaf legume shrub, distributed in Mid-Asia where the temperature can be as low as −30 °C during winter (Pang et al. 2013), the sheepgrass *Leymus chinensis*, an important perennial forage grass across the Eurasian Steppe (Chen et al. 2013), or the extremophile Antarctic hairgrass *Deschampsia antarctica*, the only natural grass species in the maritime Antarctic (Lee et al. 2013). In the second case, researchers studied the transcriptomes of tropical or other cold-sensitive plants useful in biotechnology or horticulture. Among them has been *Jatropha curcas* L., an oil-rich tropical shrub with multiple uses, including biodiesel production (Wang et al. 2013b), *Anthurium andraeanum*, one of the most popular tropical flowers (Tian et al. 2013), and the tea plant *Camellia sinensis* (Wang et al. 2013d).

In the studies, the transcriptomes have been assembled *de novo* (Table 14.6). Standard transcriptome annotation included BLAST alignments against the NCBI nonredundant (NR) database and the COG database as well as GO and KEGG annotations. However, in some works an extensive annotation has been done also by alignments against EST databases (Wang et al. 2013b, d; Lee et al. 2013), the

Table 14.6 Summary on materials and methods in studies of cold-induced plant transcriptomes

Species	Samples	Sequencing platform	Transcriptome assembly and detection of differentially expressed genes (DEG)[a]	Transcriptome annotation	Links
Jatropha curcas	Leaves 12, 24, 48 h at 12 °C	Illumina Hiseq 2000	*Assembly*: *Trinity platform* *DEG Detection*: Digital Gene Expression	*BlastX*: NR, wiss-Prot, EST *BlastN*: Nt GO in *map2slim*	SRR653198 (Wang et al. 2013b)
Anthurium andraeanum	Leaves 1, 2, 3, 10, 24 h at 6 °C	Illumina Hiseq 2000	*Assembly*: *SOAPdenovo* *DEG Detection*: Digital Gene Expression	*BlastX*: NR COG GO in *Blast2GO* KEGG pathways	E-MTAB-1955 (Tian et al. 2013)
Camellia sinensis	Leaves non-, fully-, and de-acclimated plants	Illumina Hiseq 2000	*Assembly*: *Trinity platform* *DEG Detection*: RSEM software package; DESeq package; Digital Gene expression	*BlastX*: NR, UniRef90, TAIR10, EST, COG *HMMER 3.0* *Blast2GO* KEGG pathways	SRA061043 (Wang et al. 2013d)
Ammopiptanthus mongolicus	Leaves and roots, 14 days at –4 °C	Illumina Hiseq 2000	*Assembly*: *Trinity platform*; TGICL[b] *DEG Detection*: RPKM method	*BlastX*: NR, Swiss-Prot, COG *Blast2GO*, WEGO KEGG pathways	SRA064010 (Pang et al. 2013)
Leymus chinensis	Buds –40; –15; 25 °C Spikes	454 GS FLX	*Assembly*: Newbler 2.5. *DEG Detection*: RPKM method; DESeq package	EMBOSS package *BlastP*: Swiss-Prot, NR, COG GO in *GoPipe* *KEGG* pathways *PlantGDB* for Poacea-Specific gene analysis	SRA065691 (Chen et al. 2013)
Deschampsia antarctica	Leaves at 4 °C	454 GS FLX	Assembly: TGICL[b] *DEG Detection*: Permutation *t* test after normalization of raw read counts	*BlastX*, *BlastN*: *Nr*, *tBlastX*: *EST* *Blast2GO*, *AgriGO* *KEGG pathways*	SRA051881 (Lee et al. 2013)

[a]Details about the programs and algorithms are described in Sect. 14.2
[b]From Pertea et al. (2003)

TAIR10 database (Wang et al. 2013d), or the PlantGDB database (Chen et al. 2013). For *Leymus chinensis* and *Deschampsia antarctica*, the transcriptome assembly allowed performing phylogenetic analyses (Chen et al. 2013; Lee et al. 2013). In each of these studies, thousands of genes were found to be differentially expressed under cold. Their functional annotation allowed revealing (i) pathways that were significantly affected under cold and (ii) cold-sensitive genes which were specific for the analyzed species. Specific attention has been paid to cold-sensitive transcription factors by Chen et al. (2013), Tian et al. (2013), and Wang et al. (2013d).

An interesting study of the cold-induced mRNA degradome in *Brachypodium distachyon* has been performed based on RNA-seq experiments and a parallel analysis of RNA ends (PARE) in (Zhang et al. 2013a). The authors identified specific patterns of mRNA decay in cold response. Uncapped transcripts changed significantly after cold treatment, whereas their transcript abundance remained unchanged. MiRNA-seq experiments of the similar samples showed some miRNA–mRNA pairs associated with cold response. In addition to miRNA-directed internal cleavage, the authors also revealed 90 transcripts that undergo an endogenous cleavage by an unknown mechanism through a specific and conserved motif.

14.3.2 Drought

RNA-seq analysis has been used for identifying genes that mediate the tolerance to water-limiting environments, which in the long term will contribute to improvement of plant productivity under drought. Transcriptome profiles under drought response have been analyzed in horticultural plants that are vulnerable to drought such as potato (*Solanum tubersosum*) (Massa et al. 2013; Zhang et al. 2014), rice (*Oriza sativa*) (Zong et al. 2013), and common beans (*Phaseolus vulgaris*) (Müller et al. 2013). Similarly, several drought-tolerant species such as *Agave deserti* and *Agave tequilana* have been studied based on RNA-seq experiments (Gross et al. 2013). Two varieties of quinoa (*Chenopodium quinoa* Willd.), the allotetraploid grain crop with an impressive drought tolerance and nutritional content, have been studied by RNA-seq (Raney et al. 2014).

To identify differentially expressed genes under drought stress in cotton, RNA-seq experiments have been performed in the tetraploid *Gossypium hirsutum* cotton (Bowman et al. 2013) and the diploid *Gossypium arboretum* cotton (Zhang et al. 2013c). RNA-seq experiments have also provided insight into transcriptional drought responsive in trees such as poplars (Cossu et al. 2013; Tang et al. 2013) and eucalyptus (Villar et al. 2011).

Drought stress during flowering and grain-filling stages of growth contributes to serious yield loss in common bean (Kakumanu et al. 2012; Müller et al. 2013), and the dehydration stress response of the transcriptome of *Chrysanthemum* have been studied by Xu et al. (2013b).

In some works, RNA-seq experiments have been performed for studying specific gene families that play an important role in drought response. Examples are studies

of expression patterns of ten AP2/EREB-like transcription factors in two soybean genotypes (Marcolino-Gomes et al. 2013) or the detection and analysis of LEA proteins in the tropical legume *Castanospermum austral* (Delahaie et al. 2013).

Comprehensive analyses of drought response mechanisms were performed for *Oryza sativa* by Zong et al. (2013). The authors have performed ChIP-seq and RNA-seq analyses for studying the relationships between epigenomic and transcriptional regulation in response to drought and have been found associations between the distribution pattern of histone H3K4-tri-methylation and gene expression profiles.

Drought- and salinity-responsive miRNAs have been analyzed for the emerging biofuel crop switchgrass (*Panicum virgatum*) by Xie et al. (2013). Differentially expressed miRNAs and their predicted targets have been functionally annotated and a number of interesting targets have been selected to aid in designing next-generation switchgrass for biomass and biofuel.

14.3.3 Heat and Light

In most cases, heat stress is associated with drought, but the study of heat stress defense mechanisms has a primary importance also for some of the cold-temperate species. For example, one of the most crucial factors that limits the cultivation of the Pacific Ocean kelp *Saccharina japonica* in China is its sensitivity to high temperature. The response of the *S. japonica* transcriptome to heat have been studied based on RNA-seq experiments (Liu et al. 2013), and the functional annotation of differentially expressed genes under primary heat response has showed that algae respond to heat stress by a complex network of genes rather than by a few specific stress-related genes.

The lack or excess of light can be also stressful for some plant species. An interesting comparison of the transcriptomes of an allotetraploid *Glycine* and its diploid progenitors has been published by Coate et al. (2013). Allopoliploidy is often associated with increased photosynthetic capacity as well as enhanced stress tolerance. In this work it has been shown that, under chronic excess of light, a photoprotective mechanism was higher in an allopolyploid *Glycine dolichocarpa* than in its diploid progenitors *G. tomentella* and *G. syndetika*.

14.3.4 Soil Pollutants

Toxic heavy metals in the soil can be absorbed and accumulated by plant roots, significantly suppressing their growth and making them a potential source for human health risks, especially in vegetables with edible roots. The radish response to lead stress has been studied based on RNA-seq experiments by Wang et al. (2013e). In this work, the radish transcriptome has been *de novo* assembled, and thousands of differentially expressed genes between control roots and roots

subjected to lead stress have been detected. Their functional annotation revealed that the upregulated genes have been predominately involved in defense responses in the cell wall and glutathione metabolism, whereas downregulated genes have been mainly related to carbohydrate metabolism pathways.

Al^{3+} tolerance mechanisms in rice roots have been studied by Arenhart et al. (2013). The central role of the ASR5 transcription factor for regulating the transcriptional response to Al^{3+} has been demonstrated by the fact that most of the genes differentially expressed under Al^{3+} stress have not been differentially expressed in plants with suppressed ASR5.

Boron-induced transcriptomes in barley have been studied by Tombuloglu et al. (2013), wherein the authors identified critical boron-induced transcription factors of the MYB family, which are well-known regulators of stress response. Boron-induced transcriptome changes in barley have been also studied based on miRNA-seq experiments by Ozhuner et al. (2013), whereby boron-induced miRNAs and their potential targets have been identified and partially validated by quantitative polymerase chain reaction (qPCR).

14.3.5 Mineral Deficiency

Response to deficiency of macro- and micronutrients has been also studied by RNA-seq experiments in plants. One example is the study of the response to potassium starvation in two watermelon genotypes (Fan et al. 2014).

The response to phosphate deficiency has been intensively studied by RNA-Seq experiments in plants. Plants utilize different morphological and physiological strategies to adapt to phosphate starvation in the soil. Studies of phosphate-deficient transcriptomes or the transcriptomes of plants highly tolerant to the lack of phosphate may elucidate the molecular basis of the response to phosphate starvation, and such studies have been performed by RNA-seq experiments in four rice cultivars (Oono et al. 2013). As a result, a set of core transcripts responsive to phosphate deficiency in the four rice cultivars has been identified.

The response to phosphate deficiency was studied in white lupin (O'Rourke et al. 2013), which has evolved unique adaptations for growth in phosphate-deficient soils, including the development of cluster roots to increase the root surface area. As a result, 12 genes have been found differentially expressed in response to phosphate deficiency in *Arabidopsis thaliana*, potato, and white lupin, making these genes ideal candidates to monitor the phosphate status of plants.

The expression of mRNA of ribosomal proteins has been studied based on RNA-seq experiments in phosphate- and iron-deficient plants of *A. thaliana* (Wang et al. 2013c), and three and 81 differentially expressed genes have been identified, respectively. At the protein level, many more ribosomal proteins were accumulated in response to phosphate than in response to iron, suggesting that phosphate and iron starvation provoke an altered composition of ribosomes and a biased translation, which can be an important mechanism of adaptation to changing environmental conditions.

14.3.6 Salinity

Salinity is one of the major abiotic factors affecting productivity of horticultural plants. Recently RNA-seq experiments have been performed for studying the response to salinity stress in lucerne (alfalfa) (Postnikova et al. 2013), potato (Massa et al. 2013), and barley (Ziemann et al. 2013). Among cereal crops, barley is considered as notably salt tolerant, but an interesting task is to analyze the transcriptomes of halophytic plants. For example, global transcriptome profiling has been performed in *Salicornia europaea*, an edible plant well adapted to extreme saline environments (Ma et al. 2013). Among other halophytic plant species, RNA-seq experiments have been also performed on the desert poplar *Populus euphratica* (Tang et al. 2013) and the Inner Mongolia endemic shrub *Reaumuria trigyna*, which has unique morphological characteristics that allow it to tolerate stress imposed by semidesert saline soil (Dang et al. 2013).

Salinity stress has been also studied in wild halophyte relatives of horticultural plants such as *Porteresia coarctata,* a wild rice that is capable of tolerating high salinity and submergence (Garg et al. 2014), and *Gossypium* species with a remarkable tolerance to saltwater immersion such as *Gossypium aridum* (Xu et al. 2013a).

Effects of saline-alkaline soils have been studied by RNA-seq experiments in several plant species. Time series of transcriptomes of roots of the halophyte wood *Tamarix hispida* stressed by $NaHCO_3$ have been studied by Wang et al. (2013a). Early transcriptomic adaptation to sodium carbonate in maize has been studied by Zhang et al. (2013b), where the authors analyzed shared and distinctive targets in $Na_2CO_3^-$, $NaCl^-$, and high-pH-induced transcriptomes.

Acknowledgments We thank A.V. Kochetov, I. Lemnian, and N.A. Omelyanchuk for fruitful discussions and Dynasty Foundation (grant for young biologists), DFG (grant no. GR 3523/2), RAS program 6.6, and RFBR Foundation (grant no. 12-04-33112) for financial support.

References

Alamancos GP, Agirre E, Eyras E (2014) Methods to study splicing from high-throughput RNA sequencing data. Methods Mol Biol 1126:357–397. doi:10.1007/978-1-62703-980-2_26

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

Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. Genome Res 22:2008–2017. doi:10.1101/gr.133744.111

Anders S, McCarthy DJ, Chen Y et al (2013) Count-based differential expression analysis of RNA sequencing data using R and bioconductor. Nat Protoc 8:1765–1786. doi:10.1038/nprot.2013.099

Anders S, Pyl PT, Huber W (2014) HTSeq A Python framework to work with high-throughput sequencing data. BioRxiv doi:10.1101/002824

Arenhart RA, de Lima JC, Pedron M et al (2013) Involvement of ASR genes in aluminium tolerance mechanisms in rice. Plant Cell Environ 36:52–67. doi:10.1111/j.1365-3040.2012.02553.x

Birol I, Jackman SD, Nielsen CB et al (2009) De novo transcriptome assembly with ABySS. Bioinformatics 25:2872–2877. doi:10.1093/bioinformatics/btp367

Boley N, Stoiber MH, Booth BW et al (2014) Genome-guided transcript assembly by integrative analysis of RNA sequence data. Nat Biotechnol 32:341–346. doi:10.1038/nbt.2850

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193

Bonnet E, He Y, Billiau K, Van de Peer Y (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. Bioinformatics 26:1566–1568. doi:10.1093/bioinformatics/btq233

Bowman MJ, Park W, Bauer PJ et al (2013) RNA-Seq transcriptome profiling of upland cotton (*Gossypium hirsutum* L.) root tissue under water-deficit stress. PLoS One 8:e82634

Chen S, Huang X, Yan X et al (2013) Transcriptome analysis in sheepgrass (*Leymus chinensis*): a dominant perennial grass of the Eurasian Steppe. PLoS One 8:e67974. doi:10.1371/journal.pone.0067974

Coate JE, Powell AF, Owens TG, Doyle JJ (2013) Transgressive physiological and transcriptomic responses to light stress in allopolyploid *Glycine dolichocarpa* (Leguminosae). Heredity (Edinb) 110:160–170. doi:10.1038/hdy.2012.77

Cossu RM, Giordani T, Cavallini A, Natali L (2013) High-throughput analysis of transcriptome variation during water deficit in a poplar hybrid: a general overview. Tree Genet Genomes 10:53–66. doi:10.1007/s11295-013-0661-5

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCF tools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Dang Z, Zheng L, Wang J et al (2013) Transcriptomic profiling of the salt-stress response in the wild recretohalophyte *Reaumuria trigyna*. BMC Genomics 14:29 (doi:10.1186/1471-2164-14-29)

Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM (2013) An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS One 8:e85024. doi:10.1371/journal.pone.0085024

Delahaie J, Hundertmark M, Bove J et al (2013) LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. J Exp Bot 64:4559–4573. doi:10.1093/jxb/ert274

DePristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. doi:10.1038/ng.806

Dillies M-A, Rau A, Aubert J et al (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform 14:671–683. doi:10.1093/bib/bbs046

Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. doi:10.1093/bioinformatics/bts635

Fan M, Huang Y, Zhong Y et al (2014) Comparative transcriptome profiling of potassium starvation responsiveness in two contrasting watermelon genotypes. Planta (Berl) 239:397–410. doi:10.1007/s00425-013-1976-z

Fasold M, Langenberger D, Binder H et al (2011) DARIO: a ncRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 39:W112–W117. doi:10.1093/nar/gkr357

Fonseca NA, Rung J, Brazma A, Marioni JC (2012) Tools for mapping high-throughput sequencing data. Bioinformatics 28:3169–3177. doi:10.1093/bioinformatics/bts605

Garg R, Verma M, Agrawal S et al (2014) Deep transcriptome sequencing of wild halophyte rice, *Porteresia coarctata*, provides novel insights into the salinity and submergence tolerance factors. DNA Res 21:69–84. doi:10.1093/dnares/dst042

Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80. doi:10.1186/gb-2004-5-10-r80

Git A, Dvinge H, Salmon-Divon M et al (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA 16:991–1006. doi:10.1261/rna.1947110

Gomes CPC, Cho J-H, Hood L et al (2013) A review of computational tools in microRNA discovery. Front Genet 4:81. doi:10.3389/fgene.2013.00081

Gross SM, Martin JA, Simpson J et al (2013) De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. BMC Genomics 14:563. doi:10.1186/1471-2164-14-563

Guttman M, Garber M, Levin JZ et al (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28:503–510. doi:10.1038/nbt.1633

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. doi:10.1038/nprot.2013.084

Hackenberg M, Rodríguez-Ezpeleta N, Aransay AM (2011) miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res 39:W132–W138. doi:10.1093/nar/gkr247

Hardcastle TJ, Kelly KA (2010) BaySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422. doi:10.1186/1471-2105-11-422

Hertel J, Hofacker IL, Stadler PF (2008) SnoReport: computational identification of snoRNAs with unknown targets. Bioinformatics 24:158–164. doi:10.1093/bioinformatics/btm464

Hoffmann S, Otto C, Doose G et al (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. Genome Biol 15:R34. doi:10.1186/gb-2014-15-2-r34

Jean G, Kahles A, Sreedharan VT et al (2010) RNA-Seq read alignments with PALMapper. Curr Protoc Bioinformatics. doi:10.1002/0471250953.bi1106s32

Jia J, Zhao S, Kong X et al (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature (Lond) 496:91–95. doi:10.1038/nature12028

Kakumanu A, Ambavaram MMR, Klumas C et al (2012) Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant Physiol 160:846–867. doi:10.1104/pp. 112.200444

Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27:757–763. doi:10.1093/bioinformatics/btr010

Koenig D, Jiménez-Gómez JM, Kimura S et al (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci USA 110:E2655–E2662. doi:10.1073/pnas.1309606110

Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39:D152–D157. doi:10.1093/nar/gkq1027

Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am J Bot 99:248–256. doi:10.3732/ajb.1100340

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi:10.1038/nmeth.1923

Lawrence M, Huber W, Pagès H et al (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118. doi:10.1371/journal.pcbi.1003118

Lee LW, Zhang S, Etheridge A et al (2010) Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA 16:2170–2180. doi:10.1261/rna.2225110

Lee J, Noh EK, Choi H-S et al (2013) Transcriptome sequencing of the Antarctic vascular plant *Deschampsia antarctica* Desv. under abiotic stress. Planta (Berl) 237:823–836. doi:10.1007/s00425-012-1797-5

Leng N, Dawson JA, Thomson JA et al (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29:1035–1043. doi:10.1093/bioinformatics/btt087

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. doi:10.1186/1471-2105-12-323

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

Li H, Homer N (2010) A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform 11:473–483. doi:10.1093/bib/bbq015

Li H, Ruan J, Durbin R (2008a) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18:1851–1858. doi:10.1101/gr.078212.108

Li R, Li Y, Kristiansen K, Wang J (2008b) SOAP: short oligonucleotide alignment program. Bioinformatics 24:713–714. doi:10.1093/bioinformatics/btn025

Li J, Witten DM, Johnstone IM, Tibshirani R (2012) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. Biostatistics 13:523–538. doi:10.1093/biostatistics/kxr031

Liao Y, Smyth GK, Shi W (2014) Feature counts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. doi:10.1093/bioinformatics/btt656

Liu F, Wang W, Sun X et al (2013) RNA-Seq revealed complex response to heat stress on transcriptomic level in *Saccharina japonica* (Laminariales, Phaeophyta). J Appl Phycol. doi:10.1007/s10811-013-0188-z

Lohse M, Bolger AM, Nagel A et al (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res 40:W622–W627. doi:10.1093/nar/gks540

Luo R, Liu B, Xie Y et al (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18. doi:10.1186/2047-217X-1-18

Ma J, Zhang M, Xiao X et al (2013) Global transcriptome profiling of *Salicornia europaea* L. shoots under NaCl treatment. PLoS One 8:e65877

Marcolino-Gomes J, Rodrigues FA, Oliveira MCN et al (2013) Expression patterns of GmAP2/EREB-like transcription factors involved in soybean responses to water deficit. PLoS One 8:e62294. doi:10.1371/journal.pone.0062294

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10. doi:10.14806/ej.17.1.200

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. doi:10.1038/nrg3068

Martin J, Bruno VM, Fang Z et al (2010) Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics 11:663. doi:10.1186/1471-2164-11-663

Massa AN, Childs KL, Buell CR (2013) Abiotic and biotic stress responses in group Phureja DM1-3 516 R44 as measured through whole transcriptome sequencing. Plant Genome 6:1–10. doi:10.3835/plantgenome2013.05.0014

McGettigan PA (2013) Transcriptomics in the RNA-seq era. Curr Opin Chem Biol 17:4–11. doi:10.1016/j.cbpa.2012.12.008

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46. doi:10.1038/nrg2626

Motameny S, Wolters S, Nürnberg P, Schumacher B (2010) Next generation sequencing of miRNAs: strategies, resources and methods. Genes (Basel) 1:70–84. doi:10.3390/genes1010070

Müller BSDF, Sakamoto T, Silveira RDD et al. (2013) Differentially expressed genes during flowering and grain filling in common bean (*Phaseolus vulgaris*) grown under drought stress conditions. Plant Mol Biol Rep 438–451. doi:10.1007/s11105-013-0651-7

O'Rourke JA, Yang SS, Miller SS et al (2013) An RNA-Seq transcriptome analysis of orthophosphate-deficient white lupin reveals novel insights into phosphorus acclimation in plants. Plant Physiol 161:705–724. doi:10.1104/pp. 112.209254

Oono Y, Kawahara Y, Yazawa T et al (2013) Diversity in the complexity of phosphate starvation transcriptomes among rice cultivars based on RNA-Seq profiles. Plant Mol Biol 83:523–537. doi:10.1007/s11103-013-0106-4

Ozhuner E, Eldem V, Ipek A et al (2013) Boron stress responsive microRNAs and their targets in barley. PLoS One 8:e59543. doi:10.1371/journal.pone.0059543

Pang T, Ye C-Y, Xia X, Yin W (2013) De novo sequencing and transcriptome analysis of the desert shrub, *Ammopiptanthus mongolicus*, during cold acclimation using Illumina/Solexa. BMC Genomics 14:488. doi:10.1186/1471-2164-14-488

Pantano L, Estivill X, Martí E (2010) SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res 38:e34. doi:10.1093/nar/gkp1127

Pertea G, Huang X, Liang F et al (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652. doi:10.1093/bioinformatics/btg034

Postnikova OA, Shao J, Nemchinov LG (2013) Analysis of the alfalfa root transcriptome in response to salinity stress. Plant Cell Physiol 54:1041–1055. doi:10.1093/pcp/pct056

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. doi:10.1093/bioinformatics/btq033

Raney JA, Reynolds DJ, Elzinga DB et al (2014) Transcriptome analysis of drought-induced stress in *Chenopodium quinoa*. Am J Plant Sci 2014:338–357

Rapaport F, Khanin R, Liang Y et al (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol 14:R95. doi:10.1186/gb-2013-14-9-r95

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. doi:10.1093/bioinformatics/btp616

Rumble SM, Lacroute P, Dalca AV et al (2009) SHRiMP: accurate mapping of short color-space reads. PLoS Comput Biol 5:e1000386. doi:10.1371/journal.pcbi.1000386

Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6:e17288. doi:10.1371/journal.pone.0017288

Schweikert G, Zien A, Zeller G et al (2009) mGene: accurate SVM-based gene finding with an application to nematode genomes. Genome Res 19:2133–2143. doi:10.1101/gr.090597.108

Silva GG, Dutilh BE, Matthews TD et al (2013) Combining de novo and reference-guided assembly with scaffold_builder. Source Code Biol Med 8:23. doi:10.1186/1751-0473-8-23

Simpson JT, Wong K, Jackman SD et al (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123. doi:10.1101/gr.089532.108

Smyth G (2005) Limma: linear models for microarray data. Bioinform Comput Biol Sol R Bioconductor. doi:10.1007/0-387-29362-0

Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14:91. doi:10.1186/1471-2105-14-91

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637–644. doi:10.1093/bioinformatics/btn013

Steijger T, Abril JF, Engström PG et al (2013) Assessment of transcript reconstruction methods for RNA-seq. Nat Methods 10:1177–1184. doi:10.1038/nmeth.2714

Tang S, Liang H, Yan D et al (2013) *Populus euphratica:* the transcriptomic response to drought stress. Plant Mol Biol 83:539–557. doi:10.1007/s11103-013-0107-3

Teune J-H, Steger G (2010) NOVOMIR: de novo prediction of microRNA-coding regions in a single plant-genome. J Nucleic Acids. doi:10.4061/2010/495904

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. doi:10.1093/bib/bbs017

Tian D-Q, Pan X-Y, Yu Y-M et al (2013) De novo characterization of the *Anthurium* transcriptome and analysis of its digital gene expression under cold stress. BMC Genomics 14:827. doi:10.1186/1471-2164-14-827

Tombuloglu H, Kekec G, Sakcali MS, Unver T (2013) Transcriptome-wide identification of R2R3-MYB transcription factors in barley with their boron responsive expression analysis. Mol Genet Genomics 288:141–155. doi:10.1007/s00438-013-0740-1

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111. doi:10.1093/bioinformatics/btp120

Trapnell C, Williams BA, Pertea G et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515. doi:10.1038/nbt.1621

Trapnell C, Hendrickson DG, Sauvageau M et al (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31:46–53. doi:10.1038/nbt.2450

Villar E, Klopp C, Noirot C et al (2011) RNA-Seq reveals genotype-specific molecular responses to water deficit in eucalyptus. BMC Genomics 12:538. doi:10.1186/1471-2164-12-538

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wang C, Gao C, Wang L, et al. (2013a) Comprehensive transcriptional profiling of $NaHCO_3$-stressed *Tamarix hispida* roots reveals networks of responsive genes. Plant Mol Biol 84(1-2): 145–157. doi: 10.1007/s11103-013-0124-2

Wang H, Zou Z, Wang S, Gong M (2013b) Global analysis of transcriptome responses and gene expression profiles to cold stress of *Jatropha curcas* L. PLoS One 8:e82817. doi:10.1371/journal.pone.0082817

Wang J, Lan P, Gao H et al (2013c) Expression changes of ribosomal proteins in phosphate- and iron-deficient *Arabidopsis* roots predict stress-specific alterations in ribosome composition. BMC Genomics 14:783. doi:10.1186/1471-2164-14-783

Wang X-C, Zhao Q-Y, Ma C-L et al (2013d) Global transcriptome profiles of *Camellia sinensis* during cold acclimation. BMC Genomics 14:415. doi:10.1186/1471-2164-14-415

Wang Y, Xu L, Chen Y et al (2013e) Transcriptome profiling of radish (*Raphanus sativus* L.) root and identification of genes involved in response to lead (Pb) stress with next generation sequencing. PLoS One 8:e66539

Wen M, Shen Y, Shi S, Tang T (2012) MiREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments. BMC Bioinformatics 13:140. doi:10.1186/1471-2105-13-140

Witkos TM, Koscianska E, Krzyzosiak WJ (2011) Practical aspects of microRNA target prediction. Curr Mol Med 11:93–109

Wu J, Liu Q, Wang X et al (2013) mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. RNA Biol 10:1087–1092. doi:10.4161/rna.25193

Xie F, Stewart CN, Taki FA et al (2013) High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. Plant Biotechnol J 159:1–13. doi:10.1111/pbi.12142

Xie Y, Wu G, Tang J et al (2014) SOAP denovo-trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666. doi:10.1093/bioinformatics/btu077

Xu P, Liu Z, Fan X et al (2013a) De novo transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress. Gene (Amst) 525:26–34. doi:10.1016/j.gene.2013.04.066

Xu Y, Gao S, Yang Y et al (2013b) Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. BMC Genomics 14:662. doi:10.1186/1471-2164-14-662

Yang X, Li L (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics 27:2614–2615. doi:10.1093/bioinformatics/btr430

Zavolan M, Kondo S, Schonbach C et al (2003) Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. Genome Res 13:1290–1300. doi:10.1101/gr.1017303

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. doi:10.1101/gr.074492.107

Zhang J, Mao Z, Chong K (2013a) A global profiling of uncapped mRNAs under cold stress reveals specific decay patterns and endonucleolytic cleavages in *Brachypodium distachyon*. Genome Biol 14:R92. doi:10.1186/gb-2013-14-8-r92

Zhang L-M, Liu X-G, Qu X-N et al (2013b) Early transcriptomic adaptation to Na_2CO_3 stress altered the expression of a quarter of the total genes in the maize genome and exhibited shared and distinctive profiles with NaCl and high pH stresses. J Integr Plant Biol 55:1147–1165. doi:10.1111/jipb.12100

Zhang X, Yao D, Wang Q et al (2013c) mRNA-seq analysis of the *Gossypium arboreum* transcriptome reveals tissue selective signaling in response to water stress during seedling stage. PLoS One 8:e54762. doi:10.1371/journal.pone.0054762

Zhang N, Liu B, Ma C et al (2014) Transcriptome characterization and sequencing-based identification of drought-responsive genes in potato. Mol Biol Rep 41:505–517. doi:10.1007/s11033-013-2886-7

Ziemann M, Kamboj A, Hove RM et al (2013) Analysis of the barley leaf transcriptome under salinity stress using mRNA-Seq. Acta Physiol Plant 35:1915–1924. doi:10.1007/s11738-013-1230-0

Zong W, Zhong X, You J, Xiong L (2013) Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress. Plant Mol Biol 81:175–188. doi:10.1007/s11103-012-9990-2