Dual RNA-seq to Elucidate the Plant–Pathogen Duel

Sanushka Naidoo[1]*, Erik Andrei Visser[1], Lizahn Zwart[1], Yves du Toit[1], Vijai Bhadauria[2] and Louise Simone Shuey[1]*

[1]Department of Genetics, Forestry and Agricultural Biotechnology Institute, Genomics Research Institute, University of Pretoria, Pretoria, South Africa.

[2]Crop Development Centre and Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada.

*Correspondence: Sanushka.Naidoo@fabi.up.ac.za and Louise.Shuey@fabi.up.ac.za

https://doi.org/10.21775/cimb.027.127

Abstract

RNA-sequencing technology has been widely adopted to investigate host responses during infection with pathogens. Dual RNA-sequencing (RNA-seq) allows the simultaneous capture of pathogen-specific transcripts during infection, providing a more complete view of the interaction. In this review, we focus on the design of dual RNA-seq experiments and the application of downstream data analysis to gain biological insight into both sides of the interaction. Recent literature in this area demonstrates the power of the dual RNA-seq approach and shows that it is not limited to model systems where genomic resources are available. Sequencing costs continue to decrease and single cell transcriptomics is becoming more feasible. In combination with proteomics and metabolomics studies, these technological advances are likely to contribute to our understanding of the temporal and spatial aspects of dynamic plant–pathogen interactions.

A dual approach *in planta*

The interaction between plants and pathogens is an active and dynamic process that can be likened to a duel. Plants have complex defence mechanisms that can be rendered ineffective when pathogens interfere with one of the various processes required for host defence. These processes include penetration resistance, recognition by Pattern Recognition Receptors (PRRs), phytohormone signalling pathways, secretory pathways, secondary metabolite production, and plant cell death (Dou and Zhou, 2012). Until recently, transcriptomic approaches have been applied in the host and pathogen separately to obtain the gene expression profile of each organism and gain insight into infection biology or host defence mechanisms.

RNA sequencing (RNA-seq) is a powerful technology that does not rely on any prior knowledge of transcripts and can generate vast quantities of data with much smaller costs involved than for older techniques such as microarrays (Pareek *et al.*, 2011; Wilhelm and Landry, 2009). An advantage of RNA-seq in the field of plant–pathogen interactions is that both plant and pathogen transcripts can be detected simultaneously and accurately in the same sample. This tactic, known as dual RNA-seq, *in planta* RNA-seq, simultaneous RNA-seq, or comparative RNA-seq, is a relatively new technique both in the plant and medical fields. In plants, it allows for the study of plant–pathogen interactions in herbaceous crops (Chen *et al.*, 2013; Kunjeti *et al.*, 2012; Lowe *et al.*, 2014) as well as trees (Hayden *et al.*, 2014; Liang *et al.*, 2014; Teixeira *et al.*, 2014). This review outlines technical considerations for dual RNA-seq experiments, summarizes recent insights drawn from such approaches in plant–pathogen interactions, and provides an

overview of the next generation of dual approaches. Since this technique is most useful to study interactions with pathogens with complex prokaryotic and eukaryotic genomes, viral pathogens have not been included in this review.

It's all in the design

Experimental design considerations for a dual RNA-seq experiment can be divided into three broad categories: sample generation, data generation and data analysis. An overview of the process can be found in Fig. 8.1.

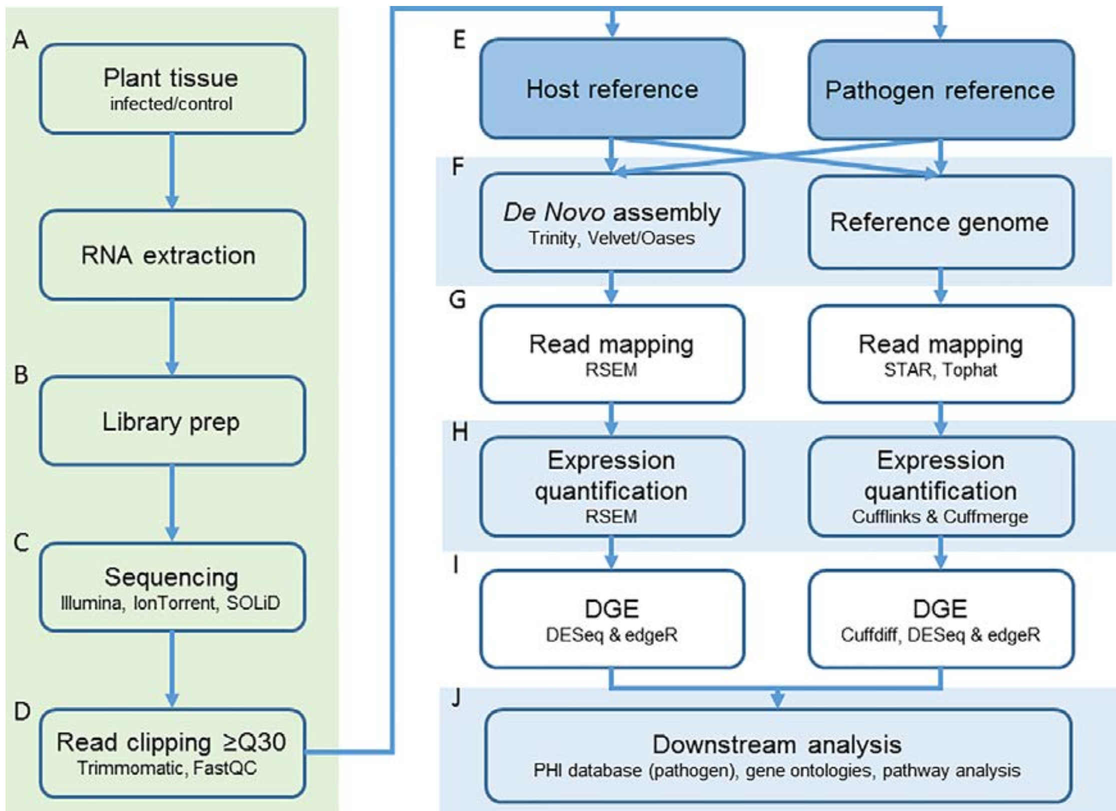

Figure 8.1 Flow chart of a dual RNA-seq experiment with example software programs. (A) Experimental design considerations can include comparisons between resistant and susceptible interactions, normalized to a mock inoculated control. (B) Different library preparation options include enrichments for mRNA, small RNAs, stranded RNA and total RNA. (C) The sequencing platform can vary based on availability and aim of the study. Paired-end sequencing on the Illumina platform is a common approach for RNA-seq. Deep sequencing is required for dual RNA-seq approaches. Read length can vary depending on the application and sequencing platform used. (D) Downstream read quality control can be implemented. Filtering for a minimum Phred quality score of Q30 is generally optimal, but the threshold is data dependent. (E) Dual RNA-seq is performed by mapping to both the host and pathogen reference genome sequences, or to the host first with the remaining reads mapped to the pathogen reference. Endophyte contamination can be removed by mapping to common contaminant sequences obtained from databases such as Refseq. (F) If a reference is not available, a *de novo* transcriptome can be assembled from the reads. (G) The mapping approach will differ based on the type of reference used and different methods can be used when mapping to the host or the pathogen. (H) Other programs commonly used for expression quantification include HTSeq, featureCounts (Liao *et al*., 2014) and Limma (Ritchie *et al*., 2015). (I) Differential Gene Expression (DGE) analysis can be performed using a number of methods. The two examples listed here can be used for both transcriptome and genome-based DGE analysis. (J) Genes identified as differentially expressed can be used in programs and databases such as BinGO (Maere *et al*., 2005), MapMan (Thimm *et al*., 2004) and Kegg (Ogata *et al*., 1999) to predict biological significance (Maere *et al*., 2005).

Sample generation

When considering experimental design for sample generation, the main factors include trial design, sample harvesting approach and sample handling (reviewed in Yang and Wei, 2015).

Two important trial design and sample harvesting considerations for dual RNA-seq experiments are the predicted gene number in the pathogen and host genomes, and the relative amounts of pathogen and host cells within a given sample (Westermann *et al.*, 2012). Both of these factors influence the amount of pathogen RNA relative to host RNA within a sample. A lower ratio of pathogen to host RNA requires greater sequencing depth to capture the full extent of biological variation within the pathogen.

A dual RNA-seq experiment considering an interaction between a eukaryotic host and prokaryotic pathogen requires approximately 10 to 20 times as many reads than would usually be required. This is partly due to the smaller amount of cellular RNA within prokaryotic cells relative to eukaryotes (Westermann *et al.*, 2012). While the relative amounts of cellular RNA between host and pathogen are more similar in eukaryote–eukaryote interactions, the higher read coverage is still necessary due to the lower quantity of pathogen versus host cells, which results in less pathogen RNA per sample.

An important trial design consideration specific to dual RNA-seq experiments is the inclusion of a control for pathogen gene expression. This can be done by comparing *in planta* expression of pathogen genes to i*n planta* gene expression of a non-pathogenic strain and/or pathogen gene expression in an agar culture or spore suspension (Kawahara *et al.*, 2012). Synthetic RNA spike-ins can also be included to quantify both pathogen and host RNA (Box 8.1).

Data generation

The main experimental design considerations for data generation include the level of sample replication, library construction and sequencing (Liu *et al.*, 2014).

Sample replication

One of the first factors to consider in experimental design is the level of sample replication (Auer and Doerge, 2010). Sample replication is divided into technical replication, which is defined as performing the same analysis multiple times on the same sample, and biological replication, a study dependent term that can be loosely defined as harvesting the same type of sample from the same type of organism from the same conditions.

Technical variation arises when errors occur in the experimental procedure and can be accounted for through technical replication. Illumina sequencing produces negligible technical variability, removing the need for technical replication in RNA-seq experiments (Marioni *et al.*, 2008). However, when coverage is low for certain transcripts, technical variation can still arise (McIntyre *et al.*, 2011). Thus, technical replication should be considered for dual RNA-seq experiments where there is low representation of pathogen RNA within a sample, resulting in low coverage of pathogen transcripts.

Box 8.1 Total RNA quantification

It is not always possible to accurately predict the amounts of host and pathogen RNA that will be present in a sample. While it is possible to measure the amount of host and pathogen DNA in a sample using qRT-PCR, this is not always an accurate reflection of total host and pathogen RNA. This problem can be circumvented by the addition of RNA spike-ins to samples. An RNA spike-in for RNA-seq is a mixture of synthesized RNA transcripts of known sequence, concentration and abundance. While inclusion of an RNA spike-in could increase the cost of sequencing due to increased coverage requirements, it can be used to measure sensitivity and accuracy of sequencing as well as to detect biases that can occur during RNA-seq (Jiang *et al*., 2011). Furthermore, standard curves can be generated from RNA spike-ins. This allows for more accurate quantification of transcript abundance (Jiang *et al*., 2011). In dual RNA-seq experiments, it is possible to use these standard curves to estimate host and pathogen RNA levels within a sample. However, it is important to ensure that none of the spike-in sequences are present in the genome of either host or pathogen, as this could preclude accurate quantification of genes containing similar sequences and the use of those spike-in sequences (Jiang *et al*., 2011).

While technical replication can be excluded due to reliability of the technology, biological replication remains crucial to all RNA-seq experiments. Besides accounting for biological variation (Hansen *et al.*, 2011; Nettleton, 2014), biological replication significantly affects the power and accuracy of differential expression analyses. Liu *et al.* (2014) showed that increasing the number of biological replicates sequenced increased the number of accurately identified differentially expressed genes, whereas increased read depth produced diminishing returns for both statistical power and the precision with which differential expression is detected. This is especially important in dual RNA-seq experiments where biological variation is introduced from both pathogen and host.

Library construction and sequencing

The main factors to consider during library construction and sequencing are depletion methods, strandedness, insert size, read length, and read depth. The use of strand-specific rather than non-strand-specific libraries [reviewed in Levin *et al.* (2010)], allows the accurate detection of anti-sense transcription and can allow accurate expression quantification of overlapping transcripts. Thus, strand-specific sequencing in dual RNA-seq experiments could enable detection of evidence suggesting host–pathogen interaction through anti-sense transcription.

The choice of insert size is dependent on the complexity of the transcriptome and the target RNA species (reviewed in Head *et al.*, 2014). Insert size selection can be a limiting factor in which RNA species can be analysed because inclusion of a size selection step during library preparation results in loss of transcripts shorter than the selected insert size. Insert size selection also imposes an upper limit on read length, since reads longer than the insert size will sequence into adapters, providing no new information.

Apart from insert size, the choice of read length is dependent mainly on the objectives of the study and the quality of the reference sequence used for mapping. If a high quality and well annotated reference sequence is available, increasing read length above 50 bp is unnecessary for accurate detection of differential expression (Chhangawala *et al.*, 2015). Similarly, sequencing of paired-end instead of single-end reads does not significantly affect detection of differential expression in these cases (Chhangawala *et al.*, 2015). Conversely, when studying organisms with less well defined reference sequences, sequencing of longer paired-end reads increases the accuracy of splice junction detection (Chhangawala *et al.*, 2015). When no reference sequence is available, it is often assumed that longer reads equate to increased accuracy for *de novo* assembly. Similar to the detection of differential expression, however, there seems to be a species-specific threshold beyond which increasing read length becomes redundant (Chang *et al.*, 2014).

In cases where a high quality reference is available, less coverage is required for accurate transcript identification and quantification, compared to cases where a reference is missing. This is because gaps in an assembly arising from low coverage can be filled using the underlying reference sequence. For studies relying on *de novo* assembly, a predicted minimum of 30× total reference coverage is required (Martin and Wang, 2011), while genome-guided assembly can be accomplished with coverage below 10× (Denoeud *et al.*, 2008).

For an RNA-seq experiment to be representative, it is important to make sure that the number of reads is sufficient to account for the least represented RNA species. This is also referred to as sufficient sequencing depth. To obtain adequate depth for a dual RNA-seq experiment, enough reads need to be sequenced to have at least 1× coverage of the least represented pathogen transcript in the sample with the lowest level of pathogen to host RNA. However, it is almost impossible to know this information when performing *de novo* RNA-seq experiments.

Techniques to deplete or enrich certain RNA species, such as RNA fractionation and poly(A) selection, can enhance detection of transcripts with low expression in eukaryotes (Sims *et al.*, 2014). Depletion of the rRNA fraction can further reduce the required sequencing depth of an experiment and, unlike poly(A) selection, allow for detection of non-poly(A) transcripts. Although depletion-based methods allow for selection of non-poly(A) RNA species, these methods can bias quantification of abundant transcripts and decrease exon coverage and power to detect splice junctions due to the presence of sequenced introns from pre-mRNA in eukaryotes (Martin and Wang, 2011; Sims *et al.*, 2014).

Data analysis

As with sample and data generation, data analysis considerations are dependent on the underlying biological questions. Data analysis for the majority of RNA-seq experiments follows three sequential steps: (1) quality control, (2) mapping, expression quantification and DE analysis, and (3) downstream analysis. Due to the variety of tools and platforms that can be used for RNA-seq data analysis (Grant *et al.*, 2011), programs typically used for RNA-seq data analysis may be created for specific analysis types and tested within a specific experimental context. Thus, it is often advisable to repeat an analysis using different programs and compare the outputs.

Quality control

Quality control for dual RNA-seq studies is similar to that used for traditional RNA-seq studies. However, contaminant filtering becomes more complicated as reads from both host and pathogen need to be retained. While reads originating from the host and pathogen can be separated by mapping to the host and pathogen reference sequences (Schulze *et al.*, 2015; Westermann *et al.*, 2012), contamination of various forms should be considered in order to improve the accuracy and efficiency with which genes and transcripts are mapped and quantified. Contamination may occur in two main forms: non-mRNA species (which constitute the majority of the total RNA extracted) and reads representing mRNA extracted from organisms (saprophytes and endophytes) other than the organisms of interest. These forms of contamination may skew the quantification of genes and transcripts when assembling and mapping reads to the reference. Westermann *et al.* (2012) provide insight into dealing with contaminating RNA which is species and study dependent.

Contamination in the form of RNA extracted from endophytic or saprophytic organisms is especially important in plant–pathogen interaction studies. Saprophytes may be present at the sites of wounding due to the degradation of tissue that occurs, while endophytes colonize areas below the surface of the plant tissue without causing symptoms. Thus RNA from these types of organisms can be present in RNA-seq libraries. While surface sterilization could be used to decrease the presence of saprophytes, the process is time consuming and may result in damage to host RNA. Surface sterilization could also result in decreased pathogen representation, which is counterproductive for a dual RNA-seq experiment. Therefore, removal of these contaminating sequences requires bioinformatics intervention. This can be accomplished through stringent mapping of data to a database of common contaminant cDNA sequences constructed from databases such as RefSeq, UniRef100 and GenBank (Ikeue *et al.*, 2015). In cases where reference genomes are available for known endophytes and saprophytes, stringent alignment to these references could also be used to filter reads (Zuluaga *et al.*, 2015).

Mapping, expression quantification and differential expression analysis

Mapping is the reconstruction of the transcriptome through alignment of reads to a reference sequence. In dual RNA-seq experiments, reads are mapped to the host reference sequence and the unaligned sequences are retained and mapped to the pathogen reference sequence (Teixeira *et al.*, 2014). A common program used for read alignment to a reference is the short read aligner Bowtie, which is part of the Tophat package of the Tuxedo pipeline (Trapnell *et al.*, 2012). Box 8.2 describes mapping and splice site determination. Bowtie allows the user to set the number of mismatches between the query and reference sequence, effectively setting a stringency threshold for the alignment (Langmead and Salzberg, 2012). This affects the stringency with which reads will be aligned and effectively assigned to the host or pathogen reference.

Once the reads have been assembled and filtered into host and pathogen libraries, transcript abundance quantification and differential expression analysis can be performed (Boxes 8.3 and 8.4, respectively). Expression levels are quantified by counting the number of reads mapped to each gene/transcript, normalized across the length of the gene/transcript to account for bias across abundant gene regions, relative to the number of reads in the original library. Programs like Cufflinks (Trapnell *et al.*, 2012) and RSEM (Li and Dewey, 2011) can be used to accurately quantify relative numbers of genes and transcripts. Differential expression analysis is commonly performed using packages such as Cuffdiff, DESeq and EdgeR (Anders and Huber, 2010; Robinson *et al.*, 2010; Trapnell *et al.*, 2013).

Box 8.2 Mapping and splice site determination

One of the most important aspects when mapping RNA-seq data to a reference genome is the detection of splice site junctions, and this should be considered when choosing an appropriate program. In many programs, the accuracy of estimating isoform abundances is still debatable. It is important to detect splice site junctions with as much accuracy as possible (Chandramohan *et al*., 2013), particularly when *a priori* splice site annotations are unavailable. Programs discussed here include Tophat (Kim *et al*., 2013), MapSplice (Wang *et al*., 2010), SOAPsplice (Huang *et al*., 2011), GSNAP (Wu and Nacu, 2010) and CRAC (Philippe *et al*., 2013). CRAC is the only known splice junction mapper that includes the discovery of transcriptomic splice junctions and chimeras, single nucleotide variants, and indels in a single analysis step. MapSplice specializes in determining splice site junctions from empirical methods rather than from the canonical intronic dinucleotide (ITDN) detecting methods, while TopHat makes use of the ultrafast short read aligner Bowtie and performs canonical splice site detection. SOAPsplice is capable of detecting true splice site junctions when there is low coverage, and provides efficient use of computational memory by making use of the Burrows–Wheeler transformation. GSNAP makes use of both single-end and paired-end reads, and can tolerate a variety of read lengths from very short to long length reads. GSNAP detects splice junctions using a probabilistic model as well as user defined splice site databases. Once successful detection of splice site junctions is achieved, the quantification of gene/transcript expression is the next step.

Box 8.3 Gene and transcript quantification

Programs discussed here – Cufflinks (Trapnell *et al*., 2012), IsoEM (Nicolae *et al*., 2011), HTSeq (Anders *et al*., 2015) and RSEM (Li and Dewey, 2011), have proven to estimate gene abundances with similar accuracy, however, their accuracy for transcript abundance estimation varies. Cufflinks is capable of calculating both gene and isoform level abundance estimates, although it has proven to be less efficient when considering computational memory, making it a time-consuming but accurate choice. HTSeq does not retain reads that map to multiple genes, which can prevent accurate expression quantification of overlapping genes. RSEM allows accurate quantification of gene and isoform abundances for organisms that do not have a reference genome. Isoform abundance estimations can be improved with paired-end read libraries when provided with an annotation of expected isoforms. IsoEM outperforms its competitors in its ability to estimate isoform abundances by making use of paired-end libraries and annotated lists of expected isoforms.

If a reference sequence is incomplete or unavailable, *de novo* assembly can be performed to create a reference sequence. Yazawa *et al.* (2013) describe a method for *de novo* assembly of the *in planta* pathogen transcriptome in the *Sorghum bicolor* and *Bipolaris sorghicola* interaction. Several putative defence and virulence related genes were identified. In this case, however, a reference sequence was available for the host, and host sequences could easily be removed from the assembled pathogen transcriptome by alignment to the host reference sequence, resulting in 232,760 non-host transcripts (Yazawa *et al.*, 2013). In cases where neither reference sequence is available, high quality annotation of *de novo* assembled sequences is crucial because separation of host and pathogen sequences will have to be done through alignment-based homology. When *de novo* assembly is required, it is important to consider the complexity of the transcriptome to be assembled. Increased transcriptome complexity, such as a large number of alternative splicing events, complicates the alignment of reads (Chang *et al.*, 2014). The accuracy of *de novo* assemblies increases with read length, which reduces mapping uncertainty. While there are algorithms that account for mapping uncertainty, thereby allowing the use of shorter read lengths (Li *et al.*, 2010),

Box 8.4 Differential expression analysis

When considering differential expression analysis, a number of confounding issues can arise, such as inherent errors in next-generation sequencing technologies, biases introduced due to length and composition of genes and transcripts, the existence of multiple gene isoforms, and the problem of overdispersion (Zhang *et al*., 2014). Overdispersion can occur when biological replicates are increased to account for biological variation. Using a Poisson distribution to model this variation results in underestimates. When considering differential gene expression, using a variance stabilizing transformation decreases overdispersion (Soneson and Delorenzi, 2013). However, if a negative binomial (NB) distribution is used, it effectively eliminates overdispersion (Anders and Huber, 2010; Zhang *et al*., 2014). Programs such as Cuffdiff, edgeR (Robinson *et al*., 2010) and DESeq (Anders and Huber, 2010) use very different statistical approaches to account for these confounding issues. EdgeR employs the empirical Bayes and weighted likelihood methods which allow for gene-specific variation while generalized linear models are used to account for complex experimental designs (Chen *et al*., 2014). DESeq has the option to use a variance stabilizing transformation, which results in differentially expressed gene candidates being balanced over the large dynamic range which is typical of RNA-seq experiments. Finally, Cuffdiff, a popular choice for differential gene expression analysis, estimates expression at transcript-level resolution and stabilizes variability across different biological replicates. Cuffdiff2 is capable of identifying differentially expressed transcripts and genes as well as promoter-preference changes where applicable (Trapnell *et al*., 2013).

these algorithms assume the availability of a high quality reference sequence.

Downstream analysis

Potential genes and pathways involved in the host–pathogen interaction can be identified during downstream analysis by comparing annotations of the identified differentially expressed genes to databases. These resources utilize controlled vocabularies provided by the Gene Ontology Consortium (GOC), which defines how genes function biologically within three categories: biological process, molecular function and cellular localization (Ashburner *et al*., 2000). Once potential pathways or biological functions have been identified through GOC, a Gene Set Enrichment Analysis (GSEA) can be conducted. This may aid in identifying pathways that are perturbed in the host and pathogen during the interaction.

The pathogen–host interactions database (www.phi-base.org) provides curated information on genes that affect the interaction between host and pathogens based on functional genetics studies with links to experimental evidence (Urban *et al*., 2015). A BLAST analysis against the nucleotide or predicted amino acid sequences in the database provides clues about whether pathogen genes expressed *in planta* are putative effectors, virulence factors or pathogenicity factors. *PathoPlant* (www.pathoplant.de) (Bülow *et al*., 2004) is a plant–pathogen interactions database with information on 350 interactions from model plant–pathogen systems. The database also allows the identification of signal transduction pathway components potentially induced during pathogen challenge.

For insight into host responses, the TreeGenes database (Wegrzyn *et al*., 2008), Phytozome database (http://phytozome.jgi.doe.gov) (Goodstein *et al*., 2012), The Plant Genome Integrative Explorer (PlantGenIE; www.plantgenie.org) (Sundell *et al*., 2015), and Gramene (www.gramene.org) (Liang *et al*., 2008) provide tools for analyses and visualization. Gramene contains the Plant Reactome database, which can be used to identify plant metabolic and regulatory pathways (Tello-Ruiz *et al*., 2016).

Weapons in plant–pathogen duels

The ultimate goal of a dual RNA-seq experiment is to gain biological insight into the interaction that is being studied, specifically those factors that determine disease severity. Since the dual RNA-seq approach is relatively new, it has not yet been applied to many systems. Several recent examples, representing interactions between a variety of plant and pathogen species, are listed in Table 8.1. This

Table 8.1 Summary of dual RNA-seq experiments showing genomic resources utilized, biological replicates (BR), mapping statistics and biological insight for host and pathogen

Host	Pathogen	Pathogen type	Time-point(s)	BR	Host	Pathogen	% reads mapped to pathogen	Main finding: host	Main finding: pathogen	Reference
Populus trichocarpa × *deltoides*	*Melampsora larici-populina* (virulent and avirulent isolates)	Fungus, biotroph	18, 21, 24 hpi (incompatible) 18, 24, 48 hpi (compatible)	1	*P. trichocarpa* genome v2.2	Genome v1.0	1.30%	Host sulfate transporter with greatly increased transcription after infection	19 candidate effectors identified	Petre *et al*., 2012
Oryza sativa	*Magnaporthe oryzae* (compatible, incompatible isolates)	Fungus, hemibiotroph	24 hpi	2	*Oryza sativa* build 5 genome	*Magnaporthe grisea* release 6 genome	0.1–0.2%	PR and phytoalexin genes in plant. More pronounced changes in incompatible interactions	Secreted protein genes were identified among the 240 up-regulated pathogen genes	Kawahara *et al*., 2012
Triticum aestivum	*Zymoseptoria tritici*	Fungus, hemibiotroph	7, 13, 56 dpi	3	–	*Mycosphaerella graminicola* genome v2.0	Not reported	–	PCWDEs had different expression profiles at the biotrophic, necrotrophic and saprotrophic stages	Brunner *et al*., 2013
Sorghum bicolor	*Bipolaris sorghicola*	Fungus, necrotroph	12, 24 hpi	4	*Sorghum* genome	*De novo* assembled transcriptome	Not reported	Pathogen-induced receptor, signalling, transcription factor and redox genes were expressed	LysM domain containing protein gene, PCWDEs, TFs identified	Yazawa *et al*., 2013
Theobroma cacao	*Moniliophthora perniciosa*	Fungus, hemibiotroph	30 dpi	5	*Theobroma cacao* genome v0.9	*M. perniciosa* genome	0.15–0.44%	Genes associated with carbohydrate degradation, secondary metabolism and cell wall modification were up-regulated, genes involved in starch biosynthesis and nitrogen assimilation were down-regulated	Putative effectors identified, model for biotrophic phase proposed	Teixeira *et al*., 2014

Notholithocarpus densiflorus	*Phytophthora ramorum*	Oomycete, hemibiotroph	1, 5 dpi	3	*De novo* transcriptome	*P. ramorum* genome	0.1–4.5%	Many R genes and pathogenesis-related (PR) genes were transcribed in the host, as well as genes related to cell death and signalling	Pathogenicity factor genes with roles in necrosis and host cell damage were expressed by the pathogen	Hayden *et al.*, 2014
Brassica napus	*Leptosphaeria biglobosa*	Fungus, necrotroph	7, 14 dpi	2	*B. napus*, *B. rapa* and *B. oleracea* exon array	Genome	5–44%	Jasmonic acid (JA) and salicylic acid (SA) pathway genes as well as cell wall remodelling genes were expressed	Expressed more cell wall degrading enzyme genes than the hemibiotroph	Lowe *et al.*, 2014
	Leptosphaeria maculans	Fungus, hemibiotroph	7, 14 dpi	2	*B. napus*, *B. rapa* and *B. oleracea* exon array	*Genome*	6–61%	SA related gene was expressed	Might employ carbohydrate binding enzymes (CAZy) such as LysM to evade the PAMP-triggered immunity and also expressed avirulence genes	Lowe *et al.*, 2014
Arabidopsis thaliana	*Hyaloperonospora arabidopsidis* (virulent, avirulent isolates)	Fungus, biotroph	1, 3, 5 dpi	3	Genome	Genome	0.022–1.452%	SA responses required for defence	Pathogen suppresses host SA-related defence	Asai *et al.*, 2014
Eucalyptus nitens	*Phytophthora cinnamomi*	Oomycete, hemibiotroph	5 dpi	3	*E. grandis* v1.1	*P. cinnamomi* draft genome	1%	PR-9 gene highly expressed	Crinkler family effector highly expressed	Meyer *et al.*, 2016

demonstrates the versatility of dual RNA-seq and the importance of experimental design in this type of study. Experimental design directly affects the kind of results that can be obtained from a dual RNA-seq experiment. The number of biological replicates among the studies in Table 8.1 ranges from one to five. With fewer biological replicates, very few differentially expressed genes were identified, but the information could still be used to guide functional characterization studies of the interaction. Knowledge of disease progression and the physical interaction between the host and pathogen is valuable when designing the study and interpreting the biological significance of the results. Examples include information about the timing of hyphal penetration of host cells or the switch between biotrophic and necrotrophic stages in hemibiotrophic pathogens, which can then be used to select the most prudent time points for RNA-seq.

Different sample groups can be chosen to understand those responses that are most likely to be involved in a particular phenotype, such as increased or reduced host resistance or pathogen virulence. The approaches listed in Table 8.1 included comparisons between resistant and susceptible hosts inoculated with a virulent pathogen, as well as comparisons between resistant hosts inoculated with virulent or avirulent pathogen strains.

The percentage of reads mapping to the pathogen reference sequence varies drastically, but it is typically very low compared to that of the host. Another technical aspect that becomes apparent is that several studies were possible without the use of an existing reference sequence, which expands the possible number of interactions that can be studied.

While the interactions are quite complex, researchers have been able to use the information to subsequently target genes for functional characterization. For example, Petre *et al.* (2016) and Yokotani *et al.* (2014) proceeded to characterize candidate effectors and a transcription factor, respectively. Many new approaches are becoming available for functional characterization. For example, the CRISPR-Cas9 system appears to be a good alternative to T-DNA knock-outs in fungi and its feasibility was recently demonstrated in *Trichoderma reesei* (Liu *et al.*, 2015) and *Pyricularia oryzae* (Arazoe *et al.*, 2015) following codon usage optimization of the Cas9 for use in filamentous fungi. This method may be favoured over conventional knock-out and knock-down systems due to its efficiency and the ability to produce knock-downs using a less efficient Cas9 nuclease (Doudna and Charpentier, 2014).

Next generation of dual approaches

The decrease in sequencing costs will promote dual RNA-seq approaches for plant–pathogen interaction studies and could allow multiple time-points and replicates to be investigated. Recent multiplexing approaches in RNA-seq such as RNA-tagSeq (Shishkin *et al.*, 2015) and Ligation Mediated RNA sequencing (Hou *et al.*, 2015) are set to further reduce costs of sequencing. Additionally, the number of sequenced genomes of plant hosts and their pathogens is rapidly increasing, facilitating dual RNA-sequencing studies [e.g. Joint Genome Initiative: 1000 fungal genomes project, http://1000.fungalgenomes.org (Kovalchuk *et al.*, 2013; Spatafora *et al.*, 2013)]. The sequencing of new plant transcriptomes [1000 Plant Transcriptomes (1KP Project), www.iplantcollaborative.org] and improvement of annotations is constantly occurring, providing more informative resources for dual RNA-seq experiments.

Laser capture microdissection coupled with an expression profiling technique may become a valuable tool for understanding the temporal and spatial dynamics of plant–pathogen interactions (Podgorny and Lazarev, 2016). Laser capture microdissection was applied to stomatal and surrounding areas of grapevine leaves in contact with the downy mildew pathogen *Plasmopara viticola* during early stages of infection, allowing the researchers greater resolution of transcriptional responses compared to whole leaves (Lenzi *et al.*, 2016). Laser capture coupled with microarray expression profiling was also used to determine the expression of *Melampsora larici* genes in the palisade mesophyll of poplar leaves and in the uredinium, revealing genetic signatures indicative of biotrophy and the switch to sporulation, respectively (Hacquard *et al.*, 2010).

We are currently in an age where single cell RNA-seq transcriptomics is possible [reviewed in Saliba *et al.* (2014)]. There are two technologies which

will allow direct RNA-sequencing using nanopores. One is where each nucleotide is read while the RNA molecule is threaded through the pore. The second strategy is RNA exosequencing where RNA is successively cleaved using polynucleotide phosphorylase, whereafter each nucleotide is read separately in the nanopore (Ayub *et al.*, 2013). Dual RNA-seq of infected single cells together with an intracellular bacterial pathogen was demonstrated with the latter approach (Hardwick *et al.*, 2011).

While we have reviewed a dual RNA-seq approach to dissect the interplay between plants and their pathogens, insight into the interaction is also possible at the level of proteomics and metabolomics (Allwood *et al.*, 2010; Rees *et al.*, 2015). Protein–protein interaction networks can reveal important triggers governing host resistance (Wang *et al.*, 2013). We predict that a combination of these approaches, together with gene function studies, will provide a systems level understanding of the interaction biology between the two organisms, which will allow more effective management of plant–pathogen duels for plant protection and improvement.

Glossary

Biological replicate: Repetition of measurements using different samples obtained under the same experimental conditions to account for biological variance.

Biological variance: Variation between data sets obtained from different samples within an experimental group arising from genetic differences or environmental influences on gene expression.

Coverage: The average number of reads that represent any given nucleotide in a reference or *de novo* assembled genome sequence. For a transcriptome assembly or expression study, coverage should be calculated based on the least expressed transcript.

Dual RNA-seq: A technique that makes use of high-throughput next-generation sequencing platforms to assess gene expression changes in the pathogen and its host simultaneously.

Fragment size: The combined length of the insert and the adapters within a sequencing library.

Insert size: The length of the DNA (or RNA) target between adapters within a sequencing library.

Library complexity: The number of unique molecules present within a sequencing library.

Overdispersion: The presence of higher variability within a dataset than predicted/accounted for by a statistical model.

Poly(A) selection: Selective amplification of poly(A) RNA species during library preparation by using poly(dT) primers for cDNA synthesis.

Read depth: The average read depth of a sequencing experiment is the read length multiplied by the number of reads, divided by the sum of the size of all genes being considered.

Read length: The number of bases sequenced per read.

rRNA depletion: Selective removal of the majority of ribosomal RNA species during library preparation.

Sequencing complexity: The number of unique molecules sequenced relative to the expected number of unique molecules.

Strand-specific library: A double-stranded cDNA library where information about which strand represents the original mRNA has been preserved.

Technical replicate: Repetition of measurements using the same sample to account for technical variance.

Technical variance: Variation between data sets obtained from the same sample using the same procedure arising from artefacts introduced by the process.

Unstranded library: A double-stranded cDNA library where information regarding which strand represents the original mRNA was not preserved. This results from the standard library preparation approach where synthesis of double stranded cDNA is primed randomly before the addition of adaptors.

References

Allwood, J.W., Clarke, A., Goodacre, R., and Mur, L.A. (2010). Dual metabolomics: a novel approach to understanding plant-pathogen interactions. Phytochemistry *71*, 590–597. http://dx.doi.org/10.1016/j.phytochem.2010.01.006

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

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq – a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. http://dx.doi.org/10.1093/bioinformatics/btu638

Arazoe, T., Miyoshi, K., Yamato, T., Ogawa, T., Ohsato, S., Arie, T., and Kuwata, S. (2015). Tailor-made CRISPR/Cas system for highly efficient targeted gene replacement in the rice blast fungus. Biotechnol. Bioeng. *112*, 2543–2549. http://dx.doi.org/10.1002/bit.25662

Asai, S., Rallapalli, G., Piquerez, S.J., Caillaud, M.C., Furzer, O.J., Ishaque, N., Wirthmueller, L., Fabro, G., Shirasu, K., and Jones, J.D. (2014). Expression profiling during *Arabidopsis*/downy mildew interaction reveals a highly-expressed effector that attenuates responses to salicylic acid. PLOS Pathog. *10*, e1004443. http://dx.doi.org/10.1371/journal.ppat.1004443

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000). Gene Ontology: tool for the unification of biology. Nat. Gen. *25*, 25–29.

Auer, P.L., and Doerge, R.W. (2010). Statistical design and analysis of RNA sequencing data. Genetics *185*, 405–416. http://dx.doi.org/10.1534/genetics.110.114983

Ayub, M., Hardwick, S.W., Luisi, B.F., and Bayley, H. (2013). Nanopore-based identification of individual nucleotides for direct RNA sequencing. Nano Lett. *13*, 6144–6150. http://dx.doi.org/10.1021/nl403469r

Brunner, P.C., Torriani, S.F., Croll, D., Stukenbrock, E.H., and McDonald, B.A. (2013). Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. Mol. Biol. Evol. *30*, 1337–1347. http://dx.doi.org/10.1093/molbev/mst041

Bülow, L., Schindler, M., Choi, C., and Hehl, R. (2004). PathoPlant: a database on plant-pathogen interactions. In Silico Biol. *4*, 529–536

Chandramohan, R., Wu, P.Y., Phan, J.H., and Wang, M.D. (2013). Benchmarking RNA-Seq quantification tools. Conf. Proc. IEEE Eng. Med. Biol. Soc. *2013*, 647–650. http://dx.doi.org/10.1109/EMBC.2013.6609583

Chang, Z., Wang, Z., and Li, G. (2014). The impacts of read length and transcriptome complexity for de novo assembly: a simulation study. PLOS ONE *9*, e94825. http://dx.doi.org/10.1371/journal.pone.0094825

Chen, X.R., Xing, Y.P., Li, Y.P., Tong, Y.H., and Xu, J.Y. (2013). RNA-seq reveals infection-related gene expression changes in *Phytophthora capsici*. PLOS ONE *8*. http://dx.doi.org/10.1371/journal.pone.0074588

Chen, Y., Lun, A.T.L., and Smyth, G.K. (2014). Diferential Expression Analysis of Complex RNA-seq Experiments Using edgeR. In Statistical Analysis of Next Generation Sequencing Data, S. Datta, and D. Nettleton, eds. (Cham, Switzerland: Springer International Publishing), pp. 51–74.

Chhangawala, S., Rudy, G., Mason, C.E., and Rosenfeld, J.A. (2015). The impact of read length on quantification of differentially expressed genes and splice junction detection. Genome Biol. *16*, 131. http://dx.doi.org/10.1186/s13059-015-0697-y

Denoeud, F., Aury, J.M., Da Silva, C., Noel, B., Rogier, O., Delledonne, M., Morgante, M., Valle, G., Wincker, P., Scarpelli, C., *et al.* (2008). Annotating genomes with massive-scale RNA sequencing. Genome Biol. *9*, R175. http://dx.doi.org/10.1186/gb-2008-9-12-r175

Dou, D., and Zhou, J.M. (2012). Phytopathogen effectors subverting host immunity: different foes, similar battleground. Cell Host Microbe *12*, 484–495. http://dx.doi.org/10.1016/j.chom.2012.09.003

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096. http://dx.doi.org/10.1126/science.1258096

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., *et al.* (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. *40*, D1178–D1186. http://dx.doi.org/10.1093/nar/gkr944

Grant, G.R., Farkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B., and Pierce, E.A. (2011). Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics *27*, 2518–2528. http://dx.doi.org/10.1093/bioinformatics/btr427

Hacquard, S., Delaruelle, C., Legue, V., Tisserant, E., Kohler, A., Frey, P., Martin, F., and Duplessis, S. (2010). Laser capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptional switch between biotrophy and sporulation. Mol. Plant-Microbe Interact. *23*, 1275–1286. http://dx.doi.org/10.1094/mpmi-05-10-0111

Hansen, K.D., Wu, Z., Irizarry, R.A., and Leek, J.T. (2011). Sequencing technology does not eliminate biological variability. Nat. Biotechnol. *29*, 572–573. http://dx.doi.org/10.1038/nbt.1910

Hardwick, S.W., Chan, V.S., Broadhurst, R.W., and Luisi, B.F. (2011). An RNA degradosome assembly in *Caulobacter crescentus*. Nucleic Acids Res. *39*, 1449–1459. http://dx.doi.org/10.1093/nar/gkq928

Hayden, K.J., Garbelotto, M., Knaus, B.J., Cronn, R.C., Rai, H., and Wright, J.W. (2014). Dual RNA-seq of the plant pathogen *Phytophthora ramorum* and its tanoak host. Tree Genet. Genomes *10*, 489–502. http://dx.doi.org/10.1007/s11295-014-0698-0

Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., and Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. BioTechniques *56*, 61–64, 66, 68, passim. http://dx.doi.org/10.2144/000114133

Hou, Z., Jiang, P., Swanson, S.A., Elwell, A.L., Nguyen, B.K., Bolin, J.M., Stewart, R., and Thomson, J.A. (2015). A cost-effective RNA sequencing protocol for large-scale gene expression studies. Sci. Rep. *5*, 9570. http://dx.doi.org/10.1038/srep09570

Huang, S., Zhang, J., Li, R., Zhang, W., He, Z., Lam, T.W., Peng, Z., and Yiu, S.M. (2011). SOAPsplice:genome-wide *ab initio* detection of splice junctions from RNA-Seq data. Front. Genet. *2*, 46. http://dx.doi.org/10.3389/fgene.2011.00046

Ikeue, D., Schudoma, C., Zhang, W., Ogata, Y., Sakamoto, T., Kurata, T., Furuhashi, T., Kragler, F., and Aoki, K. (2015). A bioinformatics approach to distinguish plant parasite and host transcriptomes in interface tissue by classifying RNA-Seq reads. Plant Methods *11*, 34. http://dx.doi.org/10.1186/s13007-015-0066-6

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res. *21*, 1543–1551. http://dx.doi.org/10.1101/gr.121095.111

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., and Minami, E. (2012). Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. PLOS ONE *7*, e49423. http://dx.doi.org/10.1371/journal.pone.0049423

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36. http://dx.doi.org/10.1186/gb-2013-14-4-r36

Kovalchuk, A., Keriö, S., Oghenekaro, A.O., Jaber, E., Raffaello, T., and Asiegbu, F.O. (2013). Antimicrobial defenses and resistance in forest trees: challenges and perspectives in a genomic era. Annu. Rev. Phytopathol. *51*, 221–244. http://dx.doi.org/10.1146/annurev-phyto-082712-102307

Kunjeti, S.G., Evans, T.A., Marsh, A.G., Gregory, N.F., Kunjeti, S., Meyers, B.C., Kalavacharla, V.S., and Donofrio, N.M. (2012). RNA-Seq reveals infection-related global gene changes in *Phytophthora phaseoli*, the causal agent of lima bean downy mildew. Mol. Plant Pathol. *13*, 454–466. http://dx.doi.org/10.1111/j.1364-3703.2011.00761.x

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. http://dx.doi.org/10.1038/nmeth.1923

Lenzi, L., Caruso, C., Bianchedi, P.L., Pertot, I., and Perazzolli, M. (2016). Laser microdissection of grapevine leaves reveals site-specific regulation of transcriptional response to *Plasmopara viticola*. Plant Cell Physiol. *57*. http://dx.doi.org/10.1093/pcp/pcv166

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods *7*, 709–715. http://dx.doi.org/10.1038/nmeth.1491

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. http://dx.doi.org/10.1186/1471-2105-12-323

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics *26*, 493–500. http://dx.doi.org/10.1093/bioinformatics/btp692

Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E.S., Casstevens, T., Hurwitz, B., McCouch, S., Ni, J., Pujar, A., *et al.* (2008). Gramene: a growing plant comparative genomics resource. Nucleic Acids Res. *36*, D947–D953. http://dx.doi.org/10.1093/nar/gkm968

Liang, H., Staton, M., Xu, Y., Xu, T., and LeBoldus, J. (2014). Comparative expression analysis of resistant and susceptible *Populus* clones inoculated with *Septoria musiva*. Plant Sci. *223*, 69–78. http://dx.doi.org/10.1016/j.plantsci.2014.03.004

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. http://dx.doi.org/10.1093/bioinformatics/btt656

Liu, R., Chen, L., Jiang, Y., Zhou, Z., and Zou, G. (2015). Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. Cell Discovery *1*, 15007. http://dx.doi.org/10.1038/celldisc.2015.7

Liu, Y., Zhou, J., and White, K.P. (2014). RNA-seq differential expression studies: more sequence or more replication? Bioinformatics *30*, 301–304. http://dx.doi.org/10.1093/bioinformatics/btt688

Lowe, R.G.T., Cassin, A., Grandaubert, J., Clark, B.L., Van de Wouw, A.P., Rouxel, T., and Howlett, B.J. (2014). Genomes and transcriptomes of partners in plant-fungal- interactions between canola (*Brassica napus*) and two *Leptosphaeria* species. PLOS ONE *9*. http://dx.doi.org/10.1371/journal.pone.0103098

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448–3449. http://dx.doi.org/10.1093/bioinformatics/bti551

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517. http://dx.doi.org/10.1101/gr.079558.108

Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. Nat. Rev. Genet. *12*, 671–682. http://dx.doi.org/10.1038/nrg3068

McIntyre, L.M., Lopiano, K.K., Morse, A.M., Amin, V., Oberg, A.L., Young, L.J., and Nuzhdin, S.V. (2011). RNA-seq: technical variability and sampling. BMC Bioinf. *12*, 293. http://dx.doi.org/10.1186/1471-2164-12-293

Meyer, F.E., Shuey, L.S., Naidoo, S., Mamni, T., Berger, D.K., Myburg, A.A., van den Berg, N., and Naidoo, S. (2016). Dual RNA-sequencing of *Eucalyptus nitens* during *Phytophthora cinnamomi* challenge reveals pathogen and host factors influencing compatibility. Front. Plant Sci. *7*, 191. http://dx.doi.org/10.3389/fpls.2016.00191

Nettleton, D. (2014). Design of RNA Sequencing Experiments. In Statistical Analysis of Next Generation Sequencing Data, S. Datta, and D. Nettleton, eds. (Cham, Switzerland: Springer International Publishing), pp. 93–113.

Nicolae, M., Mangul, S., Măndoiu, I.I., and Zelikovsky, A. (2011). Estimation of alternative splicing isoform frequencies from RNA-Seq data. Algorithms Mol. Biol. *6*, 9. http://dx.doi.org/10.1186/1748-7188-6-9

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *27*, 29–34. http://dx.doi.org/10.1093/nar/27.1.29

Pareek, C.S., Smoczynski, R., and Tretyn, A. (2011). Sequencing technologies and genome sequencing. J. Appl. Genet. *52*, 413–435. http://dx.doi.org/10.1007/s13353-011-0057-x

Petre, B., Lorrain, C., Saunders, D., Win, J., Sklenar, J., Duplessis, S., and Kamoun, S. (2016). Rust fungal effectors mimic host transit peptides to translocate into chloroplasts Cell. Microbiol. *18*, 453–465. http://dx.doi.org/10.1111/cmi.12530

Petre, B., Morin, E., Tisserant, E., Hacquard, S., Da Silva, C., Poulain, J., Delaruelle, C., Martin, F., Rouhier, N., Kohler, A., *et al.* (2012). RNA-seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers PtSultr3;5, a fungal-induced host sulfate transporter. PLOS ONE *7*. http://dx.doi.org/10.1371/journal.pone.0044408

Philippe, N., Salson, M., Commes, T., and Rivals, E. (2013). CRAC: an integrated approach to the analysis of RNA-seq reads. Genome Biol. *14*, R30. http://dx.doi.org/10.1186/gb-2013-14-3-r30

Podgorny, O.V., and Lazarev, V.N. (2016). Laser microdissection: A promising tool for exploring microorganisms and their interactions with hosts. J. Microbiol. Methods http://dx.doi.org/10.1016/j.mimet.2016.01.001

Rees, M.A., Kleifeld, O., Crellin, P.K., Ho, B., Stinear, T.P., Smith, A.I., and Coppel, R.L. (2015). Proteomic characterization of a natural host pathogen interaction: repertoire of *in vivo* expressed bacterial and host surface-associated proteins. J. Proteome Res. *14*, 120–132. http://dx.doi.org/10.1021/pr5010086

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y.F., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*. http://dx.doi.org/10.1093/nar/gkv007

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. http://dx.doi.org/10.1093/bioinformatics/btp616

Saliba, A.E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. *42*, 8845–8860. http://dx.doi.org/10.1093/nar/gku555

Schulze, S., Henkel, S.G., Driesch, D., Guthke, R., and Linde, J. (2015). Computational prediction of molecular pathogen-host interactions based on dual transcriptome data. Front. Microbiol. *6*, 65. http://dx.doi.org/10.3389/fmicb.2015.00065

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., *et al.* (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nat. Methods *12*, 323–325. http://dx.doi.org/10.1038/nmeth.3313

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. *15*, 121–132. http://dx.doi.org/10.1038/nrg3642

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

Spatafora, J., Stajich, J., and Grigoriev, I. (2013). 1000 Fungal Genomes project. Phytopathology *103*, 137–137

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y.C., Sjödin, A., Van de Peer, Y., Jansson, S., Hvidsten, T.R., and Street, N.R. (2015). The Plant Genome Integrative Explorer Resource: PlantGenIE.org. New Phytol. *208*, 1149–1156. http://dx.doi.org/10.1111/nph.13557

Teixeira, P.J., Thomazella, D.P., Reis, O., do Prado, P.F., do Rio, M.C., Fiorin, G.L., Jose, J., Costa, G.G., Negri, V.A., Mondego, J.M., *et al.* (2014). High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. Plant Cell *26*, 4245–4269. http://dx.doi.org/10.1105/tpc.114.130807

Tello-Ruiz, M.K., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., Amarasinghe, V., Dharmawardhana, P., Jiao, Y., Mulvaney, J., *et al.* (2016). Gramene 2016: comparative plant genomics and pathway resources. Nucleic Acids Res. *44*, D1133–D1140. http://dx.doi.org/10.1093/nar/gkv1179

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y., and Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. *37*, 914–939. http://dx.doi.org/10.1111/j.1365-313X.2004.02016.x

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53. http://dx.doi.org/10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578. http://dx.doi.org/10.1038/nprot.2012.016

Urban, M., Pant, R., Raghunath, A., Irvine, A.G., Pedro, H., and Hammond-Kosack, K.E. (2015). The Pathogen-Host Interactions database (PHI-base): additions and future developments. Nucleic Acids Res. *43*, D645–D655. http://dx.doi.org/10.1093/nar/gku1165

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., *et al.* (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. *38*, e178. http://dx.doi.org/10.1093/nar/gkq622

Wang, Y.C., Lin, C., Chuang, M.T., Hsieh, W.P., Lan, C.Y., Chuang, Y.J., and Chen, B.S. (2013). Interspecies protein-protein interaction network construction for characterization of host-pathogen interactions: a *Candida albicans*-zebrafish interaction study. BMC Syst. Biol. 7. http://dx.doi.org/10.1186/1752-0509-7-79

Wegrzyn, J.L., Lee, J.M., Tearse, B.R., and Neale, D.B. (2008). TreeGenes: A forest tree genome database. Int J Plant Genomics *2008*, 412875. http://dx.doi.org/10.1155/2008/412875

Westermann, A.J., Gorski, S.A., and Vogel, J. (2012). Dual RNA-seq of pathogen and host. Nat. Rev. Microbiol. *10*, 618–630. http://dx.doi.org/10.1038/nrmicro2852

Wilhelm, B.T., and Landry, J.R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods *48*, 249–257. http://dx.doi.org/10.1016/j.ymeth.2009.03.016

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881. http://dx.doi.org/10.1093/bioinformatics/btq057

Yang, C., and Wei, H. (2015). Designing microarray and RNA-Seq experiments for greater systems biology discovery in modern plant genomics. Mol. Plant. *8*, 196–206. http://dx.doi.org/10.1016/j.molp.2014.11.012

Yazawa, T., Kawahigashi, H., Matsumoto, T., and Mizuno, H. (2013). Simultaneous transcriptome analysis of sorghum and *Bipolaris sorghicola* by using RNA-seq in combination with de novo transcriptome assembly. PLOS ONE *8*, e62460. http://dx.doi.org/10.1371/journal.pone.0062460

Yokotani, N., Tsuchida-Mayama, T., Ichikawa, H., Mitsuda, N., Ohme-Takagi, M., Kaku, H., Minami, E., and

Nishizawa, Y. (2014). OsNAC111, a blast disease-responsive transcription factor in rice, positively regulates the expression of defense-related genes. Mol. Plant-Microbe Interact. *27*, 1027–1034. http://dx.doi.org/10.1094/mpmi-03-14-0065-r

Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R., *et al.* (2014). A comparative study of techniques for differential expression analysis on RNA-Seq data. PLOS ONE *9*, e103207. http://dx.doi.org/10.1371/journal.pone.0103207

Zuluaga, A.P., Sole, M., Lu, H., Gongora-Castillo, E., Vaillancourt, B., Coll, N., Buell, C.R., and Valls, M. (2015). Transcriptome responses to *Ralstonia solanacearum* infection in the roots of the wild potato *Solanum commersonii*. BMC Genomics *16*, 246. http://dx.doi.org/10.1186/s12864-015-1460-1