HortFlora Research Spectrum
Vol. 5, Issue 1; 1-6 (March 2016)
www.hortflorajournal.com
ISSN: 2250-2823

TRANSCRIPTOME PROFILING ASSOCIATED TO PLANT DISEASES: A REVIEW

B. Tanuja Priya[1*], B. N.S. Murthy[2] and B. Divya[1]

[1]*College of Horticulture, University of Horticultural Sciences*
[2]*Division of Fruit Science, Indian Institute of Horticultural Research*
**Corresponding Author's E-mail : tpriyahort@gmail.com*

ABSTRACT : Transcriptome profiling involves estimation of transcript's relative abundance and focuses on differentially expressed genes among various groups, which helps in identification of potential genes responsible for susceptible and resistant reaction of plant diseases. The transcriptome study enriches knowledge on host-pathogen interaction and also discloses the crucial biochemical pathways involved in defense mechanism of plants against various diseases.

Keywords : *Transcriptome profiling, plant diseases, RNA sequencing, NGS.*

The plant diseases are caused as a result of three interactive components: host factors, pathogen factors, and environmental factors. The relationship between these factors and the disease situation they create can best be visualized as a triangle, the disease triangle. The disease triangle is a conceptual model that can be used to predict epidemiological outcomes in plant health. Disease resistance/susceptibility in plants is often characterized by matching of genes in host and pathogen. The plant disease resistance is a heritable characteristic which enables it to overcome, either completely or in-part, the effect of a pathogen or other damaging factor. Many major resistance genes operate in a *gene-for-gene* way. For each resistance gene in the host, there is a corresponding avirulence gene in the pathogen (Flor, 6), and only the corresponding avirulence gene can initiate the hypersensitive reaction (HR) leading to incompatibility. Resistance (R) genes play a central role in recognising effectors from pathogens and in triggering downstream signalling during plant disease resistance. To date, more than 112 R genes and 104,310 putative R -genes present in a wide variety of plants species and conferring resistance to 122 pathogens. The known R proteins can be grouped into several super-families based on the presence of a few structural motifs, including nucleotide-binding sites (NBSs), leucine-rich repeat (LRR) domains, Toll/ Interleukin-1 receptor (TIR) domains, coiled-coil (CC) domains and transmembrane (TM) regions. The identification of genes that controls disease resistance in plants provides the basis for new progress in genetic improvement of crop species, complementing traditional methods based on assisted crosses.

Article's History:
Received : 08-01-2016 ***Accepted : 29-01-2016***

Molecular biology provides several techniques to assess gene functions and indeed one such technique is transcriptomics that have been applied to the development of novel strategies to demonstrate the differential expression of a particular genes conferring resistance. Transcriptomics refers to the study dealing with whole transcriptome. The initial product of genome expression is the transcriptome, the collection of RNA molecules derived from those protein-coding genes whose biological information is required by the cell at a particular time. The RNA molecules of the transcriptome, as well as many other RNAs derived from genes that do not code for proteins, are synthesized by the process called transcription. The combination of NGS and potential of modern computational biology opens up new opportunities for studying the transcriptomes, including those of non-model species, that ensures progressive advance in many areas of biological science (Zhukov *et al.,* 7).

Table 1: Historical perspective of transcriptome analysis

Year	Milestone
1965	Sequence of the first RNA molecule determined
1977	Development of the Northern blot technique and the Sanger sequencing method
1989	Reports of RT-PCR experiments for transcriptome analysis
1991	First high throughput EST sequencing study
1992	Introduction of Differential Display (DD) for the discovery of differential expressed genes
1995	Report of the microarray and Serial Analysis of Gene Expression (SAGE) methods
2001	Draft of the Human Genome completed
2005	First next generation sequencing technology (454/Roche) introduced to the market

2006	First transcriptome sequencing studies using next generation technology (454/Roche) and Genome Analyzer (1st Solexa Sequencer) Launches
2007	Illumina Acquires Solexa
2010	Helicos platform
2011	SMRT platform
2012	GridION platform

VARIOUS SEQUENCING TECHNOLOGIES USED FOR TRANSCRIPTOME ANALYSIS

Pre Genome and Genome Era RNA Sequencing Technologies

Northern Blot Technique: A northern blot is a laboratory method used to detect specific RNA molecules among a mixture of RNA. In Northern blot technique, RNA was denatured and separated in to single strands based on its sizes using gel electrophoresis method. Following separation, the RNA was transferred on to a blotting membrane that was treated with a small piece of complementary DNA or RNA called a *probe* which was typically labeled with radioactive atom or a fluorescent dye. The *probe* further hybridized, to a specific RNA fragment on the membrane and allowed the RNA molecule of interest to be detected from among the many different RNA molecules on the membrane. To gain insight into the transcriptome of the *Xanthomonas campestris* pv. *vesicatoria* (*Xcv*) strain 85–10, Schmidtke *et al.* (16) took a differential RNA sequencing (dRNA-seq) approach in gram-negative plant-pathogenic bacterium *Xcv*, elucidated the mechanisms involved in the interaction with the host and reported 1421 putative TSSs (transcription start site) in the *Xcv* genome. Northern blot analyses confirmed 16 intergenic small RNAs and seven cis encoded antisense RNAs in *Xcv*. The advantages of this techniques are to identify infectious agents present in the sample and inherited disease, can be applied to mapping restriction sites in single copy gene, widely accepted method, adaptable protocol - it allows the usage of many types of probes and disadvantages are complex and time consuming, requires electrophoretic separation, only one gene can be analysed at a time, gives information about presence of DNA, RNA or proteins but does not give information about regulation and gene interactions.

Sanger Sequencing Technique: The first method of sequencing the genetic code was devised by Fred Sanger. To sequence the DNA, two strands were separated and the strand to be sequenced was copied using chemically altered bases. These altered bases caused the copying process to stop each time one particular letter was incorporated into the growing DNA chain. This process was carried out for all four bases, and then the fragments were put together like a jigsaw to reveal the sequence of the original piece of DNA. The advantages of this method are simple – no preliminary extension is required, results are better with fewer artifacts, larger sequence can be read, works on a small scale as well as a large one, 300 nucleotides can be determined and disadvantages are contaminant fragments, band “pile-ups” can occur due to loop formation under gel conditions, and is usually depicted as numerous bands in the same position and not completely accurate.

Expressed sequence tags : Firstly, total RNA was isolated and then mRNA was extracted by using different kits (*e.g.,* RNAgent Promega) and subjected to reverse transcription to form cDNA library. From the cDNA library 5’ or 3’-ESTs were generated by cDNA end sequencing. The constructed ESTs could then be assembled separately in multimember sequence assembly, bridged sequence assembly and small clusters on the basis of size of ESTs to find out the actual sequences of bases. Newcomb *et al.* (13) had produced a substantial expressed sequence tag collection from various tissues of apple. They observed representatives from protein families that indicated the presence of many genes involved in disease resistance. Riju and Arunachalam (15) explored the single nucleotide polymorphims (SNPs) in expressed sequence tags (ESTs) of cocoa and observed a relative increase in the proportion of transversions (1268) over transitions (950) in bean and leaves and defense related EST sequence libraries. The main advantages are gene discovery, digital insights into transcriptome, genome analyses esp. annotation of genomic DNA and disadvantages are inherent low quality due to single pass nature, not 100 % full length cDNA clones and redundant sequencing of abundant transcripts.

Microarray Technique : A typical microarray experiment involved the hybridization of an mRNA molecule to the DNA template from which it was originated. Many DNA samples were used to construct an array. The amount of mRNA bound to each site on the array indicated the expression level of the various genes. This number may run in thousands. All the data was collected and a profile was generated for gene expression in the cell. Solanum commersonii transcriptional response to Ralstonia solanacearum was analyzed by Zuluaga *et al.* (18) and reported specific genes that are differentially expressed in respond to bacterial wilt. Their findings highlighted that a high proportion of *S. commersonii*-specific transcripts were altered by *R. solanacearum* only in F118

(resistant genotype) while phythormone-related genes were highly induced in F97 (susceptible genotype), suggesting a markedly different response by the genotypes to the pathogen. Choi *et al.* (5) carried out a transcriptome analysis of chrysanthemum in response to three different viruses including *cucumber mosaic virus (CMV), tomato spotted wilt virus (TSWV)* and *potato virus X (PVX)* using microarray analysis. They identified a total of 125, 70 and 124 differentially expressed genes for *CMV, TSWV* and *PVX*, respectively and stated that their analysis revealed several genes related to hormone mediated viral stress response and DNA modification. The advantages are gene discovery, gene expression, and disease diagnosis and disadvantages are insufficient quality and sensitivity, lack of reproducibility and costly technique.

SAGE Technique: Serial analysis of gene expression (SAGE) is a powerful tool that allows digital analysis of overall gene expression patterns. mRNA was isolated from an input sample. A small chunk of sequence from a defined position of each mRNA molecule was extracted. These small pieces of sequence were linked together to form a long chain (or concatamer) and were cloned into a vector. The formed chains using modern high-throughput DNA sequences were sequenced. This data was then processed with a computer to count the small sequence tags. Asamizu *et al.* (2) performed a comprehensive transcript analysis on the early stage of root nodulation in the model legume *Lotus japonicus* by serial analysis of gene expression (SAGE) and indicated the effectiveness of SAGE in discriminating different gene family members. They also identified genes for 44 unknown tags by means of reverse SAGE and found 11 antisense tags that increased during nodulation, indicating that regulation of gene expression by antisense transcripts may occur in an organ-dependent manner. The advantages are no hybridization of mRNA output to probes, so quantifying gene expressions is more exact in SAGE, mRNA sequences need not to be known in priory, so genes or gene variants which are not known can be discovered and disadvantages are large-scale studies do not typically use SAGE and costly process.

Next Generation Sequencing Technologies

Roche/454 Pyrosequencing: The basis of pyrosequencing depends on the detection of pyrophosphate molecules released during chain synthesis. Here, DNA fragments were connected to beads coated with oligonucleotides that are complementary to adapter sequences at the ends, and amplification operation was performed through emulsion PCR. The beads carrying thousands of copies of a different DNA fragment on their surface were inserted into a plate consisting of millions of wells. Wells also contained enzymes that provide chemiluminescent detection. Sequencing was accomplished, respectively, by the addition of DNA polymerase and a type of nucleotide. When a nucleotide is added to template chains that are fixed to the beads by DNA polymerase, ATP sulfurylase enzyme converts released pyrophosphate to ATP. By converting ATP to a light-generating molecule in direct proportion to the amount of luciferase enzyme, the type and number of base that is added are determined per cycle. Passos *et al.* (14) employed 454 GS-FLX Titanium technology to determine the sequence of gene transcripts in genotypes of *Musa acuminata* sp. *burmannicoides* Calcutta 4 and *M. acuminata* subgroup Cavendish cv. Grande Naine, contrasting in resistance to the fungal pathogen *Mycosphaerella musicola*, causal organism of Sigatoka leaf spot disease. Genes from a number of defense-related pathways were observed in transcripts from each cDNA library and a subset of 95 potential defense-related gene-derived simple sequence repeat (SSR) loci were validated for specific amplification and polymorphism across *M. acuminata* accessions. The advantages are longest readings (up to 1000 base pairs), which facilitates reference genome alignment of sequenced DNA fragments and *de novo* (without reference sequence) binding and disadvantages are it has the lowest total output (700 Mb) and the highest cost per base.

Illumina/Solexa Genome Analyzer: The DNA templates that have adapter sequences on both ends were connected to them via a glass surface that is coated with primer complementary sequences on one end. During amplification, the free end was connected to the closest complementary primer and takes the form of a bridge. After each amplification cycle, in which DNA polymerase synthesizes the chain, chains were denatured. The sequencing of the resulting DNA clusters is performed through the cyclic reversible termination method. With the addition of primer, DNA polymerase, and different fluorescently labeled terminator nucleotides, sequencing begins. With the addition of each nucleotide, synthesis pauses, and fluorescence light is recorded. After the removal of the terminator chemical group that is connected to the nucleotide, the next synthesis cycle takes place. RNA samples from different organs of the Cavendish cultivar were pooled for deep sequencing using the Illumina technology. Analysis of the banana transcriptome led to identification of over 842 genes that were not annotated by the Musa genome project. A large

number of simple nucleotide polymorphisms (SNPs) and short insertions and deletion (indels) were also identified from the transcriptome data (Li *et al.,* 10). *Fusarium wilt,* caused by the fungal pathogen *Fusarium oxysporum f. sp. cubense* tropical race 4 (*Foc TR4*), is considered the most lethal disease of Cavendish bananas in the world. Li *et al.* (11) and Bai *et al.* (4) performed *de novo* sequencing technology to investigate defense mechanism in banana against the mentioned pathogen, but, former compared the middle resistant cultivar 'Nongke No 1' and latter compared highly resistant cultivar 'Yueyoukang 1' with the susceptible cultivar 'Brazilian'. The results of Li *et al.* (11) indicated that basal defense mechanisms are involved in the recognition of pathogen associated molecular patterns and high levels of defense-related transcripts. Further, Bai *et al.* (4) provided valuable insights for understanding the compatible and incompatible interactions between banana and *Foc TR4.* Martinelli *et al.* (12) studied host responses of citrus to infection with *Candidatus* Liberibacter *asiaticus (CaLas)* using next-generation sequencing technologies and revealed that huanglongbing (HLB) strongly affected pathways involved in source-sink communication, including sucrose and starch metabolism and hormone synthesis and signaling. They also reported that *CaLas* infection triggered a response via both the salicylic acid and jasmonic acid pathways and increased the transcript abundance of several members of the WRKY family of transcription factors. Hong *et al.* (7) studied the host responses of the mango fruit against *C. gloeosporioides* using Illumina paired-end sequencing technology, identified 35 potential defense-related genes and further analyzed expression profiles using qRT-PCR. The advantage and disadvantage of this method are highest-scale next-generation sequencing platform and the reading length is shorter (2x100 bp), respectively.

Applied Biosystems/Solid : In this sequencing method, ligation was performed to produce short sequences, also known as interrogation probes. It consisted of 2 bases specific to the probe and 6 variable bases. Furthermore, the 5' end of each probe was marked with one of four different fluorescence molecules. The sequencing reaction blend consisted of interrogation probes that include 16 different combinations that might be created by the first interrogated base pair. When an interrogated probe hybridized with the template sequence, fluorescent light was recorded that helped in detection of nucleotide sequences. A RNA-seq time course (12, 32 and 67 dpi) study, monitoring gene expression in SACMV-challenged susceptible (T200) and tolerant (TME3) cassava landraces, was performed using the Applied Biosystems (ABI) SOLiD next-generation sequencing platform by Allie *et al.* (1). The multiplexed paired end sequencing run produced a total of 523 MB and 693 MB of paired-end reads for SACMV-infected susceptible and tolerant cDNA libraries, respectively. Alterations in the expression of other interesting genes such as transcription factors, resistance (R) genes, and histone/DNA methylation-associated genes, were observed. KEGG pathway analysis uncovered important altered metabolic pathways, including phenylpropanoid biosynthesis, sucrose and starch metabolism, and plant hormone signaling. Jambagi and Dunwell (8) investigated differentially expressed genes (log2 fold changes =5) between control and 1 day after inoculation of powdery mildew (*Podosphaera aphanis*) in both F. vesca ssp. vesca Hawaii 4 and *F. vesca* f. *semperflorens* Yellow Wonder 5AF7 and identified a large number of genes related to secondary metabolism, signal transduction; transcriptional regulation and disease resistance were highly expressed. The advantage and disadvantage of this method are low error rate and costlier process, respectively.

Ion Torrent platform : In this method, no fluorescent light or chemical modification is used, and the hydrogen ion that is released when each nucleotide is added during DNA synthesis causes a pH change in the solution. The pH change is detected by an ion detector and recorded. The transcriptome sequencing was performed in Musa B genome accession Attikol using the Ion Torrent platform by Backiyarani *et al.* (35). They included that RNA sequencing of the Musa B genome would provide a vast array of transcriptomic information that could lead to the development of trait-specific markers and the discovery of new genes and regulatory sequences involved in resistance mechanisms. They also reported that this method led to the generation of about 4.5 million paired-end reads, which were assembled using the MIRA assembler. The assembly also produced 82,413 unique transcripts with a mean length of approximately 113 bp. Functional annotation against Plant CYC pathway database also identified 20,696 unique transcripts, which were mapped to 455 pathways. The advantage is no fluorescent light or chemical modification is used and disadvantage is the highest cost per base.

Single-molecule DNA template sequencing technologies : The fluorescent detection methods that are used in the next-generation sequencing platforms mentioned above and are designed to detect amplified

signals, hence they require template chains to be amplified. However, apart from these technologies, devices that do not require amplification phases and more sensitive to low fluorescent signals have been developed. Some of these technologies are as follows

1. Helicos platform : Single DNA molecule is used as a template and are more sensitive to low fluorescent signals and it is developed by Heliscope Biosciences, Cambridge, USA

2. SMRT platform : Single-molecule real-time sequencing-by-synthesis is capable of real time recording for millisecond fluorescent signals and it is developed by Pacific Biosciences, Menlo Park, USA

3. GRIDION platform : Single chain DNA molecules passed through the nanometric diameter wells where electric current is applied to detect nucleotides. This has been designed by Oxford Nanopore technologies, Oxford, United Kingdom.

It was stated by Kanter and Kalisky (9) that traditionally, gene expression measurements were performed on “bulk” samples containing populations of thousands of cells. Recent advances in genomic technology have made it possible to measure gene expression in hundreds of individual cells at a time. As a result, cellular properties that were previously masked in “bulk” measurements can now be observed directly. The advantages are it do not require an amplification phase and initial DNA sample amount (<1 ìg) is less than other methods fication phase and disadvantage is error ratio in the sequencing may increase.

CONCLUSION

Whole transcriptome analysis is growing importance in understanding how altered expression of genetic variants contributes to complex diseases of plants. The experiments on transcriptome analysis can characterize all transcriptional activity (coding and non-coding), focus on a subset of relevant target genes and transcripts, or profile thousands of genes at once to create a global picture of cell function. Moreover, analysis of genome-wide differential RNA expression provides researchers with greater insights into biological pathways and molecular mechanisms that regulate cell fate, development, and disease progression. Next-generation sequencing (NGS) offers comprehensive high-quality gene expression and transcriptome analysis data for a broad range of sample types. It detects and quantifies any active gene or transcript, including novel transcripts. Expression analysis also allows the validation study of transcriptome profiling by measuring the relative activity of known, predefined genes and transcripts. Thereby, it is now predictable that transcriptome profiling of plant diseases helps in understanding plant pathogen interaction and routes to effective implementation of crop improvement programme.

REFERENCES

1. Allie, F., Pierce, E. J., Okoniewski, M.J. and Rey, C. (2014). Transcriptional analysis of *South African cassava mosaic virus*-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. *BMC Genomics,* **15** : 1006
2. Asamizu, E., Nakamura, Y. and Tabata, S. (2005). Comparison of the transcript profiles from the root and the nodulating root of the model legume *Lotus japonicus* by serial analysis of gene expression. *Mole Plant-Microbe Interact.,* **18**(5): 487 – 498.
3. Backiyarani, S., Uma, S., Saraswathi, M.S., Saravanakumar, A.S. and Chandrasekar, A. (2015). Transcriptome analysis of banana (*Musa balbisiana*) based on next-generation sequencing technology. *Turk. J. Agric. For.* **39** : 705-717
4. Bai, T.T., Xie, W.B., Zhou, P.P., Wu, Z.L. and Xiao, W.C. (2013). Transcriptome and expression profile analysis of highly resistant and susceptible banana roots challenged with *Fusarium oxysporum* f. sp. *cubense* tropical race 4. *PLoS one,* **8**(9): 1-11.
5. Choi, H., Jo, Y., Lian, S., Jo, K.M., Chu, H., Yoon, J.Y., Choi, S.K., Kim, K.H. and Cho, W.K. (2015). Comparative analysis of chrysanthemum transcriptome in response to three RNA viruses: *Cucumber mosaic virus, tomato spotted wilt virus and potato virus X. Plant Mol. Biol.,* **88**(3): 233-248.
6. Flor, H.H. (1956). The complementary genic system in flax and flax rust. *Adv. Genet.,* **8**: 29-54.
7. Hong, K., Gong, D., Zhang, L., Hu, H., Jia, Z., Gu, H. and Song, K. (2015). Transcriptome characterization and expression profiles of the related defense genes in postharvest mango fruit against *Colletotrichum gloeosporioides. Gene,* **40947** : 1 – 9.
8. Jambagi, S. and Dunwell, J.M. (2015). Global transcriptome analysis and identification of differentially expressed genes after infection of *Fragaria vesca* with powdery mildew

(Podosphaera aphanis). *Transcriptomics,* **2**(1): 4 - 10

9. Kanter, I. and Kalisky, T. (2015). Single cell transcriptomics: methods and applications. *Fronteirs in Oncology,* **5**(53): 1 - 8
10. Li, C., Shao, J., Wang, Y., Li, W., Guo, D., Yan, B., Xia, Y. and Peng, M. (2013). Analysis of banana transcriptome and global gene expression profiles in banana roots in response to infection by race 1 and tropical race 4 of *Fusarium oxysporum* f. sp. cubense. *BMC Genomics,* **14** : 851
11. Li, C.Y., Deng, G.M., Yang, J., Viljoen, A., Jin, Y., Kuang, R.B., Zuo, C.W., Lv, Z.C., Yang, Q.S., Sheng, O., Wei, Y.R., Hu, C.H., Dong, T. and Yi, G.J. (2012). Transcriptome profiling of resistant and susceptible Cavendish banana roots following inoculation with *Fusarium oxysporum* f. sp. *cubense* tropical race 4. BMC Genomics, **13** : 374.
12. Martinelli, F., Uratsu, S.L., Albrecht, U., Reagan, R.L., Phu, M.L., Britton, M., Buffalo, V., Fass, J., Leicht, E., Zhao, W., Lin, D., D'souza, R., Davis, C.E., Bowman, K.D. and Dandekar, A.M. (2012). Transcriptome profiling of citrus fruit response to huanglongbing disease, PLoS one, **7**(10): 1-16.
13. Newcomb, R.D., Crowhurst, R.N., Gleave, A.P., Rikkerink, E.H.A., Allan, A.C., Beuning, L.L., Bowen, J.H., Gera, E., Jamieson, K.R., Janssen, B.J., Laing, W.A., Mcartney, S., Nain, B., Ross, G.S., Snowden, K.C., Souleyre, E.J.F., Walton, E.F. and Yauk, Y.K. (2006). Analyses of expressed sequence tags from apple. *Plant Physiol.*, **141** (1) : 147–166.
14. Passos, M.A.N., Oliveira De Cruz, V., Emediato, F.L., Camargo De Teixeira, C., Rennó Azevedo, V.C., Brasileiro, A.M.C., Amorim, E.P., Ferreira, C.F., Martins, N.F., Togawa, R.C., Pappas, G.J., Bonfim Da Silva, O. and Miller, R.G.N. (2013). Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. *BMC Genomics,* **14** : 78.
15. Riju, A. and Arunachalam, V. (2010). Electronic sorting of SNP /Indel Sites in expressed sequence tag libraries of cocoa (*Theobroma cacao* L.). *Genes, Genomes and Genomics,* **4** (1) : 37 - 40
16. Schimidtke, C., Findei, S., Sharma, C.M., Kuhfu, J., Hoffmann, S., Vogel, J., Stadler, P.F. and Bonas, U. (2011). Genome-wide transcriptome analysis of the plant pathogen Xanthomonas identifies sRNAs with putative virulence functions. *Nucleic Acids Res.,* 1–12
17. Zhukov, V.A., Ksulaeva, O.A., Zhernakov, I.A. and Tikhonovich (2015). Next generation sequencing for studying transcriptome profiles of tissues and organs of garden pea (*Pisum sativum* L.). *Agric. Biol.,* **50** (3) : 278 – 287.
18. Zuluaga, A.P., Sole, M., Lu, H., Castillo, E.G., 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.

❑

Citation : Priya B.T., Murthy N.S. and Divya B. (2016). Transcriptome profiling associated to plant diseases : a review. *HortFlora Res. Spectrum,* **5**(1) : 1-6.