Review Article

TRANSCRIPTOMICS -ROLE IN CROP IMPROVEMENT

ABSTRACT

Traditional breeding methods have been practiced for years, while more contemporary initiatives using markers-assisted breeding have made significant advancements that have helped us meet the world's food demands. With the depletion of natural resources, the threat of population growth, and the effects of climatic change, the demand for food will be extremely high in the future. But with urgent needs for food and other farm related produce, there is an extreme need to utilize genomic resources for precise crop breeding fully. Plant breeding can benefit in integrating transcriptome studies to improve crops for various biotic and abiotic challenges. Transcriptome studies can effectively link the phenotype with the genotypes of the observed traits, regardless of the genome availability and/or complexity of the crops facilitating the practical selection of superior genotypes for us.

Keywords: Transcriptomics and genomics.

INTRODUCTION

Global food security is threatened by rapid population expansion, global climate change, environmental degradation, drought, new illnesses, and saline soils (Nejat *et al*., 2018). Traditional breeding relies primarily on phenotypic selection based on the experiences of breeders, produces a large variety of high-yielding kinds. However, today's conventional plant breeding has significant challenges due to characteristics including high labor intensity, time requirements, ineffectiveness, and reliance on the environment, among others. The use of marker-assisted selection (MAS) in breeding programs has boosted breeding efficiency to some extent for many years (Xu and Crouch, 2008). A few of the MAS techniques that have been developed are selection for quantitative traits using markers at multiple loci, marker-assisted backcrossing or introgression of significant genes or quantitative trait loci (QTL), accumulation of favorable alleles in early generations, and selection for qualitative traits using markers at single loci (Eathington *et al*., 2007). To counteract the negative effects of these obstacles on agricultural output and increase crop yield and stress tolerance in plants, we must go beyond conventional and molecular plant breeding because both approaches have many limitations. To prevent unexpected consequences in modified crop plants that have a negative impact on human

health, discovering transcriptional regulatory elements and comprehending the mechanisms underlying transcriptional control are essential. Because of its high throughput, improved precision, and cost effectiveness during the past two decades, genomic sequence databases have increased massively (Lathe *et al*., 2008; Jain, 2012; Karsch-Mizrachi *et al*., 2018). Transcriptomics has been extensively studied in a variety of organisms and provides critical insights into gene structure, expression, and regulation (Lowe *et al*., 2017). In recent years, transcriptomics research has grown tremendously due to rapid progress in sequencing technologies (Wang *et al*., 2016).

REVOLUTIONIZING THE STUDY OF GENOME

Global scientists made outstanding progress toward genomic studies, to make it even more accurate and to experiment with its applications in agriculture, from ground-breaking discoveries to bold scientific breakthroughs. The past decade have seen enormous advances in our understanding of the many omics platforms, including genomics, epigenomics, metabolomics, transcriptomics, and proteomics, as sequencing technologies have developed tremendously. Additional research on the genome-wide transcriptional landscape of each gene has revealed the functional processes behind phenotypic variances. As a result, biological studies on high-throughput omics data began to run from the genomic level into transcriptional level; thereby achieving considerable attention towards transcriptome analysis.

UNRAVELING THE TRANSCRIPTOME

Transcriptome data includes biological information of gene transcriptional activities in a certain cell, a tissue, or an individual or a population of cells at certain status. The term “Transcriptome” was first attributed by scientist Charles Auffaray, to signify an entire set of transcripts i.e., ribonucleic acid molecule (RNA) which includes coding RNA (mRNA) and regulatory non-coding RNAs (tRNA, rRNA, etc.) (Morozova *et al*., 2009; Piétu *et al*., 1999). Transcriptome data provides entire gene sets as well as the way that they are expressed, regulated, and function in various cell populations. It includes spatio-temporal bioinformation influenced by diverse tissue types, developmental stages, and environmental and experimental events. Therefore, transcriptome data is more complex than genome data. Researchers can acquire a deeper understanding of the identification of novel genes, splice junctions, cleavage and splice variations, and differentially expressed genes (DEGs). The tools and strategies used to achieve food safety and sustainability align significantly with recent advances in transcriptome studies. This increased interest in the topic. Hence, scientists are developing comprehensive and sophisticated conception of the transcriptome in agriculture and allied fields. This chapter explores the broadening scope of plant transcriptomics over the decades, major objectives, techniques involved, and application of transcriptome-based plant selection in plant breeding.

BRIEF HISTORY OF TRANSCRIPTOMICS

The first attempts to study the whole transcriptome began in the early 1990s. Prior to the sequencing era, genetic and QTL mapping, chromosome walking, were the tools that were used for the identification of genes. Various transcriptomic approaches are in use today to identify genes that are differentially regulated globally in a genome. These approaches help to understand the molecular and cellular changes occurring in different tissues and cells in response to diverse factors. There are many key contemporary techniques in the field: traditional methods viz., Northern blot/RNA blot, nylon membrane assays, and later reverse transcriptase quantitative PCR (RT-qPCR) provided insights into a tiny subsection of a transcriptome. However, these methods were largely overtaken by advanced techniques viz., semi quantitative and quantitative real-time PCR (qRT-PCR). The efforts on this direction were further augmented with the advent of hybridization based approaches, such as micro-array technology, tiling-arrays, and next-generation sequencing-based approaches, for example expressed sequence tags (EST) sequencing, serial analysis of gene expression (SAGE), massive parallel signature sequencing (MPSS), and RNA-Sequencing (RNA-seq) (Rodrigues *et al*., 2014). The genome-wide transcription profiling have made it possible to characterize the transcriptome and to untangle molecular basis of gene regulation during stress to a very higher extend. This can be achieved by using medium- and high-throughput methods. Medium-throughput methods include complementary DNA (cDNA) clones and ESTs, whereas microarray and RNA-seq are high-throughput methods (Nagaraj *et al*., 2007). Although micro-arrays are useful for analyzing large transcriptome data; they can detect only known sequences, and cannot be used for gene discovery. RNA-seq, also called whole transcriptome shotgun sequencing, have taken central and expanding role in transcriptome studies using deep-sequencing techniques to profile trancriptome. This approach consist of converting RNA molecules to a library of cDNA fragments with adaptors, these fragments are sequenced, and resulting reads are either aligned to a reference genome, or assembled de novo (Wang *et al*., 2009).

Non-coding (ncRNAs) which were previously thought to be a part of junk genome, play several key roles in gene regulation including transcriptional and posttranscriptional regulation (Pertea, 2012). This discovery was possible only due to the advent of high-throughput RNA-seq analysis that provide insight into coding and non-coding transcripts, examine splicing and novel splicing alterations, identify single nucleotide polymorphisms (SNP), allele-specific expression, and gene fusions, and also performs quantification of the data generated (Bussotti *et al*., 2016). The complexity of the organism studied; the availability of a sequenced genome or prior knowledge of the transcriptome; and the ability to store, retrive and analyze of data and cost are the factors influencing the choice of techniques in transcriptome analysis.

TRANSCRIPTOME ANALYSIS FOR PLANT BREEDING

Since the early 2000s traditional breeding techniques have been employed to identify specific genes within a biological framework in typical plant biology investigations. Since high-throughput sequencing methods have been available, it is now possible to map transcripts onto

the genome and learn important details about gene structure, splicing patterns, and other transcriptional changes (Wang *et al*., 2009). This has led to the discovery of multiple genes, alleles, and splice variants in various organisms. RNA-seq for the study of gene expression, whole genome molecular marker development, and identification of markers in known function genes, particularly those conferring defense mechanisms against biotic and abiotic stresses, are all important uses of transcriptomics in the field of plant breeding. Associating a significant amount of genomic data with systemic characterization of phenotypes for a variety of traits and situations is a critical component of breeding that uses high-throughput sequencing. As a result, the genome sequencing of significant crops is turning into the first stage in determining the genome and evolution. The sequence can also be used to modify individual genes with the aid of CRISPR/Cas9, ZFNs, TALENs, and other genome editing tools, or it can be used to find the right mutations to create a new allelic form (Vlk & Epková, 2017). Only phenotypic selection is used in conventional agricultural breeding. Linkage maps are used in classical QTL analysis to find genetic areas linked to population-level phenotypic differences. Nowadays, a method known as "Genetic genomics" has been used to find expressed QTLs (eQTLs), which are genes and genetic regulator loci that code for the observed variation (Joosen *et al*., 2009). RNA-seq has been used successfully as an alternative to whole genome sequencing for crop species with complex genomes, including coffee, wheat, sugarcane, and several "orphan crops," including sweet potato, chickpea, and pearl millet. It unravels genes and regulatory regions and provides a variety of molecular markers (M. Perez-de-Castro *et al*., 2012).

WHEN TO USE TRANSCRIPTOMICS IN PLANT BREEDING?

In the absence of the complete genome sequence, transcriptome analysis would improve our understanding of gene function. The use of transcriptome methods in agricultural research has been widespread and interdisciplinary. Worldwide, the main threats to sustainable agricultural productivity and food security are rapid population expansion, global climate change, environmental degradation, drought, emerging diseases, and saline soils (Nejat *et al*., 2018). We need to go beyond conventional and molecular plant breeding in order to reduce the adverse effects of these risks to agricultural output, increase crop yield, and improve plant stress tolerance. The most sophisticated methods for determining how the plasticity of gene expression respond to a specific environmental or developmental stimulation are transcriptomics. In order to avoid unexpected consequences in transformed crop plants, which could ultimately have negative impacts on human health, it is essential to identify transcriptional regulatory components and understand the mechanisms underlying transcriptional regulation. The study of transcriptomes has a profound impact on many areas of biological sciences because it enables quantitative and qualitative analysis of variations in the gene expression of several mRNAs (Tan & Yiap, 2009). Transcriptomics can now be the beginning of any large-scale sequencing endeavor rather than its end. It is now simple, affordable, and quick. Comparison of transcriptomes allows the identification of genes that are differentially expressed in distinct cell populations, or in response to different treatments. Researchers can quantify the variations in

gene expression between various biological contexts and identify which sets of genes are switched on or off in a given state through transcriptome investigations. The main goals of transcriptomics are to catalog all transcripts, including mRNAs, noncoding RNAs, short RNAs, the 52 and 32 end sites of the genome, posttranscriptional modifications, and splicing patterns to ascertain the transcriptional status of genes. According to Wang *et al*. (2009), transcriptomics also aims to quantify changes in gene expression levels caused by different stress conditions and developmental stages.

APPLICATION OF TRANSCRIPTOMICS IN PLANT BREEDING

In plant research, there has been a lot of interest in the identification and characterization of differential gene expression from tissues subjected to stress. In addition to elucidating the genes and their network involved in various forms of stress tolerance, several developmental processes of plant components have also been studied. The data generated by transcriptome analysis provided marker information on regulatory gene networks, which can subsequently be used for breeding improved varieties. Table 1 shows case studies on the use of transcriptome studies in plant breeding, along with the methodology and methodologies applied.

Transcriptome assembly and profiling is the widespread use of transcriptome sampling strategies is a complementary approach to genome sequencing, and results in a large collection of expressed sequence tags (ESTs) for almost all the important plant species (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). The plant EST database has recently passed the five million sequence landmark. More than 50 plant species, each with >5000 ESTs, are represented. Small RNA characterization is also one of the major application have a role in development, genome maintenance and plant responses to environmental stresses. Most sRNAs belong to two major groups: 1) microRNAs (miRNA) are about 21 nt and usually have a post-transcriptional regulatory role by directing cleavage of a specific transcript 2) short interfering RNAs (siRNA) are usually 24 nt-long and influence de novo methylation or other modifications to silence genes. The finding of their prevalence in low-molecular-weight fractions of total RNA in animals and plants predated the development of NGS. Another major application expression (eQTL). Metabolite, protein and transcript profiles can also be directly mapped onto a segregating population to provide information on loci that control gene expression levels, protein modification or levels of a particular secondary metabolite. The QTLs associated with such traits are known as expression (eQTL), protein (pQTL) or metabolite (mQTL)

LIMITATION OF TRANSCRIPTOMICS

To create a whole transcriptome, transcripts must be accurately identified and quantified using approaches that are objective, highly effective, and economical. While transcriptomics has enormous potential benefits, it also raises enormous ethical issues. The key limitations of this

approach is the humongous amount of data generated. The most crucial step is to simulate the networks and pathways predicted by transcriptome analysis in order to determine the precise functions and locations of all the genes in a plant. A full-scale availability and maintenance of the public repositories and databases containing the data generated by transcriptome studies is the need of the hour. Utilizing the relevant information associated with their own experiments can be helpful to the researchers.

CONCLUSION AND FUTURE GOALS

Transcriptomics has revolutionized our understanding of genome expression and how the genes are identified and expressed. Over the past three decades, new technologies have redefined what is possible to look at, and integrations with other "omics" employing more complex methods, like system biology, are providing an increasingly integrated view of the complexity of cellular life. This approach will be a great promise of “transcriptomics-assisted breeding” in the 21st century. The gradual diversification of plant transcriptomics' uses for studying various aspects of plant development has set up opportunities in which knowledge from various fields—including physiology, biochemistry, and evolution—is being combined to produce significant conclusions that will advance ongoing research projects. Small laboratories can now conduct transcriptomics studies because of their falling cost, and large-scale transcriptomics-consortia may conduct studies comparing the transcriptomes of thousands of different organisms, tissues, or environmental conditions. With the advancement of sequencing tools, this tendency is likely to persists.

Table 1: Influence of different stresses for crop improvement

Crop & Type of stress	Research goal	Methodology	Findings	References

Rice Drought stress	To examine the functions of OsC3H10 gene	RNA sequencing	Over expression of OsC3H10 results in increased expression of certain gene account for stress related responses viz., PATHOGENESIS RELATED GENEs (PRs) LATE EMBRYOGENESIS ABUNDANT (LEAs) PROTEINS and GERMIN-LIKE PROTEINS (GLPs)	(Seong *et al.*, 2020)
Chickpea Biotic stress	To learn how a necrotrophic chickpea pathogen, Ascochyta rabei, survives in the face of oxidative stress	RNA seq	The transcriptome profiling found five genes linked to stress and fungal pathogencity: ST47 g10291, ST47 g9396, ST47 g10294 and ST47 g4395	Maurya *et al.*, 2020
Barley Salt stress	To uncover mechanisms underlying salt tolerance	RNA-seq analysis	In comparision to H.vulgare, ion transporters such HKT1;1, HKT1;5, HKT2;2 and SOS1 were found the most important factors in salt stress tolerance to H. marinum.	(Huang *et al.*, 2018)

Mustard Biotic stress	To learn more about the involvement of NAC TFs in plant defese against the ALternaria brassicicola-caused 'Black Spot' disease in oilseed mustard	qRT- PCR	Despite being evolutionarily conserved in modulating desiccation and wound response, NACs in B. juncea and S. alba demonstrate variance in early signaling upon pathogen detection.	(Mondal *et al.*, 2020)
Cotton Salt stress	Investigated by identification and followed with study of NHX genes	qRT-PCR	GhNHXs revealed diverse expression fashion in different level of salinities.	Fu *et al.*, 2020
Wheat Salt	To detect abiotic stress tolerance regulated transcriptional factor (TF) genes in wheat	RNA-sequencing	A salt stress responsive gene TabZIP15 (TraesCS7A02G488600), was identified	Bi *et al.*, 2021
Ground nut Biotic stress	To investigate the mechanisms of Sclerotium rolfsii resistance in peanuts (Arachis hypogea L.)	RNA-seq	PAMP-triggered immunity has been discovered as a potential important resistance mechanism against sclerotium rolfsii in peanut with the jasmonic acid signaling pathway being the most likely defence mechanism	Bosamia *et al.*, 2020

REFERENCES

Bhide, A., Schliesky, S., Reich, M., Weber, A. P. M and Becker, A. 2014. Analysis of the floral transcriptome of Tarenaya hassleriana (Cleomaceae), a member of the sister group to the Brassicaceae: Towards understanding the base of morphological diversity in Brassicales. *BMC Genomics*. 15(1): 1–17. https://doi.org/10.1186/1471-2164-15-140

Bi, C., Yu, Y., Dong, C., Yang, Y., Zhai, Y., Du, F., Xia, C., Ni, Z., Kong, X and Zhang, L. 2021. The bZIP transcription factor TabZIP15 improves salt stress tolerance in wheat. *Plant Biotechnology Journal*. 19(2): 209.

Bosamia, T. C., Dodia, S. M., Mishra, G. P., Ahmad, S., Joshi, B., Thirumalaisamy, P. P., Kumar, N., Rathnakumar, A. L., Sangh, C and Kumar, A. 2020. Unraveling the mechanisms of resistance to Sclerotium rolfsii in peanut (Arachis hypogaea L.) using comparative RNA-Seq analysis of resistant and susceptible genotypes. *PLoS One*. 15(8): e0236823.

Bussotti, G., Leonardi, T., Clark, M. B., Mercer, T. R., Crawford, J., Malquori, L., Notredame, C., Dinger, M. E., Mattick, J. S and Enright, A. J. 2016. Improved definition of the mouse transcriptome via targeted RNA sequencing. Genome Research, 26(5): 705–716.

Cheng, L., Li, S., Yin, J., Li, L and Chen, X. 2013. Genome-Wide Analysis of Differentially Expressed Genes Relevant to Rhizome Formation in Lotus Root (*Nelumbo nucifera* Gaertn). *PLoS ONE*: 8(6). e67116. https://doi.org/10.1371/ journal.pone.0067116

Eathington, S.R., Crosbie, T.M., Edwards, M.D., Reiter, R.S and Bull, J.K. 2007. Molecular markers in a commercial breeding program. *Crop Sci*. 47:154–163

Fu, X., Lu, Z., Wei, H., Zhang, J., Yang, X., Wu, A., Ma, L., Kang, M., Lu, J and Wang, H. 2020. Genome-wide identification and expression analysis of the NHX (sodium/hydrogen antiporter) gene family in cotton. *Frontiers in Genetics*. 964.

Guggisberg, A., Lai, Z., Huang, J and Rieseberg, L. H. 2013. Transcriptome divergence between introduced and native populations of Canada thistle, Cirsium arvense. New Phytologist, 199(2): 595–608. https://doi.org/10.1111/nph.12258

Huang, L., Kuang, L., Li, X., Wu, L., Wu, D and Zhang, G. 2018. Metabolomic and transcriptomic analyses reveal the reasons why Hordeum marinum has higher salt tolerance than *Hordeum vulgare*. Environmental and Experimental Botany. 156: 48–61.

Jain, M. 2012. Next-generation sequencing technologies for gene expression profiling in plants. *Brief. Funct. Genomics.* 11: 63–70.

Joosen, R., Ligterink, W., Hilhorst, H and Keurentjes, J. 2009. Advances in Genetical Genomics of Plants. *Current Genomics*. 10(8): 540–549.

Karsch-Mizrachi, I., Takagi, T and Cochrane G. 2018. International nucleotide sequence database collaboration. the international nucleotide sequence database collaboration. *Nucleic Acids Res.* 46: D48–D51.

Kim, S. H., Kumpatla, S. P., Buyyarapu, R., Abdurakhmonov, I. Y and Mammadov, J. A. 2012. Genomics-assisted plant breeding in the 21st century: technological advances and progress. In Plant breeding. Intechopen.

Lathe, W., Williams, J., Mangan, M and Karolchik D. 2008. Genomic data resources: challenges and promises. *Nat. Educ.* 1:2.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S and Shafee T. 2017. Transcriptomics technologies. *PloS Comput. Biol.* 13: e1005457

M. Perez-de-Castro, A., Vilanova, S., Canizares, J., Pascual, L., M. Blanca, J., J. Diez, M., Prohens, J and Pico, B. 2012. Application of Genomic Tools in Plant Breeding. *Current Genomics.* 13(3): 179–195. https://doi.org/10.2174/ 138920212800543084

Maurya, R., Singh, Y., Sinha, M., Singh, K., Mishra, P., Singh, S. K., Verma, S., Prabha, K., Kumar, K and Verma, P. K. 2020. Transcript profiling reveals potential regulators for oxidative stress response of a necrotrophic chickpea pathogen Ascochyta rabiei. 3 Biotech. 10(3): 1–14.

Mondal, B., Mazumder, M., Mukherjee, A., Ghosh, S., De, A., Bose, R and Basu, D. 2020. Association of Alternaria brassicicola induced NAC transcription factors with desiccation and wound responses in Indian mustard. *Physiological and Molecular Plant Pathology*. 112: 101540.

Morozova, O., Hirst, M and Marra, M. A. 2009. Applications of new sequencing technologies for transcriptome analysis. Annual Review of Genomics and Human Genetics. 10: 135–151. https://doi.org/10.1146/annurev-genom-082908- 145957

Nagaraj, S. H., Gasser, R. B and Ranganathan, S. 2007. A hitchhiker's guide to expressed sequence tag (EST) analysis. Briefings in Bioinformatics. 8(1): 6–21. https://doi.org/10.1093/bib/bbl015

Nejat, N., Ramalingam, A and Mantri, N. 2018. Advances in transcriptomics of plants. *Plant Genetics and Molecular Biology*. 161–185.

Pertea, M. 2012. The human transcriptome: An unfinished story. *Genes*. 3(3): 344–360. https://doi.org/10.3390/genes3030344

Piétu, G., Mariage-Samson, R., Fayein, N. A., Matingou, C., Eveno, E., Houlgatte, R., Decraene, C., Vandenbrouck, Y., Tahi, F., Devignes, M. D., Wirkner, U., Ansorge, W., Cox, D., Nagase, T., Nomura, N and Auffray, C. 1999. The genexpress IMAGE knowledge base of the human brain transcriptome: A prototype integrated resource for functional and computational genomics. *Genome Research*. 9(2): 195–209. https://doi.org/10.1101/gr.9.2.195

Rodrigues, C. M., Mafra, V. S and Machado, M. A. 2014. Transcriptomics. Omics in Plant Breeding. 33–57.

Sangwan, R. S., Tripathi, S., Singh, J., Narnoliya, L. K and Sangwan, N. S. 2013. De novo sequencing and assembly of Centella asiatica leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. *Gene*. 525(1): 58–76. https:// doi.org/10.1016/j.gene.2013.04.057

Seong, S. Y., Shim, J. S., Bang, S. W and Kim, J.K. 2020. Overexpression of OsC3H10, a CCCH-zinc finger, improves drought tolerance in rice by regulating stress-related genes. Plants. 9(10): 1298.

Sharma, R., Agarwal, P., Ray, S., Deveshwar, P., Sharma, P., Sharma, N., Nijhawan, A., Jain, M., Singh, A. K., Singh, V. P., Khurana, J. P., Tyagi, A. K and Kapoor, S. 2012. Expression dynamics of metabolic and regulatory components across stages of panicle and seed development in *indica* rice. *Functional and Integrative Genomics*. 12(2): 229-248. https://doi.org/10.1007/ s10142-012-0274-3

Tan, S. C and Yiap, B. C. 2009. DNA, RNA and protein extraction: The past and the present. *Journal of Biomedicine and Biotechnology*. 2009. https:/ /doi.org/10.1155/2009/574398.

Vlk, D and Øepková, J. 2017. Application of next-generation sequencing in plant breeding. *Czech Journal of Genetics and Plant Breeding*. 53(3): 89-96. https://doi.org/10.17221/192/2016-CJGPB

Wang, B., Tseng, E., Regulski, M., Clark, T.A., Hon, T and Jiao Y. 2016. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun*. 7:11708.

Wang, Z., Gerstein, M and Snyder, M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics. 10(1): 57–63. https://doi.org/ 10.1038/nrg2484.

Xu, Y and Crouch J.H. 2008. Marker-assisted selection in plant breeding: from publication to practice. *Crop Sci*. 48: 391-407.

Xue, L. J., Zhang, J. J and Xue, H. W. 2012. Genome-wide analysis of the complex transcriptional networks of rice developing seeds. PLoS ONE. 7(2): e31081. https://doi.org/10.1371/journal.pone.0031081