



OPEN ACCESS

EDITED BY
Md. Anwar Hossain,
University of Rajshahi, Bangladesh

REVIEWED BY
Jlabao Ye,
Yangtze University, China
Haidong Yan,
University of Georgia, United States

*CORRESPONDENCE
Rajiv Ranjan
rajivranjanbt@gmail.com

SPECIALTY SECTION
This article was submitted to
Plant Biotechnology,
a section of the journal
Frontiers in Plant Science

RECEIVED 29 August 2022
ACCEPTED 27 October 2022
PUBLISHED 15 December 2022

CITATION
Tyagi P, Singh D, Mathur S, Singh A
and Ranjan R (2022) Upcoming
progress of transcriptomics studies on
plants: An overview.
Front. Plant Sci. 13:1030890.
doi: 10.3389/fpls.2022.1030890

COPYRIGHT
© 2022 Tyagi, Singh, Mathur, Singh and
Ranjan. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Upcoming progress of transcriptomics studies on plants: An overview

Parul Tyagi, Deeksha Singh, Shivangi Mathur, Ayushi Singh and Rajiv Ranjan*

Plant Molecular Biology Lab, Department of Botany, Faculty of Science, Dayalbagh Educational Institute, Agra, India

Transcriptome sequencing or RNA-Sequencing is a high-resolution, sensitive and high-throughput next-generation sequencing (NGS) approach used to study non-model plants and other organisms. In other words, it is an assembly of RNA transcripts from individual or whole samples of functional and developmental stages. RNA-Seq is a significant technique for identifying gene predictions and mining functional analysis that improves gene ontology understanding mechanisms of biological processes, molecular functions, and cellular components, but there is limited information available on this topic. Transcriptomics research on different types of plants can assist researchers to understand functional genes in better ways and regulatory processes to improve breeding selection and cultivation practices. In recent years, several advancements in RNA-Seq technology have been made for the characterization of the transcriptomes of distinct cell types in biological tissues in an efficient manner. RNA-Seq technologies are briefly introduced and examined in terms of their scientific applications. In a nutshell, it introduces all transcriptome sequencing and analysis techniques, as well as their applications in plant biology research. This review will focus on numerous existing and forthcoming strategies for improving transcriptome sequencing technologies for functional gene mining in various plants using RNA-Seq technology, based on the principles, development, and applications.

KEYWORDS

NGS, transcriptome analysis, gene function, molecular markers, secondary metabolites

Introduction

Since the advent of the post-genomic era, numerous omics approaches have been developed, such as genomics, transcriptomics, metabolomics, and proteomics. As one of these technologies, transcriptomics is the second earliest and most commonly used (Lockhart and Winzler, 2000; Levy and Myers, 2016). Transcriptomics studies focus on

the transcriptome. 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). However, the complex mapping of a genome to various phenotypes, tissues, developmental stages, and environmental factors continues to pose a major challenge in molecular biology. A more in-depth understanding of gene regulation transcripts and expression is not only difficult but is also at the core of the problem. 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; Abdel-Ghany et al., 2016).

In plant research, there are some significant differences between transcriptomics and genomics. First, genome assembly is more complex and costly in plant research compared to transcriptomics (RNA-Seq); in the absence of a reference genome, the transcriptome can be used to assess an organism or plant's overall transcriptional activity. Second, the transcriptome changes over time and space because it includes information on secondary metabolic pathways as well as variations in gene expression at different times and spatial regions. According to research, the accumulation of biologically active compounds in different plant tissues depends on the gene expression levels of that tissue as well as the time they are produced. This is because plants have various growth conditions and periods, even within the same species. As a result, the transcriptome outperforms the genome in terms of identifying genes related to therapeutic plant components (Wang et al., 2009). These differences are important for studies on plant functional genome mining, gene regulatory domains development, genetic diversity (dominant and recessive genes), and bioactive compounds (Tyagi and Ranjan, 2021; Tyagi et al., 2022).

Transcriptome study approaches have progressed from basic DNA microarrays platform to RNA-Seq technology in recent times due to the continual developments in sequencing technology (Mironova et al., 2015). It has several perks including high sensitivity, high throughput, and efficacy, which may be used to analyses a complete transcriptome without a genomic reference sequence. RNA-Seq technology is a popular sequencing approach in molecular biology, biotechnology, and bioinformatics (Sun and Wei, 2018). This method has been widely exploited in model plants, including *Arabidopsis thaliana* (Zhang et al., 2019), *Oryza sativa* (Li et al., 2012; Pradhan et al., 2019; Yang et al., 2021), *Zea mays* (Xu B et al., 2014; Liu et al., 2019; Xu et al., 2021), *Rehmannia glutinosa* (Ma et al., 2021), *Polygonum cuspidatum* (Wang et al., 2021), *Asarum sieboldii* (Chen et al., 2021), *Calotropis gigantea* (Hoopes et al., 2018). Transcriptome can be defined as the total number of RNA molecules transcribed from a particular tissue or cell at a given

functional or developmental stage, which includes mRNA (messenger RNA) and nc-RNA (non-coding RNA). They are referred to as “bridges” because they accurately manage the transport of genetic information from DNA to protein (Costa et al., 2010), whereas non-coding RNA influences gene expression, protein synthesis, and various cellular processes at several levels (Kumar et al., 2016). Understanding transcriptomics thus enhances research with reference to the functions of cells, tissues, and organisms. RNA-Seq is a fairly new technology that measures all the biological amounts of the transcriptome. This makes it easier to study the transcriptome (Wang et al., 2009).

Computing and analysing RNA-seq still remains a challenging task. Accurate alignment of sequencing reads, and accurate expression level inference are some of these challenges. In RNA-seq, the central computational challenge is accurately and efficiently assigning short sequencing reads to their transcripts and determining gene expression based on that information. There have been some benchmark studies conducted, but they have generally been conducted using simulated RNA-seq datasets or RNA-seq datasets focusing only on long RNAs, such as messenger RNAs and long noncoding RNAs. Therefore, they did not assess the suitability of these tools for measuring total RNA in datasets that included small RNAs like transfer RNAs and small nucleolar RNAs. This review paper focuses on various methodologies, including transcriptome sequencing. Additionally, it provides an overview of the advancements made in using the RNA-Seq approach in plant research, including the molecular markers identification, functional gene expression analysis, biosynthetic pathways of secondary metabolite, and mechanisms for plant development.

Techniques of next generation sequencing (NGS) and their potential applications

RNA-seq: A history and evolution

Sequencing of events is a crucial comprehension method for a variety of reasons. Researchers of various abilities can efficiently organize material and ideas using sequencing structures. It is crucial to problem-solving in all subject areas, including science and social studies. So, differences of sequencing are coming according to generation, in first-generation sequencing platform, enabled sequencing of clonal DNA populations, in second generation sequencing, massively increased throughput by parallelizing many reactions, and in third generation sequencing, allow direct sequencing of single DNA molecules. In the 1970s, Walter Fiers and colleagues marked the start of RNA-Seq studies by sequencing the MS2

bacteriophage (3,569 nucleotides) whole transcriptome (Crowgey and Mahajan, 2019). RNA is highly unstable and susceptible to breakdown by cell membrane-associated RNases due to its single-stranded structure. Therefore, extensive efforts are required to sequence the whole transcriptome accurately. The discovery of reverse transcriptase (Zhang, 2019) enables the transformation of messenger RNA and non-coding RNA into stabilized DNA. The reverse-synthesized DNA is referred to as complementary DNA (cDNA). Sanger, a British scientist, invented DNA sequencing technology in 1975 (Zhang, 2019; Sanger and Coulson, 1975). Iscoe (Hwang et al., 2018) used the technique of RNA-Sequencing PCR, which exponentially amplifies cDNA. As a result of these pioneering efforts, microarray (chip) technology was developed (Bolón-Canedo et al., 2019). Through Arrayed technology, hundreds of known partial DNA sequences can be mounted on nylon membranes or solid support slides using molecular hybridization technology. Numerous genes can be quantitatively detected through hybridization. The reverse synthesis of cDNA can rapidly be used to sequence the transcriptomes of different biological samples, making microarrays an excellent way of exploring gene function (Clark et al., 2002). As of now, the technology is primarily used to study known genes, rather than unknown genes. Additionally, the microchip has difficulty detecting numerous transcripts generated by alternative splicing.

Sanger sequencing is an expensive, laborious, and time-consuming technique. NGS sequencing, however, has met a number of critical needs for molecular biologists since 2006, when it was introduced. In comparison to first-generation sequencing technology, NGS is fast, high-throughput, and economical. It is possible to sequence millions or billions of nucleic acids simultaneously, enabling the analysis of transcriptomes and genomes of all species using NGS. As of today, NGS is used for constructing different organism genomes for the purpose of collecting the complete gene sequences of various species of plants, including SARS-CoV-2 (Li et al., 2020) and *Chosenia arbutifolia* (Khoar), a Korean plant (Chen et al., 2014; Mei et al., 2016; Feng et al., 2019). Next generation sequencing is predominantly employed to sequence and analyze messenger RNA (mRNA) and small RNA (small RNA) from the transcriptome. As indicated in Table 1 (Zhang et al., 2016), the Roche 454, Illumina Solexa, and ABI SOLiD are all prominent next-generation sequencing platforms.

In 2005, Roche introduced the high-throughput Roche 454 sequencing technology, based on pyro-sequencing technique (Margulies et al., 2005). A longer read length and short run time make 454 sequencing stand out from other NGS technologies. This approach is non-fluorescent and does not require nucleic acid probes. However, it has the potential to introduce errors such as deletions or insertions during the

TABLE 1 Comprehensive overview of various next-generation sequencing (NGS) technologies.

Name of Sequencer	Sanger Sequencing	Roche 454 sequencing	Solexa-Illumina	ABI SOLiD	SMRT	Nanopore
Based on Principle	"Dideoxy chain termination"	"Sequencing by synthesis"	"Sequencing by synthesis"	"Sequencing by ligation"	"Sequencing by synthesis"	"Electrical signal sequencing"
Advantage	Higher sensitivity to detect low-frequency variants	Long read length, ferments are generated in large number	Enables a wide variety of applications, allowing researchers to ask virtually any question related to the genome, transcriptome, or epigenome of any organism	High accuracy, Coverage is more than 30x, Detecting targeted re-sequencing, and transcriptome sequencing	Short time-consuming and no need for PCR amplification	Ability to produce ultra-long reads
Disadvantage	Sequencing could only small DNA sequence, higher cost, and minimum high throughput	More than 6 error rates with polybases, high cost, and low throughput	Short reads length, low multiplexing capability, only highly trained person can operate	Short read length, prone to chain decoding error	High cost and minimum high throughput	High single-base error rate and long sequencing time
Length of base pair	15-40bp	300-800 bp	2×150bp	~75 bp	8-15 to 40-70kb	500bp-2.3Mb
Developed by and year	Frederick Sanger and colleagues in 1977	Shankar Balasubramania and David Klenerman in 2005	Shankar Balasubramanian and David Klenerman in 2006	Robert C. Martin (also known as Uncle Bob) in 2006	Christian Henry, and John F. Milligan in 2004	Hagan Bayley Clive G. Brown et al., in 2005
Run time	4-5hrs	24 hrs	6-7 days	6-7 days	~6 hrs	10 minutes
Output data	1.9~84 Kb	~7 Gb	600 Gb	120 Gb	5~10Gb	More than 50 Gb
References	Sanger et al., 1977	Zhang et al., 2016	Wang et al., 2009	Xu, 2018	Li et al., 2018	Ma et al., 2019

process of sequencing (Liang et al., 2017). The method utilized by Solexa (Illumina) primarily rely on the principle of “sequencing by synthesis”, in which DNA fragments are randomly linked on a flow cell. Following extension and amplification, the surface of the glass generates millions of clusters containing thousands of identical DNA fragments. Sequencing of labeled dNTPs (fluorescence) is performed on the stretched DNA strand. This technology permits the use of synthetic probes and reference sequences in genome-wide expression analysis (Liu et al., 2012).

There are a few drawbacks to the platform, including its shorter read length as well as its complexity in assembling reads from scratch. With solid sequencing (2007), magnetic beads are used to sequence data in a highly parallel manner. This approach allows massive DNA amplification without incurring high costs because of constant ligation and development of fluorescently labeled oligonucleotides (Li and Xu, 2019). While high precision is the significant benefit of this technology, the primary disadvantage is that it is prone to chain decoding errors if an error occurs, as mentioned in Table 1.

Third-generation sequencing has been enabled ‘by improvements in sequencing technologies’. In addition to amplification requirements for NGS platforms, template relocation and mismatches in nucleotides and GC percentages are common challenges related to these approaches. These challenges reduce the accuracy and completeness of our sequencing data. Because third-generation sequencing offers such long reads, it has been widely used in structural variation, methylation, transcriptome and genomic analysis, among others. SMRT (Single- molecule real-time) and Nanopore sequencing have recently become key technologies for third-generation sequencing (Ma et al., 2019). As a sequencing method, the SMRT employs the “synthesis-based principle.” SMRT technology has two unique properties compared to classical methods. This was followed by adding and linking the fluorescent group to the phosphoric acid (H_3PO_4) group to eliminate the background noise. As a second benefit, there is no need for amplification and the SMRT allows for more precise measurements through self-correction (Li et al., 2018). On the other hand, SMRT sequencing technology has several disadvantages, one of which being the integration of random errors. By using circular consensus sequencing (CCS) and improving sequence coverage, it is now possible to increase the accuracy of a single read to 99.8% in comparison to second generation sequencing (Wenger et al., 2019).

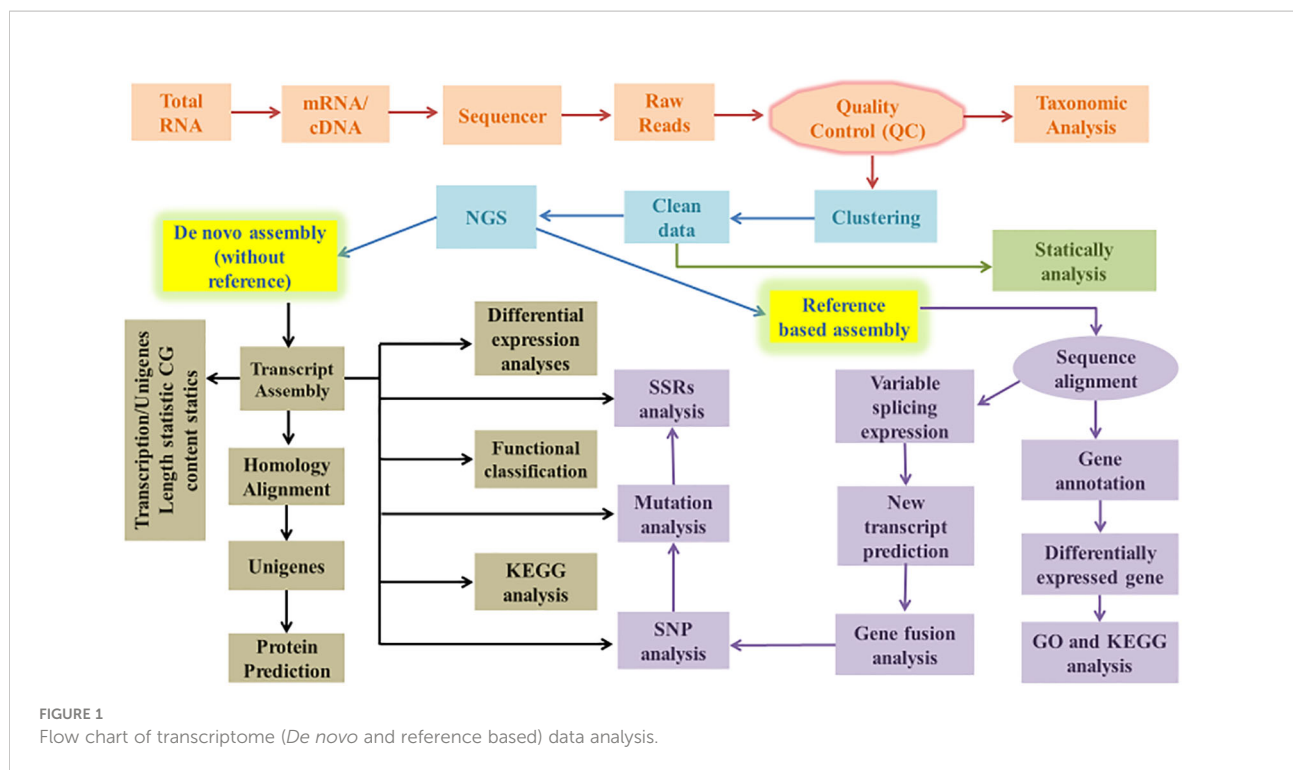
Nanopore is a next-generation SMRT sequencing platform that determines the base composition mostly through changes in electrical impulses. Nanopore sequencing is more cost-effective and offers longer reads than other platforms. However, the major drawback of this technology is its high error rate which ranges from 5-20% (Huang et al., 2021). Despite its advantages, it has

some disadvantages, including a high probability of random errors and single-base errors as mentioned in Table 1. Using the above approaches without a reference sequence can be cost-prohibitive and time-consuming. RNA-Seq is a crucial part of NGS technology and an important tool for transcriptome analysis. It overcomes the limitations of microarray analysis and provides a greater understanding of transcriptome research.

Sample preparation and assembly for RNA-seq

The total RNA from biological samples is isolated using kits or manually (TRIzol) protocols (Drygin et al., 2021). m-RNA was extracted from total RNA “Experiments should be employed in triplicate form.” The enriched mRNA was fragmented and converted into first-strand cDNA, which was then followed by second-strand generation, A-tailing, adaptor ligation, and a limited number of PCR amplifications of the adaptor-ligated libraries. The amplified libraries were analyzed on the Bio-analyzer according to the instructions of the manufacturer. Quantification and qualification of the library were performed using any sensitivity kit by the Nanodrop spectrophotometer. NGS for cDNA of all plants was performed by library on any sequencing platform according to samples. The raw reads were first filtered to exclude the reads containing adaptors or with ambiguous nucleotides (‘N’). Below than 20% Q < 20 nucleotide bases were also trimmed. The obtained high-quality (HQ) clean reads were used to construct a sequence assembly using Trinity (Tyagi and Ranjan, 2021) or any other software. After removing duplicate Trinity-generated sequences with the TGICL, clusters and unigenes were recovered, as illustrated in Figure 1. Cogent creates a k-mer profile of non-redundant (NR) transcripts, calculates pairwise distances, and then categorizes transcripts into families based on k-mer similarity. Using any of the graph techniques, each transcript family was further reconstructed into one or more distinct transcript model (s).

Transcriptome assembly is required for a variety of downstream analyses. Raw transcriptome data is error-prone due to the vast amount of transcriptome data (Nagalakshmi et al., 2008). Thus, selecting assemblies based on distinct transcriptome data and research ideas is important. The primary thing to consider when assembling is whether there is a reference sequence available; assembly can be classified into two types, reference-based assembly and *de novo* assembly as shown in Figure 1. It is impossible to carry out the reference assembly for most plants because their genomic sequence data is frequently lacking; therefore, many non-model species are not provided with genomic sequence information. The only assembly approach appropriate for non-model plants is *de novo* assembly, which is depicted in Figure 1.



The assembly and function annotation process

Software such as Cufflinks and Scripture (Guo et al., 2021) is used to assemble genomes, whereas Oases, Trinity, Rnnotator (Martin et al., 2010), and SOAPdenovo-Trans (Madritsch et al., 2021), are used to assemble *de novo* genomes (Martin et al., 2010; Schulz et al., 2012; Grabherr et al., 2013; Trapnell et al., 2013; Xie et al., 2014). Assembling a single K-mer method with SOAPdenovo-Trans and Trinity is one stage of the *de novo* (without reference) assembly, while multiple K-mer methods are assembled with Oases and Rnnotator. Numerous k-mer methods can be utilized for obtaining huge transcripts of data, but when the findings of numerous k-mer methods are merged, a huge amount of duplicated data is generated, thus increasing the error rate and data complexity. As a result, at present time, a single K-mer assembly method is the most significant for improved precision that can be adopted, as mentioned in Table 2 (Strickler et al., 2012). SOAPdenovo-Trans connect to the contigs using an unknown nucleotide base (N). This approach is less accurate than the other three software programs in terms of assembly correctness, does not produce longer transcripts, and has a lower average accuracy rate (Lu, 2013). In recent years, Trinity has become an increasingly popular *de novo* assembly program due to its ability to produce accurate and efficient assembly outputs. In that respect, Scripture and Cufflinks use sensitive and conservative assembly methods, respectively. Despite the superior quality of

Cufflinks' assembly, Scripture's transcript quantity greatly exceeds that of Cufflinks, as shown in Table 2. Cufflinks can provide higher accuracy and quality assembly results by using high-quality reference sequences.

Gene function annotation is the process of identifying the functions related with individual gene based on existing data by comparing unidentified gene sequences to those published in public databases. GenOntology (GO) (Yu, 2020) and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Yousef et al., 2021) are the two widely used approaches for functionally categorizing gene functions. A GO is divided into three categories, namely biological processes (BP), cellular compounds (CC) and molecular functions (MF). Unigene sequences are commonly annotated by databases such as Non-Redundant (NR), Protein Sequence Database (Zhao et al., 2021), GO, KEGG, Clusters of Orthologous Groups (COG) (Lee and Lee, 2021), Pfam (Mistry et al., 2021), NCBI Nucleotide Sequence Database (Schoch et al., 2020), and Swiss-Prot database (Liu FX et al., 2018).

Application of transcriptome sequencing

The RNA-Seq method provides insight into gene expression under a variety of environments and enables the discovery of new genes (Clamp et al., 2007; Mortazavi et al., 2008; Shendure, 2008; Pickrell et al., 2010; Touch et al., 2010; Filichkin et al.,

TABLE 2 List of softwares used in *De novo* assembly pipeline with their functions.

Name of Software	Raw data	Data filtration	Output
Through Sequencer (Leng et al., 2012)	RNA-Seq Raw Reads	Short Reads - Obtained from the common NGS platforms, including Illumina, SOLiD and 454, are often very short bases (35-500 bp). Long Reads - Oxford Nanopore/PacBio Sequencers can sequence up to long 5 to 100Kb reads.	Massively parallel millions to billions sequence that offers high- throughput, scalability, and takes lesser time.
FastQc (Andrews, 2010)	Quality Check (QC)	Quality assessment -Evaluate the raw read quality, identify the adaptor contaminations, and identify low quality samples	Good/Bed - According to Phred Score quality (Q-value)
Trimalore	Read clean up (If contaminated)	Trimming - Removes the bad bases (adaptor sequences and low-quality bases) at start and end of the reads Filtering - Removes contaminants, low complexity reads (repeats), short reads less than 20 bases	K-mer - Shorter nucleotides than the read length De Bruijn graph - Several transcriptome assembly programs. Every path in the graph denotes a potential transcript for transcriptome assembly.
De novo assembly (Zhao et al., 2011)	Trinity (Henschel et al., 2012)	Quality/Phred Score quality (Q-score) - prediction of the probability (P) of an error in base calling. $Q(\text{phred}) = -10 \log_{10}P$ Or $P = 10^{-Q/10}$	De novo Assembler - A novel method for the efficient and robust <i>de novo</i> reconstruction of transcriptomes. Software modules - Inchworm, Chrysalis, and Butterfly
RSEM (Li and Dewey, 2011)	Transcript Abundance Estimation	Assembly Statistics	N50 length is defined as the shortest sequence length at 50% of the transcriptome
Cd-hit (Suzek et al., 2007)	Transcript clustering	Generate unigenes	Group of transcript sequences
edgeR (Robinson et al., 2010)	Differential Expression Analysis	<i>edgeR</i> can be applied to differential expression at the gene, exon, transcript, or tag level. In fact, any genetic feature can be utilized to calculate read counts. There are two testing methods: likelihood ratio tests and quasi-likelihood F-tests.	The key abilities of package, and then gives several fully worked case studies, from counts to list of genes
TransDecoder (Wang et al., 2009)	Coding DNA Prediction	CDS prediction from unigenes	Segments of a gene's (mRNA) that code for protein.
Blast2GO (Martin et al., 2004)	Gene Ontology (GO)	Mapping and annotation	In detail, describe a gene/gene product, including three main characteristics: molecular function (MF), Biological process (BP), cellular compound (CC)
Trinotate (Wang et al., 2009)	Functional Annotation	(COG) Clusters of Orthologous Groups (for prediction of individual proteins function), Ven diagram (to identify common genes of all software), Pfam domain (to identification of protein family), Volcano plot (gene expression), Scattered plot (for normalization of obtained values), Heatmap (for highly significant differential expressed genes)	Portion identification and Gene prediction (process of collecting information about and describing a gene's)
KAAS (Moriya et al., 2007)	Pathway Prediction	Pathway analysis against KEGG databases	Identification of biological functions
DESeq2 (Love et al., 2017)	Differential Expression Analysis	Normalization, differential analysis and visualization of high- dimensional count data	Count matrices can be collapsed using collapse Replicates, which helps to combine counts from technical replications into single columns.

2010), which in turn help us understand how cells function and how they function metabolically. The primary benefit of RNA-Seq is that it allows the comparison of gene expression patterns across samples. Before the advent of deep sequencing technology, the main means of measuring the expression levels of different genes was the microarray (Barrett et al., 2013; Anita and Caterina, 2018). However, hybridization techniques have low sensitivity, making it difficult to identify low-abundance targets and are incapable of finding tiny changes in the expression level of the target gene (Birney et al., 2007). Therefore, RNA-Seq is more accurate than microarray. RNA-Seq, in general, may measure the absolute number of each molecule in a cell

population and directly compare the results between tests. Second, RNA-Seq promotes new gene discovery. The annotations to transcripts in existing databases may not be comprehensive. The RNA-Seq results might self-assemble without the need for known genome annotations, facilitating the discovery of new genes (Zhao et al., 2014).

Third, RNA-Seq has shown satisfactory results in detecting sequence differences, including the identification of fusion candidate genes and the investigation of coding sequence polymorphisms. Because of alternative splicing, a single gene can produce many mRNA transcripts, each of which can be translated into a distinct protein with a different function.

Alternative splicing is universal, in eukaryotic species. The introduction or removal of various introns and exon regions during the splicing process forms different mRNA precursor obtained during gene transcription. Including the sequences spanning the splice junction region, sequences of all transcripts could be found with appropriate sequencing depth in RNA-Seq. The depth is defined as the ratio of the total number of bases (bp) retrieved by sequencing to the genome size, and it is one of the indicators used to evaluate the quantity of sequencing. Lastly, the main aspect of transcriptomics study is the finding and analysis of non-coding RNA (nc-RNA), as shown in Figure 1.

Single cell RNA sequencing technology (scRNA-seq)

In recent years, single-cell RNA sequencing (scRNA-seq) technologies have revolutionized the way we think about biological systems, as they are able to provide a high degree of spatial and temporal resolution of analyses (Zhang et al., 2019). Through scRNA-seq of plants, it is possible to identify new cell types and reveal how the different cell types interact spatially and developmentally revealing both common and rare cell types and cell states. By using single-cell RNA sequencing technology (scRNA-seq), stem cells can be systematically studied at a cellular and molecular level to gain insight into their differentiation trajectory (Islam et al., 2014; Macosko et al., 2015). An early developmental stage transcriptome was described in 2009 based on a next-generation sequencing platform as the first example of single-cell transcriptome analysis (Tang et al., 2009). High-resolution global views of single-cell heterogeneity have become increasingly popular since this study. A critical aspect of this study is analyzing the gene expression differences (Hwang et al., 2018). When analysis is conducted on individual cells, it may be possible to detect rare populations that may not be detectable in pooled analysis. Furthermore, recent advances in immunological research techniques and bioinformatics pipelines have made it possible for researchers to deconvolute highly diverse immune cell populations (Shalek et al., 2013). Additionally, scRNA-seq is increasingly used to analyze myoblast differentiation (Trapnell et al., 2014), lymphocyte fate (Stubbington et al., 2016), and early development (Petropoulos et al., 2016).

Advances in scRNA-seq technologies have opened new possibilities for uncovering the cellular and molecular differentiation trajectory of plant stem cells. The first application of single-cell technologies is to identify cell subtypes in heterogeneous populations of cells. To date, the Arabidopsis primary root tip is the most studied tissue using single-cell RNA sequencing. By analysing briTRIPLE mutants at single-cell resolution, Graeff et al., 2021 studied the impact of brassinosteroids (BR) signalling in Arabidopsis root tissue. In

this study, the researchers found that BR signalling does not affect cell proliferation or cellular development, rather it promotes cellular anisotropy and cell division plane orientation. It is possible to identify phenotypic variations among cell types through single-cell sequencing profiling. ScRNA-seq was used to examine the phenotypes of epidermal cells in the mutant, including root hair deficient cells (rhd6) and glabrous2 cells (gl2). Cell identity phenotypes were identified based on generated data. A study of rice radicals using single-cell sequencing and chromatin accessibility was conducted by Zhang et al., 2021. The development trajectories of epidermal cells and ground tissues were reconstructed using root tip cell profiling, which provided insight into the mechanism that controls cell fate determination in these lineages. Transcriptome profiles and marker genes for these cell types were uncovered through further analysis.

Furthermore, scRNA-seq has been applied to crop improvement in maize by highlighting transcriptional differentiation at high resolution in maize cells. An analysis of 12,525 maize ear cells using scRNA-seq technology was recently published by Xu et al. (2021). A scRNA-seq map of an inflorescence was generated as a result of this profiling. By identifying genetic redundancy in maize, establishing gene regulatory networks at the cell level, and identifying key loci with high ear-yielding characteristics, the generated data can help promote maize genetics. A similar study was conducted by Bezruczyk and colleagues in the same year, using scRNA-seq to study bundle sheath (BS) differentiation in maize. In maize, single-cell sequencing profiling helped identify cells with unique characteristics on the adaxial side, enhancing the possibility of bioengineering the plant (Bezruczyk et al., 2021). As single-cell transcriptomics is increasingly applied to several plant species, future research will indeed introduce it to crop species. This will pave the way for its incorporation into applied plant research, which could benefit our agricultural systems in the future.

Functional gene extraction from plants

To withstand external pressures and adapt to their environments, medicinal plants have evolved a plethora of regulatory mechanisms. A functional gene mining technique identifies key enzymes, pathways, and regulatory mechanisms in plants, thus helping us better understand their molecular biology. Xu WR et al., (2014) analyzed *Vitis amurensis* (Amur grape) transcriptome using the Illumina GA-II sequencing platform and found that in cold regulation a total of 6,850 transcripts involved. There were 3,676 as well as 3,147 copies of transcripts that were upregulated and downregulated, respectively, and 38 key TF families that were implicated in cold regulation. The results of this study provide a foundation for further research into the mechanisms involved in *Vitis* species' cold stress tolerance. Pragati et al., 2018 reported that

a total of 221,792 and 161,733 transcripts in which 141,310 and 113,062 unigenes were obtained from leaf and root tissues of *Aloe vera* (Gwarpatha), respectively, were used on the Illumina platform. It has been determined that 16 genes are involved in the production of lignin, carotenoids saponins and anthraquinone. The results of this study will be useful in future research into genes involved in secondary metabolite biosynthesis and metabolic regulation in *A. vera* and other Aloe species.

The transcriptome of *Paeonia suffruticosa* was sequenced and examined in 2012 by Mutasa-Gottgens and colleagues, and 81,725 copies of unigenes associated with drought resistance were found. It has been predicted that genes associated with hormone signaling pathways are important for drought adaptation and setting framework to study *P. suffruticosa*'s drought stress response mechanism in the future. Singh et al. (2017) sequenced the *Trillium govanianum* transcriptome using the Illumina sequencing platform, collecting 69,174 transcripts, and discovering many genes involved in steroidal saponin production and biosynthetic pathways of various secondary metabolites. Researchers identified tissues (leaf and fruit) as the primary sites for producing steroidal saponins in the biosynthesis of terpenoids, brassinosteroids, carotene, flavonoids, steroids and phenylpropane. Genetic manipulation is valuable for the identification of biologically active metabolites, as well as for the development of molecular markers that are functionally related to the identified metabolites. A transcriptome sequence of *Polygonum minus* was published by Loke et al. (2017), which revealed 188,735 transcripts. They also reported 163,200 (86.5%) *P. minus* transcript similarity matches, the vast majority of which were with *Arabidopsis* transcripts (58.9%). Root and leaf tissues have improved metabolite pathways. The findings will contribute to the development of this species' genetic resources.

Callerya speciosa genome sequencing was conducted by Li et al. (2016) using Illumina's platform through which 161,926 unigenes and 4,538 differentially expressed genes were obtained. Store roots may be implicated in starch synthesis, cell wall loosening and light signaling. Additionally, they may play a role in the development of store roots. Using these findings, subsequent research was conducted on growing *C. speciosa* roots, producing therapeutic substances, and breeding the plant. In a study by Hou et al. (2018), 56,392 unigenes and 4,585 significant DEGs were found in *Cornus officinalis* leaf and fruit tissues using next-generation sequencing (NGS). A total of 1,392 genes were up-regulated in fruit tissues, while 3,193 genes were down-regulated. Most DEGs are involved in the regulation of secondary metabolism and the production of terpenoids. This knowledge contributes to the understanding of plant metabolism and gene expression. The phenolic compound rosmarinic acid has antimicrobial and antioxidant properties and is physiologically active. It was reported that *Dracocephalum tanguticum* revealed 151,463 unigenes in its transcriptome by

Li et al. (2017). A total of 22 genes are predicted to be involved in the biosynthesis of rosmarinic acid, providing references for future research on rosmarinic acid biosynthesis genes.

Development of plant molecular markers

SSR markers are a frequently deployed form of microsatellite DNA marker. A tandem repetition sequence consists of 1-6 nucleotide base pairs, with the most common sequence being dinucleotide repeats. These markers were polymorphic and different numbers of tandem repetitions were associated with them. Historically, SSRs have been used in creating genetic maps, defining genetic diversity, mapping genes, and identifying parental ties due to their high polymorphism, simplicity, codominance, and easy detection. To understand the genetic diversity within asparagus species, Kapoor et al. (2020) used SSR markers to study asparagus varieties grown in different regions of northwest India. There were more than 120 alleles amplified, ranging between 3 and 8, with an average number of five alleles per marker. The lengths of the alleles ranged from 90 to 680 bp. Based on genetic diversity analysis, most Asparagus varieties have a conservative genetic base, except for *A. adscendens*, which indicates that this species has an extensive genetic base.

Future hybridization and conservation of Asparagus species are likely to be affected by these findings. Bhandari et al. (2020) used Illumina paired-end sequencing technology to create new SSR markers for *Salvadora oleoides* (Bada Peelu). From 21,055 microsatellite repeats, they developed 14,552 SSR markers, and randomly selected and confirmed 7,101 SSRs; 94 primers exhibited polymorphisms, and 34 primers failed to amplify. This study provides a foundation for future research on *S. oleoides*.

A HiSeq 4000 sequencing platform was used to analyze transcripts from *Populus alba* (root, leaf, and stem) (Dinh et al., 2020). A total of 11,343 EST-SSRs were identified, of which 101 primer pairs (forwards and reverses) were selected for polymorphism validation. Polymorphisms in populations were discovered by amplifying DNA fragments with 20 primers. Conservation, restoration, and management strategies can benefit greatly from this conclusion. Wang et al. (2020) sequenced and analyzed the transcriptome of *Gastrodia elata* and identified 34,322 unigenes. In 2,007 unigenes (5.85%), at least one SSR was present.

There were 498 detections (21.67%) of a repeating pattern of AG/CT among these SSRs. As a result of this research, Lade et al. (2020) have gained deeper insight into the molecular mechanisms that regulate the growth, development and metabolism of *G. elata* (Tianma). Total 96 sample of *Tinospora cordifolia* were gathered from 10 different geographically diverse locales of the India. In *T. cordifolia*, a total of 268,149 transcripts were assembled. Amongst them,

7,611 SSRs were identified. Tc16, Tc17, Tc31, Tc38, Tc59, Tc60, Tc129, Tc106, Tc130, and Tc131 were shown to contain genetic diversity potential. The potential markers SSR-18, TcSSR-37, TcSSR-59, TcSSR-92, TcSSR-123, and TcSSR-126 have been identified. Genetic enhancement of *T. cordifolia* will be assisted by these components and the newly discovered SSR markers. A comprehensive genetic analysis of two *Menispermum* species was conducted by Hina et al. (2020) using Illumina's transcriptomic platform and *de novo* assembly. Sum of 521 polymorphic EST-SSRs were found out of a total of 53,712 and 78,921 unigenes. The newly designed EST-SSR marker was also shown to be highly transferrable throughout the *Menispermum* species studied. In order to genetically map *Menispermum* populations, these unique microsatellites will be used. A total of 86,195 unigenes were identified in *P. lactiflora* by using microsatellite software, while 21,998 SSR sites were identified dispersed over 17,567 unigenes. Among the 100 primer pairs, 45 were selected at random and amplified bands of polymorphism. For the cluster analysis of sixteen *P. lactiflora* variations, these 45 primer pairs were used. Molecular markers-assisted breeding with *P. lactiflora* will be facilitated by the novel SSR marker.

The metabolic pathways of secondary metabolites

Secondary metabolites are often the most significant components of plants. They play an essential part in the process by which plants adapt to their surroundings and build up a defense mechanism against the impact of various stress. There are many factors that affect the accumulation of secondary metabolites, including the growing environment as well as the developmental stage of the part from which it produced. In various stages of development, transcriptomics is used to investigate pathways of biosynthesis, which lead to secondary metabolites, and to mine genes involved in biogenesis. A scientific foundation has been laid for determining how plants accumulate and utilize active components. *Entada phaseoloides* has been used for medical purposes for centuries. Traditional medicine makes extensive use of the stems due to the wind-dampness-eliminating and anti-inflammatory properties that these stems possess. The triterpenoid saponins found in *E. phaseoloides* are the compounds with the highest level of physiological activity. *E. phaseoloides* root, stem, and leaf tissues to uncover 26 candidate genes for cytochrome P450 and 17 uridine diphosphate glycosyltransferases that are involved in the production of triterpene saponin (Liao et al., 2020). As can be seen in Supplementary Table 1, the findings were beneficial to both the production of triterpenoid saponin and research into functional genomics.

There are several physiologically active compounds in *Lantana camara* (Lantana), including steroidal saponins,

flavonoids, and glycosides. Using tools for sequencing the transcriptome, Shah et al. (2020) put together *L. camara* leaves and roots from scratch. It was found that 72,877 and 513,985 unigenes were present in leaves and roots, respectively. Of these, 229 and 943 genes were responsible for the production of phenyl-propanoic acid. As a broad-spectrum antibiotic, *Tetrastigma hemsleyanum* extract is used to treat fever and sore throats. An in-depth analysis of the metabolome and transcriptome of purple and green *T. hemsleyanum* leaves was performed by Yan et al. (2020). 4211 transcripts have been identified in the purple and green leaves, 209 metabolites have been found to be differentially expressed, and 16 chemicals have been associated with 14 transcripts implicated in the pathway for anthocyanin synthesis. The sesquiterpene lactones produced by *Saussurea lappa* have a high medicinal value. In a study conducted by Bains et al. (2019), the leaf transcriptome of *S. lappa* was sequenced to identify flavonoid and sesquiterpene-producing transcripts. Transcripts from genes implicated in alkaloid metabolism have been identified in a small number of cases. As a result of these insights, scientists will be able to learn more about how plants' functional genomes work. Transcriptome analysis of *Arisaema heterophyllum* Blume and its leaf, tuber, and root tissues identified 47783, 43363, and 35686 unigenes, respectively, implicating genes involved in isoflavone biosynthesis. Experimental confirmation of 87 candidate genes encoding isoflavone-producing enzymes was accomplished (Wang et al., 2018). The findings of this study pave the way for further research on the pharmacological action of *Arisaema*. The antioxidative and anti-inflammatory properties of flavonoids, along with their application in the treatment of diseases, are illustrated in Supplementary Table 1.

A significant percentage of flavonoids may be found in the leaves of *Ginkgo biloba*. Wu et al. (2018) identified 37,625 unigenes from transcriptome sequencing of *G. biloba* with various flavonoid concentrations. According to the research, several potential genes are involved in the manufacture, transport, and regulation of flavonoids, according to the research. It was found that MYB transcription factor and dihydroflavonol-4-reductase, two of the fourteen flavonoid transport genes, participate in flavonoid transport. It is anticipated that the discoveries will help expand the current *G. biloba* gene database, broaden Ginkgo species research, and provide crucial information for the development of Ginkgo-related pharmaceuticals. *De novo* transcriptome sequencing of *Abrus mollis* leaves enabled analysis of flavonoid synthesis routes and associated precursors (Yuan et al. (2018)). Liu MM et al. (2018) found 99,807 unigenes in *Artemisia argyi* leaf, root, and stem tissues, including many genes that encode terpene-synthesis enzymes and transcription factors. It is anticipated that the findings will be used to investigate the molecular pathways of *A. argyi*. An analysis of *Panax ginseng* root tissues using 454-sequencing technology was conducted by Jayakodi et al. (2014). There were 17 percent difference in transcript levels

between adventitious and common roots, as well as a 21 percent difference in ginsenoside-producing genes of *P. notoginseng* (Luo et al., 2011; Liu MH et al., 2015). The transcripts of *P. notoginseng* differed by 17% between adventitious and common roots, as did 21 ginsenoside-producing genes (Luo et al., 2011; Liu MH et al., 2015). Jayakodi et al., 2015 studied the transcriptomic profile of *P. ginseng* leaves, roots, and flowers and identified 107,340 unigenes, including 9,908 metabolic pathways and 270 triterpene saponin-producing genes. Among 32 genes expressed specifically in annual ginseng roots, seven genes were expressed specifically in 6-year-old ginseng roots, and 38 genes were implicated in triterpene saponin synthesis, as shown in [Supplementary Table 1](#).

The transcriptomic process

Transcriptomic approaches have been widely used to identify genes involved in plant growth and development, to identify genes that are expressed differently under abiotic stress, and to study their resistance to it. Identifying key influencing components can be helpful in plant cultivation and breeding, as well as simplifying the selection of improved varieties (Rastogi et al., 2019). Using transcriptome sequencing, Liu MM et al. (2018) studied the response of *A. argyi* leaves to cold, drought, waterlogging, and salt stress. Cold stress was the most damaging condition to the plants. Treatments with abiotic stress also reduced eugenol synthesis. The discovery of several stress-tolerance genes in *A. argyi* has enabled transgenic or polymerized plants to become stress-tolerant. According to Li et al. (2020), transcriptome sequencing was used to investigate the molecular mechanisms behind *Salvia miltiorrhiza* tissues' responses to mild abiotic stress (drought). In total, 58,085 unigenes were discovered, of which 28,846 could be identified as such. Significant enrichment in metabolic processes and catalytic activities were found among differential transcripts in roots and leaves based on GO enrichment studies. The expression of genes that encode enzymes involved in the synthesis of phenylpropanoids and terpenoids increased in response to moderate drought stress. These findings have provided a solid foundation for further research into the process of manufacturing therapeutic components in *S. miltiorrhiza* as well as irrigation techniques for successful cultivation, as shown in [Supplementary Table 1](#).

Using the third sequencing technique, Feng et al. (2019) sequenced the full-length transcriptome of *Angelica sinensis*, and differential expression sequencing was done on the wild-type transcriptome. An analysis of NGS data identified 25,463 transcripts with differential expression. There was a significant difference between transcripts in the pathway for plant-pathogen interaction and signal transduction of plant hormones. The purpose of this study is to lay out a platform for screening and developing *A. sinensis*. The transcriptome sequencing of callus

tissue of *S. laniceps* was used to identify genes associated with frost resistance (Xu, 2018). In the GO enrichment investigation, 155 substances related to low temperatures, oxidative stress, and plant hormone responses were identified. Based on the KEGG enrichment analysis, several pathways were significantly enriched during low-temperature responses, including ribosomes, fatty acid metabolism, and unsaturated fatty acid biosynthesis ([Supplementary Table 1](#)). The findings of this study provide a framework for future research on genes associated with frost resistance in *Saussurea laniceps*.

Discussion

In the biological sciences, transcriptome sequencing is an important sequencing technique that can be used without genomic reference sequences. Transcriptomic analysis can be used to analyze a wide range of plants and has many applications. The application of transcriptomics for obtaining genetic information about plants is growing rapidly due to its fast, high coverage, efficiency, and high throughput characteristics. This technology has been applied to mining new functional genes, analyzing secondary metabolite pathways, identifying plant developmental pathways, and obtaining helpful information for plant breeding (Li et al., 2018). Understanding secondary metabolite synthesis pathways and associated genes will benefit secondary metabolism regulatory network analysis and secondary metabolism studies in plants. Numerous plants have been analyzed by transcriptomic analysis, and the technique has been applied in many different fields. The use of transcriptomics to gather genetic information on plants is expanding as a fast, high-coverage, high-throughput, and high-efficiency analytical approach. Using the technology, researchers have been able to identify plant developmental pathways, mine novel functional genes, and analyze secondary metabolite synthesis pathways (Li et al., 2018).

Some plants produce multiple secondary metabolites to survive with biological and abiotic stresses, some of which can be used to treat a variety of human ailments. These plants are commonly referred to as plants because of their medicinal potential. Approximately 270,000 plants species have been identified globally, with less than 40,000 species having putative medical significance (Mamedov and Nazim, 2012; Tan et al., 2015). Except for a few model plants that are important research tools and sources of genomic data, most plants are largely underexplored in terms of biological information. There is a significant gap in plant genomic data due to a lack of research knowledge about the characterization of plant transcriptomes. This hinders research on important topics like the identification of significant differentially expressed genes (DEGs) and pathways related to secondary metabolite biosynthesis. In order to gather data for future plant research,

plant transcriptome research should be encouraged (Navin and Hicks, 2011). In the future, transcriptome data analysis and study will assist in identifying functional genes associated with secondary metabolism pathways.

NGS techniques help us understand the transcriptome's complexity. In several transcriptome researches on plants, NGS reads are limited by the requirement for assembly or reference genomes. The short read length of NGS technology makes it difficult to study full-length transcripts in plants. It is usually possible to investigate only the local structure of the gene and alternative splicing mechanisms for full-length transcripts. Third-generation (3G) sequencing technology is becoming more common with the advancement of sequencing techniques. Due to its relatively high sequencing cost and low throughput, third-generation sequencing is limited in its utility for transcription at this time. RNA-Seq research is mostly conducted using NGS technology, with the third generation serving as a backup. Genomic, transcriptomic, metabolomic, and proteomic technologies are also available for collaborative research. Biological functions are currently hot research topics for high-throughput sequencing, and the data will be useful for mining genes and algorithms, for fast analysis, and for showing that these topics are very useful for mining genes and algorithms.

A transcriptome profiling approach is required to identify isolated cells in a population of plants or animals, and the observations are often reflective of the number of cells predominating (Ranzoni et al., 2019). Due to cell heterogeneity, phenotypic characteristics may appear to be the same, but genetic information will vary dramatically. Consequently, transcriptome sequencing typically results in the loss of a great deal of low-abundance information (Navin and Hicks, 2011). It is predicted that single-cell transcriptome research will move into a new phase with the development of sequencing technology and the sharp decline in sequencing costs. It is possible to systematically track the dynamic changes of individual cells using single-cell transcriptome sequencing, which can effectively support the heterogeneity of single-cell gene expression that conventional sequencing ignores. We can thus gain a better understanding of cell state, genetic makeup, gene expression, and regulation, as well as develop herbal medicines.

Conclusion

As a precursor to NGS developments, first-generation sequencing technologies and pioneering computing and bioinformatics tools generated the initial sequencing data and information within a structural and functional genomics framework. With NGS, high-throughput sequencing options are substantially cheaper, friendlier, and more flexible, allowing us to generate much more data on genomics and transcriptomics that can be used to further explore proteomics and metabolomics. A variety of NGS platforms have been

released. During the first three decades of sequencing, Sanger sequencing dominated, but cost and time were major obstacles. The emergence of the second generation sequencing in 2005 and subsequent years has set the stage for breaking through the limitations of the first-generation sequencing. Sequencing by synthesis and sequencing by ligation are the two approaches proposed so far for second-generation sequencing.

There are currently more advantages to “third-generation sequencing” platforms (compared to first- and second-generation NGS platforms), including longer run times, complete transcript sequencing, and faster turnaround times. Their high mismatch rate, however, limits their use in transcriptome sequencing. On the other hand, third-generation (3G) sequencing technology can be used in conjunction with NGS technology to repair errors and offer genotyping recognition. As the cost of third-generation sequencing decreases and the accuracy of the technology improves, third-generation sequencing will become more frequently used in transcriptome research for accurate and complete transcriptome sequencing. The economic crop research model should lead to the widespread use of transcriptome sequencing in traditional plants. The majority of plant research does not rely solely on RNA-Seq at the moment. As RNA-Seq technology develops, multi-omics-related techniques will play a significant role, along with metabolomics and proteomics. Plant research would be modernized with the advancement of transcriptomic approaches, including developing metabolomics and proteomics techniques.

NGS has become a science that integrates biological information systems with big data, but many challenges remain for NGS data acquisition, analysis, storage, interpretation and integration. In order to continue producing comprehensive, high-throughput data for analysis and production, new technologies and large-scale collaboration efforts will be needed in the future. With the advent of affordable benchtop sequencers and third-generation sequencing tools, smaller laboratories and individual scientists can participate in the genomic revolution and contribute new knowledge to structural genomics and functional genomics in the life sciences.

Author contributions

PT, DS, SM, and AS wrote the manuscript and RR edited the same for further improvement.

Acknowledgments

We are grateful thanks to the Director of Dayalbagh Educational Institute, Dayalbagh, Agra for the support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1030890/full#supplementary-material>

References

- Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., and Schilke, F. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7, 11706. doi: 10.1038/ncomms11706
- Ai, Y., Zhang, Q., Wang, W., Zhang, C., Cao, Z., and Bao, M. (2016). Transcriptomic analysis of differentially expressed genes during flower organ development in genetic Male sterile and Male fertile *Tagetes erecta* by digital gene-expression profiling. *PloS One* 11 (3), e0150892. doi: 10.1371/journal.pone.0150892
- Andrews, S. (2010) *FastQC: A quality control tool for high throughput sequence data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Anita, A., and Caterina, M. (2018). Whole transcriptome profiling of late-onset alzheimer's disease patients provides insights into the molecular changes involved in the disease. *Sci. Rep.* 8, 4282. doi: 10.1038/s41598-018-22701-2
- Bains, S., Thakur, V., Kaur, J., Singh, K., and Kaur, R. (2019). Elucidating genes involved in sesquiterpenoid and flavonoid biosynthetic pathways in *Saussurea lappa* by *de novo* leaf transcriptome analysis. *Genomics* 111, 1474–1482. doi: 10.1016/j.ygeno.2018.09.022
- Barrett, T., Wilhite, S. E., and Ledoux, P. (2013). NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193
- Bezruczyk, M., Zöllner, N. R., Kruse, C. P., Hartwig, T., Lautwein, T., Köhrer, K., et al. (2021). Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. *Plant Cell* 33 (3), 531–547. doi: 10.1093/plcell/koaa055
- Bhandari, M. S., Meena, R. K., Shamooin, A., Saroj, S., and Pandey, S. (2020). First *de novo* genome specific development, characterization, and validation of simple sequence repeat (SSR) markers in genus *Salvadora*. *Mol. Biol. Rep.* 47, 6997–7008. doi: 10.1007/s11033-020-05758-z
- Birney, E., Stamatoyannopoulos, J. A., and Dutta, A. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874
- Bolón-Canedo, V., Alonso-Betanzos, A., López-de-Ullibarri, I., and Cao, R. (2019). "Challenges and future trends for microarray analysis," in *Microarray bioinformatics* (New York, NY: Humana), 283–293.
- Chen, Y., Liu, Y. S., and Zeng, J. G. (2014). Progresses on plant genome sequencing profile. *Life Sci. Res.* 18, 66–74. doi: 10.16605/j.cnki.1007-7847.2014.01.006
- Clamp, M., Fry, B., Kamal, M., Xie, X. H., Cuff, J., Lin, M. F., et al. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19428–19433. doi: 10.1073/pnas.0709013104
- Clark, T. A., Sunget, C. W., and Ares, M. J. (2002). Genome wide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296, 907–910. doi: 10.1126/science.1069415
- Costa, V., Angelini, C., Feis, I. D., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-seq. *J. BioMed. Biotechnol.* 2010, 853916. doi: 10.1155/2010/853916
- Crowgey, E. L., and Mahajan, N. (2019). "Advancements in next-generation sequencing for detecting minimal residual disease," in *Minimal residual disease testing* (Switzerland: Springer, Cham), 159–192.
- Dinh, D. V., Syed, N. S., Mai, P. P., Van, T. B., Minh, T. N., and Thi, P. N. (2020). *De novo* assembly and transcriptome characterization of an endemic species of Vietnam, *Panax vietnamensis* ha et grushv., including the development of EST-SSR markers for population genetics. *BMC Plant Biol.* 20, 159–138. doi: 10.1186/s12870-020-02571-5
- Drygin, Y. F., Butenko, K. O., and Gasanova, T. V. (2021). Environmentally friendly method of RNA isolation. *Analytical Biochemistry* 620, 114113. doi: 10.1016/j.ab.2021.114113
- Feng, C. H., Hei, C. Y., Wang, Y., Zeng, Y. F., and Zhang, J. G. (2019). Phylogenetic position of *Chosenia arbutifolia* in the salicaceae inferred from whole chloroplast genome. *For. Res.* 32, 73–77. doi: 10.13275/j.cnki.lykxyj.2019.02.011
- Filichkin, S. A., Priest, H. D., and Givan, S. A. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., and Amit, I. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Graeff, M., Rana, S., Wendrich, J. R., Dorier, J., Eekhout, T., Fandino, A. C. A., et al. (2021). A single-cell morpho-transcriptomic map of brassinosteroid action in the arabidopsis root. *Mol. Plant* 14 (12), 1985–1999. doi: 10.1016/j.molp.2021.07.021
- Guo, J., Huang, Z., Sun, J., Cui, X., and Liu, Y. (2021). Research progress and future development trends in medicinal plant transcriptomics. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.691838
- Henschel, R., Lieber, M., Wu, L., Nista, P. M., Haas, B. J., and LeDuc, R. D. (2012). "Trinity RNA-seq assembler performance optimization," in *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond*. (New York: The association for computing machinery).
- Hina, F., Yisilam, G., Wang, S., Li, P., and Fu, C. X. (2020). *De novo* transcriptome assembly, gene annotation and SSR marker development in the moon seed genus *Menispermum* (Menispermaceae). *Front. Genet.* 11. doi: 10.3389/fgenet.2020.00380
- Hoopes, G. M., Hamilton, J. P., Kim, J., Zhao, D., Wiegert-Rininger, K., Crisovan, E., et al. (2018). Genome assembly and annotation of the medicinal plant *calotropis gigantea*, a producer of anticancer and antimalarial cardenolides. *G3: Genes Genomes Genet.* 8 (2), 385–391. doi: 10.1534/g3.117.300331
- Hou, D. Y., Shi, L. C., Yang, M. M., Li, J., Zhou, S., and Zhang, H. X. (2018). *De novo* transcriptomic analysis of leaf and fruit tissue of *Cornus officinalis* using illumina platform. *PloS One* 13 13, e0192610. doi: 10.1371/journal.pone.0192610
- Huang, N., Nie, F., Ni, P., Luo, F., Gao, X., and Wang, J. (2021). NeuralPolish: a novel nanopore polishing method based on alignment matrix construction and orthogonal bi-GRU networks. *Bioinformatics* 37 (19), 3120–3127. doi: 10.1093/bioinformatics/btab354
- Hwang, B., Lee, J. H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50 (8), 1–14. doi: 10.1038/s12276-018-0071-8
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., et al. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11 (2), 163–166. doi: 10.1038/nmeth.2772
- Jain, M. (2012). Next-generation sequencing technologies for gene expression profiling in plants. *Brief. Funct. Genomics* 11, 63–70. doi: 10.1093/bfpg/blr038
- Jayakodi, M., Lee, S. C., Lee, Y. S., Park, H. S., Kim, N. H., and Jang, W. J. (2015). Comprehensive analysis of *Panax ginseng* root transcriptomes. *BMC Plant Biol.* 15, 138. doi: 10.1186/s12870-015-0527-0

- Jayakodi, M., Lee, S. C., Park, H. S., Jang, W. J., Lee, Y. S., and Choi, B. S. (2014). Transcriptome profiling and comparative analysis of *Panax ginseng* adventitious roots. *J. Ginseng. Res.* 38, 278–288. doi: 10.1016/j.jgr.2014.05.008
- Kapoor, M., Mawal, P., Sharma, V., and Gupta, R. C. (2020). Analysis of genetic diversity and population structure in *Asparagus* species using SSR markers. *J. Genet. Eng. Biotechnol.* 18, 50. doi: 10.1186/s43141-020-00065-3
- Karsch-Mizrachi, L., Takagi, T., and Cochrane, G. (2018). International nucleotide sequence database collaboration. the international nucleotide sequence database collaboration. *Nucleic Acids Res.* 46, D48–D51. doi: 10.1093/nar/gkx1097
- Kumar, S., Razaq, S. K., and Vo, A. D. (2016). Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip Rev RNA* 7, 811–823. doi: 10.1002/wrna.1382
- Lade, S., Pande, V., Rana, T. S., and Yadav, H. K. (2020). Estimation of genetic diversity and population structure in *Tinospora cordifolia* using SSR markers. *3 Biotech.* 10, 413–425. doi: 10.1007/s13205-020-02300-7
- Lathe, W., Williams, J., Mangan, M., and Karolchik, D. (2008). Genomic data resources: challenges and promises. *Nat. Educ.* 1, 2.
- Lee, D. G., and Lee, S. H. (2021). Investigation of COGs (Clusters of orthologous groups of proteins) in 1,309 species of prokaryotes. *J. Life Sci.* 31 (9), 834–839. doi: 10.5352/JLS.2021.31.9.834
- Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. L., Smits, B. M., et al. (2012). An empirical bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29, 1035–1043. doi: 10.1093/bioinformatics/btt087
- Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115. doi: 10.1146/annurev-genom-083115-022413
- Liang, S. B., Liu, J. Y., Yang, J. T., Liu, J., Li, J. L., and Zhang, Y. M. (2017). Next-generation sequencing applications for crop genomes. *China Biotechnol.* 37, 111–120. doi: 10.13523/j.cb.20170216
- Liao, W. F., Mei, Z. N., Miao, L. H., Liu, P. L., and Gao, R. J. (2020). Comparative transcriptome analysis of root, stem, and leaf tissues of *Entada phaseoloides* reveals potential genes involved in triterpenoid saponin biosynthesis. *BMC Genomics* 21, 639. doi: 10.1186/s12864-020-07056-1
- Li, X. L., Bo, B., Wu, J., Deng, Q. Y., and Zhou, B. (2012). Transcriptome analysis of early interaction between rice and *Magnaporthe oryzae* using next-generation sequencing technology. *Hereditas* 34, 104–114. doi: 10.3724/SP.J.1005.2012.00102
- 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. doi: 10.1186/1471-2105-12-323
- Li, H., Fu, Y., Sun, H., Zhang, Y., and Lan, X. (2017). Transcriptomic analyses reveal biosynthetic genes related to rosmarinic acid in *Dracocephalum tanguticum*. *Sci. Rep.* 7, 74. doi: 10.1038/s41598-017-00078-y
- Li, Y. M., Li, S. X., Li, X. S., and Li, C. Y. (2018). Transcriptome studies with the third-generation sequencing technology. *Life Sci. Instrum.* 16, 114–121.
- Liu, X., Bi, B., Xu, X., Li, B., Tian, S., Wang, J., et al. (2019). Rapid identification of a candidate nicosulfuron sensitivity gene (Nss) in maize (*Zea mays* L.) via combining bulked segregant analysis and RNA-seq. *Theor. Appl. Genet.* 132 (5), 1351–1361. doi: 10.1007/s00122-019-03282-8
- Liu, L., Li, Y., Li, S., Hu, N., He, Y. M., and Pong, R. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2008, 251364. doi: 10.1155/2012/251364
- Liu, M. H., Yang, B. R., Cheung, W. F., Yang, K. Y., Zhou, H. F., and Kwok, J. L. (2015). Transcriptome analysis of leaves, roots, and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. *BMC Genomics* 16, 265. doi: 10.1186/s12864-015-1477-5
- Liu, F. X., Yang, W. G., and Sun, Q. H. (2018). Transcriptome sequencing data analysis and high throughput GO annotation. *J. Anhui Agric. Univ.* 46, 88–91. doi: 10.13989/j.cnki.0517-6611.2018.31.027+100
- Liu, M. M., Zhu, J. H., Wu, S. B., Wang, C. K., Guo, X. Y., and Wu, J. W. (2018). *De novo* assembly and analysis of the *Artemisia argyi* transcriptome and identification of genes involved in terpenoid biosynthesis. *Sci. Rep.* 8, 1236–1243. doi: 10.1038/s41598-018-24201-9
- Li, X., Wang, J., Ming, L., Li, L., and Li, Z. (2016). Transcriptome analysis of storage roots and fibrous roots of the traditional medicinal herb *Callerya speciosa* (Champ.) ScHot. *PLoS One* 11, e0160338. doi: 10.1371/journal.pone.0160338
- Li, Y., and Xu, X. Y. (2019). Research progress of high-throughput sequencing technology. *China Med. Eng.* 27, 32–37. doi: 10.19338/j.issn.1672-2019.2019.03.009
- Li, X. Y., Zhou, J. W., Yan, Z. Y., and Chen, X. (2020). Sequencing and analysis of transcriptome to reveal regulation of gene expression in *Salvia miltiorrhiza* under moderate drought stress. *Zhong Cao Yao* 51, 1600–1608. doi: 10.7501/j.issn.0253-2670.2020.06.029
- Lockhart, D. J., and Winzler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature* 405 (6788), 827–836. doi: 10.1038/35015701
- Loke, K. K., Rahnamaie-Tajadod, R., Yeoh, C. C., Goh, H. H., and Noor, N. M. (2017). Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. *PeerJ* 5, e2938. doi: 10.7717/peerj.2938
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. *PLoS Comput. Biol.* 13, e1005457. doi: 10.1371/journal.pcbi.1005457
- Lu, X. (2013). “A comparison of transcriptome assembly software for next-generation sequencing technologies,” in *Ph.D. Thesis* (Gansu: University of Lanzhou).
- Luo, H., Sun, C., Sun, Y. Z., Wu, Q., Li, Y., and Song, J. Y. (2011). Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics* 12, S5. doi: 10.1186/1471-2164-12-S5-S5
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161 (5), 1202–1214. doi: 10.1016/j.cell.2015.05.002
- Ma, L., Dong, C., Song, C., Wang, X., Zheng, X., Niu, Y., et al. (2021). *De novo* genome assembly of the potent medicinal plant *Rehmannia glutinosa* using nanopore technology. *Computational and Structural Biotechnology Journal* 19, 3954–3963. doi: 10.1016/j.csbj.2021.07.006
- Madritsch, S., Burg, A., and Sehr, E. M. (2021). Comparing *de novo* transcriptome assembly tools in di- and autotetraploid non-model plant species. *BMC Bioinf.* 22 (1), 1–17. doi: 10.1186/s12859-021-04078-8
- Mamedov, N., and Nazim, N. (2012). Medicinal plants studies: history, challenges and prospective. *Aromat. Plants* 1, 1–2. doi: 10.4172/2167-0412.1000e133
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., and Bembel, L. A. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959
- Martin, J., Bruno, V. M., Fang, Z., Meng, X., Blow, M., Zhang, T., et al. (2010). Rnnotator: an automated *de novo* transcriptome assembly pipeline from stranded RNA-seq reads. *BMC Genomics* 11 (1), 1–8. doi: 10.1186/1471-2164-11-663
- Martin, D., Berriman, M., and Barton, G. J. (2004). GOTcha: a new method for prediction of protein function assessed by the annotation of seven genomes. *BMC Bioinf.* 5, 178. doi: 10.1186/1471-2105-5-178
- Ma, L. N., Yang, J. B., Ding, Y. F., and Li, Y. K. (2019). Research progress on three generations sequencing technology and its application. *China Anim. Husb. Vet. Med.* 46, 2246–2256. doi: 10.16431/j.cnki.1671-7236.2019.08.007
- Mei, C. G., Wang, H. C., Zan, L. S., Cheng, G., Li, A. L., and Zhao, C. P. (2016). Research progress on animal genome research based on high-throughput sequencing technology. *J. Northwest A&F Univ.* 44, 43–51. doi: 10.13207/j.cnki.jnwafu.2016.03.007
- Mironova, V. V., Weinholdt, C., and Grosse, I. (2015). “RNA-seq data analysis for studying abiotic stress in horticultural plants,”. *Abiotic Stress Biol* 1, 197–220. doi: 10.1007/978-4-431-55251-2_14
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49 (D1), D412–D419. doi: 10.1093/nar/gkaa913
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M. (2007). KAA5: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Mortazavi, A., Williams, B. A., and McCue, K. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., and Gerstein, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349. doi: 10.1126/science.1158441
- Navin, N., and Hicks, J. (2011). Future medical applications of single-cell sequencing in cancer. *Genome Med.* 3, 1–12. doi: 10.1186/gm247
- Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., et al. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* 165 (4), 1012–1026. doi: 10.1016/j.cell.2016.03.023
- Pickrell, J. K., Marioni, J. C., and Pai, A. A. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464, 768–772. doi: 10.1038/nature08872
- Pradhan, S. K., Pandit, E., Nayak, D. K., Behera, L., and Mohapatra, T. (2019). Genes, pathways and transcription factors involved in seedling stage chilling stress tolerance in indica rice through RNA-seq analysis. *BMC Plant Biol.* 19 (1), 1–17. doi: 10.1186/s12870-019-1922-8

- Pragati, C., Muniya, R., Sangwan, R. S., Ravinder, K., Anil, K., and Vinod, C. (2018). *De novo* sequencing, assembly, and characterization of *Aloe vera* transcriptome and analysis of expression profiles of genes related to saponin and anthraquinone metabolism. *BMC Genomics* 19, 427. doi: 10.1186/s12864-018-4819-2
- Ranzoni, A. M., Strzelecka, P. M., and Cvejic, A. (2019). Application of single-cell RNA sequencing methodologies in understanding haematopoiesis and immunology. *Essays Biochem.* 63, 217–225. doi: 10.1042/EBC20180072
- Rastogi, S., Shah, S., Kumar, R., Vashisth, D., Akhtar, M. Q., and Dwivedi, U. N. (2019). *Ocimum* metabolomics in response to abiotic stresses: cold, flood, drought, and salinity. *PLoS One* 14, e0210903. doi: 10.1371/journal.pone.0210903
- 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. doi: 10.1093/bioinformatics/btp616
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* 74 (12), 5463–5467.
- Schoch, C. L., Ciufu, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., et al. (2020). NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database* 2020, 1–21. doi: 10.1093/database/baaa062
- Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi: 10.1093/bioinformatics/bts094
- Shah, M., Alharby, H. F., Hakeem, K. R., Ali, N., Rahman, I. U., and Munawar, M. (2020). *De novo* transcriptomics analysis of *Lantana camara* L. revealed candidate genes involved in phenylpropanoid biosynthesis pathway. *Sci. Rep.* 10, 467–486. doi: 10.1038/s41598-020-70635-5
- Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498 (7453), 236–240. doi: 10.1038/nature12172
- Shendure, J. (2008). The beginning of the end for microarrays. *Nat. Methods* 5, 585–587. doi: 10.1038/nmeth0708-585
- Singh, P., Singh, G., Bhandawat, A., Singh, G., Parmar, R., and Seth, R. (2017). Spatial transcriptome analysis provides insights of key gene(s) involved in steroidal saponin biosynthesis in medicinally important herb *Trillium govanianum*. *Sci. Rep.* 7, 45295. doi: 10.1038/srep45295
- Singh, S. K., Wu, Y., Ghosh, J. S., Pattanaik, S., Fisher, C., Wang, Y., et al. (2016). RNA-Sequencing reveals global transcriptomic changes in *Nicotiana tabacum* responding to topping and treatment of axillary-shoot control chemicals. *Sci. Rep.* 5, 18148. doi: 10.1038/srep18148
- Strickler, S. R., Bombarely, A., and Mueller, L. A. (2012). Designing a transcriptome next-generation sequencing project for a non-model plant species. *Am. J. Bot.* 99, 257–266. doi: 10.3732/ajb.1100292
- Stubbington, M. J., Lönnberg, T., Proserpio, V., Clare, S., Speak, A. O., Dougan, G., et al. (2016). T Cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* 13 (4), 329–332. doi: 10.1038/nmeth.3800
- Sun, H. J., and Wei, H. J. (2018). The application of RNA-seq technology in the study of the transcriptome. *chin. Foreign Med. Res.* 16, 184–187. doi: 10.14033/j.cnki.cfmr.2018.20.089
- Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23, 1282–1288. doi: 10.1093/bioinformatics/btm098
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6 (5), 377–382. doi: 10.1038/nmeth.1315
- Tan, X. M., Zhou, Y. Q., Chen, J., and Guo, S. X. (2015). Advances in research on diversity of endophytic fungi from medicinal plants. *Chin. Pharm. J.* 50, 1563–1580. doi: 10.11669/cpj.2015.18.001
- Touch, B. B., Laborde, R. R., and Xu, X. (2010). Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. *PLoS One* 5 5, e9317. doi: 10.1371/journal.pone.0009317
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat. Biotechnol.* 32 (4), 381. doi: 10.1038/nbt.2859
- 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. doi: 10.1038/nbt.2450
- Tyagi, P., and Ranjan, R. (2021). Comparative study of the pharmacological, phytochemical, and biotechnological aspects of *Tribulus terrestris* Linn. and *Pedalium murex* Linn: An overview. *Acta Ecologica Sinica*. doi: 10.1016/j.chnaes.2021.07.008
- Tyagi, P., Singh, A., Gupta, A., Prasad, M., and Ranjan, R. (2022). Mechanism and function of salicylate in plant toward biotic stress tolerance. *Emerging Plant Growth Regulat. Agriculture*. 131–164 doi: 10.1016/B978-0-323-91005-7.00018-7
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Wang, X., Hu, H., Wu, Z., Fan, H., Wang, G., Chai, T., et al. (2021). Tissue-specific transcriptome analyses reveal candidate genes for stilbene, flavonoid and anthraquinone biosynthesis in the medicinal plant *Polygonum cuspidatum*. *BMC Genomics* 22 (1), 1–17. doi: 10.1186/s12864-021-07658-3
- Wang, Y. S., Shahid, M. Q., Ghouri, F., and Baloch, F. S. (2020). *De novo* assembly and annotation of the juvenile tuber transcriptome of a *Gastrodia elata* hybrid by RNA sequencing: detection of SSR markers. *Biochem. Genet.* 58, 914–934. doi: 10.1007/S10528-020-09983-W
- 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. doi: 10.1038/ncomms11708
- Wang, C., Zhu, J., Liu, M., Yang, Q. S., Wu, J. W., and Li, Z. G. (2018). *De novo* sequencing and transcriptome assembly of *Arisaema heterophyllum* blume and identification of genes involved in isoflavonoid biosynthesis. *Sci. Rep.* 8, 17643. doi: 10.1038/s41598-018-35664-1
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., and Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 11551162. doi: 10.1038/s41587-019-0217-9
- Wu, Y. Q., Guo, J., Zhou, Q., Xin, Y., Wang, G. B., and Xu, L. A. (2018). *De novo* transcriptome analysis revealed genes involved in flavonoid biosynthesis, transport, and regulation in *Ginkgo biloba*. *Ind. Crop Prod.* 124, 226–235. doi: 10.1016/j.indcrop.2018.07.060
- Xie, Y., Wu, G., Tang, J., Luo, R., Jordan, P., and Liu, S. (2014). SOAPdenovo-trans: *de novo* transcriptome assembly with short RNA-seq reads. *Bioinformatics* 12, 1660–1666. doi: 10.1093/bioinformatics/btu077
- Xu, Y. (2018). “Transcriptome analysis of freezing tolerance mechanism for Tibetan saussurea laniceps callus,” in *Ph.D. Thesis* (Beijing: Beijing Forestry University).
- Xu, X., Crow, M., Rice, B. R., Li, F., Harris, B., Liu, L., et al. (2021). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell* 56 (4), 557–568. doi: 10.1016/j.devcel.2020.12.015
- Xu, W. R., Li, R. M., Zhang, N. B., Ma, F., Jiao, Y. T., and Wang, Z. P. (2014). Transcriptome profiling of *Vitis amurensis*, an extremely cold-tolerant Chinese wild vitis species, reveals candidate genes and events that potentially connected to cold stress. *Plant Mol. Biol.* 86, 527–541. doi: 10.1007/s11103-014-0245-2
- Xu, B., Zhang, W. Q., Feng, X. X., Wang, C. Y., Zhang, H. S., and Xu, H. T. (2014). Application progress of transcriptome sequencing technology in maize. *J. Maize Sci.* 22, 67–72. doi: 10.13597/j.cnki.maize.science.2014.01.014
- Yang, J., Guo, Z., Luo, L., Gao, Q., Xiao, W., Wang, J., et al. (2021). Identification of QTL and candidate genes involved in early seedling growth in rice via high-density genetic mapping and RNA-seq. *Crop J.* 9 (2), 360–371. doi: 10.1016/j.cj.2020.08.010
- Yan, J. L., Qian, L. H., Zhu, W. D., Qiu, J. R., Lu, Q. J., and Wang, X. B. (2020). Integrated analysis of the transcriptome and metabolome of purple and green leaves of *Tetragonia hemsleyana* reveals gene expression patterns involved in anthocyanin biosynthesis. *PLoS One* 15, e0230154. doi: 10.1371/journal.pone.0230154
- Yousef, M., Ülgen, E., and Sezerman, O. U. (2021). CogNet: classification of gene expression data based on ranked active-subnetwork-oriented KEGG pathway enrichment analysis. *PeerJ Comput. Sci.* 7, e336. doi: 10.7717/peerj.cs.336
- Yu, G. (2020). “Gene ontology semantic similarity analysis using GOSemSim,” in *Stem cell transcriptional networks* (New York, NY: Humana), 207–215.
- Yuan, X., Li, K., Huo, W., and Lu, X. (2018). *De novo* transcriptome sequencing and analysis to identify genes involved in the biosynthesis of flavonoids in *Abrus mollis* leaves. *Russ. J. Plant Physiol.* 65, 333–344. doi: 10.1134/S1021443718030147
- Zhang, H. (2019). “The review of transcriptome sequencing: principles, history and advances,” in *IOP conference series: Earth and environmental science*, vol. 332, No. 4. (Bristol, UK: IOP Publishing), 042003.
- Zhang, T. Q., Chen, Y., Liu, Y., Lin, W. H., and Wang, J. W. (2021). Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat. Commun.* 12 (1), 1–12. doi: 10.1038/s41467-021-22352-4
- Zhang, H., Tan, J., Zhang, M., Huang, S., and Chen, X. (2020). Comparative transcriptomic analysis of two bottle gourd accessions differing in fruit size. *Gene* 11 (4), 359. doi: 10.3390/genes11040359
- Zhang, T. Q., Xu, Z. G., Shang, G. D., and Wang, J. W. (2019). A single-cell RNA sequencing profiles the developmental landscape of arabidopsis root. *Mol. Plant* 12 (5), 648–660. doi: 10.1016/j.molp.2019.04.004

Zhang, D. Y., Zhang, T. X., and Wang, G. X. (2016). Development and application of second-generation sequencing technology. *Environ. Sci. Technol.* 39, 96–102. doi: 10.3969/j.issn.1003-6504.2016.09.017

Zhao, Q. Y., Wang, Y., Kong, Y. M., Lou, D., Li, X., Hao, P., et al. (2011). Optimizing de novo transcriptome assembly from short-read RNA-seq data: a comparative study. *BMC Bioinf.* 12, 644–652. doi: 10.1186/1471-2105-12-S14-S2

Zhao, X., Yu, H., Kong, L., Liu, S., and Li, K. (2014). Comparative transcriptome analysis of two oysters, *Crassostrea gigas* and *Crassostrea hongkongensis* provides insights into adaptation to hypo-osmotic conditions. *MolEcolResour* 14, 139–149.

Zhao, B., Katuwawala, A., Oldfield, C. J., Dunker, A. K., Faraggi, E., Gsponer, J., et al. (2021). DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Res.* 49 (D1), D298–D308. doi: 10.1093/nar/gkaa931