Editorial   **Open Access**

Transcriptomics in Aid to the Establishment of Secondary Metabolic Pathways in Non-Model Plants

Neelam S Sangwan[*]

CSIR-Central Institute of Medicinal and Aromatic Plants, PO CIMAP, India

[*]**Corresponding author:** Neelam S Sangwan, Senior Principal Scientist, Central Institute of Medicinal & Aromatic Plants, India, Tel: 09415766696; E-mail: **nsangwan5@gmail.com**, sangwan.neelam@gmail.com

Introduction

Medicinal and aromatic plants are rich resource of secondary metabolites which are commercially and industrially important. Several plants are being used in various traditional and modern medicine systems. In fact these plants used to be the medicine since ancient times, when modern medicine was not evolved. Recent studies have authenticated and validated to large scale the claims put forward by traditional systems for the use of medicinal plants. Increasingly the traditional medicinal plants are being put up for such scrutiny to evaluate and further validate their potential in modern pharmacological aspects. As such these plants possess a wide spectrum of secondary molecules such as terpenoids, alkaloids, and flavonoids etc. which are essential ingredient of various drugs. These secondary metabolites are occasionally expressed in plants in very specific manner in terms of plant type, stages or development and plant part or organ. Scientists have found it intriguing as to why these secondary metabolites are synthesized by plants in to a diverse scale of unprecedented homology. As such more than 200, 000 such structures are reported to occur from plants.

These metabolic pathways are very intricate and complex. In several instances even most of the steps are not known and characterized and are putatively hypothesized based on information on analogous system. The major bulk of information associated with medicinal plants is generated from various sources which are scattered and generally unlinked, limiting the potential to identify new sources for drugs and other products. Of late there is sudden upsurge in broadening the knowledge base on genomic, transcriptomic and other omic approaches in revealing the metabolic essentialities of the secondary metabolism. Expectedly huge information in this area is about to conglomerate in near future. Earlier approaches including forward genetics yielded although useful progress in imagining the power of the area and also necessity for the inception of faster, efficient and effective methods for knowing what is hidden in each specific and characteristic genome of a medicinal and aromatic plants. Initial and pioneering work basically related to establishing the biosynthetic routes as pinpointed by radio-labelled precursor studies. Many of the landmark and thumb rules evolved and made the basis of secondary metabolism followed even till date. Studies suggested role of various specific and distinct catalytic proteins and enzyme functionalities and specificities. Later on these studies/reports guided to draw a biosynthetic hypothesis or frame to be mapped in terms of annotation and assignment of gene(s) of the relevant pathway.

There existed a huge gap between the timing for interest to study medicinal and aromatic plants and between advent of genomic tools and techniques to a level where getting acquainting the genome of these hard to study complex and too specific metabolome could be eyed/sighted. However, in depth indications could continue to line up utilizing biochemical and enzymic approaches.

In early 2000, After the Arabidopsis genome was out in public domain, a great interest was noticed among biologists. A possible way out to reach to genetic basis for various agronomical, superior and quality traits could be queued for formulation. The genomic information was being lined up when transcriptomic approaches made the way to enter the major labs all across the globe. People started project on transcriptomes of various non-model plants at a gigantic levels never seen or thought before.

Presently, information on a large number of transcriptomes are available of tissues of native conditions as well as in challenged conditions to elucidate various transcriptional events involved in the life cycle of a plant. An increasingly rapid, affordable and high resolution analysis of plant transcriptomics have been driven by next generation DNA sequencing technologies (NGS) through sequencing of their associated cDNA (complementary DNA) populations. This specialized platform is also known as RNA-sequencing (RNA-seq). Despite entering the arena of whole genome profiling technologies which came into light few years back only, RNA-seq has been established as a powerful tool with surprisingly diverse range of applications. The applications utilizing transcripts profiles and abundance may range from detailed studies on biological processes at cell specific level, to other important studies such as gaining insights into fundamental questions in plant biology on evolutionary basis. Major applications include generation of genomic data for therefore to species still to be sequenced to expand the boundaries of so called model organisms, for elucidation of structural and regulatory gene networks controlling various cellular processes in the minutest detail. Furthermore, this would hugely impact in getting the details and in revealing plant's response to developmental cues and the related environment. Additionally, to decipher a better understanding of the gene-product relationship, unifying the "omics" approaches viz transcriptomics, proteomics, and metabolomics into the present system biology paradigm would be imperative step. The advent of RNA-seq methodologies have provided a powerful tool to address questions encompassing cell type-specific transcriptomics, transcript secondary structure and gene mapping as well.

Majorly, huge data in terms of base sequences is the end product of transcript profiling or trasncriptome sequencing in de novo manner. Before, EST based and other sequence information was utilized for the microarray based analysis to identify the transcriptional overtones and regulation of the targeted metabolic pathways. Though, being part of various illustrious publications and reports, micro-array approach did not go so far. On the contrary, omic approaches specifically transcriptomic approaches delivered a huge number of data sets to

virtual screen the entire proposed metabolic pathways, and exercise tremendous flexibility in analyzing data. Slowly micro-array data approaches faded out of the attention of the biologists. Another advantage of, transcriptomic data approaches is in tune to metabolomic data wherein a direct synchronization with the metabolomic data can be perceived. Ultimately transcriptional data points should relate well to other omics approaches to understand the metabolite biosynthetic platform in the widest manner.

High throughput techniques have become extremely important source and meaningful in drawing the righteous conclusions in the absence of a reference genome sequence. High throughput techniques in the absence of a reference genome sequence have become a boon to gather to the maximum level. Transcriptome data is being used to identify putative genes and networks involved in secondary metabolite production in medicinal plants. In such non-model plants, transcriptome studies are helpful for gene discovery, transcript quantification, marker discovery, and also small RNA discovery and regulation. Thus transcriptomic approaches slowly provided accurate and meaningful comparisons, replacing EST based microarray platforms which were quite tedious in experimentation, data analysis and prone to experimental errors in a significant manner. Although transcriptome sequencing itself poses some restrictions and challenges such as production and assembly of large number of short reads, and annotation of huge amount of data. Moreover, generated EST data was also proven to be expensive and time consuming and biased against low-abundance transcripts which forms usual scenario for non model organism viz medicinal and aromatic plants. Contrarily, RNA-seq provides accurate resolution of splice junctions and also reveals alternative splicing events. Furthermore, mining of RNA-seq data in search of transcription start site variation is also improving gene structure annotation as appearing in newer publications. The output may now be achieved as the full-length genes, comparatively defined transcript sequences, comprehensive and genome wide information on SNPs, insertion/deletions, splice variants, characterization of temporal, spatial, regulatory, and evolutionary transcriptome landscapes and mapping. Earlier similar information was unthinkable without exerting too large amounts of technical efforts, un- economical inputs and technical inconsistencies. Under the umbrella of the 1000 plant initiative (oneKP), the transcritome sequencing and mining of genes for medicinal plants are proposed at a mammoth magnitude. Several medicinal plants including Papaver species, Camptotheca acuminata, Catharanthus roseus, Dioscorea villosa, Colchicum autumnale, Gloriosa superba, Taxus baccata, Centella asiatica and Withania somnifera have been analyzed for transcriptomes so far. Also the pace with which transcriptomes are being reported is of towering magnitude and it is expected that soon many key pathways could be established and future synthetic biology approaches would occupy the centre stage of biological research in coming era.