FUNCTIONAL ANALYSIS OF MEDICINAL PLANTS USING SYSTEMS BIOLOGY APPROACHES

YATHISHA NEERAGUNDA SHIVARAJ, SHARATHCHANDRA RAMASANDRA GOVIND1, SUDISHA JOGAIAH2, DEVARAJA SANNANINGAIAH3
1Department of Studies and Research in Environmental Science, 2Department of Studies and Research in Biochemistry, Tumkur University, Tumkururu, Karnataka 572103, 3Department of Microbiology and Biotechnology, Karnataka University, Dharwad Email: rgschandra@gmail.com

Received: 25 Aug 2015 Revised and Accepted: 30 Sep 2015

ABSTRACT

Plant derived medicine is an important source of life saving drugs, but the genome information of most important medicinal plants is still unavailable. The need of the hour is to identify more functional genes and enzymes that control secondary metabolite production in medicinal plants, metabolite levels at different geographic locations, natural growth environments, or cultivation conditions. It can help us understand the metabolic pathways for the production of these bioactive compounds generate metabolic fingerprinting of medicinal plants for the authentication and quality control, classify medicinal plants, and establish a quantitative version of chemotaxonomic analysis to advance our knowledge of the evolutionary relationship of medicinal plants. In addition recent advances in metabolomics have enabled rapid identification and quantification of as yet unknown metabolites [11]. Rapid identification of existing compounds, and generating new knowledge of the pharmacological and toxic effects of the plant under study and the chemical ecological aspects of the metabolite in specific is dealt with very efficiently using different metabolomic tools. MS and NMR methods are the backbone of any metabolomic tools. MS and NMR methods are the backbone of any metabolomic tools.
determination of the quantity of gingoic acids from Ginkgo leaves and in six commercial Ginkgo products [14]. Metabolic profiling of Angelica cutíloba (Sieb.  &Zucc.) Kitag. roots, has been carried out using gas chromatography-time-of-flight-mass spectrometry which enabled quantification of a number of metabolites in a tissue specific manner in addition multivariate pattern recognition could also be established based on the taxonomy of the metabolites identified [15]. Comparative metabolomics strategy coupled with cell and gene-based assays was used for species classification and anti-inflammatory bioactivity validation of medicinal Echinacea species, i.e. Echinacea purpurea (L) Moench, E. pallida Nutt. and E. angustifolia DC [16]. More metabolic profiling have been conducted in the medicine (Medicago truncatula) which is also a model organism for legume biology [17-19]. Popular methods used to perform metabolomic experiments are mass spectrometry (MS) and nuclear magnetic resonance (NMR). More focus on lipid compounds has been noticed and studied all in a particular physiological state is called lipidomic technology is a valuable high through put method that can potentially be used to identify novel biomarkers and also for quantification of lipids and other related metabolites [20].

Proteomics

Proteomics is the information of a whole proteome. It also refers to the complete complement of proteins including the modifications that proteins undergo in a particular physiological state. Some of the post-translational modifications are studies include methylation, acetylation, glycosylation, oxidation, and nitrosylation [21]. Most of the published proteomics data of medicinal plants are most obtained from Medicago the focus of these studies is on the protein expression and phosphorylation changes at various conditions [22-26]. The foremost reason for less number of publications in proteomics is the lack annotation of protein and gene sequence information on medicinal plants. However, recent improvements in technologies like tandem mass spectrometry which enable de novo sequencing of proteins more reports on the full complement of proteins is expected to rise in the near future.

Genomics and Epigenomics

Genomics is the study of an organism’s whole genome. Genome refers to all of the DNA sequences in an organism. So far, there are very few genomes of medicinal plants that have been fully sequence. A recent draft of Neem (Azadirachta indica) has been published [27]. The Azadirachta indica (neem) tree is a source of a wide range of natural products, including the potent biopesticide azadirachtin. In spite of its widespread applications in agriculture and medicine, the molecular basis of the biosynthesis of neem terpenoids remain largely unexplored. The current report describes the draft genome and four transcriptomes of A. indica and attempts to contextualize the sequence information in terms of its molecular phylogeny, terpene expression pattern and terpenoid biosynthesis pathways. A. indica is the first member of the family Meliaceae to be sequence using next generation sequencing approach. The genome and transcriptomes of A. indica were sequenced using multiple sequencing platforms and libraries. The A. indica genome is AT-rich, bears little repetitive DNA elements and comprises about 20,000 genes. Molecular phylogenetic analyses grouped A. indica together with Citrus sinensis from the Rutaceae family validating its conventional taxonomic classification. Comparative transcript expression analysis showed either exclusive or enhanced expression of known genes involved in neem terpenoid biosynthesis pathways compared to other sequences in angiosperms. Genome and transcriptome analyses in A. indica led to the identification of repeat elements, nucleotide composition and expression profiles of genes in various organ [27]. However, the number of publications associated with identification and authentication of medicinal plants at the DNA level has increased exponentially. DNA-based techniques that do not require whole genome information like PCR, RFLP, AFLP, RAPD, and sequencing are employed to resolve ambiguities in plant identification and discrimination [27; 29:30:31]. Since a number of medicinal plants fall under the endangered species, category it is essential to select one or more model medicinal plants to sequence their entire genomes which will profoundly enhance the research of medicinal plants. Epigenomics is the whole genome level study of epigenetic elements. Epigenomics is important for us to understand the mechanisms of gene expression changes in medicinal plants. "Omics" Data analyses

Omics technologies usually produce huge data sets that require statistical tools and several software tools for analysis. Most of the statistical tools are based on clustering algorithms which can be used to cluster genes that have similar behaviours and identified in a microarray experiment the genes in the same cluster should have similar functions. Such algorithms can be utilized to predict secondary metabolites or pathways that are involved in similar biological activities. Clustering algorithms can be subdivided into unsupervised and supervised algorithms. Other algorithms such as classification algorithms can be employed to classification, authentication and quality control. In addition, it can be utilized to evolution and systematics of the medicinal plant. Classification algorithms can help in identifying a list of efficient markets to precisely distinguish different restorative materials using data from chemical composition, quantity and morphology. There are many feature selection methods including support vector machine recursive feature elimination (SVM-RFE), chi-square, info gain, gain ratio, relief, wrapper, and CSF. Generally used for classification algorithms include decision tree J48, random forest (RF), Nave Bayes (NB), simple logistic (SL), RBF Neural Nets, MLP neural nets and support vector machines (SVMs). These algorithms use microarray data and can be used for classification and clustering of genes, proteins and metabolites into an integrated map. Many such algorithms can be integrated by writing software application tools which are user friendly. Such tools with a user friendly graphical user interface help biologists and other application scientists in classifying data based on different functional categories [32].

CONCLUSION

Integrating transcriptomics, proteomics and metabolomics data can help us to predict gene function particularly for genes involved in complicated pathways that can produce bioactive constituents. Through integrated metabolite and transcript profiling, a biosynthetic mechanism for hispidol in Medicago truncate, cell cultures was characterized [33]. The major challenge for systems biologist is to construct a transcriptional regulatory network and over lay that of a metabolite and post-translational modification map by using reverse engineering algorithms. By collaborating with computer scientists, an ensemble learning approach to reverse engineering transcriptional regulatory networks from time-series gene expression data has to be developed which can then be used with gene ontology mapping methods to analyse genomic, proteomics and metabolomic data in tandem Such an integrated approach will enable discovery of novel biomarkers based on functional and epigenetic data. The data thus generated can be overlaid on heat maps to predict patterns of gene expression, protein modifications and metabolite profiles. Such integration will enable us to take a decision on the nature and structure of genetic interventions necessary for production of desired metabolites.

CONFLICT OF INTERESTS

Declared None

REFERENCES

distinct stages of nodule organogenesis are accompanied by biosynthesis in Medicago truncatula cell cultures. Plant Physiol 2008;146:387-402.


"Drugs from Nature: Plants as an important source of pharmaceutically important metabolites" Guest Editor: Dr. Dhananjaya Bhadrapura Lakkappa