Innovare Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 7, Supple 1, 2015

Review Article

FUNCTIONAL ANALYSIS OF MEDICINAL PLANTS USING SYSTEMS BIOLOGY APPROACHES

YATHISHA NEERAGUNDA SHIVARAJ, SHARATHCHANDRA RAMASANDRA GOVIND[1], SUDISHA JOGAIAH[2], DEVARAJA SANNANINGAIAH[3]

[1]Department of Studies and Research in Environmental Science, [3]Department of Studies and Research in Biochemistry, Tumkur University, Tumakuru, Karnataka 572103, [2]Department of Microbiology and Biotechnology, Karnatak 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 medical plants, develop new methods for systematics, engineer resistance to number of biotic and abiotic stresses, and develop new conservation strategies, more genomics, proteomics and metabolomics information needs to be produced. In this review, a brief overview of various omic technologies and its applications to medicinal and aromatic plants are discussed.

Keywords: Medicinal Plants, Omic Approaches, Transcriptomics, Metabolomics, Genomics and Epigenomics**,** Bioinformatics.

INTRODUCTION

To understand the biological and other metabolic processes of medicinal plants it is essential to study these plants using a combination of omic approaches. In order to analyse the data generated from such studies in bioinformatics and other associated tools are essential [1]. Use of all omic data combined with chemometric analysis tools like bioinformatics is termed as systems biology. Systems biology helps is complete understanding of a system or a physiological state by integrating different "omics" data sources for the analysis of networks and regulation. The ultimate aim is to how the system responds as a whole to a particular physiological state [2].

Transcriptomics

Transcriptome can be described as the set of all RNA molecules, including mRNA, tRNA, rRNA, and non-coding RNA produced by a single cell or a group of cells in response to physiological stimuli. Transcriptomics therefore, is the study of global analysis of gene expression profile at the mRNA level. Transcriptomic technologies generally involve genome wide gene expression methods like cDNA-AFLP, SAGE, DNA microarray and oligo-microarray. Currently however, the most popular transcriptomics method is oligo-microarray because it offers the advantage of complete genome coverage. Microarray is also one of the most popular tools and can be used to measure gene expression changes of medicinal plants which could be later related to metabolite expression levels using systems biology. Different organs of the same plant, the same organ at different development stages, *in vitro* cultured plants or cells are used as a source for extraction of DNA/RNA for analysis. Transcriptomic studies on medicinal plants include measuring metabolite levels at different geographic locations, natural growth environments, or cultivation conditions. However, scientific investigations using high throughput gene expression profiles to study gene expression changes in medicinal plants are very rare. One of the most interesting investigations used cDNA microarray examine the gene expression profiles of hairy root of Salvia miltiorrhiza Bunge at different stages [3]. Such experiments aim to identify genes that are in charge of regulating production of secondary metabolites with medicinal value. Many authors have reported using transcriptome approaches for identifying specific gene functions from medicinal plants.

Examples include Identification of Ethylene responsive element binding protein genes for S. militiorrhiza [4]. In addition DNA microarray can also be used for systematic authentication of medicinal plants by using 5Sribosomal RNA genes [5] or 18SRNAgenes across closely related species. Microarray experiments use whole genome sequence information for a specific organism. However, the whole genome of any single medicinal plant has been sequenced with the exception of *Azadirecthaindica*. In order to bridge this lacunae transcriptome information through generating expression sequence tags (ESTs) is the choice of many scientific gropus around the world. ESTs from several medicinal plants such as *Panax quinquefolius* L. [6, 7], *Huperziaserrata* (Thunb. ex Murray) Trev [8]. *P. notoginseng* (Burk.) [7]. *Rehmanniaglutinosa* Libosch [6]. and *Catharanthusroseus* (L.) G. Don [9] are available. Recently developed next generation sequencing technologies, such as illumine, 454, and solid approaches have revolutionized the EST generation. Millions of EST sequences of Northern bobwhite (*Colinusvirginianus* L.) and red earthworm have been generated from the 454 platform. Assembled unique sequences are used to design oligo-microarrays [10].

Metabolomics

Metabolomics is the science of profiling all the metabolites produced in an organism. More than 50, 000 metabolites have been characterized from medicinal and aromatic plants. Metabolomics is an extremely useful, due to the fact that one of the main objectives of studying medicinal plants is to find active compounds, which, in fact, are metabolites. Therefore, fast development of metabolomics technology provides us a valuable opportunity to advance the studies of medicinal plants. One advantage of conducting metabolomics is that genomics information is not needed. Metabolomics can be used to compare metabolite quantitative changes in medicinal materials between different organisms, ages, origins, organs, developmental stages, environmental cultivation and culture conditions, and processing methods. 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 evolutional 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 metabolic fingerprinting study and they have been successfully employed to evaluate the quality of herbal material and phyto pharmaceutical [12]. Some of the examples of, metabolomic profiling include differentiation of 12 Cannabis sativa cultivars [13, 14] and

determination of the quantity of gingolic acids from Ginkgo leaves and in six commercial Ginkgo products [14]. Metabolic profiling of Angelica cutiloba (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 study of all lipids 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 that 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 (Azadirechtaindica) has been published [27]. The Azadirachtaindica (neem) tree is a source of a wide number of natural products, including the potent biopesticideazadirachtin. In spite of its widespread applications in agriculture and medicine, the molecular aspects 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 contextualise the sequence information in terms of its molecular phylogeny, transcript expression 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 sequences 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 organs [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

Omic 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), Naivy 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 transcirptomics, 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 biologists 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 epigenomic 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

1. Kann MG. Advances in translational bioinformatics: computational approaches for the hunting of disease genes. Briefings Bioinf 2009;11:96-110.
2. Ray LB, Chong LD, Gough NR. Computational biology: Sci STKE; 2002. p. 10.
3. Cui GH, Huang LQ, Qiu DY, Yuan Y, Fu GF. Functional genomics studies of Salvia miltiorrhiza II-gene expression profiling of different stage of hairy root. China J Chin Mater Med 2007;32:1267-72.
4. Xu B, Huang L, Cui G, Mao Y, Zhang H. Functional genomics of Salvia militiorrhiza IV--analysis of ethylene responsive element binding protein gene. China J Chin Mater Med 2009;34:2564-6.
5. Carles M, Cheung MK, Moganti S, Dong TT, Tsim KW, Ip NY, *et al.* A DNA microarray for the authentication of toxictraditional Chinese medicinal plants. Planta Med 2005;71:580-4.

6. Wu Q, Song JY, Sun YQ, Suo FM, Li CJ, Luo HM, *et al.* Transcriptprofiles of panax quinquefolius from flower leaf and root bring new insights into genes related to ginsenosides biosynthesis andtranscriptional regulation. Physiol Plant 2010;138:134-49.
7. Chen SL, Sun YQ, Song JY, Li Y, Li CJ, Hu SN, *et al.* Analysis of expressed sequence tags (EST) from panax quinquefolium root. Acta Pharm Sin B 2008;43:657-63.
8. Luo HM, Sun C, Li Y, Wu Q, Song JY, Wang DL, *et al.* Analysis of expressed sequence tags from the Huperziaserrata leaf for gene discovery in the areas of secondary metabolite biosynthesis and development regulation. Physiol Plantarum 2009;139:1-12.
9. Shukla AK, Shasany AK, Gupta MM, Khanuja SP. Transcriptome analysis in Catharanthus roses leaves and roots for comparative terpenoid indole alkaloid profiles. J Exp Bot 2006;57:3921-32.
10. Gong P, Pirooznia M, Guan X, Vera JC, Liang C, Deng YP, *et al.* l. A strategy for high throughput microarray oligo probe design using 454 and sanger sequencing data of Eisenia Fetida expressed sequence tags. The 2008 World Congress in Computer Science Computer Engineering and Applied Computing, USA; 2008. p. 1040-3.
11. Wang LJ, Wang SZ, Li YH, Paradesi MS, Brown SJ. The model organism database for Triboliumcastaneum. Beetle Base. Nucleic Acids Res 2007;35:D476-D479.
12. Van der KF, Maltese F, Choi YH, Kim HK, Verpoorte R. Qualitycontrol of herbal material and phytopharmaceuticals with MS and NMR based metabolic fingerprinting. Planta Med 2009;75:763-75.
13. Choi YH, Kim HK, Hazekamp A, Erkelens C, Lefeber AW, Verpoorte R. Metabolomic differentiation of Cannabissativa cultivars using 1H-NMR spectroscopy and principal component analysis. J Nat Prod 2004;67:953-7.
14. Choi YH, Choi HK, Peltenburg Looman AM, Lefeber AW, Verpoorte R. Quantitative analysis of ginkgolic acids from Ginkgo leaves and products using 1H-NMR. Phytochem Anal 2004;5:325-30.
15. Tianniam S, Tarachiwin L, Bamba T, Kobayashi A, Fukusaki E. Metabolic profiling of Angelica acutiloba roots utilizing gas chromatography-time-of-flight-mass spectrometry for quality assessment based on cultivation area and cultivar via multivariatepattern recognition. J Biosci Bioeng 2008;105:655-9.
16. Hou CC, Chen CH, Yang NS, Chen YP, Lo CP, Wang SY, *et al.* Comparative metabolomics approach coupled with cell-and gene-based assays for speciesclassification and anti-inflammatory bioactivity validation of Echinacea plants. J Nutr Biochem 2010;21:1045-59.
17. Farag MA, Huhman DV, Dixon RA, Sumner LW. Metabolomics reveals novel pathways and differentialmechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in Medicago truncatula cell cultures. Plant Physiol 2008;146:387-402.
18. Barsch A, Tellstrom V, Patschkowski T, Kuster H, Niehaus K. Metabolite profiles of nodulated alfalfa plants indicate that distinct stages of nodule organogenesis are accompanied by global physiological adaptations. Mol Plant Microbe Interact 2006;19:998-1013.
19. Barsch A, Carvalho HG, Cullimore JV, Niehaus K. GC-MS based metabolite profiling implies three interdependent ways of ammonium assimilation in Medicago truncatula root nodules. J Biotechnol 2006;127:79-83.
20. Mao JH, Shen Wu M, Roth M, Welti R, Deng YP. Lipidomic profiles reveal the potential useful plasma biomarkers associated with type II diabetes in black and white patients. Biocomputing 2009;2:774-9.
21. Blackstock WP, Weir MP. Proteomics. Quantitative and physical mapping of cellular proteins. Trends Biotechnol 1999;17:121-7.
22. Castillejo MA, Maldonado AM, Dumas Gaudot E, Fernandez paricio M, Susin R, Diego R, *et al*. Differential expressionproteomics to investigate responses and resistance to Orobanchecrenata in Medicago truncatula. BMC Genomics 2009;10:294.
23. Dam S, Laursen BS, Ornfelt JH, Jochimsen B, Staerfeldt HH, Friis C, *et al.* The proteome of seed development in the model legume Lotus japonicus. Plant Physiol 2009;149:1325-40.
24. Aloui A, Recorbet G, Gollotte A, Robert F, Valot B, Gianinazzi Pearson V, *et al.* On the mechanisms of cadmium stress alleviation in Medicagotruncatula by particular mycorrhizae symbiosis. A root proteomic study. Proteomics 2009;9:420-33.
25. Van Noorden GE, Kerim T, Goffard N, Wiblin R, Pellerone FI. Overlap of proteome changes in Medicago truncatula in response to auxin and Sinorhizobiummeliloti. Plant Physiol 2007;144:1115-31.
26. Prayitno J, Imin N, Rolfe BG, Mathesius U. Identification of ethylene mediated protein changes during nodulation in Medicago truncatula using proteome analysis. J Proteome Res 2006;5:3084-95.
27. Krishnan. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm Azadirachtaindica. BMC Genomics 2012;13:464.
28. Titanji VP, Ngwa AA, Ngemenya M. Applications of biotechnology techniques to the study of medicinal plants. Afr J Med Med Sci 2007;36:23-9.
29. Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, *et al.* Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 2010;5:8613.
30. Song JY, Yao H, Li Y, Li XW, Lin YL, Liu C, *et al*. Authentication of the family Polygonaceae in Chinesepharmacopoeia by DNA bar coding technique. J Ethnopharmacol 2009;124:434-9.
31. Yu HH, Yang ZL, Yang X, Tan ZF, Shu X, Liu RN. Advancesin studies on ITS sequence of medicinal plants germplasmresources. Chin Tradit Herb Drugs 2009;41:491-6.
32. Liu QZ, Sung AH, Chen ZX, Liu JZ, Huang XD, Deng YP. Feature selection and classification of MAQC-II breast cancerand multiple myeloma microarray gene expression data. PLoS One 2009;4:8250.
33. Farag MA, Deavours BE, de FA, Naoumkina M, Dixon RA, SumnerLW. Integrated metabolite and transcript profiling identify the biosynthetic mechanism for hispidol in Medicago transcultural cultures. Plant Physiol 2009;151:1096-113.