

# A Review on Application of Bioinformatics in Medicinal Plant Research

**Harishchander A\***

Department of Bioinformatics, Sathyabama University, India

**\*Corresponding author:** Harishchander Anandaram, Department of Bioinformatics, Sathyabama University, India, Email: harishchander.a@gmail.com

**Mini review**

Volume 1 Issue 1

**Received Date:** September 05, 2017

**Published Date:** October 02, 2017

## Abstract

Plants serve as a source of medicine from historic times and more commercially important drugs are based on the origin of plants. The traditional approach in discovering plant-based drugs involves a significant amount of expenditure and time. In labor-intensive approach, more struggles were involved and hence there was a rapid development of high-throughput technologies. In the era of post genomics, generation of data was high-throughput and hence, bioinformatics plays a crucial role. In general, rational analysis is vital for drug designing and discovery. However, more attention is required to address the potential application of bioinformatics with respect to plant-based knowledge. In this chapter, a review bioinformatics studies were focused to identify the contribution in medicinal plants research. In particular, specific areas were highlighted in medicinal plant research where the potential application of methodologies in bioinformatics may result in a rapid and cost-effective lead generation towards finding remedies from plants.

**Keywords:** Medicinal plants; Bioinformatics; Lead generation; Drug discovery

## Introduction

Plants are a valuable resource for a variety of products. Plant materials are used for many purposes including food and medicine. In case of medicine, the uses of plant-based materials were also present in ancient civilizations. There are several ancient records that provide evidence regarding use of plant sources of remedies [1,2]. The knowledge from ancient systems of plant-based remedies has also been used by the modern pharmaceutical industry. There is thus an immense potential for discovery of new drugs from plants based on the ethno-medicinal data [3,4]. About one-third of currently available drugs come from natural products that have a plant origin [5]. Even though plant-based remedies have much potential towards advancing modern medical treatments, research continues to lag behind (especially when compared to the interest in developing synthetic drugs for commercial use) [6].

This may be partly because conventional plant drug discovery methodologies can be slow and expensive [7]. Nonetheless, there may be utility to increase research in the area of medicinal plants. The available literature and resources in this area is generally scattered, which hinders the ability to readily leverage available information about medicinal plants. There are several computational approaches for analyzing the diversity of compounds. These approaches have played a significant role in computer-aided drug design [8]. The field of drug design and discovery from medicinal plant requires the application of such approaches for quicker and efficient progress so as to cope up with the continually demanding pharmaceutical needs.

Bioinformatics offers a suite of essential techniques for analyzing and interpreting huge volumes of information generated using molecular biology-based techniques. With the advancement of high-throughput

techniques, such approaches have become essential in analyzing and integrating data to infer knowledge from a whole systems point of view. To increase our understanding of cellular processes associated with plants, an in depth analysis of genomic, proteomic and metabolomic information is required. Bioinformatics approaches offer essential tools for the identification of genes and pathways that may be associated with important bioactive secondary metabolites from medicinal plants [9].

In this review the focus is on the potential applications of computational methodologies for the overall advancement of plant-based drug discovery. Different areas are explored where use of such approach can lead to valuable findings in a cost and time efficient manner. Aspects related to the integration of scattered information, analysis of molecular data, drug discovery and design, authentication and toxicology are discussed with focus on computational methods.

### Bioinformatics & Plant Research

A limited number of plants have whole-genome sequence data available. To date, the majority of genomics resources for plants have come from ESTs. Transcript-level information could be valuable to molecular biology-based research relative to medicinal plants. Transcriptome data has been used to identify putative genes and networks involved in secondary metabolite production in medicinal plants [10-12]. Analysis of transcriptome data can also be helpful in predicting transcription factors, response elements and effector genes involved in bioactive metabolite synthesis [13-15]. For example, ethylene responsive element binding genes were analyzed in *Salvia miltiorrhiza* [16]. Another example is the identification of miRNAs, their targets and transcription factors involved in secondary metabolism pathways from *Salvia sclarea* L. [17]. Once EST data are generated and assembled, an essential next step is annotation. There are several resources like KEGG genes, SwissProt, TAIR, NCBI's nonredundant and nucleotide databases that provide a platform for annotation of sequence data. EST data can also be used for mining of molecular markers [18-20]. Identification of molecular markers can be used in studies involving linkage mapping, comparative genomics, identification of different species and distribution of genes on chromosomes [21-24]. Compared to other EST-based markers, Simple Sequence Repeat (SSR) markers have been shown to be most advantageous because of their multi-allelic nature, reproducibility, codominant inheritance, high abundance and extensive genome coverage [25]. SSR Locator is an example of a computational approach for detection and characterization of SSRs and mini-satellite motifs [26].

Bioinformatics approaches can be used to create coexpression networks from transcriptome data, providing possible leads to gene discovery in related plant species. In particular, the use of comparative genomics provides basis for exchange of information among the different species. Plant-specific data sets can be retrieved from PLEXdb [27], GEO [28] and EBI ArrayExpress [29]. Coupled with the study of co-expression networks, it may be possible to discover genes of interest and their function. For example, transcriptome data from barley have been collected and used to create a coexpression network [30]. Results from coexpression analyses were further used to derive subnetworks ('modules') associated with biological functions, with particular emphasis given to identifying modules related to drought stress and cellulose biosynthesis. This genome scale sequence comparisons have been shown to reveal several Triticeae species-specific genes that are related to specific regulatory networks [31].

Pathway analysis can be valuable approach for identifying potential functional roles of genes. The KEGG is a resource that provides a platform for pathway analysis of secondary metabolites from several organisms [32]. The KEGG Drug database further provides information related to two types of molecular networks: (i) interaction of drugs with target molecules and (ii) biosynthetic pathways of natural products in various organisms. KEGG Drug contains chemical structures or components of prescription and Over-The-Counter (OTC) drugs as well as drugs from TCM [33]. This information could potentially be used for drug discovery from the genomes of plants. Another resource for pathway analysis of secondary metabolites indexed in KEGG is PathPred [34]. PathPred is web server that predicts pathways of multi-step reaction for a given query compound, starting with a similarity search against the KEGG COMPOUND database. This server was designed for pathways associated with microbial biodegradation of environmental compounds and biosynthesis of secondary plant metabolites. Nonetheless, PathPred reflects generalized reactions shared among structurally related compounds.

With a myriad of advances in 'omic' technologies, bioinformatics plays essential role in facilitating systems level understanding of metabolic processes. Integration of transcriptomic and metabolomic data facilitated by data mining techniques offers many opportunities to study metabolic pathways [35]. Expression patterns of intensities of ESTs and mass peaks classified by batch-learning self-organizing maps revealed regulatory linkages among nutrient deficiency, primary metabolism and glucosinolate metabolism [36]. Gene-metabolite coexpression analysis led to identification of terpene synthase genes involved in volatile compound formation in cucumber [37].

Rischer, et al. [38] analyzed a gene-metabolite coexpression network of the medicinal plant *Catharanthus roseus* to identify possible genes and metabolites associated with the biosynthesis of terpenoid indole alkaloids. Integrated gene-metabolite expression analyses have thus shown potential for examining metabolic regulation of nonmodel plants of potential medicinal value.

Bioinformatics provides essential mechanisms to analyze bulk information generated from high throughput techniques. In particular, such approaches have made it possible for the identification of putative genes, pathways and networks involved in synthesis of bioactive metabolites in medicinal plants. In addition to facilitating the analysis of high-throughput data, bioinformatics approaches can be important for connecting scattered pieces of evidence into meaningful hypotheses thereby generating potential leads for experimental validation.

## Conclusion

Plants can be a valuable source of pharmacologically important compounds. Bioinformatics approaches may provide an essential set of tools for designing efficient and targeted searches for plant-based remedies. This review highlighted the different aspects associated with medicinal plant research where bioinformatics strategies could be employed to attain significant progress. The combination of bioinformatics strategies may enable a new era of plant-based drug discovery.

## References

1. Cowan MM (1999) Plant products as antimicrobial agents. *Clin Microbiol Rev* 12(4): 564-582.
2. Patwardhan B, Warude D, Pushpangadan P, Narendra B (2005) Ayurveda and traditional Chinese medicine: a comparative overview. *eCAM* 2: 465-473.
3. Fabricant DS, Farnsworth NR (2001) The value of plants used in traditional medicine for drug discovery. *Environ Health Perspect* 109(1): 69-75.
4. Clarkson C, Maharaj VJ, Crouch NR, Grace OM, Pillay P, et al. (2004) *In vitro* antiplasmodial activity of medicinal plants native to or naturalised in South Africa. *J Ethnopharmacol* 92: 177-191.
5. Strohl WR (2000) The role of natural products in a modern drug discovery program. *Drug Discov Today* 5(2): 39-41.
6. Miller J (2011) The discovery of medicines from plants: a current biological perspective. *Econ Botany* 65(4): 396-407.
7. DiMasi JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. *J Health Econ* 22(2): 151-185.
8. Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303(5665): 1813-1818.
9. Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol* 61: 463-489.
10. Li Y, Luo HM, Sun C, Song JY, Sun YZ, et al. (2010) EST analysis reveals putative genes involved in glycyrrhizin biosynthesis. *BMC Genomics* 11: 268.
11. Chen S, Luo H, Li Y, Sun Y, Wu Q, et al. (2011) 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Rep* 30(9): 1593-1601.
12. Luo H, Li Y, Sun C, Wu Q, Song J, et al. (2010) Comparison of 454-ESTs from *Huperzia serrata* and *Phlegmariurus carinatus* reveals putative genes involved in lycopodium alkaloid biosynthesis and developmental regulation. *BMC Plant Biol* 10: 209.
13. Yuan D, Tu L, Zhang X (2011) Generation, annotation and analysis of first large-scale expressed sequence tags from developing fiber of *Gossypium barbadense* L. *PloS One* 6(7): e22758.
14. Kavitha K, Venkataraman G, Parida A (2008) An oxidative and salinity stress induced peroxisomal ascorbate peroxidase from *Avicennia marina*: molecular and functional characterization. *PPB/Societe Francaise De Physiologie Vegetale* 46: 794-804.
15. Cabral A, Stassen JH, Seidl MF, Jaqueline B, Jane EP, et al. (2011) Identification of *Hyaloperonospora arabidopsidis* transcript sequences expressed during infection reveals isolate-specific effectors. *PloS One* 6: e19328.
16. Xu B, Huang L, Cui G (2009) [Functional genomics of *Salvia miltiorrhiza* IV-analysis of ethylene responsive element binding protein gene] *Zhongguo Zhong Yao Za Zhi = Zhongguo Zhongyao Zazhi = China Journal Of Chinese Materia Medica* 34: 2564-2566.
17. Legrand S, Valot N, Nicole F, Moja S, Baudino S, et al. (2010) One-step identification of conserved miRNAs, their targets, potential transcription

- factors and effect of genes of complete secondary metabolism pathways after 454 pyrosequencing of calyx cDNAs from the Labiate *Salvia sclarea* L. *Gene* 450(1-2): 55-62.
18. Joshi RK, Kar B, Nayak S (2011) Exploiting EST databases for the mining and characterization of short sequence repeat (SSR) markers in *Catharanthus roseus* L. *Bio information* 5(9): 378-381.
  19. Victoria FC, da Maia LC, de Oliveira AC (2011) *in silico* comparative analysis of SSR markers in plants. *BMC Plant Biol* 11: 15.
  20. Zeng S, Xiao G, Guo J, Fei Z, Xu Y, et al. (2010) Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics* 11: 94.
  21. Wang CM, Liu P, Yi C, Gu K, Sun F, et al. (2011) A first generation microsatellite- and SNP-based linkage map of *Jatropha*. *PLoS One* 6(8): e236-232.
  22. Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, et al. (2011) A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.) *BMC Plant Biol* 11: 111.
  23. Korotkova N, Borsch T, Quandt D, Taylor NP, Müller KF, et al. (2011) What does it take to resolve relationships and to identify species with molecular markers? An example from the epiphytic *Rhipsalideae* (Cactaceae) *Am J Botany* 98(9): 1549-1572.
  24. Venuprasad R, Bool ME, Quiatchon L, Atlin G N (2011) A QTL for rice grain yield in aerobic environments with large effects in three genetic backgrounds, TAG. *Theoretical and applied genetics. Theoretische Und Angewandte Genetik* 124(2): 323-332.
  25. Varshney RK, Graner A, Sorrells ME. (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23(1): 48-55.
  26. da Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FIF, et al. (2008) SSR locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int J Plant Genomics* 412696.
  27. Wise RP, Caldo RA, Hong L, Shen L, Cannon E et al. (2007) Barley Base/PLEXdb. *Methods Mol Biol* 406: 347-363.
  28. Barrett T, Edgar R (2006) Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol* 411: 352-369.
  29. Rocca-Serra P, Brazma A, Parkinson H, Sarkans U, Shojatalab M, et al. (2003) Array Express: a public database of gene expression data at EBI. *Comptes Rendus Biologies* 326(10-11): 1075-1078.
  30. Mochida K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, Shinozak K, et al. (2011) Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops. *Plant Cell Physiol* 52(5): 785-803.
  31. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M, et al. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38:D355-360.
  32. Kanehisa M (2009) Representation and analysis of molecular networks involving diseases and drugs. *Genome Inform Int Conf Genome Inform* 23: 212-213.
  33. Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, et al. (2010) PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res* 38: W138-143.