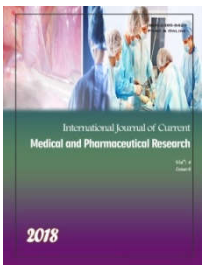
Research Article

INTERNATIONAL JOURNAL OF CURRENT MEDICAL AND PHARMACEUTICAL RESEARCH

*ISSN: 2395-6429, Impact Factor: 4.656*
*Available Online at www.journalcmpr.com*
*Volume 4; Issue 6(A); June 2018; Page No. 3318-3326*
*DOI: http://dx.doi.org/10.24327/23956429.ijcmpr20180453*




# MEDICINAL PLANT GENOME: A SOURCE OF FINDING NEW ENZYMES, METABOLIC PATHWAYS AND DRUG DISCOVERY

**Prasanta Chakraborty***

Kalpana Chawla Center for Space and Nanosciences, Kolkata Indian Institute of Chemical Biology (retd.), Kolkata-700032, India

**ARTICLE INFO**

***Article History:***
Received 12th March, 2018
Received in revised form 10th April, 2018
Accepted 7th May, 2018
Published online 28th June, 2018

***Key words:***
Medicinal plant, genomics, omics, enzymes, drugs

**ABSTRACT**

New drugs may be developed from secondary metabolites of medicinal plant metabolism. Identification of new metabolic pathways, unknown enzymes of vast unexplored medicinal plants will be a really very challenging area for future research. Information on underlying genes of unknown enzymes/metabolic pathways are hidden away in the dark matter of plant genomes. Genome sequencing and analysis by breakthrough sequencing technology like inexpensive NGS (Next-generation sequencing) technology may accelerate/fecilitate the gene identification. The emergence of this genomic research along with other omics research data including transcriptomic, metabolomics data may find out gene-metabolite linkage and discovery of new potential molecules. This review discusses the recent developments on medicinal plant genomic research and their importance in new drug discovery.



## INTRODUCTION

Medicinal plants have got enormous potential, and many of our drugs from those plants are products from secondary metabolism of the plant biosynthetic pathway (structures of some drug molecules given in Fig.1, modified from Chang *et al*.,2016, Luca *et al*.,2012). There are thousands of medicinal plants, and they may be the fertile source of many biologically active molecules/drugs. The vast majority of chemical potential of unexplored medicinal plants awaits discovery and is hidden away in plant genomes. Identification of underlying genes for the enzymes and metabolic pathways for the biosynthesis of new products/molecules inside medicinal plant, requires genome sequencing. Now, breakthroughs in sequencing technology, and the use of inexpensive next-generation sequencing (NGS) technology will definitely accelerate the ability to find enzymes and pathways for the biosynthesis of new natural products.

The World Health Organization has listed 21,000 medicinal plants (Modak *et al*.2007), however, till date, only few of plant-derived compounds are in clinical pipeline, either due to low production levels in plant species or due to loss of source for extinction (Brower,2008). The solution to overcome this problem lies on technologies such as metabolic engineering of effective plant and microbial production platform (like, Fig.2), so that more and more plant-derived compounds having enormous structural diversity and biological activities enter the clinical pipeline.

Considering thousands of unexplored medicinal plants and their chemical biodiversity, its now worth researching on the agricultural traits, genetic background and the medicinal quality and values of those plants. Emerging genomics research, with fast and inexpensive high throughput sequencing technologies, together with transcriptomic, proteomic and metabolomic data can altogether be used to predict the secondary metabolic pathways of medicinal plants. In future, it is hoped that the discovery of previously unknown pathways/enzymes of unexplored medicinal plants will help to find out new pharmaceutical agents.

### *Current status of the Genome research/sequencing of medicinal plants*

The genomic studies of medicinal plants lag behind those of model plants and important crop plants, however, knowledge on high-throughput sequencing of medicinal plants, very much important as it shed light on the biosynthetic pathways of medicinal compounds, especially secondary metabolites, and also play a major role in the molecular breeding of high-yielding medicinal cultivars and molecular farming of transgenic medicinal strains. High-throughput sequencing or next-generation sequencing (NGS) comprises different modern sequencing technologies (Illumina sequencing, Roche 454 sequencing, Iontorrent:Proton/PGM sequencing, SOLiD

**Corresponding author:* **Prasanta Chakraborty**
Kalpana Chawla Center for Space and Nanosciences, Kolkata Indian Institute of Chemical Biology (retd.), Kolkata-700032, India

sequencing) and these technologies (Morey *et al*.,2013, Metzker,2010) revolutionized genomic research with their quick and economical performance, e.g. entire human genome could be sequenced within a single day. Also, development of NGS coupled with the advancement of computational methods, has allowed researchers to access even the transcriptomes of recalcitrant genomes such as those of medicinal plant species. However, as there is a lack of comprehensive molecular genetic studies on most medicnal plants, it is vital to have some preliminary genome evaluations done before whole-genome sequencing, such as DNA barcoding techniques(Hao *et al*.,2012, Techen *et al*.,2014 ) to authenticate the candidate species, karyotypes determination through metaphase chromosomes(Hao *et al*.,2015), and flow cytometry and pulsed-field gel electrophoresis(Hao *et al*. 2015,Hao *et al*.,2011 ) to determine the ploidy level and genome size. While, these preliminary informations are known, attempts can be made for whole-genome sequencing of medicinal plants. In this regard, Chen *et al*.(2010,2011) initiated a project, "Herb Genome Programme" for the whole-genome sequencing of various medicinal plants and post genomic functional analysis of various secondary metabolite biosynthetic pathways, also very recent papers by Chakraborty(2018,2018) described how herbal genomics could be used for identification of various unknown pathways. Although whole-genome sequencing of medicinal plants hampers due to large genome size, polyploidy, duplication events, heterozygosity, and abundance of repetitive sequences, still, many medicinal plant genome have been sequenced successfully. The highly heterozygous genome of *ZiziphusJujuba* Mill having highly repetitive content was sequenced effectively by Liu *et al*.(2014). *Z. Jujuba* has significant medicinal value, and the plant contains various therapeutically important alkaloids, flavonoids and phenolics and has got bioactivities against cancer, ulcer, and various microbes (Yang *et al*.,2013a, Mahajan,2009 ). De novo assembly of its complex genome was made possible despite having highly repetitive contents. The fruits of this plant are highly rich in vitamin C and sugar. Combined genomics and transcriptomics data established that L-galactose pathway is the major synthesis pathway for vitamin C and consistently higher expression of the genes for the enzymes, GDP-D-mannose 3,5 epimerase, and GDP-L-galactose phosphorylase contributes for sugar metabolism(Li *et al*.,2014c). *AzadirachtaIndica* (Neem) is another important tree for its huge medicinal value and bioactivities against malaria, diabetes, tumor etc.(Biswas *et al*.,2002) have also been sequenced. Genome sequence of the plant revealed genome composition and predicted approximately about 20,000 genes (Krishnan,2012). Contrary to common genome evolution in plants, genome of *A.Indica* was found to be less complex in terms of smaller number of repetitve elements. Overall genomic data showed that *Z.Jujuba* got more complex genomes whereas the Neem genome exhibited simplicity.

Recently, in the context of more genomic references, for large no. of medicinal plants, e.g. for genomic references of 1000 medicinal plants, Illumina and IMPLAD (Institute for medicinal plant development) have signed a deal( Illumina news, March 14,2017 ) and hopefully this deal will give many sequenced chloroplast genome in future. However, recently, Yang He *et al*.(2016 ) reported the complete chloroplast genome sequences of the medicinal plant *Pogostemoncablin*. *P.cablin*, the natural source of patchouli alcohol, is an important herb in the Lamiaceae family (the mint family of flowering plants) is composed of more than 7000 species. It is an annual herb native to the Phillipines and has been widely cultivated in tropical and subtropical areas of Asia. Chemical and pharmacological studies of *P.cablin* indicates more than 40 major components, including monoterpenoids, and sesquiterpenoids, triterpenoids, and steroids, flavonoids and alkaloids and phenylpropanoid glycosides. In addition to its application in perfumes, soaps and cosmetic products, it also exerts wide range of medicinal effects including anti-inflammatory activity, inhibition of platelet aggregation, antidepressant and so on. Molecular sequences of the herb provide vast information not only about genes and its encoded proteins, but also functional implications and the evolutionary relationships. The development of next-generation sequencing technologies has allowed for the sequencing of entire chloroplast genomes. This genome, with 38.24% GC content, is 152,460 bp in length. Genome encodes 127 genes, of which 107 genes are single- copy, including 79 protein-coding genes, four rRNA genes, and 24 tRNA genes. Phylogenetic analysis reveals that *P.cablin* diverged from the *Sentellarioideae clade* about 29.45 million year ago. Complete sequences and annotation of *P.cablin* genome will facilitate phylogenetic, population and genetic engineering research investigations involving this particular species. *Salvia miltiorrhiza* is another important medicinal plant with great economic and medicinal value, whose complete chloroplast genome been sequenced by Qian J. *et al*. (2013) and draft sequence of whole genome by Xu *et al*.(2016). It is a significant traditional Chinese medicinal herb widely cultivated in China. The dried roots of this plant, commonly known as 'Chinese Sage' or 'red sage' in western countries are widely used in the treatment of several diseases including cardiovascular, cerebrovascular, and hyperlipidemia disease. More than 70 compounds have been isolated and structurally identified from the root of this plant e.g. hydrophilic phenolic acids including rosmarinic acids, salvianolic acids and lipophilic components diterpenoids and tanshinones. While several early acting CYPs(Cytochrome P450 mono-oxygenase) for tanshinone biosynthesis in *S.miltiorrhiza* have been identified (Gao *et al*.,2009, Guo *et al*.,2013), the majority of overall biosynthetic pathway, and the relevant regulatory components associated for tanshinone production remains unexplored. To identify these pathways and many others, draft sequence of the plant was made which shows that the plant genome size is very small(~600Mb),and it contains 30,478 protein- coding genes,and 1620 genes for transcription factors, and several of these transcription factors reavealed to be involved in the biosynthesis of tanshinone and phenolic acid. Xu *et al*. also identified 82 terpene synthase genes involved in hemi- mono- sesqui- and diterpene production, and 427 CYPs involved in the catalysis of various oxidation reaction. In addition to their small genomic size, *S.miltiorrhiza* have short life cycle and genetic transformability. These characteristics make this species an exemplary starting point to investigate the mechanism of medicinal plant secondary metabolism. Whereas, chloroplast genome of this plant is 151,328 bp in length and it contains 114 unique genes including 80-protein coding genes, 30-tRNA genes and four rRNA genes. The complete chloroplast genome sequences will definitely facilitate population, phylogenetic and genetic engineering studies of this medicinal plant.

Genome sequences of another medicinal plant, *Catharanthusroseus* helps in the genome guided investigation of hundreds of biologically active monoterpene-derived indole

alkaloid(MIA) metabolites. The plant is the sole source of the potent, expensive anti-cancer compound vinblastine/vincristine. Although ample transcriptomic and proteomic resources are now available for *C.roseus* (Champagne *et al*., 2012, Gongora-Castillo *et al*., 2012, Verma *et al*.2014), and the information has dramatically accelerated the discovery of MIA biosynthetic genes, a whole genome sequence will provide additional and important insights into the production, regulation and evolution of these valuable metabolites and hence Kellner *etal*.(2015 ) generated a genome assembly for *C.roseus* that provides a near comprehensive representation of the genic space that reveals the genomic context of key points within the MIA biosynthetic pathway including physically clustered genes, tandem gene duplication, expression subfunctionalization and putative neo-functionalization. The genome also revealed localization of enzyme-rich genic regions and transporters near known biosynthetic enzymes and highlighted how even a draft genome sequence could empower the study of high-value specialized metabolites.

*Artemisia annua* is an another important medicinal plant which produces artemisinin, an active ingredient in the most effective treatment for malaria. Artemisinin is a sesquiterpene, which in addition to their anti-malarial activity exhibits anti-cancer, antiviral and anti-inflammatory activity (Graham *et al*.,2010, Jiang *et al*.,2016, Hao *et al*.,2017). In the past decades, in the context of identification of many enzymes and intermediate compounds leading to artemisinin production, many genes encoding enzymes have been cloned and characterized (Bertea *etal*.2005, Zhang *etal*.2008, Mercke *etal*.2000, Teoh *etal*.,2006). However, little is known about regulatory aspects of sesquiterpene metabolism due to limited genomic information available and sequencing of limited number of randomly selected cDNA clones. As, whole genome or transcriptome sequencing enables functional genomic studies based on global gene expression, global transcriptomes of *A.annua* glandular trichome were characterized( Wang *et al*.,2009). It enabled putative function assignment to 28,573 unigenes, including previously undescribed enzymes involved in sesquiterpenebiosynthesis. Recently, considering the great medicinal value of artemisinin and its derivatives, and the unstable supply of the plant *A.annua*, a new synthetic biology approach was taken to transfer entire metabolic pathway of artemisinin from *A.annua* to chloroplast genome of tobacco plant , Nicotianatabacum(Fuentes *et al*.,2016).The construction vectors were designed in such a way that they contained all four genes required for the canonical artemisinic acid biosynthetic pathway in *A.Annua*. The work produced significant amount of artemisinicacid(120mg) per kg of biomass tobacco crop, and hopefully will meet the growing demand of artemisinin and access to the poorest people.

Genome sequencing of herb Tulsi, *Ocimum sanctum* L. family Lamiaceae is an important achievement by Indian researchers who with a view to understand the full metabolic potential of this plant whole nuclear and chloroplast genomes sequenced combining the sequence data from 4 libraries and three NGS platforms(Rastogi *et al*.,2015, ). The saturated draft assembly of the genome is ~386 Mb, along with the plastid genome of 142,245 bp, smallest in Lamiaceae. Pathway analysis indicated the abundance of phenylpropanoids/terpenoid pathway genes in *O.sanctum*. Previous transcriptome data analysis indicated several cytochrome P450s and transcription factor families important to characterize genes related to secondary metabolism and its regulation(Vinogradov,1999,Carovic-Stanko *et al*.,2010). Gene model prediction revealed the similarity of *O.sanctum* genome to Nicotianatabacum and Solanum lycopersicum, all sharing same sub-class. Comparison of the chemical compounds and genes availability in O.sanctum and S.miltiorrhiza indicated the potential for the discovery of new active molecules. Genome sequencing of another type of Tulsi, Krishna Tulsi, *Ocimuumtenuiflurum* with Illumina Hiseq 1000 showed assembled genome was about ~374Mb (Upadhyay,2015). Expression of anthocyanin biosynthesis related genes were observed to be relatively high explaining the purple coloration of leaves of Krishna Tulsi. The expression of six important genes identified from the genome data were validated by q-RT-PCR in different tissues of five different species showed high extent of urosolic-acid producing genes in young leaves of Rama, another subtype of Tulsi. In addition, presence of eugenol and urosolic acid in this plant implied as potential drugs in the cure of many diseases including cancer.

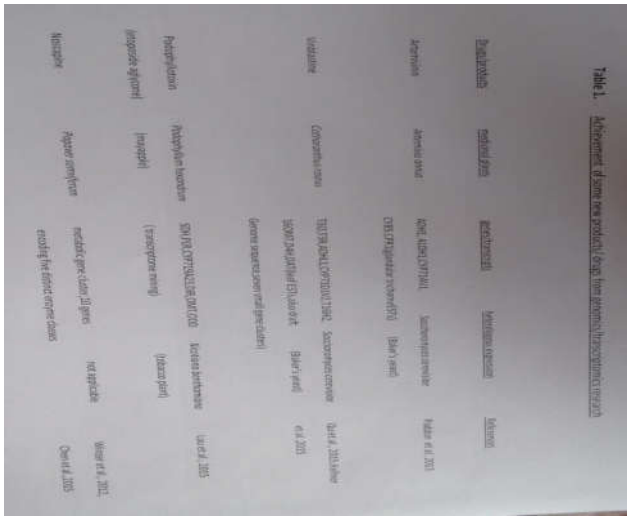
Finally, it is worth mentioning, that before starting genome sequencing of medicinal plant, genome authentication of the candidate species is very much important because this only will help in the isolation of pure and high molecular weight DNA. In the context of authentication of candidate species,the Canadian researchers have done significant amount of work(Hollingsworth *et al*.,2011, Ivanova *et al*.,2016). They used DNA barcoding technique and next-generation sequencing technologies for authentication of several herbal species. Though, establishing a standardised DNA barcoding system in plants is challenging, this technique, a shared community resource of DNA sequences have been used for organismal identification and taxonomic clarification. Ivanova *etal*. used next-generation sequencing(NGS) technology for taxonomic authentication of five medicinal plants, *Echinacea purpurea*, *Valerianaoffcinalis*, *Ginkobiloba*, *Hypericumperforatum*, and *Trigonellafoenum-graecum*. NGS revealed a diverse community of fungi, known to be associated with live plant material. As the efficacy of the drug decreases, if it is adulterated and sometimes lethal if the source is contaminated/substituted with toxic adulterants, recently, Techen *et al*.(2014) in their work emphasised on careful investigation of barcoding medicinal plants their substitutes and adulterants and also challenges on genomic regions selected to provide barcode of medicinal plants. They and others (Theodoridis *et al*.,2012, Schori *et al*.,2011) also elaborated that In the identification of species of medicinal plants, genomic regions plays a vital role and analyzed various regions. Collectively, though it is a general recommendation that the genomic regions *mat*K+*rbc*L for barcoding, often other genomic regions such as *psb*A-*trn*H, ITS(Internal Trascribed Spacer),ITS-2,could be more useful for medicinal material identification.

***Elucidation of new metabolic pathways/unknown enzymes through omics research***

The vast majority of the chemical/drug potential of the unexplored medicinal plant kingdom awaits discovery and is hidden away in the dark matter of plant genomes. Emerging herbal genomics research along with transcriptomics in combination with metabolomics/metabolites from different tissues will definitely aid in the identification of candidate genes for new enzymes/metabolic pathways for the biosynthesis of new natural products.

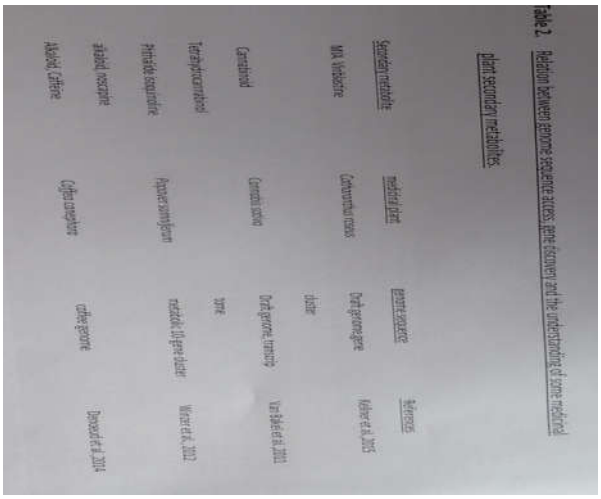
***Genomics of medicinal plants:*** Next-generation sequencing (NGS) technology revolutionized the study of genomics and molecular biology in all fields including genomes of medicinal plants. The reduced cost of this sequencing technology accelerating the genome research and consequently the ability to find enzymes and pathways for the biosynthesis of new products by identifying the underlying genes. As we discussed in the previous section, there are very few genomes (Neem and *Z.Jujuba*) of medicinal plants that have been fully sequenced. In spite of widespread applications of neem in agriculture and medicine, the molecular aspects of the biosynthesis of neem terpenoids remain largely unexplored. The genomes and trascriptomes analysis of neem shows that its genome is AT-rich, bears little repetitive DNA elements and comprises about 20,000 genes (Krishnan,2012). 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 analysis of neem also led to the identification of repeat elements, nucleotide composition and expression profiles of genes in various organs (Krishnan,2012, Shivaraj *et al*.,2015).

For plant genomes, there are now multiple examples in which the genes encoding certain natural product pathways have been found to be grouped together in biosynthetic gene clusters(table 1).

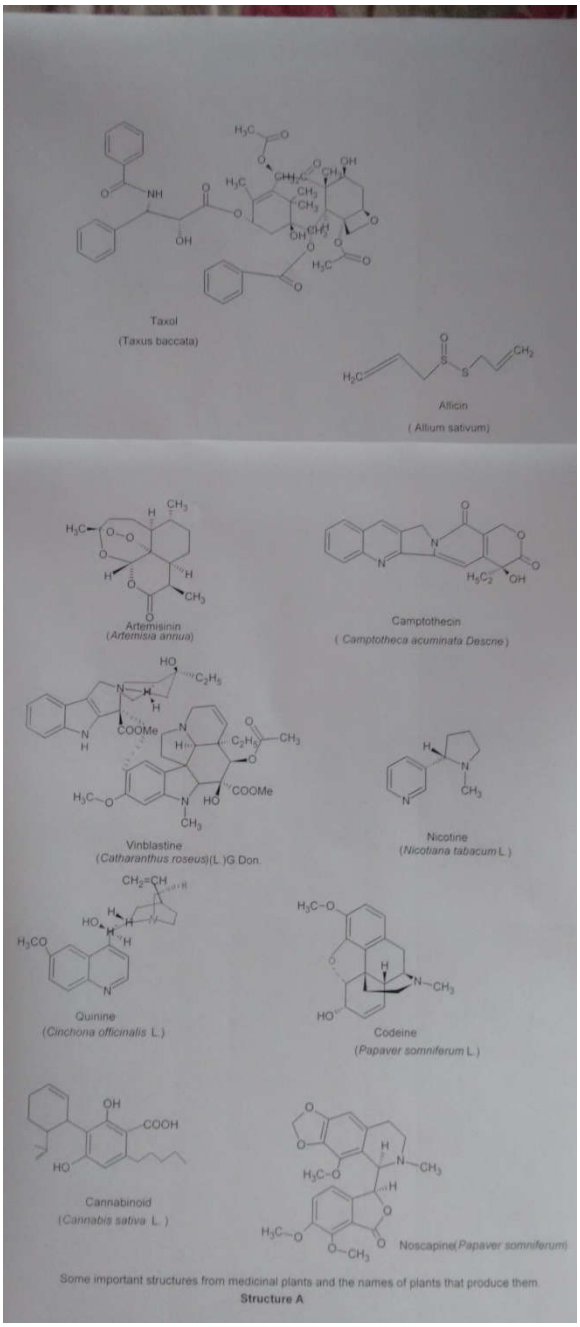


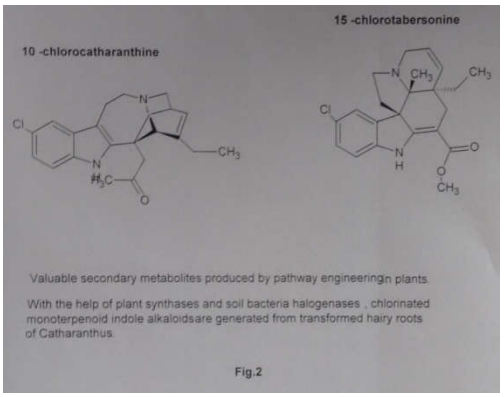
Analysis of a draft genome sequence of *Catharanthusroseus* provided evidence for partial clustering of genes for the biosynthesis of the monoterpene indole alkaloids (MIAs) vinblastine and vincistrine (Kellner *et al*., 2015). Vinblastine/vincristine pathway is a part of much larger and more complex biosynthetic MIA pathway network that gives rise to a wealth of other diverse products. With the help of bacterial artificial chromosome (BAC) sequencing, Kellner *et al*. showed seven small clusters each of two to three genes that contained genes encoding enzymes for vinblastine/vincristine biosynthesis pathway and other genes for other pathway. Whether these small clusters are dispersed throughout the genome of *C.roseus* or they form larger cluster, and how these genes are distributed relative to those required for the synthesis of other types of MIAs yet to know.

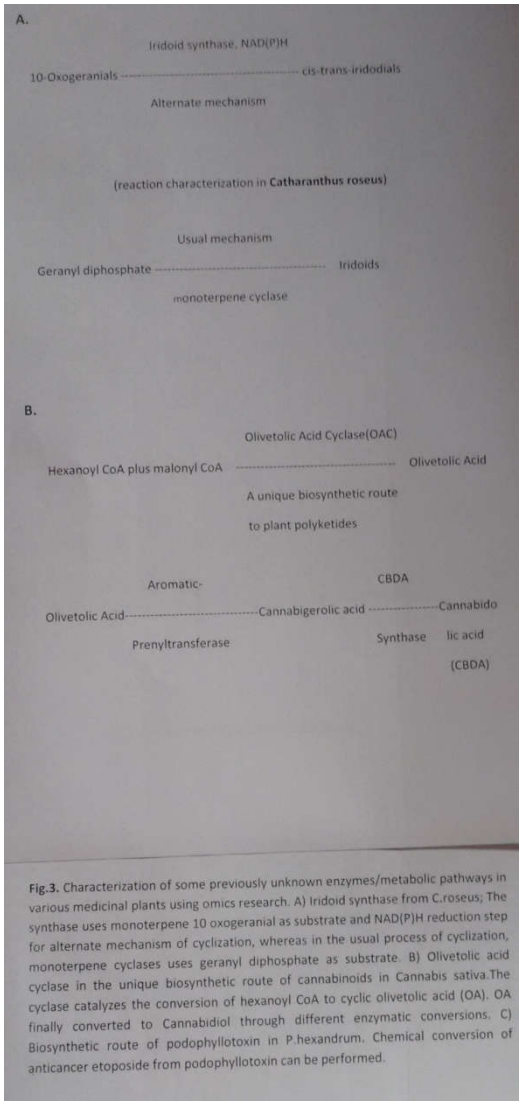
Genome sequence data of *Cannabis sativa* generated through NGS shed light on the psychoactive drug cannabinoid, biosynthetic pathway and also the understanding of chemodiversity of this medicinal plant (Van Bakel *etal*.,2011, Sharma *et al*., 2016). The cannabinoids, cannabidolicacids (CBDA) or tetrahydrocannabinolic acids(TCHA) are processed through same precursor cannabigerolic acid (CBGA) by two different enzymes THCA synthase or CBDA synthase (Taura *et al*.,2007). Marijuana and Hemp are the two chemotypes in cannabis based on the cannabinoids content and type(El-Alfy *et al*.,2010 ). The first draft genome of *cannabis* revealed genetic and epigenetic basis of major cannabinoids produced differentially in marijuana and hemp (Van Bakel *et al*. ,2011). Genome sequences of marijuana variety and hemp established high copy number of cannabinoids related genes as one of the reasons for their higher expression in female flowers of marijuana strains. Evolutionary processes and genomic interactions with environment sometimes may contribute in the abundance of particular bioactive molecules to one variety or gender of the plant species.



***Transcriptomics:*** A database and analytical platform for transcriptomic and metabolomic data for 14-medicinal plants are now available for hypothesis development of genes' function (Wurtele *et al*., 2012). From recently available transcriptomic data of *Catharanthusroseus*, a plant-derived iridoid biosynthetic pathway enzyme iridoid synthase was discovered (Geu-Flores *et al*., 2012 ). The iridoids comprise a large family of distinctive bicyclic monoterpenes that possess anticancer, anti-inflammatory and anti-bacterial activities (Dinda *etal*., 2009, 2011 Abdullah *etal*.,2017, Kirmizibekmez *et al*., 2017). During iridoid cyclization step, iridoid synthase uses linear monoterpene 10-oxogeranial as substrate which is in contrast to all monoterpene cyclases that uses geranyl diphosphate as substrate, and the enzyme synthase probably couples an initial NAD(P)H-dependent reduction via a Diels-Alder cycloaddition or a Michael addition( Fig.3A). As in the *C.roseus* transcriptome, hundreds of NAD(P)H-dependent enzymes are encoded, they applied a co-regulation criterion to reduce the number of candidates and observed two transcripts among the 20 best co-regulated transcripts coding for NADPH-using enzymes. The expression profile of iridoid synthase was similar to geraniol 10-hydroxylase, the closest characterized enzyme upstream of the cyclization step. This work not only suggest alternative biochemical mechanism for the biosynthesis of cyclic terpenes but with anticipation that this will enable the large scale heterologous production of iridoids in plants and microorganisms for agricultural(Dobler *et al*., 2011, Soe *et al*.,2004 )and pharmaceutical (Dinda *et al*.,2011,vanderHeijden *et al*.,2004) applications.




Transcriptomics of glandular trichomes from female cannabis (*Cannabis sativa*) flowers, identified olivetolic acid cyclase (OAC), a polyketide cyclase-like enzyme(Gagne *et al*.,2012 ). OAC transcripts were present at high levels in glandular trichomes, an expression profile that parallels other cannabinoid pathway enzymes. Olivetolic acid (OA) is proposed to be the first intermediate in the cannabinoid biosynthetic pathway and forms the polyketide nucleus of the cannabinoids(Fig.3B). During searching for polketide cyclase-like enzyme, that could assist in OA cyclization, the enzyme, OAcyclase was discovered. Identification of this enzyme may play an overlooked role in generating plant chemical diversity. Recently, using the transcriptome data from the plant mayapple (*Podophyllumhexandrum*) and selecting candidate genes to combinatorially express in tobacco (*Nicotianabethamiana*), six pathway enzymes of phodophyllotoxin pathway to etoposide aglycone were identified(Lau *et al*.,2015).

Podophyllotoxinis the natural product precursor of the chemotherapeutic 'unnatural' anticancer etoposide, however, till date only part of its biosynthetic pathway is known( Fig.3C). These works not only shows the expression of genes of etoposide precursor in a different plant species but also circumvent the need for cultivation of mayapple.

***Metabolomics:*** It is a key component for the analysis of phytochemicals/metabolites in plants including medicinal plants. Out of 200,000 to 1-million plant metabolites estimated, 50,000 are from medicinal and aromatic plants (Afendi,2012, Shivaraj *et al*.,2015). One of the major backbone for current metabolomics analysis is detection of metabolite peaks through mass spectra(MS) and nuclear magnetic resonance(NMR) spectroscopy. There are now

several databases for plant metabolites and their mass spectra available. These databases and tools have successfully been employed to evaluate the quality of herbal material and phytochemical(Vander *et al*.,2009),including determination of quantity of ginkgolic acids from Ginkgo leaves and in several commercial Ginkgo products(Choi *et al*.,2004). Recently, metabolic profiling of Angelica cutilobaKitag. roots has been carried out using gas chromatography-time-of-flight-mass spectrometry that enabled quantification of a number of metabolites in a tissue specific manner (Shivaraj *et al*., 2015). Scientists around the globe are now trying to make a correlations between gene- to- metabolite mostly by an integrated analysis of transcriptomes and metabolomes. As expression profiles of metabolites in different tissues/organelles, obviously depend upon the environmental conditions and stresses, metabolomic analysis under drought, cold, and other stresses, reveals the correlation of metabolites and genes responsible for the synthesis of particular sets of metabolites. For example, several inducible hemiterpenoid glycosides in leaves and scopolin and coniferin in roots were observed under nitrogen deprivation condition(Kusano *et al*.,2011, Ward *et al*.,2011). Combination of metabolic profiling and genetics study also help in identifying novel genes involved in the biosynthesis of bioactive specialized metabolites in *Arabidopsis*, major crops and in medicinal plants (Yonekura-Saka *et al*.,2012 ). For glucosinolates and flavonoids, metabolic quantitative loci analysis indicated possible association of genetic loci(quantitative trait loci,QTLs) with metabolic characters in *Arabidopsis thaliana*(Chan *et al*.,2011, Routaboul *etal*.,2012). Applying similar approaches in medicinal plants, two cytochrome P-450 genes for saponin biosynthesis in *Glycyrrhizauralensis*(Seki *et al*.,2011), and lysine decarboxylase gene for quinolizidine alkaloids biosynthesis in *Lupinusangustifolius* (Bunsupa *et al*.,2012 )were identified.

***Achievement of new products/drugs from omics research***

21st century omics technologies can advance the synthesis and production of natural products and hand over new drugs to the poor people. The perfect example is artemisinin, the antimalarial compound traditionally derived from *Artemisia annua*. Omics technologies are being employed for higher yield of artemisinin, through identification of key genes, characterization of transcription and profile expression and metabolite level(Covello,2008, Wang *et al*.,2009, Ma *et al*.,2009, Arsenault *etal*.,2010) and heterologous expression of all key genes of artemisinin biosynthetic pathway in *Nicotianatabacum* (Fuentes *et al*.,2016 ). The development of artemisinin and related antimalarial compound is a revolution in omics research and indicates the importance of traditional medicines in drug discovery (table2). In future, it can be expected for additional discoveries like isoquinoline N-oxide alkaloids as leads for new drug discovery (Dembitsky *et al*.,2015) and others of similar importance as out of 250,000-500,000 estimated plant species, only a fraction of it scientifically investigated for biological activity (Hostettmann *et al*.,2002, Lewis *etal*.,2003). It is also unfortunate, that still some plants well known sophisticated structures such as codeine, vinblastine, taxol and camptothecin remain well beyond the reach of commercially feasible total chemical syntheses, in spite of metabolic engineering research started since 1990 (Songstad *et al*.1990,1991, Kutchan,1995). In 2013, Paddon *et al*. through synthetic biology approach, transferred complete biosynthetic pathway of artemisinic acid, a precursor of artemisinin and developed strains of *Saccharomyces cerevisiae* (baker's yeast) for high-yielding biological production of artemisinic acid(Paddon *et al*.,2013).Furthermore, they have developed a practical, efficient chemical process for the conversion of artemisinic acid to artemisinin. In 2016, Fuentes *et al*. through a new synthetic biology approach, combinatorial super transformation of transplastomic recipient lines (COSTREL), introduced the complete pathway for artemisinic acid into high-biomass crop, tobacco plant(Fuentes *et al*.,2016). They isolated plants that produced more than 120mg of artemisinic acid per kilogram of biomass.

Recently, Qu *et al*. (Qu *et al*.,2015) engineered complete seven-gene vindoline pathway in yeast to produce vindoline, the anticancer drug precursor from tabersonine. The biosynthesis of vindoline in the medicinal plant *Catharanthusroseus* from monoterpenoid indole alkaloid (MIA) precursortabersonine is well understood at the molecular and biochemical levels. Although in high demand, the valuable anticancer drugs, vinblastine and vincristine only accumulate in trace amounts in *C.roseus* leaves. These anticancer molecules are condensed from MIA precursorscatharanthine and vindoline. The elucidation of the biosynthetic pathway of vindoline helped in engineering the pathways in microorganisms to allow industrial production of such a huge medicinally relevant compound.

## CONCLUDING REMARKS

Drug discovery/new molecules from medicinal plants depends upon the systematic research on information hidden in plant genomes. Till date, very few medicinal plant genomes have been sequenced and very little is known about the location of the genes encoding the specialized metabolic pathways in plant genomes. Until recently, only a handful of plant specialized metabolic pathways have been fully characterized in terms of both their biochemistry and genomic locations of the pathway genes. As more and more medicinal plant genome sequences become available, it will be possible to gain wider overview of the organization of the specialized metabolism in plants. Now emerging herbal genomics research and the technological advances in sequencing like inexpensive NGS revolution coupled with transcriptomics and metabolomics data will definitely set up a platform to find out about the organization of metabolic pathways in plant genomes and for accelerating the discovery and elucidation of new natural product pathways. Genetics-driven, trait-based approaches like QTL(quantitative trait loci) indicates association of genetic loci with metabolic characters, DNA-bar coding genomic regions such as rbcL+matK, ITS, ITS-2 could be useful for the identification of medicinal plant species and increased knowledge of plant metabolic gene cluster and their systematic analysis will enhance genome to natural product discovery pipeline. Last but by no means least, it can be said that as only a small fraction of the vast diversity of plant metabolism has been explored, for the speedy discovery of unknown metabolic pathways/enzymes, medicinal plant genomics could be used as tools for dissecting the pathways and future drug discovery.

**Conflict of Interests**

The author declared no conflict of interest with respect to the authorship and /or publication of this article.

**Acknowledgments**

The author gratefully acknowledges the overwhelming support from American Center Library, Kolkata, and National Library, Kolkata.

## References

Chang,C., Bowman,J.L., Meyerowitz,E.M., 2016. Field guide to plant model systems. *Cell*, 167, pp.325-329.

Luca,V.D., Salim,V., Masada,S., Yu,F.,2012. Mining the biodiversity of plants: a revolution in the making. *Science*, 336, pp. 1658-1661.

Modak,M., Dixit,P., Londhe,J., Ghaskadbi,S. Devasagayam,T.P.A.,2007. Indian herbs and herbal drugs used for the treatment of diabetes. *J. Clin.Biochem.Nutr.,* 40, pp. 163-173.

Brower,V.,2008. Back to Nature : extinction of medicinal plants threatens drug discovery. *J. Natl. CancerInstt*., 100 , pp. 838- 839.

Morey,M., Fernandez-marmiesse, A., Castineiras,D., *et al*., 2013. A glimpse into past, present and

future DNA sequencing. *Mol.GenetMetabol*., 110, pp. 3-24.

Metzker, M.L., 2010. Sequencing technologies- the next generation. *Nat.Rev.Genet*., 11, pp.31-46.

Hao, D.C., Xiao, P.G., Peng, Y., *et al*., 2012. Evaluation of the chloroplast barcoding markers by mean and smallest interspecific distances. *Pak.J.Bot*., 44, pp.1271-74.

Techen, N., Parveen,I., Pan, Z., *et al*., 2014. DNA barcoding of medicinal plant material for identification. *Curr.Opin.Biotech*., 25, pp. 103-110.

Hao,D.C., Xiao,P.G., 2015. Genomics and evolution in traditional medicinal plants: road to a healthier life. *Evol.Bioinform*., 11, pp. 197-212.

Hao,D.C., Vautrin,S., Song, C. *et al*., 2015. The first insight into the Salvia(Lamiaceae) genome via BAC library construction and high-throughput sequencing of target BAC clonones. *Pak.J.Bot*., 47, pp.1347-57.

Hao, D.C., Yang, L., Xiao, P., 2011.The first insight into the Taxus genome via fosmid library construction and end sequencing. *Mol.Genet.Genomics*, 285, pp. 197-205.

Chen,S., Sun,Y.Z., Xu, J., *et al*., 2010. Strategies of the study on herb genome program.Yao XueXueBao, 45, pp.807-812.

Chen,S., Li Xiang Xu, Guo,Q.L., 2011. An introduction to the medicinal plant genome project. *Front. Med*., 5, pp. 178-184.

Chakraborty, P., 2018. Herbal genomics as tools for dissecting new metabolic pathways of unexplored medicinal plants and drug discovery. *Biochimie Open*, 6, pp. 9-16.

Chakraborty, P., 2018. Search of new molecules/prospects of drug discovery from herbal medicines. *J. Complement Med. Alt. Healthcare*, 5(3).

Liu, M-J.,Zhao, J., Cai , Q-L . *et al*., 2014. The complex jujube genome provides insights into fruit tree biology. Nat.Commun.,doi : 10.1038/n comms 6315.

Yang, B., Yang, H., Chen, F. *et al*., 2013a. Phytochemical analyses of ZiziphusJujuba Mill. Var. spinosa seed by ultrahigh performance liquid chromatography-tandem mass spectrometry and gas chromatography-mass spectrometry. *Analyst*, 138, pp. 6881-6888.

Mahajan, RTCM., 2009. Phyto-pharmacology of ZiziphusJujuba Mill –a plant review. *Pharmacogn.Rev.,* 3, pp. 320-329.

Li,Y., Xu,C. ,Lin, X. , 2014c. De novo assembly and characterization of the fruit transcriptome of Chinese jujube(ZiziphusJujuba Mill) using 454 pyrosequencing and the development of novel trinucleotide SSR markers. PLoS One, doi: 10.1371/journal.Pone.0106438.

Biswas,K., Chattopadhyay,I., Banerjee, R.K. *et al*.,2002. Biological activities and medicinal properties of neem (Azadirachtaindica). *Curr.Sci.,* 82, pp. 1336-1345.

Krishnan, A., 2012.A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm Azadirachtaindica. *BMC Genomics*, 13, pp. 464-471.

Chen, S., 2017. Illumina, IMPLAD partner to sequence medicinal plants. Illumina News, March,14.

He, Y., Xiao,H., Deng, C., *et al*., 2016. The complete chloroplast genome sequences of the medicinal plant, Pogostemoncablin. *Int.J.Mol.Sci*., 17, pp. 820-830.

Qian, J., Song, J., Gao, H., *et al*., 2013. The complete chloroplast genome sequence of the medicinal plant, Salvia miltiorrhiza.PLoS One, 8, e57607.

Xu, H., Song, J., Luo,H., *et al*., 2016. Analysis of the genome sequence of the medicinal plant, Salvia miltiorrhiza. *Mol. Plant*, 9, pp. 949-952.

Gao, W., Hilwig, M.L., Huang, L. *et al*., 2009. A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *Org. Lett*., 11, pp.5170-5173.

Guo, J., Zhou, Y.J., Hilwig, M.L. *et al*., 2013. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proc.Natl.Acad.Sci*., USA, 110, pp. 12108-12113.

Champagne, A., Rischer, H., Oksman-Caldentey, K-M.*et al*., 2012. In-depth proteome mining of cultured Catharanthusroseus cells identifies candidate proteins involved in the synthesis and transport of secondary metabolites. Proteomics, 12, pp. 2536-2547.

Gangora-Castillo,E., Childs, K.L., Fedewa, G. *et al*., 2012. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species.PLoS One, 7, e 52506.

Verma, M., Ghangal, R., Sharma, R. *et al*., 2014.Transcriptome analysis of Catharanthusroseus for gene discovery and expression profiling. PLoS One, 9, e 103583.

Kellner, F., Kim, J., Clavijo-B, J. *et al*., 2015.Genome-guided investigation of plant natural product biosynthesis. *Plant J*., 82, pp. 680-692.

Graham, I.A., Besser, K., Blumer,J. *et al*., 2010. The genetic map of Artemisia annua L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science*, 327, pp. 328-331.

Jiang, W., Fu, X., Pan, Q. *et al*., 2016. Overexpression of AaWRKY1 leads to an enhanced content of artemisinin in Artemisia annua. Biomed. *Res. Int.,* 2016, doi.org/10.1155.

Hao, X.,Zhong, Y., Fu, X. *et al*., 2017. Transcriptome analysis of genes associated with the artemisinin biosynthesis by jasmonic acid treatment under the light in Artemisia annua. *Front. Plant Sci.* doi.org/10.3389.

Bertea, C,M., Freije, J.R., Woude, H. *et al*., 2005. Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in Artemisia annua. *Planta Medica*, 71, pp. 40-47.

Zhang, Y., Teoh, K.H., Reed, D.W., *et al*., 2008.The molecular cloning of artemisinic aldehyde Delta 11(13) reductase and its role in glandular trichome dependent biosynthesis of artemisinin in Artemisia annua. *J.Biol.Chem*., 283, pp. 21501-21508.

Mercke,P., Bengtsson, M , Bouwmeester, H.J., *et al*., 2000. Molecular cloning expression and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinic biosynthesis in Artemisia annua L., Archives Biochem. *Biophys*. 381, pp. 173-180.

Teoh, K.H., Polichuk, D.R., Reed,D.W., *et al*., 2006. Artemisia annua L.(Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett.,* 580, pp. 1411-1416.

Wang,W., Wang, Y., Zhang, Q., *et al*., 2009. Global characterization of Artemisia annua glandular trichometranscriptome using 454 pyrosequencing. BMC genomics, 10, pp. 465 .

Fuentes, P., Zhou, F., Erbon, A., *et al*., 2016. A new synthetic biology approaches allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. e Life, 5, e 13664.

Rastogi,S.,Kalra,A.,Gupta,V.,*etal*.2015. Unravelling the genome of Holy basil : an "incomparable"elixir of life" of traditional Indian medicine. BMC Genomics, 16 , pp. 413-431.

Vinogradov, A. E., 1999. Intron-genome size relationship on a large evolutionary scale. J MolEvol., 49, pp. 376-384.

Carovic-Stanko, K., Liber, Z., Besendorfer,V., *et al*. 2010. Genetic relations among basil taxa (Ocimum L.) based on mlecular markers, nuclear DNA content, and chromosome number. *Plant SystEvol*., 285, pp. 13-22.

Upadhyay, A.K., Chacko, A.R., Gandhimathi, A., *et al*., 2015. Genome sequencing of herb Tulsi (Ocimumtenuiflorum) unravels key genes behind its strong medicinal properties. *BMC Plant Biol*., 15, pp. 212.

Hollingsworth,P.M., Graham,S.W., Damon,P., *et al*., 2011. Choosing and using a plant DNA barcode. PLoS One, 6, e 19254.

Ivanova,N.V., Kuzmina,L.M., Thomas,W.A., *et al*., 2016. Authentication of herbal supplements using next-generation sequencing. PLoS One, 11, e 0156426.

Theodoridis, S., Stefanaki, A., Tezcan, M., *et al*., 2012. DNA barcoding in native plants of the Labiatae(L. amiaceae) family from Chios island(Greece) and the adjacent Cesme-karaburum Peninsula(Turkey). MolEcolResour, 12, pp. 620-633.

Schori, M., Showalter, A.M., 2011. DNA barcoding as a means for identifying medicinal plants of Pakistan. *Pak J Bot*. , 43(S1), pp. 1-4.

Shivaraj, Y., Govind, S., Jogaiah,S., *et al*., 2015. Functional analysis of medicinal plants using system biology approaches. *Int. J. Pharm. Pharmaceu Sci*., 7(S1), pp. 41-43.

Van Bakel, H., Stout, J.M., Cote, A.G., *et al*., 2011. The draft genome and transcriptome of Cannabis sativa. *Genome Biol*. ,12-10-r102.

Sharma, S., Shrivastava, N., 2016. Renaissance in phytomedicines : promising implications of NGS technologies. *Planta* , 244, pp. 19-38.

Taura, F., Srikantaramas, S., Shoyama, Y., *et al*., 2007. Cannabidolic acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. *FEBS Lett.,* 581, pp. 2929-2934.

El-Alfy, A.T., Ivey, K., Robinson, K., *et al*., 2010.Anti depressant-like effect of delta 9-tetrahydrocannabinol and other cannabinoids isolated from Cannabis sativa L., *Pharmacol. Biochem.Behav*., 95, pp. 434-442.

Wurtele, E., Chappell, J., Jones, A., *et al*., 2012. Medicinal plant: a public resource for metabolomics and hypothesis development. *Metabolites*, 2, pp. 1031-1059.

Geu-Flores, F.,Sherden,N.H., Courdavault,V., *et al*.,2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, 492, pp. 138-142.

Dinda,B., Chowdhury,D.R., Mohanta, *et al*., 2009. Naturally occurring iridoids, secoiridoids and their bioactivity. *Chem. Pharm. Bull*. (Tokyo), 57, pp. 765-796.

Dinda,B., Debnath,S., Banik,R., 2011. Naturally occurring iridoids and secoiridoids. *Chem. Pharm. Bull*. (Tokyo), 59, pp. 803-833.

Abdullah,F.O., Hussain, F.H.S., Clericuzio,M., *et al*., 2017. A new iridoid dimer and other constituents from the traditional Kurdish plant Pterocephalusnestorianus. *Chem. Biodiversity*, 14, e 1600281.

Kirmizibekmez, H., Tiftik,K., Klisz,N., *et al*., 2017. Three new iridoid glycosides from the aerial parts of Asperulainvolucrata. *Chem.Biodiversity*, 14, e 1600288.

Dobler,S., Petschenka,G., Pankoke, H., 2011. Coping with toxic plant compounds-the insect's perspectives on iridoid glycosides and cardenolides. *Phytochemistry*, 72, pp. 1593-1604.

Soe, A.R.B., Bartram, S., Gatto, N., et. al., 2004. Are iridoids in leaf beetle larvae synthesized de novo or derived from plant precursors ?A methodological approach. Isotopes *Environ.Health Stud*., 40, pp. 175-180.

vanderHeijden,R., Jacobs, D.I., Snoeijer,W., *et al*., 2004. The Catharanthus alkaloids: pharmacognosy and biotechnology. *Curr.Med.Chem*., 11, pp. 607-628.

Gagne,S. J., Stout,J.M., Liu,E., 2012. Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. *Proc. Natl. Acad. Sci*., USA, 109, pp. 12811-12816.

Lau,W., Sattely,E.S., 2015. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science*, 349,, pp. 1224-1228.

Afendi, F.M., Okada, T., Yamazaki, M., *et al*., 2012.KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant cell physiol*., 53, e1.

Van der, K.F., Maltese, F., Choi, Y.H., *et al*., 2009. Quality control of herbal material and phytopharmaceuticals with MS and NMR based metabolic finger printing. *Planta Med.,* 75, pp. 763-775.

Choi, Y.H., Choi, H.K., Peltenburghooman, A.M., *et al*., 2004. Quantitative analysis of ginkgolic acids from Ginkgo leaves and products using $^{1}$*H-NMR. Phytochem. Anal*., 5, pp. 325-330.

Kusano, M.,Fukushima, A., Redestig, H., *et al*., 2011. Metabolomic approaches toward understanding nitrogen metabolism in plants. *J. Exp. Bot*., 62, pp. 1439-1453.

Ward, J.L., Baker, J.M., Llewellya, A.M., *et al*., 2011. Metabolomic analysis of Arabidopsis reveals

hemiterpenoid glycosides as products of a nitrate ion-regulated, carbon flux overflow. *Proc. Natl. Acad. Sci.,* USA, 108, pp. 10762-10767.

Yonekura-Sakakibara, K., Fukushima, A., Saito, K., 2012. Transcriptome data modeling for targeted plant metabolic engineering. *Curr. Opin. Biotechnol*., http://dx.doi.org/10.1016/j.copbio 2012.10.018.

Chan, E.K.F., Rowe, H.C., Corwin, J.A., *et al*., 2011.Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thalina. PLoS Biol., 9, e 1001125.

Routaboul, J.M., Dubos, C., Beck, G., *et al*., 2012.Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis. *J. Exp. Biol.*, 63, pp. 3749-3764.

Seki, H., Sawaj, S., Ohyama, K., *et al*., 2011. Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant cell,* 23, pp. 4112-4123.

Bunsupa, S., Katayama, K., Ikeura, E., *et al*., 2012. Lysine decarboxylase catalyses the first step of quinolizidine alkaloid biosynthesis and coevolved with alkaloid production in Leguminosae. *Plant cell,* 24, pp. 1202-1216.

Covello, P.S., 2008. Making artemisinin. *Phytochem*., 69, pp.2881-2885.

Ma, D., Pu, G., Lei, C., *et al*., 2009. Isolation and characterization of AaWRkY1, an Artemisia annua transcription factor that regulates the amorpha-4,11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant cell physiol*., 50, pp. 2146-2161.

Arsenault, P.R., Vail, D., Wobbe, K.K., *et al*., 2010. Reproductive development modulates gene expression and metabolite levels with possible feedback inhibition of artemisinin in Artemisia annua. *Plant Physiol*., 154, pp. 958-968.

Dembitsky, V.M., Gloriozova, T.A., Poroikov, V.V., 2015. Naturally occurring plant isoquinoline N-oxide alkaloids: their pharmacological and SAR activities. *Phytomed*., 22, pp. 183-202.

Hostettmann, K., Marston, A., 2002. Twenty years of research into medicinal plants: Results and perspectives. *Phytochem.Rev.,* 1, pp. 275-285.

Lewis, W.H., Elvin-Lewis, MPF, 2003. Medical Botany: Plants affecting human health. *J. Wiley*.

Songstad, D.D., De Luca, V., Brisson, N., *et al*., 1990. High levels of tryptamine accumulation in transgenic tobacco expressing tryptophan decarboxylase. *Plant Physiol*. 94, pp. 1410-1413.

Songstad, D.D, Kurz, W.G.W., Nessler,C.L., 1991. Tyramine accumulation in Nicotianatabacum transformed with a chimeric tryptophan decarboxylase gene. Phytochem., 30, pp. 3245-3246. Kutchan, T.M., 1995. Alkaloid Biosynthesis- The basis for metabolic engineering of medicinal plants. *The Plant cell*, 7, pp. 1059-1070.

Paddon, C.J., Westfall, P.J., Pitera, D.J., *et al*., 2013.High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496, pp. 528-532.

Qu, Y, Easson, M.L.A, Froese, J., *et al*., 2015.Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl. Acad. Sci.,* USA, 112, pp.6224-6229.

**How to cite this article:**

Prasanta Chakraborty (2018) 'Medicinal Plant Genome: A Source Of Finding New Enzymes, Metabolic Pathways And Drug Discovery', *International Journal of Current Medical And Pharmaceutical Research,* 04(6), pp. 3318-3326.

*******