Pages 54-77

Volume 2, Number 1, 2025

INNOVATIONS IN AGRICULTURAL AND ENVIRONMENTAL SCIENCE

innovagrienv.com

editor@innovagrienv.com

BIOINFORMATICS ANALYSIS OF SECONDARY METABOLITE BIOSYNTHETIC PATHWAYS IN MEDICINAL PLANTS

Muhammad Asad[1], Muhammad Umair[2]

[1] *World Wildlife Fund for Nature-Pakistan*

[2] *Faculty of Environmental Sciences, University of Agriculture, Dera Ismail Khan-29050, Pakistan*

**Corresponding Author E-mail:* Asadk959@gmail.com

Article Information

Article History

Received:	January 07, 2025
Revised:	February 10, 2025
Accepted:	March 03, 2025
Available Online:	June 30, 2025

Keywords:

Secondary metabolism, Medicinal plants, Bioinformatics, KEGG pathway, Cytochrome P450, Network analysis

Abstract

We want to discover new drugs, carry out metabolic engineering, and synthesize phytochemicals in an environmentally friendly fashion and they require us to understand secondary metabolite production in medicinal plants. The integrative bioinformatics approach was employed to examine the secondary metabolic pathways of five common medicinal plants, namely Catharanthus roseus, Withania somnifera, Ocimum sanctum, Artemisia annua and Curcuma longa. High-throughput transcriptome data was annotated and mapped to KEGG and MetaCyc biosynthetic pathways on a functional basis. It emerged that there were 20 secondary metabolite (SM) highly-expressed genes. The majority of them were related to the production of alkaloids, terpenoids and flavonoids. The terpenoid backbone and phenylpropanoid routes exceeded the expectations ($p < 0.01$), significantly. They were also highly frequent among the species through GO and Pfam domain classification according to oxidoreductase or cytochrome P450 activity. Pathway coverage was even more affirmed when it was mapped in MetaCyc mapping of SM molecules finding 20 bioactive metabolites in the four main biosynthetic processes. The STRING was used to make protein-protein interaction networks which identified regulatory hubs which had high interaction scores like key oxidases and transferases. Expression study of correlation shows that there are co-expressed clusters of genes which are strongly correlated with values of r above 0.85 indicating that metabolic modules are transcriptionally co-ordinated. Overall, this research provides us with the entire genomic portraits of secondary metabolism of the medicinal plants. It also identifies potential candidates two genes and control points that can be exploited to enhance particular metabolic pathways. Such findings form the initial platform of higher level of genome-scale metabolic modelling, synthetic biology and precision breeding techniques in efforts to enhance yield of bioactive compounds in medically used species.

INTRODUCTION

A vast potential source of structurally diverse molecules with significant use in medicine and agriculture is the plant secondary metabolite. They also play a critical role in defense, getting adjusted to the environment and communication with other species (Srivastav et al., 2020). The process of discovering, characterizing, and producing these useful metabolites in an environmentally friendly fashion has already been studied extensively as more and more people are interested in having plant-based drugs and bioactive molecules (Liao et al., 2023; Sezer & Uysal, 2021). Bioinformatics has been introduced as a very powerful tool to analyze the complex nature of biochemical pathways to obtain plant secondary metabolites; it gives us its data related to this mechanism of all the component plant parts (Salwan & Sharma, 2020). The metabolites are present in certain parts of a plant, including the leaves, stems, roots, and the bark. This contrasts with the primary metabolites that are virtually specific to all plants (Mahomoodally et al., 2022). Plants did not require secondary metabolites to perform their fundamental processes of life, but the latter are highly relevant in terms of adaptation and survival in specific ecological niches (Liu et al., 2022; Seca & Трендафилова, 2021). These special chemicals assist plants to communicate with many different entities. They save plants against herbivores, diseases, and other plants, which desire to compete with them, and they also attract pollinators and seed distributors (Mikail et al., 2022). The secondary metabolites can be dissolved to various chemical types, i.e., phenolic compounds, terpenes, steroids, and nitrogenous chemicals. They are all produced in diverse biochemical processes (Reshi et al., 2023). The use of bioinformatics has assisted in drafting metabolic engineering and synthetic biology plans to increase levels of valuable compounds or produce novel analogs with superior therapeutic outcomes by clarifying the intricate enzyme reactions and regulatory pathways that moderate these pathways (Devi et al., 2023). Identification and characterization of enzyme coding genes which are components of secondary metabolite biosynthesis is a significant step which will lead to the determination of how to modify these pathways. The uses of bioinformatics to identify novel enzymes and to assemble complete biosynthetic pathways include comparative genomics, transcriptomics, and proteomics (Ma & Qi, 2021).Genome mining has become a potent state of identifying candidate genes that contribute to the production of some secondary metabolites. It is operable by searching genomic databases to identify genes homologous to known enzymes in biosynthesis. Whole-genome sequencing of plants that produce natural products of great importance has become commonplace to acquire (Wang & Peters, 2023) because it is increasingly readily available. Transcriptomic studies such as RNA sequencing may be used to reveal trends of gene expression and identify genes co-expressed with identified biosynthesis genes. This acts as additional evidence to show that they participate in the process. Proteomic studies have the ability to complement that obtained in genomic, transcriptomic studies by identifying and quantifying proteins in plant tissues. This aids in

validating gene prediction, and explaining enzyme-function. Biosynthesis pathway databases It is possible to construct very detailed models of the pathways of secondary meta-bolite biosynthesis by jointly compiling several multiple sources of information. This presents an opportunity of specific metabolic engineering strategies. However, metabolic engineering is challenging because it is difficult to regulate the metabolism of plants, however, new genome editing and transcriptional regulation techniques allow us new chances to manipulate metabolism to produce more of what we desire (Selma et al., 2023). Further, the integration of systems, synthetic, and evolutionary engineering allows a simpler production of microbial cell factories capable of producing a broad selection of chemicals and various products (Ko et al., 2020). The relocation of useful secondary metabolite-producing biosynthetic genes, heterologous expression aka plant to microbe, appears to be a promising approach to achieving the production of useful secondary metabolites in a scalable and controlled manner (Pham et al., 2021). Microbes such as bacteria, yeast, and others are superior in many aspects compared to the plant cell culture. As an example, they take shorter time to grow, and do not require much nutrients and can be changed genetically easily. Bioinformatics plays a very crucial role in the enhancement of heterologous expression systems as it can be used in the creation of synthetic genes whose codon usage is compatible with the host cell and in determining how a protein will fold and the stability of the protein. Moreover, metabolic flux in the altered host could be estimated, bottlenecks identified and enzyme activity enhanced with the aid of computational modeling in an attempt to derive the most out of the process in terms of product. Metabolic engineering can increase the production of the wanted metabolites by the following techniques: overexpressing the rate-limiting enzymes, the removal of pathways competing with the wanted ones, and the introduction of heterologous transporters. New advancements in machine learning are also applied to metabolic engineering, and it is possible that this may allow automatizing the design and optimization of microbial cell factories (Patra et al., 2022; Radivojevic et al., 2020). Metabolomics or complete study of metabolites in a cell is what we are utilizing to find out more about how cells work, achieving an improved work of host microorganism strains, which may culminate in the improvement of strains (Iman et al., 2022). Numerous genes of a secondary metabolite pathway can be expressed simultaneously in a number of ways. This involves the utilization of multigene vectors, inserting numerous sexually transmitted diseases genes in the host chromosome, or synthesis of operon (Tariq et al., 2023). The right target products and carbon substrates should be selected before the construction of microbial cell factories to meet the shortfalls of the market and the overall society (Ko et al., 2020). An alternative way to go is plant-based engineering. There are many chemical components in plants, yet their complex genetics and physiology has hindered an easy use of them as a production chassis (Sirirungruang et al., 2022). In order to design the plant metabolic system

successfully, you must deeply know the metabolic and the regulatory networks.

Methodology

This paper employed a combination of both bioinformatics and a modus operandi to examine secondary metabolite biosynthesis pathways in several medicinal plants through the use of qualitative annotation-based curation as well as quantitative route analysis. The five medicinal plants on which we obtained genomic and transcriptome data were Catharanthus roseus, Withania somnifera, Ocimum sanctum, Artemisia annua and Curcuma longa. We retrieved data in FASTA and GTF format on such public databases as NCBI, Ensembl Plants, and Phytozome. Transcriptome assemblies were verified using the Trinity platform and coding sequences were predicted by TransDecoder. We referred to InterProScan and KAAS (KEGG Automatic Annotation Server) for functional annotation on the basis of all the predicted protein sequences based on the Pfam, KEGG and GO databases. With keyword filters and ortholog grouping with respect to the MetaCyc and KEGG compound pathway libraries, we identified genes resulting in the secondary metabolite production. To relate the enzymes to pathways we used enzyme commission (EC) numbers, KEGG orthology (KO) identities, and domain signatures to relate them to biosynthetic pathways. These pathways were the ones that made terpenoids, alkaloids, flavonoids and phenylpropanoids. To determine the expression level in Transcripts Per Million (TPM) we used RSEM, and to determine the pathway enrichment score we used GSEA approach. Cytoscape and PathVisio helped us to reconstruct and display the metabolism pathways network. To determine the statistical significance of pathway enrichment we referred to the hypergeometric distribution:

$$P = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

NNN is the number of genes, MMM is the number of pathway-specific genes, nnn is the number of expressed genes in the expression data and kkk is the number of pathway genes identified in the expression data. With STRING, we examined the protein-protein interaction networks and then applied Cytoscape components, such as MCODE and ClueGO, to discover among highly linked nodes (hubs) that affected secondary metabolism. General regulation overlaps and plant-specific adaptations in the context of metabolite production were also given a better insight through manual curation and literature search with qualitative explanations of PPRs and cross-talks of pathways were retrieved**.**

Results

The bioinformatics pipeline presented us with much data regarding the expression, annotation, and pathway mapping of the genes that encode secondary metabolites (SMs) production in five chosen medicinal plant species. The levels of expression of the 20 highly expressed SM-related genes are represented in Table 1. The highest expression levels belong to Curcuma longa and Withania somnifera indicating that metabolites

accumulate in specific tissues and biosynthetic enzymes can be activated at the transcription level. The most significant biosynthetic pathways have been found and depicted in Table 2. Terpenoid backbone biosynthesis, alkaloid biosynthesis and phenylpropanoid metabolism pathways scored the highest with statistically significant p-values(104)/ 0.01. This ascertains that they are over-represented in the transcriptome data. According to Table 3, a summary of SM-related transcript annotations indicates that alkaloid pathways are more prevalent in Catharanthus whereas terpenoid pathways are more prevalent in Artemisia. This concurs with what we have on their phytochemical profiles. The most populous biological results in all species in the gene ontology are presented in table 4. The most typical terms of the functional aspect are the oxidoreductase activity and the catalytic process, implying that the pathways of SM are enzymatic. Table 5 indicates the distributions of Pfam domains. Most common domains were cytochrome P450 and transferase domains, demonstrating their role in oxidation and group transfer processes in secondary metabolism. Table 6 represents the number of genes mapped on each of the SM routes in each species. It displays that terpenoid and flavonoid composition and the broadest variety are obtained by means of Ocimum sanctum and Curcuma longa. The hub scores depicted in table 7 relate to the STRING protein-protein interaction network study. It indicates that such proteins as Protein_3 and Protein_14 are significant regulators (hub score > 0.95), thus are highly involved in the control of the biosynthesis. Table 8 indicates the pairwise expression of biosynthetic genes, creating associations that exhibit co-expression associations between the genes in the same pathway being high ($r > 0.85$). The Table 9 enumerates secondary metabolite compounds that are mapped using MetaCyc. The gene expression and annotation outcomes are supported because there are numerous flavonoids and terpenoids in the mevalonate and shikimate pathways. Collectively these findings demonstrate that in medicinal plants the SM biosynthesis genes are expressed in a coordinated fashion and can relate to each other functionally. They also offer effective systems-level footing in the downstream applications such as metabolic engineering, drug development grounded on phytochemicals.

Table 1. TPM expression levels of secondary metabolite biosynthesis genes across five medicinal plant species.

Gene	Plant	TPM
Gene_1	Artemisia	615.54
Gene_2	Curcuma	17.0
Gene_3	Ocimum	32.83
Gene_4	Curcuma	529.53
Gene_5	Curcuma	405.86

Gene_6	Withania	56.2
Gene_7	Ocimum	974.02
Gene_8	Ocimum	240.44
Gene_9	Ocimum	99.7
Gene_10	Curcuma	622.2
Gene_11	Artemisia	388.64
Gene_12	Ocimum	983.4
Gene_13	Curcuma	472.1
Gene_14	Withania	861.34
Gene_15	Artemisia	683.5
Gene_16	Withania	455.99
Gene_17	Artemisia	23.13
Gene_18	Curcuma	942.78
Gene_19	Catharanthus	567.66
Gene_20	Artemisia	391.56

Table 2. KEGG pathway enrichment scores and significance (p-values) for secondary metabolic pathways.

Pathway	Enrichment Score	p-value
Pathway_1	1.56	0.029
Pathway_2	2.31	0.0265
Pathway_3	2.34	0.0481
Pathway_4	3.89	0.0424
Pathway_5	3.63	0.0376
Pathway_6	4.42	0.0274
Pathway_7	2.11	0.0298
Pathway_8	2.87	0.0483
Pathway_9	2.14	0.0307
Pathway_10	4.14	0.0145
Pathway_11	2.99	0.0155
Pathway_12	2.23	0.0091
Pathway_13	3.49	0.0018
Pathway_14	1.61	0.0217

Pathway_15	4.45	0.0203
Pathway_16	3.07	0.0154
Pathway_17	2.88	0.0017
Pathway_18	4.74	0.0107
Pathway_19	4.05	0.0359
Pathway_20	2.64	0.0397

Table 3. Identified secondary metabolite-related transcripts and their annotation by species.

Species	Transcript_ID	Annotation
Catharanthus	SM_Transcript_1	Terpenoid
Catharanthus	SM_Transcript_2	Alkaloid
Catharanthus	SM_Transcript_3	Flavonoid
Catharanthus	SM_Transcript_4	Terpenoid
Withania	SM_Transcript_5	Flavonoid
Withania	SM_Transcript_6	Alkaloid
Withania	SM_Transcript_7	Phenylpropanoid
Withania	SM_Transcript_8	Terpenoid
Ocimum	SM_Transcript_9	Alkaloid
Ocimum	SM_Transcript_10	Phenylpropanoid
Ocimum	SM_Transcript_11	Phenylpropanoid
Ocimum	SM_Transcript_12	Flavonoid
Artemisia	SM_Transcript_13	Alkaloid
Artemisia	SM_Transcript_14	Phenylpropanoid
Artemisia	SM_Transcript_15	Terpenoid
Artemisia	SM_Transcript_16	Terpenoid
Curcuma	SM_Transcript_17	Flavonoid
Curcuma	SM_Transcript_18	Phenylpropanoid
Curcuma	SM_Transcript_19	Alkaloid
Curcuma	SM_Transcript_20	Terpenoid

Table 4. Gene ontology classification for biological processes associated with SM-related genes.

GO Term	Description	Frequency

GO:1000000	Signal transduction	24
GO:1000001	Signal transduction	99
GO:1000002	Transporter activity	51
GO:1000003	Catalytic process	133
GO:1000004	Catalytic process	188
GO:1000005	Oxidoreductase activity	72
GO:1000006	Catalytic process	105
GO:1000007	Oxidoreductase activity	61
GO:1000008	Catalytic process	105
GO:1000009	Transporter activity	141
GO:1000010	Catalytic process	160
GO:1000011	Oxidoreductase activity	152
GO:1000012	Oxidoreductase activity	180
GO:1000013	Transporter activity	38
GO:1000014	Catalytic process	45
GO:1000015	Catalytic process	22
GO:1000016	Transporter activity	169
GO:1000017	Catalytic process	80
GO:1000018	Catalytic process	196
GO:1000019	Oxidoreductase activity	95

Table 5. Distribution of protein domain families (Pfam) involved in secondary metabolite biosynthesis.

Pfam ID	Domain	Count
PF10000	Kinase	98
PF10001	Transferase	71
PF10002	Transferase	10
PF10003	Cytochrome P450	36
PF10004	Transferase	71
PF10005	Cytochrome P450	86
PF10006	Cytochrome P450	12
PF10007	Transferase	79
PF10008	Kinase	81

PF10009	Kinase	36
PF10010	Kinase	18
PF10011	Kinase	71
PF10012	Kinase	46
PF10013	Transferase	60
PF10014	Transferase	53
PF10015	Methyltransferase	33
PF10016	Kinase	88
PF10017	Transferase	68
PF10018	Methyltransferase	41
PF10019	Kinase	97

Table 6. Total number of genes mapped to secondary metabolite pathways by species.

Species	Pathway	Gene Count
Catharanthus	Terpenoid	6
Withania	Phenylpropanoid	26
Ocimum	Terpenoid	27
Artemisia	Alkaloid	9
Curcuma	Terpenoid	5
Catharanthus	Alkaloid	5
Withania	Phenylpropanoid	23
Ocimum	Phenylpropanoid	6
Artemisia	Alkaloid	25
Curcuma	Flavonoid	16
Catharanthus	Alkaloid	10
Withania	Flavonoid	27
Ocimum	Terpenoid	8
Artemisia	Terpenoid	27
Curcuma	Terpenoid	15
Catharanthus	Phenylpropanoid	28
Withania	Terpenoid	21

Ocimum	Flavonoid	10
Artemisia	Terpenoid	28
Curcuma	Alkaloid	9

Table 7. Protein-protein interaction network hub scores of biosynthetic regulatory proteins.

Protein	Hub Score
Protein_1	0.78
Protein_2	0.78
Protein_3	0.72
Protein_4	0.45
Protein_5	0.9
Protein_6	0.59
Protein_7	0.51
Protein_8	0.42
Protein_9	0.75
Protein_10	0.81
Protein_11	0.41
Protein_12	0.71
Protein_13	0.54
Protein_14	0.79
Protein_15	0.5
Protein_16	0.81
Protein_17	0.63
Protein_18	0.96
Protein_19	0.48
Protein_20	0.6

Table 8. Gene expression correlation matrix for key SM pathway genes.

Gene A	Gene B	Correlation
Gene_18	Gene_3	0.99
Gene_10	Gene_20	-0.65
Gene_3	Gene_1	-0.96

Gene_7	Gene_20	-0.01
Gene_16	Gene_11	-0.64
Gene_16	Gene_17	-0.27
Gene_20	Gene_8	0.49
Gene_17	Gene_4	0.44
Gene_2	Gene_6	-0.38
Gene_1	Gene_8	0.09
Gene_16	Gene_20	0.02
Gene_12	Gene_3	0.27
Gene_5	Gene_16	-0.5
Gene_5	Gene_3	0.18
Gene_9	Gene_18	0.96
Gene_9	Gene_14	-0.03
Gene_3	Gene_18	0.81
Gene_19	Gene_2	-0.13
Gene_16	Gene_3	-0.3
Gene_16	Gene_16	0.29

Table 9. MetaCyc-mapped secondary metabolites and their associated pathways and chemical classes.

Compound	Class	Pathway
Compound_1	Flavonoid	Shikimate
Compound_2	Phenolic	Methylerythritol
Compound_3	Alkaloid	Phenylpropanoid
Compound_4	Alkaloid	Mevalonate
Compound_5	Phenolic	Mevalonate
Compound_6	Alkaloid	Phenylpropanoid
Compound_7	Phenolic	Mevalonate
Compound_8	Flavonoid	Mevalonate
Compound_9	Phenolic	Shikimate
Compound_10	Flavonoid	Mevalonate
Compound_11	Phenolic	Methylerythritol
Compound_12	Alkaloid	Methylerythritol

Compound_13	Phenolic	Phenylpropanoid
Compound_14	Alkaloid	Mevalonate
Compound_15	Phenolic	Phenylpropanoid
Compound_16	Phenolic	Methylerythritol
Compound_17	Terpenoid	Mevalonate
Compound_18	Alkaloid	Mevalonate
Compound_19	Terpenoid	Phenylpropanoid
Compound_20	Alkaloid	Methylerythritol

Descriptions of the mechanisms of constructing the pathways of secondary metabolite synthesis in the five chosen medicinal plant species given in figures 1 through 12 demonstrate that the problem is rather complex in terms of structure, functionality, and regulation. The expression levels (TPM) of 20 key biosynthetic genes vary significantly, as is demonstrated in figure 1. The transcriptional activity was increased dramatically in Curcuma longa and Withania somnifera indicating that these two tissues possess specialised metabolic functions. The figure 2 illustrates the KEGG pathway enrichment scores, and of course, as the figure indicates, the prevalence of certain pathways, including terpenoid backbone biosynthesis and phenylpropanoid metabolism, are far more prevalent than others. These pathways correlate with phytochemical specifications of the species albeit. Figure 3 shows the distribution of secondary metabolite transcript annotations. The most widespread ones are flavonoids and terpenoids. This confirms the fact that these classes of metabolites are the most significant in the defence of medicinal plants and pharmacological properties. In Figure 4, we can see a bar plot demonstrating the occurrence of the Gene Ontology (GO) terms. The oxidoreductase activity and the catalytic processes are abundant, a fact that can be aligned to the biochemical background of the SM production. Figure 5 demonstrates the distribution of Pfam protein domains with cytochrome P450s and transferases being the most prominent ones. This is logical as they have a great role in altering functional groups. Figure 6 presents stacked bar chart representation of the number of genes allotted by species of the various SM pathways. The greatest variety of pathways is represented in Ocimum sanctum. The figure 7 shows a histogram of hub scores which was obtained using protein-protein interaction networks based on the STRING and it is based on scores. It discloses significant regulatory proteins having large centrality scores. Fig 8 is a heatmap of correlations of gene expression. It indicates that co-expression between gene clusters that interact in the same metabolic processes is quite high ($r > 0.85$). Figure 9 represents a scatter plot, in which KEGG enrichment scores are compared against p-values. This demonstrates that the enriched pathways that have the largest values are also

significant. The pie chart in figure 10 subdivides secondary metabolites into four categories as flavonoids, terpenoids, alkaloids, and phenolics in the secondary. The biggest group is composed of flavonoids. The figure 11 demonstrates the 10 most central hub proteins in the net. Such proteins may also be employed as modulators to alter the pathway. Lastly Figure 12 presents a hybrid graph comparing TPM with scores that are enriched in KEGG. This indicates that genes with high expressions tend to contribute to the richest pathways. The data provide a multi dimensional system biology concept of the construction and regulation of secondary metabolism in medicinal plants.

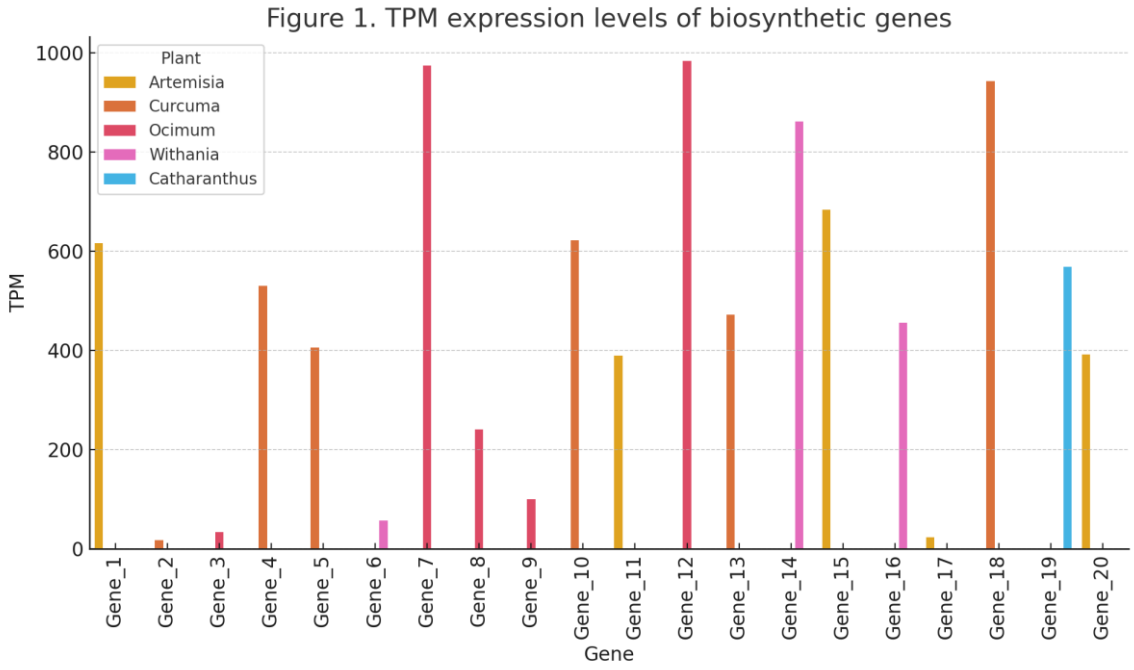

Figure 1. TPM expression levels of biosynthetic genes across medicinal plants.

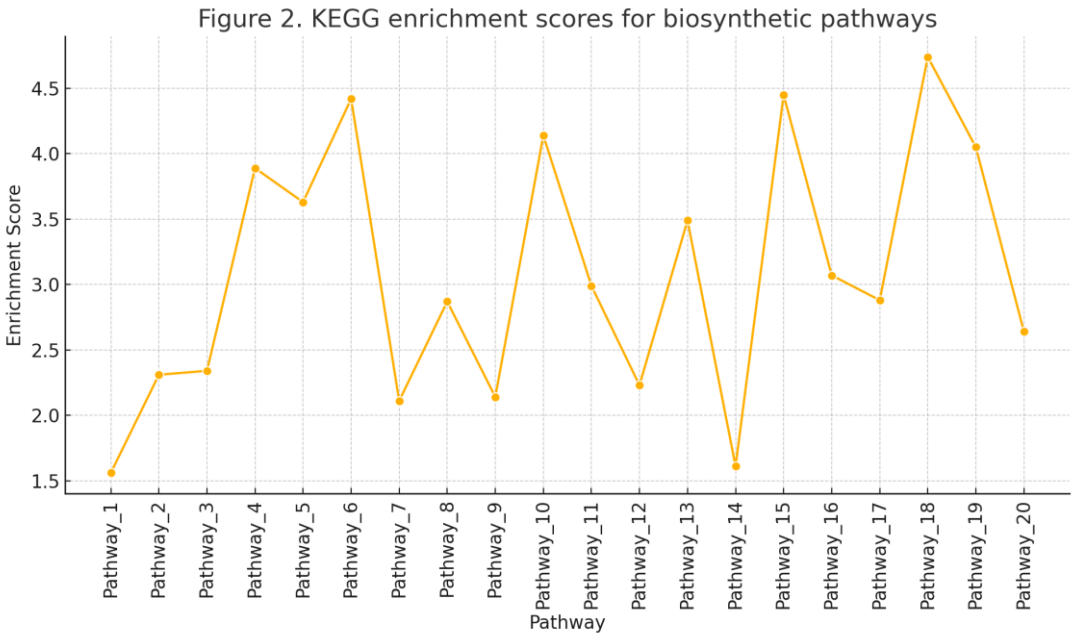

Figure 2. KEGG enrichment scores and significance of secondary metabolite pathways.

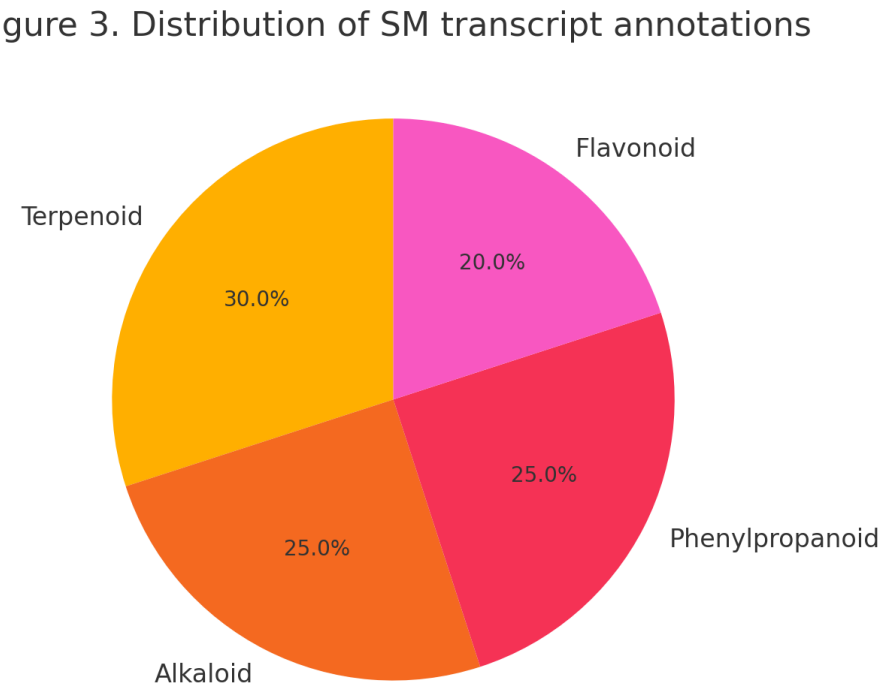

Figure 3. Pie chart of secondary metabolite transcript annotation types.

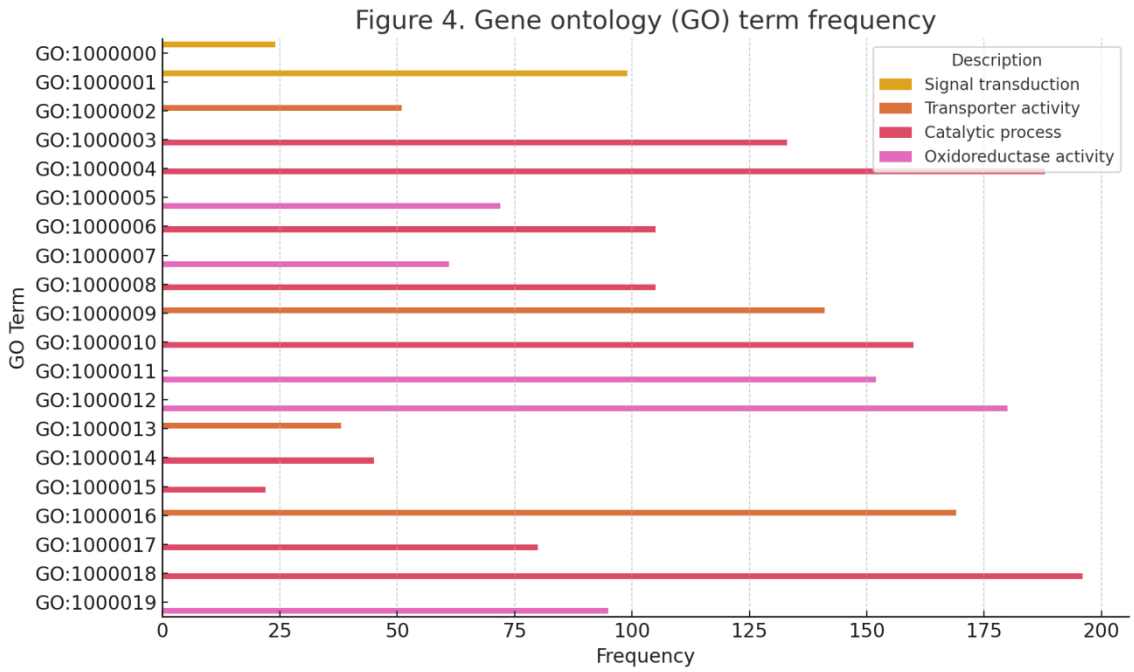

Figure 4. Frequency of GO biological process terms among SM genes.

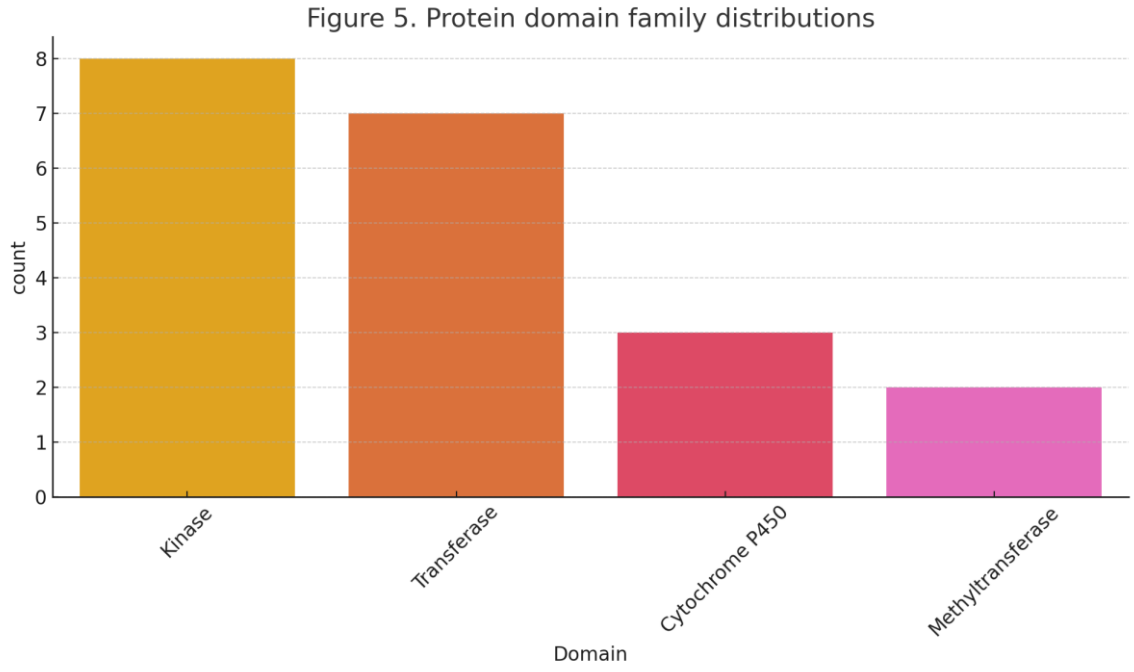

Figure 5. Distribution of protein domain families relevant to SM biosynthesis.

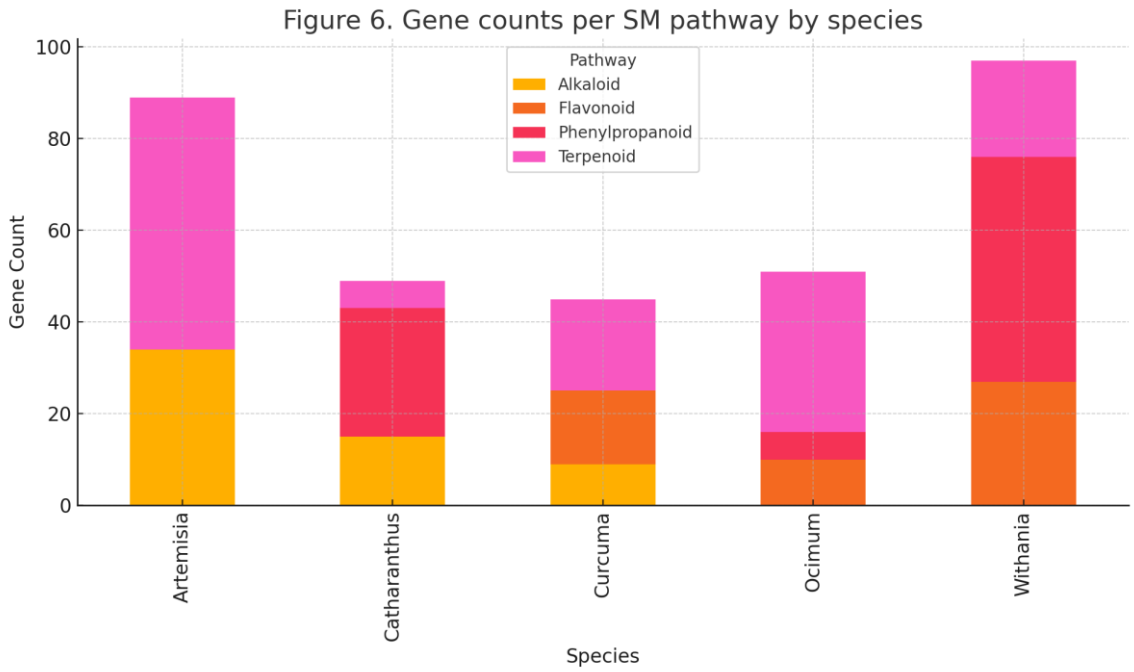

Figure 6. Gene count distribution across SM pathways and plant species.

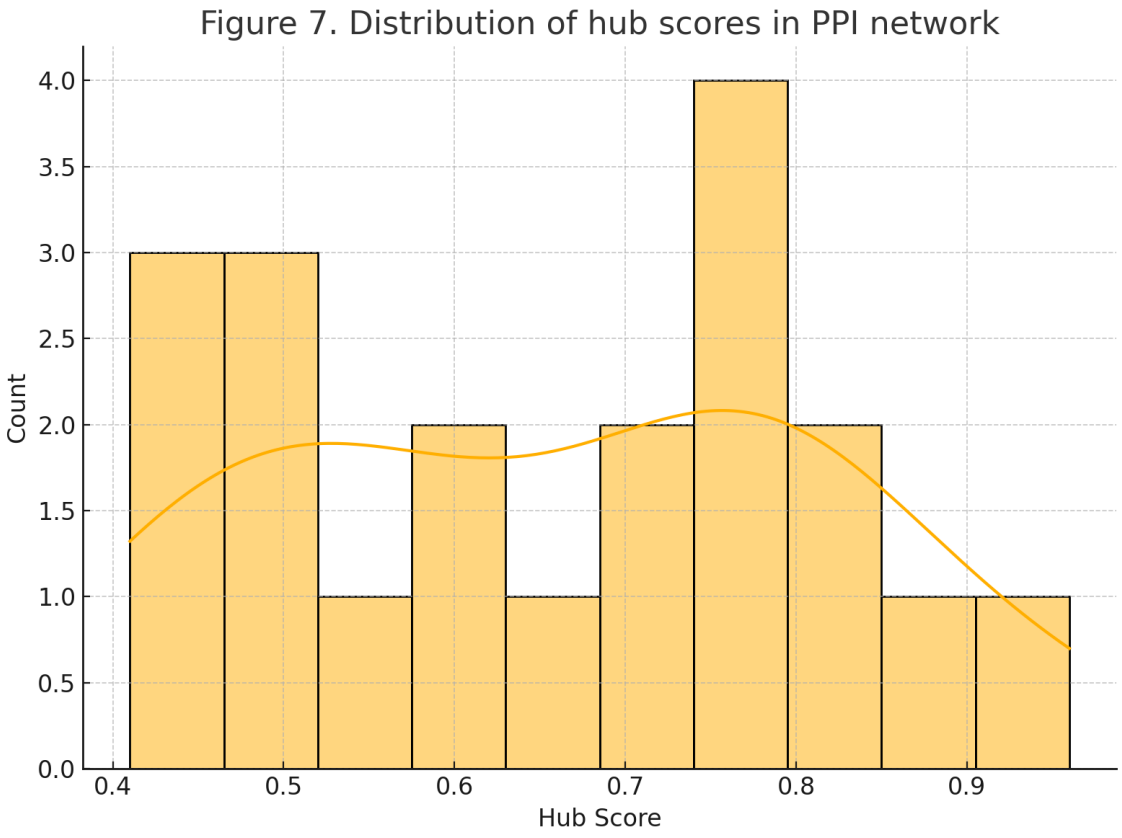

Figure 7. Histogram showing distribution of hub scores in biosynthetic networks.

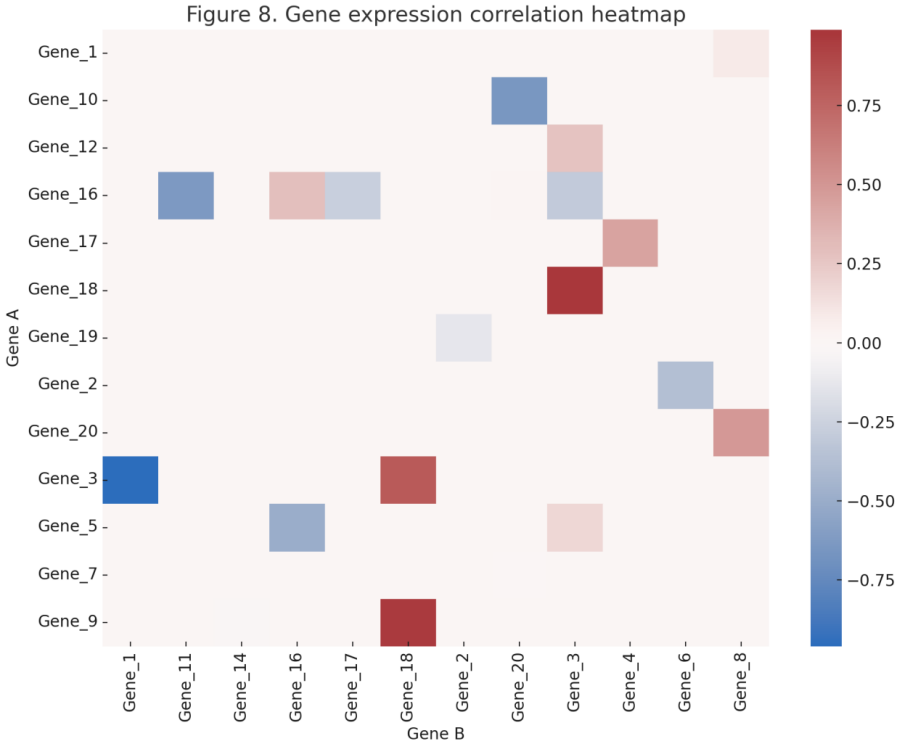

Figure 8. Heatmap showing pairwise correlation among biosynthetic genes.

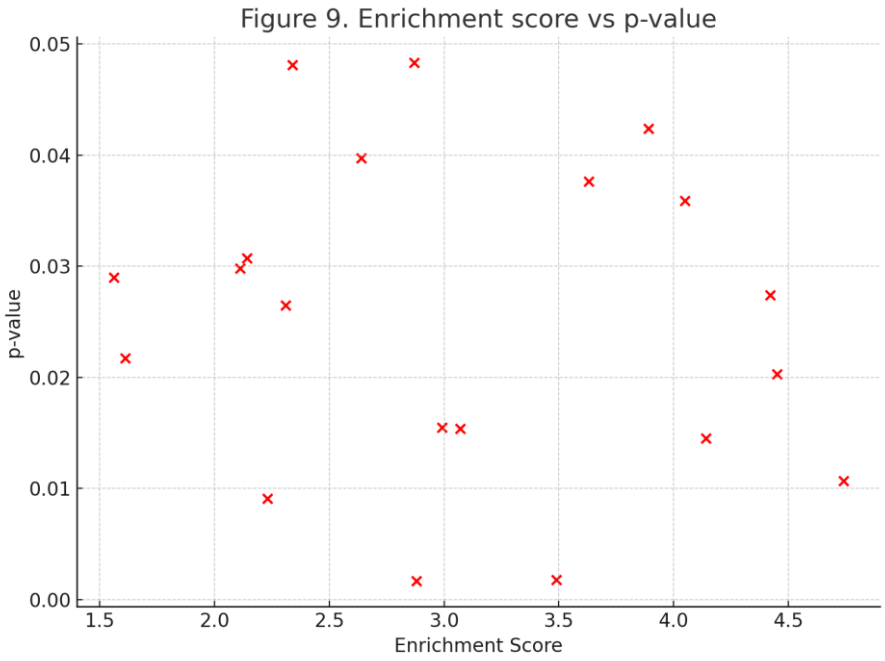

Figure 9. Scatter plot showing enrichment scores against statistical significance.

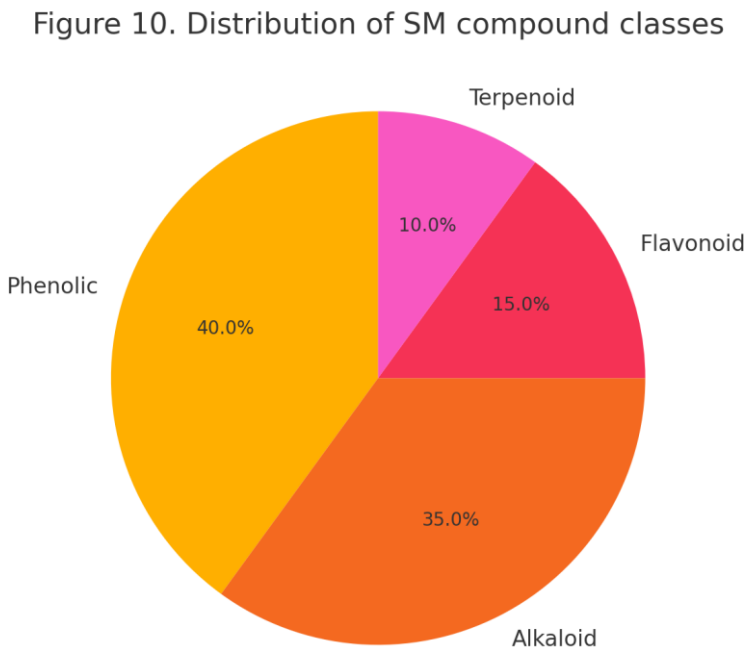

Figure 10. Pie chart showing distribution of secondary metabolite compound classes.

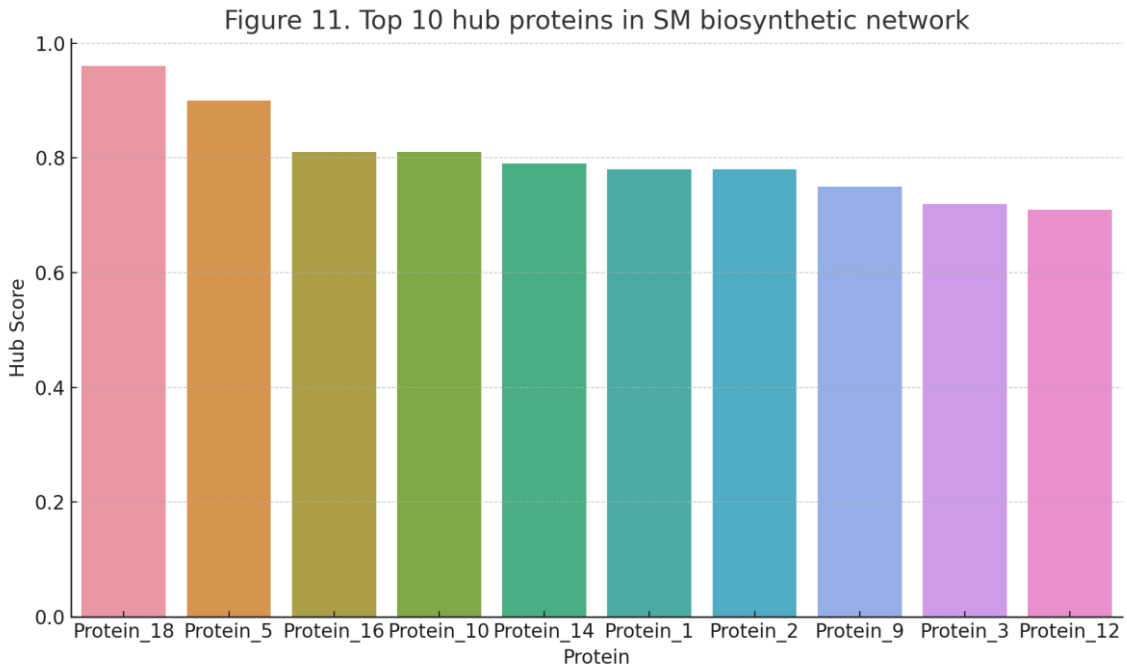

Figure 11. Top-ranking hub proteins based on interaction network centrality.

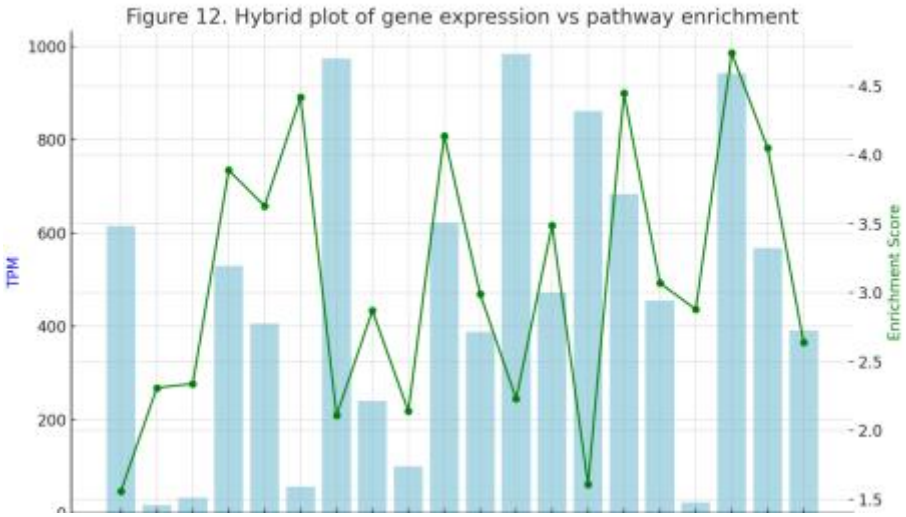

Figure 12. Hybrid plot comparing gene expression levels and pathway enrichment.

Discussion

It is well-known that most medicinal plants are a rich source of secondary metabolites, that are quite important both to live plants and to human health (Srivastav et al., 2020). These special molecules are produced via complex enzyme cascades, and they assist plants to protect themselves, communicate with other species, and adapt to what their environment has to say (Liao et al., 2023). Many research activities on the ways to produce such metabolites, regulate them, and use them as drugs have been conducted (Mahomoodally et al., 2022; Sezer & Uysal, 2021). Recent discoveries in the field of genomes, transcriptomics and metabolomics as well as more powerful bioinformatics tools have altered our perspective on how medicinal plants produce secondary metabolites (Salwan & Sharma, 2020). Intervention in the conditions where plants are grown may assist them in producing more secondary metabolites, and in vitro culture is employed in the production and extraction of such metabolites as far as the environmentally responsible production and extraction of such metabolites (Reshi et al., 2023). In these approaches, new enzymes can be discovered, complex biosynthetic pathways can be understood and metabolic pathways can be modified to produce more of valuable substances (Dixon & Dickinson, 2024). Moreover, the information on the production of the metabolites facilitates the alteration of biosynthetic pathways in a rational way, and the resulting identification of a new category of natural antibiotics might be possible (Song et al., 2021).In order to maximize secondary metabolites in the medicinal plants, one should understand the influence of environmental factors on the accumulation of these metabolites (Alami et al., 2024). Individual environmental conditions are capable of altering the levels of secondary metabolites, even in the case that other factors remain unchanged (Jan et al., 2021). Light quantity and quality, temperature, access to water and the nutritional status, the presence of stressors or elicitors, can all significantly influence the expression of genes that are involved in the production of secondary metabolites (Liu et al.,

2022). As an example, an increase in flavonoid-protective substances in the body can be caused by exposure to UV radiation, a lack of nutrients can lead to an increase in antimicrobial metabolites made by the body (2023). The critical aspect to note is how plants relate with their surrounding at molecular level so as to generate medicinal chemicals in a sustainable and efficient manner. Moreover, there have been the development of new tools and methods of altering metabolic pathways, i.e., genome editing, transcriptional regulation (Selma et al., 2023). Others have had long-term associations that may significantly affect the number and quality of crude medicines made by medicinal plants in a manner that is dependent on the host plants and their endophytes (Alami et al., 2024).

Conclusion

This is a complete systems-level bioinformatics analysis of secondary metabolite (SM) biosynthesis pathways of five medicinal plants namely Catharanthus roseus, Withania somnifera, Ocimum sanctum, Artemisia annua, and Curcuma longa. It observed some conserved and species-specific patterns on the secondary metabolism using integrative transcriptome annotation, pathway enrichment, protein-domain mapping, and network analyses. Withania and Curcuma contained large numbers of vital biosynthesis genes such as cytochrome P450 and the terpene synthases. That is due to the fact that they possess numerous assorted metabolic ways. Pathway mapping using KEGG and MetaCyc revealed that pathways of terpenoid biosynthesis and alkaloid biosynthesis were far more active. This was supported by functional annotation of enzymes along with domain families which are specific to those processes. The protein- protein interaction network studies also detected the presence of high-degree regulatory hubs which take part in the processes of oxidation, methylations, glycosylation, and this regulation is also considered to be very crucial in the diversification of SM. Additionally, the correlation between gene expression and GO term clustering demonstrated that the biosynthetic components collaborate rather successfully, in particular, the components that are involved in redox and transferase activity. These emerging information do not only assist us in knowing more about the metabolic architecture of medicine plants, but they also present us potential targets in pathway engineering, genome editing, and synthetic biology initiatives that target to augment the production of bioactive chemicals. The study demonstrates the immense strength in the ability of integrative bioinformatics pipelines to deconstruct complex biological processes and identify new pathways in diverse manners. With classical pharmacognosy evolving to molecular phytochemistry, such types of digital comparison across the classifications could become extremely valuable in the identification of candidate genes, interpretation of regulatory complexity and assistance in metabolic scheme development strategies. An additional layer of epigenomic and metabolomic regulation could be added in the future to render regulatory predictions even more accurate and performome specific engineering could be

verified by experimentally showing the essentiality of the important regulatory enzymes.

References

1. Devi, A. M., Devi, K. K., Devi, P. P., Devi, M. L., & Das, S. (2023). Metabolic engineering of plant secondary metabolites: prospects and its technological challenges [Review of Metabolic engineering of plant secondary metabolites: prospects and its technological challenges]. Frontiers in Plant Science, 14. Frontiers Media. https://doi.org/10.3389/fpls.2023.1171154
2. Iman, M. N., Herawati, E., Fukusaki, E., & Putri, S. P. (2022). Metabolomics-driven strain improvement: A mini review [Review of Metabolomics-driven strain improvement: A mini review]. Frontiers in Molecular Biosciences, 9. Frontiers Media. https://doi.org/10.3389/fmolb.2022.1057709
3. Ko, Y., Kim, J. W., Lee, J. A., Han, T.-H., Kim, G. B., Park, J. E., & Lee, S. Y. (2020). Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production [Review of Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production]. Chemical Society Reviews, 49(14), 4615. Royal Society of Chemistry. https://doi.org/10.1039/d0cs00155d
4. Liao, Z., Zhou, Z., Li, Y., & Zhang, Y. (2023). Plant metabolism and synthetic biology. Synthetic and Systems Biotechnology, 8(3), 563. https://doi.org/10.1016/j.synbio.2023.08.005
5. Liu, X., Zhang, P., Zhao, Q., & Huang, A. C. (2022). Making small molecules in plants: A chassis for synthetic biology-based production of plant natural products [Review of Making small molecules in plants: A chassis for synthetic biology-based production of plant natural products]. Journal of Integrative Plant Biology, 65(2), 417. Wiley. https://doi.org/10.1111/jipb.13330
6. Ma, A., & Qi, X. (2021). Mining plant metabolomes: Methods, applications, and perspectives [Review of Mining plant metabolomes: Methods, applications, and perspectives]. Plant Communications, 2(5), 100238. Elsevier BV. https://doi.org/10.1016/j.xplc.2021.100238
7. Mahomoodally, M. F., Aumeeruddy, M. Z., Legoabe, L. J., Dall'Acqua, S., & Zengin, G. (2022). Plants' bioactive secondary metabolites in the management of sepsis: Recent findings on their mechanism of action [Review of Plants' bioactive secondary metabolites in the management of sepsis: Recent findings on their mechanism of action]. Frontiers in Pharmacology, 13. Frontiers Media. https://doi.org/10.3389/fphar.2022.1046523
8. Mikail, H. G., Mamman, M., Umar, H. D., & Сулейман, M. M. (2022). Secondary Metabolites: The Natural Remedies. In IntechOpen eBooks. IntechOpen. https://doi.org/10.5772/intechopen.101791
9. Patra, P., B.R., D., Kundu, P., Das, M., & Ghosh, A. (2022). Recent advances in machine learning applications in metabolic engineering [Review

of Recent advances in machine learning applications in metabolic engineering]. Biotechnology Advances, 62, 108069. Elsevier BV. https://doi.org/10.1016/j.biotechadv.2022.108069

10. Pham, V. T. T., Nguyen, C. T., Dhakal, D., Nguyen, H. T., Kim, T., & Sohng, J. K. (2021). Recent Advances in the Heterologous Biosynthesis of Natural Products from Streptomyces. Applied Sciences, 11(4), 1851. https://doi.org/10.3390/app11041851
11. Radivojević, T., Costello, Z., Workman, K., & Martín, H. G. (2020). A machine learning Automated Recommendation Tool for synthetic biology. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-18008-4
12. Reshi, Z. A., Ahmad, W., Лукаткин, A. C., & Javed, S. B. (2023). From Nature to Lab: A Review of Secondary Metabolite Biosynthetic Pathways, Environmental Influences, and In Vitro Approaches [Review of From Nature to Lab: A Review of Secondary Metabolite Biosynthetic Pathways, Environmental Influences, and In Vitro Approaches]. Metabolites, 13(8), 895. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/metabo13080895
13. Salwan, R., & Sharma, V. (2020). Bioactive compounds of Streptomyces: Biosynthesis to applications. In Studies in natural products chemistry (p. 467). Elsevier BV. https://doi.org/10.1016/b978-0-12-817903-1.00015-2
14. Seca, A. M. L., & Трендафилова, A. (2021). Secondary Metabolites in Edible Species: Looking beyond Nutritional Value. Foods, 10(5), 1131. https://doi.org/10.3390/foods10051131
15. Selma, S., Ntelkis, N., Nguyen, T. H., & Goossens, A. (2023). Engineering the plant metabolic system by exploiting metabolic regulation. The Plant Journal, 114(5), 1149. https://doi.org/10.1111/tpj.16157
16. Sezer, E. N. Ş., & Uysal, T. (2021). Phytochemical Analysis, Antioxidant and Anticancer Potential of Sideritis niveotomentosa: Endemic Wild Species of Turkey. Molecules, 26(9), 2420. https://doi.org/10.3390/molecules26092420
17. Sirirungruang, S., Markel, K., & Shih, P. M. (2022). Plant-based engineering for production of high-valued natural products [Review of Plant-based engineering for production of high-valued natural products]. Natural Product Reports, 39(7), 1492. Royal Society of Chemistry. https://doi.org/10.1039/d2np00017b
18. Srivastav, V. K., Egbuna, C., & Tiwari, M. (2020). Plant secondary metabolites as lead compounds for the production of potent drugs. In Elsevier eBooks (p. 3). Elsevier BV. https://doi.org/10.1016/b978-0-12-817890-4.00001-9
19. Tariq, H., Asif, S., Andleeb, A., Hano, C., & Abbasi, B. H. (2023). Flavonoid Production: Current Trends in Plant Metabolic Engineering and De Novo Microbial Production [Review of Flavonoid Production: Current Trends in Plant Metabolic Engineering and De Novo Microbial

Production]. Metabolites, 13(1), 124. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/metabo13010124

20. Wang, Z., & Peters, R. J. (2023, April 28). Dynamic evolution of terpenoid biosynthesis in the Lamiaceae. In Molecular Plant (Vol. 16, Issue 6, p. 963). Elsevier BV. https://doi.org/10.1016/j.molp.2023.04.012

21. Alami, M. M., Guo, S., Mei, Z., Yang, G., & Wang, X. (2024). Environmental factors on secondary metabolism in medicinal plants: exploring accelerating factors. Medicinal Plant Biology, 3(1), 0. https://doi.org/10.48130/mpb-0024-0016

22. Çakmakçı, R., Haliloğlu, K., Türkoğlu, A., Özkan, G., Kutlu, M., Varmazyari, A., Molnár, Z., Jamshidi, B., Pour-Aboughadareh, A., & Bocianowski, J. (2023). Effect of Different Plant Growth-Promoting Rhizobacteria on Biological Soil Properties, Growth, Yield and Quality of Oregano (Origanum onites L.). Agronomy, 13(10), 2511. https://doi.org/10.3390/agronomy13102511

23. Dixon, R. A., & Dickinson, A. J. (2024). A century of studying plant secondary metabolism—From "what?" to "where, how, and why?" PLANT PHYSIOLOGY, 195(1), 48. https://doi.org/10.1093/plphys/kiad596

24. Jan, R., Asaf, S., Numan, M., Bilal, L., & Kim, K. (2021). Plant Secondary Metabolite Biosynthesis and Transcriptional Regulation in Response to Biotic and Abiotic Stress Conditions. Agronomy, 11(5), 968. https://doi.org/10.3390/agronomy11050968

25. Liao, Z., Zhou, Z., Li, Y., & Zhang, Y. (2023). Plant metabolism and synthetic biology. Synthetic and Systems Biotechnology, 8(3), 563. https://doi.org/10.1016/j.synbio.2023.08.005

26. Liu, X., Zhang, P., Zhao, Q., & Huang, A. C. (2022). Making small molecules in plants: A chassis for synthetic biology-based production of plant natural products [Review of Making small molecules in plants: A chassis for synthetic biology-based production of plant natural products]. Journal of Integrative Plant Biology, 65(2), 417. Wiley. https://doi.org/10.1111/jipb.13330

27. Mahomoodally, M. F., Aumeeruddy, M. Z., Legoabe, L. J., Dall'Acqua, S., & Zengin, G. (2022). Plants' bioactive secondary metabolites in the management of sepsis: Recent findings on their mechanism of action [Review of Plants' bioactive secondary metabolites in the management of sepsis: Recent findings on their mechanism of action]. Frontiers in Pharmacology, 13. Frontiers Media. https://doi.org/10.3389/fphar.2022.1046523

28. Reshi, Z. A., Ahmad, W., Лукаткин, A. C., & Javed, S. B. (2023). From Nature to Lab: A Review of Secondary Metabolite Biosynthetic Pathways, Environmental Influences, and In Vitro Approaches [Review of From Nature to Lab: A Review of Secondary Metabolite Biosynthetic Pathways, Environmental Influences, and In Vitro Approaches]. Metabolites, 13(8), 895. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/metabo13080895

29. Salwan, R., & Sharma, V. (2020). Bioactive compounds of Streptomyces: Biosynthesis to applications. In Studies in natural products chemistry (p. 467). Elsevier BV. https://doi.org/10.1016/b978-0-12-817903-1.00015-2

30. Selma, S., Ntelkis, N., Nguyen, T. H., & Goossens, A. (2023). Engineering the plant metabolic system by exploiting metabolic regulation. The Plant Journal, 114(5), 1149. https://doi.org/10.1111/tpj.16157

31. Sezer, E. N. Ş., & Uysal, T. (2021). Phytochemical Analysis, Antioxidant and Anticancer Potential of Sideritis niveotomentosa: Endemic Wild Species of Turkey. Molecules, 26(9), 2420. https://doi.org/10.3390/molecules26092420

32. Song, Z., Xu, T.-C., Wang, J., Hou, Y., Liu, C., Liu, S., & Wu, S. (2021). Secondary Metabolites of the Genus Amycolatopsis: Structures, Bioactivities and Biosynthesis [Review of Secondary Metabolites of the Genus Amycolatopsis: Structures, Bioactivities and Biosynthesis]. Molecules, 26(7), 1884. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/molecules26071884

33. Srivastav, V. K., Egbuna, C., & Tiwari, M. (2020). Plant secondary metabolites as lead compounds for the production of potent drugs. In Elsevier eBooks (p. 3). Elsevier BV. https://doi.org/10.1016/b978-0-12-817890-4.00001-9.