In Silico Analysis of Secondary Metabolite Biosynthesis Clusters in the Genome of *Panicum virgatum*

[1]Supratik Adhikary, [2]Aditi Nayak

Department of Life Science
Guru Nanak Institute of Pharmaceutical Science and Technology
157/F, Nilgunj Rd, Sahid Colony, Panihati, Kolkata
West Bengal-700114, India.

***Abstract*- A biosynthetic gene cluster (BGC) is a group of genes in a genome that collectively encode the enzymes responsible for producing a specific natural product, such as a secondary metabolite or a bioactive compound with diverse biological activities, including antibiotics, antifungals, antiviral agents. Understanding the structure and function of these biosynthetic gene clusters is important for developing new pharmacological compounds. By studying these clusters, it would be beneficial in identifying the potential targets for drug development and strategies to prevent the spread of antibiotic resistance. In this study, we identified biosynthetic gene clusters in *Panicum virgatum*. The plant possess various medicinal properties. Using KNApSAck database, we identified the secondary metabolites present in the plant. PlantiSMASH is a bioinformatics tool that can be used to identify biosynthetic gene clusters (BGCs) in plant genomes. Using PlantiSMASH, BGCs in *Panicum virgatum* genomes were predicted. This study gives an insights into the secondary metabolism of *Panicum virgatum*, and identify potential targets for genetic engineering to improve the nutritional value or other properties of the plant. Hence, the Biosynthetic gene clusters in *Panicum virgatum* offer a promising avenue for improving the productivity, nutritional value, and health benefits.**

Keywords: **Biosynthetic Gene Cluster (BGC), Secondary metabolites, *Panicum virgatum*, PlantiSMASH, KNApSAck database**

1. INTRODUCTION

Plants contain a pool of metabolites having diverse role in medicine and agriculture. Now a days, genome mining has provided a great platform for the discovery of these natural products. The most valuable gift of this genome mining to scientific research is the biosynthetic gene clusters (BGCs). In case of bacteria and fungi, the genes that are involved in the metabolic pathways of the metabolites are found to be clustered and are termed as biosynthetic gene clusters (BGCs) (1-4). In BGCs, the whole genome sequence of the organism is provided which then identify the natural molecules by extracting the structural, chemical, metabolism, and expression data computationally. The discovery of the antiSMASH database, a genome mining platform, has opened up a new road for the discovery of natural compounds. This came into force in the year 2010 (5) and has been continuously developing each year (6,7). The main purposes of this genome mining approach are to search the biosynthetic genes coding for the natural compounds as well as to recognize novel metabolites. This can assist in large scale production of those particular strains industrially. Therefore, this approach is well known as 'gene cluster revolution' (1).

Recently, it has been found that along with microbes, the biosynthetic pathways in plants are also highly clustered. The initial studies on the gene clusters of cyclic hydroxamic acid 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) and avenacin (8,9) have found around thirty plant BGCs (10,11). Most important thing is that all together, the clusters are found to code for a vast range of metabolites like di and triterpenes, cyanogenic glucosides, cyclic hydroxamic acids, steroidal and benzylisoquinoline alkaloids, and polyketides. In the model plant species, *Arabidopsis thaliana*, around four BGCs have been identified those are associated with particular metabolites (12). Developments are ongoing in discovering the genomic sequences of plants at larger scales (13). The complete genome sequence of around 100 plant species are available publicly. Hence, identification of BGCs is a key for the discovery of novel plant natural products.

Tools for genome mining for microbes may not be suitable for plants since the genome of plants are more complicated, bigger, variable and contain sequence for unique enzymes (14) (15,16). Therefore, scientist have introduced a special genome mining tool for plants i.e. antiSMASH for plants or 'plantiSMASH' (plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters). It displays the presence of numerous complex biosynthetic loci in varied plant species. *Panicum virgatum L.* is commonly known as 'Switchgrass', a perennial grass. It is basically used for soil conservation and forage. It is also used as a source of biomass for biofuel production (7-10). Several studies on Switchgrass are there such as from roots and leaves the stress-induced volatile terpene biosynthesis (11), Lignin biosynthesis (12), in generating Bioenergy feedstocks (13), and in logistics also (14). In this study, we have identified the secondary metabolites present in this plant with their chemical structures. Around 53 BGCs are found in *Panicum virgatum*. Hence, this study has opened up a new avenue for the discovery of novel natural products in Switchgrass through *in silico* approach.

2. MATERIALS AND METHODS

2.1. Identification of secondary metabolites present in the plant *Panicum virgatum*

To identify the secondary metabolites in *Panicum virgatum* (PV), "KNApSAck Family" database was used. After entering the web page, there are various options present for the detection of the subject of interest such as core system, search engine, and graph clustering. Not only these there are several different divisions and subdivisions of sections such as food and health, crude drug and biological functional sp. search. For identification of the secondary compounds present in the plant PV, the protocol of Weber, *et al*. was followed (17).

2.2. Identification of Biosynthetic gene clusters (BGCs) present in the plant *Panicum virgatum*

To identify the Biosynthetic gene clusters (BGCs) present in PV, PlantiSMASH database was used. After entering the database search engine with the plant name or NCBI accession number, all the BGCs present in the plants are displayed. Not only this but also the details of size, location, and other details of all BGCs can be accessed through this database (15).

2.3. Determination of secondary metabolic mediated pathway

To analyze the secondary metabolite-mediated pathway, the 'Kyoto Encyclopedia of Genes and Genomes' database was used. As per the protocol by Ma *et al.*, several metabolisms involved in a secondary metabolite-mediated pathway were studied through this database (16).

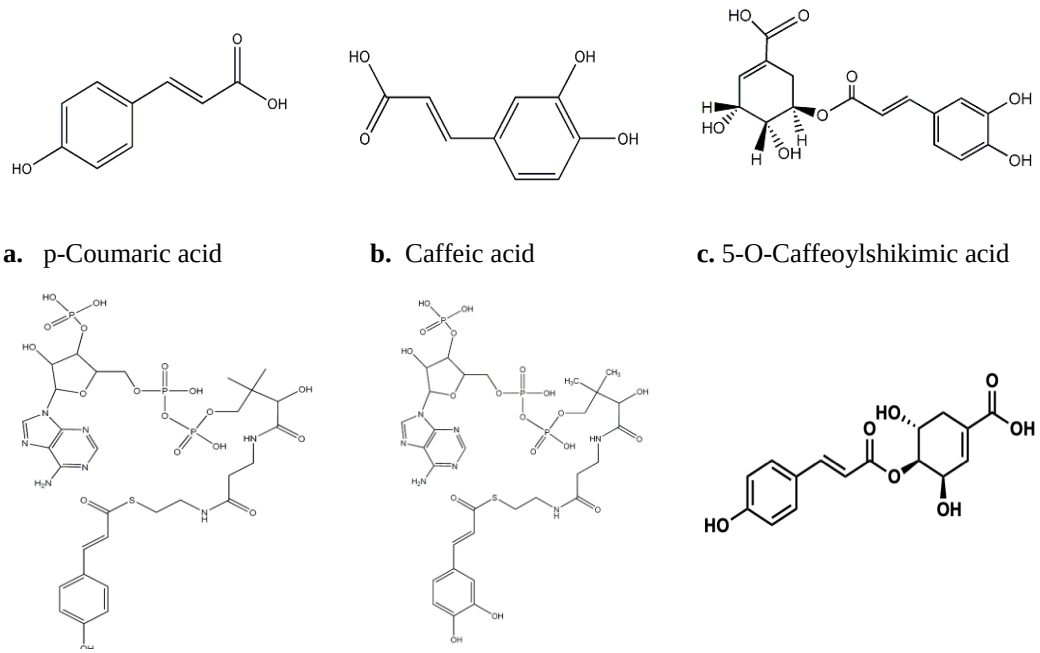
3. RESULTS

3.1. Identification of novel secondary metabolites of *Panicum virgatum*

First of all, we investigated the types of secondary metabolites present in *Panicum virgatum* through our *in silico* study (KNApSAck Family). Interestingly, a total of six secondary metabolites were found in PV (Table 1). Their structures are represented in figure 1.

Table 1: List of secondary metabolites found in *Panicum virgatum* as per the *in silico* study through KNApSAcK**.**

Sl. No.	C_ID	CAS ID	Metabolites	Molecular formula
1	C00000152	7400-08-0	p-Coumaric acid	$C_9H_8O_3$
2	C00000615	501-16-6	Caffeic acid	$C_9H_8O_4$
3	C00002720	73263-62-4	5-O-Caffeoylshikimic acid	$C_{16}H_{16}O_8$
4	C00007280	30802-00-7	p-Coumaroyl CoA	$C_{30}H_{42}N_7O_{18}P_3S$
5	C00007281	53034-79-0	Caffeoyl-CoA	$C_{30}H_{42}N_7O_{19}P_3S$
6	C00056037	196496-50-1	4-Coumaroylshikimate	$C_{16}H_{16}O_7$



a. p-Coumaric acid **b.** Caffeic acid **c.** 5-O-Caffeoylshikimic acid

d. p-Coumaroyl CoA **e.** Caffeoyl-CoA **f.** 4-Coumaroylshikimate

Figure 1. Chemical structure of secondary metabolites found in *Panicum virgatum*. (a) p-Coumaric acid, (b) Caffeic acid, (c) 5-O-Caffeoylshikimic acid, (d) p-Coumaroyl CoA, (e) Caffeoyl-CoA, (f) 4-Coumaroylshikimate.

3.2. Biosynthetic Gene Clusters (BGCs) analysis of *Panicum virgatum*

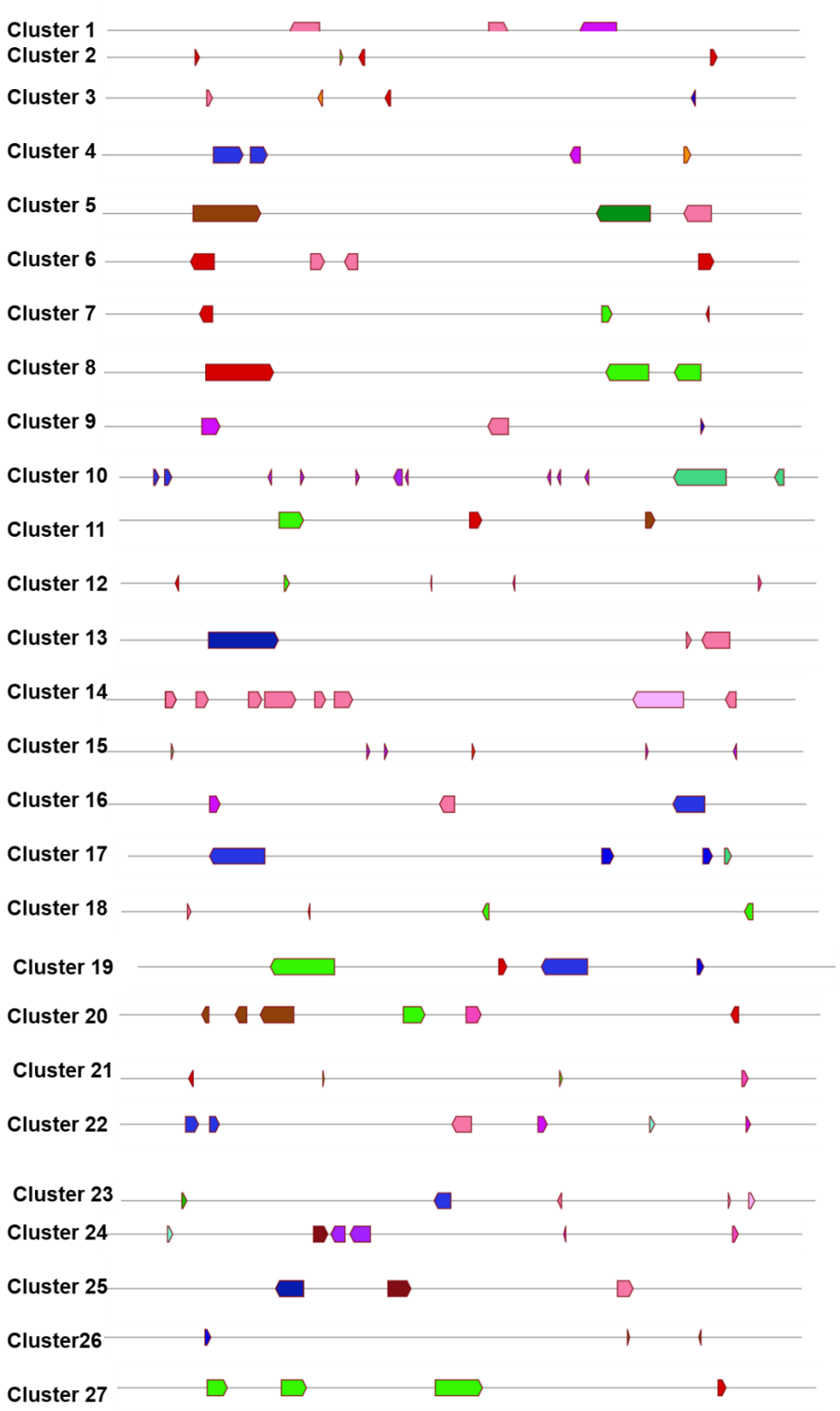
After finding the novel metabolites through our *in silico* study (Figure 1, Table 1), we searched for the potential biosynthetic gene clusters in *Panicum virgatum*. Interestingly, a total of fifty-three BGCs were identified (Figure 2, Table 2). Those BGCs are enlisted in Table 2, with their location, size, and CD-HIT Clusters details.

In the table 2, the number of CD-HIT clusters for each gene cluster signifies the number of functionally different protein (sub) families it encodes. This can be used to highlight gene clusters with complex architectures that produce diverse types of enzymes related to specialized metabolite biosynthesis. An example can be seen with *Panicum virgatum*, containing a predicted saccharide cluster in the chromosome with 3 CD-HIT clusters as depicted in Fig. 2. The legend of each cluster is represented in table 3.

Table 2: Biosynthetic Gene clusters and their details in *Panicum virgatum*

Cluster	Type of metabolites	Location	Size (kb)	CD-HIT Clusters
Cluster 1	Saccharide	2196872 – 2232128	35.26	3
Cluster 2	Terpene	16301953 – 16552887	250.93	4
Cluster 3	Saccharide	50212613 – 50467538	254.93	4
Cluster 4	Putative	52689816 – 52807526	117.71	4
Cluster 5	Saccharide	52870634 – 52923784	53.15	3
Cluster 6	Saccharide	61277895 – 61360091	82.20	4
Cluster 7	Terpene	66344610 – 66509550	164.94	3
Cluster 8	Terpene	71100351 – 71192875	92.52	3
Cluster 9	Saccharide	23693629 – 23783321	89.69	3
Cluster 10	Putative	25688406 – 25891203	202.80	5
Cluster 11	Terpene	40130922 – 40221592	90.67	3
Cluster 12	Saccharide-Terpene	67391985 – 67753211	361.23	5
Cluster 13	Saccharide	68016557 – 68154613	138.06	3
Cluster 14	Saccharide	69810590 – 69905799	95.21	3
Cluster 15	Putative	74619574 – 74851784	232.21	4
Cluster 16	Saccharide	61810255 – 61888027	77.77	3
Cluster 17	Putative	2209644 – 2290091	80.45	4
Cluster 18	Saccharide-Terpene	20745796 – 21066743	320.95	3
Cluster 19	Terpene	56513575 – 56659662	146.09	4
Cluster 20	Putative	3178568 – 3320497	141.93	5
Cluster 21	Terpene	12363497 – 12744451	380.95	4
Cluster 22	Saccharide-Terpene	19747332 – 19945495	198.16	5
Cluster 23	Saccharide-Terpene	22403374 – 22576175	172.80	5

Cluster 24	Polyketide	54820989 – 55061029	240.04	5
Cluster 25	Saccharide	69018376 – 69088333	69.96	3
Cluster 26	Lignan-Alkaloid	56178228 – 56350065	171.84	3
Cluster 27	Terpene	9508472 – 9618651	110.18	3
Cluster 28	Putative	47351443 – 47463092	111.65	6
Cluster 29	Terpene	53983647 – 54156072	172.43	3
Cluster 30	Terpene	39663750 – 39771897	108.15	3
Cluster 31	Alkaloid	49643003 – 49948561	305.56	4
Cluster 32	Saccharide	17255228 – 17373788	118.56	3
Cluster 33	Saccharide	18114999 – 18306185	191.19	3
Cluster 34	Polyketide	20771942 – 21003825	231.88	4
Cluster 35	Lignan	24977166 – 25461974	484.81	4
Cluster 36	Terpene	40077589 – 40584062	506.47	3
Cluster 37	Lignan	2903099 – 2950638	47.54	3
Cluster 38	Saccharide	12244836 – 12354840	110.00	3
Cluster 39	Saccharide	24990865 – 25058503	67.64	3
Cluster 40	Lignan	43890404 – 44009109	118.70	4
Cluster 41	Saccharide-Alkaloid	51186199 – 51298355	112.16	3
Cluster 42	Lignan	1162828 – 1222364	59.54	3
Cluster 43	Saccharide	14362566 – 14413834	51.27	3
Cluster 44	Polyketide	16168880 – 16312724	143.84	3
Cluster 45	Saccharide-Polyketide	9455839 – 9687341	231.50	6
Cluster 46	Saccharide-Polyketide	24061492 – 24171183	109.69	3
Cluster 47	Saccharide	42210215 – 42327302	117.09	5
Cluster 48	Lignan	5214821 – 5260670	45.85	3
Cluster 49	Polyketide	10763597 – 10837503	73.91	3
Cluster 50	Putative	22875320 – 22964668	89.35	4
Cluster 51	Saccharide-Alkaloid	44637564 – 44957188	319.62	5
Cluster 52	Saccharide-Terpene	11796647 – 11944119	147.47	3
Cluster 53	Saccharide	6982 – 26299	19.32	3


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Cluster 13
Cluster 14
Cluster 15
Cluster 16
Cluster 17
Cluster 18
Cluster 19
Cluster 20
Cluster 21
Cluster 22
Cluster 23
Cluster 24
Cluster 25
Cluster26
Cluster 27

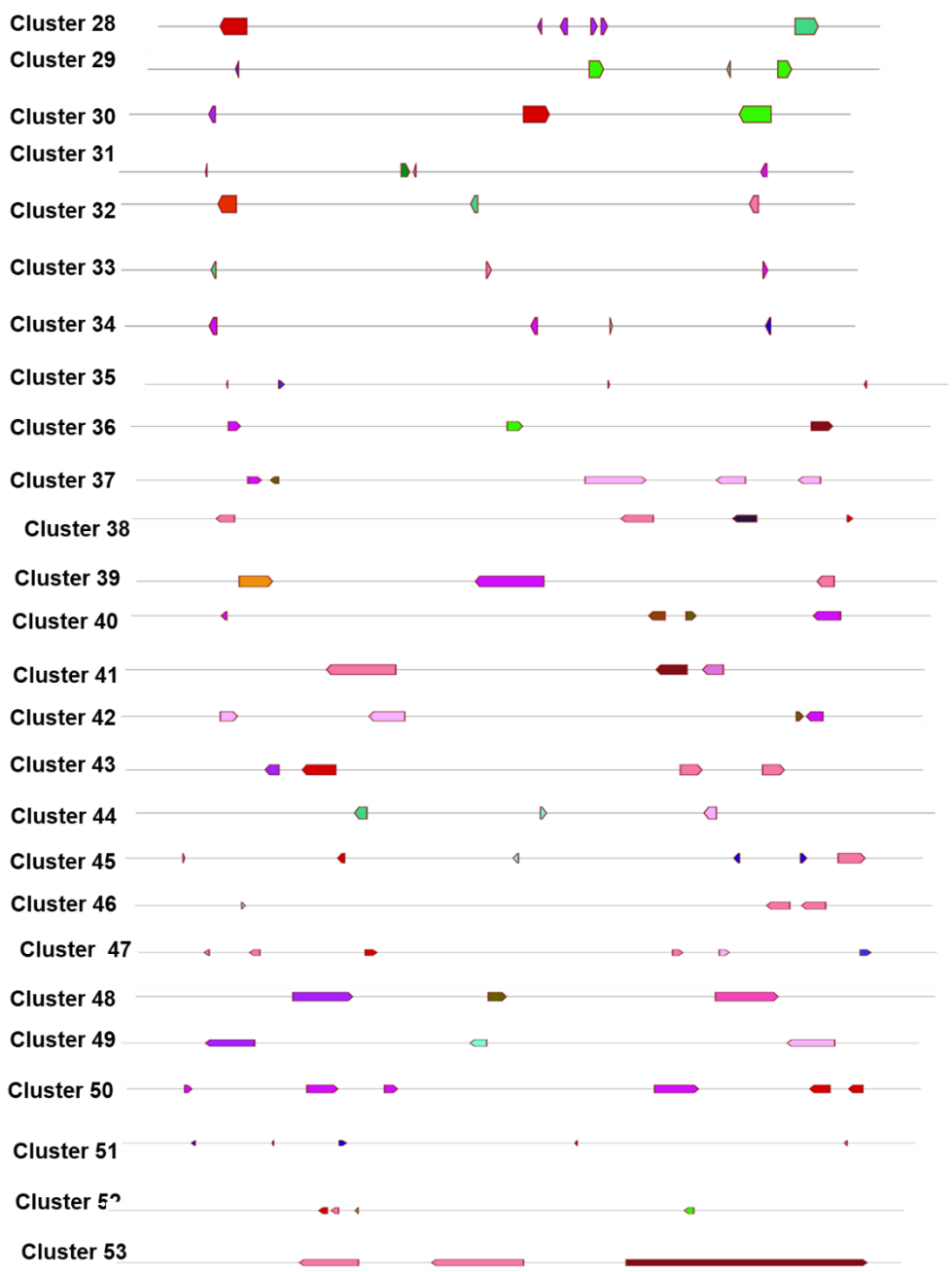


Figure 2: 53 Biosynthetic Gene clusters in different locations in the chromosome of *Panicum virgatum*. The legend of each cluster is represented in table 3.

Table 3: The Legends of Biosynthetic Gene clusters found in *Panicum virgatum*

Cluster	Legend
1	Glycosyltransferase; Methyltransferase; (Other) Biosynthetic Genes; Other Genes
2	Cytochrome 450; Terpene synthase; (Other) Biosynthetic Genes; Other Genes
3	Cytochrome 450; Glycosyltransferase; BAHD acyltransferase; Fatty acid desaturase; (Other) Biosynthetic Genes; Other Genes
4	Methyltransferase; Scl acyltransferase; Fatty acid desaturase; (Other) Biosynthetic Genes; Other Genes
5	Glycosyltransferase; CoA-ligase; Cellulose synthase-like; (Other) Biosynthetic Genes; Other Genes
6	Cytochrome 450; Glycosyltransferase; (Other) Biosynthetic Genes; Other Genes
7	Cytochrome 450; Terpene synthase; (Other) Biosynthetic Genes; Other Genes
8	Cytochrome 450; Terpene synthase; (Other) Biosynthetic Genes; Other Genes

9	Glycosyltransferase, Methyltransferase, BAHD acyltransferase, (Other) Biosynthetic Genes, Other Genes
10	COesterase, Scl acyltransferase, Oxidoreductase, (Other) Biosynthetic Genes, Other Genes
11	Cytochrome 450, Terpene synthase, CoA-ligase, (Other) Biosynthetic Genes, Other Genes
12	Cytochrome 450, Terpene synthase, Glycosyltransferase, Methyltransferase, Epimerase, (Other) Biosynthetic Genes, Other Genes
13	Glycosyltransferase, Lipoxygenase, (Other) Biosynthetic Genes, Other Genes
14	Glycosyltransferase, Dioxygenase, (Other) Biosynthetic Genes, Other Genes
15	COesterase, Oxidoreductase, Aminotransferase, (Other) Biosynthetic Genes, Other Genes
16	Glycosyltransferase, Methyltransferase, Scl acyltransferase, (Other) Biosynthetic Genes, Other Genes
17	BAHD acyltransferase, Scl acyltransferase, Oxidoreductase, (Other) Biosynthetic Genes, Other Genes
18	Cytochrome 450, Terpene synthase, Glycosyltransferase, (Other) Biosynthetic Genes, Other Genes
19	Cytochrome 450, Terpene synthase, BAHD acyltransferase, Scl acyltransferase, (Other) Biosynthetic Genes, Other Genes
20	Cytochrome 450, Terpene synthase, Epimerase, CoA-ligase, (Other) Biosynthetic Genes, Other Genes
21	Cytochrome 450, Terpene synthase, Epimerase, (Other) Biosynthetic Genes, Other Genes
22	Glycosyltransferase, Ketosynthase, Methyltransferase, Scl acyltransferase, (Other) Biosynthetic Genes, Other Genes
23	Glycosyltransferase, Scl acyltransferase, Dioxygenase, PRISE enzymes, (Other) Biosynthetic Genes, Other Genes
24	Ketosynthase, COesterase, Methyltransferase, Epimerase, (Other) Biosynthetic Genes, Other Genes
25	Glycosyltransferase, Lipoxygenase, (Other) Biosynthetic Genes, Other Genes
26	Pictet-Spengler enzyme (Bet v1), BAHD acyltransferase, Dirigent enzymes, (Other) Biosynthetic Genes, Other Genes
27	Cytochrome 450, Terpene synthase, (Other) Biosynthetic Genes, Other Genes
28	Cytochrome 450, COesterase, Oxidoreductase, (Other) Biosynthetic Genes, Other Genes
29	Terpene synthase, Scl acyltransferase, Oxidoreductase, (Other) Biosynthetic Genes, Other Genes
30	Cytochrome 450, Terpene synthase, COesterase, (Other) Biosynthetic Genes, Other Genes
31	Methyltransferase, Strictosidine synthase-like, Cellulose synthase-like, (Other) Biosynthetic Genes, Other Genes
32	Glycosyltransferase, Oxidoreductase, Aminotransferase, (Other) Biosynthetic Genes, Other Genes
33	Glycosyltransferase, Methyltransferase, Oxidoreductase, (Other) Biosynthetic Genes, Other Genes
34	Ketosynthase, Methyltransferase, BAHD acyltransferase, (Other) Biosynthetic Genes, Other Genes

35	Cytochrome 450 Methyltransferase Scl acyltransferase Dirigent enzymes (Other) Biosynthetic Genes Other Genes
36	Terpene synthase Methyltransferase (Other) Biosynthetic Genes Other Genes
37	Methyltransferase Dioxygenase Dirigent enzymes (Other) Biosynthetic Genes Other Genes
38	Cytochrome 450 Glycosyltransferase Amino oxidase (Other) Biosynthetic Genes Other Genes
39	Glycosyltransferase Methyltransferase Fatty acid desaturase (Other) Biosynthetic Genes Other Genes
40	Methyltransferase CoA-ligase Dirigent enzymes (Other) Biosynthetic Genes Other Genes
41	Glycosyltransferase Strictosidine synthase-like (Other) Biosynthetic Genes Other Genes
42	Methyltransferase Dioxygenase Dirigent enzymes (Other) Biosynthetic Genes Other Genes
43	Cytochrome 450 Glycosyltransferase COesterase (Other) Biosynthetic Genes Other Genes
44	Ketosynthase Oxidoreductase Dioxygenase (Other) Biosynthetic Genes Other Genes
45	Cytochrome 450 Glycosyltransferase Ketosynthase BAHD acyltransferase Dioxygenase (Other) Biosynthetic Genes Other Genes
46	Glycosyltransferase Ketosynthase (Other) Biosynthetic Genes Other Genes
47	Cytochrome 450 Glycosyltransferase Scl acyltransferase Dioxygenase (Other) Biosynthetic Genes Other Genes
48	COesterase Epimerase Dirigent enzymes (Other) Biosynthetic Genes Other Genes
49	Ketosynthase COesterase Dioxygenase (Other) Biosynthetic Genes Other Genes
50	Cytochrome 450 Methyltransferase (Other) Biosynthetic Genes Other Genes
51	Cytochrome 450 Pictet-Spengler enzyme (Bet v1) Glycosyltransferase BAHD acyltransferase (Other) Biosynthetic Genes Other Genes
52	Cytochrome 450 Terpene synthase Glycosyltransferase (Other) Biosynthetic Genes Other Genes
53	Glycosyltransferase (Other) Biosynthetic Genes Other Genes

3.3. Secondary metabolite-mediated biosynthesis pathway in *Panicum virgatum*

After finding the novel secondary metabolites in *Panicum virgatum* along with their BGCs, we analyzed the Biosynthetic pathways of secondary metabolites through KEGG pathway study. Secondary metabolite-mediated biosynthesis pathway of Caffeic acid, 5-O-Caffeoylshikimic acid, p-Coumaroyl CoA, caffeoyl-CoA, and 4-Coumaroylshikimate are represented in figure 3, 4, 5, 6, and 7 respectively.

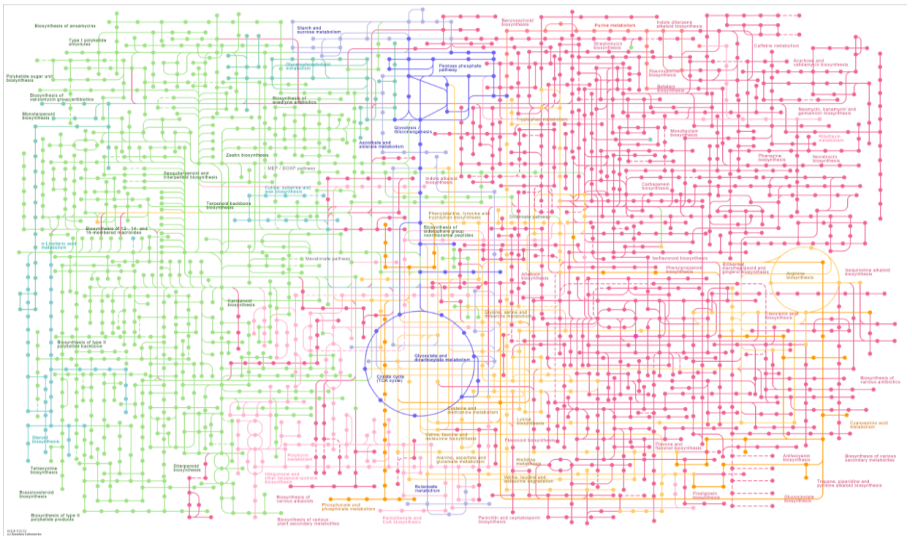


Figure 3: Secondary metabolite-mediated biosynthesis pathway of Caffeic acid

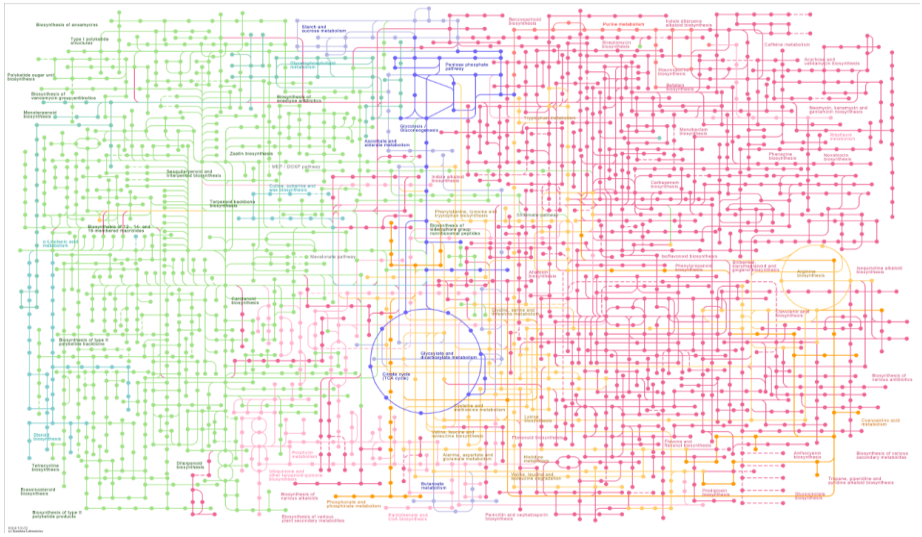


Figure 4: Secondary metabolite-mediated biosynthesis pathway of 5-O-Caffeoylshikimic acid

Figure 5: Secondary metabolite-mediated biosynthesis pathway of p-Coumaroyl CoA

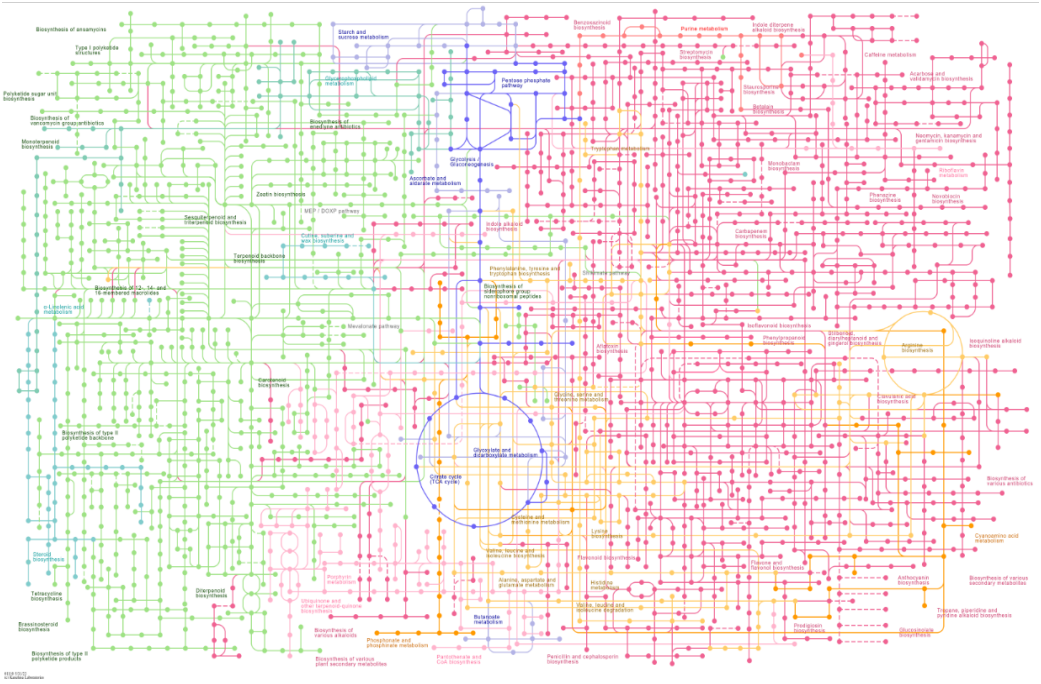


Figure 6: Secondary metabolite-mediated biosynthesis pathway of caffeoyl-CoA

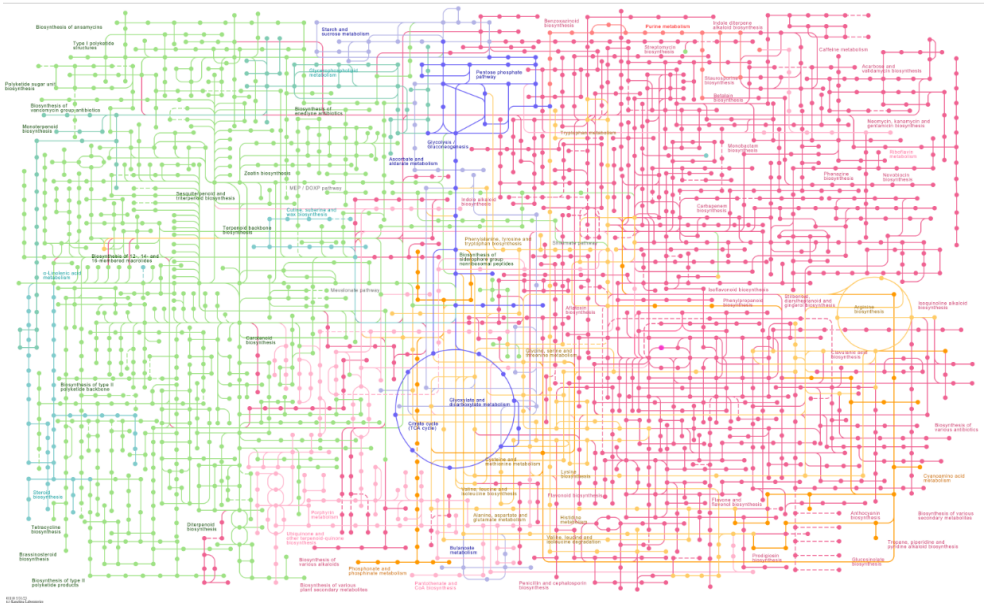


Figure 7: Secondary metabolite-mediated biosynthesis pathway of 4-Coumaroylshikimate

From our analysis, we found that caffeic acid is involved in glyoxylate cycle, threonine biosynthesis, histidine biosynthesis, phenylalanine biosynthesis, and beta-Carotene biosynthesis. oumaroyl coa is associated with tryptophan biosynthesis, jasmonic acid biosynthesis, ascorbate biosynthesis, and cysteine biosynthesis. Caffeoyl-CoA is found to be involved in glyoxylate cycle, castasterone biosynthesis, ethylene biosynthesis, and abscisic acid biosynthesis. Finally, Coumaroylshikimate is found to be involved in glyoxylate cycle, valine/isoleucine biosynthesis, shikimate pathway, and monolignol biosynthesis.

4. DISCUSSION

This *in silico* study identified 6 novel metabolites in *Panicum virgatum.* This was calculated through the latest release of the KNApSAcK Core DB, which contains 101,500 species-metabolite relationships involving 50,048 metabolites and 20,741 species (18). With the use of this database and protocol by Watts *et al*, we tried to estimate the total number of secondary metabolites (six) and their chemical structures (19). Additionally, understanding the metabolic pathways are also very important for analyzing their metabolic networks in plants (20). KEGG pathway database has the existence of all over 223,300 plant species available on Earth and expected metabolites from 136,436 to 367,328 (21). Our BGCs study using PlantiSMASH found a wide variety of candidate BGCs across the plant taxonomy. A recent investigation utilized PlantiSMASH to examine recovered genome bins from Lake Stechlin in north-east Germany for the presence of secondary metabolites (22). After conducting a detailed examination of the

BGCs within individual genome bins, researchers discovered an unclassified bacterium containing a polyketide synthase (PKS) cluster. This cluster exhibited three associated domains related to the enediynes polyketides pathway. Enediynes polyketides are a type of secondary metabolite known for their remarkable anticancer and antibiotic properties, primarily attributed to the cytotoxic characteristics of their enediyne core (23). The majority of these molecules are believed to be advantageous. Within the BGCs, there are clusters of thiopeptides that encode antibiotics, and several of these antibiotics exhibit structural resemblances to compounds currently undergoing clinical trials (24).

5. CONCLUSION

Secondary metabolites (SMs) have played a crucial role in pharmaceuticals, research, and the industry, offering extensive possibilities for expanding our comprehension of chemical diversity and the functions of these metabolites. Numerous studies have uncovered a wealth of biosynthetic gene clusters (BGCs) responsible for both known and potentially undiscovered SMs. These findings reinforce the idea that rich biodiversity often correlates with a wide range of chemical diversity. In the future, there is expected to be further advancements in bioinformatics software to overcome current limitations in genome mining, metagenomic data analysis, and compound characterization. Additionally, new or improved techniques for culturing and screening may emerge to explore the vast array of SMs.

We observed novel metabolites in *Panicum virgatum*. However, further *in vitro* and *in vivo* studies of these compounds are required to confirm these results. In addition, several vegetal sources might be further studied or characterized in order to analyze their potential against several diseases.

Acknowledgement

This study was partially supported by the Guru Nanak Institute of Pharmaceutical Science and Technology in Kolkata. S.A. and A.N. acknowledge the Director and Principal, GNIPST, Kolkata for lab facilities. S.A. received an Institutional fellowship from GNIPST, Kolkata.

Author contributions

S.A. and A.N conceived the study and designed the approach for execution. S.A. accomplished the analyses and construed the data. S.A. prepared the figures and tables. A.N. transcribed the main manuscript text with S.A. and all the authors finally reviewed this.

Competing interests

The author(s) state no competing interests.

REFERENCES:

1. Medema, M. H., and Osbourn, A. (2016) Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. *Natural product reports* **33**, 951-962
2. Kersten, R. D., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.-J., Fenical, W., Fischbach, M. A., Moore, B. S., and Dorrestein, P. C. (2011) A mass spectrometry–guided genome mining approach for natural product peptidogenomics. *Nature chemical biology* **7**, 794-802
3. Keller, N. P., Turner, G., and Bennett, J. W. (2005) Fungal secondary metabolism—from biochemistry to genomics. *Nature Reviews Microbiology* **3**, 937-947
4. Palazzotto, E., and Weber, T. (2018) Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Current opinion in microbiology* **45**, 109-116
5. Ren, H., Wang, B., and Zhao, H. (2017) Breaking the silence: new strategies for discovering novel natural products. *Current opinion in biotechnology* **48**, 21-27
6. Kouprina, N., and Larionov, V. (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. *Nature protocols* **3**, 371-377
7. McLaughlin, S. B., and Kszos, L. A. (2005) Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States. *Biomass and bioenergy* **28**, 515-535
8. Bouton, J. H. (2007) Molecular breeding of switchgrass for use as a biofuel crop. *Current opinion in genetics & development* **17**, 553-558
9. Schmer, M. R., Vogel, K. P., Mitchell, R. B., and Perrin, R. K. (2008) Net energy of cellulosic ethanol from switchgrass. *Proceedings of the National Academy of Sciences* **105**, 464-469
10. Yuan, J. S., Tiller, K. H., Al-Ahmad, H., Stewart, N. R., and Stewart, C. N. (2008) Plants to power: bioenergy to fuel the future. *Trends in plant science* **13**, 421-429
11. Muchlinski, A., Chen, X., Lovell, J. T., Köllner, T. G., Pelot, K. A., Zerbe, P., Ruggiero, M., Callaway III, L., Laliberte, S., and Chen, F. (2019) Biosynthesis and emission of stress-induced volatile terpenes in roots and leaves of switchgrass (Panicum virgatum L.). *Frontiers in plant science* **10**, 1144
12. Rao, X., Chen, X., Shen, H., Ma, Q., Li, G., Tang, Y., Pena, M., York, W., Frazier, T. P., and Lenaghan, S. (2019) Gene regulatory networks for lignin biosynthesis in switchgrass (Panicum virgatum). *Plant Biotechnology Journal* **17**, 580-593
13. Zheljazkov, V. D., Stewart Jr, C. N., Joyce, B., Baxter, H., Cantrell, C. L., Astatkie, T., Jeliazkova, E. A., and Poovaiah, C. R. (2018) Dual utilization of medicinal and aromatic crops as bioenergy feedstocks. *Journal of agricultural and food chemistry* **66**, 8744-8752
14. Sokhansanj, S., Mani, S., Turhollow, A., Kumar, A., Bransby, D., Lynd, L., and Laser, M. (2009) Large-scale production, harvest and logistics of switchgrass (Panicum virgatum L.)–current technology and envisioning a mature technology. *Biofuels, Bioproducts and Biorefining* **3**, 124-141
15. Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A., and Medema, M. H. (2017) plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic acids research* **45**, W55-W63

16. Ma, W., Xu, L., Gao, S., Lyu, X., Cao, X., and Yao, Y. (2021) Melatonin alters the secondary metabolite profile of grape berry skin by promoting VvMYB14-mediated ethylene biosynthesis. *Horticulture research* **8**
17. Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., and Darusman, L. K. (2012) KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant and Cell Physiology* **53**, e1-e1
18. De Luca, V., and St Pierre, B. (2000) The cell and developmental biology of alkaloid biosynthesis. *Trends in plant science* **5**, 168-173
19. Watts, D. J., and Strogatz, S. H. (1998) Collective dynamics of 'small-world'networks. *nature* **393**, 440-442
20. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A.-L. (2000) The large-scale organization of metabolic networks. *Nature* **407**, 651-654
21. Macel, M., Van Dam, N. M., and Keurentjes, J. J. (2010) Metabolomics: the chemistry between ecology and genetics. *Molecular ecology resources* **10**, 583-593
22. Cuadrat, R. R., Ionescu, D., Dávila, A. M., and Grossart, H.-P. (2018) Recovering genomics clusters of secondary metabolites from lakes using genome-resolved metagenomics. *Frontiers in Microbiology* **9**, 251
23. Van Lanen, S. G., and Shen, B. (2008) Biosynthesis of enediyne antitumor antibiotics. *Current topics in medicinal chemistry* **8**, 448-459
24. Donia, M. S., Cimermancic, P., Schulze, C. J., Brown, L. C. W., Martin, J., Mitreva, M., Clardy, J., Linington, R. G., and Fischbach, M. A. (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402-1414