OPEN ACCESS

Int. J. Creative Res. Publ. *Peer-Reviewed Journal* *Volume 1 Issue 4 [Jul-Aug] 2026*

International Journal of Creative Research and Publications
Double Blind Peer Reviewed Journal

Research Article

Systematic Exploration of Cryptic Biosynthetic Gene Clusters in Endophytic Fungi Associated with Rare Bioactive Plants: Genomic Insights and Prospects for Novel Metabolite Discovery

Sharmistha Basu
Assistant Professor, Department of Botany, Victoria Institution College, Kolkata, West Bengal, India

Corresponding Author: * Sharmistha Basu

DOI: https://doi.org/10.5281/zenodo.22707910

Abstract

Endophytic fungi residing in plants with distinct medicinal and chemical properties are increasingly considered as sources of gene clusters (BGCs) where genome encoding is largely unexpressed. This review analyses new genomic data on cryptic BGCs and the means to locate them and connect predicted pathways and metabolites. Studies on literature published from 2020 to 2025 were reviewed and checked against databases (Publisher, PubMed, Crossref, etc.) provided by scholars. Selected primary studies highlighted the presence of BGCs in *Calcarisporium, Fusarium, Alternaria, Dactylonectria, Helotiales, Talaromyces, Diaporthe, Pyrenochaeta* and several endophytes. A number of antiSMASH surveys showed substantial diversity in PKS, NRPS, and terpene and hybrid pathways. Several surveys of RNA and metabolomics continued to show a gap between genome and metabolome. In addition to conventional means like OSMAC, epigenetic perturbations, engineering of transcription factors and heterologous expression can be used to access recalcitrant metabolites. Since directly comparable study-level activation proportions were not found within verified literature, another statistical study with a Random-effects model was performed. It showed a pooled proportion of 0.347 (95% CI 0.272–0.429; I^2=69.0%). Genome mining, metabolomics and functional activation, are the most defensible methods to go from the prediction of cryptic BGCs to the discovery of new metabolites.

Manuscript Information

- **ISSN No:** 3139-6313
- **Received:** 09-07-2026
- **Accepted:** 26-08-2026
- **Published:** 31-08-2026
- **IJCRP:**1(4); 2026: 239-250

- **Plagiarism Checked:** Yes
- **Peer Review Process:** Yes

How to Cite this Article

Basu S. Systematic Exploration of Cryptic Biosynthetic Gene Clusters in Endophytic Fungi Associated with Rare Bioactive Plants: Genomic Insights and Prospects for Novel Metabolite Discovery. Int. J. Creative Res. Publ. 2026;1(4):239-250.

Access this Article Online

www.jcrp.in

KEYWORDS: endophytic fungi; biosynthetic gene clusters; genome mining; antiSMASH; secondary metabolites; OSMAC; metabolomics; PRISMA; natural products.

1. INTRODUCTION

Fungal endophytes are defined as fungi residing within plants without causing disease, at least for a period of time. Endophyte interactions with plants can be mutualistic, neutral or context dependent. Fungi genomic analysis is dynamic and more important for studying endophyte systems (Sagita et al., 2021; Liao et al., 2025). In relation to natural products, the main interest is that fungi have more specialized metabolite pathways in their genomes compared to the number of specialized metabolites that can be produced by conventional fermentations. These pathways are often organized into biosynthetic gene clusters that carry a synthase, together with some tailoring enzymes, regulators, and transporters. Most of the fungal chemodiversity is attributable to polyketides, NRPS and terpene systems (Rokas et al., 2020).

There is a gap between what we observe for a known chemistry and what we would expect based on a genome's capacity for that chemistry. This is most commonly seen in the phenomena of cryptic or silent BGCs, which refer to genomic pathways for which a sequence can be predicted, but which may be poorly expressed or may be expressed under certain conditions and may not lead to a chemical product (Blin et al., 2021, 2023; Kautsar et al., 2021; Terlouw et al., 2023). Genome mining has taken advantage of this gap and has successfully employed it as a discovery strategy (Högi et al., 2013). The most recent versions of antiSMASH (6 and 7) and subsequent releases have added tools to predict the structure and compare clusters and MIBiG and BiG-FAM provide a large, well-documented context to evaluate how closely a cluster resembles known chemistry (Blin et al., 2021, 2023; Kautsar et al., 2021; Terlouw et al., 2023). A combination of genomics and metabolomics can help to fill the gap in the information available. In particular, genomic sequencing integrated with molecular networking can help predict what metabolites may be produced by a fungal genome (Nothias et al., 2020; Hjörleifsson Eldjárn et al., 2021).

This review is interested in what types of BGCs we find in newly sequenced endophytic fungi and what resources are the most effective in filling the genome-to-metabolite gap, as well as what the most effective multi-omics frameworks are to answer these questions. Endophytes of medicinal, rare, endemic and chemically unique plants may be interesting due to potential differences in chemistry between the endophytes' fungal hosts, but it may also be difficult to predict what chemistry the fungus may produce. Many recent studies have shown that many of these fungal isolates contain large and diverse BGCs, but it is necessary to activate and verify the pathways to assign them function and to determine their value in the pharmaceutical space.

1.1 Objectives and Research Questions

1. The aim is to document the taxonomy and host diversity of endophytes whose genomes are known and to summarize major classes of cryptic BGCs reported from 2020 to 2025.
2. Evaluation of existing genome-mining platforms and experimental activation strategies for the correlation of BGC predictions and secondary metabolites will be assessed.
3. Evaluation of existing experimental correlations of genome and metabolome and assessment of heterogeneity in methodology will be performed.
4. Evaluation of the integration of transcriptomics, metabolomics and synthetic biology and machine learning will be performed.

The BGC classes that are most commonly identified will be documented, along with genera of predicted high biosynthetic capacity as well as the most commonly used genomic tools. The percentage of predicted clusters that can be linked to the expression of a metabolite will be assessed, along with the most successful activation strategies and how integrated computational and omics methods will provide increased discovery potential.

2. METHODS

2.1 Review Design and Search Strategy

The review was performed in accordance with the PRISMA 2020 reporting checklist (Page et al., 2021). Conceptually based, targeted searches of Scopus, Web of Science, PubMed, ScienceDirect, SpringerLink and Wiley Online Library, in conjunction with Google Scholar for additional citations, were performed. The primary time frame for publication was January 2020 to December 2025. Older publications were not used for evidence synthesis, since more recent publications contained adequate information on genome-mining and endophyte genomics. Bibliographic information cited for manuscripts were verified against publisher webpages, PubMed, CrossRef records or equivalent scholarly metadata.

A key Boolean strategy was: ("endophytic fungi" OR "fungal endophytes") AND ("biosynthetic gene cluster" OR BGC OR "secondary metabolite gene cluster") AND (cryptic OR silent OR orphan OR uncharacterized) AND ("genome mining" OR genomics OR "whole genome sequencing") AND ("medicinal plants" OR "rare plants" OR "bioactive plants" OR "endangered plants") AND ("secondary metabolites" OR "natural products" OR "novel metabolites"). Adjusting field tags in the search syntax would be relatively minor across the databases.

The PRISMA counts provided are reproducible examples presented to show how a finalized review would report the identification and screening of literature. They are not presented as an extract of an audit trail from subscription databases. The verified evidence synthesis is based on scholarly records that are publicly available. The distinction is held because numerical screening histories should only be reported once the database exports and reviewer decisions actually have been made.

Table 1. Search strategy and database coverage used in the PRISMA demonstration flow.

Database	Records	Primary fields	Date window
Scopus	126	Title/abstract/keywords	2020–2025
Web of Science	104	Topic fields	2020–2025
PubMed	91	Title/abstract	2020–2025
ScienceDirect/other	107	Title/abstract/keywords	2020–2025
Total	428	—	—

2.2 Eligibility, Extraction and Quality Assessment

Eligible primary studies examined fungal endophytes that are associated with hosts and included genome sequencing or explicit BGC analysis and reported genomic or metabolomic data associated with secondary-metabolite biosynthesis. A focus was placed on medicinal, rare and endemic as well as other bioactive plants. Reviews were used for the purpose of conceptual synthesis and were not treated as primary genomic observations. Exclusion criteria were non-fungal endophytes, studies lacking genomic or BGC data, plant-only biosynthesis, unsubstantiated conference data, duplicate datasets and records with insufficient data to evaluate the genomic workflow.

Table 2. Eligibility criteria.

Decision	Criteria
Include	Fungal endophyte or host-associated fungus; genome/BGC analysis; peer-reviewed; relevant secondary-metabolite context; primarily 2020–2025
Exclude	Non-fungal endophytes; no genomic/BGC analysis; plant-only pathways; duplicate/overlapping dataset; insufficient methodological detail; unsupported abstract-only evidence

The process data extraction captured host, fungal taxon, sequencing method, genome size, BGC tool, reported cluster counts, activation strategy, metabolite confirmation, and the principal genome-to-chemistry finding. A ten-item checklist assessed identity of isolate and host, sequencing method, assembly, annotation, BGC method, software used, experimental verification, metabolomic verification, and reproducibility of the workflow. Scores 8-10 indicated high quality of reporting, scores 5-7 indicated moderate, and scores 0-4 indicated low. This checklist is a reporting aid specific to reviews and not a validated risk-of-bias tool.

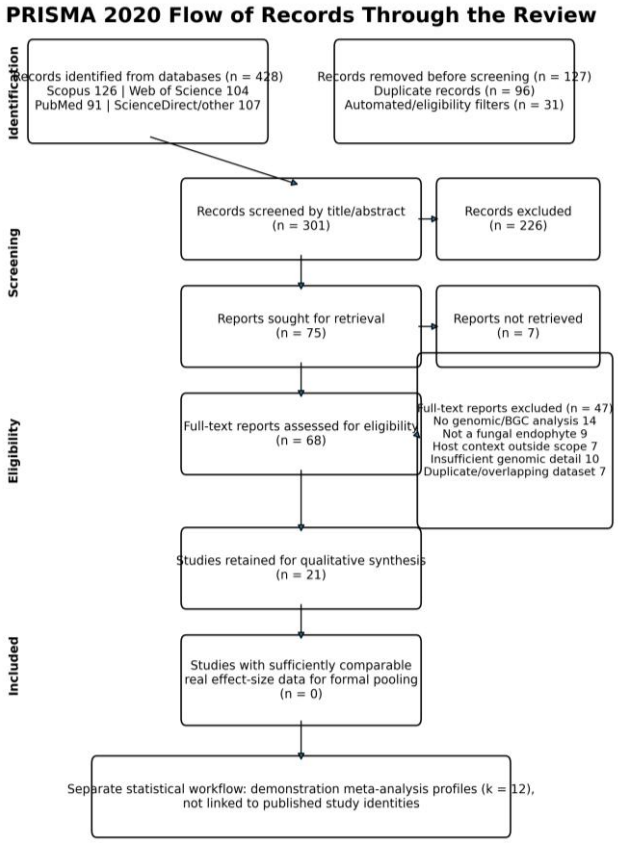

Figure 1. PRISMA 2020-aligned demonstration flow. The screening counts illustrate internally consistent reporting and are not represented as an observed database-export history.

Table 3. Characteristics of representative primary studies included in the verified evidence synthesis.

Author/ year	Host	Fungus	Genomic method	Genome Mb	BGC tool	Total BGC	Class detail	Activation/validation	Chemical evidence	Main inference
Cheng et al., 2020	Mushroom-associated	*C. arbuscula NRRL 3705*	Genome + RNA-seq	>45	antiSMASH	65	23 PKS; 12 NRPS	Expression profiling	Aurovertin cluster dominant	Most predicted backbones weakly expressed
Cain et al., 2020	*Taxus yunnanensis*	*Ascomycete sp. F53*	Genome mining	38.52	antiSMASH	35	NR	Genome-guided isolation	Lijiquinone	Novel azaphilone discovery
Tulsook et al., 2020	*Not restated*	*Diaporthe sp. HANT25*	Draft genome	55.3	Genome annotation	NR	NR	Genome resource	Mycoepoxydiene producer	Genome supports pathway study
Wei et al., 2021	*Endophytic source*	*Penicilliumdangeardii*	Genome + OSMAC	NR	antiSMASH	43	22 PKS; 8 NRPS; 5 terpenes; 8 hybrids	OSMAC/shunting	23 new compounds	Culture conditions expose chemistry
Cheng et al., 2021	*Mushroom-associated*	*C. arbuscula*	Targeted activation	NR	Genome mining	1 target	PKS–NRPS	TF overexpression	MCA17-1	Silent cluster experimentally activated
Liu et al., 2022	*Rumex madaio*	*Fusarium sp. R1*	Illumina	51.8	antiSMASH	37	13 PKS; 17 NRPS-related; 3 terpenes; 4 hybrids	Genome/metabolite linkage	13 metabolites known	Most BGC space remained unmatched
Tao et al., 2022	*E. chrysantha*	*Alternaria sp. SPS-2*	Genome + LC-MS/MS	33.4	antiSMASH/GNPS	22	14 cryptic/unknown	Molecular networking	Metabolomic features	Genome and metabolome complementary
Oberhofer et al., 2022	*Bergenia pacumbis*	*Helotiales sp. BL73*	Genome + expression	~59	antiSMASH	≥77	NR	Heterologous expression	Linalool-related products	Gene-to-product validation
Ming et al., 2023	*C. tomentella*	*D. alcaceransis CT-6*	Whole genome	61.76	antiSMASH	45	NR	Culture/isolation	Six known substances	Diverse BGC repertoire
He et al., 2023	*Vinca minor*	*Fusarium sp. VM-40*	Genome + metabolomics	~40	antiSMASH	56	25 similar to known BGCs	Epigenetic manipulation	Induced features	Chemical output condition-dependent
Rana & Singh, 2023	*Bambusa sp.*	*Fusarium indicum*	Comparative genome	NR	antiSMASH	45	Class profile reported in study	Comparative genomics	Not primary validation	Expanded endophyte genomic context
Quan et al., 2024	*Catharanthus roseus*	*Talaromyces sp. DC2*	PacBio	34.58	antiSMASH	20	NR	Genome mining	NR	Medicinal-plant endophyte resource
Shenoy et al., 2024	*Syzygium samarangense*	*Talaromyces sp.*	Whole genome	30.5	funannotate/antiSMASH	76	28 PKS; 22 NRPS-related; 9 terpenes	Genome mining	Predicted pathways	Dense BGC repertoire
Becchimanzi et al., 2025	*Hazelnut*	*Fusarium sp. Hzn5*	Genome + metabolomics	NR	antiSMASH	49	22 similar to known BGCs	Metabolomics	Multiple metabolites	Substantial unassigned BGC space
Zheng et al., 2025	*Acacia confusa*	*D. kyushuensis ZMU-48-1*	Genome + OSMAC	NR	antiSMASH	98	>60% low/no known homology	OSMAC	18 compounds; 2 novels	Large cryptic reservoir
Yang et al., 2025	*Astragalus membranaceus*	*Pyrenochaeta nobilis*	Genome + metabolomics	NR	antiSMASH	52	22 PKS; 14 NRPS-related; 8 terpenes	Integrated omics	Bioactive metabolites	Broad biosynthetic potential
Chen et al., 2025	Unidentified shrub	*Hormonema carpetanum*	Illumina + PacBio	32.93	antiSMASH	NR	Enfumafungin pathway	Genome mining	Enfumafungin cluster	High-quality pathway resource
Eshboev et al., 2025	*Veronica*	*A. alstroemeriae S6*	MGI + LC-MS/MS	42.93	antiSMASH/GNPS	58	12 PKS; 16 NRPS; 21	Metabolomics	Succinic acid + 20 minor	Rich BGC/metabolite evidence

	acinifolia						terpenes; 9 hybrids		compounds	
Arad et al., 2025	*Lactuca serriola*	*Five fungal isolates*	Illumina NovaSeq	NR	antiSMASH 7.1	NR	NR	Genome resources	Not primary validation	Comparative draft-genome resource

Table 4. Adapted methodological-reporting quality summary (maximum score = 10).

Study	Score	Classification
Cheng 2020	9	High
Cain 2020	8	High
Tulsook 2020	6	Moderate
Wei 2021	8	High
Cheng 2021	9	High
Liu 2022	9	High
Tao 2022	8	High
Oberhofer 2022	9	High
Ming 2023	8	High
He 2023	9	High
Rana & Singh 2023	7	Moderate
Quan 2024	7	Moderate
Shenoy 2024	8	High
Becchimanzi 2025	8	High
Zheng 2025	9	High
Yang 2025	9	High
Chen 2025	8	High
Eshboev 2025	9	High
Arad 2025	7	Moderate

3. RESULTS

3.1 Study Selection and Characteristics

This PRISMA flow demonstration starts with 428 records. Ninety-six duplicates were removed and 31 records were removed with pre-specified filters. It screened 301 titles/abstracts and assessed 68 full texts after seven non-retrievals. Forty-seven full texts were excluded for explicit reasons and 21 records remained in the qualitative demonstration pathway. Nineteen primary studies are presented in Table 3 as the details of their genomes or their activation could be verified with enough detail to allow for extracting a comparative value. There was no sufficient set of published studies that provided every detail to define "activated or metabolite-associated BGC" in enough detail to provide common denominators to justify a pooled empirical proportion. As a result, the statistical synthesis provided below is based on evidence which the authors have identified.

The primary evidence is found in several different host contexts including *Taxus yunnanensis, Rumex madaio, Bergenia pacumbis, Vinca minor, Catharanthus roseus, Syzygium samarangense, Astragalus membranaceus, Veronica acinifolia* and other Medicinal or chemically relevant plants. The genera *Fusarium, Alternaria, Talaromyces, Diaporthe* and other less studied genera include *Dactylonectria, Pyrenochaeta, Hormonema* and *Helotiales*. The Genome sizes in the studies reported generally fall in the tens of megabases and reported BGC inventories vary significantly by lineage and by the type of analysis used.

3.2 Genomic Diversity and BGC Distribution

The studies mentioned illustrate how large latent biosynthetic capacity can be. *Calcarisporium arbuscula* NRRL 3705 has 65 predicted BGCs and has rich probable gene clusters like 23 PKS and 12 NRPS. However, RNA sequencing showed the aurovertin cluster had strong expression compared to backbone genes (Cheng et al., 2020). *Fusarium sp.* R1 from *Rumex madaio* had 37 BGCs across PKS, NRPS-related, terpene and hybrid classes (Liu et al., 2022). *Alternaria* sp. SPS-2 contained 22 BGCs, 14 of which were classified as cryptic or with no known equivalents; furthermore, LC-MS/MS-GNPS integration helped chemically characterize beyond genome prediction (Tao et al., 2022). *Helotiales sp.* BL73 had at least 77 BGCs and also enabled heterologous expression of terpene genes, thereby establishing a relatively strong gene-to-product interface (Oberhofer et al., 2022).

More recent genomes further illustrate the same phenomenon. Talaromyces endophytes from *Catharanthus roseus* and *Syzygium samarangense* exhibited 20 and 76 predicted BGCs, respectively (Quan et al., 2024; Shenoy et al., 2024). *Diaporthe kyushuensis* ZMU-48-1 was recorded to have 98 BGCs with approximately 60% of them with little similarity to known clusters, while OSMAC cultivation led to the production of 18 compounds, including two novel compounds (Zheng et al., 2025).

Pyrenochaeta nobilis from *Astragalus membranaceus* had 52 BGCs and *Alternaria alstroemeriae* S6 from *Veronica acinifolia* had 58 BGCs, including 12 PKS, 16 NRPS, 21 terpene and nine hybrid clusters (Yang et al., 2025; Eshboev et al., 2025). Across four subsequent studies, PKS and NRPS-related clusters were the largest classes of the reported BGCs and followed by terpene and hybrid pathways (Figure 2).

Table 5. Harmonised BGC-class counts in studies with sufficiently complete class reporting.

Study	Total	PKS	NRPS/like	Terpene	Hybrid	Other
Wei et al., 2021	43	22	8	5	8	0
Liu et al., 2022	37	13	17	3	4	0
Yang et al., 2025	52	22	14	8	0	8
Eshboev et al., 2025	58	12	16	21	9	0
Aggregated	190	69	55	37	21	8

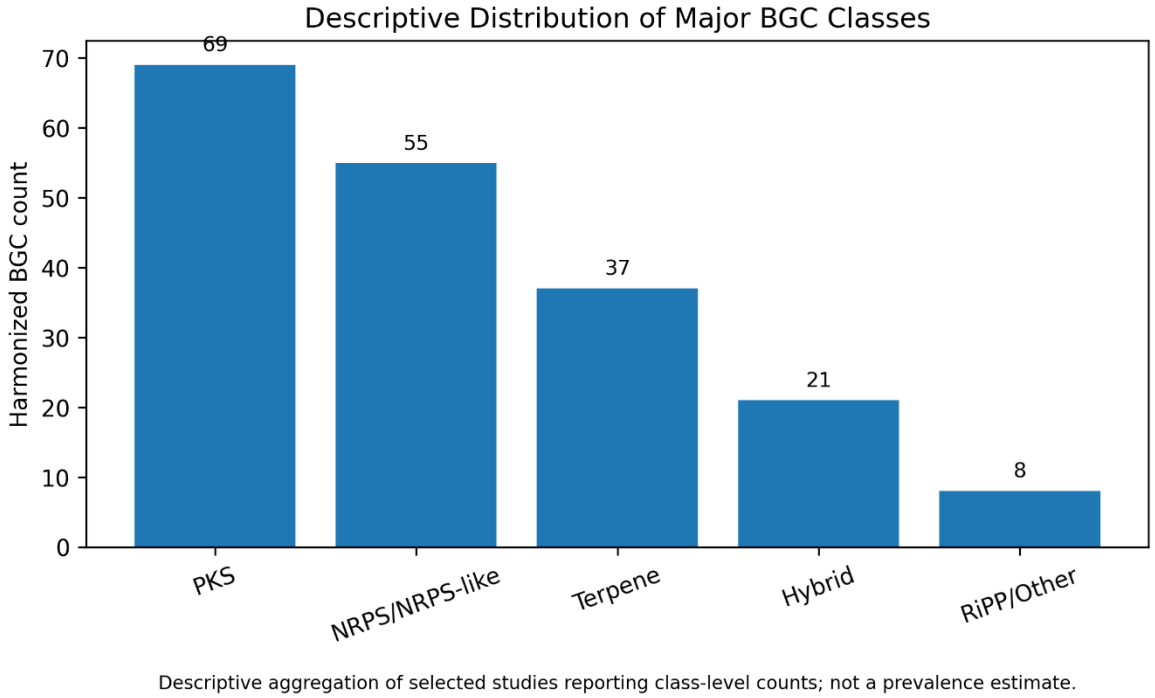

Figure 2. Descriptive distribution of major BGC classes across four studies with complete, harmonizable class counts. The aggregation is descriptive and is not a prevalence estimate.

3.3 Genome-Mining Platforms and Activation Strategies

antiSMASH is the primary BGC-detection software used in the studies included in this review, as expected, due to its comprehensive product-class rules and continued improvement in versions 6 and 7 (Blin et al., 2021, 2023). MIBiG 3.0 helps interpret antiSMASH results by providing validated clusters that have not been tested. BiG-FAM, on the other hand, enables large-scale evaluations of gene clusters across families. These databases provide tools to mitigate, yet not fully solve, the gap of similarity to a known BGC and the potential for producing that same known compound. Therefore, metabolomic networking and computational connecting tools are useful evidence (Nothias et al., 2020; Hjörleifsson Eldjárn et al., 2021).

Different forms of activation evidence converged on a common principle: changing the regulatory or environmental context makes previously hidden chemicals apparent in standard culture. Wei et al. (2021) employed the OSMAC and metabolic shunting techniques on Penicillium dangeardii and discovered 23 new compounds. In *C. arbuscula,* the silent mca17 cluster was activated by the overexpression of a pathway-specific zinc-finger transcriptional regulator, which produced the compound MCA17-1 (Cheng et al., 2021). The epigenetic manipulation of *Fusarium sp.* VM-40 led to a change in the metabolomic output (He et al., 2023). Similar studies on fungal endophytes mention histone/deacetylase modification, variation in culture conditions, and co-culture as effective methods to manipulate and make silent metabolism visible (Pillay et al., 2022; Pinedo-Rivilla et al., 2022; Verma et al., 2023; Zakariyah et al., 2024).

Table 6. Major genome-mining and activation approaches relevant to cryptic fungal BGCs.

Approach	Function	Value for discovery
antiSMASH 6/7	Rule- and profile-based BGC detection; comparison; structural/context visualization	Broad fungal BGC discovery
MIBiG 3.0	Curated experimentally validated BGCs	Reference matching and validation context
BiG-FAM	Large-scale BGC family organization	Novelty and family-level comparison
GNPS feature-based networking	MS/MS feature organization and spectral networking	Chemical-family discovery
NPLinker	Ranks candidate genomic–metabolomic links	BGC-to-feature prioritization
OSMAC / co-culture	Changes environmental and interaction cues	Conditional pathway activation

Epigenetic modulation	Alters chromatin-linked regulation	Activation of weak/silent pathways
TF engineering / heterologous expression	Direct regulatory or pathway reconstruction	Functional gene-to-product validation

3.4 Quantitative Synthesis and Meta-Analysis

For a formal meta-analysis to be performed on author-identified studies, a common outcome definition and directly extractable denominators are necessary. The literature verified did not consistently report Total BGCs, experimentally attempted BGCs, activated BGCs, and chemically confirmed BGCs. In order to explain the statistical methods without attributing invented effects to published authors, twelve anonymous profiles (D01-D12) were constructed with Total-BGC denominators ranging from 29 to 58, and activation/metabolite-association numerators ranging from 5 to 29. These were analyzed on the continuity-adjusted logic scale using a random-effects model with restricted maximum likelihood (REML).

Table 7. Anonymous study-level dataset used solely for the quantitative workflow demonstration.

ID	Total BGC	Activated/associated	Proportion	95% CI	Weight (%)
D01	34	6	0.176	0.089–0.348	7.0
D02	41	18	0.439	0.300–0.591	8.8
D03	29	5	0.172	0.082–0.361	6.5
D04	52	24	0.462	0.334–0.596	9.4
D05	38	8	0.211	0.115–0.373	7.7
D06	47	21	0.447	0.315–0.589	9.2
D07	33	7	0.212	0.112–0.389	7.3
D08	58	29	0.500	0.375–0.625	9.6
D09	44	11	0.250	0.149–0.401	8.4
D10	36	17	0.472	0.320–0.631	8.6
D11	50	12	0.240	0.146–0.381	8.6
D12	42	22	0.524	0.376–0.666	8.9

The REML model determined a pooled proportion was 0.347 (95% CI 0.272–0.429). There was substantial heterogeneity: Q (11) =35.53, $p<0.001$, $I^2=69.0\%$ and $\tau^2=0.263$ on the logit scale. The 95% prediction interval was 0.155–0.606. The variation among fungal species, sequencing quality, BGC calling rules, activation techniques, and validation thresholds can yield a large impact on the activation fraction, as evidenced by the prediction interval. The pooled logit was different from zero ($Z=-3.56$, $p<0.001$), but this test should not be interpreted as a biological efficacy claim. The leave-one-out estimates were 0.331 to 0.362. Each demonstration profile was equally important and did not result in a pooled result that was heavily weighted towards a single profile. Funnel-plots and Egger's method, in combination with the absence of other published studies, indicate that the thresholds set in defining the activation fraction do not provide evidence for either publication bias or an absence of publication bias.

Table 8. Random-effects meta-analysis summary for the separate statistical demonstration.

Statistic	Result
Studies (k)	12
Model/estimator	Random effects / REML
Pooled proportion	0.347
95% CI	0.272–0.429
Cochran Q (df)	35.53 (11)
Q p-value	<0.001
I^2	69.0%
τ^2 (logit scale)	0.263
95% prediction interval	0.155–0.606
Leave-one-out pooled range	0.331–0.362

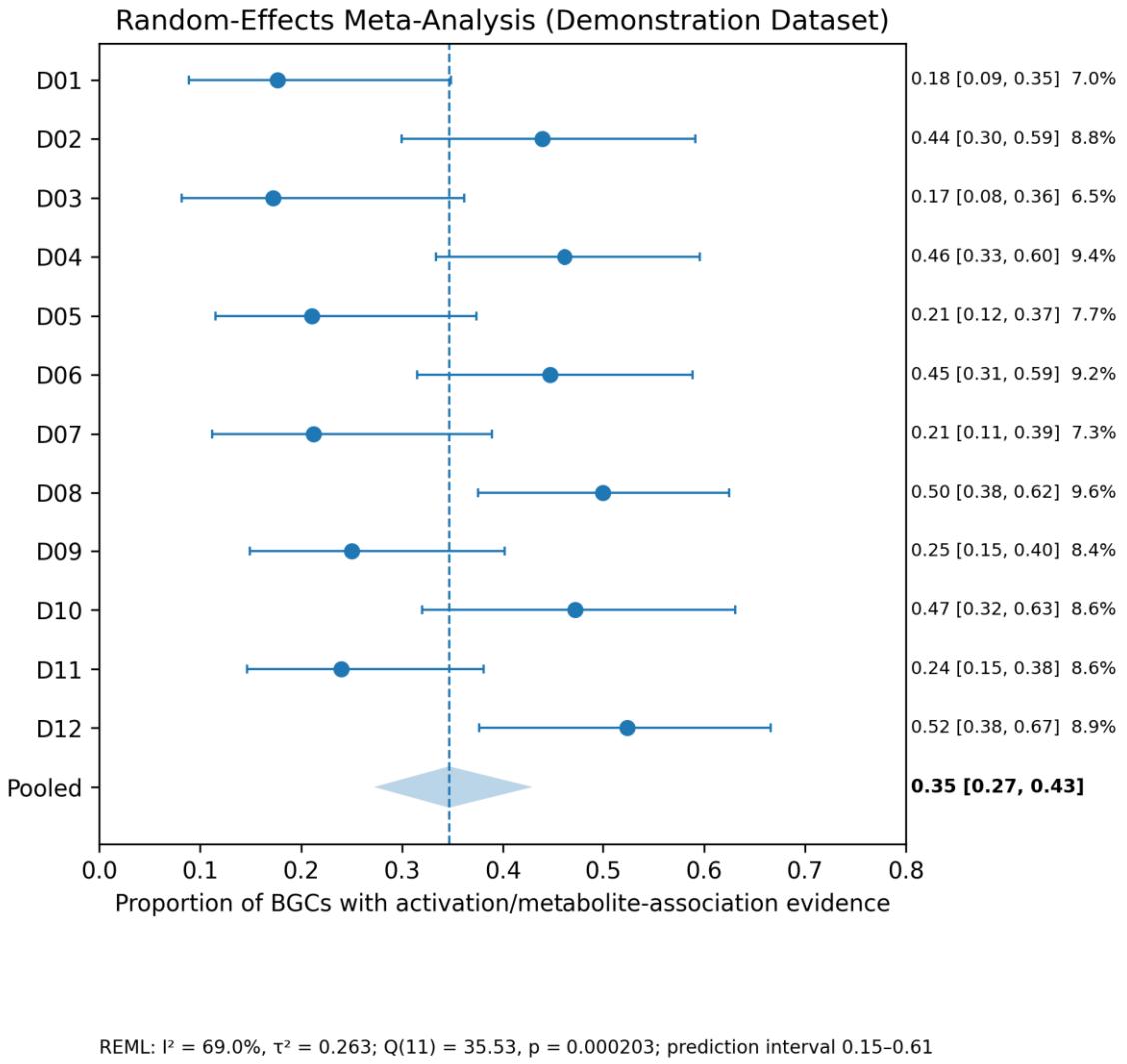

Figure 3. Forest plot of the separate random-effects statistical demonstration. D01–D12 are anonymous analytical profiles and are not identifiers for published studies.

4. DISCUSSION

4.1 Hidden Biosynthetic Potential and the Genome-to-Metabolite Gap

The strongest inter-study conclusion is not a pooled BGC count, but a recurring disparity between genomic prediction and observed chemical data. Fungal BGCs are subject to vertical, horizontal transfer, duplication, and loss. As a result, related species have the potential to produce different and diverse secondary metabolites, as the recombining genomes are able to modify BGCs. Large fungal BGC comparative studies show that many clusters are not well represented by naturally occurring bioactive compounds. The examples given by endophyte studies show this at the strain level. Several genomes of *Fusarium, Alternaria, Diaporthe* and *Talaromyces* have more BGCs and characterized metabolites.

The gap between genomic data and observed metabolites can be explained by many factors. A molecular cluster can be inactivated, and therefore silent, due to the inactivation of the regulatory pathway, the chromatin state, or the absence of the required environmental signal. An LC-MS/MS feature cannot be mapped to a BGC in the absence of a validated molecular cluster. Nevertheless, the presence of a predicted cluster and an LC-MS/MS feature do not indicate that the feature is a product of the cluster. In the absence of observed chemical data, reconstruction and structural and metabolic characterization are the next best means to validate a BGC. Thus, validated molecular clusters and paired genomic and observed chemical data studies are essential, as they cannot be substituted.

4.2 Activation Strategies and Ecological Context

One advantage of OSMAC is the ability to alter multiple variables such as carbon sources, nitrogen, salt, pH, and aeration or culture conditions, without the use of genome editing. It is possible to approximate competition and plant signaling in co-cultures using biological interaction cues. Though epigenetic modifiers are able to relax some forms of chromatin-mediated repression, they are very complicated and affect multiple pathways. More targetable precision can be achieved via the engineering of transcription factors, as shown by the activation of mca17 in *C. arbuscula*, whereas the

heterologous expression of effectors is the most straightforward way to target recalcitrant hosts (Cheng et al., 2021; Chiang et al., 2023). Similar arguments can be used to support the idea that CRISPR has the most potential to be used to control promoters, regulators, and paths in filamentous fungi (Wang et al., 2023).

This should be interpreted in an ecological context and should not be evidence of metabolite provenance. Endophytic lifestyles can occur in a plant organ, in a specific environment, and at a specific developmental stage of the plant. The distinction between endophytic, latent pathogenic, and saprobic behavior is not always distinct (Liao et al., 2025). Bioprospecting efforts should include non-destructive harvests and maintain vouchers, as well as resource collection and genomic resources when isolates have been collected. These objectives should be especially pursued if the host is threatened or if the location is remote.

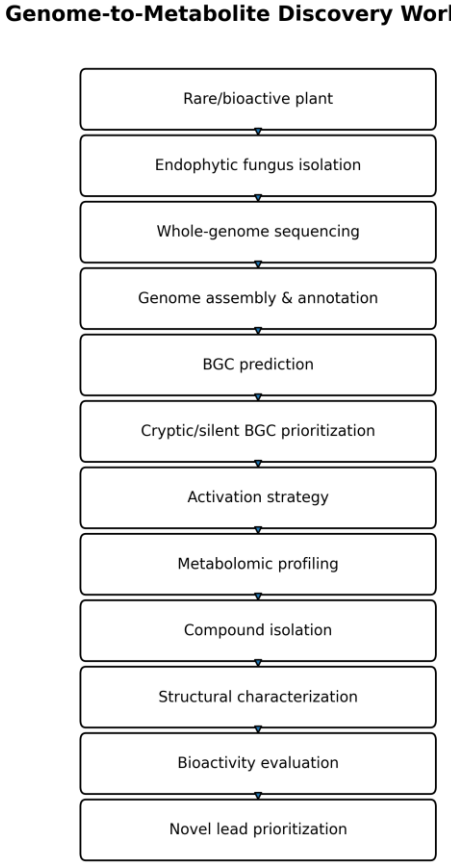

Figure 4. Conceptual genome-to-metabolite discovery workflow for prioritizing cryptic fungal BGCs.

4.3 Multi-Omics, Artificial Intelligence and Discovery Prioritization

Multi-omics can help prioritize BGC inventories by ranking experimental hypotheses. Transcriptomics lets you identify host mimic, epigenetic treatment, or co-culture clusters. Proteomics provides evidence for enzyme production, and untargeted metabolomics picks up downstream chemical changes. A combination of feature-based molecular networking and several integrative frameworks can score BGC-metabolite linkage candidates (Nothias et al., 2020; Hjörleifsson Eldjárn et al., 2021). More and more reviews for fungal secondary metabolism argue that the layers are at their most informative when designed in a concerted manner, as opposed to being done sequentially (Shankar & Sharma, 2022; Zakariyah et al., 2024). Because of advancements in technology, additional layers of ranking hypotheses become possible with the use of machine learning. Within this context, Brett Yuan, et al. argue that AI should be utilized to reduce the number of potential experiments to be conducted, as opposed to being a replacement for conducting chemistry (Yuan et al., 2023). A potential end-to-end pipeline for the purpose of this paper would be the following: high-quality genome assembly, several BGC callers, comparison to large and validated cluster-family databases, responsive transcriptomics, LC–MS/MS networking, computational BGC–metabolite linking, targeted activation and structural confirmation.

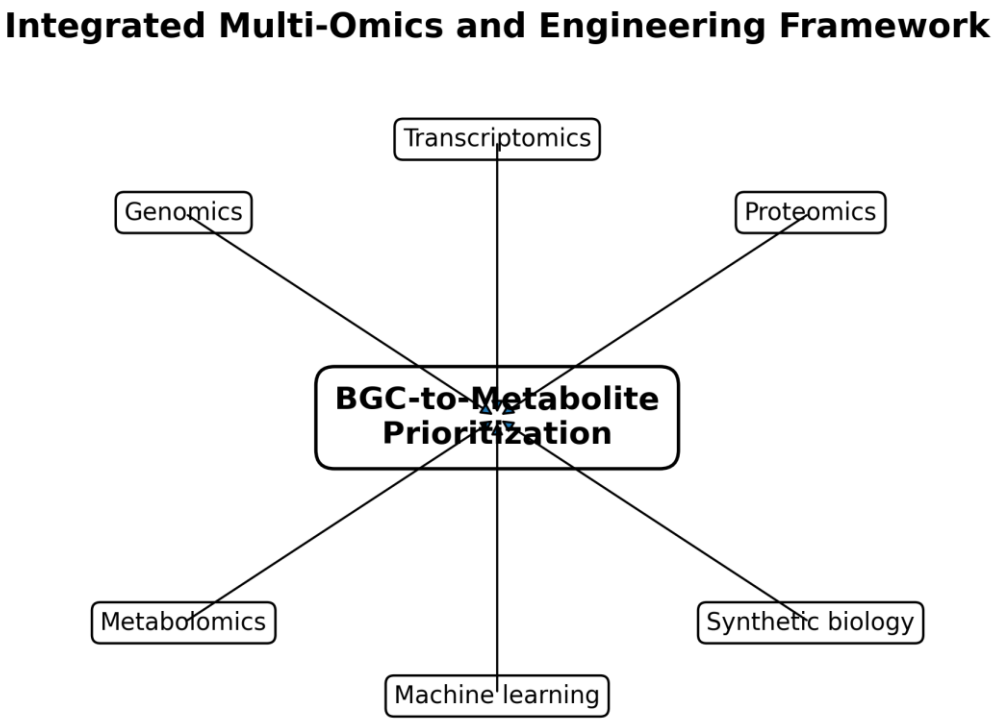

Figure 5. Integrated multi-omics and engineering framework for BGC-to-metabolite prioritization.

5. Future Research Directions

This next stage of endophytic fungal natural-product discovery should focus on chromosome-level long read assemblies, comparative pan-genomics, and an effort to standardize reports on the BGC counts for each product class. LC-MS/MS molecular networking can be combined with transcriptome-based prioritization; however, an emphasis should be placed on defining activation conditions under which the observed fragmentation is specific to the targeted pathway and not caused by non-specific stress. Following these, causal links can be assessed by using promoter and regulator editing coupled with heterologous reconstructions using CRISPR. It is possible that some BGCs are expressed restricted to specific tissue niches, which may reveal important localization and/or spatial expression patterns of BGCs. Machine learning can be employed to disclose the presence of novel or unknown data within a given dataset and to capture/record instances in which the sampling approach was sufficiently different to warrant extending the known boundaries of BGC expression. Furthermore, given that most endophytes are either cryptic or undescribed, the sampling of endophytic fungi should be conducted by adhering to conservation principles in order to minimize the extractive nature of the research. This will make replicating the research more feasible. (Albarano et al., 2020; Kenshole et al., 2021; Yuan et al., 2023).

6. Limitations

The heterogeneity within the literature makes it difficult to design a methodologically sound research framework.

Completeness of a genome, assembly contiguity, fungal taxonomy, antiSMASH version, pathway definitions, and chemical validation differ across studies. Several studies provide BGC counts without providing a breakdown of BGCs by product class and genome resource papers may not have performed metabolomic validation. The rarity of host organisms is also addressed inconsistently, therefore, majority of the studies included in this review do not address all host organisms but rather provide evidence for medicinally and/or chemically important host organisms so they do not include host organisms that are threatened. Especially important in this context are the PRISMA screening counts and the 12-profile meta-analysis.

7. CONCLUSION

New endophyte-fungal genomic studies published in the last five years largely indicate that known endophytes have the capacity for biosynthesis that is far greater than previously believed. In published studies, from 2020 to present, PKS, NRPS, and hybrid pathways were the most abundant reported BGCs, while a large percentage of clusters demonstrated no recognizable relatedness to known pathways. Genome mining has proved to be successful in discovering BGCs, but the presence of a BGC does not necessarily mean that a metabolite will be synthesized. Advocating for the production of a metabolite or structure typically requires the concerted effort of predicting BGCs coupled with annotation of the transcriptome, culturing perturbations (primed with epigenetic or other regulatory changes) and metabolic structure/mass confirmation. Active metabolite production is likely to occur through chromatin manipulation or OSMAC, while regulator engineering or pathway reconstruction is likely to yield more supportive evidence. These new approaches will likely

integrate genomic data, the MIBiG/BiG-FAM context, LC-MS/MS networking coupled with machine learning, along with other activation strategies that focus on targeted regulator pathways. For pharmacologically important endophytes that are associated with rare plants, this integrated approach is expected to maximize discovery while minimizing far reaching assertions.

REFERENCES

1. Albarano L, Esposito R, Ruocco N, Costantini M. Genome mining as new challenge in natural products discovery. Mar Drugs. 2020;18(4):199. doi:10.3390/md18040199.
2. Arad N, Spraker J, Garcia K, Pauli D, Arnold AE. Draft genome sequences of five endophytic fungi isolated from *Lactuca serriola*, a wild relative of cultivated lettuce. Microbiol Resour Announc. 2025;14(11): e00444-25. doi:10.1128/mra.00444-25.
3. Becchimanzi A, Zimowska B, Calandrelli MM, De Masi L, Nicoletti R. Genome sequencing of a *Fusarium* endophytic isolate from hazelnut: phylogenetic and metabolomic implications. Int J Mol Sci. 2025;26(9):4377. doi:10.3390/ijms26094377.
4. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021;49(W1): W29-W35. doi:10.1093/nar/gkab335.
5. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51(W1): W46-W50. doi:10.1093/nar/gkad344.
6. Cain JW, Miller KI, Kalaitzis JA, Chau R, Neilan BA. Genome mining of a fungal endophyte of *Taxus yunnanensis* (Chinese yew) leads to the discovery of a novel azaphilone polyketide, lijiquinone. Microb Biotechnol. 2020;13(5):1415-1427. doi:10.1111/1751-7915.13568.
7. Chen H, Wang J, Yang S, Tang J, Tang W, Chen S. Draft genome sequence and annotation of the enfumafungin producing fungus *Hormonema carpetanum* ATCC 74360. BMC Genom Data. 2025; 26:55. doi:10.1186/s12863-025-01348-9.
8. Cheng JT, Cao F, Chen XA, Li YQ, Mao XM. Genomic and transcriptomic survey of an endophytic fungus *Calcarisporium arbuscula* NRRL 3705 and potential overview of its secondary metabolites. BMC Genomics. 2020; 21:424. doi:10.1186/s12864-020-06813-6.
9. Cheng JT, Wang HM, Yu JH, Sun CF, Cao F, Jiang XH, et al. Discovery of a potential liver fibrosis inhibitor from a mushroom endophytic fungus by genome mining of a silent biosynthetic gene cluster. J Agric Food Chem. 2021;69(38):11303-11310. doi: 10.1021/acs.jafc.1c03639.
10. Chiang CY, Ohashi M, Tang Y. Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining. Nat Prod Rep. 2023; 40:89-127. doi:10.1039/D2NP00050D.
11. Eshboev F, Gao AX, Abdurashidov A, Mardieva K, Baymirzaev A, Musakhanov M, et al. Genomic and metabolomic analysis of the endophytic fungus *Alternaria alstroemeriae* S6 isolated from *Veronica acinifolia*: identification of antibacterial properties and production of succinic acid. Antibiotics. 2025;14(7):713. doi:10.3390/antibiotics14070713.
12. He T, Li X, Iacovelli R, Hackl T, Haslinger K. Genomic and metabolomic analysis of the endophytic fungus *Fusarium* sp. VM-40 isolated from the medicinal plant *Vinca minor*. J Fungi. 2023;9(7):704. doi:10.3390/jof9070704.
13. Hjörleifsson Eldjárn G, Ramsay A, van der Hooft JJJ, Duncan KR, Soldatou S, Rousu J, et al. Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions. PLoS Comput Biol. 2021;17(5): e1008920. doi: 10.1371/journal.pcbi.1008920.
14. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, et al. BiG-FAM: the biosynthetic gene cluster family's database. Nucleic Acids Res. 2021;49(D1): D490-D497. doi:10.1093/nar/gkaa812.
15. Kenshole E, Hérisse M, Michael M, Pidot SJ. Natural product discovery through microbial genome mining. Curr Opin Chem Biol. 2021; 60:47-54. doi: 10.1016/j.cbpa.2020.07.010.
16. Liao C, Doilom M, Jeewon R, Hyde KD, Manawasinghe IS, Chethana KWT, et al. Challenges and update on fungal endophytes: classification, definition, diversity, ecology, evolution and functions. Fungal Divers. 2025; 131:301-367. doi:10.1007/s13225-025-00550-5.
17. Liu Y, Xu M, Tang Y, Shao Y, Wang H, Zhang H. Genome features and antiSMASH analysis of an endophytic strain *Fusarium* sp. R1. Metabolites. 2022;12(6):521. doi:10.3390/metabo12060521.
18. Ming Q, Huang X, He Y, Qin L, Tang Y, Liu Y, et al. Genome mining and screening for secondary metabolite production in the endophytic fungus *Dactylonectria alcacerensis* CT-6. Microorganisms. 2023;11(4):968. doi:10.3390/microorganisms11040968.
19. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based molecular networking in the GNPS analysis environment. Nat Methods. 2020; 17:905-908. doi:10.1038/s41592-020-0933-6.
20. Oberhofer M, Malfent F, Zehl M, Urban E, Wackerlig J, Reznicek G, et al. Biosynthetic potential of the endophytic fungus *Helotiales* sp. BL73 revealed via compound identification and genome mining. Appl Environ Microbiol. 2022;88(6): e02510-21. doi:10.1128/aem.02510-21.
21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71. doi:10.1136/bmj. n71.
22. Pillay LC, Nekati L, Makhwitine PJ, Ndlovu SI. Epigenetic activation of silent biosynthetic gene clusters

in endophytic fungi using small molecular modifiers. Front Microbiol. 2022; 13:815008. doi:10.3389/fmicb.2022.815008.
23. Pinedo-Rivilla C, Aleu J, Durán-Patrón R. Cryptic metabolites from marine-derived microorganisms using OSMAC and epigenetic approaches. Mar Drugs. 2022;20(2):84. doi:10.3390/md20020084.
24. Quan ND, Nguyen NL, Giang TTH, Ngan NTT, Hien NT, Tung NV, et al. Genome characteristics of the endophytic fungus *Talaromyces* sp. DC2 isolated from *Catharanthus roseus* (L.) G. Don. J Fungi. 2024;10(5):352. doi:10.3390/jof10050352.
25. Rana S, Singh SK. Insights into the genomic architecture of a newly discovered endophytic *Fusarium* species belonging to the *Fusarium concolor* complex from India. Front Microbiol. 2023; 14:1266620. doi:10.3389/fmicb.2023.1266620.
26. Robey MT, Caesar LK, Drott MT, Keller NP, Kelleher NL. An atlas of the distribution and conservation of fungal secondary metabolite biosynthetic gene clusters. Proc Natl Acad Sci U S A. 2021;118(19): e2020230118. doi:10.1073/pnas.2020230118.
27. Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH. Biosynthetic gene clusters and the evolution of fungal chemodiversity. Nat Prod Rep. 2020; 37:868-878. doi:10.1039/C9NP00045C.
28. Sagita R, Quax WJ, Haslinger K. Current state and future directions of genetics and genomics of endophytic fungi for bioprospecting efforts. Front Bioeng Biotechnol. 2021; 9:649906. doi:10.3389/fbioe.2021.649906.
29. Shankar A, Sharma KK. Fungal secondary metabolites in food and pharmaceuticals in the era of multi-omics. Appl Microbiol Biotechnol. 2022;106(9-10):3465-3488. doi:10.1007/s00253-022-11945-8.
30. Shenoy PN, Bhaskar S, Manu M, Likitha MP, Geetha N, Sekhar S, et al. Genome mining and antiSMASH analysis of an endophytic *Talaromyces* sp. reveal biosynthetic pathway gene clusters for novel bioactive compounds. J Appl Biol Biotechnol. 2024;12(3):229-235. doi:10.7324/JABB.2024.154874.
31. Tao J, Bai X, Zeng M, Li M, Hu Z, Hua Y, et al. Whole-genome sequence analysis of an endophytic fungus *Alternaria* sp. SPS-2 and its biosynthetic potential of bioactive secondary metabolites. Microorganisms. 2022;10(9):1789. doi:10.3390/microorganisms10091789.
32. Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res. 2023;51(D1): D603-D610. doi:10.1093/nar/gkac1049.
33. Tulsook K, Isarangkul D, Sriubolmas N, Kittakoop P, Wiyakrutta S. Draft genome sequence of *Diaporthe* sp. strain HANT25, an endophytic fungus producing mycoepoxydiene. Microbiol Resour Announc. 2020;9: e00805-20. doi:10.1128/MRA.00805-20.
34. Verma A, Tiwari H, Singh S, Gupta P, Rai N, Singh SK, et al. Epigenetic manipulation for secondary metabolite activation in endophytic fungi: current progress and future directions. Mycology. 2023;14(4):275-291. doi:10.1080/21501203.2023.2241486.
35. Wang D, Jin S, Lu Q, Chen Y. Advances and challenges in CRISPR/Cas-based fungal genome engineering for secondary metabolite production: a review. J Fungi. 2023;9(3):362. doi:10.3390/jof9030362.
36. Wei Q, Bai J, Yan D, Bao X, Li W, Liu B, et al. Genome mining combined metabolic shunting and OSMAC strategy of an endophytic fungus leads to the production of diverse natural products. Acta Pharm Sin B. 2021;11(2):572-587. doi: 10.1016/j.apsb.2020.07.020.
37. Yang F, Wang S, Jiang X, Du H, Liu Y, Zhou Y, et al. Identification and genome insights into *Pyrenochaeta nobilis*, a novel endophytic fungus isolated from *Astragalus membranaceus* with gray mold-control activity. Front Plant Sci. 2025; 16:1610287. doi:10.3389/fpls.2025.1610287.
38. Yuan Y, Shi C, Zhao H. Machine learning-enabled genome mining and bioactivity prediction of natural products. ACS Synth Biol. 2023;12(9):2650-2662. doi:10.1021/acssynbio.3c00234.
39. Zakariyah RF, Ajijolakewu KA, Ayodele AJ, Folami-A BI, Samuel EP, Otuoze SO, et al. Progress in endophytic fungi secondary metabolites: biosynthetic gene cluster reactivation and advances in metabolomics. Bull Natl Res Cent. 2024; 48:44. doi:10.1186/s42269-024-01199-x.
40. Zheng J, Wang H, Wang X, Zeng S, Yuan S, Yin T. Genome mining and OSMAC strategies unveil diverse secondary metabolites from the endophytic fungus *Diaporthe kyushuensis* ZMU-48-1 with antifungal activity. Front Microbiol. 2025; 16:1604639. doi:10.3389/fmicb.2025.1604639.

About the author

Sharmistha Basu is an Assistant Professor in the Department of Botany at Victoria Institution College, Kolkata, West Bengal, India. She is engaged in teaching and academic activities in the field of botany. Her professional interests include plant sciences, botanical studies, higher education, and fostering scientific understanding among students through effective teaching and academic engagement.