

# Activation of Biosynthetic Gene Clusters in Fungi

Subjects: **Chemistry, Applied**

Contributor: Ji Yun Hur , Eunju Jeong , Young Chan Kim , Seoung Rak Lee

In the early 2000s, technological advances in genome sequencing and bioinformatics on filamentous fungi began to reveal a discrepancy between the number of biosynthetic gene clusters (BGCs) encoding the biosynthesis of fungal secondary metabolites and the actual number of identified fungal compounds from the target strain. The discovery of cryptic BGCs in microorganisms, including fungi, has spurred the development of new experimental methodologies for identifying the secondary metabolites of these clusters, which led to the realization that they have the potential to produce novel specialized metabolites, giving rise to a new field of research called genome-guided natural product discovery.

fungi

biosynthetic gene cluster

natural products

## 1. Introduction

Natural products have been recognized as crucial sources for new drug discovery. Over the past 38 years, more than half of the clinical drugs that have been approved by the FDA were derived from natural sources, and natural products still hold promising potential for discovering novel drug candidates and bioactive chemical templates <sup>[1][2]</sup>. Fungi, in particular, offer an incredibly prolific and diverse array of bioactive secondary metabolites, making them an important natural resource for producing unique chemical compounds to combat a variety of diseases <sup>[3][4][5]</sup>. Notably, a multitude of fungal natural products exhibiting various biological effects have been discovered, suggesting that fungi play a role in communicating with other organisms and adapting to different environments <sup>[6][7]</sup>. Some of these identified fungal natural products have already been utilized in the health-functional food, agrochemical, cosmetic, and pharmaceutical industries <sup>[8]</sup>.

In the early 2000s, technological advances in genome sequencing and bioinformatics on filamentous fungi began to reveal a discrepancy between the number of biosynthetic gene clusters (BGCs) encoding the biosynthesis of fungal secondary metabolites and the actual number of identified fungal compounds from the target strain <sup>[9][10]</sup>. This fact suggested that fungi have a great potential for identifying structurally and/or biologically novel secondary metabolites. However, many BGCs are not actively expressed in the normal laboratory growth environment. These are so-called cryptic or silent BGCs <sup>[11][12]</sup>. It is estimated that there are over 5 million fungal species on earth, and each of these species is capable of producing a variety of secondary metabolites, including bioactive compounds, pigments, and toxins <sup>[9][12]</sup>. These secondary metabolites are produced by specialized biosynthetic pathways, which are encoded by clusters of genes known as BGCs. Despite the availability of over 1000 fully sequenced

fungal genomes and the identification of tens of thousands of BGCs, only a small fraction (<3%) of these clusters have been linked to specific secondary metabolites in part because of the cryptic BGCs of fungi [9].

*Neurospora crassa*, a member of the Ascomycota phylum, serves as a model organism for the study of fungal genetics, physiology, and development. It has been widely employed to investigate fundamental processes, such as circadian rhythm and gene regulation [13]. *N. crassa* is known to produce a variety of secondary metabolites, including carotenoids, melanins, and mycotoxin sterigmatocystin [13][14]. The sequencing information of *N. crassa* was completed in 2003 and has been found to contain numerous BGCs, many of which are predicted to encode secondary metabolites [15]. Recently, about 70 BGCs including polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), terpene synthases, and siderophore synthetases were reported from the sequencing data of the fungus [14][15]. However, only a few of BGCs of *N. crassa* have been linked to specific secondary metabolites or characterized in detail. Bioinformatics-based predictions of the chemical structures based on the uncharacterized BGCs suggested that many of them were likely to have novel structures. Experimental characterization of these novel metabolites is often challenging since many BGCs are weakly expressed under laboratory conditions and may require specific environmental cues, growth conditions, and extraction and isolation techniques to induce production [16][17][18].

After the completion of genome sequencing on *N. crassa*, the genomes of many fungi, including those of both Ascomycota and Basidiomycota phyla, have been found to contain numerous cryptic BGCs [19]. *Aspergillus nidulans* is one of the most well-studied secondary metabolite producers. *A. nidulans* has been shown to produce a diverse array of secondary metabolites, including emericellamides, terrain, asperfuranone, fumitremorgins, gliotoxin, and aspernidine A [20]. Several studies have used computational methods to predict the number of BGCs in the *A. nidulans* genome. One such study, published in 2015, identified 52 BGCs in *A. nidulans* using a combination of genome mining and phylogenetic analysis [21]. Another study, published in 2018, identified 63 BGCs in the strain using a similar approach [21].

The discovery of cryptic BGCs in microorganisms, including fungi, has spurred the development of new experimental methodologies for identifying the secondary metabolites of these clusters, which led to the realization that they have the potential to produce novel specialized metabolites, giving rise to a new field of research called genome-guided natural product discovery [22]. Aside from pinpointing the genomics-driven approach, traditional approaches for identifying and characterizing natural products, such as fractionation and purification followed by structural elucidation using techniques such as NMR spectroscopy and mass spectrometry, can be time-consuming and require large amounts of material. To address these challenges, newer approaches such as metabolomics, transcriptomics, and proteomics have been developed to more efficiently identify and characterize natural products from cryptic BGCs [23].

## 2. Organization of Biosynthetic Gene Clusters of Fungi and Their Regulation

Fungi can produce various secondary metabolites with diverse biological activities, such as antibiotics, antifungals, immunosuppressants, and anticancer agents. These secondary metabolites are often encoded by BGCs, which are physically co-localized on the fungal genome and contain all the genes necessary for the biosynthesis of the corresponding secondary metabolite [21][24]. The organization of BGCs in fungi can differ depending on the type of secondary metabolite being produced, but there are some common features. Typically, BGCs are composed of a core set of genes that encode enzymes responsible for the biosynthesis of secondary metabolites, as well as regulatory genes that control gene expression and coordinate the biosynthesis process [24][25]. In many cases, BGCs are found within mobile genetic elements such as transposable elements, plasmids, or integrative and conjugative elements, which can facilitate their transfer between different fungal strains or even different fungal species. The structure of BGCs can also be highly variable, with some BGCs containing only a few genes, while others can harbor dozens of genes that are organized into sub-clusters or modules. These sub-clusters may be responsible for the synthesis of different parts of the secondary metabolite, which are then combined to form the final product [26][27].

Fungal BGCs can be quite large, often exceeding 100 kb in size [19][28]. This fact presents a challenge for researchers who want to study the activity of these gene clusters by expressing them heterologously in a different host organism, such as *E. coli* or yeast. Fungal BGCs are classified based on the type of secondary metabolite they encode, including polyketides, non-ribosomal peptides, terpenoids, saccharides, and ribosomally synthesized and post-translationally modified peptides (RiPPs) [29][30][31]. The organization of BGCs in fungi is highly complex and dynamic, reflecting the diverse functions and ecological roles of the secondary metabolites they produce. Polyketide synthases (PKSs) are a class of enzymes found in fungi and other organisms that are responsible for the biosynthesis of polyketides. PKSs are modular enzymes that utilize a repeating cycle of catalytic domains to assemble complex polyketide chains from simple building blocks, such as acetate and malonate. Each module typically contains several different domains that are responsible for different steps in the biosynthesis process, such as chain initiation, chain elongation, and chain termination. Fungal NRPs utilize a repeating cycle of catalytic domains to assemble complex peptides from simple amino acid building blocks. Each module in NRPs harbors several various domains, which lead to the biosynthesis processes including amino acid activation, amino acid incorporation, and peptide bond formation.

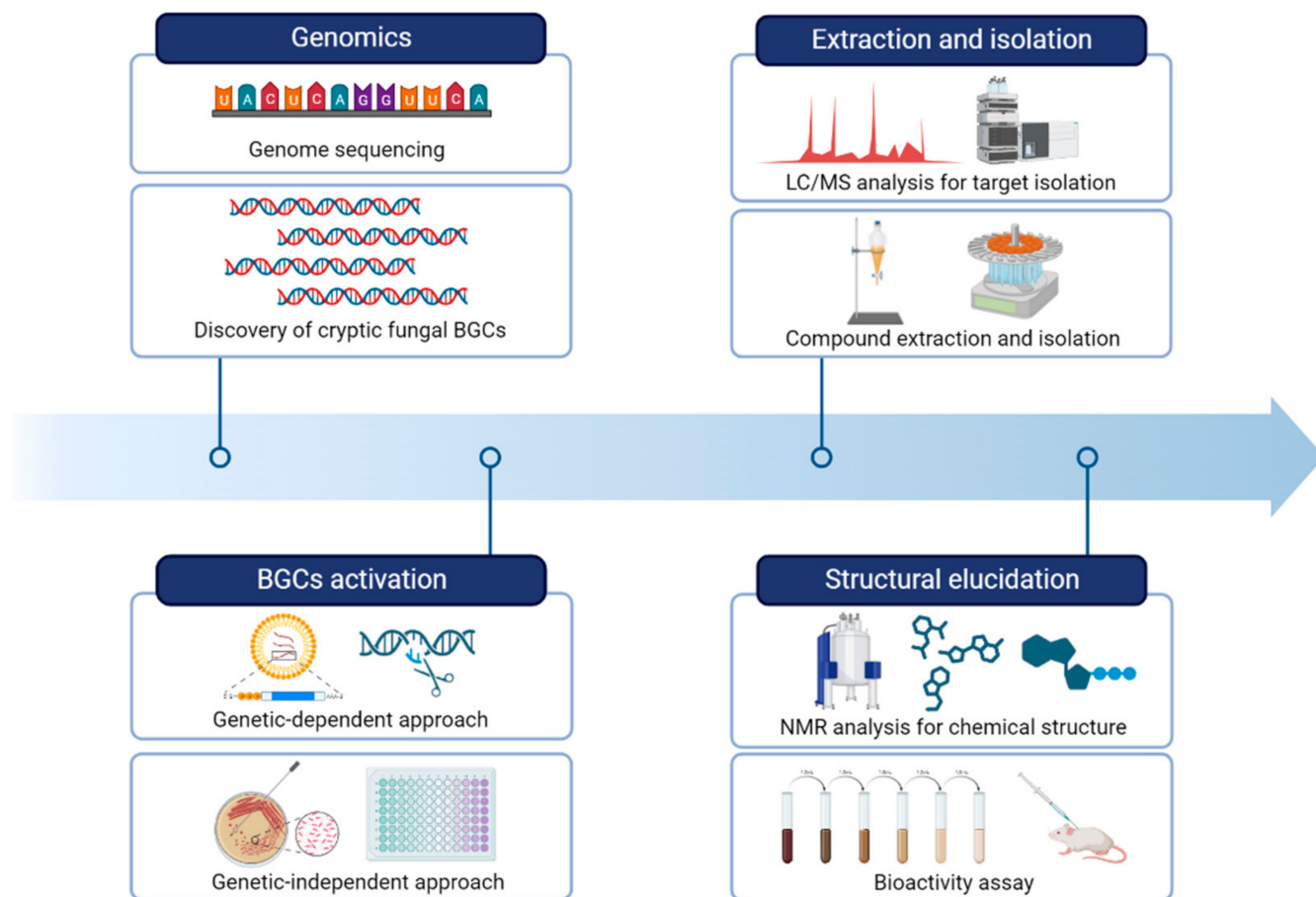
The position of fungal BGCs is usually observed proximal to the telomeres in the genome and often within heterochromatin regions [32]. Heterochromatic regions are generally considered to be silent regions of the genome with low gene density and reduced recombination [32][33]. This may provide a more stable genomic environment for the BGCs, which are often under positive selection due to their role in fungal survival. The expression of BGCs is tightly regulated via a complex interaction of genetic, epigenetic, and environmental factors. The regulation of BGCs is important for ensuring that these clusters are expressed under appropriate conditions and that the products of biosynthesis are synthesized and utilized efficiently. A transcription factor is a protein that can bind to specific DNA sequences and activate or repress gene expression. Many BGCs are controlled by transcription factors that are specific to the biosynthetic pathway and that respond to environmental signals to activate or repress expression [19]. The structure of chromatin has a significant impact on gene expression in fungi. The presence of histone modifications such as methylation, acetylation, and phosphorylation controls the accessibility

of DNA, and therefore the expression of the genes within BGCs [34]. Many BGCs are expressed in response to specific environmental triggers, such as nutrient availability or the presence of competing organisms. These signals affect transcription factors or other regulatory elements that modulate the expression of BGCs. In some cases, BGCs can be acquired through horizontal gene transfer, which involves the transfer of genetic material from one organism to another [35].

Fungi are highly adaptable organisms that live in diverse and complex natural environments, and their growth and metabolism are influenced by a variety of biotic and abiotic factors. However, laboratory growth conditions are usually simple and standardized and may not accurately reflect the physical structure, nutrient availability, and microbial diversity of actual natural environments. Additionally, laboratory conditions may not accurately mimic the environmental stresses that fungi encounter in the wild, such as changes in temperature, pH, osmotic pressure, and competition with other microorganisms. As a result, fungal fermentations in the general laboratory may not accurately represent the full range of metabolic and biosynthetic capabilities that fungi exhibit in their natural habitats [36]. Therefore, understanding the regulatory processes that modulate the growth, metabolism, and biosynthetic capabilities of fungi is critical for unlocking their full potential as sources of bioactive compounds. This requires a multidisciplinary approach that combines microbiology, biochemistry, and genetics to fully understand the complex regulation of fungi.

### 3. Characterization of Biosynthetic Gene Clusters and Natural Product Discovery

Next-generation sequencing (NGS) technologies have revolutionized the field of genomics by allowing rapid and cost-effective acquisition of genomic data. The rapid pace of technological advancements in NGS has led to an exponential increase in the amount of fungal genomic data generated, which has in turn fueled the development of new analytical tools and computational approaches to handle and analyze these data [27][37]. It is now common for researchers to identify BGCs responsible for the production of fungal secondary metabolites. By obtaining a draft genome sequence of fungi, researchers utilize a variety of bioinformatic tools to identify and analyze potential BGCs involved in the biosynthesis of a particular compound of interest (**Figure 1**).



**Figure 1.** A workflow of strategies for natural product research by activating fungal BGCs.

Once a draft genome sequence has been acquired, researchers start to use tools such as antiSMASH and MIBiG to identify potential BGCs within the genome [38][39]. These tools analyze the genome sequence for specific gene clusters known to be involved in the biosynthesis of secondary metabolites, such as polyketides, non-ribosomal peptides, and terpenes. By comparing the identified BGCs to known BGCs in databases, researchers can predict the structure and function of the secondary metabolites. antiSMASH is one of the most widely used BGC detection tools [38]. It is a web-based tool that allows users to input draft genome sequences and predict the location of BGCs in the genome. The tool uses a variety of algorithms to identify BGCs, including hidden Markov models (HMMs), Pfam domains, and Clusters of Orthologous Groups (COGs). The tool also provides annotations of the predicted BGCs, including predictions of the chemical structure of the metabolite produced. Alternative tools for BGCs identification are PRISM and BAGEL [40][41]. PRISM uses a machine learning algorithm to predict BGCs in microbial genomes. It integrates multiple data sources, including gene co-occurrence patterns, gene expression data, and functional annotations, to identify BGCs. PRISM also includes tools for visualizing and exploring predicted BGCs, including interactive network visualizations. BAGEL is a tool specifically designed for the identification of bacteriocin gene clusters, which are BGCs involved in the biosynthesis of antimicrobial peptides produced by bacteria. BAGEL uses a combination of HMMs and machine learning algorithms to predict bacteriocin gene clusters in microbial genomes.

When the above tools are applied to the genome sequence of a specific fungal strain, they are expected to identify cryptic BGCs that might be responsible for the production of unknown compounds and would be a promising starting point for natural product discovery. The type of NGS technology utilized impacts the capability to characterize the full complement of BGCs in a particular genome [42]. Different NGS technologies have different strengths and limitations, and the choice of technology will depend on factors such as the size and complexity of the genome, the sequencing depth needed, and the availability of bioinformatics tools and resources [43]. Some NGS technologies, such as PacBio and Oxford Nanopore, can generate long reads that can span entire BGCs and provide more complete sequence information than short-read technologies such as Illumina [44]. In addition to the choice of NGS technology, the ability to identify the full complement of BGCs in a particular genome also depends on the quality of the genome assembly, the bioinformatics tools used for BGC identification, and the expertise of the researchers involved [43][44]. Genome assembly is a critical step in NGS-based approaches for BGC identification, as errors or gaps in the assembly can lead to missed or incomplete BGCs.

Although NGS-based approaches have been utilized to identify the full complement of putative specialized metabolite BGCs, there are limitations to the application of natural product discovery. It is still challenging to predict the chemical structures of the metabolites produced by these BGCs based only on genomic and bioinformatic information [45][46]. The identification of BGCs is just the first step in the process of discovering and characterizing natural products. After a specific BGC is identified, it is necessary to express and characterize the genes involved in the biosynthesis of the secondary metabolite, to produce and purify the metabolite itself, and to test its relevant biological activity. These steps require significant resources and expertise, and may not be feasible for all BGCs identified from genomic data (**Figure 1**).

---

## References

1. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 2020, 83, 770–803.
2. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 2016, 79, 629–661.
3. Helaly, S.E.; Thongbai, B.; Stadler, M. Diversity of biologically active secondary metabolites from endophytic and saprotrophic fungi of the ascomycete order Xylariales. *Nat. Prod. Rep.* 2018, 35, 992–1014.
4. Demain, A.L. Valuable secondary metabolites from fungi. In *Biosynthesis and Molecular Genetics of Fungal Secondary Metabolites*; Springer: New York, NY, USA, 2014; pp. 1–15.
5. Bills, G.F.; Gloer, J.B. Biologically active secondary metabolites from the fungi. *Microbiol. Spectr.* 2017, 4, 1087–1119.

6. Atanasov, A.G.; Zotchev, S.B.; Dirsch, V.M.; Supuran, C.T. Natural products in drug discovery: Advances and opportunities. *Nat. Rev. Drug Discov.* 2021, 20, 200–216.
7. Fischer, M.S.; Glass, N.L. Communicate and fuse: How filamentous fungi establish and maintain an interconnected mycelial network. *Front. Microbiol.* 2019, 10, 619.
8. Dias, D.A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* 2012, 2, 303–336.
9. Robey, M.T.; Caesar, L.K.; Drott, M.T.; Keller, N.P.; Kelleher, N.L. An interpreted atlas of biosynthetic gene clusters from 1000 fungal genomes. *Proc. Natl. Acad. Sci. USA* 2021, 118, e2020230118.
10. Clevenger, K.D.; Bok, J.W.; Ye, R.; Miley, G.P.; Verdan, M.H.; Velk, T.; Chen, C.; Yang, K.; Robey, M.T.; Gao, P. A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nat. Chem. Biol.* 2017, 13, 895–901.
11. Hoskisson, P.A.; Seipke, R.F. Cryptic or silent? The known unknowns, unknown knowns, and unknown unknowns of secondary metabolism. *MBio* 2020, 11, e02642-20.
12. Amos, G.C.; Awakawa, T.; Tuttle, R.N.; Letzel, A.-C.; Kim, M.C.; Kudo, Y.; Fenical, W.; Moore, B.S.; Jensen, P.R. Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. USA* 2017, 114, E11121–E11130.
13. Honda, S.; Eusebio-Cope, A.; Miyashita, S.; Yokoyama, A.; Aulia, A.; Shahi, S.; Kondo, H.; Suzuki, N. Establishment of *Neurospora crassa* as a model organism for fungal virology. *Nat. Commun.* 2020, 11, 5627.
14. Keller, N.P. Fungal secondary metabolism: Regulation, function and drug discovery. *Nat. Rev. Microbiol.* 2019, 17, 167–180.
15. Dunlap, J.C.; Borkovich, K.A.; Henn, M.R.; Turner, G.E.; Sachs, M.S.; Glass, N.L.; McCluskey, K.; Plamann, M.; Galagan, J.E.; Birren, B.W. Enabling a community to dissect an organism: Overview of the *Neurospora* functional genomics project. *Adv. Genet.* 2007, 57, 49–96.
16. Liu, Z.; Lin, Z.; Nielsen, J. Expression of fungal biosynthetic gene clusters in *S. cerevisiae* for natural product discovery. *Synth. Syst. Biotechnol.* 2021, 6, 20–22.
17. Reen, F.J.; Romano, S.; Dobson, A.D.; O’Gara, F. The sound of silence: Activating silent biosynthetic gene clusters in marine microorganisms. *Mar. Drugs* 2015, 13, 4754–4783.
18. Almeida, H.; Tsang, A.; Diallo, A.B. Improving candidate Biosynthetic Gene Clusters in fungi through reinforcement learning. *Bioinformatics* 2022, 38, 3984–3991.
19. Mózsik, L.; Iacovelli, R.; Bovenberg, R.A.; Driessen, A.J. Transcriptional activation of biosynthetic gene clusters in filamentous fungi. *Front. Bioeng. Biotechnol.* 2022, 10, 1199.

20. Chiang, Y.-M.; Szewczyk, E.; Nayak, T.; Davidson, A.D.; Sanchez, J.F.; Lo, H.-C.; Ho, W.-Y.; Simityan, H.; Kuo, E.; Praseuth, A. Molecular genetic mining of the *Aspergillus* secondary metabolome: Discovery of the emericellamide biosynthetic pathway. *Chem. Biol.* 2008, 15, 527–532.
21. Drott, M.; Bastos, R.; Rokas, A.; Ries, L.; Gabaldón, T.; Goldman, G.; Keller, N.; Greco, C. Diversity of secondary metabolism in *Aspergillus nidulans* clinical isolates. *MSphere* 2020, 5, e00156-20.
22. Scherlach, K.; Hertweck, C. Mining and unearthing hidden biosynthetic potential. *Nat Commun.* 2021, 12, 3864.
23. Machado, H.; Tuttle, R.N.; Jensen, P.R. Omics-based natural product discovery and the lexicon of genome mining. *Curr. Opin. Microbiol.* 2017, 39, 136–142.
24. Kwon, M.J.; Steiniger, C.; Cairns, T.C.; Wisecaver, J.H.; Lind, A.L.; Pohl, C.; Regner, C.; Rokas, A.; Meyer, V. Beyond the biosynthetic gene cluster paradigm: Genome-wide coexpression networks connect clustered and unclustered transcription factors to secondary metabolic pathways. *Microbiol. Spectr.* 2021, 9, e00898-21.
25. Medema, M.H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M.A. A systematic computational analysis of biosynthetic gene cluster evolution: Lessons for engineering biosynthesis. *PLoS Comput. Biol.* 2014, 10, e1004016.
26. Calvo, A.M.; Wilson, R.A.; Bok, J.W.; Keller, N.P. Relationship between secondary metabolism and fungal development. *Microbiol. Mol. Biol. Rev.* 2002, 66, 447–459.
27. Manzoni, M.; Rollini, M. Biosynthesis and biotechnological production of statins by filamentous fungi and application of these cholesterol-lowering drugs. *Appl. Microbiol. Biotechnol.* 2002, 58, 555–564.
28. Gluck-Thaler, E.; Haridas, S.; Binder, M.; Grigoriev, I.V.; Crous, P.W.; Spatafora, J.W.; Bushley, K.; Slot, J.C. The architecture of metabolism maximizes biosynthetic diversity in the largest class of fungi. *Mol. Biol. Evol.* 2020, 37, 2838–2856.
29. Arnison, P.G.; Bibb, M.J.; Bierbaum, G.; Bowers, A.A.; Bugni, T.S.; Bulaj, G.; Camarero, J.A.; Campopiano, D.J.; Challis, G.L.; Clardy, J. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 2013, 30, 108–160.
30. Le Govic, Y.; Papon, N.; Le Gal, S.; Bouchara, J.-P.; Vandeputte, P. Non-ribosomal peptide synthetase gene clusters in the human pathogenic fungus *Scedosporium apiospermum*. *Front. Microbiol.* 2019, 10, 2062.
31. Cox, R.J.; Simpson, T.J. Fungal type I polyketide synthases. *Methods Enzymol.* 2009, 459, 49–78.



32. Palmer, J.M.; Keller, N.P. Secondary metabolism in fungi: Does chromosomal location matter? *Curr. Opin. Microbiol.* 2010, 13, 431–436.
33. Tamaru, H. Confining euchromatin/heterochromatin territory: Jumonji crosses the line. *Genes Dev.* 2010, 24, 1465–1478.
34. Lai, Y.; Wang, L.; Zheng, W.; Wang, S. Regulatory roles of histone modifications in filamentous fungal pathogens. *J. Fungi* 2022, 8, 565.
35. Tran, P.N.; Yen, M.-R.; Chiang, C.-Y.; Lin, H.-C.; Chen, P.-Y. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl. Microbiol. Biotechnol.* 2019, 103, 3277–3287.
36. Jouhten, P.; Ponomarova, O.; Gonzalez, R.; Patil, K.R. *Saccharomyces cerevisiae* metabolism in ecological context. *FEMS Yeast Res.* 2016, 16, fow080.
37. Salem-Bango, Z.; Price, T.K.; Chan, J.L.; Chandrasekaran, S.; Garner, O.B.; Yang, S. Fungal whole-genome sequencing for species identification: From test development to clinical utilization. *J. Fungi* 2023, 9, 183.
38. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; Van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 2021, 49, W29–W35.
39. Kautsar, S.A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J.C.; Terlouw, B.R.; Van Der Hooft, J.J.; Van Santen, J.A.; Tracanna, V.; Suarez Duran, H.G.; Pascal Andreu, V. MIBiG 2.0: A repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 2020, 48, D454–D458.
40. Russell, A.H.; Truman, A.W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput. Struct. Biotechnol. J.* 2020, 18, 1838–1851.
41. Alam, K.; Hao, J.; Zhong, L.; Fan, G.; Ouyang, Q.; Islam, M.; Islam, S.; Sun, H.; Zhang, Y.; Li, R. Complete genome sequencing and in silico genome mining reveal the promising metabolic potential in *Streptomyces* strain CS-7. *Front. Microbiol.* 2022, 13, 3751.
42. Koboldt, D.C.; Steinberg, K.M.; Larson, D.E.; Wilson, R.K.; Mardis, E.R. The next-generation sequencing revolution and its impact on genomics. *Cell* 2013, 155, 27–38.
43. Slatko, B.E.; Gardner, A.F.; Ausubel, F.M. Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* 2018, 122, e59.
44. Amarasinghe, S.L.; Su, S.; Dong, X.; Zappia, L.; Ritchie, M.E.; Gouil, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 2020, 21, 30.
45. Louwen, J.J.; Medema, M.H.; van der Hooft, J.J. Enhanced correlation-based linking of biosynthetic gene clusters to their metabolic products through chemical class matching. *Microbiome* 2023, 11, 13.

46. Van Der Hooft, J.J.; Mohimani, H.; Bauermeister, A.; Dorrestein, P.C.; Duncan, K.R.; Medema, M.H. Linking genomics and metabolomics to chart specialized metabolic diversity. *Chem. Soc. Rev.* 2020, 49, 3297–3314.
- 

Retrieved from <https://encyclopedia.pub/entry/history/show/101936>