Unveiling Novel Secondary Metabolite Gene Clusters in a *Paecilomyces* sp. Strain From The Brazilian Amazon Through Computer Aided Genomic Analysis

Daniel V. de Oliveira[1,2], Thiago F. Sousa[2], Sabrina S. P. Sousa[2], Miguel V. de Oliveira[3], Gleucinei dos S. Castro[2], Maria Aparecida de Jesus[4], Gilvan F. da Silva[2]

[1]Instituto de Ciências Biológicas – Universidade Federal do Amazonas (UFAM)
Av. Rodrigo Otávio, 6200, Setor Sul do Campus Universitário
Manaus, Amazonas, Brasil

[2]Embrapa Amazônia Ocidental
Manaus, Amazonas, Brasil

[3]Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Av. Rodrigo Otávio, 6200, Setor Norte do Campus Universitário
Manaus, Amazonas, Brasil

[4]Instituto de Pesquisa da Amazônia – INPA
Av. André Araújo, 2936
Manaus, Amazonas, Brasil

daniel.oliveira@ufam.edu.br, fernandesthiago9620@gmail.com,
sabrynasinara@gmail.com, gleucineii@gmail.com,
miguel.oliveira@icomp.ufam.edu.br, ranhna@gmail.com,
gilvan.silva@embrapa.br

Abstract. *The genus Paecilomyces represents a promising source of bioactive secondary metabolites, yet it remains largely underexplored from chemical and genomic perspectives. To address this, a comprehensive genome mining approach was employed to characterize a novel Paecilomyces sp. strain, CMI-INPA 1390, isolated from the Brazilian Amazon. Utilizing long-read sequencing, a high-quality, near-chromosome level genome assembly was achieved, comprising six linear scaffolds totaling 31.7 Mb and a circular mitochondrial genome of 52094 bp, with a high BUSCO completeness of 98.7%. Genomic analysis identified 42 biosynthetic gene clusters (BGCs), which is consistent with the genus average. Notably, 100% of these BGCs showed low to no correlation with known biosynthetic pathways, revealing the high potential of this species for the biosynthesis of novel molecules. Gene Cluster Family (GCF) analysis, conducted with 10 reference genomes for the Paecilomyces genus, revealed that among the NRPS-related GCFs, this new Amazonian species presents a singleton associated with nidulanin A biosynthesis. This singleton contained all necessary tailoring genes, but the core nonribosomal peptide synthetase (NRPS) gene exhibited a loss of several modules and domains, retaining only a single module out of the four canonical ones related to nidulanin A biosynthesis. This loss suggests a potential evolutionary divergence, possibly resulting in the production of a shorter or structurally simplified molecule. These findings*

underscore that Paecilomyces sp. CMI-INPA 1390 is a promising source of diverse and potentially novel secondary metabolites and demonstrate the role of genome mining in unveiling the full biosynthetic potential of filamentous fungi.

1. Introduction

The genus *Paecilomyces* encompasses a diverse group of filamentous fungi with significant ecological and biotechnological importance. Known for their wide ecological distribution, these fungi thrive in various niches, exhibiting saprophytic, endophytic, and pathogenic lifestyles [Moreno-Gavíra et al. 2020]. This versatility is underpinned by their remarkable metabolic capabilities, particularly the production of bioactive secondary metabolites. These compounds play crucial roles in their ecological interactions and serve as a rich source for diverse applications in medicine, agriculture, and industry. Genomic studies of *Paecilomyces* species reveal genomes typically ranging from 27 to 36 Mb [Visagie et al. 2024], reflecting their considerable genetic diversity and harboring numerous biosynthetic gene clusters (BGCs). These BGCs encode the enzymatic machinery responsible for synthesizing a vast array of natural products [Wang et al. 2020]. However, many BGCs often remain transcriptionally silent under standard laboratory conditions, making their metabolic products undetectable through conventional chemical analysis [Mózsik et al. 2022]. To overcome this limitation, genome mining, empowered by advanced bioinformatics tools, has emerged as a powerful and cost-effective strategy. This approach enables the prediction and characterization of BGCs directly from genomic data, accelerating the discovery pipeline [Albarano et al. 2020, Fedorova et al. 2012]. In this study, we employed a comprehensive genome mining approach to analyze a novel *Paecilomyces sp*. strain, CMI-INPA 1390. This strain was isolated from decaying hardwood samples collected in the Adolpho Ducke Forest Reserve in Manaus, Amazonas, Brazil, and is currently undergoing taxonomic description. Our primary focus was to characterize the biosynthetic potential of this Amazonian isolate through a detailed genomic analysis.

2. Materials and Methods

Strain Isolation and Cultivation *Paecilomyces sp*. CMI-INPA 1390 was isolated from a decaying wood trunk in the *Reserva Florestal Adolpho Ducke* (RFAD), Brazil. Monosporic cultures were maintained on Malt Extract Agar (MEA) at 28 °C for 48h. For DNA extraction, liquid cultures were grown in Malt Extract Broth (MEB) at 28 °C with 180 rpm shaking for 7 days.

Genome Sequencing and Assembly Genomic DNA was extracted from mycelia using the CTAB method [Doyle and Doyle 1987] and quantified (Qubit, Nanodrop). Long-read sequencing was performed on the Oxford Nanopore PromethION 2 Solo (P2 Solo) platform using R10.4.1 Flow Cells and the Rapid Sequencing Kit (SQK-RBK114.96). Basecalling was performed using Guppy (v3.0). De novo assembly was carried out using Flye (v2.9.6) [Kolmogorov et al. 2020] with default parameters.

Genome Quality Assessment and Annotation Genome completeness and duplication (an indicator of potential contamination or heterozygosity) were assessed with BUSCO (v6.0.0) [Manni et al. 2021] against the *eurotiales_odb12* dataset. Gene prediction was performed using Augustus (v3.3.3) [Stanke et al. 2006] with *Aspergillus nidulans* as the training model. Functional annotation utilized BLASTp [Altschul et al. 1990] against NCBI NR and HMMER (v3.4) [Eddy 2011] against Pfam (v37.4) [Paysan-Lafosse et al. 2025].

Secondary Metabolite Gene Cluster (BGC) Analysis BGCs were predicted and characterized using FungiSMASH (v8.0.2) [Blin et al. 2025] with all functions activated and 'loose' strictness. Gene cluster families (GCFs) were defined using BiG-SCAPE (v2.0.0) [Navarro-Muñoz et al. 2020] against MIBiG (v3.1) [Terlouw et al. 2023].

3. Results and Discussion

3.1. Assembly Quality

The whole-genome assembly of the *Paecilomyces sp.* CMI-INPA 1390 resulted in a highly contiguous sequence, consisting of six linear scaffolds and one circular scaffold corresponding to the mitochondrial DNA, for a total of 31,690,111 base pairs, with a mitochondrial genome size of 52 kb. The assembly exhibited an N50 of approximately 5 Mb and a total percentage of gaps of 0.0%. Further supporting the high quality and completeness of this assembly, BUSCO analysis revealed 98.7% completeness, with 4,310 out of 4,365 expected BUSCO genes found as complete (4,301 single-copy and 9 duplicated). Only 6 BUSCO genes were fragmented and 49 were missing. This high level of contiguity and completeness suggests a near-chromosome level assembly. Furthermore, the obtained six scaffolds are highly compatible with previously inferred chromosome numbers in *Paecilomyces*. Analyses of complete genomes of *Paecilomyces* species deposited in the NCBI database consistently show chromosome numbers varying from 6 to 10 for those with chromosome-level assemblies. To our knowledge, the only study reporting chromosome number estimations for *Paecilomyces* is [Inglis et al. 2005], which performed telomeric fingerprinting and estimated seven chromosomes for *Paecilomyces lilacinus* and between six and nine for *Paecilomyces fumosoroseus*. However, these taxa have since been reclassified (*P. lilacinus* as *Purpureocillium lilacinum* and *P. fumosoroseus* as *Cordyceps fumosorosea*), meaning no PFGE-based or related approach are available for chromosome number determination in true *Paecilomyces* species.

3.2. Secondary Metabolite Gene Cluster Analysis

In totality, 10 *Paecilomyces* genomes were compared, including 9 strains containing reference genomes deposited in the NCBI genome database and the novel focus strain CMI-INPA 1390. The number of regions containing biosynthetic gene clusters range from 32 to 66 in the reference *Paecilomyces spp.* genomes, with an average of 41.5 regions. The analysis presented 42 regions, fitting well within the expected range and average.

Analysis based on number of BGCs across *Paecilomyces spp.* revealed an average of 25 BGC classes containing an average of 17.52 representatives each. The strain on which this work is focused, 1390, showed 12 BGC classes with an average of 3.58 representatives each. The majority of secondary metabolites predicted were classified as terpene and T1PKS. Suggesting that strain 1390 is a promising source of diverse terpenoids and polyketides, which are secondary metabolites with a wide range of biological activities.

BiG-SCAPE analysis identified a singleton cluster containing all tailoring genes involved in nidulanin A biosynthesis; however, the NRPS core gene appears largely inactive, having lost the N-terminal portion corresponding to the first 3,819 amino acids and retaining only the C-terminal 2,340 amino acids, which include a single complete NRPS module. In the canonical nidulanin A pathway, the NRPS contains four complete modules

with four adenylation domains, each responsible for incorporating one amino acid into the final product. This loss of modules suggests that the pathway may no longer produce the full-length peptide, potentially resulting in a shorter scaffold or a structurally simplified analog of nidulanin A as shown in figure 1C. In *Aspergillus puulaauensis* MK2, the upstream and downstream regions of the cluster are highly conserved compared to *Paecilomyces sp.* CMI-INPA 1390. However, while the core NRPS gene in *A. puulaauensis* appears intact and of the expected size, in CMI-INPA 1390 it is altered and retains only part of the adenylation and condensation domains . This contrast highlights a possible evolutionary divergence in NRPS functionality, where the *Paecilomyces* strain may have lost modules or domains while conserving the surrounding tailoring genes, suggesting either a repurposing of the pathway as can be seen in figure 1B.

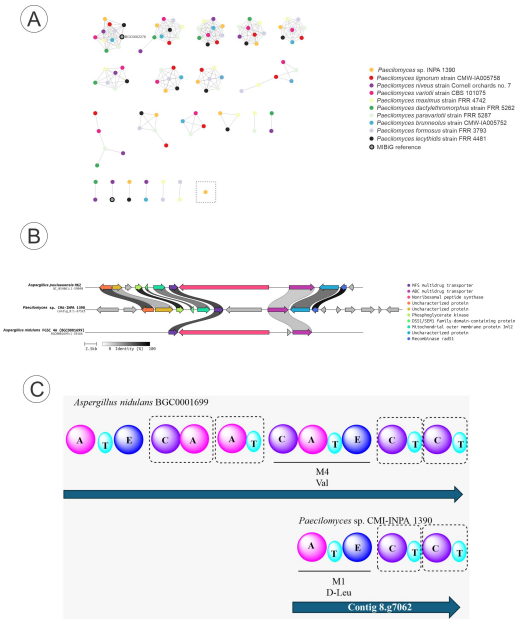

Figure 1. Figure 1. Comparative analyses of Nonribosomal Peptide Synthetase (NRPS) gene cluster families and the nidulanin A biosynthesis pathway. (A) Gene Cluster Family (GCF) Network of NRPS BGCs in Paecilomyces species. Notably, the biosynthetic gene cluster (BGC) associated with the nidulanin A biosynthesis pathway in *Paecilomyces* sp. CMI-INPA 1390 forms a unique singleton GCF, highlighted within the dashed box, indicating its distinct evolutionary trajectory compared to other characterized clusters within the genus.(B) Synteny comparison focusing on gene organization and composition of NRPS biosynthetic regions. This panel presents a detailed comparison of the gene organization and composition within the NRPS biosynthetic regions of *Paecilomyces* sp. CMI-INPA 1390, *Aspergillus puulaauensis* MK2, and the MIBiG reference cluster BGC0001699 (related to nidulanin A). It highlights the conserved order of genes and structural elements, providing insights into the evolutionary relationships and potential functional divergences among these pathways.(C) Structural comparison of the core nidulanin A NRPS gene, highlighting the loss of domains/modules. This panel provides a structural comparison of the core Nonribosomal Peptide Synthetase (NRPS) gene involved in nidulanin A biosynthesis. It contrasts the canonical and complete NRPS structure (typically associated with MIBiG BGC0001699 and production in Aspergillus nidulans) with the modified NRPS gene found in *Paecilomyces* sp. CMI-INPA 1390. The *Paecilomyces* NRPS exhibits a significant loss of multiple domains and modules compared to the four canonical modules, although it retains a single complete module. This structural alteration suggests

an evolutionary divergence, which likely leads to the production of a shorter or structurally simplified nidulanin A analog. Key NRPS domains are indicated: adenylation (A), thiolation (T), condensation (C), and epimerization (E). For instance, while the canonical *A. nidulans* pathway incorporates Valine (Val) at M4, the altered assembly line in *Paecilomyces* sp. incorporates Leucine (Leu) at M1, demonstrating functional changes resulting from the modified architecture.

4. Conclusions

This comprehensive genomic analysis of *Paecilomyces sp.* CMI-INPA 1390 has yielded several key conclusions that contribute significantly to the fields of mycology and natural product discovery. The successful generation of a high-quality, near-chromosome level genome assembly, a valuable resource that provides a robust foundation for future studies on this novel Amazonian isolate. This high-quality data provides confidence that the subsequent BGC analysis is both accurate and comprehensive. The study establishes that while the total number of BGCs in the CMI-INPA 1390 strain falls within the average range for the genus, the composition of these clusters is unique. The identification of a lower diversity of gene families, with a particular prevalence of terpene and polyketide-related clusters, suggests a specialized biosynthetic repertoire with a high potential for yielding novel chemical structures. This finding highlights that a species' true biosynthetic value may not be found in the total number of BGCs, but in their unique composition and evolutionary state. Finally, the discovery and characterization of the modified NRPS gene, which has lost most of its functional modules while retaining its tailoring genes, represents a powerful illustration of how genome mining can unveil novel and evolutionarily significant biosynthetic pathways. This finding suggests that the organism has found a way to repurpose an existing pathway to potentially generate structurally simplified analogs of known compounds, which could possess new or improved biological activities and functions. This demonstrates the effectiveness of integrating advanced sequencing and bioinformatics to overcome the limitations of traditional cultivation-based screening methods. The work provides a clear case for why a deeper exploration of underexplored environments, such as the Brazilian Amazon, is crucial for identifying novel genetic and chemical resources. In summary, this *Paecilomyces sp.* CMI-INPA 1390 strain is a promising subject for future research, confirming that genomic analysis is an indispensable tool for unlocking the full biosynthetic potential of filamentous fungi.

References

Albarano, L., Esposito, R., Ruocco, N., and Costantini, M. (2020). Genome mining as new challenge in natural products discovery. *Marine drugs*, 18(4):199.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410.

Blin, K., Shaw, S., Vader, L., Szenei, J., Reitz, Z. L., Augustijn, H. E., Cediel-Becerra, J. D., de Crécy-Lagard, V., Koetsier, R. A., Williams, S. E., et al. (2025). antismash 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation¡? mode pagerangestyle?¿. *Nucleic acids research*, page gkaf334.

Doyle, J. J. and Doyle, J. L. (1987). A rapid dna isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin*.

Eddy, S. R. (2011). Accelerated profile hmm searches. *PLoS computational biology*, 7(10):e1002195.

Fedorova, N. D., Moktali, V., and Medema, M. H. (2012). Bioinformatics approaches and software for detection of secondary metabolic gene clusters. In *Fungal secondary metabolism: methods and protocols*, pages 23–45. Springer.

Inglis, P. W., Sarmento, R. B., Gavião, C. F., and Valadares-Inglis, M. C. (2005). Dna fingerprinting of paecilomyces strains of potential use for the biological control of pests. *World Journal of Microbiology and Biotechnology*, 21(8):1487–1492.

Kolmogorov, M., Bickhart, D. M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T. P., et al. (2020). metaflye: scalable long-read metagenome assembly using repeat graphs. *Nature methods*, 17(11):1103–1110.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). Busco update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution*, 38(10):4647–4654.

Moreno-Gavíra, A., Huertas, V., Diánez, F., Sánchez-Montesinos, B., and Santos, M. (2020). Paecilomyces and its importance in the biological control of agricultural pests and diseases. *Plants*, 9(12):1746.

Mózsik, L., Iacovelli, R., Bovenberg, R. A., and Driessen, A. J. (2022). Transcriptional activation of biosynthetic gene clusters in filamentous fungi. *Frontiers in Bioengineering and Biotechnology*, 10:901037.

Navarro-Muñoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., De Los Santos, E. L., Yeong, M., Cruz-Morales, P., Abubucker, S., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nature chemical biology*, 16(1):60–68.

Paysan-Lafosse, T., Andreeva, A., Blum, M., Chuguransky, S. R., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Llinares-López, F., Meng-Papaxanthos, L., et al. (2025). The pfam protein families database: embracing ai/ml. *Nucleic acids research*, 53(D1):D523–D534.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). Augustus: ab initio prediction of alternative transcripts. *Nucleic acids research*, 34(suppl_2):W435–W439.

Terlouw, B. R., Blin, K., Navarro-Munoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., Lee, S., Meijer, D., Recchia, M. J., Reitz, Z. L., et al. (2023). Mibig 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic acids research*, 51(D1):D603–D610.

Visagie, C. M., Cruywagen, E. M., and Duong, T. A. (2024). A new paecilomyces from wooden utility poles in south africa. *Fungal Systematics and Evolution*, 13(1):163–181.

Wang, X., Peng, J., Sun, L., Bonito, G., Guo, Y., Li, Y., and Fu, Y. (2020). Genome sequencing of paecilomyces penicillatus provides insights into its phylogenetic placement and mycoparasitism mechanisms on morel mushrooms. *Pathogens*, 9(10):834.