



## **Previsão *in silico* e caracterização de aglomerados de genes biossintéticos em uma linhagem hidrolítica de *Trichoderma harzianum***

**Palavras-Chave:** Genômica; Metabólitos secundários; Aglomerados de genes biossintéticos; Bioinformática; *Trichoderma harzianum*

**Autores(as):**

**Fabricio Aparecido Ruela, FCF – UNICAMP**

**Rafaela Rossi Rosolen, IB – UNICAMP**

**Profa. Dra. Anete Pereira de Sousa, IB – UNICAMP**

---

### **INTRODUÇÃO:**

Os fungos filamentosos são conhecidos por sua capacidade de produzir uma vasta gama de compostos químicos distintos, também conhecidos como metabólitos secundários (MSs) (Keller, 2019). Tais compostos são derivados de vias metabólicas centrais e *pools* de metabólitos primários, onde acil-CoA é o bloco de construção inicial para a síntese de MSs de policetídeos (por exemplo, aflatoxina) e terpenos (por exemplo, caroteno), enquanto os aminoácidos são utilizados para a síntese de MSs de peptídeos não ribossomais (por exemplo, penicilina) (Keller, 2019). Em contraste com os genes necessários para a síntese de um metabólito primário, os quais estão dispersos por todo o genoma fúngico, os genes que codificam as atividades enzimáticas para produzir os MSs são arranjados de forma contígua como um agrupamento de genes biossintéticos (BGC) (Rokas et al., 2020). Comumente, os BGCs codificam as enzimas que são responsáveis pela biossíntese do esqueleto do metabólito, incluindo peptídeos sintetases não ribossomais (NRPS), sintases de policetídeos (PKS) e fusões de PKS e NRPS (híbridos PKS-NRPS/NRPS-PKS), bem como outras enzimas envolvidas em outras modificações do esqueleto do metabólito (Keller, 2019).

Os fungos do gênero *Trichoderma* (Hypocreales, Pezizomycotina, Ascomycota) apresentam alta versatilidade nutricional e ambiental, além de abranger espécies com uma ampla gama de estilos de vida (Kubicek et al., 2019). *T. harzianum* CBMAI-0179 (Th0179) é uma linhagem originária da Coleção Brasileira de Microrganismos Ambientais e Industriais (CBMAI), a qual está localizada no Centro de Pesquisas Pluridisciplinares Químicas, Biológicas e Agropecuárias (CPQBA) da Universidade Estadual de Campinas (UNICAMP), Brasil. Recentemente, foi realizado o sequenciamento, montagem e anotação funcional do genoma de Th0179.

Diante do exposto, **neste projeto de IC investigamos e caracterizamos os BGCs ao longo do genoma de Th0179. As análises genômicas foram complementadas com dados de transcriptômica para avaliar o perfil de expressão gênica dos MSs em condições de degradação de celulose e glicose. Além disso, análises filogenéticas de enzimas e proteínas chaves foram realizadas para melhor entender as características evolutivas associadas a produção de MSs dentro o gênero *Trichoderma*, incluindo as linhagens Th3844, Th0711 e Ta0020. É válido ressaltar que este relatório final compreende o período da vigência da bolsa PIBIC/CNPq, iniciada em março/2024**

e finalizada em agosto/2024.

## **METODOLOGIA:**

### **Identificação molecular de Th0179**

A linhagem Th0179 foi obtida da CBMAI, a qual está localizada no CPQBA da UNICAMP. A identidade dos isolados de *Trichoderma* foi autenticada por tal instituição com base em estudos filogenéticos de *internal transcribed spacer (ITS) region*, *translational elongation factor 1 (tef1)* e *RNA polymerase II (rpb2)* (Rosolen et al., 2023).

### **Previsão de BGCs de MSs no genoma de Th0179**

O genoma de Th0179 foi recuperado do NCBI através do número BioProject PRJNA781962 e do número BioSample SAMN23309298 (Rosolen et al., 2023). Em seguida, os BGCs foram previstos pela busca de genes que codificam enzimas de *backbone* e outros domínios de proteínas associados a agrupamentos usando o programa antiSMASH v.6.0 (Blin et al., 2021), o qual é, atualmente, a ferramenta mais utilizada para detectar e caracterizar BGCs em bactérias e fungos (Blin et al., 2021).

### **Caracterização de BGCs alvos**

Após a predição de BGSs, *clusters* de interesse foram escolhidos e analisados quanto ao seu perfil dos genes que os compõe. Para isso, foi realizada a anotação funcional de tais alvos de acordo com a metodologia descrita em (Rosolen et al., 2023). De forma geral, foram escolhidos para uma investigação mais detalhada BGCs contendo genes que codificam transportadores, FTs e enzimas hidrolíticas. Tais critérios foram adotados visando um maior entendimento dos *clusters* potencialmente interessantes do ponto de vista biotecnológico.

### **Análises filogenéticas**

A partir da análise dos BGCs, foram escolhidos alvos, tais como enzimas PKSs, NRPSs e PKS-NRPSs, especificamente os genes Th0179\_007420 (enzima beta xilanase) e Th0179\_005503 (fator de transcrição com domínio MFS) para análise filogenética. As sequências de nucleotídeos alvo de *Trichoderma* spp. foram recuperadas do banco de dados NCBI (<https://www.ncbi.nlm.nih.gov/>). *Ilyonectria* sp., *Clonostachys* e *Cladobotryum mycophilum* foram utilizadas como *outgroup*. O alinhamento de múltiplas sequências será realizado usando ClustalW (Thompson et al., 1994) e a árvore filogenética será criada usando o software Molecular Evolutionary Genetics Analysis (MEGA) v7.0 (Kumar et al., 2016). 1.000 réplicas *bootstrap* (Felsenstein, 1985) serão aplicadas em cada análise. As árvores serão visualizadas e editadas usando Interactive Tree of Life (iTOL) v6 (<https://itol.embl.de/>).

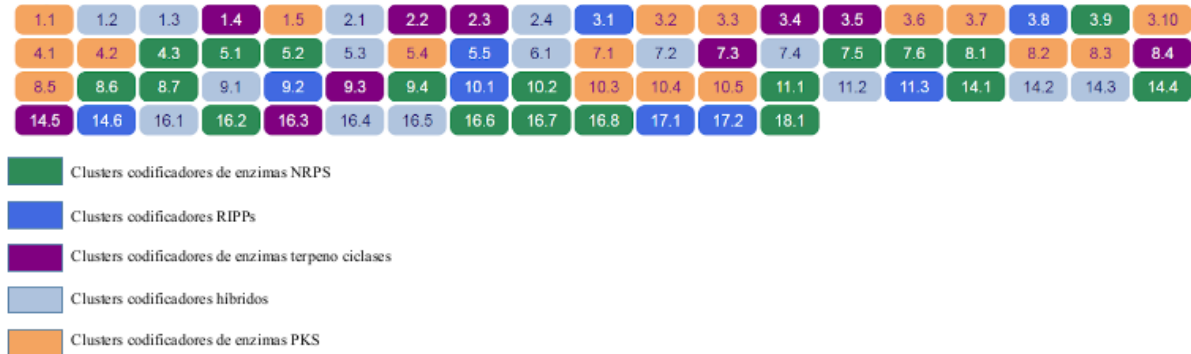
### **Nível de expressão gênica**

O perfil de expressão gênica dos genes alvo identificados nos BGCs de interesse foi investigado de acordo com a metodologia descrita em (Rosolen et al., 2023).

## **RESULTADOS E DISCUSSÃO:**

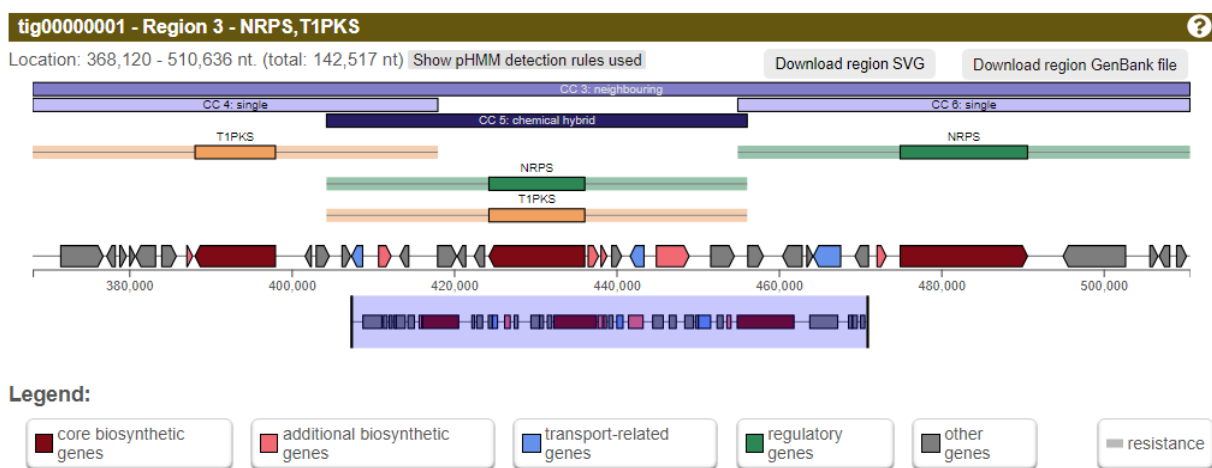
O AntiSMASH previu a presença de diferentes tipos de BGCs que codificam potenciais MSs.

No total, 70 BGCs foram previstos para o genoma de Th0179. De forma geral, foi observada a presença de 17 BGCs codificadores de enzimas PKS (**Figura 1**), as quais estão associadas à síntese de policetídeos com propriedades antifúngicas, antibióticas e outros efeitos biológicos relevantes (Keller, 2019). Além disso, foram observados 15 BGCs de genes codificadores híbridos que possuem mais de uma enzima biossintética principal codificada, incluindo NRPS e PKS (**Figura 1**).



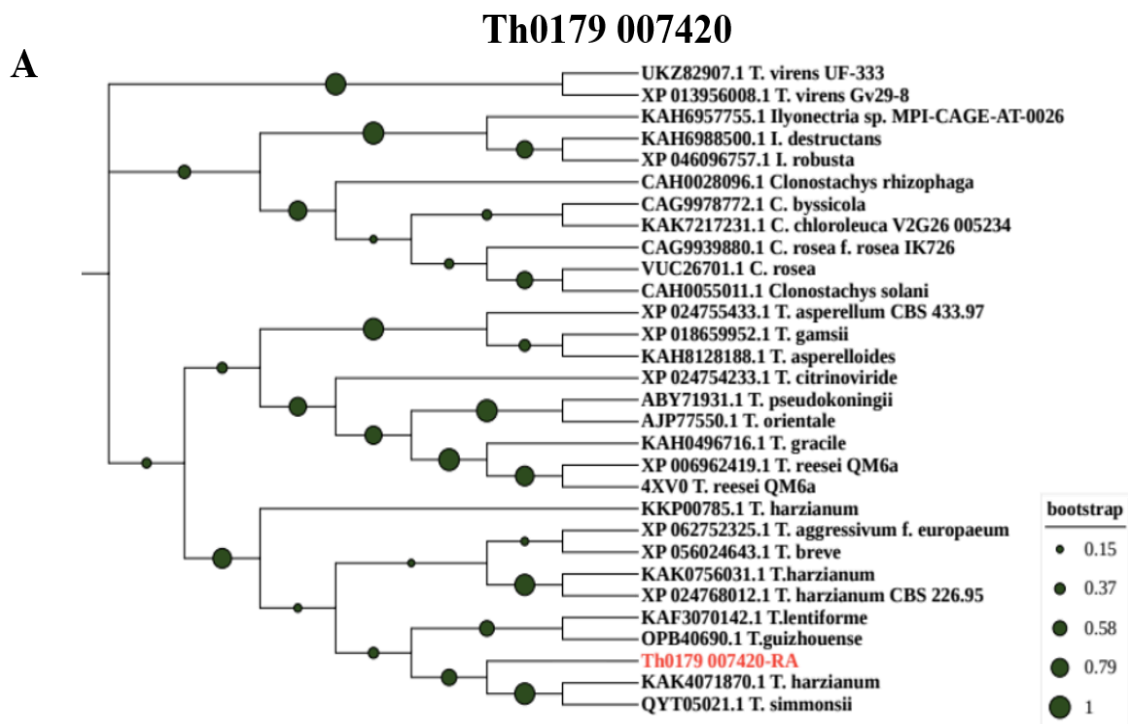
**Figura 1: Visão geral dos clusters de BGCs previstos pelo antiSMASH** (Blin et al., 2021). Cada cor representa uma classe de enzimas diferente. Azul claro: *Clusters* codificadores de enzimas híbridas (incluindo NRPS e PKS, NRPS e RIPP, etc.). Verde: *Clusters* codificadores de enzimas NRPS (Síntases de peptídeos não ribossomais). Azul royal: *Clusters* codificadores de enzimas RIPPs-like. Roxo escuro: *Clusters* codificadores de enzimas Terpeno Ciclases. Laranja: *Clusters* codificadores de enzimas PKS (Policetídeos Síntases).

De forma geral, este estudo destacou a notável diversidade metabólica de Th0179, evidenciada pela identificação de múltiplos BGCs de genes biossintéticos, incluindo PKS, NRPS, RIPP-like e terpeno ciclases. Tais *clusters* sugerem um amplo potencial para a síntese de compostos bioativos com diferentes propriedades entre si, os quais poderiam ser utilizados biotecnologicamente. Para uma análise mais detalhada, selecionamos *clusters* visando traçar regiões gênicas mais interessantes para trabalhos futuros. O BGC 1.3, por exemplo, possui uma ampla composição de elementos genômicos que se destacaram com relação aos outros *clusters* presentes no genoma analisado. Como enzima codificadora principal, temos aquelas codificadoras de enzimas NRPS e PKS. Além disso, genes adicionais à síntese proteica e genes transportadores também estão presentes nesse *cluster* (**Figura 2**).



**Figura 2: Apresentação do BGC 1.3 pela análise via antiSMASH.** A figura mostra a identificação e anotação de *clusters* de genes biossintéticos (BGCs) na região "Region 3 - NRPS, T1PKS" do *contig* 1. As barras coloridas representam diferentes tipos de genes: genes biossintéticos principais (vermelho), genes biossintéticos adicionais (rosa), genes relacionados ao transporte (azul claro), genes reguladores (verde), outros genes (cinza) e genes de resistência (roxo). A análise revela a sobreposição e interação entre *clusters* NRPS (verde) e T1PKS (amarelo).

Para a análise filogenética, escolhemos os genes Th0179\_007420 (codificador de enzima beta xilanase) e Th0179\_005503 (codificador de fator de transcrição com domínio MFS). Para inferir as relações filogenéticas das proteínas de interesse de Th0179 (**Figura 3**), foram utilizadas 26 sequências completas de proteínas relacionadas a 007420. Na árvore filogenética, observou-se uma proximidade genética entre Th0179\_007420, *Trichoderma simmonsii* e outra linhagem de *T. harzianum* (KAK4071870.1). *Ilyonectria* sp. e *Clonostachys* foram usados como grupos externos na análise. A. de forma geral, uma grande distância genética foi observada entre as sequências de aminoácidos de Th0179 e aquelas de *T. atroviride* e *T. reesei*. Tais resultados estão de acordo com a filogenia molecular das espécies analisadas (Rosolen et al., 2022; Rosolen et al., 2023).



## CONCLUSÃO:

A disponibilidade do genoma de Th0179 permitiu a identificação de BGCs putativos e a investigação detalhada de *clusters* de interesse; fornecendo, portanto, *insights* sobre aspectos da organização do genoma, expressão gênica e seus papéis potenciais na indústria biotecnológica.

## BIBLIOGRAFIA



Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, Marnix H., Weber, T., 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49, W29-W35. <https://doi.org/10.1093/nar/gkab335>.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *evolution* 39, 783-791.

Keller, N.P., 2019. Fungal secondary metabolism: regulation, function and drug discovery. *Nature Reviews Microbiology* 17, 167-180. <https://doi.org/10.1038/s41579-018-0121-1>.

Kubicek, C.P., Steindorff, A.S., Chenthamara, K., Manganiello, G., Henrissat, B., Zhang, J., Cai, F., Kopchinskiy, A.G., Kubicek, E.M., Kuo, A., Baroncelli, R., Sarrocco, S., Noronha, E.F., Vannacci, G., Shen, Q., Grigoriev, I.V., Druzhinina, I.S., 2019. Evolution and comparative genomics of the most common *Trichoderma* species. *BMC Genomics* 20, 485. <https://doi.org/10.1186/s12864-019-5680-7>.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 33, 1870-1874.

Rokas, A., Mead, M.E., Steenwyk, J.L., Raja, H.A., Oberlies, N.H., 2020. Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Natural product reports* 37, 868-878.

Rosolen, R.R., Aono, A.H., Almeida, D.A., Ferreira Filho, J.A., Horta, M.A.C., De Souza, A.P., 2022. Network Analysis Reveals Different Cellulose Degradation Strategies Across *Trichoderma harzianum* Strains Associated With XYR1 and CRE1. *Frontiers in Genetics* 13, <https://doi.org/10.3389/fgene.2022.807243>.

Rosolen, R.R., Horta, M.A.C., de Azevedo, P.H.C., da Silva, C.C., Sforca, D.A., Goldman, G.H., de Souza, A.P., 2023. Whole-genome sequencing and comparative genomic analysis of potential biotechnological strains of *Trichoderma harzianum*, *Trichoderma atroviride*, and *Trichoderma reesei*. *Molecular Genetics and Genomics* 298, 735-754. <https://doi.org/10.1007/s00438-023-02013-5>.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22, 4673-4680.