

Genômica comparativa e Potencial biossintético de *Streptomyces* MAD1003, isolada de sedimentos do Rio Madeira, para produção de metabólitos secundários

Costa, Gerodes Vasconcelos¹; Queiroz, Claudia Afras de Queiroz²; Mendes, Valdir da Costa¹; Raposo, Débora de Sena³; Sousa, Thiago Fernandes⁴; Bandeira, Izabel Correa⁴; Koolen, Hector Henrique Ferreira⁴; Silva, Gilvan Ferreira²

¹Instituto Nacional de Pesquisas da Amazônia, ²Embrapa Amazônia Ocidental, ³Universidade do Estado do Amazonas, ⁴Universidade Federal do Amazonas
Email: gilvan.silva@embrapa.br

Resumo

O gênero *Streptomyces* tem sido uma rica fonte de produtos naturais a anos, com a capacidades biossintéticas de produzir centenas compostos bioativos, são bactérias com grande potencial para produção de antibióticos, e outras moléculas, com aplicações médicas, industriais e agrícolas, incluindo antifúngicos, sideróforos, anticancerígenos, antioxidantes. Com os recentes avanços na tecnologia de sequenciamento de DNA, a mineração do genoma tornou-se uma ferramenta para explorar produtos naturais, facilitando a identificação de agrupamentos de genes biossintéticos (BGC). O presente estudo teve como objetivo explorar o potencial do isolado de *Streptomyces* sp. MAD1003. O genoma montado tem 6.642.326 Mb, com 6.293 possíveis regiões codificadoras (CDS), 66 genes de RNA de transferência (tRNA). A análise filogenômica indica que MDA1003 é filogeneticamente relacionada com *Streptomyces bauhinae* Bv016T. As similaridades entre as espécies usando vários índices de medidas como, identidade média de nucleotídeos (ANI) foi de 96.30%, identidade média de aminoácidos (AAI) de 96,24% e valores de hibridização digital DNA-DNA (dDDH) de 68.1% são indicativos que MDA1003 representa uma espécie. A mineração genômica combinando análise de bioinformática para identificação de BGCs e anotação manual, permitiu a identificação de 23 clusters envolvidos na produção de metabólitos secundários, destes apenas 26% (6) BGCs estão relacionados a moléculas e clusters caracterizados. A análise de sintenia dos clusters de MAD1003 com vias de biossíntese para o sideróforos desferrionamine B e para os agentes antimicrobiano lomofugin revelam o potencial para a produção de compostos quimicamente semelhantes mas não iguais. O estudo usando o genoma completo desta linhagem revela que a maioria dos BGC identificados não estão envolvidos na produção de metabólitos secundários conhecidos, indicando que esta possível nova espécie de *Streptomyces* pode ser usada na prospecção de novos produtos naturais bioativos.

Palavras-Chave: Clusters; biossintético; Metabólitos.

Introdução

As bactérias são uma importante fonte de produtos naturais, o gênero *Streptomyces* tem sido uma rica fonte de produtos naturais, baseado na descoberta anual de novas moléculas de *Streptomyces* a cada ano, estima-se que em se mantendo a tendência atual de descoberta o gênero pode sintetizar cerca de 150.000 compostos antimicrobianos a mais do que os atualmente conhecidos, sugerindo que o *Streptomyces* está longe de ser um recurso esgotado (Lacey e Rutledge 2022). Além dos antibióticos, outras moléculas, com aplicações médicas, industriais e agrícolas, incluindo compostos antibacterianos, antifúngicos, sideróforos, anticancerígenos, antioxidantes, imunossuppressores, compostos antitumorais, compostos antiparasitários e a produção de enzimas microbianas para a indústria (Giordano 2020).

As *Streptomyces* são bactérias Gram-positivas, que pertencem ao filo das actinobactérias, sendo responsáveis por cerca de 10.100 dos 22.500 metabólitos microbianos bioativos descobertos até agora, são bactérias com genoma com alto teor de C+G de 69-78%, descrita pela primeira vez por Waksman e Henrici (1943) com características fisiológicas que se assemelham às de muitas espécies de fungos, podendo ser encontradas em diferentes ambientes, especialmente no solo e ambientes aquáticos, sendo capazes de sobreviver em diversos habitats (Quinn et al. 2020; Donald et al. 2022; Kuncharoen et al. 2022; Liu et al. 2022).

Com os recentes avanços na tecnologia de sequenciamento de DNA, a mineração do genoma tornou-se uma nova maneira de explorar produtos naturais produzidas por esses microrganismos, facilitando a identificação de agrupamentos de genes biossintéticos (BGC), tornando possível a detecção de novas moléculas, e possibilitando o entendimento de sua base molecular, biossintética e funcional (Sandoval-Powers et al. 2021; Alam et al. 2022; Liu et al. 2022).

O presente estudo teve como objetivo realizar a identificação filogenômica de *Streptomyces* sp. MAD1003 e explorar o potencial biossintético do isolado por meio de mineração genômica.

Material e Métodos

Microrganismo e condições de cultivo

O isolado MAD1003 analisado neste trabalho faz parte da coleção de cultura do Laboratório de Biologia Molecular da Embrapa Amazônia Ocidental. Cadastrado no SISGEN Acesso No AB6B14F. Esta linhagem foi isolada de uma amostra de sedimento do Rio Madeira, e cultivada em meio ISP2.

Análise filogenômica

A identificação taxonômica baseada no genoma completo foi realizada por meio da plataforma Type (Strain) Genome Server (TYGS) (<https://tygs.dsmz.de>), uma plataforma de alto rendimento para taxonomia de procariotos que se baseia (LPSN- List of Prokaryotic names with Standing in Nomenclature) disponível em <https://lpsn.dsmz.de> que é uma base de dados acurada por especialistas para nomenclatura procariótica (Meier-Kolthoff et al. 2021).

A análise filogenômica foi avaliada usando vários índices de medidas, para a delimitação de similaridade entre as espécies mais relacionados com base na identidade média de nucleotídeos (ANI) e identidade média de aminoácidos (AAI), onde os cálculos dos valores médios de identidade nucleotídica (ANI) foram determinados através software OrthoANI (Yoon et al. 2017). E valores de hibridização DNA-DNA digital (dDDH), calculada com o Genome-to-Genome Distance Calculator (GGDC 2.1) (Sandoval-Powers et al. 2021).

Anotação do genoma

A sequência do genoma foi predita e anotada automaticamente seguindo o pipeline do servidor RAST (<https://rast.nmpdr.org/>), usando a base de dados SEED Viewer version 2.0. (Aziz et al. 2008). Comparando a qualidade da anotação utilizando o servidor BV-BRC (Bacterial and Viral Bioinformatics Resource Center), combinando os dados e ferramentas dos recursos do software PATRIC (<https://www.patricbrc.org>) através de uma análise abrangente do genoma. Todas as sequências do genoma das espécies tipos comparados foram baixados do banco de dados GenBank (<https://www.ncbi.nlm.nih.gov/genome>). A análise de pan genoma e as proteínas preditas foram comparadas com a espécie mais relacionadas de *Streptomyces bauhinae* e *Streptomyces*

seoulensis e plotadas na plataforma Orthovenn2 (Xu et al. 2019). Os valores baseados em BLAST (ANIb) e MUMmer (ANIm) foram avaliados na ferramenta web JSpeciesWS <http://jspecies.ribohost.com/jspeciesws> (Richter et al. 2016).

Predição de clusters gênicos biossintéticos e sintenia

O genoma foi analisado no AntiSMASH (antibiotics & Secondary Metabolite Analysis Shell) versão 6.0 (Blin et al. 2021) para a anotação dos clusters genes biossintéticos (BGC). Todas as sequências dos agrupamentos preditos apresentaram similaridade com BGCs depositados no banco de dados MIBiG (Minimum Information About a Biosynthetic Gene cluster) versão 2.0 <https://mibig.secondarymetabolites.org/> (Terlouw et al. 2022).

A análise de sintenia foi feita através do alinhamento dos clusters de genes de cada BGC depositados no banco de dados MIBiG, antiSMASH, ou NCBI, através de gráficos de sintenia plotados no programa clinker, uma ferramenta baseada em Python e clustermap.js, (Gilchrist e Chooi 2021).

Resultados e Discussão

Identificação filogenômica

O isolado MAD1003 havia sido identificado previamente como *Streptomyces* com base no sequenciamento da região 16S do rDNA. E a análise filogenética baseado no genoma completo confirmou o gênero e revelou que MAD1003 forma um clado com três espécies de *Streptomyces* (*S. bauhinae*, *S. griseoluteus*, *S. seoulensis*) sendo *Streptomyces bauhinae* Bv016T a espécie filogeneticamente mais relacionada (Fig.2).

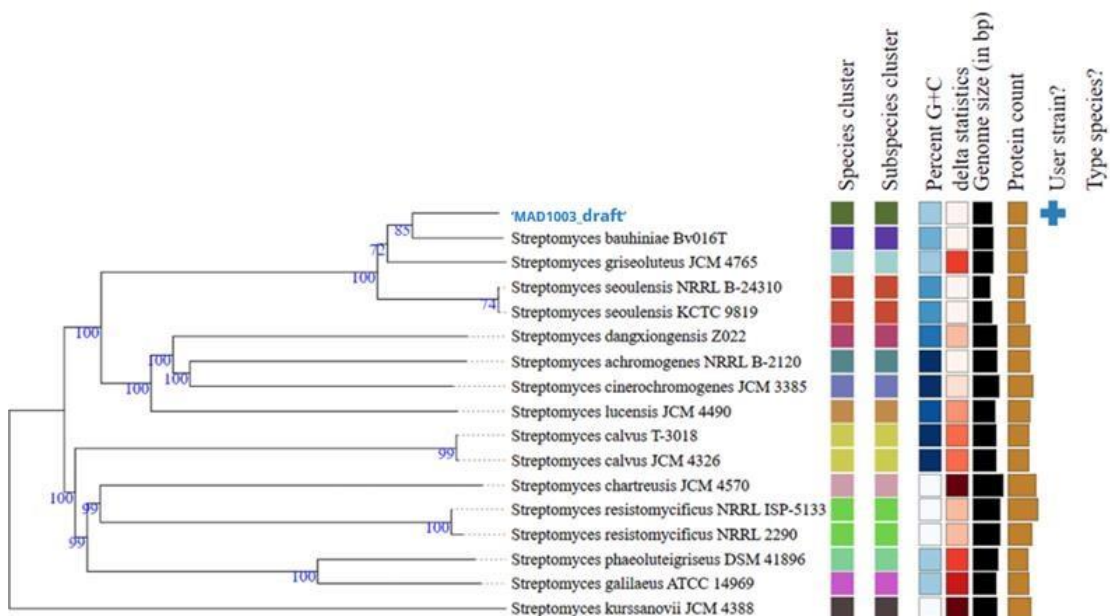


Figura 1 - Árvore filogenética obtida pelo TYGS (Type Strain Genome Server) baseada no genoma completo do isolado MAD1003 e espécies tipo mais proximamente relacionadas

As espécies tipos comparadas foram obtidas do banco de dados LSNP (List of Prokaryotic names with Standing in Nomenclature) que mantém uma lista acurada de nomes revisados por especialistas em taxonomia de procaríotos. O gênero *Streptomyces* é o maior gênero da classe das Actinobacteria, com 668 espécies caracterizadas com nomes validamente publicados (Maiti e Mandal 2021).

Os valores de ANI entre o isolado e duas das espécies intimamente relacionadas ficaram 96,13 e 95,13% (Fig. 3), onde mostra uma relação com espécie tipo descritiva, com valores próximos aos considerados como o limite da espécie entre (95-96%). Segundo Lee et al. (2016), esses são valores recomendados como critério para a identidade nucleotídica média (ANI).

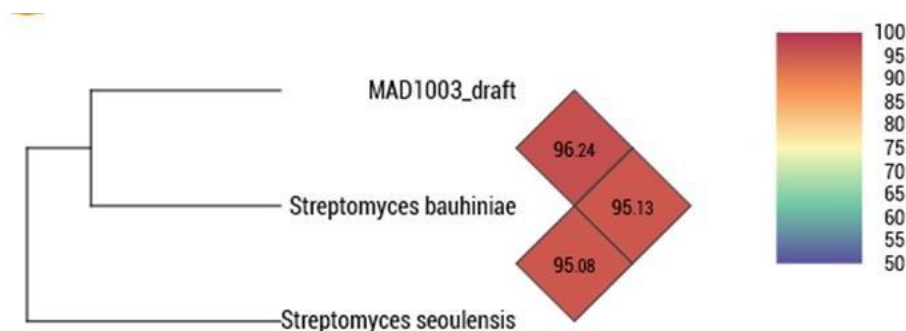


Figura 2 - Mapa de calor mostrando a identidade média de nucleotídeos por ortologia (OrthoANI) entre o Isolado MAD1003 e as linhagens das espécies tipo relacionadas.

A linhagem MAD1003 quando comparada com *S. bauhiniae* que é a linhagem mais próxima, apresenta valores de AAI de 96.24%. Vários estudos iniciais recomendaram um valor de ANI de aproximadamente 95-96% como limite para a demarcação de espécies (Goris et al. 2007; Richter e Rosselló-Móra 2009). Com o OrthoANI também foi recomendado pelos autores uma faixa semelhante de demarcação de espécies em 95 ~ 96 (Lee et al. 2016). Contudo, recentemente novo limite para delimitação de espécies do gênero *Streptomyces* foi estabelecido em 96,7%, este limite foi baseado na análise de 80 espécies de *Streptomyces* cujo o valor de dDDH de 70% não correspondeu a um valor de ANIm de 95~96%.

MAD1003 apresentou dDDH foi de 68.1%, este valor é mais um indicativo que esta linhagem pode representar uma nova espécie, conforme a literatura a ponte de corte para delimitação de espécies é dDDH igual ou maior que 70%, podendo ser consideradas novas espécies valores abaixo deste ponto de corte (Meier-Kolthoff et al 2013). Todos os valores estão descritos (Tabela 1).

Tabela 1. Valores de médias da identidade nucleotídica (ANI), identidade média de aminoácidos (AAI) e valores de hibridização digital DNA-DNA I (dDDH).

Isolado	Espécie tipo	dDDH (d2, em %)	dDDH (d4, em %)	AAI (%)	ANIm (%)	ANIB (%)
MAD1003	<i>Streptomyces bauhiniae</i>	68.1	68,1	96,24	96.36	95.65

AAI- identidade média de aminoácidos ANIm- MUMmer ANIB- Blast+ Anotação do genoma e análise de genes ortólogos

O genoma sequenciado possui um tamanho de 6,642,326 Mb, com 6.293 possíveis regiões codificadoras (CDS), 66 genes de RNA de transferência (tRNA) e 8 genes de RNA ribossômico (rRNA). o conteúdo G + C do isolado foi encontrado em uma faixa de 71.5%.

Alguns dos genes anotados estão relacionados e classificadas em subsistemas relacionados a fatores de virulência, cada subsistema tem um conjunto de proteínas que, juntas, implementam um processo biológico específico ou complexo estrutural, como: aquisição e Metabolismo do Ferro com genes relacionado para produção de sideróforo, a produção de metabólitos secundários e resistência a antibióticos. Verificou-se que o isolado MAD1003 continha genes 25 relacionados a Resistência a antibióticos e compostos tóxicos,

42 genes intimamente envolvidos na aquisição de ferro no metabolismo, com genes relacionados a produção do sideróforo Desferrioxamin B.

Foi evidenciado por Zhang et al. (2015) que a lomofungina é um antibiótico produzidos por *Streptomyces* sp., com atividade antibacteriana de amplo espectro, sintetizados pelo agrupamento de genes *phz* altamente conservado, com atividade contra bactérias Gram-positivas e Gram-negativas, bem como fungos patogênicos.

O OrthoVenn2 revelou 6293 proteínas para o isolado MAD1003, 649 clusters ortólogos, o pan genoma mostrou o genoma central intimamente relacionado com as espécies tipos *Streptomyces bauhinae* e *Streptomyces seoulensis* contendo 4842 clusters de genes de cópia única.

Identificação de Clusters gênicos biossintéticos

Foram identificados 23 diferentes BGCs relacionados à produção de metabólitos secundários no genoma da MAD1003. Diferentes classes de metabólitos, como: NRPS (8), PKS-I (2), PKS-II (1), RiPPs (1), 2 agrupamentos de genes contendo PKS-III, 3 classes de terpenos, entre outras moléculas (Figura 3). Notadamente o metabolismo secundário da MAD1003 está voltado para a biossíntese de diferentes tipos de peptídeos (8/23) e policetídeos (5/23).

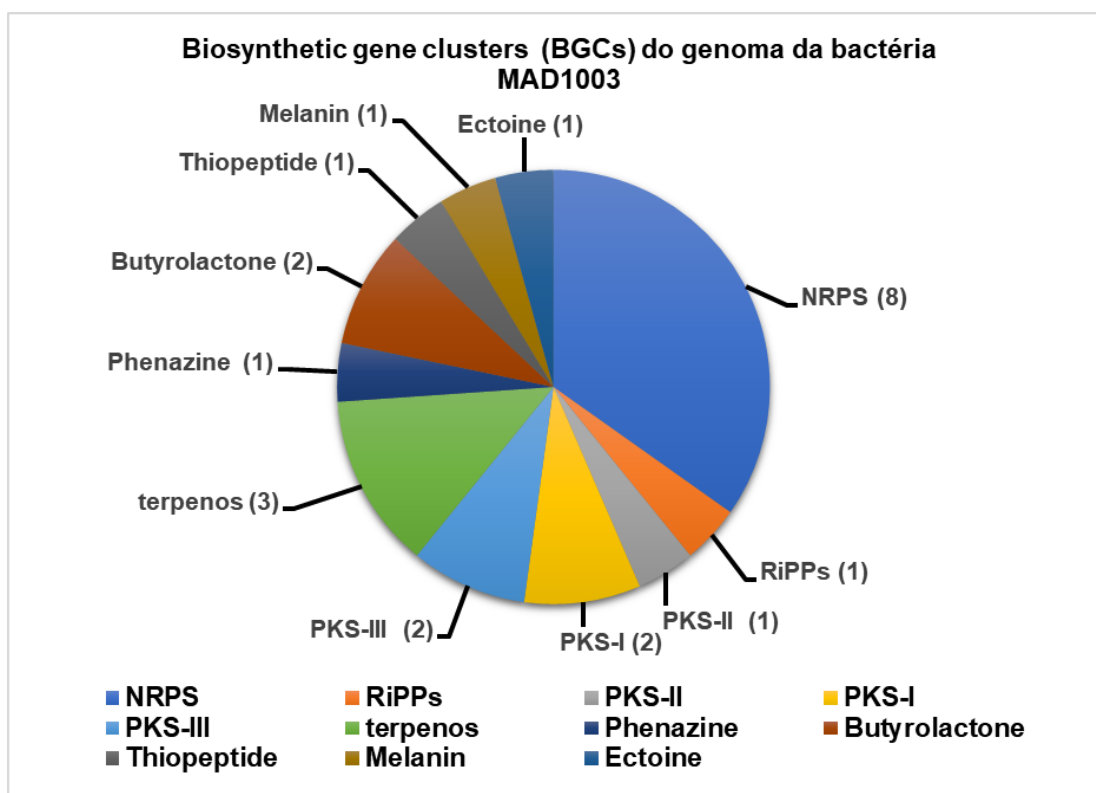


Figura 3 - Clusters gênicos biossintéticos identificados na linhagem MAD1003 relacionados a produção de metabólitos secundários, tendo como as principais classes: NRPS (peptídeo sintetase não ribossomal), PKS T1, T2 e T3 (policetídeo sintase tipo I, tipo II e tipo 3), terpeno e RiPPs (peptídeos sintetizados ribossomalmente e modificados pós-traducionalmente).

Foram identificados no genoma de MAD1003 Clusters gênicos biossintéticos (BGCs) com 100% de identidade com vias de biossintese já caracterizadas indicando que a linhagem é geneticamente capaz de produzir as seguintes moléculas: scabichelin (BGC 1.1), albaflavenone (24.1), ectoine (42.1), e algumas PKS T3 (policetídeo sintase tipo III) com 100% de semelhança predita na plataforma AntiSMASH.

Interessantemente a NRPS do BGC 74.1, apresentou 39% de similaridade com o BGC relacionada biossintese do sideróforo a desferrioxamine B e desferrioxamine E (figura 4), usado para medir a absorção do ferro, que foi aprovada para uso clínico pela Food and Drug Administration (FDA) em 1968 (Barona-Gómez et al. 2004).

O BGC 1.2 apresentou 39% com o cluster da lomofungin, um atibiótico que pertence a classe "phenazine", as fenazinas são antibióticos de amplo espectro. E estudos de Zhang et al. (2015) relatam que as fenazinas têm muitas funções biológicas como atividades antimicrobianas, antifúngicas, antitumorais, antimaláricas e antiparasitárias.

A análise de sintenia mostrou que a composição de genes dos clusters de MAD1003, *S. bauhiniae* e os genes da via de biossíntese da lomofungin depositada no banco de dados MIBiG, indicam a habilidade para produção de uma fenazina diferente visto que só foram identificados os genes lphzB, lphzC, lphzgG, que se mostraram alta similaridade com os genes da biossíntese de fenazina.

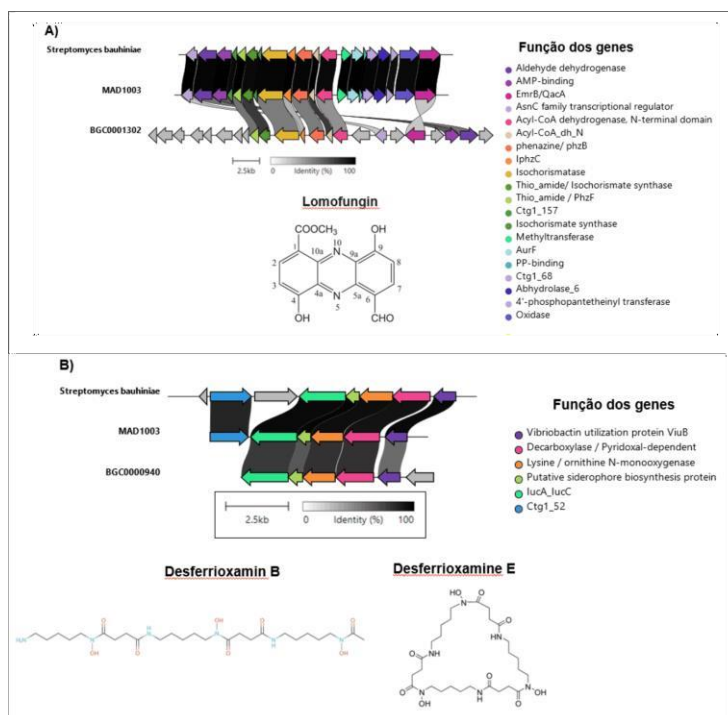


Figura 4 - Sitenia entre os BGC 1.2 e o BGC 74.1 de *Streptomyces* sp. MAD1003. A) representa a sitenia entre *Streptomyces* sp. MAD1003 e da *Streptomyces bauhiniae* como os clusters gênicos biossintéticos BGC0001302 junto com a estrutura química da lomofungin. B) representa a sitenia entre *Streptomyces* sp. MAD1003 e *Streptomyces bauhiniae* como os clusters gênicos biossintéticos BGC0000940 e as estruturas química de desferrioxamin B e desferrioxamine E. Toda a Predição da função gênica foi a partir da base de dados MIBiG.

Conclusões

Este estudo forneceu uma análise de genômica comparativa, indicando que o isolado MAD1003 representa uma possível nova espécie de *Streptomyces* proximamente relacionada com *S. bauhiniae* (dDDH de 68.1 e ANI 96,3, ANIm 96,34). A mineração gênica e outro ponto importante que o trabalho permitiu foi a identificação dos clusters biossintéticos para a produção de sideróforos como a desferrioxamine B, e do agente antimicrobianos lomofungin, facilitando a caracterização através da sintenia, indicando que o isolado estudado pode ser usado no processo para a produção de novos produtos naturais bioativos.

Agradecimentos

Os autores agradecem às agências de fomento: FAPEAM (Fundação de Amparo à Pesquisa do Estado do Amazonas) CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Este trabalho contou com o suporte financeiro da FAPEAM Editais: (PROSPAM 08/2021, CT&I ÁREAS PRIORITÁRIAS 010/2021, PRODUTIVIDADE-CT&I-013/2022, Biodiversa N° 007/2021), CAPES editais (Procad AmazonMicro, Amazônia Legal e pela concessão de bolsa de estudo ao primeiro autor).

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pela e concessão de Bolsas de estudo ao primeiro autor no PDPG Amazônia Legal. Agradeço ao Instituto Nacional de Pesquisas da Amazônia- INPA, A Fundação de Amparo à Pesquisa do Estado do Amazonas- FAPEAM, agradecemos a Universidade do Estado do Amazonas- UEA, por todo apoio durante as pesquisas, e a EMBRAPA - Amazônia Ocidental por todo apoio.

Referências

- Alam, K.; Hao, J.; Zhong, L.; Fan, G.; Ouyang, Q.; Mahmudul, I.S.; et al. 2022. Complete genome sequencing and in silico genome mining reveal the promising metabolic potential in *Streptomyces strain CS-7*. *Frontiers in Microbiology* 3751.
- Aziz, R.K.; Bartels, D.; Best, A.A.; Dejongh, M.; Disz, T.A.; Edwards, R.; et al. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9(1): 1-15.
- Barona-Gómez, F.; Wong, U.; Giannakopoulos, A.E.; Derrick, P.J.; Challis, G.L. 2004. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces c oelicolor* M145. *Journal of the American Chemical Society* 126(50): 16282-16283.
- Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-powers, Z.; Van wezel Gilles, P.; Medema, M.H.; Weber, T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49(W1): W29-W35.
- Donald, L.; Pipite, A.; Subramani, R.; Owen, J.; Keyzers, R.A.; Taufa, T. 2022. *Streptomyces*: Still the biggest producer of new natural secondary metabolites, a current perspective. *Microbiology Research* 13(3): 418-465.
- Gilchrist, C.L.; Chooi, Y. H 2021. Clinker & clustermap. js: Automatic generation of gene cluster comparison figures. *Bioinformatics* 37(16): 2473-2475.

- Goris, J.; Konstantinidis, K.T.; Klappenbach, J.A.; Coenye, T.; Vandamme, P.; Tiedje, J.M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* 57(1): 81-91.
- Lacey, H.J.; Rutledge, P.J. 2022. Recently discovered secondary metabolites from *Streptomyces* species. *Molecules* 27(3): 887.
- Lee, I.; Kim, Y.O.; Park, S.C.; Chun, J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology* 66(2): 1100-1103.
- Lin, Z.; Qu, X. 2022. Emerging diversity in polyketide synthase. *Tetrahedron Letters* 110: 154183.
- Liu, T.; Ren, Z.; Chunyu, W.X.; Li, G.D.; Chen, X.; Zhang, Z.T.L.; et al. 2022. Exploration of Diverse Secondary Metabolites From *Streptomyces* sp. YINM00001, Using Genome Mining and One Strain Many Compounds Approach. *Frontiers in Microbiology* 13: 831174.
- Kuncharoen, N.; Yuki, M.; Kudo, T.; Okuma, M.; Booncharoen, A.; Mhuantong, W.; Tanasupawat, S. 2022. Comparative genomics and proposal of *Streptomyces radidis* sp. nov., an endophytic actinomycete from roots of plants in Thailand. *Microbiological Research* 254: 126889.
- Maiti, P.K.; Mandal, S. 2021. *Streptomyces cupreus* sp. nov., an antimicrobial producing actinobacterium isolated from *Himalayan soil*. *Archives of Microbiology* 203(4): 1601-1609.
- Meier-Kolthoff, J.P.; Carbasse, J.S. Peinado-Olarte, R.L.; Göker, M. 2021. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Research* 50(D1): D801-D807.
- Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.P.; Göker, M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* (14): 1-14.
- Olson, R.D.; Assaf, R.; Brettin, T.; Conrad, N.; Cucinell, C.; Davis, J.J.; et al. 2023. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research* 51(D1): D678-D689.
- Quinn, G.A.; Banat, A.M.; Abdelhameed, A.M.; Banat, I.M. 2020. *Streptomyces* from traditional medicine: Sources of new innovations in antibiotic discovery. *Journal of Medical Microbiology* 69(8): 1040.
- Sandoval-Powers, M.; Králová, S.; Nguyen, G.S.; Fawwal, D.V.; Degnes, K.; Lewin, A.S.; et al. 2021. *Streptomyces poriferorum* sp. nov., a novel marine sponge-derived Actinobacteria species expressing anti-MRSA activity. *Systematic and Applied Microbiology* 44(5): 126244.

Richter M, Rosselló-Móra R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, 106(45), 19126-19131.

Richter, M.; Rosselló-Móra, R.; Oliver Glöckner, F.; Peplies, J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32(6): 929-931.

Terlouw, B.R.; Blin, K.; Navarro-Muñoz, J.C.; Avalon, N.E.; Chevrette, M.G.; Egbert, S.; et al. 2022. MIBiG 3.0: a community- driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research* 51(D1): D603-D610.

Xu, L.; Dong, Z.; Fang, L.; Luo, Y.; Wei, Z.; Guo, H.; et al. 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* 47(W1): W52-W58.

Yoon, S.H.; Ha, S.M.; Kwon, S.; Lim, J.; Kim, Y.; Seo, H.; Chun, J. 2017. Introducing EzBioCloud: a taxonomically united database of 16s rRNA gene sequences and whole- genome assemblies. *International Journal of Systematic And Evolutionary Microbiology* 67(5): 1613-1617.

Zhang, C.; Sheng, C.; Wang, W.; Hu, H.; Peng, H.; Zhang, X. 2015. Identification of the Lomofungin Biosynthesis Gene Cluster and Associated Flavin-Dependent Monooxygenase Gene in *Streptomyces lomondensis* S015. *PLoS ONE* 10(8): e0136228.