

Prospecção de produtos naturais de interesse biotecnológico com base na análise genômica de *Streptomyces* sp. MAD 27

Neves, Kiandro de Oliveira Gomes¹; Silva, José Carlos Ipuchima¹; Queiroz, Claudia Afras²; Yamagishi, Michel Eduardo Beleza³; Koolen, Hector Henrique¹; Silva, Gilvan Ferreira²

¹Universidade do Estado do Amazonas, ²Embrapa Amazônia Ocidental, ³Embrapa Agricultura Digital

Email: gilvan.silva@embrapa.br

Resumo

As bactérias do gênero *Streptomyces* são responsáveis pela biossíntese de aproximadamente 75% dos antibióticos existentes, bem como outras moléculas de interesse econômico. Recentemente, o sequenciamento de nova geração (NGS) tem facilitado a descoberta de novos produtos naturais, por meio da exploração da capacidade biossintética dos organismos por meio da análise de genomas completos. A abordagem de mineração de genomas permite analisar o genoma completo e prever agrupamentos gênicos, conhecidos como BGC (*Biosynthetic gene cluster*), responsáveis pela produção de metabólitos secundários (MS). Desta forma, este estudo teve como objetivo realizar a mineração genômica de *Streptomyces* sp. MAD 27 visando a identificação de BGCs relacionados à produção de MS de interesse biotecnológico. Foi possível identificar 45 BGCs, desses, 80% (36) que apresentam de moderada a nenhuma similaridade com cluster gênicos depositados na base de dados MIBiG (*Minimum Information about a Biosynthetic Gene cluster*). Entre os MS preditos há dois policetídeos (PK): *tetracenomycin C* e *flaviolin*, um peptídeo sintetizado por ribossomo e modificado pós-tradução (RIPP): thioholgamide A e B, e um terpeno que foi a *geosmin*. Estes MS possuem aplicação medicinal, associados a atividade antitumoral e de auto-oxidação, além de ambiental, como controle de contaminação hídrica. Dessa forma, os resultados obtidos demonstram o potencial biotecnológico de *Streptomyces* sp. MAD 27 e fornecem suporte para futuros estudos metabolômicos e de engenharia genética voltados à criação de protocolos para produção de bioinsumo e a caracterização de novos MS.

Palavras-Chave: Metabólitos; Genoma; Mineração.

Introdução

Os microrganismos são foco de estudos relacionados à produção de moléculas bioativas, estas também são conhecidas como metabólitos secundários (MS) ou produtos naturais (PN) e estão normalmente atuando na adaptação ao meio e/ou na defesa contra possíveis predadores ou competidores (Demain e Fang 2000; Machado et al. 2017).

Os MS são amplamente utilizados na produção de fármacos e de produtos de interesse socioambiental (Newman e Cragg 2020). Atualmente as tecnologias de sequenciamento de nova geração (NGS) combinadas com desenvolvimento e aprimoramento de ferramentas de bioinformática, tornou possível a predição de MS, a partir do genoma completo microrganismos por meio de mineração genômica (Albarano et al. 2020).

Estudos com mineração genômica indicam que em microrganismos a maioria dos MS é sintetizado por um conjunto de proteínas cujos genes estão fisicamente agrupados. A este conjunto de genes, dá-se o nome de BGC que vem do inglês *Biosynthetic gene clusters* (BGCs), e inclui genes que codificam enzimas relacionadas à via de biossíntese, e podem conter também genes para proteínas reguladoras como fatores de transcrição e/ou repressores, bem como transportadores relacionados à molécula sintetizada pelo BGC (Hahk-Soo e Eung-Soo 2021).

A identificação dos BGCs de um genoma microbiano permite conhecer todo o conjunto de metabólitos secundários que o referido microrganismo é capaz de sintetizar, mesmo aqueles que são pouco expressos ou silenciados em condições de laboratório. A caracterização de BGCs além de permitir a identificação de MSs, conectando os genes à química por meio da similaridade de genes pertencentes às vias metabólicas já conhecidas e depositadas no banco de dados MIBiG (Terlouw et al. 2023).

Também é possível identificar novos produtos naturais relacionados aos BGC sem nenhuma similaridade aos já conhecidos e caracterizados. Além disso, as informações genéticas contidas nos BGCs permitem explorar o arsenal metabólico por meio do desbloqueio das vias silenciadas, com base na expressão heteróloga, engenharia genética ou refatoração dos BGCs, ampliando dessa forma a descoberta e utilização de moléculas de interesse (Li et al. 2021; Sharma e Salwan 2021; Chiang et al. 2023).

Dentre os microrganismos mais estudados, em relação ao potencial biotecnológico presente em seu genoma, é possível destacar as bactérias do gênero *Streptomyces*, pois estão associadas à produção de 75% dos antibióticos disponíveis no mercado (Lee et al. 2020). Entre suas características genômicas, as espécies de *Streptomyces* apresentam BGCs que codificam enzimas que

contribuem para produção de MS, das mais diversas classes como terpenos, policetídeos, peptídeos não ribossomais entre outras (Nett et al. 2009).

Desta forma, o objetivo deste estudo foi realizar a mineração genômica de *Streptomyces* sp. MAD 27 visando a identificação de BGCs relacionados à produção de MS de interesse biotecnológico.

Material e Métodos

Neste estudo, o isolado de *Streptomyces* sp. MAD27, obtido de sedimentos do Rio Madeira, Amazonas, Brasil, foi selecionado com base na produção de compostos antimicrobianos identificados previamente *in vitro*. Cadastro de Acesso ao patrimônio genético N°. AB6B14F.

O genoma completo foi sequenciado usando a plataforma Illumina com *read length*: 2 x 150bp (Paired End). A montagem "de novo" foi realizada com o montador SPAdes (Prjibelski et al. 2020), e resultou em 8,6 milhões de pares de bases (pb) com cobertura de 294X e 907 scaffolds com N50 de 33243 pb e L50 de 80 pb. O maior scaffold apresentou tamanho de 188267 pb.

A identificação de BGCs foi realizada em janeiro de 2023 por meio da plataforma AntiSMASH (<https://antismash.secondarymetabolites.org/#!/start>), versão 7.0 (beta). Os dados obtidos foram agrupados de acordo com as classes de MS e relacionados à literatura para identificação da função das substâncias preditas na linhagem *Streptomyces* sp. MAD 27. O desenho das estruturas químicas das substâncias preditas foi realizado por meio do *software* ChemDraw Pro 12.0, a partir dos *smiles* presentes no banco MIBiG.

A análise de sintenia foi realizada usando o pipeline *clinker & clustermap.js* (Gilchrist e Chooi 2021). Para isso, os arquivos Genbank de BGCs já caracterizados foram extraídos pelo repositório MIBiG. Os arquivos genbank dos BGCs de MAD 27 foram obtidos por meio do antiSMASH. A comparação nos BGCs e as buscas de similaridade foram baseadas no algoritmo TBLASTX em *clinker* e a visualização da sintenia foi realizada pelo *clustermap.js*.

Resultados e Discussão

A partir da análise genômica, foram identificados 45 BGCs, desses, 20% (9) possuem alta similaridade ($\geq 70\%$) com vias de biossíntese já caracterizadas, 9% (4) apresentam similaridade moderada (30-69%), 42% (19) apresentam

baixa similaridade (1-29%) e 29% (13) não apresentaram nenhuma similaridade (0%) com os BGCs presentes base de dados MIBiG.

Segundo Blin et al. (2021), por meio do BGC é possível prever o potencial biossintético de uma linhagem. No genoma da linhagem MAD 27, foram identificados BGCs relacionados a 12 classes de metabólitos secundários (Figura 1). As três classes mais abundantes foram: policetídeos sintase (PKS), peptídeos não ribossomais sintase (NRPS), peptídeos sintetizados por ribossomos e modificados pós-traducionalmente (RIPP) e terpenos, estas classes correspondem a 69% (31) dos BGCs.

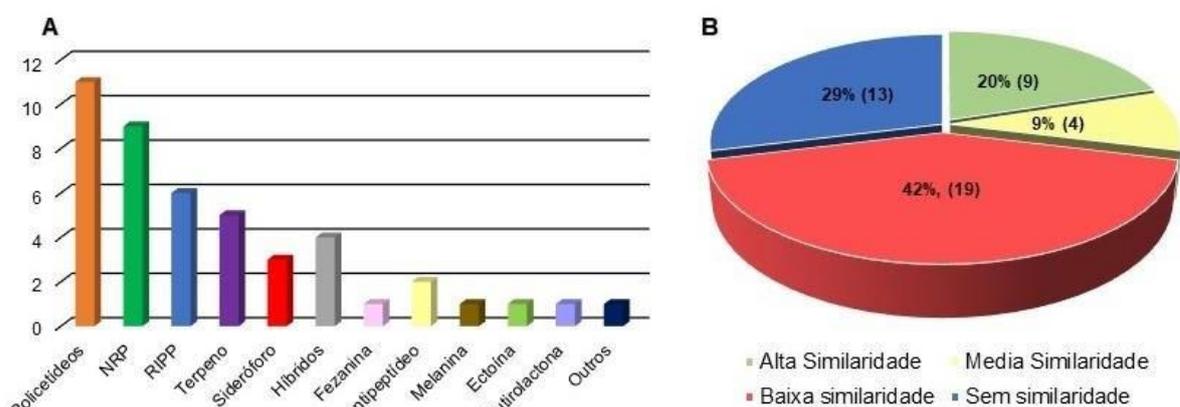


Figura 1 - Mineração genômica de *Streptomyces* sp. MAD 27. Em (A) Distribuição dos 45 clusters gênicos biossintéticos (BGC) identificados e distribuídos com bases na classe de produtos naturais relacionados ao gene core. Em (B) similaridade mensurada com base no número de genes que apresentam correspondência com vias de biossíntese já caracterizadas e depositadas na base de dados MIBiG (*Minimum Information about a Biosynthetic Gene cluster*).

As PKS correspondem a 25% (11) dos BGCs, muitos deles associados a substâncias descritas na literatura com atividade citotóxica, antimicrobiana e herbicida (Tabela 1). Dentro desta classe há a predição para a tetracenomicin C, um importante anticancerígeno aromático que inibe a tradução peptídica por meio da ligação à subunidade ribossômica grande (Nguyen et al. 2021), além de flaviolin, um pigmento produto da auto-oxidação do THN (Izumikawa et al. 2003).

Tabela 1 - Clusters gênicos biossintéticos (BGC) de metabólitos do tipo PKS previstos no genoma de *Streptomyces* sp. MAD27 usando antiSMASH 7.0 (beta).

Tipo BGC	Tamanho (pb)	Metabólito e ^a(% similaridade do BGC)	MIBiG ID
T1PKS,hgIE-KS	42,274	hexacosalactone A (13%)	BGC0002497
T1PKS	25,628	Desconhecido	-
T2PKS	45,666	tetracenomycin C (100%)	BGC0000275
T1PKS,T2PKS	44,088	spore pigment (83%)	BGC0000271
T1PKS	36,232	catenulisporolides (3%)	BGC0001963
T3PKS	34,367	flaviolin	BGC0002127
T1PKS	17,402	cyphomycin (2%)	BGC0001877
T1PKS	17,341	filipin (38%)	BGC0000059
T1PKS	15,805	ionostatin (43%)	BGC0002446
T1PKS	13,208	filipin (69%)	BGC0000059
T1PKS	8,037	4-hexadecanoyl-3-hydroxy-2-(hydroxymethyl)-2H-furan-5-one (18%)	BGC0000140

^aPorcentagem de semelhança entre os genes do BGC identificado na linhagem MAD 27 e genes localizados em cluster já caracterizado e depositado na base de dados MIBiG (Liu *et al.*, 2021). MIBiG ID- representa o número do BGC cuja similaridade foi identificada.

Dentre os NRP, para 21% (9), não houve predição para nenhum MS (Tabela 2), embora tenha sido encontrado cluster similares a MS com atividade antibiótica e ligadas ao sistema imune. Como não houve similaridade completa entre os BGCs do tipo NRPS com a base de dados MIBiG, porém, não significa que a linhagem não produz MS a partir de NRPS, sugerindo estudos químicos para corroborar com os resultados obtidos.

Tabela 2 - Aglomerados de genes biossintéticos de metabólitos do tipo NRPS previstos no genoma de *Streptomyces* sp. MAD27 usando antiSMASH 7.0 (beta).

Tipo de BGC	Tamanho (pb)	Metabólito e ^a (% similaridade)	MIBiG ID
NRPS-like	22,578	neoabyssomicin, abyssomicin (6%)	BGC0001694
NRPS	52,893	friulimicin A, friulimicin B, friulimicin C, friulimicin D (21%)	BGC0000354
NRPS	22,568	naphthomycin A (9%)	BGC0000106
NAPAA,NRPS	33,967	stenothricin (13%)	BGC0000431
NRPS	33,538	meridamycin (15%)	BGC0001011
NRPS	18,691	Desconhecido	-
NRPS-like	12,058	Desconhecido	-
NRPS	10,916	Desconhecido	-
NRPS	2,749	Desconhecido	-

^aPorcentagem de semelhança entre os genes do BGC identificado na linhagem MAD 27 e genes localizados em cluster já caracterizado e depositado na base de dados MIBiG (Liu *et al.*, 2021). MIBiG ID- representa o número do BGC cuja similaridade foi identificada.

Os RIPP e Terpenos (Tabela 3) representam duas classes de metabólitos com 13% (6) e 11% (5), respectivamente, dos BGCs de *Streptomyces* sp. MAD 27, com predição, no caso dos RIPP, para thioholgamide A e B, um MS de interesse medicinal para o combate de diferentes tipos de câncer (Dahlem *et al.* 2020), e, na classe dos terpenos, o MS geosmin associado a contaminação de efluentes e o característico cheiro de "terra molhada" (Cane *et al.* 2006).

Tabela 3 - Aglomerados de genes biossintéticos de metabólitos do tipo RIPP e terpenos previstos no genoma de *Streptomyces* sp. MAD27 usando antiSMASH 7.0 (beta).

RIPP			
Tipo de BGC	Tamanho (pb)	Metabólito e ^a(% similaridade)	MIBiG ID
lanthipeptide-class-v, thioamitides	38,789	thioholgamide A, thioholgamide B (100%)	BGC0001802
lassopeptide	22,333	Desconhecido	-
RiPP-like	10,216	informatipeptin (28%)	BGC0000518
LAP,thiopeptide	13,417	Desconhecido	-
RiPP-like	11,107	Desconhecido	-
RiPP-like	6,865	Desconhecido	-
TERPENO			
terpene	26,712	hopene (92%)	BGC0000663
terpene	21,086	julichrome Q3-3, julichrome Q3-5 (25%)	BGC0002012
furan	14,771	asukamycin (24%)	BGC0000187
terpene	14,583	geosmin (100%)	BGC0001181
terpene	4,263	ebelactone (5%)	BGC0001580

^aPorcentagem de semelhança entre os genes do BGC identificado na linhagem MAD 27 e genes localizados em cluster já caracterizado e depositado na base de dados MIBiG (Liu *et al.*, 2021). MIBiG ID- representa o número do BGC cuja similaridade foi identificada.

Na classe das PKS, NRPS, RIPP e terpeno encontramos a predição para quatro MS. A partir da análise de sintenia entre os BGCs da *Streptomyces* sp. MAD 27 e os BGCs de referência na base de dados MIBiG (Figura 2), foi possível comparara composição gênica entre as vias de biossíntese e inferir que a linhagem estudada tem todos os genes necessários para a biossíntese do MS predito.

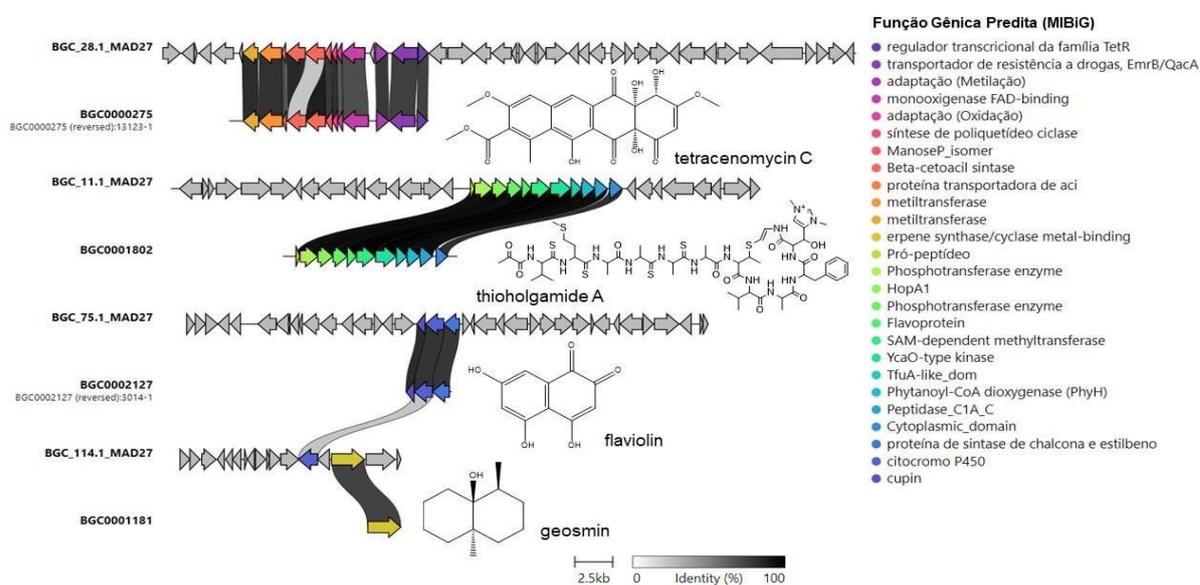


Figura 2 - Análise de sintenia entre clusters identificados em *Streptomyces* sp. MAD 27 e aqueles relacionados a vias de biossíntese já caracterizadas. (1) Sintenia entre o BGC 28.1 de MAD 27 e o BGC0000275 de *Streptomyces glaucescens*. (2) Estrutura química de tetracenomycin C. (3) Sintenia entre o BGC 11.1 de MAD 27 e o BGC0001802 de *Streptomyces malaysiense*. (4) Estrutura química de thiohologamide A. (5) Sintenia entre o BGC 75.1 de MAD 27 e o BGC0002127 de *Streptomyces coelicolor* A3(2). (6) Estrutura química de flaviolin. (7) Sintenia entre o BGC 114.1 de MAD 27 e o BGC0002127 de *S. coelicolor* A3(2). (8) Estrutura química de geosmin. (9) Predição da função gênica a partir da base de dados MIBiG.

A partir da análise de sintenia é possível prever que há chance de biossíntese por *Streptomyces* sp. MAD 27, para os MS thiohologamide A e B, flaviolin e geosmin, pois apresentam maior nível de identidade gênica (> 80%). Apesar de tetracenomycin C apresentar 100% de semelhança predita na plataforma AntiSMASH, o mesmo não é visto na análise de sintenia, podendo ser resultado de um erro de predição do AntiSMASH, o que pode ser verificado com outras abordagens químicas e de bioinformática. Todos os clusters gênicos associados a metabólitos secundários, com aplicação biotecnológica, são importantes, sendo a análise do genoma o primeiro passo para a descoberta de novos bioativos (Hahk-Soo e Eung-Soo 2021). Dessa forma, as informações com base na mineração genômica apresentadas aqui são, portanto, uma etapa tanto para seleção quanto para decidir os BGCs a serem explorados na busca de novos produtos naturais e de moléculas que possam trazer soluções aos setores agrícola, clínico e farmacêutico e fomentar a utilização dos metabólitos secundários nos setores sociais aos quais possuem aplicação (Albarano et al. 2020).

Além disso, a caracterização dos BGCs identificados fornece subsídios para o uso de abordagens como: engenharia genética voltada à refatoração de genes visando desbloqueio de vias, expressão heteróloga de BGCs que combinadas com ensaios químicos e biológicos tornarão possível identificar novos PNs e suas aplicações biotecnológicas.

Conclusão

Neste trabalho foi possível identificar os BGCs da linhagem *Streptomyces* sp. MAD 27 e prever com base no genoma que a linhagem (MAD 27) é capaz de realizar a síntese de: tetracenomicin C, thioholgamide A e B, flaviolin e geosmin, que são produtos naturais com atividades biológicas já reportada. Embora alguns BGCs tenham sido conectados às moléculas mencionadas, 80% do genoma, possui pouca ou nenhuma similaridade com vias já caracterizadas, indicando a possibilidade de estarem relacionados à biossíntese de novos produtos naturais ou podem estar ligados a moléculas já conhecidas cujos genes responsáveis pela rota metabólica ainda não foram identificados.

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 e PRODUTIVIDADE-CT&I- 013/2022), CAPES editais (Procad AmazonMicro, AmazôniaLegal, e pela concessão de bolsa de estudo ao primeiro autor).

Referências

- Albaro, L.; Esposito, R.; Ruocco, N.; Costantini, M. 2020. Genome mining as new challenge in natural products discovery. *Marine Drugs* 18(4): 199.
- Blin, K.; Shaw, S.; Kloosterman, A.; Charlop-Powers, Z.; Wezel, G.; Medema, M.; Weber, T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49(W1): W29-W35.

- Chiang, C.; Ohashi, M.; Tang, Y. 2023. Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining. *Natural Product Reports* 40: 89-127.
- Dahlem, C.; Siow, W.; Lopatniuk, M.; Tse, W.; Kessler, S.; Kirsch, S. 2020. Thioholgamide a New Anti-Proliferative Anti-Tumor Agent, Modulates Macrophage Polarization and Metabolism. *Cancers (Basel)* 12(5): 1288.
- Demain, A.; Fang, A. 2000. The natural functions of secondary metabolites. *History of Modern Biotechnology* I 1-39.
- Gilchrist, C.; Chooi, Y. 2021. Clinker & clustermap.js: geração automática de números de comparação de clusters de genes. *Bioinformatics* 37(16): 2473-2475.
- Hahk-Soo, K.; Eung-Soo, K. 2021. Recent advances in heterologous expression of natural product biosynthetic gene clusters in *Streptomyces* hosts. *Current Opinion in Biotechnology* 69: 118-127.
- Izumikawa, M.; Shipley, P.; Hopke, J.O.; Hare, T.; Xiang, L.; Noel, J.; Moore, B. 2003. Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from *Streptomyces coelicolor* A3(2). *Journal of Industrial Microbiology and Biotechnology* 30(8): 510-515.
- Lee, N.; Hwang, S.; Kim, J.; Cho, S.; Palsson, B.; Cho, B. 2020. Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Computational and Structural Biotechnology Journal* 18: 1548-1556.
- Li, L.; MacIntyre, L.; Brady, S. 2021. Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. *Current Opinion in Biotechnology* 69: 145-152.
- Machado, H.; Tuttle, R.; Jensen, P. 2017. Omics-based natural product discovery and the lexicon of genome mining. *Current Opinion in Microbiology* 39: 136-142.
- Nett, M.; Ikeda, H. 2009. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Natural Products Reports* 26: 1362-1384.
- Newman, D.; Cragg, G. 2020. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of Natural Products* 27;83(3): 770-803.
- Nguyen, J.; Riebschleger, K.; Brown, K.; Gorgijevska, N.; Nybo, E. 2021. A BioBricks toolbox for metabolic engineering of the tetracenomycin pathway. *Biotechnology Journal* 17(3): e2100371.
- Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. 2020. Using 539SPAdes de novo assembler. *Current Protocols in Bioinformatics* 70: e102.540.

Sharma, V.; Kaur, R.; Salwan, R. 2021. *Streptomyces*: host for refactoring of diverse bioactive secondary metabolites. *3 Biotech* 11(7): 1-32

Terlouw, B.; Blin, K.; Navarro-Muñoz, J. 2023. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research* 51(D1): D603-D610.