

MINERAÇÃO DE GENOMAS A PARTIR DE SEQUENCIAMENTO DE METAGENOMAS DE AMOSTRAS DA AMAZÔNIA: UMA BUSCA POR PROTEASES

Data de submissão: 03/02/2024

Data de aceite: 01/03/2024

Eric de Lima Silva Marques

Instituto de Ciências Farmacêuticas,
Universidade Federal de Alagoas
Maceió – AL
<http://lattes.cnpq.br/9592694184543133>

Adriana Barros de Cerqueira e Silva

Faculdade de Ciências Farmacêuticas,
Universidade Federal do Amazonas
Manaus – AM
<http://lattes.cnpq.br/6754157321525845>

Mariana Barros de Cerqueira e Silva

Faculdade de Ciências Farmacêuticas,
Universidade Federal do Amazonas
Manaus – AM
<http://lattes.cnpq.br/8916215916089270>

RESUMO: O presente estudo investigou os genomas montados a partir de sequenciamentos de metagenomas de amostras da Amazônia em busca de potenciais novas proteases e peptidases. Utilizando técnicas de mineração de genomas/metagenomas depositados em bases de dados, identificamos 23 potenciais proteases/peptidases com tamanhos variados classificadas como proteínas hipotéticas. Uma protease em particular mostrou homologia com a

superfamília da Peptidase S1, destacando-se pela similaridade de 60% com enzimas conhecidas considerando os aminoácidos quimicamente similares. A análise comparativa revelou a presença de um domínio de peptidase tipo tripsina. Esse trabalho demonstra a importância da Amazônia como uma fonte promissora de enzimas para aplicações biotecnológicas. Futuras pesquisas devem centrar-se na síntese e avaliação experimental destas enzimas, especialmente aquelas associadas a proteínas hipotéticas, a fim de confirmar suas atividades e explorar seu potencial uso em diversos setores industriais. Este estudo destaca não apenas a diversidade biotecnológica única da Amazônia, mas também a necessidade de práticas sustentáveis e colaboração (direta ou indiretamente) entre pesquisadores para maximizar o entendimento e aproveitamento desses recursos.

PALAVRAS-CHAVE: bioprospecção, enzimas, peptidase, potencial biotecnológico

GENOME MINING FROM METAGENOME SEQUENCING OF AMAZONIAN SAMPLES: A SEARCH FOR PROTEASES

ABSTRACT: The present study investigated genomes assembled from metagenome sequencing of Amazonian samples in search of potential new proteases and peptidases. Utilizing techniques for mining genomes/metagenomes deposited in databases, we identified 23 potential proteases/peptidases with varying sizes classified as hypothetical proteins. One particular protease exhibited homology with the Peptidase S1 superfamily, standing out with a 60% similarity to known enzymes considering chemically similar amino acids. Comparative analysis revealed the presence of a trypsin-like peptidase domain. This work underscores the significance of the Amazon as a promising source of enzymes for biotechnological applications. Future research endeavors should focus on synthesizing and experimentally evaluating these enzymes, especially those associated with hypothetical proteins, to confirm their activities and explore their potential use across diverse industrial sectors. This study highlights not only the unique biotechnological diversity of the Amazon but also emphasizes the need for sustainable practices and collaboration (directly or indirectly) among researchers to maximize the understanding and utilization of these resources.

KEYWORDS: bioprospection, enzymes, peptidase, biotechnological potential

INTRODUÇÃO

Genoma, por definição, é o conjunto de genes de um organismo. Em uma publicação de 1988 (Handelsman *et al.*, 1998), o termo ‘metagenoma’ foi utilizado pela primeira vez com o prefixo ‘meta-’ significando além, depois. Dessa forma, ‘metagenoma’ é utilizado para se referir ao conjunto de genes de uma comunidade microbiana em um determinado ambiente sem a necessidade de isolar ou cultivar os microrganismos. Até o início do século XXI, estudos metagenômicos não eram tão frequentes devido aos altos custos. No entanto, com o desenvolvimento de novas técnicas de sequenciamento de DNA, ocorreu um barateamento significativo da técnica, permitindo seu uso em grande quantidade por diversos grupos de pesquisa (Akaçin *et al.*, 2022).

Com o uso do sequenciamento de genomas e metagenomas em larga escala, muitos estudos vêm sendo realizados e as sequências são depositadas em bancos de dados, como o Genbank do NCBI (<https://www.ncbi.nlm.nih.gov/>). De maneira geral, esses estudos buscam responder pelo menos uma das perguntas a seguir: “quem são?”, “o que estão fazendo?” e “por que estão nesse ambiente?” (Kumar *et al.*, 2015; Pandey e Singhal, 2023). A primeira pergunta está comumente associada à diversidade e abundância de membros da comunidade microbiana; a segunda pergunta está associada ao potencial metabólico da comunidade, e a terceira pergunta está relacionada às relações ecológicas entre membros da comunidade.

No âmbito do “o que estão fazendo?”, há uma área interessante de mineração de genomas que vem se destacando (Bauman *et al.*, 2021; Bhattacharjee *et al.*, 2023) e pode gerar uma pergunta derivada: “o que podem fazer?”. A mineração de genoma

busca encontrar genes, operons ou grupos de genes específicos para uma ou mais funções determinadas. Muitas vezes, é utilizada para buscar potenciais genes de interesse biotecnológico. A mineração de genoma pode ser realizada em qualquer genoma de um organismo específico ou metagenoma de uma comunidade, utilizando ferramentas de bioinformática específicas para identificar esses genes.

A mineração de genoma pode ser feita com genomas/metagenomas sequenciados pelo grupo de estudos ou com os disponíveis em uma base de dados de acesso livre, como o Genbank do NCBI, que foram sequenciados por outro grupo de pesquisa (Barriuso e Martínez, 2015). O uso de dados de bancos de dados é comum na mineração de genoma, pois muitos estudos focam apenas em uma questão ou em análises de genes muito específicos para o ambiente em estudo, não explorando o potencial biotecnológico da sequência do genoma/metagenoma. Dessa forma, muitos genomas e metagenomas já estudados apresentam potenciais novos genes que não foram explorados como aqueles associados com vias de biossíntese antimicrobianas e antitumorais, por exemplo, descobertos em vias de polipeptídeos e peptídeos não ribossômicos (Bhattacharjee *et al.*, 2023). Além disso, enzimas podem ser identificadas por meio da mineração de genomas. Hidrolases, como proteases, lipases, celulasas e amilases, são enzimas comumente prospectadas em microrganismos cultivados e podem ser identificadas em sequenciamentos de genomas e metagenomas, assim como qualquer outra enzima de interesse biotecnológico, como ligases, polimerases, nucleases, entre outras.

Como é possível avaliar as enzimas *in silico* e sintetizar os genes *in vitro*, ou mesmo desenhar primers para tentar amplificar e clonar o gene de interesse identificado (se houver alíquota da amostra de DNA que foi utilizada para o sequenciamento armazenada), a mineração de genoma apresenta um potencial muito grande para novas descobertas.

Nesse sentido, um ambiente com grande potencial para novas descobertas e que possui alguns estudos de genoma e metagenoma é a Amazônia. O bioma amazônico é o habitat de uma diversidade de animais, plantas e microrganismos, muitos deles ainda pouco conhecidos e que (Pereira *et al.*, 2017) e representam possibilidades de descobertas biotecnológicas interessantes. Dessa forma, o presente estudo visa avaliar os dados metagenômicos na base de dados do NCBI e minerar o metagenoma buscando potenciais novas proteases produzidas pelos microrganismos da Amazônia.

METODOLOGIA

Ao buscar sequenciamento de metagenomas associados à Amazônia, foram encontrados três artigos que realizaram a montagem de genomas a partir desses metagenomas (Mandro *et al.*, 2022; Parks *et al.*, 2017; Venturini *et al.*, 2022). Dessa forma, a busca por proteases foi realizada em genomas já montados provenientes de metagenomas.

Genomas montados a partir de amostras coletadas associadas atividades ligadas aos seres humanos (direta ou indiretamente) não foram utilizados. Os genomas foram acessados na base de dados do NCBI, e aqueles que não tinham anotação associado também foram excluídos do presente estudo.

A anotação dos genomas restantes foi avaliada em relação as proteínas anotadas e associadas a uma função/atividade e as proteínas hipotéticas. As proteínas hipotéticas foram separadas e analisadas no pBLAST, comparando com a base de dados do Protein Data Bank (PDB) e no MEROPS (Barrett, Rawlings e O'Brien, 2001). Aquelas que apresentaram resultado significativo para protease em pelo menos uma das análises foram posteriormente analisadas no InterPro (Paysan-Lafosse *et al.*, 2023) para classificar a família de proteínas e no SignalP (Nielsen, Brunak e Heijne, von, 1999) para confirmar a presença de peptídeo sinal. Proteínas que apresentaram resultado positivo no InterPro foram submetidas a alinhamento com outras similares.

RESULTADOS E DISCUSSÃO

Na última década, o avanço da tecnologia de sequenciamento de DNA permitiu a utilização do sequenciamento em larga escala em pesquisas, abrangendo genomas e metagenomas, com custos cada vez mais reduzidos. Como resultado, diversas seqüências genômicas estão disponíveis em bases de dados, como o GenBank do NCBI. O desenvolvimento da mineração de genoma surgiu para possibilitar a busca por genes e vias biossintéticas de interesse biotecnológico nessas bases de dados, principalmente na região amazônica, ainda muito pouco explorada nesse sentido (Bergmann *et al.*, 2014).

A busca por seqüências de genomas montados a partir de metagenomas na Amazônia resultou na identificação de 19 genomas. No entanto, devido aos critérios de exclusão utilizados, 12 genomas foram excluídos por estarem associados a amostras de pastagem. Dos 7 genomas restantes, 6 foram associados a amostras de floresta (Mandro *et al.*, 2022; Venturini *et al.*, 2022), e um a amostra de rio (Parks *et al.*, 2017) (*Sphingobium* sp. UBA5915). As informações sobre os genomas podem ser visualizadas na Tabela 1. Observa-se que nenhum genoma foi montado completamente, mas o número de contigs permite a prospecção de genes de interesse biotecnológico, neste caso, proteases/peptidases. Em análises de metagenoma, a utilização de genomas montados a partir de metagenomas ou de sequenciamentos oriundos de plataformas que sequenciam fragmentos grandes de DNA (> 10kb) aumenta a possibilidade de ter sequenciado e montado o gene completo e, com isso, ter a seqüência do peptídeo completo, principalmente quando se trata de um operon ou de uma via biossintética, possibilitando a identificação de todos os genes completos.

ORGANISMO	ACESSO	TAMANHO DO GENOMA	CONTIGS	NÚMERO DE PROTEÍNAS HIPOTÉTICAS
<i>Massilia</i> sp. Bin.003_Forest	GCA_024508955.1	7.5MB	77	*
<i>Mesorhizobium</i> sp. Bin.005_Forest	GCA_024509055	6.7MB	588	833
<i>Nocardioidekongjuensis</i> Bin.002_Forest	GCA_024508895	5.5MB	12	809
<i>Paraburkholderia</i> sp. Bin.013_Forest	GCF_024281035	5.6MB	78	637
<i>Sphingobium</i> sp. UBA5915	GCA_002431195	4.6MB	184	747
<i>Sphingomonas</i> sp. Bin.006_Forest	GCF_024508715	3.2MB	93	451
<i>Xanthobacteraceabacterium</i> Bin.001_Forest	GCA_024508855.1	5.2MB	113	*

* Anotação do genoma não disponível no NCBI. Genoma excluído de análises subsequentes.

Tabela 1 – Genomas montados a partir de sequenciamento de metagenomas que foram identificados na base de dados do NCBI oriundos de amostras do bioma amazônico.

Nas páginas associadas à montagem de cada genoma no NCBI, há a possibilidade de acessar os genes que foram anotados para esse genoma. Ao avaliar isso, verificou-se um grande número de genes associados a proteínas hipotéticas para cada um dos genomas anotados, variando entre 451 e 833 (Tabela 1). Considerando apenas os genes anotados, é possível analisar em busca de identificar potenciais genes de interesse biotecnológico, lembrando que essas amostras são de um sequenciamento de metagenoma, portanto, o microrganismo não foi cultivado.

Ao analisar os genes anotados e associados a uma função específica, observa-se que esses microrganismos possuem protease e lipase, conforme esperado, e três deles também apresentaram anotações para amilases e celulasas (Tabela 2), valores esses dentro do esperado (Ficarra *et al.*, 2016). Por serem de origem metagenômica e não ter o microrganismo isolado, essas enzimas podem ser avaliadas quanto ao potencial biotecnológico *in silico* e, se possível, sintetizadas e clonadas para uma avaliação *in vitro* daquelas consideradas promissoras, uma vez que já foram identificadas por meio das ferramentas *in silico* (Cornish *et al.*, 2022; Guan *et al.*, 2020).

ORGANISMO	PROTEASE	LIPASE	AMILASE	CELULASE
<i>Mesorhizobium</i> sp. Bin.005_Forest	32	12	3	2
<i>Nocardioidekongjuensis</i> Bin.002_Forest	26	11	0	1
<i>Paraburkholderia</i> sp. Bin.013_Forest	22	13	0	0
<i>Sphingobium</i> sp. UBA5915	26	3	2	0
<i>Sphingomonas</i> sp. Bin.006_Forest	23	3	3	3

Tabela 2 – Presença de algumas hidrolases já identificadas na anotação dos genomas oriundos de sequenciamento de metagenomas amazônicos

Além das proteases identificadas pela anotação do genoma, proteínas hipotéticas também podem ser associadas com proteases ainda não conhecidas. Dessa forma, ao analisar as sequências de proteínas hipotéticas nas bases de dados do MEROPS e do RDP, foram identificadas 23 potenciais proteases/peptidases com tamanhos variando entre 122 e 535 aminoácidos. Dessas 23 potenciais proteases/peptidases, uma apresentou homologia com a superfamília da Peptidase S1 (clã PA), com função molecular predita de atividade endopeptidase tipo serina protease (Gene Ontology:0004252). Esta proteína hipotética com homologia à peptidase possui 449 aminoácidos e pertence à *Nocardioidekongjuensis* bin002Forest, com sequência MRRSLGVVAAAIALLTGLPVVAPVVAPALAE PQAAEPGPLAP ARHGAVVQDLESPAGTAYWTRRRRMRAALPLDLDP SGAVITSSADAAAPAQPAVTPPGLRS TGKLFRRDPRTSQGYVCSASSVNTPERNLVV TAGHCVVYSTRDGCVLLCTARHYFSDFVFP SYDHGTAPYGQWTGVRAITQQAWIADEDDGHDQAFLAVAPVGGVNLVDVVGGNGLAW NYPAREDGVRVVGWPAQAPYD GQSRQEC SGATTVSQVTDPTDAQISCLTGGASGGPW FLRMASADTGFVFAVTSRRPADGTPLLFAMPFDSSIETLLAAARSAAVPVSARVPARPGAR PRLRLVASAASVGFGESYQLVAETRVRRIVLQVRTVPGAPWRRVARARVHQGVTVFDQA LPPGTRWYRVRERGTTRRHSKPVVVTGACPLPLDRSPGVVSNTRCTSPVG. Ao analisar no signalP foi identificado que possui um peptídeo de sinal de 30 aminoácidos sublinhado na sequência acima.

Ao realizar o BLAST com a base de dados de proteínas não redundantes, foi identificado um domínio de peptidase tipo tripsina (WP215815756) similar a de *Pimelobacter* sp. que, assim como Nocardioide, é uma Actinomycetota (antiga Actinobacteria) da família Nocardioideaceae. O alinhamento dessas sequências (Figura 1) mostrou a similaridade entre as sequências de 48%, sendo que ao analisar aminoácidos quimicamente similares a similaridade sobe para 60%.

A identificação da superfamília e a similaridade de 60% considerando aminoácidos quimicamente similares demonstra o potencial para descoberta de uma potencial nova protease de origem amazônica ainda não estudada com aproximadamente 44,5 KDa.

Proteína hipotética (WP273517200)	59	AYWTRRRMRAALPLDLDPGAVITS-SADAAAPAQPAVTPPGLRSTGKLFRRDPRTSQGY	117
Peptidase tipo tripsina (WP215815756)	48	+YWT +RMR A P+D GA +A A RSTGKLF D + Y SYWTPQRMRAEPVDRSADGATGAGPTARRTTAGARARAVVAPRSTGKLFSD--GASDY	105
Proteína hipotética (WP273517200)	118	VCSASSVNTPERNLVVTAGHCVYSTRDGCVLLCTARHYFSDFVFVPSYDHGTAPYGQWGTG	177
Peptidase tipo tripsina (WP215815756)	106	VCS ++VNTPERN+V+TAGHCV S +L C YF+ F+FVP Y G PYG WGTG VCSGAAVNTPERNVLLTAGHCVNSGATRGLLGCRRGRYFTRFLFVPGYAGGARPYGAWTG	165
Proteína hipotética (WP273517200)	178	VRAITQQAWIADEDD-GHDQAF LAVAPVGGVNLVDVVGGNGLAWNYPAREDGVRVVGWPA	236
Peptidase tipo tripsina (WP215815756)	166	V A + Q W+ + HDQA L VAP+GG LVDVVGGN +AW YP REDGVRV+GWPA VTALAQPDWVNQCNGYSHDQAVL TVAPLGGRRLLVDVVGGNAVWGYPQREDEGVRVIGWPA	225
Proteína hipotética (WP273517200)	237	QAPYDQGSRQECSGATTVSQVTDPTDAQISCLPTGGASGGPWFLRMASADTGFVFAVTSR	296
Peptidase tipo tripsina (WP215815756)	226	+APYDG+S Q+C+G+T S + DA ++CPL GGASGGPWFL M SAD GF++AVTSR EAPYDGESAQQCTGST--SAFPETGDAHLACPLNGGASGGPWFLTMSADVGFIVAVTSR	283
Proteína hipotética (WP273517200)	297	RPADGTPLL FAMPFDSSIETLLAAARSAAVPVSARVPARPGARPR--LRLVASAASVGGF	354
Peptidase tipo tripsina (WP215815756)	284	R +G P+L A P D +I TLLAA RS A + AR R L + A A G G RTTEGPPILLAWPLDRITLTLAAVRSPAARAARPAQPAVVSARGRGLAVAAVPAVTGRG	343
Proteína hipotética (WP273517200)	355	ESYQLVAETRRRIVLQVRTVPGAPWRRVARARV-HQGVTVFDQ-ALPPGTRWYVRER	412
Peptidase tipo tripsina (WP215815756)	344	+ Y+L R ++V+QVR A W R+ R+ GV + A P G W+R+++ QPYRLRLRARPGTKVWVQVRRSASARWARLGAVRIPASGVVELTRPATVPGPAWFLRKLQ-	402
Proteína hipotética (WP273517200)	413	GTRRHSPVWVTVGACPLPLDRSPGVVSNTRCTSPV	448
Peptidase tipo tripsina (WP215815756)	403	+R K V V ACPLP DRS VV+ T CTSP -GKRVGKTKVWRACPLPADRS GAVVAATGCTSP	437

Figura 1 – Alinhamento entre sequências de aminoácidos da proteína hipotética do presente estudo de *Nocardioide*s e um domínio de peptidase tipo tripsina de *Pimelobacter* sp. (+) indica aminoácido quimicamente similar na posição entre as sequências. (-) indica um gap.

Este estudo preliminar evidencia a presença de potenciais novas proteases e peptidases nos microrganismos da Amazônia, destacando a relevância dessa região como fonte promissora de enzimas de interesse biotecnológico. A identificação específica de uma protease relacionada à superfamília da Peptidase S1, com similaridade a enzimas conhecidas, enfatiza a importância desses microrganismos na busca por novas ferramentas biotecnológicas. Além das proteases, a prospecção em genomas e metagenomas depositados em bases de dados pode revelar outras hidrolases e enzimas valiosas para diversas aplicações industriais. É interessante realizar futuras investigações para sintetizar e avaliar experimentalmente essas proteases, especialmente aquelas associadas a proteínas hipotéticas, a fim de confirmar a atividade proteolítica/peptídica e explorar seu potencial uso. Essa pesquisa não apenas destaca a diversidade proteolítica na Amazônia, mas também enfatiza a importância da colaboração e compartilhamento de dados para avançar nossa compreensão e aplicação sustentável desses recursos.

REFERÊNCIAS

AKAÇIN, İ.; ERSOY, Ş.; DOLUCA, O.; GÜNGÖRMÜŞLER, M. **Comparing the significance of the utilization of next generation and third generation sequencing technologies in microbial metagenomics.** *Microbiological Research*, v. 264, p. 127154, nov. 2022.

BARRETT, A. J.; RAWLINGS, N. D.; O'BRIEN, E. A. **The MEROPS Database as a Protease Information System.** *Journal of Structural Biology*, v. 134, n. 2–3, p. 95–102, maio 2001.

BARRIUSO, J.; MARTÍNEZ, M. J. **In Silico Metagenomes Mining to Discover Novel Esterases with Industrial Application by Sequential Search Strategies.** Journal of Microbiology and Biotechnology, v. 25, n. 5, p. 732–737, 28 maio 2015.

BAUMAN, K. D.; BUTLER, K. S.; MOORE, B. S.; CHEKAN, J. R. **Genome mining methods to discover bioactive natural products.** Natural Product Reports, v. 38, n. 11, p. 2100–2129, 2021.

BERGMANN, J. C.; COSTA, O. Y. A.; GLADDEN, J. M.; SINGER, S.; HEINS, R.; D'HAESELEER, P.; SIMMONS, B. A.; QUIRINO, B. F. **Discovery of two novel β -glucosidases from an Amazon soil metagenomic library.** FEMS Microbiology Letters, v. 351, n. 2, p. 147–155, fev. 2014.

BHATTACHARJEE, A.; SARMA, S.; SEN, T.; DEVI, M. V.; DEKA, B.; SINGH, A. K. **Genome mining to identify valuable secondary metabolites and their regulation in Actinobacteria from different niches.** Archives of Microbiology, v. 205, n. 4, p. 127, 21 abr. 2023.

CORNISH, K. A. S.; LANGE, J.; AEVARSSON, A.; POHL, E. **CPR-C4 is a highly conserved novel protease from the Candidate Phyla Radiation with remote structural homology to human vasohibins.** Journal of Biological Chemistry, v. 298, n. 5, p. 101919, maio 2022.

FICARRA, F. A.; SANTECCHIA, I.; LAGORIO, S. H.; ALARCÓN, S.; MAGNI, C.; ESPARIZ, M. **Genome mining of lipolytic exoenzymes from *Bacillus safensis* S9 and *Pseudomonas alcaliphila* ED1 isolated from a dairy wastewater lagoon.** Archives of Microbiology, v. 198, n. 9, p. 893–904, 7 nov. 2016.

GUAN, F.; HAN, Y.; YAN, K.; ZHANG, Y.; ZHANG, Z.; WU, N.; TIAN, J. **Highly efficient production of chitooligosaccharides by enzymes mined directly from the marine metagenome.** Carbohydrate Polymers, v. 234, p. 115909, abr. 2020.

HANDELSMAN, J.; RONDON, M. R.; BRADY, S. F.; CLARDY, J.; GOODMAN, R. M. **Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products.** Chemistry & Biology, v. 5, n. 10, p. R245-9, out. 1998.

KUMAR, S.; KRISHNANI, K. K.; BHUSHAN, B.; BRAHMANE, M. P. **Metagenomics: Retrospect and Prospects in High Throughput Age.** Biotechnology Research International, v. 2015, p. 1–13, 17 nov. 2015.

MANDRO, J. A.; NAKAMURA, F. M.; GONTIJO, J. B.; TSAI, S. M.; VENTURINI, A. M. **Metagenome-Assembled Genomes from Amazonian Soil Microbial Consortia.** Microbiology Resource Announcements, v. 11, n. 11, 17 nov. 2022.

NIELSEN, H.; BRUNAK, S.; HEIJNE, G. VON. **Machine learning approaches for the prediction of signal peptides and other protein sorting signals.** Protein Engineering, Design and Selection, v. 12, n. 1, p. 3–9, jan. 1999.

PANDEY, M.; SINGHAL, B. **Metagenomics: adding new dimensions in bioeconomy.** Biomass Conversion and Biorefinery, v. 13, n. 9, p. 7461–7480, 19 jun. 2023.

PARKS, D. H.; RINKE, C.; CHUVOCHINA, M.; CHAUMEIL, P.-A.; WOODCROFT, B. J.; EVANS, P. N.; HUGENHOLTZ, P.; TYSON, G. W. **Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life.** Nature microbiology, v. 2, n. 11, p. 1533–1542, nov. 2017.

PAYSAN-LAFOSSE, T. *et al.* **InterPro in 2022**. *Nucleic Acids Research*, v. 51, n. D1, p. D418–D427, 6 jan. 2023.

PEREIRA, J. O.; SOUZA, A. Q. L. DE; SOUZA, A. D. L. DE; CASTRO FRANÇA, S. DE; OLIVEIRA, L. A. DE. **Overview on Biodiversity, Chemistry, and Biotechnological Potential of Microorganisms from the Brazilian Amazon**. *Em: Diversity and Benefits of Microorganisms from the Tropics*. Cham: Springer International Publishing, 2017. p. 71–103.

VENTURINI, A. M.; GONTIJO, J. B.; MANDRO, J. A.; PAULA, F. S.; YOSHIURA, C. A.; FRANÇA, A. G. DA; TSAI, S. M. **Genome-resolved metagenomics reveals novel archaeal and bacterial genomes from Amazonian forest and pasture soils**. *Microbial genomics*, v. 8, n. 7, jul. 2022.