

# ANOTAÇÃO FUNCIONAL DE SEQUÊNCIAS GENÔMICAS DE POLIHIDROXIALCANOATO EM *PRIESTIA MEGATERIUM* ISOLADA DE SOLO

Data de aceite: 01/07/2024

### **Ryan Fernandes Vieira de Souza**

Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados  
Dourados - Mato Grosso Do Sul  
<https://lattes.cnpq.br/5252229894109187>

### **Kauanny Victoria Plenz**

Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados  
Dourados - Mato Grosso Do Sul  
<http://lattes.cnpq.br/6400558585644143>

### **Mateus Marques Souza**

Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados  
Dourados - Mato Grosso Do Sul  
<http://lattes.cnpq.br/7525718845564368>

### **Wellinton Jhon Cupozak Pinheiro**

Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados  
Dourados - Mato Grosso Do Sul  
<http://lattes.cnpq.br/6370351706655222>

### **Maricy Raquel Lindenbah Bonfá**

Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados  
Dourados - Mato Grosso do Sul

<http://lattes.cnpq.br/5670504878145026>

### **Rodrigo Matheus Pereira**

Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados  
Dourados - Mato Grosso do Sul  
<http://lattes.cnpq.br/8155952045293310>

**RESUMO:** Os plásticos convencionais derivados do petróleo causam sérios problemas ambientais, devido principalmente à poluição e à degradação lenta. Alternativas sustentáveis, como os polihidroxicanoatos (PHAs), que são produzidos por microrganismos como *Priestia megaterium* (anteriormente conhecido como *Bacillus megaterium*), são vistos como substitutos promissores desses materiais pelas suas capacidades de biodegradabilidade e biocompatibilidade. A genômica e a bioinformática adentram esse cenário desempenhando um papel essencial na identificação de genes e proteínas de PHAs no genoma da bactéria *Priestia megaterium* cepa E1, isolada a partir de amostras de solo. As sequências genômicas da bactéria em questão foram anotadas pelas ferramentas de anotação

funcional Prokka e eggNOG-mapper e aquelas sequências com anotação de PHA foram submetidas a alinhamentos locais em bancos de dados secundários Swiss-Prot e Reference Sequence Protein. As correspondências das anotações feitas nessas diferentes ferramentas foram observadas através do Diagrama de Venn. Em seguida, foi realizada uma modelagem 3D das proteínas PhaC sintases pelas ferramentas Phyre 2 e Swiss Model. Com isso, foram anotadas 25 sequências de PHA, distribuídas entre diferentes proteínas como PhaA, PhaB e PhaI, entre outras. As modelagens 3D obtiveram resultados que se assemelham aos estudos já realizados, o que mostrou mais confiabilidade ao processo de anotação e corroborou com evidências científicas na espécie como produtora de PHA.

**PALAVRA CHAVE:** PHA; genômica; enzimas.

## FUNCTIONAL ANNOTATION OF GENOMIC SEQUENCES OF POLYHYDROXYALKANOATE IN *PRESTIA MEGATERIUM* ISOLATED FROM SOIL

**ABSTRACT:** Conventional plastics derived from petroleum cause serious environmental problems, mainly due to pollution and slow degradation. Sustainable alternatives, such as polyhydroxyalkanoates (PHAs), which are produced by microorganisms such as *Priestia megaterium* (formerly known as *Bacillus megaterium*), are seen as promising substitutes for these materials due to their biodegradability and biocompatibility capabilities. Genomics and bioinformatics enter this scenario, playing an essential role in identifying PHA genes and proteins in the genome of the bacterium *Priestia megaterium* strain E1, isolated from soil samples, for biotechnological application to the bacterium. The genomic sequences of the bacterium in question were annotated by the functional annotation tools Prokka and eggNOG-mapper and those sequences with PHA annotation were subjected to local alignments in databases. Swiss-Prot and Reference Sequence Protein. The correspondences of the notes made in these different tools were observed using the Venn Diagram. Next, 3D modeling of the PhaC synthase proteins was carried out using the Phyre 2 and Swiss Model tools. As a result, 25 PHA sequences were noted, distributed among different proteins such as PhaA, PhaB and PhaI, among others. The 3D modeling obtained results that are similar to studies already carried out, which showed more reliability to the annotation process and corroborated scientific evidence on the species as a producer of PHA. and corroborated scientific evidence on the species as a producer of PHA.

**KEYWORD:** PHA; genomics; enzymes.

## 1 | INTRODUÇÃO

Os plásticos derivados de petróleo são amplamente utilizados devido a suas propriedades vantajosas. No entanto, o seu uso extensivo tem levantado sérias preocupações ambientais, dado que a produção e o descarte inadequados contribuem à poluição e impactos adversos ao meio ambiente. Além disso, a decomposição lenta dos plásticos convencionais, que podem persistir no ambiente por séculos, agrava ainda mais o cenário (Melchor Martínez et al., 2022).

Diante dessas preocupações, cresce o interesse em alternativas sustentáveis, como os bioplásticos, derivados de fontes renováveis, como amido e celulose, ou produzidos

por microrganismos (Kumar et al., 2023). Os PHAs destacam-se nesse cenário pela sua capacidade de biodegradabilidade e biocompatibilidade, apresentando-se como alternativas ecologicamente viáveis aos plásticos descartáveis (SUPs), como polietileno e propileno (Wen et al., 2021). Essa classe de poliésteres, produzida por microrganismos como *Priestia megaterium* (anteriormente conhecido como *Bacillus megaterium*) (Liu et al., 2023), apresenta uma gama versátil de propriedades, tendo diversas aplicações biotecnológica, desde tecnologia na área da saúde até embalagens e na agricultura (Jaffur et al., 2023).

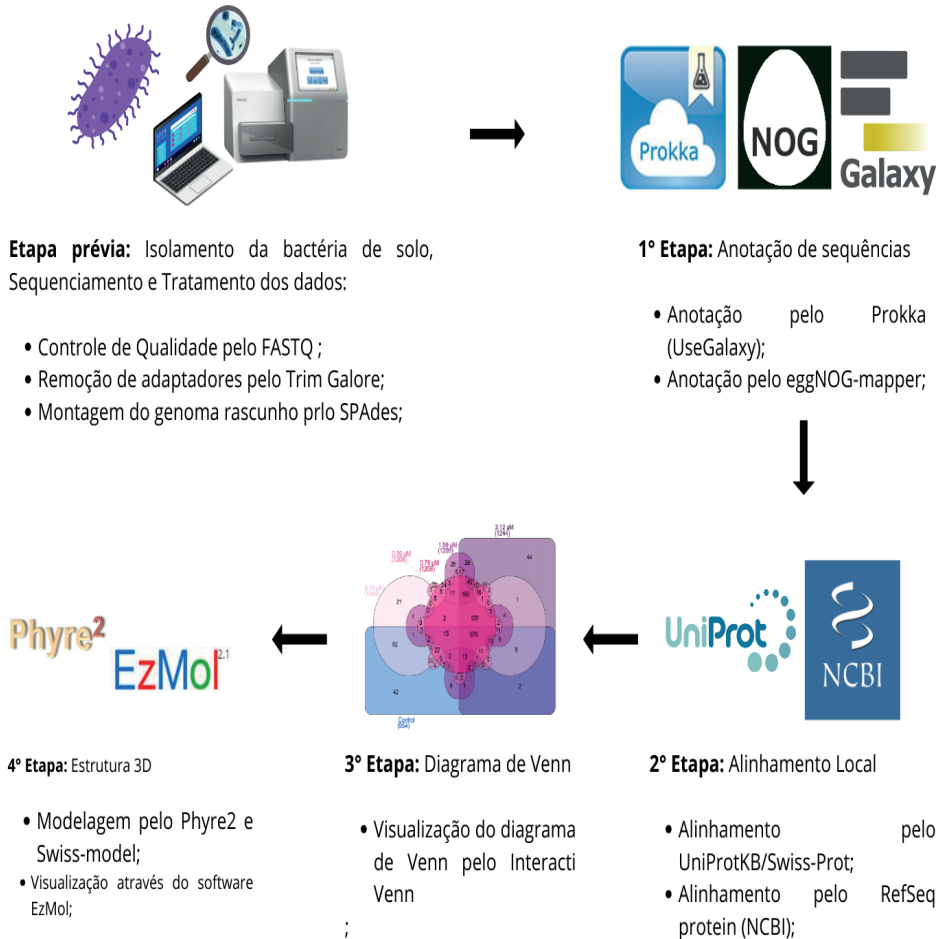
A biossíntese de PHA é mediada por uma série de enzimas, como PhaA, PhaI, PhaB, entre outras, as quais são transcritas por genes que são homólogos entre diferentes organismos. À medida que a tecnologia avança, surgem novas metodologias para a prospecção dessas enzimas e os genes relacionados com as suas sínteses. A genômica, que envolve o estudo do material genético completo de um organismo, destaca-se como uma abordagem bastante abrangente (Yu et al., 2022). Essa área da bioinformática proporciona uma visão ampla e detalhada da composição de um genoma, através de anotações, alinhamentos e análises filogenéticas, permitindo uma identificação mais precisa de enzimas e potenciais alvos biotecnológicos (Wang, 2023).

Nesse contexto, diversos programas desempenham papéis essenciais, entre eles o PROKKA (Seemann 2014; Afgan et al., 2018) e o Egg-Nog mapper (Cantalapiedra et al., 2021) são amplamente utilizados para executar tarefas específicas para identificar e anotar sequências de interesses. O uso destas diferentes ferramentas, juntamente com alinhamentos locais feitos em bancos de dados secundários, permite a anotação funcional de sequências genômicas de um organismo específico, possibilitando identificar o seu potencial biotecnológico. Isso permite identificar possíveis genes e proteínas relacionadas ao metabolismo dos polihidroxialcanoatos presentes no genoma da bactéria *Priestia megaterium* cepa E1 isolada de solo.

## 2 | MATERIAL E MÉTODOS

A bactéria *Priestia megaterium* cepa E1 foi isolada do solo da Fazenda Experimental da Universidade Federal da Grande Dourados (UFGD) (Prado 2017). Seu DNA genômico foi extraído e sequenciado pelo, seguindo os protocolos do fabricante. Os dados foram processados, incluindo controle de qualidade com FASTQ (Cock et al., 2010), remoção de adaptadores pelo Trim Galore (Krueger, 2021) e montagem do genoma rascunho com SPAdes (Bankevich et al., 2012), usando parâmetros padrão. O estudo iniciou com a anotação funcional dessas sequências genômicas.

**Figura 1.** Fluxograma das atividades de bioinformática desenvolvidas no presente trabalho.



Utilizando o software Prokka v. 1.14.6 (Torsten 2014; Afgan et al., 2018) com parâmetros padrões, foi realizada a predição e anotação funcional das seqüências montadas do sequenciamento da bactéria *Priestia megaterium* cepa E1. Um conjunto de seqüências de aminoácidos foi exportado em formato Excel (.xlsx). Pesquisas com termos como “PHA” e “PHB” foram realizadas, e as seqüências encontradas foram selecionadas. Essas seqüências foram analisadas pelo eggNOG-mapper v. 2 (Cantalapiedra, C. P. et al., 2021) para confirmação das anotações. Em seguida, as seqüências foram alinhadas nos bancos de dados Swiss-prot v. 7.0 (Bairoch; Apweiler, R, 2000) e RefSeq protein v. 259 (Pruitt et al., 2007), usando parâmetros padrão. As anotações das quatro ferramentas foram inseridas na plataforma Interacti Venn (<https://www.interactivenn.net>, Heberle et al., 2015).

Por fim, as seqüências das proteínas PhaC sintase foram modeladas usando os softwares Phyre v. 2.0 (Kelley et al., 2015) e Swiss-model v. 2024.2 (Schwede et al., 2003),

que utilizam o modelo oculto de Markov. Após cada análise, foi gerado um arquivo PDB (Protein Data Bank), visualizado no software EzMol v. 2.1 (Reynolds, 2018) para observação da estrutura nos formatos nuvem eletrônica e cartoon. As outras proteínas anotadas não foram modeladas devido à falta de artigos científicos que prevejam suas estruturas 3D.

### 3 | RESULTADOS E DISCUSSÕES

O PROKKA retornou um conjunto de 23.145 sequências anotadas a partir do genoma rascunho da bactéria *Priestia megaterium* cepa E1 (Tabela 1). Dentre elas, foram selecionadas 25 sequências de PHA, incluindo sete proteínas PhaC sintases, cinco Acetyl-CoA transferases (PhaA), três Acetoacetyl-CoA reductases (PhaB), duas Enoyl-CoA hidratases (PhaJ) e oito proteínas hipotéticas, que foram anotadas por ferramentas posteriores. Os tamanhos dessas sequências variam de 246 a 1683 aminoácidos. A PhaC sintase é a principal enzima no processo de produção de PHA, presente em todos os microrganismos bioacumuladores de PHA, transformando monômeros de (R)-3-hydroxybutyryl-CoA no produto final de PHA (Neoh et al., 2022).

Outras enzimas importantes são PhaA e PhaB, que atuam na via I de produção de PHA, fornecendo monômeros de (R)-3-hydroxybutyryl-CoA às PhaC sintases (Madison et al., 2022). Proteínas da superfamília MaoC, como a PhaJ (Enoyl-CoA hidratase), também são cruciais na produção de PHA, pois atuam na rota metabólica da  $\beta$ -oxidação, fornecendo monômeros para a biossíntese de PHA (Wan 2023).

**Tabela 1:** As 25 sequências filtradas com anotação de PHA feitas pela ferramenta PROKKA.

Locus_tag	PROKKA		
	Tamanho	Descrição	Genes
EAMNKLGA_05793	1185	Acetyl-CoA transferase	<i>phaA_1</i>
EAMNKLGA_05917	1086	Poly(3-hydroxyalkanoate) polymerase PhaC	<i>phaC_1</i>
EAMNKLGA_07463	1683	Poly(3-hydroxyalkanoate) polymerase PhaC	<i>phaC_2</i>
EAMNKLGA_07465	1680	Poly(3-hydroxyalkanoate) polymerase PhaC	<i>phaC_3</i>
EAMNKLGA_10818	1179	Acetyl-CoA transferase	<i>phaA_2</i>
EAMNKLGA_13058	1644	Poly(3-hydroxyalkanoate) polymerase PhaC	<i>phaC_4</i>
EAMNKLGA_13061	747	Acetoacetyl-CoA reductase	<i>phaB_1</i>
EAMNKLGA_13677	246	Acetyl-CoA transferase	<i>phaA_3</i>
EAMNKLGA_13678	261	Acetyl-CoA transferase	<i>phaA_4</i>
EAMNKLGA_16837	741	Acetoacetyl-CoA reductase	<i>phaB_2</i>
EAMNKLGA_17817	471	(R)-specific enoyl-CoA hydratase	<i>phaJ_1</i>
EAMNKLGA_17957	741	Acetoacetyl-CoA reductase	<i>phaB_3</i>
EAMNKLGA_20212	408	(R)-specific enoyl-CoA hydratase	<i>phaJ_2</i>
EAMNKLGA_20844	1182	Acetyl-CoA transferase	<i>phaA_5</i>
EAMNKLGA_22690	1152	Poly(3-hydroxyalkanoate) polymerase PhaC	<i>phaC_5</i>

EAMNKLGA_22716	1065	Poly(3-hydroxyalkanoate) polymerase PhaE	<i>phaE</i>
EAMNKLGA_22717	1068	Poly(3-hydroxyalkanoate) polymerase PhaC	<i>phaC_6</i>
EAMNKLGA_17958	552	hypothetical protein	-
EAMNKLGA_05951	456	hypothetical protein	-
EAMNKLGA_07460	411	hypothetical protein	-
EAMNKLGA_13384	348	hypothetical protein	-
EAMNKLGA_16952	492	hypothetical protein	-
EAMNKLGA_07461	972	hypothetical protein	-
EAMNKLGA_07464	1005	hypothetical protein	-
EAMNKLGA_11085	885	hypothetical protein	-

O eggNOG-mapper confirmou as anotações feitas pelo PROKKA de 17 sequências de PHA e das 8 sequências inicialmente apresentadas como hipotéticas. Essas foram identificadas como uma proteína repressora (PhaR), cinco proteínas fasinias (PhaI) e duas PHB despolimerases (PhaZ), conforme mostra a Tabela 2. No geral, as 25 sequências apresentaram boas pontuações de *e-value* e *score*, indicando a baixa probabilidade de ocorrência ao acaso e a alta qualidade das correspondências, respectivamente (Pearson et al., 2013).

**Tabela 2:** As 25 sequências filtradas com anotações de PHA feitas pelo PROKKA e Egg-NOG-mapper.

Locus_tag	Lenght_pb	PROKKA		Egg-NOG mapper	
		Genes	Genes	<i>E-value</i>	<i>Score</i>
EAMNKLGA_05793	1185	<i>phaA_1</i>	<i>phbA</i>	1.77e-280	767.0
EAMNKLGA_05917	1086	<i>phaC_1</i>	<i>phaC</i>	4.58e-248	682.0
EAMNKLGA_07463	1683	<i>phaC_2</i>	<i>phaC2</i>	0.0	1100.0
EAMNKLGA_07465	1680	<i>phaC_3</i>	<i>phaC1</i>	0.0	1117.0
EAMNKLGA_10818	1179	<i>phaA_2</i>	<i>atoB</i>	2.07e-262	721.0
EAMNKLGA_13058	1644	<i>phaC_4</i>	<i>phaC</i>	7e-268	749.0
EAMNKLGA_13061	747	<i>phaB_1</i>	<i>phaB</i>	2.4e-141	403.0
EAMNKLGA_13677	246	<i>phaA_3</i>	<i>phbA</i>	7.74e-22	94.4
EAMNKLGA_13678	261	<i>phaA_4</i>	<i>phaA</i>	7.32e-33	124.0
EAMNKLGA_16837	741	<i>phaB_2</i>	<i>phbB</i>	5.41e-171	478.0
EAMNKLGA_17817	471	<i>phaJ_1</i>	<i>phaJ</i>	6.1e-101	293.0
EAMNKLGA_17957	741	<i>phaB_3</i>	<i>phbB</i>	1.39e-173	484.0
EAMNKLGA_20212	408	<i>phaJ_2</i>	<i>phaJ</i>	3.6e-92	269.0
EAMNKLGA_20844	1182	<i>phaA_5</i>	<i>atoB</i>	5.6e-272	745.0
EAMNKLGA_22690	1152	<i>phaC_5</i>	<i>phbC</i>	8.12e-285	777.0
EAMNKLGA_22716	1065	<i>phaE</i>	-	1.68e-219	609.0
EAMNKLGA_22717	1068	<i>phaC_6</i>	<i>phbC</i>	2.4e-256	702.0

EAMNKLGA_17958	552	-	<i>phaR</i>	1.2e-106	310.0
EAMNKLGA_05951	456	-	<i>phal</i>	9.11e-96	280.0
EAMNKLGA_07460	411	-	<i>phal</i>	1.16e-84	250.0
EAMNKLGA_13384	348	-	<i>phal</i>	4.05e-35	124.0
EAMNKLGA_16952	492	-	<i>phal</i>	4.76e-106	306.0
EAMNKLGA_07461	972	-	<i>phaF</i>	1.52e-71	230.0
EAMNKLGA_07464	1005	-	<i>phaZ</i>	5.52e-240	659.0
EAMNKLGA_11085	885	-	<i>phaZ</i>	0.0	885.0

As proteínas Phal pertencem à família das proteínas fasinás que são codificadas pelos genes *phal* e *phaF* (Choi, 2020). Elas se encontram associadas à superfície do grânulo e são responsáveis pela estabilização do PHA (Aghaali et al., 2023). Já as PHA despolimerases (PhaZ) são expressas para degradar os grânulos presentes no citoplasma da célula, geralmente pelo gene *phaZ* (Zhou et al., 2023). Outras proteínas também bastante comuns são as PhaR, codificadas pelo gene *phaR*, que atua reprimindo a síntese de PHA, geralmente associada à superfície do grânulo de PHA (Aghaali et al., 2023).

Os alinhamentos local no banco de dados secundário SwissProt (UniProt) das 25 sequências de PHA apresentou, de modo geral, valores de identidades superiores a 60%, com exceção de seis sequências com identidade inferior a 50%. Já o alinhamento local no banco de dados no banco de dados RefSeq\_protein (NCBI) das mesmas 25 sequências apresentaram boas identidades e porcentagem de positivos superior a 83%. Além disso, as proteínas que retornaram melhores similaridades confirmam as anotações e o alinhamento feito no banco de dados Swiss-Prot, como mostra a Tabela 4. Além disso, foi montado uma tabela consenso (Tabela 3) que exhibe as correspondências de anotações feitas pelas quatro ferramentas utilizadas até então.

**Tabela 3.** Anotações correspondentes entre as ferramentas de anotação funcional, Prokka e eggNOG-mapper, e os bancos de dados secundários, Swiss-prot e RefSeq Protein.

Genes	Cópias	Proteínas	Prokka	eggNOG-mapper	Swiss-prot	RefSeq Protein
<i>phaA</i>	5	PhaA	•	•	•	•
<i>phaB</i>	3	PhaB	•	•	•	•
<i>phaC</i>	7	PhaC	•	•	•	•
<i>phaJ</i>	2	PhaJ	-#	•	•	•
<i>phal</i>	5	Proteínas fasinás	•	•	•	•
<i>phaR</i>	1	Proteína reguladora	-#	•	•	•
<i>phaZ</i>	2	PhaZ	-#	•	•	•

• - Sequência que obteve anotação correspondente entre todas as ferramentas; -# - Sequência que não foi anotada pela ferramenta.

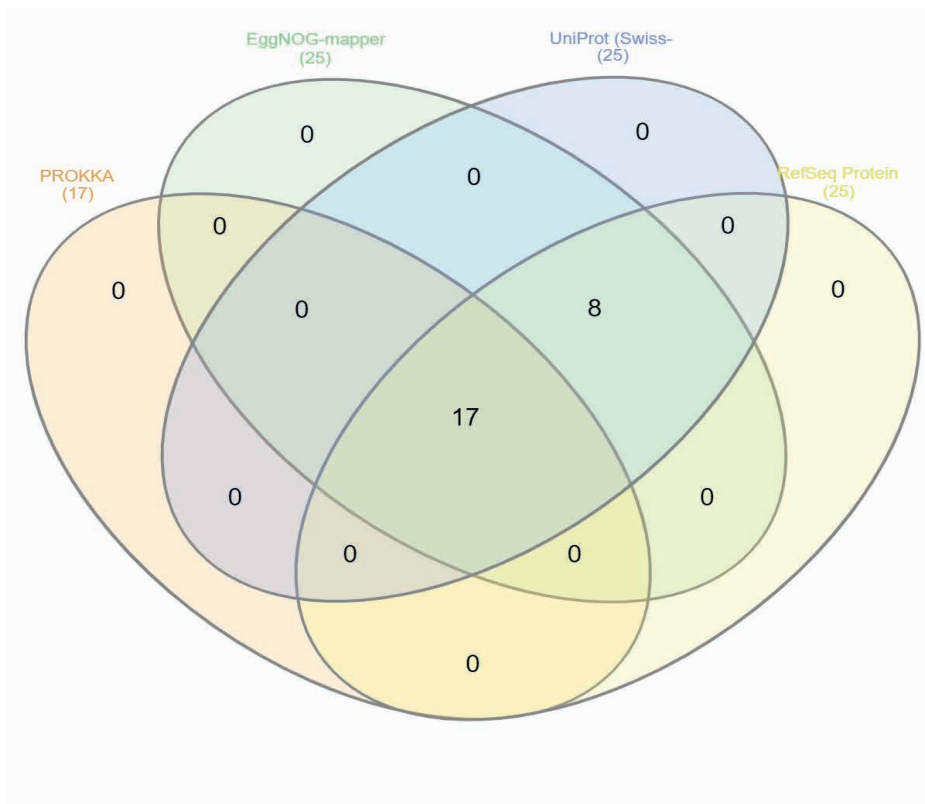
**Tabela 4:** As 25 sequências filtradas com anotação de PHA alinhadas nos bancos de dados secundários **RefSeq** **Proteine** **Swiss-Prot**.

Query	RefSeq Protein		Swiss-Prot	
	Descrição	Positivos	Descrição	Identidade
EAMNKLGA_05793	Acetyl-CoA C-acetyltransferase	100%	Belongs to the thiolase family	78.6%
EAMNKLGA_05917	Alpha/beta fold hydrolase	98%	alpha/beta hydrolase fold	27.7%
EAMNKLGA_07463	Class II PhaC synthase	98%	PHAB polymerase (PhaC)	70.7%
EAMNKLGA_07465	Class II PhaC synthase	97%	poly(R)-hydroxyalkanoic acid synthase	82.1%
EAMNKLGA_10818	Acetyl-CoA C-acetyltransferase	96%	Belongs to the thiolase family	83.7%
EAMNKLGA_13058	Alpha/beta fold hydrolase	78%	Poly-beta-hydroxybutyrate polymerase	39.5%
EAMNKLGA_13061	Acetoacetyl-CoA reductase	94%	Acetoacetyl-CoA reductase	56.6%
EAMNKLGA_13677	Acetyl-CoA C-acetyltransferase	86%	Belongs to the thiolase family	60%
EAMNKLGA_13678	Acetyl-CoA C-acyltransferase	83%	Belongs to the thiolase family	69.8%
EAMNKLGA_16837	Acetoacetyl-CoA reductase	100%	Acetoacetyl-CoA reductase	64.2%
EAMNKLGA_17817	MaoC family dehydratase	99%	N-terminal half of MaoC dehydratase	45%
EAMNKLGA_17957	Acetoacetyl-CoA reductase	98%	Dehydrogenases reductases	53.7%
EAMNKLGA_20212	MaoC family dehydratase	92%	Acyl dehydratase	33.8%
EAMNKLGA_20844	Acetyl-CoA C-acetyltransferase	99%	Coenzyme A transferase	78.6%
EAMNKLGA_22690	Alpha/beta fold hydrolase	99%	Poly-beta-hydroxybutyrate(PhaC)	29.4%
EAMNKLGA_22716	Class III PhaC subunit PhaE	95%	PhaC synthase subunit PhaE PhaE	26.3%
EAMNKLGA_22717	Class III PhaC	99%	PhaC synthase subunit PhaE PhaE	99.2%
EAMNKLGA_17958	PHA repressor PhaR	97%	synthesis repressor, PhaR	95.6%
EAMNKLGA_05951	Phasin family protein	100%	PHA granule associated protein	93.4%
EAMNKLGA_07460	Phasin family protein	99%	PHA granule associated protein	94.1%
EAMNKLGA_13384	Phasin family protein	98%	Phasin protein	91.3%
EAMNKLGA_16952	Phasin family protein	99%	PHA granule associated protein	96.9%
EAMNKLGA_07461	Phasin family protein	98%	Granule-associated protein	52.6%
EAMNKLGA_07464	PHA depolymerase	98%	PHA depolymerase	98.6%
EAMNKLGA_11085	PHA depolymerase	99%	PHA de-polymerase	99.1%



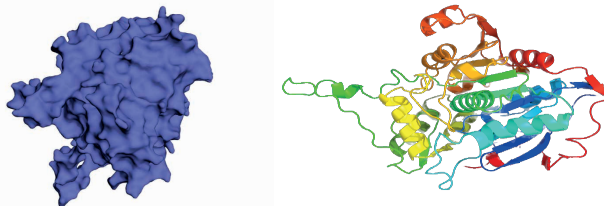
Vale ressaltar que nem todas as ferramentas forneceram o nome específico de cada proteína e seus genes codificadores, mas ao menos indicaram a família a que pertencem. O Diagrama de Venn (Figura 1) mostra as correspondências das 25 sequências analisadas. Das 25, 17 apresentaram resultados comuns nas quatro ferramentas, enquanto oito sequências foram exclusivas das ferramentas EggNOG-mapper, Swiss-Prot (UniProt) e RefSeq protein (NCBI). O PROKKA não anotou essas oito sequências, classificadas como hipotéticas, ou seja, previstas pela análise genômica sem função biológica determinada experimentalmente. As outras três ferramentas identificaram essas oito sequências como uma proteína repressora (PhaR), cinco proteínas fasinias (PhaI) e duas proteínas despolimerases (PhaZ).

**Figura 1.** Representação gráfica da interação das 25 sequências do genoma da bactéria *Priestia megaterium* cepa E1 nos programas PROKKA, EggNOG-mapper, UniProt (Swiss-Prot) e RefSeq Protein.

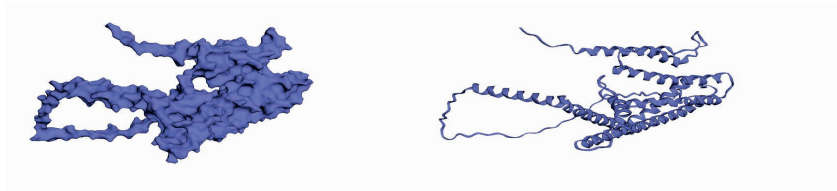


A modelagem 3D das PhaC sintases pelo Phyre2 (Figura 2) resultou em seis sequências modeladas com 100% de confiança e uma cobertura variando de 67% a 87%. No entanto, a sequência da subunidade PhaE (EAMNKLGA\_22716) não alcançou um nível significativo de cobertura e confiança. Por isso, foi empregado o Swiss-model para modelar essa sequência (Figura 3).

**Figura 2.** Representação da estrutura 3D de seis seqüências de PhaC sintase do genoma da bactéria *Priestia megaterium* cepa E1, modeladas pelo programa Phyre2. A visualização é apresentada no programa EzMol, exibindo o modelo de nuvem de elétrons à esquerda e o modelo cartoon à direita.

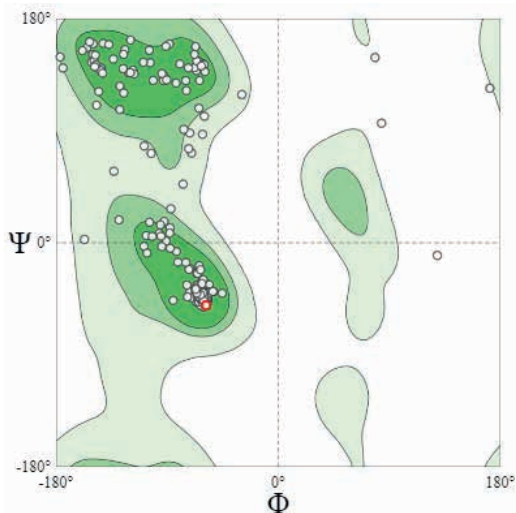


**Figura 3.** Visualização da modelagem da proteína subunidade PhaE (EAMNKLGA\_22716) de classe III do genoma da bactéria *Priestia megaterium* cepa E1 realizada pelo programa Swiss-model exibida no programa EzMol no modelo de nuvem de elétrons, à esquerda, e no modelo cartoon, à direita.



A ferramenta Swiss-model não exibe o valor de confiança, ela apresenta a confiabilidade da modelagem através do gráfico de Ramachandran (Figura 4), que apresentou 91.84% de resíduos dentro da zona de favorecimento à modelagem, enquanto outros 8%, aproximadamente, esteve fora dessa zona, indicando possíveis problemas na estrutura (Schwede, T. et al., 2003).

**Figura 3.** Gráfico de Ramachandran da modelagem da proteína subunidade PhaE de poli(3-hidroxicanoato) polimerase de classe III (EAMNKLGA\_22716) realizada pelo Swiss-model.



A modelagem das PhaC sintases apresentaram estruturas de alfa-hélice e folha-beta pregueadas em todas as sequências, uma estrutura característica de uma dobra  $\alpha/\beta$ -hydrolase, semelhante à das esterases (Jendrossek, 2009). Composta por domínios C-terminal e N-terminal, as estruturas cristalográficas do domínio catalítico C-terminal foram resolvidas com sucesso em algumas PhaC sintases (Wittenborn et al., 2016). No entanto, a função exata do domínio N-terminal ainda não foi determinada devido à dificuldade em visualizar sua estrutura por meio da cristalização. Além disso, a modelagem da proteína subunidade PhaE também demonstrou semelhança com as estruturas das PhaC sintases, como apontado por Neoh e colaboradores (2022).

## 4 | CONCLUSÃO

Foram anotadas 25 sequências de polihidroxialcanoato (PHA) no genoma da espécie *Priestia megaterium* cepa E1. Estas sequências foram anotadas como cinco referentes a PhaA (Acetyl-CoA transferase), sete PhaC (Poly(3-hydroxyalkanoate) polymerase), três PhaB (Acetoacetyl-CoA reductase), duas PhaJ (Enoyl-CoA hydratase), cinco Phal (proteínas fasinase), uma PhaR (proteína repressora) e duas PhaZ (poly(3-hydroxyalkanoate) depolymerase).

As anotações automáticas realizadas pelos programas PROKKA e eggNOG-mapper demonstraram similaridades entre as anotações feitas das 25 sequências nos dois programas. Os alinhamentos realizados nos bancos de dados Swiss-Prot (UniProt) e RefSeq Protein com as sequências do genoma da bactéria *Priestia megaterium* cepa E1 apresentaram valores bons de e-value e fortes identidades, com exceção de duas sequências (EAMNKLGA\_13677 e EAMNKLGA\_13678) anotadas como sendo PhaAs, que tiveram os piores valores de score, e-value e identidade - possivelmente sejam sequências parciais. Além disso, as modelagens 3D das PhaC sintases feitas pelas ferramentas Phyre2 e Swiss-model demonstraram boa confiabilidade e se assemelham com estudos já feitos. Todos esses resultados e indícios, forneceram mais confiabilidade ao processo de anotação de sequências genômicas de PHAs feita na bactéria *Priestia megaterium* cepa E1 isolada, o que corrobora com evidências científicas já feitas na espécie desde 1926.

## REFERÊNCIAS

Afgan, E.; Baker, D.; Batut, B.; Beek, M.; Bouvier, D.; Cech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B. A.; Guerler, A.; Hillman-Jackson, J.; Hiltmann, S.; Jalili, V.; Rasche, H.; Soranzo, N.; Goecks, J.; Taylor, J.; Nekrutenko, A.; Blankenberg, D. **The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update**. *Nucleic Acids Res.* 46(W1):W537-W544. 2018.

Aghaali, Z.; Naghavi, M. R. **Biotechnological Approaches for Enhancing Polyhydroxyalkanoates (PHAs) Production: Current and Future Perspectives**. *Curr Microbiol* 80, 345. 2023.

Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; Pyshkin, A. V.; Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A. **SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing**. *J Comput Biol*. 19(5):455-77. 2012.

Bairoch, A.; Apweiler, R. **The SWISS-PROT protein sequence database and its supplement TrEMBL**. *Nucleic Acids Res*. 1;28(1):45-8. 2000.

Cantalapiedra, C. P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. **eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale**. *Mol Biol Evol*. 9;38(12):5825-5829. 2021.

Choi, S. Y.; Cho, I. J.; Lee, Y.; Kim, Y. J.; Kim, K. J.; Lee, S. Y. **Microbial Polyhydroxyalkanoates and Nonnatural Polyesters**. *Adv. Mater*. 2020, 32, e1907138.

Cock, P. J.; Fields, C. J.; Goto, N.; Heuer, M. L.; Rice, P. M. **The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants**. *Nucleic Acids Res*.38(6):1767-71. 2010.

Jaffur, B. N.; Kumar, G; Pratima; Seeram, J; Bhatia, S. C. **Current advances and emerging trends in sustainable polyhydroxyalkanoate modification from organic waste streams for material applications, International**. *Journal of Biological Macromolecules*, Volume 253, 2023

Jendrossek, D. Polyhydroxyalkanoate granules are complex subcellular organelles (carbonosomes). *J. Bacteriol*. 191 (10), 3195-3202. 2009.

Krueger, F. **Trim Galore**. In GitHub repository. 2021. Acesso em: <<https://github.com/FelixKrueger/TrimGalore.com>> Acessado em: 17 Apr 2024.

Kumar, R.; Dongyi, L.; Liwen, L.; Manu, M. K.; Jun, Z.; Rajeshwar, D. T.; Jonathan, W.C. **Genome-centric polyhydroxyalkanoate reconciliation reveals nutrient enriched growth dependent biosynthesis in Bacillus cereus IBA1**. *Bioresource Technology*. Volume 382 , 129210, ISSN 0960-8524. 2023.

Madison, L. L.; Huisman, G. W. **Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic**. *Microbiol*. 63 ,21–53.1999.

Melchor Martínez, E. M.; Macias-Garbett, R.; Malacara, A.; Iqbal, H; Sosa-Hernández, J.; Parra, R. **Environmental impact of emerging contaminants from battery waste: A mini review**. *Case Studies in Chemical and Environmental Engineering*. 2021.

Neoh, S. Z; Chek, M. Z; Hua, T. T.; Javier, A.; Ardra, N. **Polyhydroxyalkanoate synthase (PhaC): The key enzyme for biopolyester synthesis**. *Current Research in Biotechnology*, Volume 4, Pages 87-101, ISSN 2590-2628. 2022.

Prado, C. C. A. do. **Degradação de fipronil por bactérias bioprospectadas de solo**. 2017. 69 f. TCC (Graduação) – Curso de Biotecnologia, FCBA Universidade Federal da Grande Dourados, Dourados, 2017. Disponível em: <https://repositorio.ufgd.edu.br/jspui/bitstream/prefix/3244/1/CaioCesarAchillesPrado.pdf>. Acesso em: 20 nov. 2023.

Pearson, W. R. **An introduction to sequence similarity (“homology”) searching**. *Curr Protoc Bioinformatics*. Chapter 3:3.1.1-3.1.8. 2013.

Seemann, T. **Prokka: rapid prokaryotic genome annotation**. *Bioinformatics*, Volume 30, Issue 14, Pages 2068–2069. 2014.

Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. **Nucleic Acids Res.** 1;31(13):3381-5. 2003.

Torsten, S. **Prokka: anotação rápida do genoma procariótico**. *Bioinformatics*, Volume 30, Edição 14, páginas 2068–2069. julho de 2014.

Wan, J. H.; Ng, L. M.; Neoh, S. Z. **Sequência completa do genoma de Aquitalea pelogenes USM4 (JCM19919), produtor de polihidroxialcanoato**. *Arch Microbiol* 205, 66. 2023.

Wang, R. C.; Wang, Z. **Precision Medicine: Disease Subtyping and Tailored Treatment**. *Cancers* (Basel). 2023 Jul 28;15(15):3837.

Wen, Q.; Liu, B.; Li, F.; Chen, Z. **Substrate strategy optimization for polyhydroxyalkanoates producing culture enrichment from crude glycerol**. *Bioresource Technology*, 311, 123516. 2020.

Wittenborn, E. C.; Jost, M.; Wei, Y.; Stubbe, J. C. L. Structure of the catalytic domain of the class I polyhydroxybutyrate synthase from *Cupriavidus necator*. **J. Biol. Chem.** 291 (48), 25264–25277. 2016.

Yu, S.; Ali, J.; Zhou, S.; Ren, G.; Xie, H.; Xu, J.; Yu, X.; Zhou, F.; Peng, S.; Ma, L.; Yuan, D.; Li, Z.; Chen, D.; Zheng, R.; Zhao, Z.; Chu, C.; You, A.; Yu, W.; Zhan, Q. **From Green Super Rice to green agriculture: reaping the promise of functional genomics research**. *Molecular Plant*, v. 15, n. 1, p. 9-26, 2022.

Zhou, W., Bergsma, S.; Colpa, I. D.; Euverink, G.; Krooneman, J. **Polyhydroxyalkanoates (PHAs) synthesis and degradation by microbes and applications towards a circular economy**, *Journal of Environmental Management*, Volume 341, 118033, ISSN 0301-4797. 2023.