



Universidade do Minho

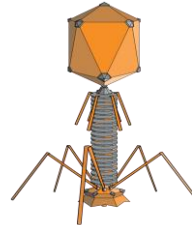
# PhageReLife: resistência e classificação do ciclo de vida de fagos

Projeto de Bioinformática

Mestrado de Bioinformática

Karyna Lysenko, pg51191

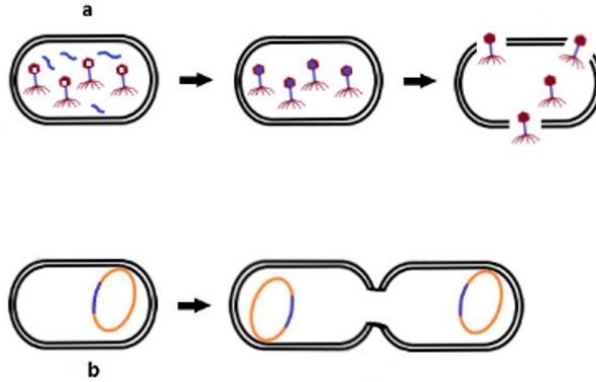
2022/2023



# Bacteriófagos (fagos) e Terapia

Introdução

a) Fagos **líticos/virulentos**

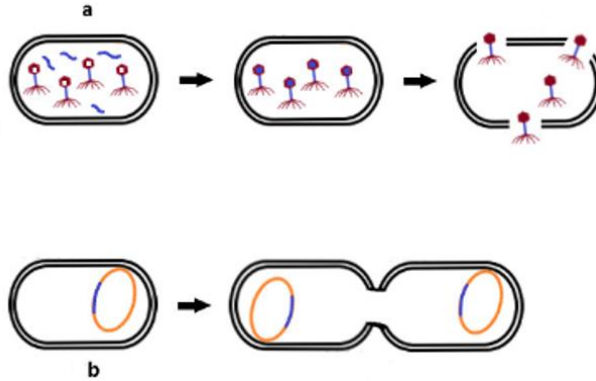


b) Fagos **lisogênicos/temperados**

# Bacteriófagos (fagos) e Terapia

Introdução

## a) Fagos líticos/virulentos

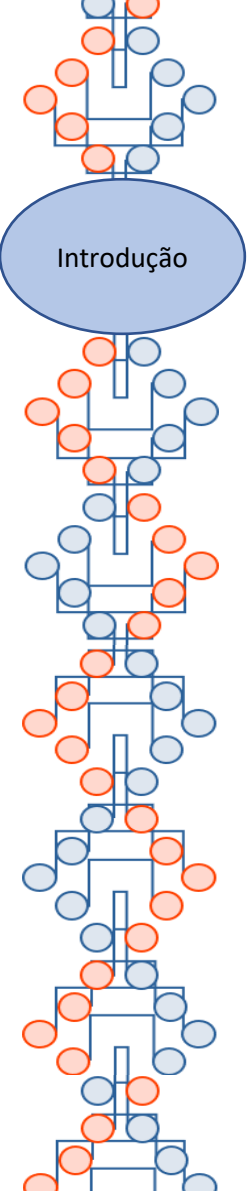


## b) Fagos lisogênicos/temperados

### Vantagens

- Fagos não são tóxicos nas células dos mamíferos;
- Hospedeiro-específico;
- Auto-doseados;
- Auto-eliminados.

# Ferramentas bioinformáticas para classificação de fagos terapêuticos



Introdução

IMG/VR v4 contém >15 milhões  
genomas e fragmentos de vírus

Fagos virulentos podem ter  
marcadores lisogênicos

Estudos *in  
silico*

5 ferramentas recentes:

1. Agrupamento dos dados/classe  
(por Homologia ou *Machine Learning*)
2. Previsão (*ML*)

# Ferramentas bioinformáticas para classificação de fagos terapêuticos

Introdução

	Nome, Ano	Funções
Baseado em homologia	<b>PhaTYP</b> , 2022	Classificação do ciclo de vida
	<b>PhageLeads</b> , 2022	Procura por marcadores lisogênicos, genes virulentos e de resistência antimicrobiana
	<b>BACPHLIP</b> , 2021	Classificação do ciclo de vida
Baseado em ML	<b>PhageAI</b> , 2020	Classificação do ciclo de vida
	<b>PHACTS</b> , 2012	Classificação do ciclo de vida

# Ferramentas bioinformáticas para classificação de fagos terapêuticos

Introdução

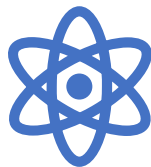
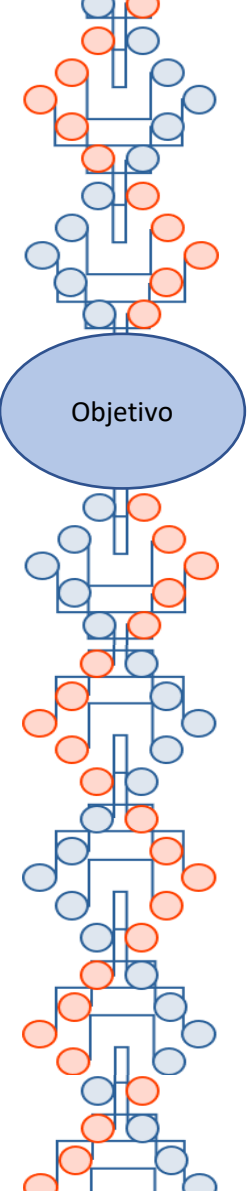
	Nome, Ano	Funções	Interface	Input
Baseado em homologia	<b>PhaTYP</b> , 2022	Classificação do ciclo de vida	Web; Terminal	DNA (contigs ou completo)
	<b>PhageLeads</b> , 2022	Procura por marcadores lisogênicos, genes virulentos e de resistência antimicrobiana	Web; Terminal	DNA (completo)
	<b>BACPHLIP</b> , 2021	Classificação do ciclo de vida	Terminal	DNA (completo)
Baseado em ML	<b>PhageAI</b> , 2020	Classificação do ciclo de vida	Web; Terminal	DNA (completo)
	<b>PHACTS</b> , 2012	Classificação do ciclo de vida	Terminal	AA (completo)

# Ferramentas bioinformáticas para classificação de fagos terapêuticos

## Introdução

	Nome, Ano	Funções	Interface	Input	Output	Base de dados	Precisão
Baseado em homologia	<b>PhaTYP</b> , 2022	Classificação do ciclo de vida	Web; Terminal	DNA (contigs ou completo)	Classificação, algumas features, Proteínas homologas (blastp)	3474 genomas completos 142434 contigs (inclusive teste)	94-98%
	<b>PhageLeads</b> , 2022	Procura por marcadores lisogênicos, genes virulentos e de resistência antimicrobiana	Web; Terminal	DNA (completo)	Nome dos marcadores ou genes	<b>7686</b> genomas completos, 188080 sequencias proteicas teste: 10-fold cross-validation	96.20%
	<b>BACPHLIP</b> , 2021	Classificação do ciclo de vida	Terminal	DNA (completo)	GFF3 file*	634 genomas completos teste: 183 virulentos e 240 temperados	98.30%
Baseado em ML	<b>PhageAI</b> , 2020	Classificação do ciclo de vida	Web; Terminal	DNA (completo)	Classificação, algumas features	<b>&gt;10000</b> genomas completos teste: 103 virulentos and 42 temperados	97.18%
	<b>PHACTS</b> , 2012	Classificação do ciclo de vida	Terminal	AA (completo)	Classificação, probabilidade e std	654 genomas completos, teste: 79 virulentos e 148 temperados (inclusive teste)	99%

\*GFF3 file format includes: prophage region and boundaries, quality score, genomic features (type of molecule)



# Análise crítica das ferramentas bioinformáticas de classificação do ciclo de vida de fagos e de genes associados à virulência



Caso de estudo: Fagos de *Acinetobacter*





## Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

Zedzedvirus × Metrivirus × Exceevirus × Lazarusvirus × Hadassahvirus × Lasallevirus × Acainonavirus × Obolenskivirus × Saclayvirus ×  
Friunavirus × Pettyvirus × Daemvirus × Vieuvirus × Presleyvirus × Lokivirus × Enter one or more taxonomic names

Filters complete\_genome +

STATUS

☐ Reference genomes

☐ Annotated genomes

☐ Annotated by NCBI RefSeq

☐ Annotated by GenBank submitter

☐ Exclude atypical genomes

SEARCH WITHIN RESULTS

Enter taxon name or modifier, assembly name or submitter

ASSEMBLY LEVEL

contig scaffold chromosome complete

YEAR RELEASED

1980 2023

Gêneros de fagos de Acinetobacter

169 genomas  
2 de maio 2023

Métodos

## Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

Zedzedvirus × Metrivirus × Exceevirus × Lazarusvirus × Hadassahvirus × Lasallevirus × Acainonavirus × Obolenskivir × Saclayvirus ×

Friunavirus × Pettyvirus × Daemvirus × Vieuvirus × Presleyvirus × Lokivirus × Enter one or more taxonomic names

Filters complete\_genome +

STATUS

☐ Reference genomes

☐ Annotated genomes

☐ Annotated by NCBI RefSeq

☐ Annotated by GenBank submitter

☐ Exclude atypical genomes

SEARCH WITHIN RESULTS

ASSEMBLY LEVEL

contig scaffold chromosome complete

YEAR RELEASED

1980 2023

Gêneros de fagos de Acinetobacter

169 genomas

2 de maio 2023

Ubuntu

Instalação das ferramentas

API

Anaconda

Biopython\*

PhageAI

PhageTYP

PHACTS

PhageLeads

BACPHLIP

Python scripts automatizados

Manual

Análise com pandas

\*Bio.SeqIO para tradução: Tabela 11

# Script exemplo: PhaTYP

Métodos

```
import os
import re

main_dir_path = '/home/karyna/Project/data'
rootpth = '/home/karyna/Project/PhaTYP/phatyp_results/'
midfolder = '/midfolder'
# Path to the PhaTYP executable
PhaTYP_path = '/home/karyna/PhaBOX/PhaTYP_single.py'

# Loop through all files in the folder
for root, dirs, files in os.walk(main_dir_path):
    for file in files:
        # Get the full path of the file
        file_path = os.path.join(root, file)
        if file.endswith('.fna'):
            GCA_name = re.search(r'^(GCA_[^_]+)', file)
            if GCA_name:
                GCA_id = GCA_name.group(1)
                print(file_path)
                os.system(f"python {PhaTYP_path} --contigs {file_path} --rootpth {rootpth+GCA_id}")
```

# Comparação técnica

	Script	Tempo execução/genoma	Output	Classificador	Interface Web
PhageAI	Limite 100/dia				
PhaTYP	Sem limite				
PHACTS	Sem limite				
PhageLeads					

Resistência a antibióticos

169 genomas

PhageLeads

PhageAI

123 virulentos

46 temperados

PhaTYP

161 virulentos

8 temperados

PHACTS

Pandas do python

# Comparação técnica

	Script	Tempo execução/genoma	Output	Classificador	Interface Web
PhageAI	Limite 100/dia				
PhaTYP	Sem limite				
PHACTS	Sem limite				
PhageLeads					

Resistência a antibióticos

## 169 genomas

PhageLeads

PhageAI

123 virulentos

46 temperados

PhaTYP

161 virulentos

8 temperados

PHACTS

# Comparação de Precisão da previsão do ciclo temperado

Genome Assembly accession_id	PhageLeads	Taxa de precisão			
		PhageAI	PhaTYP	PHACTS	
				Precisão	sd
GCA_028515195.1	<i>Predicted 2 genes</i>	72.6	0.9916	0.526	0.0384
GCA_027582695.1	<i>Predicted 1 genes</i>	84.2	0.9141	0.512	0.0367
GCA_028515145.1	<i>Predicted 2 genes</i>	79.6	0.5026	0.521	0.0331
GCA_014656645.1	<i>Predicted 2 genes</i>	87.7	0.9989	-	-
GCA_002605545.1	<i>Predicted 2 genes</i>	90.8	0.9998	0.531	0.0359
GCA_028514995.1	<i>Predicted 2 genes</i>	78.8	0.9997	0.515	0.0460
GCA_004989725.1	<i>Predicted 2 genes</i>	86.6	0.9842	0.507	0.0345
GCA_028515085.1	<i>Predicted 1 genes</i>	93.9	0.5752	0.500	0.0438
<b>Média %</b>	-	84.3	87.1	51.6	3.76

Precisão lítico < Precisão – (2\* sd)

Precisão lítico > Precisão + (2\* sd)

# Comparação de Precisão da previsão

Genome Assembly accession_id	PhageLeads	PhageAI	PhaTYP	Taxa de precisão		
				PHACTS		
				Precisão	sd	confiança
GCA_028515195.1	Predicted 2 genes	72.6	0.9916	0.526	0.0384	×
GCA_027582695.1	Predicted 1 genes	84.2	0.9141	0.512	0.0367	×
GCA_028515145.1	Predicted 2 genes	79.6	0.5026	0.521	0.0331	×
GCA_014656645.1	Predicted 2 genes	87.7	0.9989	-	-	-
GCA_002605545.1	Predicted 2 genes	90.8	0.9998	0.531	0.0359	×
GCA_028514995.1	Predicted 2 genes	78.8	0.9997	0.515	0.0460	×
GCA_004989725.1	Predicted 2 genes	86.6	0.9842	0.507	0.0345	×
GCA_028515085.1	Predicted 1 genes	93.9	0.5752	0.500	0.0438	×
<b>Média %</b>	-	84.3	87.1	51.6	3.76	×

Precisão lítico < Precisão – (2\* sd)

Precisão lítico > Precisão + (2\* sd)<sub>16</sub>

Resultados  
e  
Discussão



# Conclusão

Falta de ferramentas online robustas que permitam:

→ classificar vários genomas de forma automatizada:

- **PhageAI** é única

→ prever marcadores de virulência e/ou resistência para vários genomas de forma automatizada:

- **PhageLeads** ≈

## Perspetivas futuras

Desenvolver uma ferramenta online que possa incorporar ferramenta(s) do tipo PhageAI (já tem API) e PhageLeads com aplicação a vários genomas.







Universidade do Minho

# PhageReLife: resistência e classificação do ciclo de vida de fagos

Obrigada pela atenção!

