Pseudomonas aeruginosa prophages: a new and simplified bioinformatic analysis

Rodrigo Monteiro[1,2], Maria Vieira[1], Jelena Erdmann[3], Susanne Häussler[3], Joana Azeredo[1,4*]

*jazeredo@deb.uminho.pt

[1]CEB, Centre of Biological Engineering, Laboratory of Research in Biofilms Rosário Oliveira (LIBRO), University of Minho, Braga, Portugal, [2]Department of Medical Microbiology - Molecular Bacteriology, UMCG, University Medical Center Groningen, Groningen, Netherlands, [3]Twincore, Centre for Experimental and Clinical Infection Research, a joint venture of the Helmholtz Centre for Infection Research and the Hannover Medical School, Hannover, Germany, [4]LABBELS – Associated Laboratory, Braga, Guimarães, Portugal

University of Minho
School of Engineering

CENTRE OF BIOLOGICAL ENGINEERING

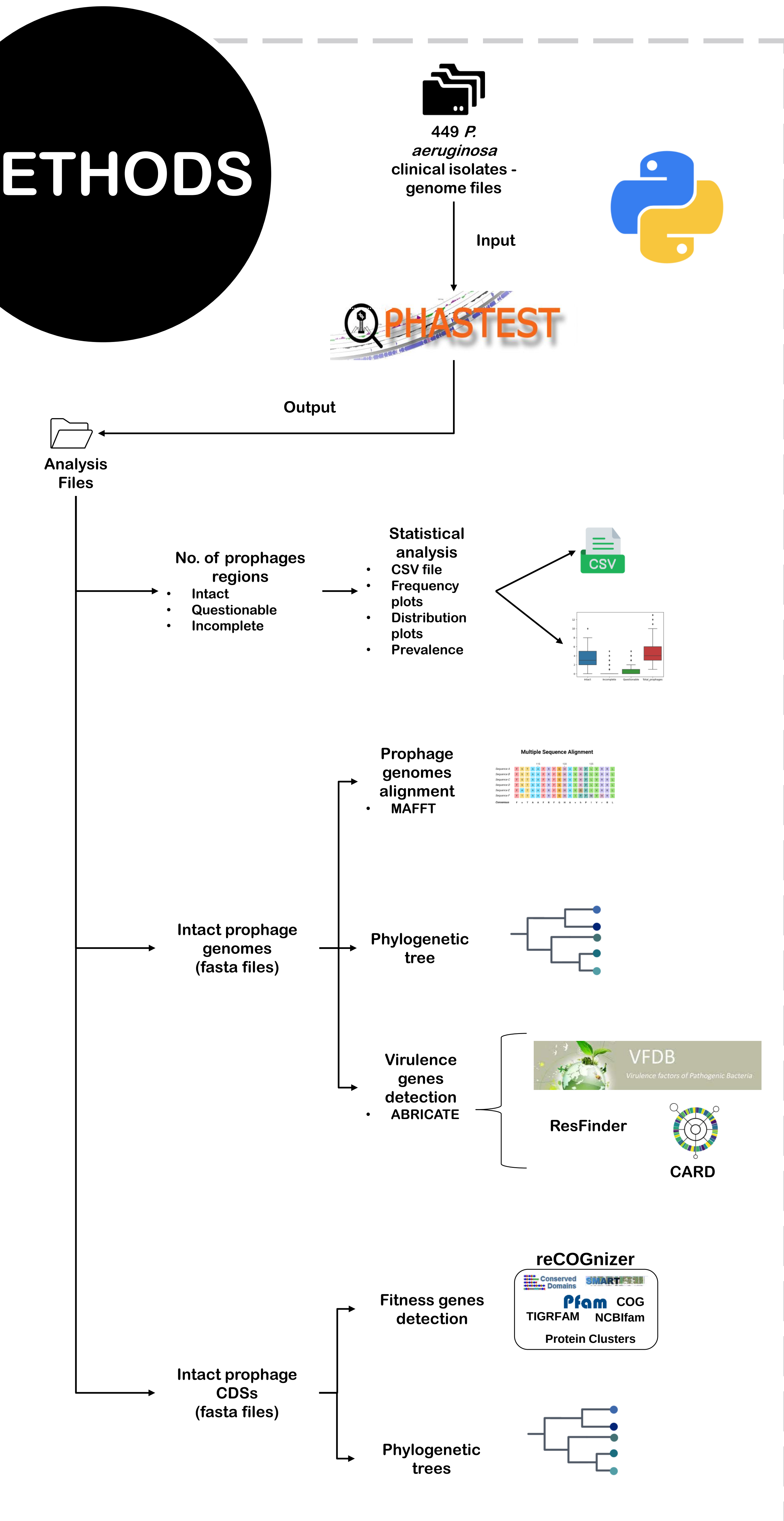
AZEREDO LAB

BACTERIOPHAGE | AZEREDO LAB | BIOTECHNOLOGY

INTRODUCTION

ESKAPE pathogens are a global threat due to the emergence of antibiotic resistance caused by the misuse of antibiotics. One major challenge researchers are facing is the monitoring of antibiotic resistance genes that are actively exchanged between different bacterial cells. Prophages, which are parasites inserted within the bacterial genome, play a role in transferring these virulence and fitness-related genes by transduction. Knowing the importance of this type of studies for monitoring problematic pathogens, we developed a fast and simple pipeline using Python for the analysis of prophages in large bacterial genome datasets combining different freely available bioinformatic tools which detect, quantify, and characterize prophages within the dataset.

METHODS



449 *P. aeruginosa* clinical isolates - genome files

Input

PHASTEST

Output

Analysis Files

No. of prophages regions
- Intact
- Questionable
- Incomplete

Statistical analysis
- CSV file
- Frequency plots
- Distribution plots
- Prevalence

Intact prophage genomes (fasta files)

Prophage genomes alignment
- MAFFT

Phylogenetic tree

Virulence genes detection
- ABRICATE

VFDB
Virulence factors of Pathogenic Bacteria

ResFinder

CARD

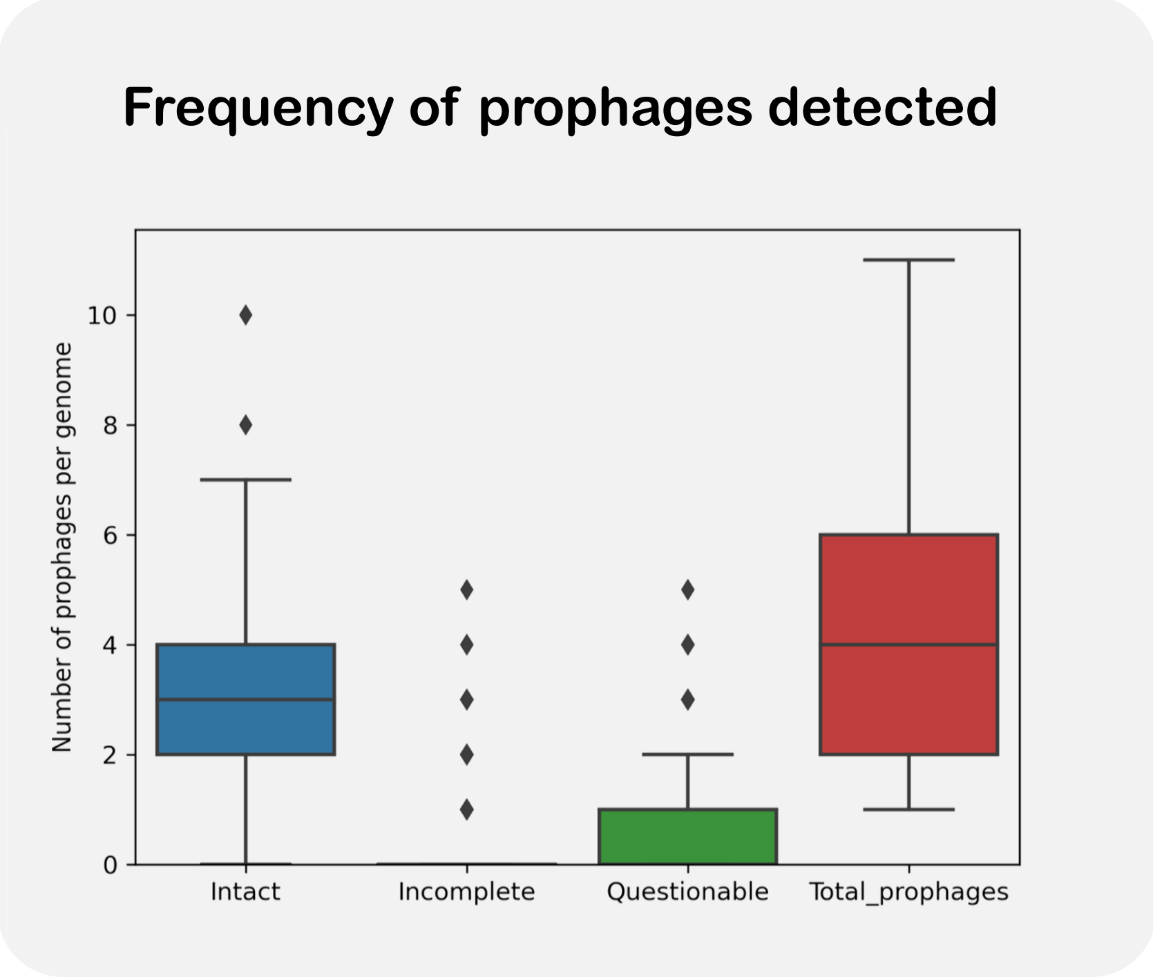
Intact prophage CDSs (fasta files)

Fitness genes detection

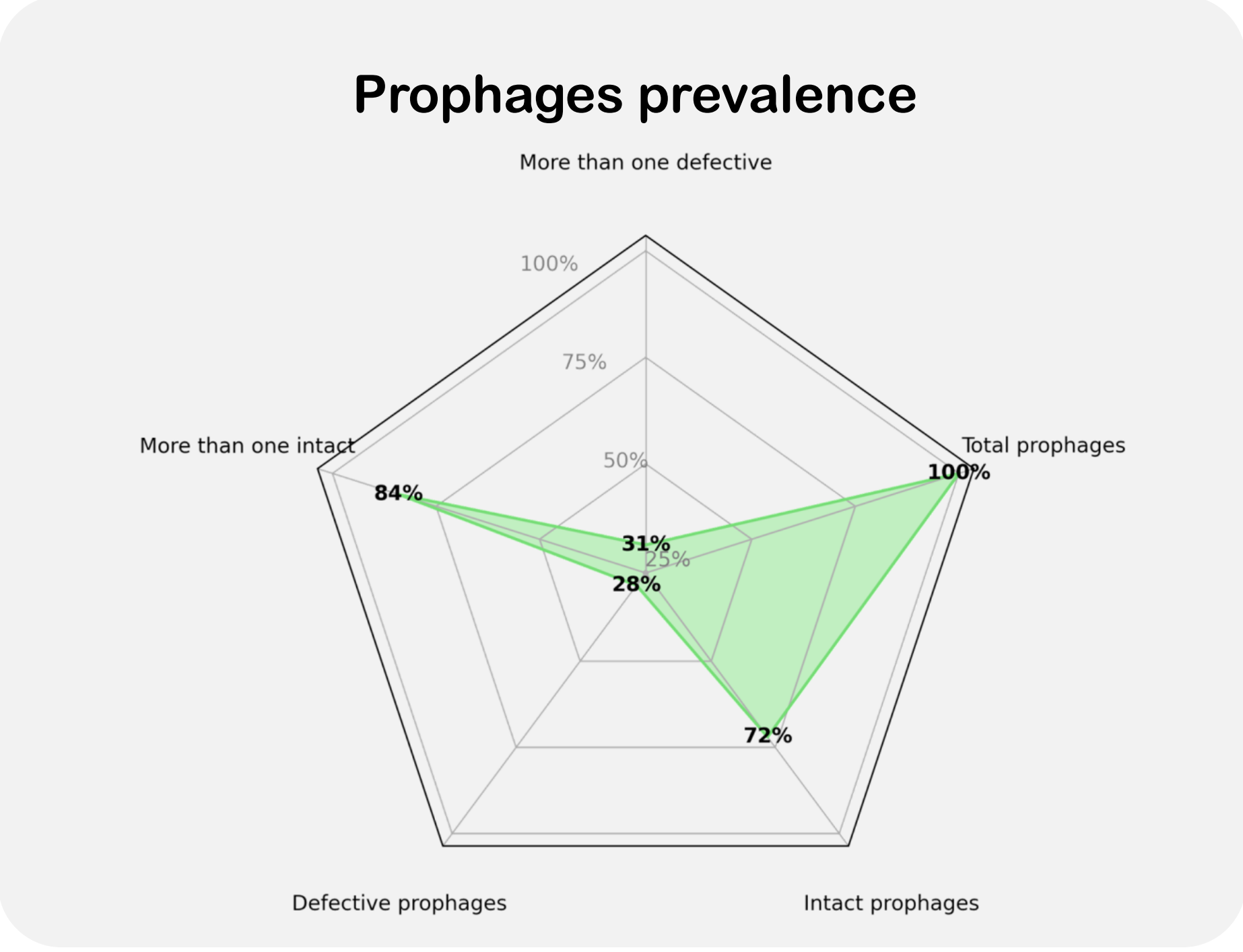
reCOGnizer
Conserved Domains
SMART
Pfam
COG
TIGRFAM
NCBIfam
Protein Clusters

Phylogenetic trees

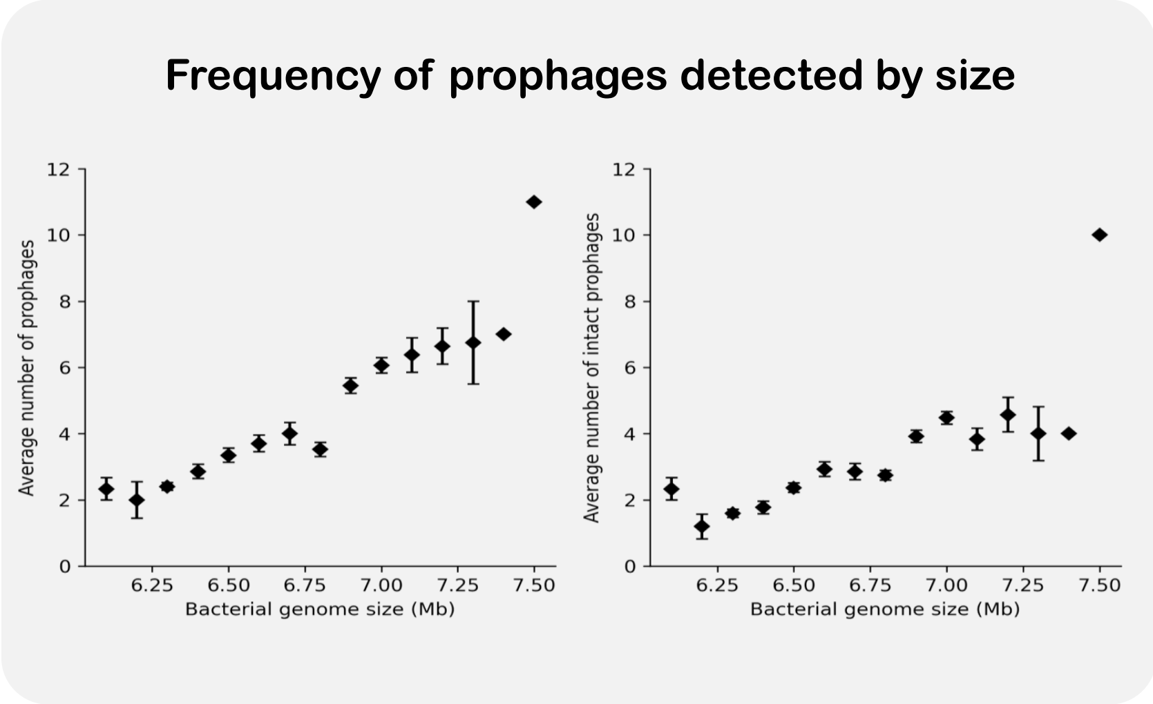
RESULTS



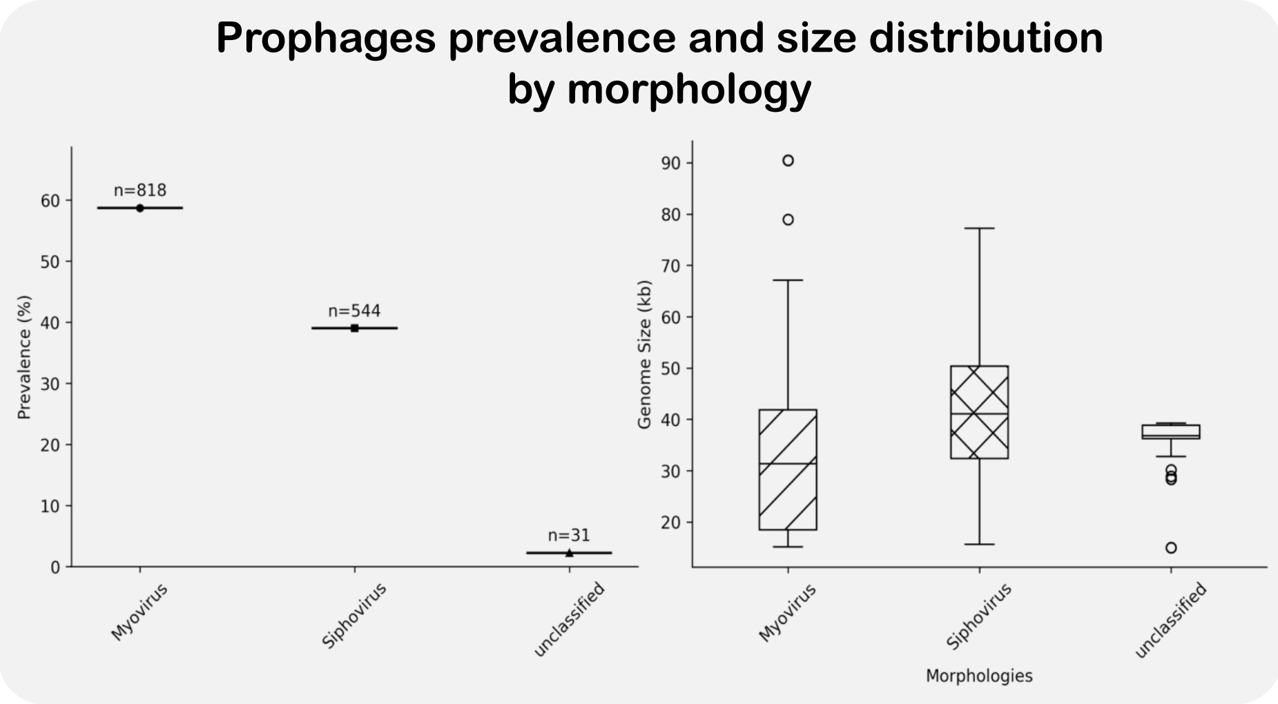
Frequency of prophages detected



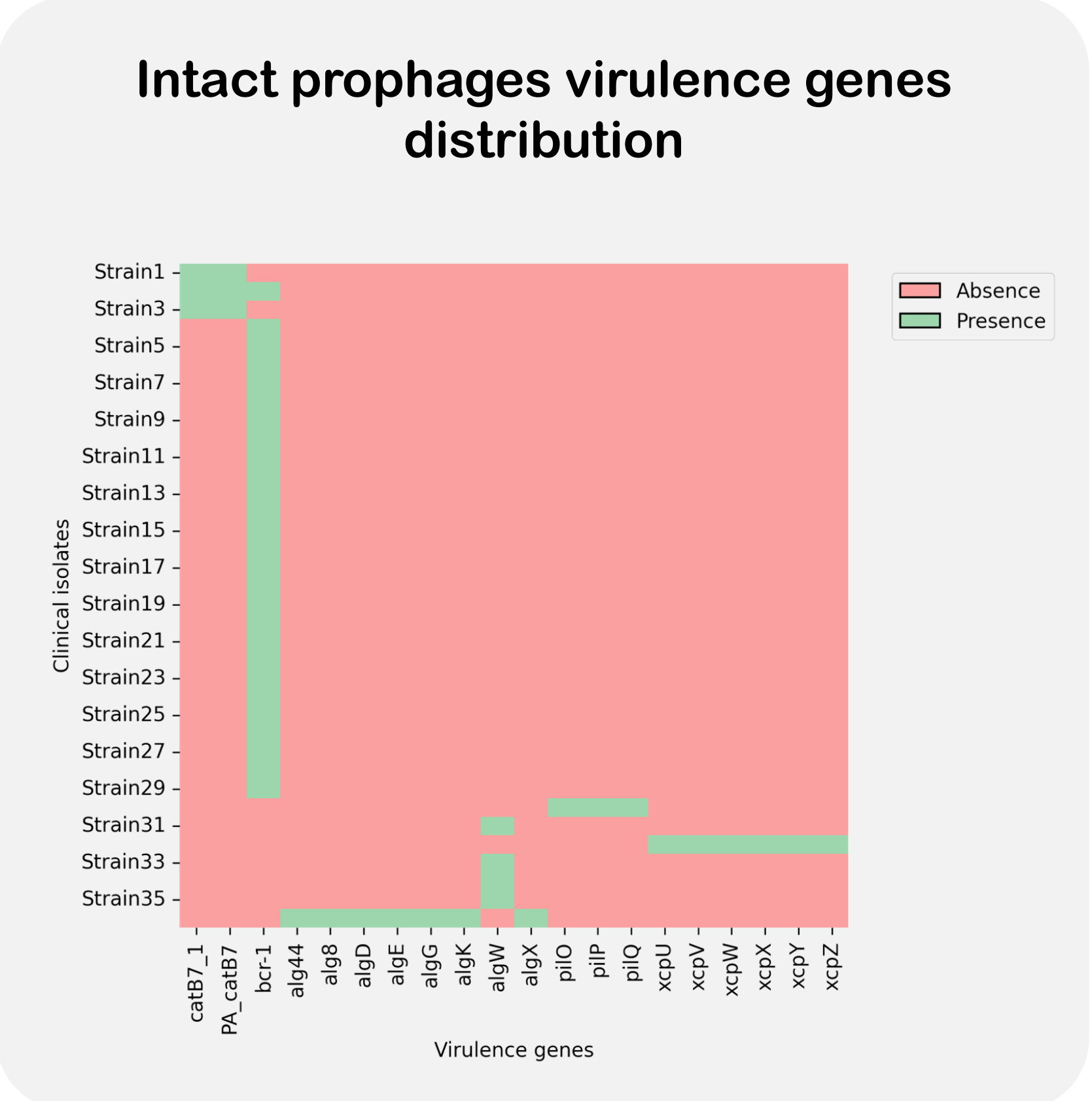
Prophages prevalence



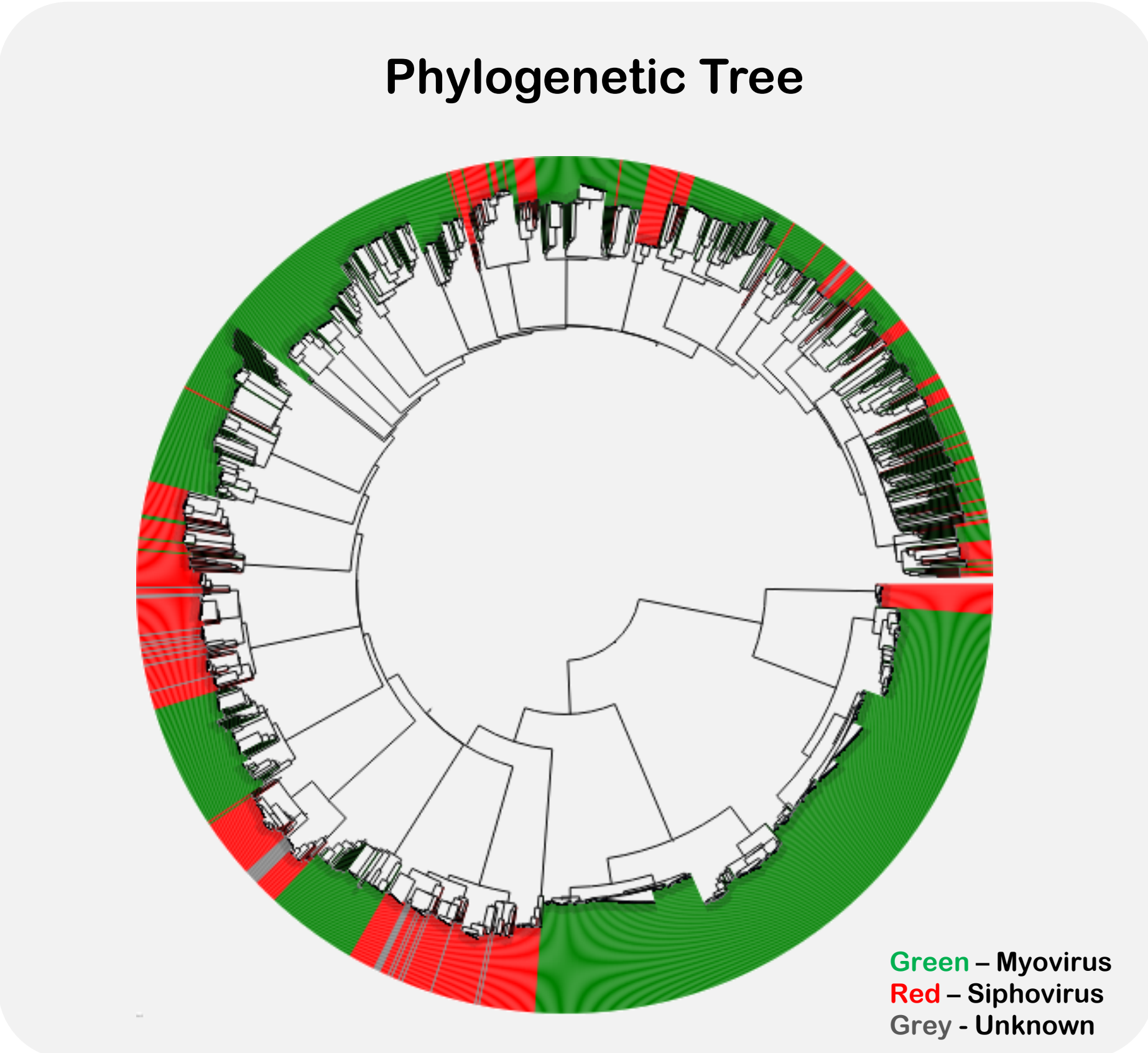
Frequency of prophages detected by size



Prophages prevalence and size distribution by morphology



Intact prophages virulence genes distribution



Phylogenetic Tree

Green – **Myovirus**
Red – **Siphovirus**
Grey - **Unknown**

- On **average** there were **three intact prophages** per strain;
- There was a **low number of incomplete prophages**;
- **99.5%** of analyzed strains had at least **one prophage, 84%** had at least **one intact**;
- The number of prophages increases with the host genome size;
- **Most prophages** detected were predicted to be **myo** (58%) and **siphoviruses** (40%);
- 35 strains contained intact **prophages encoding for virulence genes**;
- There is a **high prevalence** of **metabolic-associated genes** (n=1915) according to the COG classification.

CONCLUSIONS

Our pipeline enabled a fast detection and characterization of 1949 prophage regions from 449 genomes. We showed that prophages are widespread among *P. aeruginosa* clinical isolates, with a significant proportion of them potentially inducible. Prophages were highly prevalent within bacterial genomes since almost all strains analyzed contained at least one. Furthermore, a high prevalence of fitness-related genes, mostly to metabolic pathways, were found in prophages' genomes. Our study demonstrates the great diversity and prevalence of inducible prophages in clinical strains of *P. aeruginosa* carrying virulence genes, highlighting the importance of characterizing prophages in bacterial genomes to understand bacterial evolution and virulence. Our pipeline provides a fast and simple way to analyze large bacterial genomes datasets in one go, outputting a complete report on prophages' presence and their relevance for the strains analyzed.

Acknowledgements
This study was co-funded by FSE under "Programa Operacional Regional Norte", QREN, FEDER and the Project LCF/PR/HP21/52310017. Rodrigo Monteiro is supported by the FCT through the grant SFRH/BD/143639/2019.