Microbes and Infection 27 (2025) 105429

Contents lists available at ScienceDirect

Microbes and Infection

journal homepage: www.elsevier.com/locate/micinf

# Screening and *in silico* characterization of prophages in *Helicobacter pylori* clinical strains

Rute Ferreira [a,b], Graça Pinto [a], Eva Presa [a], Mónica Oleastro [c], Catarina Silva [d,e], Luís Vieira [d,e], Cláudia Sousa [a], Diana P. Pires [a,f], Ceu Figueiredo [b,g,h], Luís D.R. Melo [a,f,*]

[a] CEB - Centre of Biological Engineering, University of Minho, Braga, Portugal
[b] i3S - Instituto de Investigação e Inovação em Saúde, University of Porto, Porto, Portugal
[c] Department of Infectious Diseases, National Institute of Health Doctor Ricardo Jorge (INSA), Lisbon, Portugal
[d] Unit of Technology and Innovation, Department of Human Genetics, National Institute of Health Doctor Ricardo Jorge (INSA), Lisbon, Portugal
[e] Centre for Toxicogenomics and Human Health (ToxOmics), NOVA Medical School, Faculdade de Ciências Médicas, University Nova de Lisboa, Lisbon, Portugal
[f] LABBELS –Associate Laboratory, Braga, Guimarães, Portugal
[g] Ipatimup - Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal
[h] Department of Pathology, Faculty of Medicine, University of Porto, Porto, Portugal

ARTICLE INFO

*Keywords:*
*Helicobacter pylori*
Prophages
PCR
Virulence factors
Antimicrobial resistance genes

ABSTRACT

The increase of antibiotic resistance calls for alternatives to control *Helicobacter pylori*, a Gram-negative bacterium associated with various gastric diseases. Bacteriophages (phages) can be highly effective in the treatment of pathogenic bacteria. Here, we developed a method to identify prophages in *H. pylori* genomes aiming at their future use in therapy. A polymerase chain reaction (PCR)-based technique tested five primer pairs on 74 clinical *H. pylori* strains. After the PCR screening, 14 strains most likely to carry prophages were fully sequenced. After that, a more holistic approach was taken by studying the complete genome of the strains. This study allowed us to identify 12 intact prophage sequences, which were then characterized concerning their morphology, virulence, and antibiotic-resistance genes. To understand the variability of prophages, a phylogenetic analysis using the sequences of all *H. pylori* phages reported to date was performed. Overall, we increased the efficiency of identifying complete prophages to 54.1 %. Genes with homology to potential virulence factors were identified in some new prophages. Phylogenetic analysis revealed a close relationship among *H. pylori*-phages, although there are phages with different geographical origins. This study provides a deeper understanding of *H. pylori*-phages, providing valuable insights into their potential use in therapy.

## 1. Introduction

*Helicobacter pylori* chronically infects nearly half of the world population and is a major risk factor for several diseases, including gastroduodenal ulcers and gastric cancer. Current standards for *H. pylori* treatment rely on the use of various antibiotics, with alarming rates of resistance worldwide, thus emphasizing the need for alternative therapeutic approaches [1].

In recent years, bacteriophages (phages) have gained recognition as an effective therapy to combat bacterial infections [2]. Phages are viruses of bacteria that are classified according to their lytic (virulent) or lysogenic (temperate) cycle. Virulent phages replicate inside bacteria, causing cell lysis to release progeny, while temperate phages integrate into the host bacterial genome as prophages, without causing bacterial lysis [2]. Under specific stimuli, prophages can be induced, leading to activation of their lytic cycle. By mediating horizontal gene transfer, phages can influence bacterial phenotypes, providing new traits, such as antibiotic resistance and metabolic factors [3]. Despite temperate phages being usually avoided for therapy due their ability of mediating gene transfer between bacteria by specialized transduction, current advances in synthetic biology create new prospects to explore their therapeutic use against bacterial infections [3].

The analysis of *H. pylori* genomes is compatible with the presence of prophage genes in about 20 % of strains [4,5], but only few complete phages have been induced [6,7]. The lack of deep genomic analyses and the limited characterization of *H. pylori* (pro)phages remain major challenges in the prospective use of phage therapy against *H. pylori*. Therefore, the aim of this study was to determine the prevalence and

* Corresponding author. CEB - Centro de Engenharia Biológica, Universidade do Minho, Campus de Gualtar, 4710-057, Braga, Portugal.
*E-mail address:* lmelo@ceb.uminho.pt (L.D.R. Melo).

https://doi.org/10.1016/j.micinf.2024.105429
Received 16 May 2024; Received in revised form 2 October 2024; Accepted 2 October 2024
Available online 3 October 2024

characterize new prophages in a panel of Portuguese *H. pylori* clinical strains.

## 2. Materials and methods

### 2.1. H. pylori strains

Seventy-four *H. pylori* strains isolated from human gastric biopsies belonging to the collection of bacterial strains from INSA – Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal, were used in this study. Strain metadata can be seen in Table S1.

### 2.2. PCR screening of prophages and sequencing of H. pylori strains

Screening of prophages was performed by polymerase chain reaction (PCR) in a collection of 74 *H. pylori* clinical strains (Table S1), using four different primer sets (Table S2). These targeted two regions of the integrase gene, the holin gene, and the insertion site (IS) region between the S-adenosylmethionine synthetase (S-adeno) and the UDP-3-*O*-[3-hydroxymyristoyl] glucosamine N-acyltransferase (UDP) genes [8]. Strains were confirmed as *H. pylori* by amplification of the housekeeping *glmM* gene.

To prepare the DNA template for PCR, 10 μL of cryopreserved cells were diluted in 50 μL of nuclease-free water (New England Biolabs, USA). Afterwards, the cells were treated at 100 °C for 5 min followed by an incubation at −20 °C for 5 min. Finally, the cells were centrifuged (5000 *g*, 5 min, 4 °C) and the supernatant was used as a DNA template in the PCR reaction.

PCR mixtures contained 10 μL of DreamTaq DNA Polymerase MasterMix (Thermo Fisher Scientific, USA), primers (0.5 μM each), 2 μL DNA template and 6 μL of nuclease-free water to complete (New England Biolabs). PCR was performed with a pre-denaturation of 3 min at 95 °C, followed by 30 cycles of 30 s at 95 °C, 30 s at 55–59 °C, and 1 min at 72 °C, and a final extension step of 7 min at 72 °C. Amplicons were visualized after electrophoresis in 1.5 % agarose gels.

### 2.3. Whole-genome sequencing of H. pylori strains

Illumina and MinION Nanopore sequencing platforms were the sequencing strategies used for all the strains. Whole genome sequencing was performed as previously described [9]. DNA was extracted with a Bioline ISOLATE II Genomic DNA kit (Meridian Bioscience Inc., USA). Quantification and quality assessment of the purified DNA was performed using the DNA HS Assay Kit (Thermo Fisher Scientific) in the Qubit Fluorometer and agarose gel electrophoresis (0.8 %), respectively. For long-reads Nanopore sequencing, DNA purity was assessed by spectrophotometry using NanoDrop One (Thermo Fisher Scientific). High-quality DNA samples were then used to prepare dual-indexed Nextera XT Illumina libraries that were subsequently subjected to cluster generation and paired-end sequencing (2 × 150 bp) on a NextSeq 550 equipment (Illumina, USA) available at INSA. Genomic libraries for long-read Nanopore sequencing were prepared using the Rapid Barcoding Sequencing Kit (SQK-RBK004; Oxford Nanopore Technologies-ONT) according to the manufacturer's instructions. The sequencing run was performed on an Mk1C device (ONT) using one FLO-MIN106D flow cell for a total run time of 50 h. The MinKNOW software (v.21.11.6, ONT) was used to program and configure run parameters and to acquire the raw signal data during the sequencing run. A fast basecalling model (Guppy v5.1.12), was used to obtain basecalled reads in FASTQ files. Several run statistics were obtained using NanoPlot (v1.30.1) [10]. Quality control and bacterial de novo assembly were performed using the INNUca v4.01 pipeline (https://github.com/BUMMI/INNUca), which consists of several integrated modules for reads QA/QC, de novo assembly, and post-assembly optimization steps. The sequencing data were uploaded to the Galaxy web platform, and the public server at usegalaxy.org was used to analyse the data [11]. Genomes were further assembled with Unicycler v. 0.4.8 using default settings [12].

### 2.4. Identification of prophage sequences in H. pylori genomes

PHASTER [13], Prophage Hunter [14], and Phigaro [15] with default settings were used to predict prophage regions. Prophage Hunter data were excluded due to the high number of prophages proposed, differences in prophage locations and variances in integrase, holin, and IS genes sites. Prophages from Phigaro and PHASTER were compared and considered the same if they shared locations. For similarity searches and structured prediction, proteins were annotated using phage-limited BLASTp (limited to Caudoviricetes (taxid:2731619)) and HHmer (with *E-value* cut-off of $1.0 \times 10^{-5}$).

### 2.5. Prophage classification

The complete prophages identified were classified *in silico* into their respective phage morphology based on the prophage structural head-neck-tail proteins using VIRFAM [16] webserver.

### 2.6. Identification of potential virulence factors encoded by prophages

A subset of 12 complete prophages were analysed for encoding putative antimicrobial resistance genes and virulence factors. For this, genome sequences of each prophage were analysed using the Resistance Gene Identifier (RGI) of the Comprehensive Antibiotic Resistance Database (CARD), selecting criteria of perfect, strict, and loose hits. Moreover, the DNA prophage sequences were scanned using ResFinder 4.1 database [17] and VirulenceFinder 2.0 [18–20] to identify virulence genes.

### 2.7. Phylogenetic analyses

The 12 prophages identified in this study were aligned with 40 *H. pylori* phages deposited in GenBank. The entire analysis was carried out by the VICTOR web service (https://victor.dsmz.de), a method for the genome-based phylogeny and classification of prokaryotic viruses [21]. All pairwise comparisons of the amino acid sequences were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method [22] under settings recommended for prokaryotic viruses [21]. The resulting intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR post-processing [23] for formula D6. Branch support was inferred from 100 pseudo-bootstrap replicates each. Tree was rooted at the midpoint [24] and visualized with ggtree [25]. Taxon boundaries at the species, genus and family level were estimated with the OPTSIL program [26], the recommended clustering thresholds [21] and an F value (fraction of links required for cluster fusion) of 0.5 [27].

The software excluded prophage sequence Pt-4497-U due to its similarity with Pt-4472-G.

The OrthoVenn3 [28] software was used to compare and find orthologous gene clusters and proteins among the phage genomes from different genus. *E-value* of 1.0 x $10^{-5}$ and the default inflation value of 1.5 were used.

## 3. Results and discussion

### 3.1. PCR-based screening of prophage genes in H. pylori strains

A PCR-based screening of 74 *H. pylori* clinical strains was conducted to identify prophages in their genomes. Screening was performed targeting four different regions, including the integrase and holin genes, which encode proteins necessary for the integration of the phage into the bacterial genome, and for controlling bacterial cell lysis, respectively. The presence of these genes has been previously used as a basis to infer

the presence of complete prophages in *H. pylori* [4]. Since prophage sequences are very heterogeneous, in addition to the previously described primers for integrase (integrase 1), we designed a novel primer set targeting a different region of the integrase gene (integrase 2). We further designed a new primer set directed to the region between the S-adeno and UDP genes, which is known as a typical integration zone for prophages classified as hpNEurope and hpAfrica1 [8].

In our study, the integrase gene was detected in 18 (24.3 %) *H. pylori* strains with integrase 1 primers (Table S3), which was comparable to the detection rate of 21.4 % reported in a study of 341 worldwide *H. pylori* strains characterized with the same primers [5]. Adding the holin gene, which was present in 4 (5.4 %) of *H. pylori* genomes, did not improve the rate of phage detection, since all holin-positive strains were also positive for integrase. These results are also in keeping with those of other studies reporting positivity rates for both integrase and holin genes between 4.2 % and 4.7 % [4,29]. In contrast, the use of the newly developed integrase 2 primers, led to integrase detection in 40 (54.1 %) *H. pylori* strains. Although integrase 1 and 2 primers target different regions of the gene, only 14 (18.9 %) strains were positive for both sets, a result likely explained by the genetic variability of integrase. Eleven (14.9 %) strains did not show amplification of the IS region, suggesting that they have a higher probability of having an integrated prophage genome, since the S-adeno and the UDP genes are usually contiguous in the *H. pylori* genome [8].

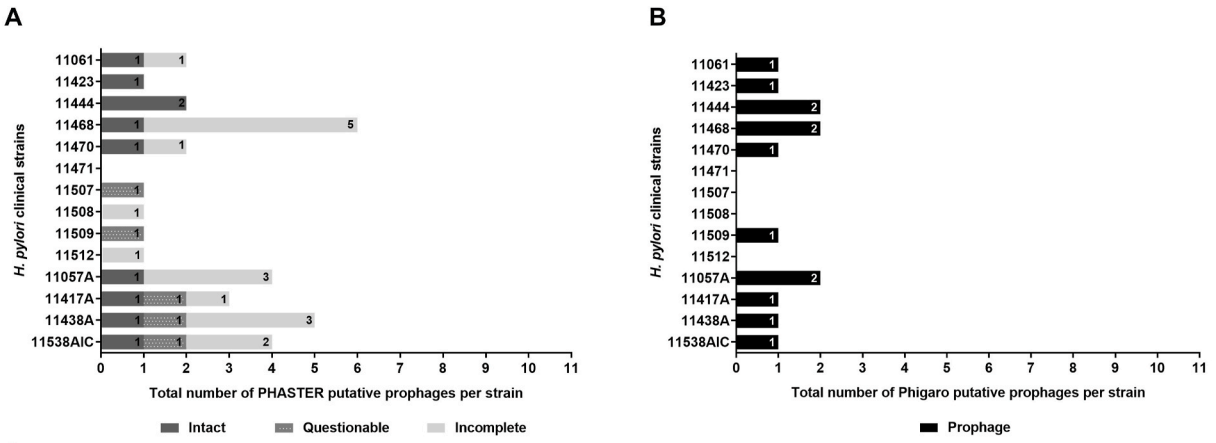
The choice for the strains to be sequenced was based on the results of the different primer sequences tested. Fourteen *H. pylori* strains were selected for sequencing, including the four strains positive for holin, seven strains positive for both integrase 1 and 2, two strains with opposing results for integrase 1 and 2, and the propagation strain of the previously isolated phage HPy1R [6]. The selection included strains with and without amplification of the IS region (Table S3).

### 3.2. Identification and characterization of prophages in H. pylori strains

Overall, the 14 *H. pylori* genomes had a size between 1.52 and 1.73 Mbp and a GC content between 38.40 % and 39.10 % (Table S4).

Prophage identification in these genomes revealed 33 prophages (10 intact, five questionable, and 18 incomplete) using PHASTER, and 13 prophages using Phigaro (Fig. 1A and B). The identification of prophages as intact, questionable, or incomplete by PHASTER is based on the number of coding DNA sequences (CDS) of a region attributable to prophages and on the presence of phage-derived genes [13]. Here, "complete" phages represent prophages identified by PHASTER as intact or questionable, or by Phigaro with an acceptable genome length (> 20 kb) and with phage-derived genes, such as the ones coding for tail fiber, integrase, or portal protein.

Twelve new prophages were identified, considering those that were simultaneously detected by the two tools (Fig. 1C), showing that at least 10 out of the 14 (71.4 %) sequenced *H. pylori* strains contain at least one intact prophage. None of the tools detected prophages in strain 11471




C

| Prophage | Identification tool | Location* (bp) | | Size (Kb) | GC content (%) | Morphology** | Virulence and fitness proteins*** | Accession number |
|---|---|---|---|---|---|---|---|---|
| Pt11061_1 | PHASTER | 771947 | 817304 | 45.36 | 38.80 | Podovirus | Putative ABC binding cassette transporter (ORF 24) | PQ159895 |
| Pt11423_1 | PHASTER | 396069 | 421250 | 25.18 | 38.10 | ND | Putative ABC binding cassette transporter (ORF 20) | PQ159897 |
| Pt11444_1 | PHASTER | 22602 | 70801 | 32.45 | 37.80 | ND | Putative ABC binding cassette transporter (ORF 32 and 36) | PQ159899 |
| Pt11444_2 | PHASTER | 75 | 27998 | 28.64 | 37.10 | Podovirus | Putative ABC binding cassette transporter (ORF 16) | PQ159900 |
| Pt11468_1 | PHASTER | 451924 | 497959 | 46.04 | 36.00 | Podovirus | Lipopolysaccharide ABC transporter (ORF 108-110) | PQ159901 |
| Pt11470_1 | PHASTER | 290868 | 327628 | 36.76 | 36.60 | Podovirus | TA type II system (HicAB system) (ORF 26 and 27) | PQ159902 |
| Pt11470_2 | Phigaro | 665024 | 686618 | 21.60 | 36.60 | Podovirus | TA type II system (HicAB system) (ORF 21 and 22) | PQ159903 |
| Pt11509_1 | PHASTER | 117749 | 140566 | 22.82 | 35.50 | ND | - | PQ159904 |
| Pt11057A_1 | PHASTER | 1426135 | 1469910 | 43.78 | 38.10 | Podovirus | Barnase/EndoU/colicin E5-D/RelE system (ORF 48)<br>TA type II system (HicAB system) (ORF 85) | PQ159894 |
| Pt11417A_1 | PHASTER | 832331 | 876110 | 43.78 | 38.10 | Podovirus | TA type II system (HicAB system) (ORF 30 and 76)<br>Barnase/EndoU/colicin E5-D/RelE system (ORF 44) | PQ159896 |
| Pt11438A_1 | PHASTER | 1182474 | 1221149 | 38.68 | 38.20 | ND | - | PQ159898 |
| Pt11538AIC_1 | PHASTER | 918866 | 949628 | 30.76 | 37.40 | ND | Barnase/EndoU/colicin E5-D/RelE system (ORF 15) | PQ159905 |

**Fig. 1. Prophage distribution in *H. pylori* genomes. A)** PHASTER and **B)** Phigaro analysis results. PHASTER separates the identified prophages into intact, questionable, and incomplete according to criteria that consider the number of coding DNA sequences (CDSs) of a region attributable to prophage CDSs, and the presence of phage-related genes. Phigaro uses a combination of homology-based and machine learning-based methods to identify prophage regions in bacterial genomes. **C)** Complete prophages identified in this study. *Location according prophage identification tools; **Morphology according to the VIRFAM tool; ***Results provided by BLASTp.

(positive with integrase 1 primers), and Phigaro did not detect prophage sequences in strains 11507 (negative with integrase and holin primers), 11508 and 11512 (positive with integrase 1 and 2 primers).

The comparison of prophage distribution with the PCR-based screening indicates that the selection of strains for sequencing based solely on the integrase 1 and holin primers, may exclude strains with prophages (e.g., strain 11423 with a complete prophage was PCR-negative for integrase 1 and holin and positive for integrase 2). Furthermore, only a positive result for integrase 1 was inaccurate, since in strain 11471 no prophages sequences were detected. In the case of strains 11508 and 11512 (PCR-positive for integrase 1 and 2), incomplete phages were detected (Fig. 1A), however not considered for the present analysis. The screening approach was not flawless, as strain 11507 would have not been selected for sequencing (negative for integrase 1 and 2 and holin) but was chosen due to its key role as propagation strain of *H. pylori* phage HPy1R [6]. Still, the use of an additional integrase region can be a useful approach, resulting in increased probability of selecting strains with prophages to 54.1 % (in contrast to the 20 % previously reported [4]).

An *in silico* primer hybridization analysis revealed that all prophage-positive strains carry the integrase gene, and three strains (11468, 11057, and 11417) also possess holin. These observations were confirmed, with one exception, in prophage sequence annotations. Concerning the insertion sites, these were not conserved among the *H. pylori* prophages identified (Table S5). However, gene annotation showed the presence of bacterial genes in the beginning and end of the phage sequences, including S-adeno and UDP in four prophages, indicating that prophage limits generated by PHASTER may not be correct. The IS region PCR did not fully overlap gene annotation results of the prophage insertion site, which may be due to population origins different than those of hpNEurope and hpAfrica1. Noteworthy, 33.3 % of the identified prophages are inserted in this typical integration zone [8].

Regarding phage characterization, the results obtained with VIRFAM were not surprising (Fig. 1C), since most reported *H. pylori* prophages have a podovirus-like morphology [6,7]. A similarity search with BLASTp identified virulence and fitness proteins, namely putative ABC binding cassette transporter, TA type II system (HicAB system), and Barnase/EndoU/colicin E5-D/RelE system genes (Fig. 1C; Table S6). ABC transporters are associated with the need of acquiring nutrients in order to adapt to environmental conditions, although in bacteria, this gene also acts on the uptake of molecules as opposed to efflux [30]. The HicAB system is found in many bacteria and archaea and is involved in virulence and adaptation to extracellular stresses [31]. Barnase/EndoU/colicin E5-D/RelE is a system used by bacteria to compete with neighbouring cells [31].

Antibiotic resistance genes were additionally identified including mutant *lpx* in three prophages and *fabI* in five prophages (Table S7), however with a "Loose" cut-off by the tool. The “Loose” algorithm works outside of the detection model cut-offs and provides more distant

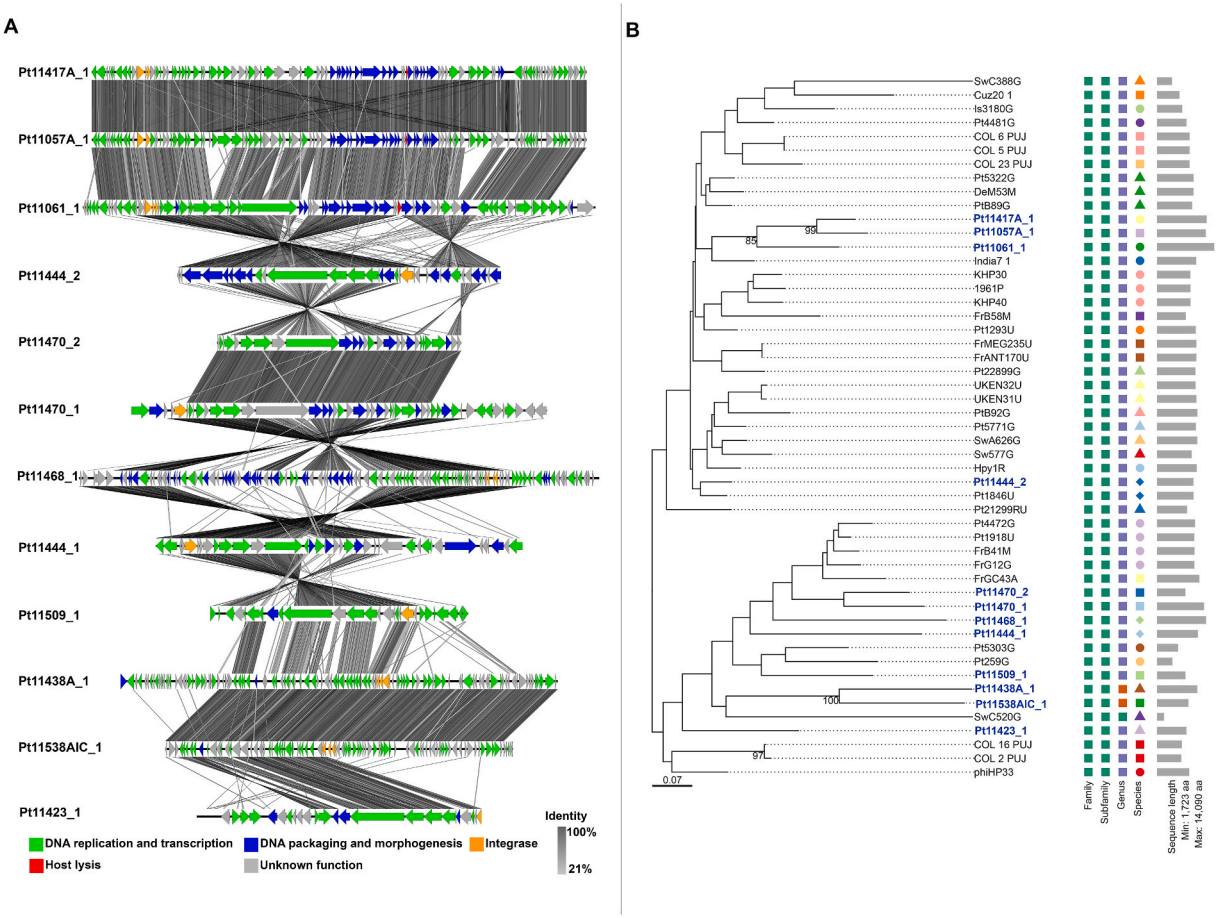



**Fig. 2. A)** Genomic comparison between the 12 *H. pylori* prophages identified using the EasyFig software. Prophage genomes were presented by linear visualization with coding regions shown as arrows. The direction of the arrow indicates the transcription direction of each ORF. Selected ORFs were coloured in relation to their functions. The percentage of sequence similarity according to BLASTn is indicated by the intensity of the grey colour. **B)** Phylogenetic tree of the 12 prophages identified in this study and for all *H. pylori* phages deposited in GenBank. The Genome BLAST Distance Phylogeny (GBDP) tree was inferred using formula D6 and yielding average support of 20 %. The numbers below branches were GBDP pseudo-bootstrap support values from 100 replications. The branch lengths of the resulting VICTOR trees were scaled in terms of the respective distance formula used. The OPTSIL clustering yielded forty species clusters. At the genus level, two clusters resulted. The number of clusters determined at the family and subfamily level was one.

homologs of antimicrobial resistance (AMR) genes. Still, it also catalogues homologous sequences and spurious partial hits that may not have a role in AMR [32].

Prophage genome organization and transcription directions were dissimilar, as assessed by Easyfig (Fig. 2A). High level of similarity in integrase and holin genes was observed. Portal protein, terminase, and tail fiber genes involved in DNA packaging and morphogenesis were identified in almost all prophages (> 40 %; Table S6).

For phylogenetic analysis, a tree with all annotated phages to date and the 12 new prophages identified herein was constructed. Additionally, the prophage sequences of strains India 7 and Cuz20, which were previously shown to contain prophages [8], were determined using PHASTER, and were also included in the tree. The tree evidenced that the large majority of *H. pylori* phages were phylogenetically related, since they all belong to the same family and genus (Fig. 2B; Table S8). We observed two exceptions: i) phage Sw-C520-G with a shorter sequence length, potentially causing bias in bioinformatics analysis; ii) phages Pt11438A_1 and Pt11538AIC_1 sharing 78 % orthologues between them, and 55 % and 44 % orthologue proteins, respectively with a phage from another genus (Pt11417_1, Fig. S1). Original research work about *H. pylori* phages reported that they belong to different families, including the phage KHP30, which was referenced as a new type of spherical phage [33]. However, it should be noted that recently, the International Committee on Taxonomy of Viruses updated the virus taxonomy [34] and the majority of phages are included in the *Schmidvirus* genus, confirming our observations. Interestingly, and although there were phages isolated from various geographical origins and belonging to different population structures [8,29], our analysis suggests a certain degree of conservation in *H. pylori* prophages.

In conclusion, the PCR screening method with the inclusion of a new primer set for the integrase gene can be a useful approach in the detection of prophages on *H. pylori* genomes. The identification of genes with virulence potential expands the data of prophages in *H. pylori* clinical strains and their possible role in disease development. Phylogenetic analysis of all *H. pylori* phages shows conservation in genomes, even though they have different geographic origins. Overall, our findings reveal a novel area for future investigations, which include the application of prophage induction strategies to the most promising strains, and consequently use phages in therapy-related applications.

## CRediT authorship contribution statement

**Rute Ferreira:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Graça Pinto:** Writing – review & editing, Software, Investigation, Formal analysis. **Eva Presa:** Writing – review & editing, Investigation. **Mónica Oleastro:** Writing – review & editing, Resources, Investigation. **Catarina Silva:** Writing – review & editing, Software, Investigation. **Luís Vieira:** Writing – review & editing, Software, Investigation, Data curation. **Cláudia Sousa:** Writing – review & editing, Investigation. **Diana P. Pires:** Writing – review & editing. **Ceu Figueiredo:** Writing – review & editing, Visualization, Validation. **Luís D.R. Melo:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

None.

## Acknowledgements

This work was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UIDB/04469/2020 unit, and by LABBELS – Associate Laboratory in Biotechnology, Bioengineering, and Microelectromechanical Systems, LA/P/0029/2020 and Project Helicophage PTDC/SAU-PUB/29182/2017 [POCI-01-0145-FEDER-029182]. Rute Ferreira acknowledges the FCT grant (https://doi.org/10.54499/SFRH/BD/146496/2019). Luís D. R. Melo acknowledges funding from the FCT through the Scientific Employment Stimulus Program (https://doi.org/10.54499/2021.00221.CEECIND/CP1664/CT0002).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.micinf.2024.105429.

## References

[1] Sousa C, Ferreira R, Azevedo NF, Oleastro M, Azeredo J, Figueiredo C, et al. *Helicobacter pylori* infection: from standard to alternative treatment strategies. Crit Rev Microbiol 2021;48:376–96. https://doi.org/10.1080/1040841X.2021.1975643.

[2] Melo LDR, Oliveira H, Pires DP, Dabrowska K, Azeredo J. Phage therapy efficacy: a review of the last 10 years of preclinical studies. Crit Rev Microbiol 2020;46:78–99. https://doi.org/10.1080/1040841X.2020.1729695.

[3] Monteiro R, Pires DP, Costa AR, Azeredo J. Phage therapy: going temperate? Trends Microbiol 2019;27:368–78. https://doi.org/10.1016/j.tim.2018.10.008.

[4] Vale FF, Vadivelu J, Oleastro M, Breurec S, Engstrand L, Perets TT, et al. Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. Sci Rep 2015;5:14333. https://doi.org/10.1038/srep14333.

[5] Lehours P, Vale FF, Bjursell MK, Melefors O, Advani R, Glavas S, et al. Genome sequencing reveals a phage in *Helicobacter pylori*. mBio 2011;2:e00239. https://doi.org/10.1128/mBio.00239-11. 11.

[6] Ferreira R, Sousa C, Gonçalves RFS, Pinheiro AC, Oleastro M, Wagemans J, et al. Characterization and genomic analysis of a new phage infecting *Helicobacter pylori*. Int J Mol Sci 2022;23:7885. https://doi.org/10.3390/IJMS23147885. 2022;23:7885.

[7] Luo C-H, Chiou P-Y, Yang C-Y, Lin N-T. Genome, integration, and transduction of a novel temperate phage of *Helicobacter pylori*. J Virol 2012;86:8781–92. https://doi.org/10.1128/JVI.00446-12.

[8] Vale FF, Nunes A, Oleastro M, Gomes JP, Sampaio DA, Rocha R, et al. Genomic structure and insertion sites of *Helicobacter pylori* prophages from various geographical origins. Sci Rep 2017;7:42471. https://doi.org/10.1038/srep42471.

[9] Silveira L, Nunes A, Pista A, Isidro J, Belo Correia C, Saraiva M, et al. Characterization of multidrug-resistant isolates of *Salmonella enterica* serovars heidelberg and Minnesota from fresh poultry meat imported to Portugal. Microb Drug Resist 2021;27:87–98. https://doi.org/10.1089/MDR.2019.0384.

[10] De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 2018;34:2666–9. https://doi.org/10.1093/bioinformatics/bty149.

[11] Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 2016;44:W3–10. https://doi.org/10.1093/NAR/GKW343.

[12] Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 2017;13. https://doi.org/10.1371/JOURNAL.PCBI.1005595.

[13] Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 2016;44:W16–21. https://doi.org/10.1093/nar/gkw387.

[14] Song W, Sun H-X, Zhang C, Cheng L, Peng Y, Deng Z, et al. Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Res 2019;47:W74–80. https://doi.org/10.1093/nar/gkz380.

[15] Starikova EV, Tikhonova PO, Prianichnikov NA, Rands CM, Zdobnov EM, Ilina EN, et al. Phigaro: high-throughput prophage sequence annotation. Bioinformatics 2020;36:3882–4. https://doi.org/10.1093/bioinformatics/btaa250.

[16] Lopes A, Tavares P, Petit MA, Guérois R, Zinn-Justin S. Automated classification of tailed bacteriophages according to their neck organization. BMC Genom 2014;15:1–17. https://doi.org/10.1186/1471-2164-15-1027/FIGURES/5.

[17] Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 2012;67:2640–4. https://doi.org/10.1093/JAC/DKS261.

[18] Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinf 2018;19. https://doi.org/10.1186/S12859-018-2336-6.

[19] Tetzschner AMM, Johnson JR, Johnston BD, Lund O, Scheutz F. *In Silico* genotyping of *Escherichia coli* isolates for extraintestinal virulence genes by use of whole-genome sequencing data. J Clin Microbiol 2020;58. https://doi.org/10.1128/JCM.01269-20/SUPPL_FILE/JCM.01269-20-S0002.PDF.

[20] Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol 2014;52:1501–10. https://doi.org/10.1128/JCM.03617-13.

[21] Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 2017:3396–404. https://doi.org/10.1093/bioinformatics/btx440.

[22] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinf 2013;14:60. https://doi.org/10.1186/1471-2105-14-60.
[23] Lefort V, Desper R, Gascuel O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program: table 1. Mol Biol Evol 2015;32:2798–800. https://doi.org/10.1093/molbev/msv150.
[24] Farris JS. Estimating phylogenetic trees from distance matrices. Am Nat 1972;106:645–68. https://doi.org/10.1086/282802.
[25] Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinformatics 2020;69. https://doi.org/10.1002/cpbi.96.
[26] Göker M, García-Blázquez G, Voglmayr H, Tellería MT, Martín MP. Molecular taxonomy of phytopathogenic fungi: a case study in peronospora. PLoS One 2009;4:e6319. https://doi.org/10.1371/journal.pone.0006319.
[27] Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, et al. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 2014;9:2. https://doi.org/10.1186/1944-3277-9-2.
[28] Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res 2023;51:W397–403. https://doi.org/10.1093/nar/gkad313.
[29] Muñoz AB, Trespalacios-Rangel AA, Vale FF. An American lineage of *Helicobacter pylori* prophages found in Colombia. Helicobacter 2021;26:e12779. https://doi.org/10.1111/hel.12779.
[30] Maqbool A, Horler RSP, Muller A, Wilkinson AJ, Wilson KS, Thomas GH. The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. Biochem Soc Trans 2015;43:1011–7. https://doi.org/10.1042/BST20150135.
[31] Gucinski GC, Michalska K, Garza-Sánchez F, Eschenfeldt WH, Stols L, Nguyen JY, et al. Convergent evolution of the barnase/EndoU/colicin/RelE (BECR) fold in antibacterial tRNase toxins. Structure 2019;27:1660–1674.e5. https://doi.org/10.1016/j.str.2019.08.010.
[32] Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. Card 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 2017;45:D566–73. https://doi.org/10.1093/nar/gkw1004.
[33] Muñoz AB, Stepanian J, Trespalacios AA, Vale FF. Bacteriophages of *Helicobacter pylori*. Front Microbiol 2020;11:12. https://doi.org/10.3389/fmicb.2020.549084.
[34] Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, et al. Recent changes to virus taxonomy ratified by the international committee on taxonomy of viruses (2022). Arch Virol 2022;167:2429–40. https://doi.org/10.1007/s00705-022-05516-5.