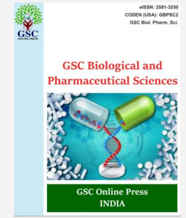
GSC Biological and Pharmaceutical Sciences

eISSN: 2581-3250 CODEN (USA): GBPSC2

Cross Ref DOI: 10.30574/gscbps

Journal homepage: https://gsconlinepress.com/journals/gscbps/



(Research Article)

Helicobacter pylori infection in Bamako (Mali): Polymorphism, diversity and prophage

David Guindo [1, 2, *], Alpha Seydou YARO [1, 3], Abdelaye KEITA [4], Yaya BOUARE [2], Youssouf Faya KEITA [1] And Bernard SODIO [1]

[1] Entomology and Parasitology Laboratory, Department of Biology, Faculty of Sciences and Techniques, University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali.
[2] Laboratoire de Microbiologie Appliquée, Department of Biology, Faculty of Sciences and Techniques, University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali.
[3] International Center for Excellence in Research, University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali.
[4] National Institute of Public Health of Bamako, Ministry of Health and Social Development of Mali, Bamako, Mali.

GSC Biological and Pharmaceutical Sciences, 2023, 24(01), 360–372

Publication history: Received on 08 May 2023; revised on 13 July 2023; accepted on 16 July 2023

Article DOI: https://doi.org/10.30574/gscbps.2023.24.1.0228

Abstract

Helicobacter pylori (*Hp*) is a bacteria associated with gastritis and gastric cancer (GC). Its prevalence is higher than 80% in Mali and affects the incidence of GC, which is a major cancer in the department of digestive surgery. VacA and CagA toxins are responsible of *Hp* strain polymorphism. With prophage, they are also associated with a region depending diversity. The aim of this study was to identify the variants of *Hp* strains and their prophage identity in Bamako (Mali). A total of 61 gastric biopsies were isolated in patients with gastric diseases. The extracted DNA was amplified by PCR targeting *cagA*, *vacA* (*s, m*) and phage integrase (*int*) genes. *cagA* and *int* amplicons were sequenced according to Sanger method. *Hp* was detected in 23 (37.7%) samples. A predominance of housewives, commercial workers, farmers and students was observed, confirming the importance of socio-economic conditions and the associated factors in *Hp* infection. Carcinogenic *vacAs1m1* was the second dominant polymorph. This would explain the prevalence of *Hp*'s GC in the country. The phylogenetic tree from *cagA* sequences has revealed the specific variants strain of *Hp* that causes gastritis in Mali (*Hp*ML). Their prophage integration protein shares more than 90% of its sequence with species of the *Schmidvirus* genus. Like the *Hp*ML probable strains, Malian's *Hp* prophage (*Hp*MLpΦ) is geographically clustered with hspWestAfrica1 (hpAfrica1). These results confirm the existence of probable *Hp* and bacteriophage specific strains to Mali and suggest that using the molecular diagnostics method for effective pathogen identification and rapid patient management.

Keywords: *Helicobacter pylori*; Polymorphism; Gastric cancer; Prophage; Diversity; Mali

1. Introduction

Helicobacter pylori (*Hp*) is one of the causative agents of gastric diseases such as chronic gastritis, ulcers, GC and extra gastric diseases whose causal relationship remains controversial. It was first isolated by Marshall and Warren between 1982 and 1983 [1-5]. The pathogen's diversity and virulence have changed during human migration [6, 7]. Globally, the prevalence averages around 50%. The infection rate can be higher than 90% in developing countries because the infection is strongly associated with socioeconomic conditions. In Mali, it varies between 21% and 89% in gastritis patients when serological or histological diagnostic method is used. The GC rate is estimated to be 20,2 per 100,000 men population and *Hp* infection has been recognized as its second major cause in the country. As opposed to Asian

* Corresponding author: David Guindo

Copyright © 2023 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

countries that have a very high GC incidence, African countries have less subjects that develop GC. This phenomenon is called the "African enigma" [8-12]. In addition, the prevalence of infection and GC is affected by factors like age, environment, ethnicity, region of origin, smoking, alcohol consumption, immune status, diet, and the performance of diagnostic methods [13-15]. The vacuolization cytotoxin (VacA), cytotoxin-associated gene (CagA) and other pro-inflammatory proteins are responsible for regional polymorphism. They contribute to bacterial virulence [15-17]. Thus, strains with *vacA s1m1* alleles are more carcinogenic than *s1m2*. The *s2m2* are non-virulent and *s2m1* have rarely been observed in the clinic. Because of this ability to develop GC, *Hp* was classified as a type I carcinogen by the WHO [18]. Their role in the evolution provides information on the geographic origin of strains, and consequently the carriers. They help identify native, nomadic and immigrant populations [15]. Therefore, *Hp* strains were grouped by geographical region. Groups can also include sub-populations: hpAfrica1 (includes hspWestAfrica1, hspSouthAfrica1, hspCentralAfrica1...) identified in West, South and Central Africa; hpNEAfrica identified in Northeast Africa, hpAfrica2 identified in South Africa, hpEurope identified in Europe, hpEAsia and hpAsia2 identified in Asia (also Amerindians), hpSahul in Oceania... [19-20]. The prophages found in *Hp's* genome also contribute to this strain diversification and virulence by over-expressing the CagA toxin [21]. After the prophage excision, lytic bacteriophages are proposed to be used in therapy to combat the antimicrobial resistance problem [22-26].

The objective of this research is to identify the diversity of *Hp* strains implicated in gastric pathologies in Bamako (Mali), and determine the presence of prophage in their genomes.

2. Material and methods

2.1. The site and period of the study

Gastric biopsies (antrum) were collected by clinicians from 61 patients in the gastroenterology department of two medical clinics in Bamako ("Clinique *LA PROMENADE DES ANGEVINS*" in Bagadadji and the "Polyclinique *GUINDO*" in Badala-Est). They were collected from January through December 2020. The 61 subjects included 31 men, 28 women and 2 whose gender was not disclosed. The two sites are frequented by the population category with higher or average incomes.

2.2. Gastric biopsy DNA extraction

The DNA was extracted in accordance with the protocol recommended by the manufacturer of the used kit (Wizard® Genomic DNA Purification Kit, Promega). They are summarized as follows: 30 mg of the gastric biopsy was crushed in 600 µl of a nucleic acid lysis solution before incubation for 30 min at 65 °C in the water bath (cell lysis); RNA was degraded with by adding 3 µl of RNase for 30 min at 37 °C; protein debris was precipitated by adding 200 µl of the protein precipitation solution (vortexed for 20 sec) and incubated for 5 min at 4 °C; the DNA-containing supernatant was transferred to 600 µL of isopropanol after centrifugation at 14,000 × g for 3 min; the DNA was precipitated after centrifugation at the same speed for 1 min; the pellet is washed in 600 µl of 70% ethanol by centrifuging the product as before; the resulting DNA pellet is dried and re-hydrated in 100 µl of DNA rehydration solution for 60 min at 65 °C before being stored in the refrigerator.

2.3. Amplification of target genes by the polymerase chain reaction (PCR)

After DNA extraction, some *Hp* polymorphism genes were amplified by.

The following reaction mixture was used in a final volume of 25 µl: 25-50 ng of template DNA, 0.5 ng of each primer (**Table 1**), 12.5 µl of GoTaqR G2 Hot Start Green Master Mix from Promega, and the necessary volume of water to obtain the 25 µl. The programs used for the PCR were: pre-denaturation 95 °C for 5 min; 30 cycles (denaturation 94 °C 1 min, annealing 52 °C 50 sec, extension 72 °C 50 sec) for *cagA*, 30 cycles (denaturation 95 °C 45 sec, annealing 57 °C 45 sec for *s1b*, 56 °C for *s2*, 58 °C for *m1b* and 48 °C for *m2*, extension 72 °C 70 sec) for *vacA*, 35 cycles of (denaturation 95 °C 30 sec, annealing 58 °C 30 sec, elongation 72 °C 1 min) for *int*; final extension 72 °C for 7 min. PCR amplicons were migrated in 1% agarose gel [Agarose LE, Molecular Biology Grade, Ultrapure Thermo Scientific de Thermo Fisher; tampon TBE 0.5X (Tris-borate-EDTA 10X, Thermo Fisher); 0.5 µg/ml of ethidium bromure, Promega] for 75 min at 90 volts.

The BenchTop pGEM® DNA Markers (36 à 2,645 pb) was used to determine fragment size

Table 1 Forward and reverse primers to amplify the target gene [21, 27]

Gene	Forward (5’-3’)	Reverse (5’-3’)	Size (bp)
cagA	GAT AAC AGC CAA GCT TTT GAG G	CTG CAA AAG ATT GTT TGG CAG A	350
vacAs1b	AGC GCC ATA CCG CAA GAG	CTG CTT GAA TGC GCC AAA C	187
vacAs2	GCT AAC ACG CCA AAT GAT CC	CTG CTT GAA TGC GCC AAA C	199
vacAm1b	GGC CCC AAT GCA GTC ATG GA	GCT GTT AGT GCC TAA AGA AGC AT	291
vacAm2	GGA GCC CCA GGA AAC ATT G	CAT AAC TAG CGC CTT GCA	352
int	AAG YTT TTT AGM GTT TTG YG	CGC CCT GGC TTA GCA TC	529

cagA- cytotoxin-associated gene A, *vacA*- vacuolizing gene A (*m*- middle region and *s*- signal), *int*- phage integrase

2.4. Sequencing and data analysis

The Chi-2 test, Anova test and Pearson correlation were used for statistical analyses using R 4.2.3 statistical software. The probability values were interpreted as follows: '***' *p*-value≤0.001 highly significant; '**' 0.001<*p*-value≤0.01 very significant; '*' 0.01<*p*-value≤0.05 significant; '.' 0.05<*p*-value≤0.1 weakly significant; ' ' *p*-value>0.1 no presumption against H_0 which is accepted, therefore the observed differences are not significant. The average, standard deviation, maximum, and minimum were also determined with the same software.

Sequencing of PCR products (18 for *cagA* and 1 for *int*) was realized by *Inquaba Biotec laboratory*, Pretoria-South Africa, according to their internal protocols and with the materials sent: minimum 10 µl of primers with a concentration of 10 µM (+1 µl for each reaction), minimum 15µl of PCR product. Sanger method was used for sequence reading (ABI 3500XL genetic analyzer, POP7™, BrilliantDye™ Terminator v3.1).

The SnapGene 6.0.2 software was used for visualization and correction of the obtained sequences. The alignment of the *cagA* sequences with those of the South and West African, Europe, Asia and America regions (hspSouthAfrica1, hspWestAfrica1, hpEurope, hpEastAsia, hpAsia2, hspAmerind), the construction of the phylogenetic tree by the FastME method (same principle like NJ method: inclusion of the evolutionary distance), were realized at the following website: https://ngphylogeny.fr/ [28]. The prophage was identified by comparing the obtained sequence with others in the NCBI (National Center for Biotechnology Information) database (https://www.ncbi.nlm.nih.gov/). All the sequences used for the study are accessible with special codes in NCBI (Appendix 1 and Appendix 2).

3. Results

3.1. Frequency of infection and associated factors

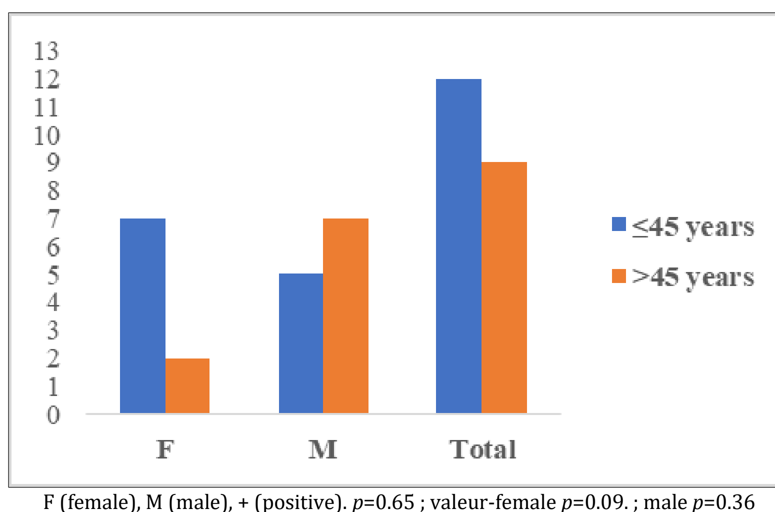
After DNA extraction, *Hp* was detected through amplification of *vacAs1b* and *cagA*. Result are summarized in **Table 2**.

Table 2 Presence of polymorphism gene in samples

		Age (year)			***Hp* specific gene**				
Sexe	***Hp+* (%)**	**Min**	**Aver (Sd)**	**Max**	***cagA***	***vacAs1b***	***vacAm1b***	***vacAm2***	***Int***
M	11/28 (47.8%)	22	38.33 ±15.35	67	N: 43 P: 18	P: 22 N: 39	N: 16 P: 6 NA: 39	N: 12 P: 9 NA: 40	N: 13 P: 7 NA: 41
F	12/31 (52.2%)	35	50.50 ±12.09	75					
Gene	23/61 (100%)	22	45.28 ±14.91	75					
	0.054.	*0.06.*			*0.00137***	*0.029**	*0.033**	*0.5*	*0.25*
	p-value								

'***' highly significant; '**' very significant; '*' significant; '.' weakly significant; ' ' not significant; *cagA*- cytotoxin-associated gene A, *vacA*- vacuolizing gene A (*m*- middle region and *s*- signal), *int*- phage integrase, *Hp+ Helicobacter pylori* positive, P- positive, N- negative, NA- non-acquired, Aver- average, Sd- standard deviation, Max- maximum, and Min- minimum.

In total, *Hp* was found in 23 of the 61 samples, or 37.7% of the subjects tested (p=0.054·). The average age of the subjects was 45.28 years (±14.91) (**Table 2**) with a predominantly under -45 age group (p=0.06·) (**Figure 1**).



F (female), M (male), + (positive). p=0.65 ; valeur-female p=0.09. ; male p=0.36

Figure 1 Frequency of *Helicobacter pylori* (*Hp*) infection by age group and sex

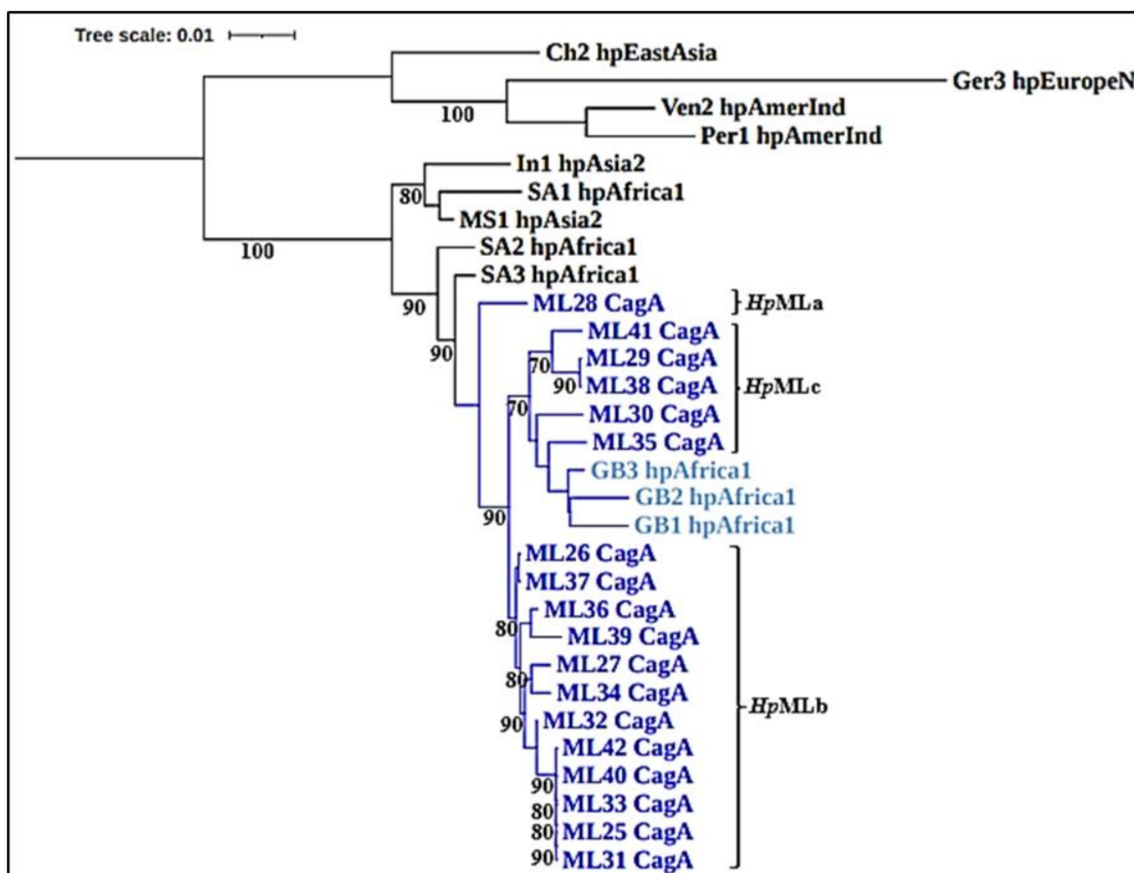
Housewives were the most represented (five positive cases), followed by farmers, commercial workers and students with three cases per occupation or profession (r=0.84; p=4e-9***) (**Table 3-a**). A predominance of multiple antral ex-ulceration (12 cases) was observed in *Hp*+ subjects (r=0.97; p=3e-11***) (**Table 3-b**).

Table 3 *Helicobacter pylori* (*Hp*) infectivity according to patients' profession or occupation (**a**); characterization of gastric lesions at endoscopy, and presence of *Hp* (**b**)

a	Profession or occupation	*Hp+*	b	Endoscopic examination conclusion	*Hp+* (%)
	Administrator	1		Erythematous antral gastritis	4
	Farmer	3		Erosive gastritis	4
	Bank executive	1		Multiple antral ulceration	52
	Commercial workers	3		Not acquired	9
	Accountant	1		Peptic esophagitis	4
	Student	3		Erythematous pangastritis	9
	Doctor	1		Gastro-esophageal reflux	4
	Housewife	5		Antro-fundic ulcer	4
	Military	1		Pyloro-bulbar ulcer	4
	Couturier	1		Large pre-pyloric ulcer	4
	Not acquired	3		Total	100
	Total	23		r=0.97; p=3e-11***	
	r=0.84; p=4e-9***				

3.2. *Helicobacter pylori* polymorphism, CagA sequence diversity

vacA gene alleles (*s2*, *m1b*, *m2*) also implicated in polymorphism have been amplified. The results of *vacAm1b* and *vacAm2* are summarized in **Table 2.** Six samples were *vacAm1b+* (p=0.033*) and 9 *vacAm2+*(p=0.5). No samples were *vacAs2*-positive. The combination of polymorphic genes allowed the identification of 6/18 samples with *cagA+ vacAs1bm1b* genes (26%); 8/18 *cagA+ vacAs1bm2* (35%); 3/18 samples *cagA+ vacAs1bm1bm2* (13%) and 5/23 samples *cagA- vacAs1b+* (22%).



Internal scale=0.01, Bootstrap=100. ML (Mali)- sequences from this study, Per (Pérou), Ch (Chine), In (Inde), MS (Malaisie), SA (South Africa), GB (Gambia), Ger (Germania), Ven (Venezuela) and *Hp* (*Helicobacter pylori*). NCBI accession code, ML (OP574206- OP574216), GB1-GB3 (NC_017371.1, NZ_KB636623.1, NZ_KB642259.1 and KY366700.1), SA1-SA3 (NZ_CBOS010000014.1, NZ_CBNE010000001.1 and NZ_CBPU010000011.1), MS1 (NZ_AUSN01000003.1), In1 (NZ_AJFA02000091.1), Ch2 (NZ_JH791774.1), Tai3 (NZ_JQNY01000023.1), Per1 (NC_017359.1), Ven2 (NC_017355.1), Ger3 (NZ_CP032908.1). The NGPhylogeny program was used to construct the tree [28].

Figure 2 Phylogenetic tree of partial *cagA* sequences based on the FastME method

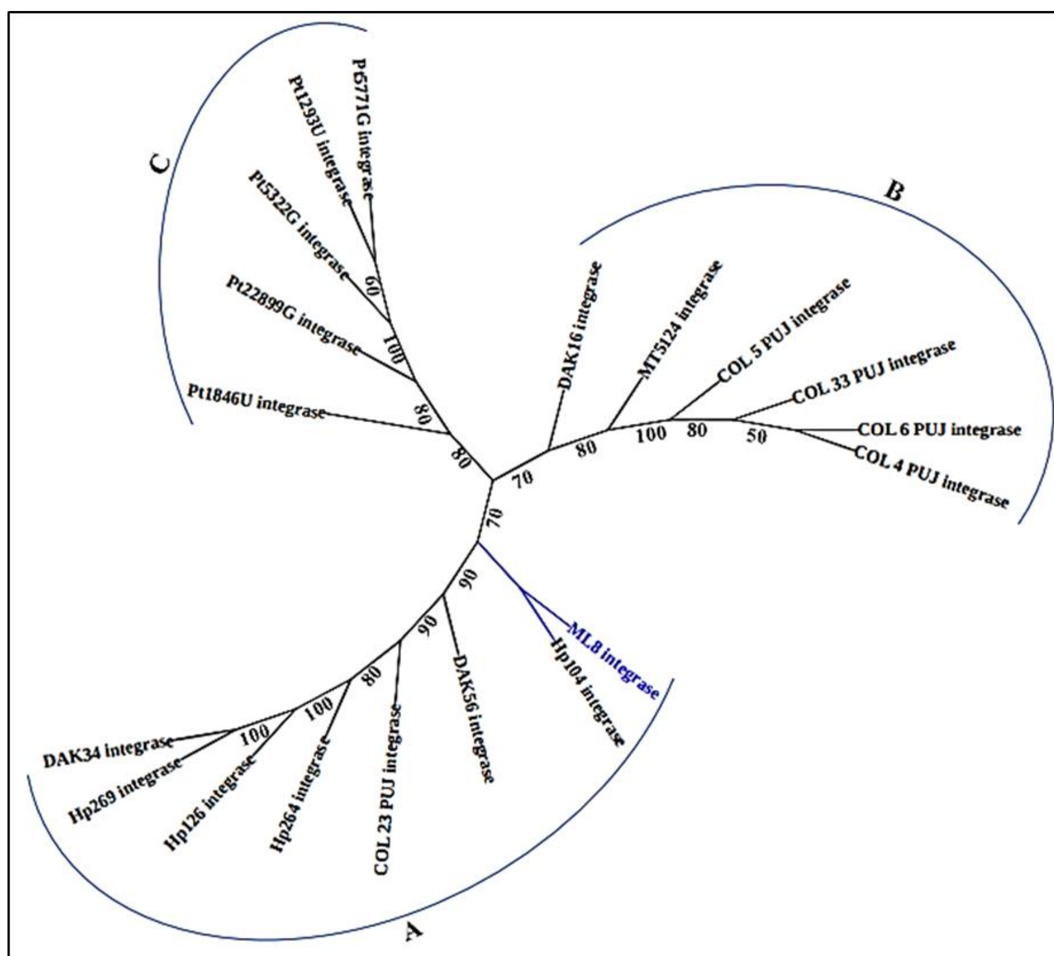
Based on the degree of sequence similarity, two groups and one isolate from Mali were observed: the first group (*Hp*MLb) includes ML34, ML27, ML29, ML40, ML33, ML32, ML31, ML25, ML42, ML26, ML37, ML36 and ML39; the second (*Hp*MLc) ML35, ML41, ML38, ML29, ML30; the third (*Hp*MLa) ML28. In addition, sequence identity between samples ML40, ML42, ML33, ML31, and ML25 was noted, also between ML29 and ML38, as well as ML26 and ML37 samples. This indicates that multiple patients have the same variants of *Hp*ML strain. It should be noted that South African strains form a different group. There proximities with Gambia strains indicate that ML variants are grouped with hspWestAfrica1, even if *Hp*MLa remain very near of hspSudAfrica1 (**Figure 2**).

Furthermore, ML35, ML38, ML39, ML40, ML41 and ML42 are *cagAvacAs1bm1b*; ML30, ML31, ML32, ML33, ML37, ML39, ML40 and ML42 are *cagAvacAs1bm2* polymorph. The ML39, ML40 and ML42 sequences are in *cagAvacAs1bm1bm2* polymorphism group. *cagA+vacAs1bm1b* is in majority *Hp*MLc variant and *cagA+vacs1bm2* the *Hp*MLb.

3.3. Identification of the isolated prophage integrase gene (*int*)

Prophages are also associated with diversification of *Hp* and can be used clinically for phage treatment (if they are lytic after excision). The *int* gene fragment was identified in 7/23 samples (p=0.25) (**Table 2**). One of these amplicons was sequenced and a resulting “ML8 integrase” sequence was analysed and compared with others from the NCBI database.

It shows more than 90% similarity with the prophage integrase isolated from *Hp's* DNA (**Appendix 3**). These prophages are classified in the Viruses superkingdom, the Duplodnaviria clade, the Heunggongvirae kingdom, the Uroviricota phylum, the Caudoviricetes class and the *Schmidvirus* genus. To determine the most evolutionary-related species, a phylogenetic tree (FastME method) was constructed with sequences of Gambia, Senegal (West Africa), Colombia (America) and Portugal (Europe) (**Figure 3**).



Bootstrap=100. A: West Africa, B: Portugal, C: Columbia. Pt for sequences isolated in Portugal (Europe), Col and MT from Colombia (South America), *Hp* for the sequences isolated in Gambia, DAK sequence from Dakar (Senegal), ML8 from Mali and *Hp* for *Helicobacter pylori*. NCBI accession code, COL 33-PUJ, 23-PUJ, 6-PUJ, 5-PUJ, 4-PUJ and *Hp* MT5124 (MW160256.1, MW247147.1, MW247145.1, MW247144.1, MW160250.1 and CP071986.1); Pt22899, 5322, 1293U, 1846U, and 5771G (KX119175.1, KX119198.1, KX119202.1, KX119176.1, and KX119199.1); *Hp*104, 126, 264 and 269 (KY366700.1, KY366706.1, KY366710.1 and KY366711.1); DAK 16, 34 and 56 (KM275907.1, KM275910.1 and KM275911.1); ML8 (OP697052). The NGPhylogeny program was used for tree construction [28].

Figure 3 Phylogenetic tree of prophage integrase gene sequences constructed using the FastME method

ML8 (*Hp*MLpΦ) evolve in the same group of prophages that have been isolated in Senegal (DAK) and Gambia (*Hp*), but is closer to the *Hp*104 prophage identified in Gambia (hspWestAfrica1). In addition, the prophages of Portugal and Colombia form separates groups.

4. Discussion

4.1. *Hp*ML infection and associated factors

In Africa, *Hp* infection prevalence is higher than the international average because it is strongly related to poor socioeconomic conditions, quality of daily food intake, family proximity, stress, and other predisposing factors [12]. As mentioned, in Mali, the prevalence varies between 21% and 89% in gastric patients when the diagnostic method is serologic or histologic [8, 10, 17]. The molecular method is the one used here. A prevalence of 37.7% was found in both medical clinics examined. These are frequented by a class of the population with higher income than the average. This can explain this low prevalence rate among the subjects tested. Housewives account for 21.7% of cases, in accord with the findings of earlier epidemiological studies. However, the number of cases increases for students (13%) and farmers (13%) [8, 29, 30]. This confirms the elevated risk of infection amongst the most vulnerable in terms of exposure and socio-economic conditions. The gastric microbiota serves as a protective barrier against pathogens. Its balance is negatively impacted by diet, the quality of which may depend on socio-economic conditions, as well as other factors such as stress (adrenaline), antibiotic therapy, anti-inflammatory drugs, smoking, alcohol consumption and immune status. Dysbiosis affects the gastric mucosa and favors the multiplication of *Hp* [31-33]. All these factors (stress, diet,

dysbiosis of the intestinal microbiota...) confirm the high risk of infection in the most vulnerable, both in terms of exposure and socio-economic conditions. The average age of the *Hp*+ population investigated was 45.28 years (±14.91). It is in the range of the mean previously seen in the country [8, 30]. In contrast to developed countries, those under 45 were found to be the most affected. This is the case in a number of other African countries [34-36]. The difference can be explained by cohort effect and higher life expectancy in developed countries [34].

4.2. *Hp*ML polymorphism and virulence

The causal relationship between the presence of CagA and VacA in polymorphism and the severity of *Hp*-related pathologies is no longer in question. It is also known that *vacAs1m1* strains are more virulent and likely to progress pathology to GC than *vacAs1m2* and *vacAs2m2*. 78.3% of the *Hp*+ samples in the study are *cagA*+. According to several studies, the prevalence of *cagA* varies from one country to another and is estimated between 29 and 91%. Mali's case remains similar to those in West Africa (Senegal and Ghana) [17]. About 26% of the samples were *vacAs1m1*, and 34.8% *vacAs1m2*. These data, in addition to other factors, would explain why the country has a low GC prevalence relative to Asia. But may also justify *Hp's* second position in GC development in Mali [9, 11]. A national study with more samples can be realized to confirm these results.

4.3. *Hp*ML diversity (*cagA* and *int*)

Three variants of *Hp* probable strains (*Hp*MLa, b and c) based on the divergence in their *cagA* sequences have been identified in this study (*Hp*MLc are associated in ulcer symptom). These are classified in the sub-population of *Hp* strains from West Africa known as "hspWestAfrica1". The majority of the sequences in the sub-population have been isolated in Gambia and Senegal [19-20]. Indeed, Mali and the Gambia have beyond the geographic proximity, various commonalities (culture, part of the history as well as culinary habits). The Gambia also includes several ethnic groups whose languages are considered national languages in Mali: The Malinke who represent 38% of their population, the Fulani 14%, the Soninke about 9% as well as other minorities such as the Bambara and the Khassonke [37-39]. All these factors contribute to the proximity of the strains in circulation in the two countries. Moreover, the hspWestAfrica1 where the *Hp*ML resides, is of the same group (hpAfrica1) as the hspSouthAfrica1. Strains of this group have been isolated from individuals speaking Niger-Congo languages and the majority of Malian languages are of this family. This diversity in the *cagA* sequence can also generate various antibodies against CagA. Thus, besides the omission of *cagA*-variants, serology tests based on the detection of the CagA protein or its antibodies can be biased. These data support the Bamako prevalence study's hypothesis in 2013 [10]. The research was conducted with the serological identification of *Hp* CagA and the prevalence obtained was among the lowest in the country (21%). It was concluded that CagA diversity would account for the lowest prevalence in comparison with the national average [10].

Hp prophage inserts the integral protein into the host chromosome, evolves with it, and permits the identification of the origin of the host [40]. The results of the integrase sequence analysis in this study show that ML8 *int* shares over 90% of its sequence with those of the genus *Schmidvirus*. From an evolutionary point of view, it shares a common ancestor with the Gambia and Senegal prophage *int*. It is therefore geographically associated with hspWestAfrica1 prophages. This re-affirms the possibility of traceability of human migration. In addition to evolution, *Hp* diversity can be used to see the story of the mixture in a region. This is the case in Gambia, where hpNEAfrica (North-East Africa) has been identified, although the country is mostly colonized by hpAfrica1 strains [20]. In the United States, the diversity of strains is even more informative, because it retraces the nation's immigration history. For example, similar strains of hspWestAfrica1 have been found in Afro-American populations [15]. This method may be used to investigate the origin of the various ethnic groups present in Mali and to provide another element that will reduce the contradictions resulting from the oral.

As indicated, other prophages have been identified in *Hp*. Some of them (HP1, KHP30, PhiHp33, HPy1R...), after excitation can infect and induce *Hp* lysis, even under conditions similar to those of the gastric environment [21, 41]. The question remains whether Malian prophages (*Hp*MLpΦ) are genetically diverse on all genes and if after excision of the bacterial genome the virions produced have lytic activity on the *Hp*ML strains in the extreme conditions of the stomach. A complementary study will be required to further investigate these preliminary data.

5. Conclusion

In conclusion, *Hp* infection remains a public health issue in Mali. The presence of carcinogenic strains in the country can explain the incidence of GC. This study report that there are at least four variants of the probable *Hp*ML strain which are involved in the development of gastric pathologies in Mali (3 *cagA+* and 1 *cagA-*). Some of them have a prophage in their chromosome. Boths *Hp*ML and *Hp*MLpΦ are geographically associated with hspWestAfrica1 (hpAfrica1) sub-group. Their diversity might be a key to tracing the history of the origins of populations in the country. Thus, detection by the

molecular method must be adapted and favored in the Malian context, for an accurate diagnosis of *Hp* and an effective management of patients.

Compliance with ethical standards

Acknowledgments

We thank the agents of the gastroenterology department of the "Polyclinique *GUINDO*" of Bala-Est (Mali) and the "Clinique *LA PROMENADE DES ANGÉVINS*" of Bagadadji (Mali) for agreeing to provide the gastric biopsies. We also thank Malian Government "Training of Trainers Program (TTP) " for its partial support to this work.

Disclosure of conflict of interest

The authors declared no competing conflict of interest.

Statement of ethical approval

Patient recruitment and sampling was done by the gastroenterologists responsible for the patients. Informed consent was obtained from participants at both centers [42]. The results were anonymized and kept confidential. The data generated by the study were utilized in a strictly scientific context.

Statement of informed consent

Informed consent was obtained from all individual participants included in the study.

Data availability statement

All sequences identified in this study have been deposited in the GenBank. The accession numbers are OP574206 to OP574216 for *cagA* and OP697052 for *int*.

Funding

This research was partially supported by the "Training of Trainers Program (TTP) of Mali".

References

[1] Robin Warren J, Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. The Lancet. juin 1983, 321(8336):1273-5.

[2] Marshall BJ, Royce H, Annear DI, Goodwin CS, Pearman JW, Warren JR, and Armstrong JA. Original isolation of Campylobacter pyloridis from human gastric mucosa. Microbios. Lett. 1984, 25 :83-88.

[3] Itoh T, Yanagawa Y, Shingaki M, Takahashi M, Kai A, Ohashi M, et al. Isolation of Campylobacter pyloridis from human gastric mucosa and characterization of the isolates. Microbiology and Immunology. 1987, 31(7):603-14.

[4] Richy F, Mégraud F. L'infection par *Helicobacter pylori* responsable d'affections extra-digestives : mythe ou réalité ? Gastroenterol Clin Biol. 2003, (27):459-66.

[5] Gravina AG, Zagari RM, Musis CD, Romano L, Loguercio C, Romano M. *Helicobacter pylori* and extragastric diseases: A review. WJG. 2018, 24(29):3204-21.

[6] Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 2007, 445(7130):915-8.

[7] Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, et al. Age of the association between *Helicobacter pylori* and man. Ochman H, éditeur. PLoS Pathog. 2012, 8(5): e1002693.

[8] Konate A, Diarra M, Soucko-Diarra A, Dembele M, Bah N, Kalle A, et al. Chronic gastritis at the era of *Helicobacter pylori* in Mali. Acta Endosc. 2007, 37(3):315-20.

[9] Togo A, Diakite I, Togo B, Coulibaly Y, Kante L, Dembele BT, et al. Gastric cancer at the CHU Gabriel-Touré : epidemiological and diagnostic aspects. J Afr Cancer. 2011, 3(4):227-31.

[10] Austarheim I, Inngjerdingen KT, Togola A, Diakite C, Diallo D, Paulsen BS. Chromatographic immunoassays for *Helicobacter pylori* detection - are they reliable in Mali, West Africa? Pan Afr Med J [Internet]. Pan Afr Med J. 2013, 2. https://doi.org/10.11604/pamj.2013.14.72.2131.

[11] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clin. 2021, 71(3):209-49.

[12] Smith SI, Ajayi A, Jolaiya T, Onyekwere C, Setshedi M, Schulz C, et al. *Helicobacter pylori* infection in Africa: Update of the current situation and challenges. Dig Dis. 2022, 40(4):535-44.

[13] Zamani M, Ebrahimtabar F, Zamani V, Miller WH, Alizadeh-Navaei R, Shokri-Shirvani J, et al. Systematic review with meta-analysis: the worldwide prevalence of *Helicobacter pylori* infection. Aliment Pharmacol Ther. 2018, 47(7):868-76.

[14] Smith S, Fowora M, Pellicano R. Infections with *Helicobacter pylori* and challenges encountered in Africa. WJG. 72019, 25(25):3183-95.

[15] Muñoz-Ramirez ZY, Pascoe B, Mendez-Tenorio A, Mourkas E, Sandoval-Motta S, Perez-Perez G, et al. A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. ISME J. 2021, 15(1):78-92.

[16] Jeyamani L, Jayarajan J, Leelakrishnan V, Swaminathan M. CagA and VacA genes of *Helicobacter pylori* and their clinical relevance. Indian J Pathol Microbiol. 2018, 61(1):66.

[17] Ofori EG, Adinortey CA, Bockarie AS, Kyei F, Tagoe EA, Adinortey MB. *Helicobacter pylori* infection, virulence genes' distribution and accompanying clinical outcomes: The West Africa situation. BioMed Research International. 2019, 2019:1-13.

[18] IARC Working Group. *Helicobacter pylori*. World Health Organization, 2012 p. 385-435.

[19] Linz B, Vololonantenainab CRR, Seck A, Carod JF, Dia D, Garin B, et al. Population genetic structure and isolation by distance of *Helicobacter pylori* in Senegal and Madagascar. Hold GL, éditeur. PLoS ONE. 2014, 9(1): e87355.

[20] Secka O, Moodley Y, Antonio M, Berg DE, Tapgun M, Walton R, et al. Population genetic analyses of *Helicobacter pylori* isolates from Gambian adults and children. Boneca IG, éditeur. PLoS ONE. 2014, 9(10): e109466.

[21] Muñoz AB, Stepanian J, Trespalacios AA, Vale FF. Bacteriophages of *Helicobacter pylori*. Front Microbiol. 2020, 11:549084.

[22] Heintschel Von Heinegg E, Nalik HP, Schmid EN. Characterisation of a *Helicobacter pylori* phage (HP1). J Med Microbiol. 1993, 38:245-9.

[23] Lehours P, Vale FF, Bjursell MK, Melefors O, Advani R, Glavas S, et al. Genome sequencing reveals a phage in *Helicobacter pylori*. Rappuoli R, éditeur. mBio. 2011, 2(6): e00239-11.

[24] Uchiyama J, Takeuchi H, Kato S ichiro, Takemura-Uchiyama I, Ujihara T, Daibata M, et al. Complete genome sequences of two *Helicobacter pylori* bacteriophages isolated from Japanese patients. J Virol. 2012, 86(20):11400-1.

[25] Uchiyama J, Takeuchi H, Kato S ichiro, Gamoh K, Takemura-Uchiyama I, Ujihara T, et al. Characterization of *Helicobacter pylori* bacteriophage KHP30. Appl Environ Microbiol. 2013, 79(10):3176-84.

[26] Zou Q, Wei W. Phage therapy: Promising for H. pylori infection. Clin Microbial. 2013, 02(4):112.

[27] Gilani A, Razavilar V, Rokni N, Rahimi E. *VacA* and *cagA* genotypes status and antimicrobial resistance properties of *Helicobacter pylori* strains isolated from meat products in Isfahan province, Iran. Iran J Vet Res. 2017, 18(2):97-102.

[28] Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Research. 2019, 47(W1): W260-5.

[29] Maiga MY, Traore HA, Diarra M, Pichard E, Dembele M, Diallo AN, et al. Anatomo-clinical study of chronic gastritis in Mali. Black African medecine. 1996 , 5.

[30] KENGNE TINE SC. Epidemiogical and histopathological study of chronic gastritis in Mali about 1089 cases. [State-certified medical doctor]. [Bamako-MALI] : Faculty of Medecine, Pharmacy and Odonto-Stomatology , 2006.

[31] Boyanova L, Markovska R, Mitov I. *Helicobacter pylori* growth stimulation by adrenaline detected by two methods. Diagnostic Microbiology and Infectious Disease. 2019, 93(1):30-2.

[32] Mărginean CO, Meliț LE, Săsăran MO. Gastric microenvironment—A partnership between innate immunity and gastric microbiota tricks *Helicobacter pylori*. JCM. 2021, 10(15):3258.

[33] Liatsos C, Papaefthymiou A, Kyriakos N, Galanopoulos M, Doulberis M, Giakoumis M, et al. *Helicobacter pylori*, gastric microbiota and gastric cancer relationship: Unrolling the tangle. WJGO. 2022, 14(5):959-72.

[34] de Korwin JD, Lehours P. *Helicobacter pylori* : fondamatal notions , epidemiology, diagnonstic methods diagnostiques. EMC - Gastro-enterology. 2010, 5(3):1-16.

[35] Andoulo FA, Noah Noah D, Tagni-Sartre M, Ndjitoyap EC, Ngu Blackett K. Epidemiology of *Helicobacter pylori* infection in Yaounde : from particularity to Africa enigma. Pan Afr Med J. 2013 , 16. https://doi.org/10.11604/pamj.2013.16.115.3007.

[36] Essadik A, Benomar H, Rafik I, Hamza M, Guemouri L, Kettani A, et al. Epidemiological and clinical aspects of *Helicobacter pylori* infection through a Moroccan study : Hegel. 2013 , 3(3) :163-9.

[37] Konate MK, Diabate I, Assima A. Dynamics of local languages the french language in Mali : insight through general population censuse (1987 et 1998). ODSEF. 2010 p. 55.

[38] Laval University. Gambia, Republique of Gambia [Internet]. © 2022 [cited 2022 Nov 02]. Available from https://www.axl.cefan.ulaval.ca/afrique/gambie.htm .

[39] Sharma GS. The ethnic groups of Mali : in the context of culture and language. International Journal of Multidsciplinary Educational Research. 2022, 11(1):3.

[40] 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):8.

[41] 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*. IJMS. 2022, 23(14):7885.

[42] ASSOCIATION (WMA). WMA Declaration of Helsinki – Ethical principles for medical research involving humans subject [Internet]. © 2023 [cited 2023 Jan 17]. Available from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Appendix

Appendix 1

Table 4 Cytotoxin associated gene (*cagA*) sequence and their references

NCBI reference	*Helicobacter pylori* (strain or isolate)	Region	Gene
OP574206.1	*Helicobacter pylori* CagA24	hpAfrica1	*cagA*
OP574207.1	*Helicobacter pylori* CagA27	hpAfrica1	*cagA*
OP574208.1	*Helicobacter pylori* CagA28	hpAfrica1	*cagA*
OP574209.1	*Helicobacter pylori* CagA29	hpAfrica1	*cagA*
OP574210.1	*Helicobacter pylori* CagA30	hpAfrica1	*cagA*
OP574211.1	*Helicobacter pylori* CagA34	hpAfrica1	*cagA*
OP574212.1	*Helicobacter pylori* CagA35	hpAfrica1	*cagA*
OP574213.1	*Helicobacter pylori* CagA36	hpAfrica1	*cagA*
OP574214.1	*Helicobacter pylori* CagA37	hpAfrica1	*cagA*
OP574215.1	*Helicobacter pylori* CagA39	hpAfrica1	*cagA*
OP574216.1	*Helicobacter pylori* CagA41	hpAfrica1	*cagA*
NZ_CBOS010000014.1	*Helicobacter pylori* SA162A	hpAfrica1	*cagA*
NZ_CBNE010000001.1	*Helicobacter pylori* SA35A	hpAfrica1	*cagA*
NZ_CBPU010000011.1	*Helicobacter pylori* SA168A	hpAfrica1	*cagA*
NC_017371.1	*Helicobacter pylori* Gambia94/24	hpAfrica1	*cagA*
NZ_KB636623.1	*Helicobacter pylori* GAM260ASi Scfld161	hpAfrica1	*cagA*
NZ_KB642259.1	*Helicobacter pylori* HP116Bi (Gambie) Scfld185	hpAfrica1	*cagA*
NZ_AUSN01000003.1	*Helicobacter pylori* UM067 contig11	hpAsia2	*cagA*
NZ_AJFA02000091.1	*Helicobacter pylori* NAB47 contig091	hpAsia2	*cagA*
NZ_JH791774.1	*Helicobacter pylori* HLJHP193 Scaffold1	hpEastAsia	*cagA*
NZ_JQNY01000023.1	*Helicobacter pylori* Taiwan-47_contig4B	hpEastAsia	*cagA*
NC_017359.1	*Helicobacter pylori* Sat464	hpAmerInd	*cagA*
NC_017355.1	*Helicobacter pylori* v225d	hpAmerInd	*cagA*
NZ_CP032908.1	*Helicobacter pylori strain 23-A-EK1*	hpEuropeN	*cagA*

Appendix 2

Table 5 Phage integrase gene (*int*) sequence and their references

NCBI reference	*Helicobacter pylori* (strain or isolate)	Continent or country	Gene
OP697052	*Helicobacter* phage ML-8	Mali	*int*
KY366700.1	*Helicobacter pylori* Hp104 prophage integrase (*int*) gene	Gambia	*int*
KY366706.1	*Helicobacter pylori* Hp126 prophage integrase (*int*) gene	Gambia	*int*
KY366709.1	*Helicobacter pylori* Hp264 prophage integrase (*int*) gene	Gambia	*int*
KY366711.1	*Helicobacter pylori* Hp269 prophage integrase (*int*) gene	Gambia	*int*
KM275907.1	*Helicobacter pylori* DAK16 prophage integrase gene	Senegal	*int*
KM275910.1	*Helicobacter pylori* DAK34 prophage integrase gene	Senegal	*int*
KM275911.1	*Helicobacter pylori* DAK56 prophage integrase gene	Senegal	*int*
MW160256.1	*Helicobacter phage* COL 33-PUJ	South America	*int*
MW247147.1	*Helicobacter* phage COL 23-PUJ	South America	*int*
MW247145.1	*Helicobacter* phage COL 6-PUJ	South America	*int*
MW247144.1	*Helicobacter* phage COL 5-PUJ	South America	*int*
MW160250.1	*Helicobacter phage* COL 4-PUJ	South America	*int*
CP071986.1	*Helicobacter pylori* MT5124 chromosome	South America	*int*
KX119176.1	*Helicobacter* phage Pt1846U	Europe	*int*
KX119199.1	*Helicobacter* phage Pt5771G	Europe	*int*
KX119202.1	*Helicobacter* phage Pt1293U	Europe	*int*
KX119198.1	*Helicobacter* phage Pt5322G	Europe	*int*
KX119175.1	*Helicobacter* phage Pt22899G	Europe	*int*

Appendix 3

Table 6 Specific *Helicobacter pylori* prophage sequences with the highest percentage of similarity to ML8 integrase

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Helicobacter phage COL 23-PUJ, complet genome	*Helicobacter* phage COL 23-PUJ	644	644	100%	8e-179	91.28%	28622	MW247147.1
Helicobacter phage Pt22899G, complet genome	*Helicobacter* phage Pt22899G	617	617	100%	4e-172	90.43%	30078	KX119175.1
Helicobacter phage COL 6-PUJ, complet genome	*Helicobacter* phage COL 6-PUJ	617	617	100%	4e-172	90.41%	26487	MW247145.1
Helicobacter phage COL 5-PUJ, complet genome	*Helicobacter* phage COL 5-PUJ	617	617	100%	4e-172	90.41%	26486	MW247144.1
Helicobacter phage Pt5322G, complet genome	*Helicobacter* phage Pt5322G	608	608	99%	2e-169	90.17%	28341	KX119198.1
Helicobacter pylori strain phage Hp104 prophage integrase (*int*), complet genome	*Helicobacter pylori*	601	601	85%	4e-167	93.75%	442	KY366700.1
Helicobacter phage Pt1293U, complet genome	*Helicobacter* phage Pt1293U	601	601	99%	4e-167	89.94%	30071	KX119202.1
Helicobacter phage COL 23-PUJ, complet genome	*Helicobacter* phage COL 23-PUJ	595	595	100%	2e-165	89.60%	1661144	CP071986.1
Helicobacter phage Pt22899G, complet genome	*Helicobacter* phage Pt22899G	592	592	99%	2e-164	89.51%	27960	KX119176.1
Helicobacter phage COL 6-PUJ, complet genome	*Helicobacter* phage COL 6-PUJ	586	586	99%	1e-162	89.32%	29801	KX119199.1

Col (Colombia), Pt (Portugal), *Hp*104 (*Helicobacter pylori* 104, Gambian strain)