EJMA

DOI: 10.14744/ejma.2022.76476
EJMA 2022;2(3):126–135

Research Article

In silico Molecular Characterization of *Helicobacter pylori* based on Tandem Repeat Number and 16S rRNA gene

Arijit Sankar Mondal,[1] Ritu Sharma,[1] Arunava Das,[2] Ashish Warghane[2]

[1]Department of Microbiology, Guru Nanak Institute of Pharmaceutical Science and Technology (Life Science), Kolkata, West Bengal, India
[2]Faculty of Life Sciences, Mandsaur University, Mandsaur, Madhya Pradesh, India

Abstract

Objectives: *Helicobacter pylori* is the major cause of gastrointestinal carcinoma. The presence of tandem repeats are mainly responsible for increasing the mutation rate, thereby enhancing the virulence and adaptation of the pathogen to its host. Similarly, 16S ribosomal RNA sequence analysis has been widely used for determining the phylogenetic and taxonomic classification of bacterial strains.

Methods: The sequences of *H. pylori* were retrieved from GenBank, the National Centre for Biotechnology Information database, in order to identify specific tandem repeats. In our study, the genomic instability in *H. pylori* due to the presence of consecutive and non-consecutive sequences has been comparatively analyzed using MEGA X software, with consecutive sequences representing the tandem repeats.

Results: Based on the occurrence of Tandem repeats in the retrieved isolates of *H. pylori*, the consecutive sequences were classified into four different classes (Class I-IV). Further, the tandem repeat and 16S rRNA sequences were classified using phylogenetic analysis into two different clades for determining species variability and evolutionary relationships, respectively.

Conclusion: The data obtained from this study can be used to characterize different *H. pylori* strains, gain a better understanding of its genome variability leading to a higher mutation rate and depict the intra-species evolutionary relationship.

Keywords: Tandem repeats, 16S rRNA, *Helicobacter pylori*, Mutation, Pathogenesis, Phylogenetic analysis

Cite This Article: *Mondal AS, Sharma R, Das A, Warghane A. In silico Molecular Characterization of Helicobacter pylori based on Tandem Repeat Number and 16S rRNA gene. EJMA 2022;2(3):126–135.*

Helicobacter pylori is a gram-negative, rod-shaped motile bacterium found to infect the humans in the late 19th century.[1] However, its intimacy with human beings as a disease causal agent have been observed around 100,000 years ago due to its co-evolution and transmission throughout the generations.[2] The *Helicobacter* genus belongs to the family ε-proteobacteria and order *Helicobacteraceae* (formerly *Camphylobacterales*). *Helicobacter* species are broadly subdivided into two major classes: gastric Helicobacter species (*Helicobacter felis*, *Helicobacter pylori*, *Helicobacter acinomychis*, and *Helicobacter mustelae*) and entero-hepatic Helicobacter species (mostly *Helicobacter hepaticus*). The gastric species being urease positive are found to colonize the gastric lumen *via* chemotactic motility.[1] Among all other gastric species, *H. pylori* alone are responsible for 50% of infections.[3] Further, *H. pylori* is also one of the most recombinogenic human infections known.[4] However, the enterohepatic species are non-gastric and found to colonize in the intestine or lower gastrointestinal tract exhibiting persistent infection.[1]

The frequency of *H. pylori* infection is comparatively lower in the developed countries than the under-developed countries except for Japan.[5] Moreover, horizontal transmission

Address for correspondence: Ashish Warghane, PhD. Faculty of Life Sciences, Mandsaur University, Mandsaur, Madhya Pradesh, India
Phone: 9359101533 **E-mail:** warghane.ashish@gmail.com

Submitted Date: July 17, 2022 **Accepted Date:** August 15, 2022 **Available Online Date:** October 10, 2022

inducing multi-strain infection has been well reported in developing countries.[3] The colonization of *H. pylori* depicts a long-term relationship with gastric carcinoma (commonly known as gastric cancer, gastric lymphoma, or gastric ulcer) resulting in millions of deaths every year[5] which has categorized this bacterium as a Group A carcinogen.[6]

The pathogenesis initiates from the entrance of the bacteria to the host stomach and the urease activity is the first step conferred by the bacteria to neutralize the pH of the acidic stomach. The flagella motility in the bacteria confirms the colonization at the gastric mucosa followed by chemotaxis and persistent infection. The bacteria release several adhesins to protect themselves from the stomach forces and to bind to the gastric epithelial cells. Finally, the bacteria release toxins to disrupt the host cells and induce bacterial replication.[7] The occurrence of gastric carcinoma initiates from the *H. pylori* interaction with the host's gastric epithelium by inducing radical oxygen species (ROS) generation thereby increasing aquaporin 3 (AQP3) expression *via* transcription factor HIF1α and other specific promoter elements. The ROS and activated AQP3 have been identified to partially elevate the pro-inflammatory signaling of the innate immune system such as IL-6, IL-8, and TNF *via* ROS- HIF1α-AQP3-ROS loop synergistic activity. Further, the bacteria sequester cholesterol *via* cholesterol-α-glucosyltransferase (cgt) and alter the pro-inflammatory signalling along with alteration in interferon-gamma (IFNγ) and IL-22 secretion.[8]

H. pylori is the first pathogenic bacteria in which the whole genome sequencing has been done in different strains due to its high mutation rate.[9] The genomes of *H. pylori* are characterized by various strains such as 26695, HP14039, G27, and J99 including 1587, 1574, 1515, and 1491 gene sequences, respectively.[1, 9, 10] The mutation rate of *H. pylori* is very high depending upon two major factors: geographical location and phage transduction. This mutation results in the sudden evolution of the bacteria into several strains and has been identified to carry various pathogenic or virulent genes responsible for its adverse pathogenicity.[4] In contrast to the wild-type strains, 26695 and G27, the HP1275 and HP0044 knockout mutant strains of *H. pylori* showed decreased Lipopolysaccharide (LPS) production, increased biofilm aggregation, and higher susceptibility to detergent and antibiotic novobiocin.[11] The gene regulation and expression in *H. pylori* are mediated by gene alterations and slipped strand mispairing which is majorly responsible for its variations induced *via* recombination. Additionally, the natural incorporation of DNA sequence from its distinct strains *via* competency increases its variation frequency with huge mutant outcomes.[12] Alongside, the increased mutation has resulted in the bacteria becoming antimicrobial-resistance (clarithromycin and fluoroquinolone), identified by analyzing Polymerase Chain Reaction (PCR) assays in 23S rRNA and gyr A genes.[13] Moreover, the pathogenicity of *H. pylori* depends on certain parameters: colonizing factors (urease, hydrogenase, chemotaxis, and flagella), cell surface proteins/adhesins (BabA, SabA, OipA, and HopQ), and virulence factors (CagA, CagY, VacA, and HtrA). Among all the virulence factors CagA and CagY are encoded by Cag PAI genes and are majorly responsible for the occurrence of persistent infection and neoplastic gastric carcinoma associating different gastrointestinal diseases by altering the host cytokine-mediated inflammatory response (interleukin (IL)1-10 and tumor necrosis factor (TNF)-α).[5, 8, 12] Moreover, CagA-negative *H. pylori* are either nonpathogenic or only mildly pathogenic.[14] Simultaneously, geography, age, sex, ethnicity, alcohol consumption, non-ulcer dyspepsia, and prior *H. pylori* therapy have become the major clinical parameters of anti-microbial resistance.[13]

Similar to other bacteria, the *H. pylori* genome also exhibits direct repetitive polymorphic repetitive sequences but the frequent mutations in their genome have led to the formation of diverse strains within a single species which makes the study more challenging with less available *in silico* data. These sequences are known as tandem repeats (TRs)[15] which are widely present in the prokaryotic and eukaryotic genomes. These repeats are responsible for mutations that occur due to various adaptation strategies towards several environmental stress and geographical locations resulting in bacterial instability. Nonetheless, the TRs are considered junk or non-functional sequences but can modulate the host-pathogen interaction[16] along with its infection frequency which drives the prophylaxis and vaccination strategies more difficult. Similarly, 16S rRNA (16S ribosomal RNA) sequence (as a universal or housekeeping genetic biomarker) analysis has been widely used for the phylogenetic and taxonomic classification of bacteria. The practice of using these gene sequences for the aforementioned statement is due to the presence of operons which make them universal markers for bacterial identification, phylogeny that can be well understood due to its long-term persistence and the large sequence length of 1500bp (16s rRNA gene) that magnifies its value for the informatics study.[2]

The clinical and phylogeographic significance of *H. pylori* infection has been a milestone in the human evolutionary migration which depicts the scramble relationship between the bacteria and human ancestors. Therefore, phylogenetic analysis can be witnessed to be used as a demographic tool for investigating the migration and evolution of humans along with the bacteria.[2] Several laboratory methods including invasive and non-invasive have been introduced for the diagnosing of bacterial presence depending on various clinical conditions. The invasive methods include: histology, culture, rapid urease test, endoscopy, etc., and

the non-invasive methods for primary treatment include: stool antigen test, serology test, urea breath test, etc.[17] Apart from these detection techniques, strain identification is manually done by using two molecular techniques: PCR and Multi Locus Sequence Analysis (MLSA).[3]

However, the strain detection technique in recent days has become advanced through *in silico* strategy by analyzing Tandem Repeat Number (TRN) and 16S rRNA sequences for rapid identification of the similarity and variations between the closely related species and strains depending on geographical locations and host-pathogen interaction. Therefore, the present study was conducted to characterize *H. pylori* using TRN and 16S rRNA sequences. To the best of our knowledge, this is the first in-silico study in which *H. pylori* strains are divided into four groups based on the presence of tandem repeats (Fig. 1).

Methods

Sequence Retrieval

The sequences for the *H. pylori* TRNs were retrieved from the National Centre for Biotechnology Information (NCBI) database (www.ncbi.nig.gov). BLASTn search was used to collect the nucleotide sequences from similar strains. Simultaneously, the 16S rRNA nucleotide sequences were retrieved from the GenBank (NCBI). The multiple sequence alignment (MSA) for the retrieved sequences was performed using CLUSTAL_W in MEGA X software.[18]

Identification of Tandem repeats

The tandem repeats were identified after the multiple sequence alignment (Pairwise check) of the nucleotide sequence was performed by MEGA X.[18] The two possible TRs ('GATTAG' and 'TGATTAGT') were identified in the master sequence of *H. pylori* strain 26695-1MET with the accession number CP010436 from the NCBI database.

Retrieval of 16S rRNA Sequence of H. pylori

Therefore, the 16S rRNA sequences for *H. pylori* were retrieved directly from the NCBI database. The sequences between 1000-1500 bp approximately were selected. Using MEGA X software, those selected sequences were trimmed from both sides to extract sequences with an equal number of base pairs with gaps followed by aligning them using CLUSTAL_W. The aligned sequences were further utilized for determining the evolutionary relationship *via* phylogenetic analysis.

Phylogenetic Analysis

The phylogenetic analysis was performed for understanding the evolutionary relationship and homology between the *H. pylori* strains. The MSA was carried out for both the cases (TRs and 16S rRNA) using the default CLUSTAL_W program in the MEGA X software. About 45 sequences (TRs) and 30 representative sequences (16S rRNA) of *H. pylori* strains were retrieved, aligned separately and inferred for individual and collective phylogenetic tree analysis,

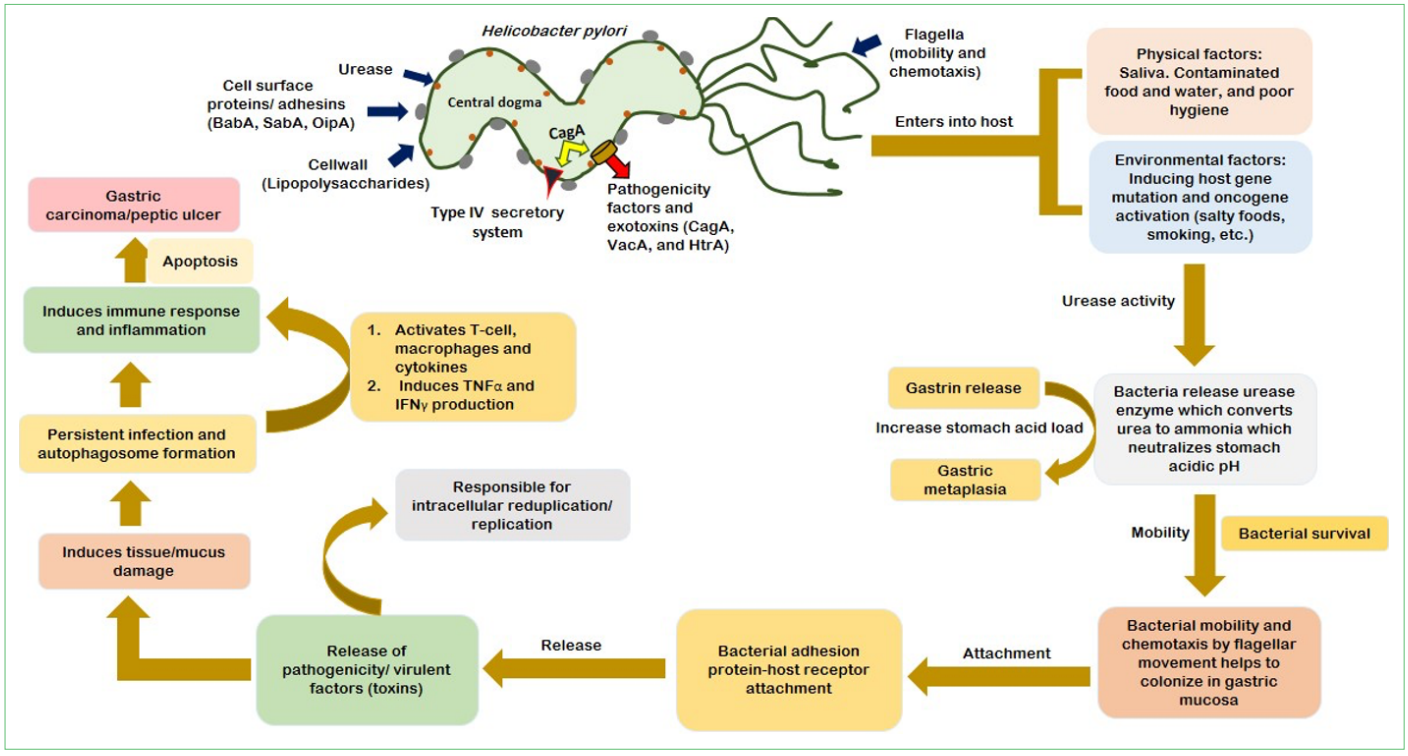

Figure 1. Pathogenesis of *Helicobacter pylori*.

respectively. The phylogeny of the nucleotide sequence alignment was constructed using the Maximum Likelihood Method and Tamura-Nei Model[19] with 100 bootstrap levels in MEGA X[20] (Fig. 2).

Results

In the present study, 100 sequences of *H. pylori* were retrieved for TRs analysis and 100 sequences of 16S rRNA of *H. pylori* were retrieved from the NCBI database to avoid sample bias across the globe. The unavailable data were represented as NA (Not Applicable) in both Table 1 and Table 2. Out of 100 similar strains obtained from BLAST analysis, only 45 strains could be identified with possible TRs. The downloaded FASTA formatted sequences of 45 strains were aligned individually in the MEGA X software for analyzing the repetition of the above two sequence repeats (Table 1).

Tandem Repeats Phylogenetic Tree Interpretation

The limited availability of the *H. pylori* sequences at NCBI lead us to the homology search of the master sequence to hunt for the "GATTAG" and "TGATTAGT" TRs using the BLASTn search to retrieve about the 100 sequences. Most of the sequences obtained were the complete nucleotide sequences. After the sequences were individually analyzed for the desired TRs with MEGA X software, a total of only 45 sequences were found to carry such repetitive TR sequences. Out of 45 sequences, 7 sequences were from Mexico City, 7 sequences from Australia, 4 sequences from Germany, 4 sequences from Peru, 2 sequences from West Africa, 1 sequence each from Texas, United Kingdom, Japan, Malaysia, and France, whereas 16 sequence locations were not available in Table 1. Considering the C.S., both the TRs were found in each strain except 5 strains (P12, PUNO-006, LIM-008, PUNO-004, SJM180) in which TGATTAGT TR was not present. On the basis of N.C.S, only 2 strains (UM037, 5-A-EK1) were identified to have both the TRs.

The two different consecutive TRs with the same TRNs had been identified in various samples such as 25 TRN (8 samples), 16 TRN (3 samples), 2 TRN (3 samples), 17 TRN (1 sample), 15 TRN (1 sample), 24 TRN (1 sample), 4 TRN (1 sample), 3 TRN (1 sample) whereas the remaining samples had both consecutive and non-consecutive TRNs. The two different TRs obtained from individual strain sequences were displayed with varied TRNs. This difference in the TRNs was responsible for increasing the intra-species mutation frequency. Few *H. pylori* strains (P12, PUNO-006, LIM-008, PUNO-004, SJM180) were not been identified with any of the TGATTAGT TRs. Moreover, the ratio of the "GATTAG" repeats with both consecutive and non-consecutive TRNs were higher compared to "TGATTAGT". Although, the TRs obtained from different samples had been identified with a common host (*Homo sapiens*), the strain HP_TH2099 (House macaque gastric juice) isolated in Japan that employed *Macaca fuscata* as its host was an exception. The strains of *H. pylori* were classified into four different classes (Class I-IV) based on the TRNs identified; TRNs ≤ 5 (Class I), TRNs > 5 but ≤ 10 (Class II), TRNs > 10 but ≤ 15 (Class III), and TRNs > 15 (Class IV); similar kind of classification of TRN was given by the Ghosh et al.[21] in *Candidatus* liberibacter asiaticus.

The selected 45 strains with "GATTAG" and "TGATTAGT" consecutive TRNs were used for the calculation of class percent-

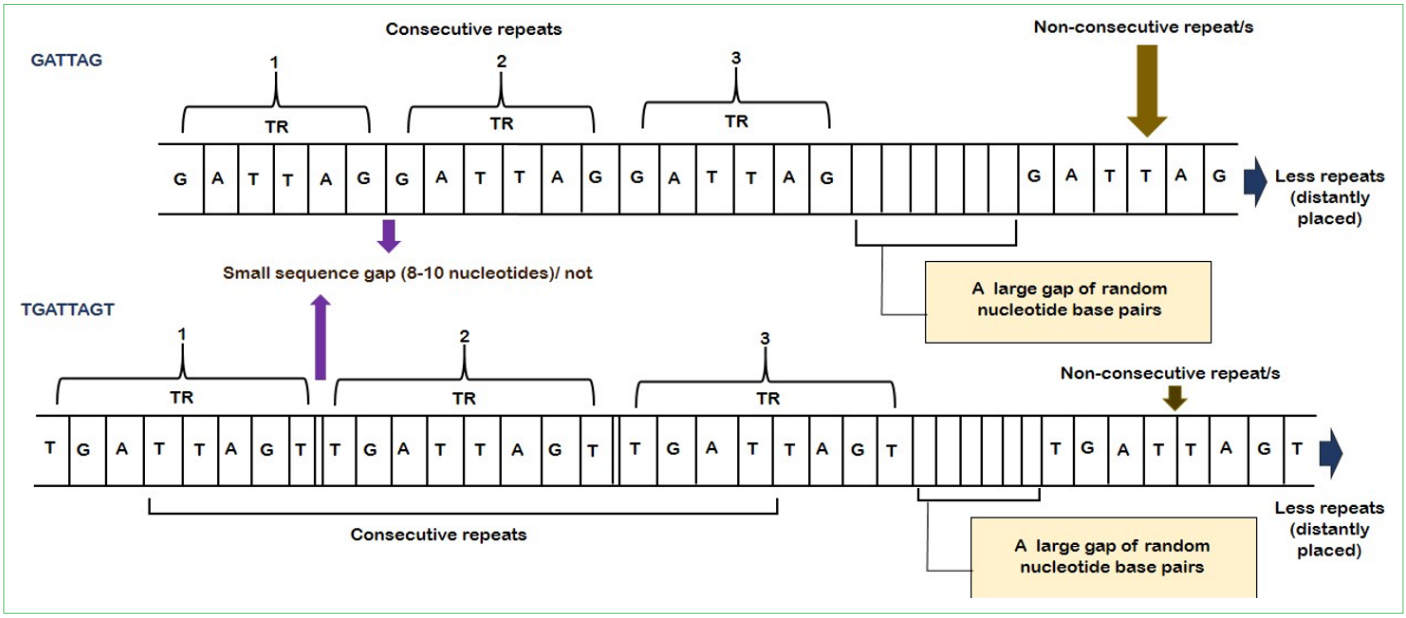

Figure 2. Occurrence of the consecutive and non-consecutive TRs found in retrieved sequences of *Helicobacter pylori*.

Table 1. Tandem Repeat Numbers (TRNs) of different strain isolates of Helicobacter pylori across the globe with their sample code, location, strain, host, source, classification, and accession numbers.

Sl. No.	Sample Code	Location	Strain	Host	Source	TRN (GATTAG)		Classification (C.S.)	TRN (TGATTAGT)		Classification (C.S.)	Accession No.
						C.S.	N.C.S.		C.S.	N.C.S.		
1.	SAMN03268365	USA: Texas	26695-1MET	*Homo sapiens*	Human stomach biopsy	25	-	Class IV	25	-	Class IV	CP010436
2.	SAMD00061016	NA	26695-1CL	NA	NA	25	-	Class IV	25	-	Class IV	AP013356
3.	SAMD00061017	NA	26695-1CH	NA	NA	25	-	Class IV	25	-	Class IV	AP013355
4.	SAMD00061015	NA	26695-1	NA	NA	25	-	Class IV	25	-	Class IV	AP013354
5.	SAMN02603763	NA	Rif2	NA	NA	25	-	Class IV	25	-	Class IV	CP003906
6.	SAMN02603762	NA	Rif1	NA	NA	25	-	Class IV	25	-	Class IV	CP003905
7.	SAMN02603761	NA	26695	NA	NA	25	-	Class IV	25	-	Class IV	CP003904
8.	SAMN02603995	United Kingdom	26695	NA	NA	25	-	Class IV	25	-	Class IV	AE000511
9.	SAMN08388637	NA	dRdM2addM2	*Homo sapiens*	NA	17	-	Class IV	17	-	Class IV	CP026515
10.	SAMN08388633	NA	26695-dR	*Homo sapiens*	Human gut	16	-	Class IV	16	-	Class IV	CP026326
11.	SAMN08388636	NA	26695-dRdM1dM2	*Homo sapiens*	Human gut	16	-	Class IV	16	-	Class IV	CP026323
12.	SAMN08388634	NA	dRdM1	*Homo sapiens*	Human gut	16	-	Class IV	16	-	Class IV	CP026325
13.	SAMN08388635	NA	26695-dRdM2	*Homo sapiens*	Human gut	15	-	Class III	15	-	Class III	CP026324
14.	SAMN02376624	Australia: Nedlands	BM012S	*Homo sapiens*	NA	24	2	Class IV	24	-	Class IV	CP006889
15.	SAMN02376623	Australia: Nedlands	BM012A	*Homo sapiens*	NA	23	2	Class IV	23	-	Class IV	CP006888
16.	SAMN02736818	Australia: Perth	BM012B	*Homo sapiens*	Human stomach	15	2	Class III	15	-	Class III	CP007605
17.	SAMN08328770	Japan	Hp_TH2099	*Macacafuscata*	Housed macaque gastric juice	24	-	Class IV	24	-	Class IV	CP025748
18.	SAMN02604194	Malaysia: KL	UM037	*Homo sapiens*	Biospy sample from patient	8	2+2	Class II	8	2	Class II	CP005492
19.	SAMD00051588	NA	ATCC 43504	NA	NA	3	2	Class I	3	-	Class I	AP017632
20.	SAMEA3178013	Australia	NCTC 11637	*Homo sapiens*	Gastric Antrum	3	2	Class I	3	-	Class I	LS483488
21.	SAMN06173311	Australia: Perth	FDAARGOS_298	*Homo sapiens*	Human gastric antrum from Royal Perth Hospital	3	2	Class I	3	-	Class I	CP028325
22.	SAMEA3138296	France	B38	NA	NA	9	-	Class II	5	2	Class I	FM991728
23.	SAMN08055036	NA	J182	NA	NA	4	-	Class I	4	-	Class I	CP024947
24.	SAMN10053382	Germany: Magdeburg	5-A-EK1	*Homo sapiens*	Gastric antrum	3	1	Class I	3	1	Class I	CP032913
25.	SAMN10053562	Germany: Magdeburg	24-A-EK1	*Homo sapiens*	Gastric antrum	5	1	Class I	5	-	Class I	CP032907
26.	SAMN09935063	Mexico: Mexico City	C-Mx-2010-5	*Homo sapiens*	NA	2	2	Class I	2	-	Class I	CP032022
27.	SAMN10053703	Germany: Magdeburg	169-A-EK5	*Homo sapiens*	Gastric antrum	2	-	Class I	2	-	Class I	CP032904
28.	SAMN10053716	Germany: Magdeburg	169-C-EK8	*Homo sapiens*	Gastric corpus Patient169	2	-	Class I	2	-	Class I	CP032477
29.	SAMN08055037	NA	B147	NA	NA	3	-	Class I	3	-	Class I	CP024946
30.	SAMN09935045	Mexico: Mexico City	G-Mx-2003-250	*Homo sapiens*	NA	5	2	Class I	5	-	Class I	CP032048
31.	SAMN09935059	Mexico: Mexico City	C-Mx-2010-8	*Homo sapiens*	NA	7	2	Class II	7	-	Class II	CP032027
32.	SAMN09935058	Mexico: Mexico City	C-Mx-2008-31	*Homo sapiens*	NA	2	2	Class I	2	-	Class I	CP032031
33.	SAMN02603595	NA	P12	NA	NA	2	-	Class I	-	-	-	CP001217
34.	SAMN03331743	Australia	SS1	*Homo sapiens*	Gastric	2	1	Class I	1	-	Class I	CP009259
35.	SAMN04362855	Australia: Sydney	PMSS1	*Homo sapiens*	Gastric tissue biopsy	2	1	Class I	1	-	Class I	CP018823
36.	SAMN03144734	Mexico	29CaP	*Homo sapiens*	Gastric biopsy	2	-	Class I	2	-	Class I	CP012907
37.	SAMN09935056	Mexico: Mexico City	C-Mx-2011-145	*Homo sapiens*	NA	2	2	Class I	2	-	Class I	CP032034
38.	SAMN02603713	West Africa	2018	*Homo sapiens*	NA	2	2	Class I	2	-	Class I	CP002572
39.	SAMN14593086	Peru	PUNO-006	*Homo sapiens*	String method	2	-	Class I	-	-	-	CP051498
40.	SAMN02603712	West Africa	2017	NA	NA	2	2	Class I	2	-	Class I	CP002571
41.	SAMN14593098	Peru	LIM-008	*Homo sapiens*	String method	2	-	Class I	-	-	-	CP051535
42.	SAMN09935054	Mexico: Mexico City	G-Mx-2011-147	*Homo sapiens*	NA	2	2	Class I	2	-	Class I	CP032037
43.	SAMN14593085	Peru	PUNO-004	*Homo sapiens*	String method	2	-	Class I	-	-	-	CP051499
44.	SAMN02603202	Peru: San Juan Miraflores shantytown, Lima	SJM180	*Homo sapiens*	Patient with gastritis	2	-	Class I	-	-	-	CP002073
45.	SAMN02604189	NA	908	NA	NA	2	2	Class I	2	-	Class I	CP002184

NAnot applicable; C.S.consecutive; N.C.S.non-consecutive.

Table 2. Collected gene sequences of Helicobacter pylori 16S rRNA for evolutionary analysis including location, strain, host, source, and accession numbers

Sl. No.	Location	Strain	Host	Source	Accession No.
1.	Brazil	LPB 36-03	*Homo sapiens*	Intestinal mucosa of patient with Chron's disease	AY593991
2.	Brazil	LPB 581-99	*Homo sapiens*	Intestinal mucosa of patient with Chron's disease	AY593986
3.	Brazil	LPB 638-99	*Homo sapiens*	Intestinal mucosa of patient with Chron's disease	AY593987
4.	Brazil	LPB 424-01	*Homo sapiens*	Intestinal mucosa of patient with Chron's disease	AY593989
5.	Venezuela	WG56	*Homo sapiens*	Stomach	HM046431
6.	Minnesota	MC238	*Homo sapiens*	NA	U01329
7.	Brazil	LPB-64B	NA	Human bile	AY304571
8.	Brazil	LPB-3B	NA	Human bile	AY304570
9.	Brazil	LPB-1B	NA	Human bile	AY304569
10.	Brazil	LPB-5V	NA	Human gallbladder	AY304551
11.	NA	LPB 582-99	*Homo sapiens*	Intestinal mucosa of patient with ulcerative colitis	AY364437
12.	NA	LPB 427-01	*Homo sapiens*	Intestinal mucosa of patient without inflammatory bowel disease	AY364439
13.	NA	USU-101	NA	NA	EU544199
14.	NA	LPB 10-02	*Homo sapiens*	Intestinal mucosa of patient with ulcerative colitis	AY364440
15.	NA	LPB 473-00	*Homo sapiens*	Intestinal mucosa of patient with ulcerative colitis	AY364438
16.	NA	NA	*Homo sapiens*	Isolated from human liver	AF361935
17.	Maryland	22694 33b	NA	NA	AY505026
18.	Maryland	20200 3b	NA	NA	AY505028
19.	Maryland	12954 6b	NA	NA	AY505031
20.	Maryland	7546 57b	NA	NA	AY505040
21.	Maryland	MS-CO49 49a	NA	NA	AY505034
22.	Maryland	62815 4a	NA	NA	AY505029
23.	Maryland	95E 34a	NA	NA	AY505027
24.	Maryland	52 30b	NA	NA	AY505025
25.	Maryland	MS61 9na	NA	NA	AY505033
26.	Maryland	11219 62a	NA	NA	AY505041
27.	Maryland	12954 5b	NA	NA	AY505030
28.	Maryland	52815 8nc	NA	NA	AY505032
29.	Maryland	Tx30a 53a	NA	NA	AY505037
30.	Maryland	MS-PA18 18f	NA	NA	AY505042

NA: not applicable.

age for each. For "GATTAG" Class I showed 55.55% presence while Class II showed 6.66%. Nevertheless, Class III and Class IV results were 4.44% and 33.33%, respectively. Similarly, for "TGATTAGT" the presence of Class I and Class II were found to be 46.66% and 4.44%, respectively. Moreover, Class III and Class IV class percentages of "TGATTAGT" were the same as for Class III and IV of GATTAG. In contrast to "GATTAG", consecutive sequences of "TGATTAGT" TRs were missing in 5 strains (P12, PUNO-006, LIM-008, PUNO-004, SJM180). Therefore, those 5 classes were considered null (0) upon 45 strains while calculating the total class percentage in each such case. Class I of "GATTAG" and "TGATTAGT" were predominant in both the TR cases. However, in the case of "GATTAG" TRs, Class III had been identified to be less predominant whereas in "TGATTAGT" both Class II and Class III were found to be less predominant concerning their other classes.

The phylogenetic tree inferred based on the 45 sequences (including the master sequence) had 10 clades with the least bootstrap value of 7 and the highest value of 100. Some clades appeared as clusters followed by clades with variability (Fig. 3). The class I TRN (of two TRs) is found to be predominant over all other classes (Fig. 4)

16S rRNA Phylogenetic Tree Interpretation

While the NCBI database had more than 450 sequences of 16S rRNA, we separately collected the 16S rRNA sequences of *H. pylori* from the database which were the partial sequences. This was because on taking a master sequence and going for a BLAST search, out of 100 sequences only 7-15 gene sequences of 16S rRNA were found while the rest displayed only the complete genome of H. pylori. On random retrieval of 100 sequences, we constructed a phylo-

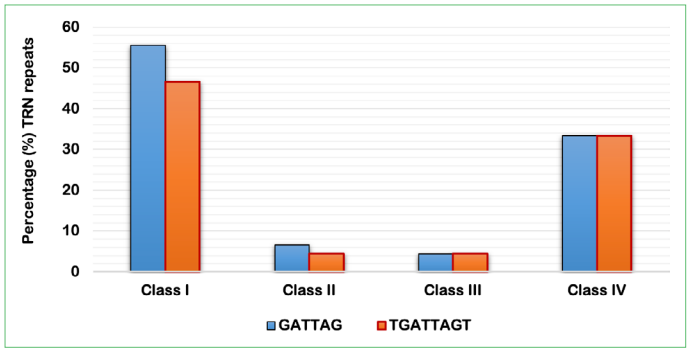

Figure 3. Percentage class distribution of *Helicobacter pylori* strains TRN of two TRs (GATTAG and TGATTAGT) retrieved from NCBI deposited data worldwide.

genetic tree which showed that among 30 sequences analyzed, 14 sequences were from Maryland, 8 sequences were from Brazil, 1 sequence from Venezuela, and 1 sequence from Minnesota while the remaining sequences had no location displayed (Table 2). Sequences with query lengths of 1000-1500 bp were selected and aligned with gaps by CLUSTAL_W using MEGA X software. Before alignment, the sequences were trimmed from both N-terminal and C-terminal ends to obtain an equal length of sequences. Phylogenetic analysis of the trimmed sequences was carried out to obtain the clades. The 16S rRNA sequences of *H. pylori* isolates were divided into 16 clades and were colored to distinguish between each other. The clades in the phylogenetic tree had a bootstrap value as low as 2 and as high as 100. This indicates a total of 16 different strains were present in the collected sample. The analysis displayed the clustered strains of the species as well as the heterogeneity within the species. It helped us to depict the intra-strain evolutionary relationship within the species (Fig. 5).

Discussion

H. pylori is indulged in different mechanisms for its survival in the gastric environment making it one of the most cunning bacteria to harbor in the gastric epithelium. It entails higher incidences of gastritis and peptic ulcers leading to gastric carcinoma causing a several-fold increase in morbidity and mortality rate.[5] The bacterium sets an appropriate example of an intimate relationship with humans.[22] In the present study, two tandem repeats ("GATTAG" and "TGATTAGT") were found in the studied sequence. BLASTn was performed to search for homologous sequences of which 45 sequences carried those tandem repeats when analyzed individually using MEGA X software. Out of the 45 sequences which carried the consecutive repeats, approximately 15 were seen to have repeats of the same sequence but in a non-consecutive manner (distantly placed). These non-consecutive sequence repeats were localized within the genome at a non-specific nucleotide position for strains including BM012A, BM012B, BM012S, ATCC 43504, NCTC 11637, FDAARGOS_298, C-Mx-2010-5, G-Mx-2003-250, C-Mx-2010-8, C-Mx-2008-31, C-Mx-2011-145, 2018, 2017, G-Mx-2011-147 and 908.

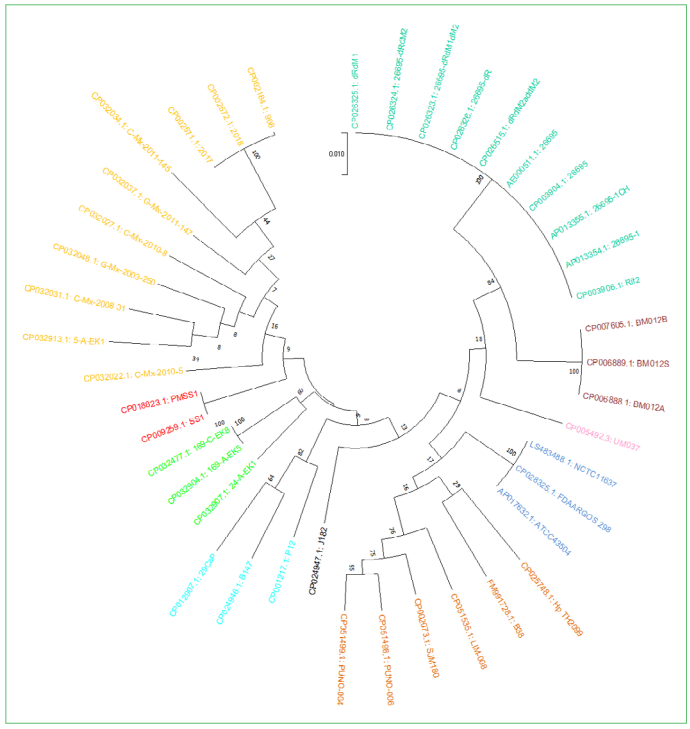

Figure 4. Phylogenetic tree of 45 strain isolates of *Helicobacter pylori* Tandem Repeat sequences using MEGA X software.

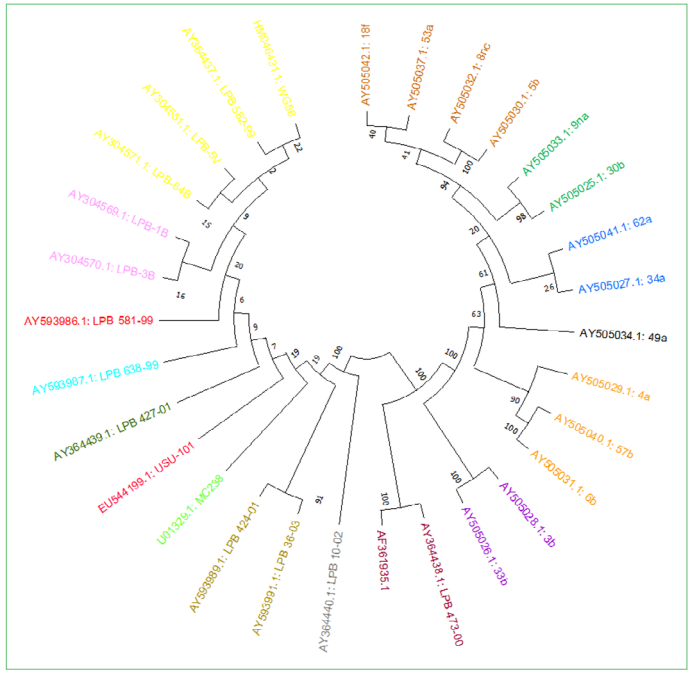

Figure 5. Evolutionary analysis of 30 16S rRNA gene sequences using MEGA X software.

Strains like BM012A, BM012B, BM012S, ATCC 43504, NCTC 11637, FDAARGOS_298, C-Mx-2010-5, G-Mx-2003-250, C-Mx-2010-8, C-Mx-2008-31, C-Mx-2011-145, 2018, 2017, G-Mx-2011-147 and 908 showed 2 TRN of "GATTAG" non-consecutive repeats and strains like 5-A-EK1, 24-A-EK1, PMSS1 and SS1 displayed only 1 TRN of the same sequence. Strain UM037 showed 8 TRN of "GATTAG" consecutive repeats with a total of 4 (2+2) TRN non-consecutive repeats after a certain distance from each other. For the "TGATTAGT" tandem repeat, UM037 and B38 showed 2 TRN non-consecutive repeats while 5-A-EK1 showed only 1 TRN. These non-consecutive repeats could be the reason for transposons. As endogenous transposable elements (insertion sequences or transposons) are predicted to be present in H. pylori, they may be responsible for genetic rearrangements or can act as a substrate for the recombination of a short segment of a homologous nucleotide sequence.[23] The occurrence of TRs is mostly identified in the DNA sequences of the organism which are commonly designated as unstable "junk/non-functional" sequences. Moreover, TRs are a type of error in which several copies of the same short sequence are repeated over and over, indicating that the DNA modifications are not random nucleotides.[24] As a result, only consecutive sequences in different *H. pylori* strains can be assumed to show the presence of TRs in our study. Therefore, in the recent era, the phenotypic modulations of microorganisms due to the presence of TRs have been identified which are responsible for their adaptation and evolution.[25]

Microsatellite tandem repeats and variation in the nucleotide sequences were observed in the genome of *H. pylori* that indulge in contributing to distinct mutations. When compared to point mutation, microsatellites mutation occurs as a result of the replication slippage process, which boosts the mutation chance by 100,000 times.[24, 26] It helps in their natural self-strain improvement making them more virulent and pathogenic. In a study conducted during the acute infection phase of H. pylori, the mutation burst of 2.2 × 10-4 - 8.4 × 10-4 changes per site per year is the excessive mutation estimate for any bacteria to date. When compared to other bacterial species, the substitution rate surpasses by more than two orders of magnitude.[27] The inability of the mismatch repair system is responsible for this rate of mutation.[26] It has been well-known that mismatch repair (MutHLS1) and homologous recombination system are absent in *H. pylori* and that could be attributed to a higher mutation rate ultimately leading to genetic diversity. Though MutS2 has been discovered in H.pylori, it has a negligible role in altering mutation rate but can act as a suppressor of recombination.[22, 28, 29]

Although the evolution of TRs contributes to the adaptation of the bacterium to its host,[26] an increase in mutation rate and several other properties also facilitates the adaptation of the pathogen to the host. These observations may obey by these mutations are the result of horizontal and vertical gene transfer[2] which may be responsible for its increased virulence and pushing it to the verge of becoming a multi-drug resistant bacterium (MDR). The effect on the host is noteworthy as it stimulates the recruitment of immunoinflammatory cells resulting in robust humoral and cellular responses along with the elevated production of proinflammatory cytokines. This in turn triggers peptic carcinoma on account of inflammation of gastric tissues.[5] 16S rRNA gene present in all bacteria makes it a perfect target for the identification of bacteria and phylogeny framework. The random sequence changes occurring in the 16S rRNA highlight the phylogenetic relationship of the organism.[2] An in-silico study was conducted by Puri et al.[3] for identification of the *H. pylori* and its relatives using 16S rRNA gene sequences of various Helicobacter species. In addition, analysis of restriction enzyme, species-specific conserved motifs, and protein sequences of hsp60 was also carried out using *in silico* approach. The result showed that some species of Helicobacter including *H. pylori* were effortlessly segregated while some were found to be heterogeneous. In contrast, our study focused on finding some specific TRs of *H. pylori* that could result in such vast mutations and their capacity to adapt to the host environment. Several consecutive and non-consecutive tandem repeats at varying locations with distinct tandem repeat numbers in the individual nucleotide sequences may be due to the existence of transposons and could be responsible for variations among the strains leading to a higher mutation rate and lowering the therapeutic index.

Moreover, the phylogenetic tree depicts that the presence of jumping genes (transposons) might be responsible for varying numbers of TRs in the nucleotide sequences of H. pylori. Also, the presence/absence of TGATTAGT repeats in a few cases is making the strains differ from each other resulting in distinct variations among the species. Therefore, the understanding and identification of the common ancestor which is varying over time are hindering the study of evolutionary analysis. In-vitro and in-vivo investigations have shown increased antimicrobial resistance due to an increase in point and spontaneous mutations against a variety of antibiotics (quinolones, rifampin, metronidazole, and clarithromycin).[4] Furthermore, recurrent random mutations of *H. pylori* strains, alongside geographical differences, have made therapy and prophylactic challenging in the global fight against the bacteria. Nevertheless, it has been well observed that the genomic studies on *H. pylori* are very scarce due to its high mutation rate invoking the

multidrug resistance characteristics and thereby making the elimination more difficult.

Conclusion

In the present study, we were able to identify variable number of consecutive and non-consecutive tandem repeats across the globe which might be due to the existence of transposons. The presence of tandem repeats in *H. pylori* could be the reason for mutation and development of distinct strains. This mutation might be correlated with antimicrobial resistance development hindering the efficacy of first and second generation antibiotics. Nevertheless, we could depict the intra-strain evolutionary relationship within the species through phylogenetic analysis. Therefore, the data obtained from the tandem repeats in our study could be useful for the development of prophylaxis and therapeutics for eradicating the *H. pylori* mediated infection and gastric cancer.

Disclosures

Acknowledgments: The authors acknowledge the Young Science Leader platform and Dr. Felix Bast for providing us the opportunity to conduct this research work. Authors are thankful to Dr. Nilesh Gawande for critically reviewing the manuscript. The authors are also thankful to the Faculty of Life Sciences, Mandsaur University & Guru Nanak Institute of Pharmaceutical Science and Technology, Kolkata for their motivation and encouragement to work on the most devastating H. pylori.

Ethics Committee Approval: As the study carried out "*in Silico*" methodology no Ethics Committee Approval is required.

Peer-review: Externally peer-reviewed.

Funding: The authors did not receive support from any organization for the submitted work.

Conflict of Interest: None declared.

Authorship Contributions: Concept – A.W.; Design – A.W., A.S.M., R.S., A.D.; Supervision – A.W., A.D.; Materials – R.S., A.S.M.; Data collection &/or processing – A.W., R.S., A.S.M.; Analysis and/or interpretation – R.S., A.S.M., A.W., N.K.; Literature search – A.S.M., R.S.; Writing – A.S.M., R.S., A.W.; Critical review – A.W., A.D.

References

1. Kusters JG, van Vilet AHM, Kuipers EJ. Pathogenesis of Helicobacter pylori infection. Clin Microbiol Rev 2006;19:449–90.
2. Idris AB, Hassan HG, Ali MAS, Eltaher SF, Idris LB, Altayb HN, et al. Molecular phylogenetic analysis of 16S rRNA sequences identified two lineages of Helicobacter pylori strains detected from different regions in Sudan suggestive of differential evolution. Int J Microbiol 2020;2020:8825718. [CrossRef]
3. Puri A, Rai A, Dhanaraj PS, Lal R, Patel DD, Kaicker A, et al. An in silico approach for identification of the pathogenic species, Helicobacter pylori and its relatives. Indian J Microbiol 2016;56:277–86. [CrossRef]
4. 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. [CrossRef]
5. Muzaheed. Helicobacter pylori oncogenicity: mechanism, prevention, and risk factors. The Sci World J 2020;2020:3018326.
6. Diaconu S, Predescu A, Moldoveanu A, Pop CS, Fierbinţeanu-Braticevici C. Helicobacter pylori infection: Old and new. J Med Life 2017;10:112–7.
7. Kao CY, Sheu BS, Wu JJ. Helicobacter pylori infection: An overview of bacterial virulence factors and pathogenesis. Biomed J 2016;39:14–23.
8. Waskito LA, Salama NR, Yamaoka Y. Pathogenesis of Helicobacter pylori infection. Helicobacter 2018;23:e121516.
9. Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, Ottemann KM, et al. The complete genome sequence of Helicobacter pylori strain G27. J Bacteriol 2009;191:447–8.
10. Lamichhane B, Chua EG, Wise MJ, Laming C, Marshall BJ, Tay CY. The complete genome and methylome of Helicobacter pylori hpNEAfrica strain HP14039. Gut Pathogens 2019;11:7.
11. Liu AN, Teng KW, Chew Y, Wang PC, Nguyen TTH, Kao MC. The effects of HP0044 and HP1275 knockout mutations on the structure and function of lipopolysaccharide in Helicobacter pylori strain 26695. Biomedicines 2022;10:145. [CrossRef]
12. Jackson LK, Potter B, Schneider S, Fitzgibbon M, Blair K, Farah H, et al. Helicobacter pylori diversification during chronic infection within a single host generates sub-populations with distinct phenotypes. PLos Pathog 2020;16:e1008686.
13. White B, Winte M, DeSipio J, Phadtare S. Clinical factors implicated in antibiotic resistance in Helicobacter pylori patients. Microorganisms 2022;10:322. [CrossRef]
14. Zhang J, Sun X, Wang J, Zhang F, Li X, Han J. Association of the IL-1RN variable number of tandem repeat polymorphism and Helicobacter pylori infection: A meta-analysis. PLos One 2017;12:e0175052. [CrossRef]
15. Subirana JA, Messeguer X. Tandem repeats in Bacillus: Unique features and taxonomic distribution. Int J Mol Sci 2021;22:5573. [CrossRef]
16. Zhou K, Aertsen A, Michiels CW. The role of variable DNA tandem repeats in bacterial adaptation. FEMS Microbiol Rev 2013;38:119–41.
17. Wang YK, Kuo FC, Liu CJ, Wu MC, Shih HY, Wang SSW, et al. Diagnosis of Helicobacter pylori infection: Current options and development. World J Gastroenterol 2015;21:11221–35.
18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 2018;35:1547–9. [CrossRef]
19. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 1993;10:512–26.

20. Warghane A, Petkar T, Preeyaa S U, Kumari N, Ranjan L. In silico characterization of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-COV-2) based on the spike protein gene. Eurasian J Med Oncol 2021;5:163–80. [CrossRef]
21. Ghosh DK, Bhose S, Motghare M, Warghane A, Mukherjee K, Ghosh DK Sr, et al. Genetic diversity of the Indian populations of 'Candidatus liberibacter asiaticus' based on the tandem repeat variability in a genomic locus. Phytopathology 2015;105:1043–9. [CrossRef]
22. Dorer MS, Sessler TH, Salama NR. Recombination and DNA repair in Helicobacter pylori. Annu Rev Microbiol 2011;65:329–48.
23. Labigne A, Jenks PJ. Mutagenesis. In: Mobley HLT, Mendz GL, Hazell SL, editors. Helicobacter pylori: Physiology and Genetics. Washington DC: ASM Press; 2001.
24. Fondon JW, Garner HR. Molecular origins of rapid and continuous morphological evolution. PNAS 2004;101:18058–63.
25. Gemayel R, Cho J, Boeynaems S, Verstrepen KJ. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. Genes 2012;3:461–80.
26. Flèche PL, Hauck Y, Onteniente L, Prieur A, Denoeud F, Ramisse V, et al. A tandem repeats database for bacterial genomes: Application to the genotyping of Yersinia pestis and Bacillus anthracis. BMC Microbiol 2001;1:2. [CrossRef]
27. Linz B, Windsor HM, McGraw JJ, Hansen LM, Gajewski JP, Tomsho LP, et al. A mutation burst during the acute phase of Helicobacter pylori infection in humans and rhesus macaques. Nat Commun 2014;5:4165.
28. Pinto AV, Mathieu A, Marsin S, Veaute X, Ielpi L, Labigne A, et al. Suppression of homologous and homeologous recombination by the bacterial MutS2 protein. Mol Cell 2005;17:113–20. [CrossRef]
29. Wang G, Alamuri P, Humayun MZ, Taylor DE, Maier RJ. The Helicobacter pylori MutS protein confers protection from oxidative DNA damage. Mol Microbiol 2005;58:166–76.