Characterization and genetic diversity of *Helicobacter pylori* type IV secretion system components CagI and CagN and their association with clinical outcomes among Iranian patients

Yasaman Azizimoghaddam
Shaheed Beheshti University of Medical Sciences Research Institute for Gastroenterology and Liver Diseases

Sadaf Kermanpour
Shaheed Beheshti University of Medical Sciences Research Institute for Gastroenterology and Liver Diseases

Nasrin Mirzaei
Shaheed Beheshti University of Medical Sciences Research Institute for Gastroenterology and Liver Diseases

Hamidreza Houri
Shaheed Beheshti University of Medical Sciences Research Institute for Gastroenterology and Liver Diseases

Abbas Yadegar (a.yadegar@sbmu.ac.ir)
Shaheed Beheshti University of Medical Sciences https://orcid.org/0000-0002-2135-7581

Navin Kumar
Gautam Buddha University

Hamid Asadzadeh Aghdaei
Shaheed Beheshti University of Medical Sciences Research Institute for Gastroenterology and Liver Diseases

Mohammad Reza Zali
Shaheed Beheshti University of Medical Sciences Research Institute for Gastroenterology and Liver Diseases

Research

Keywords: Helicobacter pylori, CagI, CagN, cagPAI, genetic diversity, virulence genotype

Posted Date: November 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-64096/v2

Abstract

Background: A number of *cag*PAI genes in *H. pylori* genome was proposed to be the most probably evolved under a diversifying selection and evolutionary pressure. Among them, CagI and CagN are described as a part of the two different-operon of *cag*PAI that are involved in the T4SS, but the definite association of these factors with clinical manifestations is unclear.

Methods: A total of 70 *H. pylori* isolates were obtained from different gastroduodenal patients. All isolates were examined for the presence of primary *H. pylori* virulence genes by PCR analysis. Direct DNA sequence analysis was performed for the *cagI* and *cagN* genes. The results were compared with reference strain.

Results: The *cagI*, *cagN*, *cagA*, *cagL*, *vacA* s1m1, *vacA* s1m2, *vacA* s2m2, *babA2*, *sabA* and *dupA* genotypes were detected in 80%, 91.4%, 84%, 91.4%, 32.8%, 42.8%, 24.4%, 97.1%, 84.3%, and 84.3% of the total isolates, respectively. The most variable codon usage in *cagI* was observed at residues 20 to 25, 55 to 60, 94, 181 to 199, 213 to 221, 241 to 268, and 319 to 320, while the most variable codon usage in CagN hypervariable motif (CagNHM) was observed at residues 53 to 63. Sequencing data analysis of *cagN* revealed a hypothetical hexapeptide motif (EAKDEN/K) in residues of 278-283 among six *H. pylori* isolates, which needs further studies to evaluate its putative function.

Conclusion: The present study demonstrated a high prevalence of *cagI* and *cagN* genes among Iranian *H. pylori* isolates with gastroduodenal diseases. Furthermore, no significant correlation between *cagI* and *cagN* variants and clinical outcomes was observed in present study. However, all patients had high prevalence of *cag*PAI genes including *cagI*, *cagN*, *cagA* and *cagL* that indicates more potential role of these genes in disease outcome.

Introduction

Helicobacter pylori (*H. pylori*) is a Gram-negative, microaerophilic bacterium that can be chronically colonized in the human stomach. This organism infects more than 50% of the world's population, and is the main cause of chronic active gastritis, gastric and duodenal ulcers, mucosa-associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma [1, 2]. *H. pylori* infection is recognized as the major risk factor for the development of gastric cancer, which is the fifth most common malignancy and the third leading cause of cancer-associated morbidity worldwide [3]. The severity of *H. pylori*-induced gastric diseases seems to be associated with several parameters, including host genetic polymorphism, inflammatory responses, environmental factors, and bacterial virulence genotype [4, 5].

H. pylori is associated with high genetic variability including virulence genes due to genetic plasticity, rearrangement of DNA and, high transformation and recombination frequency. Thereby, *H. pylori* infected patients greatly varies in the disease progression and clinical outcomes geographically. To date, several virulence factors have been identified in the genome of *H. pylori* such as CagA, VacA, BabA, SabA, and DupA [5, 6]. CagA, an oncoprotein, is the best studied virulence-associated factor of *H. pylori* that is

translocated into the host gastric epithelial cells via the type 4 secretion system (T4SS). The *H. pylori* T4SS machinery contains a cluster of gene products which harbor approximately 40 kb chromosomal region named *cag* Pathogenicity Island (*cag*PAI) [7, 8]. *cag*PAI encodes about 27-31 genes, by which a subset of these genes encodes the main components of the T4SS apparatus spanning bacterial membranes. Moreover, possibly 15 to 16 different proteins of the T4SS are required for translocation of CagA and peptidoglycan fragments into the host cells, and also secretion of IL-8 from gastric epithelial cells [9]. Once CagA is translocated then it modulates the host cell signaling which results the loss of membrane polarity, cell elongation, induction of inflammatory cytokines and development of gastric adenocarcinoma [10]. *cag*PAI encodes several unique Cag components that have no sequence similarities to any other bacterial proteins involved in T4SS. However, a number of *cag*PAI genes such as *cagI* and *cagN* were proposed to be most probably evolved under a diversifying selection and evolutionary pressure [11]. CagI, a small protein (41.5 kDa) encoded by *cagI* (*cag19*/*hp0540*) gene, does not share any sequence and topological homology to any other known proteins [12, 13], whereas CagN, a 32-35 kDa protein also termed as Cag17/HP0538 encoded by *cagN* gene (*hp0538*), is a poorly characterized component of the T4SS that appears to be localized to the bacterial inner membrane rather than the periplasm [9, 12, 14, 15].

There are some conflicting reports about the role of CagI and CagN in CagA translocation, IL-8 induction from gastric epithelial cells, and *H. pylori* T4SS machinery [14, 16-20]. Reviewing subsequent and more recent literatures have revealed that CagI is capable of binding to β1 integrins of the host cell and is essential for CagA translocation, and is also involved in pilus biogenesis of T4SS [21, 22]. On the other hand, deletion of *cagN* can reduce the phosphorylation degree of CagA into host cell and it is not considered as a substrate for the T4SS [14]. However, the putative role of CagI and CagN in translocation of CagA and *H. pylori* pathogenesis has not precisely been clarified. The oncogenic potential of *H. pylori* strains is associated with their virulence capacity, genetic diversity and specific sequence polymorphisms within the key genes involving in translocation and phosphorylation of T4SS effectors [23-26]. Therefore, the present study aimed to determine the prevalence of *cagI* and *cagN* genes and their amino acid sequence polymorphisms in Iranian *H. pylori*-infected patients with various gastroduodenal diseases. The probable association between the genetic variants of *cagI* and *cagN* and other virulence genotypes of *H. pylori* with clinical consequences were also investigated.

Methods

H. pylori clinical isolates and biopsy specimens

Gastric biopsy specimens were obtained from 70 patients who underwent upper gastroduodenal endoscopy at Research Institute for Gastroenterology and Liver Diseases in Tehran between January 2017 and May 2019. Three antral biopsies were taken from each patient and examined for culture and histopathology. The biopsy specimens were immediately placed in transport medium containing Thioglycolate supplemented with 3% yeast extract (Oxoid Ltd., Basingstoke, UK) and 1.3 g/L agar (Merck, Germany). All patients provided written informed consent. The study was approved by the Institutional

Ethical Review Committee of Research Institute for Gastroenterology and Liver Diseases at Shahid Beheshti University of Medical Sciences (Project No. IR.SBMU.RIGLD.REC.1398.023).

H. pylori culture and identification

Biopsy specimens were carefully homogenized and inoculated onto the surface of Brucella agar plates (Merck, Germany) supplemented with 7% (v/v) horse blood, 10% fetal calf serum (FCS), Campylobacter-selective supplement (vancomycin 2.0 mg, polymyxin 0.05 mg, trimethoprim 1.0 mg), and amphotericin B (2.5 mg/l). The incubation was performed at 37°C for 3-7 days under a microaerophilic atmosphere (5% O_2, 10% CO_2 and 85% N_2) in a CO_2 incubator (Innova® CO-170; New Brunswick Scientific, USA). The suspected colonies were identified as *H. pylori* based on colony morphology, Gram staining, positive reaction for oxidase, catalase and urease tests, and also by *H. pylori* gene-specific PCR following the previously described protocols [27, 28]. Pure cultures from confirmed isolates were kept in 0.5 ml of Brain heart infusion (BHI) medium (Merck, Germany) containing 15% glycerol plus 20% FCS, and stored at -80°C until further analysis.

Genomic DNA extraction

Genomic DNA was extracted from freshly harvested colonies on agar plates, using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The quality of DNA was checked by using NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, USA). The extracted DNA samples were stored at -20°C until PCR assay.

Genotyping of *H. pylori* virulence-associated genes

PCR analysis was performed to detect virulence target genes including *cagL*, *cagA*, *vacA* alleles (s1/s2 and m1/m2), *babA2*, *sabA* and *dupA* genes using specific primers (Table S1). Briefly, PCR mixtures in a volume of 25 µl consisted of 2 µl of template DNA (approximately 200 ng), 0.1 mM of each primer, 2.5 µl of a 10-fold concentrate PCR buffer, 100 mM of deoxynucleotide triphosphates, 1 mM $MgCl_2$, and 1.5 U of Super-Taq™ DNA polymerase (HT Biotechnology Ltd., Cambridge, UK). PCR amplifications were performed in a thermocycler (Eppendorf, Hamburg, Germany) under the following conditions: initial denaturation at 94°C for 4 min, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at the indicated temperature for each reaction in Table S1 for 45 s, extension at 72°C for 1 min. A final extension step was performed at 72°C for 10 min to ensure full extension of the PCR products. PCR amplicons were electrophoresed on a 1.2% TBE agarose gel, stained with ethidium bromide, and examined under a UV transilluminator. *H. pylori* J99 (CCUG 47164) and a no-template mixture served as positive and negative controls in each PCR experiment, respectively.

Primer designation for *cagI* and *cagN* genotyping

The NCBI GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) and the DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp/) were searched for all available complete and partial *cagI* and *cagN*

sequences of *H. pylori* strains. Based on pairwise and multiple nucleotide sequence alignments of *cagI* and *cagN* genes from different *H. pylori* strains and using the complete relevant sequence of *H. pylori* P12 (CP001217.1) as the reference strain, two pairs of specific primers were designed from the conserved regions for detection of complete related sequences using CLC Sequence Viewer 8 software (https://www.qiagenbioinformatics.com/). The selected primer target sites were compared to all available complete and partial *cagI* and *cagN* sequences of *H. pylori* strains with the Basic Local Alignment Search Tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Analysis of *cagI* and *cagN* diversity by PCR sequencing

For DNA sequencing of *cagI* and *cagN*, PCR amplification was carried out in a final volume of 25 μl using designed specific primers including 5′-CATTTGACTTACCTTGATTAC-3′ (*cagI*-F) and 5′-TTTGAGCACTTGTTGGTTGG-3′ (*cagI*-R), 5′-GAGCGACAAAACAACTATGC-3′ (*cagN*-F) and 5′-GATCCCTAGAACAAAGTAAGC-3′ (*cagN*-R) yielding DNA fragments of about 1377 and 1192 bp in length, respectively. The PCR products were purified using the Silica Bead DNA Gel Extraction Kit (Thermo Scientific, Fermentas, USA) followed by sequencing on both strands using an automated sequencer (Macrogen, Seoul, Korea). DNA sequences were edited by Chromas Lite version 2.5.1 (Technelysium Pty Ltd, Australia) and BioEdit version 7.2.5 [29]. The *cagI* and *cagN* nucleotide and amino acid sequences were aligned to *H. pylori* strain P12 as a reference strain (GenBank: CP001217.1). The single nucleotide variations and codon usage of the sequences were examined using BioEdit version 7.2.5.

Phylogenetic analysis

Phylogenetic trees were generated for CagI and CagN nucleotide and amino acid sequences using Molecular Evolutionary Genetics Analysis version 7.0 (MEGA7) [30]. Evolutionary history was inferred by the Maximum Likelihood trees using Tamura 3-parameter model and Poisson correction method for nucleotide and amino acid sequences, respectively.

Nucleotide sequence accession numbers

The complete and partial nucleotide sequences of *cagI* and *cagN* genes from *H. pylori* strains determined in this study were deposited in the NCBI GenBank database under the accession numbers MG573078-MG573107 (*cagI*) and MG559675-MG559720 (*cagN*).

Statistical analysis

The statistical associations between *H. pylori* virulence genotypes and different clinical status were determined by the Chi-square and Fisher's exact tests. A two-sided *P* value of less than 0.05 was regarded statistically significant. The IBM SPSS Statistics for Windows version 21.0 (Armonk, NY: IBM Corp.) was used for all statistical analyses.

Results

Demographic and clinical characteristics of patients

The median age of the patients was 45.6 years (ranging from 14 to 75 years). Of the study cohort, 32.8% (*n* = 23) was male and 67.1% (*n* = 47) was female. According to the endoscopic and histopathology findings, 39 (55.7%) patients were diagnosed with were non-ulcer dyspepsia (NUD), 23 (32.8%) patients had peptic ulcer disease (PUD), 7 (10%) patients had intestinal metaplasia (IM), and one (1.4%) had gastric cancer. Three patients (4.3%) suffered from gastritis and duodenitis simultaneously. Table S2 indicates the demographic characteristics and clinical status of the included subjects. In each of the 70 cases, *H. pylori* was isolated by culture and the isolates were approved by detection of the *glmM* and 16s rRNA genes.

Virulence genotypes and variants

The molecular analysis revealed that the *cagA*, *cagI*, *cagN*, *cagL*, *vacA* s1m1, *vacA* s1m2, and *vacA* s2m2 positive strains had a prevalence of, respectively, 84% (*n*=59), 80% (*n*=56), 91.4% (*n*=64), 91.4% (*n*=64), 32.8% (*n*=23), 42.8% (*n*=30), and 24.4% (*n*=17) in our study, whereas *babA2*, *dupA*, and *sabA* were detected in, respectively, 97.1% (*n*=68), 84.3% (*n*=59), and 84.3% (*n*=59) of the isolates included in this investigation (Table 1). There was no statistically significant association between the *H. pylori* virulence genotypes and clinical status of the patients (P >0.05). In the present study, 100% (23/70) of the PUD and 94.9% (37/70) of the NUD strains were positive for *babA2* gene by PCR. Furthermore, the prevalence of *cagN* and *cagL* genes for PUD strains is attributed to 95.6% (22/70) and 91.3% (21/70), respectively. In the meantime, patients suffering from NUD showed the frequency of 89.7% (35/70) and 94.9% (37/70) for the same genes as PUD. When it comes to *vacA* allelic combinations, *vacA* s1m2 was found to be the most common allele among the strains recovered from the PUD patients (52.2%), whereas 42.8% and 33.3% of allelic combinations were assigned to *vacA* s1m1 and *vacA* s2m2, within the IM and NUD strains, respectively.

cagI variants in patients with different clinical status

Out of 56 *cagI*-positive *H. pylori* strains, the *cagI* gene of 30 strains were randomly selected and sequenced. The full-length *cagI* gene was successfully sequenced in 27 *H. pylori* strains. Moreover, the *cagI* gene was partially sequenced in three strains due to poor quality of sequence data or sequencing errors. According to our sequencing data, there was no insertion or deletion in the full-length *cagI* fragment from 27 *H. pylori* studied, and sequence alignments were therefore straight forward. In addition, we performed in-frame translation for *cagI* gene into amino acid sequences, and investigated rates and locations of CagI variants. The distribution of amino acid polymorphisms in CagI of *H. pylori* strains are represented in Figure S1 and Table 2. The most variable codon usage was observed at residues G20 to I25, Q55 to E60, G94, M181 to A199, K213 to T221, and Q241 to A268. As we expected, the SKVIVK hexapeptide motif (376-381) located at the C-terminal of CagI was completely conserved among the *cagI* sequenced *H. pylori* strains.

cagN variants in patients with different clinical status

Regarding *cagN* sequence analysis, 46 strains were randomly sent for direct DNA sequencing from 64 *cagN*-positive *H. pylori* strains. The complete *cagN* gene was successfully sequenced in 43 *H. pylori* strains. Furthermore, the *cagN* gene fragments of three strains were partially sequenced as the same reasons for the *cagI* gene. The *cagN* sequencing findings showed a high level of variability in CagN nucleotide and protein sequences. The most variable codon usage was observed at residues 53 to 63, so-called as CagN hypervariable motif (CagNHM). Moreover, a hypothetical hexapeptide (EAKDEN/K) was inserted in residues 278-283 among six *H. pylori* strains. Interestingly, this motif was observed two times in a row in one of these clinical strains (EAKDENEAKDEN). The other insertion sequences were detected between residues 224-225 and 234-235 for KV and KN amino acids in one of the strains. The sequencing data analysis revealed that these insertion sequences in *cagN* gene caused no frameshift mutations as compared to the P12 reference strain. Figure S2 and Table 3 showed the distribution of amino acid polymorphisms of CagN among 43 *H. pylori* strains in this study.

Phylogenetic analysis of *H. pylori* CagI and CagN

The phylogenetic trees of *cagI* nucleotide and amino acid sequences from *H. pylori* isolates are illustrated in Figure 1 and Figure 2, respectively. Generally, no characteristic clusters were observed between DNA and amino acid sequences of CagI and different clinical status. Furthermore, on the basis of the CagN nucleotide and amino acid sequences, a phylogenetic tree was reconstructed by using the Maximum Likelihood method, which are illustrated in Figure 3 and Figure 4, respectively. Similar to CagI sequences, the CagN phylogenetic analysis indicated no characteristic clusters with regard to the clinical status.

Discussion

Virulent *H. pylori* strains harbor the *cag*PAI (*cag*$^+$) encoding a type IV secretion apparatus, which has been shown to inject CagA and possibly also other virulence effectors into infected gastric epithelial cells [31]. It has been well documented that *cag*$^+$ *H. pylori* strains augment the risk for severe gastritis, peptic ulceration, atrophic gastritis, dysplasia, and gastric adenocarcinoma compared to strains that lack the *cag*PAI (*cag*$^-$) [32-34]. Previously, it has been described that CagI forms a functional protein complex at the bacterial cell surface by interacting with CagL, which is another important Cag secretion apparatus component. Accordingly, some evidence suggested that CagI can interact with CagL protein and let to bind to integrin receptors on the target cell surface [8, 17]. CagI and CagL proteins contain N-terminal signal peptide, and therefore they can be supposed to be transported to the periplasm, however, the two proteins are not distributed equally on the bacterial cell surface [35]. Regarding different views on CagI, Kumar et al. [36] found that CagI does not participate in CagA translocation from cytoplasm to bacterial cell surface. On the other hand, it has been discovered that mutation in *cagN* did not interrupt CagA delivery or IL-8 secretion and the CagN-deficient *H. pylori* strains could cause an infection similar to wild-type *H. pylori* strains. Some experiments also have indicated that CagN is not conclusively required for *H. pylori* T4SS function [16]. In another study conducted by Kutter et al., CagN was established to interact with two other *cag*PAI proteins, including CagV and CagY [35]. Thus, the biological function of CagN is yet to be investigated. In the current study, the attempts were made to detect possible variants of CagI

and CagN, as uncharacterized *cag*PAI-encoded factors, on both nucleotide and amino acid sequence levels among *H. pylori* isolates in Iran. We also investigated the distribution and variations in *H. pylori* virulence factors. Our findings revealed that 80% of *H. pylori* isolates harbored *cagI* gene, whilst 91.4% of strains had *cagN* gene. To the best of our knowledge, the *cagI* and *cagN* variants in *H. pylori* isolates in the subset of patients with different gastroduodenal diseases are not available in the literature. Based on our molecular findings, CagI E22, E221, and V268 amino acid polymorphisms occurred at higher rate in *H. pylori* isolates from NUD individuals compared to that isolated from PUD patients. On the other hand, CagI amino acid changes A23, S57, and S94 were detected at higher rates in *H. pylori* isolates from PUD patients compared to NUD subjects.

Despite the fact that Olbermann et al. found that *cagN* and *cagM* were demonstrated to be conserved in the *cag*PAI throughout all *cag*$^{+}$ *H. pylori* strains that have been sequenced so far [11], a high level of variability in CagN nucleotide and protein sequences was observed in present study. Furthermore, the most variable region in CagN amino acid sequence, so-called here as CagNHM, was found at residues 53 to 63 and contained many missense mutations. This region is postulated to contain GDEEITEEEKK sequence in the P12 reference strain, but varied among the sequenced strains in the current study.

Our findings revealed that there was no significant correlation between clinical outcomes and *cagI* and *cagN* variants at both nucleotide and amino acid levels (P >0.05), which is in consistent with previous study reported by Ogawa et al. [25]. Pham et al., stated that C-terminal motif (SKVIVK) in CagI is essential for T4SS function, and thus is completely conserved among *H. pylori* strains. Remarkably, the C-terminal motif of CagI is reported to be similar to the C-terminal motifs of CagL SK(I/V)IVK and CagH TKIIVK, representing the possibility that the amino acid sequences essentially act as binding motifs for a common interaction partner of all three proteins [17]. In agreement with above mentioned study, our findings also confirmed that the CagI C-terminal motif was completely conserved among all *H. pylori* isolates. Ogawa et al. discovered complete RGD motifs in CagL sequences were observed from all isolates, which possibly imply the importance of the RGD motif for CagL function [25]. A recent investigation on this topic was performed by Yadegar et al. in Iran, in which almost 97% of *H. pylori* clinical strains contained *cagL* gene [28]. Furthermore, their findings highlighted the importance of a common CagL hypervariable motif (CagLHM) such as NEIGQ along with multiple C-type EPIYA repeats, which was linked to PUD, GE, and GC with more severity compared to NUD. In fact, it is believed that the over mentioned CagLHM motif played a key role in the pathogenesis of *H. pylori* strains. Besides, sequencing analysis of the present study also showed that a hypothetical hexapeptide motif (EAKDEN/K) was detected in residues 278-283 in CagN among 13.9% of *H. pylori* isolates. Although Bats et al. [37] implied that the mutations and truncations in CagN sequence was irrelevant to folding properties or the overall shape of CagN, further studies are required to assess the impact of this hexapeptide motif on CagN protein structure and its role in *H. pylori* T4SS activity. Despite the alterations in various *cag* sequences, it is noticeable that all patients had high prevalence of *cag*PAI genes including *cagI*, *cagN*, *cagA* and *cagL* that indicates more potential role of these genes in disease outcome.

In the preset study, we also investigated the presence of various *H. pylori* virulence genotypes. In accordance with our previous studies in Iranian populations, we detected a high prevalence of *vacA* s1 (77.1%) and *vacA* m2 (65.7%) allelic genotypes [38, 39]. The *vacA* s1 allele has been reported to be associated with more severe atrophic gastritis in peptic ulcer patients [40, 41]. In our study, the *vacA* s1 genotype was found to be more prevalent among PUD patients, however, there was no significant association between the presence of other virulence genes and clinical disease outcomes. The mosaic combination of s- and m-region allelic genotypes also has been established to be associated with the pathogenicity of *H. pylori* [42, 43]. Accordingly, type s1m1 *H. pylori* strains express large amounts of VacA toxin and are strongly associated with a higher level of inflammation and mucosal ulceration, while *vacA* s1m2-harboring strains produce moderate amount of toxin and *vacA* s2m2 strains are virtually non-toxic and rarely associated to clinical outcome [44]. A majority of *H. pylori* strains in the current study contained *vacA* s1m2 genotype and this was mainly observed in NUD patients. On the contrary, allelic combination s1m1 or s2m2 genotypes were detected among the majority of clinical isolates of *H. pylori* in other parts of the world, and the hypervirulent *vacA* s1m1 genotype was commonly associated with PUD patients [45]. Hence, it can be inferred that correlation between *H. pylori* genotyping and clinical outcome of the patients vary in different geographical regions.

Conclusion

In summary, a large body of evidence indicates that certain *cag*PAI components are correlated with the risk of gastric carcinogenesis. Here, we investigated the diversity of CagI and CagN sequences in clinical *H. pylori* isolates from Iranian patients with different clinical status. We detected several putative variants of CagI and CagN sequences in *H. pylori* isolates, however, there was no significant relevance between these variants and clinical phenotypes. Our findings also demonstrated that the C-terminal SKVIVK motif within the CagI protein is conserved among all tested *H. pylori* strains. Meanwhile, the motif EAKDEN was a typical attribute identified in C-terminal sequence of CagN protein among some of the *H. pylori* strains, however its potential impact on T4SS activity and translocation of effectors requires further investigations. Despite the present study has successfully demonstrated the genetic diversity of *cagI* and *cagN* genes, it has certain limitations in terms of insufficient sample size. Accordingly, the possible effects of CagI and CagN variants on the T4SS activity as well as their possible interactions with other *cag*PAI components in a large number of *H. pylori* isolates needs to be explored. Also, the probable relevance of overmentioned variants with different clinical outcomes should not be ignored.

Declarations

Acknowledgments

The authors wish to thank the laboratory staff of Foodborne and Waterborne Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Authors' contributions

YA and SK cultured the isolates and performed the PCR test. AY worked on concepts and designed the study. AY, NM and HH participated in data analysis and wrote the manuscript. NK, HAA and MRZ critically revised the manuscript. All authors approved the final version of the manuscript.

Funding

This study was supported financially by a grant (no. RIGLD 969) from Foodborne and Waterborne Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Availability of data and materials

The available data used and/or analyzed during the current study are all included in the manuscript.

Ethics approval and consent to participate

This work deals with clinical bacterial strains isolated from human gastric biopsies. No tissue material or other biological material was stored from the patients, only subcultured bacterial isolates. Informed consent was obtained from all individual participants included in the study. All procedures performed were in accordance with the ethical standards retrieved from the Institutional Ethical Review Committee of Research Institute for Gastroenterology and Liver Diseases (RIGLD) at Shahid Beheshti University of Medical Sciences (Project No. IR.SBMU.RIGLD.REC.1398.023).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Amieva MR, El-Omar EM. Host-bacterial interactions in *Helicobacter pylori* infection. Gastroenterol. 2008; 134(1):306-23.
2. Naumann M, Sokolova O, Tegtmeyer N, Backert S. *Helicobacter pylori*: A paradigm pathogen for subverting host cell signal transmission. Trends Microbiol. 2017; 25(4):316-28.
3. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY, Chan FKL, Sung JJY, Kaplan GG, NG SC. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. Gastroenterol. 2017;153(2):420-29.

4. Malfertheiner P, Link A, Selgrad M. *Helicobacter pylori*: perspectives and time trends. Nat Rev Gastroenterol Hepatol. 2014;11(10):628-38.
5. Polk DB, Peek Jr RM. *Helicobacter pylori*: gastric cancer and beyond. Nat Rev Cancer. 2010;10(6):403-14.
6. Backert S, Neddermann M, Maubach G, Naumann M. Pathogenesis of *Helicobacter pylori* infection. Helicobacter. 2016;21 Suppl 1:19-25.
7. Hatakeyama M. *Helicobacter pylori* CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. Cell Host Microbe. 2014;15(3):306-16.
8. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, Misselwitz R, Berger J, Sewald N, König W, Backert S. *Helicobacter* exploits integrin for type IV secretion and kinase activation. Nature. 2007;449(7164):862-66.
9. Backert S, Tegtmeyer N, Fischer W. Composition, structure and function of the *Helicobacter pylori cag* pathogenicity island encoded type IV secretion system. Future Microbiol. 2015;10(6):955-65.
10. Tegtmeyer N, Wessler S, Backert S. Role of the *cag*-pathogenicity island encoded type IV secretion system in *Helicobacter pylori* pathogenesis. FEBS J. 2011;278(8):1190-202.
11. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, Vauterin M, Suerbaum S, Achtman M, Linz B. A global overview of the genetic and functional diversity in the *Helicobacter pylori cag* pathogenicity island. PLOS Genetic. 2010;6(8):e1001069.
12. Cendron L, Zanotti G. Structural and functional aspects of unique type IV secretory components in the *Helicobacter pylori cag*-pathogenicity island. FEBS J. 2011;278(8):1223-31.
13. Merino E, Flores-Encarnacion M, Aguilar-Gutierrez GR. Functional interaction and structural characteristics of unique components of *Helicobacter pylori* T4SS. FEBS J. 2017;284(21):3540-49.
14. Bourzac KM, Satkamp LA, Guillemin K. The *Helicobacter pylori cag* pathogenicity island protein CagN is a bacterial membrane-associated protein that is processed at its C terminus. Infect Immun. 2006;74(5):2537-43.
15. Ta LH, Hansen LM, Sause WE, Shiva O, Millstein A, Ottemann KM, Castillo AR, Solnick JV. Conserved transcriptional unit organization of the *cag* pathogenicity island among *Helicobacter pylori* strains. Front Cell Infect Microbiol. 2012; 2:46.
16. Fischer W, Püls J, Buhrdorf R, Gebert B, Odenbreit S, Haas R. Systematic mutagenesis of the *Helicobacter pylori cag* pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol. 2001;42(5):1337-48.
17. Pham KT, Weiss E, Soto LFJ, Breithaupt U, Haas R, Fischer W. CagI is an essential component of the *Helicobacter pylori* Cag type IV secretion system and forms a complex with CagL. PloS One. 2012;7(4):e35341.
18. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature. 2010;464(7286):250-55.

19. Li SD, Kersulyte D, Lindley IJ, Neelam B, Berg DE, Crabtree JE. Multiple genes in the left half of the *cag* pathogenicity island of *Helicobacter pylori* are required for tyrosine kinase-dependent transcription of interleukin-8 in gastric epithelial cells. Infect Immun. 1999;67(8):3893-99.
20. Selbach M, Moese S, Meyer TF, Backert S. Functional analysis of the *Helicobacter pylori cag* pathogenicity island reveals both VirD4-CagA-dependent and VirD4-CagA-independent mechanisms. Infect Immun. 2002;70(2):665-71.
21. Jimenez-Soto LF, Kutter S, Sewald X, Ertl C, Weiss E, Kapp U, Rohde M, Pirch T, Jung K, Retta SF, Terradot L, Fischer W, Haas R.. *Helicobacter pylori* type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner. PLoS Pathog. 2009;5(12):e1000684.
22. Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, Hennig EE, McClain MS, McDonald WH, Cover TL. *Helicobacter pylori* exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface. PLoS Pathog. 2011;7(9):e1002237.
23. Gorrell RJ, Zwickel N, Reynolds J, Bulach D, Kwok T. *Helicobacter pylori* CagL hypervariable motif: a global analysis of geographical diversity and association with gastric cancer. J Infect Dis. 2016;213(12):1927-31.
24. Linz B, Windsor HM, Gajewski JP, Hake CM, Drautz DI, Schuster SC, Marshall BJ. *Helicobacter pylori* genomic microevolution during naturally occurring transmission between adults. PloS One. 2013;8(12):e82187.
25. Ogawa H, Iwamoto A, Tanahashi T, Okada R, Yamamoto K, Nishiumi S, Yoshida M, Azuma T. Genetic variants of *Helicobacter pylori* type IV secretion system components CagL and CagI and their association with clinical outcomes. Gut Pathog. 2017;9(1):21.
26. Rizzato C, Torres J, Plummer M, Muñoz N, Franceschi S, Camorlinga-Ponce M, Fuentes-Pananá EM, Canzian F, Kato I. Variations in *Helicobacter pylori* cytotoxin-associated genes and their influence in progression to gastric cancer: implications for prevention. PLoS One. 2012;7(1):e29605.
27. Yadegar A, Alebouyeh M, Lawson AJ, Mirzaei T, Mojarad EN, Zali MR. Differentiation of non-pylori *Helicobacter* species based on PCR-restriction fragment length polymorphism of the 23S rRNA gene. World J Microbiol Biotechnol. 2014;30(6):1909-17.
28. Yadegar A, Mohabati Mobarez A, Zali MR. Genetic diversity and amino acid sequence polymorphism in *Helicobacter pylori* CagL hypervariable motif and its association with virulence markers and gastroduodenal diseases. Cancer Med. 2019;8(4):1619-32.
29. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: Nucleic Acids Symp Ser. 1999: [London]: Information Retrieval Ltd., c1979-c2000.; 1999: 95-98.
30. Kumar S, Stecher G, Tamura K: MEGA7. Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870-74.
31. Noto JM, Peek RM. The *Helicobacter pylori cag* pathogenicity island. Methods Mol Biol. 2012;921: 41-50.

32. Crabtree J, Taylor J, Heatley R, Shallcross T, Rathbone B, Wyatt J, Tompkins D. Mucosal IgA recognition of *Helicobacter pylori* 120 kDa protein, peptic ulceration, and gastric pathology. Lancet. 1991;338(8763):332-35.
33. Crabtree J, Wyatt J, Sobala G, Miller G, Tompkins D, Primrose J, Morgan A. Systemic and mucosal humoral responses to *Helicobacter pylori* in gastric cancer. Gut. 1993;34(10):1339-43.
34. Torres J, Pérez-Pérez GI, Leal-Herrera Y, MUNoz O. Infection with CagA+ *Helicobacter pylori* strains as a possible predictor of risk in the development of gastric adenocarcinoma in Mexico. Int J cancer. 1998;78(3):298-300.
35. Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R, Fischer W. Protein subassemblies of the *Helicobacter pylori* Cag type IV secretion system rRevealed by localization and interaction studies. J Biotechnol. 2008;190(6):2161-71.
36. Kumar N, Shariq M, Kumari R, Tyagi RK, Mukhopadhyay G. Cag type IV secretion system: CagI independent bacterial surface localization of CagA. PLoS One. 2013;8(9):e74620.
37. Bats SH, Bergé C, Coombs N, Terradot L, Josenhans C. Biochemical characterization of the *Helicobacter pylori* Cag type 4 secretion system protein CagN and its interaction partner CagM. Int J Med Microbiol. 2018;308(4):425-37.
38. Yadegar A, Mobarez AM, Alebouyeh M, Mirzaei T, Kwok T, Zali MR. Clinical relevance of cagL gene and virulence genotypes with disease outcomes in a *Helicobacter pylori* infected population from Iran. World J Microbiol Biotechnol. 2014;30(9):2481-90.
39. Farzi N, Yadegar A, Sadeghi A, Asadzadeh Aghdaei H, Marian Smith S, Raymond J, Suzuki H, Zali MR. High prevalence of antibiotic resistance in Iranian *Helicobacter pylori* Isolates: importance of functional and mutational analysis of resistance genes and virulence genotyping. J Clin Med. 2019;8(11).
40. Chiurillo MA, Moran Y, Cañas M, Valderrama E, Granda N, Sayegh M, Ramírez JL. Genotyping of *Helicobacter pylori* virulence-associated genes shows high diversity of strains infecting patients in western Venezuela. Int J Infect Dis. 2013;17(9):e750-e56.
41. van Doorn L-J, Schneeberger P, Nouhan N, Plaisier A, Quint W, De Boer W. Importance of *Helicobacter pylori cagA* and *vacA* status for the efficacy of antibiotic treatment. Gut. 2000;46(3):321-26.
42. Chauhan N, Tay ACY, Marshall BJ, Jain U. *Helicobacter pylori* VacA, a distinct toxin exerts diverse functionalities in numerous cells: An overview. Helicobacter. 2019;24(1):e12544.
43. Palframan SL, Kwok T, Gabriel K. Vacuolating cytotoxin A (VacA), a key toxin for *Helicobacter pylori* pathogenesis. Front cell Infect Microbiol. 2012;2:92.
44. Atherton JC, Cao P, Peek RM, Tummuru MK, Blaser MJ, Cover TL. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori* association of specific *vacA* types with cytotoxin production and peptic ulceration. J Biol Chem. 1995;270(30):17771-7.
45. Miernyk K, Morris J, Bruden D, McMahon B, Hurlburt D, Sacco F, Parkinson A, Hennessy T, Bruce M. Characterization of *Helicobacter pylori cagA* and *vacA* genotypes among Alaskans and their correlation with clinical disease. J Clin Microbiol. 2011;49(9):3114-21.

Tables

Table 1. Distribution of virulence genotypes in relation to clinical status among 70 *H. pylori* strains.

Virulence genotypes	Clinical status				Total
	NUD (*n*=39)	PUD (*n*=23)	IM (*n*=7)	GC (*n*=1)	(*n*=70)
cagI-positive	32 (82%)	17 (73.9%)	6 (85.7%)	1 (100%)	56 (80%)
cagN-positive	35 (89.7%)	22 (95.6%)	6 (85.7%)	1 (100%)	64 (91.4%)
cagA-positive	33 (84.6%)	19 (82.6%)	6 (85.7%)	1 (100%)	59 (84.3%)
cagL-positive	37 (94.9%)	21 (91.3%)	5 (71.4%)	1 (100%)	64 (91.4%)
vacA s1m1	12 (30.8%)	8 (34.8%)	3 (42.8%)	0 (0%)	23 (32.8%)
vacA s1m2	14 (35.9%)	12 (52.2%)	3 (42.8%)	1 (100%)	30 (42.8%)
vacA s2m2	13 (33.3%)	3 (13%)	1 (14.3%)	0 (0%)	17 (24.3%)
babA2-positive	37 (94.9%)	23 (100%)	7 (100%)	1 (100%)	68 (97.1%)
sabA-positive	32 (82%)	20 (87%)	6 (85.7%)	1 (100%)	59 (84.3%)
dupA-positive	32 (82%)	20 (87%)	6 (85.7%)	1 (100%)	59 (84.3%)

GC, gastric cancer; IM, intestinal metaplasia; NUD, nonulcer dyspepsia; PUD, peptic ulcer disease.

Table 2. The frequency of amino acid substitutions of CagI among clinical strains of *H. pylori* (*n* = 27) from patients with different clinical status

Residue [a]	Reference	Variant	NUD (*n*=19)	PUD (*n*=7)	IM (*n*=1)
2	K	N	1	_ [b]	-
3	C	S/F	1/1	-	-
6	S	D	1	-	-
9	S	F	1	-	-
12	T	I	1	-	-
22	E	G	1	3	-
23	V	A/I	1/1	3/1	-
25	I	M	2	1	-
34	I	N	1	-	-
40	A	V	1	-	-
44	T	A	1	-	-
51	A	V	2	-	-
57	N	S	9	5	1
94	G	S	10	6	1
116	A	G	1	-	-
162	A	T	1	-	-
166	A	V	-	1	-
182	E	K	1	-	-
187	A	T	1	-	-
190	S	N	-	1	-
192	S	F	1	-	-
195	A	T	2	1	-
199	A	T	1	-	-
204	G	S	1	-	-
213	K	E	3	-	-
221	T	E	4	-	-
243	A	T	4	2	-
246	A	V	-	1	-

254	S	N	2	-	-
257	A	T	1	-	-
262	I	F	1	-	-
263	E	Q	1	-	-
268	A	V/E	5/1	1/-	1/-
305	D	G/N	-/1	1/1	-
319	G	E	1	2	-
320	E	Q	1	2	-
351	L	F	1	-	-
353	K	T	1	-	-
368	T	M/K	1/1	-	-
375	S	G	-	2	1

NUD, nonulcer dyspepsia; PUD, peptic ulcer disease; IM, intestinal metaplasia.

[a]Positions of amino acid residues correspond to the *H. pylori* P12 reference strain.

[b]Positions of amino acid residues similar to the *H. pylori* P12 reference strain.

Table 3. The frequency of amino acid substitutions of CagN among clinical strains of *H. pylori* (*n*=43) from patients with different clinical status

Residue [a]	Reference	Variant	NUD (*n*=24)	PUD (*n*=14)	IM (*n*=4)	GC (*n*=1)
8	L	I	-	1	_ [b]	-
15	S	F	2	-	-	-
17	V	A/I	3/1	1/1	-	-
18	I	V	11	7	2	-
32	S	N	-	1	-	-
33	E	K	1	-	-	-
36	E	K	9	1	2	-
38	A	V	24	14	4	1
39	A	V	-	1	-	-
46	K	T	-	1	1	-
48	L	F	8	6	1	1
49	H	Y	7	4	1	1
52	H	R	-	1	-	-
53	G	D	24	14	4	1
54	D	N	1	-	-	-
55	E	K	7	3	-	-
57	I	V	16	10	2	-
59	E	K	17	13	3	-
61	E	K	3	-	-	-
63	K	E	16	12	3	-
80	A	V	-	2	1	-
98	V	I	18	9	1	1
102	A	V	2	-	-	-
103	A	T/S	8/-	5/1	1/-	-
106	K	R	3	1	-	-
114	I	T	3	6	-	-
117	T	N/H	7/14	3/9	3/1	1/-
118	P	S	-	1	-	-

121	D	N	2	-	-	-
125	S	G	3	2	-	-
129	A	T	16	11	4	1
134	N	H	2	-	-	-
137	D	G	1	-	-	-
140	D	N	2	2	-	-
148	E	G	13	6	2	-
149	A	S	7	4	2	-
154	A	T/V	2/1	4/-	1/-	1/-
155	A	T	-	1	-	-
160	N	D	18	11	4	-
161	E	K	-	1	-	-
170	I	V	3	-	-	-
174	C	G	1	-	-	-
182	D	N	-	1	1	-
191	G	D	-	2	1	-
194	D	E	1	-	-	-
199	A	T/V	5/-	5/1	2/-	-
203	E	K	1	-	1	-
208	I	V	1	-	-	-
221	S	N	2	-	-	-
224	K	R	1	-	-	-
225	L	F	1	-	-	-
226	A	V	1	-	-	-
227	L	F	-	2	1	1
228	N	H	-	1	-	-
232	N	S	-	1	-	-
233	R	K	-	1	-	-
241	T	A	22	13	4	1

248	K	R	-	1	-	-
259	T	I	1	-	-	-
262	A	T	2	1	-	-
263	S	G	-	-	1	-
264	K	E	23	14	4	1
267	T	A	15	11	2	1
268	T	A	1	12	-	-
273	N	S	1	-	-	-
279	T	A/V	7/1	5/-	2/-	1/-
280	F	S	1	1	1	-
284	R	H	4	2	-	-
285	S	F/P	2/-	-	-/1	-
287	S	F	1	1	-	-
288	E	D	1	-	-	-
302	A	V	1	-	-	-
304	E	G	24	14	4	1

NUD, nonulcer dyspepsia; PUD, peptic ulcer disease; IM, intestinal metaplasia; GC, gastric cancer.

[a]Positions of amino acid residues correspond to the *H. pylori* P12 reference strain. The inserted sequences are not indicated in the table.

[b]Positions of amino acid residues similar to the *H. pylori* P12 reference strain.

Figures

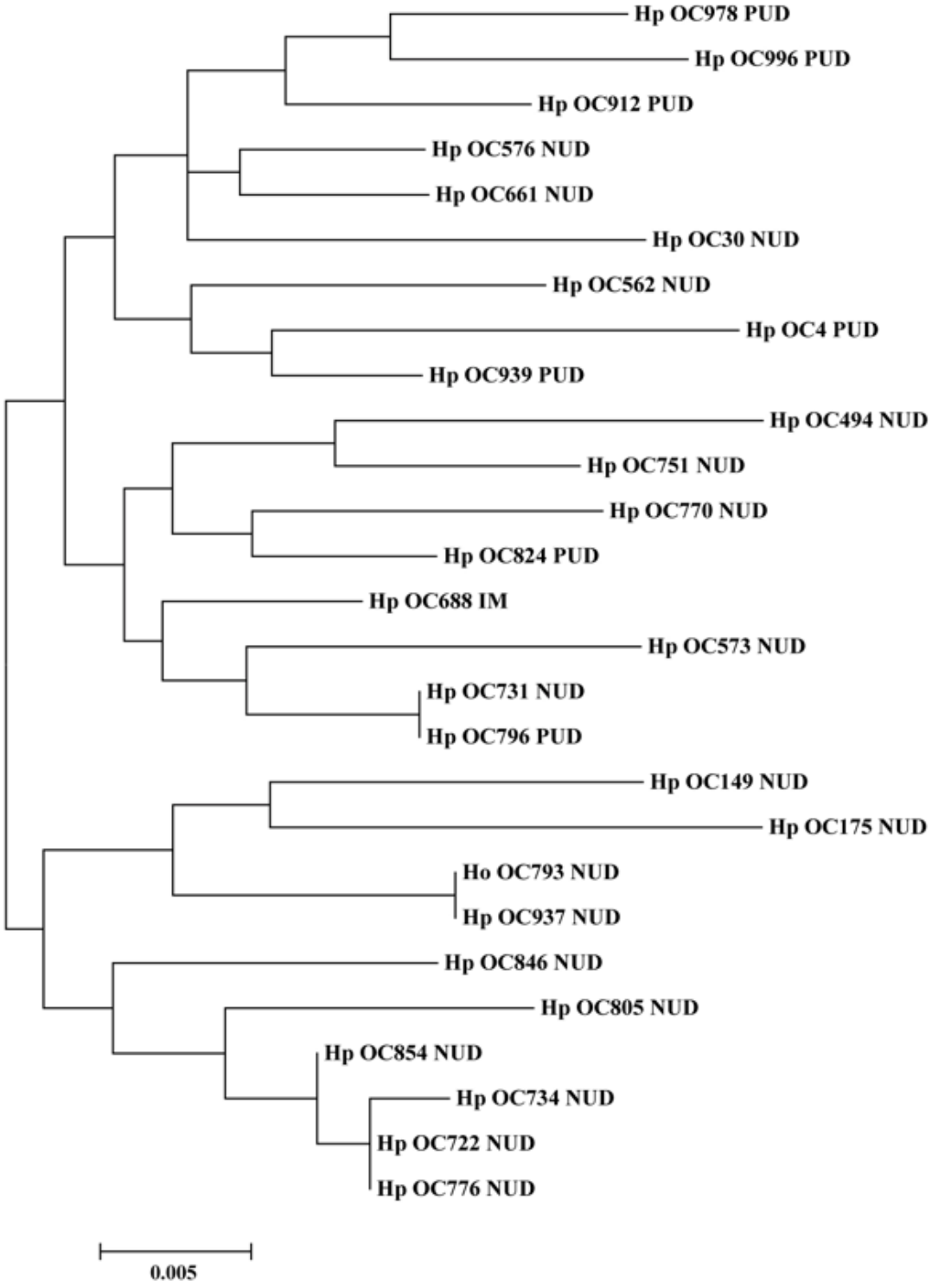

Figure 1

Phylogenetic tree of H. pylori clinical strains (n=27) based on cagI nucleotide sequences. Maximum likelihood tree of concatenated sequences was constructed using MEGA7 software with bootstrap method at 1000 replications. The evolutionary distances were computed using the Tamura 3-parameter model.

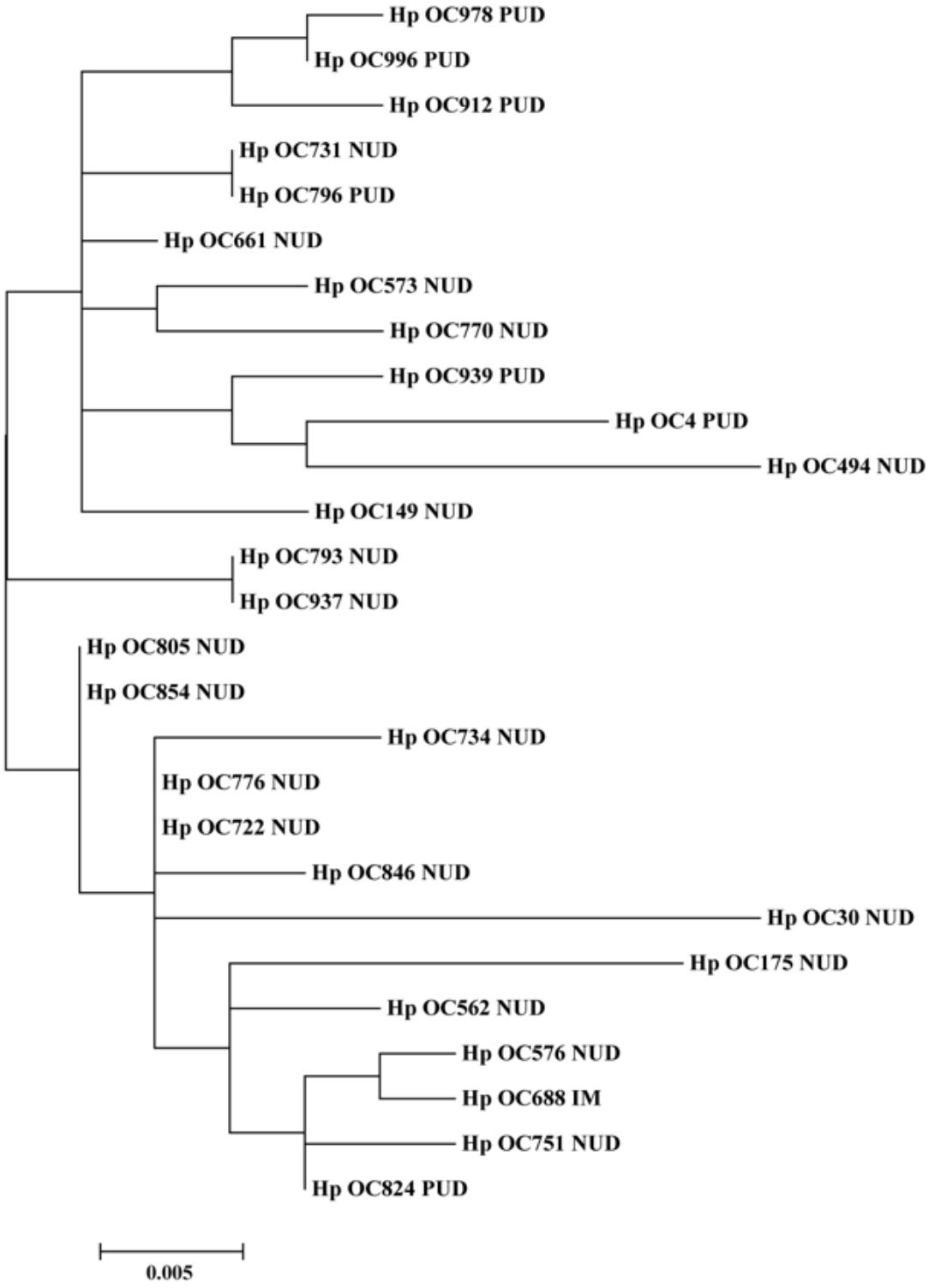

Figure 2

Phylogenetic tree of H. pylori clinical strains (n=27) based on translated CagI amino acid sequences. Maximum likelihood tree of concatenated sequences was constructed using MEGA7 software with bootstrap method at 1000 replications. The evolutionary distances were computed using the Poisson correction method.

Figure 3

Phylogenetic tree of H. pylori clinical strains (n=43) based on cagN nucleotide sequences. Maximum likelihood tree of concatenated sequences was constructed using MEGA7 software with bootstrap method at 1000 replications. The evolutionary distances were computed using the Tamura 3-parameter model.

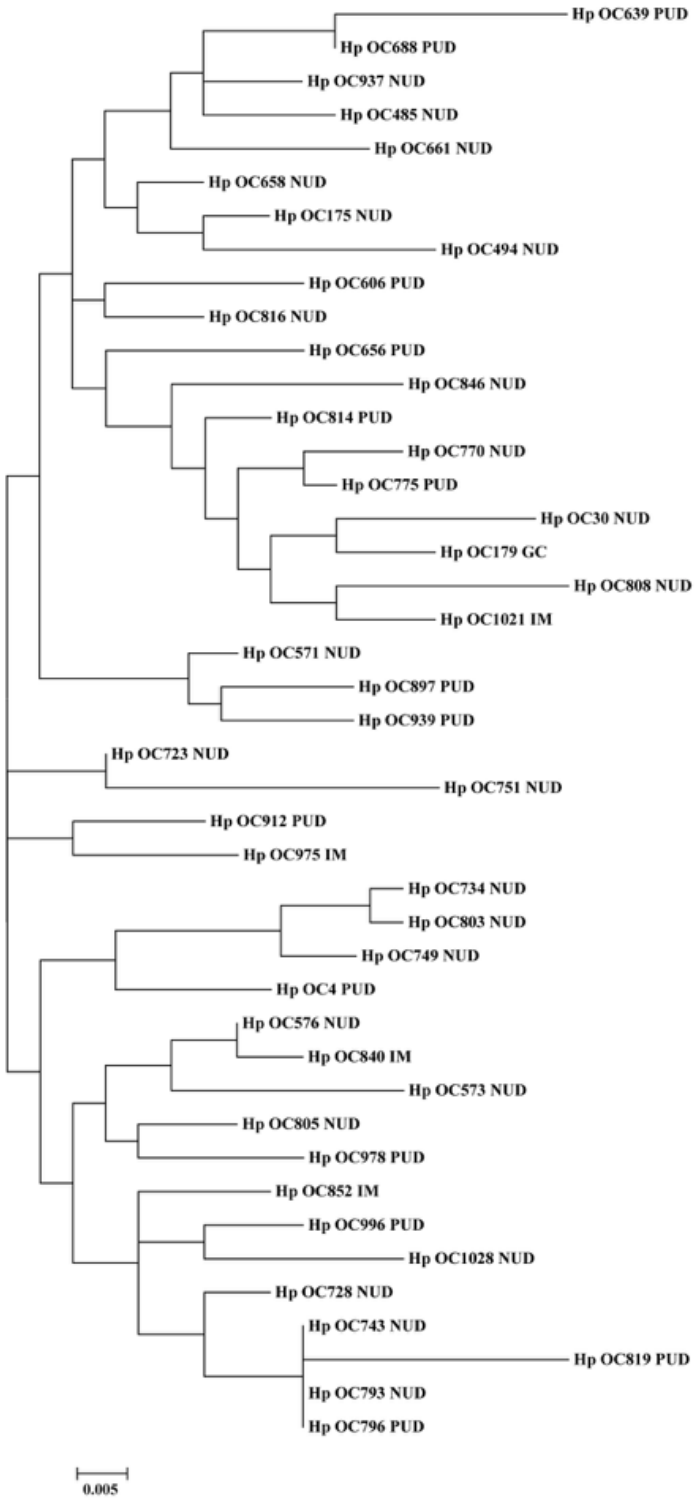

Figure 4

Phylogenetic tree of H. pylori clinical strains (n=43) based on translated CagN amino acid sequences. Maximum likelihood tree of concatenated sequences was constructed using MEGA7 software with bootstrap method at 1000 replications. The evolutionary distances were computed using the Poisson correction method.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx