

Genetic Determinants of Biofilm Formation and Antibiotic Resistance of Helicobacter Pylori using Whole Genome Sequencing

Kartika Afrida Fauzia

Oita University

Hafeza Aftab

Dhaka Medical College and Hospital

Muhammad Miftahussurur

Airlangga University

Langgeng Agung Waskito

Airlangga University

Vo Phuoc Tuan

Cho Ray Hospital

Takashi Matsumoto

Oita University

Michiyuki Yurugi

Oita University

Phawinee Subsomwong

Hirosaki University

Evariste Tshibangu-Kabamba

Osaka City University

Junko Akada

Oita University

Yoshio Yamaoka (✉ yyamaoka@oita-u.ac.jp)

Oita University

Research Article

Keywords: whole genome sequences, ARIBA, SNP, variants, biofilm formation, antibiotic resistance

Posted Date: August 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-800655/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The nucleotide polymorphisms (SNPs) associated with the biofilm formation phenotype of *Helicobacter pylori* were investigated. Fifty-six *H. pylori* isolates from Bangladeshi patients were included in this cross-sectional study. Crystal violet was used to classify the phenotypes into high- and low-biofilm formers. Whole genome sequences were analyzed using the “Antimicrobial Resistance Identification By Assembly” (ARIBA) pipeline. The results indicated 19.6% high- and 81.4% low-biofilm formers. These phenotypes were not related to specific clades in the phylogenetic analysis. Biofilm formation was significantly associated with SNPs of *alpA*, *alpB*, *cagE*, *cgt*, *csd4*, *csd5*, *futB*, *gluP*, *homD*, and *murF* ($P < 0.05$). Among the SNPs reported in *alpB*, strains encoding the N156K, G160S, and A223V mutations were high-biofilm formers. Mutations associated with antibiotic resistance can be detected. This study revealed the potential role of SNPs to biofilm formation, and propose a method to detect mutation in antibiotic resistance and biofilm from whole genome sequences.

1. Background

Helicobacter pylori infection remains a public health problem worldwide, affecting half of humans and inducing various diseases of the gastrointestinal tract [1]. The acidic and hostile conditions of the human stomach constitute the natural niche of *H. pylori*, while living bacteria have also been found in water resources, suggesting high adaptation abilities [2, 3]. Biofilm formation is a well-known adaptation mechanism in bacteria. Biofilm-forming *H. pylori* strains have been observed *in vivo* on the surface of the gastric mucosa [4]. The biofilm is a complex compound that protects and maintains life in the presence of external stress [5]. The biofilm also provides protection against antibiotics, resulting in a decline in the cure rate of *H. pylori* eradication therapy [6, 7]. Understanding the mechanism of biofilm formation is important for improving *H. pylori* elimination strategies.

Several approaches that enable the investigation of genetic mechanisms sustaining a given phenotype are currently available. The knockout of the target gene, which alters the phenotype, could be an approach to identify the responsible gene. For example, knockout of the *alpA* gene and the *luxS* gene inhibits the adhesion and initiation of biofilm formation [8, 9]. Another approach is to investigate the variation in the genome that is eventually associated with a certain phenotype. This approach has been rapidly developed, especially with the improvement of whole-genome sequencing methods. One of the notorious results is the discovery of mutations associated with antibiotic resistance, enabling clinicians to predict resistance only by gene analysis [10]. Other comparative genomic studies have also found the presence of several genes associated with biofilm density [11].

Biofilm formation ranges from low to high levels, indicating the potential involvement of a certain genotype or variant in the genome. However, the absence of a gene is not the only cause of the phenotypic alterations. As reported in a genome-wide association study, a nucleotide alteration variant might also result in a phenotype shift and an increased risk of gastric cancer [12]. Various methods and

pipelines have been developed to assess single nucleotide variants (SNVs). “Antimicrobial Resistance Identification By Assembly” (ARIBA) is a comprehensive pipeline that can be used to assemble a targeted gene and detect the presence or absence of genes and nucleotide variants^[13]. Therefore, in this study, we investigated the nucleotide variant from the whole genome sequence associated with the biofilm phenotype of clinical isolates obtained from subjects from Bangladesh.

2. Results

2.1. Distribution of biofilm formation among *H. pylori* strains from Bangladesh

Biofilm formation was quantified and classified into two groups: high- and low-biofilm formers (Fig. 1). The category of high-biofilm formers included 19.6% of strains (11/56) with a mean crystal violet staining optical density (OD) of 0.85 ± 0.4 . The prevalence of low-biofilm formers was higher, 80.4% (45/56), and this category of strain had a mean crystal violet staining OD of 0.24 ± 0.06 .

2.2. The phylogenetic similarity among the biofilm formers

A phylogenetic tree (Fig. 2) inferred by the maximum likelihood algorithm was used to assess the eventual association between the biofilm formation phenotype and different genetic *H. pylori* lineages. *H. pylori* in this study belong to the similar cluster with the hpEurope and hpAsia2. We noted that there was no specific phylogenetic branch associated with biofilm formation.

2.3. The association of presence and absence of genes to biofilm formation

We also investigated the presence and absence of targeted genes, which theoretically determine the phenotypes. Hence, we investigated the presence and absence of the genes based on a minimum coverage of 50% and a minimum identity of 90%. Using these criteria, we found that 32 of the 46 genes were present in all isolates. The genes that were absent in some isolates are listed in Table 1. The genes that were absent in more than 10% of the isolates were *vapD*, *cagD*, *cagE*, and *csd5*. However, no significant association was found between gene-presence-absence and biofilm formation.

Table 1
The association of the presence and absence of the gene with the biofilm formation

Genes	High			Low			P Value
	Present	Absent	Present Rate	Present	Absent	Present Rate	
<i>hp0968</i>	11	0	1.00	43	2	0.96	1
<i>cagE</i>	8	3	0.73	33	12	0.73	1
<i>aibB</i>	10	1	0.91	45	0	1.00	0.2
<i>cagD</i>	8	3	0.73	33	12	0.73	1
<i>csd5</i>	9	2	0.82	41	4	0.91	0.58
<i>luxS</i>	11	0	1.00	44	1	0.98	1
<i>futA</i>	11	0	1.00	44	1	0.98	1
<i>napA</i>	10	1	0.91	41	4	0.91	1
<i>ccmA</i>	11	0	1.00	44	1	0.98	1
<i>vapD</i>	5	6	0.45	26	19	0.58	0.51
<i>homB</i>	10	1	0.91	44	1	0.98	0.36
<i>tlpB</i>	10	1	0.91	43	2	0.96	0.49
<i>chePep</i>	10	1	0.91	44	1	0.98	0.36
<i>futB</i>	10	1	0.91	44	1	0.98	0.36

2.4. The variants associated with biofilm formation

The ARIBA pipeline assembles the gene according to the sequences in the references. The genes investigated in this study played a role in biofilm formation. The total number of polymorphisms that could be extracted from the analysis was 960, including frameshift, insertion, deletion, multiple site polymorphism, and single nucleotide polymorphism, as summarized in Fig. 3. The association of each novel variant with the biofilm level as the phenotype was analyzed using Fisher's exact test. The variants that were statistically significant for the biofilm phenotype are listed in Table 2. The genes that reported adhesion have been reported, such as *alpA*, *alpB*, *homD*, *cagE*, and *futB*. Other genes that have roles in metabolism and cell shape regulation, such as *gluP*, *cgt*, *csd4*, *csd5*, *murF*, and *amiA*, also showed significant associations. In the *alpB* gene, three SNPs were significantly associated with the high-biofilm former, A223V, G160S, and N156K.

Table 2
The significant variants associated with biofilm formation

Gene	SNP	P Value	Frequency in low biofilm	Frequency in high biofilm
<i>csd4</i>	C12Y	0.041	0.711	0.364
<i>cagE</i>	S419C	0.045	0.424	0.875
<i>gluP</i>	T85S	0.032	0.267	0.636
<i>cgt</i>	V34A	0.020	0.178	0.545
<i>alpB</i>	A223V	0.017	0.311	0.727
	G160S	0.002	0.378	0.909
	N156K	0.010	0.089	0.455
	T127A	0.040	0.911	0.636
<i>csd5</i>	M125I	0.017	0.146	0.556
	P43S	0.026	0.098	0.444
	V110A	0.015	0.073	0.444
<i>alpA</i>	G196N	0.045	0.156	0.455
<i>homD</i>	A570T	0.040	0.089	0.364
	V249fs	0.029	0.133	0.455
<i>murF</i>	V250I	0.026	0.333	0.000
<i>futB</i>	E98fs	0.033	0.159	0.500
	L317I	0.010	0.432	0.000
<i>amiA</i>	T168V	0.022	0.067	0.364

2.4. The genes associated with antibiotic resistance

Biofilm-related antibiotic resistance is the main concern in biofilm studies. Screening genes for antibiotic resistance and biofilm formation can improve personalized treatment strategies. Mutations associated with antibiotic resistance can be detected by this pipeline. A well-known mutation, which has been proven by a natural transformation study, was obtained (Table 3). Mutations such as A2147G of the *23S-rRNA* gene and D91N and E679D of *gyrA* were significantly associated with clarithromycin and levofloxacin resistance, respectively.

Table 3
The SNPs associated with antibiotic resistance

Antibiotic	Resistant rate (%)	Associated gene	Mutation	P Value
Amoxicillin	2/54 (3.7)	<i>pbp1A</i>	V45I	0.019
Clarithromycin	17/54 (31.5)	<i>23SrRNA</i>	A2147G	0.00017
Metronidazole	52/54 (96.3)	<i>rdxA</i>	NA*	
Tetracycline	0/54 (0.0)	<i>16srRNA</i>	NA**	
Levofloxacin	35/54 (64.8)	<i>gyrA</i>	D91N	0.039
			E679D	0.001
			G733E	0.006
		<i>gyrB</i>	A343V	0.039
*No variants or mutations were significantly associated with metronidazole resistance				
**All strains in this study were sensitive to tetracycline				

2.5. The predicted effect of mutation on the protein structure

Amino acid alterations can affect protein stability. Protein modeling can predict the structure according to the template available in the database. Here, we evaluated the protein structure prediction of the *gluP* and *cgt* genes. Analysis showed that the presence of the mutation had a destabilizing effect, as shown in Fig. 4. Mutations in the CGT gene showed a decrease in hydrogen bonds and hydrophobic contact, while mutation in GluP showed a decrease in the halogen bond.

3. Discussion

Biofilm formation is a beneficial mechanism. Its construction requires complex regulation not only within a single bacterium but also with other bacteria within a biofilm body. In *H. pylori*, *in vitro* observations of mono-species biofilms showed that high biomass was obtained after 72 h [14, 15]. However, evaluation of several strains under the same conditions showed that biofilm formation was significantly higher in some strains. This shows that variations exist in biofilm formation. These variations were also observed in the present study. Among the strains, 19.6% overproduced the biofilm. High-biofilm formers are likely to be more resistant to antibiotic exposure [7, 16]. However, the prediction of genetic determinants for biofilm formation and antibiotic resistance is necessary.

Phylogenetic analysis showed that there was no association of biofilm formation to a certain lineages inferred using the SNPs based core genome alignment tree. In some studies on other *Staphylococcus*

aureus species, biofilm formation has been reported to be associated with specific lineages^[17] whereas other studies reported no phylogenetic link^[18]. In a previous study, the *H. pylori* population of Bangladesh was determined using Structure software, which showed that Bangladeshi isolates were separated into two populations, hpEurope and hpAsia2. The whole genome phylogenetic tree in this study also confirmed that Bangladeshi isolates were clustered near hpEurope strains and hpAsia2 strains.

The whole genome sequence data allow researchers to investigate and screen genes and mutations related to biofilm formation. The ability of *H. pylori* to grow under a biofilm on an abiotic surface enables researchers to identify genes involved in biofilm formation through the generation of knockout mutants^[9, 19]. Hence, we tried to validate the presence or absence of specific genes of potential interest that play roles in the initiation of adhesion, shape formation, efflux pump, and even dispersion, as listed in Supplementary Table 2. Our results showed that these genes were present in almost all isolates, despite variations in the level of biofilm formation. This outcome diverges from observations made in other bacteria such as *Staphylococcus aureus*, where the presence of the *ica* gene could be noted in most of the high-biofilm formers^[20]. However, the gene marker for the high-biofilm formers remains unclear in *H. pylori*. A previous study using comparative genomic data captured several genes, such as *cagD*, *futA*, and *napA*, whose presence was associated with the biofilm level^[11]. Nevertheless, for the strains that possessed most of the targeted genes, another level of polymorphism might exist and affect the phenotype.

Additional analyses focused on amino acid variants, including insertion, deletion, missense mutation, or SNPs. To our knowledge, this is the first study to assess the SNPs associated with biofilm formation in *H. pylori*. Several SNPs have been linked to specific phenotypes of *H. pylori*, such as diseases like gastric cancer^[12]. Rather than the absence or presence of specific genes, SNPs or missense mutations are also well known to induce the antibiotic resistance phenotype in *H. pylori*^[21, 22]. The discovery of mutation from whole genome sequence approach were also used in several previous studies^[23]. In this study, by using the ARIBA pipelines, the association of the SNPs with the antibiotic resistant phenotype were observed and concordant with other reports that confirm the mutation roles by natural transformation^[21]. First, a reference database of genes encoding cellular targets for amoxicillin, clarithromycin, tetracycline, metronidazole, and clarithromycin were constructed. The mutation that has been reported to cause antibiotic resistance could be detected using this method, such as A2147G and D91N of *gyrA*^[10, 21, 24]. A relatively rare mutation in the locus V45I of *pbp1a*; E679D and E733E of *gyrA*; and A343V of *gyrB* was also observed. The analysis was then expanded to the evaluation of the SNPs associated with biofilm formation.

The database of the genes that was proposed to be associated with biofilm formation was constructed and applied to the dataset from Bangladesh. As the results, 11 genes that possessed variants significantly associated with biofilm formation were identified. Among these, five genes encode outer membrane proteins (OMPs) that are crucial for cell adhesion^[9]. The other genes are involved in cell

shape regulation [25]. This suggests that the change of the spiral shape toward unusual forms such as coccoid forms of *H. pylori* deserves a spotlight in the biofilm formation mechanism study.

The genes investigated in this study included those encoding AlpA and AlpB, which are OMPs involved in the adherence process in *H. pylori*. CagD and CagE are part of the Type IV secretion system known to be required for biofilm formation in its early establishment through a mechanism that requires further clarification [26]. The other genes encode the Csd protein complex, MurF, and AmiA, which are crucial for peptidoglycan regulation to maintain the helical shape of bacterial cells [27, 28]. The change from a helical shape to a coccoid is known to be related to the virulence of *H. pylori* [29] and constitutes a strategy for increasing survival in difficult environments [30]. Studies have established that coccoid forms are common forms of cells observed in biofilm-forming *H. pylori* [31]. These findings support further studies assessing the role of morphological changes in biofilm formation, which remains unclear to date.

The structural changes in the protein product due to the SNPs were captured by protein modeling analysis. Although this prediction was limited to proteins available in the database, structural changes could be observed in the *cgt* and *gluP* genes. Cholesterol- α -glucosyltransferase (CGT) protein acts by importing cholesterol from the host and converting it to cholesteryl glucosides, which are important for colonization, cell wall formation, and host phagocytosis evasion [32, 33]. GluP is an efflux pump that affects multidrug resistance and biofilm formation [34]. Further experimental protein analysis should be performed to confirm this functional change.

This study had some limitations. The small number of samples included in this dataset and the statistical approaches applied cannot avoid the possibility of false positive discovery among the detected SNPs. However, it could be a stepping stone for further molecular studies elucidating the genetic factors involved in biofilm formation and related molecular mechanisms.

4. Conclusions

The association analysis of the SNPs in the well-known biofilm formation genes proved the ability to screen biofilm-formation capacity from whole genome sequencing data. From these data, we observed a significant association between the SNPs including *alpA*, *alpB*, *cagE*, and *csd4*. This could be a method to decipher the mechanism of biofilm formation. Furthermore, biofilm-associated variants from the whole genome sequence could be used as biomarkers for antibiotic resistance screening alongside the well-known antibiotic-resistance genes.

5. Materials And Methods

5.1. Patient sampling and *H. pylori* isolates

The isolates used in this study were obtained from gastric biopsies of patients from a survey involving 133 subjects from Dhaka Medical College Hospital in 2014 [35]. To obtain *H. pylori* isolates, biopsy

specimens from the antrum were homogenized in phosphate buffered saline (PBS), inoculated in *H. pylori* selective plates, and incubated for 5 days in a 37°C microaerophilic environment. The colonies were then sub-cultured in Brucella agar supplemented with 7% horse blood before harvesting for genomic DNA extraction. The *H. pylori* culture obtained 56 isolates from patients with chronic gastritis (53/56) and peptic ulcers (3/56). The protocol of this study was approved by the Oita University Faculty of Medicine, Japan and the Ethics Committee of the Bangladesh Medical Research Council (BMRC), Dhaka, Bangladesh according to the Declaration of Helsinki. Informed consents were obtained from all the subjects.

5.2. Biofilm quantification and antibiotic resistance determination

Biofilm quantification was performed using the crystal violet method for *H. pylori*, as previously described, with modifications [36]. Briefly, the bacteria that grew in the blood plate were collected in 1 mL Brucella broth supplemented with 10% fetal bovine serum (FBS) and pre-cultured for 24 h under microaerophilic conditions. The bacterial suspension was adjusted to an OD of 0.4 (approximately 2.5×10^6 /mL), and 25 μ L of *H. pylori* suspension was inoculated into 24-well plates containing 1 mL medium. These plates were incubated in a microaerophilic environment with shaking (100 rpm) for 3 days. The planktonic cell suspension was then removed. The plates were washed with PBS and air-dried for 1 h to facilitate a stronger attachment of the biofilm to the wall. The biofilm was then stained with 500 μ L of 0.01% crystal violet for 1 min, washed with distilled water, and then air-dried for 15 min. The crystal violet was then diluted with a 500 μ L mixture of ethanol and acetic acid (ethanol:acetic acid = 95:5). The quantity of biofilm was obtained from the absorbance measurement with a spectrophotometer (Multiskan Go, Thermo Fischer, Japan) at a wavelength of 595 nm. The measurement of the well containing medium without bacteria was used as a negative control. The biofilm was classified as negative if the measured OD was lower than that of the control. Biofilm formation was classified into three groups: a negative biofilm former had an OD < control OD, a weak biofilm-former had an OD \geq the control OD and < 2X the control OD, and a strong biofilm former had an OD \geq 2X the control OD [37,38]. For the phenotype and genotype association analysis, the negative biofilm former group was combined with the weak biofilm group as a low-biofilm former. Finally, two phenotype groups were observed: the low-biofilm former (consisting of low and negative groups), and the high-biofilm former group (consisting of strong biofilm former strains). Because genes in reference strain 26695 were used to construct the reference database, the biofilm formation of strain 26695 was also performed but not included in the dataset. The OD of the 26695 strain biofilm was 0.35, which is less than 0.4, so it was also classified as a low-biofilm former.

The antibiotic susceptibility to amoxicillin, clarithromycin, metronidazole, levofloxacin, and tetracycline was determined by the agar dilution method on the Mueller Hinton Agar supplemented 10% horse blood as described in the previous study [35]. The resistant was determined according to the EUCAST criteria.

5.3. Genome Sequencing

Genome sequence data were used in a previous study [39]. The *H. pylori* bacteria were harvested in PBS, and the DNA was extracted using the Qiagen DNEasy Kit (Hilden, Germany) according to the manufacturer's instructions. The concentration of the DNA was then measured using the Quantifluor dsDNA System (Madison, USA) and Quantus Fluorometer (Sunnyvale, USA). After standardizing the concentration, whole genome sequencing was performed using MiSeq Illumina to obtain paired-end reads with a 300 bp length.

To analyze the genetic relatedness and the influence of population genetics, we created the whole genome alignment of all Bangladesh sequences and sequences from a previous study that identified the population [40]. The alignment was performed using the Snippy-core version 4.6.2 (<https://github.com/tseemann/snippy>). The alignment was used to construct a maximum likelihood tree using FastTree 2.0, with GTR-nt mode [41]. The tree was then visualized using Microreact [42].

5.4. Analysis of the SNP

The analysis of the SNP used in the study was performed using ARIBA software [13]. First, 42 genes of strain 26695 that were related to biofilm formation mentioned in previous studies were collected to create a reference database (Supplementary Table 1). The genes that are known to be responsible for antibiotic resistance with different variants were also prepared in another reference database. Subsequently, the raw reads (fastq format) of the 56 strains were used as the input files. The metadata were set into coding sequences and new variants. The ARIBA was run with default settings using these references. The results of each strain consisted of the assembled genes and the report of variants compared to strain 26695. The reports were then summarized in a text file. The genes were considered to be present if the percent coverage was more than 50% and the percent identity was more than 90%. Then, a summary of SNVs with amino acid changes was shown, and the SNVs that presented in less than 10% of the strains were excluded. The presence and absence of amino acid alterations were calculated to determine the association with biofilm production. A Fisher's exact test result of $P < 0.05$ was considered significant. The other non-parametric correlation was analyzed using Spearman's rank correlation model. All statistical analyses and graph construction were performed using R (version 3.5.1).

5.5. Analysis of protein structure prediction

The genes with SNPs that were significantly associated with biofilm were listed and retrieved from 26695 complete genomes NCBI as fasta files. These genes were uploaded in the Swiss-Model to obtain protein structure homology [43]. The best protein prediction model [44] was used as the input file for the Dynamut software to predict protein stability changes upon mutation [44].

Declarations

Data Availability

All genome data were stored in Genbank with BioProject accession PRJDB11821. More detailed information will be available from corresponding author on the reasonable request.

Author Contribution

Study conception and design (K.A.F. and Y.Y.). Acquisition of data (K..A.F, H.A, M.M, M.Y, T.M, J.A,PS), Computational method development (K.A.F, L.A.W, V.P.T, and E.T.K.). Analysis and interpretation of data (K.A.F, E.T.K, V.P.T.). Writing manuscript (K.A.F, Y.Y.). Critical revision of the manuscript and final approval (all authors).

Funding:

This study was supported in part by grants from the National Institutes of Health (DK62813) (YY) and Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (221S0002, 16H06279, 18KK0266, 19H03473) (YY) and 18K16182 (TM), 17K09353(JA). This work was also supported by the Japan Society for the Promotion of Science Institutional Program for Young Researcher Overseas Visits and the Strategic Funds for the Promotion of Science and Technology Agency (JST) for YY. KAF is a Ph.D. student supported by the Japanese Government (MEXT) scholarship program for 2017.

Acknowledgments:

None

Conflicts of Interest:

The authors declare no conflict of interest. The funders had no role in the study design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Correa, P. & Houghton, J. Carcinogenesis of *Helicobacter pylori*., **133** (2), 659–672 (2007).
2. Dube, C., Tanih, N. & Ndip, R. *Helicobacter pylori* in water sources: a global environmental health concern. *Rev Environ Health*, **24** (1), 1–14 (2009).
3. Stark, R. M. *et al.* Biofilm formation by *Helicobacter pylori*. *Lett Appl Microbiol*, **28** (2), 121–126 (1999).

4. Carron, M. A. *et al.* Identification of *Helicobacter pylori* biofilms in human gastric mucosa. *J Gastrointest Surg*, **10** (5), 712–717 (2006).
5. Monds, R. D. & O'Toole, G. A. The developmental model of microbial biofilms: ten years of a paradigm up for review. *Trends Microbiol*, **17** (2), 73–87 (2009).
6. Yonezawa, H. *et al.* Assessment of in vitro biofilm formation by *Helicobacter pylori*. *J Gastroenterol Hepatol*, **25** (Suppl 1), 90–94 (2010).
7. Fauzia, K. A. *et al.* Biofilm Formation and Antibiotic Resistance Phenotype of *Helicobacter pylori* Clinical Isolates. *Toxins*, **12** (8), 473 (2020).
8. Cellini, L. *et al.* Biofilm formation and modulation of luxS and rpoD expression by *Helicobacter pylori*. *Biofilms*, **2** (2), 119 (2005).
9. Yonezawa, H. *et al.* Diversification of the AlpB outer membrane protein of *Helicobacter pylori* affects biofilm formation and cellular adhesion. *Journal of bacteriology*, 2017. **199**(6).
10. Tuan, V. P. *et al.* A next-generation sequencing-based approach to identify genetic determinants of antibiotic resistance in Cambodian *Helicobacter pylori* clinical isolates. *Journal of clinical medicine*, **8** (6), 858 (2019).
11. Wong, E. H. *et al.* Comparative Genomics Revealed Multiple *Helicobacter pylori* Genes Associated with Biofilm Formation In Vitro. *PLoS One*, **11** (11), 0166835 (2016).
12. Berthenet, E. *et al.* A GWAS on *Helicobacter pylori* strains points to genetic variants associated with gastric cancer risk. *BMC biology*, **16** (1), 84 (2018).
13. Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial genomics*, 2017. **3**(10).
14. Windham, I. H. *et al.* *Helicobacter pylori* Biofilm Formation Is Differentially Affected by Common Culture Conditions, and Proteins Play a Central Role in the Biofilm Matrix. 2018. **84**(14).
15. Yonezawa, H. *et al.* Outer membrane vesicles of *Helicobacter pylori* TK1402 are involved in biofilm formation. *BMC Microbiol*, **9**, 197 (2009).
16. Yonezawa, H., Osaki, T. & Kamiya, S. Biofilm Formation by *Helicobacter pylori* and Its Involvement for Antibiotic Resistance. *BioMed Research International* 2015. **2015**: p. 914791.
17. Tasse, J. *et al.* Association between biofilm formation phenotype and clonal lineage in *Staphylococcus aureus* strains from bone and joint infections. *PLoS One*, **13** (8), 0200064 (2018).
18. Song, M. *et al.* Biofilm formation and antibiotic resistance pattern of dominant *Staphylococcus aureus* clonal lineages in China. *Journal of Food Safety*, **37** (2), 12304 (2017).
19. Anderson, J. K. *et al.* Chemorepulsion from the Quorum Signal Autoinducer-2 Promotes *Helicobacter pylori* Biofilm Dispersal. *mBio*, **6** (4), 00379 (2015).
20. Cramton, S. E. *et al.* The intercellular adhesion (ica) locus is present in *Staphylococcus aureus* and is required for biofilm formation. *Infection and immunity*, **67** (10), 5427–5433 (1999).
21. Lauener, F. N. *et al.* Genetic determinants and prediction of antibiotic resistance phenotypes in *Helicobacter pylori*. *Journal of clinical medicine*, **8** (1), 53 (2019).

22. Domanovich-Asor, T. & Motro, Y. Genomic Analysis of Antimicrobial Resistance Genotype-to-Phenotype Agreement in *Helicobacter pylori*. 2020. **9**(1).
23. Tuan, V. P., Narith, D. & Tshibangu-Kabamba, E. *A Next-Generation Sequencing-Based Approach to Identify Genetic Determinants of Antibiotic Resistance in Cambodian Helicobacter pylori Clinical Isolates*. 2019. **8**(6).
24. Iwamoto, A. *et al.* Whole-genome sequencing of clarithromycin resistant *Helicobacter pylori* characterizes unidentified variants of multidrug resistant efflux pump genes. *Gut pathogens*, **6** (1), 27 (2014).
25. Hathroubi, S. *et al.* *Helicobacter pylori* Biofilm Formation and Its Potential Role in Pathogenesis. 2018. **82**(2).
26. O'Toole, G., Kaplan, H. B. & Kolter, R. Biofilm formation as microbial development. *Annu Rev Microbiol*, **54**, 49–79 (2000).
27. Sycuro, L. K. *et al.* Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization., **141** (5), 822–833 (2010).
28. Yang, D. C. *et al.* A genome-wide *Helicobacter pylori* morphology screen uncovers a membrane-spanning helical cell shape complex. *Journal of bacteriology*, **201** (14), 00724–00718 (2019).
29. Salama, N. R. Cell morphology as a virulence determinant: lessons from *Helicobacter pylori*. *Current Opinion in Microbiology*, **54**, 11–17 (2020).
30. Reshetnyak, V. I. & Reshetnyak, T. M. Significance of dormant forms of *Helicobacter pylori* in ulcerogenesis. *World J Gastroenterol*, **23** (27), 4867–4878 (2017).
31. Attaran, B., Falsafi, T. & Moghaddam, A. N. Study of biofilm formation in C57Bl/6J mice by clinical isolates of *Helicobacter pylori*. *Saudi J Gastroenterol*, **22** (2), 161–168 (2016).
32. Lebrun, A. H. *et al.* Cloning of a cholesterol- α -glucosyltransferase from *Helicobacter pylori*. *Journal of Biological Chemistry*, **281** (38), 27765–27772 (2006).
33. Du, S. Y. *et al.* Cholesterol glucosylation by *Helicobacter pylori* delays internalization and arrests phagosome maturation in macrophages. *Journal of Microbiology, Immunology and Infection*, **49** (5), 636–645 (2016).
34. Ge, X., Cai, Y. & Chen, Z. *Bifunctional Enzyme SpoT Is Involved in Biofilm Formation of Helicobacter pylori with Multidrug Resistance by Upregulating Efflux Pump Hp1174 (gluP)*. 2018. **62**(11).
35. Aftab, H. *et al.* *Two populations of less-virulent Helicobacter pylori genotypes in Bangladesh*. 2017. **12**(8): p.e0182947.
36. Yonezawa, H. *et al.* Impact of *Helicobacter pylori* biofilm formation on clarithromycin susceptibility and generation of resistance mutations. *PLoS One*, **8** (9), 73301 (2013).
37. Odeyemi, O. A. *Microtiter plate assay methods of classification of bacterial biofilm formation*. *Food Control*, 2017(73): p.245–246.
38. Reiter, K. C. *et al.* Inhibition of biofilm maturation by linezolid in methicillin-resistant *Staphylococcus epidermidis* clinical isolates: comparison with other drugs. *Journal of medical microbiology*, **62** (3),

394–399 (2013).

39. Aftab, H. *et al.* Two populations of less-virulent *Helicobacter pylori* genotypes in Bangladesh. *PLoS One*, **12** (8), 0182947 (2017).
40. Yahara, K. *et al.* Chromosome painting in silico in a bacterial species reveals fine population structure. *Mol Biol Evol*, **30** (6), 1454–1464 (2013).
41. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS one*, **5** (3), 9490 (2010).
42. Argimón, S. *et al.* *Microreact: visualizing and sharing data for genomic epidemiology and phylogeography*. *Microbial genomics*, 2016. **2**(11).
43. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, **46** (W1), 296–303 (2018).
44. Rodrigues, C. H., Pires, D. E. & Ascher, D. B. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research*, **46** (W1), 350–355 (2018).

Figures

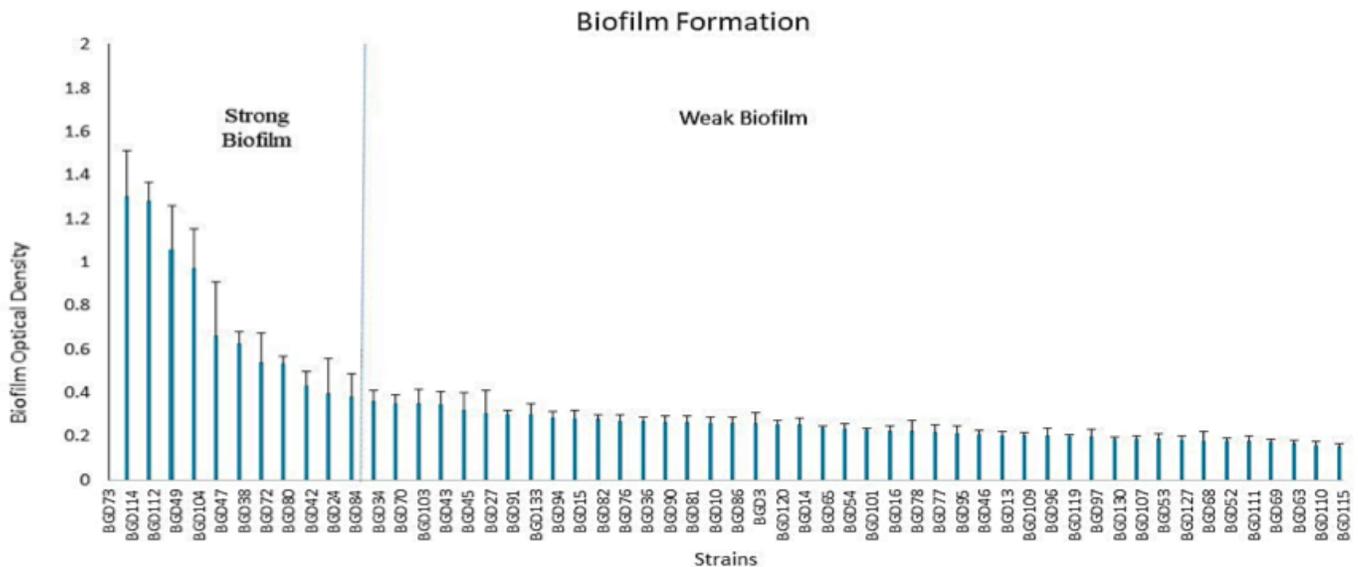


Figure 1

Distribution of biofilm formation in strains from 56 patients from Bangladesh. The x-axis represented the strains and y-axis was the optical density of biofilm formation after crystal violet staining. The cutoff (0.4) divided the strains into strong and weak biofilm former.

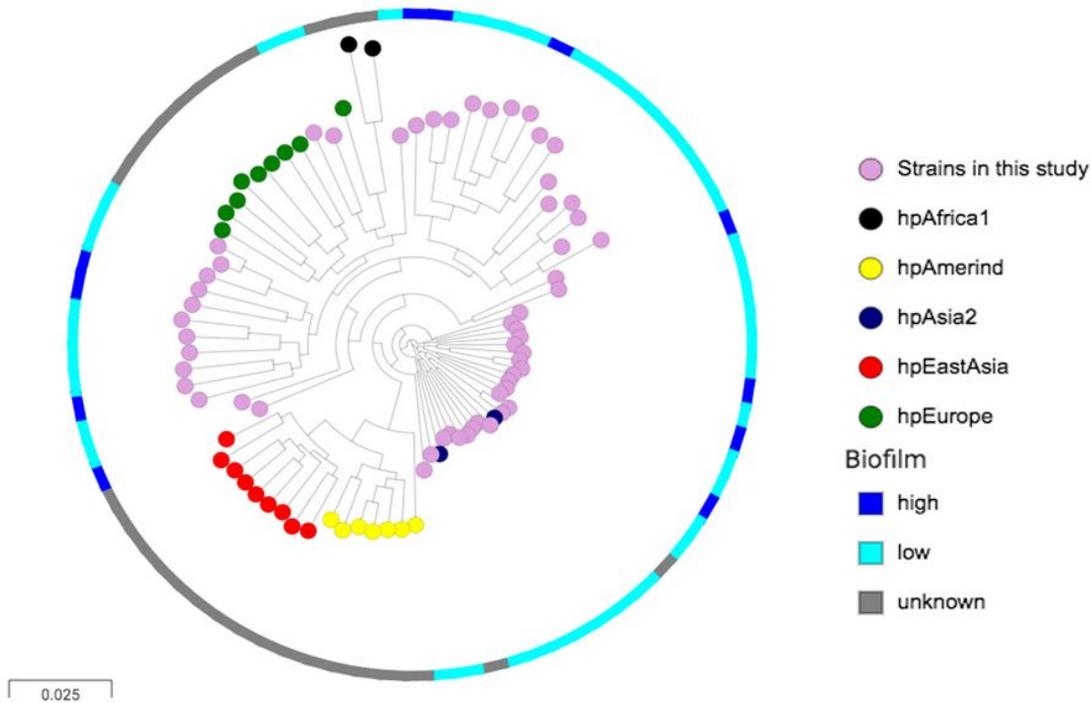


Figure 2

The phylogenetic tree depicting the population and the biofilm formation. The tree were constructed from the SNPs of Bangladesh strains and also the reference strains available from the genbank. The bar represented the biofilm phenotype. The biofilm data of the reference strains obtained from genbank were not available, hence it is written unknown. The round shapes represented the *H. pylori* genomic population.

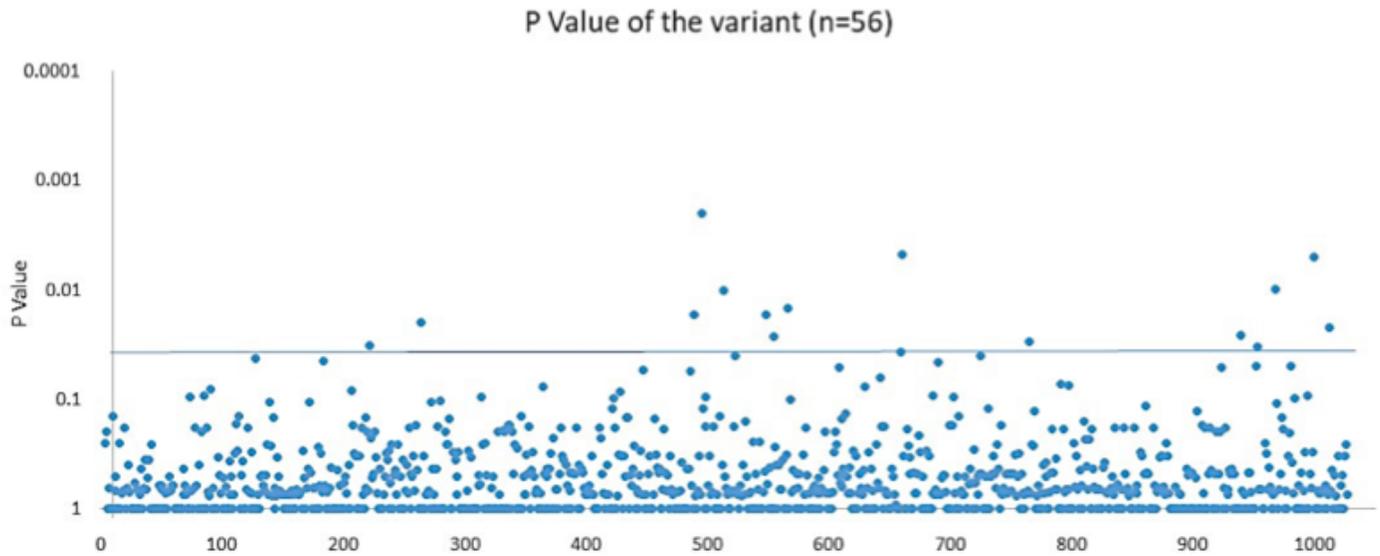


Figure 3

Scatter plot of P-values of the nucleotide polymorphisms (SNPs). The P-values observed in all nucleotide polymorphisms (SNPs) of all genes are summarized in this figure. The x-axis represents the lists of SNPs, and the y-axis represents the p-value from fisher exact tests for the genes and the biofilm phenotype. The straight horizontal line represents the cut-off for the significant P-value (0.05). The dots show the P-value of each variant, and 18 dots above the line have a P-value less than 0.05.

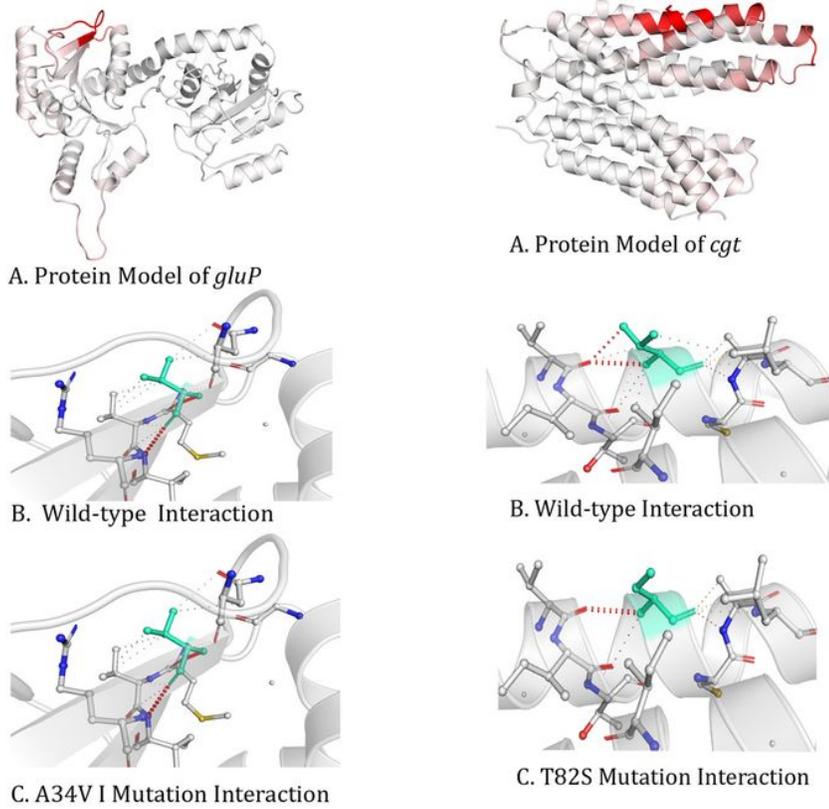


Figure 4

The effect of mutation on the amino acid interaction. The models of the CGT and GluP protein are shown. The amino acid on the target (light green) interacts with other proteins. The red dot appearing on the CGT protein indicates a hydrogen bond, and the small gray dot on GluP represents a hydrophobic bond.

Supplementary Files

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

- [SupplementaryFilesBangladeshBiofilm.xlsx](#)