JOURNAL OF MEDICAL MICROBIOLOGY

RESEARCH ARTICLE

Wilkinson *et al.*, *Journal of Medical Microbiology* 2023;72:001710

DOI 10.1099/jmm.0.001710

MICROBIOLOGY SOCIETY

OPEN ACCESS

Characterization of a *Helicobacter pylori* strain with high biofilm-forming ability

Daniel Wilkinson[1,2], Lolwah Alsharaf[1,3], Stephen Thompson[1], Andreja Paulin[1], Rhodrick Takor[1], Abed Zaitoun[4], Karen Robinson[5,6], Jonathan Thomas[1], Gareth McVicker[1] and Jody Winter[1,*]

Abstract

Introduction. *Helicobacter pylori* is highly polymorphic, and some strains are much more likely to cause disease than others. Biofilm formation can help bacteria to survive antibiotic treatment, immune attack and other stresses, promoting persistent infection.

Hypothesis/Gap Statement. We hypothesized that *H. pylori* isolates from patients with more severe *H. pylori*-associated disease would be better at forming biofilms than isolates from patients with less severe disease.

Aim. We initially aimed to determine whether or not the biofilm-forming ability of *H. pylori* isolates was associated with disease in the UK-based patients from whom the bacteria were isolated.

Methodology. Biofilm-forming ability of *H. pylori* isolates was determined using a crystal violet assay on glass coverslips. The complete genome sequence of strain 444A was generated by hybrid assembly of Nanopore MinION and Illumina MiSeq data.

Results. Although we found no associations between biofilm-forming ability of *H. pylori* and disease severity in patients, we discovered that strain 444A had particularly high biofilm-forming ability. This strain had been isolated from a patient with gastric ulcer disease and moderate to severe scores for *H. pylori*-induced histopathology. Analysis of the genome of the high biofilm-forming *H. pylori* strain 444A revealed that it possesses numerous biofilm- and virulence-associated genes and a small cryptic plasmid encoding a type II toxin–antitoxin system.

Conclusion. There is substantial variation in biofilm-forming ability in *H. pylori,* but this was not significantly associated with disease severity in our study. We identified and characterized an interesting strain with high biofilm-forming ability, including generation and analysis of the complete genome.

DATA SUMMARY

The metadata for this project can be accessed under the BioProject accession PRJNA929472. The raw sequencing reads can be accessed under SRA accessions SRR17176181 and SRR23286833 generated by the Illumina Miseq and ONT MinION sequencing platforms, respectively. The complete chromosome (444A) and plasmid (pHP444A) assemblies generated in this study can be accessed under GenBank accessions CP117024 and CP117023, respectively.

Received 01 March 2023; Accepted 11 May 2023; Published 09 June 2023

Author affiliations: [1]School of Science and Technology, Nottingham Trent University, Nottingham, UK; [2]School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK; [3]Al-Amiri Hospital, Ministry of Health, Kuwait City, Kuwait; [4]Department of Cellular Pathology, Nottingham University Hospitals NHS Trust, Queen's Medical Centre campus, Nottingham, UK; [5]Biodiscovery Institute, School of Medicine, University of Nottingham, Nottingham, UK; [6]Nottingham Digestive Diseases Biomedical Research Centre, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, UK.

***Correspondence:** Jody Winter, jody.winter@ntu.ac.uk

Keywords: Bacteria; genome; hybrid assembly; plasmid; toxin-antitoxin.

Abbreviations: bp, base pairs; BRIG, BLAST ring image generator and standard scientific units of measure; gDNA, genomic DNA; kbp, kilobase pairs; Mb, megabase.

The metadata for this project can be accessed under the BioProject accession PRJNA929472. The raw sequencing reads can be accessed under SRA accessions SRR17176181 and SRR23286833 generated by the Illumina Miseq and ONT MinION sequencing platforms respectively. The complete chromosome (444A) and plasmid (pHP444A) assemblies generated in this study can be accessed under GenBank accessions CP117024 and CP117023 respectively.

Two supplementary figures and one supplementary table are available with the online version of this article.

Impact Statement

The *Helicobacter pylori* bacterium infects many people globally and it can cause life-threatening conditions including gastric and duodenal ulcers and stomach cancer. Changes in the DNA of *H. pylori* occur frequently, so there is a lot of diversity among strains within this bacterial species. Some strains are more likely to cause ulcers and cancer than others.
This study describes a strain of *H. pylori* that is able to form biofilm (a resistant community of bacteria on a surface, within a slime matrix) much more readily than the other strains we tested. We sequenced the genome of this strain and found that it possesses many genes that are associated with bacterial biofilm-forming ability.

INTRODUCTION

Helicobacter pylori is a Gram-negative, microaerophilic bacterium that infects the stomachs of approximately half of the world's human population. Chronic infection may lead to ulceration of the gastric or duodenal lining and the development of cancer [1]. There is a high level of genetic diversity among *H. pylori* strains, which drives extensive phenotypic diversity, including dramatic variations in strain virulence [2, 3]. *H. pylori* strains that possess the *cag* pathogenicity island (*cag*PAI) enabling injection of the CagA toxin into host cells through a type IV secretion system [4] are more likely to cause disease than *cag*PAI negative strains [2]. Virtually all *H. pylori* strains possess the *vacA* gene, encoding the VacA toxin, but this gene is highly polymorphic. Strains with s1, i1, m1 alleles of *vacA* are more likely to cause disease than strains with *vacA* s2, i2 and/or m2 alleles [5, 6].

Biofilms are communities of bacteria enclosed in an extracellular polymeric substance, attached to a surface [7]. *H. pylori* forms biofilms on the gastric mucosa *in vivo* [8] and on biotic and abiotic surfaces *in vitro* including plastic, glass and cultured gastric epithelial cells [9, 10]. Formation of biofilm is associated with tolerance of, or resistance to, antibiotics including the drugs clarithromycin, amoxicillin, metronidazole, levofloxacin and tetracycline that are commonly used to treat *H. pylori* infections [11–14]. Within *H. pylori* biofilms, expression of efflux pumps (transport proteins with roles in the export of toxic substances, including antibiotics) is elevated [14] and efflux pump mutants have reduced biofilm-forming ability [12].

Although most *H. pylori* clinical isolates can form biofilm in the laboratory, there is substantial variation in biofilm-forming ability between strains [13, 15–19]. In this study, we report the discovery of very high biofilm-forming ability in a UK-derived *H. pylori* strain, 444A. Using Illumina and Nanopore data, we assembled the complete genome sequence and found that strain 444A possesses numerous genes that have previously been reported to influence biofilm formation in *H. pylori* and other bacteria, and a small cryptic plasmid encoding a type II toxin-antitoxin system.

METHODS

Bacterial isolation and culture

H. pylori strains were isolated from gastric biopsies donated by patients attending Queens Medical Centre Hospital, Nottingham, UK for investigation of dyspeptic symptoms by routine upper gastrointestinal tract endoscopy. Written informed consent was obtained, and the study was approved by the Nottingham Research Ethics Committee 2(08/H0408/195). Strains from patients who had taken antibiotics, proton pump inhibitors or >150 mg/day aspirin in the 2 weeks preceding the endoscopy were excluded from this study.

The disease status of patients was assessed visually during endoscopic examination and by histological analysis of gastric biopsy sections. These were stained using hematoxylin and eosin, blinded and scored by an expert gastrointestinal pathologist using the updated Sydney System.

The patient from whom *H. pylori* strain 444A was isolated was a 68-year-old female with gastric ulcer disease and Sydney scores indicating mild atrophy and moderate inflammation, activity and intestinal metaplasia.

The well-characterized laboratory *H. pylori* strains 60 190 and Tx30a [20] were also used in this study.

Strains were stored in isosensitest broth (Oxoid, UK) supplemented with 15% (v/v) glycerol at −80 °C until use, then routinely cultured on blood agar base no. 2 (Oxoid, UK) supplemented with 7.5% (v/v) defibrinated horse blood (Fisher Scientific, UK) at 37 °C under microaerobic conditions (5% O_2, 10% CO_2, 85% N_2) at 80% humidity. Strains were minimally passaged prior to laboratory characterization and genome sequencing.

Biofilm assays

The biofilm-forming ability of each strain was measured using a method adapted from Yonezawa *et al.* [16]. Bacteria were suspended in brain heart infusion broth (Fisher Scientific, UK) supplemented with 5% (v/v) heat-inactivated foetal calf

serum (Sigma-Aldrich, UK) and adjusted to an optical density ($OD_{600\ nm}$) of 0.3. Bacterial suspensions were pipetted into 12-well plates (sterile, non-pyrogenic, non-cytotoxic; Sarstedt, Germany) with 2 ml per well, then sterile 22×22 mm glass cover slips were added to each well. Plates were incubated under microaerobic, high-humidity conditions at 37 °C for 7 days to allow bacterial biofilms to form on the glass coverslips at the air–liquid interface. At the end of the incubation period, glass coverslips were removed for biofilm staining and samples of broth from each well were plated to blood agar plates and incubated overnight under microaerobic conditions to confirm the absence of faster-growing contaminant bacteria.

Cover slips were washed with phosphate buffered saline to remove non-adherent bacteria, then allowed to air dry for 1–2 h. Biofilms were then stained using crystal violet for 30 s, washed with water, and allowed to air dry for a further 1–2 h. Visual inspection of the biofilm quantity was possible at this point. To quantify the crystal violet dye taken up by the biofilm, each cover slip was then placed into a well of a 12-well plate and washed extensively with 1 ml of ethanol to release the dye from the biofilm. The absorbance at 594 nm was measured for triplicate 200 µl samples from each 1 ml of ethanol.

Biofilm-forming ability ($OD_{594\ nm}$) of each strain, and associations with disease, were analysed as follows. Biofilm-forming data for the *H. pylori* strain derived from each patient were compared according to the disease status of the patient [presence versus absence of ulcer disease; Sydney scores 0–1 (absent-low) versus 2–3 (moderate-severe) for inflammation, activity and intestinal metaplasia] using unpaired *t*-tests in Graphpad Prism v9.3.1. For each strain, technical replicates of at least duplicate were repeated independently on at least two separate occasions.

Genome sequencing and assembly

Helicobacter pylori genomic DNA (gDNA) was extracted using the QIAmp DNA Mini Kit (QIAGEN, Netherlands) following the manufacturer's instructions with an extended Proteinase K digestion stage of 20 h. gDNA was eluted in molecular grade water, quality checked on the NanoDrop2000 and dsDNA concentration determined on the Qubit Fluorometer 4 (ThermoFisher Scientific).

Short-read sequencing libraries were generated using the Nextera XT DNA library preparation kit (Illumina, Cambridge, UK) following manufacturer's instructions except for a higher starting input DNA amount of 1.5 ng. The library was loaded onto a v2 chemistry sequencing cartridge that was run at 2×250 bp read lengths on the Illumina MiSeq platform (Illumina, Cambridge, UK).

Nanopore sequencing library was prepared using the ligation sequencing kit SQK-LSK108 (Oxford Nanopore Technologies, UK) and native barcoding (EXP-NBD103) and loaded onto a MinION flowcell (R9.4.1).

An initial long-read *de novo* assembly was constructed as follows. FAST5 raw output data from long-read sequencing was base called through Guppy v6.1.7 (ONT, UK) employing the 'super high accuracy' model (dna_r9.4.1_450bps_sup configuration). Porechop v0.2.4 [21] was used to demultiplex FASTQ reads and trim adaptors. A *de novo* long-read assembly was created with Flye v2.9 [22] specifying ONT high-quality read input, an estimated genome size of 1.63 Mb and an initial disjointing assembly coverage of 350×. Circulator v1.5.5 [23] was run to circularize both the chromosomal and plasmid sequences. The resulting assembly was assessed for completeness through Bandage version 0.8.1 [24] and subsequently polished once using the ONT long-reads with Medaka v1.5.0 (ONT, UK) utilizing the consensus programme and running with default settings.

Illumina short-reads were used to polish the long-read medaka polished *de novo* assembly to create a hybrid long- and short-read *de novo* assembly. The paired-end short reads were trimmed with fastp [25] for quality (Phred 30) and minimum read length of 30 bp. The assembly was polished sequentially with long then curated short reads through Racon v1.4.20 [26] with default settings followed by one iteration of Pilon version 1.24 [27]. The resulting Pilon polished assembly was the completed Illumina short- and ONT long-read hybrid assembly used in this study.

Sequencing read depth was determined using Mosdepth [28] with default settings.

Chromosome annotation

The hybrid genome assembly was annotated by Prokka version 1.14.6 [29] guided by a manually curated version of the *H. pylori* strain 26695 (NC_000915.1).

Plasmid analysis

To estimate the copy number of the 6.1 kb plasmid, Unicycler version 0.5.0 [30] was run in hybrid assembly mode with both Illumina short- and ONT long-reads as input and executed with default settings.

The plasmid was manually annotated by identifying the open reading frames using SnapGene software (from Dotmatics; available at snapgene.com) and the identified features were extracted as amino acids. A custom annotation file was generated by BLASTP of the identified open reading frames to identify the genes contained in the plasmid.

Core-genome analysis

The core genome was determined by analysing a curated set of 241 publicly available *H. pylori* genomes downloaded from BIGSdb [31]. Only isolates that were sequenced on the Illumina platform were included, and QUAST [32] was used to identify and filter out genomes that contained 'N' bases and genomes with >60 contigs, in order to retain only isolates with higher-quality genome assemblies. This curated isolate list of 241 genome assemblies was passed to the genome comparator plugin on BIGSdb and the core genome was identified by comparing all genomes to the 26695 *H. pylori* reference genome (NC_000915.1) with BLASTN settings at 70% minimum identity, 90% minimum alignment and 20 bp word size. A gene was determined to be core if it was present in 99% of isolates. These settings were used to account for genes located at contig boundaries and to allow for gene sequence variation between strains [33]. The list of core genes identified using this method is in Table S1, available with the online version of this article.

Accessory genes that did not have homology to any genes in the 26695 *H. pylori* reference genome, such as HPG27_1233 (Table 1) were identified as absent or present in the 444A genome using BLASTN with the same settings as described above.

Mapping biofilm associated genes on the chromosome

The BLAST ring image generator (BRIG) version 0.95 [34] was used to draw and annotate the locations of biofilm related genes on the 444A chromosome (Table 1). Biofilm associated genes that were determined to be part of the core genome were coloured differently to biofilm genes found in the accessory genome (Fig. 5).

Literature review

A list of all genes reported to be associated with biofilm formation in *H. pylori* was compiled from the available peer-reviewed literature on NCBI PubMed by 15 December 2022. In total, 224 papers were identified for review using the search term 'pylori AND biofilm' in full-text searches and follow up citation searching. Inclusion criteria: peer-reviewed original research articles reporting *H. pylori* genes present in high biofilm-forming strains while absent from poor biofilm formers, and/or where the mutation of the gene causes a decrease in biofilm formation. Exclusions: papers describing genes or proteins upregulated in biofilm compared to planktonic cells, without subsequent use of mutagenesis to confirm their role in biofilm formation, and papers describing genes that functionally suppress biofilm formation.

RESULTS

Strain 444A is a very high biofilm-former compared with other *H. pylori* strains

The ability of strain 444A to form extensive biofilm was immediately obvious upon visual inspection of the crystal violet-stained glass coverslips (Fig. 1) and formal quantitative analysis showed that strain 444A formed significantly more biofilm than any of the other 14 strains tested in this study ($P<0.0001$, one way ANOVA with Dunnett post-hoc tests) (Fig. 2).

Biofilm-forming ability does not significantly associate with disease

Although strain 444A was isolated from a patient with gastric ulcer disease and moderate Sydney scores for inflammation, activity and intestinal metaplasia, no significant associations between bacterial biofilm-forming ability and markers of disease in the patients were detected ($P>0.05$, unpaired *t*-tests) (Fig. 3).

The genome of strain 444A encodes numerous biofilm and virulence-associated features

The complete genome of strain 444A was produced by hybrid assembly of short-read Illumina MiSeq and long-read MinION nanopore data. Assembly statistics and key characteristics of the genome are presented in Table 2. The complete genome comprised a 1.62 Mb chromosome and a 6.1 kb plasmid (Fig. S1). The ONT long-read sequencing output generated a mean sequencing coverage per base of 34X (chromosome) and 127X (plasmid) while the Illumina short-read sequencing output generated a mean coverage per base of 95X (chromosome) and 735X (plasmid). The hybrid assembly approach generated a mean coverage of 129X for the chromosome and 862X for the plasmid.

The chromosome of strain 444A encodes a complete *cagPAI* with Western-type *cagA* EPIYA motif ABC, and the *vacA* genotype is s1i1m1.

The plasmid, pHP444A, encodes a putative YafQ-family toxin and its cognate antitoxin, and five further genes. Using BLASTP to identify homologous proteins from other bacterial species, and with reference to the *H. pylori* literature, we assigned likely identities and functions to these genes. The annotated plasmid is presented in Fig. 4. The pHP444A plasmid is similar to previously reported *H. pylori* plasmids pHPG27 (87% coverage, 90% nucleotide identity) and pHel4 (also 87% coverage, 90% nucleotide identity) and contains typical *repB* and *mob* genes suggesting that it is mobilizable. It has lower GC content than the chromosome (35.83% versus 39.00%).

Table 1. Biofilm-associated genes in the 444A chromosome

Gene	Locus tag from *H. pylori* reference 26695	Presence in 444A	Core/accessory	Source
acsA	HP1045	+	Accessory	[17]
acxA	HP0695	+	Accessory	[17]
acxB	HP0696	+	Accessory	[17]
alpB	HP0913	+	Accessory	[58]
cagPAI genes	HP0520; HP0522-HP0532; HP0534-HP0547	+	Accessory	[10, 18, 49]
faaA	HP0609-HP0610	+	Accessory	[17]
fliA	HP1032	+	Core	[17]
fliK	HP0906	+	Accessory	[17, 48]
*fucT1***	HP0379	+	Accessory	[18]
*fucT2***	HP0651	+	Accessory	[18]
gluP	HP1174	+	Core	[59]
hcpE	HP0235	+	Core	[17]
homB		–	Accessory	[50]
homD	HP1453	+	Accessory	[18]
HP0168	HP0168	+	Core	[17]
HP0497	HP0497	+	Core	[12]
HP0939	HP0939	+	Core	[12]
HPG27_166	N/A	–	Accessory	[17]
HPG27_526	HP0567	+	Accessory	[17]
HPG27_715	HP0759	+	Core	[17]
HPG27_1066	HP1121	+	Core	[17]
HPG27_1233	N/A	+	Accessory	[17]
hydE	HP0635	+	Core	[17]
hypF	HP0048	+	Core	[17]
jhp_1117	HP1192	+	Accessory	[18]
K747_06625	N/A	–	Accessory	[18]
K747_09130	HP0120	++	Accessory	[18]
K747_10375	HP0338	+	Accessory	[18]
kefB	HP0471	+	Core	[12]
*luxS**	HP0105	+	Core	[10]
motB	HP0816	+	Core	[17]
napA	HP0243	+	Core	[60, 61]
spoT	HP0775	+	Core	[59, 60]
uppS	HP1221	+	Core	[17]
vlpC	HP0922	+	Accessory	[17]

A list of genes positively associated with biofilm formation in *H. pylori* was generated using the available literature. The presence (+ and green highlight) or absence (-) of each of these genes in the 444A genome is indicated in the table with references to studies that have associated each gene with biofilm formation. ++Indicates that two copies of the gene are present. *Indicates mixed reports about the associations of the gene with biofilm formation. Loss of *luxS* was associated with increased biofilm in two reports [10, 62] but decreased biofilm in another [49]. **Genes in strain 444A are homologues of the *fucT* gene reported in [18] as associated with high biofilm-forming ability.

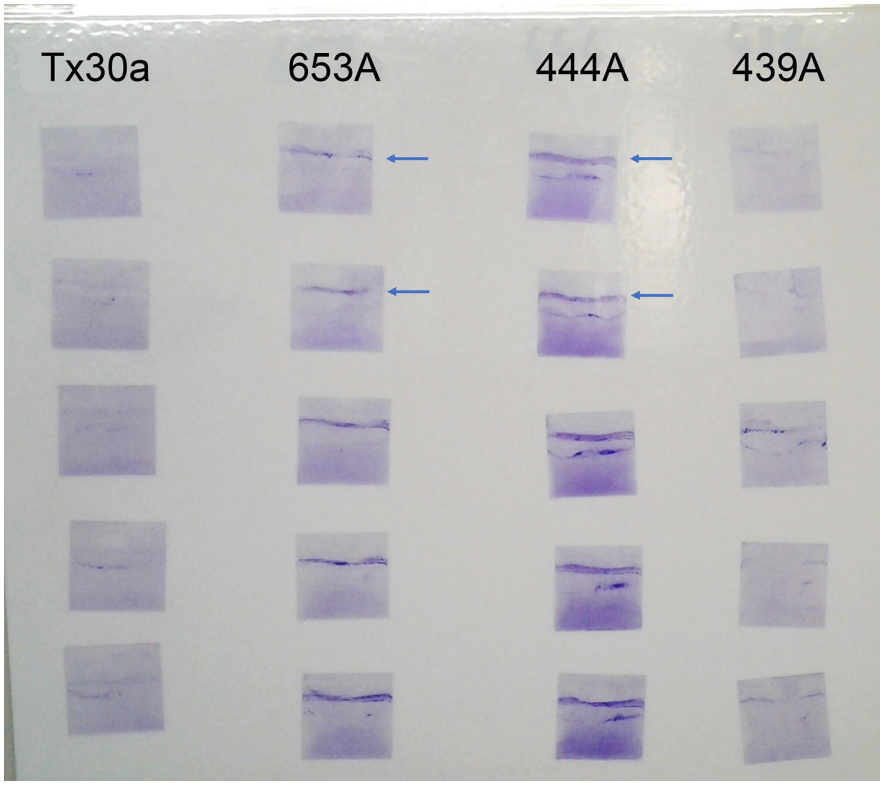

Fig. 1. Crystal violet-stained *H. pylori* biofilms on glass coverslips. Strain 444A formed more substantial biofilms on glass coverslips at the air–liquid interface, and below it, compared with other strains that were tested. Strains shown here, arranged in columns with five replicate cover slips per column, are (from left to right): Tx30a, 653A, 444A, 439A. The top of each cover slip on this image was in the air above the liquid, and the bottom of each cover slip was submerged in the liquid, within each well in the biofilm assay. Examples of biofilm formation at the air–liquid interface are indicated by the arrows.

Numerous genes that have been associated with biofilm formation are present in the genome of strain 444A (Table 1 and Fig. 5).

Identifying core and accessory biofilm-associated genes

There were 1109 genes identified as *H. pylori* core genes (Table S1) within this study, which matched closely to the 1111 core genes previously estimated by Gressmann *et al.* [35]. Sixteen of the biofilm-associated genes in strain 444A were determined to be part of the core *H. pylori* genome while the remaining 40 biofilm-associated genes (including 26 *cag*PAI genes) were accessory (Table 1 and Fig. 5).

DISCUSSION

At the time of writing, there are 370 complete *H. pylori* genomes publicly available in the NCBI database. The first complete *H. pylori* genome was for the reference strain 26695 [36]. This and many subsequent complete genomes were generated using labour intensive and/or expensive methodologies. The vast majority of the recently uploaded complete *H. pylori* genomes in the NCBI nucleotide database originate from Japan, China and the Americas, and many have used the SMRT/PacBio platform (for example, [37, 38]) or mapped short-read data to a reference genome (for example, [39]). However, there are drawbacks to both approaches; use of a reference genome may bias the assembly, since some regions may suffer from low coverage [40], and gene duplications or rearrangements may lead to errors in gene order within the assembled genome, while PacBio sequencing is still relatively expensive [41].

Nanopore–Illumina hybrid assemblies utilize the long reads of Oxford Nanopore data to assemble across regions that commonly produce contig breaks in short-read-only assemblies, such as duplicate genes and repeat regions, thus producing complete genomes; this long-read assembly can subsequently be used as a scaffold to which highly accurate short-read data may be mapped, thus correcting the traditional high error-rate from Nanopore data [42], particularly amongst homopolymer tracts. Such hybrid assembly approaches are cheaper than PacBio sequencing [41], and in addition to genome content provided by short-read assemblies, reveal the genome structure as well. The Nanopore–Illumina hybrid assembly approach has recently been successfully

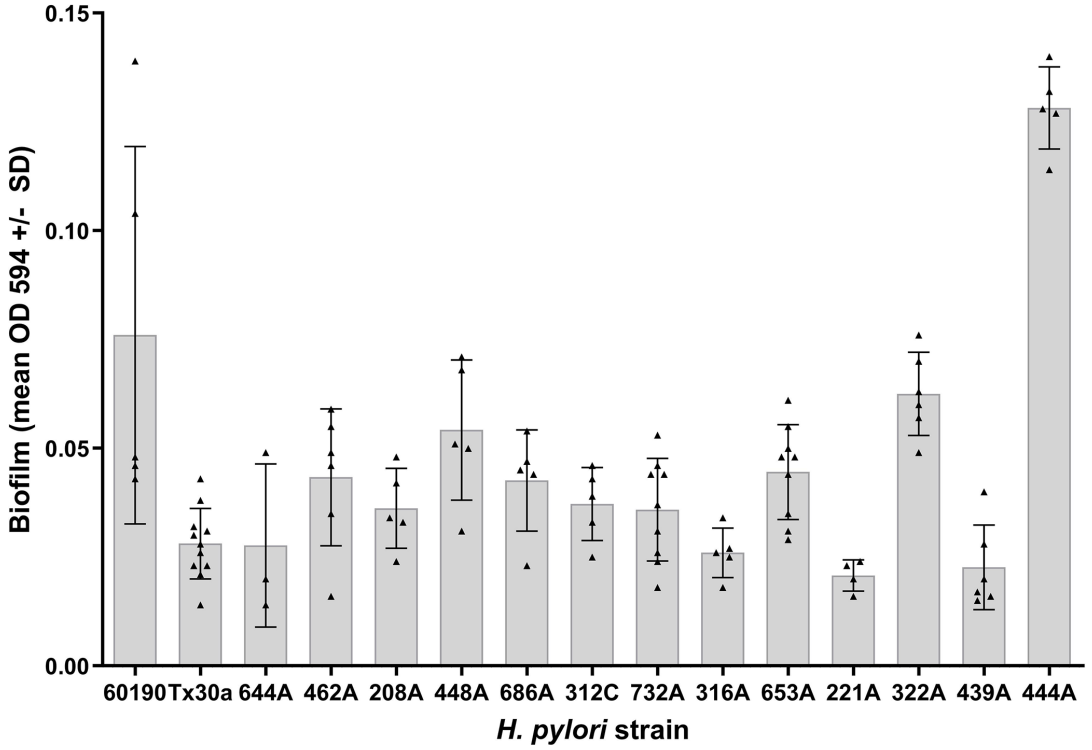

Fig. 2. Clinical isolate 444A forms significantly more biofilm than other *H. pylori* strains. A crystal violet assay was used to quantify biofilm formation on glass cover slips for 15 *H. pylori* strains. Strain 444A formed significantly more biofilm than any of the other strains tested ($P<0.0001$, one way ANOVA with Dunnett post-hoc tests). Each data point represents a separate cover slip, with assays for each strain set up in duplicate or triplicate on at least two separate occasions.

applied to *H. pylori* isolates from Vietnam [43], Japan [44], Kenya [45] and China [46]. Here we present, to our knowledge, the first Nanopore–Illumina hybrid assembly of a complete genome from a European *H. pylori* isolate.

Strain 444A was isolated in Nottingham, UK, from a patient with gastric ulcer disease and it was found to form much more extensive biofilm than other strains from the Nottingham strain collection. We did not find any significant associations between biofilm-forming ability of the bacterial isolate and markers of *H. pylori*-associated disease in the patients in our study. However, we cannot conclude from this that biofilm-forming ability is unimportant in *H. pylori* disease pathogenesis because our sample size was relatively small. We used glass coverslips and crystal violet staining to measure biofilm-forming ability at the air–liquid interface in brain heart infusion broth supplemented with 5% serum. While this technique is useful for initial screening of bacterial ability to adhere and form early-stage biofilms, more comprehensive studies using more biologically relevant model systems and larger sample sets are needed to confirm whether or not strain 444A has enhanced ability to form mature biofilms *in vivo,* compared to other strains.

Strain 444A possesses numerous biofilm-associated genes

Numerous biofilm-associated genes are present in the 444A genome, and many of these are accessory genes (not present in all *H. pylori* strains). Strain 444A has a complete *cag* pathogenicity island and the *cagPAI* genes have been linked to biofilm-forming ability in numerous studies. While loss of *cagE* caused increased biofilm formation in one early study [10], subsequent studies showed that CagA protein [47] and expression of various *cagPAI* genes [17, 48] are upregulated in biofilm compared to planktonic cells. A comparative genomics approach revealed that the *cagPAI* was present in 8/12 good biofilm formers, 7/12 moderate biofilm formers and was absent from all eight poor biofilm formers [18]. Mutating *cagA* or the *cagPAI* reduced biofilm-forming ability in strains 26695 [18] and mutating *cagY* reduced biofilm-forming ability in the J99 background [49].

Significant associations have also been previously identified between the presence of the *fucT* alpha- [1, 3] fucosyltransferase and *homD* genes and strong biofilm formation [18]. Both of these genes are present in strain 444A, although *homB*, which has been shown to induce a hyper-biofilm phenotype in strain G27 [50] is not present in 444A.

Several flagella- and motility-associated genes have been linked to biofilm-forming ability in *H. pylori* and flagellar filaments are a major component of the *H. pylori* biofilm extracellular matrix [15, 17, 48]. Flagellar genes are highly expressed in biofilms, and mutations in flagella- and motility-associated genes reduce biofilm formation in *H. pylori* [17]. Although we have not yet

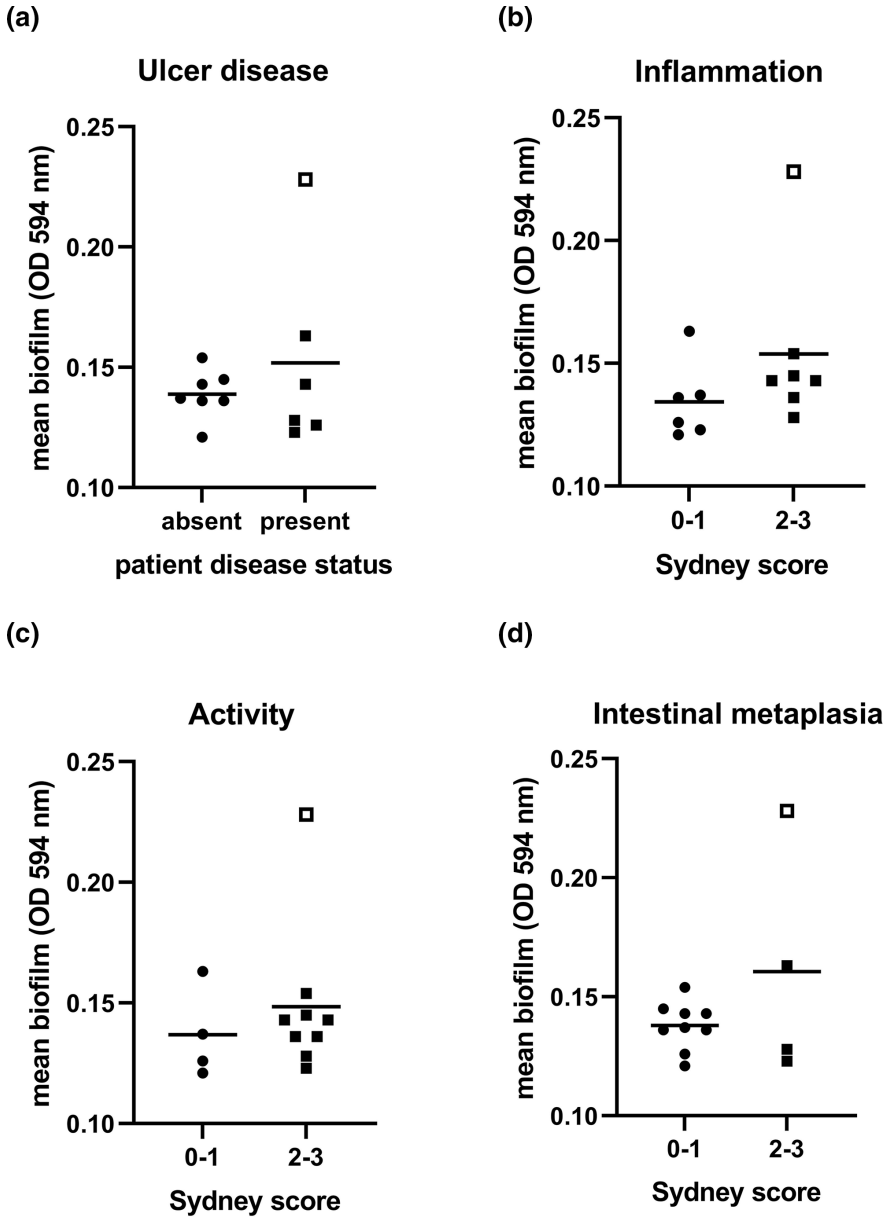

Fig. 3. *H. pylori* biofilm-forming ability was not significantly associated with disease. The mean OD_{594} for each *H. pylori* strain was determined and strains were grouped according to the disease status of the patients they were isolated from. No significant associations were seen between *H. pylori* biofilm-forming ability and the presence of ulcers in the patients (a), or between biofilm-forming ability and the presence of moderate to severe inflammation (b), activity (c) or intestinal metaplasia (d) (P>0.05, unpaired t-tests). The high biofilm-forming strain 444A is indicated on each chart by an open square and all other strains are represented by closed circles.

completed a comprehensive analysis of all of the strains in our collection, our preliminary data indicates that strain 444A is moderately motile (Fig. S2). The 444A genome includes all of the motility-associated genes that have been linked with biofilm-forming ability in the literature and the motility and biofilm-forming abilities of strain 444A are consistent with its genome content.

While the genome sequence is a useful indicator of the presence/absence of biofilm-associated genes, further studies will be needed to confirm the expression levels of each gene and the extent to which they contribute to the high biofilm-forming phenotype of strain 444A.

The 444A plasmid encodes a toxin-antitoxin system

The complete 444A genome comprises a 1.6 Mb chromosome and a 6.1 kb plasmid with an estimated copy number of 16–17 plasmid copies per cell. The pHP444A plasmid is similar to previously reported plasmids pHPG27 [51] and pHel4 [52]. Interestingly,

Table 2. Strain 444A hybrid genome assembly statistics

Assembly	444A_hybrid	444A_chromosome_only	444A_plasmid_only
No. of contigs	2	1	1
Largest contig (bp)	1621210	1621210	6106
Total length (bp)	1627316	1621210	6106
GC (%)	38.98	39.00	35.83
N50 (bp)	1621210	1621210	6106

the pHP444A plasmid encodes a putative YafQ family toxin with homology to HP0894 from the type II HP0894-HP0895 toxin-antitoxin (TA) system that has previously been described on the chromosome of strain 26695 and extensively characterized [53]. The toxin encoded on the pHP444A plasmid has 59.1% amino acid identity and 65.9% similarity to 26695 chromosomal HP0894, but the cognate antitoxin has very poor homology to HP0895 (21.3% amino acid identity, 31.9% similarity). This is not unexpected, since antitoxins tend to be intrinsically disordered [54] with highly diverse sequences [55].

HP0894 has also previously been found on small cryptic plasmids of *H. pylori* including pHel4 and pHel5 [52]. pHel4 HP0894 has 55.7% amino acid identity, 68.2% similarity to pHP444A HP0894, while pHel5 HP0894 has 53.4% amino acid identity, 68.2% similarity to pHP444A HP0894. The HP0894 toxin has RNase activity which is strongly inhibited upon binding of the HP0895 antitoxin [53, 56]. Expression of several type II TA systems including HP0894-HP0895 has been shown to be elevated in *H. pylori*

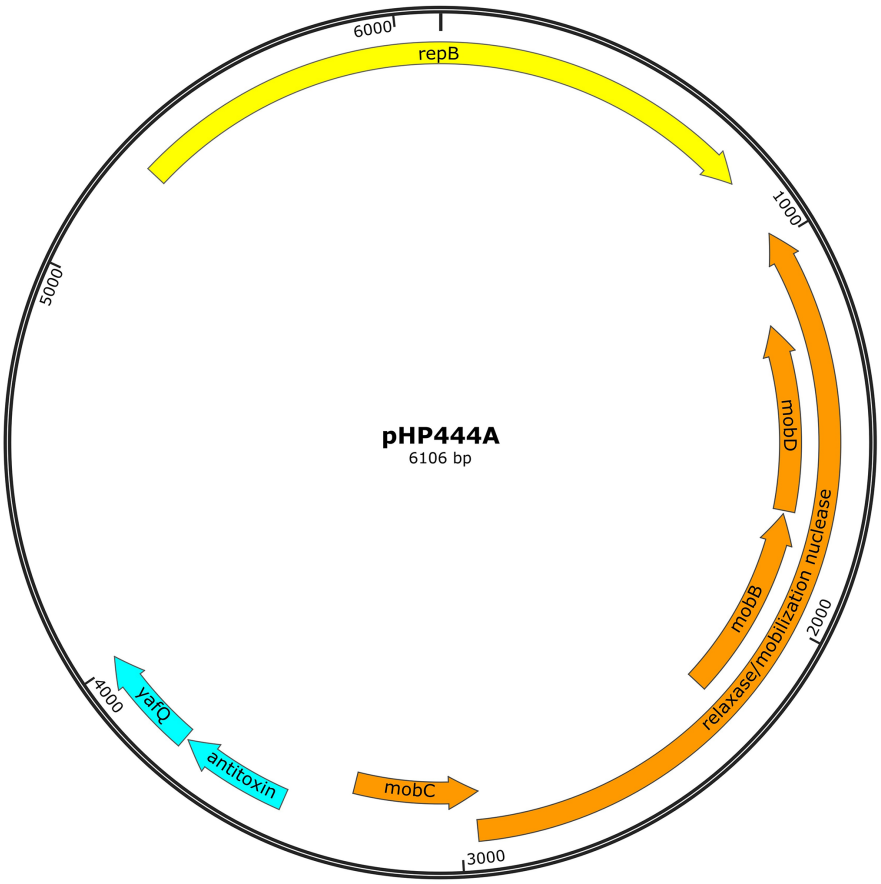

Fig. 4. Annotated plasmid from strain 444A. Genes present on the pHP444A plasmid were manually annotated after identification of encoded proteins with high homology using BLASTP. The plasmid encodes a toxin-antitoxin system (comprising a putative YafQ-family toxin with homology to HP0894, and its cognate antitoxin) shown in blue, replication associated genes shown in yellow and mobilization associated genes shown in orange. Unicycler output indicated that this is a moderate copy number plasmid with 16–17 plasmid copies per chromosome.

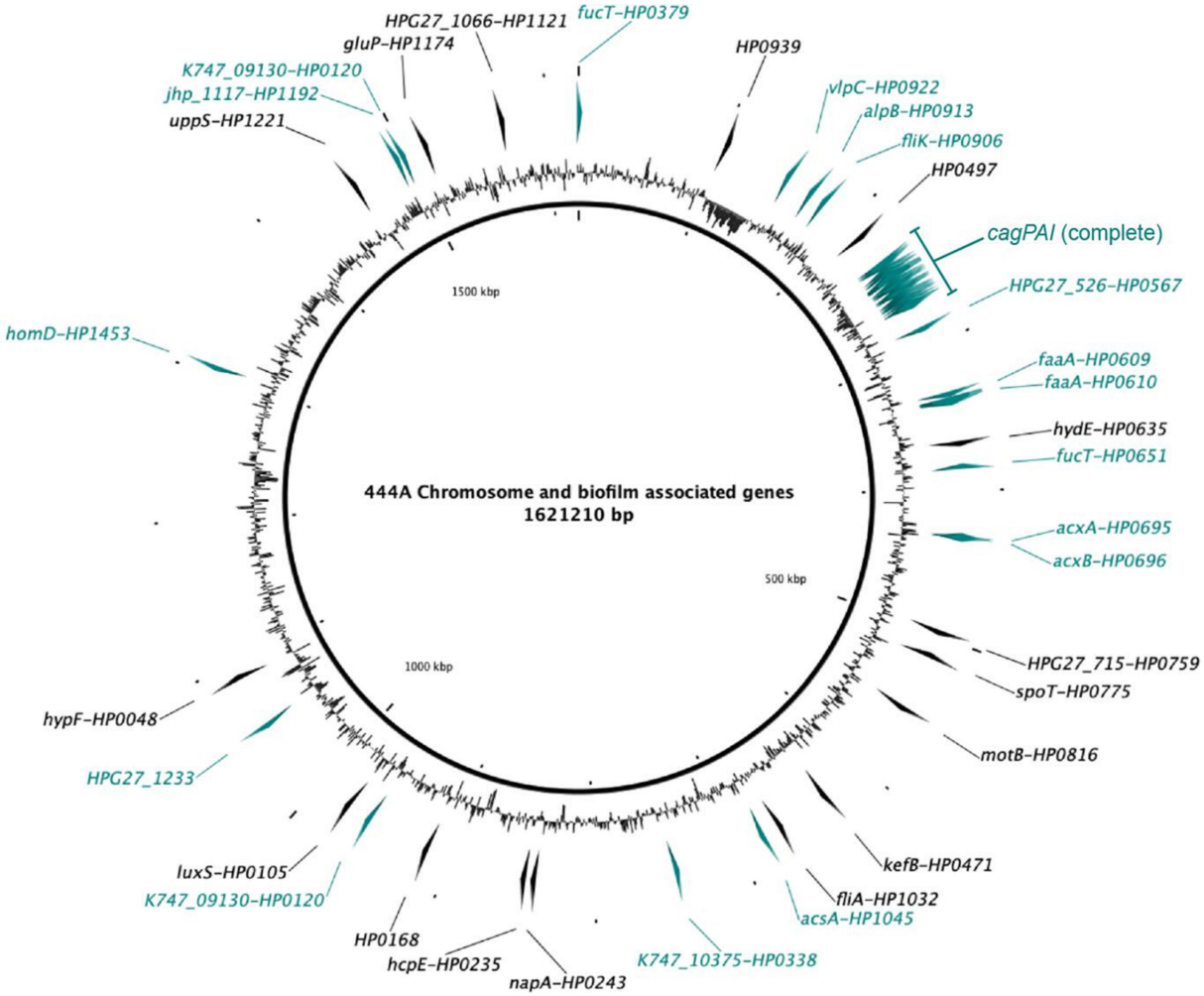

Fig. 5. Biofilm-associated genes encoded on the genome of strain 444A. The genome of strain 444A encodes numerous biofilm-associated genes. The positions, lengths and directions of these on the chromosome are shown as either black (core) or teal (accessory) arrows, labelled with the gene names. Further information about these genes is presented in Table 1. The middle black ring indicates GC content.

biofilms [57]. This raises the possibility that the plasmid borne putative YafQ family TA system of strain 444A could be involved in its high biofilm-forming ability, but further experimental work will be required to verify this.

In conclusion, we present the complete genome sequence of the high biofilm-forming *H. pylori* strain 444A including a small novel plasmid pHP444A encoding a type II toxin–antitoxin system.

Funding information
Daniel Wilkinson was supported by a Vice-Chancellor's PhD Scholarship and the School of Science and Technology at Nottingham Trent University. Lolwah Alsharaf contributed to this work while studying for postgraduate qualifications, sponsored by the Government of the State of Kuwait, represented by the Kuwait Cultural Office of the Embassy of the State of Kuwait in London. Karen Robinson is supported by the National Institute for Health Research (NIHR) through the Biomedical Research Centre at Nottingham University Hospitals NHS Trust and The University of Nottingham. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Acknowledgements
We thank Professor John Atherton and his team at The University of Nottingham for access to *H. pylori* clinical isolates and associated data, particularly Mrs Joanne Rhead and Ms Melanie Lingaya who isolated *H. pylori* from human gastric biopsies. The authors gratefully acknowledge the Technical staff within the Biosciences Department at Nottingham Trent University for technical support in this work.

Author contributions
Conceptualization and supervision: J.W., K.R. Investigation: D.W., L.A., S.T., A.P., R.T., A.Z., J.T., G.M.V. Data curation: D.W., J.T. Visualization: J.W., D.W., G.M.V. Writing – original draft: J.W. Writing – review and editing: K.R., D.W., J.T., G.M.V.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
H. pylori strains were isolated from gastric biopsies donated by patients attending Queens Medical Centre Hospital, Nottingham, UK with written informed consent and ethical approval from the Nottingham Research Ethics Committee 2(08/H0408/195).

Consent to publish
Not applicable.

References

1. Atherton JC, Blaser MJ. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. *J Clin Invest* 2009;119:2475–2487.
2. Blaser MJ, Berg DE. Helicobacter pylori genetic diversity and risk of human disease. *J Clin Invest* 2001;107:767–773.
3. Hanafiah A, Lopes BS. Genetic diversity and virulence characteristics of *Helicobacter pylori* isolates in different human ethnic groups. *Infect Genet Evol* 2020;78:104135.
4. Odenbreit S, Püls J, Sedlmaier B, Gerland E, Fischer W, *et al*. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* 2000;287:1497–1500.
5. Atherton JC, Cao P, Peek RM, Tummuru MKR, Blaser MJ, *et al*. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. *J Biol Chemist* 1995;270:17771–17777.
6. Rhead JL, Letley DP, Mohammadi M, Hussein N, Mohagheghi MA, *et al*. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* 2007;133:926–936.
7. Donlan RM. Biofilms: microbial life on surfaces. *Emerg Infect Dis* 2002;8:881–890.
8. Carron M, Tran V, Sugawa C, Coticchia J. Identification of *Helicobacter pylori* biofilms in human gastric mucosa. *J Gastrointest Surg* 2006;10:712–717.
9. Windham IH, Servetas SL, Whitmire JM, Pletzer D, Hancock REW, *et al*. *Helicobacter pylori* biofilm formation is differentially affected by common culture conditions, and proteins play a central role in the biofilm matrix. *Appl Environ Microbiol* 2018;84:e00391-18.
10. Cole SP, Harwood J, Lee R, She R, Guiney DG. Characterization of monospecies biofilm formation by *Helicobacter pylori*. *J Bacteriol* 2004;186:3124–3132.
11. Hathroubi S, Zerebinski J, Clarke A, Ottemann KM. *Helicobacter pylori* biofilm confers antibiotic tolerance in part via A protein-dependent mechanism. *Antibiotics* 2020;9:355.
12. Cai Y, Wang C, Chen Z, Xu Z, Li H, *et al*. Transporters HP0939, HP0497, and HP0471 participate in intrinsic multidrug resistance and biofilm formation in *Helicobacter pylori* by enhancing drug efflux. *Helicobacter* 2020;25:e12715.
13. Fauzia KA, Miftahussurur M, Syam AF, Waskito LA, Doohan D, *et al*. Biofilm formation and antibiotic resistance phenotype of *Helicobacter pylori* clinical isolates. *Toxins* 2020;12:473.
14. Yonezawa H, Osaki T, Hojo F, Kamiya S. Effect of *Helicobacter pylori* biofilm formation on susceptibility to amoxicillin, metronidazole and clarithromycin. *Microb Pathog* 2019;132:100–108.
15. Krzyżek P, Migdał P, Grande R, Gościniak G. Biofilm formation of *Helicobacter pylori* in both static and microfluidic conditions is associated with resistance to clarithromycin. *Front Cell Infect Microbiol* 2022;12:868905.
16. Yonezawa H, Osaki T, Kurata S, Fukuda M, Kawakami H, *et al*. Outer membrane vesicles of *Helicobacter pylori* TK1402 are involved in biofilm formation. *BMC Microbiol* 2009;9:197.
17. Hathroubi S, Hu S, Ottemann KM. Genetic requirements and transcriptomics of *Helicobacter pylori* biofilm formation on abiotic and biotic surfaces. *NPJ Biofilms Microbiomes* 2020;6:56.
18. Wong EHJ, Ng CG, Chua EG, Tay ACY, Peters F, *et al*. Comparative genomics revealed multiple *Helicobacter pylori* genes associated with biofilm formation in vitro. *PLoS One* 2016;11:e0166835.
19. Attaran B, Falsafi T. Identification of factors associated with biofilm formation ability in the clinical isolates of *Helicobacter pylori*. *Iran J Biotechnol* 2017;15:58–66.
20. Leunk RD, Johnson PT, David BC, Kraft WG, Morgan DR. Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. *J Med Microbiol* 1988;26:93–99.
21. Porechop RR. Porechop; 2017. https://github.com/rrwick/Porechop
22. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, *et al*. Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci* 2016;113:E8396–E8405.
23. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, *et al*. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015;16:294.
24. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 2015;31:3350–3352.
25. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
26. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;27:737–746.
27. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, *et al*. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9:e112963.
28. Pedersen BS, Quinlan AR, Hancock J. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 2018;34:867–868.
29. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
30. Wick RR, Judd LM, Gorrie CL, Holt KE, Phillippy AM. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
31. Jolley KA, Maiden MCJ. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;11:595.
32. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
33. van Tonder AJ, Mistry S, Bray JE, Hill DMC, Cody AJ, *et al*. Defining the estimated core genome of bacterial populations using a Bayesian decision model. *PLoS Comput Biol* 2014;10:e1003788.
34. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.
35. Gressmann H, Linz B, Ghai R, Pleissner K-P, Schlapbach R, *et al*. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet* 2005;1:e43.
36. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, *et al*. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 1997;388:539–547.
37. Mannion A, Dzink-Fox J, Shen Z, Piazuelo MB, Wilson KT, *et al*. *Helicobacter pylori* antimicrobial resistance and gene variants in high- and low-gastric-cancer-risk populations. *J Clin Microbiol* 2021;59:e03203-20.
38. Gutiérrez-Escobar AJ, Velapatiño B, Borda V, Rabkin CS, Tarazona-Santos E, *et al*. Identification of new *Helicobacter pylori*

subpopulations in Native Americans and Mestizos From Peru. *Front Microbiol* 2020;11:601839.

39. Binh TT, Suzuki R, Kwon DH, Yamaoka Y. Complete genome sequence of a metronidazole-resistant *Helicobacter pylori* strain. *Genome Announc* 2015;3:e00051-15.
40. Valiente-Mullor C, Beamud B, Ansari I, Francés-Cuesta C, García-González N, *et al*. One is not enough: on the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLoS Comput Biol* 2021;17:e1008678.
41. Tedersoo L, Albertsen M, Anslan S, Callahan B. Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Appl Environ Microbiol* 2021;87:e0062621.
42. Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. *PLoS One* 2021;16:e0257521.
43. Nguyen-Hoang TP, Nguyen Hoa T, Nguyen TH, Baker S, Rahman M, *et al*. Complete genome sequence of *Helicobacter pylori* strain GD63, isolated from a vietnamese patient with a gastric ulcer. *Microbiol Resour Announc* 2019;8:e01412-18.
44. Takahashi M, Matsumoto Y, Ujihara T, Maeda H, Hanazaki K, *et al*. Complete genome sequence of *Helicobacter pylori* strain 3401, a suitable host for bacteriophages KHP30 and KHP40. *Microbiol Resour Announc* 2021;10:e0064721.
45. Mwangi C, Njoroge S, Tshibangu-Kabamba E, Moloo Z, Rajula A, *et al*. Whole genome sequencing reveals virulence potentials of *Helicobacter pylori* strain KE21 isolated from a Kenyan patient with gastric signet ring cell carcinoma. *Toxins* 2020;12:556.
46. Jin F, Yang H, Rasko D. Complete genome sequence of *Helicobacter pylori* strain 3192, isolated from a Chinese patient with chronic nonatrophic gastritis. *Microbiol Resour Announc* 2022;11:e0038222.
47. Shao C, Sun Y, Wang N, Yu H, Zhou Y, *et al*. Changes of proteome components of *Helicobacter pylori* biofilms induced by serum starvation. *Mol Med Rep* 2013;8:1761–1766.
48. Hathroubi S, Zerebinski J, Ottemann KM. *Helicobacter pylori* biofilm involves a multigene stress-biased response, including a structural role for *Flagella*. *mBio* 2018;9:e01973-18.
49. Wong EHJ, Ng CG, Goh KL, Vadivelu J, Ho B, *et al*. Metabolomic analysis of low and high biofilm-forming *Helicobacter pylori* strains. *Sci Rep* 2018;8:1409.
50. Servetas SL, Doster RS, Kim A, Windham IH, Cha J-H, *et al*. ArsRS-dependent regulation of homB contributes to *Helicobacter pylori* biofilm formation. *Front Microbiol* 2018;9.
51. Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, *et al*. The complete genome sequence of *Helicobacter pylori* strain G27. *J Bacteriol* 2009;191:447–448.
52. Hofreuter D, Haas R. Characterization of two cryptic *Helicobacter pylori* plasmids: a putative source for horizontal gene transfer and gene shuffling. *J Bacteriol* 2002;184:2755–2766.
53. Han K-D, Matsuura A, Ahn H-C, Kwon A-R, Min Y-H, *et al*. Functional identification of toxin-antitoxin molecules from *Helicobacter pylori* 26695 and structural elucidation of the molecular interactions. *J Biol Chem* 2011;286:4842–4853.
54. Loris R, Garcia-Pino A. Disorder- and dynamics-based regulatory mechanisms in toxin-antitoxin modules. *Chem Rev* 2014;114:6933–6947.
55. Leplae R, Geeraerts D, Hallez R, Guglielmini J, Drèze P, *et al*. Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res* 2011;39:5513–5525.
56. Pathak C, Im H, Yang Y-J, Yoon H-J, Kim H-M, *et al*. Crystal structure of apo and copper bound HP0894 toxin from *Helicobacter pylori* 26695 and insight into mRNase activity. *Biochim Biophys Acta* 2013;1834:2579–2590.
57. Cárdenas-Mondragón MG, Ares MA, Panunzi LG, Pacheco S, Camorlinga-Ponce M, *et al*. Transcriptional profiling of type II toxin-antitoxin genes of *Helicobacter pylori* under different environmental conditions: identification of HP0967-HP0968 system. *Front Microbiol* 2016;7:1872.
58. Yonezawa H, Osaki T, Fukutomi T, Hanawa T, Kurata S, *et al*. Diversification of the AlpB outer membrane protein of *Helicobacter pylori* affects biofilm formation and cellular adhesion. *J Bacteriol* 2017;199:e00729-16.
59. Ge X, Cai Y, Chen Z, Gao S, Geng X, *et al*. Bifunctional enzyme SpoT is involved in biofilm formation of *Helicobacter pylori* with multidrug resistance by upregulating efflux pump Hp1174 (*gluP*). *Antimicrob Agents Chemother* 2018;62:e00957-18.
60. Zhao Y, Cai Y, Chen Z, Li H, Xu Z, *et al*. SpoT-mediated NapA upregulation promotes oxidative stress-induced *Helicobacter pylori* biofilm formation and confers multidrug resistance. *Antimicrob Agents Chemother* 2023;65:e00152-21.
61. Yang F-L, Hassanbhai AM, Chen H-Y, Huang Z-Y, Lin T-L, *et al*. Proteomannans in biofilm of *Helicobacter pylori* ATCC 43504. *Helicobacter* 2011;16:89–98.
62. Anderson JK, Huang JY, Wreden C, Sweeney EG, Goers J, *et al*. Chemorepulsion from the quorum signal autoinducer-2 promotes *Helicobacter pylori* biofilm dispersal. *mBio* 2015;6:e00379.