

Antimicrobial Resistance and Genetic Characteristics that Might be Related to Antibiotic Resistance of 112 *Helicobacter Pylori* Shanghai Isolates

Yixin Liu

Huadong Hospital Affiliated to Fudan University <https://orcid.org/0000-0002-3402-2107>

Su Wang

Huadong Hospital Affiliated to Fudan University

Feng Yang

Huadong Hospital Affiliated to Fudan University

Wenjing Chi

Huadong Hospital Affiliated to Fudan University

Li Ding

Huadong Hospital Affiliated to Fudan University

Tao Liu

Huadong Hospital Affiliated to Fudan University

Feng Zhu

Huadong Hospital Affiliated to Fudan University

Danian Ji

Huadong Hospital Affiliated to Fudan University

Jun Zhou

Huadong Hospital Affiliated to Fudan University

Yi Fang

Huadong Hospital Affiliated to Fudan University

Jinghao Zhang

Huadong Hospital affiliated to Fudan University

Ping Xiang

Huadong Hospital affiliated to Fudan University

Yanmei Zhang (✉ 15618653286@163.com)

Huadong Hospital Affiliated to Fudan University

Hu Zhao

Huadong Hospital Affiliated to Fudan University

Research Article

Keywords: *Helicobacter pylori*, antibiotic resistance, outer membrane protein, virulence, CRISPRs, resistome, variation, genetic characteristics

Posted Date: October 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-970731/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Shanghai, east China has one of the world's highest burdens of *Helicobacter pylori* infection. While multidrug regimens can effectively eradicate *H. pylori*, antibiotic-resistant (AR) *H. pylori* has been recognized by the WHO as 'high priority' for urgent need of new therapies. Moreover, the genetic characteristics of *H. pylori* AR in Shanghai is under-reported. The purpose of this study was to determine the resistance prevalence, re-substantiate resistance-conferring mutations, and investigate novel genetic elements might be related to *H. pylori* AR.

Results

We performed a whole-genomic and phenotypic analysis of 112 *H. pylori* isolated from gastric biopsy specimens from a population with different gastric diseases and residing in Shanghai. No strains were resistant to amoxicillin. Levofloxacin, metronidazole and clarithromycin resistance was observed in 39, 73 and 18 isolates, respectively. There was no association between gastroscopy diagnosis and resistance phenotypes. We reported the presence or absence of several subsystem proteins including *hopE*, *hofF*, *spaB*, *cagY* and *pflA*, and a combination of CRISPRs, which were potentially correlated with resistance phenotypes. The *H. pylori* isolates were also annotated for AR genes to define a resistome containing 80 genes. A resistome-wide association study (RWAS) was performed for the three antibiotics by correlating the phenotypes with the variation data. The RWAS correctly identified the well-known intrinsic resistance associated mutations for levofloxacin (N87T/I and/or D91G/Y mutations in *gyrA*), metronidazole (I38V mutation in *fdxB*), and clarithromycin (A2143G and/or A2142G mutations in 23S rRNA), and added 174 novel variations, with 23 nsSNPs and 48 fsIndels significantly enriched in resistant or susceptible populations. The variant-level linkage disequilibrium analysis demonstrated a series of variations with strong co-occurring correlation with known mutations, and highlighted the underlying role of a protease Lon.

Conclusion

Our study indicated *H. pylori* isolates from Shanghai exhibit multidrug antibiotic resistance, and identified specific genomic characteristics in relation to *H. pylori* AR in Shanghai. Continued surveillance of *H. pylori* AR in Shanghai is warranted in order to establish appropriate eradication treatment regimens for this population.

Introduction

Helicobacter pylori colonizes the gastric mucosa of 50–70% of the world's population and increases the risk of developing chronic gastritis, peptic ulcer, gastric cancer, and gastric mucosa-associated lymphoid tissue (MALT) lymphoma [1–3]. Currently, the recommended *H. pylori* eradication regimen is triple or quadruple therapy, comprising a proton pump inhibitor and a gastric mucosal protective agent, accompanied by one or two antibiotics including clarithromycin (CLA), metronidazole (MTZ), levofloxacin (LEV), and amoxicillin (AMX) [4, 5]. However, *H. pylori* exhibits a diverse genomic structure with a high mutation rate and, under selective antibiotic pressure, manifests growing antibiotic resistance (AR) traits, especially to MTZ and CLA [6]. The resistance rates of *H. pylori* to MTZ, CLA, and LEV in China have previously been measured as 40–70%, 20–50% and 20–50%, respectively [7]. Increasing rates of resistance have also been reported in Europe and America [5, 8], and AR *H. pylori* has been recognized by the World Health Organization (WHO) as "high priority" for urgently needed new therapies [9]. Given the limited number of antibiotics that are appropriate for *H. pylori* eradication and the worldwide increase of antibiotic resistance, the need for new efforts to understand the comprehensive mechanisms underlying *H. pylori* resistance to existing antibiotics is significant.

The genomic diversity that is characteristic of *H. pylori* is crucial for the establishment of associations between genetic elements and antibiotic resistance. A well-confirmed example of the intrinsic resistance mechanism of *H. pylori* has been elucidated showing that the emergence of resistance is mainly due to mutations in antimicrobial target genes. In particular, the single nucleotide mutations in the target genes of CLA (23S rRNA), MTZ (*rdxA*, *frxA* and *fdxB*), and LEV (*gyrA* and *gyrB*) enable *H. pylori* to evade corresponding antibiotic activity by inhibiting bacterial protein synthesis, hampering intracellular reduction-activation of MTZ, and blocking bacterial DNA replication and transcription, respectively [10, 11]. However, antibiotic resistance in *H. pylori* is a complex and multifactorial problem. Several other possible mechanisms have been described, including highly expressed outer membrane proteins that form a permeability barrier, reducing antibiotic uptake [12], loss of porins with narrow pore channels causing decreased intracellular accumulation of antibiotics [13], increase in the efflux of the antibiotics through the efflux-pump transporters [14, 15], the passive tolerance to antimicrobials by biofilm formation [16], as well as sporadic mutations with synergistic effects in the loci of some important protein synthesis factors such as translation initiation factor IF-2 and ribosomal protein L22 [17, 18]. Furthermore, many other whole genome associations with antibiotic resistance have been established in recent years. For instance, a significant association has been found between the virulence factor dupA1 genotype and the A2147G CLA resistance mutation [19] and a correlation between *H. pylori*-resistant strains and CagA-negative strains [20]. Additionally, in Gram-negative bacteria, most studies have suggested an inverse relationship of the clustered regularly interspaced short palindromic repeats (CRISPRs) loci with antimicrobial resistance in *Escherichia coli*, *Acinetobacter baumannii*, and *Klebsiella pneumoniae* [21–24], although some inconsistent observations showed either a positive or negative correlation of CRISPR-Cas with antibiotic resistance genes in some clinically important species [25]. However, although the *H. pylori* is indeed one of the few bacterial species constitutively competent for natural transformation and uptake extracellular DNA [26], little is known about the relationship between CRISPR repeats with antibiotic resistance and acquired resistance genes in *H. pylori*.

Recently, based on the tools and databases of antibiotic resistance (AR) studies, a resistome-wide association study (RWAS) has brought unprecedented insights into the global distribution of potential antibiotic-resistance determinants of many pathogenic bacteria [27], which facilitates the discovery of the molecular bases of the association between antibiotic resistance and genetic diversity. The aim of this study was to identify novel genetic elements

associated with antibiotic resistance and map genomic heterogeneity among *H. pylori* clinical isolates in Shanghai, China. Here, we surveyed the resistance prevalence and present an antibiotic resistance-related genome characterization including specific subsystem analysis at the locus level, and RWAS at variant level, for 112 clinically isolated *H. pylori* isolates from Shanghai.

Materials And Methods

Patients and Samples

A total of 112 *H. pylori* isolates (*H. pylori*-Shi) from 112 individual patients were successfully isolated from 437 gastric biopsy specimens collected during endoscopy in patients with gastric diseases who resided in Shanghai. Patients receiving treatment for *H. pylori* with any antibiotics, H₂-receptor blockers, or proton pump inhibitors within 4 weeks before the study were excluded. All the biopsy samples were identified as *H. pylori* by morphology and immunohistochemistry. The 112 gastric biopsy samples positive for *H. pylori* included 19 in chronic superficial gastritis (CSG), 66 in chronic atrophic gastritis (CAG), 8 in gastric ulcer (GU), 18 in duodenal ulcer (DU), and 1 in gastric carcinoma (GC). The demographics and clinical characteristics (gastroscopy diagnosis) of the 112 patients are shown in Supplementary Table S1. This study was approved by the Ethics Committee for Human Studies of Fudan University Huadong Hospital, with written informed consent from all subjects (Ethics Approval Number: 2020K080).

H. pylori Isolation and Culture

During endoscopy, two biopsy samples were obtained from the antrum, 2–3 cm on the front of the pylorus of each patient. One sample was shipped immediately to a pathology center for *H. pylori* detection in Huadong Hospital. Another was stored at

-80 °C and kept on dry ice during transfer. To culture *H. pylori*, the samples were homogenized and inoculated onto commercial *H. pylori* selective plates (Nissui Pharmaceutical Co. Ltd., Tokyo, Japan), followed by incubating for 3–7 days at 35 °C under microaerophilic conditions (10% O₂, 5% CO₂, and 85% N₂). Small, purple colonies typical of *H. pylori* were subsequently subjected to Gram staining, the rapid urease test (RUT), antibiotic susceptibility testing, and genomic DNA extraction.

Antibiotic Susceptibility Testing

Antibiotic resistance of *H. pylori* to MTZ, CLA, LEV, and AMX was determined by E-test (Autobio, China) and an agar dilution method according to the protocols of the Clinical and Laboratory Standards Institute (Wayne, PA, USA). Briefly, after three subcultures from the time individual strains were isolated, the culture suspension turbidity of *H. pylori* was adjusted with saline to a McFarland opacity standard of 2.0, and the suspensions were inoculated onto a Mueller-Hinton II agar plate (Becton Dickinson, Sparks, MD, USA) supplemented with 10% horse blood. The four drug E-test strips were attached to the plate and incubated at 37 °C for 3–5 days under microaerophilic conditions. Agar dilution MIC tests were performed by serial two-fold dilutions of MTZ, CLA, LEV, and AMX with each concentration ranging from 0.0039–256 mg/L. All drugs were purchased from Wako Pure Chemical Industry (Osaka, Japan). The breakpoints used to determine *H. pylori* antibiotic resistance were MIC ≥ 1 µg/ml for CLA, MIC ≥ 8 µg/ml for MTZ, MIC ≥ 1 µg/ml for LEV, and MIC ≥ 2 µg/ml for AMX. *H. pylori* ATCC 43504 was used as the control strain. Based on the antibiotic susceptibility testing results, the *H. pylori* isolates could be divided into different antibiotic-resistance categories in two patterns: resistance (R)-susceptibility (S) pattern (MTZ-S and -R, CLA-S and -R, LEV-S and -R) and non-overlapping patterns (MTZ mono-R, CLA mono-R, LEV mono-R, MTZ and CLA dual-R, MTZ and LEV dual-R, CLA and LEV dual-R, multiple drug resistant (MDR) and susceptible to four antibiotics).

H. pylori Genomic DNA Extraction and Whole Genome Sequencing

H. pylori isolates were inoculated on Columbia agar (OXOID, Thermo Fisher Scientific Inc., Waltham, MA, USA) medium containing 8% sterile defibrinated sheep blood under microaerophilic conditions at 35 °C for 3–5 days. After three subcultures from the time individual strains were isolated, the total genomic DNA of *H. pylori* was extracted by the cetyltrimethyl ammonium bromide (CTAB) method [28]. All the procedures were performed following the manufacturer's instructions.

Genome sequencing was performed using the whole genome shotgun (WGS) strategy based on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) by constructing paired-end DNA libraries. The DNA libraries (~400 base-pair insert size) for each isolate were prepared using the TruSeq DNA Sample Preparation Kit (Illumina, CA, USA). Briefly, genomic DNA was randomly uniformly fragmented in the range of 300–500 bases by sonication (Diagenode Bioruptor UCD-200). Illumina adapters were ligated to each fragment. Quality control of the libraries was identified by an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) with a DNA 1000 chip according to the manufacturer's instructions. The libraries were sequenced on the Illumina MiSeq sequencer using the 600V3 Kit (Illumina) with sequencing mode of paired-end 2×251 bp and sequencing depth of more than 400× in all the isolates. The quality of the raw reads was checked using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). To obtain high-quality reads, adapter contamination was removed by AdapterRemoval (version 2.1.7) [29], and all reads were filtered on SOAPec (version 2.0) [30] with a Kmer of 17 to trim the low-quality bases with a Phred quality score (Q value) of less than 20. The filtered clean reads were assembled by A5-miseq (version 20160825) [31] into contigs and scaffolds. The sequenced genomes of the 112 *H. pylori*-Shi isolates were submitted to the National Center for Biotechnology Information (NCBI).

H. pylori Protein Coding Genes, rRNA and CRISPRs Prediction

Predicting genome function elements included the prediction of coding genes (CDSs), non-coding RNA, and CRISPRs. Putative coding sequences were identified by Glimmer (version 3.0) [32] with the length of the open reading frame (ORF) no less than 110 bp. tRNA genes were identified by tRNAscan-SE (version 1.3.1) [33] and rRNA genes were identified by RNAmmer (version 1.2) [34]. CRISPRs were identified by the CRISPRfinder (<http://crispr.i2bc.paris-saclay.fr/Server/>), and the direct repeats (DRs) and spacers in the whole genome were predicted by CRISPRs recognition tool [35].

***H. pylori* Functional Annotation of Predicted Protein Coding Genes**

Functional annotations of all predicted protein coding genes were obtained using the Refseq database (NCBI-NR) (<https://www.ncbi.nlm.nih.gov/refseq/>) and SwissProt database (<http://www.gpmaw.com/index.html>). The Refseq database is a non-redundant gene and protein database with relatively high accuracy from the NCBI. In the SwissProt database, all sequence entries have been verified by experienced molecular biologists and protein chemists by consulting relevant literature. Each ORF was thus run through the two databases via blastall [36] with an E value less than 1e-6, at least 30% identity and 70% coverage. The functional descriptions of the best hits in the databases were assigned to the corresponding protein coding genes [37].

Functional classifications by clusters of orthologous groups (COGs) of proteins were analyzed against the eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database (version 5.0) (<https://eggnogdb.embl.de/#/app/home>) [38]. The unique genes of the indicated categories were annotated using this database via blast with E value less than 1e-6 and at least 70% identity and 70% coverage. The database assigned the best hits to the corresponding protein coding genes, which were then classified into corresponding COG classes. The genes containing more than one domain of different categories were classified as multiple class genes, whereas those that did not have a suitable hit against the database were recognized as hypothetical proteins.

***H. pylori* Genome-Wide Gene Family Analysis**

A data set of the predicted protein sequences with a minimum of 50 amino acids of all the 112 *H. pylori* isolates were constructed for further ortholog analysis. An all-VS-all blastp analysis was performed on the dataset with an E value of 1e-10 and 70% protein coverage cut-off applied for the alignments. The orthologous gene clusters of the sequences alignment results were identified using OrthoMCL (version 2.0.8) [39], with the I (Inflation) value was set to 1.5. Finally, the unique genes of indicated strain population were collected and further analyzed.

***H. pylori* Genomic Subsystem Analysis**

Virulence factor associated genes in the protein coding sequences of the 112 *H. pylori* isolates were predicted by aligning amino acid sequences in the Virulence Factors of Pathogenic Bacteria (VFDB) database (<http://www.mgc.ac.cn/VFs/>) [40], which is a comprehensive database of pathogenic bacteria virulence factors containing literature-confirmed 532 virulence factors and 2599 virulence-related genes from 74 genera of pathogenic bacteria including *H. pylori*. Each protein sequence was thus run through the database via blastp v.2.6.0+ for virulence-associated genes with E-value less than 1e-5, greater than 60% identity, 70% coverage, and less than 10% gap length. To retrieve the given outer membrane protein (OMP) family genes and efflux pump system genes, the blastx v.2.5.0+ was used with E-value less than 1e-6 and greater than 90% identity.

Phylogenetic Analysis of *H. pylori* Isolates

The determined whole genome sequence fasta file for each isolate was combined and subjected to multiple sequence alignment and quality control using MAFFT tool (<http://mafft.cbrc.jp/alignment/software/>) and Gblocks (version 0.91) (<http://molevol.cmima.csic.es/castresana/Gblocks.html>). The phylogenetic tree was constructed by the maximum likelihood method in phyl software (version 3.1).

Identification of Acquired Antimicrobial Resistance Genes

The command line version of ResFinder (version 4.1) with default parameters (90% identity and 60% coverage) was used to identify acquired antimicrobial resistance genes (ARGs) in which resistance is conferred by a complete gene [41]. The ResFinder database is composed of 1,649 acquired ARGs from 15 antibiotic drug classes. It does not identify resistance conferred by point mutations.

***H. pylori* Resistome**

The resistome of the 112 *H. pylori* isolates was obtained from databases and literature (Supplementary Table S6). First, the antibiotic-resistant (AR) genes in the genomes of the 112 *H. pylori* isolates were identified using the Comprehensive Antibiotic Research Database (<http://arpcard.mcmaster.ca>) via BLAST [42]. An E value of 1e-6, 45% identity, and 70% protein coverage cut-off were applied for the alignments. The database includes genes involved in antibiotic susceptibility and/or resistance. To specifically identify AR enzyme genes, the BLDB (β -lactamase database) (version 1.0.11) (<http://bldb.eu/>) [43], a newly updated manually curated database for AR enzymes (mainly β -lactamase), was used with an E value of 1e-6, 80% identity, and 70% protein coverage cut-off. Because there was no hit obtained in the database, no β -lactamase genes were included in the final resistome.

To obtain a more comprehensive resistome, we compiled a literature-supported list of genes whose variation or presence or absence is associated with antibiotic resistance in *H. pylori* species, including antibiotic targets and other sporadically reported genes, which were either known to confer *H. pylori* antibiotic resistance or subsequently found to be associated with *H. pylori* resistance to clinically used antibiotics.

Mapping and Variant Detection

To identify genetic alterations resulting in single nucleotide polymorphisms (SNPs) and insertions or deletions (Indels), the sequencing reads of each isolate were mapped to the *H. pylori* 26695 reference genome (NC_000915.1) using the Burrows–Wheeler Alignment Tool (version 0.7.12-r1039) [44]. The alignment files were subjected to local realignment and de-duplication using the Picard (version 1.107) (<http://www.psc.edu/index.php/user-resources/software/picard>). The accuracy of variant calling was improved using Genome Analysis Toolkit (GATK) [45]. Variant SNPs and Indels were culled from each alignment file using GATK. Variant filtering was carried out by removing variants with low mapping quality (MQ) < 40, quality depth (QD) < 2, and haplotype score > 13. The variants were annotated using ANNOVAR (Wang et al., 2010) to determine non-synonymous SNPs (nsSNPs) and frameshift Indels (fsIndels). All variants identified in this study were manually inspected using the Integrative Genomics Viewer (version 2.3.86) [46].

Linkage Disequilibrium Among Variants Potentially Associated With Antibiotic Resistance or Susceptibility

The linkage disequilibrium (LD) of SNPs/Indels potentially associated with MTZ, CLA, and LEV resistance and susceptibility ($P < 0.05$) was assessed by calculating r^2 and D' values to identify co-occurring correlation strength between each two variant sites using plink (version 1.07) [47]. The value of D' was between -1 and 1, and the D' value of 0 represented the two variants in a state of complete linkage equilibrium, and the D' value of 1 represented the two variants in a state of complete LD.

Crystal structures and Functional Domains of the Resistome-Encoding Proteins With Resistance-Associated Variations

The crystal structure of the proteins coded by the indicative genes within the *H. pylori* resistome were available from SWISS-MODEL (<https://swissmodel.expasy.org/>) [48]. The structure model with sequence identity greater than 30%, GMQE (Global Model Quality Estimation) score greater than 0.6, and QMEAN Z-score between -4 to 0 was selected from the alignment results. The detailed model evaluation including the local quality estimate and comparing the model quality score with that of other experimental structures of similar size are shown in Supplementary Figure S3. The functional domain of the indicative genes within the *H. pylori* resistome was identified using Pfam 34.0 (<http://pfam.xfam.org/>) [49]. All significantly matched entries to each gene sequence were retained and all insignificantly matched entries were discarded.

Statistical Analysis

Differences between groups were compared using the Fisher's exact tests, chi-squared test, Pearson's chi-squared tests, or Student's t-test according to the data type and the number of comparisons. When the results of Pearson's chi-squared tests showed a statistical difference, the intra-group pairwise comparison was adjusted using the Bonferroni formula. The statistical analysis was performed via RStudio (R 4.0.2) or GraphPad prism (version 8.0). A P value less than 0.05 was considered statistically significant.

Results

Antibiotic Susceptibility and General Genomic Features of the 112 *H. Pylori*-Shi Isolates

A total of 112 *H. pylori* isolates from patients with digestive diseases were categorized by antibiotic susceptibility phenotypes. Patients' mean age was 52.2 ± 14.8 (mean \pm SD) years, and 54.5% (61/112) were male (Table 1). The prevalence of resistance to MTZ was the highest (65.2%), with 33.9% MTZ mono-resistant, followed by LEV (34.8%), with 4.5% LEV mono-resistant, CLA (16.1%) with 2.7% CLA mono-resistant, and no isolate was resistant to AMX. For dual-resistant isolates, the majority (20.5%) were resistant both to MTZ and LEV. A total of eight isolates were MDR isolates that were simultaneously resistant to MTZ, CLA, and LEV, and 28 were susceptible to all four antibiotics (Table 1). Gender, age, and gastroscopy diagnosis were not significantly associated with resistance ($P > 0.05$).

Table 1
Demographic information and antibiotic resistance rates.

Group	Total, n (%)	Antibiotic-Resistance Phenotypes, n (%)						
		Amoxicillin-R/mono-R	Metronidazole-R/mono-R	Clarithromycin-R/mono-R	Levofloxacin-R/mono-R	Metronidazole-Clarithromycin dual-R	Metronidazole-Levofloxacin dual-R	Clarithromycin-Levofloxacin dual-R
Gender								
male	61 (54.5)	0 (0)/0 (0)	40 (65.6)/23 (37.7)	8 (13.1)/3 (4.9)	19 (31.1)/4 (6.6)	2 (3.3)	12 (19.7)	0 (0)
female	51 (45.5)	0 (0)/0 (0)	33 (64.7)/15 (29.4)	10 (19.6)/0 (0)	20 (39.2)/1 (2.0)	2 (3.9)	11 (21.6)	3 (5.9)
Age								
19-29	8 (7.1)	0 (0)/0 (0)	5 (62.5)/2 (25.0)	2 (25.0)/1 (12.5)	3 (37.5)/0 (0)	0 (0)	2 (25.0)	0 (0)
30-39	18 (16.1)	0 (0)/0 (0)	8 (44.4)/7 (38.9)	3 (16.7)/2 (11.1)	3 (16.7)/3 (16.7)	1 (5.6)	0 (0)	0 (0)
40-49	17 (15.2)	0 (0)/0 (0)	12 (70.6)/6 (35.3)	4 (23.5)/0 (0)	6 (35.3)/0 (0)	1 (5.9)	3 (17.6)	1 (5.9)
50-59	29 (25.9)	0 (0)/0 (0)	20 (69.0)/10 (34.5)	4 (13.8)/0 (0)	9 (31.0)/0 (0)	2 (6.9)	7 (24.1)	1 (3.4)
60-69	27 (24.1)	0 (0)/0 (0)	20 (74.1)/11 (40.7)	2 (7.41)/0 (0)	11 (40.7)/1 (3.7)	0 (0)	8 (29.6)	1 (3.7)
70-79	13 (11.6)	0 (0)/0 (0)	6 (46.2)/2 (15.4)	2 (15.4)/0 (0)	6 (46.2)/1 (7.7)	0 (0)	2 (15.4)	0 (0)
80<	2 (1.8)	0 (0)/0 (0)	2 (100)/0(0)	1 (50)/0(0)	1 (50)/0(0)	0 (0)	1 (50)	0 (0)
Gastroscopy diagnosis								
Chronic superficial gastritis	19 (17.0)	0 (0)/0 (0)	8(42.1)/5(26.3)	2(10.5)/1(5.3)	5(26.3)/2(10.5)	0(0)	2(10.5)	0(0)
Chronic atrophic gastritis	66 (58.9)	0 (0)/0 (0)	46(69.7)/20(30.3)	13(19.7)/0(0)	29(43.9)/3(4.5)	3(4.5)	16(24.2)	3(4.5)
Gastric/Duodenal ulcer	26 (23.2)	0 (0)/0 (0)	18(69.2)/12(46.2)	3(11.5)/2(7.7)	5(19.2)/0(0)	1(3.8)	5(19.2)	0 (0)
Gastric cancer	1 (0.9)	0 (0)/0 (0)	1(100)/1(100)	0 (0)/0 (0)	0 (0)/0 (0)	0 (0)/0 (0)	0 (0)/0 (0)	0 (0)/0 (0)
Total	112	0 (0)/0 (0)	73 (65.2)/38 (33.9)	18 (16.1)/3 (2.7)	39 (34.8)/5 (4.5)	4 (3.6)	23 (20.5)	3 (2.7)

The overall sequence read length, contigs, scaffolds, ORFs, and guanine-cytosine (GC) content for the genomes of the 112 *H. pylori* isolates grouped by antibiotic resistance phenotypes are shown in Supplementary Table S2. On average, the read length of the 112 *H. pylori* isolates' genomes was 1.60 billion bp long and encoded a range of 1,511 to 1,624 genes per genome. A low average guanine-cytosine content, 38.7%, was identified, and the chromosome sizes ranged from 1.52 to 1.69 Mb. All sequenced genomes harbored 36 tRNA genes and an average of two rRNA genes.

Genome-wide Gene Family Analysis and Resistance-Unique Genes of *H. pylori* Isolates

We initially conducted gene family analysis to define the core genetic content and unique-gene pool of specific antibiotic resistance. We obtained 2,439 genes from all the genomes, and among these genes, 2,194 genes were found shared between MTZ-S and -R, 2,101 genes were shared between CLA-S and -R, and 2,193 genes were shared between LEV-S and -R, accounting for 89.95%, 86.14% and 89.91% of the total genes, respectively, which constituted the core genetic content of the corresponding resistant and susceptible categories (Figure 1A-C). Moreover, we obtained 183, 9, and 23 unique genes in MTZ-, CLA-, and LEV-resistant categories, respectively, containing 75, 5, and 13 newly found genes which were not included in neither NCBI-NR, SwissProt, nor COG database (Figure 1D-F, Supplementary Table S9). Among the remaining 108, 4 and 11 unique genes defined as resistance-unique-gene pools of the three antibiotics resistance categories (Supplementary Table S3), only 31, 2, and 3 genes could be assigned to various COG functional classes, respectively (Figure 1D-F).

Distributions of the OMP Family Genes in Resistant and Susceptible Categories

To investigate whether the occurrence of genes of the OMP families was related to specific antibiotic resistance, the presence and absence of genes predicted to encode OMPs in different categories were analyzed, including the currently identified five paralogous gene families major OMP family including Hop subfamily and Hop-related (Hor) subfamily, Hof family, Hom family, iron-regulated OMPs, and efflux pump OMPs, ranging in size from 3 to 36 members, as well as 9 other putative OMPs not in paralogous families (Supplementary Table S4). We first compared the average levels of OMP genes in the resistant and the susceptible categories irrespective of the number of alleles of the genes. There were no significant differences in the average number of total OMP genes, the major OMP family genes, Hof and Hom family genes, and efflux pump OMP genes between isolates resistant and susceptible to the three antibiotics, respectively (Figure 2A-F). We intriguingly found that the average frequency of the iron-regulated OMP genes (*fecA* and *frpB*) in the CLA-resistance category was significantly lower than that in CLA-susceptible category ($P = 0.0103$). Notably, the frequencies of these genes in the isolates varied greatly. Nearly all the isolates contained the majority of the Hop subfamily genes (not including *hopE* and *hopN*), the efflux pump OMP genes, *horB*, *horG*, *horD*, *hofA* belonging to the hor or hof subfamily, and the gene coding BamA and murein lipoprotein 20 (Lpp20) (Figure 2G). However, *horA*, *horC* and *hofG* genes were present in very few strains. Remarkably, difference analysis showed the proportion of isolates carrying *hopE* in both MTZ- and LEV-susceptible categories was significantly higher than that in corresponding resistant category (Figure 2G). Another gene with significantly lower frequency in resistant categories were *hoff* (LEV).

Distributions of the Virulence-Associated Genes in Resistant and Susceptible Categories

To investigate whether the occurrence of virulence factor genes *H. pylori* carried was related to specific antibiotic resistance, the presence and absence of 108 predicted virulence-associated genes against the VFDB database in resistant and susceptible categories was analyzed, including 26 *cag* pathogenicity island (*cagPAI*) genes, 36 flagellar genes, and 47 other virulence-associated genes (Supplemental Table S5). Similarly, we compared the average levels of virulence-associated genes in the resistant and susceptible categories irrespective of the number of alleles of the genes. There were no significant differences in the number of total virulence-associated genes, *cagPAI* genes, and flagellar genes between resistant and the susceptible categories of the three antibiotics, respectively (Figure 3A-C). The presence frequency of these genes in the resistant and susceptible categories were compared, and there was no significant difference in the distribution of most of the genes (Supplemental Table S5). Dramatically, the frequencies of *cagY* (HP0527) and *pflA* (HP1274) were significantly higher ($P = 0.0240$) and lower ($P = 0.0488$) in MTZ-R category than that in MTZ-S, respectively (Figure 3D-E).

Distributions of the Efflux Pump Genes in Resistant and Susceptible Categories

To investigate whether the occurrence of efflux pump genes was related to specific antibiotic resistance, the presence and absence of 18 literature-based efflux pump genes were analyzed, including 12 RND family genes, 3 ABC family genes, and 3 MFS family genes (Figure 4). Among the 18 genes, 16 were present in almost all the 112 *H. pylori* isolates, and the remaining two genes, *spaB* (HP0600) and HP1181 were widely deleted in these isolates, detected in 25 and 31 isolates, respectively. Nevertheless, the proportion of isolates carrying *spaB* in the MTZ-R ($P = 0.0002$) and LEV-R ($P = 0.0408$) categories was obviously higher than that in the MTZ-S and LEV-S categories, respectively (Figure 4).

Correlation Analysis of CRISPR and Its Components With *H. pylori* Antibiotic Resistance

Given the natural function of the CRISPR-Cas system as an adaptive defense mechanism against foreign DNA and a controversial effect of the CRISPR-Cas system on the transfer of the acquired antibiotic resistance genes (ARGs) in bacterial genomes, we conducted predictions in the CRISPRs-Cas systems and acquired ARGs where resistance is conferred by a complete gene rather than mutations, seeking to explore the relationship among CRISPR-Cas, acquired ARGs, and antibiotic resistance in *H. pylori*. Unfortunately, the results suggested neither ARGs nor *cas* genes were found in 112 *H. pylori* isolates. However, a total of 153 CRISPR loci were predicted in 83 isolates, ranging in length from 65 to 496 bp.

Because of the different CRISPR contents of the isolates, the proportions of isolates with 0, 1, 2, 3, and 4 CRISPRs in resistant and susceptible categories of MTZ, CLA, and LEV were analyzed, respectively (Figure 5A-C). Notably, the proportion of the isolates with no CRISPR in the LEV-R category was significantly higher than that in the LEV-S category ($P = 0.0075$, Figure 5C). Because of the different spacer contents of the CRISPRs, the proportion of CRISPRs with 1, 2, 3, 4, and 6 spacers in resistant and susceptible categories was also analyzed (Figure 5D-F). Interestingly, the proportion of the CRISPRs with three spacers in the MTZ-R category was significantly higher than that in the MTZ-S category ($P = 0.0331$, Figure 5D), and the proportion of CRISPRs with two spacers in the LEV-R category was significantly lower than that in the LEV-S category ($P = 0.0152$, Figure 5F). Similarly, the proportion of the CRISPRs with two or three spacers in the resistance category of the other two antibiotics was also higher or lower than that of the corresponding susceptible category, respectively, although there was no significant difference (Figure 5D-F). In addition, detailed analysis of the DR and spacer sequences showed that a CRISPR combination containing three most common CRISPRs in *H. pylori* coexisted in seven isolates, which intriguingly were all MTZ mono-R isolates (Figure 5G). Of note, one of the DR in the CRISPR combination occurred in CRISPRs of 9 isolates, which included CRISPRs in seven MTZ mono-R isolates and 2 CRISPRs (Supplementary Figure S1) in two other MTZ-R isolates (a MTZ mono-R isolate and a MTZ and CLA dual-R isolate), causing the significantly higher proportion of the isolates with the DR in the MTZ-R category compared with that in the MTZ-S category ($P = 0.0259$, Figure 5G, Supplementary Table S10).

Phylogenetic Analysis and Relationship of Antibiotic Resistance Patterns and specific genetic loci with strain relatedness in *H. pylori* Isolates.

When constructing the maximum likelihood tree of the 112 *H. pylori*-Shi genomes with respect to different resistance patterns and the profile of the observed differently present subsystem genomic loci between susceptible and resistant categories for the corresponding antibiotic, no strain-relatedness was found to be associated to neither each single antibiotic phenotype nor the presence/absence of the differently present loci including genes coding OMPs (*hopE*, *hoff* and iron-regulated OMPs), virulence-associated genes (*virB10*/*cagY* and *pflA*), efflux pump genes (*spaB*), and CRISPRs with 2 or 3 spacers or the indicated DR sequences (Figure 6). Alternatively, these strains had distinct susceptibility and indicated loci profiles regardless of genetic relatedness among strains.

Genome-Wide Genetic Variations and Prevalence of Variations in *H. pylori* Resistome

We culled SNPs and Indels within the coding sequences and noncoding rRNAs and tRNAs from the genomes of all the *H. pylori* isolates when compared with *H. pylori* 26695 reference genome (Figure 7A), which showed high densities of the variations that could be as high as 136 Indel density and 75 SNP density corresponding to reference genome (Figure 7A). Among these genetic variations, a total of 388 non-synonymous SNPs (nsSNPs) and 1,718 frameshift Indels (fsIndels) were identified (Supplementary Table S8).

We obtained a list of 80 genes defined as the *H. pylori* resistome, whose variation, presence, or absence were known to either confer *H. pylori* antibiotic resistance or subsequently found to be associated with *H. pylori* resistance to clinically used antibiotics not limited to antibiotics with known phenotypes in this study. To investigate the prevalence of the genetic variations in the *H. pylori* resistome, we first assayed the variations in six most common resistance genes, including 23S rRNA, *gyrA* and *gyrB* genes, and *rdxA*, *frxA*, and *fdxB* genes, in which single-point mutations in specific regions are known to be responsible for CLA, LEV, and MTZ resistance, respectively (Table 2-4). Of these SNPs, a subset was significantly associated with antibiotic resistance, including A2143G, A2142G and C148T in the 23S rRNA gene, and N87T, N87I, and D91G, D91Y in *gyrA*, as well as I38V in *fdxB*. Additionally, the assay detected four Indels in 23S rRNA, three Indels in *gyrA*, and one in *frxA*. Unexpectedly, the prevalence of one of the Indels (762_763insT) in 23S rRNA was significantly higher in the CLA-S category (53.19%) compared with that in the CLA-R category (Table 2). Furthermore, we overviewed the distribution of the numbers of the SNPs and Indels present in the remaining genes of the *H. pylori* resistome of the 112 isolates along with the antibiotic resistant profile of these genes (Supplementary Figure S2). Of the remaining 74 genes of the *H. pylori* resistome, 174 variations were obtained corresponding to 54 genes, involving 29 genes detected containing nsSNPs, 36 containing fsIndels, and 10 containing both nsSNPs and fsIndels (Supplementary Table S7). Additionally, the maximum number of Indels (ranging from 0 to 15) contained in these genes was up to three times the maximum number of SNPs (ranging from 0 to 3), suggesting to a certain extent that Indel mutations make greater contributions to the high genetic diversity of the *H. pylori* resistome than SNP mutations do (Supplementary Figure S2).

Table 2
nsSNPs and fsIndels in genes conferring resistance to CLA.

Gene	Nucleotide change	No. among CLA-R (n=18)	No. among CLA-S (n=94)	P
23S rRNA	SNP			
	A2143G	17 (94.44%)	2 (2.13%)	<0.0001
	A2142G	2 (11.11%)	0 (0.00%)	0.0250
	T2182C	6 (33.33%)	31 (32.98%)	0.9770
	C148T	2 (11.11%)	0 (0.00%)	0.0250
	T1025C	0 (0.00%)	1 (1.06%)	1.0000
	G1027A	14 (77.78%)	52 (55.32%)	0.0760
	Indel			
	761delAA (GAA to G)	16 (88.89%)	84 (89.36%)	1.0000
	762_763insT (A to AT)	5 (27.78%)	50 (53.19%)	0.0482
	764_765insTTT (C to CTTT)	8 (44.44%)	54 (57.45%)	0.3093
	1036_1037insC (A to AC)	18 (100.00%)	94 (100.00%)	1.0000
	1515_1516insT (C to CT)	3 (16.67%)	9 (9.57%)	0.4054
	No change	0 (0.00%)	13 (13.83%)	0.1232

Table 3
nsSNPs and fsIndels in genes conferring resistance to LEV.

Gene	Sequence change		No. among LEV-R (n=39)	No. among LEV-S (n=73)	P
	Nucleotide position	Amino acid position			
gyrA	SNP				
	C261A	N87T	8 (20.51%)	0 (0.00%)	0.0001
	C261G	N87I	9 (23.08%)	0 (0.00%)	< 0.0001
	G271A	D91G	12 (30.77%)	0 (0.00%)	< 0.0001
	G271T	D91Y	5 (1.67%)	0 (0.00%)	0.0043
	Indel				
	2455_2456insAT (AC to AATC)	T819fs	32 (82.05%)	57 (78.08%)	0.6203
	2460_2463del (CTTCG to C)	S820fs	34 (87.18%)	62 (84.93%)	0.7460
	2472delT (AT to A)	N824fs	38 (97.44%)	66 (90.41%)	0.2577
	No change	No change	1 (2.56%)	7 (9.59%)	0.2577
gyrB	SNP				
	C1801T	L601F	37 (94.87%)	73 (100.00%)	0.3423
	No change	No change	2 (5.13%)	0 (0.00%)	0.1192

Table 4
nsSNPs and fsIndels in genes conferring resistance to MTZ.

Gene	Sequence change		No. among MTZ-R (n=73)	No. among MTZ-S (n=39)	P
	Nucleotide position	Amino acid position			
rdxA	No change	No change	73 (100.00%)	39 (100.00%)	NA
frxA	SNP				
	G499A	D167N	4 (5.48%)	4 (10.26%)	0.3497
	T577A	C193S	63 (86.30%)	39 (100.00%)	0.3238
	Indel				
	653dupA (T to TA)	X218delinsX	70 (95.89%)	39 (100.00%)	0.5503
	No change	No change	3 (4.11%)	0 (0.00%)	0.5503
fdxB	SNP				
	A112G	I38V	65 (89.04%)	39 (100.00%)	0.0488
	No change	No change	8 (10.96%)	0 (0.00%)	0.0488
ins: insert, del: delete, fs: frameshift, NA: not available					

Resistome-Wide Association Study (RWAS) of Genetic Variations

In order to explore MTZ, CLA, and LEV resistance-related or resistance-induced compensation related potential genetic variants, we compared the prevalence of the resistome-wide nsSNPs and fsIndels between resistant and susceptible isolates. We observed 23 nsSNPs and 48 fsIndels corresponding to 31 genes that were significantly enriched in antibiotic-resistant or -susceptible isolates ($P < 0.05$). We found that four variations within the integrase/recombinase gene *xerD* showed a significant association with both MTZ and CLA resistance. Some of these variations had strong linkage disequilibrium (LD). Both nsSNPs (*xerD*:D353N** and *xerD*:S275N*) were linked together but in low LD with other two *xerD* fsIndels. Similarly, the two fsIndels (*xerD*:I40/120_121 delete GG and *xerD*:L39/116_117 insert G) were also linked together (Figure 7B). A deletion within the site-specific DNA-methyltransferase gene *hsdM* (HP0850) (*hsdM*:P333/998_1002 delete AGCCG) was found to be significantly associated with both MTZ and LEV resistance, which also were in high LD with *gyrA*:N87I. Furthermore, we observed 16 variations within 15 genes simultaneously showing association with MTZ susceptibility and LEV resistance, and many of these genes code membrane-related proteins (*hetA*, *msbA*, *hofC*, *oppA*, *tetA(p)*, and *babB*) or proteins involved in recombination (*recG* and *ruvC*) (Figure 7B). Moreover, several deletions within the ATP-dependent serine protease gene *lon* (especially the deletions at *lon*:T582, *lon*:K584, and *lon*:E588) showing a significant association with MTZ susceptibility were found with high LD with more than half of the other variations, including the resistance-conferring nsSNPs 23S rRNA:A2143G, 23S rRNA:A2142G, 23S rRNA:C148T, *gyrA*:N87T, and some of the remaining variants, most of which were associated with MTZ resistance or susceptibility (Figure 7B) (Supplementary Table S11).

Resistance- or Susceptibility-Associated Variations Occurred in Functionally Important Domains of Genes Within the *H. pylori* Resistome

For the purpose of relating the resistance- or susceptibility-enriched variations to specific protein domains, we further characterized these variations in the genes within the *H. pylori* resistome with corresponding protein structure data. Localization of the nsSNPs and fsIndels in the four proteins (Lon, BabB, XerD, and TrpS) with the largest number of phenotype-associated variations were summarized (Figure 8), and the functional domains of all the genes with the phenotype-associated variations within the *H. pylori* resistome were identified. Eight variations in Lon and three partial in TrpS closely located variations were identified although they were not in any domains (Figure 8A and D, Supplementary Figure S4). In BabB, two clusters of locationally close variation sites at L59, N60, G129, S130, and N135 were all in the sialic acid binding adhesin (SabA) N-terminal extracellular adhesion domain (Figure 8B, Supplementary Figure S4). Among four variations in XerD, only one nsSNP S275N was located in the phage integrase family domain that enables the protein to covalently link to DNA. Of the remaining 27 genes, 14 contained variations totally or partially located in their functional domains, including three genes encoding efflux pump transporters with an S267P mutation in the extracellular solute-binding domain of *dppA*, an R390K mutation in the ABC transporter domain of *hetA*, and an L304F mutation in the ABC transporter transmembrane region of *msbA*; three genes required for lipopolysaccharide (LPS) biosynthesis with a V76A mutation in the mur ligase family catalytic domain of *murC*, a deletion at A350 and an insertion at I351 in the mur ligase family glutamate ligase domain of *murE*, and a deletion at P129 in the region with glycosyltransferase activity of *rfaF*; two genes related to protein translation with an S32A mutation in the elongation factor Tu GTP binding domain of *lepA* and a deletion at A297 in the tRNA synthetase domain of *iles*, as well as the metalloregulator gene *fur* with an insertion at S148 and a deletion at E149 in the ferric uptake regulator region; the OMP *hofC* with an I206V mutation and a deletion at M276 in the major functional domain; the stress-response protein coding gene *grpE* with H145N mutation in its main functional domain; the chromosomal replication initiator gene *dnaA* with A200T mutation in the whole functional region; and the restriction enzyme alleles *hsdM_2* and *hsdM_3* with a deletion at P333 in the domain the N-6 DNA methylase activity and a deletion at N49 in the N-terminal domain, respectively (Supplementary Figure S4).

Discussion

The emergence and rapid spread of AR *H. pylori* has become a major obstacle to eradicating the infection and a great threat to human health because of a sharp global decline in the efficacy of the recommended treatment that used to be 90% effective and has reached an unacceptable threshold (< 80%) compared to other bacteria [4, 50–52]. The ability of *H. pylori* to develop resistance to antibiotics is usually genetically encoded, but only few of these genetic features have been associated with specific antibiotic resistance. In this study, the first-line antibiotic (MTZ, CLA, and LEV)-resistance associated genetic features of 112 *H. pylori*-Shi isolates were analyzed to explore possible links between specific gene loci and variants with the resistance phenotype (Figure 9).

Initially we determined the antibiotic resistance rates of four antibiotics of the 112 *H. pylori*-Shi isolates. Our data demonstrated that both the mono- (65.2% and 34.8%) and dual- (20.5%) resistance of MTZ and LEV had the highest prevalence compared to other resistance categories, respectively, while the prevalence of resistance to CLA (16.1%) was slightly lower than the resistance rate range of 20–50% in China previously reported [7]. Because of the quite low AMX resistance rate of 0% or < 5% worldwide [52], which was zero in our 112 *H. pylori*-Shi isolates, it was not possible to conduct further association analysis of the resistance mechanism of AMX. In addition, the *H. pylori* strains in this study were isolated from patients with four main gastric diseases including chronic superficial gastritis, chronic atrophic gastritis, gastric/duodenal ulcer, and gastric cancer, which enabled us to examine potential differences in resistance rates in different specific gastrointestinal pathologies (except gastric cancer because of insufficient sample number). The negative results implied that the intragastric conditions of patients with different *H. pylori*-linked gastric disorders may have no significant effect on the antibiotic resistance of the pathogen colonizing the gastric mucosa.

Because different phenotypes of *H. pylori* are considered to be associated with unique sets of genetic features [53, 54], we conducted the genome-wide gene family analysis to compare resistant and susceptible categories to obtain unique-gene pools and their functional enrichment, which is rarely investigated in *H. pylori*. Putatively affected by differences in the strain number of each resistant and susceptible category, the content of the unique-gene pools of the three antibiotics varied greatly. Nevertheless, nearly or more than half of them were newly found genes in *H. pylori* with resistance to MTZ (75/183, 40.98%), CLA (5/9, 55.56%), and LEV (13/23, 56.52%), respectively, indicating the extensive intraspecies diversity and genetic heterogeneity of *H. pylori* [55].

In our analysis, despite a great deal of porin hopE labelled loss in nearly half of the 112 isolates, the gene displayed significant enrichment in susceptible categories of both MTZ and LEV compared with the resistant population. Similar situations have been reported in a large number of publications showing the loss of porins in clinical isolates [56]. It is assumed that bacterial outer membrane porins influence the influx of small-molecule antimicrobials into bacterial cells, and there have been many reports of antibiotic resistance acquired through loss or functional change of porins in a large number of organisms [57]. For example, in *E. coli*, the loss of porins OmpC and OmpF contributes to quinolone resistance [58, 59]. *A. baumannii* porins OmpA, CarO, Omp33-36, and OprD have been associated with resistance to cephalothin and carbapenem [60–63]. The loss of OmpK35 and OmpK36 porins of *K. pneumoniae* has been related to resistance to cephalosporins and increased meropenem MIC [57]. The Hop member HopE, one of the first identified porins in *H. pylori*, is a major nonselective one that is the homolog to the *E. coli* OmpF porin [64, 65]. In our study, the significant loss of hopE in the MTZ- and LEV-resistant population may implicate the underlying effect of hopE loss on resistance development to the two antibiotics with small molecular weights rather than CLA with a large molecular weight.

Concerning the relationship of virulence factors and antibiotic resistance, our analysis suggested a generally irrelevant relationship between most of the predicted virulence-associated genes and the phenotypes at the “locus” level, including the extensively studied *H. pylori* virulence factors CagA and VacA. This observation is consistent with some reports showing no correlation between antibiotic resistance and the presence of either the *cagA* or *vacA* gene [20, 66], but it contrasts with many other findings that the *cagA*-negative strains and *vacA*-containing strains were associated with CLA resistance [67, 68]. These results remain inconclusive as they showed a variable pattern, most likely due to low sample numbers in individual studies or variations in geographical regions and evolution. Moreover, we first reported the apparent differences of *cagY* and *pilA* frequencies in MTZ-resistant and -susceptible populations of *H. pylori*. CagY is an ortholog of VirB10 and is an essential gene in the ComB type IV secretion system (T4SS) [69]. Rather than a canonical type IV (pseudo-) pilus used by other Gram-negative bacteria, *H. pylori* employs the ComB system for initial DNA uptake during natural transformation [70]. Thus, the significantly higher frequency of *CagY* in the MTZ resistance category suggests that the functional ComB system with intact CagY enables the entry of resistance-associated DNA via

transformation, which promotes the development of MTZ resistance. Future studies are needed to identify and better understand the specific mechanisms behind this observation.

Previous reports have demonstrated that the AcrAB-TolC efflux pump belonging to the RND family encoding hefABC, hefDEF, and hefGHI is the most important efflux system in *H. pylori*, mainly mediating resistance to several antibiotics such as MTZ and CLA in laboratory conditions [11, 15]. However, our analysis showed that the frequencies of the RND family efflux pump genes in resistant and susceptible categories were not significantly different. We also reported here that the ABC family member *spaB* with a severe loss in the 112 isolates showed a frequency difference in the resistant and susceptible categories of both MTZ and LEV, which is consistent with a previous report by Yang et al. from China that the expression level of *spaB* in multidrug-resistant strains were significantly higher than those in susceptible strains and that the sensitivity of *spaB*-knockout *H. pylori* to four antibiotics (LEV, CLA, chloramphenicol and rifampicin) was significantly higher than that of wild-type strains (reference not shown). Similar studies on *spaB* have not been found reported in the rest of the world. Whether the transporter is involved in *H. pylori* resistance remains to be determined.

Studies investigating the characteristics of CRISPR-Cas systems in relation to antibiotic resistance in clinical *H. pylori* isolates are lacking. These are significant gaps in knowledge because the CRISPR-Cas system is hypothesized to be a natural impediment to HGT in bacteria [71], and that HGT disseminates antibiotic resistance-associated genetic elements among competent pathogens in vivo [72], and it is essential to identify strategies to reduce the spread of these elements. Nevertheless, our analysis failed to predict any acquired ARGs where resistance is conferred by a complete gene. Given the high rate of the naturally competent transformation of *H. pylori* [73], it is plausible that transfer of single nucleotide mutations conferring antibiotic resistance rather than complete ARGs plays a prominent role in the horizontally acquired resistance of *H. pylori*, which has been supported by a series of reports [26, 70, 74]. Furthermore, we found that LEV resistance and possession of complete CRISPR loci are inversely related. Similar associations showing that the CRISPR system was more likely to be found in antimicrobial-susceptible species such as *E. coli* and *Enterococci* have also been reported [71, 75, 76]. Additionally, our correlation of a different number of spacers with resistance or susceptibility to LEV or MTZ, as well as the interesting finding of a CRISPR combination and a DR sequence exclusively occurring in MTZ-resistant strains, provided an initial foundation for potentially utilizing spacer content or DR sequence to rapidly determine the phenotype of *H. pylori* isolates.

Similar to other bacteria with small-size genomes, *H. pylori* responses to environmental pressure rely mainly on variations rather than a complex regulatory network [77]. Next-generation sequencing is a powerful tool to compare whole genomes and analyze the variations found within antibiotic resistance-associated genes. When compiling the *H. pylori* resistome based on databases, no β -lactamase genes were found in the 112 *H. pylori* isolates, which was compatible with previous reports that the β -lactamase gene can be detected in AMX-resistant *H. pylori* but not in AMX-susceptible strains [78]. Regarding the antibiotic target genes, not only the well-known mutations in *gyrA* (N87T/I and D91G/Y) for LEV resistance and mutations in 23S rRNA (A2143G and A2142G) for CLA resistance were detected but also several other single nucleotide mutations and Indels in relation to antibiotic resistance or susceptibility, which could be compensatory mutations to mitigate the deleterious effects of the specific functions compromised by the resistance mutations [79]. In the other resistome-wide genetic variations associated with antibiotic resistance or susceptibility, we noticed an MTZ susceptibility-associated variations in *fur*. We have known that MTZ belonging to the nitroimidazoles class is a pro-drug that is activated by nitroreductases, and the reduction products can disrupt DNA. It was suggested that the resistance stemming from inactive nitroreductases can be eliminated by mutations in the *fur* protein [80]. The LD analysis among these variations revealed several notable deletion mutations in *lon* that were found in high LD with more than half of the other variations including the resistance-conferring nsSNPs. Given that *lon*-encoded protease mediates the selective degradation of mutants and abnormal proteins [81, 82], we speculated that deletion mutations at specific sites of *lon* indirectly participate in the resistance mechanisms of *H. pylori* by influencing the effect of *lon* in degrading mutants or proteins coded by mutants potentially related to antibiotic resistance or susceptibility of *H. pylori*, which may explain the linked co-occurrence of them to a certain extent. The further structure and functional domain analysis of the genes containing the resistance- or susceptibility-associated variations within the *H. pylori* resistome better indicated their probable involvement in the regulation of antibiotic resistance. For example, in BabB, five variations occurred in SabA N-terminal extracellular adhesion domain. This domain is conserved among SabA orthologs and BabA, and is known to function as a sugar-binding adhesion domain with conserved disulfide bonds [83]. Smiley et al previously reported that BabB transmembrane protein were up-regulated in CLA-resistant *H. pylori* [12]. The other resistome-wide functional domains containing variations included regions important for efflux (*dppA*, *hetA* and *msbA*), LPS biosynthesis (*murC*, *murE* and *rfaF*) and stress response (*grpE*). It has been reported that adhesins (especially OMPs from the Hop and Hom family) [84–86], LPS [84, 86, 87], efflux pumps [14, 84, 88] and proteins regulating a stress response [84, 85] were all associated with the biofilm formation of *H. pylori*. Thus, it may be the case that biofilm formation influenced by these genetic mutants is an important mechanism in the antibiotic resistance of *H. pylori*, which need to be substantiated via extensive experiments.

Conclusion

Our study demonstrated *H. pylori* strains isolated from Shanghai exhibited multidrug antibiotic resistance, and correctly identified the well-known mutations associated with intrinsic resistance to MTZ, CLA and LEV. Furthermore, we found antibiotic resistance phenotypes were associated with a number of novel locus- and variant-level genetic elements. The underlying co-occurrence relationships among them were also revealed. These findings provide a data basis for further resistance associated research. Moreover, thorough detection and characterization of antibiotic resistance in *H. pylori* strains is necessary to provide efficacious treatment regimens. Continued monitoring of antibiotic resistance using susceptibility assays and genotyping is warranted in Shanghai.

Declarations

Availability of data and materials

The Whole Genome Shotgun project of the 112 *H. pylori*-Shi isolates has been deposited at DDBJ/ENA/GenBank (BioProject number PRJNA690879) under the accessions shown in supplementary table S2. Additional raw data is available in supplementary table S8-S11.

Funding

This work was supported by the Shanghai Science and Technology Committee (Grant No: 18411960600,16411968000,18411950800), National High Technology Research and Development Program of China (Grant No: 2015AA021107-019), Shanghai Shenkang Hospital Development Center "New Frontier Technology Joint Research Project" (Grant No: SHDC12015107), and the Shanghai Health and Family Planning Commission Youth Project of Scientific Research Subject (Grant No: 20184Y0032). Funding was also received from Shanghai "Rising Stars of Medical Talent" Outstanding Youth Development Program of 2018 and "Huadong Hospital project" (Grant No: 2019jc019).

Acknowledgements

Not applicable.

Ethics approval and consent to participate

This study was approved by the Ethics Committee for Human Studies of Fudan University Huadong Hospital, with written informed consent from all subjects (Ethics Approval Number: 2020K080).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Authors' contributions

Hu Zhao and Yanmei Zhang conceived of and designed the study. Danian Ji, Jun Zhou and Xiang Ping conducted sample collection. Wenjing Chi, Li Ding, Tao Liu and Feng Zhu conducted the bacteria culture and antibiotic susceptibility testing. Yi Fang and Jinghao Zhang performed data reduction. Su Wang and Feng Yang conducted manual inspection of the data. Yixin Liu performed analysis and drafted the manuscript.

References

1. Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, Stemmermann GN, Nomura A: Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res.* 1995;55(10):2111-2115.
2. Hatakeyama M: *Helicobacter pylori* CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. *Cell Host Microbe.* 2014;15(3):306-316.
3. Kuipers EJ, Perez-Perez GI, Meuwissen SG, Blaser MJ: *Helicobacter pylori* and atrophic gastritis: importance of the *cagA* status. *J Natl Cancer Inst.* 1995;87(23):1777-1780.
4. Chey WD, Leontiadis GI, Howden CW, Moss SF: ACG Clinical Guideline: Treatment of *Helicobacter pylori* Infection. *Am J Gastroenterol.* 2017;112(2):212-239.
5. Nezami BG, Jani M, Alouani D, Rhoads DD, Sadri N: *Helicobacter pylori* Mutations Detected by Next-Generation Sequencing in Formalin-Fixed, Paraffin-Embedded Gastric Biopsy Specimens Are Associated with Treatment Failure. *J Clin Microbiol.* 2019;57(7).
6. Savoldi A, Carrara E, Graham DY, Conti M, Tacconelli E: Prevalence of Antibiotic Resistance in *Helicobacter pylori*: A Systematic Review and Meta-analysis in World Health Organization Regions. *Gastroenterology.* 2018;155(5):1372-1382 e1317.
7. Liu WZ, Xie Y, Lu H, Cheng H, Zeng ZR, Zhou LY, Chen Y, Wang JB, Du YQ, Lu NH et al: Fifth Chinese National Consensus Report on the management of *Helicobacter pylori* infection. *Helicobacter.* 2018;23(2):e12475.
8. Chen D, Cunningham SA, Cole NC, Kohner PC, Mandrekar JN, Patel R: Phenotypic and Molecular Antimicrobial Susceptibility of *Helicobacter pylori*. *Antimicrob Agents Chemother.* 2017;61(4).
9. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y et al: Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis.* 2018;18(3):318-327.
10. Alba C, Blanco A, Alarcon T: Antibiotic resistance in *Helicobacter pylori*. *Curr Opin Infect Dis.* 2017;30(5):489-497.
11. Gong Y, Yuan Y: Resistance mechanisms of *Helicobacter pylori* and its dual target precise therapy. *Crit Rev Microbiol.* 2018;44(3):371-392.
12. Smiley R, Bailey J, Sethuraman M, Posecion N, Showkat Ali M: Comparative proteomics analysis of sarcosine insoluble outer membrane proteins from clarithromycin resistant and sensitive strains of *Helicobacter pylori*. *J Microbiol.* 2013;51(5):612-618.
13. Co EM, Schiller NL: Resistance mechanisms in an in vitro-selected amoxicillin-resistant strain of *Helicobacter pylori*. *Antimicrob Agents Chemother.* 2006;50(12):4174-4176.
14. Cai Y, Wang C, Chen Z, Xu Z, Li H, Li W, Sun Y: Transporters HP0939, HP0497, and HP0471 participate in intrinsic multidrug resistance and biofilm formation in *Helicobacter pylori* by enhancing drug efflux. *Helicobacter.* 2020;25(4):e12715.
15. Iwamoto A, Tanahashi T, Okada R, Yoshida Y, Kikuchi K, Keida Y, Murakami Y, Yang L, Yamamoto K, Nishiumi S et al: Whole-genome sequencing of clarithromycin resistant *Helicobacter pylori* characterizes unidentified variants of multidrug resistant efflux pump genes. *Gut Pathog.* 2014;6:27.

16. Krzyżek P, Grande R, Migdał P, Paluch E, Gościński G: Biofilm Formation as a Complex Result of Virulence and Adaptive Responses of *Helicobacter pylori*. *Pathogens*. 2020;9(12).
17. Miftahussurur M, Syam AF, Nusi IA, Makmun D, Waskito LA, Zein LH, Akil F, Uwan WB, Simanjuntak D, Wibawa ID et al: Surveillance of *Helicobacter pylori* Antibiotic Susceptibility in Indonesia: Different Resistance Types among Regions and with Novel Genetic Mutations. *PLoS One*. 2016;11(12):e0166199.
18. Miftahussurur M, Shrestha PK, Subsomsong P, Sharma RP, Yamaoka Y: Emerging *Helicobacter pylori* levofloxacin resistance and novel genetic mutation in Nepal. *BMC Microbiol*. 2016;16(1):256.
19. Hussein NR, Tunjel I, Majed HS, Yousif ST, Aswad SI, Assafi MS: Duodenal ulcer promoting gene 1 (dupA1) is associated with A2147G clarithromycin-resistance mutation but not interleukin-8 secretion from gastric mucosa in Iraqi patients. *New Microbes New Infect*. 2015;6:5-10.
20. Fasciana T, Cala C, Bonura C, Di Carlo E, Matranga D, Scarpulla G, Manganaro M, Camilleri S, Giammanco A: Resistance to clarithromycin and genotypes in *Helicobacter pylori* strains isolated in Sicily. *J Med Microbiol*. 2015;64(11):1408-1414.
21. Li HY, Kao CY, Lin WH, Zheng PX, Yan JJ, Wang MC, Teng CH, Tseng CC, Wu JJ: Characterization of CRISPR-Cas Systems in Clinical *Klebsiella pneumoniae* Isolates Uncovers Its Potential Association With Antibiotic Susceptibility. *Front Microbiol*. 2018;9:1595.
22. Mangas EL, Rubio A, Alvarez-Marin R, Labrador-Herrera G, Pachon J, Pachon-Ibanez ME, Divina F, Perez-Pulido AJ: Pangenome of *Acinetobacter baumannii* uncovers two groups of genomes, one of them with genes involved in CRISPR/Cas defence systems associated with the absence of plasmids and exclusive genes for biofilm formation. *Microb Genom*. 2019;5(11).
23. Touchon M, Charpentier S, Pognard D, Picard B, Arlet G, Rocha EP, Denamur E, Branger C: Antibiotic resistance plasmids spread among natural isolates of *Escherichia coli* in spite of CRISPR elements. *Microbiology (Reading)*. 2012;158(12):2997-3004.
24. Touchon M, Cury J, Yoon EJ, Krizova L, Cerqueira GC, Murphy C, Feldgarden M, Wortman J, Clermont D, Lambert T et al: The genomic diversification of the whole *Acinetobacter* genus: origins, mechanisms, and consequences. *Genome Biol Evol*. 2014;6(10):2866-2882.
25. Shehreen S, Chyou TY, Fineran PC, Brown CM: Genome-wide correlation analysis suggests different roles of CRISPR-Cas systems in the acquisition of antibiotic resistance genes in diverse species. *Philos Trans R Soc Lond B Biol Sci*. 2019;374(1772):20180384.
26. Corbinais C, Mathieu A, Kortulewski T, Radicella JP, Marsin S: Following transforming DNA in *Helicobacter pylori* from uptake to expression. *Mol Microbiol*. 2016;101(6):1039-1053.
27. Jaillard M, van Belkum A, Cady KC, Creely D, Shortridge D, Blanc B, Barbu EM, Dunne WM, Jr., Zambardi G, Enright M et al: Correlation between phenotypic antibiotic susceptibility and the resistome in *Pseudomonas aeruginosa*. *Int J Antimicrob Agents*. 2017;50(2):210-218.
28. Jin Y, Deng J, Liang J, Shan C, Tong M: Efficient bacteria capture and inactivation by cetyltrimethylammonium bromide modified magnetic nanoparticles. *Colloids Surf B Biointerfaces*. 2015;136:659-665.
29. Schubert M, Lindgreen S, Orlando L: AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*, 2016;9:88.
30. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et al: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012;1(1):18.
31. Coil D, Jospin G, Darling AE: A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*. 2015;31(4):587-589.
32. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL: Improved microbial gene identification with GLIMMER. *Nucleic Acids Res*. 1999;27(23):4636-4641.
33. Lowe TM, Eddy SR: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5):955-964.
34. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW: RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 2007;35(9):3100-3108.
35. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P: CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*. 2007;8:209.
36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410.
37. Blake JD, Cohen FE: Pairwise sequence alignment below the twilight zone. *J Mol Biol*. 2001;307(2):721-735.
38. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldon T, Rattei T, Creevey C, Kuhn M et al: eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res*. 2014;42(Database issue):D231-239.
39. Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoekert CJ, Jr.: Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics*. 2011;Chapter 6:Unit 6 12 11-19.
40. Chen L, Xiong Z, Sun L, Yang J, Jin Q: VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res*. 2012;40(Database issue):D641-645.
41. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattori V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF et al: ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*. 2020;75(12):3491-3500.
42. McArthur AG, Waglehner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L et al: The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother*. 2013;57(7):3348-3357.
43. Naas T, Oueslati S, Bonnin RA, Dabos ML, Zavala A, Dortet L, Retailleau P, Iorga BI: Beta-lactamase database (BLDB) - structure and function. *J Enzyme Inhib Med Chem*. 2017;32(1):917-919.
44. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
45. Zhu P, He L, Li Y, Huang W, Xi F, Lin L, Zhi Q, Zhang W, Tang YT, Geng C et al: Correction: OTG-snp caller: An Optimized Pipeline Based on TMAP and GATK for SNP Calling from Ion Torrent Data. *PLoS One*. 2015;10(9):e0138824.

46. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z: Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865-2871.
47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.
48. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L et al: SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303.
49. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ et al: Pfam: The protein families database in 2021. *Nucleic Acids Res*. 2021;49(D1):D412-D419.
50. Fischbach L, Evans EL: Meta-analysis: the effect of antibiotic resistance status on the efficacy of triple and quadruple first-line therapies for *Helicobacter pylori*. *Aliment Pharmacol Ther*. 2007;26(3):343-357.
51. Malfertheiner P, Megraud F, O'Morain CA, Atherton J, Axon AT, Bazzoli F, Gensini GF, Gisbert JP, Graham DY, Rokkas T et al: Management of *Helicobacter pylori* infection—the Maastricht IV/ Florence Consensus Report. *Gut*. 2012;61(5):646-664.
52. Megraud F: *H pylori* antibiotic resistance: prevalence, importance, and advances in testing. *Gut*. 2004;53(9):1374-1384.
53. Badr MT, Omar M, Hacker G: Comprehensive Integration of Genome-Wide Association and Gene Expression Studies Reveals Novel Gene Signatures and Potential Therapeutic Targets for *Helicobacter pylori*-Induced Gastric Disease. *Front Immunol*. 2021;12:624117.
54. Gunaletchumy SP, Seevasant I, Tan MH, Croft LJ, Mitchell HM, Goh KL, Loke MF, Vadivelu J: *Helicobacter pylori* genetic diversity and gastro-duodenal diseases in Malaysia. *Sci Rep*. 2014;4:7431.
55. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI et al: Traces of human migrations in *Helicobacter pylori* populations. *Science*. 2003;299(5612):1582-1585.
56. Masi M, Winterhalter M, Pages JM: Outer Membrane Porins. *Subcell Biochem*. 2019;92:79-123.
57. Beceiro A, Tomas M, Bou G: Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin Microbiol Rev*. 2013;26(2):185-230.
58. Bekhit A, Fukamachi T, Saito H, Kobayashi H: The role of OmpC and OmpF in acidic resistance in *Escherichia coli*. *Biol Pharm Bull*. 2011;34(3):330-334.
59. Liu YF, Yan JJ, Lei HY, Teng CH, Wang MC, Tseng CC, Wu JJ: Loss of outer membrane protein C in *Escherichia coli* contributes to both antibiotic resistance and escaping antibody-dependent bactericidal activity. *Infect Immun*. 2012;80(5):1815-1822.
60. Catel-Ferreira M, Nehme R, Molle V, Aranda J, Bouffartigues E, Chevalier S, Bou G, Jouenne T, De E: Deciphering the function of the outer membrane protein OprD homologue of *Acinetobacter baumannii*. *Antimicrob Agents Chemother*. 2012;56(7):3826-3832.
61. del Mar Tomas M, Beceiro A, Perez A, Velasco D, Moure R, Villanueva R, Martinez-Beltran J, Bou G: Cloning and functional analysis of the gene encoding the 33- to 36-kilodalton outer membrane protein associated with carbapenem resistance in *Acinetobacter baumannii*. *Antimicrob Agents Chemother*. 2005;49(12):5172-5175.
62. Limansky AS, Mussi MA, Viale AM: Loss of a 29-kilodalton outer membrane protein in *Acinetobacter baumannii* is associated with imipenem resistance. *J Clin Microbiol*. 2002;40(12):4776-4778.
63. Sugawara E, Nikaido H: OmpA is the principal nonspecific slow porin of *Acinetobacter baumannii*. *J Bacteriol*. 2012;194(15):4089-4096.
64. Bina J, Bains M, Hancock RE: Functional expression in *Escherichia coli* and membrane topology of porin HopE, a member of a large family of conserved proteins in *Helicobacter pylori*. *J Bacteriol*. 2000;182(9):2370-2375.
65. Doig P, Exner MM, Hancock RE, Trust TJ: Isolation and characterization of a conserved porin protein from *Helicobacter pylori*. *J Bacteriol*. 1995;177(19):5447-5452.
66. Boehnke KF, Valdivieso M, Bussalleu A, Sexton R, Thompson KC, Osorio S, Novoa Reyes I, Crowley JJ, Baker LH, Xi C: Antibiotic resistance among *Helicobacter pylori* clinical isolates in Lima, Peru. *Infect Drug Resist*. 2017;10:85-90.
67. Brennan DE, Dowd C, O'Morain C, McNamara D, Smith SM: Can bacterial virulence factors predict antibiotic resistant *Helicobacter pylori* infection? *World J Gastroenterol*. 2018;24(9):971-981.
68. Haddadi MH, Negahdari B, Asadolahi R, Bazargani A: *Helicobacter pylori* antibiotic resistance and correlation with cagA motifs and homB gene. *Postgrad Med*. 2020;132(6):512-520.
69. Barrozo RM, Hansen LM, Lam AM, Skoog EC, Martin ME, Cai LP, Lin Y, Latoscha A, Suerbaum S, Canfield DR et al: CagY Is an Immune-Sensitive Regulator of the *Helicobacter pylori* Type IV Secretion System. *Gastroenterology*. 2016;151(6):1164-1175 e1163.
70. Damke PP, Di Guilmi AM, Varela PF, Velours C, Marsin S, Veaute X, Machouri M, Gunjal GV, Rao DN, Charbonnier JB et al: Identification of the periplasmic DNA receptor for natural transformation of *Helicobacter pylori*. *Nat Commun*. 2019;10(1):5357.
71. Price VJ, McBride SW, Hullahalli K, Chatterjee A, Duerkop BA, Palmer KL: *Enterococcus faecalis* CRISPR-Cas Is a Robust Barrier to Conjugative Antibiotic Resistance Dissemination in the Murine Intestine. *mSphere*. 2019;4(4).
72. Lermiaux NA, Cameron ADS: Horizontal transfer of antibiotic resistance genes in clinical environments. *Can J Microbiol*. 2019;65(1):34-44.
73. Moore ME, Lam A, Bhatnagar S, Solnick JV: Environmental determinants of transformation efficiency in *Helicobacter pylori*. *J Bacteriol*. 2014;196(2):337-344.
74. Levine SM, Lin EA, Emara W, Kang J, DiBenedetto M, Ando T, Falush D, Blaser MJ: Plastic cells and populations: DNA substrate characteristics in *Helicobacter pylori* transformation define a flexible but conservative system for genomic variation. *FASEB J*. 2007;21(13):3458-3467.

75. Aydin S, Personne Y, Newire E, Laverick R, Russell O, Roberts AP, Enne VI: Presence of Type I-F CRISPR/Cas systems is associated with antimicrobial susceptibility in *Escherichia coli*. *J Antimicrob Chemother.* 2017;72(8):2213-2218.
76. Palmer KL, Gilmore MS: Multidrug-resistant enterococci lack CRISPR-cas. *mBio.* 2010;1(4).
77. Kraft C, Suerbaum S: Mutation and recombination in *Helicobacter pylori*: mechanisms and role in generating strain diversity. *Int J Med Microbiol.* 2005;295(5):299-305.
78. Tseng YS, Wu DC, Chang CY, Kuo CH, Yang YC, Jan CM, Su YC, Kuo FC, Chang LL: Amoxicillin resistance with beta-lactamase production in *Helicobacter pylori*. *Eur J Clin Invest.* 2009;39(9):807-812.
79. Hanafi A, Lee WC, Loke MF, Teh X, Shaari A, Dinarvand M, Lehours P, Megraud F, Leow AH, Vadivelu J et al: Molecular and Proteomic Analysis of Levofloxacin and Metronidazole Resistant *Helicobacter pylori*. *Front Microbiol.* 2016;7:2015.
80. Hu Y, Zhang M, Lu B, Dai J: *Helicobacter pylori* and Antibiotic Resistance, A Continuing and Intractable Problem. *Helicobacter.* 2016;21(5):349-363.
81. Andrianova AG, Kudzhaev AM, Abrikosova VA, Gustchina AE, Smirnov IV, Rotanova TV: Involvement of the N Domain Residues E34, K35, and R38 in the Functionally Active Structure of *Escherichia coli* Lon Protease. *Acta Naturae.* 2020;12(4):86-97.
82. Tzeng SR, Tseng YC, Lin CC, Hsu CY, Huang SJ, Kuo YT, Chang CI: Molecular insights into substrate recognition and discrimination by the N-terminal domain of Lon AAA+ protease. *Elife.* 2021;10.
83. Pang SS, Nguyen STS, Perry AJ, Day CJ, Panjekar S, Tiralongo J, Whisstock JC, Kwok T: The three-dimensional structure of the extracellular adhesion domain of the sialic acid-binding adhesin SabA from *Helicobacter pylori*. *J Biol Chem.* 2014;289(10):6332-6340.
84. Hathroubi S, Zerebinski J, Ottemann KM: *Helicobacter pylori* Biofilm Involves a Multigene Stress-Biased Response, Including a Structural Role for Flagella. *mBio.* 2018;9(5).
85. Krzyzek P, Grande R, Migdal P, Paluch E, Gosciniak G: Biofilm Formation as a Complex Result of Virulence and Adaptive Responses of *Helicobacter pylori*. *Pathogens.* 2020;9(12).
86. Wong EH, Ng CG, Chua EG, Tay AC, Peters F, Marshall BJ, Ho B, Goh KL, Vadivelu J, Loke MF: Comparative Genomics Revealed Multiple *Helicobacter pylori* Genes Associated with Biofilm Formation In Vitro. *PLoS One.* 2016;11(11):e0166835.
87. Servetas SL, Doster RS, Kim A, Windham IH, Cha JH, Gaddy JA, Merrell DS: ArsRS-Dependent Regulation of *hombB* Contributes to *Helicobacter pylori* Biofilm Formation. *Front Microbiol.* 2018;9:1497.
88. 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.

Figures

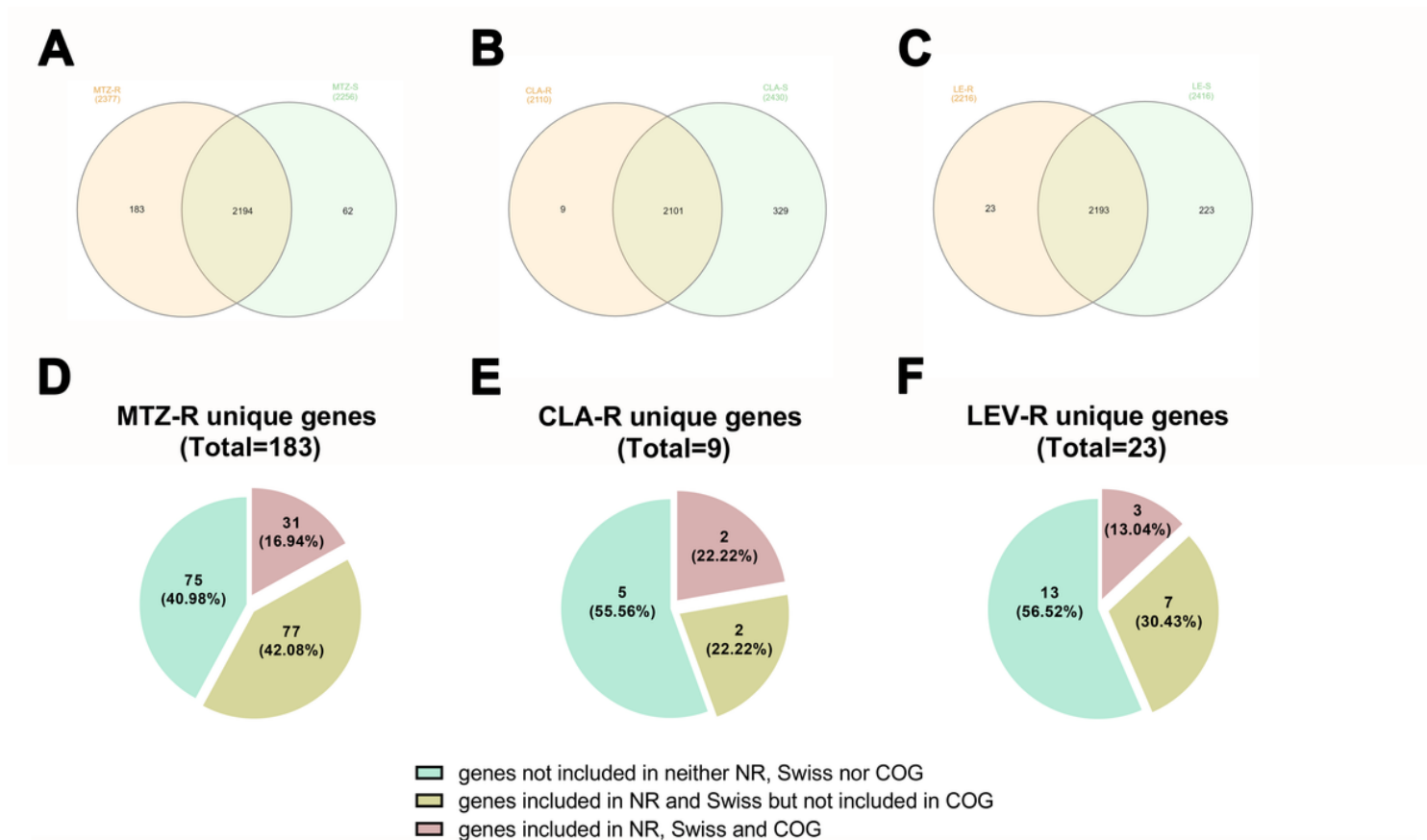


Figure 1

Genome-wide gene family analysis and functional classification of unique genes of antibiotic resistant categories. Venn diagrams showing the number of common and unique genes in MTZ- (A), CLA- (B) and LEV- (C) resistant (yellow) and susceptible (green) categories. The number in the parentheses indicates the total number of genes identified in each category. Pie diagrams revealing the status of unique genes of MTZ-R (D), CLA-R (E) and LEV-R (F) categories in NCBI-NR, SwissProt and COG databases.

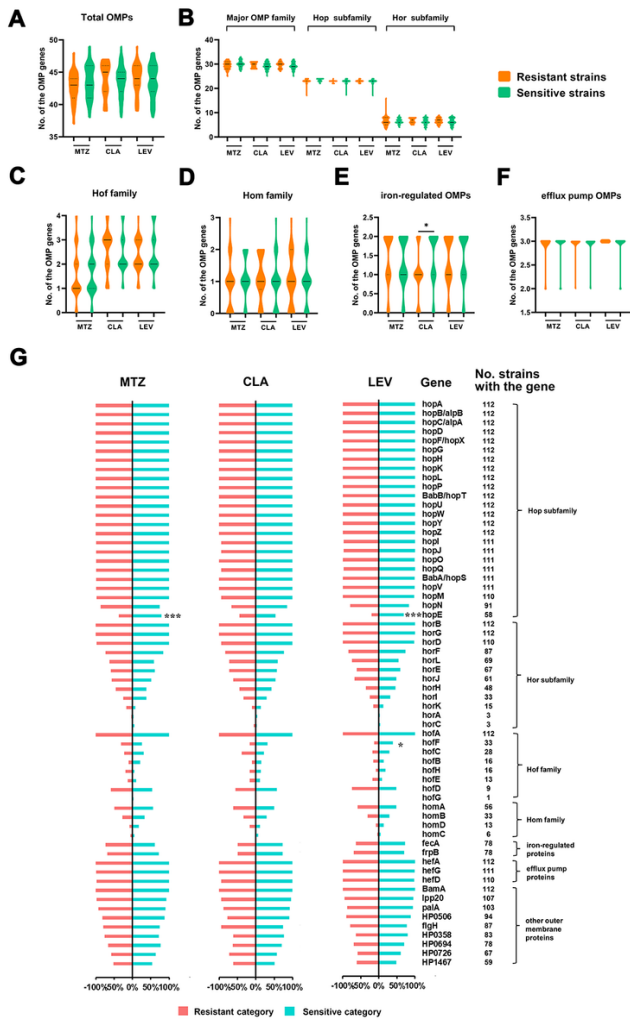


Figure 2 Analysis of OMP family genes in resistant and susceptible categories. Violin plots showing the distribution of the OMP genes in resistant and susceptible categories. The average presence levels of the total 62 OMP genes (A), the major OMP family genes along with two subfamilies genes (B), Hof family genes (C), Hom family genes (D), iron-regulated OMPs genes (F) and efflux pump OMP genes (F) in resistant and susceptible categories were compared, irrespective of the number of alleles of the genes. Solid lines indicate median levels and dotted lines indicate quartiles levels. (G) Proportion of strains with each OMP family gene in resistant and susceptible categories. Only genes with greater than 90% coverage were labelled as present. The number of strains containing the gene were listed on the right. (A-G) * $P < 0.05$, *** $P < 0.001$. Where there were no statistical analysis results labeled, there were no significant difference.

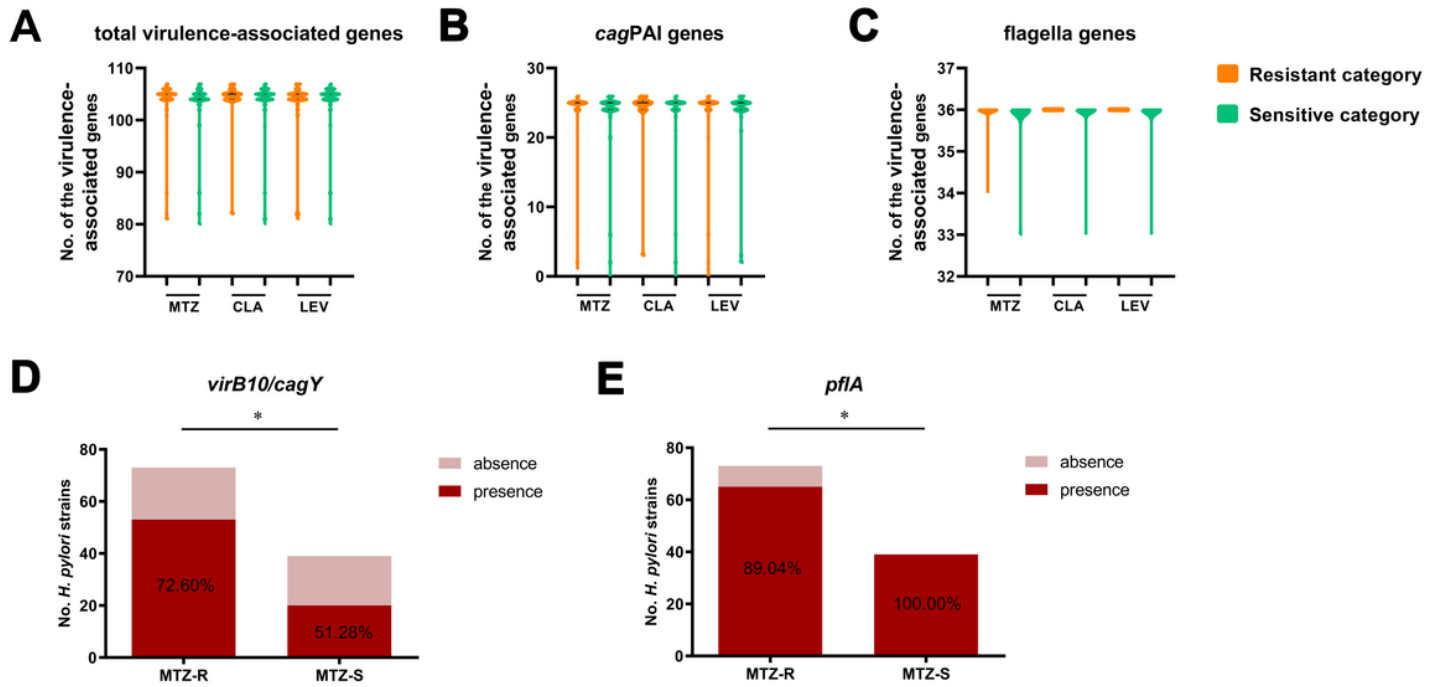


Figure 3

Analysis of virulence-associated genes in resistant and susceptible categories. Violin plots showing the distribution of the virulence-associated genes in resistant and susceptible categories. Comparisons of the average presence levels of the total virulence-associated genes (A), cagPAI genes (B) and flagellar genes (C) between resistant and susceptible categories were conducted, irrespective of the number of alleles of the genes. Solid lines indicate median levels and dotted lines indicate quartiles levels. Where there were no statistical analysis results labeled, there were no significant difference. (D) Two virulence-associated genes with significant different frequencies in MTZ-R and MTZ-S categories. * $P < 0.05$.

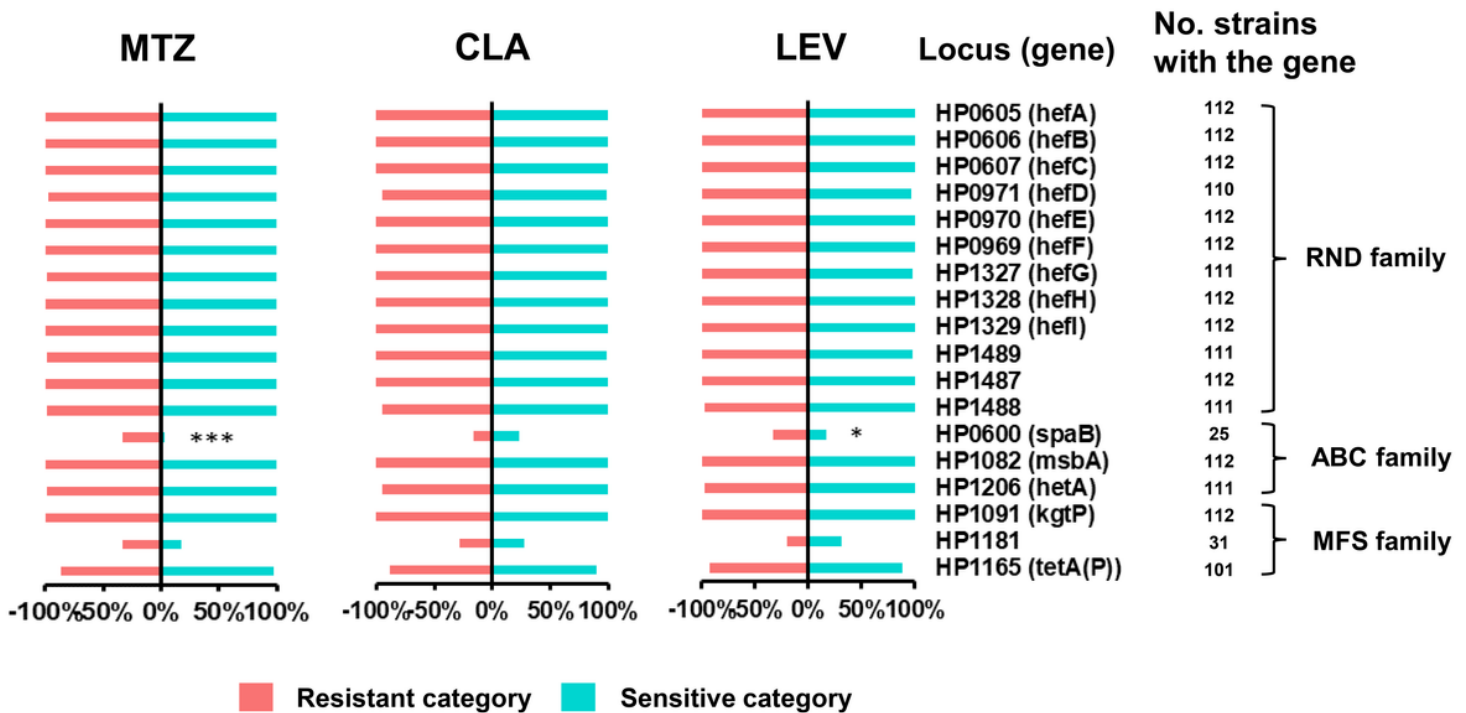


Figure 4

Analysis of efflux pump genes in resistant and susceptible categories. Proportion of strains with each efflux pump gene in resistant and susceptible categories. The number of strains containing the gene were listed on the right. * $P < 0.05$, *** $P < 0.001$. Where there were no statistical analysis results labeled, there were no significant difference.

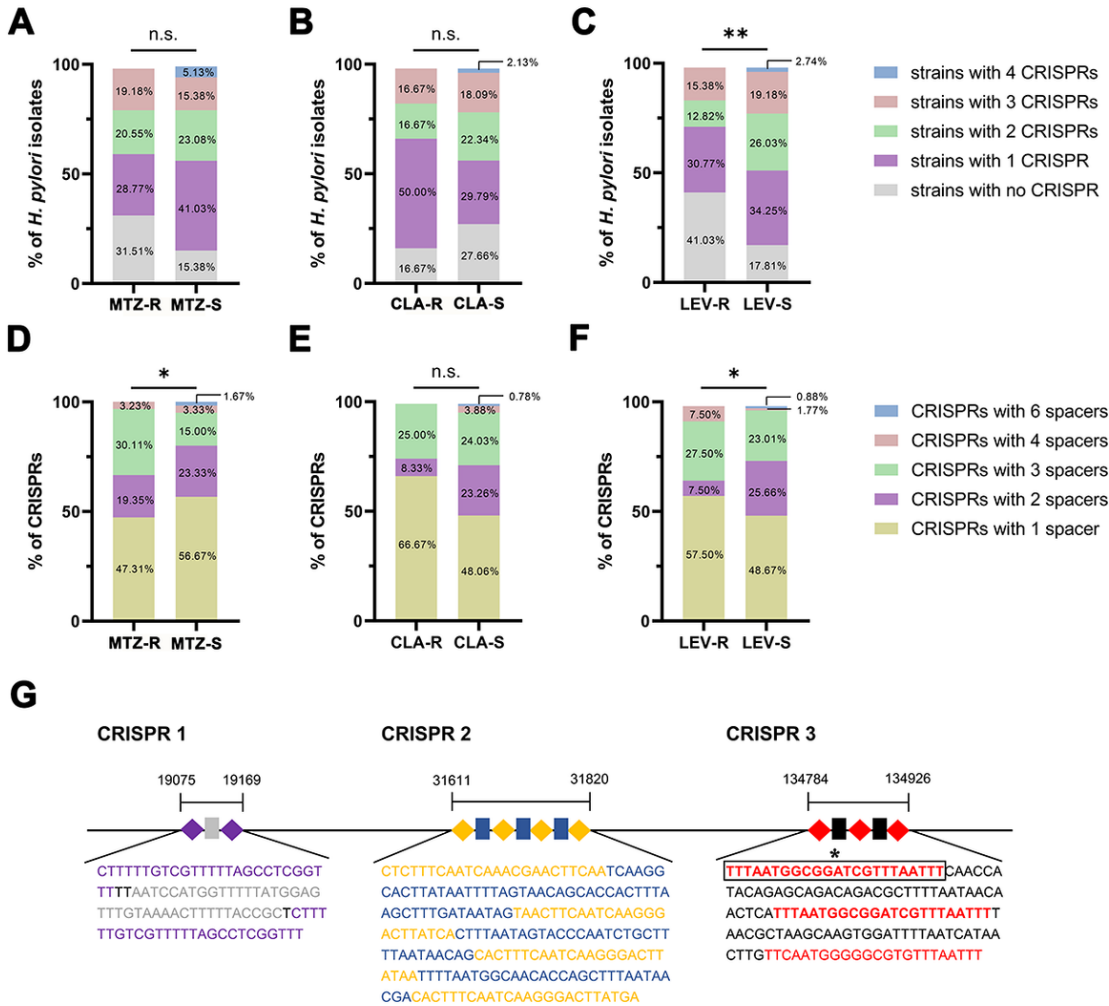


Figure 5

Distribution and constitution of CRISPRs in resistant and susceptible categories and a previously unknown CRISPRs combination present exclusively in seven MTZ mono-R strains. The proportions of strains with 0, 1, 2, 3 and 4 CRISPRs in resistant and susceptible categories of MTZ (A), CLA (B) and LEV (C), respectively. ** $P < 0.01$, n.s. not significant. The proportions of CRISPRs with 1, 2, 3, 4 and 6 spacers in resistant and susceptible categories of MTZ (D), CLA (E) and LEV (F), respectively. * $P < 0.05$, n.s. not significant. (G) The structural analysis of a previously unknown CRISPRs combination presented exclusively in seven MTZ mono-R strains, including DRs (purple, yellow and red diamonds in each CRISPR) and spacers (grey, blue and black rectangles in each CRISPR). The base sequences are shown below the CRISPR array. The DRs and spacers in the colored characters correspond with the colors of the respective diamonds and rectangles. The DR sequence in box, which exclusively presented in nine MTZ-R strains, indicated that there was a significant difference of the DR frequencies between MTZ-R and MTZ-S categories. * $P < 0.05$.

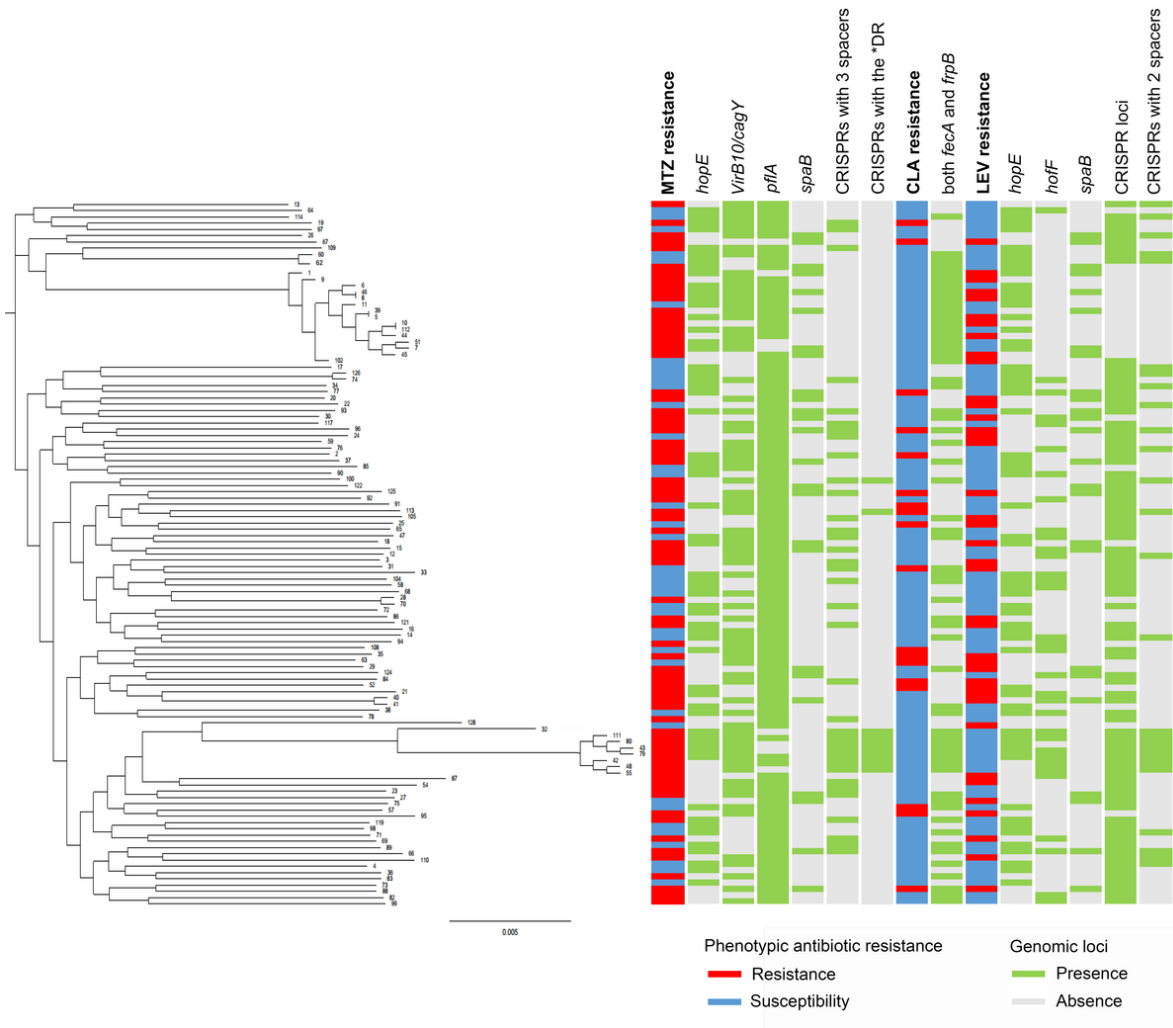
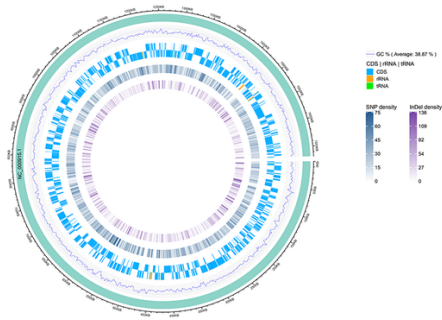


Figure 6
 Phylogenetic analysis with respect to antibiotic resistance patterns and specific genetic loci profiles in *H. pylori* isolates. Maximum likelihood phylogenetic tree based on whole genome sequences of the 112 *H. pylori* isolates, the resistance patterns, and the significantly different genomic loci observed above between susceptible and resistant categories for the corresponding antibiotic. The susceptible and resistant patterns are denoted by blue and red rectangles, respectively. The presence and absence of the specific loci are denoted by green and grey rectangles, respectively.

A



B

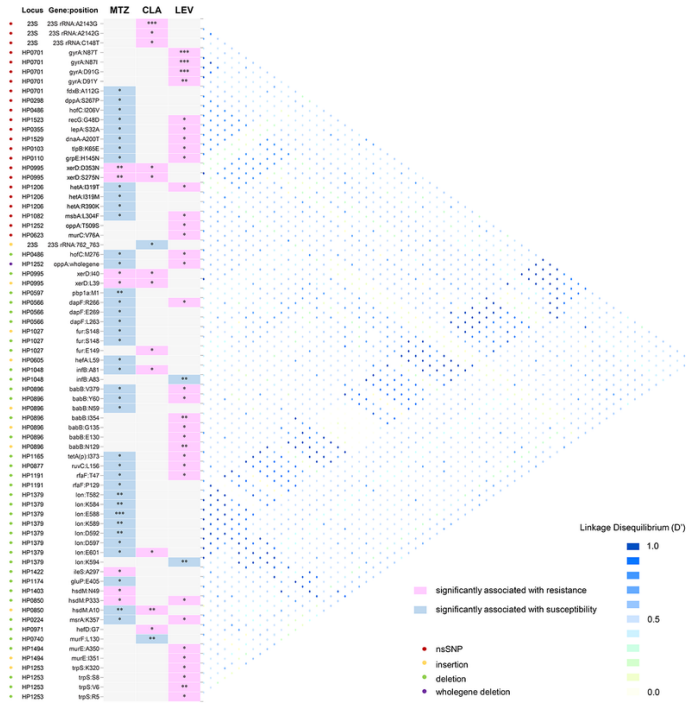


Figure 7

Synoptic representation of the risistome-wide association study (RWAS) of genetic variations and linkage disequilibrium analysis. (A) The densities of total SNPs and Indels identified in the 112 *H. pylori* isolates against *H. pylori* 26695 genome are overviewed. The numbers next to the rulers represent the number of SNPs and Indels per unit length of *H. pylori* 26695 genome. (B) The proportions of the strains with the nsSNPs and fsIndels present in the genes within the *H. pylori* risistome in the resistant and susceptible categories of MTZ, CLA and LEV were analyzed, respectively. Only nsSNPs and fsIndels with significantly higher (red cells) or lower (blue cells) frequencies in the resistant categories of MTZ, CLA and LEV were displayed. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Conversely, grey cells indicate the variations not significantly associated with resistance or susceptibility to the antibiotics. Loci, genes, and variations are reported in rows and antibiotics in columns. The nature of the variations is color-coded on the left. The right panel represents the extent of linkage disequilibrium between variations.

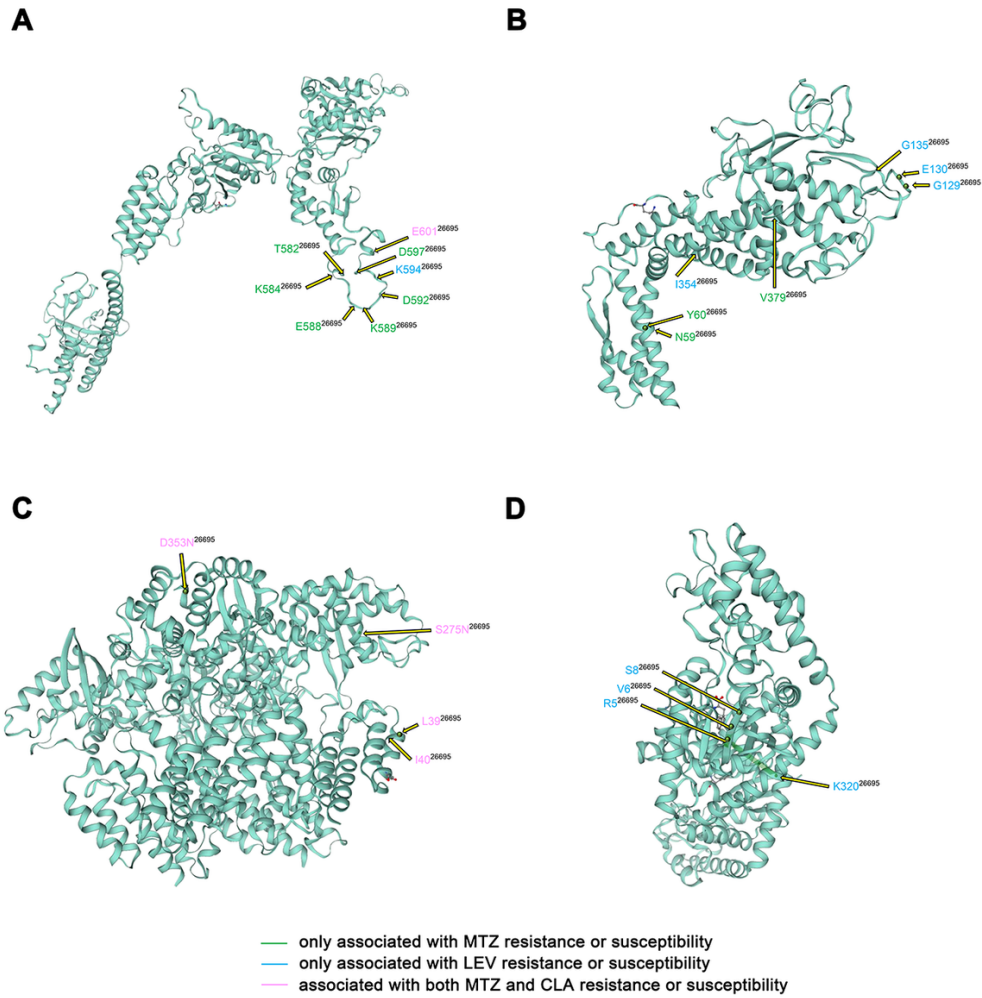


Figure 8
 Localization of the nsSNPs and fsIndels in the four proteins with the largest number of resistance-associated variations within the *H. pylori* resistome. The available crystal structure of each protein was used to indicate the sites with amino acid changes. (A) Lon, (B) BabB, (C) XerD, and (D) TrpS. The color scale represents the corresponding antibiotics of the resistance or susceptibility association.

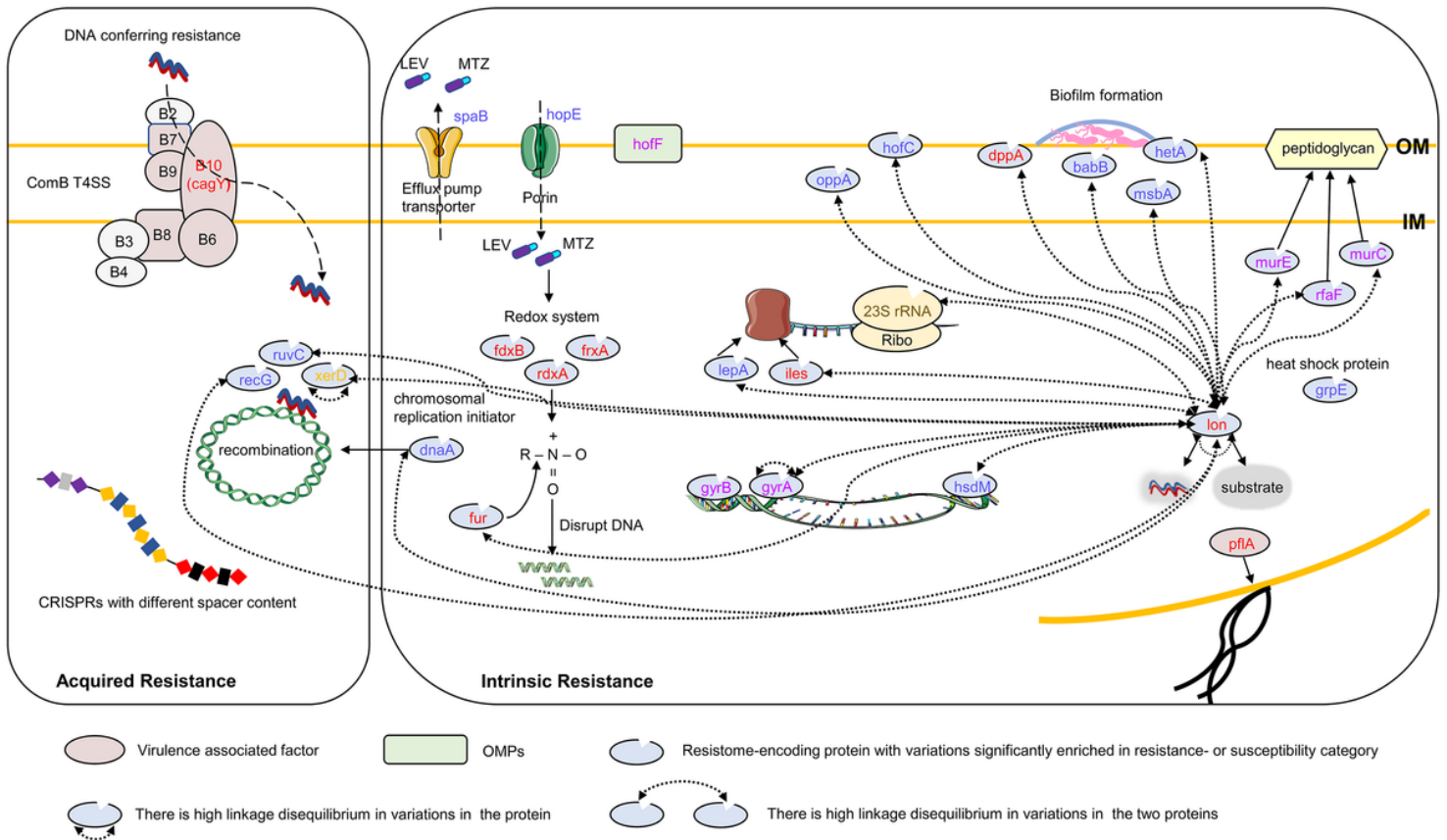


Figure 9

Cartoon model showing the putative genetic characteristics likely related to *H. pylori* resistance- or susceptibility- to MTZ, CLA and LEV at locus- and variant-levels. The intrinsic resistance or resistance-associated compensation related genetic features involves the variations in the resistome-encoding proteins including antibiotic targets, translation-associated factors, transporters, OMPs, proteins probably related to biofilm formation, and other sporadic factors including Lon protease with variations in linkage disequilibrium with many others, as well as the presence or absence of a porin, and an efflux pump transporter. In the underlying acquired resistance related genetic elements, the presence or absence of a virulence-associated factor belonging to ComB T4SS and a certain CRISPRs, as well as variations in several recombination required proteins, were demonstrated. Different colors of the genetic element names denote they are in relation to resistance or susceptibility of different antibiotics. Red (MTZ), gold (CLA), pink (LEV), orange (MTZ and CLA), blue (MTZ and LEV), brown (CLA and LEV).

Supplementary Files

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

- [SupplementaryFigures.docx](#)
- [SupplementaryTableS1S7.xlsx](#)
- [SupplementaryTableS8.xlsx](#)
- [SupplementaryTableS9.xlsx](#)
- [SupplementaryTableS10.xlsx](#)
- [SupplementaryTableS11.xlsx](#)