Supplementary materials

Section S1: Standardized molecular protocols

S1.1. Dual-phase nucleic acid extraction

Total genomic DNA and RNA were isolated from ~70 mg of thinly sectioned gastric biopsy tissue using a modified TRIzol™ (Invitrogen, Thermo Fisher Scientific) phase-separation protocol. Tissues were homogenized in 1 mL TRIzol on ice to preserve transcript integrity. After adding 200 μL chloroform and centrifuging (14,000 rpm, 15 min, 4°C), the upper aqueous phase was collected for RNA precipitation with isopropanol.

DNA was subsequently recovered from the interphase/organic phase via ethanol precipitation, followed by three washes in 0.1 M sodium acetate to remove heme and other PCR inhibitors commonly present in gastric tissue.

S1.2. Stringent quality control (QC)

Nucleic acid purity and concentration were determined using the NanoDrop-1000 (Thermo Scientific). Inclusion for Tier two transcriptional analysis was strictly gated by the following criteria:

- Genomic DNA: A_{260}/A_{280} ratio of 1.7–1.8.
- Total RNA: A_{260}/A_{280} ratio of ~2.0.
- Integrity: Verified via 1% agarose gel electrophoresis; samples showing significant degradation or ribosomal subunit smearing were excluded and re-extracted.

S1.3. Reverse transcription–quantitative PCR (RT-qPCR) and normalization strategy

RNA was reverse-transcribed into cDNA using the SuperScript™ III First-Strand Synthesis System (Invitrogen) with random hexamers. Quantitative PCR (qPCR) was performed using SYBR™ Green chemistry on the ABI PRISM 7000/7500 platform.

To account for variations in bacterial load and tissue quality, relative expression was calculated using the 2^−ΔΔCt method, normalized against a geometric mean of three endogenous reference genes: *ureA, glmM,* and *hsp60.*

S1.4. *In silico* primer design and bioinformatic validation

The design and validation of primers for *H. pylori* virulence markers and housekeeping genes followed a standardized four-step bioinformatic pipeline: (1) Sequence Retrieval and Alignment: reference sequences were retrieved from the NCBI GenBank database (Accessions: AY737319.1, AB017923.1, AF479027.1, AL445067), and multiple sequence alignments (CLUSTAL W) were performed across diverse strains to target highly conserved regions; (2) Primer Generation: primers were designed using the NCBI Primer-BLAST tool and optimized for *H. pylori* (taxid:210), with parameters set to prioritize an annealing temperature (Ta) of 60°C and a GC content of 40–60%; (3) Physicochemical Evaluation: potential pairs were validated using PCR Primer Stats and the Sequence Manipulation Suite (SMS) to confirm thermodynamic stability and the absence of secondary structures (hairpins/dimers); and (4) Validation: finalized primers were synthesized by Macrogen Inc. and analyzed using StepOne™ v2.0 software.

Supplementary Table S1. qPCR primers for *Helicobacter pylori* virulence and housekeeping Genes

Gene	Primer	Sequence (5′–3′)	Product size (bp)	Tm / Annealing (°C)	Notes
cagA	HP-F	AACAGGCGATTTCAGTAAGG	134	60	Forward
cagA	HP-R	ACAACCCATTACCGACTAGG	134	60	Reverse
vacA	HP-F	AGCATCACACCGCAACAAAG	187	60	Forward
vacA	HP-R	CTGCTTGAATGCGCCAAACT	187	60	Reverse
babA	HP-F	AGCCGTGCTTTTAGCGATCA	94	60	Virulence
babA	HP-R	GCTCTGTTGACCATTGGCAT	94	60	Virulence
iceA	HP-F	GCCATGCGGGCTATTTTGAG	183	62	Virulence
iceA	HP-R	CCTGGAACACCTAGCTGCAA	183	62	Virulence
ureA	HP-F	CAAGTTGATGCTCCACTACGC	88	60	Housekeeping
ureA	HP-R	CAAACGTACCGCTTCCACAT	88	60	Housekeeping
hsp60	HP-F	CGGGTTGATGACGCATTGAG	223	59	Housekeeping
hsp60	HP-R	CGCCATCATAACCGGCATTG	223	59	Housekeeping
glmM	HP-F	AAACGCCTTAGTGAGTGCTT	108	61	Housekeeping
glmM	HP-R	GCATCACAGCGCATGTCTTC	108	61	Housekeeping

Section S2: Technical specifications & reaction parameters

Supplementary Table S2. Reagents and volumes for cDNA synthesis

cDNA synthesis was performed using the SuperScript™ III First-Strand Synthesis System (or RevertAid™ equivalent) to generate stable complementary DNA from total RNA.

Supplementary Table S2: List of reagents used for cDNA synthesis

Reagents	Volume used in µl
RNA	1
Random Hexamer primer	1
5 x r ^{x}n Buffer	4
RNA inhibitor	1
DNTPs	2
Revert Aid RT	1
Nuclease free water	10

Supplementary Table S3. qPCR reaction composition (SYBR™ Green)

Standardized reaction conditions for quantification of *cagA*, *vacA*, *babA*, and *iceA* transcripts using the ABI PRISM 7000/7500 system.

Supplementary Table S3: Reaction mixture for cDNA amplification

Reagents	Volume per 11 µL reaction
Cdna	1 µL
RNAse free water	4 µL
2X SYBR green master	4 µL
Forward primer	1 µL
Reverse primer	1 µL
TOTAL	11 µL

Supplementary Table S4. Thermal cycling profile for cDNA synthesis

Reverse transcription conditions used for cDNA synthesis from gastric tissue RNA.

Supplementary Table S4: Temperature and time intervals for the cDNA synthesis reaction

Step	Hold 1	Hold 2	Hold 3
Temperature (°C)	25	42	80
Time Interval (Min)	5	60	5

Section S3: Study population and power analysis

S3.1. Detailed inclusion criteria

- Histologically confirmed gastric malignancy (adenocarcinoma or other gastric cancer histology; cases) or non-malignant controls.
- Availability of sufficient biopsy tissue for DNA/RNA extraction (≥ 80% Tumor cell representation)
- Detectable *H. pylori* genomic DNA (≥1 target virulence gene).
- Adequate nucleic acid quality (A_{260}/A_{280} ratio thresholds, see S1.2).

S3.2. Detailed exclusion criteria

- Prior total gastrectomy
- Severe gastric comorbidities affecting tissue integrity.
- Insufficient tissue quantity or poor nucleic acid quality.
- Prior neoadjuvant therapy causing RNA degradation.

S3.3. Sample size formula and calculation

Sample size was estimated using a two-sample t-test formula:

$$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 \cdot 2\sigma^2}{\delta^2}$$

Where:

- $Z_{1-\alpha/2} = 1.96$ (95% confidence), $Z_{1-\beta} = 0.84$ (80% power)
- $\sigma = 0.5$ (assumed standard deviation (SD) $2^{-}\Delta Ct$), $\delta = 0.3$ (expected mean difference)

This yielded a minimum of 44 samples per group. The final cohort included 159 *H. pylori*-positive samples, distributed across stages (see main text Limitations for discussion of per-stage power).

S3.4. Participant flow diagram (Supplementary Figure S1)

