A bacterial small RNA regulates the adaptation of *Helicobacter pylori* to the host environment

Ryo Kinoshita-Daitoku, Kotaro Kiga, Masatoshi Miyakoshi, Ryota Otsubo, Yoshitoshi Ogura, Takahito Sanada, Zhu Bo, Tuan Vo Phuoc, Tokuju Okano, Tamako Iida, Rui Yokomori, Eisuке Kuroda, Sayaka Hirukawa, Mototsugu Tanaka, Arpana Sood, Phawinee Subsomwong, Hiroshi Ashida, Tran Thanh Binh, Lam Tung Nguyen, Khien Vu Van, Dang Quy Dung Ho, Kenta Nakai, Toshihiko Suzuki, Yoshio Yamaoka, Tetsuya Hayashi & Hitomi Mimuro

Long-term infection of the stomach with *Helicobacter pylori* can cause gastric cancer. However, the mechanisms by which the bacteria adapt to the stomach environment are poorly understood. Here, we show that a small non-coding RNA of *H. pylori* (HPnc4160, also known as IsoB or NikS) regulates the pathogen’s adaptation to the host environment as well as bacterial oncoprotein production. In a rodent model of *H. pylori* infection, the genomes of bacteria isolated from the stomach possess an increased number of T-repeats upstream of the HPnc4160-coding region, and this leads to reduced HPnc4160 expression. We use RNA-seq and iTRAQ analyses to identify eight targets of HPnc4160, including genes encoding outer membrane proteins and oncoprotein CagA. Mutant strains with HPnc4160 deficiency display increased colonization ability of the mouse stomach, in comparison with the wild-type strain. Furthermore, HPnc4160 expression is lower in clinical isolates from gastric cancer patients than in isolates derived from non-cancer patients, while the expression of HPnc4160’s targets is higher in the isolates from gastric cancer patients. Therefore, the small RNA HPnc4160 regulates *H. pylori* adaptation to the host environment and, potentially, gastric carcinogenesis.

---

1 Department of Infection Microbiology, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan. 2 Division of Bacteriology, Department of Infectious Diseases Control, International Research Center for Infectious Diseases, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. 3 Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. 4 Department of Bacteriology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. 5 Department of Microbiology, Department of Infectious Medicine, Kurume University School of Medicine, Kurume Fukuoka, Japan. 6 Department of Environmental and Preventive Medicine, Faculty of Medicine, Oita University, Yufu Oita, Japan. 7 Department of Endoscopy, Cho Ray Hospital, Ho Chi Minh, Vietnam. 8 Department of Bacterial Pathogenesis, Infection and Host Response, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan. 9 Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. 10 Division of Nephrology and Endocrinology, The University of Tokyo School of Medicine, Tokyo, Japan. 11 Department of GI Endoscopy, 108 Central Hospital, Hanoi, Vietnam. 12 Present address: Laboratory of Pharmaceutical Integrated Omics, Department of Pharmaceutical Engineering, Facility of Engineering, Toyama Prefectural University, Toyama, Japan. Email: mimuro@biken.osaka-u.ac.jp
The gastric pathogen Helicobacter pylori infects approximately half of the world’s population and increases the risk of developing peptic ulcers, chronic gastritis, intestinal metaplasia, and gastric cancer. The genomic diversity characteristic of H. pylori is crucial for its establishment of persistent infections in hosts and confers its adaptability to extreme gastric environments. A well-known example of H. pylori’s strong diversity is the presence of outer membrane proteins (OMPs), which are highly antigenic cell surface proteins and are thought to allow H. pylori to escape from host immunity during chronic infection. Mutations in OMP genes are characterized by simple repetitive sequences such as mononucleotide repeats (e.g., poly-T) and dinucleotide repeats (e.g., CT-repeats). The expansion and contraction of these simple repetitive sequences result in a genetically heterogeneous bacterial population, with phase variations controlled by the ON/OFF gene expression of the proteins. However, whether the H. pylori adaptation to the host is controlled only by the diversity of mRNA containing simple repetitive sequences remains unclear. Because of the prevalence of small regulatory RNAs (sRNAs) in H. pylori, it is also possible that a heterogeneously mutated sRNA changes the expression levels of its target mRNAs or a mutation in the target sequence results in dysregulation by an sRNA. To understand the adaptation mechanisms of H. pylori, we analyzed bacterial gene mutations acquired by H. pylori over the course of infection using an experimental animal infection system in hosts of identical genetic background.

Here, we show that a small non-coding RNA HPrnc4160 of H. pylori regulates the pathogen’s adaptation to the host environment as well as bacterial oncoprotein production. We revealed that during the infection, a T-repeat length upstream of the coding region for HPrnc4160 is elongated in H. pylori and that this elongation leads to decreased expression of HPrnc4160. Through RNA-seq and iTRAQ analysis we identified several targets of HPrnc4160, including cagA, a known carcinogenic protein. Importantly, we found that HPrnc4160 expression is lower in clinical isolates from gastric cancer patients than in isolates derived from non-cancer patients, while the expression of HPrnc4160’s targets is higher in the isolates from gastric cancer patients. Our findings demonstrate that the small RNA HPrnc4160 regulates H. pylori adaptation to the host environment and, potentially, gastric carcinogenesis.

Results
Identification of mutations accumulated during infection. To analyze the bacterial gene mutations acquired by H. pylori during persistent infection, Mongolian gerbils (n = 10) were inoculated with the H. pylori ATCC 43504 wild-type strain for 8 weeks with stable gastric colonization. H. pylori were isolated from the infected stomachs (4 clones per gerbil, total 40 clones; Fig. 1a) and then analyzed by comparative whole-genome sequencing (Supplementary Fig. 1a, Supplementary Data 1 and 2). By integrating the genomic positions of these mutations, we identified 13 regions (R1, R3–R5, R7–R8, R10–R16) in which mutations were introduced in 50% or more of the strains (Supplementary Fig. 1a, Supplementary Data 2).

To investigate whether these mutated regions affect gene expression, transcripts in the mutated regions were quantified in isolates obtained from gerbils. Among the 15 corresponding coding (CDS) and non-coding RNA sequences (HP0947, babA, tpiA, jhp1163, HP0811, HPrnc4160, HPrnc4170, jhp0540, araS, pldA, saba, HP1354, hopZ, tlpB, and HPB8_818), the expression of HPrnc4160 showed the greatest fluctuation (Fig. 1b, Supplementary Fig. 1a, b). Similar results were obtained from strains (2 clones per mouse, total 10 clones) isolated from C57BL/6 mice infected with H. pylori ATCC 43504 wild-type strain for 8 weeks (Supplementary Fig. 1a–c, Supplementary Data 3 and 4).

HPrnc4160 (IsOB) was previously proposed as a cis-acting sRNA for HPrnc4170 (aapB small ORF homolog) (Supplementary Fig. 2a). HPrnc4160 and its upstream T-repeat region are conserved in H. pylori strains, but the T-repeat length in the promoter region is highly variable (Supplementary Fig. 2b). One to four additional thymidine bases were inserted into the T-repeat of our isolates from rodents, with the repeat length increasing with duration of period (Fig. 1c, Supplementary Fig. 2c–f). However, expansion of this region was not observed in long-term in vitro culture of wild-type strain (Supplementary Fig. 2g). Next, we analyzed changes in HPrnc4160 expression according to the T-repeat length. In strains recovered from the H. pylori-infected stomachs of gerbils, HPrnc4160 expression decreased with the expansion of the T-repeat (Fig. 1d, e). To exclude the effects of mutations other than those of T-repeats in strains recovered from gerbils, we further analyzed the RNA expression of HP0811, HPrnc4170, and HPrnc4160 in isogenic variants of ATCC 43504 wild-type strains harboring different T-repeat lengths at the promoter region of HPrnc4160 (T0 to T20). Fig. 1f, g (Supplementary Fig. 3a–c). The expression of HPrnc4160 with varying T-repeat lengths was multiphasic: low in T1, high in T4, high in T14 (number of repeats in ATCC 43504 wild-type), low in T16, and intermediate in T18 (Fig. 1f, g). Expression levels of hpnc4170 and its downstream HP0811 were not affected by the T-repeats (Supplementary Fig. 3b–c), indicating that HPrnc4160 is unlikely to act as a cis-acting sRNA.

Identification of HPrnc4160 target genes. Helicobacter pylori utilizes RNA-mediated regulation in trans through relatively short base-pairing with multiple mRNAs from different loci despite the absence of known RNA matchers such as Hfq and ProQ. To identify the target mRNA of HPrnc4160, we generated a Δhpnc4160-hpnc4170 strain, in which both HPrnc4160 and HPrnc4170 on the complementary strand were deleted. mRNA and protein expression analysis identified eight differentially expressed factors in the mutant strain compared to the wild-type (cagA, HofC, HelpY_1262, HP0410, horB, omp14, HopE, and HP1227; P < 0.001 by RNA-seq; P < 0.01 by isobaric tag for relative and absolute quantitation labeling and liquid chromatography-tandem mass spectrometry analysis) (Fig. 2a–c, Supplementary Data 5 and 6). Notably, five of these eight factors (HofC, HP0410, HorB, Omp14, and HopE) were annotated as outer membrane proteins. More strikingly, the expression of cagA mRNA and protein, which is a known bacterial oncoprotein, was most noticeably increased (Fig. 2a, b, Supplementary Data 5 and 6).

We analyzed whether the mRNA expression of these eight factors depended on the presence of HPrnc4160. Levels of HPrnc4160 tended to decrease, with T16 exhibiting the lowest value, whereas the mRNA of the eight candidates tended to increase, with T16 exhibiting the highest value (Fig. 1f, g; Supplementary Fig. 3a–d). The expression of these target mRNAs and the presence of HPrnc4160 showed a strong inverse correlation (Spearman correlation coefficient (r) = −0.8234 to −0.7312) (Supplementary Fig. 3d). The HPrnc4160 overexpression strain (WT/pHel2-hpnc4160) showed significantly decreased expression of the target mRNAs. In contrast, in the Δhpnc4160-hpnc4170 strain, the mRNA expression of each target increased compared to in the wild-type (Supplementary Fig. 3e–h). As the Δhpnc4160-hpnc4170 strain lacks the HPrnc4170 sequence in the complementary strand of HPrnc4160, we constructed a Δhpnc4160-hpnc4170/pHel2-hpnc4160 strain complementing only HPrnc4160 to confirm the effect of the HPrnc4170 sequence on HPrnc4160 target mRNA expression. Compared to the
**H. pylori** isolates from stomachs of: Mongolian gerbils, n=40 C57BL/6 mice, n=10

**Fig. 3h**). These data indicate that the eight target candidates are negatively regulated by HPnc4160. (Supplementary Fig. 4a, b). The direct binding of HPnc4160 to the 5'UTR of the seven target genes except *cagA* were confirmed by Electrophoretic mobility shift assays (Fig. 3a). Instead, we identified five putative target sites in the CDS of *cagA* (Type 1 at 2344 nt from the start codon, and Type 2 at four positions of 2838, 2940, 3042, and 3144 nt) (Fig. 3b–d). We confirmed the direct binding of partial *cagA* CDS (positions 2778–3236 nt from start codon of *cagA*) to HPnc4160 (Fig. 3e). This binding was abolished for Non-Binding (NB)-*cagA* RNA in which the four Type 2 HPnc4160 binding sequences were mutated and the amino acid sequence of CagA was preserved (Fig. 3b, e–g). The abolished binding of NB-*cagA* RNA was

\[ \Delta \text{hpnc4160} \]

\[ \Delta \text{hpnc4170} \]

strains recovered from gerbils (T repeat number of 14, 15, 16, 17, 18, n = 6) and genetically modified T-repeat lengths (n = 3) in RT-PCR (d, f) and Northern blot (e, g). Shown are the means ± s.d (d, f). Data were processed by Grubbs’ test for outliers. *P = 0.0379, **P = 0.0088, ****P < 0.0001 by Sidak’s multiple comparison tests (two sided) (d). Data are representative of two independent experiments (d–g). Source data are provided as a Source Data file.

**Fig. 1 H. pylori acquires T-repeat extension upstream of HPnc4160 to decrease its expression during infection in vivo.** a Schematic showing experimental strategy. b RNA expression levels of ORFs or nearby genes of genome regions mutated in >50% of the strains recovered from gerbils, plotted against mutation rates. c DNA sequence around the HPnc4160 and T-repeat sequence of the strains recovered from gerbils. d–g Relative hpnc4160 expression in *H. pylori* strains recovered from gerbils (T repeat number of 14, 15, 16, 17, 18, n = 6) and genetically modified T-repeat lengths (n = 3) in RT-PCR (d, f) and Northern blot (e, g). Shown are the means ± s.d (d, f). Data were processed by Grubbs’ test for outliers. *P = 0.0379, **P = 0.0088, ****P < 0.0001 by Sidak’s multiple comparison tests (two sided) (d). Data are representative of two independent experiments (d–g). Source data are provided as a Source Data file.

In the 5'UTR of seven identified target genes except *cagA*, we found a sequence that binds complementarily to the loop region of the HPnc4160. (Supplementary Fig. 4a, b). The direct binding of HPnc4160 to the 5'UTR of the seven target genes except *cagA* were confirmed by Electrophoretic mobility shift assays (Fig. 3a). Instead, we identified five putative target sites in the CDS of *cagA* (Type 1 at 2344 nt from the start codon, and Type 2 at four positions of 2838, 2940, 3042, and 3144 nt) (Fig. 3b–d). We confirmed the direct binding of partial *cagA* CDS (positions 2778–3236 nt from start codon of *cagA*) to HPnc4160 (Fig. 3e). This binding was abolished for Non-Binding (NB)-*cagA* RNA in which the four Type 2 HPnc4160 binding sequences were mutated and the amino acid sequence of CagA was preserved (Fig. 3b, e–g). The abolished binding of NB-*cagA* RNA was

\[ \Delta \text{hpnc4160} \]

\[ \Delta \text{hpnc4170} \]

strain, the mRNA expression levels of the candidates were decreased in the HPnc4160 complemented *Δhpnc4160*-*hpnc4170* /pHel2-*hpnc4160* strain (Supplementary Fig. 3h). These data indicate that the eight target candidates are negatively regulated by HPnc4160.

Trans-acting sRNAs generally form base-pairing in the 5' untranslated region (UTR) or CDS of target mRNAs to repress or activate gene expression at the post-transcriptional level\(^\text{15,16}\). In *H. pylori*, the best-characterized sRNA RepG (HPnc5490) binds to G-repeats in 5'UTR of *tlpB* mRNA to repress translation and destabilize mRNA\(^\text{9,10}\). As *H. pylori* lacks RNase E/G homologs\(^\text{13,17}\), which promote mRNA degradation by sRNAs, other endoribonuclease such as RNase III may degrade target mRNA accompanied by translation inhibition.

\[ \text{HPnc4160} \]

5S rRNA

\[ \text{WT} \]

\[ \Delta \text{hpnc4160} \]

\[ \Delta \text{hpnc4170} \]
restored by complementary mutations in HPnc4160 at positions 64, 67, and 73 (HPnc4160 mut; Fig. 3e, h), demonstrating direct base-pairing between HPnc4160 and cagA. In addition, we found that H. pylori RNase III recombinant protein degraded the dsRNA upon binding between HPnc4160 and biotin-labeled partial cagA mRNA, but not NB-cagA RNA (Fig. 3i, Supplementary Fig. 4c). These data indicate that HPnc4160 controls cagA at the post-transcriptional level by binding to multiple binding sequences present in its CDS region and possibly promoting the degradation of mRNA by RNase III.

Effects of HPnc4160 on H. pylori pathogenicity. We further analyzed CagA, which has been shown to be deeply involved in pathogenesis. The H. pylori expressing NB-cagA, in which all five HPnc4160-binding sequences were mutated but the amino acid sequence was preserved, showed significantly increased expression of cagA mRNA and protein to the same extent as that of the Δhpnc4160-hpnc4170 strain (Supplementary Fig. 5a–c). The increased expression of cagA mRNA and protein was negated by complementary mutations in hpnc4160 (hpnc4160 mut) (Supplementary Fig. 5a–c), indicating that HPnc4160 negatively controls the expression of cagA mRNA and protein in H. pylori. H. pylori injects CagA proteins into the host epithelium via a Type IV secretion system (TFSS), and then intracellular CagA proteins are phosphorylated at tyrosine residues (pY) by the host Src/Abl kinase. Using pY-CagA-specific antibodies, we confirmed that intracellular CagA increased in NB-cagA-infected cells, accompanied by an increase in total CagA, but not in hpnc4160 mut strain-infected cells (Supplementary Fig. 5d). Intracellular CagA induces AGS cell motility (scattering/hummingbird phenotype). More elongated AGS cells were observed during infection with Δhpnc4160-hpnc4170 or NB-cagA strains vs in the wild-type or Δhpnc4160-hpnc4170/pHel2-hpnc4160 strain-infected cells (Supplementary Fig. 5e, f). The amount of IL-8 secreted from H. pylori-infected cells, which is primarily induced by intracellular CagA, was higher in NB-cagA strain-infected cells than in wild-type strain or hpnc4160 mut strain-infected cells (Supplementary Fig. 5g). These results suggest that the binding of HPnc4160 to cagA mRNA is critical for controlling the amount of functional CagA protein in H. pylori.

We further assessed changes in CagA activity dependent on T-repeat length of hpnc4160. Western blot analysis showed that 48 out of 50 strains recovered from rodents infected with ATCC 43504 at 8 weeks post infection expressed VirB7 (TFSS protein) and retained the ability to translocate CagA into host epithelium (Supplementary Fig. 5h, i). Intracellularly translocated CagA protein tends to be higher in strains with T-repeat numbers 16 and 17 with low levels of HPnc4160, and lower in T14 and T18 with high levels of HPnc4160. A similar trend of increased IL-8 expression with T-repeat length was observed in the expression of IL-8.
IL-8 in AGS cells infected with the T-repeat mutated strains (Supplementary Fig. 5).

Next, to understand the significance of the Hpnc4160 control mechanism in bacterial adaptation to the host, mice were orally inoculated with either wild-type, Δhpnc4160-hpnc4170, Δhpnc4160-hpnc4170/pHel2-hpnc4160, and NB-cagA, and the bacterial loads in the stomach were analyzed. At three days post-infection, a timepoint prior to T-repeat length extension (Supplementary Fig. 2e), levels of bacterial colonization and Cxcl2 mRNA in the stomach were found to be significantly increased for the Δhpnc4160-hpnc4170 strain compared to wild-type, whereas those of the Δhpnc4160-hpnc4170/pHel2-hpnc4160
strains were equivalent to wild-type (Fig. 4a, b). Notably, both colonization levels and Cxcl2 mRNA levels were slightly higher following NB-cagA infection than following wild-type infection. This result indicates that factors controlled by HPnc4160, other than CagA, are important for bacterial adaptation and host gastritis development (Fig. 4a, b). Importantly, the bacterial colonization levels at 8 weeks post infection was significantly decreased in the Δhpnc4160-hpnc4170-infected animals, as compared to the wild-type infection (Fig. 4c), indicating that HPnc4160 is beneficial for long-term colonization of H. pylori.

Effects of T-repeats length diversity on adaptation. Since H. pylori is known to be transmitted from human to human, we conducted reinfection experiments using isolates from mice with different T-repeats (T16 in Hp6 and T18 in Hp8) with equivalent in vitro growth rates (Supplementary Fig. 6a). Similar to the wild-type strain (Fig. 4a, c), both strains used for reinfection showed almost the same number of colonized bacteria in the stomachs from 3 days to 8 weeks after infection (Supplementary Fig. 6b). Interestingly, the strains used for reinfection displayed increased repeat lengths with increases in infection period, similar to the wild-type strain (Supplementary Figs. 2c and 6c). Furthermore, the strains with longer T-repeats (T16 and T18) were more likely to acquire the T-repeat elongation and adapt to the mice stomach than the strains with shorter T-repeats (T1 and T4) (Fig. 5a, b). These data indicate that repetitive sequence length variation of the T-repeat is advantageous for H. pylori persistent infection.

T-repeat length and HPnc4160 expression in clinical isolates. Because one of the HPnc4160 target factors was CagA, which is known as the strongest risk factor for H. pylori-related gastric cancer, we examined the expression of HPnc4160 target genes in clinical isolates from non-cancer patients and patients with cancer. Sequence analysis of clinical isolates showed that T-repeat regions were longer in strains derived from patients with cancer than in those derived from non-cancer patient (Fig. 6a, Supplementary Fig. 7, Supplementary Data 7). As shown in Fig. 6b, isolates from patients with cancer had lower levels of hpnc4160 but increased expression of six factors controlled by HPnc4160 (cagA, horB, hopE, omp14, hofC, and HP0410) compared to isolates from non-stomach cancer patients (Fig. 6b, Supplementary Fig. 7b, c). The relevance of HPnc4160 and H. pylori-related gastric cancer risk was also assessed by in vitro infection experiments. AGS cells infected with H. pylori expressing low levels of HPnc4160 displayed increased expression of CDX2, an indicator of intestinal epithelialization in the precancerous state (Supplementary Fig. 7d). In summary, our mutational analysis studies revealed a functional small RNA that regulates the pathogenicity of H. pylori and may have implications for gastric cancer development.

Discussion
We have demonstrated that mRNA expression of CagA and OMs is suppressed by HPnc4160 at the onset of H. pylori infection and that over the course of infection, thymidine repeats are inserted into the upstream region of hpnc4160, decreasing HPnc4160 expression, resulting in increased target mRNA expression; these factors contribute to bacterial adaptation to the host environment and potentiate gastritis and gastric oncogenesis (Fig. 7). Control of gene expression by varying the number of repeat sequences is a known control mechanism employed by H. pylori.23 The present study suggests that repeat sequences of the H. pylori genome are important, not only as an ON/OFF mechanism for protein expression, including that of cell adhesion factors SabA and BabA24, but also in sRNA expression. Since T-repeat is reported to affect the local DNA structure for RNA polymerase binding, by shifting the axial distance between the core promoter and upstream promoter elements, we speculate that steric distance between the promoter and the HPnc4160 transcription initiation site fluctuates as the T-repeat length increases or decreases, thereby modulating HPnc4160 expression; these factors contribute to bacterial adaptation to the host environment and potentiate gastritis and gastric oncogenesis (Fig. 7).
**Fig. 5 Time-dependent changes in bacterial-host adaptation and the T-repeat length.** a, b Time-dependent changes in the bacterial load in the infected mouse stomach (a) and length of the T-repeat region (b) in T1, T4, T16 and T18 strains recovered from mice. Data are presented as the median with interquartile range. Statistical significance determined by uncollected Dunn’s multiple comparison test (two-sided). *P = 0.0174, **P = 0.0066 (a); *P = 0.0294, **P = 0.0014, ****P < 0.0001 (b, T16); **P = 0.0012, ***P = 0.0002 (b, T18); ns, not significant. Source data are provided as a Source Data file.

**Fig. 6 T-repeat length and HPnc4160 expression in clinical isolates.** a, b T-repeat length upstream of HPnc4160 (a) and expression of indicated mRNAs (b) in clinical isolates from non-cancer (non-cancer) and patients with cancer (cancer). Data are presented as means with 95% confidence interval (a) or medians with interquartile range (b). Statistical significance determined by two-tailed Mann-Whitney test (a, b). *P = 0.0157 (a); *P = 0.0215 (hpnc4160), ****P < 0.0001 (cagA), ***P = 0.0004 (horB), *P = 0.0025 (hopE), *P = 0.0110 (omp14), *P = 0.0122 (hofC), *P = 0.0205 (HP0410) (b). Source data are provided as a Source Data file.
proteins\textsuperscript{27,28} (Fig. 3b). Generally, regulation of mRNA expression by sRNA is induced by binding at one site. Multiple binding sequences in the \textit{cagA} CDS may facilitate efficient reduction of \textit{cagA} transcript levels in response to HPnc4160. While the present manuscript was under revision, Eisenbart et al. published a paper describing that the nickel-responsive transcriptional regulator NikR regulates the expression of HPnc4160 (which the authors of that study renamed as ‘NikS’), and HPnc4160 represses multiple major virulence genes, including \textit{cagA} and \textit{vacA} in vitro\textsuperscript{29}. They showed that HPnc4160 binds to the 5'UTR of the \textit{cagA} mRNA of strain G27. Although we could not observe HPnc4160 binding to the 5'UTR region of the ATCC 43504 \textit{cagA} (Fig. 3a), they also identified a group of HPnc4160-regulating pathogenic factors that overlapped with those we found. Therefore, their report is in line with our finding that the sRNA HPnc4160 is a master regulator of \textit{Helicobacter} pathogenicity.

All identified HPnc4160 targets encode large proteins that require greater energy expenditure for production (Fig. 2a–b, Supplementary Data 5 and 6). As the growth of bacterial cells in vitro may not require these pathogenic factors, \textit{hpnc4160} expression may be higher whereas its targets are suppressed in vitro. Upon entering the host stomach, \textit{H. pylori} may activate a mechanism that decreases HPnc4160 production, thereby simultaneously increasing expression of OMPs and CagA, allowing the bacteria to efficiently adapt to environmental changes and colonize the gastric mucosa.

Because CagA is a carcinogenic protein related to gastric cancer, an increase in CagA levels achieved by down-regulation of HPnc4160 may correlate with an increased risk of gastric cancer. Additionally, we demonstrated that several OMPs are also regulated by HPnc4160. These targets are likely involved in host adaptation, given that the HPnc4160 knockout strains display increased bacterial colonization and inflammatory cytokine expression, phenotypes not solely dependent on CagA production (Fig. 4a, b). Collectively, we find HPnc4160 behaves as a master regulator of \textit{H. pylori} host adaptation and may serve to potentiate gastric carcinogenesis.

\section*{Methods}

\textbf{Data reporting.} No statistical methods were used to predetermine sample size, the experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

\textbf{Strains and culture conditions.} The isogenic mutants \textit{ΔcagA} and \textit{ΔvirB7} of \textit{Helicobacter pylori} strain ATCC 43504 were constructed in our laboratory\textsuperscript{11}. The strain PMSS1 was generously provided by Dr. Manuel R. Amieva (Stanford University)\textsuperscript{30}. \textit{H. pylori} was cultured on Trypticase soy agar with 5\% (v/v) sheep’s blood (Thermo Fisher Scientific, Waltham, MA, USA) for 2 days at 37 °C in microaerobic conditions. Bacterial colonies were suspended in Brucella broth (Thermo Fisher Scientific) supplemented with 5\% (v/v) inactivated fetal bovine serum (FBS; Thermo Fisher Scientific), adjusted to an optical density (OD) of 0.05 at 600 nm, and incubated for 15 h at 37 °C with gentle agitation under microaerobic conditions.

The AGS human gastric epithelial cell line (ATCC CRL-1739) was maintained in Dulbecco’s modified eagle medium (DMEM)/F-12 (Thermo Fisher Scientific) containing 10\% (v/v) FBS. AGS cells were seeded in six-well plates and grown to ~80\% confluence to be used for western blot analysis. For immunofluorescence microscopy, cells were seeded in six-well plates with cover glass and grown to ~80\% confluence\textsuperscript{11}.
Antibodies and immunohistochemical reagents. The anti-TyrP-CagA (anti-
P-CagA; for Western blot; 1:100; for immunofluorescence microscopy, 1:10),
anti-UreA (for Western blot; 1:2000), and anti-VariBf (for Western blot) poly
clonal antibodies were made from the antisera of rabbits immunized with recombinant proteins or peptides14. Anti-CagA polyclonal antibodies (for Western
blot; 1:1000) were purchased from ASTRAL Biologicals (CA, USA), anti-actin
monoclonal antibodies (for Western blot; 1:1000) were from MERCK (Darmstadt,
Germany), horseradish peroxidase (HRP)-labeled anti-rabbit (HRP) (for Western
blot, 1:10000) and HRP-labeled anti-mouse IgG (for Western blot, 1:5000), and
fluorescein isothiocyanate-labeled anti-rabbit IgG (for immunofluorescence
microscopy, 1:100) was from Jackson ImmunoResearch Laboratories Inc. (PA,
USA). DAPI (for staining, 300 nM) was from Sigma-Aldrich (MI, USA), and rhodamine-phalloidin (for staining, 1:100) was from Thermo Fisher Scientific (MA, USA).

In vitro passage experiment. H. pylori ATCC 43504 was recovered from frozen
stock and cultured on 6% polycrylamide/7 M urea gels in 1XTE buffer for 3 h at 250 V and was electrophoresed onto Hybond-XL nylon membrane (GE Healthcare) for 1 h at 50 V using Biometra Eco-Maxi system (Analytik-Jena). The membrane was crosslinked by 120 ml/cm² UV light. Oligonucleotides JVO-2624 and JVO-0485 to detect Hpnc1460 and 5 S rRNA, respectively, were 5'-end-labeled with [32P]-ATP by T4 polynucleotide kinase (Nippon Gene) and purified dithiothreitol and agar plate (4 μg/mL kanamycin). After pre-hybridization (Amersham), the [32P]-labeled probe was hybridized at 42 °C overnight. Membrane was washed in three steps in 5X SSC/0.1% SDS, 1X SSC/0.1% SDS and 0.5X SSC/0.1% SDS buffers for 15 min at 42 °C. Signals were visualized on Typhoon FLA7000 scanner (GE Healthcare).

Genetic manipulation. Construction of plasmids for producing gene-deficient mutants. Isogenic gene null mutants derived from ATCC 43504 were constructed by insertional mutagenesis as follows37. Using the extracted H. pylori ATCC 43504 genome as a template, DNA fragments containing the 500 bp upstream region and the 500 bp downstream region of the target gene were amplified by PCR using the primers (CagA KO up Xhol, CagA KO up EcoRI, CagA KO down BamHI/CagA KO down NotI, Hpnc1460/Hpnc4170 KO up KpnI, Hpnc1460/Hpnc4170 KO up ClaI, Hpnc1460/Hpnc4170 KO down BamHI, Hpnc1460/Hpnc4170 KO down SacI) listed in Supplementary Data 8. The DNA fragments were introduced at both sides of aphA3 (which confers kanamycin resistance) in pBlueScript II SK (+) plasmids. The fragments from the resulting plasmids were introduced into H. pylori by electroporation.

Construction of non-marker H. pylori mutants. To construct non-marker H. pylori mutants, ATCC 43504 flaA and cagA promoters and terminators were cloned into pBlueScript II SK (+) Small aphA3 Small, and sucB was cloned into the EcoRI site (pKSb plasmid).

Mid-log phase (OD600 = 0.5–0.7) of H. pylori in 20 ml liquid culture were washed twice with ice-cold 10% glycerc and resuspended by 200 μl 10% ice-cold glycerc. One microgram of pKSb vector containing the target mutation and the bacterial cells were mixed at 4 °C and electroporated by a MicroPulser (Bio-Rad) at the Ec2 (2.5 kV) setting. After 4 h of incubation at 37 °C in a microaerobic condition, cells were plated on 5% sheep blood agar TSAII plates supplemented with 4 μg/mL kanamycin and incubated 2–3 days at 37 °C under microaerobic conditions. Four single colonies were seeded on new 5% sheep's blood agar TSAII plates supplemented with 4 μg/mL kanamycin and incubated for an additional 2 days. Each colony was picked up and cultured in Brucella broth containing 5% FBS at 37 °C under microaerobic conditions until H. pylori reached the mid-log phase. The medium (0.1 ml) was plated on 5% sheep's blood agar plates supplemented with 2.5 μg sucrose and cultured for 2 days. Each colony was then seeded onto a new 5% sheep's blood agar plate without antibiotics and incubated for 2 days. At the same time, colonies were also seeded onto a different agar plate with 4 μg/mL kanamycin. The kanamycin-resistant H. pylori vaccine has been abolished. Surviving H. pylori were then transferred to liquid culture, and the genome sequence was confirmed by Sanger sequencing.

Construction of point-mutated H. pylori. The recombinant plasmids to establish various mutants T15, T16, T17, T18, and T19 H. pylori strains in the upstream region of hpnc1460 were constructed by PCR using primers from the strains isolated from gerbils after 8 weeks as a template (primers: pKSB- HPnc1460 Point mut Apal and pKSb- HPnc1460 Point mut Xhol; listed in

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-22317-7 | www.nature.com/naturecommunications
**EMSA (electrophoretic mobility shift assay).** cDNAs fragments of the entire small RNA hpn4160 region; 150-bp fragments of the 5′UTR regions [from 100 bases upstream of the ribosome binding region (RBS), to 50 bases downstream of the RBS] of hp0410 gene, hp0486 gene, torhB gene, hp0671 gene, hopE gene, cagA gene, hp1227 gene, and help_1262 gene; and 459 bp of the total cDNA containing the four hpn4160-binding regions near the 3′ tail of cagA, were amplified by PCR (primers: Small RNA Hpn4160 XhoI, Small RNA Hpn4160 EcoRI; HPO410 150 bp XhoI, HPO410 150 bp EcoRI; HP0671 150 bp XhoI, HP0671 150 bp EcoRI; HP0486 150 bp XhoI, HP0486 150 bp EcoRI; HP0635 150 bp XhoI, HP0635 150 bp EcoRI; HP0227 1555 150 bp XhoI, HP0227 1555 150 bp EcoRI; HP0126 150 bp XhoI, HP0126 150 bp EcoRI; CagA-B coding XhoI, CagA-B coding EcoRI; listed in Supplementary Data 8) and the ATCC 43504 genome as a template. The PCR products were cloned into the pBluescript SK (+) plasmid downstream of the T7 promoter region. pBlueScript SK (+) plasmid containing HPn4160 mut (compensatory used for NB-cagA mutation) was generated according to the instruction of PrimSTAR Mutagenesis Basal Kit (Takara) using pBluescript SK (+) hpn4160 as a template (primers: HPn4160_pointmut_primer_f, HPn4160_pointmut_primer_r; listed in Supplementary Data 8). The NB-cagA mutant RNA used for the gel shift assay was amplified with a T7 promoter by PCR (T7 promoter cagA-AN EMSA PCR s, T7 promoter cagA-AN EMSA PCR r; listed in Supplementary Data 8). The NB-cagA mutant RNA was added to the labeled RNA, and binding buffer (10 mM HEPES pH 7.3, 1 mM MgCl2, 50 mM KCl, 5% glycerol) was added to a culture was developed by SDS-PAGE, and the gel was stained with Coomassie Brilliant Blue to confirm that no contaminants were observed in the final product. The protein concentration was determined by absorbance at 280 nm.

**Clavage assays.** The cDNA of 720 bps of H. pylori rnasE III was amplified by PCR (primers: PGEK-6P-1 rnasEIII Kb; PGEK-6P-1 rnasEIII NotI-r, listed in Supplementary Data 8; and template: genome DNA from ATCC 43504 strain). The cDNA was cloned into a pGEX6P-1 vector (GE). E. coli BL21 transformed with the plasmids were subjected to shaking culture in LB broth containing 100 μg/ml ampicillin at 37 °C with constant shaking at 200 r.p.m. Protein expression was induced by IPTG, with IPTG to a final concentration of 0.1 mM, at 4 h. The bacteria were collected by centrifugation, and pellets were subjected to GST-fusion protein purification using Glutathione Sepharose 4B (GE) according to the manufacturer's instruction. RNase III protein was excised by PreScission Protease according to the manufacturer's instructions. The purified protein derived from 6.7 mL of bacterial culture was developed by SDS-PAGE, and the gel was stained with Coomassie Brilliant Blue to confirm that no contaminants were observed in the final product. Protein concentration was determined by absorbance at 280 nm.

**Nuclease assays using RNase III were performed using purified H. pylori recombinant RNase III. The gel shift assay protocol described above was followed, except that an RNase III-specific buffer (25 mM Tris pH 7.5, 50 mM NaCl, 50 mM KCl, 10 mM MgCl2, 1 mM DTT) was used instead of Binding Buffer. 3′-biotinyl tagged cDNA was incubated with ice at 5 μM of small RNA Hpn4160 for 20 min. RNase III was then added at a final concentration of 300 nM, and the reactions were incubated for 1 min at 37 °C. The samples were mixed with 5 μl of native loading buffer before loading on a pre-cooled native 6% poly-acrylamide (PAA), 0.5x TBE gel. Gels were run in 0.5x TBE buffer at 30 mA per gel for 2 h.

**ELISA.** AGS cells were co-incubated with H. pylori at a multiplicity of infection (MOI) of 100 for 12, 24 and 36 h at 37 °C in a 5% CO2 environment in 24 well plates. The supernatants were collected and stored at -30°C. Enzyme-linked immunosorbent assays (ELISAs) for human IL-8 were performed using the Human IL-8 ELISA Kit (ThermoFisher SCIENTIFIC) according to the manufacturer’s instructions. The results are expressed as the means ± SEM from triplicate experiments.

**Immunofluorescence microscopy.** AGS cells were infected with H. pylori at a moi of 100 for 6 h at 37°C in a 5% CO2 environment. The cells were fixed with 4% (w/v) paraformaldehyde-PBS at room temperature for 10 min. The cells were then washed with PBS 3 times, and blocked with Saponin buffer (10% (v/v) BSA) for 1 h at room temperature. Cells were then incubated with 0.5% (v/v) saponin at 4°C for 30 min. Anti-bodies and fluorescent stains used for staining were DAPI, rhodamine-phalloidin (Thermo Fisher SCIENTIFIC, MA, USA), and anti-pY-CagA11,29, Confocal laser.
scanning microscopy (CLSM) image acquisition was performed using a Zeiss LSM 800 confocal laser scanning microscope with ZEN 2.3 software (Carl Zeiss, Jena, Germany).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The raw read sequences and assembled scaffold sequences of rodents or clinical isolates used in this study are available in the DDBJ/EMBL/Genbank databases under Bioproject accession numbers SAMD00178987 to SAMD00178935, SAMD00204547, and SAMD00204646. Sequence data of the clinical isolates used in this study are available in the DDBJ/EMBL/Genbank databases with the accession codes listed in Supplementary Data 7. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary files. Source data are provided with this paper.

Received: 25 December 2020; Accepted: 10 March 2021; Published online: 09 April 2021

**References**

1. Hatakeyama, M., Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. *Cell Host Microbe* 15, 306–316 (2014).

2. Kuipers, E. J., Perez-Perez, G. I., Meuwissen, S. G. M. & Blaser, M. J. Helicobacter pylori and atrophic gastritis: importance of the cagA Status. *J. Natl Cancer Inst.* 87, 1777–1780 (1995).

3. Blaser, M. J. et al. Infection with Helicobacter pylori strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res.* 55, 2111–2115 (1995).

4. Suerbaum, S. & Josenhans, C. Helicobacter pylori evolution and phenotypic diversification in a changing host. *Nat. Rev. Microbiol.* 5, 441–452 (2007).

5. Linz, B. et al. A mutation burst during the acute phase of Helicobacter pylori infection in humans and rhesus macaques. *Nat. Commun.* 5, 4165 (2014).

6. Kennemann, L. et al. Helicobacter pylori genome evolution during human infection. *Proc. Natl Acad. Sci. USA* 108, 5033–5038 (2011).

7. Yamaoka, Y. et al. Helicobacter pylori outer membrane proteins and gastroduodenal disease. *Gut* 55, 775–781 (2006).

8. Morelli, G. et al. Microevolution of Helicobacter pylori during prolonged infection of single hosts and within families. *Plos Genet.* 6, e1001036 (2010).

9. Sharma, C. M. et al. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* 464, 250–255 (2010).

10. Pernitzsch, S. R., Tirier, S. M., Beier, D. & Sharma, C. M. A variable homopolymeric G-repeat defines small RNA-mediated posttranscriptional regulation of a chemotaxis receptor in Helicobacter pylori. *Proc. Natl Acad. Sci. USA* 111, E501–E510 (2014).

11. Mimuro, H. et al. Helicobacter pylori dampens gut epithelial self-renewal by inhibiting apoptosis, a bacterial strategy to enhance colonization of the stomach. *Cancer Res.* 55, 133–205 (2015).

12. Hör, J., Materia, G., Vogel, J., Gottesman, S. & Storz, G. Trans-acting small RNAs and their effects on gene expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus* https://doi.org/10.1128/ecosalplus.EPS-0030-2019 (2020).

13. Att-Bara, S. & Carposius, A. J. RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNAse E homologues. *RNA degradosome evolution*. *Mol. Microbiol.* 97, 1021–1135 (2015).

14. Odenbreit, S. et al. Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion. *Science* 287, 1497–1500 (2000).

15. Mueller, D. et al. c-Src and c-Ab1 kinases control hierarchical phosphorylation and function of the CagA effector protein in Western and East Asian Helicobacter pylori strains. *J. Clin. Invest.* 122, 1553–1566 (2012).

16. Segal, E. D., Cha, J., Lo, J., Falkow, S. & Tompkins, L. S. Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by Helicobacter pylori. *Proc. Natl Acad. Sci. USA* 96, 14559–14564 (1999).

17. Sim, S.-H. et al. Escherichia coli ribonuclease III activity is downregulated by osmotic stress: consequences for the degradation of bdm mRNA in biofilm formation. *Mol. Microbiol.* 75, 413–425 (2009).

18. Dugar, G. et al. The CaSR-FliW network controls polar localization of the dual-function flagellin mRNA in Campylobacter jejuni. *Nat. Commun.* 7, 11667 (2016).

19. Michaux, C. et al. RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. *Proc. Natl Acad. Sci. USA* 114, 6824–6829 (2017).

**Acknowledgements**

The authors would like to thank Manuel Amieva for providing the H. pylori strains. We acknowledge Keisuke Katsura and Junko Akada for their support. We would like to thank the members of the Division of Bacteriology, Department of Infectious Diseases...
Control, International Research Center for Infectious Diseases, The Institute of Medical Science, The University of Tokyo, and the members of the Department of Infection Microbiology, Research Institute for Microbial Diseases, Osaka University. This work was supported in part by Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan [17K19551, 18K07127, 19K22704 (to H.M.), 16K07083 (to T.S.), 17K14974 (to M.T.), 20K16244 (to R.K.-D.)], the Naito Foundation (to H.M.), the Tokyo Biochemical Research Foundation (to H.M. and P.S.), the Kao Foundation for Arts and Sciences (to K.K.), and the Smoking Research Foundation (to H.M.).

Author contributions
R.K.-D., K.K. and H.M. conceived the study, designed experiments, and wrote the manuscript; M.M. performed Northern blot analysis; R.O. performed part of the mouse experiments; Z.B. analyzed gene-expression data sets; Y.O., R.Y., K.N. and T.H. analyzed genome sequences; T.V.P. and Y.Y. provided clinical isolates and analyzed genome sequences; T.Sa., T.I., E.K., S.H., M.T., A.S., P.S., H.A., and T.Su. provided technical assistance with experiments; T.T.B., L.T.N., K.V.V., and D.Q.D.H. provided clinical isolates; H.M. supervised the study.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-22317-7.

Correspondence and requests for materials should be addressed to H.M.

Peer review information Nature Communications thanks Richard Peek Jr, Gisela Storz and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021