Helicobacter small RNA regulates host adaptation and carcinogenesis

Ryo Kinoshita-Daitoku[1,2], Kotaro Kiga[2], Ryota Otsubo[1], Yoshitoshi Ogura[3], Takahito Sanada[1,2], Zhu Bo[2], Tuan Vo Phuoc[4,5], Tokuju Okano[6], Tamako Iida[2], Rui Yokomori[7], Eisuke Kuroda[1,2], Sayaka Hirukawa[2], Mototsugu Tanaka[2,8], Arpana Sood[2], Phawinee Subsomwong[1], Hiroshi Ashida[6], Tran Thanh Binh[4,5], Lam Tung Nguyen[4], Khien Vu Van[9], Dang Quy Dung Ho[5], Kenta Nakai[7], Toshihiko Suzuki[6], Yoshio Yamaoka[4], Tetsuya Hayashi[3], and Hitomi Mimuro[1,2]

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 Department of Bacteriology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.
4 Department of Environmental and Preventive Medicine, Faculty of Medicine, Oita University, Yufu, Oita, Japan.
5 Department of Endoscopy, Cho Ray Hospital, Ho Chi Minh 749000, Vietnam.
6 Department of Bacterial Pathogenesis, Infection and Host Response, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan.
7 Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
8 Division of Nephrology and Endocrinology, The University of Tokyo School of Medicine, Tokyo, Japan

9 Department of GI Endoscopy, 108 Central Hospital, Hanoi, Vietnam.

Abstract

Type-1 carcinogenic *Helicobacter pylori* that is known to evolve during long-term infection, enters the stomach orally and causes gastric cancer using the carcinogenic protein CagA[1]. However, little is known about the adaptation mechanisms of *H. pylori* when the environment changes from the outside to the inside of the living body. Here we show that small non-coding RNA HPnc4160 is a crucial novel RNA molecule of *H. pylori* that negatively regulates bacterial-host adaptation and gastric cancer. *H. pylori* isolated from gerbil's stomachs eight weeks post-infection acquired mutations in the increased number of T-repeats within the upstream region of the HPnc4160 coding region, which leads to reduced HPnc4160 expression levels that also seen in cancer patients-derived *H. pylori*. By comparing RNA-seq and iTRAQ analysis between wild-type and *hpnc4160* deficient mutant strains, we identified eight targets of HPnc4160 including *cagA* and unknown factors. Mice infection experiment revealed that the *hpnc4160* deficient mutant had a higher number of colonized bacteria in the mice stomach than the wild-type strain, indicating that reduced expression levels of HPnc4160 was important for bacterial host adaptation. The expression level of HPnc4160 was lower in the clinical isolates derived from gastric cancer patients compared with non-cancer-derived strains, while the mRNA expression levels of target factors were higher. Our findings highlight the first discovery that HPnc4160 is an important small RNA for bacteria to adapt to the host environment leading to gastric carcinogenesis.

Main

Helicobacter pylori infection has a very high prevalence that about half of the world population is infected. Patients infected with CagA-positive *H. pylori* closely related to disease malignancy have been reported to have an increased risk of peptic ulcer, chronic gastritis, intestinal metaplasia, and gastric cancer.[2,3] The highly pathogenic *H. pylori* possess a *cag* pathogenicity island that encodes components of a Type IV secretion system (TFSS), which is an injection needle, and a carcinogenic factor CagA effector protein. The *H. pylori* that has reached the gastric epithelium injects CagA protein, peptidoglycan, and heptose-1,7-bisphosphate into the attached cells via TFSS, and stimulates signal transduction pathways such as NF-κB to promote production of chemokines such as interleukin-8 (IL-8)[4–11].

It is considered that the genetic diversity at the genomic level that is characteristic of *H. pylori* is very important to establish a persistent infection in an infected host with different backgrounds, adapting to a gastric niche with severe environmental changes[12]. The *H. pylori* gene mutations are characterized by the presence of simple repetitive sequences such as mononucleotide repeats (such as poly-T) and dinucleotide repeats (such as CT-repeats). From the analysis of clinical isolates of *H. pylori* so far, it is considered that phase variations are induced by ON / OFF control of gene expression of such as outer membrane proteins (OMPs) due to expansion and contraction of these simple repetitive sequences[12–14]. In the course of chronic infection, in order to escape from the host immunity, it is assumed that strong diversity in *H. pylori* OMPs that can serve as highly antigenic cell surface antigens will cause a strong directivity of selection[15]. Therefore, to understand the mechanism of persistent infection by *H. pylori*, we analyzed bacterial gene mutations acquired by *H. pylori* in the course of persistent infection using an

experimental animal infection system with the same host genetic background. With growing evidence that bacterial small RNA (sRNA)-mediated target gene expression in response to changes in the environment, we focused sRNAs as well[16].

Identification of HPnc4160

To analyze the bacterial gene mutation acquired by *H. pylori* during the persistent infection, Mongolian gerbils (n=10) were inoculated with *H. pylori* ATCC 43504 wild-type strain for 8 wks. *H. pylori* in the infected stomachs were isolated (n=40; Fig. 1a), and analyzed comparative whole genome sequences (Fig. 1b, Supplementary Information 1, 2). We totaled genomic positions, where these mutations were introduced, for each coding and intergenic region, and identified 13 regions (Regions R1, R3-R5, R7-R8, R10-R16) where mutations were introduced in 50% or more of the strains (Fig. 1b, Extended Data Table 1).

To investigate whether the mutated region affects RNA expression in isolates recovered from the gerbils, the expression levels of mRNA or non-coding RNA around the mutated regions were quantified by quantitative PCR. Among the corresponding 15 CDSs and non-coding RNA (HP0947, *babA*, *tpiA*, jhp1163, HP0811, HPnc4160, HPnc4170, jhp0540, *araS*, *pldA*, *sabA*, HP1354, *hopZ*, *tlpB*, HPB8_818), we found that the expression level of HPnc4160, a non-coding small RNA (sRNA) of unknown function, fluctuated the most compared to the wild-type (Fig. 1c, Extended Data Fig. 1a). Similar results were also obtained with the strains (n=10) isolated from C57BL/6 mice (n=5) (Fig. 1b, Extended Data Fig. 1a, b; Supplementary Information 3, 4).

Region R14 is the upstream region of HPnc4160 and HP0811, and is located in the CDS of HPnc4170 (*aapB* small ORF homologue) encoded by the complementary sequence of HPnc4160 (Fig. 1d) [17]. HPnc4160 and its

upstream T repeat were conserved in various *H. pylori* genome analysis strains, and T repeat length was different depending on the strain (Extended Data Fig. 2a). A repeat of 2 to 4 bases of thymidine was inserted into the repeat of isolates from rodents, and the repeat length increased depending on the infection period (Fig. 1e, Extended Data Fig. 2b-e). Importantly, sequence analysis of clinical isolates showed that T repeat lengths were longer in cancer patient-derived strains than in non-cancer patient-derived strains (Fig. 1f, Supplementary Information 5). However, expansion of the repeat was not observed in long-term *in vitro* culture (Extended Data Fig. 3). Next, we analyzed the change in HPnc4160 sRNA expression levels by T repeat length. In strains recovered from *H. pylori*-infected rodent stomachs, HPnc4160 expression level tended to decrease with expansion of T repeat length (Fig. 1g). To exclude the effects of mutations other than the T repeats, we further analyzed RNA expression levels of HP0811, HPnc4170 and HPnc4160 in various mutants in which the T repeat sequence was inserted into the HPnc4160 upstream region of wild-type strain (T15mut, T16mut, T17mut, T18mut, and T19mut). In strains in which the number of T repeats was greater than T14 of wild-type, the expression levels of HPnc4160, but not HP0811 and HPnc4170, were significantly reduced compared with wild-type (Fig. 1h, Extended Data Fig. 4a-c,). Although HPnc4160 and HPnc4170 were initially reported as the small ORF-encoding mRNA/antisense RNA family aapB/IsoB, in which the Iso transcript acts as asRNA antitoxin to modify the aap expression[18], our data indicated that HPnc4160 expression levels had no effect on HPnc4170 levels. These results indicated that expression levels of HPnc4160 sRNA decreased when the number of Region R14 T repeats increased due to persistent intragastric infection.

Identification of HPnc4160 target genes

Many sRNAs regulate the expression of a target mRNA by specifically binding to a complementary sequence in the target protein coding mRNA. To elucidate the target mRNA of HPnc4160, we made a Δ*hpnc4160*/*hpnc4170* strain, in which both the HPnc4160 and the HPnc4170 on the complementary strand were deleted, and analyzed comparative mRNA and protein expression. We identified eight factors (*cagA*, *hofC*, HELPY_1262, *hpaA*, *horB*, *omp14*, *hopE*, and HP1227) with P-values lower than 0.001 (RNA-seq analysis) and 0.01 [isobaric tag for relative and absolute quantitation (iTRAQ) labeling and LC-MS/MS analysis] (Fig. 2a-c, Extended Data Table 2a-b). Of these, *cagA* was prominent in expression levels of mRNA and protein.

We analyzed whether the mRNA expression levels of the identified eight factors depend on the presence of HPnc4160. Although the expression level of HPnc4160 showed a decreasing trend with T16mut as the lowest value, the mRNA of the eight candidates showed increasing trends with T16mut as the highest value. The Spearman correlation coefficients r between the expression levels of these target mRNAs and HPnc4160 showed a very strong inverse correlation from -0.7714 to -1.0 (Extended Data Fig. 4d).

Next, we examined the mRNA expression correlation between HPnc4160 and each factor. The HPnc4160 overexpression strain (WT / pHel2-*hpnc4160*) significantly increased the expression level of HPnc4160 compared to the wild-type strain, but significantly decreased the expression level of each factor mRNA. On the other hand, in Δ*hpnc4160*/*hpnc4170* strain, mRNA expression of each target increased compared to the wild-type. Since Δ*hpnc4160*/*hpnc4170* strain also lacks the HPnc4170 sequence present in the complementary strand of HPnc4160, we constructed a Δ*hpnc4160*/*hpnc4170* /pHel2-*hpnc4160* strain complementing only HPnc4160 to confirm the effect of the HPnc4170 sequence on HPnc4160 target mRNA expression. Compared to the Δ*hpnc4160*/*hpnc4170* strain, the mRNA expression levels of the

candidates were decreased in the HPnc4160 complemented Δ*hpnc4160*/*hpnc4170* /pHel2-*hpnc4160* (Fig. 2d, Extended Data Fig. 5a). These data indicated that the expression levels of the HPnc4160 target 8 candidates were suppressed depending on the expression level of HPnc4160.

When sRNA binds within a few bases around the 5 'UTR or start codon of target RNA, it often competes with ribosomes and inhibits translation initiation. If the sRNA binds within the CDS far downstream of the initiation codon, it causes mRNA degradation by RNase E or RNase III to suppress translational activation[19]. We confirmed by electrophoretic mobility shift assays (EMSA) that HPnc4160 binds to the 5 'UTR of seven genes other than *cagA* (Fig. 3a). In seven factors other than *cagA*, we confirmed a sequence complementary to the HPnc4160 sequence in the 5 'UTR of each gene (Extended Data Fig. 5b-c). In the *cagA* gene, we found HPnc4160-binding sequences Type 1 at one position (2344 nt), and Type 2 at four positions (2838, 2940, 3042, and 3144 nt) within the CM/CRPIA motifs in CagA C-terminal region, which is known to bind with host signal proteins[7,20] (Fig. 3b, Extended Data Fig. 5d-f). We confirmed direct binding of *cagA* partial CDS (positions 2778 to 3236 nt from start codon of *cagA*) to HPnc4160 (Fig. 3c). The binding between the two was abolished in the NB-*cagA* RNA in which the HPnc4160 binding sequence was mutated at four positions (Type 2) while the amino acid sequence of CagA was preserved (Fig. 3c, Extended Data Fig. 5g). In addition, we found that in the presence of *H. pylori* RNase III recombinant protein, the binding between HPnc4160 and biotin-labeled partial *cagA* mRNA, but not NB-*cagA* RNA, disappeared (Fig. 3d, Extended Data Fig. 5i). These data clearly demonstrated that HPnc4160 controls *cagA* at the post-transcription level by binding to multiple binding sequences present in its CDS region, and promotes degradation by RNase III.

Effects of HPnc4160 on *H. pylori* pathogenicity

Among the factors that HPnc4160 regulated the expression levels, we further analyzed CagA, which has been deeply involved in pathogenesis. First, we confirmed whether the binding of HPnc4160-*cagA* mRNA actually controls the expression levels of *cagA* mRNA and protein in *H. pylori*. The quantitative PCR showed that, in the *H. pylor*i expressing NB-*cagA* in which all five HPnc4160-binding DNA sequences were mutated but the amino acid sequence was preserved (Extended Data Fig. 5g-h), the expression level of HPnc4160 was similar to that of the wild-type, but the expression level of *cagA* mRNA was significantly increased to the same extent as that of the Δ*hpnc4160*/*hpnc4170* strain (Extended Data Fig. 6a, Fig. 4a). Using the urease UreA protein as a loading control for *H. pylori*, we confirmed that NB-*cagA* strain expressed CagA protein at a higher level than wild-type and Δ*hpnc4160*/*hpnc4170* / pHel2-hpnc4160 strains, similar to Δ*hpnc4160*/*hpnc4170* strain (Fig. 4b). Next, we analyzed Western blot of the gastric epithelial cell line AGS infected with *H. pylori*. Using β-actin as a loading control for cells, we confirmed that the amount of UreA protein, which exhibited the bacterial amounts, was the same in the lysates of cells infected with any of the strains, indicating that all of the strains showed same binding ability to AGS cells (Extended Data Fig. 6b). Some of the CagA proteins injected from *H. pylori* to host epithelium via TFSS were tyrosine phosphorylated by host Src/Abl kinase and detected by pY-CagA-specific antibody. Using the antibody, we confirmed that the amount of intracellular CagA was increased in the NB-*cagA*-infected cells, accompanied with the increase in the amount of CagA (Extended Data Fig. 6b). Injected CagA induces AGS cell motility (scattering/hummingbird). In the AGS cells infected with the Δ*hpnc4160*/*hpnc4170* or the NB-*cagA* strains, more remarkably elongated cells were observed than in the wild-type or Δ*hpnc4160*/*hpnc4170* / pHel2-*hpnc4160* strain-infected cells (Fig. 4c-d). When

we analyzed amount of IL-8 protein secreted from *H. pylori*-infected cells, which is induced mainly by intracellular CagA, we found that the *cagA*-NB strain infected cells had higher IL-8 producing ability than the wild-type infected cells (Extended Data Fig. 6c). These results suggested that binding of HPnc4160 to *cagA* mRNA is important for controlling the amount of functional CagA protein injected by *H. pylori*.

To understand the significance of the HPnc4160 control mechanism in the bacterial adaptation to the host to establish infection, mice were orally inoculated with each strain, and the number of bacteria colonized in the stomach was analyzed three days post infection. The number of colonized bacteria in the stomach was significantly increased in the Δ*hpnc4160*/*hpnc4170* strain compared to the wild-type, but the Δ*hpnc4160*/*hpnc4170* / pHel2-*hpnc4160* and the NB-*cagA* strains were equivalent to wild-type (Fig. 4e). Since Δ*hpnc4160*/*hpnc4170* infection significantly increased *Cxcl2* mRNA compared to wild-type infection, but NB-*cagA* strain was equivalent to wild-type, it is suggested that factors other than CagA controlled by HPnc4160 may be important for the bacterial adaptation as well as development of gastritis (Fig. 4f). To confirm the significance of HPnc4160 in the pathogenesis of *H. pylori*, we examined the expression levels of the HPnc4160 target genes in clinical isolates. As shown in Fig. 4g, isolates from cancer patients had lower levels of *hpnc4160*, but increased expression of six factors controlled by HPnc4160 (*cagA*, *horB*, *hopE*, *omp14*, *hofC*, and *hpaA*), compared to isolates from non-stomach cancer patients (Fig. 4g, Extended data Fig. 7). These data strongly suggested that target mRNA expression was suppressed by HPnc4160 at the onset of *H. pylori* infection, and with the course of infection, thymidine repeats were inserted into the upstream region of HPnc4160, HPnc4160 expression decreased, and target mRNA expression increased; those episodes contributed to bacterial

adaptation to the host environment, leading to gastritis and gastric cancer formation (Extended Data Fig. 8).

Discussion

Among the bacterial factors, CagA had extremely high mRNA and protein levels in the bacterial cells (Fig. 2a and b). Our data indicated that T-stretch length in the upstream region of *hpnc4160* did not elongate under *in vitro* condition (Extended Data Fig. 3). However, the T-stretch length was increased and the expression levels of CagA and OMPs increased *in vivo* in gastric infection (Fig. 1e, 2a and 2b). Adaptation to the *in vitro* culture environment does not require CagA or OMPs. Therefore, under *in vitro* conditions, it may be that the expression level of hpnc4160 is increased and the expression of target is suppressed in order to allocate energy for the growth of bacterial cells, rather than expressing genes having high expression levels such as CagA. CagA suppresses apoptosis of the gastric mucosal epithelium and contributes to persistent infection of *H. pylori*[6]. When *H. pylori* enter the stomach, the bacteria may have acquired a mechanism to decrease the expression of HPnc4160 to increase the expression of OMP and CagA simultaneously, in order to adapt to environmental changes to colonize the gastric mucosa for a scaffold for growth.

Gene expression control mechanism by variation of the number of repeat sequences is known as one of various gene expression control mechanisms by phase variation in *H. pylori*[21]. This study suggests that the repeat sequence of *H. pylori* genome is important not only as an ON / OFF mechanism of protein expression such as cell adhesion factors SabA and BabA, but also in sRNA expression. It has been reported that *H. pylori* DNA polymerase I does not have gene repairing activities like other bacteria, thus

insertions or deletions of bases called slipped strand mispairing (SSM) occur in simple repetitive sequences[22]. In fact, since there were variations in poly-T length within the upward region of *hpnc4160* in *H. pylori in vivo* (Extended Data Fig. 2a), it is conceivable that *H. pylori* become genetically heterogeneous bacterial population using SSM during the course of infection, so that a population suitable for mutation in the host is selected and propagated. In this study we mainly used strain ATCC 43504, which is a clinical isolate originally from the human antrum. The poly-T length, which did not fluctuate in the *in vitro* subculture, increased in rodent isolates ranged from 14 to 19 copies (Extended Data Fig. 2, 3), indicating that the introduction of SSM may be caused by some host stress condition.

The HPnc4160-binding sequence in the ATCC 43504 *cagA* CDS was duplicated at five sites (Fig. 3b). In the regulation of mRNA expression by general sRNA, mRNA instability can be induced by binding at one site. This is the first report of a sRNA gene regulatory mechanism having multiple binding sequences in one gene of a pathogen. The four out of five HPnc4160-binding sequences in the *cagA* CDS was located in the CM/CRPIA motif of the *cagA* gene, which is involved in maintaining host epithelial cell structure[20]. Since the number of the CM/CRPIA motif differs depending on the strain, HPnc4160 may regulate differences in pathogenicity between Western and East Asian type of *H. pylori*.

The onset of the diseases due to *H. pylori* infection is thought to be the result of persistent infection for decades after the initial infection during childhood, when a very small number of bacteria-containing aerosol were taken into the body orally[23]. However, since it is difficult to analyze *H. pylori* infection experiments using laboratory animals for decades, a relatively short time infection analysis must be performed by inoculating a large number of bacteria. Therefore, the mutation analysis in this study may correspond to

mutations acquired in the acute phase (Supplementary information 1 and 3)[15,24]. Infection by a small number of bacteria that mimic natural infection allows for more detailed analysis of the establishment of infection.

In this study, we investigated the mutations of *H. pylori*, originated from the same strain, infected in experimental animals with the same genetic background and environment. Unlike studies in individuals with widely differing genetic backgrounds, stomach environments, and infection history, our study has advantage for understanding the adaptation process of *H. pylori* to the host. We discovered a novel non-coding sRNA that is important for post-transcriptional translation control of pathogenic factors of *H. pylori*, such as CagA, which was previously considered to be the most important pathogenic factor for gastric cancer development, and putative OMPs involved in bacterial adhesion. This study is not limited to elucidating the complicated mechanism of persistent *H. pylori* infection, but its application to *H. pylori*-specific therapies that do not rely on antibiotics can also be expected.

Figure legends

Fig. 1: *H. pylori* acquire poly-T extension in upstream of HPnc4160 small RNA to decrease its expression during infection *in vivo*.

a, Experimental strategies schematic. **b**, Circular genomic map of ATCC 43504 strain recovered from stomachs of gerbils and mice. **c**, Mutation rates and expression levels of candidate RNAs (mRNA or non-coding RNA). RNA expression levels of the ORFs or nearby genes of genome regions [total 15 genes (Extended Data Fig. 1a), which mutated in more than 50% of the recovered strains from the gerbils, were assessed and plotted against the mutation rates. **d**, Schematic structures of genes around HPnc4160. **e**, Schematic diagram of genome DNA sequence around the HPnc4160 and poly-T sequence of recovered strains harboring mutations. Green indicates HPnc4160 transcribed sequence, red frame indicates poly-T mutated stretches in the upstream region of HPnc4160 of ATCC 43504 wild-type (WT) strain, and red-colored "T" indicated inserted nucleotides of each recovered strain compared with WT. **f**, The T-repeat length in the upstream region of HPnc4160 of clinical isolates. Isolated strains from cancer patients (Cancer) have a higher number of T-repeat in the upstream region of Hpnc4160 compared with the isolates from non-cancer patients (Non-Cancer). Data are presented as means with 95% confidence interval. *P* values represent the results of the two-tailed Mann-Whitney test. **g**, Expression levels of HPnc4160 in the recovered *H. pylori* strains from Mongolian gerbils (n=40). Relative expression levels of HPnc4160 were measured by real-time PCR and plotted against T-repeat length in the upstream region of HPnc4160. Data are presented as means with s.d. *P* values represent the results of two-tailed Dunn's multiple comparison test. **h**, The relative expression levels of HPnc4160 in the *H. pylori* strains genetically modified with the T-repeat length. Data are presented as means

with s.d. (n=3). *P* values are from Dunnett's multiple comparison test (at two-sided).

Fig. 2: HPnc4160 downregulates expression levels of bacterial pathogenic factors.

a, The MA plot of ratios [Δ*hpnc4160*/*hpnc4170* / wild-type (WT) *H. pylori*] versus their normalized average mRNA expression determined by RNA-sequencing (RNA-Seq). The red dots showed genes of $P<0.001$. **b**, Volcano plots of the proteins quantified by isobaric tags for relative and absolute quantification (iTRAQ) analysis comparing WT and Δ*hpnc4160*/*4170*. Each point represents the difference in expression (fold change) between the two groups plotted against the level of statistical significance. The red dots showed proteins of $P<0.01$. **c**, The Venn diagram represents the number of factors whose expression exhibited significant differences between Δ*hpnc4160*/*hpnc4170* and WT. **d**, The relative RNA expression levels of target candidates of HPnc4160 showed an inverse correlation with HPnc4160. The results represent the average of three separate experiments (each n=3). Data are presented as means ± s.d. (error bars). *P* values are from Tukey's multiple comparison test (at two-sided). ns: not significant.

Fig. 3: HPnc4160 binds to target mRNA.

a, Electrophoretic mobility shift assay (EMSA) analysis of HPnc4160 binding to the 5' UTR region of each candidate mRNA. **b**, Schematic of CagA motifs, and HPnc4160 binding sequences. **c**, EMSA analysis of HPnc4160 binding to RNA of partial *cagA* WT or HPnc4160-non-binding *cagA* (NB-*cagA*). **d**, RNase protection assay with HPnc4160, *cagA* mRNA, and recombinant RNase III.

Fig. 4: HPnc4160 controls bacterial host adaptation and pathogenesis.

a, HPnc4160, and *cagA* mRNA expression levels. The results represent the average of three separate experiments (each n=3). Data are presented as means ± s.d. (error bars). *P* values are from non-parametric Dunnett's multiple comparison test (at two-sided). ns: not significant. Experiments were repeated three times with similar results. **b**, Protein expression levels of CagA in each mutant strain. UreA protein levels serve as bacterial loading controls. **c** and **d**, Scattering phenotypes of *H. pylori*-infected AGS gastric epithelial cells. **c**, DNA (blue), F-actin (red), and anti-phosphorylated CagA antibody (pY-CagA, green) were stained. Scale bar, 50 μm. **d**, Quantification of scattering activity of AGS cells induced by *H. pylori* infection. **e** and **f**, HPnc4160 deletion mutants efficiently colonized the stomach and contributed to increasing mRNA levels of inflammatory chemokine *Cxcl2* in mice stomach. C57BL/6 mice were inoculated with *H. pylori*. At three days after infection, animals were sacrificed, and a quantitative culture assay (**e**) and a quantitative RT-PCR (**f**) were performed on gastric specimens. Data are median with interquartile range. *P* values are from non-parametric Dunn's multiple comparison test (at two-sided). ns: not significant. **g,** Clinical isolates from malignant patients downregulate HPnc4160 and upregulate expression levels of its target genes. Expression levels of indicated mRNAs in clinical isolates of non-cancer (Non-Cancer, n=39) and cancer (Cancer, n=17) patients were quantified and normalized with the levels of 23S rRNA. Data are presented as medians with interquartile range. *P* values are from the non-parametric Mann-Whitney test (at two-sided).

Methods

Data reporting

No statistical methods were used to predetermine sample size, and the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Strains and culture conditions

The *Helicobacter pylori* strain ATCC 43504, its isogenic mutants Δ*cagA* and Δ*virB7*, strains SS1 and PMSS1 have been described previously[6,25]. *H. pylori* was cultured on Trypticase soy agar with 5% (v/v) sheep blood (Thermo Fisher Scientific, Waltham, MA, USA) for 2 days at 37ºC in microaerobic conditions. Bacterial colonies were suspended in Brucella broth (Thermo Fisher Scientific) supplemented with 5% (v/v) inactivated FBS (Thermo Fisher Scientific), adjusted to Optical density 600 nm of 0.05, and incubated 15 hours at 37ºC with gentle agitation under microaerobic conditions.

The AGS human gastric epithelial cell line (ATCC CRL-1739) was maintained in 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.

Antibodies and immunohistochemical reagents

The anti-Tyr(P)-CagA, and anti-UreA polyclonal antibodies have been described previously (Mimuro MC 2002). Anti-CagA polyclonal antibody was purchased from AUSTRAL Biologicals (CA, USA), anti-actin monoclonal antibody was from MERCK (Darmstadt, Germany), Horse radish peroxidase (HRP)-labeled anti-rabbit IgG and HRP-labeled anti-mouse IgG, and FITC-

labeled anti-rabbit IgG was from Jackson ImmunoResearch Laboratories Inc. (PA, USA). DAPI was from SIGMA-ALDRICH (MD, USA), and Rhodamine Phalloidin was from Thermo Fisher SCIENTIFIC (MA, USA).

Animal infection

H. pylori infection of rodents were performed as described previously[26]. Briefly, 6-week-old male MON/Jms/GbsSlc Mongolian gerbils were orally administered with 200 μL of Vancomycin (500 mg/L) at 24 and 48 hours before *H. pylori* inoculation. On the days of *H. pylori* inoculation, 300 μL of 5% (w/v) sodium bicarbonate were orally administrated 10 minutes before bacterial inoculation. The gerbils were then intragastrically inoculated with an *H. pylori* culture containing 10^9 CFU for 2 consecutive days. As for C57BL/6 mice (SLC Japan Inc., Tokyo, Japan) were intragastrically inoculated once with *H. pylori* culture of 10^9 CFU. After indicated date, the stomach of each infected animal was opened along the greater curvature. To quantitatively isolate *H. pylori,* the stomach was excised, weighed, and homogenized. Serial dilutions were plated on *H. pylori*-selective agar plates (Eiken Chemical Co.) and incubated under microaerophilic conditions at 37°C for 4 days, after which the cfu were counted. Colonization data points of 1×10^3 cfu were the minimal detection limit of the assay.

For isolation of strains recovered from *H. pylori*-infected rodents, each colony on the *H. pylori*-selective agar plates were picked up and spread on Trypticase soy agar with 5% (v/v) sheep blood, and incubated under microaerophilic conditions at 37°C for two days. Then, the colonies were suspended in Brucella broth supplemented with 5% (v/v) inactivated FBS, adjusted to Optical density 600 nm of 0.05, and incubated 15 hours at 37ºC with gentle agitation under microaerobic conditions. The cultures were preserved with 50% (v/v) glycerol in -80°C until use.

For RNA isolation, the tissue was immediately frozen in liquid nitrogen. Animal experiments were conducted in accordance with the University of Tokyo or Osaka University guidelines for the care and use of laboratory animals and were approved by the ethics committee for animal experiments at the University of Tokyo or Osaka University.

Genomic DNA purification and sequencing

For PCR templates, genomic DNA was purified using InstaGene Matrix (Bio-Rad Laboratories, Inc., CA, USA).

For whole genome sequencing, genomic DNA was purified from mid-log phase culture of strain ATCC43504 using QIAGEN DNeasy (QIAGEN). A genomic DNA library for sequencing was prepared using the Nextera XT DNA Sample Preparation kit (Illumina, San Diego, CA, USA) and sequenced using the Illumina MiSeq (for isolates from gerbils) or HiSeq X (for isolates from mice) platform to generate 300-bp paired-end reads. Genome assembly, scaffolding, and gap-closing were performed using the Platanus assembler (Kajitani et al. 2014). Gene identification and annotation were conducted by the Microbial Genome Annotation Pipeline (MiGAP [http://www.migap.org]). The raw read sequences and assembled scaffold sequences have been submitted to the DDBJ/EMBL/Genbank under the Bioproject accession number; SAMD00178897- SAMD00178935, SAMD00179460, SAMD00178937 and SAMD00204457- SAMD00204466.

The DNA sequences mutated in more than 50% of the 40 strains recovered from Mongolian gerbils, or, in all of the 10 strains recovered from C57BL/6 mice were listed in Extended Data Table 1. We selected the genes to further analyze for their mRNA expression levels as follows. For the gene in which the mutation was in the CDS region, the mRNA expression level of the CDS was measured. While, when the mutation insertion region was an intergenic region,

we measured the mRNA expression level of an adjacent gene in which the intergenic region could be a 5'UTR region. As for HP1243 and HPG27_298, which started from 3' end of HP1243 with 33 nucleotides spaces, were regarded as a continuous gene; since both genes are annotated as *babA* gene and ribosomal binding site (RBS) is assigned only at the upstream region of HP1243.

***In vitro* passage experiment**

H. pylori ATCC 43504 was recovered from frozen stock and cultured on 5% (v/v) sheep blood agar for 2 days at 37ºC in microaerobic conditions. Bacterial colonies were suspended in 3 tubes of Brucella broth supplemented with 5% (v/v) inactivated FBS. Each bacterial suspension was adjusted to Optical density 600 nm of 0.05, and incubated 12 hours at 37ºC with gentle agitation under microaerobic conditions. Following this incubation, each fraction of the suspension was preserved by freezing in 50% (v/v) glycerol as "Original" strains. Meanwhile, each bacterial suspension was sub-cultured by resuspending in Brucella broth supplemented with 5% (v/v) inactivated FBS to adjust Optical density 600 nm of 0.05, and incubated additional 12 hours at 37ºC with gentle agitation under microaerobic conditions. The sub-cultivation was repeated for 60 passages (30 days), and each cell suspension was preserved by freezing in 50% (v/v) glycerol as "60-passaged" strains. The "Original" and "60-passaged" strains were recovered from frozen stock on 5% (v/v) sheep blood agar by 2 days incubation under microaerobic conditions, and then the colonies were suspended in Brucella broth supplemented with 5% (v/v) inactivated FBS and incubated 12 hours at 37ºC with gentle agitation under microaerobic conditions. The bacterial cells were collected and subjected to the genomic DNA purification.

RT-PCR

For preparation of total RNA from *H. pylori*, the liquid cultures of *H. pylori* were agitated under microaerobic conditions at 37°C overnight until the OD value at 600 nm reached 0.9.

Total RNA was extracted using ISOGEN (Nippon Gene, Tokyo, Japan), according to the manufacturer's instructions. The concentration of the purified total RNA was analyzed using the NanoDrop Spectrophotometer (ThermoFisher Scientific, Wilmington, DE, USA). The total RNA was reverse transcribed into cDNA with miScript II RT Kit (QIAGEN) according to the manufacturer's instructions. The levels of mRNA expression were quantified and normalized to 23SrRNA (for *H. pylori*) or *Gapdh* (for mice) expression with a THUNDERBIRD SYBR qPCR (TOYOBO) using the primer pairs described in Supplementary Information 6. The results are expressed as the means ± SEM from triplicate strain experiments.

Genetic manipulation

Construction of plasmids for producing gene-deficient mutants

Isogenic gene null mutants derived from ATCC 43504 were constructed by insertional mutagenesis as follows. Using the extracted *H. pylori* ATCC 43504 genome as a template, DNA fragments containing the upstream region 500 bp and the downstream region 500 bp of the target gene were amplified by PCR using primer (CagA KO up XhoI, CagA KO up EcoRI, CagA KO down BamHICagA KO down NotI, HPnc4160/4170 KO up KpnI, HPnc4160/4170 KO up ClaI, HPnc4160/4170 KO down BamHI, HPnc4160/4170 KO down SacI; listed in Supplementary Information 6). The DNA fragments were introduced at the both sides of the *aphA3* (which confers kanamycin resistant) in pBluescript II SK (+) plasmid. The fragments from the resulted plasmid were introduced

into *H. pylori* by electroporation.

Construction of non-marker H. pylori mutants

For constructing non-marker *H. pylori* mutants, ATCC 43504 *flaA* and *cag1* promoter and terminator were cloned into pBluescript SK(+) *Sma*I aphA3 *Sma*I, and *sacB* gene was cloned into *EcoR*I site (pKSB plasmid). Mid-log-phase (OD600 = 0.5-0.7) of *H. pylori* in 20 ml culture liquid were washed twice with ice-cold 10% glycerol and resuspended by 200 µl of ice-cold 10% glycerol. 1µg of pKSB vector containing aimed mutation and the bacterial liquid were mixed at 4℃ and electroporated by Micropulser (Bio-Rad) with Ec2 (2.5kV) setting. After 4 hours incubation at 37ºC in microaerophilic condition, cells were plated on 5% sheep blood agar plate TSAII containing 4 µg/ml Kanamycin and incubated 2-3 days at 37ºC under the microaerophilic condition. 4 single colonies were seeded on new 5% sheep blood agar plate TSAII supplemented with 4 µg/ml Kanamycin and incubated for additional 2 days. Each colony was picked up and were cultured in Brucella broth containing 5% FBS at 37ºC under the microaerophilic condition until *H. pylori* were grown to mid-log phase. 100µl of the medium were plated on 5% sheep blood agar plate supplemented with 2.5% sucrose and cultured for 2 days. Each colony was seeded on a new 5% sheep blood agar plate without antibiotics and incubated for 2 days. At the same time, the colony was seeded on a different agar plate with 4 µg/ml Kanamycin to confirm the Kanamycin resistant was disappeared. Grown *H. pylori* were transferred to liquid culture and the genome sequence was confirmed by Sanger sequencing.

Construction of point mutated H. pylori

The *H. pylori* recombination plasmids to establish various mutant strains (T15mut, T16mut, T17mut, T18mut, T19mut) in the upstream region of

hpnc4160 were constructed by PCR using *H. pylori* genome DNA from the strains isolated from gerbil after 8 weeks as a template, and primers (pKSB-HPnc4160 Point mut *Apa*I and pKSB-HPnc4160 Point mut *Xho*I; listed in Supplementary Information 6), then, the resulted DNA fragments were cloned into suicide pKSB plasmid.
H. pylori T15mut, T16mut, T17mut, T18mut and T19mut mutants were established by introducing each pKSB-based plasmid into *H. pylori* ATCC 43504 strain.

Construction of NB-cagA-expressing H. pylori
Based on the full length cagA cDNA sequence of ATCC 43504, we designed HPnc4160-unbound *cagA* gene sequence (NB-*cagA*, Extended Data Fig. 5g and h). The NB-*cagA* cDNA were artificially synthesized as pEX-K4J2-cagA mutant of 908 bps (eurofins, 99900008281-1). The cDNA fragments containing mutated *cagA* sequence were amplified using primers (pKSB-CagA-NB-ApaI, pKSB-CagA-NB-XhoI, listed in Supplemented Information 6), and cloned into a suicide vector pKSB. The resulted plasmids were introduced to *H. pylori* ATCC 43504 to obtain NB-cagA-expressing *H. pylori*.

Construction of hpnc4160 over-expressing H. pylori
The plasmid for the hpnc4160 overexpressing strain in *H. pylori* was constructed by combination of
DNA fragments of hpnc4160 regions were amplified by PCR using primers (pHel2-4160-de-4170-hed-f XhoI, pHel2-4160-de-4170-hed-r BamHI, Supplemented Information 6) and genome DNA of the ATCC 43504 strain as a template. The resulted DNA fragments included the upstream region of *hpnc4160* without including the 5' region of the hpnc4170 region. The DNA was cloned into pHel2 shuttle vector, and introduced into *H. pylori* by

electroporation.

RNA-seq

H. pylori were agitation under aerobic conditions and cultured at 37°C overnight until the OD value at 600 nm reached 0.9. Total RNA from the *H. pylori* were extracted using RNeasy (QIAGEN), according to the manufacturer's instructions. The concentration of total RNA extracted was examined using the NanoDrop Spectrophotometer (ThermoFisher Scientific, Wilmington, DE, USA), according to the manufacturer's instructions.
Ten micrograms from each total RNA sample were treated with the MICROBExpress Bacterial mRNA Enrichment kit (Ambion, Grand Island, NY, USA) and RiboMinus™ Transcriptome Isolation Kit (Bacteria) (Invitrogen, Grand Island, NY, USA) following the manufacturer's instructions. Samples were resuspended in 15 μL of RNase-free water. Bacterial mRNAs were chemically fragmented to the size range of 200-250 bp using 1 × fragmentation solution (Ambion, Grand Island, NY, USA) for 2.5 min at 94°C. cDNA was generated according to instructions given in SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen, Grand Island, NY, USA). Briefly, each mRNA sample was mixed with 100 pmol of random hexamers, incubated at 65°C for 5 min, chilled on ice, mixed with 4 μL of First-Strand Reaction Buffer (Invitrogen, Grand Island, NY, USA), 2 μL of 0.1 M DTT, 1 μL of 10 mM RNase-freed NTPmix, 1 μL of SuperScript III reverse transcriptase (Invitrogen), and incubated at 50°C for 1 h. To generate the second strand, the following Invitrogen reagents were added: 51.5 μL of RNase-free water, 20 μL of second-strand reaction buffer, 2.5 μL of 10 mM RNase-free dNTP mix, 50 U E. coli DNA Polymerase, 5 U E. coli RNase H, and incubated at 16°C for 2.5 h. The Illumina Paired End Sample Prep kit was used for RNA-Seq library creation according to the manufacturer's instructions as follows: Fragmented

cDNA was end-repaired, ligated to Illumina adaptors, and amplified by 18 cycles of PCR. Paired-end 150-bp reads were generated by high-throughput sequencing with the Illumina Hiseq 2500 Genome Analyzer instrument. After removing the low-quality reads and adaptors, RNA-Seq reads were aligned to the corresponding ATCC 43504 genome using Tophat 2.1.0 (Trapnell et al 2009), allowing for a maximum of two mismatch. If reads mapped to more than one location, only the one showing the highest score was kept. Reads mapping to rRNA and tRNA regions were removed from further analysis. After getting the reads number from every sample, edgeR with TMM normalization method was used to determine the DEGs. Significantly differentially expressed genes (FDR value < 0.05 and at least two-fold changes) were selected for further analysis.

iTRAQ

H. pylori ATCC 43504 strains of wild-type, Δ*hpnc4160*/*hpnc4170*, and Δ*hpnc4160*/*hpnc4170* / pHel2-*hpnc4160* were cultured in Brucella Broth containing 5% FCS to OD600 = 0.9. 1.5 mL of each bacterial solution was centrifuged at 5,000 xg for 10 minutes at 4ºC. The pellet was resuspended in Wash buffer (1 M KCl, 15 mM Tris-HCl, pH 7.4), centrifuged again, and the supernatant was removed. The pellet was resuspended in a Wash buffer containing 1 mM AEBSF (4- (2-Aminoethyl) benzenesulfonyl fluoride hydrochloride) and frozen at -80ºC. iTRAQ analysis was commissioned to Filgen Corporation.

EMSA (electrophoretic mobility shift assay)

cDNA fragments of small RNA HPnc4160 whole region, the fragments of 150 bp total of each 5′UTR region [from 100 bases upstream from the ribosome binding region (RBS), to 50 bases downstream of the RBS] (hp0410 gene,

hp0486 gene, *horB* gene, hp0671 gene, *hopE* gene, *cagA* gene, hp1227 gene and helpy_1262 gene), and cDNA of 459 bp total containing the hpnc4160-binding 4 region near the 3' tail of the *cagA* gene, were amplified by PCR using primers (Small RNA HPnc4160 XhoI, Small RNA HPnc4160 EcoRI; HP0410 150bp XhoI, HP0410 150bp EcoRI; HELPY_0660 150bp XhoI, HELPY_0660 150bp EcoRI; HP0671 150bp XhoI, HP0671 150bp EcoRI; HP0486 150bp XhoI, HP0486 150bp EcoRI; HPSH_00635 150bp XhoI, HPSH_00635 150bp EcoRI; HPP12_0555 150bp XhoI, HPP12_0555 150bp EcoRI; HP1227 150bp XhoI, HP1227 150bp EcoRI; HELPY_1262 150bp XhoI, HELPY_1262 150bp EcoRI; CagA-B codding XhoI, CagA-B codding EcoRI; listed in Supplementary Information 6) and the ATCC43504 genome as a template. The PCR products were cloned into the position of the downstream of the T7 promoter region of the pBluescript SK (+) plasmid. The NB-*cagA* mutant RNA used in the gel shift assay was amplified with a T7 promoter by PCR using (T7 promoter CagA-NB EMSA PCR s, T7 promoter CagA-NB EMSA PCR as) as primers and synthesized pEX-K4J2-CagA mutant (eurofins, 99900008281-1) *cagA* as a template. The *cagA* mutant RNA were prepared in the same manner except for mutations in the HPnc4160-binding 4 region. RNA was transcribed from a DNA fragment using an *in vitro* Transcription T7 kit (TAKARA).

Gel shift assays were performed using 0.04 pmol of 3'-biotin-tagged mRNA with increasing amounts of purified small RNA HPnc4160 in 20 μL reactions. Briefly, RNA was denatured (10 min, 80°C) and cooled for 5 min on ice. Yeast tRNA 1μg (ThermoFisher SCIENTIFIC) was added to the labelled RNA and the reaction was filled up to 10 μL with Binding Buffer (10 mM HEPES pH 7.3, 1 mM $MgCl_2$, 20 mM KCl, 5% glycerol). 10 μL of either labelled mRNA was added to the HPnc4160. The mixtures were incubated at room temperature for 20 min. Then the samples were mixed with 5 μL native loading buffer before loading on a pre-cooled native 6% poly-acryl amide (PAA), 0.5x TBE gel. Gels

were run in 0.5x TBE buffer at 30 mA per gel for 2 hours[27].

Cleavage assays

The cDNA of 720 bps of *H. pylori rnase* III was amplified by PCR using primers (pGEX-6P-1 RNaseIII XhoI-f, pGEX-6P-1 RNaseIII NotI-r, listed in Supplemented Information 6) and template (genome DNA from ATCC 43504 strain). The cDNA was cloned into 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 rpm. Protein expression was induced with IPTG to a final concentration of 0.1 mM, at 4°C, for 4 hours. The bacteria were collected by centrifugation and the pellets were subjected to GST-fusion protein purification using Glutathione Sepharose 4B (GE) according to the manufacture's instruction. The RNase III protein was excised by PreScission Protease according to the manufacturer's instructions. The purified protein derived from 6.7 mL of the bacterial culture was developed by SDS-PAGE, and the gel was stained with CBB to confirm that no contaminants were observed in the final product. The protein concentration was determined by absorbance at 280nm.

Nuclease assays using RNase III was performed using purified *H. pylori* recombinant RNase III. The gel shift assay protocol described above was followed, except that RNase III-specific buffer (25mM Tris pH 7.5, 50 mM NaCl, 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT) was used instead of Binding Buffer. 3'-biotin-tagged partial *cagA* mRNA was incubated on ice with either 5 μM of small RNA HPnc4160 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% PAA, 0.5x TBE gel[28].

ELISA

AGS cells were co-incubated with *H. pylori* at an MOI of 100 for 12, 24, 36 hours at 37ºC in a 5% CO_2 environment in 24 well plates. The supernatants were collected and stored at -30ºC. 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 an MOI of 100 for 6 hours at 37ºC in a 5% CO_2 environment. The cells were fixed with 4% (w/v) paraformaldehyde-PBS at room temperature for 10 min. The cells were then washed with TBS for 3 times, and blocked with Saponin buffer [10% (v/v) Blocking One (Nakalai, Japan) containing 0.2% (w/v) saponin] at 4ºC for 60 min. Antibodies used for staining were DAPI, Rhodamine Phalloidin (Thermo Fisher SCIENTIFIC, MA, USA), pyCagA. 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).

Extended Data Figures and Tables Legends

Extended Data Figure 1 | The expression levels of candidate RNAs of strains recovered from *H. pylori* ATCC43504-infected rodent stomachs. a, The list of mutation rates and expression levels of the candidate RNAs (mRNA or non-coding RNA) of strains recovered from stomachs of gerbils or mice 8 wks post-infection. The Locus tags highlighted in red indicated the candidates common in both of the strains originated from gerbils and mice. N/A, not applicable. **b,** The expression levels of candidate RNAs (mRNA or non-coding RNA) of isolates recovered from *H. pylori* ATCC43504-infected C57BL/6 mice stomachs. RNA expression levels of the genes or nearby genes of genome regions (Fig. 1b and Supplementary Information 2), which mutated in 100% of the recovered strains were assessed.

Extended Data Figure 2 | The length of the poly-T stretches upstream of the HPnc4160 coding region. a, Schematic diagram of genome DNA sequence around the HPnc4160 and poly-T sequence of genome analyzed strains. Red-colored "T" indicated stretches of poly-T sequence. **b,** Schematic diagram of genome DNA sequence around the HPnc4160 and poly-T sequence of mice-recovered strains 8 wks after infection (n=10, Hp1 to Hp10). Green indicated HPnc4160 transcribed sequence, red frame indicated poly-T mutated stretches in the upstream region of HPnc4160, and red-colored "T" indicated inserted nucleotides of each recovered strain compared with wild-type. **c-e,** Time-dependent change in the length of the poly-T stretches upstream of the HPnc4160 coding region in gerbils or mice-recovered strains. **c**, Strains from Mongolian gerbils infected with ATCC 43504. **d**, Strains from C57BL/6 mice infected with ATCC 43504. **e**, Strains from C57BL/6 mice infected with PMSS1. Data are median with interquartile range. *P* values are

from non-parametric Dunn's multiple comparison test (at two-sided). ns: not significant. **f**, Expression levels of HPnc4160 in the recovered *H. pylori* strains from mice (n=10) and *H. pylori* wild-type (T-repeat 14). Relative expression levels of HPnc4160 were measured by real-time PCR and plotted against T-repeat length in the upstream region of HPnc4160. Data are presented as means with s.d. *P* values represent the results of two-tailed Dunn's multiple comparison test.

Extended Data Figure 3 | Effect of *in vitro* cultivation on the length of the poly-T stretches upstream of the HPnc4160 coding region. a, Experimental strategies schematic. **b**, The raw data of the DNA sequence analysis of *H. pylori* genomes prepared from original culture (Original #1 - #3) and from passaged *in vitro* for 60 times (60-passaged #1 - #3).

Extended Data Figure 4 | The length of the poly-T stretches upstream of the HPnc4160 coding region and RNA expression levels. a, Growth curves of *H. pylori* ATCC43504 mutants mutated in the number of T repeat in HPnc4160 upstream region. **b-c,** The relative expression levels of *hpnc4170* (**b**) and HP0811 (**c**) in the *H. pylori* strains genetically modified with the T-repeat length. Data are presented as means with s.d. (n=3). *P* values are from Dunnett's multiple comparison test (at two-sided). ns: not significant. **d**, The relative RNA expression levels of target candidates of HPnc4160 showed an inverse correlation with HPnc4160. The total RNA from the indicated *H. pylori* ATCC 43504 strains were extracted, reverse transcribed, and provided for qPCR to assess the indicated genes. The results represent the average of three separate experiments (each n=3). Data are presented as means ± s.d. (error bars). Spearman correlation coefficients (r) were used to evaluate the relationships among relative RNA expression of HPnc4160 (Fig. 1h) and each

target.

Extended Data Figure 5 | Predicted HPnc4160 binding sites. a, Growth curves of *H. pylori* ATCC43504 mutants. **b,** Predicted the secondary structure of HPnc4160 RNA by CentroidFold. The bases in the predicted structure are colored according to base-pairing probabilities. Circles in pink and light green color indicated loop structures having probabilities of binding to target RNA sequences. **c,** Schematic diagram of predicted HPnc4160 binding sites in the corresponding 5'UTR sequence of target genes. Upper sequences indicate target mRNA sequences with base numbers, whereas lower sequences indicate HPnc4160 sequence. Colored sequences are corresponding to the loop structures indicated in (**b**). **d**, Binding prediction of HPnc4160 and 5' UTR of *cagA* mRNA. **d**, Schematic diagram of predicted HPnc4160 binding sites in the *cagA* CM/CRPIA motif of cagA CDS. e-f, Schematic diagram of predicted HPnc4160 binding sites in the corresponding CDS sequence of *cagA* TYPE 1 (**e**), TYPE 2 (**f**), and *cagA* nonbinding form (NB-*cagA*) of TYPE 2 (**g**) and TYPE 1 (**h**). Upper sequences indicate target *cagA* mRNA sequences, whereas lower sequences indicate HPnc4160 sequence with base numbers. Colored sequences are corresponding to the loop structures indicated in (**b**). Mutated nucleotides in *cagA* mRNA sequence are shown in red. **i**, Purified RNase III was separated by SDS-PAGE and stained with CBB.

Extended Data Figure 6 | Effect of *cagA*-NB on host-cell-translocated CagA activity. a, Growth curves of *H. pylori* ATCC43504 *cagA*-NB mutant compared with wild-type. **b**, Phosphorylated CagA protein levels in cell lysates of AGS cells infected with *H. pylori* ATCC43504. The whole-cell lysates of AGS cells infected with *H. pylori* strains for 6 hours were subjected to western blot against anti-CagA, anti-pY CagA, anti-UreA, and anti-Actin antibodies. The

band intensities were measured and calculated by ImageJ software. **c**, IL-8 production from AGS cells infected with *H. pylori* ATCC43504. The supernatants from AGS cells infected with *H. pylori* strains shown in the figure for the indicated time were subjected to ELISA system for IL-8 production. The results represent the average of three separate experiments (each n=3). Data are presented as means ± s.d. (error bars).

Extended Data Figure 7 | Characterization of clinical isolates. a, Clinical isolates of non-cancer (Non-Cancer, n=39) and cancer (Cancer, n=17) patients, which used in Fig. 4g, showed equal growth rate. The strains cultured on TSAII containing 5% sheep blood plates for 2 days were inoculated in Brucella broth containing 5% FCS, adjusted OD_{600nm} at 0.1, then cultured in microaerobic condition with agitation for 16 hours. The turbidity of the cultures was assessed at OD_{600nm}. The Data are presented as medians with interquartile range. *P* values represent the results of the two-tailed Mann-Whitney test. ns: not significant. **b**, The relative RNA expression levels of target candidates of HPnc4160. Spearman correlation coefficients (r) were used to evaluate the relationships among relative RNA expression of HPnc4160 and each target. **c**, Comparison of expression levels of mRNA (HELPY_1262 and HP1227) in clinical isolates of non-cancer (NC, n=39) and cancer (C, n=17) patients. The expression levels of mRNA were normalized with the levels of 23S rRNA. Data are presented as medians with interquartile range. *P* values represent the results of the two-tailed Mann-Whitney test. ns: not significant.

Extended Data Figure 8 | Infection-induced silencing of HPnc4160 upregulates target genes expression and promote bacterial host adaptation and canceration during chronic *H. pylori* infection. *H. pylori*

infection *in vivo* leads elongation of T-stretch in the upstream region of HPnc4160 sRNA coding region, which results in decreased expression levels of sRNA HPnc4160. Gene silencing of HPnc4160 results in increased levels of target genes coding OMPs and CagA, and as a result, the levels of bacterial colonization and CagA translocation into the attached host cells were increased.

Extended Data Table 1 | The list of mutated genome regions in the strains recovered from *H. pylori*-infected rodents' stomachs.

The list showed the genome regions that mutated in the strains isolated from the stomachs of rodents 8 weeks post infection. The DNA sequences in the regions listed in the table were mutated in more than 50% of the 40 strains recovered from Mongolian gerbils, or, in all of the 10 strains recovered from C57BL/6 mice. N/A, not applicable.

Extended Data Table 2 | Comparative analysis of expression levels between *H. pylori* ATCC43504 Δ*hpnc4160*/*hpnc4170* mutant and wild-type strains.

a, Comparative analysis of RNA expression levels between *H. pylori* ATCC 43504 Δ*hpnc4160*/*hpnc4170* mutant and wild-type strains by RNA-seq.

Footnote |

Normalized expression level and fold change of the strains were listed. Genes with *P*-values by Empirical Analysis of Digital Gene Expression in R (edgeR) test showed less than 0.001 were listed (17 factors). Eight genes selected by RNA-seq and iTRAQ analysis (Fig. 2c) were highlighted in red.

b, Comparative analysis of protein expression levels between *H. pylori* ATCC43504 Δ*hpnc4160*/*hpnc4170* mutant and wild-type strains by iTRAQ.

Footnote |

Proteins showing relative protein abundance with *P*-value of less than 0.01 were listed (21 factors). Eight proteins selected by RNA-seq and iTRAQ analysis (Fig. 2c) were highlighted in red.

Supplementary Information Legends

Supplementary information 1 | Summary of mutations in the isolates recovered from *H. pylori*-infected Mongolian gerbils.
The number of mutations in the isolates of 40 strains recovered from *H. pylori*-infected Mongolian gerbils' stomachs 8 weeks after post-infection were listed.

Supplementary information 2 | The mutated sequence list of 40 strains recovered from *H. pylori*-infected Mongolian gerbils' stomach.

Supplementary information 3 | Summary of mutations in the isolates recovered from *H. pylori*-infected C57BL/6 mice. The number of mutations in the isolates of 10 strains recovered from H. pylori-infected mice stomachs 8 weeks after post-infection were listed.

Supplementary information 4 | The mutated sequence list of 10 strains recovered from *H. pylori*-infected C57BL/6 mice stomach.

Supplementary information 5 | Information of *H. pylori* clinical isolates used in Fig. 1f, and Fig 4g.

Supplementary information 6 | Primers used in this study.

Acknowledgements

The authors would like to thank Manuel Amieva for providing us *H. pylori* strains. We gratefully acknowledge Keisuke Katsura for his 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 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.)], the Naito Foundation, the Tokyo Biochemical Research Foundation (to H.M. and P.S.), and the Kao Foundation for Arts and Sciences (to K.K.). This work was supported by MEXT KAKENHI (No. 221S0002).

Details of Author Contributions

Conceived and designed the experiments: RKD, KK, HM; performed the experiments: RKD, KK, RO, YO, TS, TO, TI, PS; analyzed the data RKD, KK, YO, ZB, RY, KN, TH, HM; contributed reagents/materials/analysis tools: RKD, KK, TS, TVP, EK, AH, MT, AS, HA, TTB, LTN, KVV, DQDH, TS, YY, HM; prepared the manuscript: RKD, KK, TH, HM.

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., Pérez-Pérez, G.I., Meuwissen, S.G. & Blaser, M. J. Helicobacter pylori and atrophic gastritis: importance of the cagA status. *J Natl Cancer I* **87**, 1777–80 (1995).

3. Blaser, M. *et al.* Infection with Helicobacter pylori strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res* **55**, 2111–5 (1995).

4. Mimuro, H. *et al.* Grb2 Is a Key Mediator of Helicobacter pylori CagA Protein Activities. *Mol Cell* **10**, 745–755 (2002).

5. Viala, J. *et al.* Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. *Nat Immunol* **5**, 1166 1174 (2004).

6. Mimuro, H. *et al.* Helicobacter pylori dampens gut epithelial self-renewal by inhibiting apoptosis, a bacterial strategy to enhance colonization of the stomach. *Cell Host Microbe* **2**, 250 263 (2007).

7. Suzuki, M. *et al.* Helicobacter pylori CagA phosphorylation-independent function in epithelial proliferation and inflammation. *Cell Host Microbe* **5**, 23 34 (2009).

8. Backert, S. & Naumann, M. What a disorder: proinflammatory signaling pathways induced by Helicobacter pylori. *Trends Microbiol* **18**, 479 486 (2010).

9. Zimmermann, S. *et al.* ALPK1- and TIFA-Dependent Innate Immune Response Triggered by the Helicobacter pylori Type IV Secretion System. *Cell Reports* **20**, 2384 2395 (2017).

10. Gall, A., Gaudet, R. G., Gray-Owen, S. D. & Salama, N. R. TIFA Signaling in Gastric Epithelial Cells Initiates the cag Type 4 Secretion System-Dependent Innate Immune Response to Helicobacter pylori Infection. *Mbio* **8**, (2017).

11. Stein, S. C. *et al.* Helicobacter pylori modulates host cell responses by CagT4SS-dependent translocation of an intermediate metabolite of LPS inner core heptose biosynthesis. *Plos Pathog* **13**, e1006514 32 (2017).

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

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

14. Morelli, G. *et al.* Microevolution of Helicobacter pylori during Prolonged Infection of Single Hosts and within Families. *Plos Genet* **6**, e1001036 (2010).

15. Kennemann, L. *et al.* Helicobacter pylori genome evolution during human infection. *Proc National Acad Sci* **108**, (2011).

16. Dutta, T. & Srivastava, S. Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. *Gene* **656**, 60–72 (2018).

17. Vannini, A. *et al.* Comprehensive mapping of the Helicobacter pylori NikR regulon provides new insights in bacterial nickel responses. *Sci Rep-uk* **7**, 45458 (2017).

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

19. Miyakoshi, M., Chao, Y. & Vogel, J. Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Curr Opin Microbiol* **24**, 132–139 (2015).

20. Saadat, I. *et al.* Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature* **447**, 330 333 (2007).

21. Salaün, L., Ayraud, S. & Saunders, N. J. Phase variation mediated niche adaptation during prolonged experimental murine infection with Helicobacter pylori. *Microbiology*+ **151**, 917 923 (2005).

22. Deitsch, K. W., Lukehart, S. A. & Stringer, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* **7**, 493–503 (2009).

23. Konno, M. *et al.* Predominance of Mother-to-Child Transmission of Helicobacter pylori Infection Detected by Random Amplified Polymorphic DNA Fingerprinting Analysis in Japanese Families. *Pediatric Infect Dis J* **27**, 999 1003 (2008).

24. Linz, B. *et al.* A mutation burst during the acute phase of Helicobacter

pylori infection in humans and rhesus macaques. *Nat Commun* **5**, 1 8 (1AD).

25. Hirukawa, S. *et al.* Characterization of morphological conversion of Helicobacter pylori under anaerobic conditions. *Microbiol Immunol* **62**, 221–228 (2018).

26. Park, J., Forman, D., Waskito, L., Yamaoka, Y. & Crabtree, J. E. Epidemiology of Helicobacter pylori and CagA-Positive Infections and Global Variations in Gastric Cancer. *Toxins* **10**, 163 (2018).

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

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

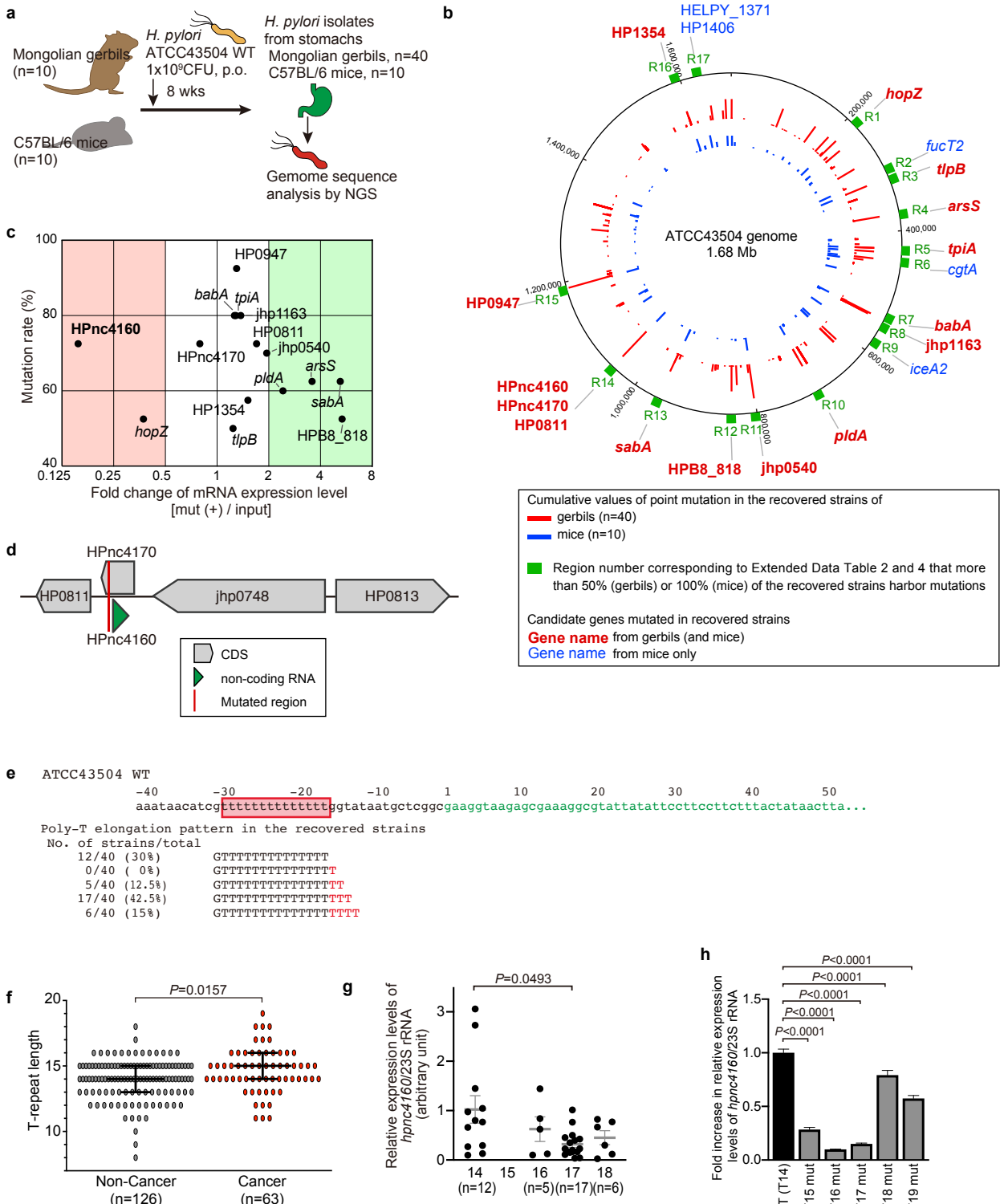


Fig. 1 Kinoshita et al.

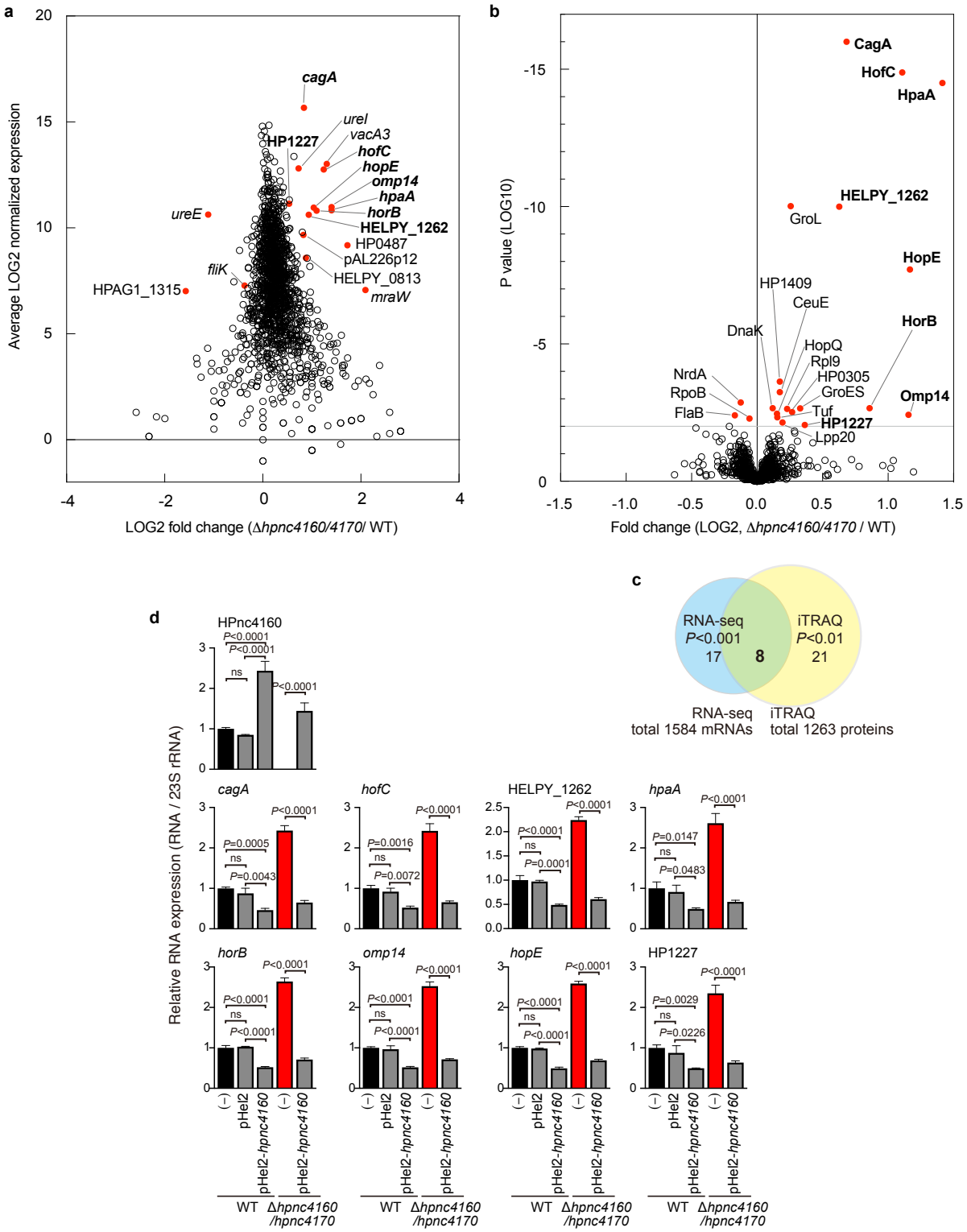


Fig. 2 Kinoshita et al.

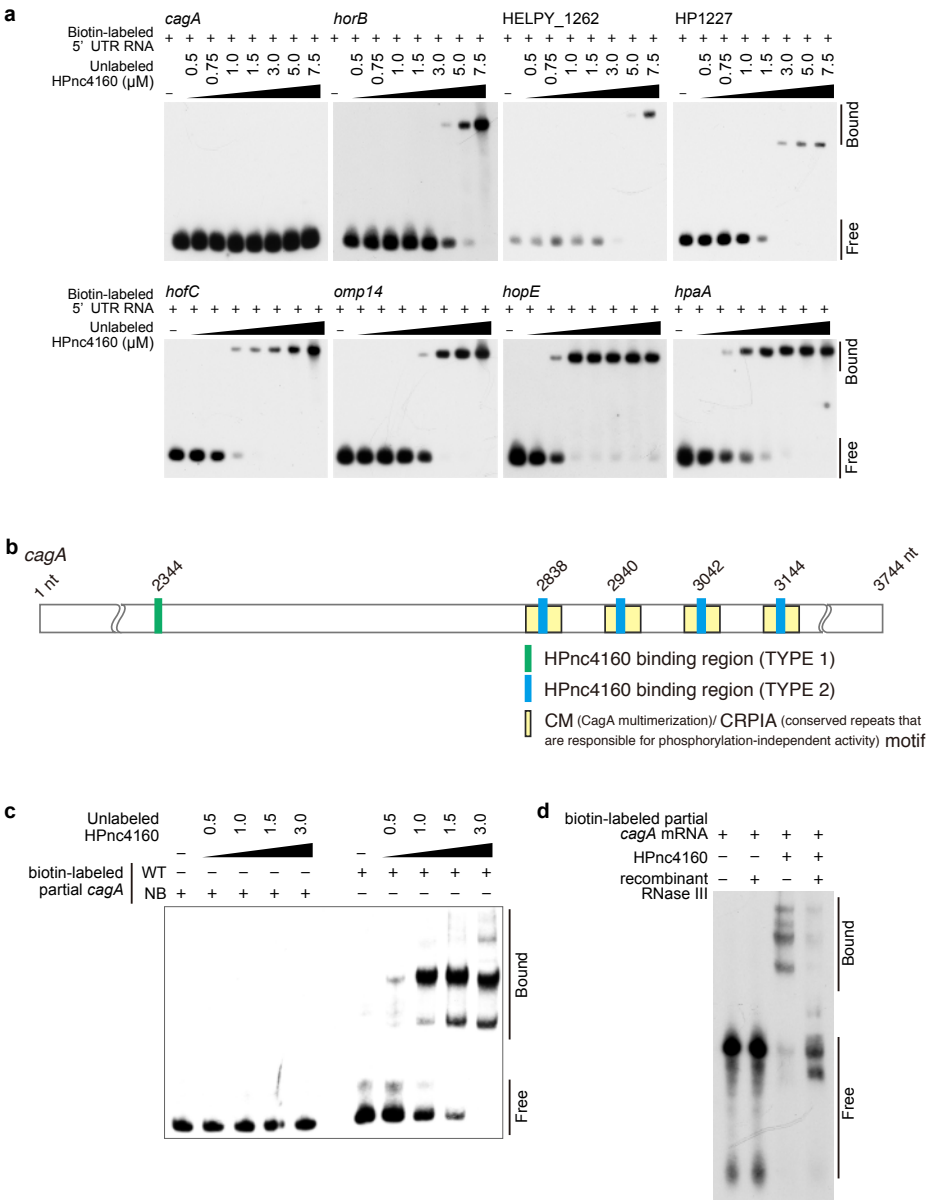


Fig. 3 Kinoshita et al.

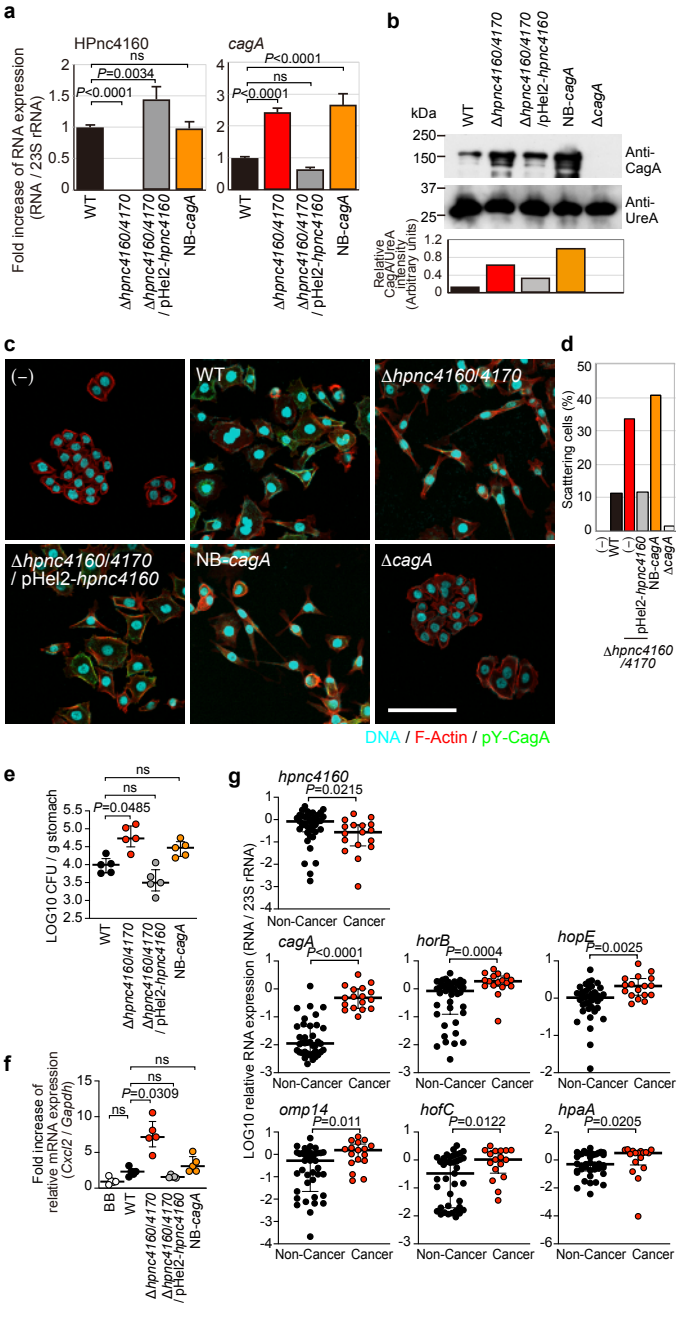
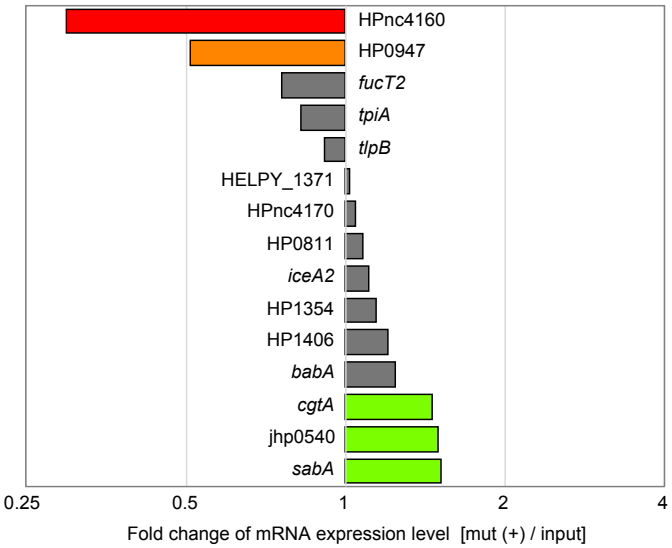


Fig. 4 Kinoshita et al.

a

Region No.	Gene symbol	Gene description	Mongolian gerbils (8 wks post infection)		C57BL/6 mice (8 wks post infection)	
			Fold change of mRNA expression level [mut(+)/input]	Mutation rate (%, /40 strains)	Fold change of mRNA expression level [mut(+)/input]	Mutation rate (%, /10 strains)
R1	*hopZ*	outer membrane protein	0.373	52.5	N/A	N/A
R2	*fucT2*	alpha-1,2-fucosyltransferase	N/A	N/A	0.758	100.0
R3	*tlpB*	methyl-accepting chemotaxis protein	1.239	50.0	0.913	100.0
R4	*arsS*	histidine kinase sensor protein	3.572	62.5	N/A	N/A
R5	*tpiA*	triosephosphate isomerase	1.286	80.0	0.825	100.0
R6	*cgtA*	beta-1,4-N-acetylgalactosamyltransferase	N/A	N/A	1.453	100.0
R7	*babA*	outer membrane protein babA	1.263	80.0	1.243	100.0
R8	jhp1163	hypothetical protein	1.374	80.0	N/A	N/A
R9	iceA2	Ulcer-associated gene restriction endonuclease	N/A	N/A	1.104	100.0
R10	*pldA*	phospholipase A1	2.420	60.0	N/A	N/A
R11	jhp0540	hypothetical protein	1.946	70.0	1.497	100.0
R12	HPB8_818	family 25 glycosyl transferase	5.346	52.5	N/A	N/A
R13	*sabA*	outer membrane protein sabA	5.215	62.5	1.512	100.0
R14	HPnc4160	mRNA/antisense RNA family IsoB	0.155	72.5	0.297	100.0
R14	HPnc4170	mRNA/antisense RNA family aapB	0.794	72.5	1.043	100.0
R14	HP0811	hypothetical protein	1.698	72.5	1.075	100.0
R15	HP0947	hypothetical protein	1.300	92.5	0.511	100.0
R16	HP1354	adenine-specific DNA methyltransferase	1.512	57.5	1.138	100.0
R17	HELPY_1371	Type III restriction enzyme R protein	N/A	N/A	1.018	100.0
R17	HP1406	biotin synthase	N/A	N/A	1.200	100.0

b



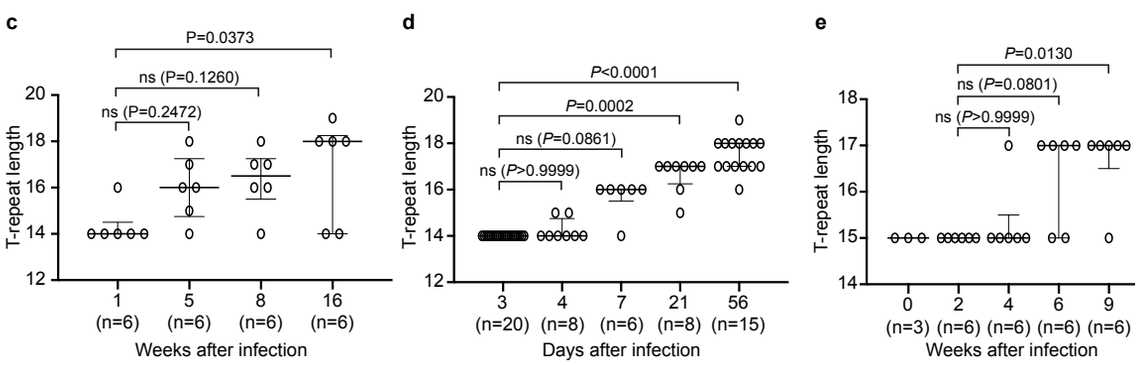
Extended Data Figure 1 Kinoshita et al.

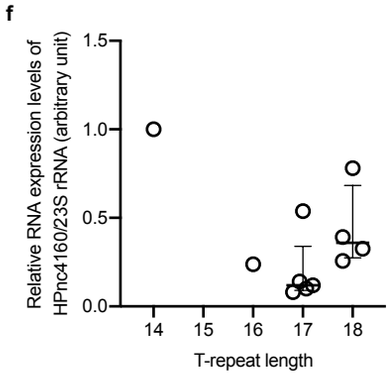
a

```
                                  -40       -30       -20            -10       1        10        20            30        40
   H. pylori ATCC43504  aaataacatcgtttttttttttttt-----ggtataatgctcggcgaaggtaagagcgaaaggcgtattata----ttccttccttctt
       H. pylori PMSS1  aaataacatc-tttttttttttt-------ggtataatgc-cgcctaaggtaagagcgaaaggcgtattata----ttccttccttctt
        H. pylori  SS1  aaataacatc-tttttttttttt------ggtataatgc-cgcctaaggtaagagcgaaaggcgtattata----ttccttccttctt
       H. pylori 26695  aaataacatcgtttttttttttttt-----ggtataatgctcggcgaaggtaagagcgaaaggcgtattata----ttccttctttctt
         H. pylori J99  aaataacatcgtttttttttttttt-----ggtataatgctcggcgaaggtaagagcgaaaggcgtattata----ttccttccttctt
       H. pylori HPAG1  aaataacatcgttttttttttttttttt--ggtataatgctcggcgaaggtaagagcgaaaggcgtattata----ttccttccttctt
         H. pylori G27  aaataacatc-tttttttttttttttt---ggtataatgc-cgcttaaggtaagagcgaaaggcgtattata----ttccttccttctt
      H. pylori Shi470  aaataacatcgtttttttttttttt-----ggtataatgctcggtgaaggtaagagcaaaaggcgtattatattccttccttccttctt
 H. acinonychis Sheeba  aaataaca---tttttttttttagtataatagtataatgt-tgttg--ggtaagggcaaagggcg-aaaata----ttccttccttctt

                                  41        50         60        70               80        90         100
   H. pylori ATCC43504  tactataac-ttagca-ttttaatcaacttttt-------cattaaaatgtcctgacgctcttacctt-aa
       H. pylori PMSS1  tactataac-ttagca-ttttaatcaacttttt-------cattaaaatgtcctgacgctcttaccttttc
       H. pylori   SS1  tactataac-ttagca-ttttaatcaacttttt-------cattaaaatgtcctgacgctcttaccttttc
       H. pylori 26695  tactataac-ttagca-ttttaatcaacttttt-------cattaaaatgtcctgacgctcttacctt-aa
         H. pylori J99  tactataac-ttagca-ttttaatcaacttttt-------cattaaaatgtcctgacgctcttaccttaaa
       H. pylori HPAG1  tactataac-ttagca-ttttaatcaacttttt-------cattaaaatgtcctgacgctcttacctt-aa
         H. pylori G27  tactataac-ttagca-ttttaatcaacttttt-------cattaaaatgtcctgacgctcttacctt---
      H. pylori Shi470  tactataac-ttagca-ttttaataaacttttt-------cattaaaatgtcctgacgctcttaccttcaa
 H. acinonychis Sheeba  tactataacattagcatttttagtaaactttttctttttacattaaaatgtcctaatgctcttacctt--a
```

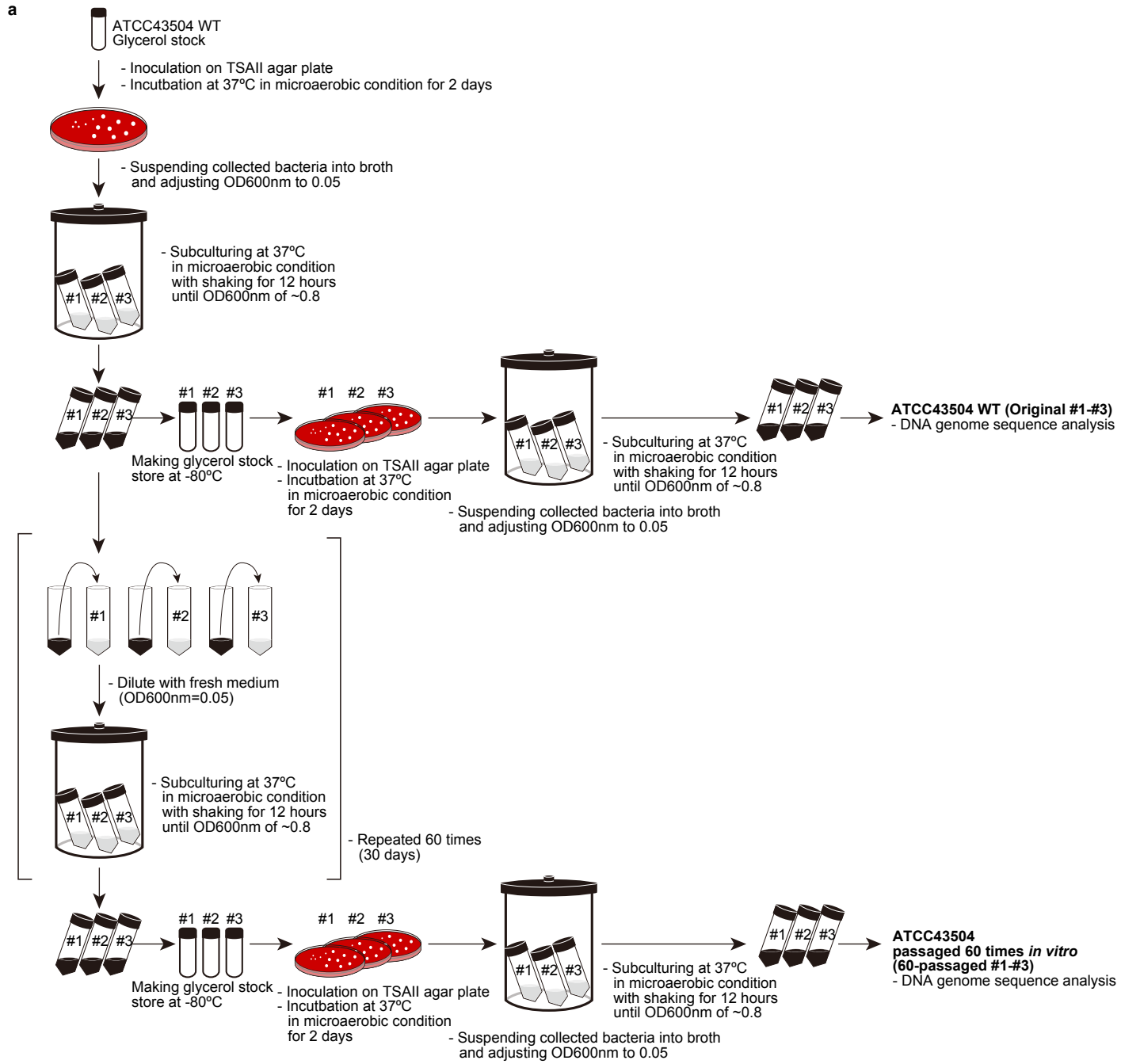
b

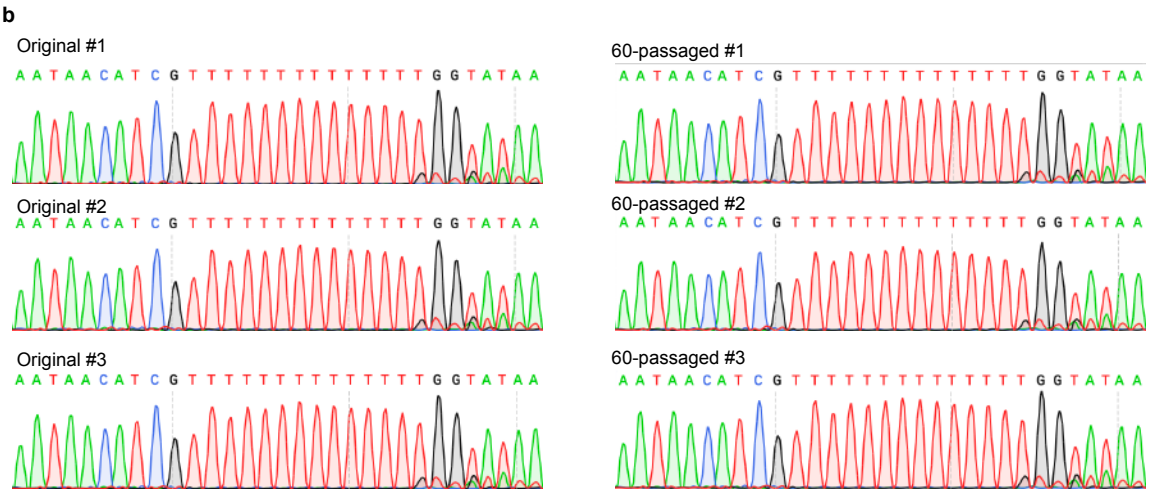
```
ATCC43504 WT
          -40       -30       -20       -10       1        10        20        30        40        50
          aaataacatcgtttttttttttttggtataatgctcggcgaaggtaagagcgaaaggcgtattatattccttccttctttactataactta...

Poly-T elongation pattern in the recovered strains from mice
 No. of strains/total
     1/10 (10%)    GTTTTTTTTTTTTTTT
     5/10 (50%)    GTTTTTTTTTTTTTTTT
     4/40 (40%)    GTTTTTTTTTTTTTTTTT
```



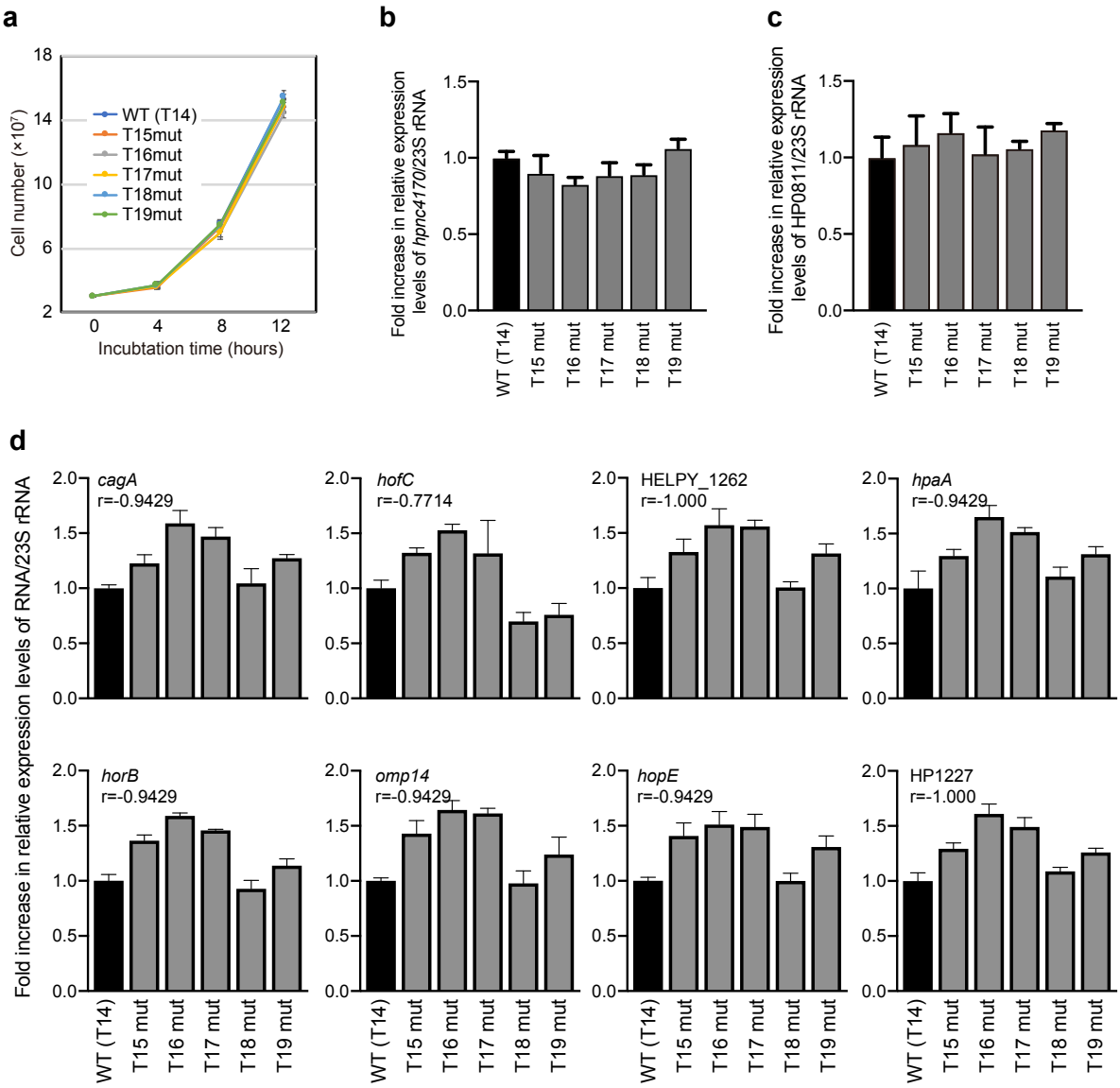


Extended Data Figure 2 Kinoshita et al.

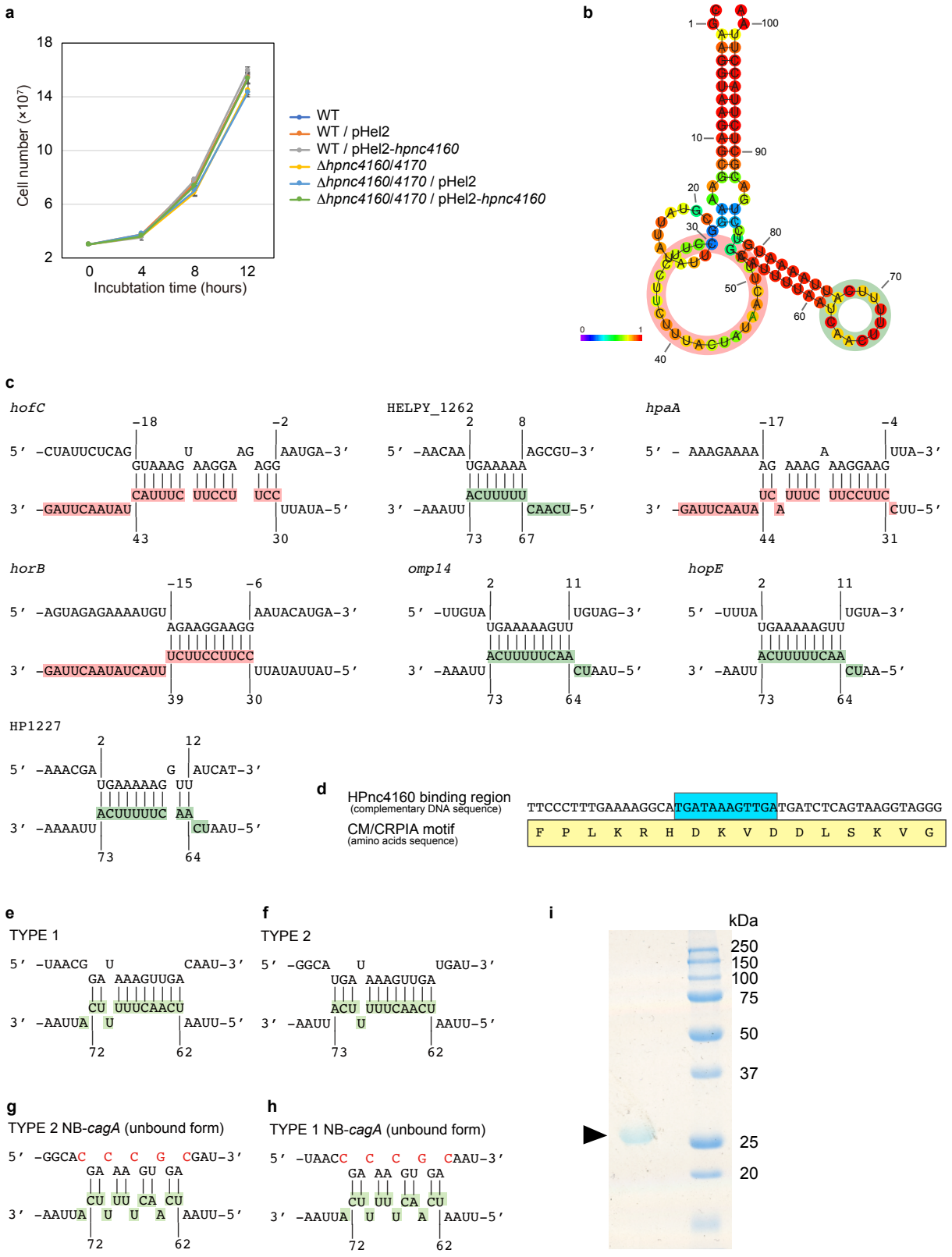




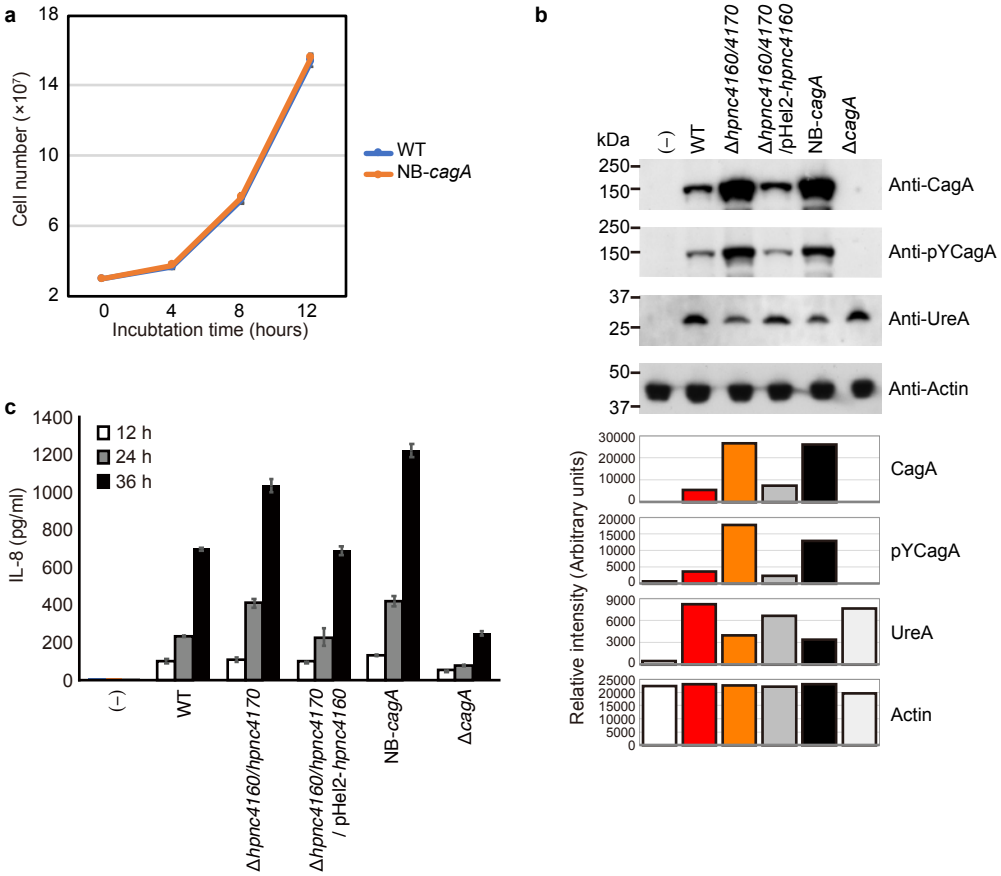
Extended Data Figure 3 Kinoshita et al.



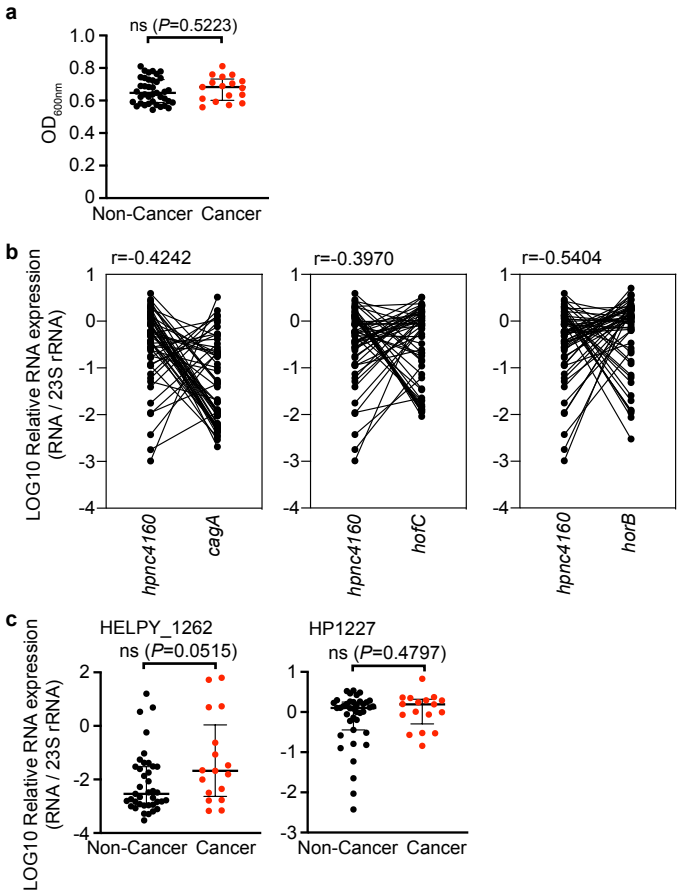
Extended Data Figure 4 Kinoshita et al.



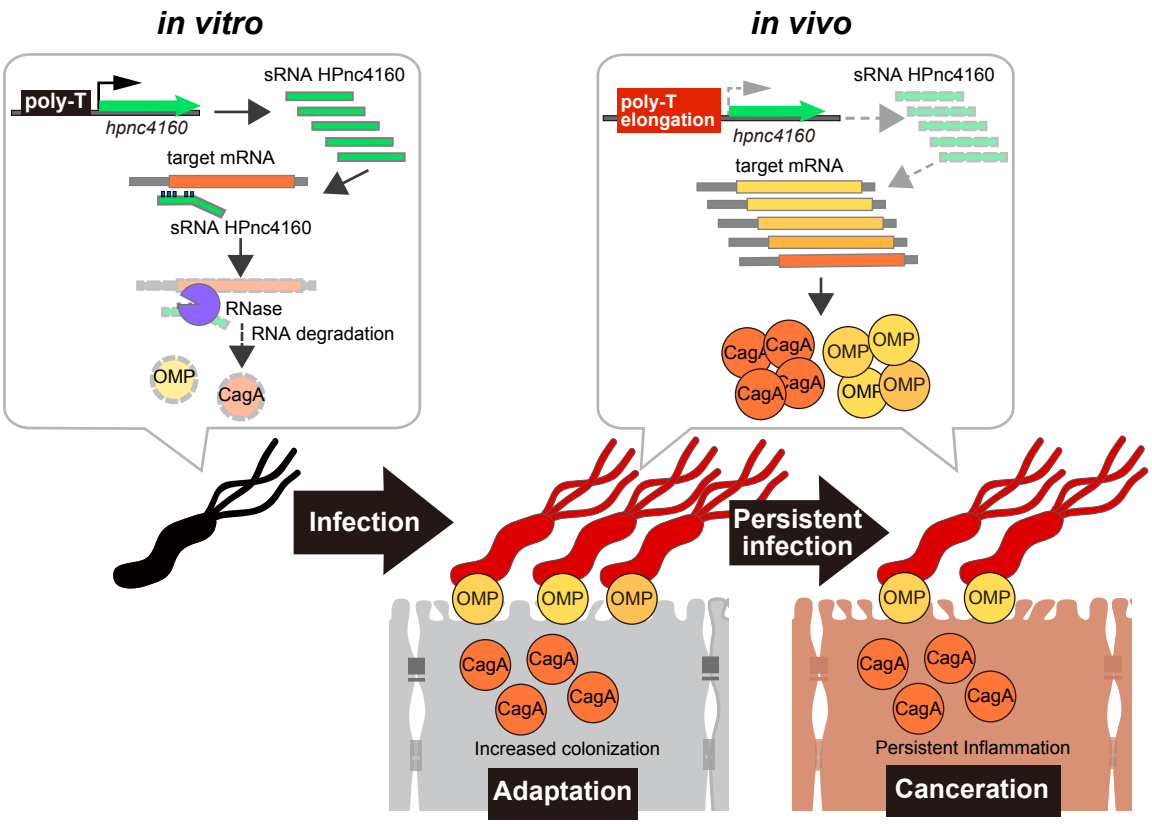
Extended Data Figure 5 Kinoshita et al.



Extended Data Figure 6 Kinoshita et al.



Extended Data Figure 7 Kinoshita et al.



Extended Data Figure 8 Kinoshita et al.

Region No.	No. of mutated strains (%)		Annotation			Gene nearby the intergenic region (#1)			Gene nearby the intergenic region (#2)		
	Mongolian gerbil (n=40)	C57BL/6 (n=10)	Gene name	Locus tag	Gene description	Gene name	Locus tag	Gene description	Gene name	Locus tag	Gene description
R1	52.5	N/A	N/A	N/A	intergenic: no annotation	hopZ	HPG27_8	outer membrane protein			
R2	N/A	100	futT2	HPP12_0096	alpha-1,2-fucosyltransferase						
R3	50	100	N/A	N/A	intergenic: no annotation	tlpB	HP0103	methyl-accepting chemotaxis protein			
R4	62.5	N/A	araS	jhp0151	signal-transducing protein, histidine						
R5	80	100	tpiA	HP0194	triosephosphate isomerase						
R6	N/A	100	HP0217	HP0217	hypothetical protein						
R7	72.5	100	N/A	N/A	intergenic: no annotation	babA	HP1243/ HPG27_298	outer membrane protein			
R8	80	N/A	N/A	N/A	intergenic: no annotation	jhp1163	jhp1163	hypothetical protein	babA	HP1243/ HPG27_298*	outer membrane protein
R9	N/A	100	N/A	N/A	intergenic: no annotation	IceA2	HPATCC43 504_00587	induced by contact with epithelium gene			
R10	60	N/A	pldA	HP0499	phospholipase A1						
R11	70	100	N/A	N/A	intergenic: no annotation	jhp0540	jhp0540	HAD superfamily			
R12	52.5	N/A	HPB8_818	HPB8_818	family 25 glycosyl transferase						
R13	62.5	100	N/A	N/A	intergenic: no annotation	sabA	HPG27_680	outer membrane protein			
R14	72.5	100	aapB	HPnc4170	small peptide	HP0811	HP0811	hypothetical protein	IsoB	HPnc4160	small RNA
R15	92.5	100	N/A	N/A	intergenic: no annotation	HP0947	HP0947	hypothetical protein			
R16	57.5	100	HP1354	HP1354	adenine-specific DNA						
R17	N/A	100	N/A	N/A	intergenic: no annotation	bioB	HP1406	biotin synthase	HELPY_1371	HELPY_1371	type III restriction enzyme R protein

Extended Data Table 1 Kinoshita et al.

a

Gene name	Average LOG2 normalized expression	LOG2 fold change (Δ*hpnc4160/4170* / WT)
cagA	15.6679564	0.8345757
vacA3	13.0181033	1.3000122
ureI	12.8122441	0.7256751
hofC	12.7562529	1.2385235
HP1227	11.1439545	0.5356426
hpaA	10.9856930	1.3994257
horB	10.9589406	1.0351050
omp14	10.8268799	1.3977712
hopE	10.8143652	1.0951572
ureE	10.6269817	-1.1164069
HELPY_1262	10.6234286	0.9346910
pAL226p12	9.6669233	0.8237894
HP0487	9.1803234	1.7213025
HELPY_0813	8.5854796	0.8765494
fliK	7.2801813	-0.3748484
mraW	7.0680550	2.0913744
HPAG1_1315	7.0164749	-1.5753123

b

Gene name	Fold change (LOG2, Δhp*nc4160*/*hpnc4170* / WT)	Pval (LOG10)
CagA	0.6831125	-16.0000000
HofC	1.1079556	-14.8875441
HpaA	1.4153801	-14.5034266
GroL	0.2561653	-10.0187654
HELPY_1262	0.6268596	-9.9953361
HopE	1.1672937	-7.7135456
HP1409	0.1735112	-3.6271647
CeuE	0.1737671	-3.2477470
NrdA	-0.1246917	-2.8700087
HorB	0.8583789	-2.6608050
DnaK	0.1192898	-2.6579389
GroES	0.3291810	-2.6528395
Rpl9	0.2299570	-2.6372507
HP0305	0.2674759	-2.5150013
HopQ	0.1519887	-2.4589011
Omp14	1.1557492	-2.4194920
FlaB	-0.1694201	-2.4003455
Tuf	0.1548424	-2.3314305
RpoB	-0.0584429	-2.2822269
Lpp20	0.1944653	-2.1385174
HP1227	0.3642362	-2.0521036

Extended Data Table 2 Kinoshita et al.

Supplementary Information 1| Summary of mutations in the isolates recovered from *H. pylori*-infected Mongolian gerbils.
Number of mutations in the isolates of 40 strains recovered from *H. pylori*-infected Mongolian gerbils' stomachs 8 weeks after post infection were listed.

Animal No.	Strain Name	Total No. of mutations	SNPs (single nucleotide polymorphysms)				indel (insertion/deletion)			SNPs rate per base per year	indel rate per base per year
			total	nonsynonymous	synonymous	intergenic	Total	genic	intergenic		
#1	S41	50	37	15	18	4	13	5	8	1.43E-04	5.04E-05
	S42	48	40	15	20	5	8	2	6	1.55E-04	3.10E-05
	S43	44	38	12	21	5	6	4	2	1.47E-04	2.33E-05
	S44	46	41	16	21	4	5	1	4	1.59E-04	1.94E-05
#2	S45	48	39	15	17	7	9	2	7	1.51E-04	3.49E-05
	S46	47	38	22	14	2	9	1	8	1.47E-04	3.49E-05
	S47	46	36	15	15	6	10	2	8	1.40E-04	3.88E-05
	S48	48	41	20	16	5	7	2	5	1.59E-04	2.71E-05
#3	S49	17	12	7	3	2	5	4	1	4.65E-05	1.94E-05
	S50	26	18	8	8	2	7	3	4	6.98E-05	2.71E-05
	S51	17	11	7	2	2	6	3	3	4.27E-05	2.33E-05
	S52	20	13	7	2	4	7	4	3	5.04E-05	2.71E-05
#4	S53	27	16	7	6	3	11	5	6	6.20E-05	4.27E-05
	S54	15	10	4	5	1	5	1	4	3.88E-05	1.94E-05
	S55	24	14	5	5	4	10	3	7	5.43E-05	3.88E-05
	S56	25	13	5	6	2	12	4	8	5.04E-05	4.65E-05
#5	S57	25	15	5	7	3	10	4	6	5.82E-05	3.88E-05
	S58	29	15	5	5	5	14	5	9	5.82E-05	5.43E-05
	S59	23	13	5	6	2	10	2	8	5.04E-05	3.88E-05
	S60	27	16	5	7	4	11	3	8	6.20E-05	4.27E-05
#6	S61	29	20	8	8	4	9	2	7	7.76E-05	3.49E-05
	S62	35	25	10	12	3	10	2	8	9.69E-05	3.88E-05
	S63	25	16	6	5	5	9	3	6	6.20E-05	3.49E-05
	S64	36	28	14	10	4	8	2	6	1.09E-04	3.10E-05
#7	S65	50	39	14	21	4	11	5	6	1.51E-04	4.27E-05
	S66	47	39	13	22	4	8	5	3	1.51E-04	3.10E-05
	S67	55	44	15	22	7	11	4	7	1.71E-04	4.27E-05
	S68	54	44	14	23	7	10	4	6	1.71E-04	3.88E-05
#8	S69	52	41	21	17	3	11	3	8	1.59E-04	4.27E-05
	S70	61	52	17	30	5	9	2	7	2.02E-04	3.49E-05
	S71	41	33	15	14	4	7	1	6	1.28E-04	2.71E-05
	S72	39	34	15	16	3	5	0	5	1.32E-04	1.94E-05
#9	S73	43	37	17	15	4	6	2	4	1.43E-04	2.33E-05
	S74	79	70	32	31	7	9	2	7	2.71E-04	3.49E-05
	S75	71	63	29	26	8	8	4	4	2.44E-04	3.10E-05
	S76	43	34	15	17	2	9	3	6	1.32E-04	3.49E-05
#10	S77	45	36	8	6	21	9	3	6	1.40E-04	3.49E-05
	S78	41	34	9	5	20	7	2	5	1.32E-04	2.71E-05
	S79	41	32	5	6	21	9	3	6	1.24E-04	3.49E-05
	S80	41	32	7	5	20	9	3	6	1.24E-04	3.49E-05
									Average	1.19E-04	3.38E-05
									SD	5.60E-05	8.51E-06

Strain name	Position in ATCC43504	Sequence in ATCC43504	Sequence in Reisolate		Type of difference	Annotation	Gene nearby the intergenic region (#1)	Gene nearby the intergenic region (#2)
S41	89186	T	C	SNP	intergenic		HP0995	
S41	99327	C	T	SNP	synonymous	jhp0935		
S41	168379	C	T	SNP	synonymous	HP1547		
S41	199597	C	T	SNP	intergenic		HP1582	jhp1488
S41	204483	CGTTCCGCTTTCAGTCTCGCTTCTTGTTCCGCTTTCA GTCTCGCTTCTTGTTCCGCTTTCAGTCTCGCTTCTT GTTCCGCTTTCA	CGTTCCGCTTTCA	indel	frameshift deletion	HPATCC43504_00192		
S41	232964	C	A	SNP	intergenic		HPPC_00115	
S41	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S41	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S41	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S41	430922	G	A	SNP	nonsynonymous	HP0194		
S41	455599	TGGGGGGGGGG	TGGGGGGGGG	indel	frameshift deletion	HP0217		
S41	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S41	672079	C	A	SNP	nonsynonymous	HP1134		
S41	691071	G	A	SNP	synonymous	HP0489		
S41	691074	A	C	SNP	synonymous	HP0489		
S41	691077	G	A	SNP	nonsynonymous	HP0489		
S41	691089	A	G	SNP	synonymous	HP0489		
S41	691091	C	A	SNP	nonsynonymous	HP0489		
S41	691101	T	C	SNP	synonymous	HP0489		
S41	691113	G	T	SNP	nonsynonymous	HP0489		
S41	691157	C	A	SNP	nonsynonymous	HP0489		
S41	691173	C	A	SNP	synonymous	HP0489		
S41	691175	A	G	SNP	nonsynonymous	HP0489		
S41	691179	C	G	SNP	synonymous	HP0489		
S41	691185	A	G	SNP	nonsynonymous	HP0489		
S41	691189	C	G	SNP	nonsynonymous	HP0489		
S41	691190	A	C	SNP	nonsynonymous	HP0489		
S41	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S41	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S41	839999	T	C	SNP	nonsynonymous	HPB8_818		
S41	873661	C	T	SNP	synonymous	HP0651		
S41	873681	G	A	SNP	synonymous	HP0651		
S41	873702	G	T	SNP	synonymous	HP0651		
S41	873717	C	T	SNP	synonymous	HP0651		
S41	879985	C	T	SNP	nonsynonymous	HP0656		
S41	952086	TTCTCTCTCTCTCTCTC	TTCTCTCTCTCTCTC	indel	intergenic		HPG27_677	
S41	958574	CGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_680	
S41	958719	G	A	SNP	intergenic		HPG27_680	
S41	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S41	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S41	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S41	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S41	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S41	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S41	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S41	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S41	1582085	TTT	TTTTTT	indel	intergenic		HELPY_1317	HP0228
S41	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S41	1592801	T	G	SNP	synonymous	HP1354		
S41	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
S42	99327	C	T	SNP	synonymous	jhp0935		
S42	168379	C	T	SNP	synonymous	HP1547		
S42	199597	C	T	SNP	intergenic		HP1582	jhp1488
S42	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S42	372979	T	C	SNP	intergenic		HP0162	
S42	374769	T	G	SNP	nonsynonymous	jhp0151		
S42	430922	G	A	SNP	nonsynonymous	HP0194		
S42	548957	C	T	SNP	intergenic		jhp1163	HPG27_298
S42	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S42	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S42	672079	C	A	SNP	nonsynonymous	HP1134		
S42	691071	G	A	SNP	synonymous	HP0489		
S42	691074	A	C	SNP	synonymous	HP0489		
S42	691077	G	A	SNP	nonsynonymous	HP0489		
S42	691089	A	G	SNP	synonymous	HP0489		
S42	691091	C	A	SNP	nonsynonymous	HP0489		
S42	691101	T	C	SNP	synonymous	HP0489		
S42	691113	G	T	SNP	nonsynonymous	HP0489		
S42	691157	C	A	SNP	nonsynonymous	HP0489		
S42	691173	C	A	SNP	synonymous	HP0489		
S42	691175	A	G	SNP	nonsynonymous	HP0489		
S42	691179	C	G	SNP	synonymous	HP0489		
S42	691185	A	G	SNP	nonsynonymous	HP0489		
S42	691189	C	G	SNP	nonsynonymous	HP0489		
S42	691190	A	C	SNP	nonsynonymous	HP0489		
S42	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S42	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S42	873661	C	T	SNP	synonymous	HP0651		
S42	873681	G	A	SNP	synonymous	HP0651		
S42	873702	G	T	SNP	synonymous	HP0651		
S42	873717	C	T	SNP	synonymous	HP0651		
S42	873807	C	G	SNP	synonymous	HP0651		
S42	874024	AGGGGGGGGG	AGGGGGGGGGG	indel	intergenic		HP0651	
S42	879985	C	T	SNP	nonsynonymous	HP0656		

S42	958574	CGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_680	
S42	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S42	1124346	CGAGAGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_1187	
S42	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S42	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S42	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S42	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S42	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S42	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S42	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S42	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S42	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S42	1592786	GT	G	indel	frameshift deletion	HP1354		
S42	1663529	GTTTTT	GTTTT	indel	1 bp deletion			
S43	99327	C	T	SNP	synonymous	jhp0935		
S43	168379	C	T	SNP	synonymous	HP1547		
S43	199597	C	T	SNP	intergenic		HP1582	jhp1488
S43	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S43	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S43	430922	G	A	SNP	nonsynonymous	HP0194		
S43	466419	TGAGAGAGAGAGAGAG	TGAGAGAGAGAGAG	indel	intergenic		HP0227	
S43	540566	GGAGATTAAACAAGAGATTAAACAAGAGATTAA ACAAGAG	GGAGATTAAACAAGAGATTAAACAAGAGATTAA ACAAGAGATTAAACAAGAG	indel	frameshift insertion			
S43	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S43	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S43	548994	A	T	SNP	intergenic		jhp1163	HPG27_298
S43	672079	C	A	SNP	nonsynonymous	HP1134		
S43	691071	G	A	SNP	synonymous	HP0489		
S43	691074	A	C	SNP	synonymous	HP0489		
S43	691077	G	A	SNP	nonsynonymous	HP0489		
S43	691089	A	G	SNP	synonymous	HP0489		
S43	691091	C	A	SNP	nonsynonymous	HP0489		
S43	691101	T	C	SNP	synonymous	HP0489		
S43	691113	G	T	SNP	nonsynonymous	HP0489		
S43	691173	C	A	SNP	synonymous	HP0489		
S43	691175	A	G	SNP	nonsynonymous	HP0489		
S43	691179	C	G	SNP	synonymous	HP0489		
S43	691185	A	G	SNP	nonsynonymous	HP0489		
S43	691189	C	G	SNP	nonsynonymous	HP0489		
S43	691190	A	C	SNP	nonsynonymous	HP0489		
S43	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S43	873661	C	T	SNP	synonymous	HP0651		
S43	873681	G	A	SNP	synonymous	HP0651		
S43	873702	G	T	SNP	synonymous	HP0651		
S43	873717	C	T	SNP	synonymous	HP0651		
S43	948328	C	A	SNP	synonymous	HPP12_0726		
S43	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S43	1318527	C	A	SNP	intergenic	jhp1031		
S43	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S43	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S43	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S43	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S43	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S43	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S43	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S43	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S43	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S43	1592801	T	G	SNP	synonymous	HP1354		
S43	1593602	TGGGGGGGGGGG	G	indel	frameshift deletion	HP1354		
S44	99327	C	T	SNP	synonymous	jhp0935		
S44	168379	C	T	SNP	synonymous	HP1547		
S44	199597	C	T	SNP	intergenic		HP1582	jhp1488
S44	232964	C	A	SNP	intergenic		HPPC_00115	
S44	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S44	430922	G	A	SNP	nonsynonymous	HP0194		
S44	549326	G	T	SNP	intergenic		jhp1163	HPG27_298
S44	672079	C	A	SNP	nonsynonymous	HP1134		
S44	691071	G	A	SNP	synonymous	HP0489		
S44	691074	A	C	SNP	synonymous	HP0489		
S44	691077	G	A	SNP	nonsynonymous	HP0489		
S44	691089	A	G	SNP	synonymous	HP0489		
S44	691091	C	A	SNP	nonsynonymous	HP0489		
S44	691101	T	C	SNP	synonymous	HP0489		
S44	691113	G	T	SNP	nonsynonymous	HP0489		
S44	691173	C	A	SNP	synonymous	HP0489		
S44	691175	A	G	SNP	nonsynonymous	HP0489		
S44	691179	C	G	SNP	synonymous	HP0489		
S44	691185	A	G	SNP	nonsynonymous	HP0489		
S44	691189	C	G	SNP	nonsynonymous	HP0489		
S44	691190	A	C	SNP	nonsynonymous	HP0489		
S44	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S44	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S44	839999	T	C	SNP	nonsynonymous	HPB8_818		
S44	873661	C	T	SNP	synonymous	HP0651		
S44	873681	G	A	SNP	synonymous	HP0651		
S44	873702	G	T	SNP	synonymous	HP0651		
S44	873717	C	T	SNP	synonymous	HP0651		

S44	879985	C	T	SNP	nonsynonymous	HP0656		
S44	958574	CGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_680	
S44	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S44	1066590	T	A	SNP	intergenic		jhp0775	
S44	1187376	ATTTTTTT	ATTTTTTTTT	indel	intergenic		HP0947	
S44	1192843	T	G	SNP	nonsynonymous	HPP12_0950		
S44	1252178	C	T	SNP	nonsynonymous	HP0407		
S44	1284281	T	C	SNP	synonymous	HPG27_1018		
S44	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S44	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S44	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S44	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S44	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S44	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S44	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S44	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S44	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S44	1592801	T	G	SNP	synonymous	HP1354		
S45	99327	C	T	SNP	synonymous	jhp0935		
S45	168379	C	T	SNP	synonymous	HP1547		
S45	199597	C	T	SNP	intergenic		HP1582	jhp1488
S45	251374	G	A	SNP	nonsynonymous	HP0045		
S45	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S45	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S45	424517	CGGGG	CGGGGG	indel	intergenic		HP0189	HP0188
S45	430922	G	A	SNP	nonsynonymous	HP0194		
S45	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S45	548957	C	T	SNP	intergenic		jhp1163	HPG27_298
S45	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S45	548994	A	T	SNP	intergenic		jhp1163	HPG27_298
S45	691071	G	A	SNP	synonymous	HP0489		
S45	691074	A	C	SNP	synonymous	HP0489		
S45	691077	G	A	SNP	nonsynonymous	HP0489		
S45	691089	A	G	SNP	synonymous	HP0489		
S45	691091	C	A	SNP	nonsynonymous	HP0489		
S45	691101	T	C	SNP	synonymous	HP0489		
S45	691113	G	T	SNP	nonsynonymous	HP0489		
S45	691157	C	A	SNP	nonsynonymous	HP0489		
S45	691173	C	A	SNP	synonymous	HP0489		
S45	691175	A	G	SNP	nonsynonymous	HP0489		
S45	691179	C	G	SNP	synonymous	HP0489		
S45	691185	A	G	SNP	nonsynonymous	HP0489		
S45	691189	C	G	SNP	nonsynonymous	HP0489		
S45	691190	A	C	SNP	nonsynonymous	HP0489		
S45	748650	T	A	SNP	intergenic		HP0514	
S45	805851	A	T	SNP	intergenic		HPG27_556	
S45	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S45	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S45	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S45	1187376	ATTTTTTT	ATTTTTTTTT	indel	intergenic		HP0947	
S45	1262251	C	A	SNP	nonsynonymous	jhp0981		
S45	1295885	GATATATATATATATATAT	GATATATATATATATATATAT,GATATATATATATATATAT	indel	intergenic		HPATCC43504_01238	
S45	1318527	C	A	SNP	intergenic	jhp1031		
S45	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S45	1332888	T	G	SNP	synonymous	HPATCC43504_01275		
S45	1332893	T	G	SNP	nonsynonymous	HPATCC43504_01275		
S45	1332897	A	G	SNP	synonymous	HPATCC43504_01275		
S45	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S45	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S45	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S45	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S45	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S45	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S45	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S45	1582072	ATT	ATTTTT	indel	intergenic		HELPY_1317	HP0228
S45	1592786	GT	G	indel	frameshift deletion	HP1354		
S46	99327	C	T	SNP	synonymous	jhp0935		
S46	168379	C	T	SNP	synonymous	HP1547		
S46	199597	C	T	SNP	intergenic		HP1582	jhp1488
S46	251374	G	A	SNP	nonsynonymous	HP0045		
S46	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S46	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S46	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S46	424517	CGGGG	CGGGGG	indel	intergenic		HP0189	HP0188
S46	430922	G	A	SNP	nonsynonymous	HP0194		
S46	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S46	691071	G	A	SNP	synonymous	HP0489		
S46	691074	A	C	SNP	synonymous	HP0489		
S46	691077	G	A	SNP	nonsynonymous	HP0489		
S46	691089	A	G	SNP	synonymous	HP0489		
S46	691091	C	A	SNP	nonsynonymous	HP0489		
S46	691101	T	C	SNP	synonymous	HP0489		
S46	691113	G	T	SNP	nonsynonymous	HP0489		
S46	691157	C	A	SNP	nonsynonymous	HP0489		
S46	691173	C	A	SNP	synonymous	HP0489		
S46	691175	A	G	SNP	nonsynonymous	HP0489		

S46	691179	C	G	SNP	synonymous	HP0489		
S46	691185	A	G	SNP	nonsynonymous	HP0489		
S46	691189	C	G	SNP	nonsynonymous	HP0489		
S46	691190	A	C	SNP	nonsynonymous	HP0489		
S46	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S46	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S46	839999	T	C	SNP	nonsynonymous	HPB8_818		
S46	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S46	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S46	1187376	ATTTTTTTT	ATTTTTTTTT	indel	intergenic		HP0947	
S46	1262251	C	A	SNP	nonsynonymous	jhp0981		
S46	1295885	GATATATATATATATATAT	GATATATATATATATATATAT	indel	intergenic		HPATCC43504_01238	
S46	1319852	A	AAG	indel	intergenic		HELPY_1075	jhp1032
S46	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S46	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S46	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S46	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S46	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S46	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S46	1482848	G	T	SNP	nonsynonymous	HPG27_298		
S46	1482850	T	G	SNP	nonsynonymous	HPG27_298		
S46	1482859	T	C	SNP	synonymous	HPG27_298		
S46	1482860	C	A	SNP	nonsynonymous	HPG27_298		
S46	1482861	G	A	SNP	nonsynonymous	HPG27_298		
S46	1483092	G	A	SNP	nonsynonymous	HPG27_298		
S46	1483202	G	A	SNP	nonsynonymous	HPG27_298		
S46	1483269	A	G	SNP	nonsynonymous	HPG27_298		
S47	99327	C	T	SNP	synonymous	jhp0935		
S47	168379	C	T	SNP	synonymous	HP1547		
S47	199597	C	T	SNP	intergenic		HP1582	jhp1488
S47	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S47	232964	C	A	SNP	intergenic		HPPC_00115	
S47	251374	G	A	SNP	nonsynonymous	HP0045		
S47	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S47	303813	A	T	SNP	nonsynonymous	HPP12_0096		
S47	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S47	424517	CGGGG	CGGGGG	indel	intergenic		HP0189	HP0188
S47	430922	G	A	SNP	nonsynonymous	HP0194		
S47	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S47	548941	C	T	SNP	intergenic		jhp1163	HPG27_298
S47	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S47	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S47	548994	A	T	SNP	intergenic		jhp1163	HPG27_298
S47	691071	G	A	SNP	synonymous	HP0489		
S47	691074	A	C	SNP	synonymous	HP0489		
S47	691077	G	A	SNP	nonsynonymous	HP0489		
S47	691089	A	G	SNP	synonymous	HP0489		
S47	691091	C	A	SNP	nonsynonymous	HP0489		
S47	691101	T	C	SNP	synonymous	HP0489		
S47	691113	G	T	SNP	nonsynonymous	HP0489		
S47	691173	C	A	SNP	synonymous	HP0489		
S47	691175	A	G	SNP	nonsynonymous	HP0489		
S47	691179	C	G	SNP	synonymous	HP0489		
S47	691185	A	G	SNP	nonsynonymous	HP0489		
S47	691189	C	G	SNP	nonsynonymous	HP0489		
S47	691190	A	C	SNP	nonsynonymous	HP0489		
S47	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S47	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S47	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTTT,GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S47	1187376	ATTTTTTTT	ATTTTTTTTT	indel	intergenic		HP0947	
S47	1262251	C	A	SNP	nonsynonymous	jhp0981		
S47	1295885	GATATATATATATATATAT	GATATATATATATATATATAT	indel	intergenic		HPATCC43504_01238	
S47	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S47	1332888	T	G	SNP	synonymous	HPATCC43504_01275		
S47	1332893	T	G	SNP	nonsynonymous	HPATCC43504_01275		
S47	1332900	G	GT	indel	frameshift deletion	HPATCC43504_01275		
S47	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S47	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S47	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S47	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S47	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S47	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S47	1582071	AATT	AATTATT	indel	intergenic		HELPY_1317	HP0228
S48	99327	C	T	SNP	synonymous	jhp0935		
S48	168379	C	T	SNP	synonymous	HP1547		
S48	199597	C	T	SNP	intergenic		HP1582	jhp1488
S48	251374	G	A	SNP	nonsynonymous	HP0045		
S48	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S48	424517	CGGGG	CGGGGG	indel	intergenic		HP0189	HP0188
S48	430922	G	A	SNP	nonsynonymous	HP0194		
S48	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S48	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S48	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S48	691071	G	A	SNP	synonymous	HP0489		
S48	691074	A	C	SNP	synonymous	HP0489		
S48	691077	G	A	SNP	nonsynonymous	HP0489		
S48	691089	A	G	SNP	synonymous	HP0489		

S48	691091	C	A	SNP	nonsynonymous	HP0489		
S48	691101	T	C	SNP	synonymous	HP0489		
S48	691113	G	T	SNP	nonsynonymous	HP0489		
S48	691149	G	C	SNP	synonymous	HP0489		
S48	691173	C	A	SNP	synonymous	HP0489		
S48	691175	A	G	SNP	nonsynonymous	HP0489		
S48	691179	C	G	SNP	synonymous	HP0489		
S48	691185	A	G	SNP	nonsynonymous	HP0489		
S48	691189	C	G	SNP	nonsynonymous	HP0489		
S48	691190	A	C	SNP	nonsynonymous	HP0489		
S48	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S48	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S48	1102177	A	T	SNP	intergenic		HP0876	HP0875
S48	1187376	ATTTTTTTT	ATTTTTTTTT	indel	intergenic		HP0947	
S48	1262251	C	A	SNP	nonsynonymous	jhp0981		
S48	1295885	GATATATATATATATATAT	GATATATATATATATATATAT,GATATATATATATATATAT	indel	intergenic		HPATCC43504_01238	
S48	1318527	C	A	SNP	intergenic	jhp1031		
S48	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S48	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S48	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S48	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S48	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S48	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S48	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S48	1482848	G	T	SNP	nonsynonymous	HPG27_298		
S48	1482850	T	G	SNP	nonsynonymous	HPG27_298		
S48	1482859	T	C	SNP	synonymous	HPG27_298		
S48	1482860	C	A	SNP	nonsynonymous	HPG27_298		
S48	1482861	G	A	SNP	nonsynonymous	HPG27_298		
S48	1483092	G	A	SNP	nonsynonymous	HPG27_298		
S48	1483202	G	A	SNP	nonsynonymous	HPG27_298		
S48	1483269	A	G	SNP	nonsynonymous	HPG27_298		
S48	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S48	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
S49	212868	C	A	SNP	nonsynonymous	HPG27_8		
S49	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S49	232964	C	A	SNP	intergenic		HPPC_00115	
S49	332580	T	A	SNP	synonymous	HP0119		
S49	332598	C	T	SNP	synonymous	HP0119		
S49	332632	G	A	SNP	nonsynonymous	HP0119		
S49	332653	C	T	SNP	nonsynonymous	HP0119		
S49	332689	G	A	SNP	nonsynonymous	HP0119		
S49	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S49	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S49	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		
S49	868876	T	C	SNP	nonsynonymous	HP0646		
S49	956664	A	G	SNP	synonymous	HPG27_680		
S49	1166858	G	T	SNP	nonsynonymous	HP0929		
S49	1185804	A	G	SNP	nonsynonymous	HP0946		
S49	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S49	1636693	TGGGG	TGGGGG	indel	frameshift insertion	HPATCC43504_01561		
S50	212868	C	A	SNP	nonsynonymous	HPG27_8		
S50	332580	T	A	SNP	synonymous	HP0119		
S50	332632	G	A	SNP	nonsynonymous	HP0119		
S50	332653	C	T	SNP	nonsynonymous	HP0119		
S50	332689	G	A	SNP	nonsynonymous	HP0119		
S50	464294	A	G	SNP	intergenic		HP0226	
S50	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S50	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S50	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		
S50	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S50	868876	T	C	SNP	nonsynonymous	HP0646		
S50	873717	C	T	SNP	synonymous	HP0651		
S50	873807	C	G	SNP	synonymous	HP0651		
S50	873808	G	A	SNP	synonymous	HP0651		
S50	873809	A	C	SNP	synonymous	HP0651		
S50	873813	G	A	SNP	synonymous	HP0651		
S50	903745	TAA	TA	indel	frameshift deletion	HPP12_0689		
S50	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S50	1166858	G	T	SNP	nonsynonymous	HP0929		
S50	1187376	ATTTTTTTT	ATTTTTTTTT	indel	intergenic		HP0947	
S50	1284512	C	T	SNP	synonymous	HPG27_1018		
S50	1284514	C	T	SNP	nonsynonymous	HPG27_1018		
S50	1284517	T	G	SNP	synonymous	HPG27_1018		
S50	1284553	C	A	SNP	nonsynonymous	HPG27_1018		
S50	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S50	1636693	TGGGG	TGGGGG	indel	frameshift insertion	HPATCC43504_01561		
S51	212868	C	A	SNP	nonsynonymous	HPG27_8		
S51	332580	T	A	SNP	synonymous	HP0119		
S51	332598	C	T	SNP	synonymous	HP0119		
S51	332632	G	A	SNP	nonsynonymous	HP0119		
S51	332653	C	T	SNP	nonsynonymous	HP0119		
S51	332689	G	A	SNP	nonsynonymous	HP0119		
S51	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S51	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S51	548978	G	T	SNP	intergenic		jhp1163	HPG27_298

S51	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		
S51	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S51	839999	T	C	SNP	nonsynonymous	HPB8_818		
S51	868876	T	C	SNP	nonsynonymous	HP0646		
S51	958719	G	A	SNP	intergenic		HPG27_680	
S51	1043704	GTTTTTTTTTTTTT	GTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S51	1166858	G	T	SNP	nonsynonymous	HP0929		
S51	1636693	TGGGG	TGGGGG	indel	frameshift insertion	HPATCC43504_01561		
S52	212868	C	A	SNP	nonsynonymous	HPG27_8		
S52	232963	A	C	SNP	intergenic		HPPC_00115	
S52	232964	C	A	SNP	intergenic		HPPC_00115	
S52	332580	T	A	SNP	synonymous	HP0119		
S52	332598	C	T	SNP	synonymous	HP0119		
S52	332632	G	A	SNP	nonsynonymous	HP0119		
S52	332653	C	T	SNP	nonsynonymous	HP0119		
S52	332689	G	A	SNP	nonsynonymous	HP0119		
S52	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S52	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S52	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S52	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		
S52	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S52	839999	T	C	SNP	nonsynonymous	HPB8_818		
S52	868876	T	C	SNP	nonsynonymous	HP0646		
S52	1043704	GTTTTTTTTTTTTT	GTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S52	1166858	G	T	SNP	nonsynonymous	HP0929		
S52	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S52	1636693	TGGGG	TGGGGG	indel	frameshift insertion	HPATCC43504_01561		
S52	1663529	GTTTTT	GTTTT	indel	1 bp deletion			
S53	658	C	T	SNP	synonymous	HP1529		
S53	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S53	232964	C	A	SNP	intergenic		HPPC_00115	
S53	303813	A	T	SNP	nonsynonymous	HPP12_0096		
S53	303816	A	T	SNP	nonsynonymous	HPP12_0096		
S53	303827	CGGGGGGGGGGGG	G	indel	frameshift deletion	HPP12_0096		
S53	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S53	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S53	431032	C	T	SNP	synonymous	HP0194		
S53	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S53	443932	ACTCTCTCTCTCTCTCTC	ACTCTCTCTCTCTCTC	indel	frameshift deletion	HPPC_01040		
S53	450628	G	A	SNP	synonymous	HP0213		
S53	546115	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTT	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTTATTTAATCTT	indel	intergenic		HP1243/HPG27_298	
S53	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S53	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S53	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S53	626404	C	T	SNP	synonymous	HP1175		
S53	627243	G	A	SNP	nonsynonymous	HP1174		
S53	701041	G	A	SNP	nonsynonymous	HP0499		
S53	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S53	839999	T	C	SNP	nonsynonymous	HPB8_818		
S53	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S53	1319683	G	C	SNP	nonsynonymous	jhp1032		
S53	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S53	1592801	T	G	SNP	synonymous	HP1354		
S53	1663529	GTTTTT	GTTTT	indel	1 bp deletion			
S53	1669898	C	T	SNP	synonymous	HP1450		
S54	658	C	T	SNP	synonymous	HP1529		
S54	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S54	431032	C	T	SNP	synonymous	HP0194		
S54	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S54	450628	G	A	SNP	synonymous	HP0213		
S54	546115	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTT	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTTATTTAATCTT	indel	intergenic		HP1243/HPG27_298	
S54	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S54	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S54	626404	C	T	SNP	synonymous	HP1175		
S54	627243	G	A	SNP	nonsynonymous	HP1174		
S54	701041	G	A	SNP	nonsynonymous	HP0499		
S54	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S54	1319683	G	C	SNP	nonsynonymous	jhp1032		
S54	1592785	TGTGG	TG	indel	frameshift deletion	HP1354		
S54	1669898	C	T	SNP	synonymous	HP1450		
S55	658	C	T	SNP	synonymous	HP1529		
S55	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S55	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S55	431032	C	T	SNP	synonymous	HP0194		
S55	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S55	450628	G	A	SNP	synonymous	HP0213		
S55	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
S55	546115	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTT	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTTATTTAATCTT	indel	intergenic		HP1243/HPG27_298	
S55	548957	C	T	SNP	intergenic		jhp1163	HPG27_298
S55	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S55	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S55	626404	C	T	SNP	synonymous	HP1175		
S55	627243	G	A	SNP	nonsynonymous	HP1174		
S55	701041	G	A	SNP	nonsynonymous	HP0499		
S55	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	

S55	839999	T	C	SNP	nonsynonymous	HPB8_818		
S55	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S55	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S55	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S55	1192576	G	T	SNP	intergenic		HP0953	
S55	1319683	G	C	SNP	nonsynonymous	jhp1032		
S55	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S55	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
S55	1669898	C	T	SNP	synonymous	HP1450		
S56	658	C	T	SNP	synonymous	HP1529		
S56	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S56	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S56	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S56	431032	C	T	SNP	synonymous	HP0194		
S56	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S56	443932	ACTCTCTCTCTCTCTCTC	ACTCTCTCTCTCTCTCTC	indel	frameshift deletion	HPPC_01040		
S56	450628	G	A	SNP	synonymous	HP0213		
S56	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
S56	546115	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTT	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTTATTTAATCTT	indel	intergenic		HP1243/HPG27_298	
S56	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S56	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S56	548986	CTTTTTTT	CTTTTTTTT	indel	intergenic		jhp1163	HPG27_298
S56	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S56	626404	C	T	SNP	synonymous	HP1175		
S56	627243	G	A	SNP	nonsynonymous	HP1174		
S56	701041	G	A	SNP	nonsynonymous	HP0499		
S56	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S56	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S56	1124346	CGAGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_1187	
S56	1302389	G	A	SNP	synonymous	HP0364		
S56	1319683	G	C	SNP	nonsynonymous	jhp1032		
S56	1410493	G	A	SNP	nonsynonymous	jhp0373		
S56	1592785	TGTGG	TG	indel	frameshift deletion	HP1354		
S56	1669898	C	T	SNP	synonymous	HP1450		
S57	658	C	T	SNP	synonymous	HP1529		
S57	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S57	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S57	431032	C	T	SNP	synonymous	HP0194		
S57	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S57	450628	G	A	SNP	synonymous	HP0213		
S57	546524	C	A	SNP	intergenic		HP1243/HPG27_298	
S57	607390	G	A	SNP	synonymous	jhp1115		
S57	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S57	626404	C	T	SNP	synonymous	HP1175		
S57	627243	G	A	SNP	nonsynonymous	HP1174		
S57	701041	G	A	SNP	nonsynonymous	HP0499		
S57	748416	T	A	SNP	intergenic		HP0514	
S57	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S57	958575	GAGA	GAGAAAGA	indel	intergenic		HPG27_680	
S57	1124346	CGAGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_1187	
S57	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S57	1272391	G	A	SNP	nonsynonymous			
S57	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S57	1555634	T	C	SNP	nonsynonymous	HELPY_0259		
S57	1582072	ATT	ATTTTT	indel	intergenic		HELPY_1317	HP0228
S57	1592786	GT	G	indel	frameshift deletion	HP1354		
S57	1592801	T	G	SNP	synonymous	HP1354		
S57	1593602	TGGGGGGGGGGG	G	indel	frameshift insertion	HP1354		
S57	1669898	C	T	SNP	synonymous	HP1450		
S58	658	C	T	SNP	synonymous	HP1529		
S58	961	G	A	SNP	nonsynonymous	HP1529		
S58	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S58	232963	A	C	SNP	intergenic		HPPC_00115	
S58	232964	C	A	SNP	intergenic		HPPC_00115	
S58	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S58	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S58	431032	C	T	SNP	synonymous	HP0194		
S58	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S58	450628	G	A	SNP	synonymous	HP0213		
S58	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
S58	546115	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTT	AATTTAATCTTATTTAATCTTATTTAATCTTATTTAA TCTTATTTAATCTTATTTAATCTTATTTAATCTTATT TAATCTT	indel	intergenic		HP1243/HPG27_298	
S58	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S58	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S58	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S58	626404	C	T	SNP	synonymous	HP1175		
S58	627243	G	A	SNP	nonsynonymous	HP1174		
S58	701041	G	A	SNP	nonsynonymous	HP0499		
S58	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S58	805851	A	T	SNP	intergenic		HPG27_556	
S58	958574	CGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_680	
S58	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S58	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S58	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S58	1284699	AGGGGGGGGGG	AGGGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S58	1555634	T	C	SNP	nonsynonymous	HELPY_0259		

S58	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S58	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift deletion	HP1354		
S58	1669898	C	T	SNP	synonymous	HP1450		
S59	658	C	T	SNP	synonymous	HP1529		
S59	27485	AGGGGGG	AGGGGGGG	indel	intergenic		HP1506	jhp1400
S59	431032	C	T	SNP	synonymous	HP0194		
S59	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S59	450628	G	A	SNP	synonymous	HP0213		
S59	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S59	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S59	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S59	626404	C	T	SNP	synonymous	HP1175		
S59	627243	G	A	SNP	nonsynonymous	HP1174		
S59	701041	G	A	SNP	nonsynonymous	HP0499		
S59	757302	G	A	SNP	synonymous	HPATCC43504_00741		
S59	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S59	854528	C	T	SNP	nonsynonymous	HP0632		
S59	958574	CGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_680	
S59	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S59	1124346	CGAGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_1187	
S59	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S59	1284699	AGGGGGGGGGG	AGGGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S59	1555634	T	C	SNP	nonsynonymous	HELPY_0259		
S59	1582224	ACTCTCTCTCTCTCTCTC	ACTCTCTCTCTCTCTC	indel	intergenic		HELPY_1317	HP0228
S59	1592800	G	GGGT	indel	frameshift insertion	HP1354		
S59	1669898	C	T	SNP	synonymous	HP1450		
S60	658	C	T	SNP	synonymous	HP1529		
S60	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S60	232963	A	C	SNP	intergenic		HPPC_00115	
S60	232964	C	A	SNP	intergenic		HPPC_00115	
S60	315327	ACCCCCCCCCCC	ACCCCCCCCCCCC	indel	intergenic		HP0103	
S60	345517	C	T	SNP	synonymous	HP0132		
S60	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S60	431032	C	T	SNP	synonymous	HP0194		
S60	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S60	450628	G	A	SNP	synonymous	HP0213		
S60	488821	C	T	SNP	nonsynonymous	jhp1233		
S60	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S60	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S60	626404	C	T	SNP	synonymous	HP1175		
S60	627243	G	A	SNP	nonsynonymous	HP1174		
S60	701041	G	A	SNP	nonsynonymous	HP0499		
S60	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S60	958574	CGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_680	
S60	958702	CAAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S60	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S60	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S60	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S60	1318527	C	A	SNP	intergenic	jhp1031		
S60	1555634	T	C	SNP	nonsynonymous	HELPY_0259		
S60	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S60	1592801	T	G	SNP	synonymous	HP1354		
S60	1669898	C	T	SNP	synonymous	HP1450		
S61	658	C	T	SNP	synonymous	HP1529		
S61	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S61	249555	G	A	SNP	nonsynonymous	HP0044		
S61	315327	ACCCCCCCCCCC	ACCCCCCCCCCCC	indel	intergenic		HP0103	
S61	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S61	413866	A	G	SNP	synonymous	HP1535		
S61	413887	A	G	SNP	synonymous	HP1535		
S61	413983	C	T	SNP	synonymous	HP1535		
S61	431032	C	T	SNP	synonymous	HP0194		
S61	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S61	450628	G	A	SNP	synonymous	HP0213		
S61	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S61	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S61	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S61	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S61	626404	C	T	SNP	synonymous	HP1175		
S61	627243	G	A	SNP	nonsynonymous	HP1174		
S61	701041	G	A	SNP	nonsynonymous	HP0499		
S61	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S61	839999	T	C	SNP	nonsynonymous	HPB8_818		
S61	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S61	1122223	C	T	SNP	intergenic		HP0895	
S61	1122228	TAAAAA	TAAAAAA	indel	intergenic		HP0895	
S61	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S61	1483092	G	A	SNP	nonsynonymous	HPG27_298		
S61	1483202	G	A	SNP	nonsynonymous	HPG27_298		
S61	1483269	A	G	SNP	nonsynonymous	HPG27_298		
S61	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S61	1669898	C	T	SNP	synonymous	HP1450		
S62	658	C	T	SNP	synonymous	HP1529		
S62	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S62	232964	C	A	SNP	intergenic		HPPC_00115	
S62	249555	G	A	SNP	nonsynonymous	HP0044		
S62	315327	ACCCCCCCCCCC	ACCCCCCCCCCCC	indel	intergenic		HP0103	

S62	413866	A	G	SNP	synonymous	HP1535		
S62	413887	A	G	SNP	synonymous	HP1535		
S62	413983	C	T	SNP	synonymous	HP1535		
S62	431032	C	T	SNP	synonymous	HP0194		
S62	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S62	450628	G	A	SNP	synonymous	HP0213		
S62	549238	AGATTTCTTTTTTAAAGGGATTTCTTTTTTAAAGGGATTTCTTTTTTAAAGG

S64	1482879	A	G	SNP	nonsynonymous	HPG27_298		
S64	1483092	G	A	SNP	nonsynonymous	HPG27_298		
S64	1483269	A	G	SNP	nonsynonymous	HPG27_298		
S64	1484872	T	C	SNP	synonymous	HPG27_298		
S64	1484881	G	A	SNP	synonymous	HPG27_298		
S64	1484897	C	A	SNP	nonsynonymous	HPG27_298		
S64	1484899	C	A	SNP	synonymous	HPG27_298		
S64	1669898	C	T	SNP	synonymous	HP1450		
S65	232964	C	A	SNP	intergenic		HPPC_00115	
S65	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S65	332580	T	A	SNP	synonymous	HP0119		
S65	332598	C	T	SNP	synonymous	HP0119		
S65	332632	G	A	SNP	nonsynonymous	HP0119		
S65	332653	C	T	SNP	nonsynonymous	HP0119		
S65	332689	G	A	SNP	nonsynonymous	HP0119		
S65	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S65	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		
S65	734196	G	A	SNP	synonymous	HPP12_0534		
S65	734198	A	G	SNP	nonsynonymous	HPP12_0534		
S65	734200	C	A	SNP	nonsynonymous	HPP12_0534		
S65	734220	C	T	SNP	synonymous	HPP12_0534		
S65	734234	A	G	SNP	nonsynonymous	HPP12_0534		
S65	734241	T	A	SNP	synonymous	HPP12_0534		
S65	734256	C	T	SNP	synonymous	HPP12_0534		
S65	734260	G	A	SNP	nonsynonymous	HPP12_0534		
S65	734271	G	A	SNP	synonymous	HPP12_0534		
S65	734277	C	T	SNP	synonymous	HPP12_0534		
S65	734283	C	T	SNP	synonymous	HPP12_0534		
S65	734286	T	C	SNP	synonymous	HPP12_0534		
S65	734309	A	C	SNP	nonsynonymous	HPP12_0534		
S65	734313	G	A	SNP	synonymous	HPP12_0534		
S65	734317	A	C	SNP	nonsynonymous	HPP12_0534		
S65	735228	A	T	SNP	synonymous	HPP12_0534		
S65	735324	C	G	SNP	synonymous	HPP12_0534		
S65	735333	A	G	SNP	synonymous	HPP12_0534		
S65	735336	G	A	SNP	synonymous	HPP12_0534		
S65	735343	C	T	SNP	synonymous	HPP12_0534		
S65	735356	T	C	SNP	nonsynonymous	HPP12_0534		
S65	735363	G	A	SNP	synonymous	HPP12_0534		
S65	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S65	805851	A	T	SNP	intergenic		HPG27_556	
S65	839999	T	C	SNP	nonsynonymous	HPB8_818		
S65	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S65	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S65	1122208	AG	A	SNP	intergenic		HP0895	
S65	1122232	A	AG	indel	intergenic		HP0895	
S65	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S65	1198606	A	C	SNP	nonsynonymous	HP0961		
S65	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S65	1306585	C	T	SNP	nonsynonymous	HP0360		
S65	1318527	C	A	SNP	intergenic	jhp1031		
S65	1483720	G	A	SNP	synonymous	HPG27_298		
S65	1484872	T	C	SNP	synonymous	HPG27_298		
S65	1484881	G	A	SNP	synonymous	HPG27_298		
S65	1484897	C	A	SNP	nonsynonymous	HPG27_298		
S65	1484899	C	A	SNP	synonymous	HPG27_298		
S65	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	frameshift insertion			
S65	1663529	GTTTTT	GTTTT	indel	1 bp deletion			
S66	143981	TGA	T	indel	intergenic		HP1400	
S66	193201	G	A	SNP	synonymous	jhp1480		
S66	332580	T	A	SNP	synonymous	HP0119		
S66	332598	C	T	SNP	synonymous	HP0119		
S66	332632	G	A	SNP	nonsynonymous	HP0119		
S66	332653	C	T	SNP	nonsynonymous	HP0119		
S66	332689	G	A	SNP	nonsynonymous	HP0119		
S66	374767	GGT	GGGG,GGGTG	indel	frameshift insertion	jhp0151		
S66	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
S66	546511	CAAAAAAAAAAAA	A	indel	intergenic		HP1243/HPG27_298	
S66	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		
S66	734196	G	A	SNP	synonymous	HPP12_0534		
S66	734198	A	G	SNP	nonsynonymous	HPP12_0534		
S66	734200	C	A	SNP	nonsynonymous	HPP12_0534		
S66	734220	C	T	SNP	synonymous	HPP12_0534		
S66	734234	A	G	SNP	nonsynonymous	HPP12_0534		
S66	734241	T	A	SNP	synonymous	HPP12_0534		
S66	734256	C	T	SNP	synonymous	HPP12_0534		
S66	734260	G	A	SNP	nonsynonymous	HPP12_0534		
S66	734271	G	A	SNP	synonymous	HPP12_0534		
S66	734277	C	T	SNP	synonymous	HPP12_0534		
S66	734283	C	T	SNP	synonymous	HPP12_0534		
S66	734286	T	C	SNP	synonymous	HPP12_0534		
S66	734309	A	C	SNP	nonsynonymous	HPP12_0534		
S66	734317	A	C	SNP	nonsynonymous	HPP12_0534		
S66	735228	A	T	SNP	synonymous	HPP12_0534		
S66	735324	C	G	SNP	synonymous	HPP12_0534		
S66	735333	A	G	SNP	synonymous	HPP12_0534		
S66	735336	G	A	SNP	synonymous	HPP12_0534		

S66	735343	C	T	SNP	synonymous	HPP12_0534		
S66	735356	T	C	SNP	nonsynonymous	HPP12_0534		
S66	735363	G	A	SNP	synonymous	HPP12_0534		
S66	805851	A	T	SNP	intergenic		HPG27_556	
S66	1122208	AG	A	SNP	intergenic		HP0895	
S66	1122235	G	A	SNP	intergenic		HP0895	
S66	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S66	1198606	A	C	SNP	nonsynonymous	HP0961		
S66	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S66	1306585	C	T	SNP	nonsynonymous	HP0360		
S66	1319853	C	A	SNP	intergenic		HELPY_1075	jhp1032
S66	1483720	G	A	SNP	synonymous	HPG27_298		
S66	1484872	T	C	SNP	synonymous	HPG27_298		
S66	1484881	G	A	SNP	synonymous	HPG27_298		
S66	1484897	C	A	SNP	nonsynonymous	HPG27_298		
S66	1484899	C	A	SNP	synonymous	HPG27_298		
S66	1592801	T	G	SNP	synonymous	HP1354		
S66	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S67	92562	T	C	SNP	intergenic		HP0993	
S67	232964	C	A	SNP	intergenic		HPPC_00115	
S67	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S67	332580	T	A	SNP	synonymous	HP0119		
S67	332598	C	T	SNP	synonymous	HP0119		
S67	332632	G	A	SNP	nonsynonymous	HP0119		
S67	332653	C	T	SNP	nonsynonymous	HP0119		
S67	332689	G	A	SNP	nonsynonymous	HP0119		
S67	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
S67	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S67	548865	C	T	SNP	intergenic		jhp1163	HPG27_298
S67	548868	T	C	SNP	intergenic		jhp1163	HPG27_298
S67	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S67	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S67	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		
S67	734196	G	A	SNP	synonymous	HPP12_0534		
S67	734198	A	G	SNP	nonsynonymous	HPP12_0534		
S67	734200	C	A	SNP	nonsynonymous	HPP12_0534		
S67	734220	C	T	SNP	synonymous	HPP12_0534		
S67	734234	A	G	SNP	nonsynonymous	HPP12_0534		
S67	734241	T	A	SNP	synonymous	HPP12_0534		
S67	734256	C	T	SNP	synonymous	HPP12_0534		
S67	734260	G	A	SNP	nonsynonymous	HPP12_0534		
S67	734271	G	A	SNP	synonymous	HPP12_0534		
S67	734277	C	T	SNP	synonymous	HPP12_0534		
S67	734283	C	T	SNP	synonymous	HPP12_0534		
S67	734286	T	C	SNP	synonymous	HPP12_0534		
S67	734309	A	C	SNP	nonsynonymous	HPP12_0534		
S67	734317	A	C	SNP	nonsynonymous	HPP12_0534		
S67	734661	C	G	SNP	synonymous	HPP12_0534		
S67	734733	C	T	SNP	synonymous	HPP12_0534		
S67	735228	A	T	SNP	synonymous	HPP12_0534		
S67	735239	C	A	SNP	nonsynonymous	HPP12_0534		
S67	735324	C	G	SNP	synonymous	HPP12_0534		
S67	735333	A	G	SNP	synonymous	HPP12_0534		
S67	735336	G	A	SNP	synonymous	HPP12_0534		
S67	735343	C	T	SNP	synonymous	HPP12_0534		
S67	735356	T	C	SNP	nonsynonymous	HPP12_0534		
S67	735363	G	A	SNP	synonymous	HPP12_0534		
S67	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S67	839999	T	C	SNP	nonsynonymous	HPB8_818		
S67	952086	TTCTCTCTCTCTCTCTC	TTCTCTCTCTCTCTC	indel	intergenic		HPG27_677	
S67	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S67	1122208	AG	A	SNP	intergenic		HP0895	
S67	1122232	A	AG	indel	intergenic		HP0895	
S67	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S67	1198606	A	C	SNP	nonsynonymous	HP0961		
S67	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S67	1306585	C	T	SNP	nonsynonymous	HP0360		
S67	1483720	G	A	SNP	synonymous	HPG27_298		
S67	1484872	T	C	SNP	synonymous	HPG27_298		
S67	1484881	G	A	SNP	synonymous	HPG27_298		
S67	1484897	C	A	SNP	nonsynonymous	HPG27_298		
S67	1484899	C	A	SNP	synonymous	HPG27_298		
S67	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S68	232964	C	A	SNP	intergenic		HPPC_00115	
S68	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S68	332580	T	A	SNP	synonymous	HP0119		
S68	332598	C	T	SNP	synonymous	HP0119		
S68	332632	G	A	SNP	nonsynonymous	HP0119		
S68	332653	C	T	SNP	nonsynonymous	HP0119		
S68	332689	G	A	SNP	nonsynonymous	HP0119		
S68	374767	GGT	GGGCGT	indel	frameshift insertion	jhp0151		
S68	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S68	548937	C	T	SNP	intergenic		jhp1163	HPG27_298
S68	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S68	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S68	548994	A	T	SNP	intergenic		jhp1163	HPG27_298
S68	701037	TGGGGGGGG	TGGGGGGG	indel	frameshift deletion	HP0499		

S68	734196	G	A	SNP	synonymous	HPP12_0534		
S68	734198	A	G	SNP	nonsynonymous	HPP12_0534		
S68	734200	C	A	SNP	nonsynonymous	HPP12_0534		
S68	734220	C	T	SNP	synonymous	HPP12_0534		
S68	734234	A	G	SNP	nonsynonymous	HPP12_0534		
S68	734241	T	A	SNP	synonymous	HPP12_0534		
S68	734256	C	T	SNP	synonymous	HPP12_0534		
S68	734260	G	A	SNP	nonsynonymous	HPP12_0534		
S68	734271	G	A	SNP	synonymous	HPP12_0534		
S68	734277	C	T	SNP	synonymous	HPP12_0534		
S68	734283	C	T	SNP	synonymous	HPP12_0534		
S68	734286	T	C	SNP	synonymous	HPP12_0534		
S68	734309	A	C	SNP	nonsynonymous	HPP12_0534		
S68	734313	G	A	SNP	synonymous	HPP12_0534		
S68	734317	A	C	SNP	nonsynonymous	HPP12_0534		
S68	734661	C	G	SNP	synonymous	HPP12_0534		
S68	734733	C	T	SNP	synonymous	HPP12_0534		
S68	735228	A	T	SNP	synonymous	HPP12_0534		
S68	735324	C	G	SNP	synonymous	HPP12_0534		
S68	735333	A	G	SNP	synonymous	HPP12_0534		
S68	735336	G	A	SNP	synonymous	HPP12_0534		
S68	735343	C	T	SNP	synonymous	HPP12_0534		
S68	735356	T	C	SNP	nonsynonymous	HPP12_0534		
S68	735363	G	A	SNP	synonymous	HPP12_0534		
S68	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S68	839999	T	C	SNP	nonsynonymous	HPB8_818		
S68	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S68	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S68	1122211	G	A	SNP	intergenic		HP0895	
S68	1122235	G	A	SNP	intergenic		HP0895	
S68	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S68	1198606	A	C	SNP	nonsynonymous	HP0961		
S68	1284699	AGGGGGGGGGG	AGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S68	1306585	C	T	SNP	nonsynonymous	HP0360		
S68	1483720	G	A	SNP	synonymous	HPG27_298		
S68	1484872	T	C	SNP	synonymous	HPG27_298		
S68	1484881	G	A	SNP	synonymous	HPG27_298		
S68	1484897	C	A	SNP	nonsynonymous	HPG27_298		
S68	1484899	C	A	SNP	synonymous	HPG27_298		
S68	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S69	99327	C	T	SNP	synonymous	jhp0935		
S69	155189	GA	G	indel	frameshift deletion	jhp1438		
S69	168379	C	T	SNP	synonymous	HP1547		
S69	199597	C	T	SNP	intergenic		HP1582	jhp1488
S69	207029	T	C	SNP	nonsynonymous	HPATCC43504_00197		
S69	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S69	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S69	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S69	430922	G	A	SNP	nonsynonymous	HP0194		
S69	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S69	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S69	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S69	548986	CTTTTTTT	CTTTTTTTT	indel	intergenic		jhp1163	HPG27_298
S69	680519	AC	ACC	indel	intergenic		HPP12_1088	
S69	691071	G	A	SNP	synonymous	HP0489		
S69	691074	A	C	SNP	synonymous	HP0489		
S69	691077	G	A	SNP	nonsynonymous	HP0489		
S69	691089	A	G	SNP	synonymous	HP0489		
S69	691091	C	A	SNP	nonsynonymous	HP0489		
S69	691101	T	C	SNP	synonymous	HP0489		
S69	691113	G	T	SNP	nonsynonymous	HP0489		
S69	691157	C	A	SNP	nonsynonymous	HP0489		
S69	691173	C	A	SNP	synonymous	HP0489		
S69	691175	A	G	SNP	nonsynonymous	HP0489		
S69	691179	C	G	SNP	synonymous	HP0489		
S69	691185	A	G	SNP	nonsynonymous	HP0489		
S69	691189	C	G	SNP	nonsynonymous	HP0489		
S69	691190	A	C	SNP	nonsynonymous	HP0489		
S69	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S69	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S69	839999	T	C	SNP	nonsynonymous	HPB8_818		
S69	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S69	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S69	1192157	T	C	SNP	nonsynonymous	HP0953		
S69	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S69	1332888	T	G	SNP	synonymous	HPATCC43504_01275		
S69	1332893	T	G	SNP	nonsynonymous	HPATCC43504_01275		
S69	1332897	A	G	SNP	synonymous	HPATCC43504_01275		
S69	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S69	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S69	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S69	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S69	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S69	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S69	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S69	1482848	G	T	SNP	nonsynonymous	HPG27_298		
S69	1482850	T	G	SNP	nonsynonymous	HPG27_298		

S69	1483092	G	A	SNP	nonsynonymous	HPG27_298		
S69	1483202	G	A	SNP	nonsynonymous	HPG27_298		
S69	1483269	A	G	SNP	nonsynonymous	HPG27_298		
S69	1592786	GT	G	indel	frameshift deletion	HP1354		
S69	1663529	GTTTTT	GTTTT	indel	1 bp deletion			
S70	99327	C	T	SNP	synonymous	jhp0935		
S70	168379	C	T	SNP	synonymous	HP1547		
S70	199597	C	T	SNP	intergenic		HP1582	jhp1488
S70	232964	C	A	SNP	intergenic		HPPC_00115	
S70	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S70	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S70	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S70	430922	G	A	SNP	nonsynonymous	HP0194		
S70	464722	T	G	SNP	synonymous	HP0227		
S70	464740	C	A	SNP	synonymous	HP0227		
S70	464743	A	G	SNP	synonymous	HP0227		
S70	464764	A	G	SNP	synonymous	HP0227		
S70	464765	T	C	SNP	nonsynonymous	HP0227		
S70	464767	G	A	SNP	synonymous	HP0227		
S70	464773	G	A	SNP	synonymous	HP0227		
S70	464779	C	T	SNP	synonymous	HP0227		
S70	464788	G	A	SNP	synonymous	HP0227		
S70	464797	A	G	SNP	synonymous	HP0227		
S70	464800	G	A	SNP	synonymous	HP0227		
S70	464803	G	A	SNP	synonymous	HP0227		
S70	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S70	548957	C	T	SNP	intergenic		jhp1163	HPG27_298
S70	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S70	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S70	548986	CTTTTTTT	CTTTTTTTT	indel	intergenic		jhp1163	HPG27_298
S70	691071	G	A	SNP	synonymous	HP0489		
S70	691074	A	C	SNP	synonymous	HP0489		
S70	691077	G	A	SNP	nonsynonymous	HP0489		
S70	691089	A	G	SNP	synonymous	HP0489		
S70	691091	C	A	SNP	nonsynonymous	HP0489		
S70	691101	T	C	SNP	synonymous	HP0489		
S70	691113	G	T	SNP	nonsynonymous	HP0489		
S70	691157	C	A	SNP	nonsynonymous	HP0489		
S70	691173	C	A	SNP	synonymous	HP0489		
S70	691175	A	G	SNP	nonsynonymous	HP0489		
S70	691179	C	G	SNP	synonymous	HP0489		
S70	691185	A	G	SNP	nonsynonymous	HP0489		
S70	691189	C	G	SNP	nonsynonymous	HP0489		
S70	691190	A	C	SNP	nonsynonymous	HP0489		
S70	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S70	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S70	839999	T	C	SNP	nonsynonymous	HPB8_818		
S70	879985	C	T	SNP	nonsynonymous	HP0656		
S70	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S70	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S70	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S70	1192157	T	C	SNP	nonsynonymous	HP0953		
S70	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S70	1332870	C	T	SNP	synonymous	HPATCC43504_01275		
S70	1332888	T	G	SNP	synonymous	HPATCC43504_01275		
S70	1332893	T	G	SNP	nonsynonymous	HPATCC43504_01275		
S70	1332897	A	G	SNP	synonymous	HPATCC43504_01275		
S70	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S70	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S70	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S70	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S70	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S70	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S70	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S70	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S70	1592801	T	G	SNP	synonymous	HP1354		
S71	99327	C	T	SNP	synonymous	jhp0935		
S71	168379	C	T	SNP	synonymous	HP1547		
S71	199597	C	T	SNP	intergenic		HP1582	jhp1488
S71	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S71	315337	C	CA	indel	intergenic		HP0103	
S71	430922	G	A	SNP	nonsynonymous	HP0194		
S71	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S71	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S71	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S71	548986	CTTTTTTT	CTTTTTTTT	indel	intergenic		jhp1163	HPG27_298
S71	691071	G	A	SNP	synonymous	HP0489		
S71	691074	A	C	SNP	synonymous	HP0489		
S71	691077	G	A	SNP	nonsynonymous	HP0489		
S71	691089	A	G	SNP	synonymous	HP0489		
S71	691091	C	A	SNP	nonsynonymous	HP0489		
S71	691101	T	C	SNP	synonymous	HP0489		
S71	691113	G	T	SNP	nonsynonymous	HP0489		
S71	691173	C	A	SNP	synonymous	HP0489		
S71	691175	A	G	SNP	nonsynonymous	HP0489		
S71	691179	C	G	SNP	synonymous	HP0489		
S71	691185	A	G	SNP	nonsynonymous	HP0489		

S71	691189	C	G	SNP	nonsynonymous	HP0489		
S71	691190	A	C	SNP	nonsynonymous	HP0489		
S71	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S71	839999	T	C	SNP	nonsynonymous	HPB8_818		
S71	879985	C	T	SNP	nonsynonymous	HP0656		
S71	958702	CAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S71	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S71	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S71	1192157	T	C	SNP	nonsynonymous	HP0953		
S71	1318502	C	T	SNP	intergenic	jhp1031		
S71	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S71	1332900	G	GT	indel	frameshift deletion	HPATCC43504_01275		
S71	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S71	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S71	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S71	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S71	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S71	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S71	1592135	G	A	SNP	nonsynonymous	HP1350		
S71	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S72	99327	C	T	SNP	synonymous	jhp0935		
S72	168379	C	T	SNP	synonymous	HP1547		
S72	199597	C	T	SNP	intergenic		HP1582	jhp1488
S72	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S72	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S72	430922	G	A	SNP	nonsynonymous	HP0194		
S72	466390	G	A	SNP	intergenic		HP0227	
S72	466443	G	A	SNP	intergenic		HP0227	
S72	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S72	691071	G	A	SNP	synonymous	HP0489		
S72	691074	A	C	SNP	synonymous	HP0489		
S72	691077	G	A	SNP	nonsynonymous	HP0489		
S72	691089	A	G	SNP	synonymous	HP0489		
S72	691091	C	A	SNP	nonsynonymous	HP0489		
S72	691101	T	C	SNP	synonymous	HP0489		
S72	691113	G	T	SNP	nonsynonymous	HP0489		
S72	691157	C	A	SNP	nonsynonymous	HP0489		
S72	691173	C	A	SNP	synonymous	HP0489		
S72	691175	A	G	SNP	nonsynonymous	HP0489		
S72	691179	C	G	SNP	synonymous	HP0489		
S72	691185	A	G	SNP	nonsynonymous	HP0489		
S72	691189	C	G	SNP	nonsynonymous	HP0489		
S72	691190	A	C	SNP	nonsynonymous	HP0489		
S72	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S72	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S72	836443	C	T	SNP	synonymous	HP0617		
S72	839999	T	C	SNP	nonsynonymous	HPB8_818		
S72	879985	C	T	SNP	nonsynonymous	HP0656		
S72	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S72	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S72	1192157	T	C	SNP	nonsynonymous	HP0953		
S72	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S72	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S72	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S72	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S72	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S72	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S72	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S72	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S73	99327	C	T	SNP	synonymous	jhp0935		
S73	168379	C	T	SNP	synonymous	HP1547		
S73	199597	C	T	SNP	intergenic		HP1582	jhp1488
S73	226261	G	A	SNP	nonsynonymous	HP0019_1		
S73	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S73	430922	G	A	SNP	nonsynonymous	HP0194		
S73	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S73	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S73	548986	CTTTTTTT	CTTTTTTTT	indel	intergenic		jhp1163	HPG27_298
S73	691071	G	A	SNP	synonymous	HP0489		
S73	691074	A	C	SNP	synonymous	HP0489		
S73	691077	G	A	SNP	nonsynonymous	HP0489		
S73	691089	A	G	SNP	synonymous	HP0489		
S73	691091	C	A	SNP	nonsynonymous	HP0489		
S73	691101	T	C	SNP	synonymous	HP0489		
S73	691113	G	T	SNP	nonsynonymous	HP0489		
S73	691173	C	A	SNP	synonymous	HP0489		
S73	691175	A	G	SNP	nonsynonymous	HP0489		
S73	691179	C	G	SNP	synonymous	HP0489		
S73	691185	A	G	SNP	nonsynonymous	HP0489		
S73	691189	C	G	SNP	nonsynonymous	HP0489		
S73	691190	A	C	SNP	nonsynonymous	HP0489		
S73	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S73	839999	T	C	SNP	nonsynonymous	HPB8_818		
S73	958719	G	A	SNP	intergenic		HPG27_680	
S73	984348	G	A	SNP	nonsynonymous	HP0751		
S73	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S73	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	

S73	1284699	AGGGGGGGGGG	AGGGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S73	1332593	A	C	SNP	nonsynonymous	HPATCC43504_01275		
S73	1332595	T	A	SNP	nonsynonymous	HPATCC43504_01275		
S73	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S73	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S73	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S73	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S73	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S73	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S73	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S73	1342802	T	C	SNP	nonsynonymous	jhp0433		
S73	1461744	C	A	SNP	intergenic		HP0349	HP0350
S73	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S73	1592801	T	G	SNP	synonymous	HP1354		
S73	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S74	99327	C	T	SNP	synonymous	jhp0935		
S74	113930	T	A	SNP	intergenic		HPATCC43504_00103	
S74	168379	C	T	SNP	synonymous	HP1547		
S74	199597	C	T	SNP	intergenic		HP1582	jhp1488
S74	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S74	226261	G	A	SNP	nonsynonymous	HP0019_1		
S74	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S74	347947	G	A	SNP	nonsynonymous	jhp0122		
S74	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S74	430922	G	A	SNP	nonsynonymous	HP0194		
S74	546511	CAAAAAAAAAAAA	A	indel	intergenic		HP1243/HPG27_298	
S74	547618	A	G	SNP	nonsynonymous	HPG27_298		
S74	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S74	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S74	691071	G	A	SNP	synonymous	HP0489		
S74	691074	A	C	SNP	synonymous	HP0489		
S74	691077	G	A	SNP	nonsynonymous	HP0489		
S74	691089	A	G	SNP	synonymous	HP0489		
S74	691091	C	A	SNP	nonsynonymous	HP0489		
S74	691101	T	C	SNP	synonymous	HP0489		
S74	691113	G	T	SNP	nonsynonymous	HP0489		
S74	691157	C	A	SNP	nonsynonymous	HP0489		
S74	691173	C	A	SNP	synonymous	HP0489		
S74	691175	A	G	SNP	nonsynonymous	HP0489		
S74	691179	C	G	SNP	synonymous	HP0489		
S74	691185	A	G	SNP	nonsynonymous	HP0489		
S74	691189	C	G	SNP	nonsynonymous	HP0489		
S74	691190	A	C	SNP	nonsynonymous	HP0489		
S74	734658	C	T	SNP	synonymous	HPP12_0534		
S74	734661	C	G	SNP	synonymous	HPP12_0534		
S74	734672	G	A	SNP	nonsynonymous	HPP12_0534		
S74	734695	A	G	SNP	nonsynonymous	HPP12_0534		
S74	734697	A	G	SNP	synonymous	HPP12_0534		
S74	734703	T	C	SNP	synonymous	HPP12_0534		
S74	734841	T	A	SNP	synonymous	HPP12_0534		
S74	734842	C	A	SNP	nonsynonymous	HPP12_0534		
S74	734852	C	A	SNP	nonsynonymous	HPP12_0534		
S74	734853	T	C	SNP	synonymous	HPP12_0534		
S74	734871	G	A	SNP	synonymous	HPP12_0534		
S74	734874	T	C	SNP	synonymous	HPP12_0534		
S74	735072	G	T	SNP	nonsynonymous	HPP12_0534		
S74	735126	T	C	SNP	synonymous	HPP12_0534		
S74	735140	A	G	SNP	nonsynonymous	HPP12_0534		
S74	735143	C	T	SNP	nonsynonymous	HPP12_0534		
S74	735162	T	A	SNP	synonymous	HPP12_0534		
S74	735163	C	A	SNP	nonsynonymous	HPP12_0534		
S74	735164	A	G	SNP	nonsynonymous	HPP12_0534		
S74	735173	C	A	SNP	nonsynonymous	HPP12_0534		
S74	735174	T	C	SNP	synonymous	HPP12_0534		
S74	735192	G	A	SNP	synonymous	HPP12_0534		
S74	735195	T	C	SNP	synonymous	HPP12_0534		
S74	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S74	805851	A	T	SNP	intergenic		HPG27_556	
S74	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S74	952086	TTCTCTCTCTCTCTCTC	TTCTCTCTCTCTCTC	indel	intergenic		HPG27_677	
S74	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S74	1122223	C	T	SNP	intergenic		HP0895	
S74	1122228	TAAAAA	TAAAAAA	indel	intergenic		HP0895	
S74	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S74	1284699	AGGGGGGGGGG	AGGGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S74	1318527	C	A	SNP	intergenic	jhp1031		
S74	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S74	1332888	T	G	SNP	synonymous	HPATCC43504_01275		
S74	1332893	T	G	SNP	nonsynonymous	HPATCC43504_01275		
S74	1332897	A	G	SNP	synonymous	HPATCC43504_01275		
S74	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S74	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S74	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S74	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S74	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S74	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S74	1332999	T	C	SNP	synonymous	HPATCC43504_01275		

S74	1362531	A	T	SNP	nonsynonymous	HPB8_1119		
S74	1456630	G	A	SNP	nonsynonymous	HP0354		
S74	1482848	G	T	SNP	nonsynonymous	HPG27_298		
S74	1482850	T	G	SNP	nonsynonymous	HPG27_298		
S74	1483092	G	A	SNP	nonsynonymous	HPG27_298		
S74	1483202	G	A	SNP	nonsynonymous	HPG27_298		
S74	1592801	T	G	SNP	synonymous	HP1354		
S75	99327	C	T	SNP	synonymous	jhp0935		
S75	122766	TTTTTTAATGAAGTT	TTT	indel	frameshift deletion	HP1003		
S75	168379	C	T	SNP	synonymous	HP1547		
S75	199597	C	T	SNP	intergenic		HP1582	jhp1488
S75	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S75	226261	G	A	SNP	nonsynonymous	HP0019_1		
S75	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S75	430922	G	A	SNP	nonsynonymous	HP0194		
S75	546511	CAAAAAAAAAAAA	A	indel	intergenic		HP1243/HPG27_298	
S75	547618	A	G	SNP	nonsynonymous	HPG27_298		
S75	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S75	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S75	691071	G	A	SNP	synonymous	HP0489		
S75	691074	A	C	SNP	synonymous	HP0489		
S75	691077	G	A	SNP	nonsynonymous	HP0489		
S75	691089	A	G	SNP	synonymous	HP0489		
S75	691091	C	A	SNP	nonsynonymous	HP0489		
S75	691101	T	C	SNP	synonymous	HP0489		
S75	691113	G	T	SNP	nonsynonymous	HP0489		
S75	691155	C	G	SNP	synonymous	HP0489		
S75	691157	C	A	SNP	nonsynonymous	HP0489		
S75	691173	C	A	SNP	synonymous	HP0489		
S75	691175	A	G	SNP	nonsynonymous	HP0489		
S75	691179	C	G	SNP	synonymous	HP0489		
S75	691185	A	G	SNP	nonsynonymous	HP0489		
S75	691189	C	G	SNP	nonsynonymous	HP0489		
S75	691190	A	C	SNP	nonsynonymous	HP0489		
S75	734658	C	T	SNP	synonymous	HPP12_0534		
S75	734661	C	G	SNP	synonymous	HPP12_0534		
S75	734672	G	A	SNP	nonsynonymous	HPP12_0534		
S75	734695	A	G	SNP	nonsynonymous	HPP12_0534		
S75	734697	A	G	SNP	synonymous	HPP12_0534		
S75	734703	T	C	SNP	synonymous	HPP12_0534		
S75	734722	G	C	SNP	nonsynonymous	HPP12_0534		
S75	735067	T	G	SNP	nonsynonymous	HPP12_0534		
S75	735072	G	T	SNP	nonsynonymous	HPP12_0534		
S75	735126	T	C	SNP	synonymous	HPP12_0534		
S75	735140	A	G	SNP	nonsynonymous	HPP12_0534		
S75	735143	C	T	SNP	nonsynonymous	HPP12_0534		
S75	735162	T	A	SNP	synonymous	HPP12_0534		
S75	735163	C	A	SNP	nonsynonymous	HPP12_0534		
S75	735164	A	G	SNP	nonsynonymous	HPP12_0534		
S75	735173	C	A	SNP	nonsynonymous	HPP12_0534		
S75	735174	T	C	SNP	synonymous	HPP12_0534		
S75	735192	G	A	SNP	synonymous	HPP12_0534		
S75	735195	T	C	SNP	synonymous	HPP12_0534		
S75	805851	A	T	SNP	intergenic		HPG27_556	
S75	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S75	960744	T	G	SNP	intergenic		HP0727	
S75	1043666	G	A	SNP	intergenic	HPnc4170	HP0811	HPnc4160
S75	1122223	C	T	SNP	intergenic		HP0895	
S75	1122228	TAAAAA	TAAAAAA	indel	intergenic		HP0895	
S75	1124346	CGAGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_1187	
S75	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S75	1284699	AGGGGGGGGGG	AGGGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S75	1318527	C	A	SNP	intergenic	jhp1031		
S75	1332593	A	C	SNP	nonsynonymous	HPATCC43504_01275		
S75	1332595	T	A	SNP	nonsynonymous	HPATCC43504_01275		
S75	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S75	1332900	G	T	SNP	synonymous	HPATCC43504_01275		
S75	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S75	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S75	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S75	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S75	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S75	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S75	1482848	G	T	SNP	nonsynonymous	HPG27_298		
S75	1482850	T	G	SNP	nonsynonymous	HPG27_298		
S75	1483092	G	A	SNP	nonsynonymous	HPG27_298		
S75	1592784	ATGTG	ATG	indel	frameshift deletion	HP1354		
S75	1592801	T	G	SNP	synonymous	HP1354		
S76	99327	C	T	SNP	synonymous	jhp0935		
S76	119015	A	C	SNP	nonsynonymous	HP1534		
S76	168379	C	T	SNP	synonymous	HP1547		
S76	199597	C	T	SNP	intergenic		HP1582	jhp1488
S76	226261	G	A	SNP	nonsynonymous	HP0019_1		
S76	276989	G	A	SNP	nonsynonymous	HPP12_0070		
S76	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S76	413866	A	G	SNP	synonymous	HP1535		
S76	413887	A	G	SNP	synonymous	HP1535		

S76	430922	G	A	SNP	nonsynonymous	HP0194		
S76	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
S76	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S76	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S76	691071	G	A	SNP	synonymous	HP0489		
S76	691074	A	C	SNP	synonymous	HP0489		
S76	691077	G	A	SNP	nonsynonymous	HP0489		
S76	691089	A	G	SNP	synonymous	HP0489		
S76	691091	C	A	SNP	nonsynonymous	HP0489		
S76	691101	T	C	SNP	synonymous	HP0489		
S76	691113	G	T	SNP	nonsynonymous	HP0489		
S76	691173	C	A	SNP	synonymous	HP0489		
S76	691175	A	G	SNP	nonsynonymous	HP0489		
S76	691179	C	G	SNP	synonymous	HP0489		
S76	691185	A	G	SNP	nonsynonymous	HP0489		
S76	691189	C	G	SNP	nonsynonymous	HP0489		
S76	691190	A	C	SNP	nonsynonymous	HP0489		
S76	830053	C	T	SNP	nonsynonymous	HPP12_0617		
S76	839999	T	C	SNP	nonsynonymous	HPB8_818		
S76	952086	TTCTCTCTCTCTCTCTC	TTCTCTCTCTCTCTC	indel	intergenic		HPG27_677	
S76	984348	G	A	SNP	nonsynonymous	HP0751		
S76	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S76	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S76	1284699	AGGGGGGGGGG	AGGGGGGGGGGG	indel	frameshift deletion	HPG27_1018		
S76	1318525	AAC	A	indel	frameshift deletion	jhp1031		
S76	1332690	C	A	SNP	synonymous	HPATCC43504_01275		
S76	1332906	C	T	SNP	synonymous	HPATCC43504_01275		
S76	1332954	T	G	SNP	synonymous	HPATCC43504_01275		
S76	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
S76	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
S76	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
S76	1332999	T	C	SNP	synonymous	HPATCC43504_01275		
S76	1592801	T	G	SNP	synonymous	HP1354		
S76	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S77	658	C	T	SNP	synonymous	HP1529		
S77	248956	C	T	SNP	nonsynonymous	jhp0037_1		
S77	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
S77	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S77	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S77	412566	T	C	SNP	nonsynonymous	HP1534		
S77	413983	C	T	SNP	synonymous	HP1535		
S77	431032	C	T	SNP	synonymous	HP0194		
S77	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S77	450628	G	A	SNP	synonymous	HP0213		
S77	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S77	548957	C	T	SNP	intergenic		jhp1163	HPG27_298
S77	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S77	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S77	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S77	626404	C	T	SNP	synonymous	HP1175		
S77	627243	G	A	SNP	nonsynonymous	HP1174		
S77	701041	G	A	SNP	nonsynonymous	HP0499		
S77	820046	C	T	SNP	nonsynonymous	HP0607		
S77	839999	T	C	SNP	nonsynonymous	HPB8_818		
S77	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S77	1124346	CGAGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_1187	
S77	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S77	1192612	C	A	SNP	intergenic		HP0953	
S77	1192627	A	G	SNP	intergenic		HP0953	
S77	1192634	A	G	SNP	intergenic		HP0953	
S77	1192636	G	A	SNP	intergenic		HP0953	
S77	1192650	T	C	SNP	intergenic		HP0953	
S77	1192686	T	G	SNP	intergenic		HP0953	
S77	1192742	C	T	SNP	intergenic		HP0953	
S77	1192744	T	C	SNP	intergenic		HP0953	
S77	1192752	A	G	SNP	intergenic		HP0953	
S77	1192763	A	G	SNP	intergenic		HP0953	
S77	1192773	A	G	SNP	intergenic		HP0953	
S77	1192779	T	G	SNP	intergenic		HP0953	
S77	1192784	T	C	SNP	intergenic		HP0953	
S77	1192786	C	T	SNP	intergenic		HP0953	
S77	1192787	A	G	SNP	intergenic		HP0953	
S77	1192792	A	G	SNP	intergenic		HP0953	
S77	1192794	C	T	SNP	intergenic		HP0953	
S77	1192820	G	A	SNP	nonsynonymous	HPP12_0950		
S77	1461824	G	T	SNP	intergenic		HP0349	HP0350
S77	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S77	1663529	GTTTTT	GTTTT	indel	1 bp deletion			
S77	1669898	C	T	SNP	synonymous	HP1450		
S78	658	C	T	SNP	synonymous	HP1529		
S78	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S78	300860	A	G	SNP	nonsynonymous	HP0090		
S78	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S78	431032	C	T	SNP	synonymous	HP0194		
S78	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S78	450628	G	A	SNP	synonymous	HP0213		
S78	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	

S78	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S78	548994	A	T	SNP	intergenic		jhp1163	HPG27_298
S78	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S78	626404	C	T	SNP	synonymous	HP1175		
S78	627243	G	A	SNP	nonsynonymous	HP1174		
S78	701041	G	A	SNP	nonsynonymous	HP0499		
S78	838963	G	T	SNP	nonsynonymous	jhp0562		
S78	839999	T	C	SNP	nonsynonymous	HPB8_818		
S78	1043704	GTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S78	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S78	1192612	C	A	SNP	intergenic		HP0953	
S78	1192627	A	G	SNP	intergenic		HP0953	
S78	1192634	A	G	SNP	intergenic		HP0953	
S78	1192636	G	A	SNP	intergenic		HP0953	
S78	1192650	T	C	SNP	intergenic		HP0953	
S78	1192686	T	G	SNP	intergenic		HP0953	
S78	1192742	C	T	SNP	intergenic		HP0953	
S78	1192744	T	C	SNP	intergenic		HP0953	
S78	1192752	A	G	SNP	intergenic		HP0953	
S78	1192763	A	G	SNP	intergenic		HP0953	
S78	1192773	A	G	SNP	intergenic		HP0953	
S78	1192779	T	G	SNP	intergenic		HP0953	
S78	1192784	T	C	SNP	intergenic		HP0953	
S78	1192786	C	T	SNP	intergenic		HP0953	
S78	1192787	A	G	SNP	intergenic		HP0953	
S78	1192792	A	G	SNP	intergenic		HP0953	
S78	1192794	C	T	SNP	intergenic		HP0953	
S78	1192820	G	A	SNP	nonsynonymous	HPP12_0950		
S78	1230615	A	T	SNP	nonsynonymous	HP0427		
S78	1362543	A	C	SNP	nonsynonymous	HPB8_1119		
S78	1608139	GT	G	indel	frameshift deletion	HP1369m		
S78	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S78	1669898	C	T	SNP	synonymous	HP1450		
S79	658	C	T	SNP	synonymous	HP1529		
S79	214663	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	intergenic		HPG27_8	
S79	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S79	431032	C	T	SNP	synonymous	HP0194		
S79	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S79	450628	G	A	SNP	synonymous	HP0213		
S79	548965	C	T	SNP	intergenic		jhp1163	HPG27_298
S79	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
S79	548994	A	T	SNP	intergenic		jhp1163	HPG27_298
S79	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S79	626404	C	T	SNP	synonymous	HP1175		
S79	627243	G	A	SNP	nonsynonymous	HP1174		
S79	701041	G	A	SNP	nonsynonymous	HP0499		
S79	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
S79	953366	G	A	SNP	nonsynonymous	HPG27_677		
S79	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
S79	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S79	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S79	1192612	C	A	SNP	intergenic		HP0953	
S79	1192627	A	G	SNP	intergenic		HP0953	
S79	1192634	A	G	SNP	intergenic		HP0953	
S79	1192636	G	A	SNP	intergenic		HP0953	
S79	1192650	T	C	SNP	intergenic		HP0953	
S79	1192686	T	G	SNP	intergenic		HP0953	
S79	1192742	C	T	SNP	intergenic		HP0953	
S79	1192744	T	C	SNP	intergenic		HP0953	
S79	1192752	A	G	SNP	intergenic		HP0953	
S79	1192763	A	G	SNP	intergenic		HP0953	
S79	1192773	A	G	SNP	intergenic		HP0953	
S79	1192779	T	G	SNP	intergenic		HP0953	
S79	1192784	T	C	SNP	intergenic		HP0953	
S79	1192786	C	T	SNP	intergenic		HP0953	
S79	1192787	A	G	SNP	intergenic		HP0953	
S79	1192792	A	G	SNP	intergenic		HP0953	
S79	1192794	C	T	SNP	intergenic		HP0953	
S79	1192820	G	A	SNP	nonsynonymous	HPP12_0950		
S79	1295885	GATATATATATATATATAT	GATATATATATATATAT	indel	intergenic		HPATCC43504_01238	
S79	1362557	G	C	SNP	synonymous	HPB8_1119		
S79	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
S79	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAGAG	indel	frameshift insertion			
S79	1669898	C	T	SNP	synonymous	HP1450		
S80	658	C	T	SNP	synonymous	HP1529		
S80	232964	C	A	SNP	intergenic		HPPC_00115	
S80	248956	C	T	SNP	nonsynonymous	jhp0037_1		
S80	267811	A	G	SNP	nonsynonymous	HP0057		
S80	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
S80	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
S80	374756	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	jhp0151		
S80	431032	C	T	SNP	synonymous	HP0194		
S80	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
S80	450628	G	A	SNP	synonymous	HP0213		
S80	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
S80	620169	C	T	SNP	nonsynonymous	HPAG1_1119		
S80	626404	C	T	SNP	synonymous	HP1175		

S80	627243	G	A	SNP	nonsynonymous	HP1174		
S80	701041	G	A	SNP	nonsynonymous	HP0499		
S80	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAA	indel	intergenic		jhp0540	
S80	820046	C	T	SNP	nonsynonymous	HP0607		
S80	958719	G	A	SNP	intergenic		HPG27_680	
S80	1043704	GTTTTTTTTTTTTT	GTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
S80	1124346	CGAGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	intergenic		HPG27_1187	
S80	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
S80	1192612	C	A	SNP	intergenic		HP0953	
S80	1192627	A	G	SNP	intergenic		HP0953	
S80	1192634	A	G	SNP	intergenic		HP0953	
S80	1192636	G	A	SNP	intergenic		HP0953	
S80	1192650	T	C	SNP	intergenic		HP0953	
S80	1192686	T	G	SNP	intergenic		HP0953	
S80	1192742	C	T	SNP	intergenic		HP0953	
S80	1192744	T	C	SNP	intergenic		HP0953	
S80	1192752	A	G	SNP	intergenic		HP0953	
S80	1192763	A	G	SNP	intergenic		HP0953	
S80	1192773	A	G	SNP	intergenic		HP0953	
S80	1192779	T	G	SNP	intergenic		HP0953	
S80	1192784	T	C	SNP	intergenic		HP0953	
S80	1192786	C	T	SNP	intergenic		HP0953	
S80	1192787	A	G	SNP	intergenic		HP0953	
S80	1192792	A	G	SNP	intergenic		HP0953	
S80	1192794	C	T	SNP	intergenic		HP0953	
S80	1192820	G	A	SNP	nonsynonymous	HPP12_0950		
S80	1640915	CGAGAGAGAGAGAGAG	CGAGAGAGAGAGAGAG	indel	frameshift insertion			
S80	1669898	C	T	SNP	synonymous	HP1450		

Supplementary Information 3 | Summary of mutations in the isolates recovered from *H. pylori*-infected C57BL/6 mice.
Number of mutations in the isolates of 10 strains recovered from *H. pylori*-infected mice stomachs 8 weeks after post infection were listed.

Animal No.	Strain Name	Total No. of mutations	SNPs (single nucleotide polymorphysms)				indel (insertion/deletion)			SNPs rate per base per year	indel rate per base per year
			total	nonsynonymous	synonymous	intergenic	Total	genic	intergenic		
#1	Hp1	37	13	5	7	1	24	8	16	5.04E-05	9.31E-05
	Hp2	36	16	5	9	2	20	7	13	6.20E-05	7.76E-05
#2	Hp3	38	16	6	8	2	22	8	14	6.20E-05	8.53E-05
	Hp4	38	18	6	9	3	20	7	13	6.98E-05	7.76E-05
#3	Hp5	34	15	6	7	2	19	6	13	5.82E-05	7.37E-05
	Hp6	33	10	4	5	1	23	7	16	3.88E-05	8.92E-05
#4	Hp7	45	25	11	12	2	20	10	10	9.69E-05	7.76E-05
	Hp8	42	19	7	10	2	23	8	15	7.37E-05	8.92E-05
#5	Hp9	46	30	11	17	2	16	5	11	1.16E-04	6.20E-05
	Hp10	36	13	5	7	1	23	8	15	5.04E-05	8.92E-05
									Average	6.79E-05	8.14E-05
									SD	2.32E-05	9.50E-06

Strain name	Position in ATCC43504	Sequence in ATCC43504	Sequence in Reisolate	Type of difference		Annotation	Gene nearby the intergenic region (#1)	Gene nearby the intergenic region (#2)
Hp1	658	C	T	SNP	synonymous	HP1529		
Hp1	146525	CAAAAAAAAAAAA	CAAAAAAAAAAAAAAA	indel	intergenic		HP1397	
Hp1	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp1	249268	C	T	SNP	nonsynonymous	HPB8_1582		
Hp1	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp1	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp1	431032	C	T	SNP	synonymous	HP0194		
Hp1	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp1	450628	G	A	SNP	synonymous	HP0213		
Hp1	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp1	483804	T	TTAAATACA	indel	intergenic		HPATCC43504_00472	
Hp1	546124	TATTTAATCTT	T	indel	intergenic		HP1243/HPG27_298	
Hp1	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp1	581837	ATCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp1	620169	C	T	SNP	nonsynonymous	HP1180		
Hp1	626404	C	T	SNP	synonymous	HP1175		
Hp1	627243	G	A	SNP	nonsynonymous	HP1174		
Hp1	649248	TG	TGG	indel	frameshift insertion	HP1156		
Hp1	701041	G	A	SNP	nonsynonymous	HP0499		
Hp1	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp1	840225	GCAAAATACGATGATCTCACCA	G	indel	nonframeshift deletion	jhp0563		
Hp1	945518	C	T	SNP	synonymous	jhp0853		
Hp1	958574	CGAGAGAGAGAGA	CGAGAGAGAGAGAGA	indel	intergenic		HPG27_680	
Hp1	958702	CAAAAAAAAAAAAAAAA	CAAAAGAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp1	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp1	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp1	1124346	CGAGAGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	intergenic		HPG27_1187	
Hp1	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp1	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp1	1230599	TTCAAGCAA	T	indel	frameshift deletion	HP0427		
Hp1	1319839	CAAAAAAAAAAAAA	CCAAAAAAAAAAAAAAA	indel	intergenic		HELPY_1075	jhp1032
Hp1	1331327	G	T	SNP	nonsynonymous	HPATCC43504_01275		
Hp1	1331328	G	A	SNP	synonymous	HPATCC43504_01275		
Hp1	1592784	ATG	A	indel	frameshift deletion	HP1354		
Hp1	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp1	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp1	1669898	C	T	SNP	synonymous	HP1450		
Hp2	658	C	T	SNP	synonymous	HP1529		
Hp2	146525	CAAAAAAAAAAAA	CAAAAAAAAAAAAAAA	indel	intergenic		HP1397	
Hp2	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp2	232964	CAAAAAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAAAAAA	indel	intergenic		HPPC_00115	
Hp2	249268	C	T	SNP	nonsynonymous	HPB8_1582		
Hp2	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp2	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp2	431032	C	T	SNP	synonymous	HP0194		
Hp2	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp2	450628	G	A	SNP	synonymous	HP0213		
Hp2	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp2	464722	T	G	SNP	synonymous	HP0227		
Hp2	464740	C	A	SNP	synonymous	HP0227		
Hp2	464743	A	G	SNP	synonymous	HP0227		
Hp2	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp2	548978	G	T	SNP	intergenic		jhp1163	HPG27_298
Hp2	581837	ATCAAATACTCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp2	620169	C	T	SNP	nonsynonymous	HP1180		
Hp2	626404	C	T	SNP	synonymous	HP1175		
Hp2	627243	G	A	SNP	nonsynonymous	HP1174		
Hp2	649248	TG	TGG	indel	frameshift insertion	HP1156		
Hp2	701041	G	A	SNP	nonsynonymous	HP0499		
Hp2	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp2	945518	C	T	SNP	synonymous	jhp0853		
Hp2	958574	CGAGAGAGAGAGA	CGAGAGAGAGAGAGA	indel	intergenic		HPG27_680	
Hp2	958702	CAAAAAAAAAAAAAAAA	CAATAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp2	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp2	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp2	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp2	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp2	1423237	G	A	SNP	nonsynonymous	HP1068		
Hp2	1592784	ATG	A	indel	frameshift deletion	HP1354		
Hp2	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp2	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp2	1640915	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	jhp1312		
Hp2	1669898	C	T	SNP	synonymous	HP1450		
Hp3	658	C	T	SNP	synonymous	HP1529		
Hp3	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp3	249555	G	A	SNP	nonsynonymous	HP0044		
Hp3	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp3	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp3	413866	A	G	SNP	synonymous	HP1535		
Hp3	413983	C	T	SNP	synonymous	HP1535		
Hp3	431032	C	T	SNP	synonymous	HP0194		
Hp3	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp3	450628	G	A	SNP	synonymous	HP0213		
Hp3	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp3	483804	TTAAATACA	T	indel	intergenic		HPATCC43504_00472	

Strain name	Position in ATCC43504	Sequence in ATCC43504	Sequence in Reisolate	Type of difference		Annotation	Gene nearby the intergenic region (#1)	Gene nearby the intergenic region (#2)
Hp3	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp3	548979	TA	TTGA	indel	intergenic		jhp1163	HPG27_298
Hp3	581837	ATCAAATACTCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp3	620169	C	T	SNP	nonsynonymous	HP1180		
Hp3	626404	C	T	SNP	synonymous	HP1175		
Hp3	627243	G	A	SNP	nonsynonymous	HP1174		
Hp3	701041	G	A	SNP	nonsynonymous	HP0499		
Hp3	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp3	958702	CAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp3	1043704	GTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp3	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp3	1122223	C	T	SNP	intergenic		HP0895	HPnc4510
Hp3	1122228	TAAAAA	TAAAAAA	indel	intergenic		HP0895	HPnc4510
Hp3	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp3	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp3	1331327	G	T	SNP	nonsynonymous	HPATCC43504_01275		
Hp3	1331328	G	A	SNP	synonymous	HPATCC43504_01275		
Hp3	1362543	ACCCCCCCCCCCCC	ACGCCCCCCCCCCCCCCCCC	indel	nonframeshift insertion	HPB8_1119		
Hp3	1482699	GAAAAAA	GAAAAAAAA	indel	intergenic		HPG27_298	
Hp3	1483269	A	G	SNP	nonsynonymous	HPG27_298		
Hp3	1592787	TGGGGGGGGGGGG		indel	nonframeshift substitution	HP1354		
Hp3	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp3	1608140	TGA	TGGA	indel	frameshift substitution	HP1369m		
Hp3	1627016	TGGGGGGGGGGGG	TGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp3	1640915	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	jhp1312		
Hp3	1669898	C	T	SNP	synonymous	HP1450		
Hp4	658	C	T	SNP	synonymous	HP1529		
Hp4	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp4	232964	C	A	SNP	intergenic		HPPC_00115	
Hp4	249555	G	A	SNP	nonsynonymous	HP0044		
Hp4	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp4	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp4	413866	A	G	SNP	synonymous	HP1535		
Hp4	413887	A	G	SNP	synonymous	HP1535		
Hp4	413983	C	T	SNP	synonymous	HP1535		
Hp4	431032	C	T	SNP	synonymous	HP0194		
Hp4	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp4	450628	G	A	SNP	synonymous	HP0213		
Hp4	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp4	483804	TTAAATACA	T	indel	intergenic		HPATCC43504_00472	
Hp4	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp4	548979	TA	TTGA	indel	intergenic		jhp1163	HPG27_298
Hp4	581837	ATCAAATACTCAAATACTCAAATACTCAAATACTCA AATAC	A	indel	intergenic		HPATCC43504_00587	
Hp4	620169	C	T	SNP	nonsynonymous	HP1180		
Hp4	626404	C	T	SNP	synonymous	HP1175		
Hp4	627243	G	A	SNP	nonsynonymous	HP1174		
Hp4	701041	G	A	SNP	nonsynonymous	HP0499		
Hp4	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp4	958702	CAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp4	1043704	GTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp4	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp4	1122223	C	T	SNP	intergenic		HP0895	HPnc4510
Hp4	1122228	TAAAAA	TAAAAAA	indel	intergenic		HP0895	HPnc4510
Hp4	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp4	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp4	1331327	G	T	SNP	nonsynonymous	HPATCC43504_01275		
Hp4	1331328	G	A	SNP	synonymous	HPATCC43504_01275		
Hp4	1362543	ACCCCCCCCCCCCC	AACCCCCCCCCCCCCCC	indel	frameshift substitution	HPB8_1119		
Hp4	1483269	A	G	SNP	nonsynonymous	HPG27_298		
Hp4	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp4	1608140	TGA	TGGA	indel	frameshift substitution	HP1369m		
Hp4	1627016	TGGGGGGGGGGGG	TGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp4	1640915	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	jhp1312		
Hp4	1669898	C	T	SNP	synonymous	HP1450		
Hp5	658	C	T	SNP	synonymous	HP1529		
Hp5	60246	G	A	SNP	synonymous	jhp1371		
Hp5	119015	A	C	SNP	nonsynonymous	HP1534		
Hp5	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp5	232964	CAAAAAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAAAAAA	indel	intergenic		HPPC_00115	
Hp5	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp5	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp5	431032	C	T	SNP	synonymous	HP0194		
Hp5	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp5	450628	G	A	SNP	synonymous	HP0213		
Hp5	455599	TGGGGGGGGGG	TTGGGGGGGGGGG	indel	nonframeshift insertion	HP0217		
Hp5	483804	TTAAATACA	T	indel	intergenic		HPATCC43504_00472	
Hp5	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp5	581837	ATCAAATACTCAAATACTCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp5	620169	C	T	SNP	nonsynonymous	HP1180		
Hp5	626404	C	T	SNP	synonymous	HP1175		
Hp5	627243	G	A	SNP	nonsynonymous	HP1174		
Hp5	701041	G	A	SNP	nonsynonymous	HP0499		
Hp5	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp5	958574	CGAGAGAGAGAGA	CGAGAGAGAGAGAGA	indel	intergenic		HPG27_680	

Strain name	Position in ATCC43504	Sequence in ATCC43504	Sequence in Reisolate	Type of difference		Annotation	Gene nearby the intergenic region (#1)	Gene nearby the intergenic region (#2)
Hp5	958702	CAAAAAAAAAAAAAAAA	CCAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp5	984279	C	T	SNP	nonsynonymous	HP0751		
Hp5	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp5	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp5	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp5	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp5	1192576	G	T	SNP	intergenic		HP0953	
Hp5	1284699	AGGGGGGGGGG	AGGGGGGGGGGG	indel	stopgain	HPG27_1018		
Hp5	1331327	G	T	SNP	nonsynonymous	HPATCC43504_01275		
Hp5	1331328	G	A	SNP	synonymous	HPATCC43504_01275		
Hp5	1362543	ACCCCCCCCCCCCC	ACCCCCCCCCCCCCC	indel	frameshift insertion	HPB8_1119		
Hp5	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp5	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp5	1669898	C	T	SNP	synonymous	HP1450		
Hp6	658	C	T	SNP	synonymous	HP1529		
Hp6	146525	CAAAAAAAAAAAA	CAAAAAAAAAAAAAAA	indel	intergenic		HP1397	
Hp6	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp6	232959	CTTTA	CTTTCAA	indel	intergenic		HPPC_00115	
Hp6	249268	C	T	SNP	nonsynonymous	HPB8_1582		
Hp6	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp6	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp6	431032	C	T	SNP	synonymous	HP0194		
Hp6	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp6	450628	G	A	SNP	synonymous	HP0213		
Hp6	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp6	483804	T	TTAAATACA	indel	intergenic		HPATCC43504_00472	
Hp6	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp6	581837	ATCAAATACTCAAATACTCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp6	620169	C	T	SNP	nonsynonymous	HP1180		
Hp6	626404	C	T	SNP	synonymous	HP1175		
Hp6	627243	G	A	SNP	nonsynonymous	HP1174		
Hp6	701041	G	A	SNP	nonsynonymous	HP0499		
Hp6	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp6	805851	AT	AATT	indel	intergenic		HPG27_556	
Hp6	958574	CGAGAGAGAGAGA	CGAGAGAGAGAGAGA	indel	intergenic		HPG27_680	
Hp6	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp6	984287	GA	G	indel	frameshift deletion	HP0751		
Hp6	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp6	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp6	1124346	CGAGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	intergenic		HPG27_1187	
Hp6	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp6	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp6	1362543	ACCCCCCCCCCCCC	ACGCGCCCCCCCCCCCCCC	indel	frameshift substitution	HPB8_1119		
Hp6	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp6	1608125	TGGGGGGGGGGGGGG	TTGGGGGGGGGGGGGG	indel	nonframeshift substitution	HP1369m		
Hp6	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp6	1669898	C	T	SNP	synonymous	HP1450		
Hp7	8299	AGGGGGGGGGGG	AGGGGGGGGGGGG	indel	frameshift insertion	HPAG1_1393		
Hp7	99327	C	T	SNP	synonymous	jhp0935		
Hp7	168379	C	T	SNP	synonymous	HP1547		
Hp7	199597	C	T	SNP	intergenic		HP1582	jhp1488
Hp7	232964	C	A	SNP	intergenic		HPPC_00115	
Hp7	251374	G	A	SNP	nonsynonymous	HP0045		
Hp7	276989	G	A	SNP	nonsynonymous	HPP12_0070		
Hp7	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	nonframeshift insertion	HPP12_0096		
Hp7	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp7	332924	TTGTCTTTTGTTTTTCTTGTTC	T	indel	nonframeshift deletion	HP0119		
Hp7	424517	CGGGG	CGGGGG	indel	intergenic		HP0189	HP0188
Hp7	430922	G	A	SNP	nonsynonymous	HP0194		
Hp7	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp7	483804	TTAAATACA	T	indel	intergenic		HPATCC43504_00472	
Hp7	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp7	581837	ATCAAATACTCAAATACTCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp7	691071	G	A	SNP	synonymous	HP0489		
Hp7	691074	A	C	SNP	synonymous	HP0489		
Hp7	691077	G	A	SNP	synonymous	HP0489		
Hp7	691088	C	CG	indel	frameshift insertion	HP0489		
Hp7	691101	T	C	SNP	synonymous	HP0489		
Hp7	691113	G	T	SNP	nonsynonymous	HP0489		
Hp7	691157	C	A	SNP	nonsynonymous	HP0489		
Hp7	691173	C	A	SNP	synonymous	HP0489		
Hp7	691175	A	G	SNP	nonsynonymous	HP0489		
Hp7	691179	C	G	SNP	synonymous	HP0489		
Hp7	691185	A	G	SNP	synonymous	HP0489		
Hp7	691189	C	G	SNP	nonsynonymous	HP0489		
Hp7	691190	A	C	SNP	nonsynonymous	HP0489		
Hp7	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp7	830053	C	T	SNP	nonsynonymous	HPP12_0617		
Hp7	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp7	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp7	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp7	1262251	C	A	SNP	nonsynonymous	jhp0981		
Hp7	1295885	GATATATATATATATATAT	GATATATATATATATATGTATAT	indel	frameshift insertion	gene_1054		
Hp7	1295885	GATATATATATATATATAT	GGTATATATATATATATATAT	indel	frameshift insertion	gene_1054		
Hp7	1332987	T	C	SNP	synonymous	HPATCC43504_01275		

Strain name	Position in ATCC43504	Sequence in ATCC43504	Sequence in Reisolate	Type of difference		Annotation	Gene nearby the intergenic region (#1)	Gene nearby the intergenic region (#2)
Hp7	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
Hp7	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
Hp7	1332999	T	C	SNP	synonymous SNV	HPATCC43504_01275		
Hp7	1362543	ACCCCCCCCCCCCC	AACCCCCCCCCCCCCCCCCCC	indel	frameshift substitution	HPB8_1119		
Hp7	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp7	1608125	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGG	indel	frameshift insertion	HP1369m		
Hp7	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp8	658	C	T	SNP	synonymous	HP1529		
Hp8	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp8	249555	G	A	SNP	nonsynonymous	HP0044		
Hp8	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp8	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp8	413866	A	G	SNP	synonymous	HP1535		
Hp8	413887	A	G	SNP	synonymous	HP1535		
Hp8	413983	C	T	SNP	synonymous	HP1535		
Hp8	431032	C	T	SNP	synonymous	HP0194		
Hp8	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp8	444106	GTTTTTTTTTTT	GTTTTTTTTTTTTTT	indel	intergenic		HP0209	HPPC_01040
Hp8	450628	G	A	SNP	synonymous	HP0213		
Hp8	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp8	483804	TTAAATACA	T	indel	intergenic		HPATCC43504_00472	
Hp8	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp8	548979	TA	TTGA	indel	intergenic		jhp1163	HPG27_298
Hp8	581837	ATCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp8	620169	C	T	SNP	nonsynonymous	HP1180		
Hp8	626404	C	T	SNP	synonymous	HP1175		
Hp8	627243	G	A	SNP	nonsynonymous	HP1174		
Hp8	701041	G	A	SNP	nonsynonymous	HP0499		
Hp8	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp8	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp8	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp8	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp8	1122223	C	T	SNP	intergenic		HP0895	HPnc4510
Hp8	1122228	TAAAAA	TAAAAAA	indel	intergenic		HP0895	HPnc4510
Hp8	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp8	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp8	1331328	G	A	SNP	synonymous	HPATCC43504_01275		
Hp8	1362543	ACCCCCCCCCCCCC	ACCCCCCCCCCCCCCC	indel	frameshift insertion	HPB8_1119		
Hp8	1482848	G	T	SNP	nonsynonymous	HPG27_298		
Hp8	1483202	G	A	SNP	nonsynonymous	HPG27_298		
Hp8	1483269	A	G	SNP	nonsynonymous	HPG27_298		
Hp8	1484980	G	A	SNP	synonymous	HPATCC43504_01426		
Hp8	1484991	ATTTTTTTT	ACTCTTTTTTTT	indel	nonframeshift insertion	HPATCC43504_01426		
Hp8	1582075	GTTTTTTTTTTTT	GATTTTTTTTTTTTTTT	indel	intergenic		HELPY_1317	HP0228
Hp8	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp8	1608140	TGA	TGGA	indel	frameshift substitution	HP1369m		
Hp8	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp8	1640915	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	jhp1312		
Hp8	1669898	C	T	SNP	synonymous	HP1450		
Hp9	8299	AGGGGGGGGGGG	AGGGGGGGGGGGG	indel	frameshift insertion	HPAG1_1393		
Hp9	99327	C	T	SNP	synonymous	jhp0935		
Hp9	168379	C	T	SNP	synonymous	HP1547		
Hp9	199597	C	T	SNP	intergenic		HP1582	jhp1488
Hp9	232964	C	A	SNP	intergenic		HPPC_00115	
Hp9	276989	G	A	SNP	nonsynonymous	HPP12_0070		
Hp9	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp9	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp9	430922	G	A	SNP	nonsynonymous	HP0194		
Hp9	444106	GTTTTTTTTTTT	GTTTTTTTTTTTTTT	indel	intergenic		HP0209	HPPC_01040
Hp9	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp9	464722	T	G	SNP	synonymous	HP0227		
Hp9	464740	C	A	SNP	synonymous	HP0227		
Hp9	464743	A	G	SNP	synonymous	HP0227		
Hp9	483804	TTAAATACA	T	indel	intergenic		HPATCC43504_00472	
Hp9	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp9	581837	ATCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp9	691071	G	A	SNP	synonymous	HP0489		
Hp9	691074	A	C	SNP	synonymous	HP0489		
Hp9	691077	G	A	SNP	synonymous	HP0489		
Hp9	691089	A	G	SNP	synonymous	HP0489		
Hp9	691091	C	A	SNP	nonsynonymous	HP0489		
Hp9	691101	T	C	SNP	synonymous	HP0489		
Hp9	691113	G	T	SNP	nonsynonymous	HP0489		
Hp9	691173	C	A	SNP	synonymous	HP0489		
Hp9	691175	A	G	SNP	nonsynonymous	HP0489		
Hp9	691179	C	G	SNP	synonymous	HP0489		
Hp9	691185	A	G	SNP	synonymous	HP0489		
Hp9	691189	C	G	SNP	nonsynonymous	HP0489		
Hp9	691190	A	C	SNP	nonsynonymous	HP0489		
Hp9	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp9	830053	C	T	SNP	nonsynonymous	HPP12_0617		
Hp9	879985	C	T	SNP	nonsynonymous	HP0656		
Hp9	958702	CAAAAAAAAAAAAAAAA	CAAAAAAAAAAAAAAAAAA	indel	intergenic		HPG27_680	
Hp9	1043704	GTTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp9	1139646	GTTTTTTTTTTTT	GTTTTTTTTTTTTT	indel	intergenic		HPP12_0910	

Strain name	Position in ATCC43504	Sequence in ATCC43504	Sequence in Reisolate	Type of difference		Annotation	Gene nearby the intergenic region (#1)	Gene nearby the intergenic region (#2)
Hp9	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp9	1192157	T	C	SNP	nonsynonymous	HP0953		
Hp9	1331328	G	A	SNP	synonymous	HPATCC43504_01275		
Hp9	1332987	T	C	SNP	synonymous	HPATCC43504_01275		
Hp9	1332996	T	G	SNP	synonymous	HPATCC43504_01275		
Hp9	1332997	G	C	SNP	nonsynonymous	HPATCC43504_01275		
Hp9	1332999	T	C	SNP	synonymous SNV	HPATCC43504_01275		
Hp9	1362543	ACCCCCCCCCCCCC	ACCCCCCCCCCCCCCC	indel	frameshift insertion	HPB8_1119		
Hp9	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp9	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp10	658	C	T	SNP	synonymous	HP1529		
Hp10	8299	AGGGGGGGGGGG	AGGGGGGGGGGGG	indel	frameshift insertion	HPAG1_1393		
Hp10	214663	CGAGAGAGAGAGAGA	CGAGAGAGAGAGAGAGA	indel	frameshift insertion	HPG27_8		
Hp10	232959	CTTTA	CTTTCA	indel	intergenic		HPPC_00115	
Hp10	303827	CGGGGGGGGGGGG	CGGGGGGGGGGGGG	indel	frameshift insertion	HPP12_0096		
Hp10	315327	ACCCCCCCCCC	ACCCCCCCCCCC	indel	intergenic		HP0103	
Hp10	431032	C	T	SNP	synonymous	HP0194		
Hp10	437765	G	C	SNP	intergenic		HP0204	HELPY_0206
Hp10	450628	G	A	SNP	synonymous	HP0213		
Hp10	455599	TGGGGGGGGGG	TGGGGGGGGGGG	indel	frameshift insertion	HP0217		
Hp10	483804	TTAAATACATAAATACATAAATACA	T	indel	intergenic		HPATCC43504_00472	
Hp10	546114	TATTTA	T	indel	intergenic		HP1243/HPG27_298	
Hp10	546140	A	AATCTTATTTG	indel	intergenic		HP1243/HPG27_298	
Hp10	546511	CAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HP1243/HPG27_298	
Hp10	581837	ATCAAATACTCAAATACTCAAATAC	A	indel	intergenic		HPATCC43504_00587	
Hp10	620169	C	T	SNP	nonsynonymous	HP1180		
Hp10	626404	C	T	SNP	synonymous	HP1175		
Hp10	627243	G	A	SNP	nonsynonymous	HP1174		
Hp10	701041	G	A	SNP	nonsynonymous	HP0499		
Hp10	801836	GAAAAAAAAAAAA	GAAAAAAAAAAAAA	indel	intergenic		jhp0540	
Hp10	958702	CAAAAAAAAAAAAAAAA		indel	intergenic		HPG27_680	
Hp10	1043704	GTTTTTTTTTTTTT	GTTTTTTTTTTTTTTTTT	indel	frameshift insertion/intergenic	HPnc4170	HP0811	HPnc4160
Hp10	1102157	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HP0876	HP0875
Hp10	1124501	GTTTTTTTTTTTT	GTTTTTTTTTTTT	indel	intergenic		HPG27_1187	
Hp10	1187376	ATTTTTTT	ATTTTTTTT	indel	intergenic		HP0947	
Hp10	1302389	G	A	SNP	synonymous	HP0364		
Hp10	1319683	G	C	SNP	nonsynonymous	jhp1032		
Hp10	1319839	CAAAAAAAAAAAAA	CAAAAAAAAAAAAA	indel	intergenic		HELPY_1075	jhp1032
Hp10	1331328	G	A	SNP	synonymous	HPATCC43504_01275		
Hp10	1362543	ACCCCCCCCCCCCC	ACCCCCCCCCCCCCCC	indel	frameshift insertion	HPB8_1119		
Hp10	1410493	G	A	SNP	nonsynonymous	jhp0373		
Hp10	1592785	TGTG	T	indel	nonframeshift deletion	HP1354		
Hp10	1593602	TGGGGGGGGGGG	TGGGGGGGGGGGG	indel	frameshift insertion	HP1354		
Hp10	1608125	TGGGGGGGGGGGGGG	TGGGGGGGGGGGGGGGGG	indel	nonframeshift insertion	HP1369m		
Hp10	1627016	TGGGGGGGGGGGGG	TGGGGGGGGGGGGGGGG	indel	intergenic		HP1406	HELPY_1371
Hp10	1669898	C	T	SNP	synonymous	HP1450		

Supplementary Information 5 | Information of *H. pylori* clinical isolates used in Fig. 1f and Fig 4g

Clinical isolates obtained from gastric cancer patients (C)				
Strain name	Isolation country	Host pathology	Source	No. of T-repeat
CPY6271	Japan	Cancer	GCA_000274665.1	11
F57	Japan	Cancer	GCA_000270065.1	11
HE_C13	Sweden	Cancer	SAMN08381257	11
ELS37	ElSalvador	Cancer	GCA_000255955.1	12
VN0464	Vietnam	Cancer	Registration in progress	12
XZ274	China	Cancer	GCA_000262655.1	12
HE_C52	Sweden	Cancer	SAMN08381260	13
SJM180	Peru	Cancer	GCA_000148855.1	13
VN0232	Vietnam	Cancer	Registration in progress	13
VN0246	Vietnam	Cancer	Registration in progress	13
VN0274	Vietnam	Cancer	Registration in progress	13
VN0390	Vietnam	Cancer	Registration in progress	13
132	Singapore	Cancer	SAMN05771043	14
132A	Singapore	Cancer	SAMN05804595	14
C664	China	Cancer	SAMN03013094	14
CPY6261	Japan	Cancer	SAMN00777167	14
CPY6311	Japan	Cancer	SAMN00777266	14
GC30_HL	France	Cancer	SAMN08381231	14
H9	Singapore	Cancer	SAMN05804593	14
HLJ039	China	Cancer	SAMN02597427	14
VN0219	Vietnam	Cancer	Registration in progress	14
VN0224	Vietnam	Cancer	Registration in progress	14
VN0227	Vietnam	Cancer	Registration in progress	14
VN0240	Vietnam	Cancer	Registration in progress	14
VN0352	Vietnam	Cancer	Registration in progress	14
VN0355	Vietnam	Cancer	Registration in progress	14
VN0401	Vietnam	Cancer	Registration in progress	14
VN0403	Vietnam	Cancer	Registration in progress	14
VN0405	Vietnam	Cancer	Registration in progress	14
VN0448	Vietnam	Cancer	Registration in progress	14
1177	Singapore	Cancer	SAMN05791070	15
22402	Colombia	Cancer	SAMN05405368	15
26093	Colombia	Cancer	SAMN05394518	15
26100	Colombia	Cancer	SAMN05393257	15
GC69_HL	France	Cancer	SAMN08381240	15
HE_C30	Sweden	Cancer	SAMN08381272	15
PeCan18	Peru	Cancer	SAMN02603196	15
PeCan4	Peru	Cancer	SAMN02603203	15
VN0212	Vietnam	Cancer	Registration in progress	15
VN0220	Vietnam	Cancer	Registration in progress	15
VN0228	Vietnam	Cancer	Registration in progress	15
VN0229	Vietnam	Cancer	Registration in progress	15
VN0272	Vietnam	Cancer	Registration in progress	15
VN0411	Vietnam	Cancer	Registration in progress	15
VN0472	Vietnam	Cancer	Registration in progress	15
IID6299	Japan	Cancer	LC497495	15
178	Singapore	Cancer	SAMN05771044	16
30950	Belgium	Cancer	SAMN08381243	16
98-10	Japan	Cancer	SAMN02472069	16
H30	Singapore	Cancer	SAMN05806843	16
S468A	Singapore	Cancer	SAMN05804677	16
VN0235	Vietnam	Cancer	Registration in progress	16
VN0264	Vietnam	Cancer	Registration in progress	16
VN0271	Vietnam	Cancer	Registration in progress	16
VN0348	Vietnam	Cancer	Registration in progress	16
VN0361	Vietnam	Cancer	Registration in progress	16
26084	Colombia	Cancer	SAMN05395355	17
F32	Japan	Cancer	SAMD00060973	17
HE_C58	Sweden	Cancer	SAMN08381263	17
HE_C18	Sweden	Cancer	SAMN08381270	18
VN0484	Vietnam	Cancer	Registration in progress	18
VN0495	Vietnam	Cancer	Registration in progress	18
IID6298	Japan	Cancer	LC497494	19

Clinical isolates obtained from non-gastric cancer patients (NC)				
Strain name	Isolation country	Host pathology	Source	No. of T-repeat
IID6291	Japan	Gastritis	LC497489	8
IID6268	Japan	Gastritis	LC497473	9
VN0630	Vietnam	Gastritis	Registration in progress	10
462	UK	Normal	SAMN08381250	11
Hp_A-9	US	DU	SAMN00778864	11
IID6266	Japan	Gastritis	LC497471	11
VN0607	Vietnam	Gastritis	Registration in progress	11
3770	France	Gastritis	SAMN08381234	12
3824	France	Gastritis	SAMN08381236	12
HE_NC14_2	Sweden	Gastritis	Registration in progress	12
HE_NC5_3	Sweden	Gastritis	Registration in progress	12
Hp_A-20	US	DU	SAMN00778862	12
Hp_H-16	US	DU	SAMN00778876	12
Hp_H-42	US	DU	SAMN00778881	12
Hp_H-44	US	DU	SAMN00778872	12
VN0530	Vietnam	Gastritis	Registration in progress	12
VN0559	Vietnam	DU	Registration in progress	12
VN0660	Vietnam	Gastritis	Registration in progress	12
VN1154	Vietnam	Gastritis	Registration in progress	12
VN1270	Vietnam	DU	Registration in progress	12
456	UK	Normal	SAMN08381249	13
518	UK	Normal	SAMN08381251	13
HE_NC20_5	Sweden	Gastritis	SAMN08381282	13
IID6271	Japan	GDU	LC497476	13
IID6273	Japan	DU	LC497477	13
IID6292	Japan	GDU	LC497490	13
VN0510	Vietnam	Gastritis	Registration in progress	13
VN0511	Vietnam	DU	Registration in progress	13
VN0598	Vietnam	Gastritis	Registration in progress	13
VN0606	Vietnam	Gastritis	Registration in progress	13
VN0612	Vietnam	DU	Registration in progress	13
VN0638	Vietnam	DU	Registration in progress	13
VN0647	Vietnam	DU	Registration in progress	13
VN0768	Vietnam	Gastritis	Registration in progress	13
VN0899	Vietnam	DU	Registration in progress	13
VN1249	Vietnam	Gastritis	Registration in progress	13
VN1250	Vietnam	Gastritis	Registration in progress	13
VN1280	Vietnam	DU	Registration in progress	13
CPY1662	Japan	DU	SAMN01178408	14
HE_NC19_5	Sweden	Gastritis	SAMN08381281	14
HE_NC60_1	Sweden	Gastritis	SAMN08381303	14
IID6267	Japan	GU	LC497472	14
IID6274	Japan	Gastritis	LC497478	14
IID6280	Japan	Gastritis	LC497482	14
IID6286	Japan	DU	LC497485	14
IID6289	Japan	DU	LC497488	14
IID6301	Japan	GDU	LC497497	14
IID6302	Japan	DU	LC497498	14
IID6303	Japan	GU	LC497499	14
J99	US	DU	SAMN02602990	14
SSR1	Ireland	GU	SAMN08381226	14
VN0532	Vietnam	Gastritis	Registration in progress	14
VN0563	Vietnam	Gastritis	Registration in progress	14
VN0594	Vietnam	Gastritis	Registration in progress	14
VN0611	Vietnam	Gastritis	Registration in progress	14
VN0637	Vietnam	Gastritis	Registration in progress	14
VN0656	Vietnam	Gastritis	Registration in progress	14
VN0670	Vietnam	Gastritis	Registration in progress	14
VN0686	Vietnam	Gastritis	Registration in progress	14
VN0786	Vietnam	Gastritis	Registration in progress	14
VN1180	Vietnam	DU	Registration in progress	14
VN1193	Vietnam	Gastritis	Registration in progress	14
VN1196	Vietnam	Gastritis	Registration in progress	14
VN1203	Vietnam	Gastritis	Registration in progress	14
VN1204	Vietnam	Gastritis	Registration in progress	14
VN1212	Vietnam	Gastritis	Registration in progress	14
VN1213	Vietnam	Gastritis	Registration in progress	14
VN1219	Vietnam	Gastritis	Registration in progress	14
VN1225	Vietnam	Gastritis	Registration in progress	14
VN1226	Vietnam	Gastritis	Registration in progress	14
VN1246	Vietnam	DU	Registration in progress	14
VN1266	Vietnam	DU	Registration in progress	14
VN1281	Vietnam	DU	Registration in progress	14
51	SouthKorea	DU	SAMN02603300	15
CPY1313	Japan	DU	SAMN00777156	15
F30	Japan	DU	SAMD00060972	15
HE_NC38_5	Sweden	Gastritis	SAMN08381297	15
IID6270	Japan	GU	LC497475	15
IID6275	Japan	DU	LC497479	15
IID6279	Japan	Gastritis	LC497481	15
IID6285	Japan	Gastritis	LC497484	15
IID6287	Japan	Gastritis	LC497486	15
IID6293	Japan	Gastritis	LC497491	15
IID6294	Japan	Gastritis	LC497492	15
IID6296	Japan	DU	LC497493	15
IID6300	Japan	Gastritis	LC497496	15
TN2GF4	unknown	DU	SAMN08381237	15
VN0500	Vietnam	Gastritis	Registration in progress	15
VN0526	Vietnam	Gastritis	Registration in progress	15
VN0635	Vietnam	Gastritis	Registration in progress	15
VN0636	Vietnam	Gastritis	Registration in progress	15
VN0639	Vietnam	Gastritis	Registration in progress	15
VN0658	Vietnam	Gastritis	Registration in progress	15
VN0759	Vietnam	DU	Registration in progress	15
VN0760	Vietnam	Gastritis	Registration in progress	15
VN0785	Vietnam	Gastritis	Registration in progress	15
VN1156	Vietnam	Gastritis	Registration in progress	15
VN1169	Vietnam	Gastritis	Registration in progress	15
VN1192	Vietnam	Gastritis	Registration in progress	15
VN1202	Vietnam	Gastritis	Registration in progress	15
VN1205	Vietnam	DU	Registration in progress	15
VN1221	Vietnam	Gastritis	Registration in progress	15
VN1224	Vietnam	Gastritis	Registration in progress	15
VN1227	Vietnam	Gastritis	Registration in progress	15
VN1237	Vietnam	DU	Registration in progress	15
VN1258	Vietnam	DU	Registration in progress	15
VN1264	Vietnam	DU	Registration in progress	15
VN1274	Vietnam	DU	Registration in progress	15
VN1279	Vietnam	DU	Registration in progress	15
3699	France	Gastritis	SAMN08381239	16
HE_NC23_2a	Sweden	Gastritis	SAMN08381284	16
HE_NC47_5	Sweden	Gastritis	SAMN08381299	16
HE_NC89_4	Sweden	Gastritis	Registration in progress	16
IID6269	Japan	Gastritis	LC497474	16
IID6276	Japan	Gastritis	LC497480	16
IID6281	Japan	Gastritis	LC497483	16
IID6288	Japan	Gastritis	LC497487	16
VN0787	Vietnam	Gastritis	Registration in progress	16
VN1158	Vietnam	Gastritis	Registration in progress	16
VN1184	Vietnam	Gastritis	Registration in progress	16
VN1241	Vietnam	Gastritis	Registration in progress	16
VN1261	Vietnam	DU	Registration in progress	16
VN0667	Vietnam	Gastritis	Registration in progress	17
HE_NC27_4	Sweden	Gastritis	SAMN08381306	18

Supplementary Information 6 | Primers used in this study.

Purpose	Description	Primer (Forward)	Sequence (5'-3')	Primer (Reverse)	Sequence (5'-3')	Reference
RT-PCR for *H. pylori* mRNA quantitative analysis (Figures 1a, 2d, 4a, 4g; Extended Data Figures 1a, 2f, 4b, 4c, 4d, 7b, 7c)	HPnc4160	HPnc4160 det-s	AAATGTCCTGACGCTCTTACC	UNIVERSAL-as	GAATCGAGCACCAGTTACGC	This study, miScript II RT Kit (QIAGEN)
	HPnc4170	HPnc4170 det-s	CTTCGCCGAGCATTATACC	UNIVERSAL-as	GAATCGAGCACCAGTTACGC	This study, miScript II RT Kit (QIAGEN)
	jhp0540	jhp0540 det-s	TGCTTGTTTGTCCTAAGGCTTC	jhp0540 det-as	TCTGCCCTTGTTTTGCAACG	This study
	jhp1163	jhp1163 det-s	AACCGCTGAGCAACAAAACG	jhp1163 det-as	ACTCTATGATCATGCTGTGGATTTC	This study
	tpi	HELPY_0197 det-s	CAACAAGCCATCCACGCTATC	HELPY_0197 det-as	AGACGCCCTTGTTGTATGGG	This study
	tlpB	HP0103 det-s	TTACGCCAAATCCGGACACC	HP0103 det-as	TGGTCGTTCCACTCACATGC	This study
	arsS	HP0164 (jhp0151) det-s	TGAAATAGGGGATCTGGCTAACG	HP0164 (jhp0151) det-as	TAGGGGTGCGTAGTTCATGC	This study
	HP0811	HP0811 det-s	AGCGATTGTCTCTCCTATGG	HP0811 det-as	CACGCTCTGTTTAGTGAAAGGC	This study
	HP0947	HP0947 det-s	GGCCATGTATGAGCCTGATTTG	HP0947 det-as	CCGCCATCACCATTCTTTCAC	This study
	HP1354	HP1354 det-s	AGGCGCAACAGATGAAAACG	HP1354 det-as	TCAGCCCTTTGGCCATACAC	This study
	pldA	HPAG1_0475 (HP0499) det-s	AACCACTGCGCATAAATCCC	HPAG1_0475 (HP0499) det-as	CGTTATGATCGCTGGCATGG	This study
	HPB8_818	HPB8_818 det-s	CTTGTTGCGTATCTCGTGTGG	HPB8_818 det-as	AACCAATTTGTGCGCTCTCC	This study
	hopZ	HPG27_8 det-s	TCAACTCCGCCTCTGATGTG	HPG27_8 det-as	GTGCCAGCTAAAGCAATCCC	This study
	babA	HP1243/HPG27_298 det-s	GGTGGTCCACAGATGGAACC	HP1243/HPG27_298 det-as	GATTGACCAGCTCTTGTGCG	This study
	sabA	HPG27_680 det-s	GGCTACCCCACTCAATACGC	HPG27_680 det-as	ATCCTGTGGCTTGAGCTTGC	This study
	hpaA	HP0410-hpaA det-s	GCGAGTGGGGCGTTTTATAC	HP0410-hpaA det-as	GAAATGCAACTCCACGGACTC	This study
	hopE	HELPY_0660-hopE det-s	TGGCTCTCACGCCAAAATTC	HELPY_0660-hopE det-as	GCTCAGCGGCAAACTATTGG	This study
	omp14	HP0671 det-s	GCTGTATTGACGGCGGTAAG	HP0671 det-as	GCCGGCATCAAGAATTCGAC	This study
	hofC	HP0486 det-s	TACCGAGCAGGAGCAAATGG	HP0486 det-as	ACCATCAGCGATACCACGAC	This study
	horB	HPSH_00635-horB det-s	TACCCACGAGTCCCTAGAGC	HPSH_00635-horB det-as	ATCAAAGTGGGCTTATGGC	This study
	cagA	HPP12_0555-cagA det-s	CCGACTAGGGTTCCGTTCAC	HPP12_0555-cagA det-as	AGCAAGCGTTAGCCGATCTC	This study
	HP1227	HP1227 det-s	GCGATGAAGACATCAAAGCTTTAG	UNIVERSAL-as	GAATCGAGCACCAGTTACGC	This study, miScript II RT Kit (QIAGEN)
	HELPY_1262	HELPY_1262 det-s	GGGATGAGGTAAAGATTGAGCTC	UNIVERSAL-as	GAATCGAGCACCAGTTACGC	This study, miScript II RT Kit (QIAGEN)
	23SrRNA	HP23SrRNA det-s	GACTACTACTAATAGAGCGTTTGGC	UNIVERSAL-as	GAATCGAGCACCAGTTACGC	This study, miScript II RT Kit (QIAGEN)
	iceA2	HPATCC43504_00587 det-s	GATGTGGTTACAGCCACTACAC	HPATCC43504_00587 det-as	CTGATGGGCATATTGTTGCTG	This study
	tpiA	HP0194 det-s	CTAATTCCAAAGACGCGCTC	HP0194 det-as	GAAGACATTTACCTCACGCATGG	This study
	fucT2	HPP12_0096 det-s	GAGCCTAAATTATTAAAGCCAAGCC	HPP12_0096 det-as	GCTTGCGCTGATATTCTTCCTC	This study
	HP1406	HP1406 det-s	TCGCCATTGTCAAGCCATAAG	HP1406 det-as	GGCGTGAAGTGGTGTTTAAAG	This study
	HELPY_1371	HELPY_1371 det-s	CATGGAAGCTAAAGGCGAACA	HELPY_1371 det-as	GCGGTTGCAAATTCGTCAT	This study
	cgtA	HP0217 det-s	TGCCTATCAAATCGGCATGGTG	HP0217 det-as	TGCAATGGCACTTCCCTTTAG	This study
RT-PCR for mice mRNA quantitative analysis (Figure 4f)	mouse GAPDH	GAPDH det-s	GTGTCTTCACCACCATGGAG	GAPDH det-as	TCGTGGTTCACACCCATCAC	Sanada T., *et al. Nature*.2012
	mouse Cxcl2	CXCL2 det-s	CAAGGGTTGACTTCAAGAACATCC	CXCL2 det-as	CCTTGAGAGTGGCTATGACTTC	Sanada T., *et al. Nature*.2012
Cloning of *H. pylori* genome fragments for establishment of deletion mutants of *H. pylori*	cagA deletion mutant (upstream region)	CagA KO up XhoI	AAAAACCGCTCGAGGAGGGCGAATTGTTAGAAACTTG	CagA KO up EcoRI	TTTTTCCGGAATTCGTTTCCAAAAATCTTAAAGGATTAAGGAATACC	This study
	cagA deletion mutant (downstream region)	CagA KO down BamHI	AAAAACGCGGATCCGGTTTCGTTAGTCATTGTTTCTCCTTA	CagA KO down NotI	TTTTCCTTTTGCGGCCGCCTACCAATCCATCGCATGACT	This study
	hpnc4160/hpnc4170 deletion mutant (upstream region)	HPnc4160/4170 KO up KpnI	AAAAAAGGTACCCAAGCCTAATCCCTTACAAACCC	HPnc4160/4170 KO up ClaI	TTTTTTATCGATCCTTCGCCGAGCATTATACC	This study
	hpnc4160/hpnc4170 deletion mutant (downstream region)	HPnc4160/4170 KO down BamHI	AAAAAAGGATCCCTCAACTTTTCACGCAAACTAACC	HPnc4160/4170 KO down SacI	TTTTTTGAGCTCGAGCTTTGATTATGGCGATGAATTAG	This study
Construction of T repeat mutated *H. pylori*	Cloning of T repeat sequences from H. pylori isolated from Mongolian gerbils 8 wks post infection into pKSB plasmid	pKSB-HPnc4160 Point mut ApaI	AAAAGGGCCCAGCGATTTGACCTATGATGT	pKSB-HPnc4160 Point mut XhoI	AAAACTCGAGACATCTTCTTTCACGCTCTG	This study
Construction of NB-cagA *H. pylori*	pKSB-NB-cagA construct (NB-cagA fragment from synthesized NB-cagA containing plasmid)	pKSB-CagA-NB-ApaI	AAAAAAGGGCCCGCGACCTTGAAAATTCCGTTAAAG	pKSB-CagA-NB-XhoI	AAAAAACTCGAGCTGAGCCAATTCTTGATTCCTTG	This study
Establishment of hpnc4160-expressing *H. pylori*	pHel2 Plasmid construct for hpnc4160 without HPnc4170 N terminal region	pHel2-4160-de-4170-hed-f XhoI	GGGGGCTCGAGCAATTAAGGTAAGAGCGTCAGG	pHel2-4160-de-4170-hed-r BamHI	CCCCCGGATCCATCCCATCAATAGGCTTAAATTGT	This study
Construction of *H. pylori* GST-RNase III-expressing *E. coli*	pGEX6P-1-Rnase III	pGEX-6P-1 RNaseIII XhoI-f	AAAAGTCGACTCATGAAAAACAAACGCTCTCAAAATAGC	pGEX-6P-1 RNaseIII NotI-r	AAAAGCGGCCGCTCATTTGGCTTCCTTCAGTTTTTGAAGC	This study
Cloning into pBluescript plasmid for EMSA	HPnc4160	Small RNA HPnc4160 XhoI	AAAAAACTCGAGGGCGAAGGTAAGAGCGAAAG	Small RNA HPnc4160 EcoRI	AAAAAAGAATTCCAATTAAGGTAAGAGCGTCAGGA	This study
	5'UTR of hpaA	HP0410 150bp XhoI	AAAAAACTCGAGGATTGAATGGGAATGATTATCTTAAAAAATAGCG	HP0410 150bp EcoRI	AAAAAAGAATTCCCACTCGCTAATAGCGATCCTA	This study
	5'UTR of hopE	HELPY_0660 150bp XhoI	AAAAAACTCGAGTCTTATTCATAAGAGTTAAAAATATTTGCTTGATT	HELPY_0660 150bp EcoRI	AAAAAAGAATTCGAAGAGCTTAAAACTGCGGATAG	This study
	5'UTR of omp14	HP0671 150bp XhoI	AAAAAACTCGAGTTTTAAGTTTTAGGGGTGTTTTCTTTAAGA	HP0671 150bp EcoRI	AAAAAAGAATTCCCTAACACTACCGATAAACAGAGC	This study
	5'UTR of hofC	HP0486 150bp XhoI	AAAAAACTCGAGAGAAAAATTAGCTTGATTTTAAACTAATTCTATATTCT	HP0486 150bp EcoRI	AAAAAAGAATTCGTAAAACGCTTTAGCAATGTAGC	This study
	5'UTR of horB	HPSH_00635 150bp XhoI	AAAAAACTCGAGATTCTCATTCAATCGCTATATTTAATCAAAAAG	HPSH_00635 150bp EcoRI	AAAAAAGAATTCAAGCTTGTCATTGCTAGAGAGA	This study
	5'UTR of cagA	HPP12_0555 150bp XhoI	AAAAAACTCGAGAAGACATGAATTGATTACTCAAGTGTGT	HPP12_0555 150bp EcoRI	AAAAAAGAATTCTTAAAAGCCGCTTCGGTTTGT	This study
	5'UTR of HP1227	HP1227 150bp XhoI	AAAAACCGCTCGAGATGAATGTATAAAATCATATTTATTTCTCTCATTTTCACT	HP1227 150bp EcoRI	TTTTTCCGGAATTCACGCTAAAACGCCTAAAGC	This study
	5'UTR of HELPY_1262	HELPY_1262 150bp XhoI	AAAAACCGCTCGAGCTAATGGGGATTTTTTGTATTATACCTTAAAACAG	HELPY_1262 150bp EcoRI	AAAAACCGGAATTCCTAACGCCCAAAAGGGTGAA	This study
	CDS of cagA	CagA-B coding XhoI	AAAAACCGCTCGAGCAATCAAGCAGCAAGTGGT	CagA-B coding EcoRI	AAAAACCGGAATTCGCTTCTGATACCGCTTGACT	This study
PCR amplification of DNA fragments for EMSA	NB-cagA fragments	T7 promoter CagA-NB EMSA PCR s	GACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGACCGACTCAATCAAGCAGC	T7 promoter CagA-NB EMSA PCR as	GCTTCTGATACCGCTTGACTG	This study
	all kinds of DNA	T7 promoter EMSA PCR s	GACGTTGTAAAACGACGGC	T7 promoter EMSA PCR as	CCGGGCTGCAGGAATTC	This study