bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Increased *H. pylori* stool shedding and EPIYA-D *cagA* alleles are associated with gastric cancer
2 in an East Asian hospital
3
4 Short title: *H. pylori* and *cagA* quantification in stomach and stool
5
6 Sarah Talarico[1], Christina Leverich[1], Bing Wei[2], Jie Ma[2], XinGuang Cao[2], YongJun Guo[2],
7 GuangSen Han[2], Lena Yao[3], Steve Self[3], Yuzhou Zhao[2*], Nina R. Salama[1*]
8
9 [1]Human Biology Division Fred Hutchinson Cancer Research Center, Seattle, WA, USA
10 [2]The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China
11 [3]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
12
13
14 [*]Corresponding Author:
15 Email: nsalama@fredhutch.org (NRS)
16 Email:yuzhouzhao@126.com (YZ)
17

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

18 **ABSTRACT**

19 **Background:** *Helicobacter pylori* infection induces chronic inflammation and tissue damage in
20 the stomach, increasing risk for gastric cancer. Paradoxically, these tissue alterations may
21 promote loss of *H. pylori* infection during cancer progression. *H. pylori*'s role in cancer
22 progression beyond initiation is unclear. Geographic variation in gastric cancer risk has been
23 attributed to variation in carriage and type of the *H. pylori* oncogene *cagA.* **Methods:** To
24 investigate possible differences in *H. pylori* load in the stomach and shedding in stool, *H. pylori*
25 load and *cagA* genotype were assessed using droplet digital PCR assays on gastric mucosa
26 and stool samples from 49 urea breath test-positive individuals, including 25 gastric cancer and
27 24 non-cancer subjects at Henan Cancer Hospital, Henan, China. **Results:** Quantitation of *H.*
28 *pylori* DNA indicated similar gastric loads among cancer and non-cancer cases, but the gastric
29 cancer group had a median *H. pylori* load in the stool that was six times higher than that of the
30 non-cancer subjects. While the *cagA* gene was uniformly present among study subjects, only
31 70% had the East Asian *cagA* allele, which was significantly associated with gastric cancer
32 (Fisher's Exact Test, p = 0.03). **Conclusion:** *H. pylori* persists in a subset of gastric cancer
33 cases and thus may contribute to cancer progression. In this East Asian population with a high
34 prevalence of the *cagA* gene, the East Asian allele could still provide a marker for gastric cancer
35 risk. **Impact:** This study contributes to our understanding of *H. pylori* dynamics in the context of
36 pathological changes.

37

38 **INTRODUCTION**

39 The bacterial pathogen *Helicobacter pylori* infects the human stomach causing chronic
40 inflammation of the gastric mucosa. Most infected individuals remain asymptomatic but a subset
41 develop peptic ulcers, gastric adenocarcinomas, and mucosa-associated lymphoid tissue
42 (MALT) lymphoma resulting from the infection. Disease risk depends on the severity and
43 distribution of inflammation in the stomach. Those with inflammation that is predominantly in the

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

44 antrum of the stomach are at higher risk for duodenal ulcer and those with inflammation that
45 extends into the stomach corpus are at higher risk for gastric adenocarcinoma (1). In those who
46 eventually develop gastric adenocarcinoma, *H. pylori*-induced chronic inflammation initiates a
47 pathological cascade that progresses from atrophy to intestinal metaplasia and dysplasia (2).
48 Density of *H. pylori* infection has been found to be correlated with level of inflammation (3, 4) but
49 loss of *H. pylori* infection has been observed in atrophic gastric tissue (5) and increased pH in
50 the stomach can promote growth of other bacteria (6). Notably, *H. pylori* often cannot be
51 detected within tumors and the association of *H. pylori* with cancer was strongest when *H. pylori*
52 infection was assessed 10 years prior to cancer diagnosis (7).

53 Bacterial genetic factors contribute in part to differences in bacterial density,
54 inflammation, and disease development. The strain-variable *cagA* gene is associated with a
55 higher risk of gastric adenocarcinoma development (8). The *cagA* gene is located within the *cag*
56 pathogenicity island that encodes a Type IV secretion system which delivers the CagA effector
57 protein into host gastric epithelial cells (9). Inside the host cell, the CagA protein is
58 phosphorylated at EPIYA sites and the phosphorylated form is able to deregulate normal
59 cellular signaling (10-13). The *cagA* gene is grouped into two different allele types, Western
60 alleles that encode an EPIYA-C motif and East Asian alleles that encode an EPIYA-D motif.
61 *cagA* alleles having an EPIYA-D motif are predominantly found in *H. pylori* strains circulating in
62 East Asian countries and are associated with an increased risk of gastric cancer development
63 compared to *cagA* alleles having an EPIYA-C motif (14).

64 Studies investigating the role of the *cagA* gene in density of *H. pylori* infection in the
65 stomach have produced conflicting results, with some studies reporting a significantly higher
66 density of *H. pylori* infection in those with *cagA*-positive *H. pylori* (3, 15) and some finding no
67 difference in bacterial density between *cagA*-positive and *cagA*-negative strains (4, 16, 17). We
68 recently reported development of a new, non-invasive method for detection, quantification, and
69 *cagA* genotyping of *H. pylori* from stool samples that uses droplet digital PCR (ddPCR). We

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

70 tested this method using a collection of matched serum and stool samples from volunteers in
71 Costa Rica and observed a two log range of *H. pylori* loads in the stool. Furthermore, the *H.*
72 *pylori* load in the stool was significantly higher in those with a *cagA*-positive strain (18).
73 However, the extent to which the *H. pylori* load in the stool reflects the load in the stomach is not
74 clear.
75 To assess differences in *H. pylori* carriage and colonization density upon cancer
76 development and to examine factors that contribute to the observed variability in *H. pylori* loads
77 in stool, including *H. pylori* load in the stomach and *cagA* status, we applied the quantitative
78 ddPCR assays to gastric mucosal brushing samples and stool samples collected from gastric
79 cancer and non-cancer subjects from China.
80
81 **MATERIALS AND METHODS**
82 ***Study Populations and Specimen Collection***
83 To investigate possible differences in *H. pylori* carriage and load in the stomach as well as *H.*
84 *pylori* shedding in stool, C14 urea breath test (UBT, Haidewei HUBT-01P) was used to screen
85 individuals to be treated for gastric cancer or undergoing upper GI endoscopy for other
86 indications at Henan Cancer Hospital between 2015.10.27-2016.03.15. Subjects who were *H.*
87 *pylori*-positive by UBT were offered participation in this research study and all participants
88 provided written informed consent prior to participation. Participating subjects had gastric
89 mucosal brushings collected and were asked to fill out a questionnaire that covered
90 demographic characteristics, medical conditions, and medications and provide a stool sample.
91 All procedures were approved by the Henan Cancer Hospital Medical Research Institution
92 Review Board (doc # 2016oct005).
93 In total, samples were collected from 50 subjects including 25 gastric cancer subjects diagnosed
94 by pathology who were undergoing gastric resection surgery and 25 non-cancer subjects with
95 no history of gastric tumor or surgery who were undergoing upper GI endoscopy either as

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

96 indicated because of symptoms or as asymptomatic volunteers. All non-cancer cases showed
97 histologic evidence of gastritis and no ulcers. All 50 subjects had not taken antibiotics for gastric
98 disease in the past month.
99 Gastric mucosal brushing samples were collected from two different anatomical sites in
100 the stomach, the antrum and the corpus, using cytology brushes (Puritan Medical Products Co
101 LLC). Gastric mucosal brushings were collected and stored in a cryovial containing 1 ml flow
102 media (minimal essential media plus 10% DMSO, 5% fetal calf serum, 5mM HEPES (19).
103 Cryovials containing the gastric mucosal brushing samples were immediately placed and kept at
104 -80°C. The gastric mucosal brushings were collected from the gastric cancer subjects during the
105 gastric resection surgery and from non-cancer subjects during upper GI endoscopy. Stool
106 samples were collected 1-2 days before either surgery or endoscopy. Participants undergoing
107 surgery collected the stool sample at the hospital and participants undergoing endoscopy
108 collected the stool sample at home and delivered it to the hospital on the same day. Stool
109 samples were collected by the participants into a vial containing 5 ml RNAlater nucleic acid
110 preservative (Ambion) and were frozen at -20°C upon receipt. In addition to the gastric mucosal
111 brushings and stool samples, gastric biopsies were collected for histological analysis. A gastric
112 biopsy was collected from the antrum and the corpus. Gastric biopsies were formalin-fixed and
113 paraffin-embedded prior to histological analysis.
114 ***DNA Extraction***
115 For the gastric mucosal brushings, sample media and brush were transferred from the cryovial
116 to a microcentrifuge tube and centrifuged, followed by removal of the supernatant and brush.
117 DNA was then extracted from the pellet using the UltraClean Tissue and Cells DNA Isolation Kit
118 (MoBio) according to the manufacturer's instructions.
119 Stool samples were first transferred to a microcentrifuge tube and centrifuged to remove
120 the RNAlater. DNA was then extracted using the QIAamp Stool DNA Mini Kit (Qiagen)
121 according to the manufacturer's instructions, with the lysis step performed at 95°C. Stool DNA

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

122 concentration was measured using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo
123 Scientific), and the concentration was adjusted to 100 ng/µl.

124 ***Droplet Digital PCR***

125 *H. pylori*-specific droplet digital PCR (ddPCR) assays (*H. pylori* 16S assay, *cagA* detection
126 assay, and *cagA* EPIYA typing assay) were performed using the QX200 ddPCR System
127 (BioRad) as described previously (18) for stool DNA and gastric mucosal brushing DNA. Briefly,
128 each 20 µl reaction contained 1x ddPCR Supermix for Probes (BioRad), 900 nM of each primer,
129 250 nM of probes, and 10 µl DNA. Reactions were subject to thermal cycling with the following
130 conditions: 95°C for 10 minutes, 45 cycles of 94°C for 30 seconds and 55°C for 1 minute, and a
131 final incubation at 98°C for 10 minutes. Data were analyzed using the QuantaSoft software
132 version 1.6.6 (BioRad). The threshold was set to 4500 for the *H. pylori* 16S assay and 5500 for
133 the *cagA* detection assay. For the *cagA* EPIYA assay, the thresholds were set to 4000 (both
134 gastric and stool samples) for the FAM channel and 2500 (gastric samples) or 3000 (stool
135 samples) for the HEX channel. A positive control (stool DNA from an *H. pylori*-positive
136 volunteer) and negative control (molecular grade water) were included in each batch of samples
137 analyzed.

138 For the gastric brushing samples, the copy number of *H. pylori* 16S was normalized to
139 the copies of human DNA in the sample, as measured by a ddPCR assay for 18S. The following
140 primers and probe were used in the 18S assay: 18SFor (5' - CGATGCTCTTAGCTGAGTG - 3'),
141 18SRev (5' - CTTAATCATGGCCTCAGTTC - 3'), and 18S_HEX (5' hexachloro-fluorescein -
142 CCGCAGCTAGGAATAATGGAATAG – black hole quencher 3'). Reactions were subject to
143 thermal cycling with the following conditions: 95°C for 10 minutes, 45 cycles of 94°C for 30
144 seconds and 59°C for 1 minute, and a final incubation at 98°C for 10 minutes. Data were
145 analyzed using the QuantaSoft software version 1.6.6 (BioRad) and the threshold was set to
146 2000.

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

147 The concentration of the gastric mucosal brushing DNA sample was adjusted so that the
148 copy number of the ddPCR assay target was within the dynamic range of the assay. DNA
149 samples were run as either undiluted or diluted 1:10, 1:100, 1:1000, or 1:10,000. Stool DNA
150 concentration was adjusted to 100 ng/μl as needed so that no more than 1 μg stool DNA was
151 added to each ddPCR reaction.

152 ***Histological Analysis***

153 For histological analysis of the gastric biopsies, slides were made from formalin-fixed paraffin-
154 embedded (FFPE) gastric tissue samples. Slides were stained with Wright-Giemsa stain (Baso
155 Diagnostics, Inc. Zhuhai, China) according to the manufacturer's instructions. Stained slides
156 were analyzed and scored as either normal (0), mild (1), moderate (2), or marked (3) according
157 to the Sydney Classification System for *H. pylori* density, neutrophils, mononuclear cells,
158 atrophy, and intestinal metaplasia (20). For the non-cancer subjects, more than one anatomical
159 site of the stomach was analyzed for a patient (i.e. antrum and corpus) and the *H. pylori* density,
160 neutrophils, mononuclear cells, and intestinal metaplasia was scored as the average of the two
161 sites. Atrophy was scored separately for the antrum and corpus. For the gastric cancer subjects,
162 only gastric tissue from the carcinoma adjacent site was scored.

163 ***Statistical Analysis***

164 Correlation between gastric cancer and *cagA* allele was assessed using Fisher's Exact Test.
165 The Wilcoxon Two-Sample Exact Test was used for comparisons of *H. pylori* loads between
166 gastric cancer subjects and non-cancer subjects and between East Asian *cagA* allele and
167 Western *cagA* allele groups. Correlation between *H. pylori* load in the gastric mucosal brushing
168 samples from different anatomic sites and with *H. pylori* load in the stool was analyzed using
169 Spearman Correlation Coefficients with Fisher's z-Transformation. All statistical analyses were
170 performed in SAS version 9.4 (SAS Institute, Inc.).

171

172 **RESULTS**

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

173 ***Gastric Cancer cases show frequent evidence of active H. pylori infection***

174 Of 58 gastric cancer subjects screened for *H. pylori* infection by UBT, 25 (43%) were *H. pylori*-
175 positive. Of 108 non-cancer subjects screened, 25 (23%) were *H. pylori*-positive by UBT. One of
176 the 25 participating non-cancer subjects did not provide an adequate stool sample and was
177 excluded from the analysis. The remaining 49 subjects included in our study had a median age
178 of 53 years (range 27 to 76 years) and included 15 (31%) females and 34 (69%) males. Of the
179 25 gastric cancer subjects, 22 (88%) were male compared to 12 (50%) of the 24 non-cancer
180 cases. The gastric cancer cases had a median age of 59 years (range 46 – 76 years) and the
181 non-cancer cases had a median age of 49 years (range 27 – 66 years).

182 Of the 49 subjects, histological slides of gastric biopsy were available for all 25 of the
183 gastric cancer subjects and 17 of the non-cancer subjects consented to provide samples for
184 histologic scoring of inflammation, pathological changes, and *H. pylori* density by the Sydney
185 Classification System (20). Level of neutrophil and mononuclear cell infiltration, indicating acute
186 and chronic inflammation respectively, tended to be lower in the gastric cancer subjects, but
187 distribution of scores for atrophy, intestinal metaplasia, and *H. pylori* density were similar
188 between gastric cancer subjects and non-cancer subjects (Table 1). Of the 42 subjects with
189 histology slides analyzed, 13 (31%) had signs of atrophy and 18 (43%) had signs of intestinal
190 metaplasia. Despite all subjects being positive for *H. pylori* infection by UBT, three subjects (two
191 cancer cases and one non-cancer subject) did not have *H. pylori* detected by histology.

192

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

193 Table 1. Sydney Classification System scores for gastric histology slides from 25 gastric cancer
194 subjects and 17 non-cancer subjects from Henan Cancer Hospital, China

| | Score[a] | No. (%) Gastric Cancer Subjects (n = 25) | No. (%) Non-cancer Subjects (n = 17) |
|---|---|---|---|
| Neutrophils | 0 | 10 (40) | 2 (12) |
| | 1 | 9 (36) | 11 (65) |
| | 2 | 6 (24) | 4 (24) |
| | 3 | 0 | 0 |
| Mononuclear Cells | 0 | 0 | 0 |
| | 1 | 16 (64) | 4 (24) |
| | 2 | 7 (28) | 5 (29) |
| | 3 | 2 (8) | 8 (47) |
| Atrophy | 0 | 17 (68) | 12 (71) |
| | 1 | 7 (28) | 3 (18) |
| | 2 | 1 (4) | 2 (12) |
| | 3 | 0 | 0 |
| Intestinal Metaplasia | 0 | 13 (52) | 11 (65) |
| | 1 | 8 (32) | 5 (29) |
| | 2 | 3 (12) | 1 (6) |
| | 3 | 1 (4) | 0 |
| *H. pylori* | 0 | 2 (8) | 1 (6) |
| | 1 | 13 (52) | 7 (41) |
| | 2 | 6 (24) | 6 (35) |
| | 3 | 4 (16) | 3 (18) |

195 [a] 0 = normal, 1 = mild, 2 = moderate, 3 = marked; score is for the carcinoma adjacent tissue of
196 gastric cancer subjects and is the average of the antrum and corpus for non-cancer subjects

197

198 ***Cancer and non-cancer cases show similar loads of* H. pylori *in the stomach by ddPCR***

199 To investigate the load of *H. pylori* in the stomach among subjects, we employed ddPCR to
200 measure the copy number of the *H. pylori* 16S ribosomal RNA gene, a highly conserved gene

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

201 present in all *H. pylori* strains, in total DNA extracted from mucosal brush samples from both the
202 antrum and corpus for each subject. *H. pylori* load was normalized to the human 18S ribosomal
203 RNA gene copy number present in each sample. The *H. pylori* 16S gene was detected in one or
204 both gastric mucosal brushing samples of 22 (88%) of 25 gastric cancer subjects and all 24
205 (100%) non-cancer subjects. One non-cancer subject had *H. pylori* 16S detected in the sample
206 from the corpus but not the antrum. Of the three subjects that did not have *H. pylori* detected by
207 histology, one also did not have the *H. pylori* 16S gene detected by ddPCR in either gastric
208 sample. The other two had low loads of *H. pylori* 16S detected by ddPCR (gastric *H. pylori* load
209 of 10 and 0.4 16S copies per 1000 18S). As shown in Figure 1A, the load of *H. pylori* detected
210 in gastric samples varied over several logs in both cancer and non-cancer subjects. For the
211 gastric samples with detectable *H. pylori* 16S, the *H. pylori* 16S copy number per 1000 18S
212 ranged from 0.003 to 1034 with a median of 144 for the antrum and ranged from 0.84 to 1967
213 with a median of 103 for the corpus. The *H. pylori* load in the antrum and corpus were
214 significantly correlated (Spearman Correlation Estimate = 0.7, 95% CI: 0.5 – 0.8, Figure 1B) and
215 for subsequent analysis we considered the gastric load to be the average of that measured for
216 the antrum and corpus sample for each subject. The gastric *H. pylori* load between the gastric
217 cancer cases (median=139 *H. pylori* 16S copies per 1000 18S copies, range 0.9 – 386) and the
218 non-cancer subjects (median=114 *H. pylori* 16S copies per 1000 18S copies, range 0.4 - 1150)
219 was not significantly different (p=0.4, Wilcoxon Exact Test).
220
221 **Figure 1**. Gastric cancer cases and non-cancer subjects show similar variation in *H. pylori* load
222 within the stomach and high concordance between antrum and corpus samples. **A.** Copy
223 number of the *H. pylori* 16S ribosomal RNA genes was measured by ddPCR from gastric
224 brushes collected from the stomach antrum or corpus and normalized to copy number of the
225 human 18S ribosomal RNA gene. Cancer cases n=25, non-cancer subjects n=24. Box extends
226 from 25$^{th}$ to 75$^{th}$ percentiles with a line at the median value. Whiskers show minimum and

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

227 maximum values. In three cancer cases, no *H. pylori* DNA was detected and in one non-cancer
228 subject, *H. pylori* was only detected in the corpus. These points are plotted on the X-axis. **B.**
229 Correlation between corpus and antrum load (copy number of *H. pylori* 16S ribosomal gene) for
230 subjects with *H. pylori* DNA detected in both samples. Cancer cases black, n=22, non-cancer
231 subjects grey, n=23. Spearman Correlation Estimate = 0.7 (95% CI: 0.5 – 0.8). **C.** Gastric load
232 (average of corpus and antrum) in subjects with (IM+, n=18) or without (IM-, n=24) histologic
233 evidence of intestinal metaplasia (IM). Biopsy samples for histology were not available for 7
234 non-cancer cases. Points plotted on the X-axis represent subjects for which no *H. pylori* 16S
235 rRNA gene was detected.
236
237 Loss of acid secreting parietal cells and transformation of the gastric glands to an
238 intestinal metaplasia phenotype as consequence of *H. pylori* induced inflammation has been
239 suggested to favor outgrowth of other bacterial species over *H. pylori*. We saw similar
240 proportions of intestinal metaplasia in cancer cases compared to non-cancer subjects (Table1).
241 We thus checked whether gastric *H. pylori* load differed according to presence of intestinal
242 metaplasia (Fig. 1C), but saw no significant difference (p=0.3, Wilcoxon Exact Test).
243
244 ***Stool-based ddPCR detection of* H. pylori 16S *is sensitive, but shows little correlation***
245 ***with gastric load***
246 We next examined the relationship between detection and copy number of the *H. pylori*
247 16S ribosomal RNA gene by ddPCR in stool with detection and copy number in the stomach. Of
248 the stool samples from gastric cancer cases, the *H. pylori* 16S gene was detected in 21 (84% of
249 the 25 gastric cancer cases). Considering the 22 gastric cancer cases for which we detected the
250 *H. pylori 16S* gene in the stomach, 21 (95%) also had detection in the stool. The *H. pylori 16S*
251 gene was detected in 22 (92%) of the 24 stool samples from non-cancer subjects. For the stool
252 samples with detectable *H. pylori 16S*, the *H. pylori 16S* copy number per µg stool DNA ranged

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

253 from 2 to 1080 with a median of 26 (Figure 2A). The *H. pylori* load in the stool was significantly
254 higher (Wilcoxon Exact Test, p= 0.03) in the gastric cancer cases (median=46 *H.pylori* 16S
255 copies per µg stool DNA, range 0 - 560) compared to the non-cancer subjects (median=7.5
256 *H.pylori* 16S copies per µg stool DNA, range 0 - 1080). The *H. pylori* load observed in the stool
257 was only weakly correlated with the *H. pylori* load in the gastric samples of the same subject
258 (Spearman Correlation Estimate = 0.3, 95% CI:0.03 – 0.6, Figure 2B).
259
260 **Figure 2.** *H. pylori* DNA is detected in stool from both cancer and non-cancer subjects. **A.** Copy
261 number of *H. pylori* 16S DNA was measured by ddPCR of total stool DNA and normalized to µg
262 stool DNA. Box extends from 25$^{th}$ to 75$^{th}$ percentiles with a line at the median value. Whiskers
263 show minimum and maximum values. Cancer cases n=25, non-cancer cases n=24. **B.** Copy
264 number of *H. pylori* 16S DNA was measured by ddPCR and normalized to µg total stool DNA
265 (stool load) or copy number of human 18S (gastric load). Gastric load calculated as the average
266 of the corpus and antrum samples for each subject. Cancer cases n=22, non-cancer cases
267 n=24. One cancer and two non-cancer subjects had no *H. pylori* DNA detected in stool and are
268 not plotted. Spearman Correlation Estimate = 0.3, 95% CI: 0.03 – 0.6.
269
270 ***ddPCR reveals variation in* cagA *alleles and association of East Asian cagA with gastric***
271 ***cancer***
272 Previous studies suggest very high prevalence of *cagA*-positive strains and a predominance of
273 East Asian EPIYA alleles in China (21). To assess the presence of the *cagA* gene and
274 distinguish between East Asian and Western alleles, we performed our *cagA* detection and
275 *cagA* EPIYA typing ddPCR assays (18) on the antrum and corpus samples. Of the 46 subjects
276 with detectable *H. pylori* 16S, all 46 were positive for the *cagA* gene by one or both assays. Of
277 these 46 subjects, the *cagA* EPIYA typing ddPCR classified 37 (80%) as having either an East
278 Asian or Western *cagA* allele. PCR amplification and sequencing of a portion of the *cagA* gene

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

279 was performed for seven antrum samples that were negative by the *cagA* EPIYA typing assay.
280 Of these seven samples, five were determined by sequencing to have EPIYA-C (Western *cagA*
281 allele) and two had neither EPIYA-D nor EPIYA-C. The five that were negative for the *cagA*
282 EPIYA typing ddPCR assay but were determined to have EPIYA-C by sequencing all had the
283 same three nucleotide differences in the forward primer region that could explain failure of the
284 assay for these samples. One subject was determined to have an East Asian allele in the
285 corpus sample by the *cagA* EPIYA typing ddPCR assay and an allele having neither an EPIYA-
286 D nor an EPIYA-C in the antrum sample by sequencing.

287 Combining the results of the *cagA* EPIYA typing ddPCR assay and sequencing for the
288 46 subjects, 32 (70%) had the East Asian allele, 10 (22%) had a Western allele, one had only a
289 non-East Asian, non-Western allele, and three were not determined. Being infected with an *H.*
290 *pylori* strain with the East Asian *cagA* allele was significantly associated with gastric cancer
291 (Fisher's Exact Test, $p = 0.03$) (Table 2). Since we observed higher *H. pylori* loads in the stool
292 among gastric cancer cases, we examined the relationship between *H. pylori* load in the stool
293 and *cagA* allele. The median *H. pylori* stool load for subjects with an East Asian *cagA* allele was
294 33.5 *H. pylori* 16S copies per µg stool DNA (range 0 – 560) and for subjects with a Western
295 *cagA* allele was 15.5 *H. pylori* 16S copies per µg stool DNA (range 0 – 1080) (Wilcoxon Exact
296 Test $p = 0.26$).
297

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

298 Table 2. Comparison of *cagA* genotype between gastric cancer cases and non-cancer subjects
299

| | No. (%) Gastric Cancer cases n= 22 | No. (%) Non-cancer subjects n = 24 | p [a] |
|---|---|---|---|
| East Asian *cagA* | 20 (91) | 12 (50) | 0.03 |
| Western *cagA* | 2 (9) | 8 (33) | |
| Non-East Asian, non-Western *cagA* | 0 | 1 (4) | |
| Not determined | 0 | 3 (13) | |

300 [a] p-value for Fisher's Exact Test comparing presence of East Asian *cagA* allele between gastric
301 cancer and non-cancer cases
302

303 ***ddPCR-based* cagA*-detection in stool samples shows high specificity but lower***
304 ***sensitivity compared to stomach samples***

305 We extended our analysis of *cagA* genotypes by running the *cagA* EPIYA ddPCR typing assay
306 on stool samples from 48 subjects. For one endoscopy case, there was insufficient stool DNA to
307 run our *cagA* EPIYA ddPCR assay. We detected *cagA* in stool for 23/31 (73%) of subjects
308 where we detected EPIYA-D type *cagA* in the stomach and 4/5 (80%) of subjects where we
309 detected EPIYA-C type *cagA* in the stomach by the *cagA* EPIYA ddPCR assay. In all cases we
310 detected the same EPIYA type in the stomach and stool samples by the *cagA* EPIYA ddPCR
311 assay.
312

313 **DISCUSSION**

314 In this study, we leveraged recently developed ddPCR-based quantitative assays to explore
315 differences in *H. pylori* load present in the stomach and shed into stool among cancer and non-
316 cancer cases undergoing surgery or upper GI endoscopy respectively at a hospital in China's
317 Henan province, a region with high incidence of stomach cancer (38.13 per 100,000). At
318 present, all methods for detecting active *H. pylori* infection (UBT, histology, PCR, culture) have

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

319 limitations in specificity and sensitivity such that many studies require concordance between two
320 or more methods. We used the clinically accepted UBT to screen for cases with *H. pylori*
321 infection that were then analyzed by histology and ddPCR. Using UBT, our observations of 43%
322 and 23% prevalence of *H. pylori* infection in cancer and non-cancer cases falls at the low end of
323 that observed among symptomatic individuals in similar regions of China (30-80%) (22-24). Of
324 the UBT-positive cases enrolled in our study, 89% were positive by all three assays. While
325 tissue alterations associated with cancer progression have been suggested to disfavor *H. pylori*
326 colonization, we found similar loads in the stomach among cancer cases and non-cancer cases.
327 Additionally, histologic scores were not significantly different among cancer cases and non-
328 cancer cases. The high correlation between loads measured from different biopsy samples and
329 regions of the stomach within individuals suggests that patchy colonization within the stomach
330 may be the exception rather than the rule in this population. Furthermore, the continued
331 presence of *H. pylori* at high levels in the stomach may influence the rate or molecular
332 phenotype of gastric cancer in the subset of individuals that remain colonized.

333 In contrast to the similar loads observed in the stomach, gastric cancer cases showed
334 higher *H. pylori* loads in stool. Our previous population-based study in Costa Rica measured
335 higher *H. pylori* loads in stool among subjects positive for CagA by serology (18), prompting us
336 to assay *cagA* status. However, all 46 subjects that had *H. pylori* 16S detected in the gastric
337 samples also had *cagA* detected, so it was not possible to analyze the correlation between
338 presence of *cagA* and *H. pylori* load in the stool in this population. While the East Asian *cagA*
339 allele was significantly associated with gastric cancer, the allele was not associated with a
340 higher *H. pylori* load in the stool. This suggests that the higher *H. pylori* load in subjects with
341 gastric cancer is not dependent on presence of the *cagA* gene or East Asian *cagA* allele.
342 Further studies will be necessary to determine at what point during progression to gastric cancer
343 infected individuals have a higher *H. pylori* load in the stool and whether this could be used as a
344 marker either for presence of gastric cancer or risk for eventual gastric cancer development.

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

345 In this study in Henan, China, we found the *cagA* gene to be uniformly present among
346 study subjects but only 70% had the East Asian *cagA* allele, which was significantly associated
347 with gastric cancer. Previous studies examining the *cagA* gene and East Asian allele in Chinese
348 populations have found the prevalence of the *cagA* gene to be very high (75 – 90%) with the
349 East Asian allele being predominant (25, 26). While the *cagA* gene has been a good marker in
350 Western populations for cancer risk, its utility in East Asian populations has been less clear
351 since most strains have the *cagA* gene. In this population with a high prevalence of the *cagA*
352 virulence gene, the East Asian allele could still provide a marker for gastric cancer risk.

353 The limited correlation between *H. pylori* loads measured in the stomach compared to
354 the stool suggests shedding from the stomach depends on more than absolute abundance in
355 the stomach. Possible mechanisms include altered growth rate, induction of epithelial turnover,
356 and survival during transit through the lower GI tract. Further studies will need to be done to
357 investigate the factors that contribute to the observed higher *H. pylori* load in the stool among
358 gastric cancer subjects compared to non-cancer subjects despite a similar *H. pylori* load in the
359 stomach. This will further our understanding of how pathological changes associated with
360 gastric cancer development affect *H. pylori* colonization as well as what continued contribution
361 *H. pylori* has to gastric cancer beyond initiation.

362

363 **REFERENCES**

364 1. Kusters JG, van Vliet AH, Kuipers EJ. Pathogenesis of Helicobacter pylori infection. Clin Microbiol
365 Rev. 2006;19(3):449-90.
366 2. Correa P, Piazuelo MB. Natural history of Helicobacter pylori infection. Dig Liver Dis.
367 2008;40(7):490-6.
368 3. Atherton JC, Tham KT, Peek RM, Jr., Cover TL, Blaser MJ. Density of Helicobacter pylori infection
369 in vivo as assessed by quantitative culture and histology. The Journal of infectious diseases.
370 1996;174(3):552-6.
371 4. Yamaoka Y, Kodama T, Kita M, Imanishi J, Kashima K, Graham DY. Relation between clinical
372 presentation, Helicobacter pylori density, interleukin 1beta and 8 production, and cagA status. Gut.
373 1999;45(6):804-11.

bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

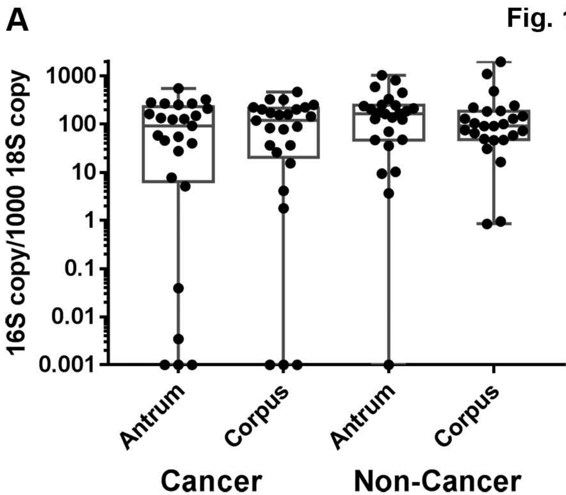
374 5. Karnes WE, Jr., Samloff IM, Siurala M, Kekki M, Sipponen P, Kim SW, et al. Positive serum
375 antibody and negative tissue staining for Helicobacter pylori in subjects with atrophic body gastritis.
376 Gastroenterology. 1991;101(1):167-74.
377 6. Wroblewski LE, Peek RM, Jr., Wilson KT. Helicobacter pylori and gastric cancer: factors that
378 modulate disease risk. Clin Microbiol Rev. 2010;23(4):713-39.
379 7. Helicobacter, Cancer Collaborative G. Gastric cancer and Helicobacter pylori: a combined
380 analysis of 12 case control studies nested within prospective cohorts. Gut. 2001;49(3):347-53.
381 8. Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, et al. Infection with
382 Helicobacter pylori strains possessing cagA is associated with an increased risk of developing
383 adenocarcinoma of the stomach. Cancer Res. 1995;55(10):2111-5.
384 9. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, et al. cag, a pathogenicity
385 island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors.
386 Proceedings of the National Academy of Sciences of the United States of America. 1996;93(25):14648-
387 53.
388 10. Segal ED, Cha J, Lo J, Falkow S, Tompkins LS. Altered states: involvement of phosphorylated CagA
389 in the induction of host cellular growth changes by Helicobacter pylori. Proceedings of the National
390 Academy of Sciences of the United States of America. 1999;96(25):14559-64.
391 11. Stein M, Bagnoli F, Halenbeck R, Rappuoli R, Fantl WJ, Covacci A. c-Src/Lyn kinases activate
392 Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs. Molecular microbiology.
393 2002;43(4):971-80.
394 12. Selbach M, Moese S, Hauck CR, Meyer TF, Backert S. Src is the kinase of the Helicobacter pylori
395 CagA protein in vitro and in vivo. The Journal of biological chemistry. 2002;277(9):6775-8.
396 13. Higashi H, Tsutsumi R, Muto S, Sugiyama T, Azuma T, Asaka M, et al. SHP-2 tyrosine phosphatase
397 as an intracellular target of Helicobacter pylori CagA protein. Science. 2002;295(5555):683-6.
398 14. Matsunari O, Shiota S, Suzuki R, Watada M, Kinjo N, Murakami K, et al. Association between
399 Helicobacter pylori virulence factors and gastroduodenal diseases in Okinawa, Japan. J Clin Microbiol.
400 2012;50(3):876-83.
401 15. Belda S, Saez J, Santibanez M, Rodriguez JC, Sola-Vera J, Ruiz-Garcia M, et al. Relationship
402 between bacterial load, morbidity and cagA gene in patients infected by Helicobacter pylori. Clin
403 Microbiol Infect. 2012;18(7):E251-3.
404 16. Warburton VJ, Everett S, Mapstone NP, Axon AT, Hawkey P, Dixon MF. Clinical and histological
405 associations of cagA and vacA genotypes in Helicobacter pylori gastritis. J Clin Pathol. 1998;51(1):55-61.
406 17. Twisk M, Kusters JG, Balk AG, Kuipers EJ, Loffeld RJ. Colonisation density and topographic
407 localisation of Helicobacter pylori do not depend on the cagA status. J Clin Pathol. 2001;54(10):771-3.
408 18. Talarico S, Safaeian M, Gonzalez P, Hildesheim A, Herrero R, Porras C, et al. Quantitative
409 Detection and Genotyping of Helicobacter pylori from Stool using Droplet Digital PCR Reveals Variation
410 in Bacterial Loads that Correlates with cagA Virulence Gene Carriage. Helicobacter. 2016;21(4):325-33.
411 19. Gall A, Fero J, McCoy C, Claywell BC, Sanchez CA, Blount PL, et al. Bacterial Composition of the
412 Human Upper Gastrointestinal Tract Microbiome Is Dynamic and Associated with Genomic Instability in
413 a Barrett's Esophagus Cohort. PLoS ONE. 2015;10(6):e0129055.
414 20. Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated
415 Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg
416 Pathol. 1996;20(10):1161-81.
417 21. Li Q, Liu J, Gong Y, Yuan Y. Association of CagA EPIYA-D or EPIYA-C phosphorylation sites with
418 peptic ulcer and gastric cancer risks: A meta-analysis. Medicine. 2017;96(17):e6620.
419 22. Aziz F, Chen X, Yang X, Yan Q. Prevalence and correlation with clinical diseases of Helicobacter
420 pylori cagA and vacA genotype among gastric patients from Northeast China. BioMed research
421 international. 2014;2014:142980.

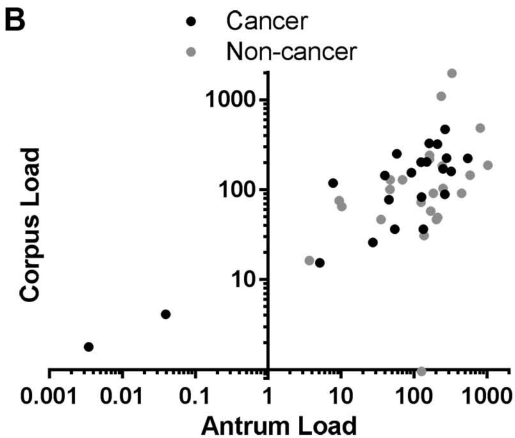
bioRxiv preprint doi: https://doi.org/10.1101/284448; this version posted March 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

422 23. Epplein M, Zheng W, Xiang YB, Peek RM, Jr., Li H, Correa P, et al. Prospective study of
423 Helicobacter pylori biomarkers for gastric cancer risk among Chinese men. Cancer Epidemiol Biomarkers
424 Prev. 2012;21(12):2185-92.
425 24. Gong YH, Wang Y, Yuan Y. Distribution of Helicobacter pylori in north China. World J
426 Gastroenterol. 2005;11(23):3523-7.
427 25. Chen CY, Wang FY, Wan HJ, Jin XX, Wei J, Wang ZK, et al. Amino acid polymorphisms flanking the
428 EPIYA-A motif of Helicobacter pylori CagA C-terminal region is associated with gastric cancer in east
429 China: experience from a single center. J Dig Dis. 2013;14(7):358-65.
430 26. Li J, Ou Z, Wang F, Guo Y, Zhang R, Zhang J, et al. Distinctiveness of the cagA genotype in
431 children and adults with peptic symptoms in South China. Helicobacter. 2009;14(4):248-55.

432

Fig. 1



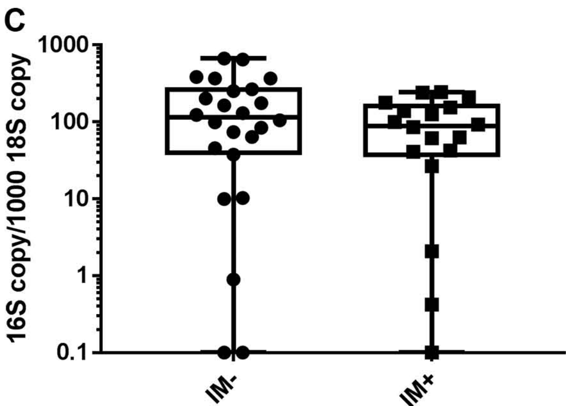



Fig. 2

A

16S copy/μg stool DNA

10000
1000
100
10
1
0.1

Cancer
Non-cancer

B

Stool Load (16S copy/μg stool DNA)

Gastric Load (16S copy/1000 18S copy)

Cancer
Non-cancer

10000
1000
100
10
1

0.1
1
10
100
1000
10000