



OPEN ACCESS

ORIGINAL ARTICLE

Pan-genomic analyses identify key *Helicobacter pylori* pathogenic loci modified by carcinogenic host microenvironments

Jennifer M Noto,¹ Abha Chopra,² John T Loh,³ Judith Romero-Gallo,¹ M Blanca Piazuolo,¹ Mark Watson,² Shay Leary,² Amber C Beckett,³ Keith T Wilson,^{1,3,4,5} Timothy L Cover,^{3,4,6} Simon Mallal,^{2,6} Dawn A Israel,¹ Richard M Peek^{1,3,5}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2017-313863>)

¹Department of Medicine, Division of Gastroenterology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

²Institute for Immunology and Infectious Diseases, Murdoch University, Murdoch, Australia

³Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

⁴Department of Veterans Affairs, Tennessee Valley Healthcare System, Nashville, Tennessee, USA

⁵Department of Cancer Biology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

⁶Department of Medicine, Division of Infectious Diseases, Vanderbilt University, Nashville, Tennessee, USA

Correspondence to

Dr. Richard M Peek, Department of Medicine, Division of Gastroenterology, Vanderbilt University Medical Center, Nashville, TN 37232, USA; richard.peek@vanderbilt.edu

Received 27 January 2017

Revised 30 June 2017

Accepted 15 July 2017

Published Online First

18 September 2017

ABSTRACT

Objective *Helicobacter pylori* is the strongest risk factor for gastric cancer; however, the majority of infected individuals do not develop disease. Pathological outcomes are mediated by complex interactions among bacterial, host and environmental constituents, and two dietary factors linked with gastric cancer risk are iron deficiency and high salt. We hypothesised that prolonged adaptation of *H. pylori* to in vivo carcinogenic microenvironments results in genetic modification important for disease.

Design Whole genome sequencing of genetically related *H. pylori* strains that differ in virulence and targeted *H. pylori* sequencing following prolonged exposure of bacteria to in vitro carcinogenic conditions were performed.

Results A total of 180 unique single nucleotide polymorphisms (SNPs) were identified among the collective genomes when compared with a reference *H. pylori* genome. Importantly, common SNPs were identified in isolates harvested from iron-depleted and high salt carcinogenic microenvironments, including an SNP within *fur* (FurR88H). To investigate the direct role of low iron and/or high salt, *H. pylori* was continuously cultured *in vitro* under low iron or high salt conditions to assess *fur* genetic variation. Exposure to low iron or high salt selected for the FurR88H variant after only 5 days. To extend these results, *fur* was sequenced in 339 clinical *H. pylori* strains. Among the isolates examined, 17% (40/232) of strains isolated from patients with premalignant lesions harboured the FurR88H variant, compared with only 6% (6/107) of strains from patients with non-atrophic gastritis alone ($p=0.0034$).

Conclusion These results indicate that specific genetic variation arises within *H. pylori* strains during in vivo adaptation to conditions conducive for gastric carcinogenesis.

INTRODUCTION

Helicobacter pylori persistently colonises the gastric mucosa of more than 50% of the world's population. *H. pylori* strains harbour remarkable genetic diversity as a species, as evidenced by variation of gene order and genetic content, the mosaic structure of genes and sequence diversity within conserved

genes.^{1–3} *H. pylori* exhibits a panmictic population structure, reflecting a high level of genetic exchange among strains,^{4 5} and recombination between strains can occur in vivo during naturally occurring mixed infections.³ Analyses of complete genomic sequences from a number of isolates indicate that *H. pylori* has incorporated genetic material from other organisms,^{4 5} and that variation occurs among single colony isolates of distinct *H. pylori* from an individual host.^{6 7}

In addition to being one of the most common bacterial infections worldwide, *H. pylori* is the strongest known risk factor for the development of gastric adenocarcinoma, and pathological outcomes are mediated by complex interactions among bacterial virulence determinants, host constituents and environmental factors. One *H. pylori* virulence determinant that increases gastric cancer risk is the *cag* pathogenicity island. *H. pylori* strains that harbour the *cag* island induce more severe gastric injury and further augment the risk for gastric cancer compared with *cag*-negative strains. The *cag* island encodes a bacterial type IV secretion system, which translocates the oncoprotein CagA into host cells to induce cellular morphological changes, proliferation and migration.^{2 8}

Two environmental factors associated with increased gastric carcinogenesis are high dietary salt intake and iron deficiency.⁹ Epidemiological studies have demonstrated a link between salt intake and increased gastric cancer risk,^{10 11} and high salt diets accelerate the development of gastric cancer in animal models of infection.^{12–14} Similarly, epidemiological studies have supported the concept that iron deficiency contributes to the risk of gastric cancer,^{15 16} and a recent study from our group demonstrated that iron depletion significantly increased the development of gastric adenocarcinoma in *H. pylori*-infected gerbils.¹⁷ Therefore, both high salt and low iron augment the severity of *H. pylori*-induced gastric inflammation and the development of premalignant and malignant lesions in rodent models of gastric cancer.

Several studies have investigated specific mechanisms through which iron deficiency or high salt may modify *H. pylori* and enhance the risk for gastric cancer. Iron deficiency and high salt-dependent



To cite: Noto JM, Chopra A, Loh JT, et al. *Gut* 2018;**67**:1793–1804.

Significance of this study

What is already known on this subject?

- ▶ *Helicobacter pylori* strains exhibit remarkable genetic diversity as a species and have a panmictic population structure, reflecting a high level of genetic exchange among strains.
- ▶ *H. pylori* is the strongest known risk factor for the development of gastric adenocarcinoma; however, the majority of *H. pylori*-infected individuals do not develop disease.
- ▶ Two environmental factors associated with increased gastric cancer risk are high dietary salt intake and iron deficiency. Both high salt and low iron significantly augment the severity of *H. pylori*-induced gastric inflammation and the development of premalignant and malignant lesions in rodent models of gastric cancer.

What are the new findings?

- ▶ Whole genome sequencing of *H. pylori* strains B128 and 7.13 and in vivo-adapted 7.13 output isolates harvested from animals maintained on iron-replete, iron-depleted, regular salt or high salt diets identified a small number of common polymorphisms that occurred within the context of iron deficiency and high salt intake.
- ▶ Specific cancer-related genetic loci were identified when *in vivo*-adapted carcinogenic *H. pylori* strain 7.13 was compared with its progenitor human *H. pylori* isolate B128, which does not cause cancer.
- ▶ Environmental-specific genetic loci were identified among *H. pylori* strains isolated from carcinogenic, iron-deficient or high salt, microenvironments.
- ▶ A single nucleotide polymorphism in *fur* was selected for by exposure to low iron or high salt conditions and was more frequently identified in clinical *H. pylori* strains from patients with premalignant lesions.

How might it impact clinical practice in the foreseeable future?

- ▶ Key *H. pylori* genetic loci that are susceptible to genetic alteration under *in vivo* environmental stress may serve as genetic biomarkers for *H. pylori* virulence potential and gastric cancer risk.

alterations of *H. pylori* per se include alterations in bacterial cell morphology, gene regulation and protein expression.^{17–26} Specifically, CagA expression and translocation into gastric epithelial cells are increased in response to both high salt and low iron conditions, and, under iron restriction, function of the *cag* type IV secretion system is enhanced.^{17 22 24} All of these modifications likely contribute to the increased risk of *H. pylori*-induced gastric cancer associated with these dietary alterations.^{14 17 22 24 26}

In this study, we tested the hypothesis that prolonged adaptation of *H. pylori* to an *in vivo* carcinogenic gastric microenvironment induced by iron deficiency or high salt would select for genetic variants important for pathogenesis. We previously identified a strain of *H. pylori*, 7.13, that reproducibly induces gastric cancer in rodent models of gastritis. This strain was derived via *in vivo* adaptation of a human clinical *H. pylori* strain, B128, which induces inflammation, but not cancer, in rodent gastric mucosa. Thus, *H. pylori* strains B128 and 7.13 are closely related genetically, but differ in oncogenic potential.²⁷ Whole genome sequence analyses were performed on *H. pylori* strains B128 and 7.13, and *in vivo*-adapted 7.13 output isolates exposed to iron-replete or iron-depleted conditions and

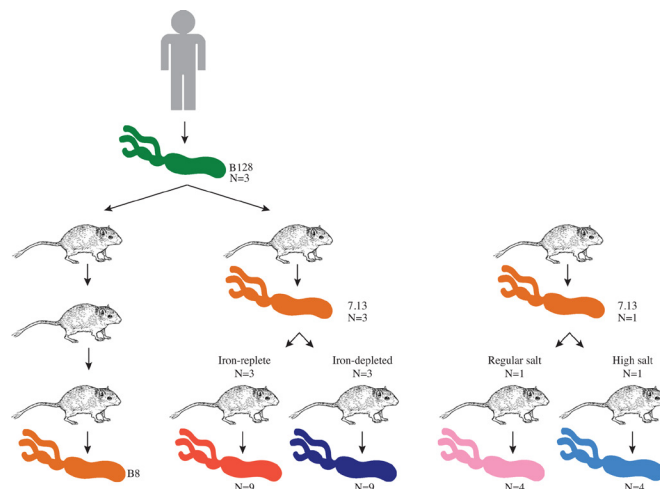


Figure 1 Experimental design. *Helicobacter pylori* strain B128 was originally isolated from a human patient.⁵⁴ Strain B128 (green) was used for orogastric challenge of Mongolian gerbils and was passed three times through gerbils in a previous study.²⁸ The *in vivo*-adapted output strain, designated *H. pylori* strain B8 (orange), was isolated from those gerbils, sequenced and used as the reference strain in this study.¹⁹ In an independent experiment, strain B128 (green) was also used for orogastric challenge of a single Mongolian gerbil, and the *in vivo*-adapted output strain isolated from this gerbil was designated *H. pylori* strain 7.13 (orange).²⁷ Carcinogenic *H. pylori* strain 7.13 was then used to infect another cohort of Mongolian gerbils maintained on either iron-replete or iron-depleted diets.¹⁷ Single colonies (n=3) from three independent gerbils maintained on iron-replete diet (n=9) and three independent gerbils maintained on iron-depleted diet (n=9) were subjected to whole genome sequence analysis. *In vivo*-adapted strain from iron-replete (red) and iron-depleted (blue) conditions was also compared with previously described *H. pylori* single colonies (n=4) harvested from one gerbil maintained on a regular salt (pink) or high salt (light blue) diet.²¹

indicate that targeted genetic variation in *H. pylori* arises during *in vivo* adaptation to conditions conducive for the development of gastric cancer.

RESULTS

***H. pylori* genetic alterations during *in vivo* adaptation**

We initially conducted whole genome sequence analyses of human progenitor *H. pylori* strain B128 and its *in vivo* gerbil-adapted carcinogenic derivative strain 7.13. Sequences were assembled using the genome sequence of *H. pylori* strain B8, a closely related strain for which its complete genome sequence is available, as the reference strain.²⁸ In addition to these strains, we also performed whole genome sequence analysis using *in vivo* gerbil-adapted *H. pylori* isolates exposed to iron-replete or iron-depleted conditions. Three single colony isolates from *H. pylori* strains B128 and 7.13 and three single colonies from 7.13-infected gerbils maintained on either iron-replete (n=3) or iron-depleted (n=3) diets were used for whole genome sequence analysis (figure 1). *In vivo*-adapted strains from iron-replete and iron-depleted conditions were also compared with *in vivo*-adapted strains isolated from regular salt or high salt conditions¹⁴ (figure 1).

Whole genome sequence analysis

The total length of the *H. pylori* reference strain B8 genome (1673 997bp) was completely covered, and the

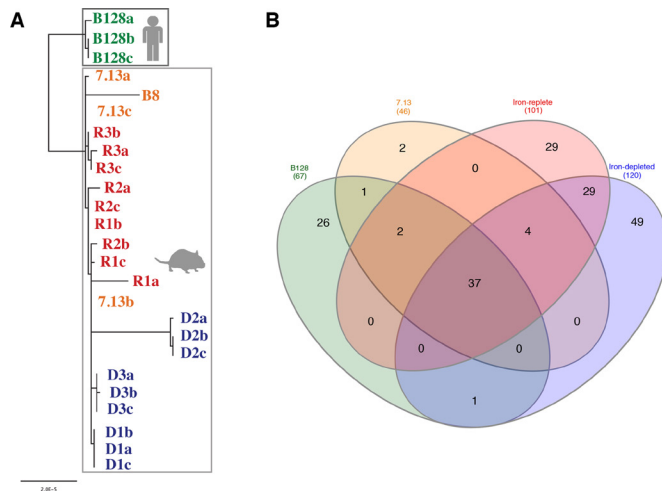


Figure 2 (A) Phylogenetic analyses of whole genome sequences from *Helicobacter pylori* strains B128 (green) and 7.13 (orange), and in vivo-adapted iron-replete (red) and iron-depleted (blue) strains. R1-3 and D1-3 designate independent gerbils maintained on either iron-replete ($n=3$) or iron-depleted ($n=3$), respectively, while a-c designate different single colonies isolated from each gerbil. *H. pylori* strains isolated from humans or gerbils are grouped within the grey boxes, respectively. The scale bar represents amino acid changes per site. (B) Venn diagram²⁰ showing the number of common and unique single nucleotide polymorphisms (SNPs) among human-adapted *H. pylori* strains B128 (green) and 7.13 (orange), and in vivo-adapted *H. pylori* strains isolated from gerbils maintained on iron-replete (red) and iron-depleted (blue) diets, all relative to the reference *H. pylori* strain B8. The number in the parentheses indicates the total number of unique SNPs identified in each strain and present exclusively in that strain, but not any other strain.

guanine-cytosine content was 38.8% for all sequenced strains. Overall sequence read count, average sequence read length, coverage and unmapped reads for *H. pylori* sequences are shown in online supplementary table S1. On average, the total read count was 645 132 bp, the mean read length was 263 bp, the fold coverage was 102 and the percentage of unmapped reads was 6.6% for all *H. pylori* strains analysed in this study.

Phylogenetic analysis confirmed that human *H. pylori* B128 isolates are more closely related to each other than to *H. pylori* strains adapted to gerbil stomachs (figure 2A). Consistent with in vivo adaptation within the same host, the reference *H. pylori* strain B8, which is also gerbil adapted, is more closely related to the gerbil-adapted strain 7.13 than to its parental human-adapted strain B128 (figure 2A). Furthermore, *H. pylori* strains isolated from gerbils under iron-replete conditions are more closely related to each other than to strains isolated from gerbils maintained on iron-depleted diets (figure 2A). These findings demonstrate that in vivo adaptations to different hosts (human vs gerbil) and to different environmental conditions (iron-replete vs iron-depleted) induce *H. pylori* genetic changes.

The number of unique single nucleotide polymorphisms (SNPs) among all sequenced *H. pylori* strains, relative to *H. pylori* reference strain B8, is shown in figure 2B.²⁹ The human-adapted strain B128 had 26 unique SNPs (ie, SNPs present exclusively in strain B128, but not in any other strain) (figure 2B, online supplementary table S2). The gerbil-adapted strain 7.13 only had two unique SNPs (figure 2B, online supplementary table S2), highlighting the similarity among strains adapted to the gerbil host. However, subsequent *H. pylori* adaptation to

iron-replete (29 SNPs) or iron-depleted (49 SNPs) conditions resulted in the greatest number of unique SNPs (figure 2B, online supplementary table S2), reinforcing the importance of genetic adaptation to different environmental conditions.

When all of the sequences were analysed, a total of 180 genetic loci with single and multinucleotide polymorphisms were identified, including 33 within intergenic regions and 147 within coding regions (online supplementary table S3). When each single colony was analysed individually for total polymorphisms, there was variation among the single colonies (online supplementary tables S4 and S5). The *H. pylori* human isolate B128 had, on average, 98 polymorphisms, 12 insertions and 10 deletions. Of the polymorphisms within coding regions of strain B128, 67% resulted in a non-synonymous change (online supplementary table S4). Among the three single colonies, gerbil-adapted *H. pylori* strain 7.13 contained, on average, 47 polymorphisms, nine insertions and three deletions. Of the SNPs within coding regions of strain 7.13, 65% resulted in a non-synonymous change (online supplementary table S4). Similarly, among the nine single colonies harvested from three gerbils maintained on iron-depleted diets, there were, on average, 52 polymorphisms, 12 insertions and five deletions, while in vivo-adapted strains isolated from gerbils maintained on iron-replete diets had, on average, 43 polymorphisms, 11 insertions and three deletions. Of the SNPs within coding regions of the iron-depleted and iron-replete strains, 54% and 52% resulted in a non-synonymous change, respectively (online supplementary table S5). Based on the number of variants identified, most of the genetic variation arose through in vivo adaptation to a new gerbil host. However, all *H. pylori* isolates were found to be unique, including isolates from a single host.

Many of the genetic alterations resulting in deletions or insertions occurred within genetic loci important for *H. pylori* pathogenesis, including polymorphisms within loci regulating chemotaxis (*cheA*), adhesion (*hopZ*, *oipA*, *sabA*), survival (*kataA*), iron acquisition or active transport (*tonB*) and *cag* type IV secretion system function (*cagY*) (table 1).

Among the 180 polymorphisms identified, 72 were located within genes with predicted function and resulted in non-synonymous amino acid changes (table 2). Similar to the insertions and deletions, many of these variants were within genes encoding proteins that function in pathogenesis, including motility (*flil*, *fliF*), adhesion (*babA*, *babB*), iron acquisition or active transport (*tonB*) and *cag* IV secretion system function (*cagY*). When we compared unique SNPs within coding regions between in vivo-adapted strains isolated from gerbils maintained on iron-depleted versus iron-replete diets, we identified 35 polymorphisms, distributed among 15 genes, including *babB*, *cagY*, *pepP*, *oppD*, *hcpD*, *hcpE*, *prfB*, and various outer membrane and hypothetical proteins. Additional SNPs were present in both groups, but were more prevalent in iron-depleted versus iron-replete strains; these included SNPs within *babA*, *babB* and *fur*.

The SNP (FurR88H) located within *fur*, a gene encoding the ferric uptake regulator, was identified previously in in vivo-adapted *H. pylori* strains isolated from a gerbil maintained on a high salt diet.³⁰ Therefore, we compared the number of SNPs from in vivo-adapted strains isolated from gerbils maintained on iron-depleted diets with those previously isolated from a gerbil maintained on a high salt diet. Specifically, we sought to identify SNPs present in these isolates that were either not present or present at a lower frequency in strains isolated from control gerbils maintained on iron-replete or regular salt diets. We identified five SNPs distributed among four genes, including HPB8_593, HPB8_716 (*cagY*), HPB8_1139 and HPB8_1145

Table 1 Deletions and insertions in *Helicobacter pylori* strains B128 and 7.13, and in vivo-adapted strains isolated from gerbils maintained on iron-depleted or iron-replete diets

Deletions		B128 single colonies			7.13 single colonies			7.13 iron-depleted single colonies									7.13 iron-replete single colonies								
Location	Gene	HPB8 locus			D1			D2			D3			R1			R2			R3					
		a	b	c	D1 a	D1 b	C	D2 a	D2 b	D2 c	D3 a	D3 b	D3 c	R1 a	R1 b	R1 c	R2 a	R2 b	R2 c	R3 a	R3 b	R3 c			
88333	Intergenic																								
132884	<i>tonB1</i>		27*	44.3–46.2	26	29	31	29	26	31	33	38													
132889	<i>tonB1</i>		27.1–27.5																						
132892	<i>tonB1</i>		26																						
132907	<i>tonB1</i>		27*																						
132926	<i>tonB1</i>				85	90	79.7–81.4	80.6–81.6	77	92				75	76										
365607	Intergenic																								
378215	<i>galE</i>			96*																					
410076	<i>cheA</i>			61.0–74.5*																					
410076	<i>cheA</i>		63.8–71.3																						
410080	<i>cheA</i>			63*																					
693451	<i>cagY</i>																								
693465	<i>cagY</i>																								
819621	<i>olpA</i>																								
919030	<i>sabA</i>		92*	91*	87*																				
1065100	Intergenic			25																					
1098589	<i>hsdR5</i>																								
1171489	Intergenic																								
1197192	Intergenic																								
1316726	<i>radA</i>																								
1445365	<i>hrgA</i>																								
1492146	Intergenic																								
1533081	Intergenic																								
1533085	Intergenic																								
1533088	Intergenic																								
1539253	Intergenic																								
1584758	<i>hopZ</i>																								
1584759	<i>hopZ</i>																								
1585383	<i>hopZ</i>																								
1621918	Intergenic																								
1621918	Intergenic																								
1621919	Intergenic																								

Insertions		B128 single colonies			7.13 single colonies			7.13 iron-depleted single colonies									7.13 iron-replete single colonies								
Location	Gene	HPB8 locus			D1			D2			D3			R1			R2			R3					
		a	b	c	D1 a	D1 b	C	D2 a	D2 b	D2 c	D3 a	D3 b	D3 c	R1 a	R1 b	R1 c	R2 a	R2 b	R2 c	R3 a	R3 b	R3 c			
88333	Intergenic																								
88334	Intergenic																								
102469	Intergenic																								
132300	Intergenic																								

Continued

Table 1 Continued

Insertions		B128 single colonies			7.13 single colonies			7.13 iron-depleted single colonies			7.13 iron-replete single colonies																
Location	Gene	HPB8 locus	a	b	c	a	b	c	D1 a	D1 b	C	D2 a	D2 b	D2 c	D3 a	D3 b	D3 c	R1 a	R1 b	R1 c	R2 a	R2 b	R2 c	R3 a	R3 b	R3 c	
149625	Intergenic																					52					
210808	Intergenic																		50								
236057	Intergenic																										
293971	Intergenic		58	34		95	91		95	97	94	94	96	93	92	91	92		96	96	95	93	95	97	95	92	95
397298	<i>fucT1</i>		55	35	36																						
471641	Intergenic		100	95	97	81	97	93	95	92	92	93	98	93	95	94	95		90	94	95	94	93	94	85	93	91
607701	Intergenic										23																
628559	Intergenic		28	28																							
631022	Intergenic			85		69	82		77	84			85			91	93		89	89	94	81	79	100	84	76	78
631033	Intergenic			80		77	83		95	62	73	96	86	78	87	87	87		75	87	78	82	78	96	91	79	84
666812	<i>nuoT74</i>											94	93	89	89												
759720	Intergenic								66			66	65			74	72				74			63	63		70
819619	<i>olpA</i>			70																							
833169	<i>fucT7</i>		83	67	67																						
883216	<i>hyuA1</i>			59	58																						
883595	<i>hyuA1</i>		100	95	97																						
1064308	<i>katA</i>					100																					
1082193	Intergenic										92																
1197792	Intergenic		42																								
1197193	Intergenic			90		59	57	59	56	48	57	51	52	54	52	55	55		57	51	57	62			55	62	
1217501	Intergenic			59															44								
1332440	Intergenic			59					46	50			46	46	47	50	50		46				48			44	
1332612	Intergenic			79	69	83.7-84.8			75	80	77	71.6-72.5	74	63	77	79	74.6-75.6		81	87	80	84	73	85	66	73	75
1414196	Intergenic																										
1458196	Intergenic		56	66	60																						
1566403	Intergenic			43					38	36			37														
1566404	Intergenic		58	54	60	45	48	44	59	46	45	53	55	55	47	44		57	51	49	47	58	57	39	59	53	
1621919	Intergenic																										
1657270	Intergenic					66	74						65	80	62				73			61			72		
1657271	Intergenic					79	76		76	63	60	70			70	67	60		68		70		63	62	60	69	

D1-3 and R1-3 designate independent gerbils maintained on either iron-depleted (n=3) or iron-replete (n=3) diets, respectively, while a-c designate different single colonies isolated from each gerbil. Colour-coded regions within each single colony represent deletions or insertions that were present in each isolate, and the number in each box represents the percentage of reads that contained each single nucleotide polymorphism. When a percentage range is shown, this represents multinucleotide polymorphisms present.

*Deletion or insertion results in a frameshift mutation.

Table 2 Unique genetic loci with polymorphisms resulting in non-synonymous amino acid changes in *Helicobacter pylori* strains B128 and 7.13, and in vivo-adapted strains isolated from gerbils maintained on iron-depleted or iron-replete diets

Variant	B128 single colonies			7.13 iron-depleted single colonies												7.13 iron-replete single colonies											
	Location	Gene	HPB8 locus	a	b	c	D1a	D1b	D1c	D2a	D2b	D2c	D3a	D3b	D3c	R1a	R1b	R1c	R2a	R2b	R2c	R3a	R3b	R3c			
118890	<i>flhI</i>						K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R	K>R			
132888	<i>tonB1</i>			K>E																							
211182	<i>babB</i>				T>I	T>I	G>R	G>R								S>L	T>I	T>I									
211347	<i>babB</i>															T>I	G>R										
211415	<i>babB</i>				G>R	G>R																					
211583	<i>babB</i>						D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N			
211592	<i>babB</i>			K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E			
211690	<i>babB</i>			N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K			
211694	<i>babB</i>						E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K			
211838	<i>babB</i>			K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q			
211844	<i>babB</i>			G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S			
212150	<i>babB</i>						G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S			
212315	<i>babB</i>			G>S	G>S	G>S	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K			
263909	<i>rpsL</i>			E>K	E>K	E>K																					
263985	<i>rpsL</i>						R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K	R>K			
264366	<i>rpsG</i>						T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M	T>M			
264374	<i>rpsG</i>			D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N	D>N			
273162	Hypothetical						D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H			
273165	Hypothetical						E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K			
342359	<i>pal</i>			T>A	T>A	T>A																					
559618	Hypothetical			E>K	E>K	E>K																					
563546	Hypothetical																										
591318	Puative OMP			T>S	T>S	T>S																					
591626	Puative OMP																										
591635	Puative OMP																										
591638	Puative OMP																										
616637	DNA methyltransferase			D>N	D>N	D>N																					
629024	<i>babA</i>			S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L	S>L			
629189	<i>babA</i>			T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I	T>I			
629257	<i>babA</i>			G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R	G>R			
629425	<i>babA</i>			D>N	D>N	D>N	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H	D>H			
629434	<i>babA</i>			K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E	K>E			
629532	<i>babA</i>			N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K	N>K			
629536	<i>babA</i>			E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K	E>K			
629680	<i>babA</i>			K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q	K>Q			
629686	<i>babA</i>			G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S			
629992	<i>babA</i>			G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S			
630157	<i>babA</i>			G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S	G>S			
692170	<i>cagY</i>																										
692262	<i>cagY</i>																										
692887	<i>cagY</i>						R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R	R>R			
693460	<i>cagY</i>						V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A	V>A			
693465	<i>cagY</i>						R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E	R>E			

Continued

Table 2 Continued

Variant	B128 single colonies			7.13 single colonies			7.13 iron-depleted single colonies			7.13 iron-replete single colonies																		
	Location	Gene	HPB8 locus	a	b	c	a	b	c	D1 a	D1 b	D1 c	D2 a	D2 b	D2 c	D3 a	D3 b	D3 c	R1 a	R1 b	R1 c	R2 a	R2 b	R2 c	R3 a	R3 b	R3 c	
693471		<i>cagY</i>	716										V>I	V>I	V>I													
693483		<i>cagY</i>	716										L>K	L>K	L>K													
693497		<i>cagY</i>	716										S>Q	S>Q	S>Q													
693505		<i>cagY</i>	716										R>R	R>R	R>R													
693508		<i>cagY</i>	716										T>W	T>W	T>W													
693513		<i>cagY</i>	716										A>K	A>K	A>K													
727820		Hypothetical	750	L>W	L>W	L>W																						
798239		Glycosyltransferase	817	A>T	A>T	A>T																						
798986		Glycosyltransferase	818																									
846179		<i>mec</i>	865																									
913806		<i>meaE</i>	927	P>S	P>S	P>S																						
1084103		Putative OMP	1104																									
1084106		Putative OMP	1104																									
1084553		Putative OMP	1104																									
1084556		Putative OMP	1104																									
1084587		Putative OMP	1104																									
1122559		<i>fur</i>	1145																									
1130504		<i>pepP</i>	1155																									
1195655		<i>flfF</i>	1220	P>L	P>L	P>L																						
1221163		Polysaccharide deacetylase	1252	A>T	A>T	A>T																						
1283634		Conserved hypothetical OMP	1309																									
1289275		<i>oppD</i>	1313																									
1302387		<i>hcpE</i>	1328																									
1369139		<i>prfB</i>	1395																									
1377477		<i>hcpD</i>	1405																									
1405574		Hypothetical	1436																									
1533082		Predicted coding region	1561																									
1533085		Predicted coding region	1561																									
1533088		Predicted coding region	1561																									

D1-3 and R1-3 designate independent gerbils maintained on either iron-depleted (n=3) or iron-replete (n=3) diets, respectively, while a-c designate different single colonies isolated from each gerbil. Colour-coded regions represent non-synonymous amino acid changes that result from specific genetic polymorphisms. Specific amino acid changes are designated within each box.

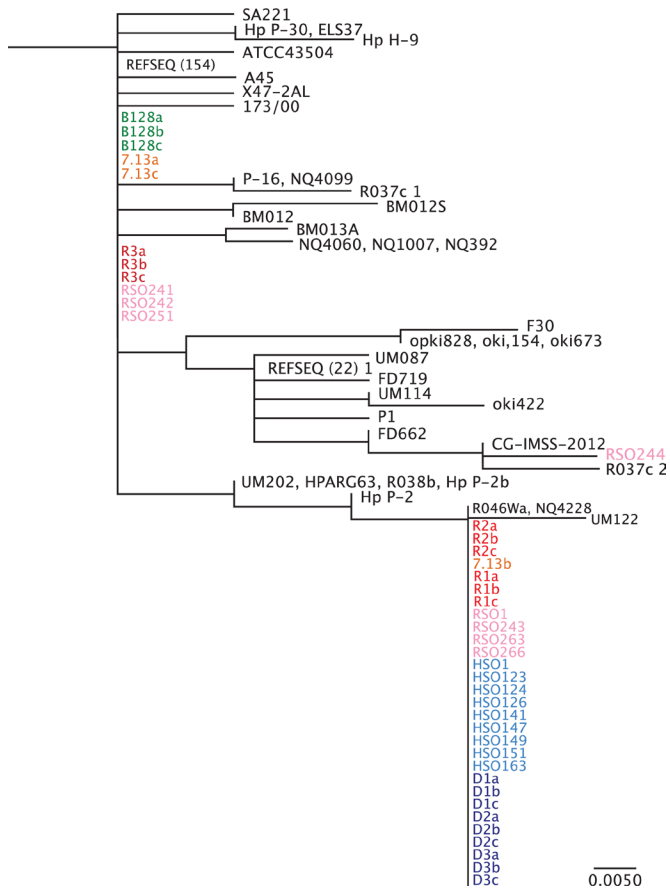


Figure 3 Phylogenetic analysis of Fur amino acid sequence. Phylogenetic analysis was performed for full-length Fur from all available *Helicobacter pylori* sequence data within the National Center for Biotechnology Information NCBI database. *H. pylori* single colonies from strains B128 (green) and 7.13 (orange), and in vivo-adapted strains isolated from gerbils maintained on iron-replete (red) and iron-depleted (blue) diets as well as in vivo-adapted strains isolated from gerbils maintained on regular salt (pink) and high salt (light blue) diets were included in these analyses. R1-3 and D1-3 designate independent gerbils maintained on either iron-replete (n=3) or iron-depleted (n=3), respectively, while a–c designate different single colonies isolated from each gerbil. HSO and RSO designate independent high salt output and regular salt output *H. pylori* strains, respectively. REFSEQ (#) represents the number of unique *H. pylori* sequences from different *H. pylori* strains that cluster under the same phylogeny. Although all Fur sequences from the NCBI database were used, this phylogenetic analysis shows only the branch to which the newly sequenced strains cluster. The scale bar represents amino acid changes per site.

(*fur*) (online supplementary table S6) in iron-depleted and high salt *H. pylori* in vivo-adapted output strains, suggesting that certain loci are subject to variation following in vivo adaptation to different host microenvironments.

Analysis of sequence variation in fur

To investigate mechanisms underpinning these changes in greater depth, phylogenetic analysis of Fur in *H. pylori* strains B128 and 7.13, and in vivo-adapted strains isolated from gerbils maintained on iron-replete, iron-depleted, high salt or regular salt diets was performed (figure 3). In vivo-adapted strains isolated from gerbils maintained on iron-depleted or high salt diets exhibited a high degree of identity within Fur, and segregated

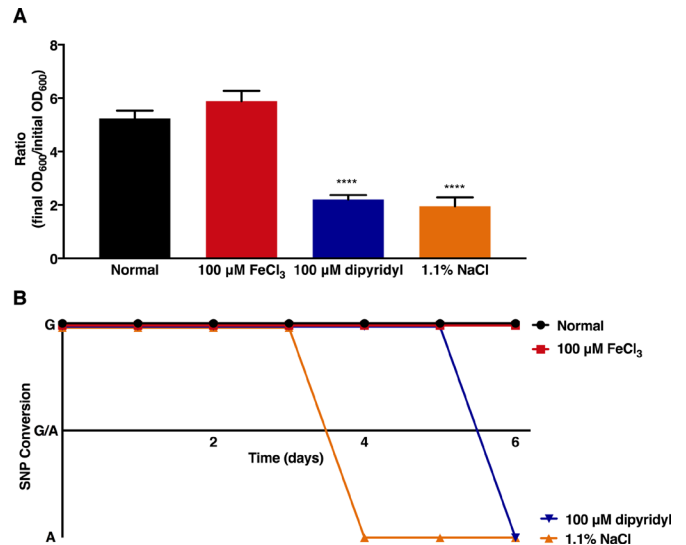


Figure 4 Analysis of genetic variation within *fur* following continuous culture in vitro under high salt or low iron conditions. (A) *Helicobacter pylori* strain 7.13 was grown in normal Brucella broth, modified Brucella broth with iron supplementation (100 μM FeCl₃), iron chelation (100 μM dipyritydil) or high salt (1.1% NaCl) for 20 hours and continuously over a 6-day time course. The growth rate or ratio of the final OD₆₀₀ to the initial starting OD₆₀₀ is represented as mean values with SE from experiments performed on at least six independent occasions. A one-way analysis of variance with multiple comparisons was used to determine statistical significance. ****p<0.0001. (B) *H. pylori* strain 7.13 was continuously cultured in vitro under these conditions for 6 days to assess the presence (CAC) or absence (CGC) of the SNP in *fur* (FurR88H). SNP, single nucleotide polymorphism.

based on the FurR88H variant, suggesting that this region may be a susceptible genetic locus subject to variation under these conditions. The presence of the *fur* SNP in some regular salt²¹ and iron-replete strains in addition to one 7.13 single colony suggests that this variant is present at a low frequency within *H. pylori* populations and is favoured under certain conditions such as high salt and low iron.

To extend these findings, we next investigated the direct effects of high salt or low iron per se on genetic variation within *fur*. *H. pylori* strain 7.13 was continuously cultured in vitro for 6 days under high salt (1.1% NaCl), in the presence of an iron chelator (100 μM dipyritydil) to simulate iron depletion, or supplementation with iron (100 μM FeCl₃). As expected, *H. pylori* growth under high salt or low iron was significantly attenuated (figure 4A). Sequence analysis revealed that adaptation to high salt and/or iron depletion in vitro resulted in selection of the *fur* SNP (FurR88H) after only 4 to 5 days, respectively, and this variant was maintained throughout the duration of growth in vitro (figure 4B). Exposure of *H. pylori* to normal growth conditions or normal growth conditions supplemented with iron did not induce sequence variation within *fur*. These data indicate that in vitro adaptation to high salt or low iron induces genetic variation in *fur* in vitro, suggesting that adaptation within the context of environmental stress may be regulated by these dietary changes in vivo.

To expand on these in vitro studies and specifically address the relationship between presence of the FurR88H variant and iron deficiency in vivo, we performed sequence analysis on 224 in vivo-adapted *H. pylori* strains isolated from gerbils maintained on iron-replete or iron-depleted diets. The FurR88H variant was

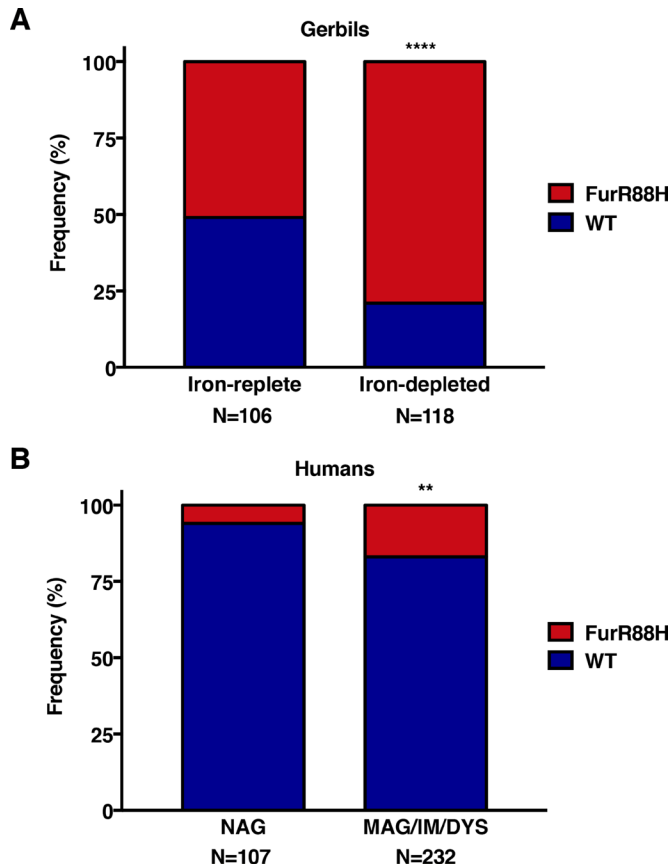


Figure 5 Analysis of genetic variation within *fur* in vivo. The frequency of the FurR88H variant was examined among (A) 224 in vivo-adapted *Helicobacter pylori* strains isolated from gerbils maintained on iron-replete or iron-depleted diets and (B) 339 clinical *H. pylori* strains isolated from three independent and well-defined patient cohorts within Colombia and the USA.^{31–33} Fischer's exact test was used to determine statistical significance. ** $p < 0.005$ and **** $p < 0.0001$. DYS, dysplasia; IM, intestinal metaplasia; MAG, multifocal atrophic gastritis; NAG, non-atrophic gastritis alone.

present at a significantly higher frequency among gerbils maintained on iron-depleted diets (93/118, 79%) compared with those maintained on iron-replete diets (54/106, 51%; $p < 0.0001$, figure 5A). These data indicate that in vivo adaptation to environmental iron deficiency leads to genetic adaptation within *fur*.

Finally, we examined genetic variation in *fur* among 339 clinical *H. pylori* strains isolated from three independent and well-defined patient cohorts within Colombia and the USA (online supplementary table S7).^{31–33} Sequence analysis of *fur* revealed that 14% (46/339) of these clinical *H. pylori* strains harboured the same FurR88H variant identified in in vivo gerbil-adapted strains isolated from conditions of high salt or iron deficiency (online supplementary table S7). Among the clinical *H. pylori* isolates examined, 17% (40/232) of strains from patients with premalignant lesions (multifocal atrophic gastritis, intestinal metaplasia, dysplasia) harboured the FurR88H variant, compared with 6% (6/107) of strains from patients with non-atrophic gastritis alone ($p = 0.0034$, figure 5B). These results demonstrate that this specific *fur* SNP is likely important for in vivo adaptation to a new host and within subpopulations of humans at increased risk for gastric cancer.

DISCUSSION

Bacterial population structures differ based on the taxa under investigation. Some species, such as *Mycobacterium tuberculosis*³⁴ and *Yersinia pestis*,³⁵ are of such recent origin that little sequence diversity has accumulated. In such largely clonal species, only a few coding sequences are variably present among different isolates.^{36–38} However, bacteria with high sequence diversity and frequent homologous recombination, such as *H. pylori*, harbour different population dynamics.³⁹

H. pylori has co-evolved with humans since their origin, and such lifelong evolution is associated with selection of host-specific variants within particular genes necessary for survival and adaptation to the human host.⁸ Survival and persistence within the human host depends on the ability of *H. pylori* to rapidly respond to changing environmental conditions and circumvent host defense mechanisms initiated during colonisation.⁸ Rearrangement of genomic DNA allows a variety of pathogens to alter expression of surface antigens and evade host immunity, and *H. pylori* has one of the highest rates of genetic recombination of any known bacterial species,⁴⁰ suggesting that this process confers a selective advantage in colonisation and survival. Indeed, *H. pylori* mutants defective in homologous recombination are spontaneously cleared,⁴¹ reinforcing the tenet that genetic recombination and variation are critical for *H. pylori* survival within the human host.

In addition to high rates of recombination, *H. pylori* displays exceptional genetic variability and intraspecies diversity.^{4, 40} *H. pylori* has a highly plastic genome, reflecting elevated rates of recombination and development of point mutations, which promotes bacterial divergence and intraspecies diversity within a population.^{42, 43} *H. pylori* exists within its gastric niche as a bacterial population in a continuous state of genetic variation, which allows rapid adaptation to changing environmental conditions in its current host as well as an ability to successfully colonise a new host.^{2, 44, 45} In this study, we hypothesised that prolonged adaptation of *H. pylori* to an in vivo carcinogenic gastric environment induced by iron deficiency would facilitate genetic changes involved in survival, persistence and increased virulence. Indeed, we demonstrated that *H. pylori* strains isolated from gerbils maintained on iron-replete diets are more closely related to each other than to strains isolated from gerbils maintained on iron-depleted diets, suggesting that specific variants are advantageous for survival and persistence in altered microenvironments. In addition, the current analysis confirmed that *H. pylori* B128 isolates recovered from a human stomach are more closely related to each other than to *H. pylori* strains adapted to gerbil stomachs, and that there are specific genetic variations that arise following adaptation to a new host. The diversity among 7.13 single colony isolates suggests that passage of B128 through a new gerbil host resulted in a more diverse population of isolates poised for further adaptation in vivo. As a result of frequent recombination and genetic variation, almost all *H. pylori* isolates from unrelated hosts and differing environments possess unique sequences.^{4, 39}

Further and consistent with other reports,^{6, 7} we have demonstrated that variability exists among single colony isolates from each individual *H. pylori* strain. As a result of this diversity, *H. pylori* has been previously termed a quasispecies, a population of genetic variants comprising a single strain that allows for rapid selection and adaptation to the changing gastric microenvironments of the host.^{45–47} Recent changes in DNA sequencing technology have afforded the opportunity to assess the genetic diversity of *H. pylori* populations in much greater depth. We

have now demonstrated that polymorphisms exist within single colonies, suggesting that even greater levels of within-host population diversity exist than previously demonstrated. Furthermore, we identified numerous genetic variants that arise under the selective pressure of iron deficiency, and some of these variants were also identified in *H. pylori* isolates exposed to high salt.

Among the various SNPs that resulted in non-synonymous variants, the majority arose within genetic loci prone to variation during adaptation to a new host or environmental condition. Many of these included outer membrane proteins that are particularly susceptible to selective pressure for survival and persistence. Adhesins that underwent genetic modification included *babA* and *babB*, which have been shown to undergo specific recombination in other models of *H. pylori* infection.^{48–50} Another genetic locus, *cagY*, which encodes an essential component of the *cag* type IV secretion system, has been shown to be particularly susceptible to variation *in vivo*.^{51 52} Another targeted locus was *fur*, which encodes a transcriptional regulator that controls expression of a large repertoire of genes and is an essential factor for survival and persistence of *H. pylori* in the stomach. *Fur* controls bacterial iron homeostasis, and also plays a central role in bacterial metabolism, energy production and immune regulation.⁵³ The specific SNP in *fur* (FurR88H) identified following adaptation to low iron conditions was also identified in another study of *in vivo* adaptation to high salt.³⁰ In this study, the *fur* SNP was directly linked to increased expression of genes controlling iron acquisition and homeostasis as well as increased resistance to high salt conditions and oxidative stress.³⁰ Due to the role of *Fur* in regulating iron homeostasis, it is likely that this SNP also promotes increased survival and persistence of *H. pylori* under low iron conditions. Collectively, our data suggest that the accumulation of polymorphisms allows *H. pylori* to rapidly adapt to changing microenvironments to promote survival and persistence within the host.

METHODS

Mongolian gerbils

Male Mongolian gerbils (Charles River Laboratories) were housed in the Vanderbilt University Animal Care Facilities and maintained on iron-depleted (TestDiet 5TWD), iron-replete (TestDiet 5STQ),¹⁷ regular salt or high salt (Purina 5001).¹⁴ Gerbils were orogastrically challenged with carcinogenic *H. pylori* strain 7.13. *H. pylori* strain 7.13 was originally isolated as a single colony following *in vivo* adaptation of the human *H. pylori* strain B128 to the Mongolian gerbil model and was then subcultured following minimal passage for single colony purification (figure 1).²⁷ Single colony purified strain 7.13 was isolated from a frozen stock, expanded on trypticase soy agar plates with 5% sheep blood (BD Biosciences) and then grown in Brucella broth (BD Biosciences) supplemented with 10% FBS (Atlanta Biologicals) for 16 hours at 37°C with 5% CO₂ prior to gerbil challenge.¹⁷ Gerbils were then euthanised 12–16 weeks postchallenge. The Vanderbilt University Institutional Animal Care and Use Committee (IACUC) approved all experiments and procedures.

H. pylori culture

H. pylori strain B128 was originally isolated from a single gastric biopsy of a human patient with gastric ulceration.⁵⁴ B128 was subsequently used for orogastric challenge of a gerbil, and the *in vivo*-adapted output strain isolated from this gerbil was designated *H. pylori* strain 7.13.²⁷ *H. pylori* strain 7.13 reproducibly induces gastric cancer in rodent models,^{14 17 27} while B128 only induces

gastritis, but not cancer in these models.²⁷ Three single colonies of B128 and 7.13 were isolated and expanded. In a more recent study, carcinogenic strain 7.13 was then used to infect another cohort of gerbils maintained on either iron-replete or iron-depleted diets.¹⁷ *In vivo*-adapted single colonies were recovered from gerbils maintained on iron-replete (n=3) and iron-depleted (n=3) diets, and single colonies were isolated and expanded.

In this study, we analysed *in vivo*-adapted output strains that were cultured from these gerbils. Specifically, we analysed three single colonies cultured from three gerbils maintained on an iron-replete (n=9) or iron-depleted (n=9) diets, and these were compared with single colonies cultured from one gerbil maintained on regular salt (n=4) or high salt (n=4). All *H. pylori* isolates were minimally passaged on trypticase soy agar plates with 5% sheep blood (BD Biosciences) and in Brucella broth (BD Biosciences) supplemented with 10% FBS (Atlanta Biologicals) for 16 hours at 37°C with 5% CO₂.

H. pylori *in vitro* adaptation to high salt or low iron conditions

H. pylori strain 7.13 was grown in modified Brucella broth containing 1% Bacto tryptone (BD Biosciences), 1% Bacto Proteose Peptone No. 3 (BD Biosciences), 0.2% yeast extract (Difco), 0.1% dextrose (Fisher Scientific) with high sodium chloride (1.1% NaCl, Sigma), dipyrindyl (100 μM 2,2-bipyridyl, Sigma) or addition of iron (100 μM FeCl₃, Sigma) approximately 20 hours at 37°C with 5% CO₂ and continuously over a 6-day time course. Approximately 5 × 10⁸ colony forming units (CFUs) were harvested at each time point during mid-log phase for DNA extraction and used to inoculate a fresh broth culture for continuous culture *in vitro*.

H. pylori whole genome sequence analysis

H. pylori cultures were harvested, and bacterial DNA was extracted using the Wizard Genomic Purification Kit (Promega). *H. pylori* DNA was sheared using Covaris, yielding an average fragment length of 550 bp. Libraries were prepared using Truseq Nano LT Kit (Illumina) with unique indexes. Postquantitation was performed using the Kapa Library Quantitation Kit (Kapa Biosystems). Libraries were sequenced on the MiSeq Sequencer using the 600V3 Kit (Illumina). For analysis, raw sequence reads were quality trimmed using CLCbio Genomics Workbench V. 8.0 (Qiagen). Alignment files were exported from the CLCbio Workbench in BAM file formats and were then imported into Visual Genomics Analysis Studio, software developed at the Institute for Immunology and Infectious Diseases at Murdoch University, and Geneious V. 6.1.8 (Biomatters) for further analysis. Analyses were performed using a 25% cut-off, which required at least 25% of the reads from at least one single colony to contain the polymorphism, to allow detection of intrapopulation variance. All sequence reads were aligned with the genome sequence of *H. pylori* strain B8 (GenBank accession number NC_014256.1), a closely related strain whose complete genome sequence is available.²⁸ B8 was chosen as the reference strain because *H. pylori* strains 7.13 and B8 each originated from the same parental *H. pylori* strain B128 and each was generated by experimentally infecting gerbils with strain B128. Although B128 (GenBank accession number ABSY0000000) and 7.13 (GenBank accession number NZ_LAQK0000000) have been previously sequenced, these strains were also sequenced in this study to directly compare these sequences within the newly derived strains using the same sequencing technology. All sequences have been deposited to GenBank (BioProject PRJNA362878). Sequence data for

H. pylori strains isolated from gerbils on a high salt diet or a regular diet²¹ were analysed using the same criteria. Specifically, we sought to identify SNPs present in isolates from gerbils on iron-depleted or high salt diets, but not present or present at a lower abundance in isolates from gerbils maintained on iron-replete or regular salt diets.

DNA extraction, PCR and sequence analysis

Approximately 5×10^8 CFUs were pelleted by centrifugation and stored at -20°C . DNA was extracted from *H. pylori* using the PureLink Pro 96 Genomic DNA Purification Kit (Invitrogen). DNA was then used as template to amplify a 500bp product using *fur*-specific primers: 5'TCCGCCAAAAAGACAAAAAC3' and 5'GGGCAAGACTTTCACCTTGA3'. PCR products were prepared for sequence analysis using ExoSAP-IT Express PCR Product Cleanup Kit (Affymetrix), quantified using spectrophotometric nucleic acid quantification (BioTek) and sequenced (GeneWiz). Geneious V. 6.1.8 (Biomatters) was used for subsequent analysis.

H. pylori clinical isolates

Clinical *H. pylori* strains were isolated from three independent and well-defined patient cohorts within Colombia and the USA.^{31–33} Online supplementary table S7 includes information on the clinical *H. pylori* strains, including strain identification number, patient country of origin, age, sex and histological diagnosis.

Statistical analysis

Mean values with SE are shown from experiments performed on at least six independent occasions. One-way analysis of variance with multiple comparisons and Fisher's exact tests were used to determine statistical significance. $p < 0.05$ was considered statistically significant.

Study approval

All animal and human studies were conducted in accordance with the Declaration of Helsinki principles and have been approved by Vanderbilt University IACUC and Institutional Review Board, respectively.

Acknowledgements We would like to acknowledge the following funding sources: R01CA190612 (KTW), R01DK053620 (KTW), P01CA028842 (KTW), R01AI039657 (TLC), R01AI118932 (TLC), Department of Veterans Affairs 2101B×0 00627 (TLC), R01CA077955 (RMP), R01DK058587 (RMP), P30DK058404 (RMP), P01CA116087 (RMP, TLC, KTW).

Contributors JMN contributed to the study design, performed experiments, gathered and analysed data, and drafted the manuscript. AC generated and analysed whole genome sequences. JTL generated and provided sequence data from *Helicobacter pylori* strains harvested from gerbils maintained on regular or high salt diets. JRG assisted with all animal experiments and isolation of in vivo-adapted *H. pylori* strains. MBP provided all patient information from the two Colombian patient cohorts and clinical *H. pylori* isolates from these subjects. MW and SL generated and analysed whole genome sequences. ACB and KTW provided feedback on the manuscript. TLC contributed to the study design and provided sequence data from *H. pylori* strains harvested from gerbils maintained on regular or high salt diets. SM contributed to the study design and the whole genome sequence analysis. DAI contributed to the study design, analysed sequence data, and helped with drafting and editing the manuscript. RMP contributed to the study design and helped with drafting and editing the manuscript.

Competing interests None declared.

Patient consent Obtained

Ethics approval Institutional Review Board, Vanderbilt University Medical Center.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Salama N, Guillemin K, McDaniel TK, et al. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci U S A* 2000;97:14668–73.
- Blaser MJ, Berg DE. *Helicobacter pylori* genetic diversity and risk of human disease. *J Clin Invest* 2001;107:767–73.
- Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat Rev Microbiol* 2007;5:441–52.
- Falush D, Kraft C, Taylor NS, et al. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc Natl Acad Sci U S A* 2001;98:15056–61.
- Linz B, Windsor HM, McGraw JJ, et al. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nat Commun* 2014;5:4165.
- Israel DA, Salama N, Krishna U, et al. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci U S A* 2001;98:14625–30.
- Björkholm B, Lundin A, Sillén A, et al. Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. *Infect Immun* 2001;69:7832–8.
- Atherton JC, Blaser MJ. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. *J Clin Invest* 2009;119:2475–87.
- Cover TL, Peek RM. Diet, microbial virulence, and *Helicobacter pylori*-induced gastric cancer. *Gut Microbes* 2013;4:482–93.
- Lee SA, Kang D, Shim KN, et al. Effect of diet and *Helicobacter pylori* infection to the risk of early gastric Cancer. *J Epidemiol* 2003;13:162–8.
- Tsugane S. Salt, salted food intake, and risk of gastric cancer: epidemiologic evidence. *Cancer Sci* 2005;96:1–6.
- Fox JG, Dangler CA, Taylor NS, et al. High-salt diet induces gastric epithelial hyperplasia and parietal cell loss, and enhances *Helicobacter pylori* colonization in C57BL/6 mice. *Cancer Res* 1999;59:4823–8.
- Bergin IL, Sheppard BJ, Fox JG. *Helicobacter pylori* infection and high dietary salt independently induce atrophic gastritis and intestinal metaplasia in commercially available outbred Mongolian gerbils. *Dig Dis Sci* 2003;48:475–85.
- Gaddy JA, Radin JN, Loh JT, et al. High dietary salt intake exacerbates *Helicobacter pylori*-induced gastric carcinogenesis. *Infect Immun* 2013;81:2258–67.
- Akiba S, Neriishi K, Blot WJ, et al. Serum ferritin and stomach cancer risk among a Japanese population. *Cancer* 1991;67:1707–12.
- Nomura A, Chyou PH, Stemmermann GN. Association of serum ferritin levels with the risk of stomach cancer. *Cancer Epidemiol Biomarkers Prev* 1992;1:547–50.
- Noto JM, Gaddy JA, Lee JY, et al. Iron deficiency accelerates *Helicobacter pylori*-induced carcinogenesis in rodents and humans. *J Clin Invest* 2013;123:479–92.
- Szcebara F, Dhaenens L, Armand S, et al. Regulation of the transcription of genes encoding different virulence factors in *Helicobacter pylori* by free iron. *FEMS Microbiol Lett* 1999;175:165–70.
- Delany I, Spohn G, Rappuoli R, et al. The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*. *Mol Microbiol* 2001;42:1297–309.
- Merrell DS, Thompson LJ, Kim CC, et al. Growth phase-dependent response of *Helicobacter pylori* to iron starvation. *Infect Immun* 2003;71:6510–25.
- Lee HW, Choe YH, Kim DK, et al. Proteomic analysis of a ferric uptake regulator mutant of *Helicobacter pylori*: regulation of *Helicobacter pylori* gene expression by ferric uptake regulator and iron. *Proteomics* 2004;4:2014–27.
- Loh JT, Torres VJ, Cover TL. Regulation of *Helicobacter pylori* *cagA* expression in response to salt. *Cancer Res* 2007;67:4709–15.
- Gancz H, Jones KR, Merrell DS. Sodium chloride affects *Helicobacter pylori* growth and gene expression. *J Bacteriol* 2008;190:4100–5.
- Loh JT, Friedman DB, Piazuolo MB, et al. Analysis of *Helicobacter pylori* *cagA* promoter elements required for salt-induced upregulation of *CagA* expression. *Infect Immun* 2012;80:3094–106.
- Voss BJ, Loh JT, Hill S, et al. Alteration of the *Helicobacter pylori* membrane proteome in response to changes in environmental salt concentration. *Proteomics Clin Appl* 2015;9:1021–34.
- Noto JM, Lee JY, Gaddy JA, et al. Regulation of *Helicobacter pylori* virulence within the context of iron deficiency. *J Infect Dis* 2015;211:1790–4.
- Franco AT, Israel DA, Washington MK, et al. Activation of beta-catenin by carcinogenic *Helicobacter pylori*. *Proc Natl Acad Sci U S A* 2005;102:10646–51.
- Farnbacher M, Jahns T, Willrodt D, et al. Sequencing, annotation, and comparative genome analysis of the gerbil-adapted *Helicobacter pylori* strain B8. *BMC Genomics* 2010;11:335.
- Heberle H, Meirelles GV, da Silva FR, et al. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 2015;16:169.

- 30 Loh JT, Gaddy JA, Algood HM, *et al.* *Helicobacter pylori* adaptation in vivo in response to a high-salt diet. *Infect Immun* 2015;83:4871–83.
- 31 Peek RM, Blaser MJ, Mays DJ, *et al.* *Helicobacter pylori* strain-specific genotypes and modulation of the gastric epithelial cell cycle. *Cancer Res* 1999;59:6124–31.
- 32 Mera R, Fontham ET, Bravo LE, *et al.* Long term follow up of patients treated for *Helicobacter pylori* infection. *Gut* 2005;54:1536–40.
- 33 Schneider BG, Mera R, Piazzuelo MB, *et al.* DNA methylation predicts progression of human gastric lesions. *Cancer Epidemiol Biomarkers Prev* 2015;24:1607–13.
- 34 Sreevatsan S, Pan X, Stockbauer KE, *et al.* Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 1997;94:9869–74.
- 35 Achtman M, Morelli G, Zhu P, *et al.* Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci U S A* 2004;101:17837–42.
- 36 Hinchliffe SJ, Isherwood KE, Stabler RA, *et al.* Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Genome Res* 2003;13:2018–29.
- 37 Tsolaki AG, Hirsh AE, DeRiemer K, *et al.* Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci U S A* 2004;101:4865–70.
- 38 Hirsh AE, Tsolaki AG, DeRiemer K, *et al.* Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci U S A* 2004;101:4871–6.
- 39 Gressmann H, Linz B, Ghai R, *et al.* Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet* 2005;1:e43.
- 40 Suerbaum S, Smith JM, Bapumia K, *et al.* Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci U S A* 1998;95:12619–24.
- 41 Loughlin MF, Barnard FM, Jenkins D, *et al.* *Helicobacter pylori* mutants defective in RuvC Holliday junction resolvase display reduced macrophage survival and spontaneous clearance from the murine gastric mucosa. *Infect Immun* 2003;71:2022–31.
- 42 Kang J, Blaser MJ. Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. *Nat Rev Microbiol* 2006;4:826–36.
- 43 Levine SM, Lin EA, Emara W, *et al.* Plastic cells and populations: DNA substrate characteristics in *Helicobacter pylori* transformation define a flexible but conservative system for genomic variation. *Faseb J* 2007;21:3458–67.
- 44 Blaser MJ. Ecology of *Helicobacter pylori* in the human stomach. *J Clin Invest* 1997;100:759–62.
- 45 Covacci A, Rappuoli R. *Helicobacter pylori*: molecular evolution of a bacterial quasi-species. *Curr Opin Microbiol* 1998;1:96–102.
- 46 Kuipers EJ, Israel DA, Kusters JG, *et al.* Quasispecies development of *Helicobacter pylori* observed in paired isolates obtained years apart from the same host. *J Infect Dis* 2000;181:273–82.
- 47 Webb GF, Blaser MJ. Dynamics of bacterial phenotype selection in a colonized host. *Proc Natl Acad Sci U S A* 2002;99:3135–40.
- 48 Pride DT, Meinersmann RJ, Blaser MJ, *et al.* Allelic variation within *Helicobacter pylori* *babA* and *babB*. *Infect Immun* 2001;69:1160–71.
- 49 Solnick JV, Hansen LM, Salama NR, *et al.* Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques. *Proc Natl Acad Sci U S A* 2004;101:2106–11.
- 50 Colbeck JC, Hansen LM, Fong JM, *et al.* Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of *Helicobacter pylori*. *Infect Immun* 2006;74:4375–8.
- 51 Aras RA, Fischer W, Perez-Perez GI, *et al.* Plasticity of repetitive DNA sequences within a bacterial (Type IV) secretion system component. *J Exp Med* 2003;198:1349–60.
- 52 Barrozo RM, Cooke CL, Hansen LM, *et al.* Functional plasticity in the type IV secretion system of *Helicobacter pylori*. *PLoS Pathog* 2013;9:e1003189.
- 53 Pich OQ, Merrell DS. The ferric uptake regulator of *Helicobacter pylori*: a critical player in the battle for iron and colonization of the stomach. *Future Microbiol* 2013;8:725–38.
- 54 Israel DA, Salama N, Arnold CN, *et al.* *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. *J Clin Invest* 2001;107:611–20.