

Full Paper

Quantum changes in *Helicobacter pylori* gene expression accompany host-adaptation

Eng-Guan Chua^{1,*}, Michael J. Wise^{1,2}, Yalda Khosravi³, Shih-Wee Seow⁴,
Arlaine A. Amoyo⁴, Sven Pettersson^{5,6,7}, Fanny Peters¹, Chin-Yen Tay¹,
Timothy T. Perkins¹, Mun-Fai Loke³, Barry J. Marshall^{1,8}, and
Jamuna Vadivelu³

¹The Marshall Centre for Infectious Diseases Research and Training, School of Pathology and Laboratory Medicine, The University of Western Australia, Nedlands, Western Australia, Australia, ²School of Computer Science and Software Engineering, The University of Western Australia, Australia, ³Department of Medical Microbiology, University of Malaya, Kuala Lumpur, Malaysia, ⁴National Cancer Centre of Singapore, Singapore, ⁵Department of Microbiology, Tumor and Cell Biology (MTC), Karolinska Institutet, Stockholm, Sweden, ⁶LKC School of Medicine, Nanyang Technological University, Singapore, ⁷SCElse Microbiome Centre, Nanyang Technological University, Singapore, and ⁸UM Marshall Centre, High Impact Research Building, University of Malaya, 50603 Kuala Lumpur, Malaysia

*To whom correspondence should be addressed. Tel. +61893464817. Fax. +61893464816. Email: eng.chua@uwa.edu.au

Edited by Prof. Masahira Hattori

Received 24 June 2016; Accepted 23 September 2016

Abstract

Helicobacter pylori is a highly successful gastric pathogen. High genomic plasticity allows its adaptation to changing host environments. Complete genomes of *H. pylori* clinical isolate UM032 and its mice-adapted serial derivatives 298 and 299, generated using both PacBio RS and Illumina MiSeq sequencing technologies, were compared to identify novel elements responsible for host-adaptation. The acquisition of a *jhp0562*-like allele, which encodes for a galactosyltransferase, was identified in the mice-adapted strains. Our analysis implies a new β -1,4-galactosyltransferase role for this enzyme, essential for Le^y antigen expression. Intragenomic recombination between *babA* and *babB* genes was also observed. Further, we expanded on the list of candidate genes whose expression patterns have been mediated by upstream homopolymer-length alterations to facilitate host adaptation. Importantly, greater than four-fold reduction of mRNA levels was demonstrated in five genes. Among the down-regulated genes, three encode for outer membrane proteins, including BabA, BabB and HopD. As expected, a substantial reduction in BabA protein abundance was detected in mice-adapted strains 298 and 299 via Western analysis. Our results suggest that the expression of Le^y antigen and reduced outer membrane protein expressions may facilitate *H. pylori* colonisation of mouse gastric epithelium.

Key words: BabA, expression, *Helicobacter pylori*, host adaptation, Lewis antigen

1. Introduction

Helicobacter pylori infects approximately half the world's population. It is a Gram-negative microaerophilic pathogen that persistently colonises the human stomach. The bacterium is associated with chronic gastritis and peptic ulceration, and less commonly, with gastric adenocarcinoma and gastric mucosa-associated lymphoid tissue (MALT) lymphoma.^{1,2} Hence, the organism was classified as a class 1 carcinogen by World Health Organisation in 1994.³

Population analysis of *H. pylori* isolates from different geographical origins has revealed a high degree of genetic diversity within the species. Multi-locus sequence typing (MLST) analysis has shown that identical alleles are extremely rare in *H. pylori* even when sampling a large population of the same geographical origin.⁴ This indicates that *H. pylori* is able to adapt to a new or rapidly changing environment by undergoing continuous genetic modification. Greater understanding of these modifications and the mechanisms by which they arise will improve the knowledge base from which we may develop new therapeutic targets.

The high genetic variation in *H. pylori* appears to be generated through intragenic and intergenic recombinations, plus extensive mutations attributed to the lack of a nucleotide mismatch repair system, such as that provided in other bacteria by MutS, MutH and MutL.⁵ There are two families of MutS homologues: MutS1 and MutS2. The MutS1 family includes *Escherichia coli* MutS, which plays a role in post-replication mismatch repair. In the *H. pylori* genome, HP0621, a member of the MutS2 family has been identified. However, rather than participating in mismatch repairing, it suppresses homeologous recombination mediated by RecA.^{6,7} Compared to *E. coli*, *H. pylori* is also missing a number of elements responsible for direct repair of DNA damage such as Ada methyltransferase, AlkB oxidative demethylase and Phr/Spl photolyase, as well as several glycosylases and endonucleases involved in base excision repair including MutM, Tag, AlkA, Mpg, YgjF, Nei and Nfo.⁸ On the other hand, the presence of a DNA polymerase I that lacks proof-reading activity also contributes to the generation of genomic plasticity in *H. pylori*.⁹

Rapid phenotype switching is another strategy that enables pathogens to survive in diverse hostile environments and to establish persistent infections in novel hosts. Phenotypic switching can be achieved by epigenetic events, such as DNA methylation and gene silencing mediated by non-coding RNAs. It can also occur via phase-variation-associated genetic modifications, such as site-specific recombination, inversion of promoter elements, insertional inactivation mediated by insertion sequences and slipped-strand mispairing of simple sequence repeats (SSRs) including homopolymeric tracts.^{10,11}

Changes in homopolymeric tract length located in coding sequences are known to generate frameshift errors and hence result in expression of truncated and often non-functional proteins.¹² While intragenic homopolymeric tract alterations turn protein expression on or off, intergenic variations may alter mRNA level to fine-tune protein expression. This has recently been shown in the adhesin-encoding gene *sabA* of *H. pylori*. Altering the length of a poly(T) tract found adjacent to its -35 promoter region changes the mRNA transcript and protein expression levels, as well as the binding to the sialyl Le^x receptor.¹³

To date, relatively little is known about *H. pylori* genetic adaptation in response to selective pressures from its host. It is important to identify these adaptive mutations to provide deeper insights into the exact mechanisms underlying *H. pylori* infection. Although *H. pylori* is highly adapted to the human gastric mucosa environment, several animal models comprising mice, gerbils and non-human primates have been

used to advance understanding of *H. pylori* colonisation and pathogenesis.^{14–16} Whilst experimental studies in humans would be most likely to reveal clinically relevant details, studies in animal models are ethically superior to human challenge experiments, and can still tell us much about the genomic changes associated with host adaptation.

Here, we report a follow-up investigation to our previously announced complete genomes of the *H. pylori* clinical strain UM032 and its mice-adapted derivatives designated 298 and 299. These were sequenced on the Pacific Biosciences RS sequencing technology using the C2 chemistry.¹⁷ PacBio RS sequencing technology produces extraordinarily long reads, thus increasing the likelihood of acquiring a complete genome sequence. As opposed to the current C5 chemistry, high error rates were observed in sequencing data generated using the earlier C2 chemistry.¹⁸ Therefore, for this study all three strains were also subjected to whole-genome sequencing using Illumina's MiSeq platform. Hence, the comparative data presented here provide comprehensive information about the changes associated with host change in *H. pylori*. We reveal *H. pylori* genetic elements potentially required for host-adaptation and epigenetic regulation of gene expression mediated by homopolymeric tracts present both in the upstream regions of coding sequences, within or in close proximity to the promoter elements, and within the coding sequences.

2. Methods

2.1. Bacterial strains

The acquisition of *H. pylori* clinical strain UM032 and its mice-adapted derivatives, designated 298 and 299 respectively, was as previously described.¹⁷

2.2. Illumina library preparation and sequencing

Preparation of MiSeq library was performed using Illumina Nextera XT DNA sample preparation kit (Illumina, San Diego, CA, USA) as previously described with minor modifications.¹⁹ In brief, 1 ng of genomic DNA was fragmented in 5 µl of Amplicon Tagment Mix and 10 µl of Tagment DNA buffer. Tagmentation reaction was performed by incubation at 55 °C for 5 min followed by neutralisation with 5 µl of Neutralise Tagment Buffer for 5 min. Tagmented DNA (25 µl) was indexed in a 50 µl limited-cycle PCR (12 cycles) as outlined in the Nextera XT protocol and subsequently purified using 25 µl of AMPure XP beads (Beckman Coulter Inc, Australia). The fragment size distribution of the purified DNA was analysed utilising a 2100 Bioanalyser with a High Sensitivity DNA assay kit (Agilent Technologies, Santa Clara, CA). DNA libraries were adjusted to 2 nM, pooled in equal volumes and then denatured with 0.2 N NaOH according to the Nextera protocol. The libraries were sequenced using 2 × 300 paired-end protocols on an Illumina MiSeq instrument (MiSeq Reagent Kit v3 for 600 cycles).

2.3. Genome correction

The generated MiSeq data were used for correction of homopolynucleotide errors present in the previously published *H. pylori* UM032, 298 and 299 complete genomes generated using PacBio C2 chemistry. In brief, MiSeq raw reads were mapped against the complete genome using the Geneious R7 read mapper with the medium sensitivity option.²⁰ Variants were subsequently called using the following parameters: minimum coverage = 10 and minimum variant frequency = 0.7. Reported errors in each reference genome sequence were corrected.

2.4. Genome annotation

Initially, the genome sequences were automatically annotated using the RAST (Rapid Annotation using Subsystem Technology) pipeline.²¹ In addition to Glimmer (version 3.02) used in the pipeline, additional ORF prediction was performed using GeneMarkS (version 4.7a) and Prodigal (version 2.60).^{22–24} By majority ruling, the predicted start codons of ORFs predicted from multiple annotation engines were manually curated to enhance start codon accuracy. Genome annotations were further compared against gene sequences from 45 completely sequenced publicly available *H. pylori* genomes to allow identification of pseudogenes. In our context, a pseudogene can be a non-functional DNA sequence with a frameshift or premature stop codon, a sequence homologous to a protein-coding gene but without ORF or a full-length reference gene broken into two or more adjacent ORFs. Our manually curated genome annotations are available in [Supplementary Tables S1–S3](#).

NCBI accession numbers of 45 public genomes used for genome annotation comparison are as follows: 35A (NC_017360), 51 (NC_017382), 52 (NC_017354), 83 (NC_017375), 908 (NC_017357), 2017 (NC_017374), 2018 (NC_017381), 26695 (NC_000915), Aklavik86 (NC_019563), Aklavik117 (NC_019560), B8 (NC_014256), B38 (NC_012973), BM012A (NC_022886), BM012S (NC_022911), Cuz20 (NC_017358), ELS37 (NC_017063), F16 (NC_017368), F30 (NC_017365), F32 (NC_017366), F57 (NC_017367), G27 (NC_011333), Gambia94/24 (NC_017371), HPAG1 (NC_008086), HUP-B14 (NC_017733), India7 (NC_017372), J99 (NC_000921), Lithuania75 (NC_017362), OK113 (NC_020508), OK310 (NC_020509), P12 (NC_011498), PeCan4 (NC_014555), PeCan18 (NC_017742), Puno120 (NC_017378), Puno135 (NC_017379), Sat464 (NC_017359), Shi112 (NC_017741), Shi169 (NC_017740), Shi417 (NC_017739), Shi470 (NC_010698), SJM180 (NC_014560), SNT49 (NC_017376), SouthAfrica7 (NC_017361), SouthAfrica20 (NC_022130), v225d (NC_017355) and XZ274 (NC_017926).

2.5. *In silico* genome analysis

The revised genome sequences of *H. pylori* UM032, 298 and 299 were compared pairwise using progressiveMauve (version 2.3.1), followed by variants calling using Geneious R7.²⁵

2.6. COG functional analysis

COG of *H. pylori* UM032 genes with amino acid substitutions detected in the mice-adapted strains 298 and 299 was annotated by rpsblast.²⁶

2.7. Western immunodetection

Whole cell lysates were prepared from bacterial cells standardised to OD₆₀₀ of 10. For each blot, 10 µl of sample was used for SDS-PAGE, followed by transfer to an Immobilon[®]-P polyvinylidene difluoride membrane (Merck Millipore). Blood group antigen-binding adhesin A (BabA) detected with polyclonal antibody, with detection by horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG antibodies (Santa Cruz Biotechnology) and visualization by chemiluminescent signal using Clarity[™] Western ECL substrate (Bio-Rad) on a Fujifilm LAS-3000 imaging system. *H. pylori* UreB was detected using rabbit anti-urease B polyclonal antibodies purchased from Sigma Aldrich. *H. pylori* Le antigen phenotypes were determined using monoclonal antibodies (Abcam) to Le^a, Le^b, Le^x, or Le^f. Bound immunoglobulin M (IgM) or IgG antibodies were

detected with HRP-conjugated goat anti-mouse antibodies (Sigma Aldrich) against IgM or IgG.

2.8. Quantitative real-time polymerase chain reaction (qRT-PCR)

Bacterial total RNA was isolated using NucleoSpin[®] RNA (Macherey-Nagel). After RNA quantification, cDNA was synthesized using 1 µg of RNA in a 20 µl reverse transcription reaction using QuantiTect Reverse Transcription Kit (Qiagen) according to manufacturer's instructions. Real time PCR analysis was subsequently performed in 20 µl reactions in 96-well plates using cDNA from 20 ng RNA as template, LuminoCt[®] SYBR[®] Green qPCR ReadyMix[™] (Sigma Aldrich) and the LightCycler 480 instrument (Roche Applied Science). Cycling conditions comprised an initial denaturation at 95 °C for 20 s, followed by 40 cycles of 95 °C for 3 s, 56 °C for 15 s and 60 °C for 30 s. Following each run a melting curve analysis was performed in which the reactions were heated slowly from 55 °C to 95 °C (0.1 °C/s). The relative expression of each target gene was calculated using the 2^{-ΔΔCT} method as described by Livak and Schmittgen²⁷, after normalisation against the *gyrA* reference gene. Primers used are listed in [Supplementary Table S4](#).

2.9. Statistical analysis

Paired Student's *t*-test was used to test for differences in relative quantification of gene expression. *P*-values of less than 0.05 were considered statistically significant.

2.10. Genome sequence update

The updated complete genomes of UM032, 298 and 299 are available in GenBank under the original accession numbers: NC_021215.3 (GI:685455742), NC_021882.2 (GI:687961717) and NC_021216.3 (GI:685456130).

3. Results and discussion

3.1. Global pairwise alignment of previously published UM032, 298 and 299 genomes revealed sequence duplications and homopolymeric tract errors

3.1.1. Duplicate sequences

Genomic rearrangement via intragenic recombination is a powerful strategy of gene regulation employed by *H. pylori* to aid in adaption to host niches and successful colonisation of the gastric mucosa. To investigate the occurrence of chromosomal shuffling events in the mice-adapted UM032 derivatives, progressive Mauve alignment was performed after assigning a common starting point for all three genomes. No genomic rearrangement was observed. However, each genome was shown to harbour an insertion of a large DNA fragment, which measured 5659 bp, 9667 bp and 6578 bp in UM032, 298 and 299, respectively ([Fig. 1](#)). BLASTN search demonstrated that each fragment could be further divided into two different regions that closely resemble the neighbouring sequences as depicted in [Figure 2](#). These insertions are probably due to misassembly of the original PacBio sequences, resulting in false segmental duplications.

3.1.2. Homopolymeric tracts

By using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP), a total of 167 pseudogenes have previously been annotated in UM032 genome compared to 29 in 298 and 25 in

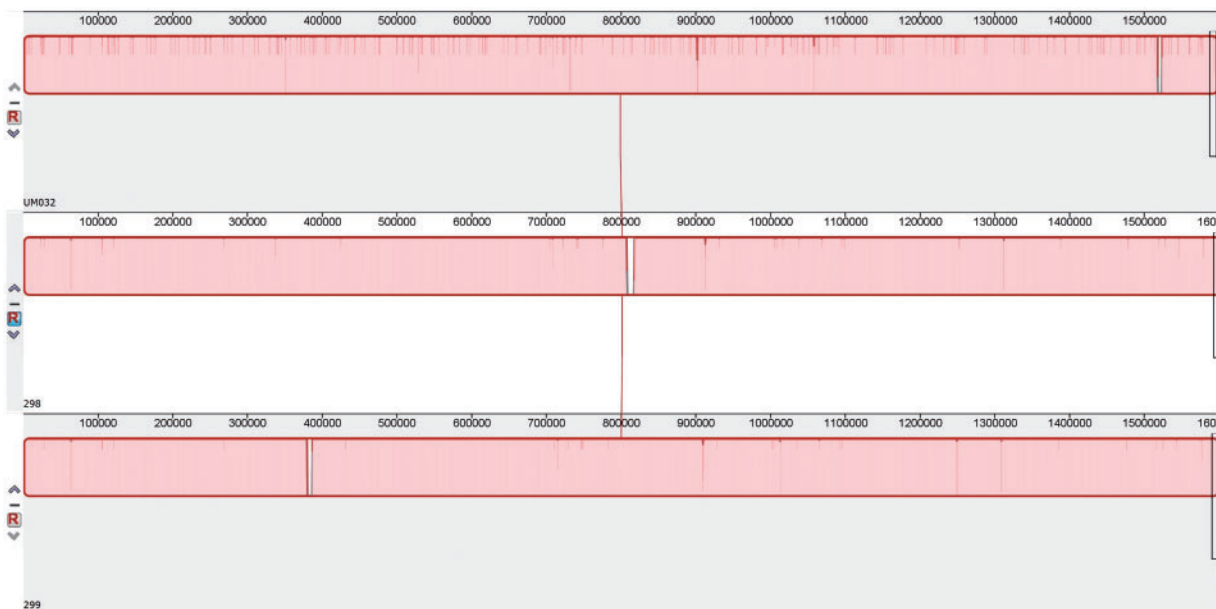


Figure 1. Whole-genome alignment of *H. pylori* strains UM032, 298 and 299 using Mauve indicates no genome shuffling. Insertions are visualised as gaps in the alignment.

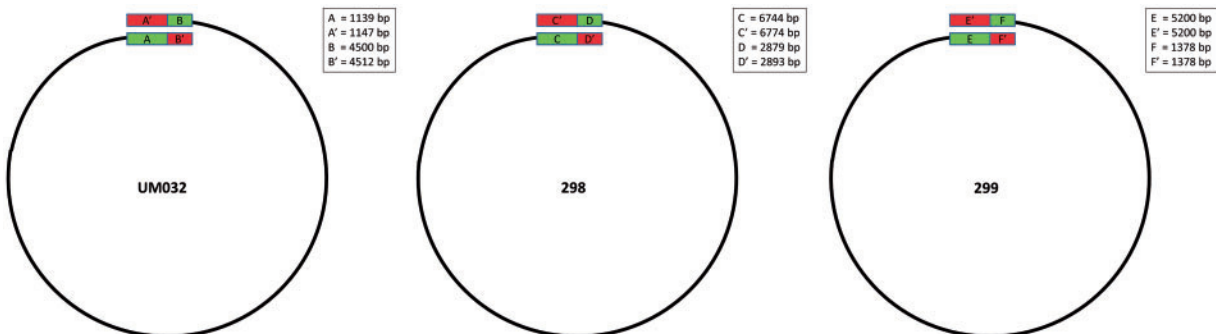


Figure 2. Duplicate sequences in UM032, 298 and 299 complete genomes. The ' character indicates the erroneous overlaps at each end of the original published contig.

299.¹⁷ We performed BLASTN pairwise alignment of the pseudogenes in UM032 against their counterparts in 298 and 299, of which 146 were shown to have deletions associated with homopolymeric sequences in 298 and 299 (Supplementary Table S5), resulting in restoration of correct reading frame and thus the functional coding sequence (CDS). The presence of homopolymeric tracts in *H. pylori* is known to play an essential role in phase variation through slipped-strand mispairing, leading to differential antigenic expression that facilitates host-adaptation and persistent colonisation.¹³ However, it is intriguing to observe such overwhelming mutational bias towards deletion in both 298 and 299 genomes, implying that there is potential sequencing and/or *de novo* assembly error in the Hierarchical Genome Assembly Process (HGAP) workflow (<http://pacbiodevnet.com/>) available in Single-Molecule Real-Time (SMRT) Analysis v2.0. Sanger sequencing was performed to verify the length of the homopolymeric tract in which variation occurs in 7 randomly chosen UM032 pseudogenes and their

functional counterparts. Homopolymer insertion errors detected in selected UM032 pseudogenes could be due to sequencing errors in the original genomes.

3.2. Whole genome resequencing using MiSeq

To remove homopolymer-length sequencing errors and duplications in the original sequences, all three genomes were resequenced using the MiSeq platform with at least 100× sequencing depth. The reads generated were subsequently mapped to the corresponding reference genome as described in Materials and Methods.

3.2.1. Below mean read depth across the duplicate sequences

The analysis of read depth can be a useful approach to check for duplications. It is expected that the read depth at any location is equal to the global average coverage, given that the sequencing reads are

Table 1. Read depth of erroneous duplicate sequences

Global read depth	Duplication sequence			Correct sequence		
	Fragment*	Size (bp)	Read depth	Fragment*	Size (bp)	Read depth*
UM032						
135x	A'	4,512	60 ± 38.4	A	4,500	103.5 ± 41.9
	B'	1,147	7.7 ± 12.3	B	1,139	178.5 ± 58.6
298						
122x	C'	6,774	28.9 ± 29.1	C	6,744	73.3 ± 35.8
	D'	2,893	16.1 ± 17.8	D	2,879	120.6 ± 45.6
299						
174x	E'	5,200	74.3 ± 28.4	E	5,200	82.9 ± 35.5
	F'	1,378	32.7 ± 22.6	F	1,378	81.1 ± 22.7

*Please refer to Figure 2 for the position of each sequence fragment described in this table.

Table 2. Read depth analysis of correct sequences against counterpart sequences following removal of duplications

Strain (global read depth)	Read depth read ratio (read depth/global read depth)		
	AB fragments	CD fragments	EF fragments
UM032 (135x)	168x 1.24	121x 0.90	113x 0.84
298 (122x)	146x 1.20	112x 0.92	102x 0.84
299 (174x)	217x 1.25	157x 0.90	148x 0.85

randomly distributed across a genome. An average read depth of 135x, 122x and 174x, respectively, was determined for UM032, 298 and 299. The erroneously duplicated sequences, however, had substantially lower read depth than the global read depth and compared to that of the closely resembled upstream and downstream sequences, as summarised in Table 1, indicating they were misassembled and incorrectly present in two copies. The false duplication in each genome was therefore removed. This was followed by remapping of the sequencing reads to determine and compare the reads ratio (regional read depth divided by global read depth) of the previously duplicated sequences and their counterparts across all three genomes. These reads ratios are expected to be comparable among each other. As anticipated, the removal of duplication restored the reads ratio of these now single-copy sequences, as shown in Table 2. It is also now clear that what seemed to be the duplicated sequences are in fact near identical overlaps found at each end of the contig, which failed to be removed during sequence assembly to generate accurate circular consensus DNA sequence.

3.2.2. Variants calling in duplication-free reference genomes

As indicated in Supplementary Table S6, 245 variants were detected in UM032, of which 243 were deletions found within homopolymeric runs. Surprisingly, there were significantly less errors in both 298 and 299 genomes, only seven and three variants being detected, respectively. Nonetheless, the reported errors, including all additional nucleotides found in the homopolymeric tracts, were corrected and the genomes were reannotated as stated in Methods. Table 3 summarises the general features of all three genomes, prior to and after the sequence error correction and reannotation. It is important to highlight that a number of false pseudogene candidates have been

efficiently removed, especially in UM032, in which the number has been significantly reduced from 167 to 42, improving the reliability of comparative analysis.

3.3. Whole genome comparisons of revised *H. pylori* UM032, 298 and 299

The revised complete genome sequences of *H. pylori* clinical isolate UM032 and its mice-adapted derivatives 298 and 299 were compared against each other. A pairwise global alignment of the genome of UM032 with that of 298 and 299 revealed 99.9% identity at the nucleotide level. The 298 strain contained 89 variants relative to strain UM032, of which 32 were indels and the remaining were substitution mutations including single-nucleotide polymorphisms (SNPs) (Supplementary Table S7). Nine indels affected ORFs. In addition to tandem repeats or homopolymeric tracts and additions or deletions to the 5' or 3' ends of ORFs, an insertion has introduced a complete ORF, which is annotated as *UM298_1363* encoding a lipopolysaccharide (LPS) biosynthesis protein. Amongst the 57 substitutions detected, 16 were intergenic and 41 were located in protein coding sequences. Only 12 of the latter resulted in amino acid changes and the remaining were synonymous substitutions. Relative to 298, only two tandem repeat insertions and 1 deletion within a homopolymeric tract were identified in 299, of which the latter restores an open reading frame designated *UM299_0755* (Supplementary Table S8). We further performed a COG (clusters of orthologous groups) analysis on genes with non-synonymous changes (Table 4). Genes without an inferred COG category were also analysed. Together it was shown that genes encoding for outer membrane and lipopolysaccharide biosynthesis proteins are enriched for substitution mutations.

Taken together, our data have demonstrated an outburst of mutations when UM032 encounters a new host environment. The second round of mice infection using the mice-adapted strain 298 however resulted only in slight genomic alterations in the output strain 299, signifying that the bacterium has undergone sufficient genomic adaptations to establish a chronic infection. It is important to mention that our findings contrast the experimental outcome of Linz *et al.* who conducted a macaque infection study using the previously macaque-adapted *H. pylori* strain J166.²⁸ In their study, J166_{output} strains accumulated up to 12 SNPs within 1 week post infection. Our output strain 299, however, showed no SNPs 2 weeks after the infection. This suggests that there was a lack of acute inflammatory

Table 3. General features of pre- and post-correction *H. pylori* UM032, 298 and 299 genomes

Features	<i>H. pylori</i>					
	UM032		298		299	
	Pre	Post	Pre	Post	Pre	Post
Size (bp)	1,599,441	1,593,537	1,604,216	1,594,544	1,601,149	1,594,569
GC content (%)	38.8	38.8	38.8	38.8	38.8	38.8
CDSs	1,415	1,458	1,553	1,456	1,576	1,457
Genes	1,624	1,543	1,624	1,544	1,644	1,544
Pseudogenes	167	43	29	46	25	45
rRNAs	6	6	6	6	6	6
tRNAs	36	36	36	36	37	36

Table 4. COG analysis of UM032 genes with non-synonymous substitutions in its mice-adapted derivatives

COG	Function	Number	Locus tag
D	Cell cycle control, cell division, chromosome partitioning	1	UM032_0202
E	Amino acid transport and metabolism	1	UM032_0124
H	Coenzyme transport and metabolism	1	UM032_1143
L	Replication, recombination and repair	1	UM032_1365
M	Cell wall/membrane/envelope biogenesis	2	UM032_0124, UM032_1363
N	Cell motility	1	UM032_0447
T	Signal transduction mechanisms	1	UM032_0447

response, perhaps due to less variability between the gastric environments of mice, than between those of individual macaques.

Of note is that an amino acid change was detected in UM298_1343 relative to UM032_1343. This protein is known as the antigenic membrane-associated tumour necrosis factor α -inducing protein. The change replaces the isoleucine at position 111 for valine. Previous studies have shown that *H. pylori* TNF α -inducing protein is able to induce interleukin-1 α , TNF α , IL-8, and macrophage inflammatory protein 1 α productions in monocytes, as well as tumorigenesis in nude mice implanted with transfected Bhas 42 cells.^{29,30}

3.4. Intragenomic recombination between *babA* and *babB*

Among 65 nucleotide substitutions detected in 298 and 299 relative to UM032, 24 were observed within ORF UM032_1223, which was identified as the *babA* gene by BLASTN search against the *Helicobacter* group in NCBI database. Such a high number of mutations could reflect the strong selective pressure for rapid adjustment of BabA-mediated adhesion to facilitate adaptation and colonisation of different host gastric niche. However, these mutations were all synonymous substitutions. Furthermore, rather than being arbitrarily distributed across the 2226 bp gene, they were confined within the C-terminal region, prompting us to investigate whether the mutations could arise due to a recombination event.

Paralogous to *babA* are both *babB* and *babC* genes. Due to their extensive sequence identity at 5' and 3' regions, intragenomic recombination can occur within the *babABC* family, resulting in phenotypic change of Lewis b antigen binding capacity.³¹⁻³⁴ The *bab* genes reside in three different chromosomal locations, downstream of the *hypD* gene (locus A), the *rpsR* gene (locus B) and *hp0318* in *H. pylori* 26695 (locus C), respectively.^{31,35} Nevertheless not all *bab* genes are present within each *H. pylori* strain. There are some strains

which do not possess the *babA* or *babC* gene and some harbour two copies of *babA* or *babB* gene.^{36,37} There has also been an interesting observation by Kawai *et al.* that the *babC* locus, corresponding to *hp0317* in *H. pylori* 26695, is empty in all the hpEastAsia strains as a result of reductive evolution in the outer membrane protein families.³⁸

To test for a recombination event, we performed a BLASTN search of the C-terminal sequence fragment in UM032 genome and identified a complete match in the C-terminal region of a silent *babB* locus designated UM032_0908. Furthermore, pairwise nucleotide sequence alignment of UM032 *babA* and *babB*, and 298 *babA*, along with their downstream sequence indicated the presence of a double crossover recombination event as highlighted in Fig. 3. As there is no conversion of any amino acid residue, no functional change is deduced. It is however unclear whether the synonymous mutations may affect the *in vivo* translational speed and thus the expression of BabA in strains 298 and 299. To evaluate the possible translational outcome of these synonymous changes in *babA*, we compared the codon usage of the *babA* genes to the codon usage of highly expressed genes, specifically the ribosomal genes and the elongation factors, in the measurement of codon adaptation index (CAI) computed using EMBOSS CAI application.³⁹ The UM032 *babA* gene achieved a CAI score of 0.693, compared to a slightly lower value of 0.689 in the recombined version found in 298 and 299. Under such condition, the latter is less optimally encoded, by 0.4%, suggesting that there would only be marginal effect on protein expression levels.

3.5. Mutations in LPS biosynthesis genes

Among the 89 total variants identified in 298 relative to UM032, 10 were located in two genes designated UM298_1086 and UM298_1395. There is no genotypic difference present in these genes between strains 298 and 299. BLASTN comparison revealed that both genes encode for fucosyltransferase (FucT) that plays a

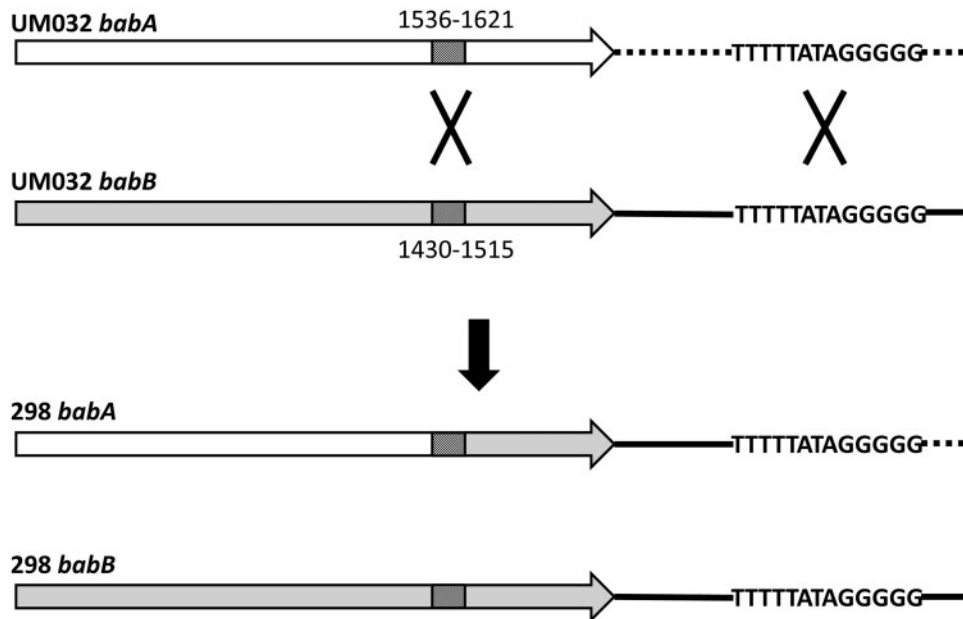


Figure 3. Recombination between *babA* and *babB* sequences.

substantial role in the synthesis of Lewis (Le) antigens of which the former gene product exhibits 75.9% amino acid sequence identity to HP0651 (FutB) whilst the latter harbours 78.3% identical amino acids to that of HP0379 (FutA). In *H. pylori* genomes, there are three phase-variable FucT-encoding genes, termed *futA*, *futB* and *futC*. FutA and FutB, both are paralogous and can fucosylate either Le^c antigen (type I carbohydrate backbone) in a α -(1,4) linkage to generate Le^a antigen, and/or N-acetyl-lactosamine (LacNAc) (type II carbohydrate backbone) in a α -(1,3) linkage to create Le^x antigen.^{40–42} Both Le^a and Le^x antigens can be further fucosylated by FutC in a α -(1,2) manner to create difucosylated Le^b or Le^y (Fig. 5).⁴³

In addition to the nucleotide variations found in FucT encoding genes, a 1-kb insertion was identified upstream of the *UM032_1363* counterpart in both the 298 and the 299 genomes, designated *UM298_1364* and *UM299_1364* respectively. *UM032_1363* shares approximately 80% nucleotide sequence identity with *HP0619* in *H. pylori* strain 26695, and *jhp0563* in *H. pylori* strain J99. *Jhp0563* is a β -(1,3)*galT* gene with a product essential for the expression of Le^c.⁴⁴ The upstream homologous glycosyltransferase-encoding gene *jhp0562*, is however absent in several *H. pylori* strains including 26695.⁴⁵ Due to a high degree of shared nucleotide sequence identity, it has been demonstrated that both *jhp0562* and β -(1,3)*galT* genes can undergo intragenic recombination within a single strain to generate functional chimeric alleles. This explains why certain strains do not possess a *jhp0562* allele.⁴⁵ Interestingly, the 1-kb insertion has introduced a gene which is 90.7% similar to *jhp0562* in the mice-derivative strains, designated 298_1363 and 299_1363 respectively. In this study, the mice were inoculated intragastrically with a pool consisting of 12 *H. pylori* clinical strains including UM032. The *jhp0562*-like allele was probably acquired by homologous recombination from one of the other strains inoculated.

To further examine how the genotypic changes above would affect Le antigen phenotypic expression, immunoblots were performed on proteinase K-treated whole cell lysates with antibodies detecting Le^a, Le^b, Le^x and Le^y (Fig. 4). All three strains reacted with the Le^b

antibody but differed in the sizes of decorated O-antigen chains. The immunoblot experiments also detected strong Le^y expression in strains 298 and 299. However, only a trace level of Le^y was detected in input strain UM032. Figure 5 shows the Le antigen synthesis model in UM032 with all anticipated glycosyltransferase-encoding genes listed. The detection of both Le^b and Le^y indicates the presence of the α -1,2-fucosylation process, which is expected since all three strains had an identical in-frame *futC* gene. Last but not least, none of these strains revealed any detectable levels of Le^a and Le^x (data not shown), suggesting that there is a relatively high activity of FutC in these strains, leading to complete α -1,2-fucosylation of these substrates once presented. In addition, terminal fucosylation could have masked internal Le^a or Le^x and therefore hindered antibody detection.

The minute expression of Le^y in strain UM032 might be either due to low level of endogenous α -1,3-FucT activity or a low LacNAc content and thus a reduced expression of Le^x. In strain UM032, the *futA* gene (*UM032_1394*) is switched off because its intragenic poly(C) tract located near the ATG start codon lacks one nucleotide, resulting in out-of-frame translation. The *futB* gene, by contrast, is in-frame, yielding correct full-length FucT enzyme in each strain. Given that FutA is inactive, the presence of both Le^b and Le^y in strain UM032 indicates that its FutB enzyme, encoded by *UM032_1086*, must contain both α -1,3 and α -1,4 properties. This is in agreement with our prediction by pairwise comparison that *UM032_1086* displays extensive amino acid identity of 93.1% with a previously documented α -1,3/4-FucT found in *H. pylori* strain UA948 (GenBank accession number AF194963) (Table 5).^{40,46}

However, it has been demonstrated that UA948FucT exhibits a greater than 5-fold affinity towards the type II chain precursor to generate Le^x. This does not support the hypothesis that a low level of α -1,3 activity is present within strain UM032. We next tested whether the minor changes in the amino acid composition could have altered the acceptor substrate affinity to make *UM032_1086* favour the type I over the type II chain. To address this question, we performed protein structure modelling comparing both UA948FucT

and UM032_1086. Figure 6 shows the models of UA948FucT and UM032_1086, each model based on the crystal structure of uncomplexed FucT (PDB ID: 2NZW). The catalytic sites accounting for interaction with the donor substrate, GDP-fucose, are highlighted. While it is accepted that structure modelling can be insensitive to minor amino acid changes, no structural differences were observed between the two models. Thus, it seems unlikely that there is distinct acceptor substrate specificity between UA948FucT and UM032_1086.

An alternative hypothesis is that the amount of LacNAc is inherently low within the UM032 strain, permitting FutB to utilise Le^c for Le^a synthesis once the LacNAc reservoir is depleted. To enhance LacNAc production and thus Le^y expression in strains 298 and 299,

an additional copy of glycosyltransferase gene must be present. This is consistent with the acquisition of a *jhp0562*-like allele in both strains. *jhp0562* has been demonstrated in previous mutagenesis and complementation studies as an essential component in both type I and type II Le synthesis pathways.⁴⁵ Despite displaying only 21% nucleotide sequence similarity to known β -(1,4)*galT* genes including *jhp0765* and *HP0826*, our findings offer further support for the idea that the product of this acquired *jhp0562*-like allele contains β -1,4-galactosyltransferase activity for the conversion of N-acetylglucosamine (GlcNAc) to LacNAc. Subsequently the expression of full-length FutA, which is expected to exhibit similar functions to that of FutB by conserving approximately 93.4% amino acids, in strains 298 and 299, provides additional α -1,3 enzymatic activity to produce Le^x that is immediately converted into Le^y by FutC.

The occurrence of genotypic and phenotypic changes in Le antigens in strain 298 relative to the input strain UM032, but not between strains 298 and 299, reflects an initial event of host-driven Le antigen expression adaptation taking place in *H. pylori* upon its first encounter to a new host species. This would aid the bacterium in establishment of long-term colonisation. The acquisition of *jhp0562*-like allele further suggests that the product of this non-phase-variable allele may further confer an increase in competitive advantage among *H. pylori* strains.

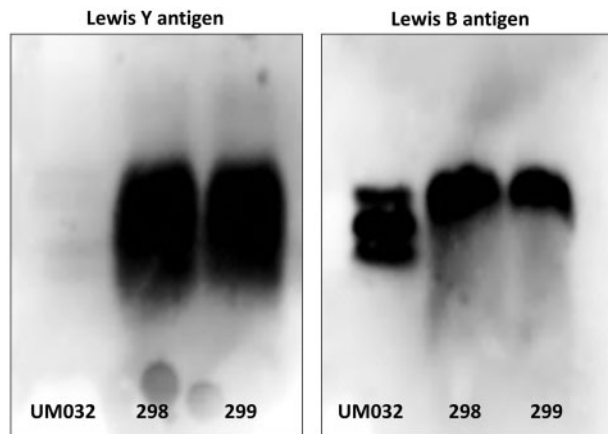


Figure 4. Immunoblot analysis of *H. pylori* strains UM032, 298 and 299 whole cell lysates with anti-Le^y and anti-Le^b antibodies.

3.6. Intergenic homopolymeric tract length alterations affect gene expression

In this study, a number of intragenic and intergenic homopolymeric tract length changes were detected in strains 298 and 299 (Tables 6 and 7). In order to verify if these variations in intergenic homopolymeric tract lengths influence the expression of adjacent genes, we performed qRT-PCR analysis to better understand how these transcriptional changes, if any, can possibly facilitate novel host

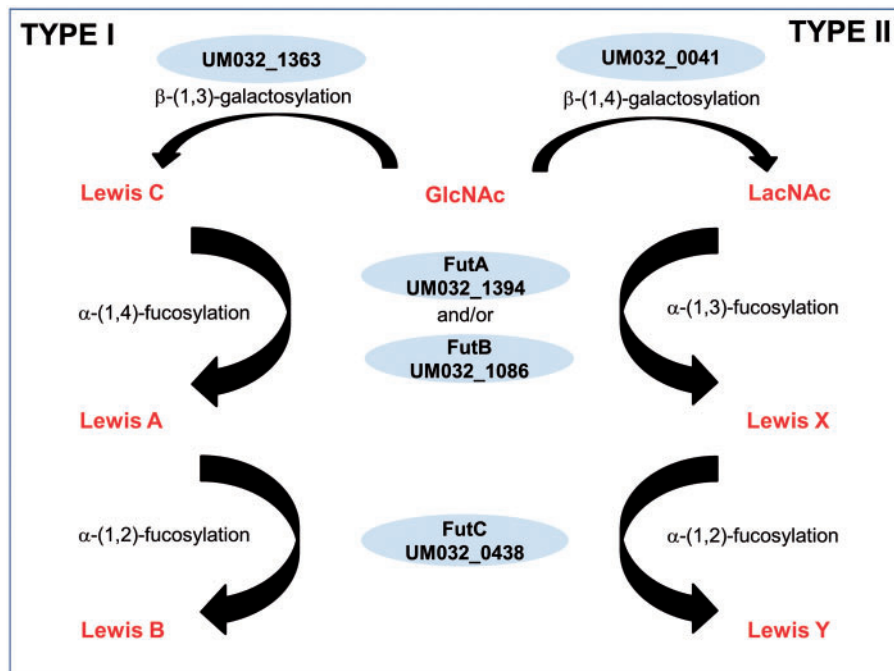


Figure 5. A schematic diagram of type I and type II Lewis antigen biosynthesis pathways in *H. pylori* strain UM032. GlcNAc, N-acetylglucosamine; LacNAc, N-acetyl-D-lactosamine.

colonisation. Data are shown as fold change relative to the transcriptional level of each corresponding gene in strain UM032, as depicted in Figure 7. Among nine tested gene candidates, five were significantly greater than four-fold down-regulated, including *UM032_0212* (hypothetical protein), *UM032_0548* (HopD), *UM032_0781* (biotin synthase), *UM032_0908* (BabB) and *UM032_1223* (BabA). A less than two-fold change was observed between *UM032_0213* (CTP synthase) and its counterparts. No amplification was detected for *UM032_0025*, *UM032_0547* and *UM032_1372* (data not shown), suggesting that these genes were either not expressed or were expressed at very low levels.

UM032_0213 mRNA levels remained unaffected by the length change in a poly(T) tract located 30 nucleotides upstream of the -35 element. However, *UM032_0212* transcription was significantly down-regulated at the poly(A) tract positioned 3 nucleotides upstream of the -35 element that was altered in length. The latter result was consistent with the findings of Åberg *et al.* that variation in the

length of homopolymeric tracts located adjacent to the -35 element modulates promoter activity by changing local DNA structure and thereby binding of the RNA polymerase.¹³ Similarly, transcriptional activity was also reduced in both outer membrane protein-encoding *UM032_0548* and *UM32_0908* as the poly(T) tract situated ~20 nucleotides upstream of each -35 element was reduced by single base pair in both strains 298 and 299. This indicates that regulation via variation in homopolymeric tract is a fairly general mechanism in *H. pylori*.

On the other hand, BabA-encoding *UM032_1223* had an altered poly(A) tract located between the -35 and -10 elements. A Western immunoblot was further conducted using anti-BabA antibody and as expected, BabA expression was nearly lost in strains 298 and 299 (Fig. 8). We also performed another Western analysis with commercial polyclonal antibody against *H. pylori* urease B (UreB) to demonstrate that comparable amounts of proteins are loaded in each lane (Fig. 9). It has been demonstrated in a number of previous studies

Table 5. Amino acid sequence identities between *H. pylori* fucosyltransferases

Fucosylation	<i>H. pylori</i> FucT	NCTC11639	UA948	DSM6709	UM032_1086	UM298_1086	UM298_1395
		Distance					
α -1,3	NCTC11639		0.22	0.14	0.22	0.22	0.2
α -1,3 and α -1,4	UA948	76.19		0.17	0.05	0.05	0.09
α -1,4	DSM6709	77.34	77.47		0.19	0.18	0.19
α -1,3 and α -1,4	UM032_1086*	75.16	93.1	77.78		0.01	0.06
α -1,3 and α -1,4	UM298_1086*	71.43	89.22	75.82	94.53		0.05
α -1,3 and α -1,4	UM298_1395*	73.03	83.94	75.55	88.26	93.39	
		% Identity					

*The enzymatic activity is predicted based on amino acid sequence comparison.



Figure 6. Protein structure modelling comparing both UA948FucT and UM032_1086. The catalytic sites accounting for interaction with the donor substrate, GDP-fucose, are highlighted.

Table 6. List of altered intragenic homopolymeric tracts

Product	UM032		298		299	
	Gene	Pseudo	Gene	Pseudo	Gene	Pseudo
Tetratricopeptide repeat family protein	UM032_0224	No	UM298_0224	Yes	UM298_0224	Yes
Oligopeptide transport system permease protein OppC	UM032_0307	No	UM298_0307	Yes	UM299_0307	Yes
Putative metal-dependent hydrolase fragment 1	UM032_0607	No	UM298_0607	Yes	UM299_0607	Yes
Hypothetical protein	UM032_0755	No	UM298_0755	Yes	UM299_0755	No
α -(1,3)-fucosyltransferase	UM032_1394	Yes	UM298_1395	No	UM299_1395	No

Table 7. Loci in strain UM032 with altered intergenic homopolymeric tracts relative to strains 298 and 299

Locus tag	Gene product	Length changes in output strains UM298 & UM299	Position	Tract length comparison in 49 <i>H. pylori</i> complete genomes excluding strains UM032, 298 and 299		
				Observed frequency	Maximum length (bp)	Minimum length (bp)
UM032_0025	Hypothetical protein	(A)14→13	≤ -35	48/49	19	7
UM032_0212	Hypothetical protein	(A)15→16	< -35	49/49	22	11
UM032_0213	CTP synthase	(T)15→16	≤ -35	49/49	22	11
UM032_0547	Putative endonuclease G	(A)16→15	-35/-10	15/49	17	11
UM032_0548	Outer membrane protein HopD	(T)16→15	< -35	49/49	21	8
UM032_0781	Biotin synthase	(G)12→13	≤ -35	46/49	14	7
UM032_0908	Outer membrane protein BabB	(T)14→13	≤ -35	27/49	22	7
UM032_1223	Outer membrane protein BabA	(A)12→13	-35/-10	44/49	15	8
UM032_1372	Hypothetical protein	(T)12→10	N/A*	49/49	18	8

*This ORF is the last gene in an operon.

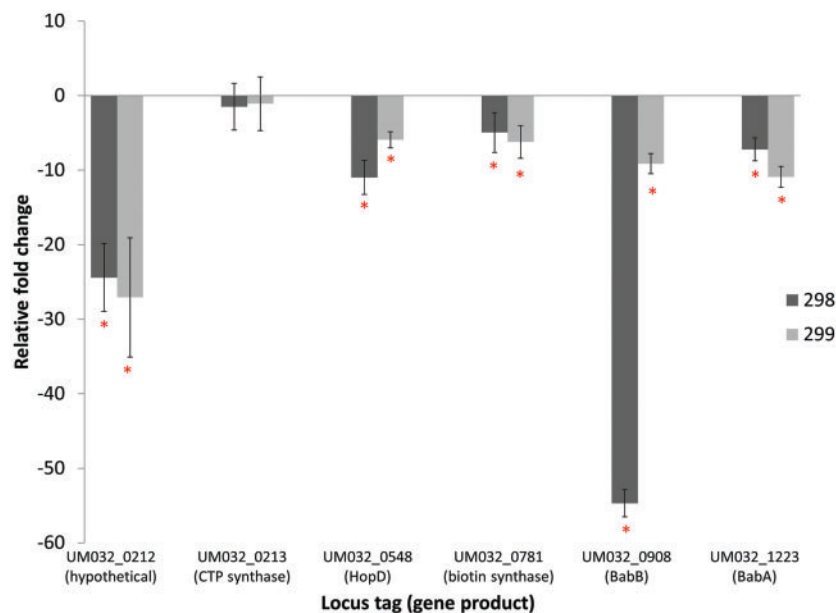


Figure 7. Real-time quantitation of genes identified with modified intergenic homopolymer-length. Data are expressed as fold change relative to strain UM032. The symbol * indicates statistical significance where $p < 0.05$.

that variation in the distance between these two consensus elements can affect RNA polymerase binding and thus the efficiency of gene transcription.^{47,48} Last but not least, the altered poly(G) tract of down-regulated *UM032_0781* was found to lie more than 150

nucleotides upstream of the -35 region. This is probably due to the presence of an unknown distal negative cis-acting element, similar to that found in the phase variable *NadA* adhesin of *Neisseria meningitidis*.⁴⁹

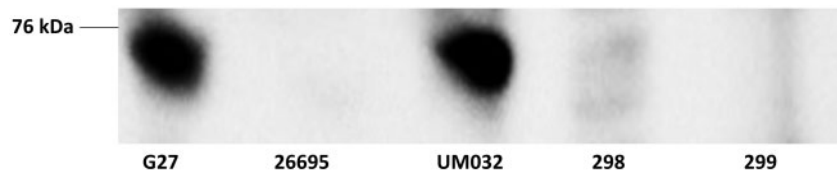


Figure 8. Western immunodetection of BabA in *H. pylori* strains UM032, 298 and 299. G27 and 26695 served as the positive and negative controls, respectively, in this assay.

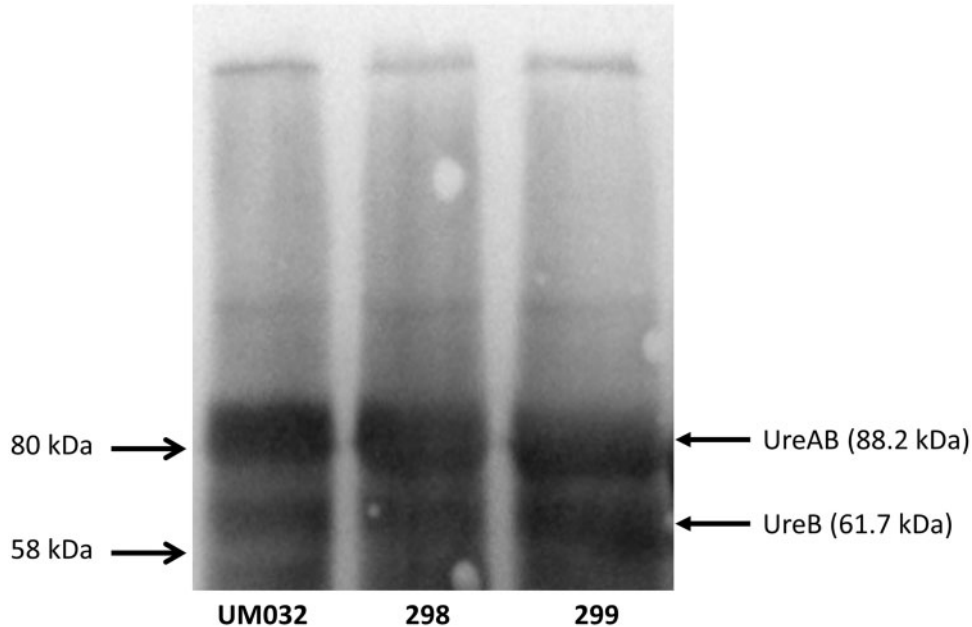


Figure 9. Western immunodetection of UreB in *H. pylori* strains UM032, 298 and 299.

It is unclear what functional roles biotin synthase and the hypothetical protein play in *H. pylori* host colonisation. Recently, it was reported that enterohemorrhagic *Escherichia coli* (EHEC) is able to regulate its adherence to intestinal epithelial cells by sensing surrounding biotin level.⁵⁰ Upon its arrival in the low-biotin large intestine, EHEC down-regulates the expression of its biotin protein ligase BirA to remove the repression on the global regulator Fur, thereby activating LEE (locus of enterocyte effacement) genes to promote bacterial adherence. In *H. pylori*, both BirA and Fur are also present (designated HP1140 and HP1027 in strain 26695, respectively).⁵¹ Nevertheless, it is uncertain in *H. pylori* if BirA may interact with Fur in a similar paradigm to that of *E. coli* as it lacks the N-terminal winged helix-turn-helix regulatory domain. It is tempting to hypothesise that the down-regulated expression of biotin synthase could possibly result in low intracellular biotin concentration, thus allowing derepression of Fur by BirA, which in turns activates genes involved in motility and chemotaxis to facilitate host colonisation.⁵² This warrants further investigation. The down-regulation of outer membrane protein genes, especially *babA*, however, is thought to increase bacterial dispersion and colonisation by preventing autoaggregation which might occur as a result of binding to LPS Le^b antigen.

3.7. Concluding remarks

When *H. pylori* encounters a hostile foreign environment, the bacterium rapidly adapts to the new environment via a series of genomic

alterations. Here, we demonstrated that a host change led to modification of the Lewis antigen profile of *H. pylori* lipopolysaccharide via acquisition of a *jhp0562*-like allele. In addition, expression levels of outer membrane proteins including BabA, BabB and HopD changed via altered homopolymeric tract lengths. These observations provide further evidence that rapid changes in membrane associated protein expressions play a major role in the early adaptation of bacterial populations to an individual host and these components are one of the key factors in *H. pylori*'s success as a pathogen.

Acknowledgements

This project was supported by the University of Malaya-Ministry of Education (UM-MoE) High Impact Research (HIR) grant (reference UM.C/625/1/HIR/MoE/CHAN13/3; Account No. H-50001-A000030), the National Health and Medical Research Council (grant no. 572723), the Vice Chancellor of the University of Western Australia, and the Western Australian Department of Commerce and Department of Health. SP was supported by SCELCE Microbiome Centre and grants from LKC School of Medicine, NTU University, Singapore. We thank Dr K. Mary Webberley for providing critical comments on this manuscript. We also thank Dr Hong Li (West China Hospital, Sichuan University) for providing blood group antigen-binding adhesin A (BabA) polyclonal antibody. We would also like to acknowledge Susana Wang, Primo Baybayan and Meredith Ashby of PacBio Biosciences (USA) and Siddarth Singh of PacBio Singapore for sequencing and assembling the original complete genomes of UM032, 298 and 299.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

References

- Marshall, B.J. and Warren, J.R. 1984, Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet*, **1**, 1311–15.
- Marshall, B.J. and Windsor, H.M. 2005, The relation of *Helicobacter pylori* to gastric adenocarcinoma and lymphoma: pathophysiology, epidemiology, screening, clinical presentation, treatment, and prevention. *Med. Clin. North Am.*, **89**, 313–44.
- 1994, Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7–14 June 1994. *IARC Monogr. Eval. Carcinog. Risks Hum.*, **61**, 1–241.
- Tay, C.Y., Mitchell, H., Dong, Q., Goh, K.L., Dawes, I.W. and Lan, R. 2009, Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. *BMC Microbiol.*, **9**, 126.
- Tomb, J. F., White, O., Kerlavage, A.R., et al. 1997, The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, **388**, 539–47.
- Pinto, A. V., Mathieu, A., Marsin, S., et al. 2005, Suppression of homologous and homeologous recombination by the bacterial MutS2 protein. *Mol. Cell*, **17**, 113–120.
- Kang, J., Huang, S. and Blaser, M. J. 2005, Structural and functional divergence of MutS2 from bacterial MutS1 and eukaryotic MSH4-MSH5 homologs. *J. Bacteriol.*, **187**, 3528–37.
- Benghezal, M. 2014, Persistence of *Helicobacter pylori* Infection: Genetic and Epigenetic Diversity. *InTech*.
- Garcia-Ortiz, M.V., Marsin, S., Arana, M.E., et al. 2011, Unexpected role for *Helicobacter pylori* DNA polymerase I as a source of genetic variability. *PLoS Genet.*, **7**, e1002152.
- van der Woude, M.W. 2011, Phase variation: how to create and coordinate population diversity. *Curr. Opin. Microbiol.*, **14**, 205–211.
- Pernitzsch, S.R., Tirier, S.M., Beier, D. and Sharma, C.M. 2014, A variable homopolymeric G-repeat defines small RNA-mediated posttranscriptional regulation of a chemotaxis receptor in *Helicobacter pylori*. *Proc Natl. Acad. Sci. USA*, **111**, E501–10.
- Appelmek, B. J., Martin, S. L., Monteiro, M. A., et al. 1999, Phase variation in *Helicobacter pylori* lipopolysaccharide due to changes in the lengths of poly(C) tracts in alpha3-fucosyltransferase genes. *Infect. Immun.*, **67**, 5361–66.
- Aberg, A., Gideonsson, P., Vallstrom, A., et al. 2014, A repetitive DNA element regulates expression of the *Helicobacter pylori* sialic acid binding adhesin by a rheostat-like mechanism. *PLoS Pathog.*, **10**, e1004234.
- Dubois, A., Berg, D.E., Incecik, E.T., et al. 1999, Host specificity of *Helicobacter pylori* strains and host responses in experimentally challenged nonhuman primates. *Gastroenterology*, **116**, 90–6.
- Day, A.S., Jones, N.L., Policova, Z., et al. 2001, Characterization of virulence factors of mouse-adapted *Helicobacter pylori* strain SS1 and effects on gastric hydrophobicity. *Digest Dis. Sci.*, **46**, 1943–51.
- Wirth, H.P., Beins, M.H., Yang, M., Tham, K.T. and Blaser, M.J. 1998, Experimental infection of Mongolian gerbils with wild-type and mutant *Helicobacter pylori* strains. *Infect. Immun.*, **66**, 4856–66.
- Khosravi, Y., Rehvathy, V., Wee, W.Y., et al. 2013, Comparing the genomes of *Helicobacter pylori* clinical strain UM032 and Mice-adapted derivatives. *Gut Pathog.*, **5**, 25.
- Chin, C. S., Sorenson, J., Harris, J.B., et al. 2011, The origin of the Haitian cholera outbreak strain. *New Engl. J. Med.*, **364**, 33–42.
- Perkins, T.T., Tay, C.Y., Thirriot, F. and Marshall, B. 2013, Choosing a benchtop sequencing machine to characterise *Helicobacter pylori* genomes. *PLoS One*, **8**, e67539.
- Kearse, M., Moir, R., Wilson, A., et al. 2012, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
- Aziz, R.K., Bartels, D., Best, A.A., et al. 2008, The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, **9**, 75.
- Besemer, J., Lomsadze, A. and Borodovsky, M. 2001, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.*, **29**, 2607–18.
- Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. 2010, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**, 119.
- Delcher, A.L., Bratke, K.A., Powers, E.C. and Salzberg, S.L. 2007, Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, **23**, 673–79.
- Darling, A.C., Mau, B., Blattner, F.R. and Perna, N.T. 2004, Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.*, **14**, 1394–403.
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., et al. 2015, CDD: NCBI's conserved domain database. *Nucleic Acids Res.*, **43**, D222–226.
- Livak, K.J. and Schmittgen, T.D. 2001, Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-Delta Delta C(T)} Method. *Methods*, **25**, 402–8.
- Linz, B., Windsor, H.M., McGraw, J.J., et al. 2014, A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nat. Commun.*, **5**, 4165.
- Yoshida, M., Wakatsuki, Y., Kobayashi, Y., et al. 1999, Cloning and characterization of a novel membrane-associated antigenic protein of *Helicobacter pylori*. *Infect. Immun.*, **67**, 286–93.
- Suganuma, M., Kurusu, M., Okabe, S., et al. 2001, *Helicobacter pylori* membrane protein 1: a new carcinogenic factor of *Helicobacter pylori*. *Cancer Res.*, **61**, 6356–59.
- Alm, R.A., Bina, J., Andrews, B.M., Doig, P., Hancock, R.E. and Trust, T.J. 2000, Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect. Immun.*, **68**, 4155–68.
- Styer, C.M., Hansen, L.M., Cooke, C.L., et al. 2010, Expression of the BabA adhesin during experimental infection with *Helicobacter pylori*. *Infect. Immun.*, **78**, 1593–600.
- Backstrom, A., Lundberg, C., Kersulyte, D., Berg, D.E., Boren, T. and Arnqvist, A. 2004, Metastability of *Helicobacter pylori* bab adhesin genes and dynamics in Lewis b antigen binding. *Proc. Natl. Acad. Sci. USA*, **101**, 16923–928.
- Nell, S., Kennemann, L., Schwarz, S., Josenhans, C. and Suerbaum, S. 2014, Dynamics of Lewis b binding and sequence variation of the babA adhesin gene during chronic *Helicobacter pylori* infection in humans. *MBio*, **5**.
- Colbeck, J.C., Hansen, L.M., Fong, J.M. and Solnick, J.V. 2006, Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of *Helicobacter pylori*. *Infect. Immun.*, **74**, 4375–78.
- Matteo, M.J., Armitano, R.I., Romeo, M., Wonaga, A., Olmos, M. and Catalano, M. 2011, *Helicobacter pylori* bab genes during chronic colonization. *Int. J. Mol. Epidemiol. Genet.*, **2**, 286–91.
- Ilver, D., Arnqvist, A., Ogren, J., et al. 1998, *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science*, **279**, 373–77.
- Kawai, M., Furuta, Y., Yahara, K., et al. 2011, Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. *BMC Microbiol.*, **11**, 104.
- Rice, P., Longden, I. and Bleasby, A. 2000, EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–77.
- Ma, B., Wang, G., Palcic, M.M., Hazes, B. and Taylor, D.E. 2003, C-terminal amino acids of *Helicobacter pylori* alpha1,3/4 fucosyltransferases determine type I and type II transfer. *J. Biol. Chem.*, **278**, 21893–900.

41. Ge, Z., Chan, N.W., Palcic, M.M. and Taylor, D.E. 1997, Cloning and heterologous expression of an alpha1,3-fucosyltransferase gene from the gastric pathogen *Helicobacter pylori*. *J. Biol. Chem.*, **272**, 21357–363.
42. Rabbani, S., Miksa, V., Wipf, B. and Ernst, B. 2005, Molecular cloning and functional expression of a novel *Helicobacter pylori* alpha-1,4 fucosyltransferase. *Glycobiology*, **15**, 1076–83.
43. Wang, G., Boulton, P.G., Chan, N.W., Palcic, M.M. and Taylor, D.E. 1999, Novel *Helicobacter pylori* alpha1,2-fucosyltransferase, a key enzyme in the synthesis of Lewis antigens. *Microbiology*, **145** (Pt 11), 3245–53.
44. Wang, G., Ge, Z., Rasko, D.A. and Taylor, D.E. 2000, Lewis antigens in *Helicobacter pylori*: biosynthesis and phase variation. *Mol. Microbiol.*, **36**, 1187–96.
45. Pohl, M.A., Kienesberger, S. and Blaser, M.J. 2012, Novel functions for glycosyltransferases Jhp0562 and GalT in Lewis antigen synthesis and variation in *Helicobacter pylori*. *Infect. Immun.*, **80**, 1593–605.
46. Rasko, D.A., Wang, G., Palcic, M.M. and Taylor, D.E. 2000, Cloning and characterization of the alpha(1,3/4) fucosyltransferase of *Helicobacter pylori*. *J. Biol. Chem.*, **275**, 4988–94.
47. Carson, S.D., Stone, B., Beucher, M., Fu, J. and Sparling, P.F. 2000, Phase variation of the gonococcal siderophore receptor FetA. *Mol. Microbiol.*, **36**, 585–93.
48. van Ham, S.M., van Alphen, L., Mooi, F.R. and van Putten, J.P. 1993, Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell*, **73**, 1187–96.
49. Metruccio, M.M., Pigozzi, E., Roncarati, D., et al. 2009, A novel phase variation mechanism in the meningococcus driven by a ligand-responsive repressor and differential spacing of distal promoter elements. *PLoS Pathog.*, **5**, e1000710.
50. Yang, B., Feng, L., Wang, F. and Wang, L. 2015, Enterohemorrhagic *Escherichia coli* senses low biotin status in the large intestine for colonization and infection. *Nat. Commun.*, **6**, 6592.
51. Chalker, A.F., Minehart, H.W., Hughes, N.J., et al. 2001, Systematic identification of selective essential genes in *Helicobacter pylori* by genome prioritization and allelic replacement mutagenesis. *J. Bacteriol.*, **183**, 1259–68.
52. Danielli, A., Roncarati, D., Delany, I., Chiarini, V., Rappuoli, R. and Scarlato, V. 2006, In vivo dissection of the *Helicobacter pylori* Fur regulatory circuit by genome-wide location analysis. *J. Bacteriol.*, **188**, 4654–62.

