

A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques

Bodo Linz^{1,2,*}, Helen M. Windsor³, John J. McGraw^{1,2}, Lori M. Hansen⁴, John P. Gajewski^{1,2}, Lynn P. Tomsho^{1,2}, Caylie M. Hake², Jay V. Solnick⁴, Stephan C. Schuster^{1,2}, Barry J. Marshall^{2,3}

¹ Dept. of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

² Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

³ School of Pathology, M504 University of Western Australia, Australia 6009

⁴ Departments of Medicine and Microbiology & Immunology, Center for Comparative Medicine, and the California National Primate Research Center, University of California, Davis, Davis, California 95616, USA

* Correspondence to:

Dr. Bodo Linz, Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, 311 Wartik Lab, University Park, PA 16802, USA. Tel: +1 814 8636053. Fax: +1 814 8636699. E-Mail: BXL29@PSU.EDU

ABSTRACT

The evolution rate and genetic changes that occur during chronic infection with *Helicobacter pylori* have been analyzed, but little is known about the genomic changes during the initial, acute bacterial infection phase. Here we analyse the rate and pattern of genome evolution in *H. pylori* from the genomes of two input strains isolated from human volunteers with asymptomatic infection, and the genomes of two output strains collected 20 and 44 days after re-infection. Similarly, we analyse genome evolution in bacteria from the genome sequences of input and output strains sequentially taken after experimental infection of a rhesus macaque. The estimated mutation rate reveals a mutation burst during the acute infection phase that is over ten times faster than the mutation rate during chronic infection, and orders of magnitude faster than mutation rates in any other bacteria. The elevated frequency of mutations in outer membrane protein genes suggests that the mutation burst facilitates rapid host adaptation of the bacteria.

The stomach bacterium *Helicobacter pylori* has infected anatomically modern humans since their origins^{1,2}, which implies perfect adaptation of the bacterial parasite to its human host. This remarkable adaptive capacity can be partly attributed to the enormous plasticity of the *Helicobacter* genome³ caused by high mutation and recombination rates⁴⁻⁶, and by the ability for aberrant genomic rearrangements and incorporation of non-homologous DNA. This results in an exceptionally high sequence diversity among strains that were isolated from unrelated human hosts^{1,2,7,8}, and in an *H. pylori* core genome that consists of only ~1100 genes while the other ~400 to 500 genes per strain are strain-specific and variably present⁹.

The initial infection with *H. pylori* induces an acute polymorphonuclear (PMN) reaction, most likely due to the activation of the innate immune system. The acute inflammatory response involves the release of reactive oxygen and nitrogen species, which were shown to trigger mutations and intra-chromosomal recombination *in vitro*¹⁰. After two to four weeks, the acute inflammation subsides and develops into a less severe chronic active gastritis, characterized by infiltration with PMNs, Th1 biased CD4+ T cells, plasma cells, and even dendritic cells, which generally persists throughout the host's life.

Numerous genomic changes have been identified in studies where sequential *H. pylori* isolates have been recovered several years apart from patients with chronic infection. Whole genome microarray analyses identified gene acquisitions and gene losses¹¹. In addition, sequencing of gene fragments^{4,12,13} and whole draft genomes^{5,6} revealed point mutations, imported DNA fragments and extensive recombination with co-colonizing strains, whereby genes encoding outer membrane proteins (OMPs) were at significantly increased frequency among the imported DNA fragments, suggesting diversifying selection⁵. During chronic infection, sequence differences accumulated in a clock-like manner⁴ at a median mutation rate of 1.55×10^{-5} changes per site per year and a recombination rate of 5.5×10^{-5} per initiation site per year^{5,6}. However, little is known about the mutation rate and genomic changes induced by the inflammatory response during initial, acute bacterial infection. We

hypothesized that the selection pressure during the acute inflammation would likely be higher than during chronic infection and result in a faster substitution rate. Here we test this hypothesis using experimental infection of human volunteers and a rhesus macaque. We determine the sequence differences that accumulate in the genomes of *H. pylori* output strains taken at several time points post infection, covering both the acute and chronic infection phases. Our genome comparisons reveal a mutation burst during the acute phase of *H. pylori* infection that is over ten times faster than the mutation rate during chronic infection and orders of magnitude faster than mutation rates in any other bacteria. We show that genetic changes in OMP genes accumulate at elevated frequencies suggesting that the mutation burst facilitates rapid adaptation of *H. pylori* to the individual human host.

RESULTS

We examined the short term genetic changes by comparing the completed genome sequences of two input strains with the genomes of two output strains collected only three and six weeks after re-infection of the original human hosts. In addition, we analyzed the genomic changes in three *H. pylori* strains each that were isolated 1 week, 1 month, 2 months and 6 months after experimental infection of a rhesus macaque.

Re-infection of human volunteers. Two human volunteers with asymptomatic *H. pylori* infections underwent gastroendoscopy and were then cured by antibiotic therapy. Eradication was confirmed one month and two months later by two consecutive, clearly negative carbon-14 urea breath tests (UBT). Three months later, the volunteers were re-infected with their own original strains BM012A and BM013A. *H. pylori* was grown from antral biopsy material that was collected during the initial endoscopic examinations as well as 20 days (strain BM013B) and 44 days (BM012B) after the re-infections. Visual inspection during the endoscopy as well as the histological analysis of the biopsies revealed moderate inflammation

of the stomach epithelium (Supplementary Fig. 1). Ultimately, *H. pylori* was eradicated with antibiotic treatment.

We sequenced the genomes of the two re-infection input strains BM012A and BM013A, and of the two output strains BM012B and BM013B using a Roche 454-Titanium FLX sequencer to >60X coverage, assembled the reads, and closed the genomes by PCR and Sanger sequencing. The original strain BM012A that was used for the re-infection experiment possessed a genome with a length of 1,660,425 bp (Fig. 1) that contained the *cag* pathogenicity island (*cag* PAI), nine copies of the insertion element IS607 and an apparently non-functional *vacA* gene that was disrupted by a frameshift. Strain BM013A had a genome of 1,604,233 bp length that did not possess the *cag* PAI and that contained an apparently non-cytotoxic *vacA* allele (s2m2) and four IS607 copies. An additional IS607 copy was located on a 16.7 kb plasmid.

Changes in *H. pylori* genomes during human re-infection. In comparison to the genome of the original strain BM012A, we identified 168 single nucleotide polymorphisms (SNPs, separated from the next sequence difference by >200 bp of identical sequence on both sides⁵) in the genome of the output strain BM012B that were scattered around the chromosome (Fig. 1, Table 1, Supplementary Tables 1 and 2). Variations in the length of homopolymeric nucleotide tracts were not included in the analysis. 22 SNPs were located in non-coding regions and 146 SNPs affected the coding regions of genes resulting in 73 synonymous and 73 non-synonymous mutations. In addition to SNPs, we identified 255 clusters of nucleotide polymorphisms (CNPs), which we defined as two or more sequence differences separated by <200 bp and flanked by >200 bp of identical sequence on both sides, similar to Kennemann et al⁵. 21 CNPs likely originated from intra-chromosomal translocations while 234 CNPs presumably resulted from imported foreign DNA (Fig. 1, Supplementary Tables 1 and 3). The incorporated DNA fragments had a mean length of 740 bp (range 9 - 6,192 bp), adding up to a total length of 173 kb, and likely resulted from 102 DNA fragments (groups of import) taken up from the environment.

The genome of strain BM013B contained 54 SNPs (Fig. 1B, Table 1, Supplementary Tables 1 and 4), four of which were non-coding while 50 were located in gene coding sequences and resulted in 15 synonymous and 35 non-synonymous mutations. Additionally, we identified 10 CNPs, only three of which originated from import of foreign DNA whereas the remaining seven CNPs likely arose by intra-chromosomal recombination (Supplementary Tables 1 and 5). All nucleotide changes were located on the bacterial chromosome, while no polymorphisms were detected in the plasmid sequence.

A mutation burst during the acute *H. pylori* infection phase. Based on the 54 SNPs that had accumulated in the genome of strain BM013B within 20 days and 168 SNPs that had accumulated in the genome of strain BM012B within 44 days after re-infection, we calculated a total mutation rate of $6.1\text{--}8.4 \times 10^{-4}$ changes per site per year (Fig. 1, Supplementary Table 1) for *H. pylori* during the acute infection phase. This mutation rate was 40 to 50 times faster than the median mutation rate of 1.55×10^{-5} (range $0.5\text{--}6.5 \times 10^{-5}$) estimated from whole genomes from paired sequential isolates during chronic infection^{5,6} (Supplementary Table 6), and orders of magnitude faster than mutation rates in other bacteria (Supplementary Table 7). In addition, the mutation rate estimate from chronic infection likely represents an overestimation because of functional inactivation of mutation repair genes *mutY* and *uvrC* by frameshift mutations in the strain with the highest mutation rate (Supplementary Table 6).

Our data suggest different mutation rates during particular stages of an *H. pylori* infection, which we attempted to analyze in more detail. Since there were no additional *H. pylori* isolates from acute human infection available to us, we used experimental inoculation of a rhesus monkey that results in persistent infection and a histological gastritis similar to that seen in humans^{14,15}. Previously, rhesus macaques were inoculated with the macaque adapted *H. pylori* strain J166, that contained a s1m1 *vacA* allele and a functional *cag* PAI that consisted of two fragments, *cagI* through *cagB* and *cagA*, separated by 12.5 kb of sequence

(Supplementary Fig. 2). Biopsies were taken under anaesthesia after an overnight fast one week, one month, two months and six months post orogastric *H. pylori* infection¹⁶. We sequenced the genome of strain J166 (1,650,561 bp) on a Roche 454-Titanium FLX sequencer and the genomes of twelve J166 macaque output (J166_{output}) strains, three individual bacterial strains cultivated from each of the four biopsies, on an Illumina MiSeq to an average coverage of 200X.

The number of point mutations in the J166_{output} strains increased in a time dependent manner, but this increase was not linear: a median of 8 SNPs accumulated within one week post infection, a median of 14 SNPs within one month, 17 SNPs within two months, and 21 SNPs within 6 months post infection (Table 2, Fig. 2A, Supplementary Fig. 2). The median number of mutations per sampling time corresponded to mutation rates of 2.5×10^{-4} changes per site per year within 1 week, of 1.0×10^{-4} within 1 month, of 6.3×10^{-5} within 2 months and of 2.6×10^{-5} within 6 months post inoculation (Table 2, Fig. 2B), revealing a peak in the mutation rate at the beginning of the infection. The mutation burst of 2.2×10^{-4} to 3.8×10^{-4} within one week was close to the estimated mutation rate of $6.1-8.4 \times 10^{-4}$ during the re-infection of the human volunteers (Fig. 1). The peak in the mutation rate was more prominent when the mutation rate was adjusted for the number of mutations that accumulated between the sampling points. On top of the 8 SNPs that accumulated between the inoculation and sampling point at day seven, the median number of SNPs consecutively increased by six, three and four additional SNPs at the sampling points after 1 month, 2 months and 6 months, respectively (Fig. 2A, Table 2). This corresponded to median mutation rates of 2.5×10^{-4} between the inoculation and 1 week, of 6.3×10^{-5} between 1 week and 1 month, of 2.2×10^{-5} between 1 month and 2 months, and of 0.74×10^{-5} between 2 months and 6 months, emphasizing the mutation burst during the acute infection phase. The mutation rates of 2.2×10^{-5} between 1 month and 2 months and of 0.74×10^{-5} between 2 months and 6 months

were within the range of those determined during chronic infection, indicating that the acute infection had likely subsided and changed into the chronic infection phase.

Diversifying selection in *H. pylori* during host adaptation. The 65 OMP encoding genes in the output strain BM012B contained 16 SNPs (Table 1), which was significantly more than expected given the total of 146 SNPs in 1591 coding genes (Fisher's exact test, $P = 0.001914$). Likewise, nine out of 73 non-synonymous SNPs (Table 1) were located in OMP genes ($P = 0.006657$). Additionally, 26 out of the 302 genes (8.6%) that were affected by the 234 CNPs in strain BM012B that arose from recombination of imported DNA (Supplementary Table 3) encoded OMPs, significantly more than the overall ratio of OMP genes in the genome (65/1591 genes, 4.1%; $P = 0.003398$). Similarly, two out of the six CNPs that affected genes in strain BM013B (Supplementary Table 5) altered OMP encoding genes ($P = 0.03823$), namely *hopM* (intra-chromosomal translocation from *hopF*) and *hopZ* (DNA import). OMP genes were also changed at high frequency in the rhesus macaque output strains, particularly the BabA adhesin encoding gene *hopS* that was altered in 11 of the 12 J166_{output} strains, indicating selection against this OMP (Supplementary Tables 8–19). In addition, frameshift mutations in the essential *cag* PAI genes *cagX* and *cagY* in J166_{output} strains isolated after 2 and 6 months apparently abolished the function of the *cag* PAI type IV secretion system that delivers the immunogenic effector protein CagA into the epithelial cells, suggesting selection against a functional *cag* PAI (Supplementary Fig. 2, Supplementary Tables 14–19).

DISCUSSION

The mutation burst of 2.2×10^{-4} to 8.4×10^{-4} changes per site per year that we estimated for the acute *H. pylori* infection phase was over one order of magnitude faster than the *H. pylori* mutation rate of 0.91×10^{-5} to 1.85×10^{-5} changes per site per year that was based on SNPs in 786 core genes in 40 *H. pylori* genome pairs⁶, and the median estimate of 1.55×10^{-5} (range 0.5×10^{-5} to 6.5×10^{-5}) determined from sequential isolates obtained from patients with chronic infection⁵.

This extraordinarily fast mutation rate during the acute infection phase is the highest mutation estimate so far for any bacteria, exceeding the substitution rate in other bacterial species by at least two orders of magnitude (Fig. 3, Supplementary Table 7). However, whole genome data on bacterial changes during the initial infection phase, such as presented here for *H. pylori*, are missing for other (pathogenic) bacteria; most of the previous mutation rates were determined from sequential isolates obtained from chronic infection or were estimated from phylogenetic comparisons of bacterial lineages obtained over many years or decades. For example, the accumulated mutations in an *Escherichia coli* clone that persisted in a family for three years gave a rate of 2.26×10^{-7} changes per site per year¹⁷. The genomic changes during chronic *Pseudomonas aeruginosa* infection in the airways of a cystic fibrosis patient for almost eight years resulted in a mutation rate of 1.3×10^{-6} changes per site per year¹⁸. Likewise, sequential *Staphylococcus aureus* isolates that were taken from the bloodstream of a patient over a period of 3 months revealed a rate of 4.9×10^{-6} changes per site per year¹⁹, which was approximately twice the population based estimate of $2.0\text{--}3.3 \times 10^{-6}$ for worldwide spreading methicillin-resistant *S. aureus* lineages^{20,21}. Mutation rates that are based on phylogenetic comparisons of bacterial populations are usually somewhat lower because of purifying selection that removes a considerable amount of non-synonymous mutations from the population. Accordingly, most changes across the bacterial population affect synonymous sites. Another example for a population based mutation rate is that from the genomes of the obligate bacterial symbiont *Buchnera aphidicola* from pea aphids that

diversified at a spontaneous mutation rate of $1.4\text{-}3.3 \times 10^{-7}$ substitutions per site per year over the past 135 years²². Somewhat higher mutation rates have been reported for *Mycoplasma gallisepticum* ($0.8\text{-}1.2 \times 10^{-5}$) isolated from house finches and poultry²³ and for the human pathogens *Neisseria gonorrhoeae*²⁴ (4.6×10^{-5}) and *Campylobacter jejuni*²⁵ (3.2×10^{-5}), but those were based on polymorphisms that have predominantly arisen by recombination between isolates rather than by mutation. If we had included recombinant sites, our mutation rate would have been $0.1\text{-}3.3 \times 10^{-2}$ changes per site per year.

The mutation burst during the acute *H. pylori* infection phase was almost as fast as the mutation rate of RNA viruses (Fig. 3) which evolve at a rate of 1.3×10^{-3} to 3.6×10^{-4} changes per site per year^{26,27}. We note however, that mutations in RNA viruses are primarily caused at the level of replication, which is not directly comparable to mutations in *H. pylori* that are likely caused by a combination of replication mistakes and chemical mutagenesis resulting from the immune response. In addition, RNA virus mutation rates have primarily been estimated from phylogenetic comparisons of viral lineages and were thus likely affected by purifying selection and possibly by population bottlenecks.

The significantly high frequency of mutational and recombinant changes in OMP genes of the human re-infection isolates and the J166_{output} strains suggested selection against bacterial surface structures during the acute infection. Similar to previous observation on J166_{output} strains from another macaque infection experiment²⁸, the *hopS* (*babA*) gene was either replaced by the closely related OMP gene *hopT* (*babB*) by intra-chromosomal translocation or contained a frameshift mutation that was caused by an extra CT dinucleotide repeat unit in the 5' part of the *hopS* gene (Supplementary Tables 8-19). As was shown previously, either gene conversion or frameshift in the poly(CT) tract functionally inactivates *babA*, resulting in loss of adherence to the Lewis^b blood group antigen on the gastric epithelium²⁸. These changes in the *H. pylori* *babA* expression in mice and in monkeys typically occurred within the first 4-8 weeks^{28,29}. In addition, the apparent loss of *cag* PAI function by frameshift mutations in the *cagX* and *cagY* genes (Supplementary Fig. 2,

Supplementary Tables 14-19) as well as by recombination in *cagY* in mice and in monkeys also happened in this time frame³⁰, implying adaptation by mutation and recombination during early stages of the infection.

Why did the *H. pylori* strains, that were isolated from asymptomatic human volunteers, trigger an inflammatory response when re-infecting the same persons? After the original infection was eradicated by antibiotic therapy, the chronic inflammation likely subsided. Following re-infection three months later, resident immune cells in the gastric mucosa such as macrophages and mastocytes detected the re-infecting *H. pylori* by recognition of general pathogen-associated molecular patterns such as LPS and by antibodies. Since several *H. pylori* OMP genes are variably expressed due to slipped strand mispairing at homopolymeric nucleotide tracts, this results in growth of clonally related *H. pylori* with many different sets of expressed OMPs, some of which were likely recognized by antibodies. Subsequently, the immune cells activated the acute inflammatory response by releasing inflammatory mediators such as TNF-alpha, which then in turn attracted neutrophils (PMNs) to the site of infection. Accordingly, histosections from biopsies taken after the re-infection revealed moderate inflammation, characterized by marked polymorphonuclear infiltration that extended from the surface into the lamina propria, while biopsies taken during the baseline endoscopy showed only mild cellular infiltration, consistent with asymptomatic, chronic *H. pylori* infection (Supplementary Fig. 1).

We cannot rule out the possibility that the patients were originally infected by a mix of strains and not a single strain. However, we cultured a mix of 8-12 colonies for antibiotic sensitivity testing in order to ensure that no resistance had developed to the antibiotics that we planned to use for eradication. All the strains (if more than one) in the stomach at the first endoscopy were sensitive to the 6 different antibiotics used (not shown), and the UBTs confirmed eradication. Nevertheless, the detection of numerous CNPs in output strain BM012B, which added up to a total of 173 kb of imported DNA (Supplementary Table 1), suggested (at least transient) co-infection with a different strain, which either was picked up

from somebody else sometime between the re-infection and the second endoscopy or was already present in the stomach during the time of the first endoscopy. However, the latter possibility is unlikely due to the two clearly negative UBTs separated by 1 month which is an excellent proof of cure, at least equal to biopsy and culture. If *H. pylori* had been present but at undetectably low levels at the time of UBT 1, the chances are very good that the infection would be picked up by the second UBT a few weeks later when more bacteria had re-established themselves.

Several factors seem to influence the extent of the acute inflammatory response and thus the level of the mutation burst, including the genetic composition of the bacterial strain such as the alleles of adhesins and the alleles of virulence factors such as *cagA* or *vacA*, the number of infecting bacteria, the genetic composition and health of the human (or primate) host and whether the immune system of the host was naïve to *H. pylori*. Consequently, the *H. pylori* mutation rate during the acute infection phase varied depending on the bacterial strain and host. While the mutation rates of 8.4×10^{-4} in the *cag* PAI positive strain BM012B and of 6.1×10^{-4} in the *cag* PAI negative strain BM013B from the human volunteers represent the highest estimates so far, the mutation rates of 3.8×10^{-4} within 1 week to 8.1×10^{-5} within 1 month in the *cag* PAI positive J166_{output} strains (Table 2) from the immunologically naïve macaque were up to one order of magnitude slower. A mutation rate of 4.5×10^{-5} was derived from 31 SNPs that accumulated in the genome of strain BM012S within 5 months following a naturally occurring transmission of *H. pylori* strain BM012A between spouses³¹. Extrapolating the accumulation of the 31 SNPs over the 5 months based on the increase in the J166_{output} strains would result in mutation rates of 5.0×10^{-4} within 1 week and of 1.6×10^{-4} within 1 month post infection. All these data contrast the accumulation of just 3 non-synonymous changes after 3 months of experimental *H. pylori* infection of another human volunteer⁵. The corresponding mutation rate of 7.5×10^{-6} , which is similar to that during chronic infection, implies the lack of an inflammatory response in this patient for reasons that are currently unknown but possibly involved a low number of 10^5 colony forming units (cfu) in the inoculum³² in comparison to 10^9 cfu in this re-infection study and to 7.5×10^8 cfu in the

macaque infection¹⁶, a *cag* PAI negative strain and infection of an immunologically naïve host³².

However, our data show that the inflammatory immune response during the acute infection phase, which is activated by the initial *H. pylori* infection of the stomach epithelium, triggers a mutation burst in the bacteria (Fig. 4). Since OMP encoding genes appear to be under constant selection by the immune system of the human host, the significantly increased frequency of mutation and recombination events in OMP genes in this (re-)infection study and during naturally occurring *H. pylori* transmission³¹ suggest that changes in immunogenic bacterial surface structures facilitate rapid adaptation of the bacterial pathogen to an individual host. This adaptation process includes evasion of the host's immune system and will ultimately result in the equilibrium of a chronic infection (Fig.4).

METHODS

Human *H. pylori* strains and ethics statement. The clinical part of this study was performed with written informed consent at the Sir Charles Gairdner Hospital, Nedlands, Western Australia, under ethics approval 2007-045 by the Human Ethics Committee of this hospital. Antral gastric biopsy specimens were taken from the study participants during gastroendoscopy and cultivated on Pylori agar plates (Biomerieux, Mercy - l'Etoile, France) from which the original “input” isolates, *H. pylori* strains BM012A and BM013A, were obtained. After single colony purification, the cultures were split. One part was stored at -80°C for the subsequent re-infection experiments, and the other part was subjected to DNA isolation for whole genome sequencing using the Qiagen Blood and Tissue Kit. After the first, “baseline” endoscopy, the human volunteers took eradication therapy with a cocktail of a proton pump inhibitor and three antibiotics (rabeprazole, amoxicillin, clarithromycin and tinidazole), and eradication was confirmed one month and two months later by two consecutive carbon-14 urea breath tests (UBT) (PYtest; Kimberly Clark, Utah, USA). According to the procedure guidelines of the Society of nuclear medicine³³ and the manufacturer, reference values of < 50 dpm are considered negative for *H. pylori*, 50 - 199 dpm as indeterminate for *H. pylori* and >200 dpm are interpreted as positive for *H. pylori*. UBTs were clearly negative for BM012A (0 dpm and 29 dpm) as well as for BM013A (0 dpm and 0 dpm). Three months after the successful *H. pylori* eradication by antibiotics therapy, the fasting volunteers drank 10^9 cfu of a two day old *H. pylori* culture suspended in 10 ml of beef broth. Two weeks after swallowing the inoculum, colonization with *H. pylori* was confirmed by positive UBTs for both BM012A (2298 dpm) and BM013A (1108 dpm). Twenty days (strain BM013B) and 44 days (strain BM012B) after the re-infection, endoscopic examination was repeated during which another antral biopsy was taken, from which the two output strains were cultivated. The bacteria were finally eradicated from the study participants. The study was recorded in the Australian Clinical Trial Registry at

http://www.anzctr.org.au/trial_view.aspx?ID=82275. Mucosal biopsies from the first and the second endoscopy were fixed in 10% formalin for histological analysis and processed routinely in the Pathology laboratory by staining with haematoxylin and eosin stain to assess cellular morphology and the presence or absence of bacteria. *H. pylori* was isolated as single colony cultures from the output strains, and purified DNA from all four strains was sent to the Pennsylvania State University for whole genome sequencing. Since no patient information was provided, the study was approved by the ethics board of the Pennsylvania State University as research not involving human participants (IRB#35170).

Experimental infection of a rhesus macaque. All animal procedures were performed at the California National Primate Research Center (CNPRC), which is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, in compliance with ethical regulations and in accordance with protocols approved by the CNPRC Research Advisory Committee and by the University of California, Davis, Chancellor's Animal Use and Care Administrative Advisory Committee. In a previous study, a 4 year old female rhesus macaque that was confirmed to be free of *H. pylori* and "Candidatus *Helicobacter heilmannii*" was inoculated orogastrically with 7.5×10^8 cfu of an overnight liquid culture of *H. pylori* J166 that was resuspended in 1.5 ml fresh brucella broth¹⁶. The monkey was biopsied at 1 week and 1, 2 and 6 months post inoculation under anesthesia after an overnight fast. DNA was isolated from J166 and from three *H. pylori* J166_{output} strains each that were grown from single colonies from antral biopsies recovered at each of the time points using the Qiagen DNeasy Blood and Tissue Kit.

454-Roche Titanium FLX whole genome sequencing. The genomes strains BM012A, BM012B, BM013A, BM013B and J166 were sequenced to >60 X coverage on a 454-Roche Titanium FLX sequencer. Prior to library preparation, the quality of the genomic DNA samples was assessed using a Bioanalyzer DNA 7500 Chip (Agilent). Sample quantitation was performed using Invitrogen's Picogreen assay. Next-generation sequencing library preparation was performed according to manufacturer's instructions (Roche Applied Science)

with the following modification. The samples were sheared on a Covaris S2 to ~700bp, following the manufacturer's recommendation. Library quantitation and emPCR were performed according to manufacturer's instructions (Roche Applied Science). The libraries were sequenced on a GS FLX instrument using GS FLX Titanium chemistry according to manufacturers instructions (Roche Applied Science). The sequence reads were assembled in NEWBLER³⁴, and genomic gaps were closed by PCR and Sanger sequencing. The genomes were annotated in RAST³⁵, followed by manual curation.

Illumina whole genome shotgun sequencing. The genomes of the twelve J166_{output} strains were sequenced on an Illumina MiSeq. Genomic DNA libraries were created on the automated Agilent Bravo NGS platform using the Illumina TruSeq Nano DNA HT Protocol. Samples were quantified using a Picogreen assay on a TBS-380 Mini-Fluorometer. 200 ng of sample was fragmented on the Covaris S2 in microtubules. Indexed adapters from the TruSeq Nano HT DNA Sample Prep Kit were ligated to the sheared DNA. Ligated products were purified and size selected to about 570 bp with Sample Selection Beads. The ligated and purified samples were then enriched by 8 PCR cycles according to the manufacturer's protocol. An aliquot of the library was run on the Agilent 2100 Bioanalyzer on a DNA 1000 chip. Library fragments averaged 570 bp, consisting of adapters of approximately 120 bp and inserts of 450 bp. Sample concentrations were quantified using KAPA Library Quantification Kit using Illumina's Sequencing Library qPCR Quantification Guide. Samples were normalized to 2nM. DNA was denatured and loaded onto the MiSeq following the protocol for Illumina MiSeq Reagent kit v2 (500 cycle). Sequencing was performed as a 250 x 250 bp paired end run with overlapping reads to an average 200X coverage. Paired-end reads were merged using FLASH³⁶ and subsequently assembled in NEWBLER³⁴.

Genome analyses. The identified nucleotide changes in strains BM012B and BM013B were validated by aligning the individual sequencing reads of strains BM012B and BM013B against the completed reference genome sequences of strains BM012A and BM013A using INGAP³⁷ and GSMAPPER implemented in NEWBLER³⁴. In addition, SNPs (separated from the next sequence difference by >200 bp of identical sequence on both sides⁵) were validated by

PCR and Sanger sequencing. We assumed that SNPs were introduced by mutation while clusters of 2 or more polymorphisms (CNPs, separated by <200 bp and flanked by >200 bp of identical sequence on both sides⁵) were likely the result of recombination. SNPs (Tables 1, 2; Supplementary Tables 1, 2, 4, 8–19) were used to calculate the mutation rate, whereby variations in the length of homopolymeric nucleotide tracts of ≥ 10 bp were not included in the analysis. Likewise, the identified nucleotide changes in the twelve J166_{output} strains were validated by aligning the individual sequencing reads against the completed reference genome sequence of strain J166.

The genomic differences (Fig. 1, Supplementary Fig. 2) were plotted using GenomeViz³⁸. For DNA fragments that were presumably introduced by recombination (Supplementary Tables 1, 3, 5, 8–19) we searched for possible intra-chromosomal DNA donor sources using the Artemis Comparison Tool³⁹. Only recombinant DNA fragments the origin of which could not be explained by intra-chromosomal recombination were assumed to have arisen by DNA import from an unrelated strain. The recombinant fragments in strains BM012B and BM013B were used to estimate the DNA import rate (Supplementary Table 1). The box and whisker plots (Fig. 2) were generated in R⁴⁰ based on the number of mutations and the corresponding mutation rates in Table 2. The band inside the box represents the median, the bottom and the top of the box represent the first and the third quartiles. The whiskers, which extend no more than 1.5 times the interquartile range from the box, indicate the range of the data. The R scripts are provided as Supplementary Notes 1 and 2.

The genomes of strains NQ315, NQ352, NQ367 and NQ392 that were isolated from patients with chronic infection⁵ were retrieved from GenBank (accession numbers CADE00000000, CADG00000000, CADL00000000, CADI00000000), the sequences of the genes involved in mutation and recombination repair (Supplementary Table 6) were extracted using BLAST searches.

References

- 1 Linz, B., *et al.* An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**, 915-918 (2007).
- 2 Moodley, Y., *et al.* Age of the association between *Helicobacter pylori* and man. *PLoS Pathog.* **8**, e1002693 (2012).
- 3 Kang, J. & Blaser, M. J. Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. *Nat. Rev. Microbiol.* **4**, 826-836 (2006).
- 4 Morelli, G., *et al.* Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. *PLoS Genet.* **6**, e1001036 (2010).
- 5 Kennemann, L., *et al.* *Helicobacter pylori* genome evolution during human infection. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5033-5038 (2011).
- 6 Didelot, X., *et al.* Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13880-13885 (2013).
- 7 Falush, D., *et al.* Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**, 1582-1585 (2003).
- 8 Moodley, Y., *et al.* The peopling of the Pacific from a bacterial perspective. *Science* **323**, 527-530 (2009).
- 9 Gressmann, H., *et al.* Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet.* **1**, e43 (2005).
- 10 Kang, J. M., Iovine, N. M., & Blaser, M. J. A paradigm for direct stress-induced mutation in prokaryotes. *FASEB J.* **20**, 2476-2485 (2006).
- 11 Israel, D. A., *et al.* *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14625-14630 (2001).
- 12 Falush, D., *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.* **98**, 15056-15061 (2001).
- 13 Prouzet-Mauleon, V., *et al.* Pathogen evolution *in vivo*: genome dynamics of two isolates obtained 9 years apart from a duodenal ulcer patient infected with a single *Helicobacter pylori* strain. *J. Clin. Microbiol.* **43**, 4237-4241 (2005).
- 14 Dubois, A., *et al.* Transient and persistent experimental infection of nonhuman primates with *Helicobacter pylori*: implications for human disease. *Infect. Immun.* **64**, 2885-2891 (1996).
- 15 Solnick, J. V., *et al.* Determination of the infectious dose of *Helicobacter pylori* during primary and secondary infection in rhesus monkeys (*Macaca mulatta*). *Infect. Immun.* **69**, 6887-6892 (2001).
- 16 Boonjakuakul, J. K., Canfield, D. R., & Solnick, J. V. Comparison of *Helicobacter pylori* virulence gene expression *in vitro* and in the Rhesus macaque. *Infect. Immun.* **73**, 4895-4904 (2005).
- 17 Reeves, P. R., *et al.* Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. *PLoS ONE* **6**, e26907 (2011).
- 18 Smith, E. E., *et al.* Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8487-8492 (2006).
- 19 Mwangi, M. M., *et al.* Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9451-9456 (2007).
- 20 Nubel, U., *et al.* A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog.* **6**, e1000855 (2010).

- 21 Harris, S. R., *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469-474 (2010).
- 22 Moran, N. A., McLaughlin, H. J., & Sorek, R. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* **323**, 379-382 (2009).
- 23 Delaney, N. F., *et al.* Ultrafast Evolution and Loss of CRISPRs Following a Host Shift in a Novel Wildlife Pathogen, *Mycoplasma gallisepticum*. *PLoS Genet.* **8**, e1002511 (2012).
- 24 Perez-Losada, M., *et al.* Temporal trends in gonococcal population genetics in a high prevalence urban community. *Infect. Genet. Evol.* **7**, 271-278 (2007).
- 25 Wilson, D. J., *et al.* Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Mol. Biol. Evol.* **26**, 385-397 (2009).
- 26 Jenkins, G. M., *et al.* Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* **54**, 156-165 (2002).
- 27 Holmes, E. C. Evolutionary history and phylogeography of human viruses. *Annu. Rev. Microbiol.* **62**, 307-328 (2008).
- 28 Solnick, J. V., *et al.* Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2106-2111 (2004).
- 29 Styer, C. M., *et al.* Expression of the BabA adhesin during experimental infection with *Helicobacter pylori*. *Infect. Immun.* **78**, 1593-1600 (2010).
- 30 Barrozo, R. M., *et al.* Functional plasticity in the type IV secretion system of *Helicobacter pylori*. *PLoS Pathog.* **9**, e1003189 (2013).
- 31 Linz, B., *et al.* *Helicobacter pylori* genomic microevolution during naturally occurring transmission between adults. *PLoS ONE* **8**, e82187 (2013).
- 32 Aebischer, T., *et al.* Correlation of T cell response and bacterial clearance in human volunteers challenged with *Helicobacter pylori* revealed by randomised controlled vaccination with Ty21a-based Salmonella vaccines. *Gut* **57**, 1065-1072 (2008).
- 33 Balon, H., *et al.* Procedure guideline for carbon-14-urea breath test. *Journal of Nuclear Medicine* **39**, 2012-2014 (1998).
- 34 Margulies, M., *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
- 35 Aziz, R. K., *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC. Genomics* **9**, 75 (2008).
- 36 Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* **27**, 2957-2963 (2011).
- 37 Qi, J., *et al.* inGAP: an integrated next-generation genome analysis pipeline. *Bioinformatics.* **26**, 127-129 (2010).
- 38 Ghai, R., Hain, T., & Chakraborty, T. GenomeViz: visualizing microbial genomes. *BMC. Bioinformatics.* **5**, 198 (2004).
- 39 Carver, T. J., *et al.* ACT: the Artemis Comparison Tool. *Bioinformatics.* **21**, 3422-3423 (2005).
- 40 Ihaka, R. & Gentleman, R. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314 (1996).

Figure legends.

Figure 1. Accumulation of SNPs and CNPs in the *H. pylori* genomes after re-infection of human hosts. (Circle contents from outside to inside). Circle 1: Genomes of *H. pylori* strains BM012A (A) and BM013A (B) with the positions of IS607 (red), the *cag* PAI, the *vacA* gene and ticks every 100 kb. Circle 2: Distribution of non-coding (black), non-synonymous (red) and synonymous (blue) SNPs in the genomes of the output strains BM012B (168 SNPs) and BM013B (54 SNPs). Circle 3: Distribution of imported (orange) and translocated (green) DNA fragments in the genomes of the two output strains. syn SNP – synonymous SNP; non-syn SNP – non-synonymous SNP.

Figure 2. Accumulation of SNPs and mutation rates in the macaque J166_{output} strains.

A) Box-and-whisker plot of the cumulative number of mutations in 3 J166_{output} strains each that were taken 1 week, 1 month, 2 months and 6 months post infection (blue boxes). The adjusted number of mutations that accumulated in addition to the median number of mutations at the previous sampling point is shown in white boxes. B) Box-and-whisker plot of the estimated mutation rate between the inoculation day and the sampling point (blue boxes). The estimated mutation rates between the inoculation and 1 week, between 1 week and 1 month, between 1 month and 2 months, and between 2 months and 6 months is shown as white boxes. The band inside the box represents the median, the bottom and the top of the box are the first and the third quartiles. The whiskers indicate the range of the data.

Figure 3. Mutation rate *versus* genome size for bacteria and other biological entities.

The estimated *H. pylori* mutation rate during the acute infection phase was orders of magnitude faster than the mutation rate in other bacteria, comparable to the evolutionary rate

of RNA viruses. Data points were taken from references in Supplementary Table 7. ssDNA – single stranded DNA; dsDNA – double stranded DNA.

Figure 4. Model of the mutation rate over the course of the infection. The initial bacterial infection of the human stomach epithelium activates an acute inflammatory response, which triggers a mutation burst in the genome of the bacterial parasite. This mutation burst facilitates rapid adaptation to the individual human host, finally resulting in the equilibrium of a chronic infection.

Table 1. SNPs in the genomes of the output strains BM012B and BM013B.

Strain	SNPs	Gene function (number of SNPs)			
		Total	Housekeeping	Hypothetical	OMP
SNPs in strain BM012B	Non-coding	22	-	-	-
	Synonymous	73	54	12	7
	Non-synonymous	73	47	17	9
	Total	168	101	29	16
SNPs in strain BM013B	Non-coding	4	-	-	-
	Synonymous	15	10	3	2
	Non-synonymous	35	22	11	2
	Total	54	32	14	4

Table 2. Number of accumulated SNPs and mutation rates in the J166_{output} strains.

Strain	Sampling time post infection	Mutation		Infection time		Adjusted mutation*	
		number	rate	from	to	number	rate
1wkA	1 week	12	3.8x10 ⁻⁴	Inoculation	1 week	12	3.8x10 ⁻⁴
1wkB	1 week	8	2.5x10 ⁻⁴	Inoculation	1 week	8	2.5x10 ⁻⁴
1wkC	1 week	7	2.2x10 ⁻⁴	Inoculation	1 week	7	2.2x10 ⁻⁴
1moA	1 month	14	1.0x10 ⁻⁴	1 week	1 month	6	6.3x10 ⁻⁵
1moB	1 month	11	8.1x10 ⁻⁵	1 week	1 month	3	3.2x10 ⁻⁵
1moC	1 month	15	1.1x10 ⁻⁴	1 week	1 month	7	7.4x10 ⁻⁵
2moA	2 months	16	5.9x10 ⁻⁵	1 month	2 months	2	1.5x10 ⁻⁵
2moB	2 months	18	6.6x10 ⁻⁵	1 month	2 months	4	2.9x10 ⁻⁵
2moC	2 months	17	6.3x10 ⁻⁵	1 month	2 months	3	2.2x10 ⁻⁵
6moA	6 months	22	2.7x10 ⁻⁵	2 months	6 months	5	9.2x10 ⁻⁶
6moB	6 months	20	2.5x10 ⁻⁵	2 months	6 months	3	5.5x10 ⁻⁶
6moC	6 months	21	2.6x10 ⁻⁵	2 months	6 months	4	7.4x10 ⁻⁶

* based on the number of mutations that accumulated in addition to the median number of mutations at the previous sampling point.

Figure-1 Linz

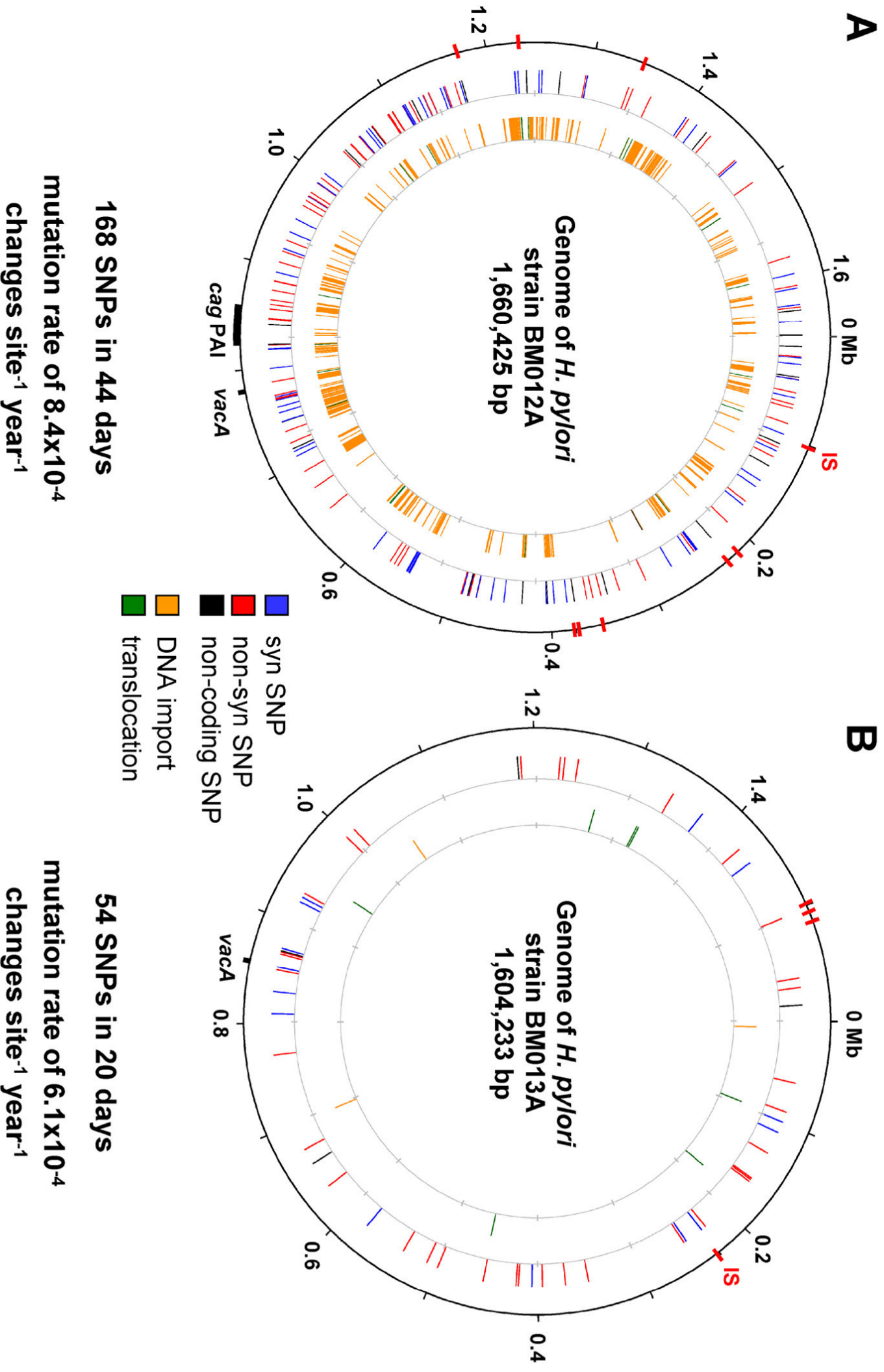
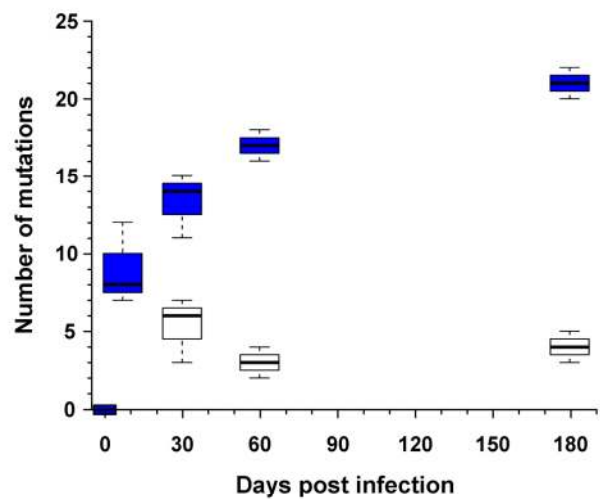


Figure-2 Linz

A



B

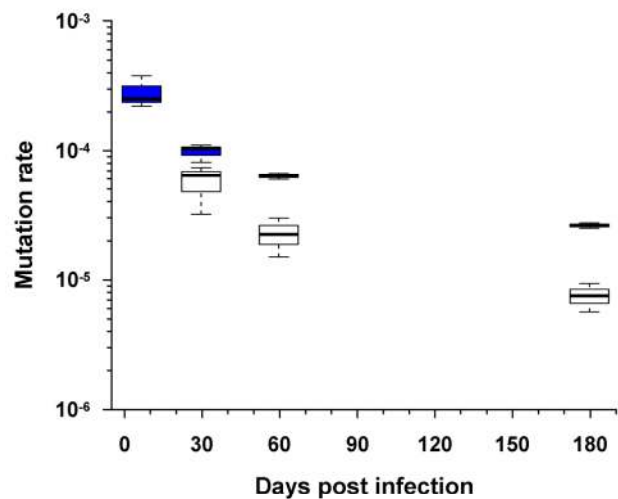


Figure-3 Linz

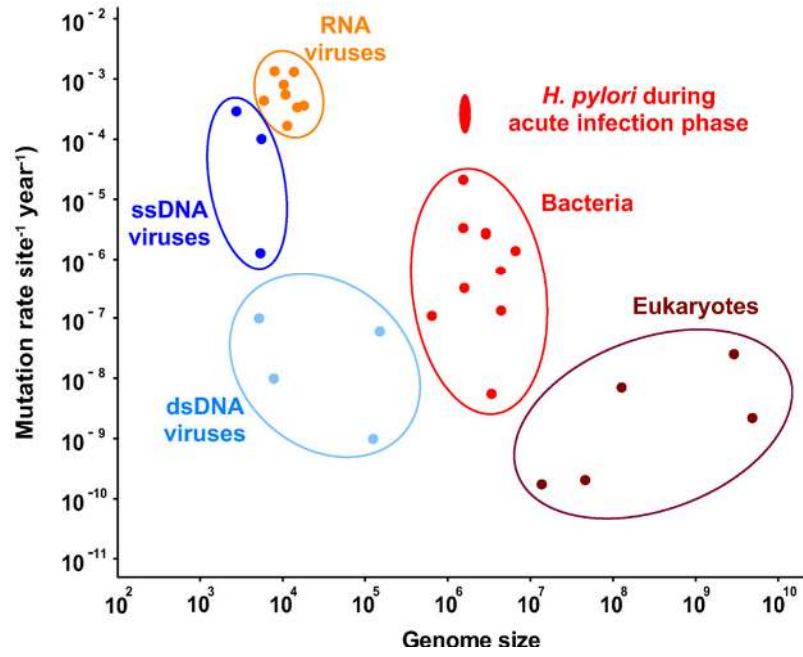


Figure-4 Linz

