Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps

Florent Ailloud, Xavier Didelot, Sabrina Woltemate, Gudrun Pfaffinger, Jörg Overmann, Ruth Christiane Bader, Christian Schulz, Peter Malfertheiner, & Sebastian Suerbaum

The human pathogen *Helicobacter pylori* displays extensive genetic diversity. While *H. pylori* is known to evolve during infection, population dynamics inside the gastric environment have not been extensively investigated. Here we obtained gastric biopsies from multiple stomach regions of 16 *H. pylori*-infected adults, and analyze the genomes of 10 *H. pylori* isolates from each biopsy. Phylogenetic analyses suggest location-specific evolution and bacterial migration between gastric regions. Migration is significantly more frequent between the corpus and the fundus than with the antrum, suggesting that physiological differences between antral and oxyntic mucosa contribute to spatial partitioning of *H. pylori* populations. Associations between *H. pylori* gene polymorphisms and stomach niches suggest that chemotaxis, regulatory functions and outer membrane proteins contribute to specific adaptation to the antral and oxyntic mucosa. Moreover, we show that antibiotics can induce severe population bottlenecks and likely play a role in shaping the population structure of *H. pylori*.
Helicobacter pylori is a Gram-negative bacterial pathogen that infects more than half of the human population worldwide. It is usually acquired during childhood and able to establish lifelong chronic infection. Infected patients are asymptomatic in most cases but clinical complications can range from gastric or duodenal ulcers to gastric atrophy and, ultimately, gastric cancer or mucosa-associated lymphoid tissue (MALT) lymphoma. The likelihood of clinical sequelae has been linked to specific virulence factors of H. pylori, most importantly the cag pathogenicity island (cagPAI) and the VacA cytotoxin. At present, recommended treatment strategies usually comprise two antibiotics, a proton-pump inhibitor, and, in some regimens, bismuth salts. These treatment protocols achieve eradication rates ~90%. Nonetheless, increasing antibiotic resistance has been reported in H. pylori, similar to other multi-drug-resistant bacteria.

H. pylori is characterized by an exceptionally high genetic diversity and variability. The underlying mechanisms include relatively inefficient DNA repair mechanisms as well as natural competence for transformation and the ability to integrate small fragments of homologous DNA into the chromosome. The species is divided into large phylogeographic populations with distinct geographical distributions, allowing the use of H. pylori genotypes to retrace historical human migration events.

The rates and patterns of genetic change of the H. pylori genome during chronic infection have been the subject of several studies, based on increasing amounts of sequencing data. In recent years, whole-genome sequence comparisons have been used extensively to study the in vivo evolution of H. pylori during the early stages of infection in challenged human volunteers and after long-term colonization in chronically infected individuals. Genes encoding outer membrane proteins (OMPs) appear to evolve faster than genes of other functional categories, suggesting adaptation to individual hosts. Overall, these studies consistently determined a mutation rate of the order of 10⁻⁵ mutations per site per year.

The extent of the genetic diversity of the H. pylori population within a single infected stomach at a given time is much less characterized, at least compared with other bacterial pathogens. The only known natural reservoir of H. pylori is the human stomach (including areas of gastric metaplasia in the duodenum). It is divided into three main anatomical regions: the pyloric antrum, connected to the duodenum; the corpus in the center and the fundus in the upper curvature, next to the cardia. The analysis of single isolates from separate gastric regions revealed that strains from the same host can be genetically differentiated. Stomach regions contain different human cell types providing distinct environmental conditions for H. pylori subpopulations, and these could lead to specific niche adaptation. For example, the differences in pH between antrum and corpus were shown to select for variant BabA adhesins with different kinetics of pH-dependent binding affinities. Differential antibiotic resistance profiles have also been observed in related isolates isolated in the antrum and in the corpus, suggesting that H. pylori populations can be structured by stomach anatomy.

The time since the most recent common ancestor (TMRCA) between isolates from the same stomach has been calculated from genomic data using pairs of strains isolated from antrum and corpus. Surprisingly, TMRCA values appeared to be independent of the host’s age and were only a few years on average. Mutation rates calculated in acute and chronic phases of the infection are comparable, thus the TMRCA of a H. pylori population was expected to increase linearly beginning at the infection during childhood. Population bottlenecks and strain turnovers are possible explanations for the low average TMRCA, which could be induced by selective pressures from the host immune system as well as common antibiotic consumption.

The aim of this study was to characterize H. pylori population diversity and dynamics within the stomach using bacterial genomics. Multiple single colonies of H. pylori were isolated from several gastric regions in naturally infected adults and analyzed by whole-genome sequencing. Our analyses reconstructed the pathways of migration of bacterial subpopulations between parts of the stomach, and identify several candidate genes with high frequencies of within-host variation and/or genetic signals of association to specific regions of the stomach, suggesting these loci participate in local adaptation to gastric niches.

Results
Extensive within-host diversity in H. pylori populations. In order to quantify the genome-wide diversity of H. pylori within the human stomach, gastric biopsies were obtained from 16 H. pylori-infected individuals (mean age, 49.9 years, nine women, seven men, Table 1). Twelve of the 16 patients displayed chronic gastritis in histology, whereas four had atrophic gastritis, including one patient with intestinal metaplasia. Five patients had a known history of one or multiple attempted eradication therapies. Gastric biopsies were taken from the antrum, corpus, and fundus regions for 10 of these patients. For the six remaining patients, only antrum and corpus were sampled. H. pylori was cultured from the biopsies, and from each biopsy, 10–15 single colonies were picked, propagated, and the genome sequences of these clones determined with an Illumina MiSeq instrument. As this short-read sequencing approach resulted in draft sequences consisting of 39 contigs on average, we selected 1–3 isolates from every patient, and additionally sequenced their genomes with Single Molecule Real-Time (SMRT) technology, which yields longer reads and permits to assemble a closed complete-genome sequence. The total data set consisted of genomes from 440 H. pylori isolates, including 414 draft sequences plus 26 closed sequences (Supplementary Data 1).

All isolates from the 16 patients could be assigned to the phylogeographic population hpEurope. The average within-host pairwise nucleotide identity ranged from 98.5 to 99.8%, and isolates from a given patient clustered tightly together in a global phylogenetic tree (Fig. 1a), suggesting we did not sample any mixed infection with unrelated H. pylori strains. The amount of genetic diversity observed within the group of H. pylori isolates from one individual patient was highly variable between patients (Fig. 1b). More than 18,000 unique SNPs were identified within the most diverse population (clones isolated from patient 476), whereas only 73 SNPs were found in the least diverse population (from patient 173).

Out of 16 patients, cagPAI-positive and babA-positive strains were isolated from 10 and 12 patients, respectively. All cagPAI-positive patients were babA-positive as well. All BabA sequences had residues in position 198 associated with generalist binding preference, except in patient 23 that had a glutamine typically found in specialist strains. In patient 476, related cagPAI-positive and -negative isolates were identified. Both cagPAI+ and cagPAI− isolates were distributed in several clades with no association with gastric regions. CagA was still present in all isolates, and a second copy with two additional EPIYA-C motifs was found in one of the several cagPAI-positive clades of this patient (Table 1). The microdiversification of the cagPAI in this patient suggests a complex history of gene gain and loss driven by fluctuating selective pressures within different niches of the stomach and likely made possible by secondary infections.
Table 1 Overview of *H. pylori* populations and their major characteristics

| Patient | Age (y) | TMRCA (y) | Patho. a | Isolates (A-C-F) | Treatment b | CLR c | CIP c | MTZ c | Unique SNPs | r/m d | Rec. events e [length(kbp)] cagPAI |
|---------|---------|-----------|----------|-----------------|-------------|-------|------|-------|-------------|-------|----------------------------------|
| 173     | 44      | 0.34      | CG       | 10-10-NA        | Y/N         | R     | R    | R     | 73          | 1.2   | [11(0.2)]                        |
| 478     | 70      | 1.59      | AG/IM    | 10-10-NA        | Y/Y         | S     | R    | R     | 225         | 0.3   | [7(0.7)]                         |
| 280     | 76      | 0.51      | CG       | 10-10-NA        | Y/Y         | S     | R    | R     | 87          | 0.9   | [7(0.4)]                         |
| 169     | 50      | 2.09      | CG       | 10-10-NA        | N/N         | R     | S    | S     | 1919        | 9.0   | [109*[45]]                      |
| 476     | 34      | 7.19      | CG       | 15-15-NA        | N/N         | S     | S    | S     | 18147       | 16.3  | [1126*[423.6]]                  |
| 479     | 69      | 4.5       | AG       | 15-15-NA        | N/N         | S     | S    | S     | 13295       | 15.7  | [562*[321.3]]                   |
| 21      | 58      | 4.04      | CG       | 10-10-10        | N/N         | S     | S    | S     | 1066        | 0.6   | [53*[3.1]]                      |
| 381     | 42      | 1.5       | AG       | 10-10-10        | Y/N         | R21S9  | S    | S     | 503         | 1.0   | [20*[2.6]]                      |
| 5       | 49      | 3.24      | CG       | 10-10-10        | N/N         | S     | S    | R     | 849         | 0.6   | [39*[2.6]]                      |
| 26      | 47      | 1.57      | AG       | 10-10-10        | NA/N        | S     | S    | S     | 409         | 0.4   | [4*[0.4]]                       |
| 23      | 30      | 1.87      | CG       | 10-10-10        | N/N         | R23S7  | S    | S     | 424         | 0.3   | [15*[0.6]]                      |
| 13      | 54      | 3.45      | CG       | 10-10-10        | N/N         | S     | S    | R     | 1108        | 0.8   | [45*[4.8]]                      |
| 20      | 55      | 1.27      | CG       | 10-10-10        | Y/Y         | R     | S    | R14S16 | 415         | 0.9   | [22*[1.7]]                      |
| 24      | 33      | 3.29      | CG       | 10-10-10        | N/N         | S     | S    | S     | 1156        | 0.3   | [29*[1.2]]                      |
| 25      | 56      | 3.31      | CG       | 10-10-10        | NA/N        | S     | S    | R18S12 | 3684        | 9.2   | [135*[87.1]]                    |
| 19      | 32      | 3.21      | CG       | 10-10-10        | N/N         | S     | S    | R     | 2632        | 3.3   | [177*[33.4]]                    |

aCG: chronic gastritis; AG: atrophic gastritis; IM: intestinal metaplasia

bTreatments undergone by the patient prior to the endoscopy. ATB: antibiotic-based eradication. PPI: proton-pump inhibitor
cCLR: Clarithromycin; CIPRO: Ciprofloxacin; MET: Metronidazole. Distribution of resistant (R) and susceptible (S) genotypes are indicated for heteroresistant populations
dr/m represent the ratio of SNPs introduced by recombination versus mutation
eTotal number of recombination events with the cumulative length (kbp) of recombined segments shown in brackets. Stars indicate that the majority of recombination events in this population are likely to have been imported during a period of mixed infection

**Fig. 1** Genetic diversity within *H. pylori* populations from 16 human individuals. a Global phylogeny including 440 *H. pylori* isolates from 16 human individuals. Isolates from the same patient are clustering together tightly indicating that no mixed infection was sampled. The scale bar represents the number of substitutions per site. Source data are provided as a Source Data file. b Every circle represents the *H. pylori* population from one individual patient. Each population consists of 20–30 clones (see text for details). Circles are ordered by decreasing genetic within-host diversity from outer to inner circle. From the outer to inner ring: patient 476, 479, 25, 19, 169, 24, 21, 13, 5, 381, 23, 26, 20, 478, 173, 280. Genetic coordinates were synchronized across all populations by mapping onto the reference strain 26695. Spontaneous mutation and recombination events are represented by red and blue lines, respectively. Regions present in 26695 but not fully covered in a population are indicated by black lines (the 23S region indicated here corresponds to the second copy, which is not always properly assembled in draft genomes, resulting in uncovered regions).
Variations of gene content and sequences are not the only ways *H. pylori* can evolve within its host. We identified large genomic rearrangements between the three closed genomes of patient 381 generated by SMRT sequencing (Supplementary Fig. 1a). Recombination between chemotaxis genes *tlpA* and *tlpC* induced the inversion of a small 9 kb fragment, whereas recombination between the 23S rRNA gene copies resulted in the inversion of a much larger 953 kb region. Finally, two inversions of 81 kb and 643 kb were owing to an exchange between copies of the miniIS605, a non-autonomous transposable element found in 643 kb were owing to an exchange between copies of the much larger 953 kb region. Finally, two inversions of 81 kb and inversion of a small 9 kb fragment, whereas recombination

HPnc058029. In the two reference genomes of patient 25, an inversion of ~ 590 kb owing to recombination between HP0488-HP1116 paralogs was also observed (Supplementary Fig. 1b). No rearrangements were observed in the other five patients where multiple closed genomes were available.

**Bacterial migration between gastric regions.** Most previous studies about *H. pylori* within-host diversity are restricted to the antrum and corpus regions. Here, fundus biopsies were also obtained for 10 of the 16 patients. In order to observe how the genetic diversity of each *H. pylori* isolate is distributed across the gastric regions, we computed recombination-corrected phylogenetic trees from whole-genome alignments. A visual inspection of these trees revealed that a structure with multiple clusters was clearly apparent in most patients. The degree to which these clusters correlated with the location from which the bacteria had been sampled was variable, and isolates with different origins were frequently found within the same cluster (Fig. 2).

This observation can be interpreted in terms of migration of *H. pylori* between the antrum, corpus, and fundus regions. We used ancestral state reconstruction methods to reconstruct the intrastomach migration routes of individual *H. pylori* isolates, as they last shared a common ancestor within a patient. By comparing three migration models of increasing complexity (see Methods) we found evidence for differences of the migration rates between parts of the stomach. There was no evidence for asymmetry in the rates between any two parts. The data fit well to the symmetric model, in which three parameters are estimated between antrum and fundus, between antrum and corpus, and finally between corpus and fundus. Migration events between corpus and fundus were significantly (based on 95% confidence intervals of the maximum likelihood estimates) more frequent than between antrum and corpus (Fig. 3a). The migration rate between antrum and fundus was negligible. Following these observations, we tested one additional meristic model where migration from antrum to fundus cannot happen directly but only in two steps via the corpus, and this model had a slightly better goodness of fit (AICC = 546.51) than the symmetric model (AICc = 548.55).

Finally, we tested an alternative approach, stochastic character mapping (see Methods), to simulate the migration along the population trees and obtain the number of transitions between each region. Following the same trend as the ancestral reconstruction method, this approach showed that the majority of the migration events happened between corpus and fundus, whereas migration between antrum and fundus represent the minority (Fig. 3b).

Overall, these results suggest that *H. pylori* gastric subpopulations are typically segregated in the antrum mucosa but are migrating more freely between regions lined by oxyntic mucosa, which includes both corpus and fundus.

We next focused on variation between patients, and to do so, repeated the ancestral state reconstruction analysis separately for each patient’s *H. pylori* population. In five patients (19, 20, 21, 23, 381), the meristic model was the best fit, with more frequent exchange between corpus and fundus than between antrum and corpus, similar to results obtained on the overall data set. The isolates of patient 26 also fit the meristic model but displayed more exchanges between antrum and corpus. However, this population was hardly differentiated according to the topology of its phylogenetic tree with no clearly identifiable clusters, suggesting that this result may be artefactual. In the four remaining patients (5, 13, 24, 25) the equal rates (ER) model had the best fit, suggesting that migration is less influenced by gastric regions for these strains. However, these results are likely explained by the low total number of migration events in most of

**Fig. 2** Clustering structure of the *H. pylori* population varies between individual patients. Representative phylogenetic trees of within-host *H. pylori* populations displaying different degrees of clustering. Trees were generated as described in the methods (source data are provided as a Source Data file). The time since the most recent common ancestor (TMRCA), expressed in years, is represented by the X axis. Pie charts correspond to the posterior probability distribution of ancestral characters (i.e., likelihood that ancestral lineages originated from antrum: blue, corpus: orange or fundus: green).

a: Phylogenetic tree of patient 381 isolates showing a well-defined cluster structure. b: Phylogenetic tree of patient 24 isolates showing a poorly defined cluster structure.
these patients, so that there is a lack of statistical power to determine the inequality of the rates of migrations between regions. Overall, frequencies of migration did not correlate with genetic diversity, indicating that the variation we observed between populations is not explained by the difference between their time of divergence and that this phenomenon is also influenced by host genetics and environment.

Despite the extensive mixing between fundus and corpus, the fundus isolates added a significant amount of diversity to our data set, therefore subsequent analyses in this study were only carried out on the subset of 10 patients including fundus isolates in order to reduce bias and to obtain a homogeneous sampling.

**Frequent within-host variation in specific functional groups.**
Within-host diversification is the net result of both random genetic drift and recurrent selection. The low levels of variability in some populations combined with the high levels of recombination in others make it challenging to quantify the contribution of each evolutionary mechanism reliably. We cross-compared genes that contained non-synonymous SNPs (NS-SNPs) between isolates from each patient, with the aim to determine whether specific loci were frequently and independently targeted by positive selection. We observed a high frequency of within-host variation in 40 genes (referred to later as host-variable genes) containing NS-SNPs in more than half of the patients, which was significantly higher (based on 95% confidence intervals) compared to the oxyntic mucosa (i.e., corpus and fundus), or vice-versa, using a Fisher’s exact test. Furthermore, we hypothesized that recently migrated isolates are likely to retain the genetic background representative of their original gastric region temporarily, which could introduce noise in the association test. To account for this, we used our ancestral state reconstruction analysis to infer recent migration events. When possible, we re-assigned the location of each isolate using the location of the corresponding ancestral node with a maximum divergence time of 6 months.

We identified 1071 polymorphisms that had significant associations with gastric regions (false discovery rate (FDR) adjusted p value < 0.01). After discarding synonymous and recombined polymorphisms, 309 non-synonymous mutations remained within a total of 232 genes (referred to later as niche-specific mutations and genotypes, respectively). Although we observed 32 niche-specific genotypes per patient on average, no significant results were obtained in patients 25 and 26. Isolates from these two patients appear shuffled together with barely any region-specific clusters, suggesting there might be limited adaptation and free migration in these patients. In addition, only one genotype with significant niche association was found in patient 24 despite clearly apparent region-specific clusters. No recent migration events were corrected by ancestral reconstruction because of the deep branches of the star-like phylogeny in this patient. In particular, one isolate sampled in the corpus was located on a long branch within a cluster containing 70% of the antrum isolates indicating a likely origin from the corpus. By increasing the maximum divergence time for re-assignment from 6 months to 1 year for patient 24, we identified nine additional genotypes with association to gastric regions (Supplementary Data 3). The gene that contained the most mutations associated with gastric regions encodes the fused beta and beta’ subunits of the DNA-dependent RNA polymerase RpoB. Niche-specific mutations were found in rpoB in five patients (5, 13, 19, 20,

**Niche-specific evolution and adaptation.** *H. pylori* is able to colonize the entire stomach despite the differing environmental conditions existing between gastric niches. The selective pressures caused by these differences might explain why specific genes evolve frequently within one host. The restricted migration of antrum isolates could also be a consequence of region-specific adaptation. Similarly, the frequent exchanges between corpus and fundus could reflect the similarities between these two regions. Consequently, we sought to determine whether polymorphisms were associated with the antrum, compared with the oxyntic mucosa (i.e., corpus and fundus), or vice-versa, using a Fisher’s exact test. Furthermore, we hypothesized that recently migrated isolates are likely to retain the genetic background representative of their original gastric region temporarily, which could introduce noise in the association test. To account for this, we used our ancestral state reconstruction analysis to infer recent migration events. When possible, we re-assigned the location of each isolate using the location of the corresponding ancestral node with a maximum divergence time of 6 months.
21), but none were located at the same position. They were specifically centered around the beta–beta’ interface region and differed from mutations described previously to be involved in rifabutin resistance or Western/East-Asian geographical differences.

Genes containing within-host NS-SNPs were cross-compared in the subset of 10 patients including fundus isolates. Genes were considered host-variable when containing NS-SNPs in more than half of the tested populations (counts are indicated in the first column). Locus tags from 26695 are given as reference.

Table 2 High-frequency host-variable genes within 10 H. pylori populations

| Counts | LocusTag | Gene   | Product                                      |
|--------|----------|--------|----------------------------------------------|
| 10     | HP0913   | hopB   | outer membrane protein (AlpA)                |
| 10     | HP1157   | hopL   | outer membrane protein                       |
| 9      | HP0099   | tlpA   | methyl-accepting chemotaxis protein          |
| 9      | HP0289   | lmaA   | immunomodulatory autotransporter, vacA-like protein |
| 9      | HP0953   | -      | hypothetical protein                         |
| 8      | HP0009   | hopZ   | outer membrane protein                       |
| 8      | HP0782   | hopE   | outer membrane protein                       |
| 8      | HP1156   | hopI   | outer membrane protein                       |
| 8      | HP0103   | tlpB   | methyl-accepting chemotaxis protein          |
| 8      | HP0599   | tlpD   | methyl-accepting chemotaxis protein          |
| 8      | HP0922   | vlpC   | toxin-like outer membrane protein            |
| 8      | HP0130   | -      | putative uncharacterized protein             |
| 8      | HP1198   | rpoB   | DNA-directed RNA polymerase subunit beta’     |
| 8      | HP0160   | hcpD   | cysteine-rich protein D                       |
| 8      | HP1117   | hcpX   | cysteine-rich protein X                      |
| 7      | HP0021   | lpxE   | lipid A 1-phosphatase                         |
| 7      | HP0252   | hopF   | outer membrane protein                       |
| 7      | HP0655   | bapA   | outer membrane protein assembly factor        |
| 7      | HP0912   | hopC   | outer membrane protein (AlpB)                |
| 7      | HP0887   | vacA   | vacuolating cytotoxin                        |
| 7      | HP1322   | -      | putative uncharacterized protein             |
| 7      | HP1664   | arsS   | sensor histidine kinase                      |
| 7      | HP0235   | hceE   | cysteine-rich protein E                       |
| 6      | HP0407   | bisC   | biotin sulfoxide reductase                   |
| 6      | HP0772   | amiA   | N-acetylmuramoyl-l-alanine amidase            |
| 6      | HP0805   | lexB   | LPS biosynthesis protein                     |
| 6      | HP0254   | hopG   | outer membrane protein (ompB)                |
| 6      | HP1453   | hopD   | outer membrane protein                       |
| 6      | HP0392   | cheAY2 | chemotaxis sensor kinase                     |
| 6      | HP0499   | pldA   | phospholipase A1                             |
| 6      | HP0219   | -      | hypothetical protein                          |
| 6      | HP0586   | -      | hypothetical protein                          |
| 6      | HP1039   | -      | O-antigen polymerase                         |
| 6      | HP1021   | -      | orphan response regulator                    |
| 6      | HP0818   | proWX  | osmoprotection protein                       |
| 6      | HP1251   | oppB   | ABC transporter, permease protein            |
| 6      | HP1252   | oppA   | ABC transporter, oligopeptide-binding protein |
| 6      | HP0471   | kefB   | potassium-efflux protein                     |
| 6      | HP0876   | tprB1  | iron-regulated outer membrane protein        |
| 6      | HP1400   | fecA3  | iron(III) dicitrate transport protein        |

Genes belonging to the OMPs, chemotaxis and motility, toxin production, and regulatory functions functional categories, were significantly over-represented (Chi-square, p < 0.05) (Fig. 4). In total, 66% of these genes showed a significant (Fisher’s exact test, FDR adjusted p value < 0.05) departure from neutrality, with a neutrality index (NI) > 1 (when it could be calculated). Here, NI > 1 indicates an excess of within-host non-synonymous polymorphisms. Considering that niche-specific mutations are fixed within local subpopulations, an excess of within-host non-synonymous substitutions can be interpreted as a sign of local adaptation. Unlike what we observed in high-frequency host-variable genes, niche-specific mutations were always unique to a
### Cell envelope

| LocusTag | Gene | MK   | N.I. | Product                          |
|----------|------|------|------|----------------------------------|
| HP0009   | hopZ | 1.98E–07 | 3.90 | Outer membrane protein           |
| HP0079   | horA | 1.00E+00 | 1.52 | Outer membrane protein           |
| HP0127   | horB | 2.87E–03 | 36.67| Outer membrane protein           |
| HP0227   | hopM | 1.66E–01 | 1.34 | Outer membrane protein           |
| HP0252   | hopF | 2.86E–06 | 39.67| Outer membrane protein           |
| HP0486   | hopC | 1.48E–03 | 8.53 | Outer membrane protein           |
| HP0506   | hdpA | 6.82E–02 | 9.86 | PG peptidase                     |
| HP0567   | –    | 1.35E–02 | N.A. | Autoinducer-2 Exporter (AI-2E)   |
| HP0638   | hopH | 4.81E–01 | 1.67 | Outer membrane protein           |
| HP0655   | bama | 1.06E–05 | 20.36| Outer membrane protein assembly factor |
| HP0671   | horF | 1.32E–02 | N.A. | Outer membrane protein           |
| HP0725   | sabA | 2.67E–03 | 2.25 | Outer membrane protein           |
| HP0788   | hopF | 6.38E–04 | 33.69| Outer membrane protein           |
| HP0912   | hopC | 1.39E–04 | 4.30 | Outer membrane protein (AlpA)    |
| HP0913   | hopB | 1.06E–05 | 4.91 | Outer membrane protein (AlpB)    |
| HP0914   | hopG | 1.65E–01 | 2.15 | Outer membrane protein           |
| HP1066   | hopD | 4.64E–02 | 14.57| Outer membrane protein           |
| HP1156   | hopI | 1.60E–09 | N.A. | Outer membrane protein           |
| HP1157   | hopL | 1.60E–09 | 8.56 | Outer membrane protein           |
| HP1167   | hopH | 1.35E–02 | 12.33| Outer membrane protein           |
| HP1453   | homD | 1.37E–03 | 7.02 | Outer membrane protein           |
| HP1469   | hopJ | 3.96E–01 | 4.33 | Outer membrane protein           |

### Cellular processes

| LocusTag | Gene | MK   | N.I. | Product                          |
|----------|------|------|------|----------------------------------|
| HP0019   | cheV1| 1.75E–03 | 31.50| Chemotaxis signal transduction protein |
| HP0103   | tlpB | 2.98E–09 | N.A. | Methyl-accepting chemotaxis protein |
| HP0232   | cheAY2| 4.44E–05 | 24.63| Chemotaxis sensor kinase         |
| HP0392   | tlpD | 2.86E–06 | 10.53| Methyl-accepting chemotaxis protein |
| HP0599   | motB | 6.19E–02 | 11.45| Flagellar motor rotation protein |
| HP0816   | fliY | 2.90E–01 | 4.44 | Flagellar motor switch protein   |
| HP1030   | pflA | 2.61E–03 | 19.89| Paralysed flagella protein       |
| HP1274   | imaA | 3.56E–01 | 1.17 | Immunomodulatory autotransporter, vacA-like protein |
| HP0289   | hefA | 1.92E–04 | 16.93| Outer-membrane protein of the hetABC efflux system |
| HP0605   | hefC | 1.04E–04 | 24.38| Cytoplasmic pump protein of the hetABC efflux system |
| HP0607   | vacA | 7.23E–01 | 1.24 | Vacuolating cytotoxin            |
| HP0887   | rdxA | 4.71E–02 | 2.83 | Oxygen-insensitive NAD(P)H nitroreductase |
| HP0954   | lptD | 2.88E–02 | 6.75 | LPS-assembly protein             |
| HP1216   | faaA | 2.61E–01 | 1.70 | Flagella-associated autotransporter A, vacA-like protein |
| HP0609   |   |      |      |                                  |

### Regulation

| LocusTag | Gene | MK   | N.I. | Product                          |
|----------|------|------|------|----------------------------------|
| HP0166   | ompR | 4.43E–03 | 36.00| Response regulator (ArsR)        |
| HP0703   | figR | 3.19E–01 | 6.23 | Response regulator               |
| HP0775   | spoT | 2.76E–05 | 15.27| ppGpp synthetase/hydrolase       |
| HP1021   | –    | 3.27E–06 | 24.73| Orphan response regulator        |
| HP1027   | tur  | 1.32E–02 | N.A. | Ferric uptake regulation protein |
| HP1032   | fliA | 8.37E–02 | 9.19 | Alternative transcription initiation factor, sigma-F |
| HP1043   | hsrA | 1.66E–01 | N.A. | Orphan response regulator        |
| HP1140   | birA | 1.00E+00 | 1.29 | Biotin ligase                    |
| HP1338   | nikR | 2.86E–06 | 56.67| Nickel responsive regulator      |
| HP1442   | csaA | 3.69E–02 | 12.00| Carbon storage regulator         |

**Fig. 4** Evidence of natural selection in niche-specific genotypes from enriched functional categories. For each *H. pylori* population, black and gray squares represent polymorphic genes with and without association to gastric regions, respectively; white squares indicate genes without non-synonymous mutations. Locus tags from *H. pylori* strain 26695 are given as reference. MK indicates the Fisher’s exact FDR adjusted $p$ value from a McDonald-Kreitman test (bolded values indicate significance at $p < 0.05$). N.I.: Neutrality Index.
patient, except in the \textit{hmf}C OMP. HofC displayed a significant
departure from neutrality and an isoleucine to valine mutation in
position 206 was observed in four patients (13, 19, 20, 21). In two
of these (13 and 20), this mutation was found associated to gastric
regions with the valine residue found predominantly in the
antrum isolates of both patients.

Several genes related to the regulation of motility through
chemotaxis contained niche-specific mutations with evidence of
natural selection. The \textit{cheA}Y2 gene (HP0392) codes for a
bifunctional protein containing both parts of two-component
regulatory system: \textit{CheA} (histidine kinase sensor) and \textit{CheY}
(response regulator)\textsuperscript{36}. All niche-specific residues were located in the
CheA region of the protein. Three out of the four methyl-
accepting chemotaxis proteins: \textit{TlpA} (HP0099), \textit{TlpB} (HP0103),
\textit{TlpD} (HP0399) were host-variable in eight to nine patients, but
only \textit{TlpB} and \textit{TlpD} had niche-specific mutations in two and one
patients, respectively. Around 70% of the mutations were clustered close to the doubleCache signaling domain in \textit{TlpA}
but none were associated with gastric regions. Mutations were
more homogeneously distributed along the sequence in \textit{TlpB} and
\textit{TlpD}, but the niche-specific mutations were located in or close to
the methyl-accepting domain.

Mosaic sequences owing to inter-strain recombination. \textit{H. pylori}
can quickly exchange substantial parts of its genome when
exposed to DNA from another \textit{H. pylori} strain owing to its nat-
ural competence\textsuperscript{11,16}. Nonetheless, it colonizes a very isolated
niche and no natural reservoir outside of the human stomach is
known, thus the only opportunity for inter-population recombina-
donation is during mixed infections. Such sudden increase of genetic
diversity might reshape the evolutionary dynamics of a popula-
tion; thus we applied ClonalFrameML\textsuperscript{37} to analyze the effect of
recombination in each patient. Recombined fragments were
observed in the \textit{H. pylori} populations from all patients, but large
differences were observed in the relative contributions of recombin-
ation and mutation to the overall diversity, as well as the
number of events and average and total recombined length. For
11 out of 16 patients, manual curation revealed that a majority of
the recombination events were located in duplicated genes or
paralogous families. Sequences identical to the recombined seg-
mants could be frequently found in the other copies of the gene
within the same isolate, suggesting these are the products of
intragenomic recombination. The most frequently involved genes
were \textit{babABC} and \textit{sabAB}, the \textit{cag} pathogenicity island gene \textit{cagY}
and the OMP gene \textit{hopQ}.

Only related isolates were found at the time of the endoscopies,
suggesting transient mixed infections can quickly increase the
generic diversity of the main population. The \textit{r/m} ratio of the five
patients (19, 25, 169, 476, 479) with clear evidence of past mixed
infections ranged from 3 to 16. Imports were \textit{465 ± 185} bp long
on average and replaced 2–26% of the genome at the population
level. Overall, an average 80% of the predicted imports introduced
at least one NS-SNP and thus had the potential to modify protein
functions. Further manual curation revealed that many imports
overlapped the same region in multiple isolates from the same
patient but had different sizes. We observed three patterns of
overlapping imports (Supplementary Fig. 3). The first pattern
suggests that some imports may not originate directly from a
mixed infection (primary import) but rather from previously
transformed isolates within the same population (secondary
import). This model is further supported by the phylogenetic
distribution of these imports, because primary and secondary
imports were found in polyphyletic clusters indicating they likely
originated from multiple events. Although the dating of internal
nodes is not fully accurate owing to the limited number of
isolates, nodes containing primary imports typically preceded
nodes containing the secondary imports. The second pattern we
observed suggest that primary imports can be partially reversed
by recombination with a wild-type isolate from the same
population (reversed import). Reversed imports could be easily
identified when the primary imports are still intact in other
isolates with a direct common ancestor. The third pattern was
observed less frequently and corresponds to multiple independent
primary imports targeting the same region. These imports are
overlapping but do not share any border. They are located in
polyphyletic clusters and have slightly different polymorphisms,
suggesting they originated from distinct lineages of the mixed
infection.

Effect of antibiotics on \textit{H. pylori} population diversity. Although
\textit{H. pylori} infection usually begins during childhood, previous
studies performed with paired isolates from antrum and corpus
showed that the within-host TMRCA of \textit{H. pylori} is typically low
(\textit{1–5} years). To determine the TMRCA values in our data set, we
used ClonalFrameML\textsuperscript{37} to generate phylogenetic trees adjusted
for recombination. Trees were then scaled to time using a fixed
mutation rate of \textit{1.38 × 10\textsuperscript{−5} per site per year}\textsuperscript{19}. The average
TMRCA of the \textit{H. pylori} populations from the 16 patients was 2.5
years (0.34–7.17). No correlation with patient age was observed.
For example, a TMRCA of 4 months was calculated for a 44-year
old woman (patient 173), whereas a TMRCA of 7 years was
estimated for a 34-year old man (patient 476). Several hypotheses
have been proposed to explain low TMRCA, including strong
genetic drift or diversity bottlenecks induced by immunity or
antibiotics. As five patients reported at least one attempted era-
dication therapy prior to the gastroscopy, we explored the role of
antibiotics in shaping the population structure of \textit{H. pylori}
in a stomach.

Clarithromycin (CLR), ciprofloxacin (CIP), and metronidazole
(MTZ) are antibiotics used for the eradication of \textit{H. pylori} as well as
for treatment of other pathogenic bacteria. The emergence of
resistant phenotypes has been associated with specific genotypes:
two specific positions of the 23S rRNA gene are responsible for
high or intermediate levels of resistance to CLR; mutations in the
gene encoding the DNA gyrase subunit A (\textit{gvrA}) can confer
resistance to CIP, whereas resistance to MTZ have been associated with either mutations or gene inactivation of the
nitroreductase gene \textit{rxa}A\textsuperscript{38}. We searched for resistance-associated
polymorphisms in the complete data set to characterize the
spread of resistance within the populations in response to either
\textit{H. pylori} eradication therapy or unrelated antibiotic intake
targeted toward other infections. The resistant genotypes we
observed were systematically confirmed phenotypically in a
subset of isolates. There was no phenotypical evidence for
resistance against either beta-lactam antibiotics or rifabutin.
This analysis revealed that the \textit{H. pylori} populations from the five
patients who had previously been prescribed eradication therapy
(patient 20, 173, 280, 381, 478), as well as two additional patients
who were not known to have undergone eradication treatment
(23, 169), contained isolates that were resistant to at least one of
the tested antibiotics (Table 1). The mean TMRCA of resistant
populations was \textit{1.6 ± 1} years, whereas the mean TMRCA for
fully susceptible populations was \textit{3.9 ± 1.7} years (\textit{p} < 0.05, Welch’s
\textit{t} test).

Resistance-associated alleles were often not fixed, with 20–60% of
the population still carrying susceptibility-associated alleles
depending on the antibiotic class. Allele fixation in some cases
was limited to specific gastric regions. For example, isolates from
the corpus, but not the antrum, were resistant to CIP in patient
173. Antrum isolates represented the only MTZ susceptible
cluster in patient 23 and, on the opposite, the only resistant one in patient 20. Although the whole population in patient 280 was resistant to CIP, the isolates had a different genotype in the antrum (D91G) compared to the corpus (D91N). To further investigate the low TMRCAs of \textit{H. pylori} populations, we obtained a new set of antrum, corpus, and fundus biopsies from patient 280, 2 years after the initial endoscopy. As before, 10 single colonies were cultured from each biopsy and sequenced on an Illumina MiSeq instrument. This patient did not decide to undergo any kind of \textit{H. pylori} eradication regimen, leading us to believe that the \textit{H. pylori} population evolved unhindered between the examinations. Nonetheless, the population recovered from the second set of biopsies appeared to have lost >98% of its genetic diversity (apparent bottleneck) and gone through a fast clonal expansion subsequent to a bottleneck. Source data are provided as a Source Data file.

Discussion

Although the existence of multiple \textit{H. pylori} strains with distinguishable but related genotypes within one stomach has been reported before, the within-host diversity of \textit{H. pylori} has not been determined in a systematic and quantitative way. Here, we used a combination of sequencing technologies to characterize the host-associated population diversity of \textit{H. pylori} in 16 naturally infected individuals. The extent of genetic diversity observed in each \textit{H. pylori} population from one stomach was highly variable. No ongoing mixed infection was detected, but strong evidence of recombination with one or multiple unrelated \textit{H. pylori} strains during a past mixed infection was found in five patients. Through natural transformation, the co-existence with another population appears to have quickly increased the overall genetic diversity in these populations. Nonetheless, the isolates were not all impacted in the same way by inter-population recombination, which is in agreement with a previous study based on paired isolates from the gastric antrum and corpus regions, which showed that even in South Africa, a high prevalence country, the contribution of recombination to diversity varied widely between different individuals\textsuperscript{19}. Individual lineages affected by recombination with a secondary infection were clearly identified and imported segments were sometimes laterally transferred to other lineages or, conversely, reverted back to wild-type by another recombination event with isolates from the original population. This phenomenon suggests that the large amount of diversity imported from a secondary infection is subsequently filtered by selection. In agreement with our previous observations\textsuperscript{19}, the TMRCAs of all populations in our data set were low and did not correlate with patients’ age despite the fact that \textit{H. pylori} infections typically are acquired during childhood\textsuperscript{39}. Five patients had undergone one or multiple \textit{H. pylori}-specific eradication therapies in the years preceding the gastroscopy. Independently of the outcome, antibiotic regimens have the potential to result in an apparent reduction of the genetic diversity of the population in the stomach. Successful eradication can be followed by re-infection with a new population since \textit{H. pylori} does not induce protective immunity\textsuperscript{40–41}. In our case, the \textit{H. pylori} isolates in all five treated patients were resistant to one or several antibiotics commonly used for eradication therapies, suggesting that the treatment failed and only induced a population bottleneck, followed by recrudescence of the infection. Such temporary reductions of the \textit{H. pylori} population have been termed “elimination” (in contrast to eradication) and are a likely cause of the low observed TMRCAs. However, antibiotic resistance was also observed in patients with no known history of eradication treatment, suggesting that antibiotic treatments for unrelated infections was selecting antibiotic-resistant \textit{H. pylori}\textsuperscript{7}. It is known from early treatment studies that monotherapies have the potential to severely reduce \textit{H. pylori} colonization\textsuperscript{44}. Within-host populations were not always homogeneously resistant to antibiotics, confirming previous reports of heteroresistance in \textit{H. pylori}\textsuperscript{45}. Although the complete fixation of resistance genotypes can be explained by intrinsic resistances or hard selective sweeps, partially fixed alleles or independent adaptation are evidence for acquired resistance and soft selective sweeps, suggesting a relatively high fitness cost of resistance. The sweeping of alleles through the population appeared to be modulated by the stomach environment, and some genotypes were restricted to specific gastric regions. In these patients, heteroresistance suggests that antibiotics exerted a weak selective pressure causing a slow spread of adaptive alleles within the stomach. Such an effect may be explained, for example, by the relative instability of clarithromycin, and macrolide antibiotics in general, at the low pH found in the stomach, compared with other antibiotics\textsuperscript{46}. Nonetheless, patients with no history of eradication and no signs of resistance still displayed low TMRCAs, implying that other events can decrease within-host diversity. These could, for example, include antibiotic treatments for any other disease, changes in gastric physiology brought about by medication, stress, or concurrent infections, and selective pressure exerted by immune responses and oxidative stress. Antibiotic-resistant populations had significantly lower TMRCAs than susceptible
populations, indicating that antibiotics have a stronger impact on population diversity than possible other uncharacterized selective events. A large reduction of genetic diversity was observed in the sequential population of patient 476, cultured from biopsies taken two years after the initial endoscopy. The patient was not treated specifically for H. pylori between the two endoscopies, and did not receive any treatment with antibiotics, yet the whole H. pylori population had become resistant to macrolides, indicating that an unrelated antibiotic regimen could have induced a population bottleneck. Through the degeneration of mutation repair systems and natural transformation, evolution has favored the accumulation of diversity in H. pylori, which is likely contributing to rapid adaptation to changing environmental conditions. In our case, the large amount of pre-existing genetic variation in this patient did not seem to have contributed to the rapid emergence of antibiotic resistance since the complete fixation of the adaptive 23S rRNA allele suggests that macrolides produced a hard selective sweep and a major reduction of diversity. Consequently, exposure to antibiotics, whether targeted toward eradication of H. pylori or not, represents an important event in the within-host evolution of these bacteria. Antibiotics are likely to have massively influenced the evolutionary dynamics of the species as a whole in recent decades as antibiotics became widely available globally and are almost ubiquitously used in almost all human populations.

The phylogenetic structure of within-host populations revealed that H. pylori isolates migrate between gastric regions. Migration happened frequently between the corpus and the fundus regions, thus ancestral lineages originating from the antrum were often easily identified, whereas corpus and fundus lineages were harder to distinguish from each other. H. pylori is a highly motile bacterium, and motility is essential for colonization;57,48 thus its ability to move between gastric regions is not surprising. However, in contrast to its vertical motility and orientation between layers of the gastric mucus, horizontal long-distance migrations of H. pylori between gastric regions have not been assessed before. The data acquired in the present study hence are a leap forward in our understanding of the development of the H. pylori population in the whole stomach. H. pylori intra-stomach localization is likely to be influenced by adhesins and the availability of cell surface receptors in different regions of the stomach, and by chemotaxis, in particular, the four Tlp chemotaxis receptor proteins, which sense and react to environmental conditions such as acidity, energy levels, lactate, urea, and other stimuli;49–53. Although corpus and fundus are anatomically distinct areas, both regions are lined by oxyntic mucosa and may thus represent similar ecological niches for H. pylori. Notably, parietal cells, which are responsible for acid production, are found in all gastric glands of the oxyntic mucosa but are only partially present in the antral mucosa. The pH and bicarbonate gradients between the lumen and the mucus are required for spatial orientation of the bacterium and motility toward the epithelium.55,56 A recent study demonstrated that the strength of the adherence mediated by the BabA OMP is pH-dependent, which could be a way for H. pylori to detect when cells have detached from the epithelial layer to be shed toward the more acidic lumen.56 Gastric units also serve to detect when cells have detached from the epithelial layer by the BabA OMP is pH-dependent, which could be a way for H. pylori to detect when cells have detached from the epithelial layer. The BabA OMP is pH-dependent, which could be a way for H. pylori to detect when cells have detached from the epithelial layer. The BabA OMP is pH-dependent, which could be a way for H. pylori to detect when cells have detached from the epithelial layer. The BabA OMP is pH-dependent, which could be a way for H. pylori to detect when cells have detached from the epithelial layer. The BabA OMP is pH-dependent, which could be a way for H. pylori to detect when cells have detached from the epithelial layer.

The isolation between the antral and oxyntic mucosa sub-populations appeared to drive independent adaptation within each niche. The independent emergence of antibiotic resistance in distinct regions observed in some patients is a clear example of such isolation and suggests that even under a global and common selective pressure, such as antibiotics, the lower frequency of migration between antral and oxyntic mucosa is hindering the selective sweep of adaptive alleles. We detected association with the antral and oxyntic mucosa in 241 genes, mainly related to chemotaxis, motility, and OMP families. With the exception of BabA and HomB, most confirmed or suspected adhesins (HopBC, HopH, HopZ, and Saba) displayed niche-specific mutations in one or more patients.59,60 This suggests that modulating adherence to the gastric epithelium could be beneficial to improve the long-term colonization of different ecological niches. The observation that genes involved in chemotaxis and motility displayed patterns of niche-specific evolution is consistent with the important role of proper orientation and directed swimming of the bacteria for adaptation to the different types of mucosa and gastric glands. In particular, TlpB and TlpD are membrane-bound and cytoplasmic chemoreceptors, respectively, that mediate attraction to urea, repulsion from acidic pH and reactive oxygen species, and its energy taxis;49,50,52,53,61,62

Only a minority of genes showed a signal of niche-specific adaptation in multiple patients, indicating there might be different ways to adapt to the same niche or alternatively that environmental variations between patients could be limiting the chance of convergent evolution. Moreover, different mutations were observed in different patients, the only niche-specific mutation existing identically in multiple patients was observed in the HofC protein, which is part of the second largest group of paralogous OMPs.60 The role of this family is not well characterized; it is known that HofC is regulated by a fur box and is able to bind lactoferrin.64 Interestingly, HofC is required for colonization in mice and was the only gene acquiring mutations during mouse adaptation in both the SS1 and SS2000 strains56. Lactoferrin is found in the gastric mucus of both antrum and corpus.67,68 The gastric levels of lactoferrin are correlated with IL-8, ammonia, and inflammation levels, thus are supposed to be increased during gastritis.69 Iron requires gastric acid for solubilization and free iron becomes less available toward the glands when bound by lactoferrin70, thus potentially different environments and oxyntic mucosa might drive the adaptation of HofC toward a more efficient binding of lactoferrin.

Here we showed that the within-host population structure of H. pylori is influenced by the physical separation of the antral and oxyntic mucosa. Consequently, it cannot be entirely dismissed that niche-specific genotypes emerged from genetic drift. Yet, functional categories relevant to host adaptation were over-represented amongst these genes and niche-specific mutations were largely non-neutral suggesting they are involved in local adaptation. Niche-specific genotypes represented only a minority of the overall diversity observed in each patient. We identified 40 high-frequency host-variable genes, which were polymorphic in >50% of the patients. Despite belonging to similar functional categories as the niche-specific genotypes, such as the OMPs and cellular processes (regulation), these host-variable genes were not so frequently displaying polymorphisms associated to gastric regions. Unlike the niche-specific genotypes, identical mutations in host-variable genes were observed in different patients. The high in vivo mutation rate of these genes may be required for antigenic variation or adaptation to uncharacterized ecological niches. For example, H. pylori colonizes preferentially gastric lesions through chemotaxis1. Damaged gastric tissues possess different characteristics than healthy ones, and the wound-healing process is slowed down by H. pylori.72,73 The diversity of
these host-variable genes means they are frequently contributing to the standing genetic variation of H. pylori and are targeted during selection events. For example, the plasticity of OMPs has been described many times in longitudinal isolates as well as in geographically distinct strains. The high within-host variability of some OMPs, such as HopB and HopL, suggests that variants could be selected from the donor inoculum diversity during person-to-person transmission or emerge from in situ adaptation.

Methods

Gastric biopsies and isolation of H. pylori. Gastric biopsies were obtained from 16 patients in the Department of Gastroenterology, Hepatology and Infectious Diseases at the Otto-von-Guericke University, Magdeburg, Germany (protocol number 80/11) as well as written informed consent from all subjects. The histological assessment was performed according to the OLGa/OLGIgM staging system. H. pylori single colonies were isolated on blood agar plates by spreading the samples directly onto the media without pre-processing.

DNA extraction, genome sequencing, assembly, and annotation. Genomic DNA was extracted using Genomic-tip 100 G columns (Qiagen, Hilden, Germany). All isolates were sequenced with >30-fold coverage on an Illumina MiSeq instrument using the MiSeq SBS Kit. 50 samples were sequenced in 5 x 2 x 300 cycles v3 reagent cartridges. Paired-end reads were assembled de novo using SPAdes 3.9.0 with default parameters. In addition, selected isolates were resequenced on a PacBio RSII or Sequel instrument in order to generate completely closed reference sequences. Genomes were annotated using Prokka 1.7.2 with the H. pylori species database and ncRNAs were identified using Infernal 11.08. Annotations of OMPs were manually curated using the 26695 and J99 reference strains as well as multiple sequence alignment and phylogenetic comparisons. For each patient, reference genomes and raw read data were submitted to NCBI under BioProject PRJNA490474. GFF Annotations files were deposited on FigShare http://dx.doi.org/10.6084/m9.figshare.7188239.

Antibiotic resistance testing. Antibiotic resistance for selected isolates was tested for MTZ, CLR, and CIP using Etest strips (Liofilchem, Roseto degli Abruzzi, Italy). EUCAST clinical breakpoints values were used to determine susceptibility (The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for EUCAST clinical breakpoints values were used to determine susceptibility (The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for EUCAST clinical breakpoints values were used to determine susceptibility (http://www.eucast.org). A meristic model, in which migration between antrum and corpus was counted as migration, was used for each host as a reference. Overlapping contours were stitched together and regions with incomplete coverage were replaced by Ns. The resulting multiple alignments were used as input for the analysis described below. The initial phylogenetic trees were generated using RAxML 8.2.84 with the GTR GAMMA model, 100 bootstrap replicates and the rooting algorithm option. Recombination events were predicted using ClonalFrameML 1.037 with default parameters, and the phylogenetic trees were corrected accordingly. Based on these recombination-corrected phylogenies, the TMRCA was calculated using a synonymous mutation rate of 1.38 x 10^{-9} per site per year as determined previously 31. For Fig 1a, the global tree was created with all the isolates using only SNPs obtained from alignment strain 26695 and FastTree 2.1.317. Figure 1b was generated with the R package Rcircos 82. Migration rates between gastric regions were computed by marginal state reconstruction using the ace function from the ape R package 83, which implements maximum likelihood ancestral state reconstruction 84. Ultrametric time-scaled trees were generated as described 85; they had to be slightly adjusted to meet the ace function requirements: multichotomies were resolved as dichotomies with the multi2di function and branches with a length of 0 were change to an insignificant length of 10^{-10} times the total tree length. Models with equal rates (one parameter), symmetric rates (three parameters), and all rates different (six parameters) were tested consecutively by order of complexity. The more complex models were only selected if they provided a significantly higher likelihood (likelihood-ratio test, p < 0.05). A meristic model, in which migration between antrum and fundus is required to happen in two steps via the corpus, was also assessed using the fitDiscrete function from the geiger 2.0.6a R package 86. In order to estimate migration rates based on the within-host trees of all individuals simultaneously, a global tree was constructed by concatenating individual time-scaled trees using long (30 years) separating branches. The number of migration events was estimated by stochastic character mapping using the make.simmap function from the phytools 0.6 R package 87, which is based on a Markov chain Monte Carlo approach 88. Character evolution was simulated for 1000 iterations using a symmetric model and 1000 simulations for each ultrametric time-scaled tree. The distribution of each type of migration event was then averaged over all trees. Genotypes associated with geographic regions were identified using a Fisher’s exact test with p values corrected for multiple testing using the FDR approach 89. To account for the high migration rate between corpus and fundus, isolates from these stomach parts were grouped together. We also accounted for recent migration events in the association analysis as follows. Using the ancestral reconstruction results, the location of each isolate was reassigned to the location of the corresponding ancestral node with a maximum divergence time of 6 months for all patients (except patient 24 for which the maximum divergence was 1 year). Genotypes with an adjusted p value < 0.01 were considered significantly associated. Enriched functional categories for genes displaying niche-associated genotypes were detected using a Chi-Square test (p < 0.05) and presented in Fig. 4. For these genes, departure from neutrality was tested subsequently using a McDonald-Kreitman test 35. In brief, the four counts (Pn, Ps, Dn, Ds) of between-host (D) vs within-host (P), and non-synonymous (n) vs synonymous (s) substitutions were computed and tested for neutrality using a Fisher’s exact test (FDR corrected p values < 0.05). The NI was calculated as follows NI = (Pn/Ps)/(Dn/Ds).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the article and Supplementary Information files. Moreover, reference genomes and raw sequencing reads that support the findings of this study have been deposited in the NCBI database with the accession code PRJNA490474. GFF Annotations files are available in the FigShare repository with the identifier https://doi.org/10.6084/m9.figshare.7188239. The source data underlying Figs. 1a, 2a-b, and 3a-b, and Supplementary Fig. 2a are provided as a Source data file.

Received: 20 November 2018 Accepted: 10 April 2019Published online: 22 May 2019

References

1. Suerbaum, S. & Michetti, P. Helicobacter pylori infection. N. Engl. J. Med. 347, 1175–1186 (2002).
2. Figueiredo, C. et al. Helicobacter pylori and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. J. Natl. Cancer Inst. 94, 1680–1687 (2002).
3. Hatakayama, M. Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. Cell Host Microbe 15, 306–316 (2014).
4. Foegeding, N. J., Caston, R. R., McClain, M. S., Ohi, M. D., & Cover, T. L. An overview of Helicobacter pylori VacA toxin biology. Toxins (Basel) 8, pii: e173 (2016).
5. Malfettheiner, P. et al. Management of Helicobacter pylori infection—the Maastricht V/Florence consensus report. Gut 66, 6–30 (2017).
6. O’Connor, A., Lamarque, D., Guibert, J. P., & O’Morain, C. Treatment of Helicobacter pylori infection 2017. Helicobacter 22, Suppl 1 (2017).
7. Megraud, F. et al. Helicobacter pylori resistance to antibiotics in Europe and its relationship to antibiotic consumption. Gut 62, 34–43 (2013).
8. Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect. Dis. 18, 318–327 (2018).
9. Suerbaum, S. & Josenshans, C. Helicobacter pylori evolution and phenotypic diversification in a changing host. Nat. Rev. Microbiol. 5, 441–452 (2007).
10. Didekot, X., Walker, A. S., Petro, T. E., Crook, D. W., & Wilson, D. J. Within-host evolution of bacterial pathogens. Nat. Rev. Microbiol. 14, 150–162 (2016).
11. 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. USA 98, 15056–15061 (2001).
12. Doner, M. S., Sessler, T. H. & Salama, N. R. Recombination and DNA repair in Helicobacter pylori. Annu. Rev. Microbiol. 65, 329–348 (2011).
13. Falush, D. et al. Traces of human migrations in Helicobacter pylori populations. Science 299, 1582–1585 (2003).
14. Linn, B. et al. An African origin for the intimate association between humans and Helicobacter pylori. Nature 415, 54–57 (2002).
15. Morelli, G. et al. Microevolution of Helicobacter pylori during prolonged infection of single hosts and within families. PLoS Genet. 6, e1001036 (2010).
16. Kellenmann, L. et al. Helicobacter pylori genome evolution during human infection. Proc. Natl. Acad. Sci. USA 108, 5033–5038 (2011).
17. Linn, B. et al. A mutation burst during the acute phase of Helicobacter pylori infection in humans and rhesus macaques. Nat. Commun. 5, 4165 (2014).
18. Nell, S. et al. Genome and methyleome variation in Helicobacter pylori with a cag pathogenicity island during early stages of human infection. Gastroenterology 154, 612–623 (2018).
21. Krebs, J., Didelot, X., Kennemann, L. & Suerbaum, S. Bidirectional genomic evolution and transmission of Helicobacter pylori in two South African families. Proc. Natl. Acad. Sci. USA 110, 13880–13885 (2013).

22. Furuta, Y. et al. Microevolution of virulence-related genes in Helicobacter pylori during naturally occurring transmission between adults. PLoS ONE 8, e82187 (2013).

23. Krebs, J., Didelot, X., Kennemann, L. & Suerbaum, S. Bidirectional genomic exchange between Helicobacter pylori strains from a family in Coventry, United Kingdom. Int. J. Med. Microbiol. 304, 1135–1146 (2014).

24. Furuta, Y. et al. Evolution of virulence-related genes in Helicobacter pylori familial infection. PLoS ONE 10, e0121977 (2015).

25. Kuipers, E. J. et al. Quasiespecies development of Helicobacter pylori observed in paired isolates obtained years apart from the same host. J. Infect. Dis. 181, 273–282 (2000).

26. Israel, D. A. et al. Helicobacter pylori genetic diversity within the gastric niche of the human stomach. Proc. Natl. Acad. Sci. USA 98, 14625–14630 (2001).

27. Salama, N. R. et al. Genetic analysis of Helicobacter pylori strain populations colonizing the stomach at different times postinfection. J. Bacteriol. 189, 3834–3845 (2007).

28. Bugaytsova, J. A. et al. Helicobacter pylori adapts to chronic infection and gastric disease via pH-responsive BabA-mediated adherence. Cell Host Microbe 21, 376–389 (2017).

29. Selgrad, M. et al. Different antibiotic susceptibility between antrum and corpus of the stomach, a possible reason for treatment failure of Helicobacter pylori infection. World J. Gastroenterol. 20, 16245–16251 (2014).

30. Aphe至於-Hurtig, M. et al. Functional adaptation of BabA, the H. pylori ABO blood group binding adhesin. Science 305, 519–522 (2004).

31. Censini, S. et al. eag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. Proc. Natl. Acad. Sci. USA 93, 14648–14653 (1996).

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

33. Heep, M., Beck, D., Bayerdorffer, E. & Lehn, N. Rifampin and rifabutin resistance mechanism in Helicobacter pylori. Antimicrob. Agents Chemother. 43, 1497–1499 (1999).

34. Lee, K. H. et al. Alanine-threonine polymorphism of Helicobacter pylori RpoB is correlated with differential induction of interleukin-8 in MKN45 cells. J. Clin. Microbiol. 42, 3218–3224 (2004).

35. Ogura, M. et al. Helicobacter pylori evolution: lineage-specific adaptations in homologs of eukaryotic Sell-like genes. PLoS Comput. Biol. 3, e151 (2007).

36. Meinke, A. et al. Composition of the ANTIGENome of Helicobacter pylori defined by human serum antibodies. Vaccine 27, 3251–3259 (2009).

37. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the rpoB locus in Helicobacter pylori. Proc. Natl. Acad. Sci. USA 98, 404–409 (2001).

38. Hauschild, M. et al. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput. Biol. 11, e1004041 (2015).

39. Hu, Y., Zhang, M., Lu, B. & Dai, J. Helicobacter pylori and antibiotic resistance, a continuing and intractable problem. Helicobacter 21, 349–363 (2016).

40. Suerbaum, S. et al. Free recombination within Helicobacter pylori. Proc. Natl. Acad. Sci. USA 95, 12619–12624 (1998).

41. Dubois, A. et al. Host specificity of Helicobacter pylori strains and host responses in experimentally challenged nonhuman primates. Gastroenterology 116, 90–96 (1999).

42. Zeneddihel, N. et al. Helicobacter pylori reinsfection rate 3 years after successful eradication. J. Gastroenterol. Hepatol. 20, 401–404 (2005).

43. Silva, F. M. et al. Helicobacter pylori reinsfection in Brazilian patients with peptic ulcer disease: a 5-year follow-up. Helicobacter 15, 46–52 (2010).

44. Take, S. et al. Reinfection rate of Helicobacter pylori after eradication treatment: a long-term prospective study in Japan. J. Gastroenterol. 47, 641–646 (2012).

45. Gisbert, J. P., Pajares, R. & Pajares, J. M. Evolution of Helicobacter pylori therapy from a meta-analytical perspective. Helicobacter 12, 50–58 (2007).

46. Kato, C. Y. et al. Persistence of Helicobacter pylori from the same patient prior to antibiotic treatment. Infect. Genet. Evol. 23, 196–202 (2014).

47. Erah, P. O., Goddard, A. F., Barrett, D. A., Shaw, P. N. & Spiller, R. C. The stability of amoxicillin, clarithromycin and metronidazole in gastric juice: relevance to the treatment of Helicobacter pylori infection. J. Antimicrob. Chemother. 39, 5–12 (1997).

48. Eaton, K. A., Suerbaum, S., Josenshan, C. & Krakowka, S. Colonization of gnotobiotic piglets by Helicobacter pylori deficient in two flagellin genes. Infect. Immun. 64, 2445–2448 (1996).

49. Josenshan, C. & Suerbaum, S. The role of motility as a virulence factor in bacteria. Int. J. Med. Microbiol. 291, 605–614 (2002).

50. Cysyk, M. A., Wheeler transform. Bioinformatics 27, 589–595 (2010).

51. Stamatkis, A. Using RaxML to infer phylogenies. Curr. Protoc. Bioinformatics 51, 14–11–14 (2015).
Acknowledgements

We thank Christine Josenhans for valuable discussions while the project was progressing as well as critical reading and commenting of the manuscript. We thank Cathrin Spröer and Boyke Bunk for performing SMRT sequencing and genome assemblies as well as Birgit Brenneke, Friederike Kops, Simone Severitt, and Nicole Heyer for their excellent technical assistance. Funding was provided from the Deutsche Forschungsgemeinschaft (grants SFB900/A1 and SFB900/Z1 to S.S.).

Author contributions

F.A. and S.S. designed the experiments. F.A., S.W., G.P., and R.C.B. performed the experiments. C.S. and P.M. carried out the clinical part of the study and provided gastric biopsies. J.O. provided PacBio sequencing. F.A., X.D., and S.S. analyzed the data. F.A., X.D., and S.S. wrote the paper. All authors provided critical comments on the manuscript and approved the final version.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10050-1.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019