

Genomic differentiation within East Asian *Helicobacter pylori*

Yuanhai You^{1†}, Kaisa Thorell^{2,3†}, Lihua He¹, Koji Yahara⁴, Yoshio Yamaoka⁵, Jeong-Heon Cha⁶, Kazunari Murakami⁷, Yukako Katsura⁸, TEAMHp⁹, Ichizo Kobayashi^{10,11,12,13,14,*}, Daniel Falush^{15,*} and Jianzhong Zhang^{1,*}

Abstract

The East Asian region, including China, Japan and Korea, accounts for half of gastric cancer deaths. However, different areas have contrasting gastric cancer incidences and the population structure of *Helicobacter pylori* in this ethnically diverse region is yet unknown. We aimed to investigate genomic differences in *H. pylori* between these areas to identify sequence polymorphisms associated with increased cancer risk. We analysed 381 *H. pylori* genomes collected from different areas of the three countries using phylogenetic and population genetic tools to characterize population differentiation. The functional consequences of SNPs with a highest fixation index (*F*_{st}) between subpopulations were examined by mapping amino acid changes on 3D protein structure, solved or modelled. Overall, 329/381 genomes belonged to the previously identified hspEAsia population indicating that import of bacteria from other regions of the world has been uncommon. Seven subregional clusters were found within hspEAsia, related to subpopulations with various ethnicities, geographies and gastric cancer risks. Subpopulation-specific amino acid changes were found in multidrug exporters (*hefC*), transporters (*frpB-4*), outer membrane proteins (*hopI*) and several genes involved in host interaction, such as a catalase site, involved in H₂O₂ entrance, and a flagellin site mimicking host glycosylation. Several of the top hits, including *frpB-4*, *hefC*, *alpB/hopB* and *hofC*, have been found to be differentiated within the Americas in previous studies, indicating that a handful of genes may be key to local geographic adaptation. *H. pylori* within East Asia are not homogeneous but have become differentiated geographically at multiple loci that might have facilitated adaptation to local conditions and hosts. This has important implications for further evaluation of these changes in relation to the varying gastric cancer incidence between geographical areas in this region.

DATA SUMMARY

All supporting data have been provided within the article or through supplementary data files. Public genome data were retrieved from the National Center for Biotechnology Information GenBank (Tables S2 and S3, available in the online version of this article). Newly sequenced genomes were deposited into GenBank with BioProject number PRJNA482300.

INTRODUCTION

Helicobacter pylori has co-evolved with human beings for at least 100 000 years and has strong association with the occurrence of gastric (stomach) cancer [1]. *H. pylori* are transmitted most effectively within households and have therefore experienced low migration rates compared to many other members of the human microbial flora. As a result, we can expect the pattern of differentiation to reflect historical human-migration patterns. The population structure of *H. pylori* worldwide has been classified into seven major groups that indeed correlate with ancient human migrations [2], of which the hpEastAsia includes at least three subgroups: hspEAsia, hspIndigenousAmerica [3] and hspMaori. The hspEAsia subgroup is thought to be ubiquitous within East Asian countries with high gastric cancer incidence, including China, Japan and Korea. These countries together make up 1/5 of the world population but account for half the global mortality from the disease [4, 5]. The hspEAsia strains have documented higher virulence than other subpopulations and have diverged from the Western strains in several proteins including virulence factors [6, 7]. Many genes have also diverged *within* this region [6].

The prevalence of gastric cancer shows geographic and ethnic variations also within East Asia. Some north and southeast areas of China such as Fujian have higher incidence [8], whereas some west regions, such as Yunnan, have low incidence (Fig. 1) [9, 10]. China also shows diverse distribution of population ethnicities; in southwest Yunnan province, more than 20 ethnicities exist. South Korea has been reported with the highest incidence in the world [11, 12] and similar is true for the Japanese main islands (including Hokkaido) [13], while in Okinawa, where the ethnic composition is different, the incidence is low [13, 14]. Previous studies [8, 15] on the high-risk regions have suggested that diet, lifestyle and *H. pylori* properties may contribute to the high

Impact Statement

East Asia, especially China, Japan and Korea, has the largest gastric cancer burden in the world. *Helicobacter pylori* in this region were thought to be homogeneous and uniformly highly virulent. However, although the incidence of gastric cancer shows considerable variation within the region, little is known about the population structure of *H. pylori*. Our whole-genome analysis implies that there is substantial regional variation within East Asian *H. pylori* subpopulation hspEAsia, including differentiation between low and high incidence regions. hspEAsia strains are differentiated geographically at a handful of genes that have facilitated adaptation to local conditions, including several genes implicated in similar studies of local differentiation within the Americas. Our analysis shows that genetic variation in *H. pylori* may be associated with gastric cancer incidence, both through direct and indirect causation.

risk. However, despite the complex human migrations and evolutionary history in these areas, variation of *H. pylori* within the hspEAsia subpopulation has been poorly explored.

In the present work, we analysed a collection of genomes of *H. pylori* strains from various places in these three countries. Our phylogenetic and population genetic analysis revealed presence of pronounced regional and ethnical population structure within hspEAsia, and specific sequence differences differentiating these subpopulations. Most of these highly region-specific variants were found on proteins involved in host interaction. Furthermore, placement of the variants on protein structure provided insight into molecular mechanisms underlying regional adaptation.

Received 09 June 2021; Accepted 24 August 2021; Published 21 February 2022

Author affiliations: ¹State Key Laboratory of Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, PR China; ²Department of Infectious Diseases, Sahlgrenska Academy, University of Gothenburg, Sweden; ³Department of Clinical Microbiology, Sahlgrenska University Hospital, Västra Götaland 12 Region, Gothenburg, Sweden; ⁴Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Tokyo, Japan; ⁵Department of Environmental and Preventive Medicine, Oita University Faculty of Medicine, Oita, Japan; ⁶Department of Oral Biology, BK21 Plus Project, Yonsei University College of Dentistry, Seoul, Republic of Korea; ⁷Department of Gastroenterology, Faculty of Medicine, Oita University, Oita, Japan; ⁸Primate Research Institute, Kyoto University, Inuyama, Japan; ⁹TEAMHp (Team for East Asian Genomics of *Helicobacter pylori*)#, ¹⁰I2BC, University of Paris-Saclay, Gif-sur-Yvette, France; ¹¹Department of Infectious Diseases, Kyorin University School of Medicine, Mitaka-shi, Tokyo, Japan; ¹²Department of Computational Biology and Medical Sciences (formerly Department of Medical Genome Sciences), Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan; ¹³Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, Japan; ¹⁴Research Center for Micro-Nano Technology, Hosei University, Koganei-shi, Tokyo, Japan; ¹⁵The Center for Microbes, Development and Health, Key Laboratory of Molecular Virology and Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, PR China.

***Correspondence:** Ichizo Kobayashi, zat14430@nifty.com; Daniel Falush, danielfalush@googlemail.com; Jianzhong Zhang, zhangjianzhong@icdc.cn

Keywords: bacterial pathogenesis; flagellin glycosylation; gastric cancer; fixation index; multidrug efflux pump; outer membrane protein; population genomics.

Abbreviations: CNP, clustered nucleotide polymorphism; Fst, fixation index; MCMC, Markov Chain Monte Carlo; MOMP, major outer membrane protein; NE, North East; PDB, protein data bank; Sg, subgroup; SNP, single nucleotide polymorphism.

TEAMHp (Team for East Asian Genomics of *Helicobacter pylori*) represents the following people. Takahiro Bino: NIBB Core Research Facilities, National Institute for Basic Biology, Okazaki, Aichi, Japan. Data Analysis. Masaki Fukuyo: Department of Molecular Oncology, Graduate School of Medicine, Chiba University, Chiba 260-8670, Japan, Kazusa DNA Research Institute, Chiba 292-0818, Japan. Rumiko Suzuki: Department of Environmental and Preventive Medicine, Oita University Faculty of Medicine. Provided strains. John Harting: Pacific Biosciences, Menlo Park, CA 94025, USA. Data Analysis. Mototsugu Kato^{1,2}: ¹Division of Endoscopy, Hokkaido University Hospital, Sapporo, Hokkaido 060-8468, Japan. ²Department of Gastroenterology, National Hospital Organization Hakodate Hospital, Hakodate, Hokkaido, Japan. Provided strains. Mutsuko Konno: Department of Pediatrics, Sapporo Kosei General Hospital, Sapporo, Hokkaido, Japan. Provided strains. Yuji Kohara: Advanced Genomics Center, National Institute of Genetics, Shizuoka 411-8540, Japan. Carried out Pacbio sequencing. Christine Lambert: Pacific Biosciences, Menlo Park, CA 94025, USA. Carried out experiments. Yohei Minakuchi: Comparative Genomics Laboratory, National Institute of Genetics, Shizuoka 411-8540, Japan. Carried out experiments. Shin Nishiumi: Department of Gastroenterology, Graduate School of Medicine, Kobe University, Chuou-ku, Kobe, Hyogo, 650-0017, Japan. Provided strains. Shuji Shigenobu: NIBB Core Research Facilities, National Institute for Basic Biology, Okazaki, Aichi, Japan. Data Analysis. Noriko Takahashi: Department of Computational Biology and Medical Sciences (formerly Department of Medical Genome Sciences), Graduate School of Frontier Sciences, University of Tokyo, Tokyo 108-8639, Japan, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan, Kyorin University School of Medicine, Mitaka, Tokyo 181-8611, Japan. Atsushi Toyoda: Comparative Genomics Laboratory and Advanced Genomics Center, National Institute of Genetics, Shizuoka 411-8540, Japan. Carried out Pacbio sequencing. Ikuo Uchiyama: Laboratory of Genome Informatics, National Institute for Basic Biology, National Institutes of Natural Sciences, Okazaki, Aichi, Japan. Data Analysis. Hirokazu Yano: Department of Computational Biology and Medical Sciences (formerly Department of Medical Genome Sciences), Graduate School of Frontier Sciences, University of Tokyo, Tokyo 108-8639, Japan, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan, Graduate School of Life Sciences, Tohoku University, Sendai 980-8577, Japan. Carried out experiments. Masaru Yoshida: Department of Gastroenterology, Graduate School of Medicine, Kobe University, Chuou-ku, Kobe, Hyogo, 650-0017, Japan. Provided strains.

†These authors contributed equally to this work

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Four supplementary tables, four supplementary figures and one supplementary text.

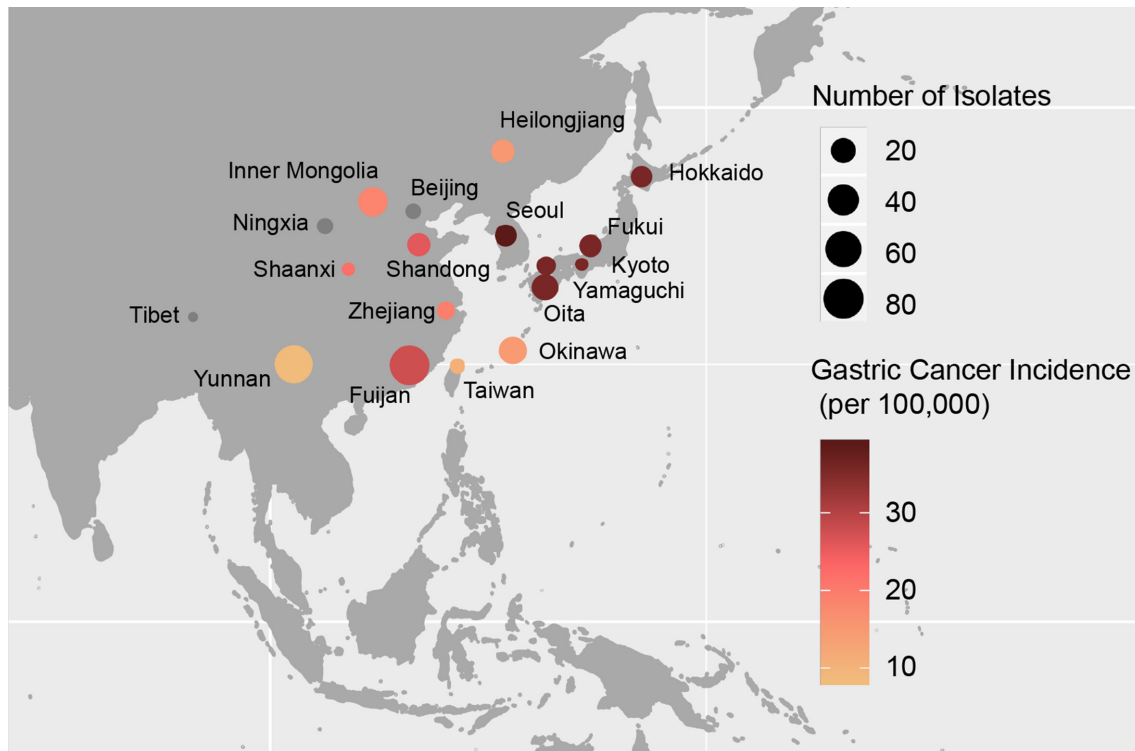


Fig. 1. Geographic distribution of the East Asian *H. pylori* strain dataset of this study. Shade of the circle fill indicates gastric cancer incidence. Grey fill indicates missing incidence data. All are age-standardized gastric cancer rates from 2014 by world population except that for Okinawa, only the data from 2016 is available.

METHODS

Strain collection across East Asia

We collected 357 *H. pylori* isolates from China, Japan and South Korea, isolated between 1999 and 2018, including 11 provinces of China and six regions of Japan (Tables S1,S2, Fig. 1). We used data from 2014 to show regional risk levels in China and compared it with data from Japan and Korea [8, 9, 11–14]. Our sample of 77 Yunnan isolates includes strains from four ethnic minorities and the Han majority individuals.

Genome sequencing

Genomic DNA was extracted using the Qiagen DNeasy Mini Kit and genomes were sequenced using Illumina or PacBio sequencers. For Illumina sequencing, paired-end libraries were created and sequenced using the Illumina HiSeq and MiSeq platform. For PacBio sequencing, quality assessment of genomic DNA, SMRTbell library preparation and data evaluation were performed, and sequenced using the PacBio RS II platform. Sequences were deposited into GenBank with BioProject number PRJNA482300. Combining publicly available genomes (available in January 2018 when the study started) with these newly sequenced genomes, we collected a dataset consisting of 381 *H. pylori* genomes for further analysis (Table S2).

Genomic comparison

We used snippy [16] to perform a whole-genome alignment with XZ274 as the reference genome and extracted 225 942 variable core-genome sites. For the phylogenetic analysis, the dataset of 381 strains were combined with 1–3 reference sequences for each of the major *H. pylori* populations (Table S3).

Population structure and phylogenetic analysis

The concatenated whole-genome SNPs and coordinates were used to prepare the haplotype file for ChromoPainter and fineSTRUCTURE analyses following the instructions from website (<http://www.paintmychromosomes.com>). In total, 13 highly clonal sequences were removed prior to the analysis, resulting in a comparison of 370 genomes. Using each genome as both donor and recipient haplotypes, we used ChromoPainter to calculate the number of genetic chunks exported from a donor to a recipient and generated a co-ancestry matrix. Then the co-ancestry matrix file was imported into fineSTRUCTURE by setting the burn-in

Table 1. High Fst SNPs distributed in each hspEAsia subgroup and involved genes*

Subgroup	Gene	Annotation	Fst	SNP coordinate
Sg1 (YunnanMP)	<i>hefC</i>	Inner pump of a multidrug efflux system	0.891	827628
	<i>porB</i>	Pyruvate-ferredoxin oxidoreductase subunit beta	0.800	1175532
	<i>hefC</i>	Inner pump of a multidrug efflux system	0.757	827633
	<i>flaA</i>	Flagellin A	0.748	833437
	<i>gltA</i>	Citrate synthase	0.731	1418093
	<i>tlpD</i>	Chemotaxis sensor	0.716	836384
	<i>katA</i>	Catalase	0.714	507178
	<i>gltA</i>	Citrate synthase	0.706	1418291
	<i>gltA</i>	Citrate synthase	0.706	1418293
	<i>flaA</i>	Flagellin A	0.705	833501
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.694	1536189
	<i>oppD</i>	Oligopeptide permease ATPase protein	0.669	283539
	<i>oppD</i>	Oligopeptide permease ATPase protein	0.663	283540
	<i>porB</i>	Pyruvate-ferredoxin oxidoreductase subunit beta	0.649	1175543
	<i>hopI</i>	Outer membrane protein HopI	0.638	1220823
	<i>gltA</i>	Citrate synthase	0.626	1418759
Sg2 (Fujian)	<i>hofC</i>	Outer membrane protein involved in adhesion and diffusion of cations including antibiotics	0.855	496438
	<i>hofC</i>	Outer membrane protein involved in adhesion and diffusion of cations including antibiotics	0.850	496437
	<i>hofC</i>	Outer membrane protein involved in adhesion and diffusion of cations including antibiotics	0.800	496436
	<i>katA</i>	Catalase	0.756	507441
	<i>hopB/alpB</i>	Outer membrane protein, Omp21	0.704	983839
	<i>flaA</i>	Flagellin A	0.651	833026
Sg3 (China North East)	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.752	1536144
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.752	1536145
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.738	1536086
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.738	1536093
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.713	1536091
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.706	1535959
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.701	1535957
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.701	1535965
	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.637	1535958
Sg5 (Korea)	<i>hypC</i>	hydrogenase expression/formation protein	0.652	959689
Sg6 (Okinawa)	<i>rhoD</i>	Rhodanese, a cyanide-detoxifying enzyme	0.646	1299845
Sg7 (Japan_Major)	<i>frpB-4</i>	TonB-dependent outer membrane Ni importer	0.658	1536096

* Criteria for showing high Fst of each subpopulation in this table is based on the following:

Fst > 0.6.

Sg4 (Mongolia) is excluded from the table as it does not belong to the hspEAsia subgroup.

SNPs of high/low and China_All are not listed because they were specific comparisons, not representing subpopulations.

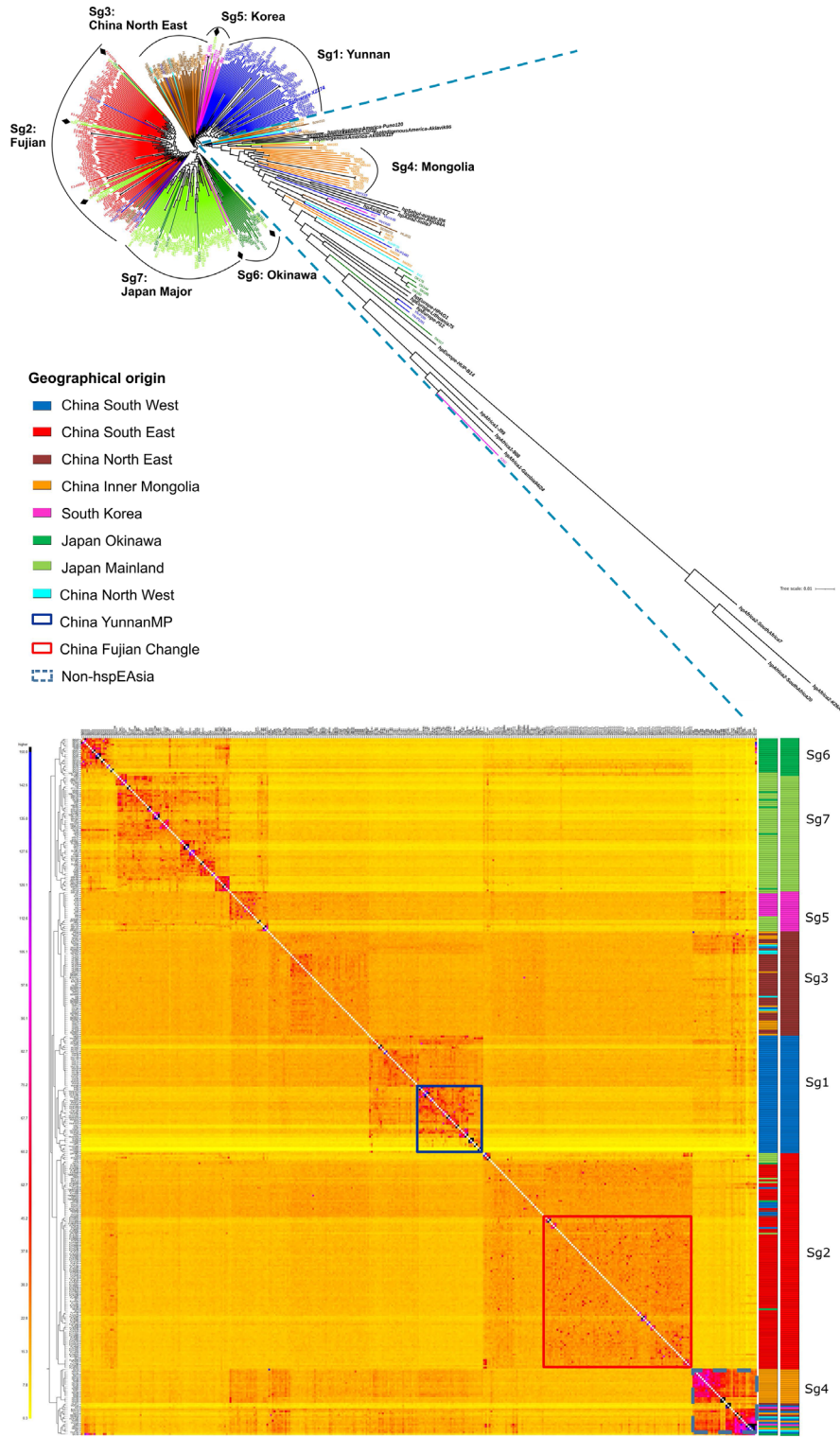


Fig. 2. Phylogenetic and population genetic analysis of East Asia *H. pylori* strains. (a) Phylogenetic tree of East Asia genomes combined with reference sequences from global *H. pylori* populations. Diamonds indicate Japanese and Korean isolates that are dispersed in other subgroups. (b) Co-ancestry matrix of East Asia isolates calculated using fineSTRUCTURE after removing isolates with highly clonal relatedness. The tree inferred on the left of the co-ancestry matrix indicates the relationships between groups with distinct ancestry profile as inferred by fineSTRUCTURE. On the right side, the first column shows the geographic origin of each isolate. The second column indicates each subgroup with a distinct colour.

Table 2. Genes with a SNP with $F_{st} > 0.5$ in the top 20 lists

Criterion	Gene	Locus tag	Figure	Category	Annotation	Residue function	
Fst > 0.6+ top 20 in different pairwise comparisons	<i>hefC</i>	HP0607	3(b)	Efflux pump	Inner membrane component of the multidrug HefABC efflux pump	Channel entrance	
	<i>frpB-4</i>	HP1512	4(a)	TonB-dependent importer	Outer membrane nickel importer	Ligand binding	
	<i>flaA</i>	HP0601	6(b)	Motility	Flagellin A of flagella, involved in immune system evasion	Glycosylation for host mimicry and more	
	<i>katA</i>	HP0875	6(a)	Sensor	Catalase sensing and destroying H ₂ O ₂	Channel entrance; Dimer-dimer interface	
	<i>hopI</i>	HP1156		Outer membrane protein	Outer membrane protein HopI		
	<i>hopB/alpB</i>	HP0913		Outer membrane protein	Hop family adhesin HopB/AlpB/Omp21.		
	<i>tlpD</i>	HP0599	6(c)	Sensor	Chemotaxis sensor sensing and destroying HOCl	Active site	
	<i>exbB-2</i>	HP1339	S3C	TonB-dependent importer	Energizer/motor in the inner membrane driven by proton	Channel entrance	
	Fst > 0.6 two or more SNPs	<i>hofC</i>	HP0486		Outer membrane protein	Hof family protein implicated in adhesion and antibiotics diffusion	
		<i>porB</i>	HP1111		Energy metabolism	Subunit of pyruvate:ferredoxin oxidoreductase, part of the microaerophilic metabolic pathway leading to acetyl ~CoA	Subunit interaction
<i>glfA</i>		HP0026	5(b)	Energy metabolism	Citrate synthase, the first enzyme in TCA cycle incorporating acetyl ~CoA	Subunit interaction	
<i>oppD</i>		HP0250	5(a)	Importer	Oligopeptide ABC transporter, ATPase subunit	Cofactor binding	

Continued

Table 2. Continued

Criterion	Gene	Locus tag	Figure	Category	Annotation	Residue function
Esi >0.5	<i>ompA-18</i>	HP1125		Outer membrane protein	OmpA family peptidoglycan-associated lipoprotein	Ligand binding
	<i>fixP</i>	HP0147	S4C	Energy metabolism	Subunit of cytochrome c oxidase in aerobic respiration	Channel entrance
	<i>hyrC</i>	HP0899			Hydrogenase expression/formation protein	(Synonymous change)
	<i>hefD</i>	HP0971	3(c)	Efflux pump	Outer membrane component of the HefABC multidrug efflux pump	Protein binding
	<i>copA</i>	HP1503	S2E	Exporter	Copper(I) exporter	Protein stability
	<i>mscS-1</i>	HP0284		Sensor	Mechano-sensor sensing tension in the membrane	Scaffolding in the periplasm
	<i>metQ</i>	HP1564	S2D	Importer	D-methionine ABC transporter, methionine-binding subunit	Ligand binding
	<i>frpB-1</i>	HP0876	S3A	TonB-dependent importer	Outer membrane haem importer	Ligand binding
	<i>fecA-1</i>	HP0686	S3B	TonB-dependent importer	Outer membrane iron(III) dicitrate importer	Ligand binding; Channel (plug)
	<i>exbD-2</i>	HP1340	S3C	TonB-dependent importer	Inner membrane energizer/motor driven by proton	Proton channel entrance
	<i>hpaA</i>	HP0797	S3F	Outer membrane protein	Neuraminylactose-binding hemagglutinin	Subunit interaction
	<i>omp</i>	HP0358		Outer membrane protein	Putative outer membrane protein	
	<i>panD</i>	HP0034	S4A (i)	Micronutrient synthesis	Vitamin B5 synthesis	Interaction between cleaved peptides
	<i>bioD</i>	HP0029		Micronutrient synthesis	Vitamin B7 synthesis	
	<i>igt</i>	HP0281	S4A (iii)	Micronutrient synthesis	Q-base synthesis; Q-base on tRNA affects translation accuracy	Active site
	<i>rhoD</i>	HP1223	S4B	Detox	Rhodanese detoxifying cyanide generated in microbiome	
	<i>rkiP</i>	HP0218	S2B	Oncoprotein	Mimic of human RKIP tumour suppressor	Interaction with signal peptide
	<i>hepX</i>	-	S2A	Effector	SLR family with repeated alpha-helix pairs	Human protein binding
	<i>dsbI</i>	HP0595	S3D	Secretion	S-S formation	Active site
	<i>jog</i>	HP1451		Secretion	Regulator of VirB11/Cag-alpha gate of Cag secretion system	Cofactor binding
	<i>lpxE</i>	HP0021	S2C	Membrane lipid modifier	Lipid A 1-phosphatase to hide it from the innate immune response	Substrate binding; active site
	<i>cfaS</i>	HP0416	S3E	Membrane lipid modifier	Cyclopropane-fatty-acyl-phospholipid synthase for acid protection	Substrate binding
	<i>fur</i>	HP1027	S4D	Transcription factor	Regulator of Fe/Ni import, redox balance and acid response	Cofactor binding

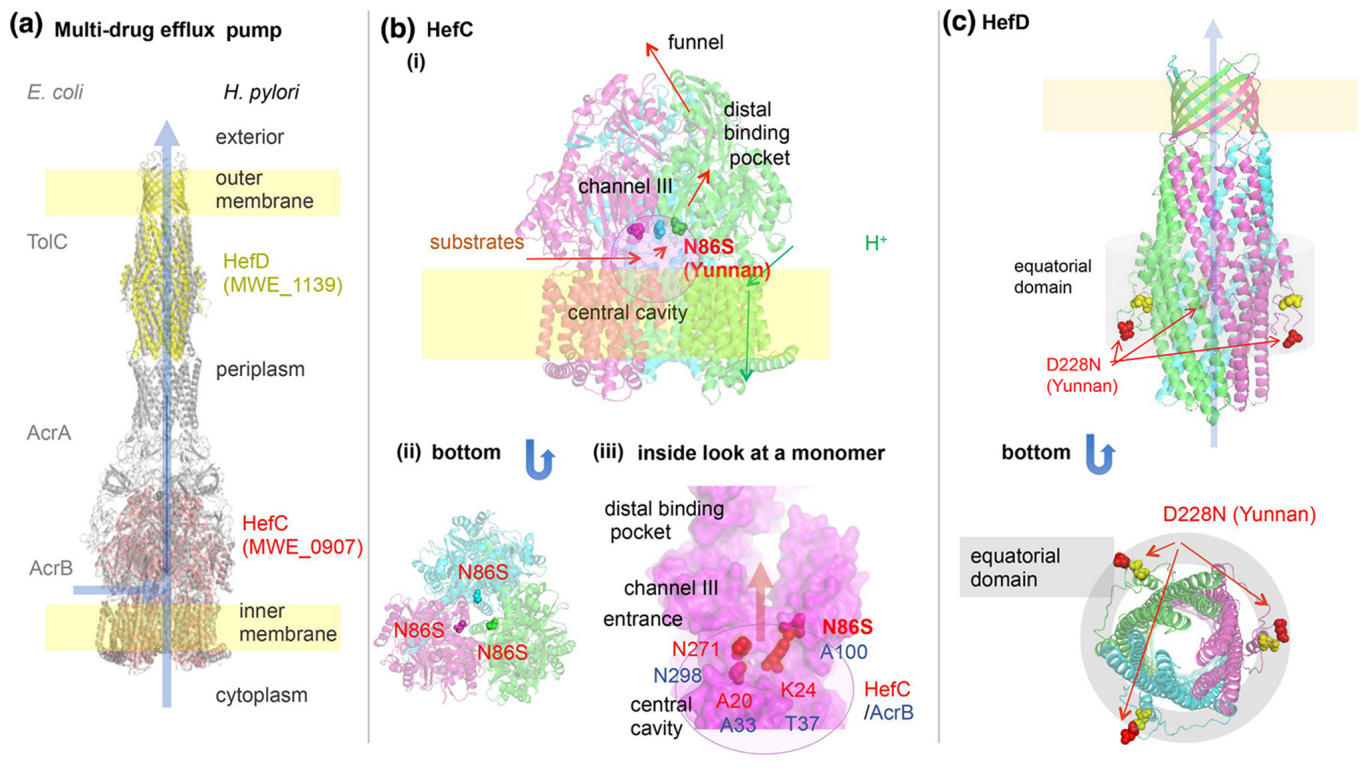


Fig. 3. Inferred structure and differentiated amino acid sites of multidrug efflux pump related proteins. (a) TolC-AcrA-AcrB of *E. coli* and its *H. pylori* homologues. (b) HefC (MWE_0907), modelled on PDB 3W9I (MexB of *Pseudomonas aeruginosa*). (iii) HefC modelled on PDB 3A0D (AcrB of *E. coli*). Yunnan-specific N86S is at the entrance of channel III. (c) HefD (MWE_1139), modelled on PDB 5BUN (ST50 from *Salmonella enterica* subsp. *enterica* serovar Typhi). Yellow spheres indicate Mongol-differentiated E219D.

and Markov chain Monte Carlo (MCMC) chain of 100 000 iterations to generate clusters for all the individual strains [17, 18]. We also generated a phylogenetic tree using FastTree (<http://www.microbesonline.org/fasttree>) [19], which was labelled using iTol (<https://itol.embl.de>) [20].

Definition of subgroups and calculation of *Fst* between subgroups

Subgroups were defined according to the population structure analysed by fineSTRUCTURE (the left-side tree). We named the subgroups assigned by fineSTRUCTURE as ‘Sg’ and also according to the geographic origin of the majority of a clade (Table S2). This criterion for subgroup definition is relatively a rough classification. To identify the SNPs attributed to the divergence of subgroups more accurately, we used a more stringent criterion for a further definition of the subgroups. Only those isolates assigned into a singular cluster in the fineSTRUCTURE tree were defined to form a subset of a subgroup and used for *Fst* calculation. For example, for the large sampling areas such as China southwest isolates, only those from Yunnan Mosuo and Pumi ethnicities, which clustered into a singular clade with low levels of admixture/recombination, were defined as a subgroup, ‘YunnanMP’. For China southeast, only those from Fujian Changle that clustered into a singular clade were defined as a subgroup, ‘Fujian’. Subgroups with mixed ethnicities or high levels of admixture were not used in the calculations.

For each SNP site, we calculated a fixation index (*Fst*) using PopGenome R package [21] for each subgroup such that *Fst* (sg1) = pairwise calculation of *Fst* of sg1 versus all other subgroups. We also compared isolates from China with those from Japan and South Korea and the two lower-incidence regions, Yunnan and Okinawa, versus high incidence regions (the remainder of hspEAsia strains).

Mapping high-*Fst* SNPs on protein structure

We located each of the SNPs with the highest *Fst* values on the reference genome of XZ274, a Tibetan strain, to identify its gene and effect on amino acid sequence. If the gene was missing or atypical in this strain, we used strain F57, a Japanese strain, or 26695, a European strain, instead as shown in Table S4. We mapped the amino acids on solved *H. pylori* protein structure or on protein structure homology-modelled by SwissModel and its repository for 26695 (<https://swissmodel.expasy.org/repository>). We analysed and presented them by PyMOL [22].

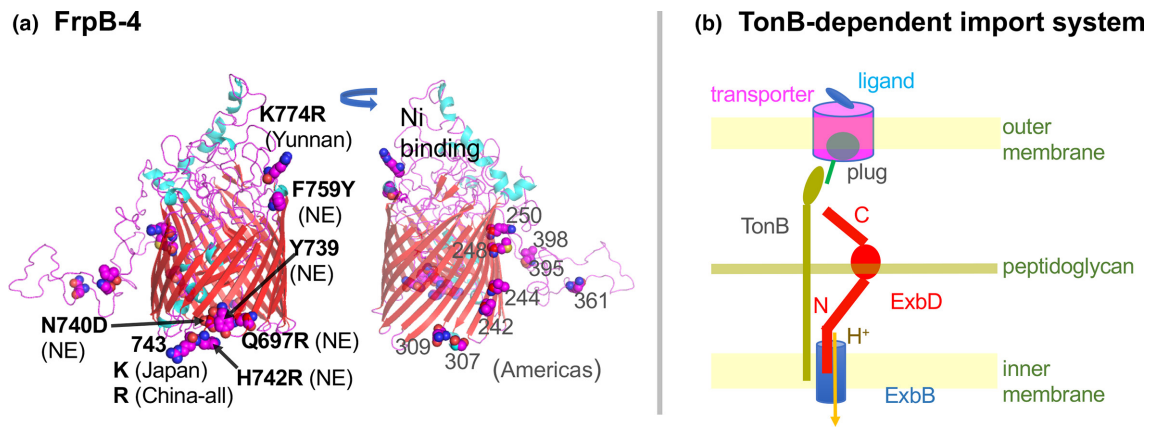


Fig. 4. (a) FrpB-4 (MWE_1700) modelled on PDB 4AIQ (FrpB of *N. meningitidis*). The left figure shows East Asia-specific sites. NE means north east China. The right one shows Americas-specific sites. (b) TonB-dependent transport system. When the transporter in the outer membrane catches a ligand, it pushes out its plug for TonB to pull and open the channel. This movement of TonB is energized by ExbBD motor in the inner membrane driven by H⁺ flow.

RESULTS

East Asian *H. pylori* population structure is associated with geography and host ethnicity

To analyse the genetic structure of *H. pylori* in China, Japan and South Korea, we constructed a phylogenetic tree (Fig. 2a) and clustered the strains using fineSTRUCTURE [17, 18] (Fig. 2b). The two methods gave broadly concordant clustering and indicated differentiation at multiple scales. By including reference genomes from the other main *H. pylori* populations, we found that the 52 most differentiated strains do not belong to hspEAsia and have had distinct evolutionary histories. One main cluster of these comprised isolates from individuals of Mongolian ethnicity (Sg4, Fig. 2b), which in the tree grouped between the references of hspIndigenousAmerica (previously called hspAmerind) and hpAsia2 populations. We also identified a cluster of Okinawan strains diverging after hpEurope, and before hpAsia2 and hpEastAsia (Fig. 2a), also shown in the lowest line of the co-ancestry matrix (Fig. 2b). This likely corresponds to group C in MLST and STRUCTURE analysis of Okinawan strains [23]. Because these populations are best analysed in the context of broader regional variation, we focused our remaining analysis on variation within the 329 hspEAsia isolates in our sample.

HspEAsia isolates are relatively homogeneous but show fine-scale differentiation that is strongly correlated with geography. The majority of Japanese isolates clustered together (Sg7), with Okinawan hspEAsia strains forming a distinct subpopulation (Sg6). The latter likely corresponds to group A in MLST [23]. Two strains from Hokkaido cluster with the Sg6 Okinawan strains. This is likely related to the evolutionary structure of Japanese people and of the Okinawa people [24]. The native ethnicity remains in Hokkaido (as Ainu ethnicity) and in Okinawa.

South Korean isolates also form a distinct subpopulation, which cluster together with isolates from the Northeast of China, concordant with its geographic location. A few Japanese strains are in the Korean cluster while a few Korean strains are in the main Japanese cluster and branching around the same time as the Okinawa cluster (Fig. 2a,b). This may reflect immigration from Korea to Northern Kyushu and Okinawa. Within China, there is clear differentiation between the Northeast, Southeast and Southwest areas. Within the Southwest, a subpopulation containing strains from Yunnan Mosuo and Pumi ethnicities could also be observed. A Tibetan isolate clustered with the Pumi population, consistent with their migration and mixture history. Within the Southeast, there was also a subpopulation (Sg2), specific to Fujian Changle, a region of high gastric cancer incidence (Fig. 2, table S1). The Southeast cluster also includes multiple clusters of Japanese/Okinawa strains (Fig. 2a,b) likely reflecting later immigration to Japan/Okinawa.

Many genetic variants show strong differentiation by subpopulations

In order to explore the genetic basis of local differentiation, we calculated fixation index, *F_{st}*, between the hspEAsia subpopulations identified by fineSTRUCTURE. To obtain more reliable SNPs associated with subgroup separation, we used the YunnanMP subcluster of Sg1 and Fujian subcluster of Sg2, along with the remaining five subgroups defined by the fineSTRUCTURE analysis. For most of the subpopulations, more than 99% of SNPs were weakly differentiated, with *F_{st}* less than 0.3 (Fig. S1). The Korean subpopulation had a smaller sample size and showed the weakest *F_{st}* values. We defined clustered nucleotide polymorphisms (CNPs) as occurring when two or more SNPs with high *F_{st}* were found in the same gene. CNPs likely result from the co-inheritance of different SNPs on the same gene fragment.

To functionally interpret differentiation between the populations, we focused on the SNPs with the top 20 *F_{st}* values in each subgroup and then removed SNPs with *F_{st}* < 0.5. This resulted in a list of 56 genes, several of which occurred more than once (Table S4, bold,

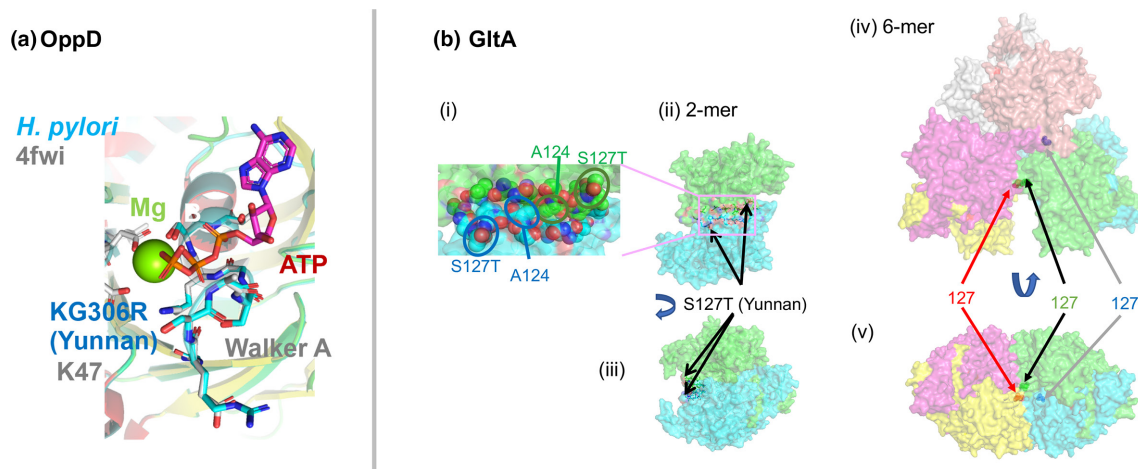


Fig. 5. (a) OppD (MWE_0327), ATP-binding subunit of an oligopeptide ABC transporter as modelled on PDB 4FWI (*Thermoanaerobacter tengcongensis*). (b) Citrate synthase, GltA (i) (ii) (iii) MWE_1570 modelled on PDB 3MSU (*Francisella tularensis* homolog). (iv) (v) On PDB 2h12 (*Acetobacter aceti*).

summarized in Table 1). In the main text, we focus on the genes with the strongest evidence for differentiation. Eight of these had one SNP with $F_{st} > 0.6$ and appeared in at least two top 20 lists in different pairwise comparisons. Another four had at least two SNPs with $F_{st} > 0.6$. Of these, HofC has multiple non-synonymous SNPs with F_{st} up to 0.86, marking it out as also being a particularly strong candidate for being differentiated by natural selection (Table 2).

To predict the effect of differentiated variants on *H. pylori* biology and pathogenesis, we functionally annotated the genes and interpreted the impact of differentiated amino acids on the protein structures, solved or based on homology modelling. The majority of genes containing the most differentiated SNPs could be grouped into four major categories; (i) transporters, (ii) outer membrane proteins, (iii) metabolism and (iv) host interaction.

Transporters

The Hef multidrug efflux pump

H. pylori carries four gene clusters that each encode a set of RND superfamily of multidrug efflux pump corresponding to TolC-AcrA-AcrB of *E. coli* (Fig. 3). They pump out endogenous bile salts and ceragenins as well as various antibiotics [25–27]. The single most differentiated SNP in our analysis is in *hefC* (HP0607) in the YunnanMP subpopulation, with an $F_{st}=0.89$. This N86S is also the most differentiated between the lower cancer incidence regions (Yunnan and Okinawa) and the remainder. This residue corresponds to the gate to channel III for planar aromatic cations in the *E. coli* homologue [28] and the regional adaptation may therefore remodel this gate to export chemicals of this type at different ratios. Also, the outer component of the efflux pump, HefD, showed YunnanMP specific residues in the equatorial domain that may be involved in the interaction with the inner component to open the aperture.

The TonB-dependent nickel importer FrpB-4

A gene containing multiple highly differentiated SNPs, especially in North Eastern China, is *frpB-4* (HP1512), encoding an outer membrane transporter of nickel of some form [29]. It is a member of TonB-dependent transporter family, which forms a trimer of 22-stranded beta barrels each filled with a ‘plug’ (Fig. 4a,b). *H. pylori* reference strain 26695 carries four *frpB* homologues: *frpB-1* (HP0876), *frpB-2/3* (HP0916/5) and *frpB-4* (HP1512).

Ligand binding lets TonB change the conformation of the plug, which opens a channel, a process energized by the ExbBD proton-driven motor [30] (Fig. 4b). Four China North East-specific F_{st} sites cluster (Fig. 4a), with three sites (739, 740, 742) presumably representing a CNP. The next residue (743) distinguishes between Japan (K) and China (R). Northeast China-differentiated F759Y (F for the remainder and Y for Northeast China) and YunnanMP-differentiated K774R are situated above the barrel in the model and may interact with the ligand. Various regions of the Americas show region-specific amino acid changes in other areas of the protein [31], out of which three are predicted to be in the decoy loop. Taken together, these changes may affect nickel transport and, consequently, urease activity, since the urease enzyme requires nickel for acid acclimation [32]. The changes could be related to regional differences in host nickel metabolism and stomach acidity.

Other transporters

Hof proteins (*Helicobacter*-specific outer membrane protein family [33]) are 18-stranded β -barrels homologous to Occ family of *Pseudomonas* and *Campylobacter jejuni* MOMP (major outer membrane protein) involved in passive diffusion of cations including

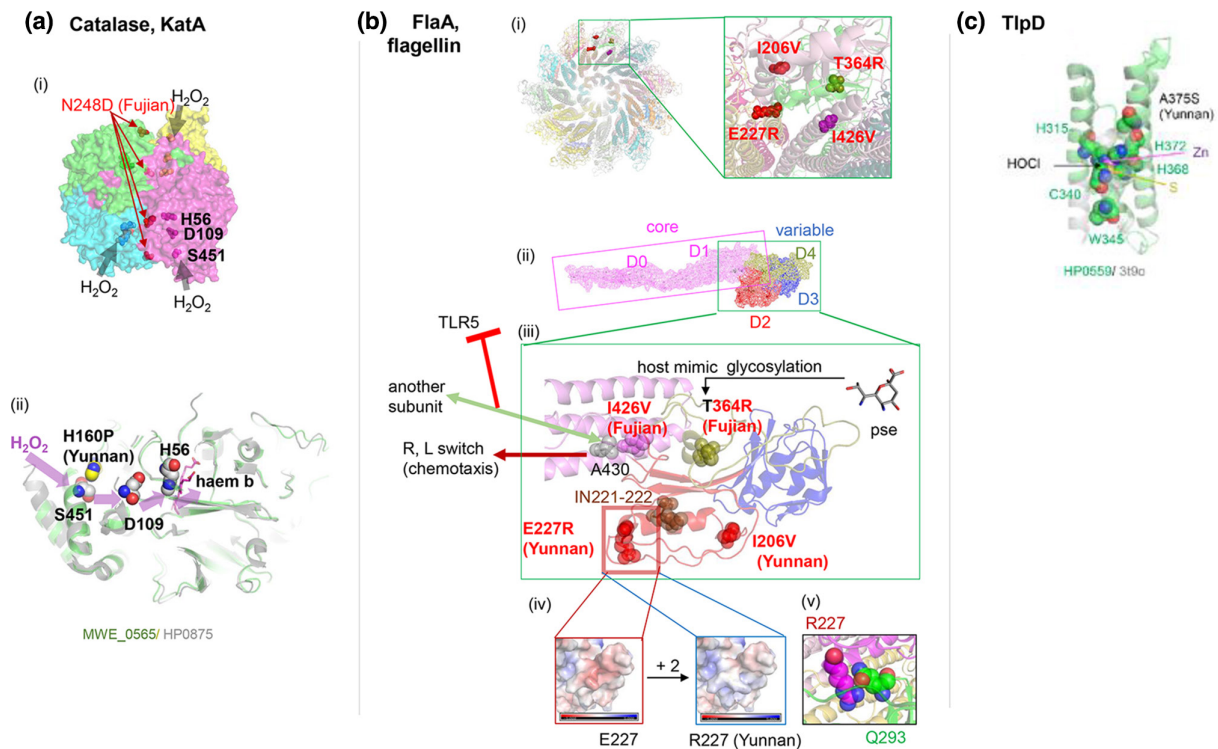


Fig. 6. (a) Catalase KatA (HP0875, PDB 2A9E). (i) Dimer of dimer. (ii) Channel for H₂O₂, Yunnan-specific P160H by its entrance S451 drastically changes local conformation and surface electric charge (mutagenesis in PyMOL). (b) Flagellin, FlaA. (i) A view from the distal end of the 22-mer model of FlaA (MWE_0913) on G508A mutant of *C. jejuni* homologue (PDB 6x80) with four population-specific amino acid changes in two interacting monomers. (ii) Monomer with five domains. (iii) Population-specific amino acid changes. pse: pseudaminic acid. (iv) Surface electric charge change by E227R. (v) R227 interaction with a neighbouring monomer. (c) TlpD, chemotaxis receptor for HOCl. HP0559 modelled on PDB 3T90, the regulatory CZB domain of DgcZ (*E. coli*).

antibiotics and in adhesion [34–36]. HofC (HP0486), required for *H. pylori* colonization in mice [37], contains the most differentiated SNPs in Fujian, a region of high cancer incidence with *Fst*=0.86. The gene is highly variable in global strains and shows many America-differentiated SNPs and region-differentiated SNPs within the Americas [31], within one narrow region. Fujian-differentiated D186S in HP0486 (166 in MWE_0556) lies at a distance from them. Fujian-differentiated V9A is in its signal peptide.

OppD (HP0250), the cytoplasmic subunit of the oligopeptide ABC transporter, has YunnanMP-differentiated KG306R within the ATP binding Walker motif A (Fig. 5a).

Outer-membrane proteins

HopB/ AlpB/ Omp21 (HP0913), an adhesin of the Hop family required for colonization, carries Fujian-differentiated N289D and N286H. According to a previous study, 24 polymorphic sites within 49 bp in AlpB are enriched for Asian ancestry in hspEuropeColumbia and 32 polymorphic sites within 65 bp were enriched for Asian ancestry in hspAfrica1Nicaragua populations [31].

Another member of the Hop family of outer-membrane proteins, HopI (HP1156), has a site (467) that distinguishes between Japan (H) and China-all (D) and YunnanMP-differentiated V633L.

Central metabolism

The region-differentiated amino acid changes involve a handful of key metabolic enzymes.

Citrate synthase GltA (HP0026, MWE_1570), is the first enzyme in the TCA cycle catalysing the conversion of acetyl-CoA and oxaloacetate to citrate. Yunnan-differentiated S127T is located between the two identical monomers and is likely involved in their association as well as in dimer–dimer association to form a 6-mer (Fig. 5b). The differentiated SNP might change the quaternary structure. Mutation of A124 in this interface was found in experimental evolution in *E. coli* [38]. In addition to GltA, two other key metabolic enzymes, PorB (HP1111), a subunit of pyruvate:ferredoxin oxidoreductase, and FixP (HP0147, MWE_0216), a subunit of cytochrome c oxidase, had high *Fst* values in YunnanMP. PorB, a key enzyme in the microaerophilic metabolism of *H. pylori*, converts pyruvate to acetyl-CoA,

the substrate of GltA, and FixP is a component in aerobic respiration. The Yunnan-differentiated residue is at the proton entrance (Fig. S4C). Together, these changes might affect the metabolic capacity of this regional *H. pylori* subpopulation.

Host interaction

In addition to the genes listed above, region-specific non-synonymous variants are present in several genes that are annotated as known virulence or host interaction factors.

Catalase KatA (HP0875) (Fig. 6a) detoxifies H₂O₂ generated by host immune cells. It also binds host vitronectin, thereby protecting against complement-mediated killing [39]. The preferred route for H₂O₂ is the channel S451-D109-H56-haem (Fig. 6a (ii)) [40]. YunnanMP-differentiated P160H by the entrance S451 drastically changes local conformation and surface electric charge. Fujian-differentiated N248D with -1 change in the electric charge takes place near the dimer-dimer interface (Fig. 6a (i)) likely changing their interaction.

The flagellar filament made of flagellin FlaA (HP0601) (Fig. 6b), is involved in motility, cell adherence and immune modulation. We have modelled it using the similar *C. jejuni* homologue [41]. The flagellin has rod-shaped domains forming a hydrophobic core, and the other domains decorating the surface of the filament are hypervariable. The flagellin is glycosylated by pseudaminic acid at several residues to stabilize the flagellum and to mimic host-cell surface, a way for the bacterium to modulate the immune response [42], but Fujian-differentiated T364R eliminates one of these sites (Fig. 6b (iii)). E227R drastically changes the surface electric charge and likely affects its interaction with a neighbouring monomer (Fig. 6b (iv)(v)). Another Fujian-differentiated residue 426 in the conservative core is next to residue 427, which is involved in evasion from TLR5-mediated innate immunity through subunit interaction [41]. Furthermore, residue 430 adjacent to 426 in the 3D structure is somehow involved in switching between R and L conformations for swimming/tumbling in chemotaxis in *Campylobacter* [43].

TlpD (HP0599, MWE_0916) (Fig. 6c) is a cytosolic chemotaxis sensor required for colonization. TlpD senses HOCl, an antimicrobial produced by neutrophils during inflammation [44]. HOCl oxidizes a conserved cysteine (C340) within a 3His/1Cys Zn-binding motif to inactivate chemo-transduction signalling. YunnanMP-differentiated A375S is right by this motif. Additional proteins in Table 2 are described in the Supplementary Material.

DISCUSSION

Analyses of *H. pylori* in East Asia have tended to emphasize their homogeneity and uniformly high virulence potential. A single subpopulation of the bacteria, hspEAsia, is prevalent in the region. The hspEAsia strains have been found to be invariably CagPAI positive, with the *cagA* gene containing the ABD EPIYA motif, which is thought to promote strong binding of the protein to SHP-2 [45]. Our large collection of genomes of *H. pylori* from multiple regions in China, Japan and Korea confirm these observations. Only a small number of these isolates, of which a majority are from Mongolia and Okinawa, belonged to other *H. pylori* populations. For the scope of this study, these were excluded from subsequent analyses. All but 13 of the total 381 isolates were *cagA* positive and 355 have the characteristic ABD EPIYA type, including 316 out of 329 hspEAsia isolates.

Our results add a layer of complexity to the picture of uniformity by demonstrating that there is differentiation of *H. pylori* strains of the hspEAsia subpopulation between regions in East Asia. Despite a large burden of gastric disease in the region, most *H. pylori* infections by hspEAsia are asymptomatic, and the gastric cancer incidence varies widely across the region, especially in China. A large part of these differences might be attributable to diet and environment. However, our results imply that there are also bacterial factors differentiating these regions, which may be significant for disease development, especially because the bacteria themselves can adapt to environmental conditions [6].

Geographic differentiation between populations accumulates progressively when migration rates between them are low. Further, adaptation of bacteria to differences in environmental conditions can greatly accelerate the process of differentiation in specific regions of the genome. Our results, in combination with a previous study of genetic variation within the Americas [31] suggest that there are a handful of loci that have undergone rapid differentiation in several regions, and therefore may be considered keys for host adaptation. These include the genes *frpB-4*, *hefC*, *alpB/hopB* and *hofC*.

Our strategy to identify geographically differentiated SNPs by dividing one population (hspEAsia) into minimal subpopulations, therefore strains with consistent population and strain labels, and comparing fixation index (Fst) site-by-site between these populations reveal numerous loci of differentiation (Table S4). Of these we discuss the 12 with the strongest evidence for being involved in local adaptation in more detail in the main text in this paper. Some of the region-specific SNPs are in genes encoding for proteins that have been implicated in host interaction and virulence in the narrow sense: attack by immune system (catalase, TlpD), host adhesion (HopB/AlpB and several outer membrane proteins), and host surface mimicry (flagellin FlaA). Several of the other genes are transporters that may have implications for antimicrobial resistance, or are involved in nutrient acquisition. These results suggest that various host-adaptive changes in many host-interaction proteins lead to population differentiation. A similar gene set was found when rapid genome changes were investigated in shorter-term, intra-body micro-evolution [46].

Epidemiological and experimental evidence suggests that iron-deficiency increases *H. pylori* virulence and risk of gastric cancer [47]. In our analyses we could see both in iron and nickel metabolism highlighted by regional changes. Apart from the above-mentioned *frpB-4*, genes encoding for the TonB motor proteins ExbB-2 and ExbD-2, haem transporter FrpB-1 [48] appear in the list of the 34 genes with Fst >0.5 and the central transcription factor ferric uptake regulator, Fur, had regional variants (Supplementary Material). The causal mechanism underlying this association is not clear but it plausibly reflects bacterial response to nutrient limitation. Simply put, it is possible that bacteria adopt more aggressive strategies in interacting with the host and its microbiome when iron and other metals such as nickel, which is necessary for urease function, is in short supply.

In other organisms, linkage means that high Fst regions often occur in large blocks, making it difficult to infer which sites are involved in local adaptation. However, *H. pylori* lineages recombine with each other, exchanging substantial fraction of their DNA in individual mixed infections [49]. The size of replacement in one event can be as short as 28 bp [50], with the result that linkage is broken down rapidly. This means that individual nucleotides can rise to high frequencies in specific populations, suggesting that local adaptation can potentially occur on a very exquisite scale.

Many of the highly differentiated amino acid changes are close to critical residues of the protein and are plausible candidates to cause important functional changes, based on 3D modelling and previous functional analyses. We have suggested possible functional consequences but validation by targeted experiments and clinical observations is necessary. Although the functional consequences of genomic differentiation of *H. pylori* within different parts of the world remain to be elucidated, the presence of this differentiation already has potential clinical utility. All else being equal, individuals who are infected by *H. pylori* that are characteristically found in high gastric cancer incidence regions are likely to be at higher risk than those associated with lower incidence regions, firstly because the bacteria may be more virulent but secondly because infection with the bacteria might also be a marker for exposure to environmental factors that underlie the high disease risk.

Funding information

This work was supported by a grant from the State Key Laboratory of Infectious Disease Prevention and Control (SKLID) (2014SKLID102) of the Chinese Centre for Disease Control and Prevention. Supported by National Science and Technology Major Project (2018Zx10712-001) and a joint project 'Isolation and sequence analysis of *Helicobacter pylori* strains collected from investigations on carriage rate'. K.T. was supported by Swedish Society for Medical research (SSMF). Parts of the bioinformatic analyses were performed on resources provided by Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under projects snic2018-8-24/upstore2017270, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. This work was also supported in part by MEXT KAKENHI (19K22543, 17H04666, 26113704, 25291080, 221S0002 to I.K., 18K14766 to Y.K.) by NIBB Collaborative Research Program (20-443,21-328) to I.K., by U.S.-Japan Cooperative Medical Sciences Program Collaborative Awards, and by Shanghai Municipal Science and Technology Major Project (No. 2019SHZDX02 to D.F.).

Acknowledgements

We thank Jonas Korlach and Primo Baybayan for help in SMRT sequencing. We would like to thank the staff of Comparative Genomics Laboratory at National Institute for Genetics, Mishima, Japan for supporting genome sequencing. We thank Dr. Chao Yang, Professor Yujun Cui and Ruifu Yang for the discussions about population genetics of *H. pylori*, and M. Zwama for discussion on HefC, Mizuki Ohno and Yosuke Kawai for discussion on human genetics.

Author group contributors

TEAMHp comprises Takahiro Bino, Masaki Fukuyo, Rumiko Suzuki, John Harting, Mototsugu Kato, Mutsuko Konno, Yuji Kohara, Christine Lambert, Yohei Minakuchi, Shin Nishiumi, Shuji Shigenobu, Noriko Takahashi, Atsushi Toyoda, Ikuro Uchiyama, Hirokazu Yano and Masaru Yoshida.

Conflicts of interest

Christine Lambert and John Harting are full-time employees at Pacific Biosciences of California, a company developing single-molecule sequencing technologies. The other authors declare that they have no conflict of interest.

References

- Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, et al. Age of the association between *Helicobacter pylori* and man. *PLoS Pathog* 2012;8:5.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, et al. Traces of human migrations in *Helicobacter pylori* populations. *Science* 2003;299:1582-1585.
- Munoz-Ramirez ZY, Pascoe B, Mendez-Tenorio A, Mourkas E, Sandoval-Motta S, et al. A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. *ISME J* 2021;15:78-92.
- GBD 2017 Stomach Cancer Collaborators. The global, regional, and national burden of stomach cancer in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease study 2017. *Lancet Gastroenterol Hepatol* 2020;5:42-54.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
- Kawai M, Furuta Y, Yahara K, Tsuru T, Oshima K, et al. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. *BMC Microbiol* 2011;11:104.
- Furuta Y, Yahara K, Hatakeyama M, Kobayashi I. Evolution of *cagA* oncogene of *Helicobacter pylori* through recombination. *Plos One* 2011;6:e23499.
- Wong BC, Lam SK, Wong WM, Chen JS, Zheng TT, et al. *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA* 2004;291:187-194.
- Chen W, Sun K, Zheng R, Zeng H, Zhang S, et al. Cancer incidence and mortality in China, 2014. *Chin J Cancer Res* 2018;30:1-12.
- He J, Chen W. *Annual Report of Cancer Registration in China*. National Cancer Center, China, 2017.

11. Jung KW, Won YJ, Kong HJ, CM O, Lee DH, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2011. *Cancer Res Treat* 2014;46:109–123.
12. Eom BW, Jung KW, Won YJ, Yang H, Kim YW. Trends in Gastric Cancer Incidence According to the Clinicopathological Characteristics in Korea, 1999–2014. *Cancer Res Treat* 2018;50:1343–1350.
13. Bray FC, Mery L, Piñeros M, Znaor A, Zanetti R, et al. CI5PLUS: Cancer Incidence in Five Continents Time Trends. 2017. <http://ci5.iarc.fr/CI5plus>
14. National Cancer Center Cancer Control Information Center. Graph database: Center for Cancer Control and Information Services, National Cancer Center Japan. <http://gdb.ganjo.jp>
15. Guo Y, ZX L, Zhang JY, JL M, Zhang L, et al. Association Between Lifestyle Factors, Vitamin and Garlic Supplementation, and Gastric Cancer Outcomes: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Netw Open* 2020;3:e206628.
16. Seemann T. Snippy: fast bacterial variant calling from NGS reads. github.com/tseemann/snippy 2015.
17. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet* 2012;8:e1002453.
18. Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, et al. Chromosome painting in silico in a bacterial species reveals fine population structure. *Mol Biol Evol* 2013;30:1454–1464.
19. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;26:1641–1650.
20. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
21. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* 2014;31:1929–1936.
22. The pymol molecular Graphics system. Schrödinger, LLC. 2000.
23. Matsunari O, Shiota S, Suzuki R, Watada M, Kinjo N, et al. Association between *Helicobacter pylori* virulence factors and gastroduodenal diseases in Okinawa, Japan. *J Clin Microbiol* 2012;50:876–883.
24. Japanese Archipelago Human Population Genetics Consortium, Jinam T, Nishida N, Hirai M, Kawamura S, et al. The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the ryukyuan populations. *J Hum Genet* 2012;57:787–795.
25. van Amsterdam K, Bart A, van der Ende A. A *Helicobacter pylori* TolC efflux pump confers resistance to metronidazole. *Antimicrob Agents Chemother* 2005;49:1477–1482.
26. Trainor EA, Horton KE, Savage PB, Testerman TL, McGee DJ. Role of the HefC efflux pump in *Helicobacter pylori* cholesterol-dependent resistance to ceragenins and bile salts. *Infect Immun* 2011;79:88–97.
27. Kutschke A, de Jonge BL. Compound efflux in *Helicobacter pylori*. *Antimicrob Agents Chemother* 2005;49:3009–3010.
28. Zwama M, Yamasaki S, Nakashima R, Sakurai K, Nishino K, et al. Multiple entry pathways within the efflux transporter AcrB contribute to multidrug recognition. *Nat Commun* 2018;9:124.
29. Fischer F, Robbe-Saule M, Turlin E, Mancuso F, Michel V, et al. Characterization in *Helicobacter pylori* of a nickel transporter essential for colonization that was acquired during evolution by gastric *Helicobacter* species. *PLoS Pathog* 2016;12:e1006018.
30. Maki-Yonekura S, Matsuoka R, Yamashita Y, Shimizu H, Tanaka M, et al. Hexameric and pentameric complexes of the ExbBD energizer in the Ton system. *Elife* 2018;7:e35419.
31. Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, et al. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genet* 2017;13.
32. Hawtin PR, Delves HT, Newell DG. The demonstration of nickel in the urease of *Helicobacter pylori* by atomic absorption spectroscopy. *FEMS Microbiol Lett* 1991;61:51–54.
33. Alm RA, Bina J, Andrews BM, Doig P, Hancock RE, et al. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect Immun* 2000;68:4155–4168.
34. Page WJ, Huyer G, Huyer M, Worobec EA. Characterization of the porins of *Campylobacter jejuni* and *Campylobacter coli* and implications for antibiotic susceptibility. *Antimicrob Agents Chemother* 1989;33:297–303.
35. Bauwens E, Joosten M, Taganna J, Rossi M, Debraekeleer A, et al. In silico proteomic and phylogenetic analysis of the outer membrane protein repertoire of gastric *Helicobacter* species. *Sci Rep* 2018;8:15453.
36. De E, Jullien M, Labesse G, Pages JM, Molle G, et al. MOMP (major outer membrane protein) of *Campylobacter jejuni*; a versatile pore-forming protein. *FEBS Lett* 2000;469:93–97.
37. Kavermann H, Burns BP, Angermüller K, Odenbreit S, Fischer W, et al. Identification and characterization of *Helicobacter pylori* genes essential for gastric colonization. *J Exp Med* 2003;197:813–822.
38. Quandt EM, Gollihar J, Blount ZD, Ellington AD, Georgiou G, et al. Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. *Elife* 2015;4:e09696.
39. Richter C, Mukherjee O, Ermert D, Singh B, YC S, et al. Moonlighting of *Helicobacter pylori* catalase protects against complement-mediated killing by utilising the host molecule vitronectin. *Sci Rep* 2016;6:24391.
40. Jha V, Chelikani P, Carpena X, Fita I, Loewen PC. Influence of main channel structure on H₂O₂ access to the heme cavity of catalase KatE of *Escherichia coli*. *Arch Biochem Biophys* 2012;526:54–59.
41. Kreutzberger MAB, Ewing C, Poly F, Wang F, Egelman EH. Atomic structure of the *Campylobacter jejuni* flagellar filament reveals how epsilon Proteobacteria escaped Toll-like receptor 5 surveillance. *Proc Natl Acad Sci U S A* 2020;117:16985–16991.
42. Schirm M, Soo EC, Aubry AJ, Austin J, Thibault P, et al. Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol Microbiol* 2003;48:1579–1592.
43. Wang F, Burrage AM, Postel S, Clark RE, Orlova A, et al. A structural model of flagellar filament switching across multiple bacterial species. *Nat Commun* 2017;8:960.
44. Perkins A, Tudorica DA, Amieva MR, Remington SJ, Guillemin K. *Helicobacter pylori* senses bleach (HOCl) as a chemoattractant using a cytosolic chemoreceptor. *PLoS Biol* 2019;17:e3000395.
45. Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, et al. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci U S A* 2002;99:14428–14433.
46. Ailloud F, Didelot X, Woltemate S, Pfaffinger G, Overmann J, et al. Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps. *Nat Commun* 2019;10:2273.
47. Noto JM, Lee JY, Gaddy JA, Cover TL, Amieva MR, et al. Regulation of *Helicobacter pylori* Virulence Within the Context of Iron Deficiency. *J Infect Dis* 2015;211:1790–1794.
48. Carrizo-Chavez MA, Cruz-Castaneda A, Olivares-Trejo Jde J. The frpB1 gene of *Helicobacter pylori* is regulated by iron and encodes a membrane protein capable of binding haem and haemoglobin. *FEBS Lett* 2012;586:875–879.
49. Cao Q, Didelot X, Wu Z, Li Z, He L, et al. Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. *Gut* 2015;64:554–561.
50. Bubendorfer S, Krebs J, Yang I, Hage E, Schulz TF, et al. Genome-wide analysis of chromosomal import patterns after natural transformation of *Helicobacter pylori*. *Nat Commun* 2016;7:11995.