HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2017 January 08.

Published in final edited form as:

Science. 2016 January 8; 351(6269): 162–165. doi:10.1126/science.aad2545.

The 5,300-year-old *Helicobacter pylori* genome of the Iceman

Frank Maixner[#1,*], **Ben Krause-Kyora**[#2], **Dmitrij Turaev**[#3], **Alexander Herbig**[#4,5], **Michael R. Hoopmann**[6], **Janice L. Hallows**[6], **Ulrike Kusebauch**[6], **Eduard Egarter Vigl**[7], **Peter Malfertheiner**[8], **Francis Megraud**[9], **Niall O'Sullivan**[1], **Giovanna Cipollini**[1], **Valentina Coia**[1], **Marco Samadelli**[1], **Lars Engstrand**[10], **Bodo Linz**[11], **Robert L. Moritz**[6], **Rudolf Grimm**[12], **Johannes Krause**[4,5,‡], **Almut Nebel**[2,‡], **Yoshan Moodley**[13,14,‡], **Thomas Rattei**[3,‡], and **Albert Zink**[1,*,‡]

[1]Institute for Mummies and the Iceman, EURAC research, Viale Druso 1, 39100 Bolzano, Italy.

[2]Institute of Clinical Molecular Biology, Kiel University, Schittenhelmstr. 12, 24105 Kiel, Germany.

[3]CUBE - Division of Computational Systems Biology, Department of Microbiology and Ecosystem Science, University of Vienna, Althanstr. 14, 1090 Vienna, Austria.

[4]Institute for Archaeological Sciences, University of Tübingen, Rümelinstr. 23, 72072 Tübingen, Germany.

[5]Max Planck Institute for the Science of Human History, Kahlaische Str. 10, 07745 Jena, Germany

[6]Institute for Systems Biology, 401 Terry Avenue North, Seattle, Washington 98109, USA.

[7]Scuola Superiore Sanitaria Provinciale “Claudiana”, Via Lorenz Böhler 13, 39100 Bolzano, Italy.

[8]Department of Gastroenterology, Hepatology, and Infectious Diseases, Otto-von-Guericke University, Leipziger Strasse 44, 39120 Magdeburg, Germany.

[9]Université de Bordeaux, Centre National de Référence des Helicobacters et Campylobacters and INSERM U853, 146 rue Léo Saignat, 33076 Bordeaux, France.

[10]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 141 83 Stockholm, Sweden.

[11]Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

[12]Robert Mondavi Institute for Food Science, University of California, Davis, California 95616, USA

[*]Correspondence to: frank.maixner@eurac.edu, albert.zink@eurac.edu.
[‡]Co-senior authors.

The authors declare no competing interests.

Supplementary Materials:
Materials and Methods
Figures S1-S17
Tables S1-S13
References (*25-93*)

[13]Department of Zoology, University of Venda, Private Bag X5050, Thohoyandou 0950, Republic of South Africa

[14]Department of Integrative Biology and Evolution, Konrad Lorenz Institute for Ethology, University of Veterinary Medicine Vienna, Savoyenstr. 1a, 1160 Vienna, Austria.

[#] These authors contributed equally to this work.

Abstract

The stomach bacterium *Helicobacter pylori* is one of the most prevalent human pathogens. It has dispersed globally with its human host resulting in a distinct phylogeographic pattern that can be used to reconstruct both recent and ancient human migrations. The extant European population of *H. pylori* is known to be a hybrid between Asian and African bacteria, but there exist different hypotheses about when and where the hybridization took place, reflecting the complex demographic history of Europeans. Here, we present a 5,300-year-old *H. pylori* genome from a European Copper Age glacier mummy. The “Iceman” *H. pylori* is a nearly-pure representative of the bacterial population of Asian origin that existed in Europe prior to hybridization, suggesting the African population arrived in Europe within the last few thousand years.

The highly recombinant pathogen *Helicobacter pylori* has uniquely evolved to live in the acidic environment of the human stomach (1). Today this Gram-negative bacterium is found in approximately half the world's human population, but less than 10% of carriers develop disease, typically stomach ulcers or gastric carcinoma (2, 3). Predominant intra-familial transmission of *H. pylori* and the long-term association with humans has resulted in a phylogeographic distribution pattern of *H. pylori* that is shared with its host (4, 5). This observation suggests that the pathogen not only accompanied modern humans out of Africa (6), but that it has also been associated with its host for at least 100,000 years (7). Thus, the bacterium has been used as a marker for tracing complex demographic events in human prehistory (4, 8, 9). Modern *H. pylori* strains have been assigned to distinct populations based on their geographic origin (hpEurope, hpSahul, hpEastAsia, hpAsia2, hpNEAfrica, hpAfrica1 and hpAfrica2) that are derived from at least six ancestral sources (4, 5, 8). The modern *H. pylori* strain found in most Europeans (hpEurope) has putatively originated from recombination of the two ancestral populations Ancestral Europe 1 and 2 (AE1 and AE2) (6). It has been suggested that AE1 originated in Central Asia, where it evolved into hpAsia2, which is commonly found in South Asia. On the other hand, AE2 appears to have evolved in northeast Africa and hybridized with AE1 to become hpEurope (4). However, the precise hybridization zone of the parental populations and the true origin of hpEurope are controversial. Early studies observed a south-to-north cline in AE2/AE1 frequency in Europe (4, 6). This finding has been attributed to independent peopling events that introduced these ancestral *H. pylori* components, which eventually recombined in Europe since the Neolithic period. More recently, it has been suggested that the AE1/AE2 admixture might have occurred in the Middle East or Western Asia between 10,000 and 52,000 years ago and that recombinant strains were introduced into Europe with the first human re-colonizers after the last glacial maximum (7).

In this study, we screened 12 biopsy samples from the gastrointestinal tract of the Iceman, a 5,300-year-old Copper Age mummy, for the presence of *H. pylori*. Stable isotope analyses showed that the Iceman originated and lived in South Europe, in the Eastern Italian Alps (10). Genetically he most closely resembles early European farmers (11-13). The Iceman's stomach was discovered in a re-appraisal of radiological data and contains the food he ingested shortly before his death (14) (Fig. 1). The study material included stomach content, mucosa tissue and content of the small and large intestines (table S1). By using direct PCR, metagenomic diagnostics and targeted genome capture (figs. S1 and S2), we determined the presence of *H. pylori* and reconstructed its complete genome.

Metagenomic analysis yielded endogenous ancient *H. pylori* DNA (15,350 reads) in all gastrointestinal tract contents (Fig. 1 and table S4). A control dataset derived from Iceman's muscle tissue was negative. The distribution of the observed read counts throughout the Iceman's intestinal tract is similar to that in modern *H. pylori*-positive humans, with abundance decreasing from the stomach towards the lower intestinal tract (15, 16). The retrieved unambiguous reads were aligned to a modern *H. pylori* reference genome (strain 26695) and showed damage patterns indicative of ancient DNA (17) (fig. S7). After DNA repair, the *H. pylori* DNA was enriched up to 216-fold using in-solution hybridization capture (Agilent) (fig. S5). From this dataset, 499,245 non-redundant reads mapped to 92.2% of the 1.6 Mb *H. pylori* reference genome with an 18.9-fold average coverage (Fig. 2). In comparison to the reference, the Iceman's ancient *H. pylori* genome had approximately 43,000 single nucleotide polymorphisms (SNPs) and 39 deletions that range from 95bp to 17kb and mainly comprise complete coding regions. Owing to deletions, the number of genomic variants is slightly below the range of what can be observed between modern *H. pylori* strains (table S13). The analysis of SNP allele frequencies does not indicate an infection by more than one strain (see supplementary materials part S6). In addition, as expected for this highly recombinant bacterium, we found evidence for gene insertions from *H. pylori* strains that differ from the reference genome (see supplementary materials part S8 for details about the InDels).

Subsequent sequence analysis classified the ancient *H. pylori* as a *cagA*-positive *vacA* s1a/i1/m1 type strain that is now associated with inflammation of the gastric mucosa (18) (fig. S11). Using multi-step solubilization and fractionation proteomics we identified 115 human proteins in the stomach metaproteome, of which six were either highly expressed in the stomach mucosa (trefoil factor 2) (19) or present in the gastrointestinal tract and involved in digestion (see supplementary materials part S10). The majority of human proteins were enriched in extracellular matrix organizing proteins (P=3.35e-14) and proteins of immune processes (P=2e-3) (fig. S13). In total, 22 proteins observed in the Iceman stomach proteome are primarily expressed in neutrophils and are involved in the inflammatory host response. The two subunits S100A8 and S100A9 of calprotectin (CP) were detected with the highest number of distinct peptide hits in both analyzed samples. Inflamed gastric tissues of modern *H. pylori*-infected patients also show high levels of CP subunit S100A8 and S100A9 expression (20, 21). Thus, the Iceman's stomach was colonized by a cytotoxic *H. pylori* type strain that triggered CP release as a result of host inflammatory immune responses. However, whether the Iceman suffered from gastric disease cannot be

determined from our analysis owing to the poor preservation of the stomach mucosa (fig. S3).

Comparative analysis of seven housekeeping gene fragments with a global multilocus sequence typing (MLST) database of 1,603 *H. pylori* strains with the STRUCTURE (22) no-admixture model assigned the 5,300-year-old bacterium to the modern population hpAsia2, commonly found in Central and South Asia (Fig. 3A and fig. S14). The detection of an hpAsia2 strain in the Iceman's stomach is rather surprising, since despite intensive sampling only three hpAsia2 strains have ever been detected in modern Europeans. Stomachs of modern Europeans are predominantly colonized by recombinant hpEurope strains. Further analysis with the STRUCTURE linkage model (23), used to detect ancestral structure from admixture linkage disequilibrium, revealed that the ancient *H. pylori* strain contained only 6.5% (95% probability intervals [PI]: 1.5%-13.5 %) of the northeast African (AE2) ancestral component of hpEurope (Fig. 3B). Among European strains, this low proportion of AE2 is unique and has thus far only been observed in hpAsia2 strains from India and Southeast Asia. In contrast, the three European hpAsia2 strains (Fig. 3B, black arrows) contained considerably higher AE2 ancestries than the *H. pylori* strain of the Iceman (Finland 13.0%, PI: 5.9-21.7; Estonia 13.2%, PI: 6.2-22.3; and the Netherlands 20.8%, PI: 11.5-31.7), although 95% probability intervals did overlap. A principal component analysis (PCA) of the MLST sequences of the hpAsia2, hpEurope and hpNEAfrica populations revealed a continuum along PC1 that correlates with the proportion of AE2 ancestry versus AE1 ancestry of the isolates (Fig. 3C). The Iceman's ancient *H. pylori* was separated from modern hpEurope strains, and its position along PC1 was close to modern hpAsia2 strains from India, reflecting its almost pure AE1 and very low AE2 ancestry.

Comparative whole-genome analyses (neighbor joining, STRUCTURE and principle component analyses) with publicly available genomes (n=45) confirmed the MLST result by showing that the Iceman's ancient *H. pylori* genome has highest similarity to three hpAsia2 genomes from India (figs. S15-S17). Although the Iceman's *H. pylori* strain appears genetically similar to the extant strains from northern India, slight differences were observed along PC2 in both MLST (Fig. 3C) and genome PCAs (fig. S17), and in the neighbor joining tree (fig. S15). To further study genomic scale introgression, we performed a high-resolution analysis of ancestral motifs using fineSTRUCTURE (24). The resulting linked co-ancestry matrix (Fig. 4) showed that the ancient *H. pylori* genome shares high levels of ancestry with Indian hpAsia2 strains (Fig. 4, green boxes), but even higher co-ancestry with most European hpEurope strains (Fig. 4, blue boxes). In contrast, the Iceman's *H. pylori* shares low ancestry with the hpNEAfrica strain, a modern representative of AE2 (Fig. 4, black box), and with European strains originating from the Iberian Peninsula, where the proportion of AE2 ancestry is relatively high (4) (Fig. 4, white box). Our sample size (n=1) does not allow further conclusions about the prevalence of AE1 in ancient Europe and the course or rate of AE2 introgression. However, the ancient *H. pylori* strain provides the first evidence that AE2 was already present in Central Europe during the Copper Age, albeit at a low level. If the Iceman *H. pylori* strain is representative of its time, the low level of AE2 admixture suggests that most of the AE2 ancestry observed in hpEurope today is a result of AE2 introgression into Europe after the Copper Age, which is later than previously proposed (4,

6). Furthermore, our co-ancestry results indicate that the Iceman's strain belonged to a prehistoric European branch of hpAsia2 that is different to the modern hpAsia2 population from northern India. The high genetic similarity of the ancient strain to bacteria from Europe implies that much of the diversity present in Copper Age Europe is still retained within the extant hpEurope population, despite millennia of subsequent AE2 introgression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge the following funding sources: the South Tyrolean grant Legge 14 (F.M, N.OS, G.C, V.C, M.S, A.Z), the Ernst Ludwig Ehrlich Studienwerk, dissertation completion fellowship of the University of Vienna (D.T), the Graduate School Human Development in Landscapes and the Excellence Cluster Inflammation at Interfaces (B.K, A.N), the European Research Council (ERC) starting grant APGREID (J.K and A.H), the National Institutes of Health from the National Institute of General Medical Sciences under Grant Nos. R01 GM087221 (R.M), S10 RR027584 (R.M) and the 2P50 GM076547/Center for Systems Biology (R.M). Emily Leproust and Owen Hardy are highly acknowledged for their help in the RNA bait design. We thank the sequencing team of the Institute of Clinical Molecular Biology at Kiel University, for support and expertise. We are grateful to Edna Hütten for proofreading of the main text. We are grateful to Olympus, Italy, for providing us with equipment for endoscopy. F.M and A.Z conceived the investigation. F.M, B.K, D.T, R.G, J.K, A.N, Y.M, T.R, and A.Z designed experiments. P.M, L.E, E.EV, M.S, F.M and A.Z were involved in the sampling campaign. F.M, B.K, MR.H, J.H, U.K, G.C performed laboratory work. F.M, B.K, D.T, A.H, MR.H, N.OS, B.L, RL.M, R.G, J.K, Y.M and T.R performed analyses. F.M wrote the manuscript with contributions from B.K, D.T, A.H, MR.H, U.K, N.OS, V.C, B.L, RL.M, R.G, J.K, Y.M, A.N, T.R and A.Z. Data are available from the European Nucleotide Archive under accession no. ERP012908.

References and Notes

1. Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. Nat. Rev. Microbiol. 2007; 5:441–452. [PubMed: 17505524]
2. Malfertheiner P, Chan FK, McColl KE. Peptic ulcer disease. Lancet. 2009; 374:1449–1461. [PubMed: 19683340]
3. Peek RM, Blaser MJ. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. Nature reviews. Cancer. 2002; 2:28–37. [PubMed: 11902583]
4. Falush D, et al. Traces of human migrations in *Helicobacter pylori* populations. Science. 2003; 299:1582–1585. [PubMed: 12624269]
5. Moodley Y, Linz B. *Helicobacter pylori* Sequences Reflect Past Human Migrations. Genome Dynamics. 2009; 6:62–74. [PubMed: 19696494]
6. Linz B, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 2007; 445:915–918. [PubMed: 17287725]
7. Moodley Y, et al. Age of the association between *Helicobacter pylori* and man. PLoS pathogens. 2012; 8:e1002693. [PubMed: 22589724]
8. Moodley Y, et al. The peopling of the Pacific from a bacterial perspective. Science. 2009; 323:527–530. [PubMed: 19164753]
9. Wirth T, et al. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: Lessons from Ladakh. Proc. Natl. Acad. Sci. U.S.A. 2004; 101:4746–4751. [PubMed: 15051885]
10. Müller W, Fricke H, Halliday AN, McCulloch MT, Wartho JA. Origin and migration of the Alpine Iceman. Science. 2003; 302:862–866. [PubMed: 14593178]
11. Haak W, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015; 522:207–211. [PubMed: 25731166]
12. Keller A, et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun. 2012; 3:698. [PubMed: 22426219]

13. Lazaridis I, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014; 513:409–413. [PubMed: 25230663]
14. Gostner P, Pernter P, Bonattie G, Graefen A, Zink AR. New radiological insights into the life and death of the Tyrolean Iceman. J. Arch. Science. 2011; 38:3425–3431.
15. Andersson AF, et al. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One. 2008; 3:e2836. [PubMed: 18665274]
16. Segata N, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biology. 2012; 13:R42. doi: 10.1186/gb-2012-13-6-r42. [PubMed: 22698087]
17. Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One. 2012; 7:e34131. [PubMed: 22479540]
18. Jones KR, Whitmire JM, Merrell DS. A Tale of Two Toxins: *Helicobacter pylori* CagA and VacA Modulate Host Pathways that Impact Disease. Frontiers in microbiology. 2010; 1:115. [PubMed: 21687723]
19. Uhlén M, et al. Tissue-based map of the human proteome. Science. 2015; 347:1260419. [PubMed: 25613900]
20. Gaddy JA, et al. The host protein calprotectin modulates the *Helicobacter pylori* cag type IV secretion system via zinc sequestration. PLoS pathogens. 2014; 10:e1004450. [PubMed: 25330071]
21. Leach ST, Mitchell HM, Geczy CL, Sherman PM, Day AS. S100 calgranulin proteins S100A8, S100A9 and S100A12 are expressed in the inflamed gastric mucosa of *Helicobacter pylori*-infected children. Canadian journal of gastroenterology = Journal canadien de gastroenterologie. 2008; 22:461–464. [PubMed: 18478131]
22. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. [PubMed: 10835412]
23. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003; 164:1567–1587. [PubMed: 12930761]
24. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. PLoS genetics. 2012; 8:e1002453. [PubMed: 22291602]
25. Spindler K. The man in the ice (Wiedenfeld and Nicoldon, London. 1994
26. Pernter P, Gostner P, Egarter Vigl E, Rühli FJ. Radiologic proof for the Iceman's cause of death (ca. 5'300 BP). Journal of Archaeological Sciences. 2007; 34:1784–1786.
27. Lynnerup N. Mummies. Am J Phys Anthropol. 2007; 45:162–190. [PubMed: 18046750]
28. Tang JN, et al. An effective method for isolation of DNA from pig faeces and comparison of five different methods. J Microbiol Methods. 2008; 75:432–436. [PubMed: 18700153]
29. Lee EJ, et al. Emerging genetic patterns of the European Neolithic: perspectives from a late Neolithic Bell Beaker burial site in Germany. Am J Phys Anthropol. 2012; 148:571–579. [PubMed: 22552938]
30. Atherton JC, et al. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific vacA types with cytotoxin production and peptic ulceration. The Journal of biological chemistry. 1995; 270:17771–17777. [PubMed: 7629077]
31. Swanston T, Haakensen M, Deneer H, Walker EG. The characterization of *Helicobacter pylori* DNA associated with ancient human remains recovered from a Canadian glacier. PLoS One. 2011; 6:e16864. [PubMed: 21359221]
32. van Doorn LJ, et al. Typing of *Helicobacter pylori vacA* gene and detection of cagA gene by PCR and reverse hybridization. J Clin Microbiol. 1998; 36:1271–1276. [PubMed: 9574690]
33. Ryberg A, Borch K, Sun YQ, Monstein HJ. Concurrent genotyping of *Helicobacter pylori* virulence genes and human cytokine SNP sites using whole genome amplified DNA derived from minute amounts of gastric biopsy specimen DNA. BMC Microbiol. 2008; 8:175. [PubMed: 18842150]

34. Mekota AM, Vermehren M. Determination of optimal rehydration, fixation and staining methods for histological and immunohistochemical analysis of mummified soft tissues. Biotech Histochem. 2005; 80:7–13. [PubMed: 15804821]
35. Mulisch, M.; Welsch, U. Romeis - Mikroskopische Techniken. Spektrum Akademischer Verlag; Heidelberg: 2010.
36. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012; 40:e3. [PubMed: 22021376]
37. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010, pdb prot5448. 2010
38. Lara-Ramirez EE, et al. New implications on genomic adaptation derived from the *Helicobacter pylori* genome comparison. PLoS One. 2011; 6:e17300. [PubMed: 21387011]
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410. [PubMed: 2231712]
40. Quast C, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012; 41:D590–596. [PubMed: 23193283]
41. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 2011; 21:1552–1560. [PubMed: 21690186]
42. Ye Y, Choi JH, Tang H. RAPSearch: a fast protein similarity search tool for short reads. BMC Bioinformatics. 2011; 12:159. [PubMed: 21575167]
43. Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics. 2011; 27:2153–2155. [PubMed: 21659319]
44. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
45. Koboldt DC, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22:568–576. [PubMed: 22300766]
46. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
47. Wickham, H. ggplot2: elegant graphics for data analysis. Springer; New York: 2009.
48. Krzywinski M, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645. [PubMed: 19541911]
49. Schwarz S, et al. Horizontal versus familial transmission of *Helicobacter pylori*. PLoS pathogens. 2008; 4:e1000180. [PubMed: 18949030]
50. Jónsson H, Ginolhac A, Schubert M, Johnson P, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013; 29:1682–1684. [PubMed: 23613487]
51. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. [PubMed: 20080505]
52. Skoglund P, Storå J, Götherström A, Jakobsson M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. J. Archaeol. Sci. 2013; 40:4477–4482.
53. Kloss-Brandstatter A, et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat. 2011; 32:25–32. [PubMed: 20960467]
54. Ermini L, et al. Complete mitochondrial genome sequence of the Tyrolean Iceman. Current Biology. 2008; 18:1687–1693. [PubMed: 18976917]
55. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012; 28:464–469. [PubMed: 22199388]
56. Tewhey R, et al. Enrichment of sequencing targets from the human genome by solution hybridization. Genome Biology. 2009; 10:R116. [PubMed: 19835619]
57. Olbermann P, et al. A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island. PLoS genetics. 2010; 6:e1001069. [PubMed: 20808891]
58. Yamaoka Y. Mechanisms of disease: *Helicobacter pylori* virulence factors. Nature reviews. Gastroenterology & hepatology. 2010; 7:629–641. [PubMed: 20938460]

59. Oliveira PH, Touchon M, Rocha EP. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. 2014; 42:10618–10631. [PubMed: 25120263]
60. Marais A, Mendz GL, Hazell SL, Megraud F. Metabolism and genetics of *Helicobacter pylori*: the genome era. Microbiology and molecular biology reviews: MMBR. 1999; 63:642–674. [PubMed: 10477311]
61. Kawai M, et al. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. BMC Microbiol. 2011; 11:104. [PubMed: 21575176]
62. Wolfe AJ. The acetate switch. Microbiology and molecular biology reviews: MMBR. 2005; 69:12–50. [PubMed: 15755952]
63. Bourzac KM, Guillemin K. *Helicobacter pylori*-host cell interactions mediated by type IV secretion. Cellular microbiology. 2005; 7:911–919. [PubMed: 15953024]
64. Censini S, et al. cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci U S A. 1996; 93:14648–14653. [PubMed: 8962108]
65. Duncan SS, Valk PL, Shaffer CL, Bordenstein SR, Cover TL. J-Western forms of *Helicobacter pylori cagA* constitute a distinct phylogenetic group with a widespread geographic distribution. J Bacteriol. 2012; 194:1593–1604. [PubMed: 22247512]
66. Gangwer KA, et al. Molecular evolution of the *Helicobacter pylori* vacuolating toxin gene vacA. J Bacteriol. 2010; 192:6126–6135. [PubMed: 20870762]
67. Ludwig W, et al. ARB: a software environment for sequence data. Nucleic Acids Res. 2004; 32:1363–1371. [PubMed: 14985472]
68. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673–4680. [PubMed: 7984417]
69. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52:696–704. [PubMed: 14530136]
70. Blaser MJ, et al. Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer research. 1995; 55:2111–2115. [PubMed: 7743510]
71. Huang JQ, Zheng GF, Sumanac K, Irvine EJ, Hunt RH. Meta-analysis of the relationship between cagA seropositivity and gastric cancer. Gastroenterology. 2003; 125:1636–1644. [PubMed: 14724815]
72. Hatakeyama M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. Nature reviews. Cancer. 2004; 4:688–694. [PubMed: 15343275]
73. Rhead JL, et al. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. Gastroenterology. 2007; 133:926–936. [PubMed: 17854597]
74. Horth P, Miller CA, Preckel T, Wenz C. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. Mol Cell Proteomics. 2006; 5:1968–1974. [PubMed: 16849286]
75. Ros A, et al. Protein purification by Off-Gel electrophoresis. Proteomics. 2002; 2:151–156. [PubMed: 11840561]
76. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics. 2008; 24:2534–2536. [PubMed: 18606607]
77. Martens L, et al. mzML--a community standard for mass spectrometry data. Mol Cell Proteomics. 2011; 10:R110 000133. [PubMed: 20716697]
78. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004; 20:1466–1467. [PubMed: 14976030]
79. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. Proteomics. 2013; 13:22–24. [PubMed: 23148064]
80. Deutsch EW, et al. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics Clin Appl. 2015; 9:745–754. [PubMed: 25631240]

81. Shteynberg D, Nesvizhskii AI, Moritz RL, Deutsch EW. Combining results of multiple search engines in proteomics. Mol Cell Proteomics. 2013; 12:2383–2393. [PubMed: 23720762]
82. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003; 75:4646–4658. [PubMed: 14632076]
83. Desiere F, et al. The PeptideAtlas project. Nucleic Acids Res. 2006; 34:D655–658. [PubMed: 16381952]
84. Kusebauch U, et al. Using PeptideAtlas, SRMAtlas, and PASSEL: Comprehensive Resources for Discovery and Targeted Proteomics. Curr Protoc Bioinformatics. 2014; 46:13, 25, 11–13, 25, 28. [PubMed: 24939129]
85. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 2000; 28:3442–3444. [PubMed: 10982861]
86. Rollo F, Ubaldi M, Ermini L, Marota I. Otzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. Proc Natl Acad Sci U S A. 2002; 99:12594–12599. [PubMed: 12244211]
87. Warinner C, et al. Pathogens and host immunity in the ancient human oral cavity. Nat Genet. 2014; 46:336–344. [PubMed: 24562188]
88. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 2015
89. Schuenemann VJ, et al. Genome-wide comparison of medieval and modern Mycobacterium leprae. Science. 2013; 341:179–183. [PubMed: 23765279]
90. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498. [PubMed: 21478889]
91. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30:2725–2729. [PubMed: 24132122]
92. Yahara K, et al. Chromosome painting in silico in a bacterial species reveals fine population structure. Mol Biol Evol. 2013; 30:1454–1464. [PubMed: 23505045]
93. Grant JR, Arantes AS, Stothard P. Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics. 2012; 13:202. [PubMed: 22621371]

One Sentence Summary

The analysis of a 5,300-year-old *H. pylori* genome from the Iceman provides novel insights into the evolutionary history of *H. pylori* populations.

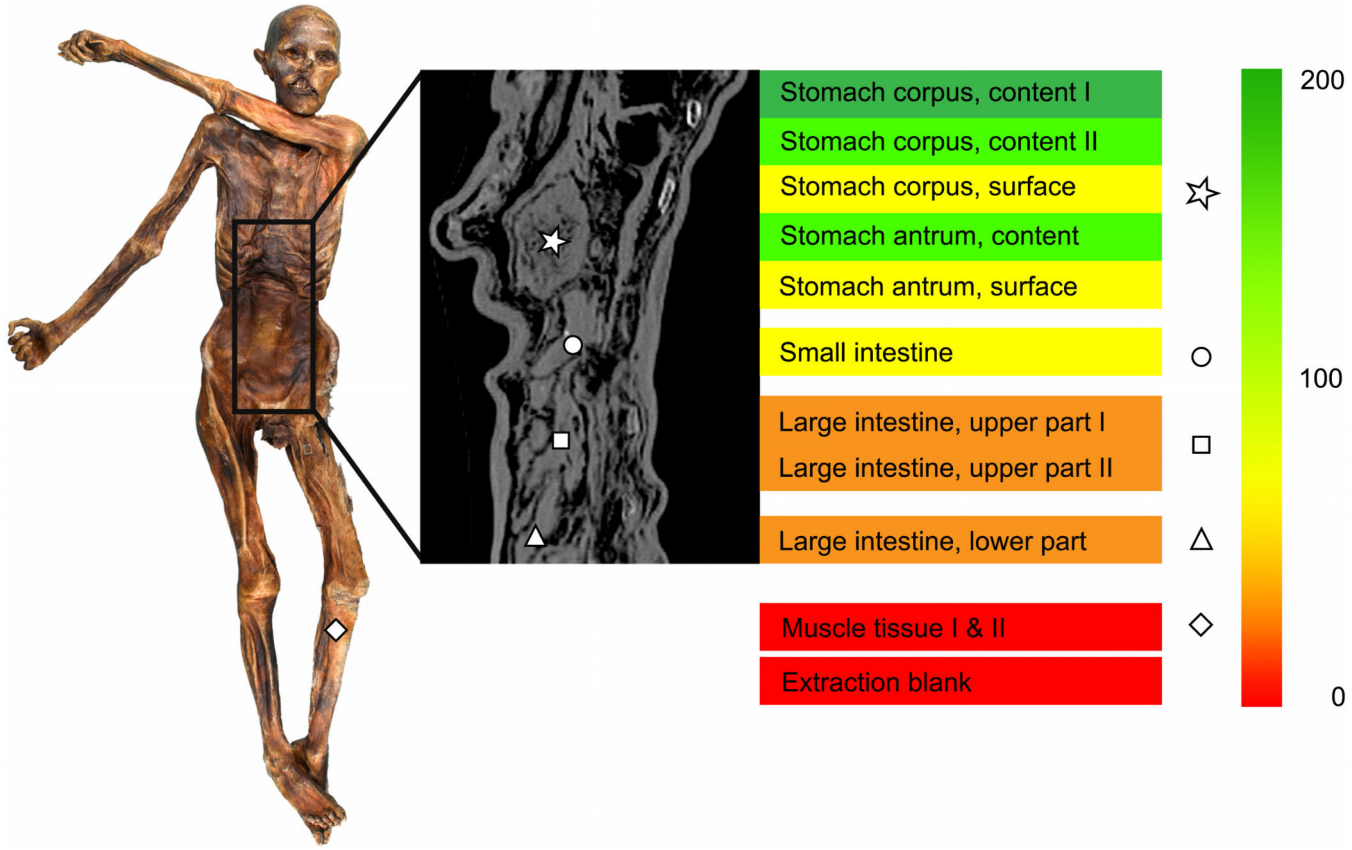

Fig. 1.

H. pylori specific reads detected in the metagenomic datasets of the Iceman's intestine content samples. The color gradient displays the number of unambiguous *H. pylori* reads per million metagenomic reads. Control metagenomic datasets of the Iceman's muscle tissue and of the extraction blank were included in the analysis. The different intestinal content sampling sites are marked in the radiographic image by the following symbols: * stomach content, ○ small intestine, □ upper large intestine, Δ lower large intestine. The sampling site of the muscle control sample is highlighted in the Iceman overview picture (◇).

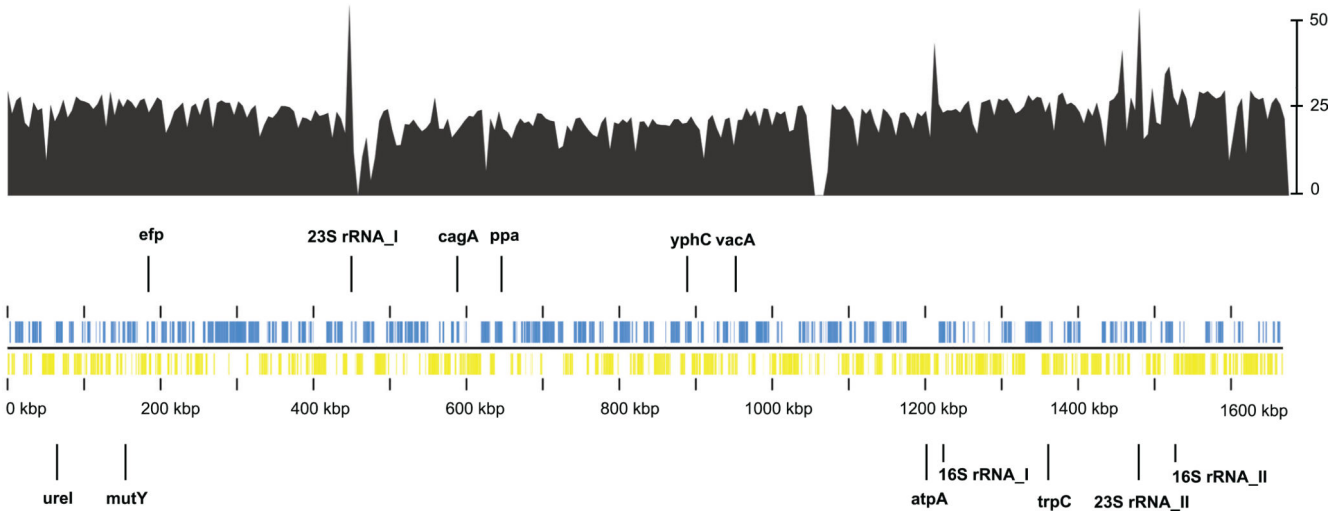

Fig. 2.

Gene coverage and distribution of the enriched and validated Iceman *H. pylori* reads mapped onto the 1.6 Mb large reference genome *H. pylori* 26695. The coverage plot displayed in black is superimposed onto the genomic plot. The bar on the right-hand side indicates a coverage of up to 50x. The gene coding sequences are shown in blue (positive strand) and yellow (negative strand) bars in the genomic plot. The loci of the ribosomal RNA genes, of two virulence genes (*vacA* and *cagA*) and of seven genes used for MLST analysis are highlighted in the genome plot.

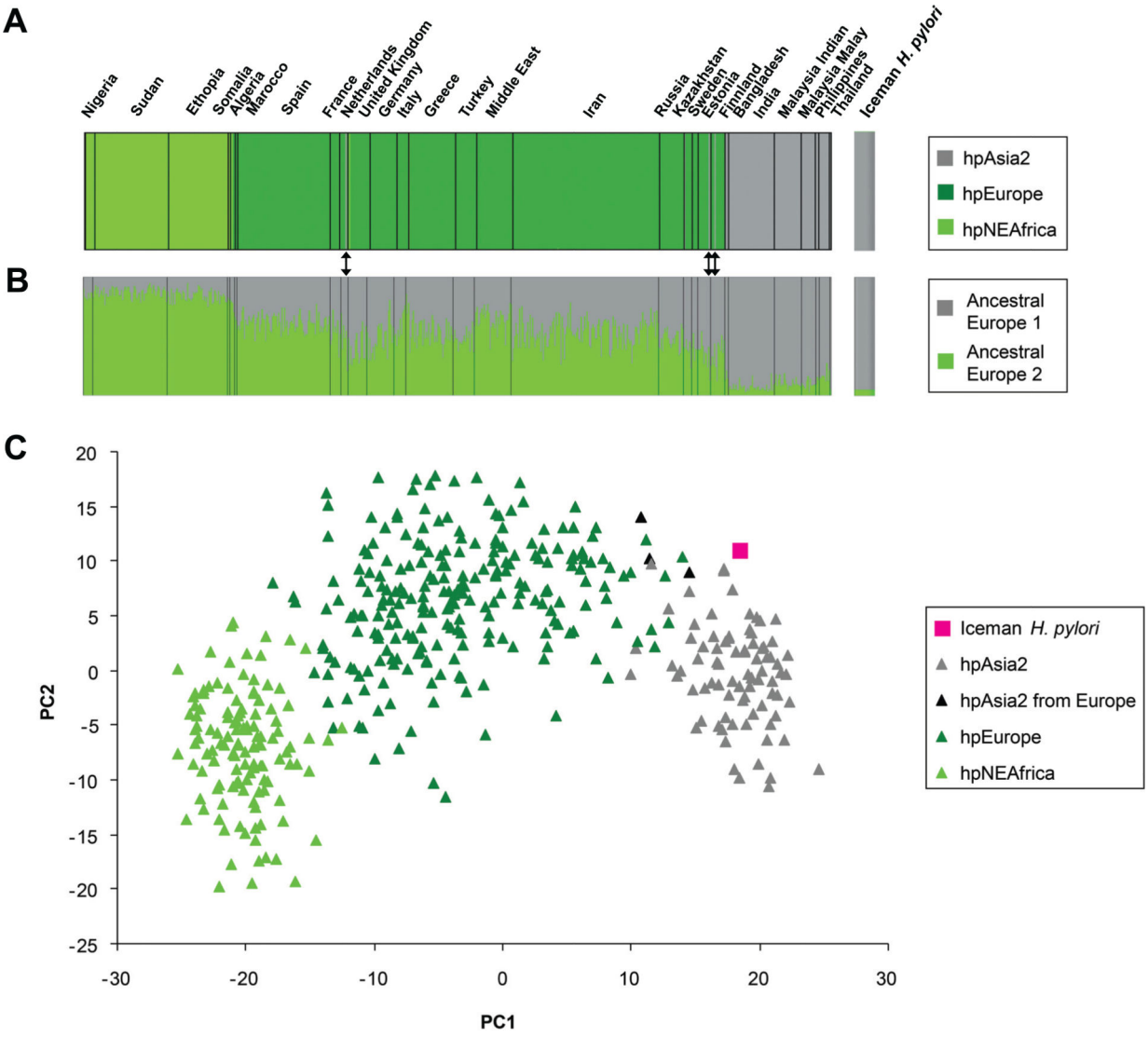

Fig. 3.

Multilocus sequence analyses (**A**) Bayesian cluster analysis performed in STRUCTURE displays the population partitioning of hpEurope, hpAsia2 and hpNEAfrica and the Iceman's *H. pylori* strain (for details about the worldwide population partitioning of 1,603 reference *H. pylori* strains please refer to fig. S14). (**B**) STRUCTURE linkage model analysis showing the proportion of Ancestral Europe 1 (from Central Asia) and Ancestral Europe 2 (from northeast Africa) nucleotides among strains assigned to populations hpNEAfrica, hpEurope and hpAsia2 and the Iceman's *H. pylori* strain on the extreme right. The black arrows indicate the position of the three extant European hpAsia2 strains (**C**) Principal component analysis of contemporary hpNEAfrica, hpEurope and hpAsia2 strains and the Iceman's *H. pylori* strain.

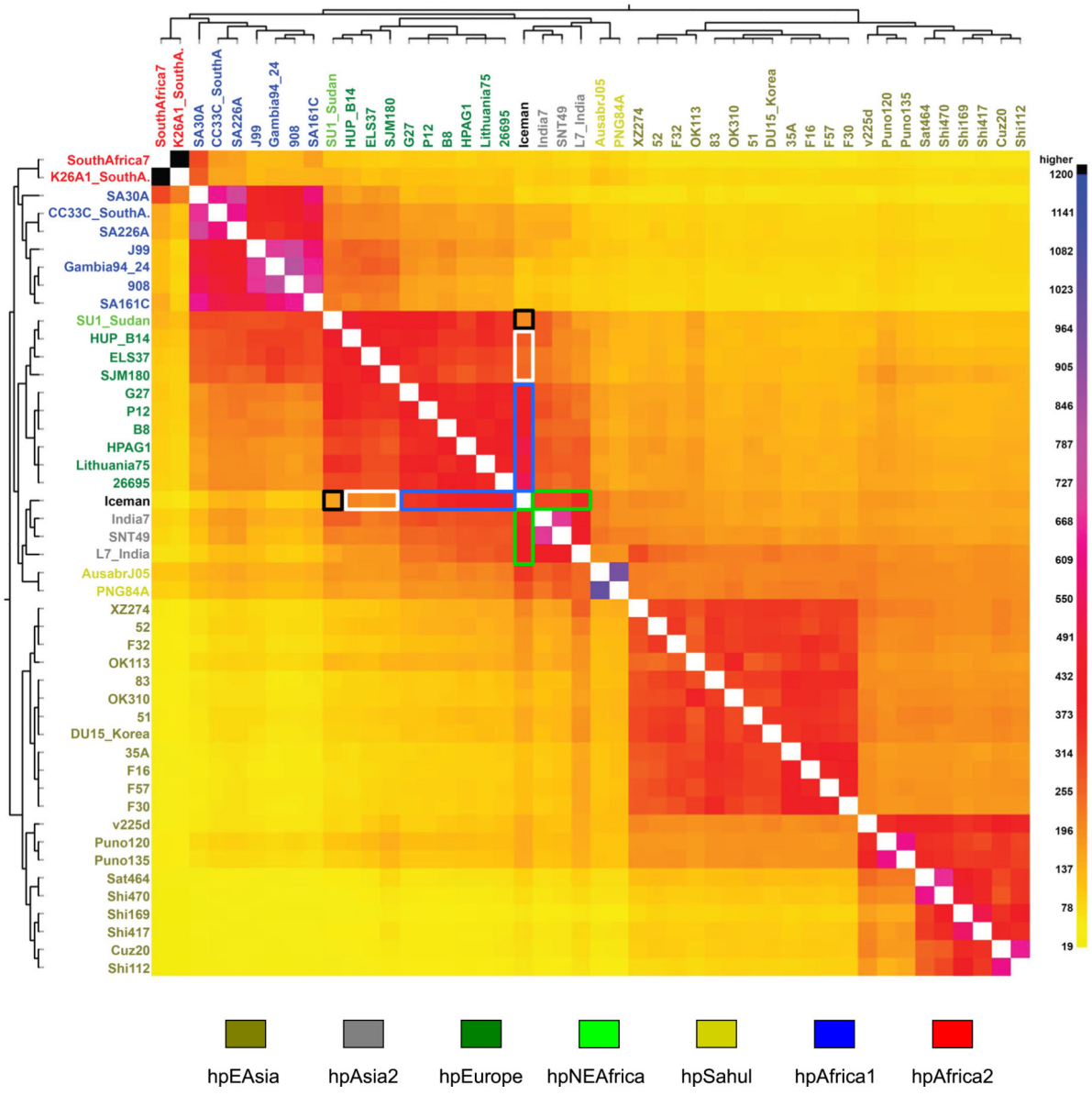

Fig. 4.

Comparative whole-genome analysis. Co-ancestry matrix showing *H. pylori* population structure and genetic flux. The color in the heat map corresponds to the number of genomic motifs imported from a donor genome (column) to a recipient genome (row). The inferred tree and the *H. pylori* strain names are displayed on the top and left of the heat map. Strain names are colored according to the *H. pylori* population assignment provided in the legend below the heat map. Signs for population ancestry are highlighted in the heat map with green, blue, black and white boxes.