The 5300-year-old *Helicobacter pylori* genome of the Iceman

Frank Maixner,1,2,*,† Ben Krause-Kyora,2† Dmitrij Turavev,3 Alexander Herbig,4,5,† Michael R. Hoopmann,4 Janice L. Hallows,6 Ulrike Kusebauch,6 Eduard Egarter Vigl,7 Peter Malfertheiner,8 Francis Megraud,9 Niall O’Sullivan,10 Giovanna Cipollini,11 Valentina Coia,1 Marco Samadelli,1 Lars Engstrand,10 Bodo Linz,11 Robert L. Moritz,6 Rudolf Grimm,12 Johannes Krause,14,22 Almut Nebel,12 Yoshan Moodley,11,14,‡ Thomas Rattel,14,† Albert Zink12

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 5300-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 before hybridization, suggesting that the African population arrived in Europe within the past few thousand years.

**REFERENCES AND NOTES**


**ACKNOWLEDGMENTS**

The data reported in this paper are available in the Qita database (http://qita.ucsd.edu) (accession numbers K0141 to 10431 and 10321) and the European Bioinformatics Institute European Nucleotide Archive (www.ebi.ac.uk/ena) (accession numbers EERP02286, EERP02879, EERP02880, and EERP02984). We thank the donors and their families for their contribution to scientific research; the STAFS Facility at SHSU and the Molecular, Cellular, and Developmental Biology Transgenic Facility at the University of Colorado, Boulder, for providing the space and opportunity for this research; N. Fierer, J. Žebkova, and J. Leff for assistance with project logistics and data processing; and the Mountain Research Station and Shortgrass Steppe Long Term Ecological Research for permission to collect soils. Mice were euthanized humanely under approved protocol no. 08-04-ACK-01 (principal investigator G.A.). This research was funded by the Office of Justice Programs National Institute of Justice Awards NU211-DN-BX-K533 (J.L.M., D.O.C., R.K.) and NU2-2012-DN-BX-K023 (S.R.B. and A.M.L.).

**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/351/6269/158/suppl/DC1

Materials and Methods
Supplementary Text
Figs. S1 to S19
Tables S1 to S20
References (16–29)

19 August 2015; accepted 25 November 2015
Published online 10 December 2015
10.1126/science.aad2646

**ANCIENT MICROBIOME**
In this study, we screened 12 biopsy samples from the gastrointestinal tract of the Iceman, a 5300-year-old Copper Age mummy, for the presence of *H. pylori*. Stable isotope analyses showed that the Iceman originated and lived in Southern 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 reappraisal of radiological data and contains the food he ingested shortly before his death (Fig. 1) (14). The study material included stomach content, mucosa tissue, and content of the small and large intestines (table S1). By using direct polymerase chain reaction (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 data set 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 toward 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 (fig. S7) (17). After DNA repair, the *H. pylori* DNA was enriched up to 216-fold by using in-solution hybridization capture (Agilent) (fig. S5). From this data set, 499,245 nonredundant 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 with the reference, the Iceman’s ancient *H. pylori* genome had ~43,000 single-nucleotide polymorphisms (SNPs) and 39 deletions that range from 95 base pair (bp) to 17 kb 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 (supplementary materials 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 (details about the InDels are provided in supplementary materials S8).

**Fig. 1.** *H. pylori*–specific reads detected in the metagenomic data sets of the Iceman’s intestine content samples. The color gradient displays the number of unambiguous *H. pylori* reads per million metagenomic reads. Control metagenomic data sets 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: asterisk, stomach content; circle, small intestine; square, upper large intestine; triangle, lower large intestine. The sampling site of the muscle control sample is highlighted in the Iceman overview picture (diamond).

**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 50×. 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.
Subsequent sequence analysis classified the ancient \textit{H. pylori} as a \textit{cagA}-positive \textit{vacA}s1a/i1/m1 type strain that is now associated with inflammation of the gastric mucosa (fig. S11) (18). Using multistep 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 (supplementary materials S10). The majority of human proteins were enriched in extracellular matrix organizing proteins ($P = 3.35 \times 10^{-14}$) and proteins of immune processes ($P = 2 \times 10^{-3}$) (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 \textit{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 \textit{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 1603 \textit{H. pylori} strains with the STRUCTURE (22) no-admixture model assigned the 5300-year-old bacterium to the modern population hpAsia2, commonly found in Central

---

**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 \textit{H. pylori} strain (details about the worldwide population partitioning of 1603 reference \textit{H. pylori} strains are available in 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 \textit{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 \textit{H. pylori} strain.

---

**Fig. 4. Comparative whole-genome analysis.** Co-ancestry matrix showing \textit{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 \textit{H. pylori} strain names are displayed on the top and left of the heat map. Strain names are colored according to the \textit{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.
and South Asia (Fig. 3A and fig. S14). The detection of an hpAsia2 strain in the Iceman's stomach is rather surprising because 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 to 13.5%) of the northeast African (AE2) ancestral component of hpEurope (Fig. 3B). Among European strains, this low proportion of AE2 is distinct and has thus far only been observed in hpAsia2 strains from India and Southwest Asia. In contrast, the three European hpAsia2 strains (Fig. 3B, black arrows) contained considerably higher AE2 ancestries than that of the H. pylori strain of the Iceman (Finland 13.0%, PI 5.9 to 21.7; Estonia 13.2%, PI 6.2 to 22.3; and the Netherlands 20.8%, PI 11.5 to 31.7), although 95% probability intervals did overlap. A principal component analysis (PCA) of the MLST sequences of the hpEurope, hpAsia2, 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 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 to 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 (Fig. 4, white box) (4). 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 from 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.

REFERENCES AND NOTES
7. Y. Moodley et al., PLOS Pathog. 8, e1002693 (2012).
ACKNOWLEDGMENTS
This study was supported by the following funding sources: the South Tyrolean grant legge 14 (F.M., N.O.S., G.C., V.C., M.S., and 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. and A.N.), the European Research Council (ERC) starting grant APGRDEED (J.K. and A.H.), the National Institutes of Health from the National Institute of General Medical Sciences under grants R01 GM087221 (R.M.). SID 90027584 (R.M.), and ZEP50 GM076547/Center for Systems Biology (R.M.). E. Leoprust and O. 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 E. 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.E.V., M.S., F.M., and A.Z. were involved in the sampling campaign. F.M., B.K., M.R.H., J.H., U.K., and G.C. performed laboratory work. F.M., B.K., D.T., A.H., M.R., N.O.S., B.L., R.L.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., M.R.H., U.K., N.O.S., V.C., B.L., R.L.M., R.G., J.K., A.N., T.R., and A.Z. Data are available from the European Nucleotide Archive under accession no. ERP012008. The authors declare no competing interests.
SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/351/6269/162/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S13
References (25–93)
15 August 2015; accepted 20 November 2015
10.1126/science.aad2545

PALEOClimATE
Reconciliation of the Devils Hole climate record with orbital forcing
Gina E. Moseley,1,2 R. Lawrence Edwards,2 Kathleen A. Wendt,1,2 Hai Cheng,2,3 Yuri Dublyansky,1 Yanbin Lu,2 Ronny Boch,1* Christoph Spötl1*

The driving force behind Quaternary glacial-interglacial cycles and much associated climate change is widely considered to be orbital forcing. However, previous versions of the iconic Devils Hole (Nevada) subaqueous calcite record exhibit shifts to interglacial values ~10,000 years before orbitally forced ice age terminations, and interglacial durations ~10,000 years longer than other estimates. Our measurements from Devils Hole 2 replicate virtually all aspects of the past 204,000 years of earlier records, except for the timing during terminations, and they lower the age of the record near Termination II by ~8000 years, removing both ~10,000-year anomalies. The shift to interglacial values now broadly coincides with the rise in boreal summer insolation, the marine termination, and the rise in atmospheric CO₂, which is consistent with mechanisms ultimately tied to orbital forcing.

Changes to Earth's orbital configuration relative to the Sun, known as the Milankovitch hypothesis, astronomical theory, or orbital forcing, have long been considered the leading theory for the primary mechanism driving Quaternary glacial-interglacial cycles (1–3) and associated climate change. The hypothesis is supported by a huge array of evidence from paleoclimate records across the globe, which show that major shifts in