

Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process

Felix M. Key^{1,2,3}✉, Cosimo Posth¹, Luis R. Esquivel-Gomez⁴, Ron Hübler¹, Maria A. Spyrou¹, Gunnar U. Neumann¹, Anja Furtwängler⁵, Susanna Sabin¹, Marta Burri¹, Antje Wissgott¹, Aditya Kumar Lankapalli¹, Åshild J. Vågene¹, Matthias Meyer⁶, Sarah Nagel⁶, Rezeda Tukhbatova^{1,7}, Aleksandr Khokhlov⁸, Andrey Chizhevsky⁹, Svend Hansen¹⁰, Andrey B. Belinsky¹¹, Alexey Kalmykov¹¹, Anatoly R. Kantorovich¹², Vladimir E. Maslov¹³, Philipp W. Stockhammer^{1,14}, Stefania Vai¹¹⁵, Monica Zavattaro¹⁶, Alessandro Riga¹¹⁵, David Caramelli¹¹⁵, Robin Skeates¹⁷, Jessica Beckett¹⁷, Maria Giuseppina Gradoli¹⁸, Noah Steuri¹⁹, Albert Hafner¹⁹, Marianne Ramstein²⁰, Inga Siebke²¹, Sandra Lösch¹²¹, Yilmaz Selim Erdal²², Nabil-Fareed Alikhan²³, Zhemin Zhou²³, Mark Achtman¹²³, Kirsten Bos¹, Sabine Reinhold¹⁰, Wolfgang Haak¹, Denise Kühnert¹⁴, Alexander Herbig¹^{1*} and Johannes Krause¹^{1*}

It has been hypothesized that the Neolithic transition towards an agricultural and pastoralist economy facilitated the emergence of human-adapted pathogens. Here, we recovered eight *Salmonella enterica* subsp. *enterica* genomes from human skeletons of transitional foragers, pastoralists and agropastoralists in western Eurasia that were up to 6,500 yr old. Despite the high genetic diversity of *S. enterica*, all ancient bacterial genomes clustered in a single previously uncharacterized branch that contains *S. enterica* adapted to multiple mammalian species. All ancient bacterial genomes from prehistoric (agro-)pastoralists fall within a part of this branch that also includes the human-specific *S. enterica* Paratyphi C, illustrating the evolution of a human pathogen over a period of 5,000 yr. Bacterial genomic comparisons suggest that the earlier ancient strains were not host specific, differed in pathogenic potential and experienced convergent pseudogenization that accompanied their downstream host adaptation. These observations support the concept that the emergence of human-adapted *S. enterica* is linked to human cultural transformations.

The transition in human lifestyle from foraging to agriculture and pastoralism during the Neolithic revolution represents possibly the biggest cultural change in human history. The transition began in the Near Eastern fertile crescent around 10,000 yr ago and subsequently Neolithic economies spread across western Eurasia¹. This Neolithization process was accompanied by a change in mobility, introduction of domesticated animals and increased contact with animal and human excrements, which facilitated a more constant and recurrent exposure to pathogens and possibly the emergence of zoonotic disease^{2–5}. However, direct

molecular evidence—such as ancient DNA—in support of this hypothesis is currently missing. Microbial palaeogenomics provides a unique window into the past human infectious disease burden, and promises to elucidate the deep evolutionary history of clinically relevant pathogens^{6,7}. Recently, ancient DNA analysis identified the human-specific bacterial pathogen *Salmonella enterica* Paratyphi C in 450-yr-old skeletons from Mexico and 800-yr-old skeleton from Norway^{8,9}. However, *S. enterica* is largely absent from the historical record because it has non-specific clinical symptoms and does not cause distinctive skeletal lesions.

¹Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany. ²Institute for Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ³Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Transmission, Infection, Diversification & Evolution Group, Max Planck Institute for the Science of Human History, Jena, Germany. ⁵Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tuebingen, Tuebingen, Germany. ⁶Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ⁷Laboratory of Structural Biology, Kazan Federal University, Kazan, Russian Federation. ⁸Samara State University of Social Sciences and Education, Samara, Russian Federation. ⁹Institute of Archaeology named after A.Kh. Khalikov of the Academy of Sciences of the Republic of Tatarstan, Kazan, Russian Federation. ¹⁰Eurasia Department, German Archaeological Institute, Berlin, Germany. ¹¹Nasledie' Cultural Heritage Unit, Stavropol, Russian Federation. ¹²Department of Archaeology, Faculty of History, Lomonosov Moscow State University, Moscow, Russian Federation. ¹³Institute of Archaeology RAS, Moscow, Russian Federation. ¹⁴Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig Maximilian University Munich, Munich, Germany. ¹⁵Department of Biology, University of Florence, Florence, Italy. ¹⁶Museum of Anthropology and Ethnology, Museum System of the University of Florence, Florence, Italy. ¹⁷Department of Archaeology, Durham University, Durham, UK. ¹⁸School of Archaeology and Ancient History, Leicester University, Leicester, UK. ¹⁹Institute of Archaeological Sciences and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland. ²⁰Archaeological Service of the Canton of Bern, Bern, Switzerland. ²¹Department of Physical Anthropology Institute of Forensic Medicine, University of Bern, Bern, Switzerland. ²²Department of Anthropology, Hacettepe University, Ankara, Turkey. ²³Warwick Medical School, University of Warwick, Coventry, UK. *e-mail: fkey@mit.edu; herbig@shh.mpg.de; krause@shh.mpg.de

Table 1 | Overview of *S. enterica*-positive samples

Sample	Site	Country	Date (cal yr BP)	Mapped reads	Endog. DNA (%)	Coverage	Reference covered (%)
MUR009	Murzikhinsky II	Russia	6,500–6,350	783,054	2.8	8.8	88.9
MUR019	Murzikhinsky II	Russia	6,490–6,320	1,405,983	7.8	16.5	90.1
IV3002	Ipatovo 3	Russia	5,580–5,080	761,643	8.9	7.0	88.9
OBP001	Oberbipp	Switzerland	5,320–5,070	138,083	0.5	1.2	59.3
IKI003	Ikiztepe	Turkey	5,290–5,050	840,560	2.7	8.3	90.0
SUA004	Su Asedazzu	Italy	4,300–4,010	2,517,870	11.5	24.0	93.2
MK3001	Marinskaja 3	Russia	2,990–2,870	79,534	0.8	0.7	37.8
ETR001	Chiusi	Italy	1,810–1,620	1,977,148	24.5	17.8	93.7

cal yr BP, calibrated years before the present (95% confidence interval), based on direct AMS dating. Number of mapped reads, percentage of endogenous DNA (Endog. DNA), mean coverage and percentage of bases covered at least once (Reference covered) are based on the alignment to *Paratyphi C* RKS4594.

The bacterial pathogen *S. enterica* encompasses over 2,500 serovars. Today 99% of salmonelloses in mammals, including up to 200,000 annual fatal human infections, are associated with *S. enterica* subsp. *enterica* (*S. enterica*) comprising 60% of all serovars^{10,11}. *S. enterica* that cause systemic infection are often restricted to single host species or have a narrow host range, while *S. enterica* causing gastroenteritis are host generalists and only rarely cause systemic infections^{12,13}. *S. enterica* Paratyphi C causes systemic disease specifically in humans, but is a member of a phylogenetic lineage ('Para C Lineage') that also includes serovars invasive for pigs (*Typhuisis*, *Choleraesuis*). Choleraesuis also rarely infects humans, and previous analyses posit that Paratyphi C arose as a zoonosis in Europe within the past 4,000 yr, possibly by a host jump of Choleraesuis from domesticated pigs to humans^{9,14}. The evolution of ancient *Salmonella* genomes might thus have been influenced by shifts in human cultural practices during the Neolithization process.

Transect of eight ancient *S. enterica* across western Eurasia

To investigate these changes, we screened for *S. enterica* DNA in 2,739 ancient metagenomes extracted from human skeletons spanning primarily the Eurasian Neolithization process until the Middle Ages. Twenty-four metagenomes yielded multiple reads mapping to different *S. enterica* serovars. Each sample was enriched for DNA using a targeted in-solution *S. enterica*-specific hybridization capture assay, which allowed the reconstruction of eight genomes with a coverage of 0.7 to 24-fold (Table 1). The human teeth from which these genomes were assembled were up to 6,500 yr old, and their locations ranged from Russia to Turkey to Switzerland (Fig. 1). The dental pulp chamber is vascularized during life, and the successful retrieval of *S. enterica* DNA suggests that bacteria were present in high levels in the blood at the time of death¹⁵. These results imply that systemic *S. enterica* infections were geographically widespread across prehistoric western Eurasia.

During the past 8,000 yr, the Neolithic lifestyle spread across western Eurasia and adapted to regionally different conditions¹⁶. We investigated differences in human subsistence practices and genomic ancestry of ancient *S. enterica*-positive individuals using both archaeological information and human ancient DNA (see also Supplementary Note 1). The oldest burial site, Murzikhinsky II in western Russia with two 6,500-yr-old individuals (MUR009 and MUR019), is archaeologically assigned to the local Eneolithic, a transitional period dominated by a foraging economy that relied on hunting, with first evidence for the adoption of pastoralism¹⁷. In addition, genomic ancestry components of MUR009 and the mitochondrial DNA from MUR019 resemble those of previously sequenced eastern European foragers (Supplementary Table 1 and Extended Data Figs. 1 and 2). In contrast, the 5,500-yr-old individuals IV3002 (southern Russia) and IKI003 (Turkey) were from

a regional Bronze Age culture that primarily practised a pastoralist economy linked to the use of domesticated sheep, goat and cattle. The other individuals, from the Neolithic (OBP001) in Switzerland, the Bronze Age (SUA004) in Sardinia, the Iron Age (MK3001) in Russia and the late Roman Empire (ETR001) in Italy, were associated with archaeological evidence for an agropastoralist economy (Supplementary Note 1). Ancient human DNA analysis for these individuals (except OBP001, with insufficient recoverable DNA) infers genomic compositions consistent with (agro-)pastoralist economies based on previously published populations with similar ancestries (Supplementary Note 1 and Extended Data Figs. 1 and 2).

In conclusion, the eight bacterial genomes transect 4,700 yr of *S. enterica* infections during the spread of an (agro-)pastoralist subsistence, and can provide an unprecedented view into the diversity and evolution of *S. enterica* during this transformative time.

Ancient *S. enterica* genomes form a previously uncharacterized super branch

We assessed the relationships of the ancient *S. enterica* genomes to the modern diversity by leveraging a representative set of 2,961 genomes from Enterobase^{9,10}. For phylogenetic reconstruction, we built a multi-sequence alignment against Paratyphi C, a close relative to several ancient genomes in our screen. To minimize any reference bias we retained only positions covered across all genomes and excluded ancient genomes with <5-fold coverage, to ensure reliable single-nucleotide polymorphism (SNP) calling. Surprisingly, all the ancient genomes clustered in a single phylogenetic branch containing a limited number of serovars (Fig. 2a), which we designate the Ancient Eurasian Super Branch (AESB). Only 60 out of the 2,961 representative, modern *S. enterica* genomes are part of the AESB (Fig. 2a), the majority of representative genomes belonging to multiple other branches^{9,10} that were not found among the ancient genomes. Those results suggest that all detected prehistoric *S. enterica* infections across western Eurasia were caused by a sub-group of serovars within the much larger *S. enterica* diversity known today.

Emergence of *S. enterica* cluster HC2600_1272 during the Neolithization process

To provide greater statistical power for detailed analyses of the AESB, we identified 403 additional modern genomes within Enterobase that belong to the AESB, and included them in subsequent analyses. A stringent phylogenetic analysis of the AESB was performed with 37,040 SNPs, which could be called in all 460 modern and six ancient genomes. The results demonstrate that the ancient genomes fall on multiple distinct lineages within the AESB (Fig. 2b). The same topology for the ancient genomes was found with a more relaxed phylogenetic analysis based on 130,036 SNPs

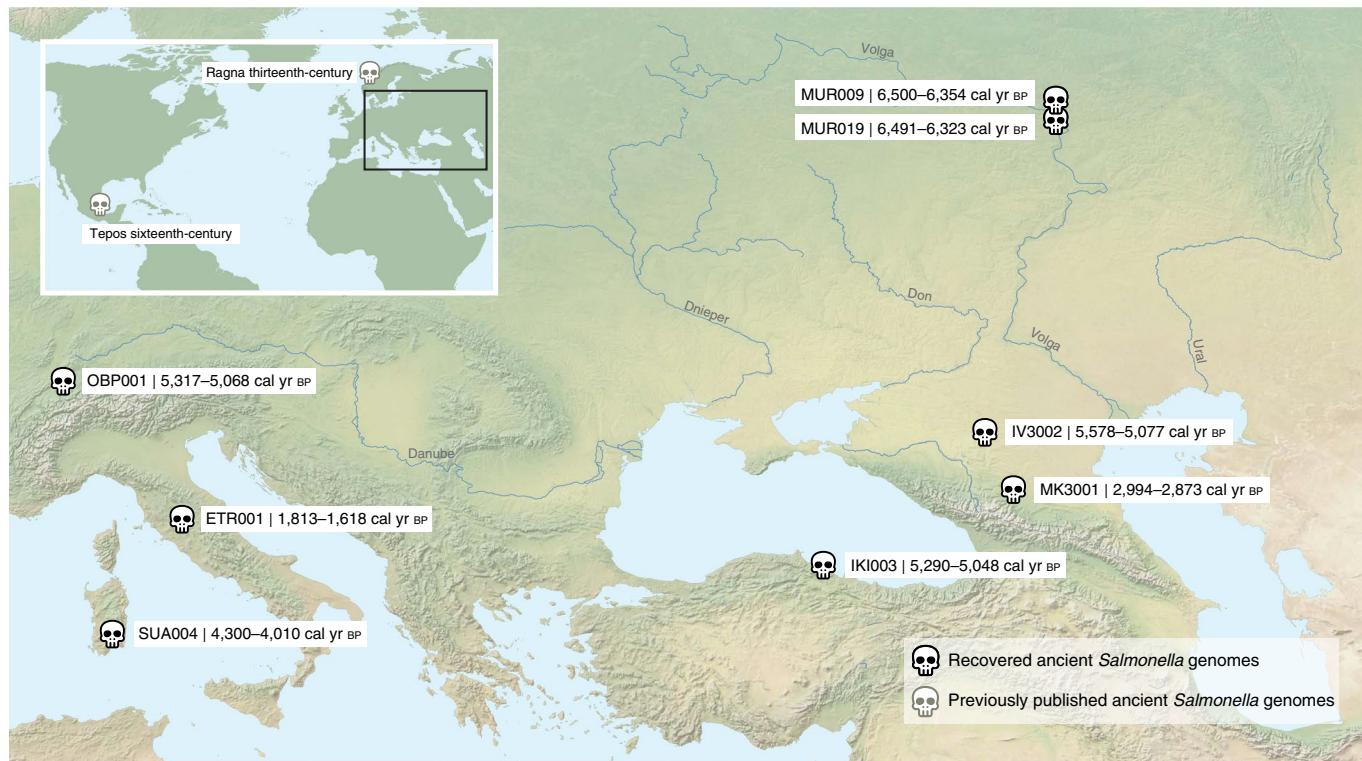


Fig. 1 | Geographic location and radiocarbon age of ancient human individuals infected with *S. enterica*. Previously published ancient genomes from thirteenth-century Norway (Ragna) and sixteenth-century Mexico (Tepos) are also shown.

that could be called from at least 95% of the genomes, except for MUR019 and a few modern clades whose topological associations varied (Extended Data Fig. 3). Such variable topological assignments might result from homologous recombination, which is common in *S. enterica*¹⁸. Indeed, we detected high levels of recombination for multiple deep lineages within the AESB, including the sub-branch on which MUR019 is located (Extended Data Fig. 4), suggesting that recombination might cause the topological differences between the stringent and relaxed phylogenies. The topology of the other ancient genomes did not differ between the two analyses, and only these were used for further phylogenetic analyses. However, exclusion of MUR019 reduces our observation of ancient *S. enterica* genomes from a transitional forager economy to a single example, MUR009.

The MUR009 *S. enterica* genome from an individual associated with a transitional forager economy is most closely related to a rare lineage containing diverse isolates (Fig. 2b). That lineage includes serovar Abortusequi, which can cause miscarriage in horses¹⁴, as well as a group of ST416/417 strains possibly adapted to marine mammals because they were exclusively recovered from the lungs of stranded harbour porpoises¹⁹. The branch also includes serovar Bispebjerg, which has been isolated from turtles and humans, and serovar Abortusovis, which can cause miscarriage in sheep¹⁴.

All *S. enterica* genomes from ancient (agro-)pastoralists are phylogenetically related to the previously designated Para C Lineage⁹. This lineage includes the invasive serovars Paratyphi C, Typhisuis and Choleraesuis, as well as the rare serovar Lomita (Fig. 2b)⁹. Of these serovars, Choleraesuis and Typhisuis are adapted to pigs. However, Paratyphi C is a major cause of enteric fever in humans, and Choleraesuis and Lomita can also cause systemic diseases in humans²⁰. The previously described ‘Tepos’⁸ (Mexico, ~500 yr before the present (bp)) and Ragna⁹ (Norway, ~800 bp) genomes are at the base of modern Paratyphi C. Of the newly reported ancient genomes, the ~1,700 bp genome from Italy (ETR001) defines a

novel fourth sub-branch of Choleraesuis. The early Bronze Age genomes SUA004 (Italy, ~4,200 bp) and Neolithic OBP001 (Switzerland, ~5,200 bp) define a new branch that is basal to the entire Para C Lineage. The other ancient genomes from (agro)-pastoralists (IV3002, IKI003 and MK3001) fall instead outside the Para C Lineage.

The original description⁹ of the Para C Lineage noted that genomes most closely related to that lineage were in serovar Birkenhead, a rare pathogen of humans²¹. Hierarchical clustering of core-genome MLST sequence types shows that the Para C Lineage and Birkenhead are both in cluster HC2600_1272, which also includes all ancient genomes from (agro-)pastoralists. Among those, the oldest genomes IV3002 (Caucasus, ~5,300 bp, early Bronze Age) and IKI003 (Turkey, ~5,200 bp, late Chalcolithic) define a new branch closely related to Birkenhead (Fig. 2b). Notably, despite the large geographic distance between these archaeological sites, the genomes differ by only 170 SNPs (out of 2.4 million aligned positions) suggesting a relatively fast dissemination during the Bronze Age, which is also confirmed by other aspects of the archaeological record (Supplementary Note 1). Finally, a low-coverage ancient *S. enterica* genome from southern Russia, MK3001 (~2,900 bp), falls basal in the HC2600_1272 cluster (Fig. 2b). These data show that *S. enterica* strains that are part of the HC2600_1272 cluster have been infecting human pastoralists and agropastoralists for >5,000 yr across western Eurasia.

Molecular dating of the ancient *S. enterica* HC2600_1272 cluster

The broad temporal transect of ancient bacterial genomes presented here has the potential to yield a better understanding of the evolutionary timing of *S. enterica* diversification. Overall, the AESB represents a very deep split within the *S. enterica* diversity (Fig. 2a), possibly even corresponding to a time to the most recent common ancestor (tMRCA) predating the arrival of humans in Eurasia. We

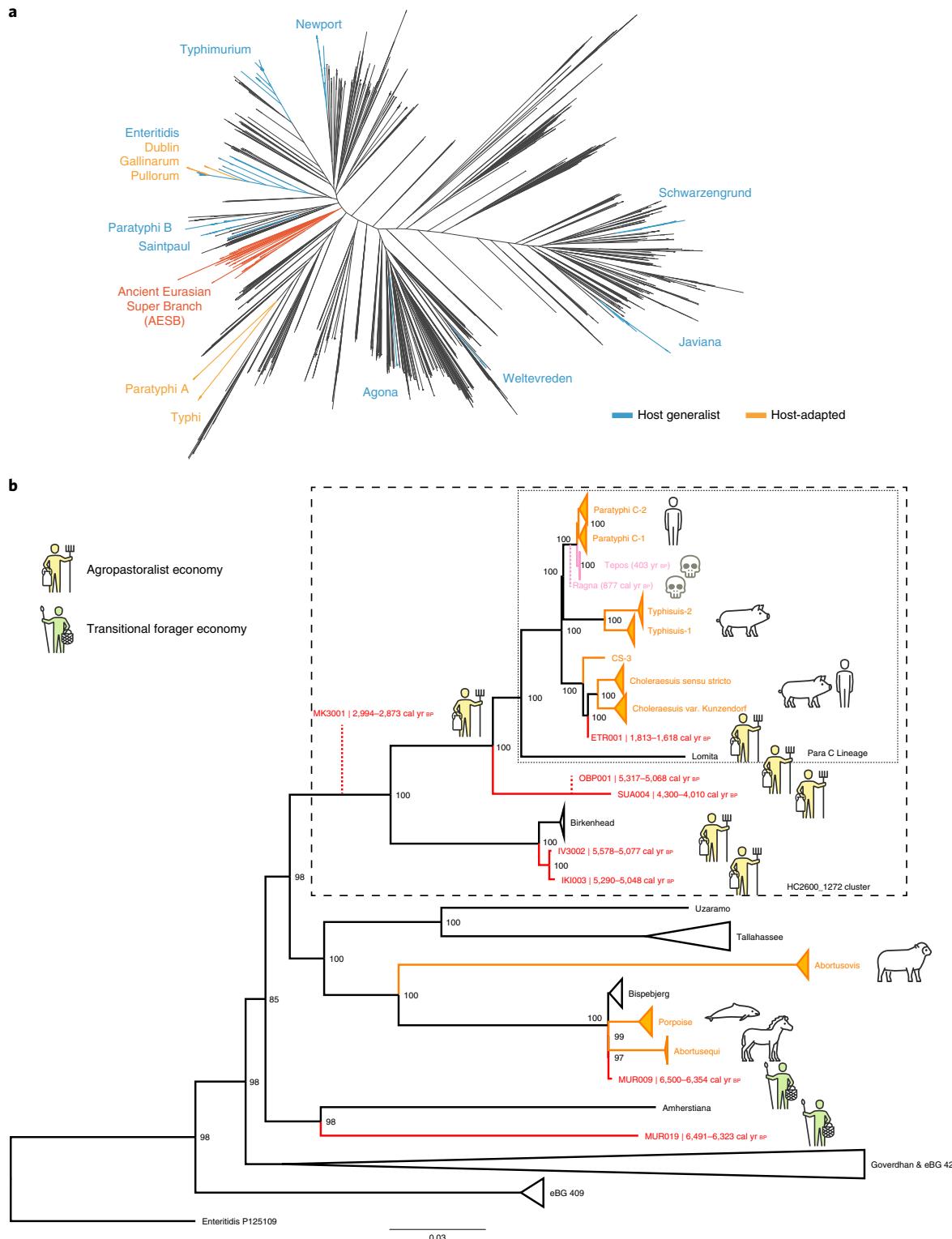


Fig. 2 | Phylogenetic relationships of reconstructed ancient and modern *S. enterica* core genomes. **a**, Maximum likelihood tree of ancient genomes (>5-fold coverage) and 2,961 modern *S. enterica* genomes, including 182,645 SNP positions in the core genome. Selected branches are identified based on predicted serotype provided by Enterobase and coloured according to host specificity (blue/orange); red denotes inclusion of ancient genomes, black denotes not specified. **b**, Maximum-likelihood tree of the AESB including ancient genomes (>5-fold coverage) and 463 *S. enterica* genomes, considering 37,040 SNP positions in the core genome. New ancient genomes are shown in red, and previously reported ancient genomes (Ragna, Tepos) in pink. Ancient human economy is indicated for all newly presented genomes based on archaeological and ancient human genetic information. Low-coverage (<5-fold) ancient genomes are phylogenetically placed (red dashed line) based on all SNP positions covered once: MK3001, 17,324; OBP001, 26,657; and Ragna, 35,465. Modern genomes are collapsed based on their predicted serovar, eBurst group (closely related sequence types) or available metadata in Enterobase. Host-adapted serovars are coloured orange (including a pictogram of host species). Bootstrap values are shown in black at each node ($n=1,000$). Black dashed rectangles denote the extents of Para C Lineage and hierarchical cluster HC2600_1272. Enteritidis P125109 is used as the outgroup.

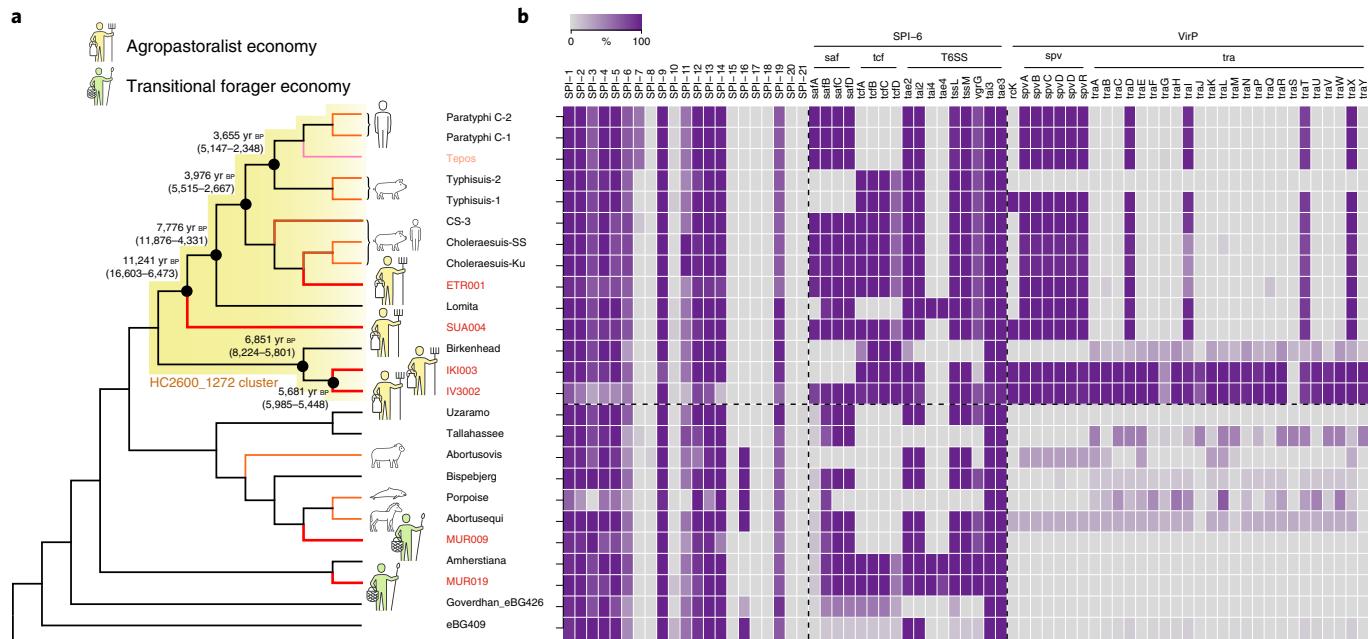


Fig. 3 | AESB topology, divergence times and gain-loss events. **a**, Topology of the AESB highlighting the hierarchical cluster HC2600_1272 (yellow), with symbols indicating ancient human economy of *S. enterica*-positive samples. Selected divergence time estimates for the HC2600_1272 cluster are shown in yr BP (95% highest posterior density intervals; see also Supplementary Table 2). **b**, Gain-loss results for all SPIs and selected genes. For SPIs, colour gradient relates to the average percent of genes, with mapped reads over 95% of the bases, across all strains per branch; for genes, colour gradient according to average percentage covered across all strains per branch. Tepos, sixteenth-century Mexican.

did not attempt to date the entire AESB, due to a weak temporal signal²² (correlation coefficient R^2 between phylogenetic distance and sampling times = 0.04). This may be due to problems in distinguishing single-step mutations from homoplasies across the long timescale of the AESB. In addition, our analyses revealed high levels of recombination on several internal branches of the AESB, which may affect branch lengths and can interfere with molecular dating²³ (Extended Data Fig. 4). Hence, we restricted our dating analysis to the HC2600_1272 cluster which yielded a stronger temporal signal ($R^2=0.16$) and successfully passed a date-randomization test²⁴ (Extended Data Fig. 5). Bayesian phylogenetic molecular dating with BEAST2²⁵ confirmed the close relationship of IV3002 and IKI003 with a tMRCA of 5,680 (5,990–5,450) yr ago (Supplementary Table 2 and Fig. 3a). The split between SUA004 and the Para C Lineage was estimated around 11,000 (16,600–6,470) yr ago. We also confirm the previous estimates⁹ of around 4,000 yr ago for the emergence of the direct progenitors of Paratyphi C. We note, however, that these estimates have large confidence intervals and that we did not date the entire HC2600_1272 because estimates are potentially affected by recombination events not detected with our approach.

Genomic architecture differentiates transitional forager and (agro-)pastoralist strains

Differences in *S. enterica* disease manifestation and virulence are not fully understood and have been linked to variation in genomic content, such as *Salmonella* pathogenicity islands (SPI) and the virulence plasmid (virP)¹⁴. The ancient and modern strains forming the AESB largely show genomic stability, with consistency in SPI presence or absence (Fig. 3b and Supplementary Note 2).

One exception is SPI-6, which encodes several putative virulence factors such as the fimbriae gene clusters *saf* and *tcf*, or the Type 6 Secretion System²⁶. These show variable gene content across different modern and ancient serovars (Fig. 3b and Supplementary Note 2). Furthermore, the virulent phenotype of virP is conferred by the *spv* locus, which is involved in bacteraemia^{27,28} and is present

primarily in the HC2600_1272 cluster (Fig. 3b and Supplementary Note 2). In addition, IV3002 and IKI003 carry the conjugal transfer operon *tra*, which probably facilitated the original horizontal transfer of *virP* into the HC2600_1272 cluster²⁹. Lastly, the virulence gene *rck*, which confers complement resistance³⁰, is present in the ancient *S. enterica* genomes IV3002, IKI003 and SUA004 but is absent in most modern genomes, possibly due to natural selection. Overall, the differential mobile gene content probably contributed to variation in pathogenicity between the HC2600_1272 cluster and other *S. enterica* within the AESB, as well as between ancient and modern *S. enterica* within the HC2600_1272 cluster.

Host range of ancient *S. enterica* strains and evolution of host specificity

Understanding the host range of the ancient *S. enterica* is informative about human and mammalian host adaptation across the AESB. Among all ancient *S. enterica*, only ETR001 is phylogenetically confined within a group of host-adapted modern *S. enterica* (pig/human-adapted Choleraesuis), whereas all other ancient genomes are basal to *S. enterica* adapted to different hosts or basal to non-adapted lineages, suggesting that those early ancient strains, too, were non-host adapted (host generalists). Accumulation of pseudogenes has been linked to host adaptation in various *S. enterica* serovars, including Paratyphi C and Choleraesuis^{31–34}. Thus, the pseudogene frequency observed in the ancient *S. enterica* can provide additional evidence about their host range even though we cannot predict definitely host specificity from genomes alone. We used an unbiased set of over 8,000 *S. enterica* pan-genes to infer pseudogenes and observed, as expected, an increase in pseudogenization from host generalists to host-adapted serovars on the AESB (Fig. 4a and Supplementary Data 1). Strikingly, all newly reported ancient genomes are in the range of host generalists, and the oldest genomes, MUR009 (14), IV3002 (15) and MUR019 (18), have the lowest pseudogene frequency within the entire AESB (Fig. 4a). Among the ancient genomes, ETR001 has, in accordance with its

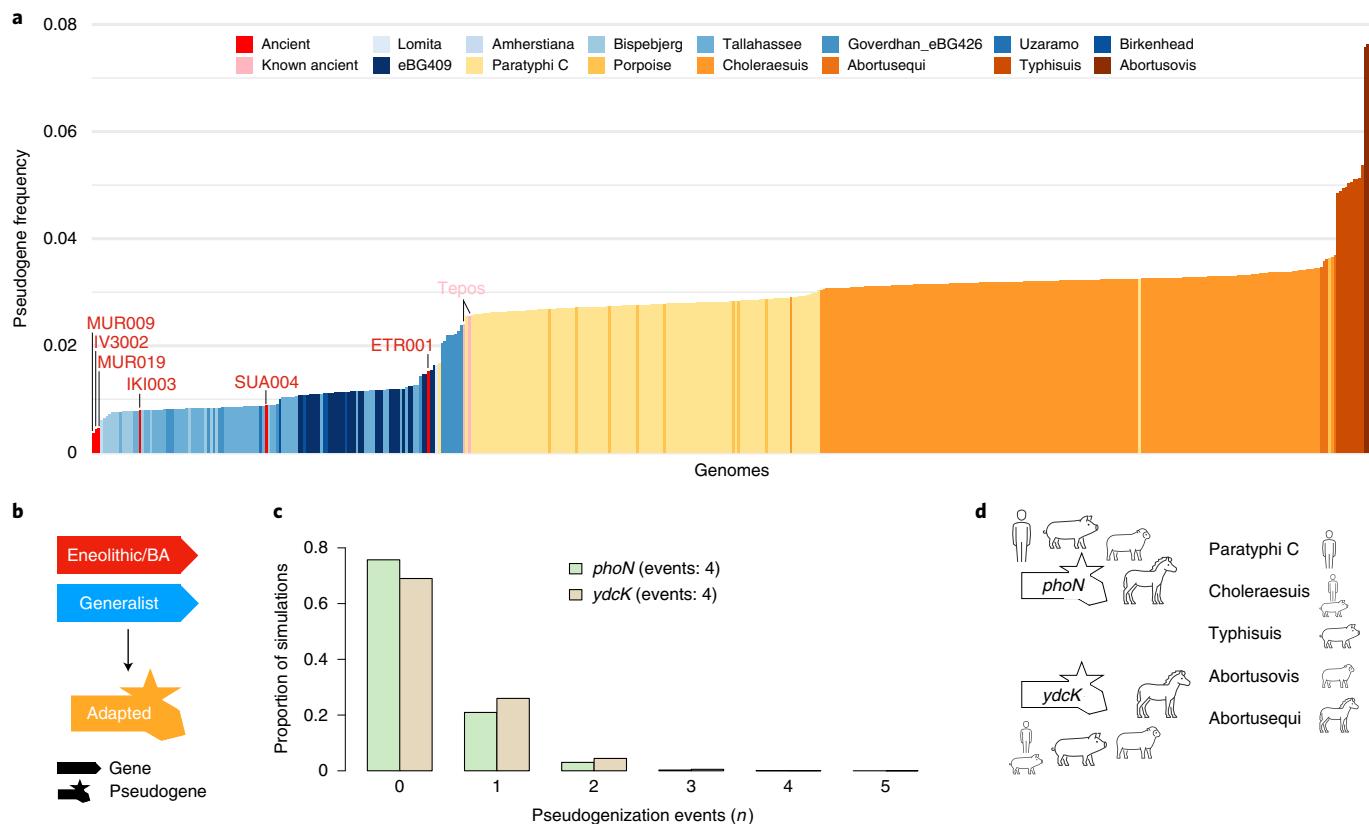


Fig. 4 | Pseudogenes and evolution of host adaptation across the AESB. **a**, Relative frequency of pseudogenes for each strain of the AESB. Ancient genomes are identified, with host generalists shown in blue and host-adapted in orange. **b**, Illustration of model used to infer evolution of host adaptation. **c**, Simulated expectation of candidate genes (*phoN* in green, *ydcK* in beige) to harbour randomly distributed pseudogenization events using 10,000 simulations. **d**, Host-adapted *S. enterica* serovars harbouring a *phoN* or *ydcK* pseudogene. BA, Bronze Age; Tepos, sixteenth-century Mexican.

archaeological age and phylogenetic placement, the highest pseudogene frequency, which is close to that of Paratyphi C strains and indicates that it might have been host adapted. Interestingly, the accumulation of pseudogenes correlates with archaeological age ($R^2=0.48$; Extended Data Fig. 6). Using linear regression, we calculated the rate of pseudogenization across ancient samples to be approximately one gene per 75 yr, which extrapolates towards frequencies observed today in human-adapted Paratyphi C. Taken together, the phylogenetic placement and low frequency of pseudogenes in early ancient genomes suggests that prehistoric systemic *S. enterica* infections in humans were caused by host generalists.

The AESB harbours six different host-adapted *S. enterica* serovars, several serovars that are host-unrestricted and the predicted host-unrestricted ancient strains older than 3,000 yr. This provides an opportunity to study the evolution of bacterial host adaptation over the past six millennia. Notably, all modern host-adapted serovars carry largely distinct sets of pseudogenes irrespective of relatedness (Extended Data Fig. 7), which is in line with previous results in *S. enterica* and suggests a primarily independent evolution of pseudogenes^{31,32}. However, shared pseudogenes that are formed independently probably represent key changes necessary for host adaptation. Here we harness the vast serovar diversity present on the AESB to test for pseudogenization events that may have facilitated the adaptation towards different mammalian hosts (Fig. 4b). We identified 29 candidate genes harbouring two or more independent pseudogenization events and that have different functional roles including metabolism, transcription and biofilm formation (Supplementary Table 3). Out of those, two genes harbour an excess of pseudogenization events based on random simulations (Bonferroni corrected $P=0.006$ (*phoN*) and $P=0.02$ (*ydcK*);

Fig. 4c and Supplementary Table 4) and are lost in four host-adapted serovars (Fig. 4d). Notably the top candidate, *phoN*, is regulated by PhoPQ and can induce a strong antibody response during systemic *Salmonella* infections in humans and mice, even though its exact role is not yet understood³³. Altogether, these results indicate that convergent evolution in terms of pseudogene formation contributed to *S. enterica* host adaptation.

Discussion

Here we elucidate the evolutionary history of *S. enterica* interwoven with the dramatic changes in human lifestyle during the spread of Neolithic economies¹ (graphical abstract in Extended Data Fig. 8). To do this, we harnessed ancient bacterial and human DNA obtained from fossils, together with their archaeological record. We present eight ancient *Salmonella* genomes from as early as 6,500 yr ago isolated from ancient pastoralists, agropastoralists and transitional foragers. All are phylogenetically confined within the AESB, suggesting that it represented a common yet diverse group of *S. enterica* that infected humans in prehistory. Apart from ancient genomes, the AESB contains the human-adapted serovar Paratyphi C and the majority of known animal-adapted *S. enterica* serovars.

All ancient genomes were recovered from the pulp chamber of human teeth. The pulp chamber is supplied by blood during life and, after death, is well insulated from environmental conditions which allows us to conclude that human-pathogenic bacteria found in the pulp chamber probably originated from systemic disease present at the time of death¹⁵. Thus, our findings highlight systemic *S. enterica* infections by strains part of the AESB as impacting human health during the past six millennia of human prehistory. It is notable that none of the current predominant causes of invasive salmonellosis

in humans were found, such as serovars Typhi and Paratyphi A. Possibly these were less common causes of systemic human disease across western Eurasia in the past.

The newly reported ancient *Salmonella* strains most probably caused systemic salmonellosis despite their lack of SPI-7, a pathogenicity island contributing³⁶ to paratyphoid fever by Paratyphi C and to typhoid fever by Typhi. Moreover, the *S. enterica* strains older than 3,000 yr that we have identified were possibly host generalists and were not specifically adapted to humans. The suggestion that they were generalists is supported by the ability of the modern descendants of some of their phylogenetic neighbours to cause disease in a variety of mammals, and by the relatively low frequency of pseudogenes, which are thought to increase in number following host adaptation^{31–34}.

The Neolithization process caused the first epidemiological transition where the introduction of domesticated animals, intensification of human–animal co-residence and changes in mobility are thought to have led to the emergence of new zoonotic diseases in humans through increased exposure to pathogens^{2,3,5,37,38}. We reconstructed two *S. enterica* genomes from ancient transitional foragers excavated from a single site in Russia, but were able to place only one of them in the phylogeny with any degree of confidence, which precludes general conclusions about the diversity of *S. enterica* infections in transitional forager populations and their geographic spread. However, we reconstructed six ancient *S. enterica* genomes from (agro-)pastoralists spanning over five millennia of human cultural evolution across western Eurasia, and all are within the HC2600_1272 cluster that includes human-adapted Paratyphi C. This suggests that progenitors to Paratyphi C evolved within pastoral and agropastoral societies during the Neolithization process, and provides evidence that the first epidemiological transition facilitated the emergence of human-specific Paratyphi C. However, we note that our ancient metagenomic data screened for *S. enterica* are biased towards human samples from the Neolithization process and that the HC2600_1272 cluster emerged possibly before the Neolithic transition in a context currently unknown. To better understand the origin and range of hosts susceptible to infections with early strains from the HC2600_1272 cluster, more ancient metagenomic data from early human groups as well as faunal remains are necessary.

The origin of human-specific Paratyphi C⁹ was previously thought to be a result of a spillover event from pigs to humans around 4,000 yr ago due to the close relationship between human- and pig-adapted lineages. Our identification of putative host generalist strains in humans as early as 5,500 yr ago (IV3002, IKI003 and SUA004), that are phylogenetically basal to human- and pig-adapted serovars, renders the pig-origin hypothesis unlikely. Our results rather support two different hypotheses: (1) that pig-adapted serovars resulted from an anthroponosis—that is, a spillover from humans to pigs³⁹ or (2) that adaptations to humans and pigs occurred via independent processes during the Neolithization process within a permissive environment that led to continuous *S. enterica* exposure and subsequent infection⁴⁰. An independent adaptive history is further supported by the observed small overlap of pseudogenes among host-adapted serovars. Ultimately, additional sampling, including ancient faunal remains, will be required to lend credence to either hypothesis and to further elucidate the evolutionary history of *S. enterica* and other pathogens.

Methods

Sequencing library preparation. Sample processing took place in dedicated ancient DNA facilities at the Max Planck Institute for the Science of Human History in Jena and the Palaeogenetics laboratory at the University of Tuebingen. Teeth were cut at the enamel–dentin junction and sampled in the pulp chamber with a dental drill. Tooth powder was extracted⁴¹ and the resulting DNA was transformed in either double- or single-stranded (only for IKI003) genetic libraries with the use of full, partial or no uracil DNA-glycosylase (UDG) treatment to fully, partially or not remove the characteristic C-to-T substitutions towards both ends

of ancient DNA molecules, respectively^{42–45}. Genetic libraries were double-indexed and amplified before shotgun sequencing. Those libraries identified as positive for *S. enterica* after screening were further captured using an in-solution target enrichment of *Salmonella enterica* subsp. *enterica* DNA⁸. The same libraries were used for a human genome-wide capture that targeted ~390,000 SNPs for ETR001 and ~1.2 million SNPs for all other samples⁴⁶. In addition, negative controls were taken along with initial library preparation and *S. enterica* capture. All sequencing runs were performed on Illumina platforms in either single- or paired-end mode.

Detection, authentication and genome reconstruction. For screening we used the software package HOPS⁴⁷, which utilizes MALT⁸, a software that aligns sequencing data to a custom-made reference database of all 6,247 complete bacterial genomes from the National Center for Biotechnology Information (NCBI; obtained December 2016) using spaced seeds. Reads were mapped with at least 85% nucleotide identity, and for taxonomic placement via the lowest common ancestor algorithm we retained only alignments within the top 1% of all alignments per read (maximum 100). For candidate detection, we interrogated the *Salmonella enterica* subsp. *enterica* taxonomic node and required a minimum of ten assigned sequencing reads showing no or only a few mismatches (a declining edit distance distribution). This step is crucial to avoid false positives due to incomplete genomic representation of unknown environmental bacteria in our database⁷. Furthermore we required the presence of typical ancient DNA damage⁴⁸—that is, C-to-T or G-to-A mismatches at the end of sequencing reads. We screened 2,739 ancient metagenomic datasets obtained from human remains excavated in Eurasia and South America that were preliminarily archaeologically classified into the following coarse groups: Eneolithic/Chalcolithic (95), Iron Age (61), Late Neolithic/Bronze Age (1,027), Medieval (382), Mesolithic (135), Neolithic (548), Palaeolithic (183), Post-Columbian (46), Pre-Columbian (127) and other (135). From those, 24 candidates had positive hits on the *Salmonella enterica* subsp. *enterica* taxonomic node.

We enriched all candidate libraries for *S. enterica* DNA using a previously presented in-solution capture reagent based on 67 *S. enterica* reference genomes⁸. Each library was sequenced for about 5 million reads and mapped to the Paratyphi C RKS4594 genome ([NC_012125](#)) using Burrows–Wheeler Aligner (BWA)⁴⁹ with non-stringent criteria (-n 0.01) within the software package EAGER⁵⁰. We use the Paratyphi C RKS4594 reference because it regularly occurred among the top hits in the screening analysis, and the overall minimal divergence (maximum 1.3% genome-wide) among *S. enterica* reference core genomes suggests no major effects by reference bias. The resulting data were evaluated for a minimum genome-wide coverage of 0.3x with 15% of the reference covered. For samples with higher genome-wide coverage, the percentage-reference covered scaled up linearly. In addition, DNA damage (>10% C>T mismatches at the 3' end) had to be observable, which led to eight samples positive for *S. enterica* (Extended Data Fig. 9). All eight archaeological specimens are shown in Extended Data Fig. 10. The remaining 16 libraries had either too little endogenous *S. enterica* DNA or were false positives due to the presence of closely related species harbouring genomic elements aligned to the *Salmonella enterica* subsp. *enterica* node by MALT. All negative controls were negative. For all positive samples (except IKI003) we generated two UDG full treated libraries (removing damage) from the initial ancient DNA extracts, which provided high-quality data for further analysis.

Library preparation of IKI003 was performed in the clean-room facilities dedicated to ancient DNA work at MPI EVA in Leipzig. Two single-stranded Illumina sequencing libraries⁴⁴ were generated from 30 µl of extract each, according to an automated version of the protocol, on an Agilent Technologies Bravo NGS Workstation without previous UDG treatment. Both libraries were double-indexed with a unique combination of seven base pair indices⁵¹.

An in-solution *S. enterica* capture was repeated for each library, and all libraries were sequenced to exhaustion (between 23 million and 133 million reads). For ETR001, SUA004, OBP001, MUR009 and MUR019, only UDG-treated data were used for downstream analysis. To maximize genome-wide coverage for IV3002 and MK3001, we merged UDG data with the initial UDG half-capture data trimmed at the 3' and 5' tails by two bases to cleave ancient DNA damage. IKI003 underwent single-stranded library preparation without any UDG treatment, and each sequencing read was trimmed by two bases to minimize the impact of ancient DNA damage in all subsequent analyses. Using this approach we generated high-quality data with as few alignment mismatches as possible due to ancient DNA damage. We reconstructed all genomes through a mapping strategy to the Paratyphi C reference genome. Again, all data were processed through EAGER but with stringent BWA mapping parameters (-n 0.1), leading to eight reconstructed genomes of which six had an average coverage of 7x and above.

Modern *S. enterica* genomes and phylogenetic analysis. For initial phylogenetic placement of the ancient genomes we used a recently published set of 2,961 *S. enterica* genome assemblies from Enterobase^{9,10}, where each genome assembly represents a random pick of a unique ribosomal sequence type (rST). The rST is a multi-locus sequence type based on 53 genes encoding ribosomal proteins and is used for efficient capture of *S. enterica* diversity⁵². Enterobase contains over 140,000 *S. enterica* genomes, and this approach allowed a drastic reduction in the number of genomes and, hence, a reduction in computational resources necessary for comprehensive analysis.

Each genome was split into k-mers of 100 bases with a step-size of one base, and mapped against the Paratyphi C RKS4594 reference following the same stringent criteria used for the ancient genomes (see above). In addition, we included two previously published sixteenth-century Mexican (*Tejos*) and one medieval Norwegian (*Ragna*) genome using the publicly available raw reads^{8,9}. An alignment of all variable sites was built for all modern and ancient genomes using the tool multivcfanalyzer (v.0.87-alpha)⁵³. Repetitive and highly conserved regions of the Paratyphi C RKS4594 reference were excluded from SNP calling to avoid spurious read mapping⁸. To avoid spurious SNP calls, every site reported per genome had to have at least 5-fold coverage and a genotype support of at least 90%. We defined the core genome of all alignments by using only sites shared across all genomes (complete deletion), which required the removal of ancient genomes with genome-wide coverage <5-fold (MK3001, OBP001 and *Ragna*) to avoid excessive loss of positions, leading to alignment with a total of 182,645 SNPs. A maximum-likelihood tree was built with RAxML⁵⁴ using the GTRCAT model and 100 bootstraps (Fig. 2a). We use the CAT approximation of rate heterogeneity compared to GAMMA because it is computationally much more tractable, with large datasets and produces comparable results⁵⁵.

The ancient genomes were phylogenetically placed in the AESB, which harbours only 60 out of the 2,961 genomes. Notably, we observed uncertainty for very deep lineages based on bootstrap support and high levels of recombination, including MUR019, which thus cannot be placed with high confidence within the vast *S. enterica* genomic diversity. For detailed analysis, from the 140,000 genomes available on Enterobase (late March 2018) we obtained all genome assemblies with one of the unique rSTs identified on the AESB, leading to a total of 465 genome assemblies from Enterobase (http://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:12971), which represents the known genetic diversity of the AESB. These include several host generalists recovered from various animal species (Tallahassee, Uzaramo and Goverdhan) and/or by food safety surveillance laboratories (for example, eBG409 and 426)^{56–61}. In addition, we added the Enteritidis P125109 reference genome as an outgroup because it is phylogenetically close to the AESB. Data processing and filtering followed the same guidelines as above, leading to a SNP alignment with 37,040 positions. A maximum-likelihood tree was built with RAxML⁵⁴ using the GTRCAT model and 1,000 bootstraps (Fig. 2b). Due to their ambiguous placement we removed two genomes: FDA00002391 and FDA00002392. We estimated the phylogenetic placement of the three ancient genomes filtered due to coverage (MK3001, OBP001 and *Ragna*) by adding them to the existent SNP alignment with relaxed SNP calling criteria. We required only 1-fold coverage per site to call a genotype, which led to the following number of callable sites—MK3001, 17,324; OBP001, 26,657; and *Ragna*, 35,465—out of the 37,040 SNPs. Missing or ambiguous sites were typed as N. In addition, we used a more relaxed filtering approach by using all SNPs shared by at least 95% of all sequences, which led to 130,036 SNPs and is shown in Extended Data Fig. 3.

Hierarchical clustering of core-genome MLST sequence types at multiple levels of pair-wise linkage distances was done using Enterobase⁶².

Quantification of recombination across the AESB. The alignment based on all SNPs shared by at least 95% of all sequences was used for recombination detection, which covered 83% of the sites in the Paratyphi C RKS4594 reference genome. A total of 130,597 SNPs were identified in these relaxed core genomic sites, and assigned using ETOKi-phyo⁶² onto branches in the tree using a maximum-likelihood method with a symmetric transition model⁶³. We found that 49,101 sites (38%) were mutated on multiple independent occasions (homoplasies), resulting in a total of 241,275 substitution events. RecHMM⁶⁴ was then applied to these substitutions to estimate the number of recombination events per substitution (Extended Data Fig. 4).

Human genetic analysis. Using EAGER⁵⁰, all sequencing reads from the human genome-wide capture data were trimmed, paired-end data were merged and mapped against the human reference genome (hg19) with a minimum mapping quality (MQ) of 30 and duplicates were removed. The typical ancient DNA damage pattern was assessed using mapDamage2.0 (ref. ⁶⁵).

Individual mitochondrial consensus sequences were reconstructed using schmutzi (base quality >20)⁶⁶; mitochondrial DNA haplogroups were assigned using HaploGrep⁶⁷, and further used for mitochondrial DNA contamination estimation. Nuclear contamination for male individuals was estimated based on heterozygosity at X-chromosome SNPs using ANGSD⁶⁸. Sex determination was achieved by comparing the coverage at targeted SNPs on the X- and Y-chromosomes relative to those on the autosomes (Supplementary Table 1).

For nuclear DNA analyses (all samples except IKI003) we minimized spurious SNP calls by minimizing the amount of ancient DNA damage in the data. Sequencing reads of fully UDG-treated libraries were kept untrimmed, sequencing reads from partially UDG-treated libraries were trimmed by 2 bp on both ends and sequencing reads from non-UDG-treated libraries were trimmed by 10 bp on both ends. The targeted set of SNPs was genotyped by pseudohaplotype calls with PileupCaller (<https://github.com/stschiff/sequenceTools>). For IKI003, the only single-stranded libraries, genotyping was performed on both 10-bp trimmed and untrimmed sequences, retaining calls deriving from untrimmed

data except for C-to-T mutations, the presence of which was inspected in the trimmed data to avoid mistaking DNA damage for real mutations. Genotype data were merged with the Human Origins dataset⁶⁹ comprising ~7,000 ancient and modern individuals. For nuclear DNA analysis we required at least 15,000 SNPs overlapping with the Human Origins dataset in each sample, which was not fulfilled for two individuals (OBP001 and MUR019), and they were thus excluded from nuclear DNA analyses. A principal component analyses was performed using smartpca⁷⁰. Principal components were computed on 65 present-day west Eurasian populations⁶⁹ onto which ancient individuals were projected (parameters lsqproject: yes, shrinkmode: yes). Model-based clustering analysis using ADMIXTURE⁷¹ was performed using the complete dataset (parameters: -s time -cv) with the number of ancestral populations (K) ranging from 3–16 and each repeated five times (Extended Data Fig. 2). The run with highest likelihood was reported for each K (Extended Data Fig. 2).

Molecular dating. The temporal signal of the full AESB was evaluated with the software TempEst²⁴, which employs a root-to-tip regression of genetic distances on sampling times. Due to the lack of a signal on the entire AESB, further analyses were done using a reduced dataset⁷. In total, we analysed two subsets of reduced replicates of the HC2600_1272 branch, both containing ETR001, CS-3, the two sixteenth-century Mexican samples (*Tejos*, Lomita, IV3002, IKI003, SUA004, all five sequences from the serovar Birkenhead, 38 randomly sampled sequences and a sequence from the serovar Bispebjerg as the outgroup. All major lineages of the HC2600_1272 cluster were represented in the subsets

The TempEst analysis yielded a temporal signal for both subsets ($R^2 = 0.17$ and 0.16). The presence of temporal signal in the subsets was further evaluated with a date-randomization test⁷². We generated ten replicates with randomized sampling dates for each alignment, and substitution rates were estimated with the software BEAST2²⁵ using a relaxed log-normal molecular clock, a general time-reversible model of nucleotide substitution with six gamma categories (GTR+ γ 6) and a Bayesian birth–death tree prior⁷³ with constant-through-time birth and death rates and an upper bound of 100,000 yr. This tree prior allowed explicit modelling of the fact that the proportion of samples is lower in the past than nearer the present. Through a monophyly constraint, the respective outgroup was enforced to be an outgroup in the analysis.

The two subsets of the HC2600_1272 branch passed the date-randomization test (Extended Data Fig. 5), and dating analysis was conducted with BEAST2 as described above. Because only SNP alignments were used, we corrected for ascertainment bias by specifying the number of invariant sites per base in relation to the Paratyphi C RKS4594 reference genome. For each dataset, multiple BEAST replicates with 400 million Markov chain Monte Carlo steps were performed and subsequently combined with the software LogCombiner, after removing a 10% burn-in. This yielded effective sample sizes above the standard threshold of 200 for all parameters. The posterior estimates for the main internal node dates and branch rates are summarized in Supplementary Table 2. The files specifying the BEAST2 analysis, the posterior distributions of all parameters and the maximum clade consensus trees are available on figshare, with a link provided in Data availability, below.

¹⁴C dating. Collagen was extracted from bone or tooth samples, purified by ultrafiltration (fraction >30 kDa) and freeze-dried. Collagen was combusted to CO₂ in an Elemental Analyzer and converted catalytically to graphite. Accelerator mass spectrometry dating was performed to estimate the ¹⁴C age of each sample—given in yr BP, meaning years before 1950. To provide absolute calendar ages, ¹⁴C ages need to be calibrated which was performed using the dataset INTCAL13⁷⁴ and the software SwissCal 1.0 (L. Wacker, ETH-Zürich). The results of the calibration were reported as ‘Cal 1-sigma’ and ‘Cal 2-sigma’ using the 1- and 2-sigma uncertainty of ¹⁴C ages, respectively. We report all ages in the format: ‘Cal 2-sigma start–end BP’ (uncalibrated radiocarbon years, laboratory number).

Genomic architecture. GenBank files for SPI-1–12 and SGI-1/2 were obtained from PAIDB⁷⁵, and all annotated CDS sequences extracted. Additional sequences for SPI-13–21 were obtained from respective references^{76–80}, as well as a SPI-6 annotation from the Para C Lineage⁹. GenBank files for plasmid annotations were obtained from NCBI. We mapped each genome (fragmented in k-mer of 50 b with step-size of 1) and the capture probes to each reference using BWA mem⁸¹, and filtered all alignments using picard tools (CleanSam, MarkDuplicates). For each annotation, we obtained the number of bases covered at least once using bedtools (genomecov -1)⁸². For analysis we considered only genes that were covered by at least 95% with the DNA capture probe set.

Pseudogene analysis. Pseudogenes were inferred based on premature stop codons or frameshift mutations relative to an intact gene across the *S. enterica* pan-genome. The previously published pan-genome of *S. enterica* (<http://enterobase.warwick.ac.uk/schemes/Salmonella.wgMLST/exemplar.alleles.fasta.gz>) is based on all CDS from 537 representative genomes, which were grouped based on sequence identity and cleaned for paralogues yielding 21,065 genes¹⁰. We filtered the complete pan-genome for genes that were 100% covered by the probe set used for DNA capture, leading to 8,726 genes used for inference. For each strain present

on the AESB, we aligned the genome to the refined pan-gene set using BWA mem⁸¹ allowing for overhang alignments (soft clipping) at the end of the reference, which is necessary for correct alignment to single genes. All alignments were filtered for duplicates, a mapping quality of 37 or above and realigned (indel realignment) before SNPs and insertion/deletion were called using GATK v3.5 (ref. ⁸²). For each strain and gene, we built a consensus sequence using GenConS (-total_coverage 2, -major_allele_coverage 2, -consensus_ratio 0.75, -punishment_ratio 0.8)⁸⁴ and inferred pseudogenes as genes with premature stop mutation or frame-shift mutation. For each genome, we considered only genes that were covered by at least 90% for analysis. We report in Fig. 4a the frequency of pseudogenes relative to the absolute number of genes analysed in each respective genome.

To test for signatures of convergent evolution in host-adapted serovars, for each gene we assessed all observed independent pseudogenization events. We selected candidate pseudogenes associated with convergent evolution of host adaptation on the AESB by filtering those with the following criteria: (1) having at least two fixed parallel (independent) pseudogenization events across the host-adapted serovars (*Abortusovis*, *Abortusequi*, *Porpoise*, *Typhisuis*, *Choleraesuis*, *Paratyphi C*) and (2) it is not a pseudogene in either the ancient host generalist from the Eneolithic and Bronze Ages (MUR019, MUR009, IV3002, IKI003, SUA004) or modern host generalist groups (Bispebjerg, Goverdhan_eBG426, eBG409, Tallahassee, Birkenhead). The probability of observing the respective number of pseudogenization events (or more) per gene with length L was simulated ($n=10,000$) by randomly distributing the observed number of pseudogenization mutations in host-adapted serovars (1,432) across the average pan-genome (size, 3,828,459 base pairs; $\alpha=69.974$; results shown in Supplementary Table 3). Bonferroni correction was applied to all simulation-derived probabilities to correct for multiple testing.

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

Data availability

Raw metagenomic data used to reconstruct ancient *S. enterica* genomes, as well as unpublished ancient human DNA, are available from the European Nucleotide Archive (Accession no. PRJEB35216; see also Supplementary Data 2). The molecular dating archive containing files specifying the BEAST analysis of the two 50-taxon (argo-)pastoralist datasets, as well as the thinned posterior distributions of all parameters and the maximum clade consensus trees, are available at <https://doi.org/10.6084/m9.figshare.10052084.v1>.

Received: 12 April 2019; Accepted: 8 January 2020;

Published online: 24 February 2020

References

- Fowler, C., Harding, J. & Hofmann, D. *The Oxford Handbook of Neolithic Europe* (OUP, 2015).
- Cockburn, T. A. Infectious diseases in ancient populations. *Curr. Anthropol.* **12**, 45–62 (1971).
- Armelagos, G. J. & Cohen, M. N. *Paleopathology at the Origins of Agriculture* (Academic Press, 1984).
- Larsen, C. S. et al. Bioarchaeology of Neolithic Çatalhöyük reveals fundamental transitions in health, mobility, and lifestyle in early farmers. *Proc. Natl Acad. Sci. USA* **116**, 12615–12623 (2019).
- Barrett, R., Kuzawa, C. W., McDade, T. & Armelagos, G. J. Emerging and re-emerging infectious diseases: the third epidemiologic transition. *Annu. Rev. Anthropol.* **27**, 247–271 (1998).
- Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).
- Key, F. M., Posth, C., Krause, J., Herbig, A. & Bos, K. I. Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends Genet.* **33**, 508–520 (2017).
- Vägene, A. J. et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
- Zhou, Z. et al. Pan-genome Analysis of Ancient and Modern *Salmonella enterica* Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Curr. Biol.* **28**, 2420–2428.e10 (2018).
- Alikhan, N.-F., Zhou, Z., Sergeant, M. J. & Achtman, M. A genomic overview of the population structure of *Salmonella*. *PLoS Genet.* **14**, e1007261 (2018).
- Kirk, M. D. et al. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *PLoS Med.* **12**, e1001921 (2015).
- Kingsley, R. A. & Bäumler, A. J. Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Mol. Microbiol.* **36**, 1006–1014 (2000).
- Kingsley, R. A. et al. Epidemic multiple drug resistant *Salmonella typhimurium* causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res.* **19**, 2279–2287 (2009).
- Barrow, P. A. & Methner, U. *Salmonella in Domestic Animals* (CABI, 2013).
- Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O. & Raoult, D. Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. *Proc. Natl Acad. Sci. USA* **95**, 12637–12640 (1998).
- Anthony, D. W. *The Horse, The Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, 2010).
- Schulting, R. J. & Richards, M. P. in *A Bronze Age Landscape in the Russian Steppes. The Samara Valley Project* (eds Anthony, D. W. et al.) 127–149 (Cotsen Institute of Archaeology Press, 2016).
- Didelot, X. et al. Recombination and population structure in *Salmonella enterica*. *PLoS Genet.* **7**, e1002191 (2011).
- Haase, J. K. et al. Population genetic structure of 4,12:a– *Salmonella enterica* strains from harbor porpoises. *Appl. Environ. Microbiol.* **78**, 8829–8833 (2012).
- Uzzau, S. et al. Host-adapted serotypes of *Salmonella enterica*. *Epidemiol. Infect.* **125**, 229–255 (2000).
- Taylor, J. & Douglas, S. H. *Salmonella birkenhead*: a new *Salmonella* type causing food poisoning in man. *J. Clin. Pathol.* **1**, 237–239 (1948).
- Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
- Hedge, J. & Wilson, D. J. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio* **5**, e02158–02114 (2014).
- Duchêne, S., Duchêne, D., Holmes, E. C. & Ho, S. Y. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol. Biol. Evol.* **32**, 1895–1906 (2015).
- Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
- Rhen, M. & Mastroeni, P. *Salmonella: Molecular Biology and Pathogenesis* (Horizon Scientific Press, 2007).
- Guiney, D. G. & Fierer, J. The role of the spv genes in *Salmonella* pathogenesis. *Front. Microbiol.* **2**, 129 (2011).
- Gulig, P. A. et al. Molecular analysis of spv virulence genes of the salmonella virulence plasmids. *Mol. Microbiol.* **7**, 825–830 (1993).
- Rötger, R. & Casadesús, J. The virulence plasmids of *Salmonella*. *Int. Microbiol.* **2**, 177–184 (1999).
- Hackett, J., Wyk, P., Reeves, P. & Mathan, V. Mediation of serum resistance in *Salmonella typhimurium* by an 11-kilodalton polypeptide encoded by the cryptic plasmid. *J. Infectious Dis.* **155**, 540–549 (1987).
- Langridge, G. C. et al. Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. *Proc. Natl Acad. Sci. USA* **112**, 863–868 (2015).
- Liu, W. Q. et al. *Salmonella paratyphi C*: genetic divergence from *Salmonella choleraesuis* and pathogenic convergence with *Salmonella typhi*. *PLoS ONE* **4**, e4510 (2009).
- Thomson, N. R. et al. Comparative genome analysis of *Salmonella enteritidis* PT4 and *Salmonella gallinarum* 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res.* **18**, 1624–1637 (2008).
- Parkhill, J. et al. Complete genome sequence of a multiple drug-resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848 (2001).
- Lee, S.-J. et al. Identification of a common immune signature in murine and human systemic salmonellosis. *Proc. Natl Acad. Sci. USA* **109**, 4998–5003 (2012).
- Seth-Smith, H. M. SPI-7: *Salmonella*'s Vi-encoding pathogenicity island. *J. Infect. Dev. Cries* **2**, 267–271 (2008).
- Omran, A. R. The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Q.* **83**, 731–757 (2005).
- Pinhasi, R. & Stock, J. T. *Human Bioarchaeology of the Transition to Agriculture* (John Wiley & Sons, 2011).
- Miller, L. & Hurley, K. *Infectious Disease Management in Animal Shelters* (John Wiley & Sons, 2009).
- Schuster, C. J. et al. Infectious disease outbreaks related to drinking water in Canada, 1974–2001. *Can. J. Public Health* **96**, 254–258 (2005).
- Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
- Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
- Gansauge, M.-T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).
- Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).

46. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
47. Hübner, R. et al. HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* **20**, 280 (2019).
48. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
50. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
51. Kircher, M. in *Ancient DNA. Methods in Molecular Biology (Methods and Protocols)* Vol. 480 (eds Shapiro, B. & Hofreiter, M.) 197–228 (Humana Press, 2012).
52. Jolley, K. A. et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* **158**, 1005–1015 (2012).
53. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
54. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
55. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
56. Moran, A. B. & Edwards, P. Three new *Salmonella* types: *S. richmond*, *S. daytona* and *S. tallahassee*. *Proc. Soc. Exp. Biol. Med.* **62**, 294–296 (1946).
57. Van der Walt, M. L., Huchzermeyer, F. & Steyn, H. C. *Salmonella* isolated from crocodiles and other reptiles during the period 1985–1994 in South Africa. *Onderstepoort J. Vet. Res.* **64**, 277–283 (1997).
58. Paton, J. & Mirfattahi, M. *Salmonella* meningitis acquired from pet snakes. *Arch. Dis. Child.* **77**, 91 (1997).
59. Pedersen, K., Sørensen, G., Szabo, I., Hächler, H. & Le Hello, S. Repeated isolation of *Salmonella enterica* Goverdhan, a very rare serovar, from Danish poultry surveillance samples. *Vet. Microbiol.* **174**, 596–599 (2014).
60. Sharma, V., Rohde, R., Garg, D. & Kumar, A. Toads as natural reservoir of salmonella. *Zentralbl. Bakteriol. Orig. A* **239**, 172–177 (1977).
61. Sharma, V. & Singh, C. *Salmonella goverdhan*, a new serotype from sewage. *Int. J. Syst. Evol. Microbiol.* **17**, 41–42 (1967).
62. Zhou, Z. et al. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* genomic diversity. *Genome Res.* **30**, 138–152 (2020).
63. Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B* **255**, 37–45 (1994).
64. Zhou, Z. et al. Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years of global spread of enteric fever. *Proc. Natl Acad. Sci. USA* **111**, 12199–12204 (2014).
65. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
66. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
67. Vianello, D. et al. HAPLOFIND: a new method for high-throughput mtDNA Haplotype assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).
68. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
69. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
70. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
71. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
72. Ramsden, C. et al. High rates of molecular evolution in Hantaviruses. *Mol. Biol. Evol.* **25**, 1488–1492 (2008).
73. Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl Acad. Sci. USA* **110**, 228–233 (2013).
74. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
75. Yoon, S. H., Park, Y.-K. & Kim, J. F. PAIDB v2. 0: exploration and analysis of pathogenicity and resistance islands. *Nucleic Acids Res.* **43**, D624–D630 (2014).
76. Fuentes, J. A., Villagra, N., Castillo-Ruiz, M. & Mora, G. C. The *Salmonella* Typhi hlyE gene plays a role in invasion of cultured epithelial cells and its functional transfer to *S. Typhimurium* promotes deep organ infection in mice. *Res. Microbiol.* **159**, 279–287 (2008).
77. Vernikos, G. S. & Parkhill, J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* **22**, 2196–2203 (2006).
78. Blondel, C. J., Jiménez, J. C., Contreras, I. & Santiviago, C. A. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics* **10**, 354 (2009).
79. Elder, J. R. et al. The *Salmonella* pathogenicity island 13 contributes to pathogenesis in streptomycin pre-treated mice but not in day-old chickens. *Gut Pathog.* **8**, 16 (2016).
80. Shah, D. H. et al. Identification of *Salmonella gallinarum* virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. *Microbiology* **151**, 3957–3968 (2005).
81. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
82. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
83. McKenna, A. et al. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
84. Fellows Yates, J. A. et al. Central European woolly mammoth population dynamics: insights from Late Pleistocene mitochondrial genomes. *Sci. Rep.* **7**, 17714 (2017).

Acknowledgements

We thank the Pathogenomics group at MPI SHH and the Lieberman laboratory at MIT for critical discussions. We thank M. O'Reilly for support in graphical design. Funding for this study came from the Max Planck Society and the European Research Council (grant agreement no. 771234 PALEoRIDER). F.M.K. was supported by DFG (grant no. KE 2408/1-1) and a generous MPI EVA guest office. I.S. is supported by SNF (grant no. CR31I3L_157024). Enterobase was supported by BBSRC (no. BB/L020319/1). N.-F.A., Z.Z. and M.A. were supported by the Wellcome Trust (grant no. 202792/Z/16/Z).

Author contributions

F.M.K., A. Herbig and J.K. conceptualized the study. The original draft was written by F.M.K., which was then reviewed and edited by F.M.K., C.P., K.B., S.R., D.K., W.H., M.A., A. Herbig and J.K. F.M.K., C.P., L.R.E.-G. and D.K. carried out the formal analysis. F.M.K., C.P., M.A.S., G.U.N., A.F., S.S., M.B., A.W., M.M., S.N., R.T. and Z.Z. handled the investigation. F.M.K., R.H., Å.J.V., A.K.L., A. Khokhlov, A.C., S.H., A.B.B., A. Kalmykov, A.R.K., V.E.M., P.W.S., S.V., M.Z., A.R., D.C., R.S., J.B., M.G.G., N.S., A. Hafner, M.R., I.S., S.L., Y.S.E., N.-F.A., Z.Z., M.A., S.R. and W.H. contributed resources. A. Herbig and J.K. supervised.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41559-020-1106-9>.

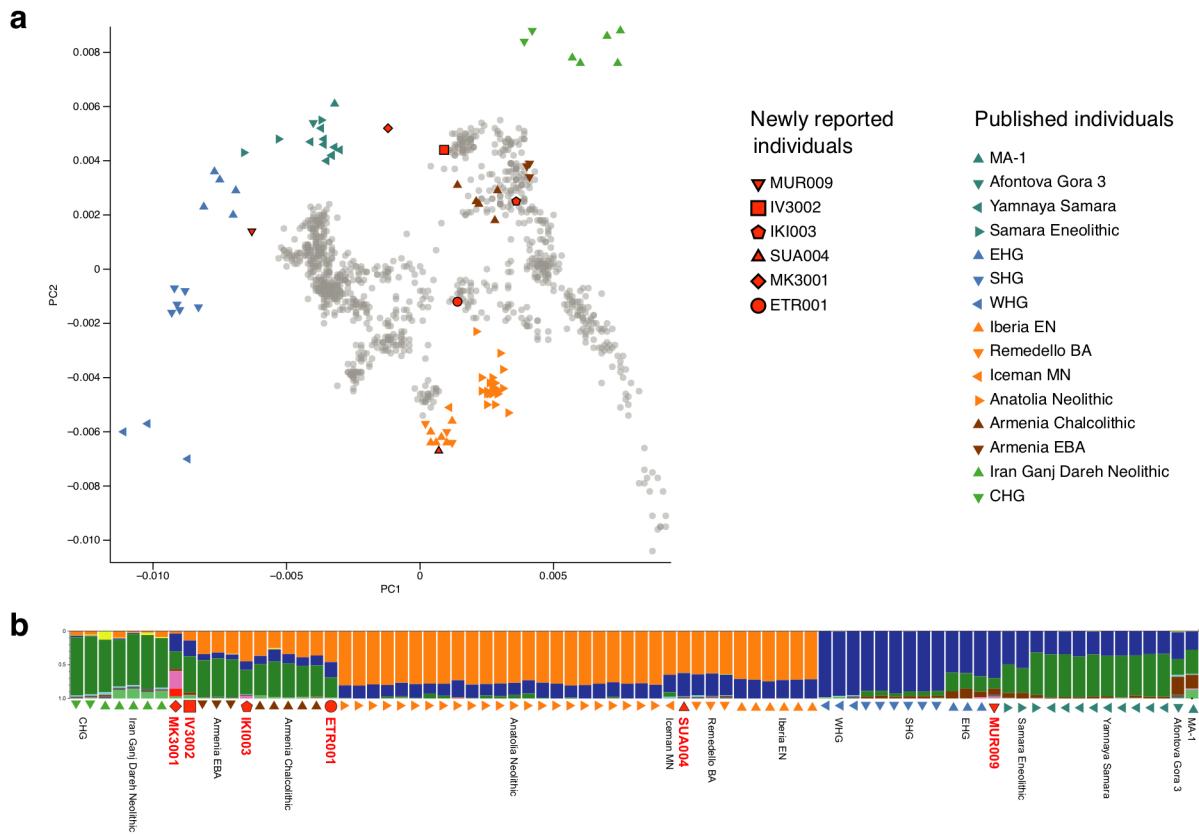
Supplementary information is available for this paper at <https://doi.org/10.1038/s41559-020-1106-9>.

Correspondence and requests for materials should be addressed to F.M.K., A. Herbig or J.K.

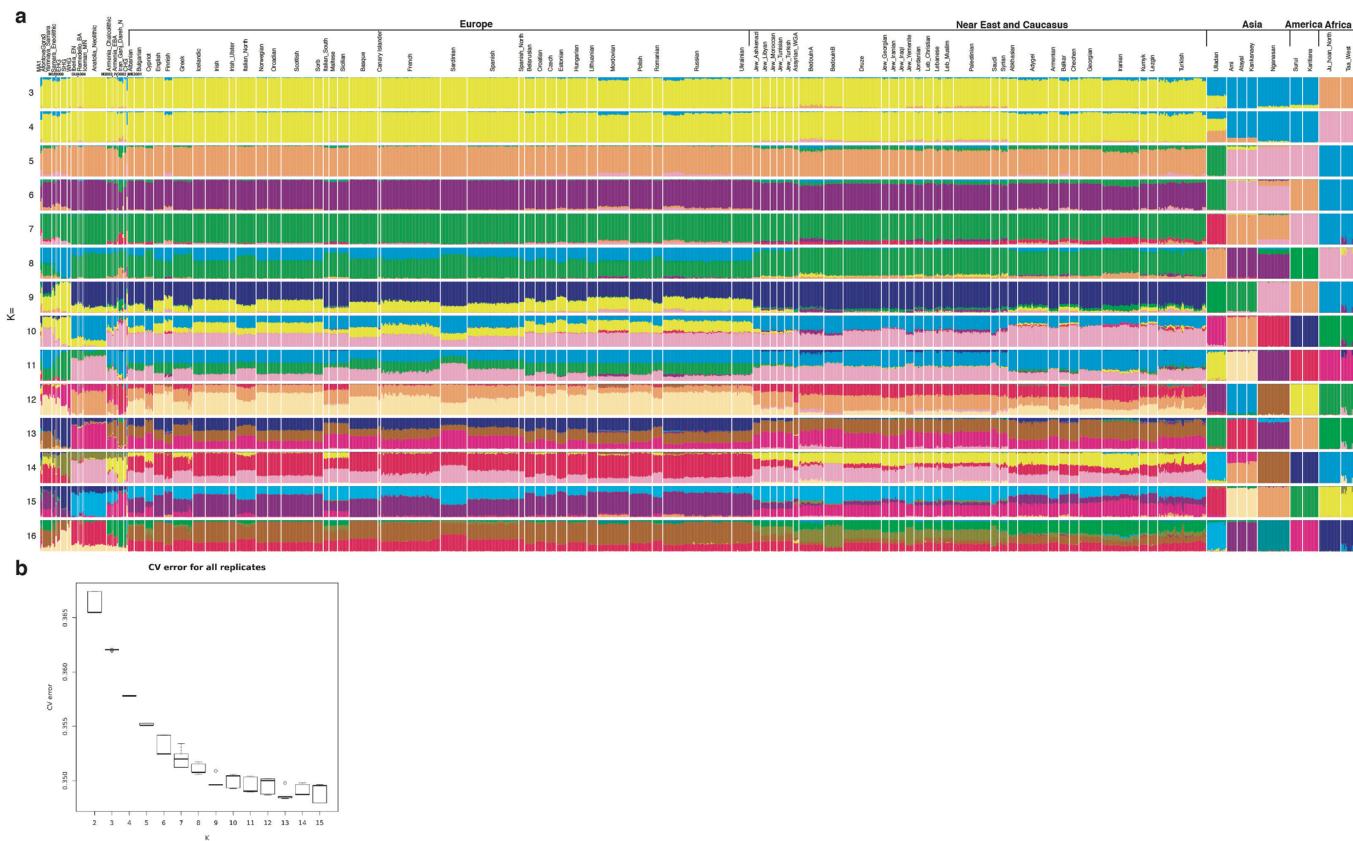
Reprints and permissions information is available at www.nature.com/reprints.

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

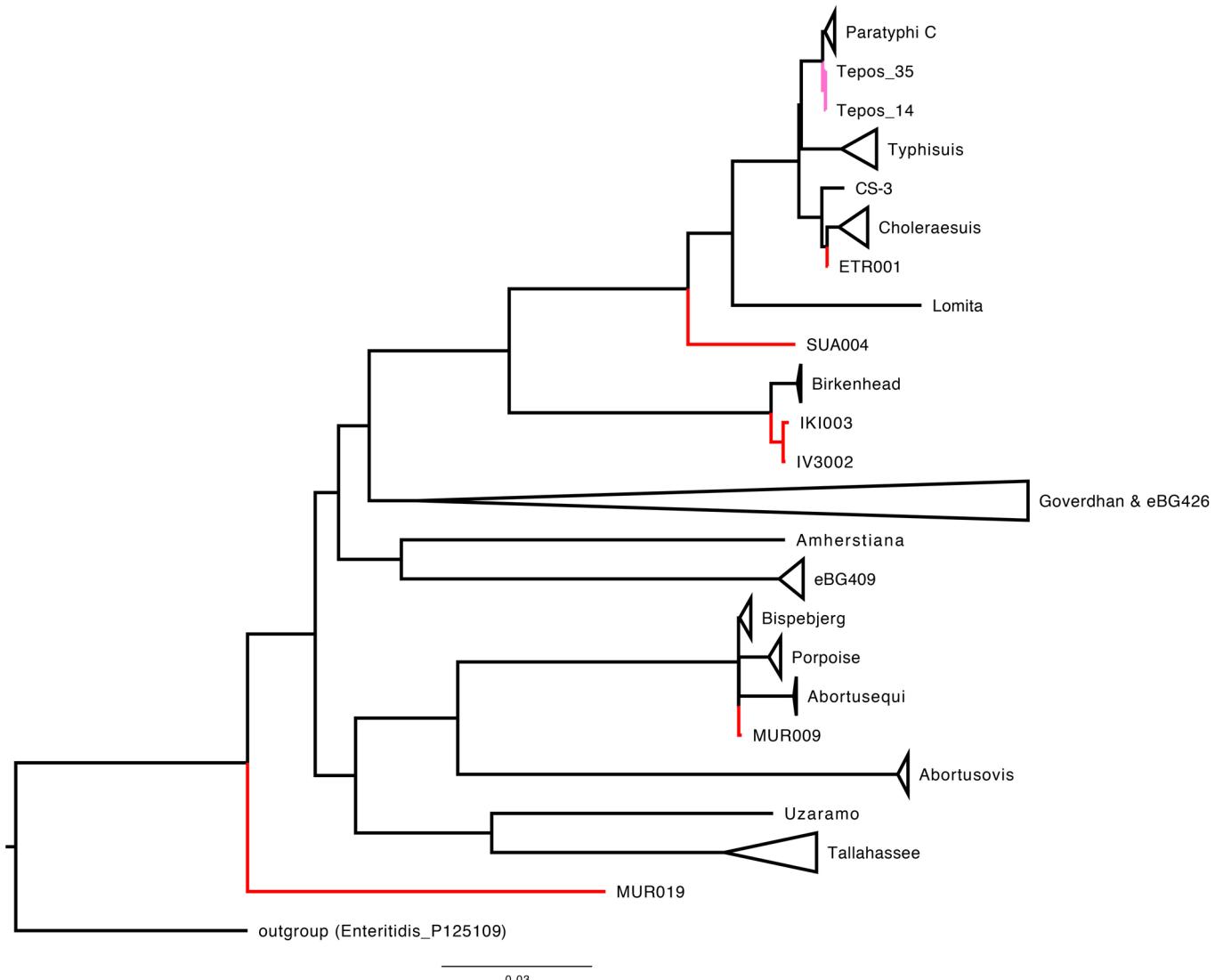
© The Author(s), under exclusive licence to Springer Nature Limited 2020



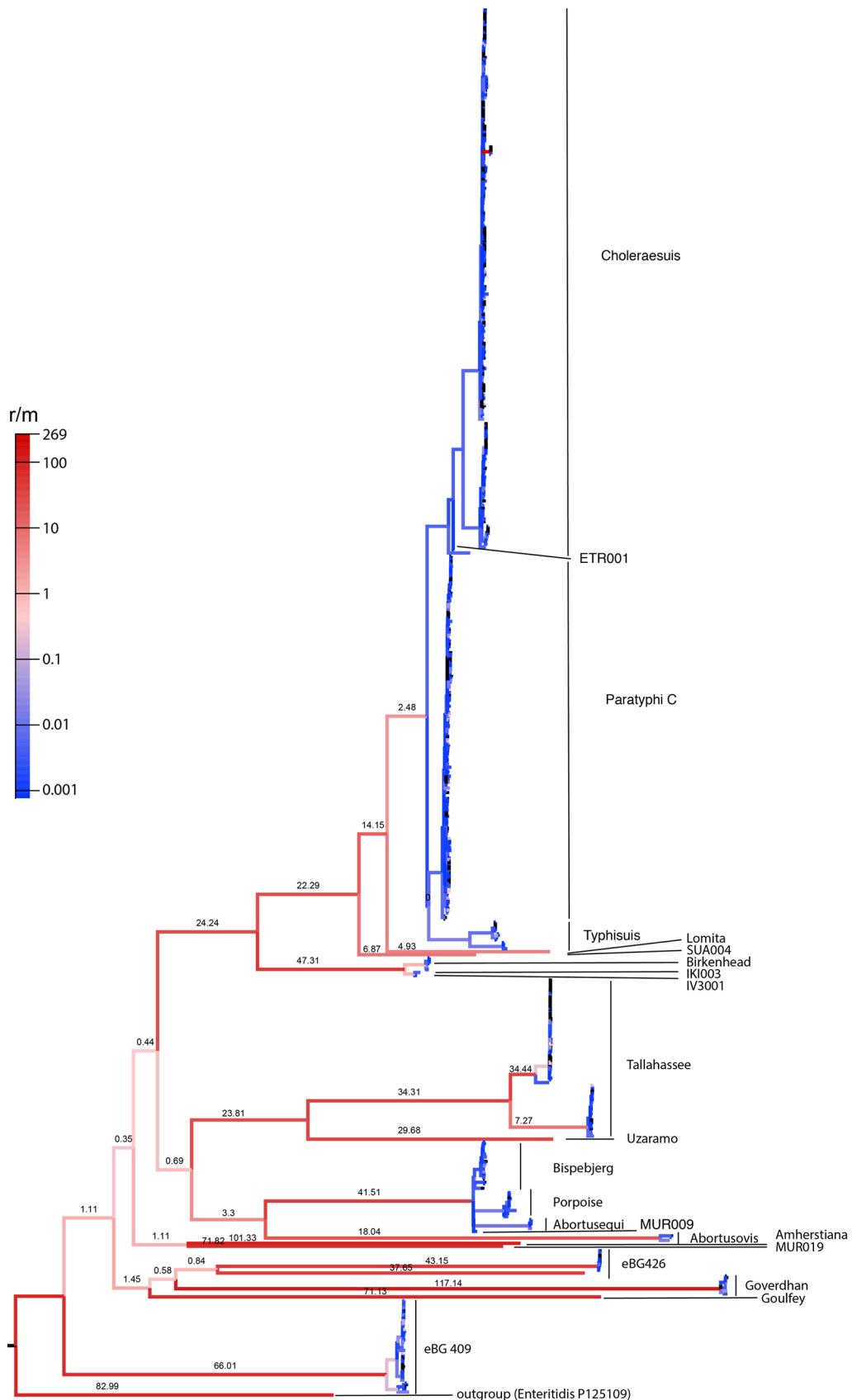
Extended Data Fig. 1 | Ancient human population genetic analysis. **a**, PCA of newly reported ancient individuals with sufficient data (in red) and selected published ancient and modern individuals are projected onto principal components built with present-day West Eurasian populations (grey dots). **b**, ADMIXTURE analysis ($K=10$) of newly reported ancient individuals and relevant published ancient and modern individuals sorted by genetic clusters. Overview ancient human genetic data Supplementary Table 1 and further analysis Extended Data Fig. 2. EHG, Eastern hunter gatherer; E, Early; M, Middle; HG, hunter-gatherer; N, Neolithic; C, Caucasus; S, Scandinavian; W, Western; BA, Bronze Age.



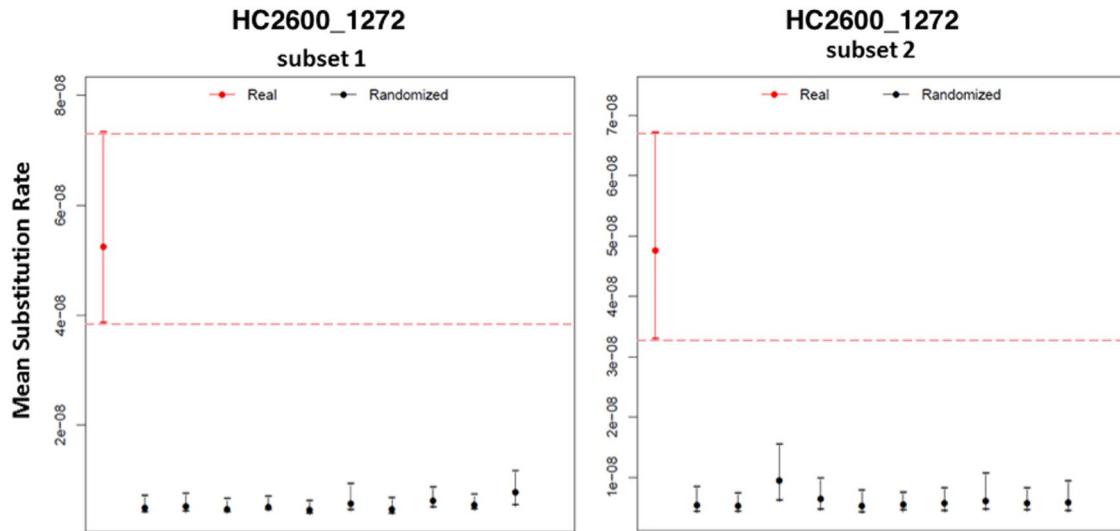
Extended Data Fig. 2 | Summary human genetic analysis. **a**, ADMIXTURE analysis ($K = 3 - 16$) of newly reported ancient individuals (bold horizontal text) and published ancient and modern individuals sorted by genetic clusters and geographic origin (Europe, Near East and Caucasus, Asia, America, Africa). Each K was run five times and the replicate with the highest likelihood is reported. Ancient MK3001 shows Asian genetic ancestry components represented by Nganasan, Kankanaey, Atayal, and Ami. **b**, Box plot of five cross-validations (CV) values for every K calculated in ADMIXTURE. EHG, Eastern hunter gatherer; E, Early; M, Middle; HG, hunter-gatherer; N, Neolithic; C, Caucasus; S, Scandinavian; W, Western; BA, Bronze Age.



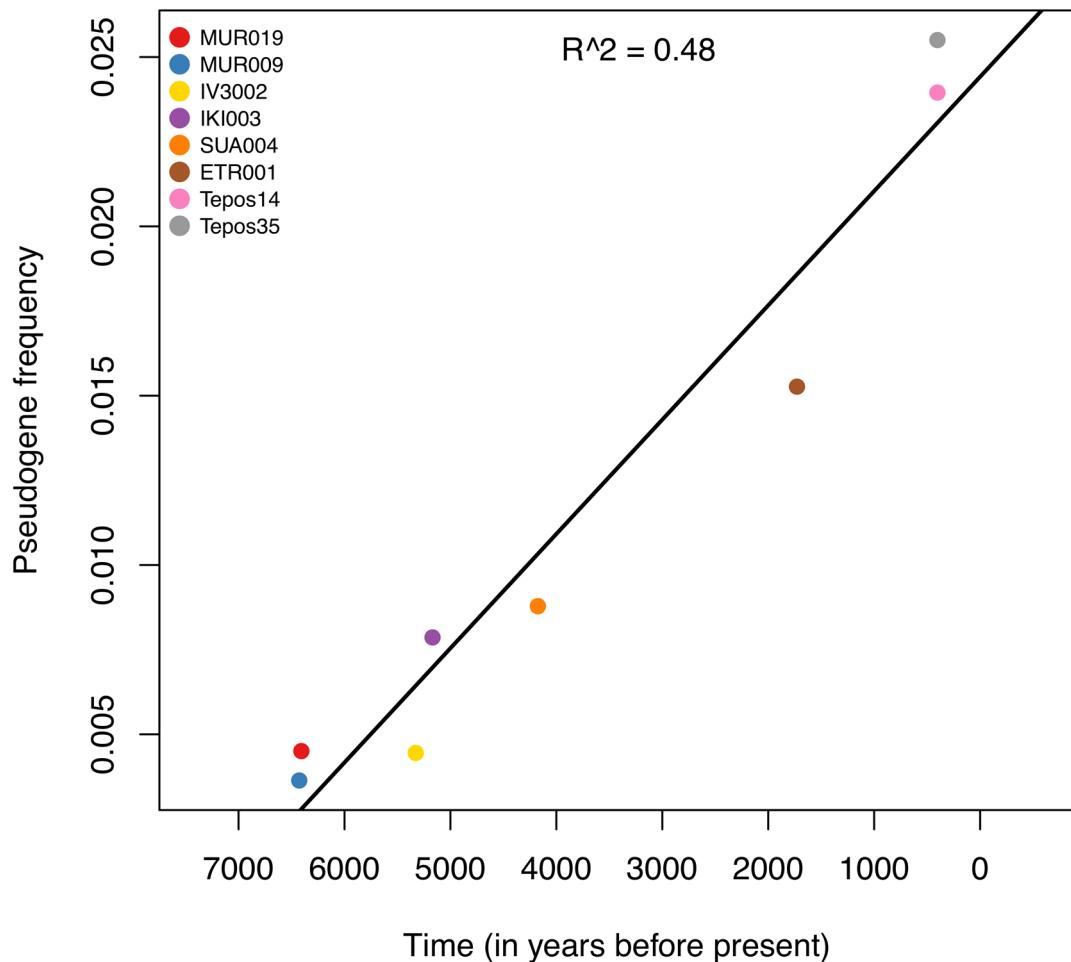
Extended Data Fig. 3 | Maximum likelihood phylogeny of the AESB based on SNPs in positions present in 95% of strains. Maximum likelihood tree of the AESB including the high coverage ancient genomes and 463 *S. enterica* genomes, considering all SNPs covered in at least 95% of strains (130,036 SNPs). New ancient genomes are shown in red, and previously reported ancient genomes (Tepos) in pink.



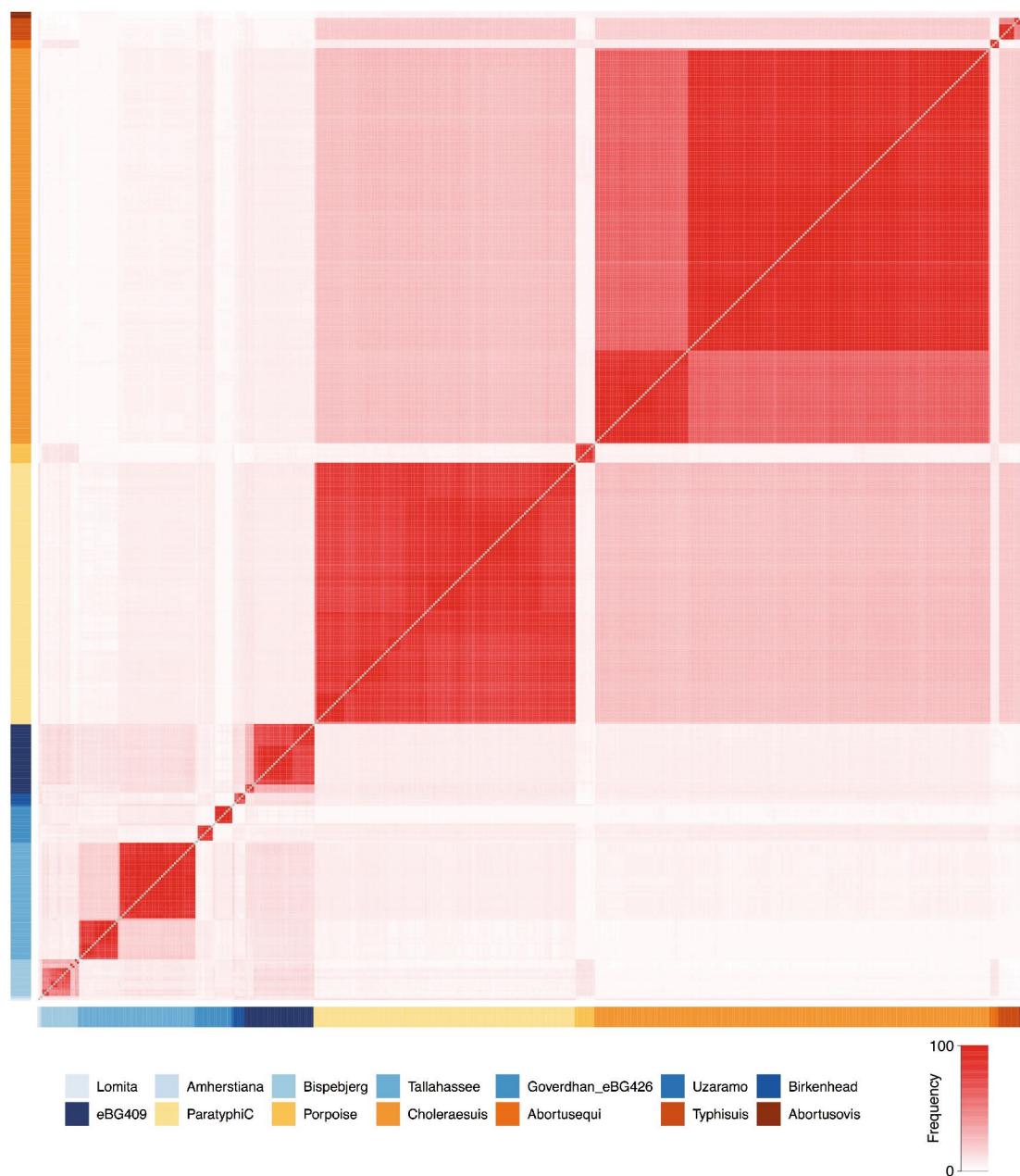
Extended Data Fig. 4 | Recombination rate estimates for the AESB. Estimated recombination rate is shown as recombination event per mutation event (r/m) and indicated on top of branch and by branch color. Recombination events have been inferred using all positions shared by 95% of strains from the AESB and are here reported for the SNPs shared by all strains on the AESB (correspond to maximum likelihood phylogeny shown in Fig. 2b). Maximum likelihood tree including all SNPs shared by at least 95% of strains from the AESB is shown in Extended Data Fig. 3.



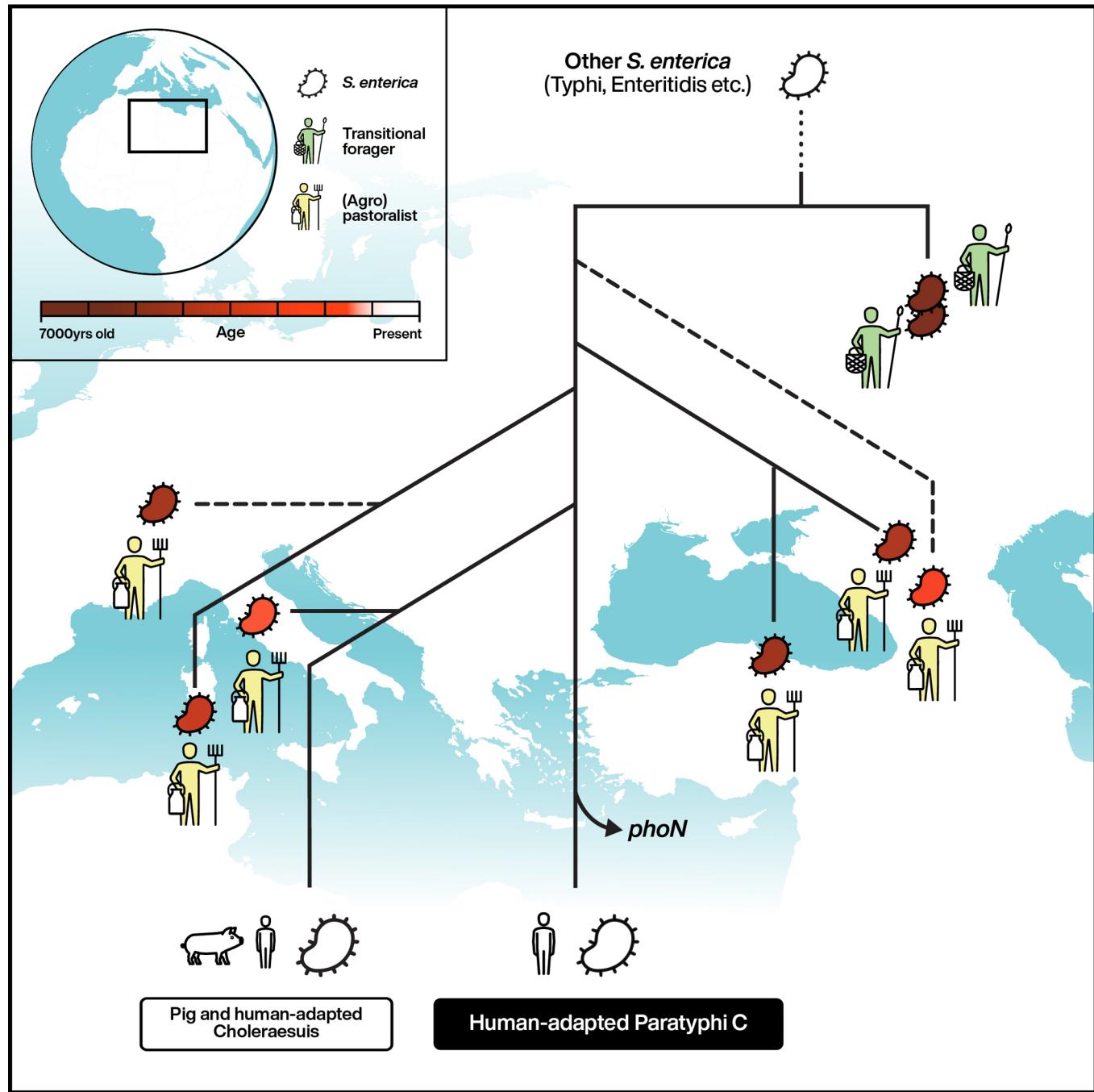
Extended Data Fig. 5 | Temporal signal analysis. Results of the date randomization test for two subsets of the HC2600_1272 cluster. Circles represent mean substitution rate estimations with error bars representing 95% highest posterior density (HPD) intervals. For each subset 10 date randomizations were done. Significant temporal signal is indicated by non-overlapping HPD intervals between real data (red) and the randomizations (black), which is the case for both subsets.



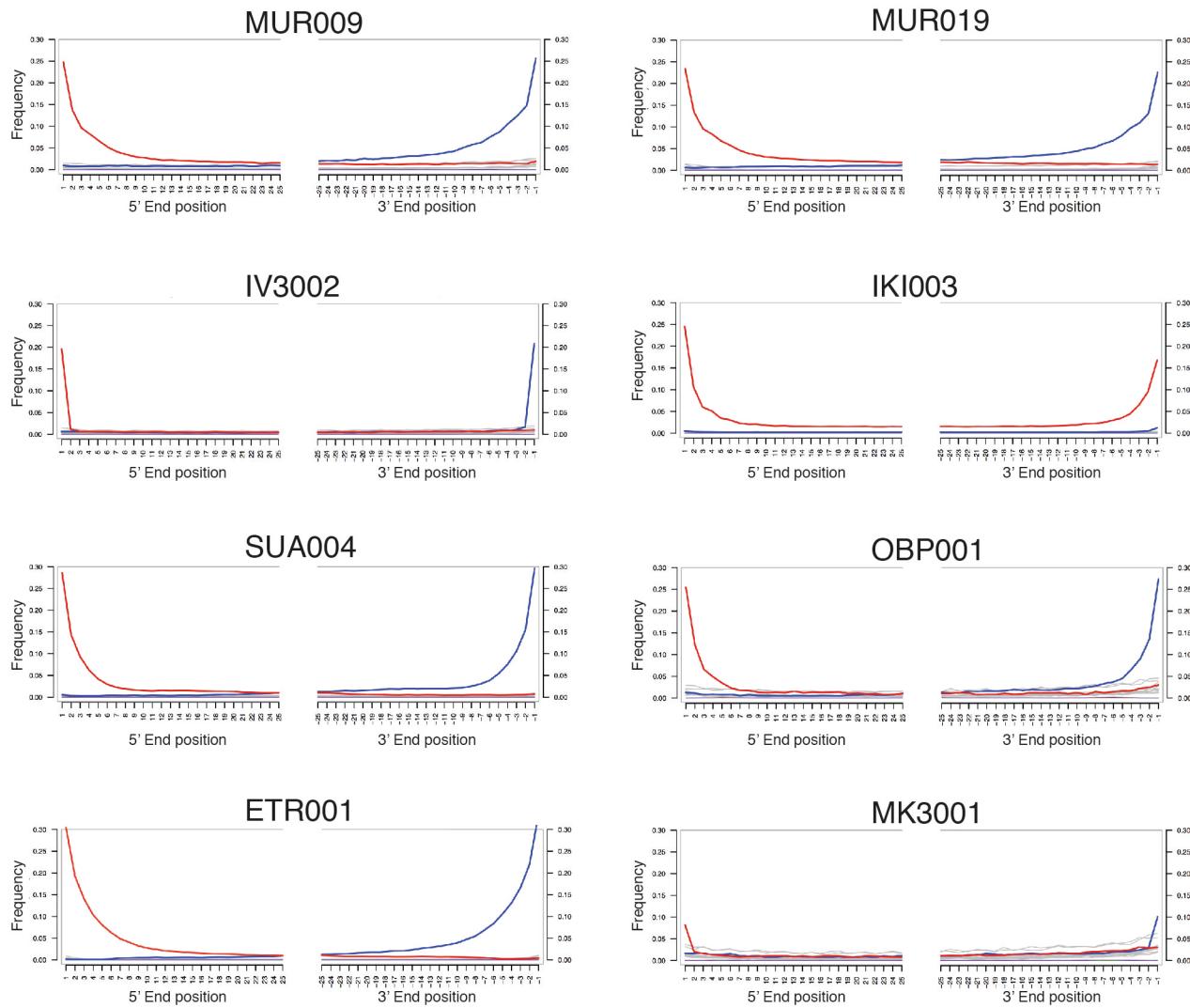
Extended Data Fig. 6 | Correlation between pseudogene frequency and time for all ancient genomes with mean genome-wide coverage above 5X.



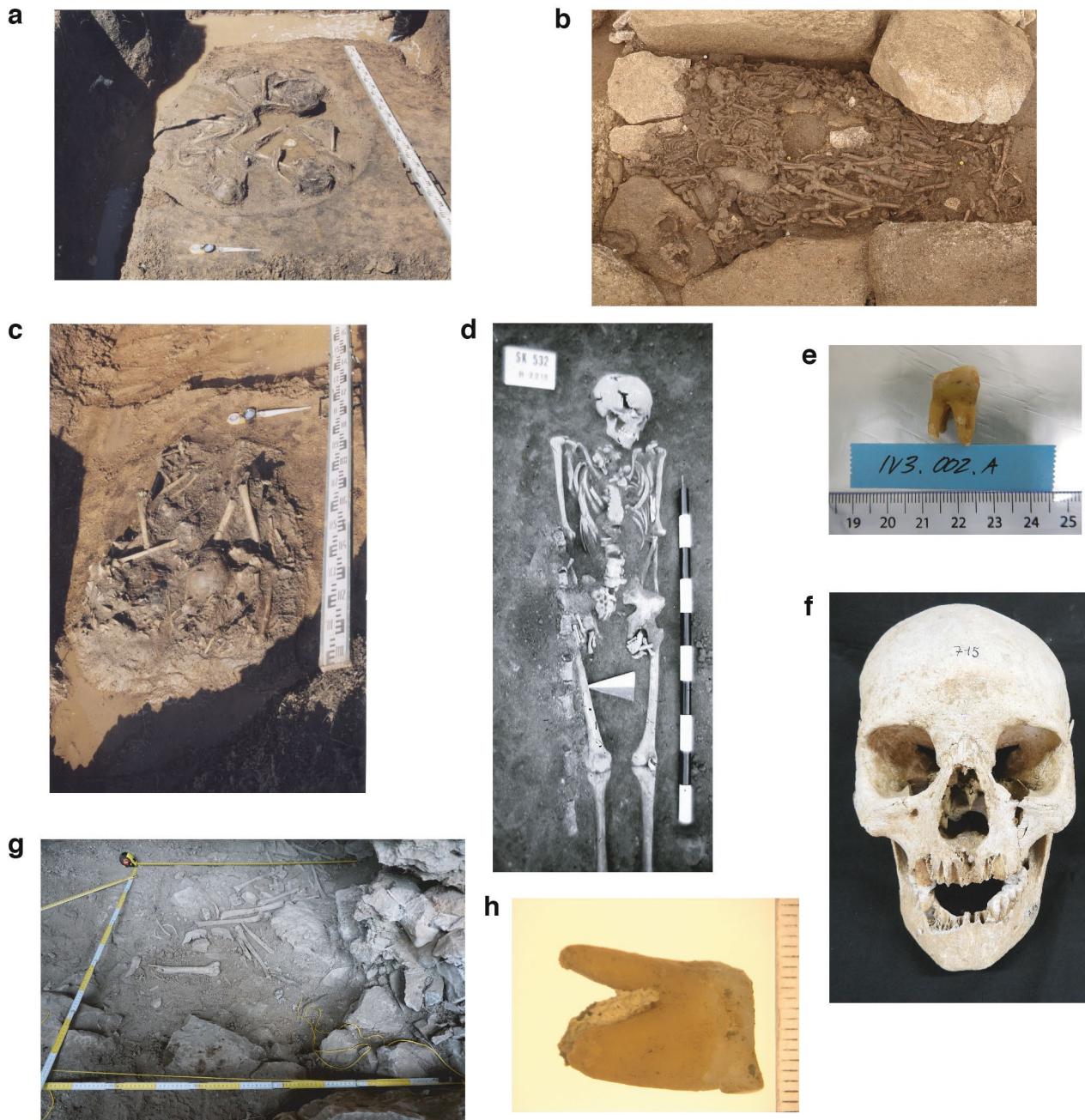
Extended Data Fig. 7 | Proportion of shared pseudogenes between strains across the AESB. Proportion of pseudogene-sharing (0–100%) between strains on the AESB is shown in tones of red. Strains are ordered by phylogenetic branch and coloured accordingly.



Extended Data Fig. 8 | Graphical abstract.



Extended Data Fig. 9 | Mismatch distribution along positions at the 5'- and 3'- end of mapped sequencing reads. C to T changes indicated in red and G to A changes in blue, all other substitutions in grey. IV3002 and MK3001 are UDG-half treated, which leads to observable damage only in the terminal positions. Plots generated with mapDamage2 (Jónsson H. et al., Bioinformatics 2013).



Extended Data Fig. 10 | Photographs of archaeological specimens that harboured ancient *S. enterica* DNA. (a) MUR009; (b) OBP001; (c) MUR019; (d) IKI003; (e) IV3002; (f) ETRO01; (g) SUA004; (h) MK3001.

Corresponding author(s): Felix M. Key

Last updated by author(s): Oct 25, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection.

Data analysis

Intcal (v13), SwissCal (v1.0), MALT (v 0.38), EAGER (v1.92.55), Multivcfanalyzer (v0.87), RAxML (v8.2.9), mapDamage (v2.0), schmutzi (v1.0), HaploGrep (v17), bwa (v0.7.12), angsd (v0.910), smartpca (v16000), admixture (v1.3.0), TempEst (v1.5), BEAST2 (2.5.0), LogCombiner (v2.5.0), Tracer (v1.6.0), GATK (v3.5), GenConS (v1.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw sequencing data will be made available upon acceptance for publication.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was required. All samples with sufficient ancient <i>S. enterica</i> DNA were analyzed. For initial placement of all ancient genomes we used 2,961 modern <i>S. enterica</i> genomes representing the known genetic diversity. A total of 469 modern <i>S. enterica</i> genomes were used for detailed analysis of the super branch containing all ancient <i>S. enterica</i> genomes. Human genetic analysis was done for all <i>S. enterica</i> positive samples that yielded sufficient ancient human DNA (mitochondrial or nuclear). For analysis we used the Human Origins reference dataset comprising around 7,000 ancient and modern human individuals.
Data exclusions	For phylogenetic, molecular dating and pseudogenization analysis of the ancient <i>S. enterica</i> we required a minimum genome-wide coverage of 5x (based on Paratyphi C RKS4594 -- NC_012125). Molecular dating was done for two random subsets of the pastoralist branch.
Replication	Robustness of the phylogenetic inference (maximum likelihood) was tested with bootstrapping (100 or 1,000 replicates). Temporal signal for molecular dating was present in both subsets of the pastoralist branch and confirmed using a date randomization test (10 replicates).
Randomization	Random subsets of the pastoralist branch were restraint to contain at least one modern genome of each major branch.
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging