

## RESEARCH ARTICLE

A 3,000-year-old, basal *S. enterica* lineage from Bronze Age Xinjiang suggests spread along the Proto-Silk Road

Xiyan Wu<sup>1,2</sup>, Chao Ning<sup>3</sup>, Felix M. Key<sup>3,4</sup>, Aida Andrades Valtueña<sup>3,5</sup>, Aditya Kumar Lankapalli<sup>3</sup>, Shizhu Gao<sup>6</sup>, Xuan Yang<sup>1</sup>, Fan Zhang<sup>1</sup>, Linlin Liu<sup>7</sup>, Zhongzhi Nie<sup>8</sup>, Jian Ma<sup>9</sup>, Johannes Krause<sup>3,5\*</sup>, Alexander Herbig<sup>3,5\*</sup>, Yinqiu Cui<sup>1,8,10\*</sup>

**1** School of Life Sciences, Jilin University, Changchun, China, **2** School of History and Culture, Henan University, Kaifeng, China, **3** Max Planck Institute for the Science of Human History, Jena, Germany, **4** Max Planck Institute for Infection Biology, Berlin, Germany, **5** Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, **6** College of Pharmacia Sciences, Jilin University, Changchun, China, **7** Department of Radiation Oncology, The Second Hospital of Jilin University, Changchun, China, **8** Research Center for Chinese Frontier Archaeology, Jilin University, Changchun, China, **9** School of Cultural Heritage, Northwest University, Xi'an, China, **10** Key Laboratory for Evolution of Past Life and Environment in Northeast Asia (Jilin University), Ministry of Education, Changchun, China

\* [krause@shh.mpg.de](mailto:krause@shh.mpg.de) (JK); [herbig@shh.mpg.de](mailto:herbig@shh.mpg.de) (AH); [cuiyq@jlu.edu.cn](mailto:cuiyq@jlu.edu.cn) (YCQ)



## OPEN ACCESS

**Citation:** Wu X, Ning C, Key FM, Andrades Valtueña A, Lankapalli AK, Gao S, et al. (2021) A 3,000-year-old, basal *S. enterica* lineage from Bronze Age Xinjiang suggests spread along the Proto-Silk Road. *PLoS Pathog* 17(9): e1009886. <https://doi.org/10.1371/journal.ppat.1009886>

**Editor:** Renée M. Tsois, University of California, Davis, UNITED STATES

**Received:** February 21, 2021

**Accepted:** August 11, 2021

**Published:** September 21, 2021

**Copyright:** © 2021 Wu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequencing data are available from the NCBI Sequence Archive (SRA) under Bioproject ID: PRJNA730407 or from the following link: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA730407>.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No. 42072018) to Y.C., the Major project of Humanities and Social Sciences Key Research Base of the Ministry of Education (Grant No. 16JJD780005) to Y.C. and the Science and technology development

## Abstract

*Salmonella enterica* (*S. enterica*) has infected humans for a long time, but its evolutionary history and geographic spread across Eurasia is still poorly understood. Here, we screened for pathogen DNA in 14 ancient individuals from the Bronze Age Quanergou cemetery (XBQ), Xinjiang, China. In 6 individuals we detected *S. enterica*. We reconstructed *S. enterica* genomes from those individuals, which form a previously undetected phylogenetic branch basal to Paratyphi C, Typhisuis and Choleraesuis—the so-called Para C lineage. Based on pseudogene frequency, our analysis suggests that the ancient *S. enterica* strains were not host adapted. One genome, however, harbors the *Salmonella* pathogenicity island 7 (SPI-7), which is thought to be involved in (para)typhoid disease in humans. This offers first evidence that SPI-7 was acquired prior to the emergence of human-adapted Paratyphi C around 1,000 years ago. Altogether, our results show that *Salmonella enterica* infected humans in Eastern Eurasia at least 3,000 years ago, and provide the first ancient DNA evidence for the spread of a pathogen along the Proto-Silk Road.

## Author summary

Recent studies that use DNA extracted from ancient human remains have shown that the bacterial pathogen *Salmonella enterica* has infected humans since prehistoric times. Expanding the knowledge about its geographic spread and genetic diversity we reconstructed six ancient *Salmonella* genomes from individuals excavated at the Bronze Age Quanergou cemetery in Xinjiang, China. Our analysis shows that this form of *Salmonella* was closely related to a type that today specifically infects humans and/or pigs. However, based on our genetic assessment we determined that our ancient strains were not host

project of Jilin Province (Grant No. 20190701077GH) to Y.C. C.N. was supported by the European Research Council (Grant No. 646612) granted to Martine Robbeets. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

adapted yet. Furthermore, we could show that in the course of its evolution the bacterium acquired some important virulence factors earlier than previously thought. Interpreting our results within an archaeological context we suggest that in Bronze Age Eastern Eurasia the spread of this pathogen was likely promoted through trade networks referred to as the Proto-Silk Road.

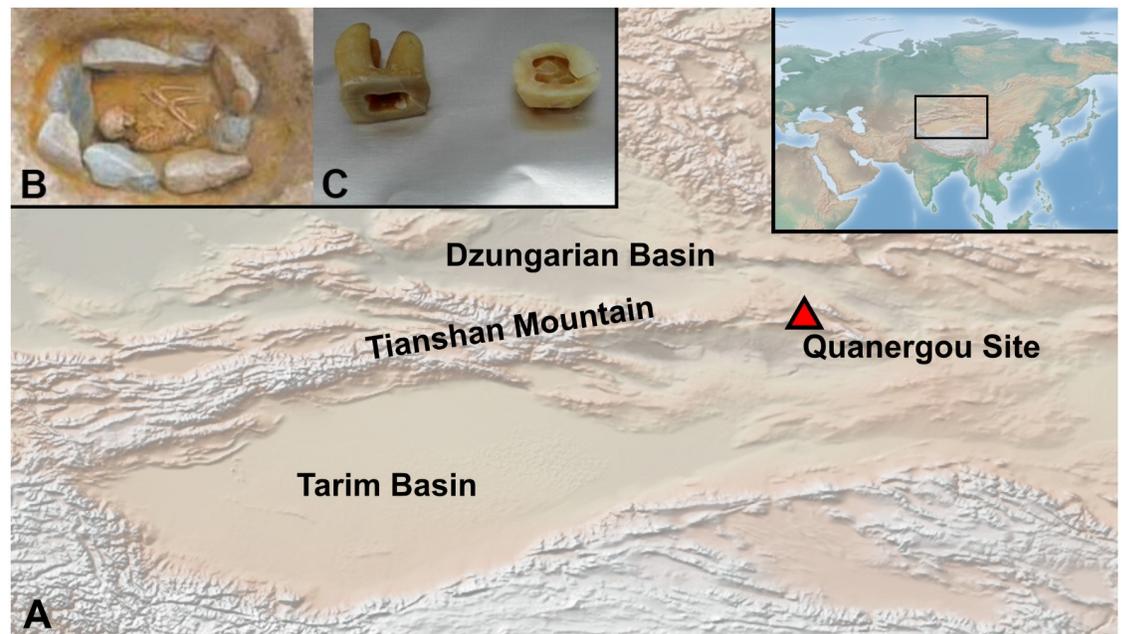
## Introduction

*Salmonella enterica* (*S. enterica*) is a Gram-negative pathogenic bacterium that can infect a wide variety of warm-blooded hosts and causes enteric fever and gastroenteritis [1]. On a global scale, at least 170 million cases of salmonellosis occur each year, resulting in up to 1 million deaths [2,3]. Traditionally, its diversity has been described based on serological differences, and approximately 2,600 *Salmonella* serovars have been identified so far [4]. Despite its worldwide distribution today, little is known about its spread and contribution to prehistoric epidemic events.

Ancient pathogen genomes are crucial for studying pathogen evolution, calibrating the molecular clock and providing time transect data to illustrate phylogenetic diversity [5]. With the development of ancient DNA technologies, pathogen genomes can be retrieved from archaeological remains, which enables us to reconstruct the phylogenetic history of pathogens and identify the causative agents of past epidemic events. Human-adapted Paratyphi C has been reconstructed from a 13th-century skeleton from Norway, providing direct evidence for the presence of this lineage in Europe about 800 years ago [6]. Vågene et al. recovered Paratyphi C genomes from a 16th-century colonial era Mexican site suggesting that this pathogen may have contributed to the collapse of native populations after its introduction to the New World [7]. Key et al. recovered eight diverse ancient *S. enterica* genomes across Western Eurasia, suggesting that the emergence of human-adapted *S. enterica* Paratyphi C was linked to the Neolithization process [8]. However, whether and when *Salmonella* had appeared in Eastern Eurasia or where and how the human-adapted Paratyphi C evolved are still largely unknown.

Geographically located in northwestern China and part of the “Silk Road”, Xinjiang has been a hub connecting eastern and western Eurasia since the Bronze Age [9]. Starting from the early Bronze Age, the Eurasian Steppe (ES) had witnessed several major cultural changes and large-scale population movements [10–13]. Millet of East Asian origin spread westward into Europe, and conversely, wheat and barley of the Near East origin spread eastward into East Asia, perhaps via what was known as the Inner Asian Mountain Corridor along the Tianshan Mountains in Xinjiang [14]. The ES pastoralists may have served as an important agent for such cereal globalization [15–17]. Furthermore, DNA studies of ancient human remains from Xinjiang suggested that populations in this region were already admixed between eastern and western Eurasians since the Bronze Age [18–20], pointing out that the Xinjiang region had served as a main corridor for trans-Eurasian contacts and likewise the transmission of certain pathogens [21,22].

The XBQ site is located in Xinjiang, to the eastern part of the Tianshan Mountains (Fig 1). The cemetery was dated to the Bronze Age and likely used by mobile pastoralists [23]. Multiple simultaneous burials were excavated from the cemetery, showing a high mortality rate for young adults and in particular children and infants without obvious skeletal trauma [23,24], which has not been observed elsewhere in the region [25]. Until now, no consensus has been reached among archaeologists on how to explain this unusual phenomenon. Here we present



**Fig 1. Geographical origin and depiction of archaeological samples.** A. Geographical location of the XBQ site in this study. Map made with Natural Earth (<https://www.naturalearthdata.com/downloads/50m-raster-data/50m-cross-blend-hypso/>). B. A child buried in a stone coffin. C. The tooth was sectioned at the cemento-enamel junction and a sample was drilled from the crown pulp chamber.

<https://doi.org/10.1371/journal.ppat.1009886.g001>

ancient DNA evidence for *S. enterica* infection in multiple individuals from XBQ suggesting systemic infectious disease that might have traveled along the Proto-Silk Road (the connection routes between the western and eastern Eurasia that preceded the Silk Road) and contributed to the suddenly high mortality.

## Results

### Identification of *S. enterica* in the Bronze Age XBQ individuals

We screened 0.4–25 million raw shotgun sequencing reads generated from dental pulp of 14 individuals recovered from the XBQ site in Xinjiang, China (see [Materials and methods](#)) to assess the possible presence of pathogen DNA. The results were evaluated by the edit distance distribution, which represented the number of nucleotide positions in a mapped DNA sequence that differ from the reference that it aligns to. [26,27]. As a result, 6 XBQ individuals showed low edit distance of the aligned reads indicating high sequence similarity with *S. enterica* ([S1 Fig](#)). Between 18 to 2,379 sequencing reads from these individuals were assigned to *S. enterica* with the two related serovars Paratyphi C and Choleraesuis as the best match. Besides, we assessed DNA deamination patterns for assigned reads. C-to-T substitutions from the 5' end and G-to-A substitutions from the 3' end were present in 6 positive samples ([S2 Fig](#)), suggesting the authenticity of our results.

### Reconstruction of the ancient *S. enterica* genomes

In order to retrieve full *S. enterica* genomes from the XBQ individuals, we performed in-solution hybridization capture to enrich the sequencing libraries for *S. enterica* DNA fragments. We generated between 13.4 to 28.37 million raw reads for the 6 positive samples (XBQM20,

**Table 1. Statistics of the *Salmonella enterica* Genome Reconstruction (including UDG library and Non-UDG library) of the XBQ individuals.**

Sample Name	Tissue Sampled	C14 dating (BP)	Library treatment	Merged reads	Mapped reads	Endogenous DNA (%)	Mean coverage	Average fragment length
XBQM20	Dental pulp	3050	Non-UDG	28372090	549836	1.938	0.6437	81.89
XBQM20	Dental pulp	3050	UDG	52625782	3875319	7.364	3.2664	67.12
XBQM90	Dental pulp	3070 to 2920	Non-UDG	23353937	485815	2.08	0.3467	106.34
XBQM90	Dental pulp	3070 to 2920	UDG	58928968	2330756	3.955	3.8519	54.26
XBQM7	Dental pulp	3210 to 3000	Non-UDG	13403597	164778	1.299	0.1979	64.75
XBQM7	Dental pulp	3210 to 3000	UDG	96358257	1195832	1.157	0.9265	54.1
XBQM83	Dental pulp	NA	Non-UDG	13827321	146472	1.059	0.0732	60.47
XBQM83	Dental pulp	NA	UDG	59022296	364829	0.618	0.2639	51.61
XBQM11	Dental pulp	3160 to 2960	Non-UDG	16362465	276052	1.687	0.2349	63.43
XBQM11	Dental pulp	3160 to 2960	UDG	8825015	106783	1.21	0.4798	52.22
XBQM64	Dental pulp	NA	Non-UDG	14806257	86420	0.584	0.0299	86.23
XBQM64	Dental pulp	NA	UDG	16991502	76625	0.451	0.0651	58.57

Notes: BP, before present; Samples without appropriate radiocarbon dating are marked with NA

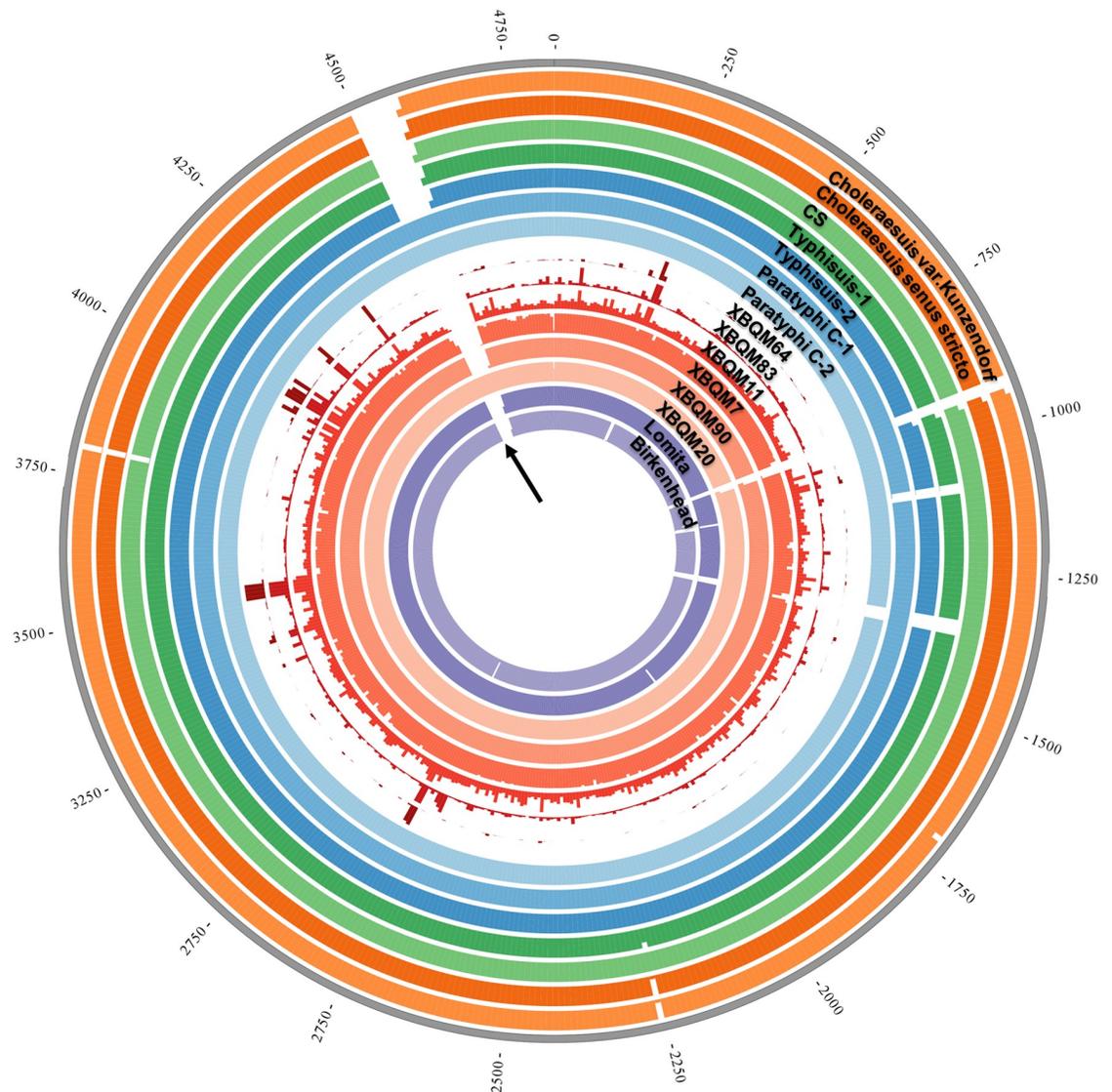
<https://doi.org/10.1371/journal.ppat.1009886.t001>

XBQM90, XBQM11, XBQM7, XBQM83 and XBQM64). After mapping to the Paratyphi C RKS4594 genome (NC\_012125.1), the data showed a clear DNA deamination pattern consistent with an ancient origin [28] (S3 Fig).

To generate high quality *S. enterica* genomes and remove deaminated cytosines owing to ancient DNA damages, we built additional libraries with uracil DNA glycosylase (UDG) treatment (see [Materials and methods](#)), which were also subjected to capture and a total of 8.83–96.36 million raw reads were generated (Table 1). Furthermore, we evaluated the in-solution capture efficiency based on the percentage of reads on target among all sequenced reads (endogenous DNA), the proportion of genome covered and the significant differences in the mean coverage (S1 Table). The percentage of endogenous DNA of XBQM20, XBQM90, XBQM11, XBQM7, XBQM83 and XBQM64 were estimated to be 228-, 104-, 80-, 118-, 41- and 73-fold higher after the in-solution capture, respectively, resulting in an average genome-wide coverage of 0.09X to 4.2X and a proportion of the genome covered of 3.69% to 86.83% (Fig 2 and S1 Table).

### Ancient *S. enterica* genomes form a previously uncharacterized branch and reveal early diversification under the Para C lineage

To determine the phylogeny of our ancient *S. enterica* strains, we mapped them together with previously published ancient and modern *S. enterica* strains to the Paratyphi C RKS4594 reference genome using strict filtering parameters (see [Materials and methods](#)). A total of 485 genomes were included to build the phylogenetic tree (including four previously published ancient genomes: Tepos14, Tepos35 [7], ragnaU [6] and ETR001 [8], 475 modern genomes as well as the six samples from this study). Given that our genomes are of relatively low genomic coverage, we evaluated SNP calls by applying the tool SNPEvaluation to prevent false positive SNP (see [Materials and methods](#)) [29]. 241,039 SNPs that passed our evaluation criteria were used to build a Maximum Likelihood (ML) tree implemented in IQ-Tree 1.6.12 with 1000 bootstrap replicates [30]. Most of XBQ genomes form a single clade and fall very close to the node that is basal to the serovars Paratyphi C, Typhisuis and Choleraesuis, all of which belong to the Para C lineage [6] (Fig 3).



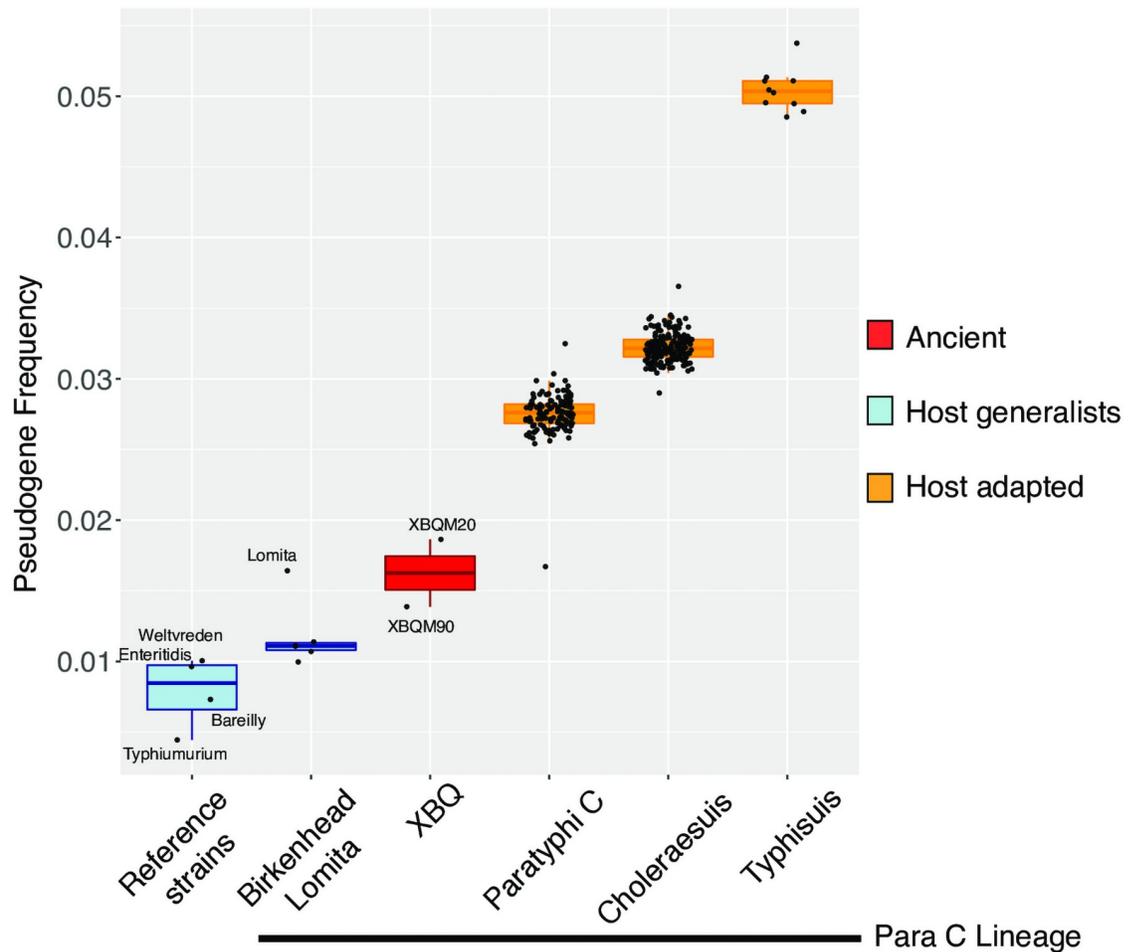
**Fig 2. Coverage overview of the modern Para C and XBQ genomes.** The average coverage was calculated for 10kb regions for the chromosome, each ring represents a maximum of 5X coverage. The figure was generated with Circos [80]. The black arrow represents the position of SPI-7.

<https://doi.org/10.1371/journal.ppat.1009886.g002>

In order to construct a more detailed phylogeny of the Para C lineage, we rebuilt the phylogenetic tree restricting our analysis to genomes under the Para C lineage. The final dataset includes 219 modern Para C lineage strains, 4 previously published ancient genomes [6–8] as well as two genomes (XBQM20 and XBQM90) from this study with an average coverage of at least 3X. In comparison to other genomes, XBQM20 and XBQM90 form a unique branch with XBQM20 falling basal to XBQM90. The relatively short branch lengths of XBQM20 and XBQM90 indicate a high genetic similarity to the direct progenitor of Paratyphi C, Typhisuis and Choleraesuis (Fig 4A). Given that the conventional phylogenetic analysis on low coverage data could produce biased results, we also adopted an alternative approach of phylogenetic placement (MGplacer) [31], whereby our ancient genomes were placed on a fixed Maximum Likelihood tree constructed by modern Para C lineage strains. As a result, the branch node of







**Fig 5. Pseudogene frequency of reference strains, XBQ samples, and strains from the Para C lineage.** Relative frequency of pseudogenes for known host generalists (blue), ancient XBQ samples (red), and known host adapted strains part of the Para C Lineage (orange). XBQ genomes with genome-wide coverage above 3X were included. Included reference strains (incl. BioSample identifier) not part of the Para C lineage are: Typhimurium (SAMN03996249), Weltvreden (SAMEA1904377), Bareilly (SAMN01823701), Enteritidis (SAMEA1705941).

<https://doi.org/10.1371/journal.ppat.1009886.g005>

analyzed the frequency of pseudogenes based on frameshift mutations and premature stop mutations, which we compare to modern as well as previously published ancient genomes [8]. In total, 1,717 genes and 32 pseudogenes were contained in XBQM20 and 2,522 genes and 35 pseudogenes were contained in XBQM90 (S3 Table). Only 8 pseudogenes (~25%) were shared between both ancient genomes, pointing to continuous acquisition of pseudogenes since their most recent common ancestor. Overall, we observe that at least 41–61% of pseudogenes are shared among genomes within each serovar (S8 Fig), confirming that pseudogenization occurred before and after serovar divergence. The pseudogene frequency of our ancient *Salmonella* was in-between the observed frequency in host generalists and host adapted serovars (Fig 5), which is in line with their phylogenetic placement basal to host adapted Paratyphi C (human), Choleraesuis (human/pig), and Typhisuis (pig). Thus, our data suggests that the ancient *S. enterica* serovars found in XBQ population were not confined to infect humans.

## Discussion

In the past decades, the study of paleopathology had largely relied on the historical reporting and morphological studies of human remains to characterize ancient epidemics events. With the development of next generation sequencing technologies [37] and the application of DNA enrichment methods [38], DNA sequences can now be retrieved from ancient human remains to verify the causative agent of diseases that cannot be achieved by other means [27,39]. More than 200 simultaneous burials were found in XBQ, as well as disproportional mortality of children. The absence of skeletal trauma suggests that they were likely not killed by violence or wars. Physical anthropological analyses showed a high mortality rate of juveniles and infants in XBQ compared with other contemporaneous cemeteries in same region [25,40]. Many archaeologists suspected this kind of unusual phenomenon could be caused by some kind of infectious diseases, but so far, no concrete conclusion could be made. Here, we used molecular methods to screen for pathogens in ancient human remains from XBQ. Six out of fourteen individuals were identified to carry detectable *S. enterica* DNA, providing an explanation for the unusual mortality event. Considering that *Salmonella* is mainly transmitted due to the intake of contaminated food and water, our findings could reflect the poor hygiene in the XBQ region such as consumption of unprocessed contaminated foods (such as meat or milk products) and water, contact with infected feces, as well as infective animals or humans.

Here we present the first evidence that *Salmonella* infected humans in East Asia roughly 3,000 years ago, which predates definite historical recordings of epidemics in China [41,42]. This further underlines that the study of ancient pathogens has great potential to track epidemic events in ancient times. In addition, all ancient genomes in this study, except for XBQM20, show an absence of SPI-7, which encodes the Vi capsular polysaccharide that could promote enteric fever in humans [33]. The presence of SPI-7 region in XBQM20 indicates that this *S. enterica* strain could have caused a systemic paratyphoid-like infection, while the XBQM90 strain might only have caused non-typhoidal salmonellosis due to the absence of SPI-7 [43]. Previous studies have suggested that SPI-7 was acquired prior to the diversification of Paratyphi C and is associated with specificity to the human host [6]. However, in our study we found the acquisition of SPI-7 was earlier than previous thought, suggesting that SPI-7 was acquired prior to the divergence of Paratyphi C and Choleraesuis. Its subsequent loss in parts of the Para C diversity could have been related to host adaptation processes.

Despite the large genetic diversity of *S. enterica* observed today, all ancient genomes reported so far are confined within the Para C Lineage or a broader group of *S. enterica*, the so-called “Ancient Eurasian Super Branch” [8], which further suggests a far-reaching global spread of those *S. enterica* serovars during prehistoric times. Although all the ancient strains in this study represent a good proxy of the direct progenitors of today’s Para C diversity, they also revealed genetic heterogeneity among each other. Firstly, XBQM20 forms a distinct branch with respect to the other XBQ samples in the phylogeny analysis (Fig 3). Secondly, we identified the presence of SPI-7 in XBQM20, but not in XBQM90 (Fig 2). Thirdly, we found there were private SNPs in our ancient samples which have passed our filtering criteria (S4 and S5 Tables). These differences indicate a genetic diversity among our ancient genomes, which suggests that multiple distinct strains were causing infections in the human population at the same time. The genomic heterogeneity of *Salmonella* observed in XBQ indicate possible recurrent epidemic events, which could imply multiple introductions of this pathogen into Xinjiang.

It has been widely accepted that the well-known historical Silk Road linking Asia, Europe and Africa, had played a key role in material trades, national reconciliation and culture exchanges. There has been growing evidence that some non-formalized long-distance trade

and communication connecting the West and East Eurasia could have occurred and were perhaps mediated by nomadic pastoralists well before the historical Silk Road (Proto-Silk Road). The observation of various archaeological cultures belonging to both East and West Eurasia [44,45] and the co-existence of diverse cereals of Near East (e.g. barley and wheat) and East Asia origin (e.g. foxtail and broomcorn millets) in Xinjiang more than 4,000 years before present [46], points out the significant role Xinjiang had played for cultural, cereal and population contacts. This dynamic crossroad could also have provided the possibility for the spread of infectious diseases such as plague, leprosy, anthrax and intestinal parasites along the Proto-Silk Road [22,47–49]. While so far there has not been a study focusing on East Asia and specifically designed to capture ancient *S. enterica* genomes, they had been identified from human skeletons in Europe dating since 6,500 BP [6,8]. Furthermore, human genetic studies demonstrated that both East and West Eurasian ancestry was detected in the XBQ individuals [50]. We propose that the emergence of *S. enterica* in this region could have been promoted by the frequent contacts between populations from the west and the east along the Proto-Silk Road, which facilitated the spread of this basal lineage that likely led to substantial mortality during Eurasian prehistory. Of course, given the limited ancient data, other possible routes cannot be excluded. It's necessary to obtain more ancient *Salmonella* DNA to reveal the accurate spreading routes in the future.

Altogether, our study shows that *S. enterica* infected human populations in East Asia at least 3,000 years ago, which long before any historical recording of pandemics in China. Our findings contribute to the known genetic diversity of *S. enterica* and reveal a previously undetected branch that falls very close to the ancestor of modern Paratyphi C, Typhisuis and Choleraesuis in the Para C lineage. Interpreting our data within the context of phylogenetic position and human population genetic background, we observed epidemic events in Xinjiang, which represents the earliest evidence that the Proto-Silk Road contributed to the spread of infectious diseases, which were likely to explain the substantial mortality in some Eurasian populations during prehistory.

## Materials and methods

### Archaeological context

The XBQ site is located in the Barkol Hakka Autonomous County, Hami District, Xinjiang Province, China (N43°30' ~ N43°32', E93°18' ~ E93°20', Fig 1). The cemetery was excavated by a joint team consisting of the Hami Bureau of Cultural Heritage, Barkol County's Cultural Relics Administration and the School of Cultural Heritage conservation of Northwest University in 2008 [23]. More than 200 burials were found in XBQ. In addition to the plant remains such as barley and wheat, some animal skeletons like cattle, sheep, horse and deer were also identified. This indicates that the XBQ population based their subsistence on cereal farming and animal husbandry, with the latter as the main contributor.

### DNA extraction, library construction and shotgun sequencing

Teeth were collected from 14 individuals excavated from XBQ. 50 mg powder per sample was removed using a dental drill following a previous study [18]. Since the preserved pathogen DNA is more likely to reside in the dried blood vessels of the pulp chamber, we only retrieved tooth powder from the dental pulp [51]. All procedures were carried out in the dedicated ancient DNA facilities in Jilin University, China. DNA extraction was performed according to a previously described protocol, with a rotation of 12–16h at 37°C during an initial lysis step [52]. A negative control was included for each step. The extraction resulted in 100 µl of DNA extract per sample. 20 µl were used for library construction as described by Meyer and Kircher

[53]. Each sample was indexed with 8bp specific indices and library amplification was performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs). Libraries were cleaned using Agencourt AMPure XP beads (Beckman Coulter) with a ratio ranging from 1:1.5 to 1:1.8 (library volume:bead volume) and were finally eluted in 26  $\mu$ l sterile water. These libraries were subsequently used for shotgun sequencing. Multiplex shotgun sequencing was carried out using an Illumina HiSeq X10 platform at Novo Inc., Beijing, China in PE 150 mode.

### Screenings methods

We used the MEGAN alignment tool (MALT), a program for the fast alignment and analysis of metagenomic DNA sequencing data [54]. The database, which includes all bacterial genomes in NCBI Refseq (December 2016) was built using the malt-build command. In order to run MALT, 14 samples were first processed with the EAGER pipeline to perform adapter clipping and paired-end read merging [55]. Merged reads were subsequently screened by malt-run. The “minimum percent identity” parameter was set to 90 (-id), the minimum support parameter was set to 1 (-sup), a top percent value of 1 (-top) was set, the maximum number of alignments per query was set to 100 (-mq), BlastN mode and SemiGlobal alignment were applied. The generated rma6 files were visual inspected in MEGAN [56] and were further processed with HOPS [57], an automated java-based pipeline incorporating the metagenomic alignment tool MALT, which focuses on screening rma6 data for the presence of a user-specified list of target species. Specifically, the MALT output (rma6) was analyzed with MAL-Texttract, a newly designed tool within HOPS that allows automatic retrieval of alignment information from rma6 files.

### Probe design, in-solution capture and sequencing

Probes were designed based on publicly available reference sequences including *Salmonella* chromosomes/assemblies and plasmids [7] (S6 Table). Due to the degradation characteristics of ancient DNA, previous studies have shown that a shorter probe length (e.g., 60 bp) works better for ancient DNA, while probes that work efficiently for modern samples can be up to 120 bp in length [58]. In order to enrich *S. enterica* genomes in a more cost-efficiency way, we designed the capture probes with a length of 100bp. Through the above design, we produced a total of 92,710 probes with 100 bp tiling. Probes were synthesized by iGeneTech Co. Ltd in China.

The libraries that were identified as positive for *S. enterica* were then enriched using an in-solution DNA capture with the probes described above. Before hybridization, the DNA library was amplified used KAPAHiFi HotStart enzyme to reach a concentration of 500ng/ $\mu$ l. RNA probes and adapter blockers were used following the kit manufacturer instructions. The hybridization condition was set to 62° for 48 hours in order to make a full hybridization between the DNA library and the probes. The Dynabeads Myone streptavidin T1 magnetic beads were used for capture because of its strong non-covalent interaction with streptavidin. Lastly, the captured libraries were re-amplified with IS5 and IS6 for 14 cycles. Subsequently, the libraries were purified by AMPure XP beads, quantified by Bioanalyzer 2100 and sequenced on a HiSeq X10 platform.

In order to avoid the incorporation of incorrect bases due to ancient DNA modifications, the positive samples were pre-treated with USER enzyme (New England Biolabs), which contains Uracil DNA Glycosylase (UDG) and endonuclease VIII (endoVIII) [59]. UDG treatment removes uracil residues, which are most common at the 5- and 3-prime end of ancient DNA molecules.

## Read processing, mapping and SNP calling

All samples were processed with the EAGER pipeline [55]. Sequencing quality for each sample was evaluated with FastQC, and adaptors clipped using the AdapterRemoval module in EAGER and reads shorter than 30 bp were discarded [60]. The preserved reads were then mapped with BWA v0.7.17 [61] to Paratyphi C RKS4594 (NC\_012125.1) as a reference genome. Non-UDG libraries (BWA parameters: -l 16; -n 0.01; -q 37) were mapped with looser parameters, while UDG treated libraries were mapped with stringent parameters (BWA parameters: -l 32; -n 0.1; -q 37). The duplicates were removed by the DeDup tool implemented in EAGER. The ancient DNA damage patterns were characterized using mapDamage [28].

To increase the depth of data, the raw fastq file including the UDG and Non-UDG data were merged using the zcat command and then processed with the EAGER pipeline using stringent alignment. In order to minimize the bias due to ancient DNA deamination, the final bam was trimmed according to the frequencies of C to T or G to A at both 3' and 5' ends to a degree that the damages at the end of the trimmed reads were identical to the baseline. The trimmed reads were performed SNP calling using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK) with high-quality base score ( $Q \geq 30$ ), using the "EMIT\_ALL\_SITES" option, which generates a call for each genomic site [62].

## SNP evaluation and phylogenetic analyses

In order to infer the phylogenetic placement of the ancient genomes, we compiled a dataset including 475 modern *S. enterica* genomes, the 4 previous published ancient *S. enterica* genomes [6–8] as well as 6 genomes from this study. All modern genomes were fragmented 100bp with 1bp step size, then aligned to the Paratyphi C genome using BWA with strict parameters (bwa aln -l 32, -n 0.1, -q 37). SNP calling was carried out with the Genome Analysis Toolkit (GATK) UnifiedGenotyper [62] using a genotype quality score of  $\geq 30$  for the ancient genomes and the modern genomes, and using the 'EMIT\_ALL\_SITES' option, providing a call for all variant or non-variant bases in the vcf file output [62]. MultiVCFanalyzer [63] was used for collating homozygous SNPs (Minimal genotyping quality set as 30, 90% of reads covering a position must be in agreement) called at a minimum of 3X coverage against the *S. Paratyphi C* RKS4594 reference [63]. SNP alignments were used for phylogenetic analysis. We evaluated SNP calls by using the SNPEvaluation tool [29]. SNPs were called true positive when meeting the following criteria within a 50-bp window: (A) The reads were mapped using stringent parameters (bwa aln -l 32, -n 0.1 and -q 37), (B) no "heterozygous" positions, and (C) no non-covered positions. All false-positive SNPs were discarded for the phylogenetic tree construction.

A maximum likelihood tree was generated by IQ-Tree 1.6.12 [30], which was run using ModelFinder with the option -m MFP. A total number of 478 models were tested, we identified GTR+F+ASC+R3 as the best substitution model. 1,000 fast bootstrap replicates were performed to assess statistical support at each node. The Arizonae serovar was set as outgroup. The result showed that our ancient genomes cluster within the Para C lineage. In order to further verify the phylogenetic tree of our ancient genomes, we constructed another maximum parsimony tree with 500 bootstraps using 219 modern Para C genomes and ancient genomes by the MEGA-proto and executed using MEGA-CC [64].

Given that conventional phylogenetic analysis on low coverage data could produce biased results, we also adopted an additional approach for phylogenetic placement (MGplacer) [31], whereby low coverage genomes were placed on a fixed Maximum Likelihood tree. 219 modern Para C group genomes were used to construct a maximum-likelihood tree, and MGplacer2.py was used to determine the branch location of our ancient genomes. Then our low coverage data

were plotted on the fixed tree. The generated tree was visualized using FigTree v1.4.3 [65] (<http://tree.bio.ed.ac.uk/software/figtree/>) and Evolview [66] (<https://evolgenius.info/evolview-v2>).

### Estimation of divergence times

The SNP alignment was used for molecular dating with BEAST v1.10.1 [67]. Before this, we investigated if the data shows a temporal signal using TempEst v1.5.1 [34]. Ancient genomes XBQM20 and XBQM90 were selected for the dating analysis as they showed sufficient genomic coverage. Modern Para C genomes, and the ancient genomes Tepos14, Tepos35 and rag-naU were included. A root to tip regression of genetic distances and sampling times showed a sufficient temporal signal and permitted us to proceed with the molecular dating analysis. For tip dating, all modern genomes were set to an age of 0, ancient genomes were set according to their C14 dating (2-sigma range). We used a relaxed lognormal molecular clock, with the GTR +G model of nucleotide substitution, with a discrete gamma distribution and six rate categories to account for rate heterogeneity across sites applied with the Bayesian skyline coalescent model. The generated XML file was run by BEAST. For our dataset, 100 million MCMC steps were computed sampling every 5,000 steps. The output log file was analyzed with Tracer with a 10 percent burn-in. A maximum clade credibility tree with a 10% burn-in was produced using the TreeAnnotator [67]. The tree was visualized with FigTree v1.4.3 [65].

### Presence/absence analysis of *S. enterica* virulence factors (SPI, virulence plasmid and phages)

We collected annotations for SPI-1 through SPI-16 as well as for the virulence plasmid and phages from the PAIDB database and previous studies [6, 68–74]. The plasmid annotation files were downloaded from NCBI. Each SPI as well as plasmid was used as a reference for mapping all reads from ancient and modern genomes part of the Para C lineage with BWA mem [75]. Alignments were filtered using Picard Tools (CleanSam, MarkDuplicates) [75]. Read depth for each gene was summarized using bedtools (bedtools coverage -a file.bed -b bamfile > depth.txt) [76]. We considered only genes that had at least 95% of its sequence covered by the capture probes. Heatmaps were plotted using the ggplot2 package in R [77].

### Pseudogene analysis

Pseudogenes are coding sequences (CDS) that are putatively inactivated by mutations including nonsense substitutions, frameshifts, or truncation by deletion or rearrangement [78]. Here, we infer the frequency of pseudogenes per genome similar to Key et al. [8]. Briefly, we use a pangenome of *S. enterica* based on all CDS from 537 representative genomes, providing 21,065 genes after filtering and cleaned for paralogs. We included only genes 100% covered by the probe set used for capture, leading to 8,726 genes used for analysis. All genomes from the Para C Lineage were included together with additional *S. enterica* genomes with known host specificity. We inferred pseudogenes using GenCons [79] to build a consensus sequence using the following two criteria: we reported every position with allele and total coverage of at least 1X and an alternative allele frequency of at least 66% (thus SNPs are only called with at least two alternative alleles in agreement). We infer frequency of pseudogenes per genome using all genes that have an allele called in at least 90% of positions.

### Supporting information

**S1 Fig. HOPS screening results for *S. enterica* positive samples.** The edit distance distribution represented the number of nucleotide positions in a mapped DNA sequence that differ

from the reference that it aligns to, eg. if there is one mismatch in the alignment the edit distance is 1. The image on the left: Edit distance distribution for all reads assigned to *S. enterica*. The image on the right: Edit distance distribution for assigned reads that show a damage signal.

(TIF)

**S2 Fig. HOPS output of DNA damage plot for assigned *S. enterica* reads.** C-to-T and G-to-A substitutions from the 5' end and the 3' end were presented in our 6 positive samples, which is typical for ancient DNA.

(TIF)

**S3 Fig. 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.

(TIF)

**S4 Fig. Maximum Likelihood tree of modern Salmonella Para C lineage and XBQ strains.** 219 modern Para C group genomes were constructed a maximum-likelihood tree, XBQ data were mapped to the tree with MGplacer. As a result, XBQ strains were placed between the node Br\_219 and Br\_220, which were basal to the position of the Paratyphi C, Typhisuis and Choleraesuis.

(TIF)

**S5 Fig. The read depth at nucleotide position of SPI-7 in XBQM20.**

(TIF)

**S6 Fig. Root-to-tip regression analysis.** Plots of the root-to-tip genetic distance against sampling time were shown. All modern genomes were set to an age of 0, ancient genomes were set according to their C14 dating. Sampling dates were given as years before the present. The dataset yielded  $R^2$  of 0.2368, which confirms the existence of temporal signal.

(TIF)

**S7 Fig. Maximum Clade Credibility tree.** The MCC tree was produced using TreeAnnotator of BEAST v1.10.1. The tree was visualized in FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). It is presented in a temporal scale between 25,000 and 0 yBP, and the main internal node dates of the Para C lineage are indicated on each corresponding node.

(TIF)

**S8 Fig. Proportion of shared pseudogenes between strains across the Para C lineage and Birkenhead.** Proportion of pseudogene-sharing (0–100%) between strains shown in tones of red. Strains are grouped by serovar.

(TIF)

**S1 Table. *Salmonella enterica* capture efficiency (capture data have been merged UDG data and Non-UDG data).**

(XLSX)

**S2 Table. Bayesian posterior estimates of the time to the most recent common ancestor and substitution rates for the sub-lineage of Para C lineage.**

(XLSX)

**S3 Table. Pseudogenes identified in XBQM20S and XBQM90S.** Pseudogenes shared between both ancient genomes are marked in yellow.

(XLSX)

**S4 Table. Information about variant positions detected in one or both of the ancient genomes (XBQM20 and XBQM90).**

(XLSX)

**S5 Table. Information about variant positions detected in one or both of the ancient genomes.**

(XLSX)

**S6 Table. Accession number information about our probes designed based on publicly available reference sequences including Salmonella chromosomes/assemblies and plasmids.**

(XLSX)

## Acknowledgments

We thank the members of the Computational Pathogenomics group at the Max Planck Institute for the Science of Human History in Jena for critical discussions. We appreciate Baoxu Ding for the bioinformatic discussion.

## Author Contributions

**Conceptualization:** Johannes Krause, Alexander Herbig, Yinqiu Cui.

**Data curation:** Xiyan Wu, Felix M. Key.

**Formal analysis:** Xiyan Wu, Felix M. Key, Aida Andrades Valtueña, Aditya Kumar Lankapalli.

**Funding acquisition:** Yinqiu Cui.

**Investigation:** Xiyan Wu, Shizhu Gao, Xuan Yang, Fan Zhang, Linlin Liu, Zhongzhi Nie, Jian Ma.

**Methodology:** Xiyan Wu, Felix M. Key.

**Project administration:** Yinqiu Cui.

**Resources:** Zhongzhi Nie, Jian Ma.

**Supervision:** Alexander Herbig, Yinqiu Cui.

**Visualization:** Xiyan Wu, Felix M. Key.

**Writing – original draft:** Xiyan Wu.

**Writing – review & editing:** Chao Ning, Felix M. Key, Aida Andrades Valtueña, Johannes Krause, Alexander Herbig, Yinqiu Cui.

## References

1. Jacobs M, Koornhof HJ, Crisp SI, Palmhert HL, Fitzstephens A. Enteric fever caused by *Salmonella* paratyphi C in South and South West Africa. *South African Medical Journal*. 1978; 54(11):434–8. PMID: [104399](https://pubmed.ncbi.nlm.nih.gov/104399/).
2. Branchu P, Bawn M, Kingsley RA. Genome Variation and Molecular Epidemiology of *Salmonella enterica* Serovar Typhimurium Pathovariants. *Infect Immun*. 2018; 86(8). <https://doi.org/10.1128/IAI.00079-18> PMID: [29784861](https://pubmed.ncbi.nlm.nih.gov/29784861/).
3. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, 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*. 2015; 12(12):e1001921. <https://doi.org/10.1371/journal.pmed.1001921> PMID: [26633831](https://pubmed.ncbi.nlm.nih.gov/26633831/).

4. Jajere SM. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Vet World*. 2019; 12(4):504–21. <https://doi.org/10.14202/vetworld.2019.504-521> PMID: 31190705.
5. Harkins KM, Stone AC. Ancient pathogen genomics: insights into timing and adaptation. *J Hum Evol*. 2015; 79:137–49. <https://doi.org/10.1016/j.jhevol.2014.11.002> PMID: 25532802.
6. Zhou Z, Lundstrom I, Tran-Dien A, Duchene S, Alikhan NF, Sergeant MJ, et al. Pan-genome Analysis of Ancient and Modern *Salmonella enterica* Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Curr Biol*. 2018; 28(15):2420–8 e10. <https://doi.org/10.1016/j.cub.2018.05.058> PMID: 30033331.
7. Vågane ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol*. 2018; 2(3):520–8. <https://doi.org/10.1038/s41559-017-0446-6> PMID: 29335577.
8. Key FM, Posth C, Esquivel-Gomez LR, Hubler R, Spyrou MA, Neumann GU, et al. Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process. *Nat Ecol Evol*. 2020; 4(3):324–33. <https://doi.org/10.1038/s41559-020-1106-9> PMID: 32094538.
9. Betts A, Jia P, Abuduresule I. A new hypothesis for early Bronze Age cultural diversity in Xinjiang, China. *Archaeological Research in Asia*. 2018; 17:204–13.
10. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015; 522(7555):207–11. <https://doi.org/10.1038/nature14317> PMID: 25731166.
11. Damgaard PB, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliusen T, et al. 137 ancient human genomes from across the Eurasian steppes. *Nature*. 2018; 557(7705):369–74. <https://doi.org/10.1038/s41586-018-0094-2> PMID: 29743675.
12. Jeong C, Wang K, Wilkin S, Taylor WTT, Miller BK, Bemann JH, et al. A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell*. 2020. <https://doi.org/10.1016/j.cell.2020.10.015> PMID: 33157037.
13. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, et al. The formation of human populations in South and Central Asia. *Science*. 2019; 365(6457). <https://doi.org/10.1126/science.aat7487> PMID: 31488661.
14. Frachetti Michael D, Anthony David W, Epimakhov A. V, et al. Multiregional Emergence of Mobile Pastoralism and Nonuniform Institutional Complexity across Eurasia. *Commentaries. Reply. Current Anthropology*. 2012; 53(1):2–38.
15. Dong G, Du L, Wei W. The impact of early trans-Eurasian exchange on animal utilization in northern China during 5000–2500 BP. *The Holocene*. 2020; 30(5):095968362094116.
16. Dong G, Yang Y, Han J, Wang H, Chen F. Exploring the history of cultural exchange in prehistoric Eurasia from the perspectives of crop diffusion and consumption. *Sci China Earth Sci*. 2017; 60(006):1110–23.
17. Betts A, Jia PW, Dodson J. The origins of wheat in China and potential pathways for its introduction: A review. *Quaternary International*. 2014; 348(oct.20):158–68.
18. Li C, Li H, Cui Y, Xie C, Cai D, Li W, et al. Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC biology*. 2010; 8(1):15. <https://doi.org/10.1186/1741-7007-8-15> PMID: 20163704.
19. Feng Q, Lu Y, Ni X, Yuan K, Yang Y, Yang X, et al. Genetic History of Xinjiang's Uyghurs Suggests Bronze Age Multiple-Way Contacts in Eurasia. *Mol Biol Evol*. 2017; 34(10):2572–82. <https://doi.org/10.1093/molbev/msx177> PMID: 28595347.
20. Ning C, Wang CC, Gao S, Yang Y, Zhang X, Wu X, et al. Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan. *Curr Biol*. 2019; 29(15):2526–32 e4. <https://doi.org/10.1016/j.cub.2019.06.044> PMID: 31353181.
21. Zhang L, Krist G. *Archaeology and Conservation Along the Silk Road: Vandenhoek & Ruprecht*; 2018.
22. Yeh HY, Mao R, Wang H, Qi W, Mitchellet PD, et al. Early evidence for travel with infectious diseases along the Silk Road: Intestinal parasites from 2000 year-old personal hygiene sticks in a latrine at Xuanquanzhi Relay Station in China. *J Archaeol Sci*. 2016; 9:758–764.
23. Ma J. The archaeology discovery on the Hongshankou Site in Barkol, Xinjiang in 2008. *Cultural relics*. 2014. [In Chinese].
24. Ma J. A preliminary study on the early settlement forms in Eastern Tianshan Mountain. *Chinese Social Sciences Today*. 2017. [In Chinese].
25. Wei D, Zen W, Tuohuti Tulahong. A preliminary study on the physical characteristics of ancient remains from the Baiqier Cemetery, Xinjiang. *Research of China's Frontier Archaeology*. 2010; (1):258–70. [In Chinese].

26. Key FM, Posth C, Krause J, Herbig A, Bos KI. Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication. *Trends in Genetics*. 2017; 33(8):508–20. <https://doi.org/10.1016/j.tig.2017.05.005> PMID: 28688671.
27. Spyrou MA, Bos KI, Herbig A, Krause J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat Rev Genet*. 2019. <https://doi.org/10.1038/s41576-019-0119-1> PMID: 30953039.
28. Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*. 2011; 27(15):2153–5. <https://doi.org/10.1093/bioinformatics/btr347> PMID: 21659319.
29. Keller M, Spyrou MA, Scheib CL, Neumann GU, Kropelin A, Haas-Gebhard B, et al. Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). *Proc Natl Acad Sci USA*. 2019; 116(25):12363–72. <https://doi.org/10.1073/pnas.1820447116> PMID: 31164419.
30. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015; 32(1):268–74. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430.
31. Kay GL, Sergeant MJ, Zhou Z, Chan JZ-M, Millard A, Quick J, et al. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat Commun*. 2015; 6(1):1–9. <https://doi.org/10.1038/ncomms7717> PMID: 25848958.
32. Barrow PA, Methner U. *Salmonella* in domestic animals. CABI. 2013.
33. Wilson RP, Winter SE, Spees AM, Winter MG, Nishimori JH, Sanchez JF, et al. The Vi capsular polysaccharide prevents complement receptor 3-mediated clearance of *Salmonella enterica* serotype Typhi. *Infect Immun*. 2011; 79(2):830–7. <https://doi.org/10.1128/IAI.00961-10> PMID: 21098104.
34. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016; 2(1):vew007. <https://doi.org/10.1093/ve/vew007> PMID: 27774300.
35. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012; 29(8):1969–73. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748.
36. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, et al. Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. *Proc Natl Acad Sci USA*. 2015; 112(3):863–8. <https://doi.org/10.1073/pnas.1416707112> PMID: 25535353.
37. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437(7057):376–80. <https://doi.org/10.1038/nature03959> PMID: 16056220.
38. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 2010; 7(2):111–8. <https://doi.org/10.1038/nmeth.1419> PMID: 20111037.
39. Stone AC, Lewis CM Jr., Schuenemann VJ. Insights into health and disease from ancient biomolecules. *Philos Trans R Soc Lond B Biol Sci*. 2020; 375(1812):20190568. <https://doi.org/10.1098/rstb.2019.0568> PMID: 33012226.
40. Wei D. The study on the changes and communication models of ancient people in Hami, Xinjiang from the Bronze Age to the Early Iron Age: Science Press; 2017. [In Chinese].
41. Zhang W. Historical Epidemic Disease in China and Prevention. Party and Government Forum. 2003; (06):41–2. [In Chinese].
42. Zhang Z. Chronology of Epidemics in Ancient China: Fujian Science and Technology Publishing House. 2007. [In Chinese].
43. Seth-Smith HM. SPI-7: *Salmonella*'s Vi-encoding Pathogenicity Island. *J Infect Dev Ctries*. 2008; 2(4):267–71. <https://doi.org/10.3855/jidc.220> PMID: 19741287.
44. Wang T, Wei D, Chang X, Yu Z, Zhang X, Wang C, et al. Tianshanbeilu and the Isotopic Millet Road: Reviewing the late Neolithic/Bronze Age radiation of human millet consumption from north China to Europe. *National ence Review*. 2017;(5):5.
45. Wei P, Jia M, Betts AVG. A re-analysis of the Qiemu'erqieke (Shamirshak) cemeteries, Xinjiang, China. *Journal of Indo-European Studies*, 2010; 38(4).
46. Zhou X, Yu J, Spengler RN, Shen H, Li X. 5,200-year-old cereal grains from the eastern Altai Mountains redate the trans-Eurasian crop exchange. *Nat Plants*. 2020; 6(2). <https://doi.org/10.1038/s41477-019-0581-y> PMID: 32055044.
47. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat Genet*. 2009; 41(12):1282–9. <https://doi.org/10.1038/ng.477> PMID: 19881526.

48. Schmid BV, Büntgen U, Easterday WR, Ginzler C, Walløe L, Bramanti B, et al. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci USA*. 2015; 112(10):3020–5. <https://doi.org/10.1073/pnas.1412887112> PMID: 25713390.
49. Simonson TS, Okinaka RT, Wang B, Easterday WR, Huynh L, U'Ren JM, et al. *Bacillus anthracis* in China and its relationship to worldwide lineages. *BMC Microbiol*. 2009; 9(1):1–11. <https://doi.org/10.1186/1471-2180-9-71> PMID: 19368722.
50. Wu X. Genomics and Evolution History Study of Ancient Pathogen from Xinjiang [phd thesis]: Jilin University; 2020. [In Chinese].
51. La VD, Aboudharam G, Raoult D, Drancourt M. Dental Pulp as a Tool for the Retrospective Diagnosis of Infectious Diseases. *Paleomicrobiology*: Springer; 2008;175–96.
52. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA*. 2013; 110(39):15758–63. <https://doi.org/10.1073/pnas.1314445110> PMID: 24019490.
53. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. 2010; 2010(6):pdb. <https://doi.org/10.1101/pdb.prot5448> PMID: 20516186.
54. Herbig A, Maixner F, Bos KI, Zink A, Krause J, Huson DH. MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *BioRxiv*. 2016;050559.
55. Peltzer A, Jager G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. *Genome Biol*. 2016; 17:60. <https://doi.org/10.1186/s13059-016-0918-z> PMID: 27036623.
56. Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition—Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol*. 2016; 12(6):e1004957. <https://doi.org/10.1371/journal.pcbi.1004957> PMID: 27327495.
57. Huebler R, Key F. M., Warinner C., Bos K. I., Krause J., & Herbig A. HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol*. 2019; 20(1):280. <https://doi.org/10.1186/s13059-019-1903-0> PMID: 31842945.
58. Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am J Hum Genet*. 2013; 93(5):852–64. <https://doi.org/10.1016/j.ajhg.2013.10.002> PMID: 24568772.
59. Gansauge M-T, Meyer M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res*. 2014; 24(9):1543–9. <https://doi.org/10.1101/gr.174201.114> PMID: 25081630.
60. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016; 9:88. <https://doi.org/10.1186/s13104-016-1900-2> PMID: 26868221.
61. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010; 26(5):589–95. <https://doi.org/10.1093/bioinformatics/btp698> PMID: 20080505.
62. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20(9):1297–303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199.
63. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*. 2014; 514(7523):494–7. <https://doi.org/10.1038/nature13591> PMID: 25141181.
64. Kumar S, Stecher G, Peterson D, Tamura K. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*. 2012; 28(20):2685–6. <https://doi.org/10.1093/bioinformatics/bts507> PMID: 22923298.
65. Rambaut A, Drummond A. FigTree v 1.4.3. online at: <http://tree.bio.ed.ac.uk/software/figtree/>. 2012.
66. Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res*. 2012; 40(Web Server issue):W569–72. <https://doi.org/10.1093/nar/gks576> PMID: 22695796.
67. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007; 7:214. <https://doi.org/10.1186/1471-2148-7-214> PMID: 17996036.
68. Yoon SH, Park YK, Kim JF. PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. *Nucleic Acids Res*. 2015; 43(Database issue):D624–30. <https://doi.org/10.1093/nar/gku985> PMID: 25336619.
69. Blondel CJ, Jimenez JC, Contreras I, Santiviago CA. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics*. 2009; 10:354. <https://doi.org/10.1186/1471-2164-10-354> PMID: 19653904.

70. Elder JR, Chiok KL, Paul NC, Haldorson G, Guard J, Shah DH. The Salmonella pathogenicity island 13 contributes to pathogenesis in streptomycin pre-treated mice but not in day-old chickens. *Gut Pathog.* 2016; 8:16. <https://doi.org/10.1186/s13099-016-0098-0> PMID: 27141235.
71. Hayward MR, Jansen V, Woodward MJ. Comparative genomics of Salmonella enterica serovars Derby and Mbandaka, two prevalent serovars associated with different livestock species in the UK. *BMC Genomics.* 2013; 14:365. <https://doi.org/10.1186/1471-2164-14-365> PMID: 23725633.
72. Desai PT, Porwollik S, Long F, Cheng P, Wollam A, Bhonagiri-Palsikar V, et al. Evolutionary Genomics of Salmonella enterica Subspecies. *MBio.* 2013; 4(2). <https://doi.org/10.1128/mBio.00579-12> PMID: 23462113.
73. Espinoza RA, Silva-Valenzuela CA, Amaya FA, Urrutia IM, Contreras I, Santiviago CA. Differential roles for pathogenicity islands SPI-13 and SPI-8 in the interaction of Salmonella Enteritidis and Salmonella Typhi with murine and human macrophages. *Biol Res.* 2017; 50(1):5. <https://doi.org/10.1186/s40659-017-0109-8> PMID: 28202086.
74. Zou QH, Li QH, Zhu HY, Feng Y, Li YG, Johnston RN, et al. SPC-P1: a pathogenicity-associated prophage of Salmonella paratyphi C. *BMC Genomics.* 2010; 11:729. <https://doi.org/10.1186/1471-2164-11-729> PMID: 21192789.
75. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv. 2013;13033997.
76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278.
77. Allaire J. RStudio: integrated development environment for R. Boston, MA. 2012;770:394.
78. Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R, Bhutta ZA, et al. Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. *BMC Genomics.* 2009; 10:36. <https://doi.org/10.1186/1471-2164-10-36> PMID: 19159446.
79. Fellows Yates JA, Drucker DG, Reiter E, Heumos S, Welker F, Munzel SC, et al. Central European Woolly Mammoth Population Dynamics: Insights from Late Pleistocene Mitochondrial Genomes. *Sci Rep.* 2017; 7(1):17714. <https://doi.org/10.1038/s41598-017-17723-1> PMID: 29255197.
80. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19(9):1639–45. <https://doi.org/10.1101/gr.092759.109> PMID: 19541911.