

# Genomic perspective on the bacillus causing paratyphoid B fever

Received: 7 June 2024

Accepted: 11 November 2024

Published online: 10 December 2024

 Check for updates

Jane Hawkey <sup>1,25</sup>, Lise Frézal <sup>2,25</sup>, Alicia Tran Dien<sup>2,24</sup>, Anna Zhukova <sup>3</sup>, Derek Brown <sup>4</sup>, Marie Anne Chattaway<sup>5</sup>, Sandra Simon<sup>6</sup>, Hidemasa Izumiya <sup>7</sup>, Patricia I. Fields <sup>8</sup>, Niall De Lappe<sup>9</sup>, Lidia Kaftyreva<sup>10</sup>, Xuebin Xu<sup>11</sup>, Junko Isobe<sup>12</sup>, Dominique Clermont <sup>13</sup>, Elisabeth Njamkepo <sup>2</sup>, Yukihiro Akeda<sup>7</sup>, Sylvie Issenhuth-Jeanjean<sup>2</sup>, Mariia Makarova <sup>10</sup>, Yanan Wang <sup>14,15</sup>, Martin Hunt<sup>16,17,18,19</sup>, Brent M. Jenkins <sup>8</sup>, Magali Ravel<sup>2</sup>, Véronique Guibert<sup>2</sup>, Estelle Serre<sup>2</sup>, Zoya Matveeva<sup>10</sup>, Laëtitia Fabre<sup>2</sup>, Martin Cormican <sup>9,20</sup>, Min Yue<sup>21,22</sup>, Baoli Zhu <sup>15</sup>, Masatomo Morita <sup>7</sup>, Zamin Iqbal <sup>16,23</sup>, Carolina Silva Nodari<sup>2</sup>, Maria Pardos de la Gandara <sup>2</sup> & François-Xavier Weill <sup>2</sup> ✉

Paratyphoid B fever (PTB) is caused by an invasive lineage (phylogroup 1, PG1) of *Salmonella enterica* serotype Paratyphi B (SPB). However, little was known about the global population structure, geographic distribution, and evolution of this pathogen. Here, we report a whole-genome analysis of 568 historical and contemporary SPB PG1 isolates, obtained globally, between 1898 and 2021. We show that this pathogen existed in the 13th century, subsequently diversifying into 11 lineages and 38 genotypes with strong phylogeographic patterns. Following its discovery in 1896, it circulated across Europe until the 1970s, after which it was mostly reimported into Europe from South America, the Middle East, South Asia, and North Africa. Antimicrobial resistance recently emerged in various genotypes of SPB PG1, mostly through mutations of the quinolone-resistance-determining regions of *gyrA* and *gyrB*. This study provides an unprecedented insight into SPB PG1 and essential genomic tools for identifying and tracking this pathogen, thereby facilitating the global genomic surveillance of PTB.

At the turn of the 20th century, investigators in Europe and North America showed that Eberth's bacillus (now known as *Salmonella enterica* serotype Typhi) was not the only organism causing enteric fever, a severe systemic infection causing prolonged high fever, fatigue, headache, and abdominal pain, exclusively in humans<sup>1–3</sup>. The other causal bacteria identified were paratyphoid bacilli of three different types: A, B, and C. The first two cases involving the paratyphoid B bacillus (now known as *S. enterica* serotype Paratyphi B and referred to hereafter as SPB) were described by Achard and Bensaude in Paris, France, in 1896 (refs. 1,4,5). Following the introduction of appropriate laboratory tools – initially based on O- and H-antigen serotyping<sup>6</sup>

(recognising the antigenic formula 1,4[5],12:b:1,2) and then on an absence of *d*-tartrate (*d*-Tar) fermentation by the cultured bacteria (SPB strains)<sup>7</sup> – SPB was more frequently detected in Europe (Supplementary Note “Epidemiology of paratyphoid B fever during the first half of the 20th century”). PTB disease was milder than typhoid fever, with a lower incidence of complications and a lower mortality<sup>8,9</sup>. It occurred as sporadic cases with occasional outbreaks, rarely due to case-to-case transmission in institutions (such as garrisons, hospitals, and children's homes), but was mostly foodborne, particularly in natural or synthetic cream, unpasteurised milk, and bakery products<sup>9</sup>. PTB cases remained frequent across Europe in the 1950s and 1960s<sup>10–12</sup>, but

A full list of affiliations appears at the end of the paper. ✉ e-mail: [francois-xavier.weill@pasteur.fr](mailto:francois-xavier.weill@pasteur.fr)

the end of the 1970s<sup>13</sup> saw a progressive decline in SPB<sup>-</sup> isolation accompanied by an increase in the isolation of zoonotic *d*-tartrate-fermenting (SPB<sup>+</sup>) strains, also known as variant Java strains. Epidemiologically, PTB also shifted from being locally acquired to being an imported disease. In the UK, 43.2% (152/352) of PTB cases were considered to have been contracted locally between 1973 and 1977, whereas 56.8% (200/352) were considered to have been contracted abroad, particularly in Mediterranean and Middle Eastern countries<sup>14</sup>.

A phage typing scheme was developed for the surveillance of PTB as early as World War II (WWII)<sup>15,16</sup>. This scheme subtyped SPB<sup>-</sup> isolates into 10 different phage types (PTs) – 1, 2, 3a, 3a1, 3b, BAOR (British Army of the Rhine), Jersey, Beccles, Dundee, and Taunton – and reports based on tens of thousands of isolates phage-typed across the world were regularly published from the 1940s to the 1990s<sup>10–12,17,18</sup>. A particular geographic distribution was observed, with PTs 1 and 2 reported to be autochthonous to the UK (whereas Taunton was found in imported cases); BAOR was prevalent in Central Europe; Dundee was prevalent in France, and the PT 3 series (3a, 3a1, and 3b) was more common in North America than elsewhere. However, despite their different epidemiological characteristics and pathogenicity, SPB<sup>-</sup> and SPB<sup>+</sup> strains were grouped together as a single serotype, SPB<sup>6,19</sup>, and their PTs were ultimately combined without distinction in world surveys, rendering these surveys less informative.

From the 1990s onwards, population genetics tools were used to distinguish between SPB<sup>+</sup> and invasive SPB<sup>-</sup> strains, initially by multilocus enzyme electrophoresis<sup>20</sup> (Pb1 group for SPB<sup>-</sup>) and then by multilocus sequence typing (MLST)<sup>21</sup> (sequence type ST86 and five single-locus variants [SLVs] of ST86 for SPB<sup>-</sup>). In 2003, the molecular basis of the *d*-Tar<sup>-</sup> character of SPB<sup>-</sup> strains was elucidated: a single-nucleotide variant (SNV) leading to the loss of the start codon of a gene involved in the *d*-tartrate fermentation pathway<sup>22</sup>.

In 2017, an estimated 10.9 million cases of typhoid fever and 3.4 million cases of paratyphoid fever (all types), resulted in 116.8 and 19.1 thousand deaths worldwide, respectively<sup>23</sup>. However, by contrast to *S. enterica* serotypes Typhi (STY)<sup>24</sup>, Paratyphi A (SPA)<sup>25</sup>, and Paratyphi C<sup>26</sup>, little is known about the population structure of SPB<sup>-</sup>. Connor and coworkers<sup>27</sup> were the first to try to elucidate the structure in a study of 191 isolates of SPB (SPB<sup>+</sup> and SPB<sup>-</sup>) from around the world. Their phylogenomic analysis identified 10 distinct lineages, named phylogroups (PGs 1–10). All 34 SPB<sup>-</sup> isolates were grouped into the invasive PG1 lineage, which was derived from the closely related lineages PG2 to –5, containing SPB<sup>+</sup> gastrointestinal isolates. Microbial genomic surveillance conducted in the UK since 2014 led to the identification of phylogenetic clades of SPB<sup>-</sup> strains associated with travel to South America, Iraq, and Pakistan<sup>28,29</sup>. However, the use of a limited number of global SPB<sup>-</sup> isolates or of contemporary routine SPB<sup>-</sup> isolates ruled out any deeper global phylogenomic analysis of this pathogen.

Here, we studied 568 genomes from a spatially and temporally diverse set of SPB<sup>-</sup> isolates, to determine the global population structure, geographic distribution, and evolution of this pathogen. We also developed a hierarchical SNV-based genotyping scheme implemented within Mykrobe open-source software that splits SPB<sup>-</sup> isolates into 38 distinct and often phylogeographically informative genotypes, thereby facilitating the global genomic surveillance of PTB.

## Results

### Phylogenomics of *S. enterica* serotype Paratyphi B PG1

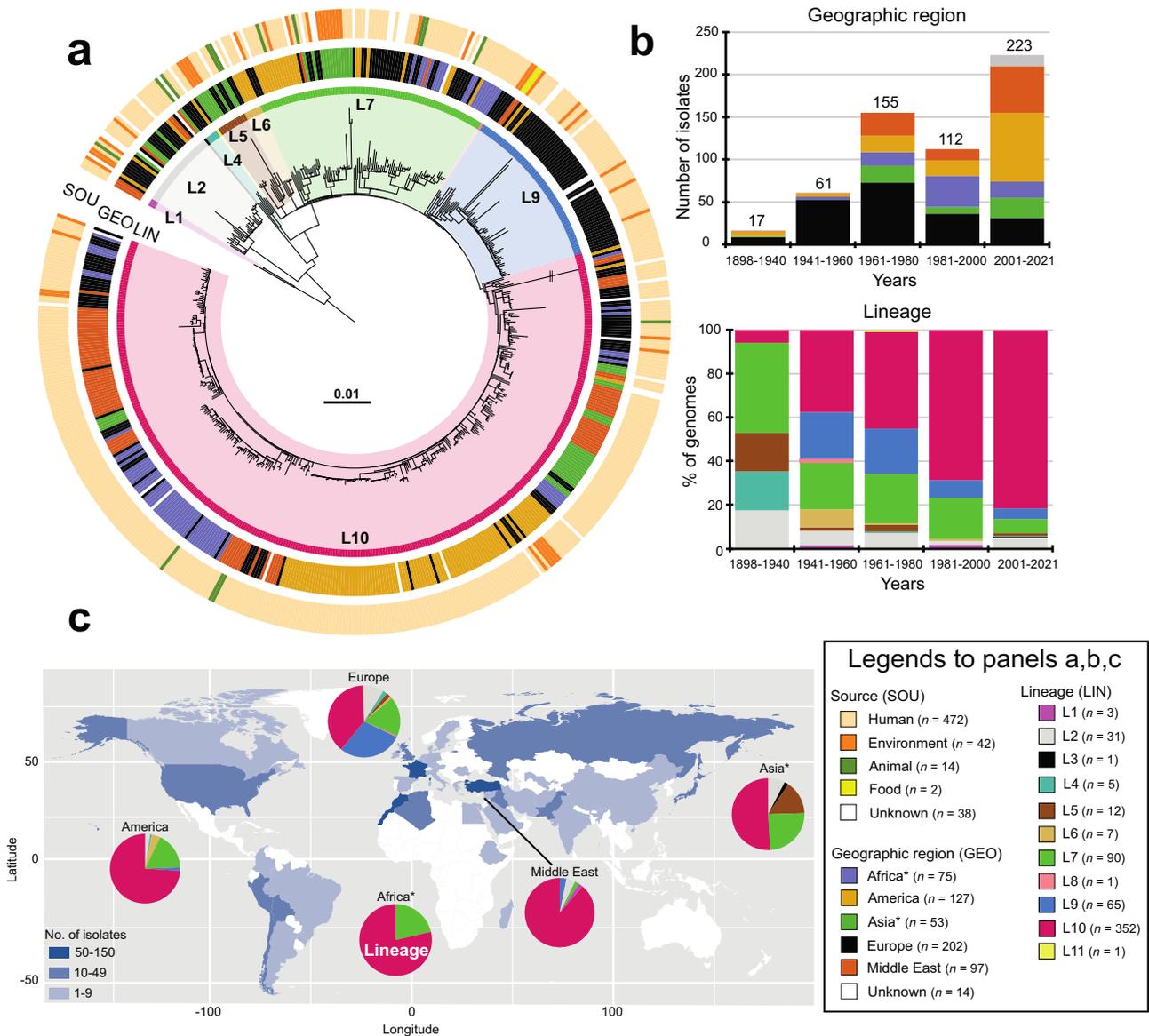
We assembled a set of 568 genomes (the “diversity dataset”), including 446 generated specifically for this study, from the widest possible temporal and geographic distribution of available SPB<sup>-</sup> isolates. These isolates originated from different sources (humans, environment, food, animals), geographic areas (41 countries spanning four continents) and were collected between 1898 and 2021. The number of historical isolates was significant, with 41% (233/568) collected between 1898 and 1980 (Fig. 1, Supplementary Data 1). We ensured that

this diversity dataset comprised exclusively genomes (i) with the correct in silico serotype, (ii) containing the specific SNV associated with the loss of *d*-Tar fermentation in SPB<sup>-</sup> (ref. 22), and (iii) belonging to the invasive lineage, PG1 (ref. 27). This was achieved in a straightforward manner with the HC400\_1620 cluster of the Enterobase *Salmonella* core-genome MLST scheme (<https://enterobase.warwick.ac.uk/species/index/senterica>) used as a proxy (Supplementary Note “Validation of the SPB<sup>-</sup> PG1 diversity dataset”, Supplementary Fig. 1).

A maximum likelihood (ML) phylogeny of these 568 SPB<sup>-</sup> PG1 genomes was constructed from 15,995 single-nucleotide variants (SNVs) distributed over the non-repetitive, non-recombinant core genome (Fig. 1a, Supplementary Fig. 2, Supplementary Data 2). Eleven lineages were identified (L1 to L11), one of which (L10) predominated, accounting for 62% (352/568) of the isolates. The frequency of L10 increased sharply over the study period, from 6.1% (2/33) between 1898 and 1950 to 81.6% (182/223) for the 2001–2021 period (Fig. 1b). L10 was found worldwide, accounting for 38.6% (78/202) of the European isolates (i.e., isolates recovered from infections acquired in Europe) to 88.7% (86/97) of the Middle Eastern isolates (Fig. 1c). Lineage L7 was also widely distributed and was detected in isolates from all geographic regions other than the Middle East (Fig. 1c). Lineage L5 was more frequent in Asia (particularly East Asia), whereas L2 and L9 were more frequently identified in Europe. The only strain (116 K) of serotype Onarimon (an O:9 antigen variant of SPB<sup>-</sup>) – isolated in Japan in 1935 – belonged to lineage L5.

Lineages were further subdivided into well-supported monophyletic groups at various hierarchical levels, including clades, subclades, and an additional higher-resolution group on the basis of both hierBAPS and visual inspection (Fig. 2a). In total, we defined 38 hierarchical genotypes with a phylogenetically informative nomenclature of the form [lineage].[clade].[subclade].[higher-resolution group] (Fig. 2a,b). Strong geographic patterns with differences from country to continent level were observed for 17 genotypes, whereas two genotypes were more widespread, isolated from two continents (Supplementary Data 1). These two genotypes were genotype 7.2 found in Europe (26.3%, 5/19) and Asia (73.7%, 14/19, mostly East Asia), and genotype 10.3.8.5 found in Europe (71.4%, 20/28) and North Africa (25%, 7/28). One genotype, 7.3.2 – also found in Europe (41.2%, 7/17) and North Africa (52.9%, 9/17) – was associated with a particular PT, BAOR (see Comparison of phylogenomics data with other typing schemes) (Supplementary Data 1). This geographic or PT information was added to the genotype nomenclature as an alias, to make it more informative.

The vast majority ( $n = 27$ , 71%) of the 38 genotypes comprised European isolates. These European isolates were found in the following genotype groups (each containing three or more isolates): 2.0, 2.1, 4, 7.3, 9.0, 9.1 France, 10.1.1\_Europe1, 10.2, and 10.3.8.5\_EuropeNorthAfrica (Table 1). Our oldest strain (CIP A214) – probably isolated by H. Conradi in Germany in 1898 – belonged to genotype 4. Middle Eastern isolates were found in 50% (19/38) of the genotypes, with particularly high frequencies in genotype 1, 2.1.1\_Turkey1, 10.1, 10.3.2\_MiddleEast1, 10.3.3\_Turkey2, 10.3.5\_MiddleEast2, 10.3.8.2\_Turkey3, 10.3.8.3\_MiddleEast3, and 10.3.8.4\_MiddleEast4 groups. American isolates belonged to 34.2% (13/38) of the genotypes and were particularly frequent in the genotype 6, 7.1.1\_Chile1, 10.3.4\_Chile2, and 10.3.6\_SouthAmerica groups. African (excluding Egypt) and Asian (excluding Middle Eastern countries; see the footnotes of Table 1) isolates were each present in 23.7% (9/38) of the genotype groups. The African isolates, which were almost exclusively from North Africa (only three were from East Africa), were most frequently of genotypes 7.3.1\_NorthAfrica1, 7.3.2\_BAOR, 10.3.7\_NorthAfrica2, and 10.3.9\_NorthAfrica3. Two East African isolates from Madagascar collected in 1962 and 2001 were of genotype 10.2; whereas a third isolate collected from Ethiopia in 2018 was of genotype 10.0 (Supplementary Data 1). The South Asian isolates belonged exclusively to the genotypes of lineage L10 (in particular



**Fig. 1 | Phylogeny, temporal, geographic, and source distribution of the 568 SPB PGI isolates from the diversity set (1898 – 2021).** **a** A circular maximum likelihood phylogeny (rooted on ancestral lineage L1 genome NCTC 8299) for the 568 SPB PGI isolates from the diversity dataset. The double slash (//) indicates an artificial shortening of this branch for visualisation. The rings show the associated information for each isolate, according to its position in the phylogeny, from the innermost to outermost, in the following order: (1) lineage (LIN); (2) geographic region (GEO); (3) and source (SOU). Lineages are labelled LX, where X is the lineage number. Lineages L3, L8, and L11, which contain only singletons are not labelled. The tips of the tree are highlighted according to lineage with a lighter hue of the colour used in the innermost ring (LIN). The scale bar indicates the number of

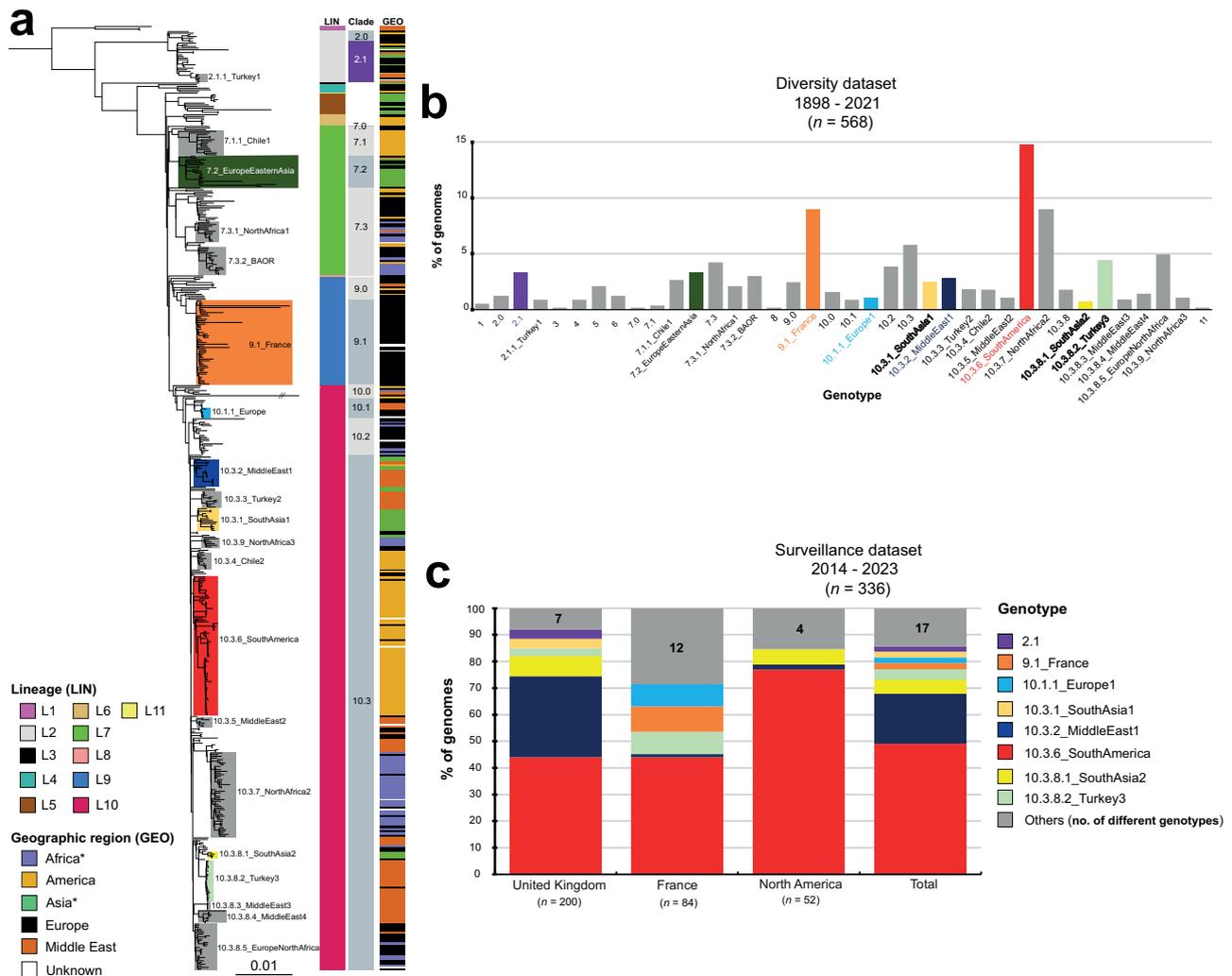
substitutions per variable site (SNVs). **b** The stacked bar chart on the left shows the distribution of the 568 isolates by geographic region and time period, and the stacked bar chart on the right shows the frequencies of the lineages for the same time periods. **c** Number of isolates per country (map) and frequencies of the lineages by world region (pie charts). An asterisk indicates the reassignment of some African and Asian countries to the Middle East (see Table 1). The map was drawn in R with the “ggplot2” package world map data from Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>. Source data are provided as a Source Data file.

10.3.1\_SouthAsia1 and 10.3.8\_SouthAsia2), whereas almost all the East Asian isolates belonged to genotypes from older lineages (2.1, 5, and 7.2\_EuropeEasternAsia). The genotype distribution for animal isolates did not differ significantly from that for human isolates (Supplementary Note “Genotypes found in non-human SPB PGI isolates”).

After confirming the presence of a strong temporal signal in our dataset (Supplementary Fig. 3), we applied a Bayesian phylogenetic approach to a spatially and temporally representative subset of 256 isolates to estimate the nucleotide substitution rates and divergence times of the different lineages (Table 2) and to construct a dated phylogeny (Fig. 3 and Supplementary Fig. 4). We estimated the

genome-wide substitution rate at  $1.2 \times 10^{-7}$  substitutions site<sup>-1</sup> year<sup>-1</sup> (95% credible interval (CI) =  $9.6 \times 10^{-8}$  to  $1.5 \times 10^{-7}$ ), giving a most recent common ancestor (MRCA) for all SPB PGI in our collection dating back to 1274 CE (common era) (95% CI, 915–1583). The MRCAs of the different lineages were estimated to have existed in the 18th century or first half of the 19th century (Table 2).

**Comparison of phylogenomics data with other typing schemes**  
World surveys of SPB PTs were regularly reported over several decades following World War II (Supplementary Fig. 5). However, the phylogenetic value of the typing scheme used at the time has never been



**Fig. 2 | Identification of the 38 hierarchical genotypes of SPB' PGI and their distribution between the diversity and surveillance datasets.** **a** Maximum likelihood phylogeny for the 568 SPB' PGI isolates (as in Fig. 1a, but not circular). The main genotypes are labelled and coloured. Columns on the right indicate the lineage (LIN) (see inset legend), clade and geographic origin (GEO) (see inset legend) of the isolates. An asterisk indicates the reassignment of some African and

Asian countries to the Middle East (see Table 1). **b** Frequencies of the 38 genotypes for the 568 genomes of the diversity dataset. The colours are similar to those used in panel “a”. **c** Stacked bars indicate the relative abundance of each genotype – coloured as in the legend, inset – for the 336 recent isolates from the UK, France and North America (surveillance dataset). Source data are provided as a Source Data file.

assessed. We, therefore, analysed the correlation between phage-typing and genomic data for the 254 SPB' isolates from the diversity dataset for which phage-typing results were available (Fig. 4). Good concordance was observed for the isolates of four PTs: 92.8% (13/14) of PT 1 isolates belonged to lineage L2, 95.8% (91/95) of those typed as Taunton belonged to L10, and 100% of BAOR (11/11) and Jersey (6/6) isolates belonged to lineage L7 (Supplementary Data 1). The isolates of the BAOR and Jersey PTs even belonged to single genotypes, 7.3.2\_BAOR and 7.3, respectively. By contrast, isolates typed as PT Taunton belonged to 10 different genotypes within L10 (Supplementary Data 1). The only PT 2 isolate belonged to L1. The isolates of the remaining PTs were not assigned to a predominant lineage and were instead considered to belong to four lineages. For example, 50.1% (32/63) of the isolates typed as Dundee belonged to L9 (all isolates were from Europe, and all but one belonged to genotype 9.1\_France), 46% (29/63) belonged to L10 (from multiple genotypes, mostly corresponding to isolates from the Middle East and North Africa), and 1.6% (2/63) belonged to L2.

Using MLST7, Achtman and coworkers<sup>21</sup> showed that ST86 and five SLVs of ST86 contained only SPB isolates containing the *d-Tar*

-specific SNV. We used this same MLST7 scheme – now implemented in Enterobase – to analyse the 568 genomes of the diversity dataset. The vast majority of these genomes (96.7%, 549/568) belonged to ST86 (Supplementary Data 1). However, 17 of these genomes belonged to 13 different SLVs of ST86. These 17 isolates included the three lineage L1 genomes, all three belonging to ST5113. One genome (13–80) could not be typed by MLST due to an incomplete *purE* gene. Finally, one genome (ERR129867) described by Connor and coworkers<sup>27</sup> was a triple-locus variant of ST86 (ST2134).

With a higher-resolution cgMLST HC level, such as HC200, it was possible to identify lineages L1 to L4 (L1, HC200\_137805; L2, HC200\_17706; L3, HC200\_301037; L4, HC200\_12575) (Supplementary Fig. 6, Supplementary Data 1). However, all other isolates belonging to lineages L5 to L11 were assigned to the same HC200 cluster (HC200\_1620), with the exception of three isolates previously found to be outliers in the root-to-tip analysis, which were each assigned to a unique HC200 cluster. With HC100 to HC50, it was not possible to recognise SPB' population structure as determined by our core-genome SNV-based phylogenetic analysis. For example, the isolates belonging to the emerging South American genotype 10.3.6 were

**Table 1 | Geographic distribution of the 38 genotypes found in the diversity dataset**

Genotype	n	Africa* n (%)	America n (%)	Asia** n (%)	Europe n (%)	Middle East n (%)	Unknown n (%)
1	3					<b>3 (100%)</b>	
2.0	7		1 (14.3%)		<b>6 (85.7%)</b>		
2.1	19		2 (10.5%)	4 (21.1%)	<b>11 (57.9%)</b>	1 (5.3%)	1 (5.3%)
2.1.1_Turkey1	5				1 (20.0%)	<b>3 (60.0%)</b>	1 (20.0%)
3	1			<b>1 (100%)</b>			
4	5		1 (20.0%)		<b>4 (80.0%)</b>		
5	12			<b>8 (66.7%)</b>	4 (33.3%)		
6	7		<b>5 (71.4%)</b>		2 (28.6%)		
7.0	1				<b>1 (100%)</b>		
7.1	2				<b>2 (100%)</b>		
7.1.1_Chile1	15		<b>15 (100%)</b>				
7.2_EuropeEasternAsia	19			<b>13 (68.4%)</b>	5 (26.3%)	1 (5.3%)	
7.3	24		6 (25.0%)		<b>17 (70.8%)</b>		1 (4.2%)
7.3.1_NorthAfrica1	12	<b>7 (58.3%)</b>			4 (33.3%)	1 (8.3%)	
7.3.2_BAOR	17	<b>9 (52.9%)</b>	1 (5.9%)		7 (41.2%)		
8	1				<b>1 (100%)</b>		
9.0	14		2 (14.3%)		<b>10 (71.4%)</b>	2 (14.3%)	
9.1_France	51				<b>48 (94.1%)</b>		3 (5.9%)
10.0	9	1 (11.1%)	2 (22.0%)		4 (44.4%)	2 (22.2%)	
10.1	5				2 (40.0%)	<b>3 (60.0%)</b>	
10.1.1_Europe1	6				<b>4 (66.7%)</b>	1 (16.7%)	1 (16.7%)
10.2	22	5 (22.7%)			<b>16 (72.7%)</b>		1 (4.5%)
10.3	33		2 (6.1%)	7 (21.2%)	15 (45.5%)	9 (27.3%)	
10.3.1_SouthAsia1	14			<b>13 (92.9%)</b>		1 (7.1%)	
10.3.2_MiddleEast1	16		1 (6.3%)	3 (18.8%)		<b>12 (75.0%)</b>	
10.3.3_Turkey2	10					<b>10 (100%)</b>	
10.3.4_Chile2	10		<b>10 (100%)</b>				
10.3.5_MiddleEast2	6					<b>5 (83.3%)</b>	1 (16.7%)
10.3.6_SouthAmerica	84		<b>79 (94.0%)</b>		3 (3.6%)		2 (2.4%)
10.3.7_NorthAfrica2	51	<b>41 (80.4%)</b>			8 (15.7%)		2 (3.9%)
10.3.8	10	1 (10.0%)			4 (40.0%)	5 (50.0%)	
10.3.8.1_SouthAsia2	4			<b>4 (100%)</b>			
10.3.8.2_Turkey3	25				1 (4.0%)	<b>24 (96.0%)</b>	
10.3.8.3_MiddleEast3	5					<b>5 (100%)</b>	
10.3.8.4_MiddleEast4	8					<b>8 (100%)</b>	
10.3.8.5_EuropeNorthAfrica	28	6 (21.4%)			<b>20 (71.4%)</b>	1 (3.6%)	1 (3.6%)
10.3.9_NorthAfrica3	6	<b>5 (83.3%)</b>			1 (16.7%)		
11	1				<b>1 (100%)</b>		
<b>Total</b>	<b>568</b>	75	127	53	202	97	14

n number of isolates. \*the isolates from Egypt were assigned to the Middle East. \*\*the isolates from Turkey, Iraq, Lebanon, Syria, Mandatory Palestine, and Saudi Arabia were assigned to the Middle East. If a genotype is found at a percentage >50% in a particular geographic region, the data are indicated in bold.

found in three different HC50 clusters. The predominant cluster, HC50\_1857, was not even specific to 10.3.6 (also found in isolates of five other genotypes) and the 10.3.6 isolates within HC50\_1857, were further split between 15 different HC20 clusters, making them hard to track by cgMLST (Supplementary Figs 6 and 7, Supplementary Data 1).

### Evolution of antimicrobial resistance

The emergence of antimicrobial resistance (AMR) in SPB PGI is recent (Fig. 5, Supplementary Data 1). Between 1898 and 2000, only one isolate (0.3%, 1/345) had antibiotic resistance genes (ARGs). This human isolate (B73-1117), collected in France in 1973, displayed resistance to ampicillin (*bla*<sub>TEM-1D</sub>), streptomycin (*strAB*, *aadA1*, and *aadA2b*), sulfonamides (*sulI*), chloramphenicol (*cmIA1*), and tetracycline (*tetA*). Between 2001 and 2021, 23.1% (52/223) of isolates had

ARGs. One isolate acquired in Turkey in 2001 (01-7995) produced a CTX-M-3 extended-spectrum beta-lactamase<sup>30</sup>, whereas another isolate (P7704) acquired in South America in 2019 produced an OXA-48 carbapenemase<sup>31</sup>. One isolate (83282), acquired in South America in 2014, contained the *mph(A)* gene encoding a macrolide 2' phosphotransferase, a common determinant of azithromycin resistance (however, the antimicrobial susceptibility pattern of this isolate was unavailable). All these rare isolates contained AMR plasmids (Supplementary Data 1); however, the mechanisms of resistance (21.5%, 48/223) most prevalent during the 2001–2021 period involved mutations of the quinolone resistance-determining regions of the chromosomal *gyrA* and *gyrB* DNA gyrase genes leading to resistance to nalidixic acid and/or decreased susceptibility or resistance to ciprofloxacin (minimum inhibitory concentration [MIC] between 0.125 and 0.5 mg/L). The

**Table 2 | Dating of the main lineages and genotypes of SPB<sup>+</sup> PG1 with BEAST2**

Main lineages and genotypes	Time (year) of the MRCA (95% HPD)
All 256 SPB <sup>+</sup> PG1	1274 (915–1583)
Lineage 1	1757 (1592–1894)
Lineage 2	1813 (1745–1874)
Lineage 4	1740 (1641–1831)
Lineage 5	1727 (1660–1787)
Lineage 6	1853 (1797–1899)
Lineage 7	1765 (1709–1819)
7.1.1_Chile1	1831 (1776–1887)
7.2_EuropeEasternAsia	1840 (1791–1882)
7.3.1_NorthAfrica1	1896 (1861–1931)
7.3.2_BAOR	1896 (1866–1925)
Lineage 9	1768 (1711–1819)
9.1_France	1913 (1887–1933)
Lineage 10	1793 (1740–1842)
10.3.1_SouthAsia1	1926 (1900–1950)
10.3.2_MiddleEast1	1913 (1886–1937)
10.3.4_Chile2	1946 (1921–1969)
10.3.5_MiddleEast2	1930 (1904–1952)
10.3.6_SouthAmerica	1918 (1888–1942)
10.3.7_NorthAfrica2	1932 (1916–1947)
10.3.8.1_SouthAsia2	1993 (1978–2005)
10.3.8.2_Turkey3	1994 (1985–2002)
10.3.9_NorthAfrica3	1930 (1907–1951)

MRCA, most recent common ancestor; HPD, highest posterior density interval.

first isolate harbouring such a mutation was isolated in 2004. Diverse mutations were observed, with *gyrA*\_S83F in 20 isolates (9.0%, 20/223), *gyrA*\_D87Y in 12 isolates (5.4%, 12/223), *gyrA*\_D87N in six isolates (2.7%, 6/223), *gyrA*\_D82N in three isolates (1.3%, 3/223), *gyrB*\_S464F in three isolates (1.3%, 3/223), *gyrA*\_D87G in two isolates (0.9%, 2/223), a combination of *gyrA*\_D87N and *gyrB*\_S464F in one isolate (0.4%, 1/223), and a combination of *gyrA*\_S83F and *gyrA*\_D87G in one isolate (0.4%, 1/223). For this last isolate (CLA-76-5), the antimicrobial susceptibility pattern of which was unavailable, we cannot rule out resistance to ciprofloxacin (MIC > 0.5 mg/L). These *gyrA* and *gyrB* mutations occurred over the entire phylogenetic tree, in lineages L2 ( $n = 4$ ), L3 ( $n = 1$ ), L5 ( $n = 2$ ), L7 ( $n = 3$ ), L9 ( $n = 5$ ), and L10 ( $n = 33$ ) (Fig. 5). The lineage L10 isolates containing such *gyrA* and *gyrB* mutations were further classified into eight different genotypes. One cluster contained seven *gyrA*\_D87Y isolates acquired in Turkey between 2009 and 2017 (genotype 10.3.8.2\_Turkey3) (Supplementary Data 1).

### Pan-genome analysis and *sopE* prophages

A pan-genome analysis of the 568 SPB<sup>+</sup> PG1 genomes studied (Supplementary Fig. 8) identified a closed pan-genome ( $\alpha = 1.3$ ) containing a total of 5,681 genes, including a core genome of 4044 genes (Supplementary Data 3).

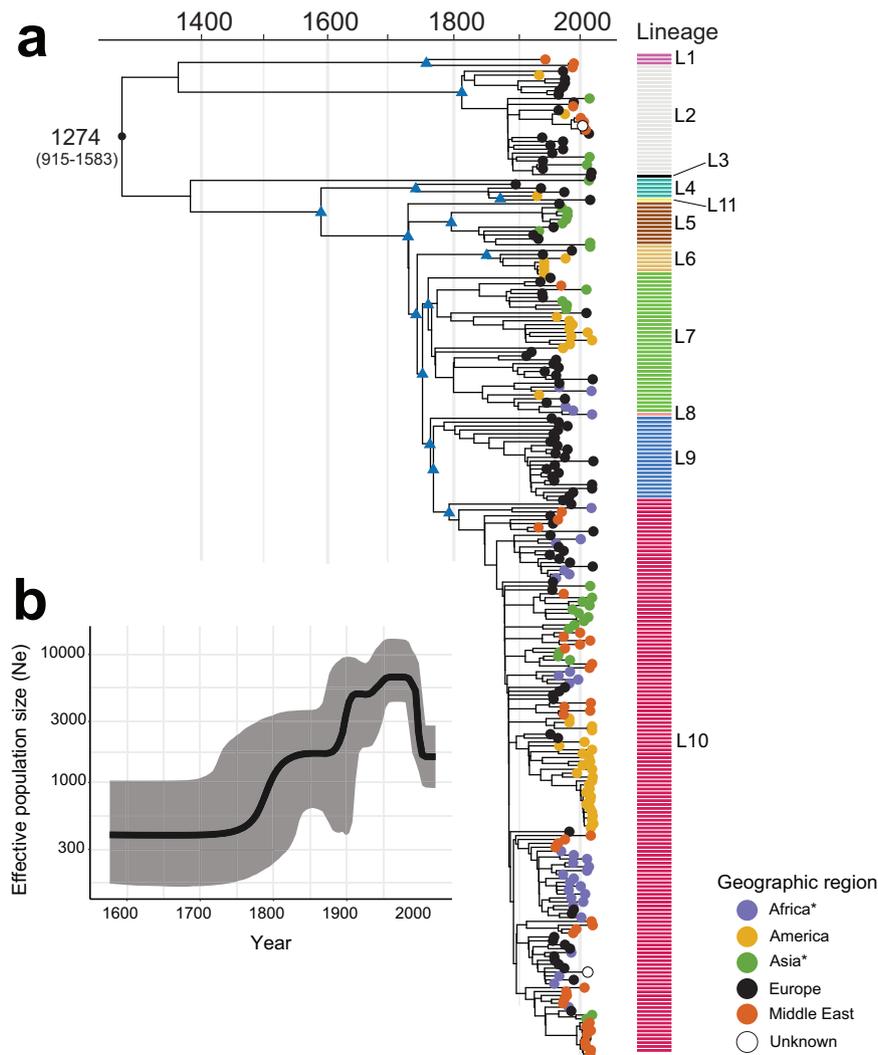
Of the 1506 accessory genes present in <95% of the genomes, 696 (46.2%), 242 (16.1%), and 28 (1.9%) were found to belong to prophages, plasmids (with or without AMR genes), and transposases, respectively (Supplementary Data 4). The mapping of these accessory genes onto the core genome phylogeny revealed an absence of phylogenetic clustering for plasmid genes, whereas some prophage genes were (sub)lineage-specific (Supplementary Fig. 9). We then evaluated the occupancy of the 10 prophage insertion sites – identified in the 14 complete genomes – with short-read assemblies from the entire diversity dataset (Supplementary Note “Prophages of SPB<sup>+</sup> PG1”, Supplementary Fig. 10).

Three (sites #1 to 3) of the 10 sites were occupied by prophages containing *sopE*, a virulence gene that encodes an effector translocated into eukaryotic cells that promotes bacterial invasion through cytoskeleton rearrangement. The nucleotide sequence of the *sopE* gene was 100% identical in the 14 complete genomes, regardless of the prophage into which it was inserted, with the exception of the *sopE* gene inserted at site #3 of the B624 genome, which differed from the *sopE* consensus sequence of SPB<sup>+</sup> by eight (five being non-synonymous) of the 723 nucleotides. Two types of *sopE* prophages, both belonging to class *Caudoviricetes*, were identified; the first was 41,250 bp to 44,615 bp in size and displayed 89.8% to 95.9% nucleotide identity (38% to 55% coverage) to the *Salmonella Brunovirus* SEN34 (GenBank accession no. [NC\\_028699](https://doi.org/10.1038/s41467-024-54418-4)); the second was 34,723 bp long and displayed ~98% nucleotide identity (83% coverage) to the enterobacterial *Xuanwuivirus* P88 (GenBank accession no. [NC\\_026014](https://doi.org/10.1038/s41467-024-54418-4)) (Supplementary Table 1).

The *sopE*-containing SEN34-like prophages occupied sites #1 and #2, whereas the P88-like prophage occupied site #3 (Fig. 6a). This *sopE*-prophage content was correlated with SPB<sup>+</sup> phylogeny, with a *sopE*-containing SEN34-like prophage inserted at site #1 in lineages L1 to L4 and at site #2 in lineages L5 to L11 (Fig. 6b). The P88-like prophage was not seen alone but always in addition to the SEN34-like prophage and its presence was strongly associated with the Dundee PT (Chi squared test,  $p < 0.0001$ ) (Supplementary Fig. 10). All but two (173/73 and 73/67) of the 568 genomes had at least one *sopE* prophage: most genomes (75.1%, 425/566) had one, some (17.8%, 101/566) had two and a few (7.1%, 40/566) had three (Fig. 6). All genomes containing three *sopE* prophages belonged to genotype 9.1\_France (associated with the Dundee PT).

### Development of a new SNV-based genotyping tool and analysis of recent SPB<sup>+</sup> isolates

We identified marker SNVs unique to each genotype (38 SNVs in total) (Supplementary Data 5) and implemented this genotyping scheme on the open-source Mykrobe platform (<https://www.mykrobe.com/>). Mykrobe checks (i) for the presence of *invA* to ensure that the genome belongs to the genus *Salmonella*, (ii) and then for the presence of the STM 3356 *d-Tar*-specific SNV to confirm that the genome belongs to SPB<sup>+</sup>, and (iii) finally assigns genomes to genotypes based on presence of the 38 genotype-specific SNVs. After validation of the Mykrobe-implemented version of the scheme on the 568 genomes from the diversity dataset (Supplementary Note “Development of a new SNV-based genotyping tool for SPB<sup>+</sup> PG1”), this genotyping tool was used on the 336 genomes from the surveillance dataset – originating from public health laboratories in the UK ( $n = 200$ ), France ( $n = 84$ ), USA ( $n = 39$ ), and Canada ( $n = 13$ ), with isolation dates between 2014 and 2023 – to identify genotypes recently isolated in these four countries (Fig. 2c, Supplementary Data 6). During the 2014 – 2023 period, 25 genotypes were observed. Genotype diversity was highest among French isolates (17 genotypes for 84 isolates) and lowest in the UK (13 genotypes for 200 isolates). The most frequent genotype was 10.3.6\_SouthAmerica, found in 49.1% (164/336) of the isolates from all four countries. Its frequency ranged from 44% (125/284) for the isolates collected in Europe to 76.9% (40/52) for the isolates collected in North America. The second most frequent genotype, 10.3.2\_MiddleEast1 (18.8%, 63/336), was found mostly in the UK. The other genotypes were found in less than 10% of the isolates. There was an overrepresentation of genotypes linked to South Asia in the UK (e.g., 10.3.1\_SouthAsia1 and 10.3.8.1\_SouthAsia2, together accounting for 11% (22/200) of the isolates), and of genotypes linked to North Africa in France (e.g., 7.3.1\_NorthAfrica1, 7.3.2\_BAOR, 10.3.7\_NorthAfrica2 and 10.3.9\_NorthAfrica3, together accounting for 13.1% (11/84) of the isolates). Interestingly, older genotypes – such as 2.1, accounting for 3.5% (7/200) of UK isolates, and 9.1\_France, accounting for 9.5% (8/84) of French isolates – are still being isolated.



**Fig. 3 | Timed phylogeny of a representative subsample of 256 SPB<sup>+</sup> PG1 isolates.** **a** Maximum clade credibility tree produced with BEAST2 (optimised relaxed clock model; Bayesian skyline) with the tips coloured according to the geographic origin of the isolates (see inset). Selected nodes supported by posterior probability values > 0.9 are shown as blue triangles. The estimated age of the MRCA (with 95% confidence intervals in parentheses) is shown. The lineage of these isolates is

indicated at the right side of the tree. The scale bar indicates the number of substitutions per variable site (SNVs). An asterisk indicates the reassignment of some African and Asian countries to the Middle East (see Table 2). **b** Bayesian skyline plot showing temporal changes in effective population size ( $N_e$ ) (black curve) with 95% confidence intervals (grey shading).

## Discussion

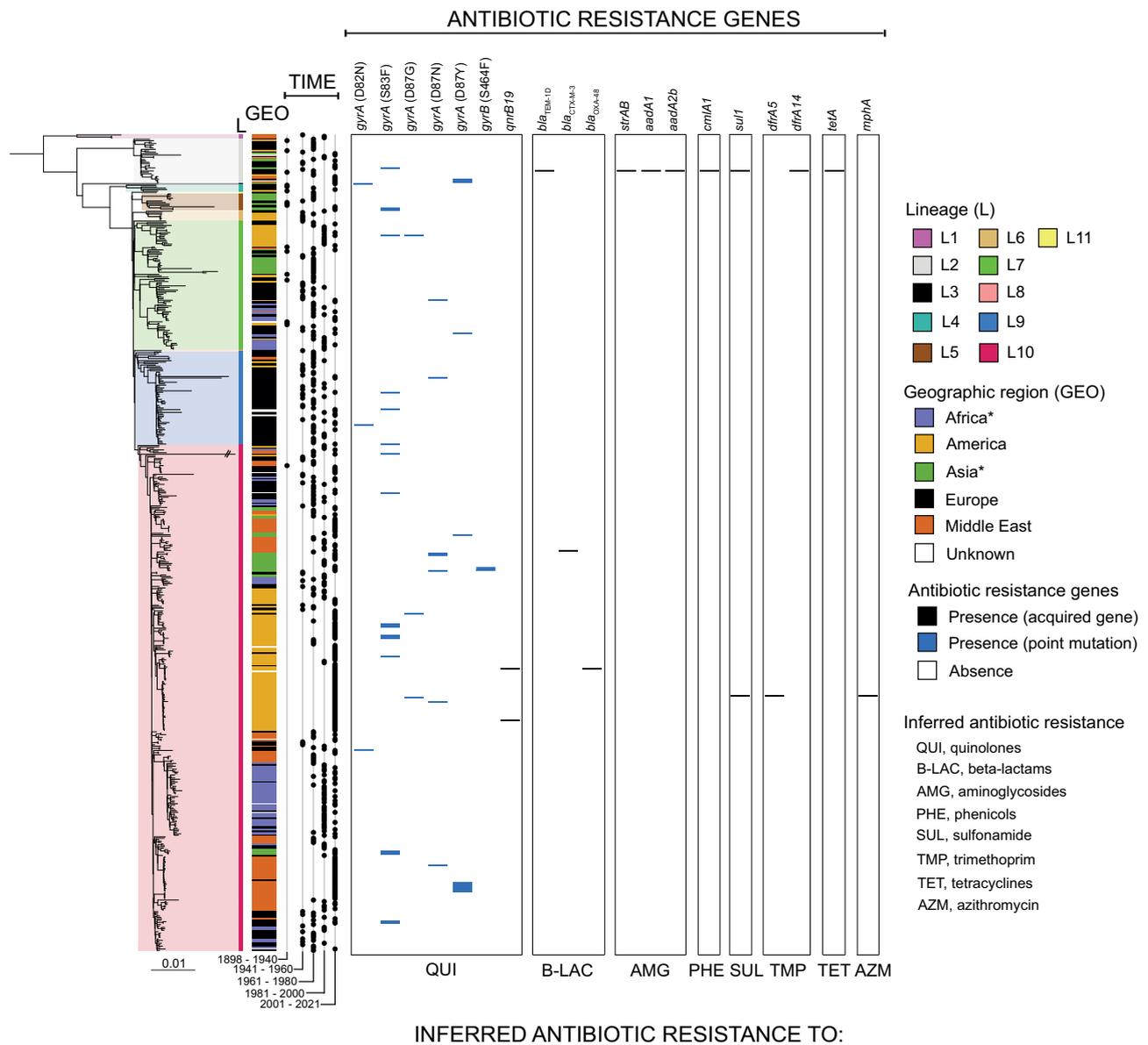
Global phylogenomic studies of bacterial pathogens can be strongly affected by sampling biases, such as a lack of bacterial strains from certain geographic areas and periods of time. We tried to minimise these sampling biases, by ensuring that we studied a spatiotemporally representative set of SPB<sup>+</sup> strains selected after (i) a comprehensive search of the medical and scientific literature since the first report of PTB, (ii) an analysis of SPB<sup>+</sup> strains available from an international network of reference laboratories with collections of historical isolates, and (iii) a search for SPB<sup>+</sup> genomes among the >400,000 *Salmonella* genomes present in the EnteroBase database.

Based on our dataset of 568 SPB<sup>+</sup> PG1 genomes, we estimated the age of this pathogen at ~750 years (1274 CE; 95% CI, 915–1583), which is very close to the previous median date of origin estimated by Connor and coworkers<sup>27</sup> (1188 CE; 95% CI, 469 BC – 1799 CE), who used only 25 SPB<sup>+</sup> PG1 genomes (i.e., those with a known year of isolation). SPB<sup>+</sup> is older than SPA, which is estimated to have originated 450–700 years ago<sup>25</sup>. SPA was discovered two years after SPB (in the USA in 1898)<sup>1,2</sup> but is currently the most frequent agent of paratyphoid fever<sup>28</sup>. Due to lineage extinction, in particular, times to the MRCA are often

underestimated and the inclusion of ancient DNA in the analysis would increase precision and make it possible to establish dates of origin further in the past<sup>25</sup>. The dating of a representative collection of modern isolates of SPC estimated the origin of SPC to 456–664 years ago. When a draft SPC genome from an 800-year-old Norwegian skeleton was added to the analysis, the time to the MRCA increased to 1162–1526 years<sup>26</sup>. Unfortunately, no ancient DNA is currently available for SPB<sup>+</sup> strains.

SPB<sup>+</sup> strains, first identified in Europe in 1896, became a common cause of enteric fever across Europe thereafter, suggesting that they were circulating in this particular region of the world. Indeed, SPB<sup>+</sup> genetic diversity was greatest among European isolates and those collected in Turkey. Turkey, part of which lies in Europe, has a long shared history with the rest of Europe through the rule of the Ottoman Empire over most of the countries in south-eastern Europe for several centuries. SPB<sup>+</sup> isolates were, however, less frequently reported in other parts of the world (the Americas, East Asia, North Africa) early in the 20th century. In Morocco, North Africa, SPB<sup>+</sup> strains were reported to have been introduced around the city of Fes by the French troops during the Rif war in 1925 (ref. 32). Two genotypes may have been





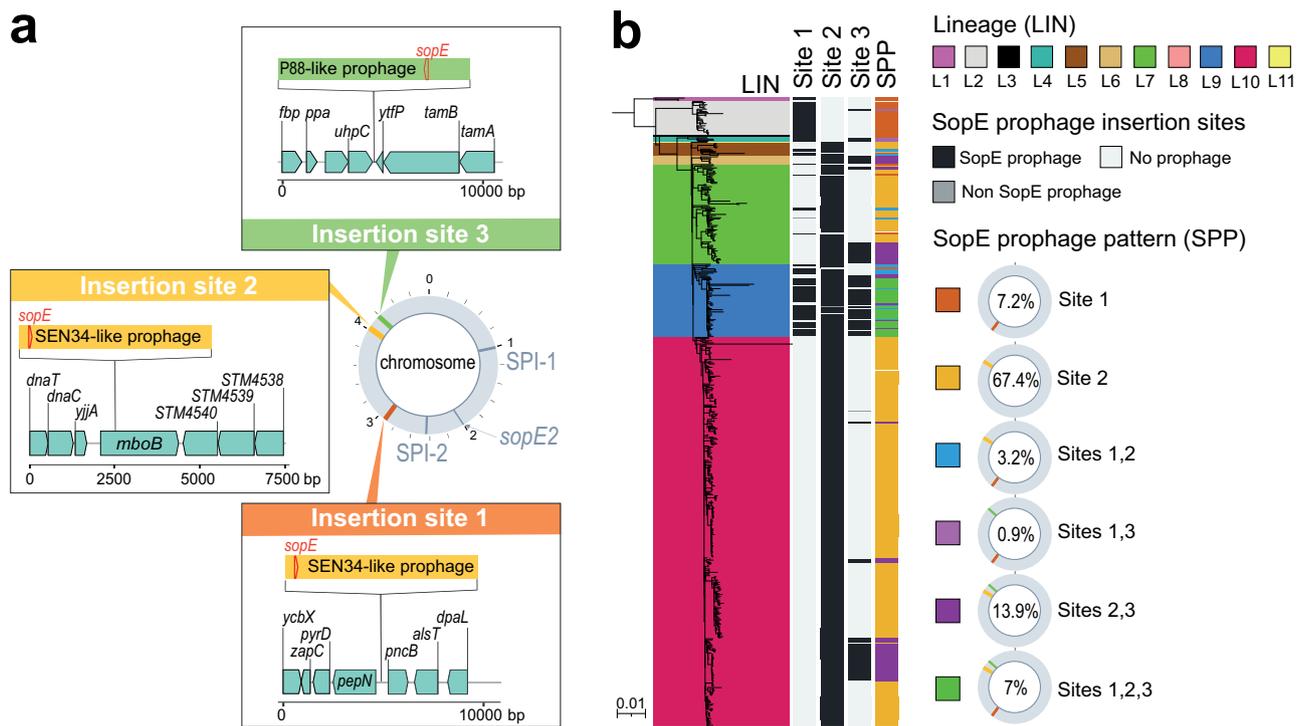
**Fig. 5 | Genomic characterization of antibiotic resistance genes in SPB PGI.** Distribution of antibiotic resistance genes by phylogeny (as in Fig. 3a), geography, and time period. Genes acquired via horizontal gene transfer are indicated in black

and those acquired via chromosomal mutation are indicated in blue. An asterisk indicates the reassignment of some African and Asian countries to the Middle East (Table 1).

not very frequent and have only recently been detected in South Asia – where selection pressures exerted by quinolones and, later, fluoroquinolones, began early – potentially accounting for this lower level of AMR than in STY and SPA strains.

The presence of a particular *sopE* bacteriophage ( $\Phi$ SopE309) was previously used by Prager and coworkers<sup>37</sup> to distinguish systemic SPB<sup>+</sup> isolates from enteric SPB<sup>+</sup> isolates. By contrast, Connor and coworkers<sup>27</sup> suggested that the *sopE* gene was not a suitable marker for identifying SPB<sup>+</sup> PGI isolates because this gene was not present in all of their PGI isolates and some of these isolates had a gene homologous to *sopE* also found in all other SPB PGs. As  $\Phi$ SopE309 was initially isolated from SPB<sup>+</sup> strain B309 – also included in our study – we were able to determine that it was actually the Brunovirus SEN34-like prophage. Our in-depth analysis supports the recommendation of Prager and coworkers<sup>37</sup> to use *sopE* as a marker gene for SPB<sup>+</sup> because the *sopE*-carrying SEN34-like prophage was present in almost all the 568 SPB PGI isolates from our diversity dataset (including 28 genomes from the study by Connor and coworkers<sup>27</sup>). Only two isolates were devoid of

*sopE* and each of the three insertion sites for *sopE* prophages were empty in these isolates. These two isolates were collected more than 50 years ago and we cannot therefore rule out the possibility that the *sopE* prophage was excised during storage or subculture. Furthermore, according to their nearest neighbours on the phylogenetic tree, these isolates would have contained only one *sopE* prophage. The gene homologous to *sopE* unexpectedly found in some SPB<sup>+</sup> PGI genomes by Connor and coworkers<sup>27</sup> is almost certainly *sopE2*, a chromosomal gene present in all *Salmonella* lineages and encoding an effector protein, Sophe2, 69% identical to Sophe<sup>38</sup>. By contrast, the *sopE* gene, carried by different bacteriophages (lambda-like Gifsy-2, P2-like, mTmV), was previously found in only some serotypes of *Salmonella* (e.g., Gallinarum, Typhi, Dublin, Heidelberg) or some strains of certain serotypes (e.g., epidemic *S. enterica* serotype Typhimurium DT204 in the 1970s or the currently dominant monophasic *S. enterica* serotype Typhimurium ST34)<sup>39–41</sup>. It has been suggested that the acquisition of *sopE* by lysogenic conversion increases the fitness of *S. enterica* serotype Typhimurium, thereby contributing to the emergence of epidemic



**Fig. 6 | Lineage-specific accumulation of *sopE* prophages in SPB PGI.** **a** Three insertion sites occupied by *sopE* prophages were identified in the 14 complete genomes of SPB PGI isolates. Prophages of the SEN34 family [40.89–44.3 kb] were found at insertion sites #1 and/or #2. The P88-family prophage [34.3 kb] was found at insertion site #3. Further details concerning the insertion sites are available from Supplementary Data 8. Gene arrow maps were generated with the *gggenes* v.0.5.0

and *ggplot2* v.3.4.2 packages of R v.4.1.2 software. **b** The 568 genomes from the diversity dataset were screened for an absence of insertions at sites #1, #2 and #3. Presence/absence is colour-coded in black and light grey, respectively; dark grey indicates the presence of a potential *sopE*-free prophage. Six types of prophage insertion were recorded across the 11 lineages (Supplementary Data 9).

strains<sup>42</sup>. SopE and SopE2 are G-nucleotide exchange factors that are translocated into the host cell, where they activate host cellular Rho-GTPases – such as Cdc42 and Rac1 for SopE and Cdc42 alone for SopE2 – which act as key regulators of diverse activities, such as the expression of pro-inflammatory cytokines and the organisation of the actin cytoskeleton, ultimately promoting bacterial uptake and even intracellular replication<sup>42,43</sup>. In addition to their chromosomal *sopE2* gene and their SEN-34-like prophage *sopE* gene, 17.8% and 7.1% of SPB PGI isolates contain one and two additional copies of *sopE* (carried by either the SEN-34-like or P88-like prophages), respectively. Three copies of *sopE* were observed exclusively in the PT Dundee strain (genotype 9.1 France) that was epidemic in France after WWII and is still isolated, even now, from long-term carriers. Similarly, two copies of *sopE* were previously described in *S. enterica* serotype Heidelberg isolates from a large multistate outbreak linked to turkey meat in the US<sup>44</sup> and three copies were found in a cluster of monophasic *S. enterica* serotype Typhimurium ST34 isolates from humans and pigs in the UK<sup>41</sup>. An increase in *sopE* copy number is observed in ~25% of SPB PGI isolates, but additional studies are required to determine the consequences of this increase for the expression of *sopE* and interaction with host cells.

We also provide a new robust framework for identifying and tracking the causal agent of PTB. Firstly, SPB isolates can be assigned to the 10 known PGs described by Connor and coworkers<sup>27</sup> with the Enterobase cgMLST scheme, with the HC400\_1620 cluster considered a signature of SPB PGI. This cluster has been shown to be more informative than the traditional MLST7 ST86 criterion. Even though ST86 is an excellent predictor of SPB PGI isolates, the existence in such isolates of other SLVs of ST86 (such as ST772, ST2340, ST5113, ST6558, ST7999, ST8505, ST8506, ST8931, ST8932, ST9224, ST10005, ST10013, and ST10014 and potentially others), a triple-locus variant of

ST86 (ST2134), or even the possibility of non-typability by MLST might complicate the use of MLST7 as a unique tool for identifying SPB PGI. Furthermore, the SLVs of ST86 are also observed in other SPB PGs, such as ST43 in PG3 and PG4, and ST149 in PG5 (ref. 27). If MLST7 is used, we recommend its use in combination with detection of the specific *d-Tar* SNV within STM 3356, this SNV being crucial for the formal assignment of non-ST86 genomes to SPB PGI. We have also developed a hierarchical SNV-based genotyping scheme for tracking the different SPB PGI strains, an approach previously successfully used for the surveillance of the two main agents of enteric fever, STY<sup>45</sup> and SPA<sup>46</sup>. As already reported for *S. sonnei*<sup>47</sup>, we showed that our SNV-based genotyping scheme performs better than cgMLST hierarchical clustering at higher levels of resolution for tracking epidemiologically relevant strains. Our scheme, now implemented in the open-source Mykrobe software (<https://github.com/mykrobe-tools/mykrobe>), can identify 38 different genotypes with a phylogenetically – and sometimes phylogeographically – informative nomenclature. It is, however, important to apply this scheme only to confirmed SPB PGI genomes – identified by the cgMLST HC400\_1620 cluster or using MLST7 plus the specific *d-Tar* SNV – as Mykrobe may otherwise assign the rare genotype 7.0 and to a lesser extent genotype 7.3 to most non-PGI *Salmonella* genomes (for non-PGI genomes the tool will, however, always yield “Unknown” instead of “Salmonella Paratyphi B” under the column “species” of the output table). We anticipate that the use of this scheme and its universal nomenclature will improve the laboratory surveillance of PTB. It is now easier to track the different SPB PGI populations at global scale. We were able to identify the different genotypes detected recently in travellers from or migrants to four high-income countries in North America and Europe (surveillance dataset). However, more global studies involving countries experiencing PTB across the globe and performed regularly over time – as for

the PT surveys – would be helpful to monitor the diversity, spread and evolution (of AMR in particular) of this pathogen. New genotypes with geographic information (in cases of new emerging genotypes in a defined area) or new updates on alias names (in cases of the establishment of known genotypes in new areas) should be added to the scheme in the future. The use of a common nomenclature would also be helpful during outbreak investigations. It is now straightforward to define the bacterial types of SPB PG1 and to share this information during transborder outbreak investigations. This should also facilitate the identification of transmission chains associated with particular geographic regions, particularly in areas in which PTB is not endemic. For example, genomic surveillance in the UK identified an imported SPB outbreak in travellers coinciding with a mass gathering in Iraq in 2021 (ref. 29). The isolates from these patients clustered in one of the two clades labelled “travel to Iraq”. According to our genotyping scheme, this clade corresponds to genotype 10.3.2\_MiddleEast1. The oldest isolates belonging to this genotype were collected in Iran in 1965 and Iraq in 1975, suggesting that this strain has been endemic in the region for many decades. The second clade labelled “travel to Iraq” identified by UKHSA corresponded to genotype 10.3.8.3\_MiddleEast3. In endemic regions, this genotyping scheme might also be useful for determining the genotype of a potential outbreak strain if several genotypes are known to circulate regionally.

In conclusion, using a carefully selected set of genomes from historical and contemporary isolates, we were able to unravel the population structure and evolution of SPB PG1, the agent of PTB. This pathogen, which emerged at least 750 years ago, initially thrived in Europe, but is now active in other parts of the world, frequently in areas lacking enteric fever surveillance systems. We anticipate that the widespread use of our genotyping scheme by public health laboratories will improve our understanding of the global epidemiology of this pathogen.

## Methods

### Ethics statement

This study was based exclusively on bacterial isolates (including historical and reference strains) and associated metadata. The vast majority of these isolates ( $n = 362$ ) were obtained from *Institut Pasteur* (reference laboratories or *Collection de l'Institut Pasteur* (CIP)); 115 isolates corresponded to historical or reference strains collected between 1898 and 1971 and 247 were bacterial isolates collected under the mandate for laboratory-based surveillance awarded by the French Ministry of Health to the National Reference Centre for *Escherichia coli*, *Shigella* and *Salmonella* (NRC-ESS) since 1972. Data collection and storage by the NRC-ESS was approved by the French National Commission for Data Protection and Liberties (“*Commission Nationale Informatique et Libertés* (CNIL)”; approval number: 1474659). For other human bacterial isolates collected (or the genome sequences derived from them) by participating reference laboratories under local mandates for laboratory-based surveillance of salmonellosis in line with local laws and regulations, the associated metadata did not contain any personal identifiable information and were restricted to year and country of isolation, and international travel information. As a result, neither informed consent nor approval from an ethics committee was sought given this is not research conducted on human participants.

### *S. enterica* serotype paratyphi B sequence data collections

**Diversity dataset.** We first studied a diversity dataset of 568 SPB genomes, 446 of which were generated specifically for this study, 109 had already been published<sup>27–29,31,48–50</sup>, and the other 13 were unpublished but deposited in EnteroBase (<https://enterobase.warwick.ac.uk/species/index/senterica>) or GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) (Supplementary Data 1). This diversity dataset also contained the reference strain (116 K) of *S. enterica* serotype Onarimon

(1,9,12:b:1,2) considered an O-antigen variant of SPB (Supplementary Note “Validation of the SPB- PG1 diversity dataset”).

The 446 isolates were obtained from the bacterial collections of *Salmonella* reference laboratories located at the *Institut Pasteur* (IP), Paris, France ( $n = 351$ ), NHS Greater Glasgow and Clyde (NHSGGC), Glasgow, UK ( $n = 19$ ), the National Institute of Infectious Diseases (NIID), Tokyo, Japan ( $n = 18$ ), the Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA ( $n = 15$ ), the Pasteur Institute of St Petersburg (PISP), St Petersburg, Russian Federation ( $n = 12$ ), the UK Health Security Agency (UKHSA), Colindale, UK ( $n = 10$ ), the Robert Koch Institute (RKI), Wernigerode, Germany ( $n = 7$ ), University College Hospital (UCH), Galway, Ireland ( $n = 3$ ), or from the *Collection de l'Institut Pasteur* (CIP) ( $n = 11$ ).

The 568 genomes from the diversity dataset were obtained from isolates collected between 1898 and 2021 from humans (472/568, 83.1%), environmental samples (42/568, 7.4%), animals (14/568, 2.5%), food items (2/568, 0.3%) or unknown sources (38/568, 6.7%) (Supplementary Data 1). For the human isolates, 43% (203/472) were obtained from blood, 39.2% (185/472) from stools, 3.8% (18/472) from other sources (urine, cerebrospinal fluid, pus, cysts, wounds, gallbladders, bile) and 14% (66/472) were of unknown origin. The 568 isolates and strains were isolated locally or from travellers and originated from 41 countries in Europe (202/568, 35.6%), Asia (144/568, 25.3%), the Americas (127/568, 22.3%), and Africa (81/568, 13.2%). Some historical or laboratory strains were of unknown geographic origin (14/568, 2.5%). The European isolates accounted for 57.9% (135/233) of the total isolates collected in the 1898–1980 period, decreasing to 20% (67/335) in the 1981–2022 period.

**Surveillance dataset.** We assembled a surveillance dataset of 336 SPB genomes routinely obtained by four public health microbiology laboratories (CDC, UKHSA, IP and Canada’s National Microbiology Laboratory) and submitted to EnteroBase between August 27th 2015 and May 2nd 2023. These 336 genomes included 111 already present in the “diversity dataset” (Supplementary Data 6).

### Antimicrobial drug susceptibility testing

Antimicrobial drug susceptibility was determined at IP for 273 SPB isolates from the diversity dataset (Supplementary Data 1) by disk diffusion on Mueller-Hinton (MH) agar in accordance with the guidelines of the Antibiogram Committee of the French Society for Microbiology (CA-SFM) / European Committee on Antimicrobial Susceptibility Testing (EUCAST)<sup>51</sup>. The following antimicrobial drugs (Bio-Rad, Marnes-la-Coquette, France) were tested: amoxicillin, ceftriaxone, ceftazidime, streptomycin, kanamycin, amikacin, gentamicin, nalidixic acid, ofloxacin, ciprofloxacin, sulfonamides, trimethoprim, sulfamethoxazole-trimethoprim, chloramphenicol, tetracycline, and azithromycin. *Escherichia coli* CIP 76.24 (ATCC 25922) was used as a control. The minimal inhibitory concentrations (MICs) of nalidixic acid and ciprofloxacin were determined for 84 isolates (all 24 isolates resistant to nalidixic acid in disk diffusion tests and 60 susceptible isolates chosen on the basis of year of isolation and antibiotic resistance gene content) with Etest strips (bioMérieux, Marcy L’Etoile, France). As a means of distinguishing *Salmonella* isolates susceptible to ciprofloxacin (minimum inhibitory concentration [MIC]  $\leq 0.25$  mg/L), which are wild-type (WT), from those that are non-WT, we defined two categories based on the epidemiological cutoffs used by the CLSI: decreased susceptibility to ciprofloxacin (MIC between 0.12 and 0.5 mg/L) and true susceptibility to ciprofloxacin (MIC  $\leq 0.06$  mg/L)<sup>52</sup>. Please note that due to clinical evidence of poor response to ciprofloxacin in systemic infections caused by *Salmonella* spp. isolates displaying such decreased susceptibility to ciprofloxacin, the clinical breakpoint to define ciprofloxacin resistance in non-enteric isolates of this species is MIC  $> 0.06$  mg/L<sup>51</sup>.

## Short-read sequencing

Illumina short-read sequencing was performed by the IP genomics platforms (PF1 and the Mutualised Platform for Microbiology (P2M)) for 389 isolates (362 from IP plus 12 from PISP and 15 from the CDC). At PF1 (150 isolates sequenced), total DNA was extracted with the Wizard Genomic DNA Kit (Promega, Madison, WI, USA) and fragmented with a Covaris E220 ultrasonicator. Sequencing libraries were then prepared with the NEXTflex PCR-free DNA-Seq Kit (Bioo Scientific Corporation, Austin, TX, USA) and sequencing was performed with the HiSeq 2500 system (Illumina, San Diego, CA, USA). At P2M (239 isolates sequenced), total DNA was extracted with the MagNA Pure DNA isolation kit (Roche Molecular Systems, Indianapolis, IN, USA). Sequencing libraries were prepared with the Nextera XT kit (Illumina), and sequencing was performed with the NextSeq 500 system (Illumina). The other 57 SPB<sup>-</sup> isolates were sequenced by the participating laboratories, in accordance with their usual practices (Supplementary Methods “Short-read sequencing”).

Paired-end reads varied in length according to the sequencing platform/site, from 95 to 300 bp, yielding a mean coverage of 124-fold for each isolate (minimum 25-fold, maximum 350-fold) (Supplementary Data 1).

Taxonomic read classification with Kraken<sup>53</sup> v.2.1.1 was used to confirm that sequencing reads originated from *Salmonella* and not from a contaminant. All short-read data are publicly available. Their genome accession numbers are provided in Supplementary Data 1.

## Long-read sequencing and complete genome circularisation

At the time of the study, there was no complete SPB<sup>-</sup> PG1 genome that could be used as a reference genome. We therefore sent total DNA extracted from strain CIP 54.115 (ref. 27) to GATC Biotech (now Eurofins Genomics) for long-read sequencing on the Pacific BioSciences RSII platform. The high-quality de novo assembly was performed by GATC Biotech after the hierarchical genome-assembly process (HGAP)<sup>54</sup>. We performed an additional step of assembly polishing with short reads and the Pilon<sup>55</sup> v.1.23 tool.

Long-read sequences for another 11 isolates – selected to provide a maximal representation of diversity in terms of lineages and phage types – were generated with Oxford Nanopore Technology (ONT) (Supplementary Data 1). Bacterial DNA was extracted from cultures grown in Trypto-casein soy broth (Bio-Rad) at 37 °C with shaking at 200 rpm. We used Genomic-tip 100/G columns (Qiagen) according to the manufacturer’s instructions, for DNA extraction. DNA integrity and the absence of RNA were checked by agarose gel electrophoresis and by the determination of A<sub>260</sub>/A<sub>230</sub> and A<sub>260</sub>/A<sub>280</sub> ratios with a NanoDrop™ 2000 spectrophotometer. DNA concentrations were measured with the Qubit system and the dsDNA BR Assay Kit (Invitrogen). Libraries were prepared from total DNA with the SQK-LSK109 ligation sequencing kit and the EXP-NBD104/114 barcoding kit according to the ONT procedure (Native Barcoding Amplicons protocol version ACDE\_9064\_v109\_revP\_14Aug2019, <https://doi.org/10.17504/protocols.io.bgzjx7n>). Libraries were sequenced with R9.4.1 flow cells and a Mk1C MinION sequencer. Base calling was performed with Guppy<sup>56</sup> v.4.3.4 or v.5.0.13. The filtlong tool (<https://github.com/rrwick/Filtlong/>) v.0.2.1 was used to filter reads according to their length (min\_length 800 bp) and quality (keep\_percent 90). Read lengths ranged from 3827 to 20,347 bp (mean of 9,923 bp), with a mean of 262-fold coverage per genome (minimum 54-fold, maximum 676-fold). Complete de novo genome assemblies were generated with the Tricycler<sup>57</sup> v.0.5.0 pipeline using default parameters. For each isolate, long reads were subsampled into 12 sets, which were subsequently used to generate 12 independent assemblies (four sets per assembler) with Flye<sup>58</sup> v.2.9, raven<sup>59</sup> v.1.6.0 or miniasm<sup>60</sup> v.0.3-&Minipolish<sup>60</sup> v.0.1.3. The consensus assembly was first long read-polished with medaka v.1.4.4 (<https://github.com/nanoporetech/medaka>) and then short read-polished four times with pilon<sup>55</sup> v.1.23. The final assemblies were annotated with bakta<sup>61</sup> v.1.5.0.

## Other genomes

SPB<sup>-</sup> PG2 strain 201906085 (ENA accession no. ERR12749341) and SPB<sup>-</sup> PG3 strain SPB7 (GenBank accession no. [NC\\_010102.1](https://doi.org/10.1038/s41467-024-54418-4)) were used as outgroups to identify the ancestral lineage of SPB<sup>-</sup> PG1 genomes (Supplementary Data 1).

## Genomic typing methods

A two-step in silico serotype confirmation procedure was used: O-antigen determination by a fast kmer-alignment method from KMA<sup>62</sup> v.1.4.14 for the alignment of raw paired-end reads against SPB-specific sequences within the *rfb* cluster (Supplementary Table 2) and *fljC* (encoding the H1 flagellin) and *fljB* (encoding the H2 flagellin), with calling by the NCBI BLAST +<sup>63</sup> v.2.14.1 blastn command line tool on assemblies against SPB reference sequences (Supplementary Table 2). The specific STM 3356 SNV present in *d-Tar* isolates<sup>22</sup> was sought with the NCBI BLAST+ v.2.14.1 blastn command line tool on assemblies against SPB<sup>-</sup> and SPB<sup>+</sup> reference sequences (Supplementary Table 2).

Multilocus sequence typing (MLST)<sup>21</sup>, and core genome MLST (cgMLST) were performed with various tools integrated into Enterobase<sup>64</sup> (<https://enterobase.warwick.ac.uk/species/index/senterica>). The Enterobase *Salmonella* cgMLST scheme (“cgMLST V2 + HierCC VI”) – based on 3002 core genes – assigns bacterial genomes to single-linkage hierarchical clusters (HCs) at 13 fixed levels of resolution, ranging from HCO (high-resolution clusters consisting of identical genomes with no allelic differences) to HC2850 (low-resolution clusters consisting of genomes with up to 2850 allelic differences). Evaluations by Zhou and coworkers<sup>65</sup> found that, in the genus *Salmonella*, cluster assignments at the HC2850, HC2000, and HC900 levels could be used to distinguish subspecies, super-lineages and eBurst groups<sup>21</sup>, respectively, and that epidemic outbreaks could be distinguished with HC2, HC5, or HC10. Sequence type and cgMLST clustering results at the HC2, HC5, HC10, HC20, HC50, HC100, HC200, HC400, HC900, HC2000, HC2600 and HC2850 levels are shown in Supplementary Data 1 for the genomes from the diversity dataset and in Supplementary Data 6 for the genomes from the surveillance dataset.

We also constructed a cgMLST tree – based on core genome allelic distances – inferred with the NINJA neighbour-joining algorithm (present in the “cgMLST V1+HierCC VI” scheme) and visualised with GrapeTree<sup>66</sup> for two different datasets. The first dataset corresponded to the 166 genomes from Connor and coworkers<sup>27</sup> present in Enterobase ([https://enterobase.warwick.ac.uk/species/senterica/search\\_strains?query=workspace:86472](https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86472)), as 12 genomes did not pass the quality control of Enterobase and one was discarded due to a probable inversion (Supplementary Data 7). The second corresponded to the 567 genomes from our diversity dataset that were present in Enterobase ([https://enterobase.warwick.ac.uk/species/senterica/search\\_strains?query=workspace:86468](https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86468)); one unpublished draft genome (ATCC 10719, [SRR955214](https://doi.org/10.1038/s41467-024-54418-4)) sequenced with 454 technology could not be uploaded by Enterobase, which accepts only complete genomes or Illumina short reads. The resulting cgMLST trees used for Supplementary Figs. 1 and 6 are publicly available from [https://enterobase.warwick.ac.uk/ms\\_tree?tree\\_id=92077](https://enterobase.warwick.ac.uk/ms_tree?tree_id=92077) and [https://enterobase.warwick.ac.uk/ms\\_tree?tree\\_id=92095](https://enterobase.warwick.ac.uk/ms_tree?tree_id=92095), respectively. We found that some of the strains described in the Supplementary Table 1 of the article by Connor et al.<sup>27</sup> were assigned to incorrect STs, PGs (Supplementary Data 7). We have corrected the assignments for this study.

Enterobase was searched on April 9th 2021, with the HC400\_1620 criterion (see Supplementary Note “Validation of the SPB<sup>-</sup> PG1 diversity dataset”) used to identify additional genomes, thereby extending the geographic and time coverage of our diversity dataset. Enterobase was also searched on April 10th 2023 with the same HC400\_1620 criterion, to assemble the surveillance dataset of 336 genomes ([https://enterobase.warwick.ac.uk/species/senterica/search\\_strains?query=workspace:91222](https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:91222)).

## Detection of antimicrobial resistance determinants

Antimicrobial resistance (AMR) genes were detected with SRST2 (ref. 67) v.0.2.0 using default parameters and the CARD<sup>68</sup> v.3.0.8 AMR gene database. We also determined the presence or absence of mutations in the quinolone resistance-determining regions (QRDRs) by extracting the relevant SNV calls for codons 83 and 87 of *gyrA*, codon 464 of *gyrB*, and codon 80 of *parC*, from the SPB<sup>+</sup> PG1 reference genome CIP 54.115 (GenBank accession no. [CP147898](https://doi.org/10.1093/nar/31/1/115)) (Supplementary Data 2). We searched for AMR plasmids with PlasmidFinder<sup>69</sup> v.2.0 (<https://cge.food.dtu.dk/services/PlasmidFinder-2.0/>).

## Phylogenetic analyses

The paired-end reads and simulated paired-end reads were mapped onto the complete SPB<sup>+</sup> PG1 reference genome CIP 54.115 (GenBank accession no. [CP147898](https://doi.org/10.1093/nar/31/1/115)) with RedDog (<https://github.com/katholt/reddog-nf>). Genomes with a depth of  $\leq 10\times$  across the reference genome and a ratio of heterozygous SNVs/homozygous SNVs  $> 0.7$  were excluded from the analysis, leaving 568 genomes for downstream analysis. We excluded SNVs present in fewer than 99% of genomes, and SNVs in repeat regions, such as insertion sequences or phages. Recombination was masked with Gubbins<sup>70</sup> v.2.3.2 using the `weighted_robinson_foulds` parameter. The resulting SNV alignment of 15,995 SNVs (see Supplementary Data 2 for a full list of locations) was used to create a maximum likelihood phylogeny with RAxML<sup>71</sup> v.8.2.9, with a GTR + G base substitution model and 1000 bootstraps. The final tree was rooted on the lineage L1 SPB<sup>+</sup> PG1 genome NCTC 8299 and visualised with iTOL<sup>72</sup> v.6 (<https://itol.embl.de>).

We first assessed the temporal signal in the diversity dataset (with and without the five outliers) by “root-to-tip” regression (linear regression of the number of substitutions accumulated from the root to the tips of the ML phylogenetic tree as a function of sampling times) with TempEst<sup>73</sup> v.1.5.3 (Supplementary Fig. 3 and Supplementary Data 1). We confirmed the temporal signal in a subset of 256 genomes (see below) by performing eight date randomisations<sup>74</sup> in BEAST<sup>75</sup> v.2.7.1. The eight runs with randomised dates gave significantly different estimates for substitution rate compared to the run with real dates (Supplementary Fig. 3), indicating a strong temporal signal in the real data.

For the generation of a timed phylogeny, we selected a subset of 256 genomes representing the diversity of the phylogeny and the full range of isolation dates and geographies (Supplementary Data 1). We used all genomes in lineages L1 to L6 and L8, and lineages L7, L9 and L10 were subsampled by selecting the oldest and newest genomes in each lineage, together with a random selection of the remaining genomes in the lineage, keeping the numbers of genomes sampled proportional to the full dataset (Supplementary Data 1). We used BEAST v.2.7.1 with a GTR substitution model, the optimised relaxed clock model (setting the `uclMean` prior to a lognormal distribution with a mean of 0.0001 and a standard deviation of 2, an initial value of 0.0001, and ensuring the ‘mean in real space’ option was checked) and the Bayesian Skyline population model. We performed three independent runs of 600 million states. We removed a 10% burn-in for each run and combined the runs together to create the consensus file. A consensus tree was drawn with the ‘majority rule’ option from `sumtrees.py` in `dendropy`<sup>76</sup> v.4.5.2, setting branch lengths to the median length. The skyline plot was drawn with `Tracer`<sup>77</sup> v.1.7.2.

## Defining lineages and genotypes

We used `fastbaps`<sup>78</sup> v.1.0.8 and visual inspection to define lineages in the maximum likelihood tree, using the `best_baps_partition` function with the phylogeny as a prior.

Genotypes within lineages were defined following visual inspection. For the selection of marker SNVs for each genotype, all SNVs were mapped onto the phylogeny with `SNPPar`<sup>79</sup> v.1.1. On branches leading to genotypes, we selected marker SNVs, prioritising SNVs that were

synonymous mutations within genes for which the ratio of non-synonymous SNVs to synonymous SNVs was as close to zero as possible. The final selection of marker SNVs, their coding consequences and genome location relative to the reference can be found in Supplementary Data 5.

## Implementation and validation of the genotyping scheme

We implemented the genotyping scheme in Mykrobe<sup>80</sup> v.0.12.2. The scheme is therefore supported by this version onwards and is installed when running the Mykrobe commands “`panels update_metadata`” and “`panels update_species all`”. We used the pre-existing *invA* probe<sup>81</sup> to detect genomes belonging to *S. enterica*. An additional probe was created to distinguish between SPB PG1 genomes and other serotypes, including non-PG1 Paratyphi B genomes – this probe detects the *d-Tar* specific SNV (i.e., a G  $\rightarrow$  A change in the start codon of STM 3356)<sup>22</sup> (Supplementary Table 2). We tested Mykrobe implementation on all 568 genomes from the diversity dataset (Supplementary Data 1), using Illumina reads as input, to confirm genotypes were assigned correctly. We then tested the scheme on the 336 genomes of the surveillance dataset (Supplementary Data 6). All these genomes were mapped onto the reference genome as described above and included in the ML phylogeny (which was constructed with IQTree<sup>82</sup> v.2, with a GTR substitution model, on 17,338 SNVs), resulting in 793 genomes (Supplementary Fig. 11). We assigned a genotype to each genome with Mykrobe, and genotypes were validated against the phylogeny to ensure they were correct.

## De novo assembly

Assemblies were generated from Illumina paired-end reads with the `fq2dna/21.06` script (<https://gitlab.pasteur.fr/GIPhy/fq2dna> strategy B; default settings).

Assemblies were generated from 454-GS-FLX reads with SPAdes<sup>83</sup> v.3.15.5 (k-values of 21, 37, 53, 69, 77, default settings for other criteria).

## Pan-genome analysis

The pan-genome analysis was performed on the diversity dataset, which included 566 Illumina paired-end reads, one published complete genome (P7704), and one publicly available draft genome from 454-GS-FLX reads (CFSANO24725) (Supplementary Data 1). We used `panaroo`<sup>84</sup> v.1.3.0, with default parameters and genome assemblies annotated with `bakta`<sup>61</sup> v.1.5.0. We estimated the pan-genome openness level (open if  $\alpha \leq 1$  and closed if  $\alpha > 1$ ) for the 568 genomes, using Heaps’ law with the `R micropan`<sup>85</sup> v.2.1 package. The gene presence/absence matrix of the pan-genome created is shown in Supplementary Data 3. We increased the accuracy of accessory gene assignment to prophages or plasmids by performing a second pan-genome analysis on 581 genomes constructed from 567 short-read assemblies and 14 complete genomes (see next paragraph). After manual curation of the prophage content in the complete genomes (see next paragraph), we selected the 1,506 accessory gene IDs of the 568-pan-genome in the 581-pan-genome output, and we were able to assign the 1,506 accessory CDS to prophage and plasmid regions. An additional manual curation of short-read assemblies was performed for plasmid contigs potentially absent from the 14 complete genomes (Supplementary Data 4).

## Prophage content analysis and *sopE* copy-number variation

Unlike short-read assemblies, complete genome assemblies provide complete information about prophage content and location. We therefore used 14 complete SPB<sup>+</sup> genomes: the genome included in our diversity dataset (P7704), and 13 additional genomes from isolates with short reads included in the “diversity dataset”: SARA41\_FB\_1, which was publicly available (GenBank accession no. [CP074225.1](https://doi.org/10.1093/nar/31/1/115)) and 12 genomes newly generated for this study (Materials section “Long-read sequencing and complete genome circularisation”).

(Supplementary Data 1). The prophage regions were detected in these genomes and taxonomically classified with the PHASTER<sup>86</sup> tool in June 2023. Prophages annotated as intact and absent in at least one of the 14 complete genomes were retained for further analysis. Ten prophage regions were then precisely delineated from the alignment of the 14 genomes around each insertion site (Supplementary Data 8). Finally, we screened for occupancy of the 10 insertion sites in the 568 genomes of the diversity dataset with the blastn algorithm (BLAST +<sup>63</sup> v.2.14.1).

By combining the annotation of the 14 SPB<sup>+</sup> complete genomes and the prophage delineation described above, we were able to locate the *sopE* gene in SEN34-like and P88-like prophages. We used the blastn algorithm and the *sopE* reference sequence (GenBank accession no. L78932) from *S. enterica* serotype Dublin<sup>37</sup> to confirm our identification of the prophage-borne *sopE* gene.

As the number of *sopE*-containing prophages differed between genomes, we obtained a good estimate of the *sopE* prophage copy number per genome (*sopE*-CN) by combining the insertion site occupancy and short-read coverage data. The short reads were first aligned with the complete genome of the B62 isolate (GenBank accession no. CP147902), with the very-sensitive-local option of bowtie2 (ref. 87) v.2.3.5.1, and coverage was estimated across the entire genome (cov\_g), at the *sopE* locus (cov\_s) and the *Salmonella* Pathogenicity Island 1 locus (GenBank accession no. CP147902; coordinates 981997–1036581) (cov\_pi) with samtools<sup>88</sup> v.1.13. The mean value of the cov\_s/cov\_pi and cov\_s/cov\_g ratios was chosen as an estimate of the *sopE*-CN. The sequencing platform has a strong and significant effect on *sopE*-CN values (Kruskal-Wallis test  $p = 4.8E-50$ , large magnitude effect  $\eta^2[H] = 0.403$ ). We therefore split the dataset according to the Illumina platform used (HiSeq,  $n = 274$ ; MiSeq,  $n = 45$ ; or NextSeq,  $n = 242$ ) before searching for outliers in the *sopE*-CN distribution with Dunn's pairwise comparison test with Bonferroni correction (Supplementary Fig. 12). If the *sopE*-CN and the number of insertion sites occupied were not correlated, we visually inspected both the long- and short-read mappings onto the genomes of isolates B62 and B2590, with Tablet<sup>89</sup> v.1.21.02.08. The complete data are summarised in Supplementary Data 9.

Prophages inserted in the vicinity of the *hin-fljB-fljA* cluster are described in Supplementary Data 8. The assemblies from the 568 diversity dataset genomes were screened for the presence and structure of the *hin-fljB-fljA* gene cluster with the blastn<sup>90</sup> algorithm.

### Structure of the *hin-fljB-fljA* gene cluster

We used the gggenes<sup>91</sup> v.0.5.0 and ggplot2 (ref. 92) v.3.4.2 packages of R<sup>93</sup> v.4.1.2 to visualise prophage gene content.

### Statistics

Statistical analysis was performed with R v.4.1.2, R packages rstatix<sup>94</sup> v.0.7.2, and ggpubr<sup>95</sup> v.0.6.0.

### Data collection

The data were entered into an Excel (Microsoft) version 16.76 spreadsheet or tabulation-separated text files (tsv).

### Reporting summary

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

### Data availability

The publicly available sequences used in this study are available from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession numbers NC\_028699, NC\_026014, NC\_010102.1, CP074225.1, L78932, NZ\_CP065185, NZ\_CP065186.1, NZ\_CP065187.1, NZ\_CP065188.1, and JWQX00000000.1.

The long-read sequence data generated in this study are available from GenBank under accession numbers CP147895, CP147896,

CP147897, CP147898, CP147899, CP147900, CP147901, CP147902, CP147903, CP147904, CP147905, CP147906, and CP147907. The short-read sequence data generated in this study were submitted to Enterobase (<https://enterobase.warwick.ac.uk/>) and to the European Nucleotide Archive (ENA, <https://www.ebi.ac.uk/ena/>) under study numbers PRJDB11608, PRJEB18998, PRJEB28356, PRJEB30317, PRJEB67705, PRJNA248792, PRJEB68323, PRJEB49424, PRJEB71958. All the accession numbers of the short-read sequences produced and/or used in this study are listed in Supplementary Data 1 and Supplementary Data 6.

The list of genomes studied (and their assembled short-read data) can be obtained from Enterobase at: [https://enterobase.warwick.ac.uk/species/senterica/search\\_strains?query=workspace:86468](https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86468) (diversity dataset, Supplementary Data 1) and: [https://enterobase.warwick.ac.uk/species/senterica/search\\_strains?query=workspace:91222](https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:91222) (surveillance data set, Supplementary Data 6).

The list of 166 genomes (and their assembled short-read data) published by Connor et al.<sup>27</sup> and present in Enterobase can be obtained from [https://enterobase.warwick.ac.uk/species/senterica/search\\_strains?query=workspace:86472](https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86472) (Supplementary Data 7).

The cgMLST GrapeTrees shown in Supplementary Figs. 1 and 6 can be visualised with Enterobase at: [https://enterobase.warwick.ac.uk/ms\\_tree?tree\\_id=92077](https://enterobase.warwick.ac.uk/ms_tree?tree_id=92077) and [https://enterobase.warwick.ac.uk/ms\\_tree?tree\\_id=92095](https://enterobase.warwick.ac.uk/ms_tree?tree_id=92095), respectively. Source data are provided with this paper.

### Code availability

Mykrobe is available for download at <https://github.com/mykrobe-tools/mykrobe>. Instructions for running Mykrobe for SPB<sup>+</sup> PG1 are available in the Mykrobe documentation at <https://github.com/Mykrobe-tools/mykrobe/wiki/AMR-Prediction>. Briefly, the option --species paratyphiB should be used when running the Mykrobe predict command. The genotyping panel developed here is available from [https://figshare.com/articles/dataset/Mykrobe\\_panel\\_paratyphi\\_B\\_version\\_20230627/24925506](https://figshare.com/articles/dataset/Mykrobe_panel_paratyphi_B_version_20230627/24925506)

The fq2dna script (genome de novo assembly from raw paired-end FASTQ files) can be found at <https://gitlab.pasteur.fr/GIPhy/fq2dna>.

### References

- Achard, C. & Bensaude, R. Infections paratyphoïdiques. *Bull. Mem. Soc. Hop. Paris* **13**, 820–833 (1896).
- Gwyn, N. B. On infection with a para-colon bacillus in a case with all the clinical features of typhoid fever. *Bull. Johns. Hopkins Hosp.* **9**, 54–56 (1898).
- Hirschfeld, L. A new germ of paratyphoid. *Lancet* **193**, 296–297 (1919).
- Pratt, J. H. On paratyphoid fever and its complications. *Boston Med. Surg.* **148**, 137–142 (1903).
- Proescher, F. & Roddy, J. A. Bacteriological studies on paratyphoid A and paratyphoid B. *Arch. Intern. Med.* **3**, 263–312 (1910).
- Grimont, P. A. D. & Weill F. X. Antigenic formulae of the *Salmonella* serovars, 9th ed. World Health Organization Collaborating Center for reference and research on *Salmonella*, Institut Pasteur (2007). Available online at: [https://www.pasteur.fr/sites/default/files/veng\\_O.pdf](https://www.pasteur.fr/sites/default/files/veng_O.pdf) (accessed June 7, 2024).
- Kristensen, M. & Bojlen, K. Vergärungsmässig definierte Typen des Paratyphus B-Bazillus. *Zentralbl. Bakteriol. Parasitenk. Infektionskr. I. Abt. Orig.* **114**, 86–108 (1929).
- Vogelsang, T. M. Paratyphoid B in Western Norway: bacteriology, symptomatology, morbid anatomy and epidemiology of infection with bacillus paratyphosus B. *J. Infect. Dis.* **55**, 276–298 (1934).
- Savage, W. Paratyphoid fever: an epidemiological study. *J. Hyg.* **42**, 393–410 (1942).
- Felix, A. World survey of typhoid and paratyphoid-B phages types. *Bull. World Health Organ.* **13**, 109–170 (1955).

11. Nicolle, P. Rapport sur la distribution des lysotypes de *Salmonella typhi* et de *S. paratyphi B* dans le monde, d'après les résultats fournis par les centres nationaux membres du comité international de la lysotypie entérique à l'occasion du congrès international de microbiologie, Stockholm, 1958. *Ann. Inst. Pasteur* **102**, 389–409 (1962).
12. International Committee for Enteric Phage-Typing (ICEPT). The geographical distribution of *Salmonella typhi* and *Salmonella paratyphi A* and *B* phage types during the period 1 January 1966 to 31 December 1969. *J. Hyg.* **71**, 59–84 (1973).
13. Braddick, M. R. & Sharp, J. C. Enteric fever in Scotland 1975–1990. *Public Health* **107**, 193–198 (1993).
14. World Health Organization. Surveillance of *Salmonella paratyphi A* and *B*. *Wkly Epidemiol. Rec.* **53**, 257 (1978). Available online at: [https://iris.who.int/bitstream/handle/10665/222048/WER5334\\_257-257.PDF?sequence=1&isAllowed=y](https://iris.who.int/bitstream/handle/10665/222048/WER5334_257-257.PDF?sequence=1&isAllowed=y) (accessed June 7, 2024).
15. Felix, A. & Callow, B. R. Typing of paratyphoid B bacilli by Vi bacteriophage. *Br. Med. J.* **2**, 127–130 (1943).
16. Felix, A. & Callow, B. R. Paratyphoid-B Vi-phage typing. *Lancet* **2**, 10–14 (1951).
17. Nicolle, P. Rapport sur la distribution des lysotypes de *Salmonella typhi* et de *S. paratyphi B* dans le monde, d'après les résultats fournis par les centres nationaux membres du comité international de la lysotypie entérique à l'occasion du congrès international de microbiologie, Stockholm, 1958 Deuxième partie. *Ann. Inst. Pasteur* **102**, 580–595 (1962).
18. International Federation for Enteric Phage-Typing (IFEPT). The geographical distribution of *Salmonella typhi* and *Salmonella paratyphi A* and *B* phage types during the period 1 January 1970 to 31 December 1973. *J. Hyg.* **88**, 231–254 (1982).
19. Le Minor, L., Veron, M. & Popoff, M. Proposition pour une nomenclature des *Salmonella*. *Ann. Microbiol.* **133B**, 245–254 (1982).
20. Selander, R. K. et al. Genetic population structure, clonal phylogeny, and pathogenicity of *Salmonella paratyphi B*. *Infect. Immun.* **58**, 1891–1901 (1990).
21. Achtman, M. et al. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* **8**, e1002776 (2012).
22. Malorny, B., Bunge, C. & Helmuth, R. Discrimination of d-tartrate-fermenting and -nonfermenting *Salmonella enterica* subsp. *enterica* isolates by genotypic and phenotypic methods. *J. Clin. Microbiol.* **41**, 4292–4297 (2003).
23. GBD 2017 Typhoid and Paratyphoid Collaborators. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect. Dis.* **19**, 369–381 (2019).
24. Carey, M. E. et al. Global diversity and antimicrobial resistance of typhoid fever pathogens: Insights from a meta-analysis of 13,000 *Salmonella Typhi* genomes. *Elife* **12**, e85867 (2023).
25. 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).
26. 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).
27. Connor, T. R. et al. What's in a name? Species-wide whole-genome sequencing resolves invasive and noninvasive lineages of *Salmonella enterica* serotype Paratyphi B. *mBio* **7**, e00527-16 (2016).
28. Chattaway, M. A. et al. Phylogenomics and antimicrobial resistance of *Salmonella Typhi* and Paratyphi A, B and C in England, 2016–2019. *Microb. Genom.* **7**, 000633 (2021).
29. Chattaway, M. A. et al. Genomic sentinel surveillance: *Salmonella Paratyphi B* outbreak in travellers coinciding with a mass gathering in Iraq. *Microb. Genom.* **9**, mgen000940 (2023).
30. Weill, F. X., Fabre, L., Grandry, B., Grimont, P. A. & Casin, I. Multiple-antibiotic resistance in *Salmonella enterica* serotype Paratyphi B isolates collected in France between 2000 and 2003 is due mainly to strains harboring *Salmonella* genomic islands 1, 1-B, and 1-C. *Antimicrob. Agents Chemother.* **49**, 2793–2801 (2005).
31. Balandraud, A. et al. Sepsis caused by *Salmonella Paratyphi B* producing an OXA-48 carbapenemase in a traveller. *J. Glob. Anti-microb. Resist.* **26**, 219–221 (2021).
32. Melnotte, P. Les affections typhoïdes au Maroc. *Bull. Soc. Path. Ex.* **25**, 447–460 (1932).
33. Nicolle, P. & Hamon, Y. Distribution des lysotypes du bacille typhique et du bacille paratyphique B en France, dans les territoires d'Outre-Mer et dans quelques autres pays. *Rev. Hyg. Med. Soc.* **2**, 424–463 (1954).
34. Ministerio de Salud Argentina. Alerta epidemiológica. Situación de fiebre paratifoidea en la provincia de Salta. Semana epidemiológica 51/2022 (2022). Available online at: <https://bancos.salud.gob.ar/sites/default/files/2021-11/Alerta-fiebre-paratifoidea-VFF.pdf> (accessed June 7, 2024).
35. Roumagnac, P. et al. Evolutionary history of *Salmonella Typhi*. *Science* **314**, 1301–1304 (2006).
36. Wong, V. K. et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella Typhi* identifies inter- and intracontinental transmission events. *Nat. Genet.* **47**, 632–639 (2015).
37. Prager, R. et al. Molecular properties of *Salmonella enterica* serotype Paratyphi B distinguish between its systemic and its enteric pathovars. *J. Clin. Microbiol.* **41**, 4270–4278 (2003).
38. Friebe, A. et al. SopE and SopE2 from *Salmonella typhimurium* activate different sets of RhoGTPases of the host cell. *J. Biol. Chem.* **276**, 34035–34040 (2001).
39. Prager, R. et al. Prevalence and polymorphism of genes encoding translocated effector proteins among clinical isolates of *Salmonella enterica*. *Int. J. Med. Microbiol.* **290**, 605–617 (2000).
40. Mirold, S., Rabsch, W., Tschäpe, H. & Hardt, W. D. Transfer of the *Salmonella* type III effector sopE between unrelated phage families. *J. Mol. Biol.* **312**, 7–16 (2001).
41. Tassinari, E. et al. Whole-genome epidemiology links phage-mediated acquisition of a virulence gene to the clonal expansion of a pandemic *Salmonella enterica* serovar Typhimurium clone. *Microb. Genom.* **6**, mgen000456 (2020).
42. Ehrbar, K., Mirold, S., Friebe, A., Stender, S. & Hardt, W. D. Characterization of effector proteins translocated via the SPI1 type III secretion system of *Salmonella typhimurium*. *Int. J. Med. Microbiol.* **291**, 479–485 (2002).
43. Vonaesch, P. et al. The *Salmonella Typhimurium* effector protein SopE transiently localizes to the early SCV and contributes to intracellular replication. *Cell Microbiol.* **16**, 1723–1735 (2014).
44. Hoffmann, M. et al. Comparative genomic analysis and virulence differences in closely related *Salmonella enterica* serotype Heidelberg isolates from humans, retail meats, and animals. *Genome Biol. Evol.* **6**, 1046–1068 (2014).
45. Wong, V. K. et al. An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. *Nat. Commun.* **7**, 12827 (2016).
46. Tanmoy, A. M. et al. Paratype: a genotyping tool for *Salmonella Paratyphi A* reveals its global genomic diversity. *Nat. Commun.* **13**, 7912 (2022).
47. Hawkey, J. et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, *Shigella sonnei*. *Nat. Commun.* **12**, 2684 (2021).
48. Wang, Y. et al. The temporal dynamics of antimicrobial-resistant *Salmonella enterica* and predominant serovars in China. *Natl Sci. Rev.* **10**, nwac269 (2022).
49. Baker, K. S. et al. The Murray collection of pre-antibiotic era Enterobacteriaceae: a unique research resource. *Genome Med.* **7**, 97 (2015).

50. Higginson, E. E. et al. Improving our understanding of *Salmonella enterica* serovar Paratyphi B through the engineering and testing of a live attenuated vaccine strain. *mSphere* **3**, e00474-18 (2018).
51. CA-SFM., EUCAST. Comité de l'Antibiogramme de la Société Française de Microbiologie Recommandations 2018. Available on line at: [https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/CASFMV2\\_SEPTEMBRE2018.pdf](https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/CASFMV2_SEPTEMBRE2018.pdf) (accessed June 7, 2024).
52. CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 30th edn. Supplement M100 (Clinical and Laboratory Standards Institute, 2020).
53. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
54. Chin, C. S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
55. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
56. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
57. Wick, R. R. et al. Tricycler: consensus long-read assemblies for bacterial genomes. *Genome Biol.* **22**, 266 (2021).
58. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
59. Vaser, R. & Šikić, M. Time- and memory-efficient genome assembly with Raven. *Nat. Comput. Sci.* **1**, 332–336 (2021).
60. Wick, R. R. & Holt, K. E. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res.* **8**, 2138 (2021).
61. Schwengers, O. et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb. Genom.* **7**, 000685 (2021).
62. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinforma.* **19**, 307 (2018).
63. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
64. Zhou, Z., Alikhan, N. F., Mohamed, K. & Fan, Y. Agama Study Group & Achtman, M. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia core* genomic diversity. *Genome Res.* **30**, 138–152 (2020).
65. Zhou, Z., Charlesworth, J. & Achtman, M. HierCC: a multi-level clustering scheme for population assignments based on core genome MLST. *Bioinformatics* **37**, 3645–3646 (2021).
66. Zhou, Z. et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* **28**, 1395–1404 (2018).
67. Inouye, M. et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6**, 90 (2014).
68. Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
69. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Anti-microb. Agents Chemother.* **58**, 3895–3903 (2014).
70. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
71. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
72. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
73. 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).
74. 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).
75. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
76. Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571 (2010).
77. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
78. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **47**, 5539–5549 (2019).
79. Edwards, D. J., Duchene, S., Pope, B. & Holt, K. E. SNPPar: identifying convergent evolution and other homoplasies from microbial whole-genome alignments. *Microb. Genom.* **7**, 000694 (2021).
80. Hunt, M. et al. Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. *Wellcome Open Res.* **4**, 191 (2019).
81. Ingle, D. J., Hawkey, J., Dyson, Z. A. & Holt, K. E. Genotyphi implementation in Mykrobe - preliminary technical report. Zenodo (2022). Available online at <https://zenodo.org/record/4785179> (accessed March 25, 2024).
82. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
83. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
84. Tonkin-Hill, G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
85. Snipen, L. & Liland, K. H. micropan: an R-package for microbial pangenomics. *BMC Bioinforma.* **16**, 79 (2015).
86. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
87. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
88. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Giga-Science* **10**, giab008 (2021).
89. Milne, I. et al. 2013. Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinforma.* **14**, 193–202 (2013).
90. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
91. Wilkins, D. gggenes: Draw Gene Arrow Maps in 'ggplot2'. R package version 0.5.0 (2023). Available on line at: <https://wilcox.org/gggenes/> (accessed June 7, 2024).
92. Wickham, H. ggplot2: *Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4 (2016). Available on line at: <https://ggplot2.tidyverse.org> (accessed June 7, 2024).
93. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021. Available on line at: <https://www.R-project.org/> (accessed March 28, 2024).
94. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2 (2023). Available online at: <https://rpkgs.datanovia.com/rstatix/> (accessed June 7, 2024).
95. Kassambara, A. ggpubr: "ggplot2" Based Publication Ready Plots <https://CRAN.R-project.org/package=ggpubr> (2023). Available online at accessed June 7, 2024.

## Acknowledgements

We thank Prof. Jacques Ravel, Prof. David A. Rasko, Luke Tallon, Kranthi Vavikolanu, and Michael Pietsch for submitting archived or new short reads to a public repository, Susan Van Duyne for ensuring the safe shipping of strains, Anthony M. Smith and Chien-Shun Chiou for reviewing their data, Paul O'Dette for his support, and the sequencing team at the *Institut Pasteur* (PF1 & P2M-Plateforme de Microbiologie Mutualisée) for sequencing the samples. We also thank all the corresponding laboratories of the French National Reference Centre for *Escherichia coli*, *Shigella*, and *Salmonella*.

This research was funded by the *Fondation Le Roch-Les Mousquetaires* (to F.-X.W.); *Institut Pasteur* (to F.-X.W.); *Santé publique France* (to F.-X.W.); and by the French Government's *Investissement d'Avenir* programme, *Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases'* (grant no. ANR-10-LABX-62-IBEID to F.-X.W.). M.H. was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University in partnership with the UK Health Security Agency (NIHR200915), and the NIHR Biomedical Research Centre, Oxford. M.A.C. is affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) (NIHR200892) in Genomics and Enabling Data at University of Warwick in partnership with the UK Health Security Agency (UKHSA), in collaboration with Universities of Cambridge and Oxford. M.A.C. is based at UKHSA. The views expressed are those of the authors and not necessarily those of the Centers for Disease Control and Prevention, the NIHR, the Department of Health and Social Care or the UK Health Security Agency. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

F.-X.W. designed and oversaw the study. D.B., M.A.C., S.S., H.I., P.I.F., N.d.L., L.K., X.X., J.I., D.C., Y.A., M.M.A., Y.W., B.M.J., Z.M., M.C., M.Y., B.Z., M.M.O., C.S.N., M.P.G., and F.-X.W. selected and provided isolates or genomes with their basic metadata. L.FR., A.T.D., S.I.J., M.R., V.G., E.S., and L.F.A. subcultured the bacteria, performed phenotypic experiments, and extracted DNA. J.H., L.FR., A.Z., E.N., L.F.A., and F.-X.W. analysed and/or interpreted the data. J.H., M.H. and Z.I. implemented the genotyping scheme in Mykrobe. J.H., L.FR., and F.-X.W. wrote the manuscript. All the authors contributed to the editing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54418-4>.

**Correspondence** and requests for materials should be addressed to François-Xavier Weill.

**Peer review information** *Nature Communications* thanks Thomas Connor, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

<sup>1</sup>Department of Infectious Diseases, School of Translational Medicine, Monash University, Melbourne, VIC 3004, Australia. <sup>2</sup>Institut Pasteur, Université Paris Cité, Unité des Bactéries pathogènes entériques, Paris F-75015, France. <sup>3</sup>Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris F-75015, France. <sup>4</sup>Scottish Microbiology Reference Laboratories (SMiRL), Glasgow G31 2ER, UK. <sup>5</sup>Gastrointestinal Bacteria Reference Unit (GBRU), United Kingdom Health Security Agency, London NW9 5EQ, UK. <sup>6</sup>Unit of Enteropathogenic Bacteria and Legionella (FG11)/National Reference Centre for Salmonella and Other Bacterial Enteric Pathogens, Robert Koch-Institute, Wernigerode 38855, Germany. <sup>7</sup>Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo 162-8640, Japan. <sup>8</sup>Division of Foodborne, Waterborne and Environmental Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA. <sup>9</sup>National Salmonella, Shigella and Listeria Reference Laboratory, Galway University Hospitals, Galway SW4 671, Ireland. <sup>10</sup>Pasteur Institute of St Petersburg, St Petersburg 197101, Russia. <sup>11</sup>Department of Microbiology, Shanghai Municipal Centre for Disease Control and Prevention, Shanghai 200336, China. <sup>12</sup>Department of Bacteriology, Toyama Institute of Health, Toyama 939-0363, Japan. <sup>13</sup>Institut Pasteur, Université Paris Cité, Collection of Institut Pasteur (CIP), Paris F-75015, France. <sup>14</sup>International Joint Research Centre for National Animal Immunology, College of Veterinary Medicine, Henan Agricultural University, Zhengzhou, Henan 450046, China. <sup>15</sup>CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences (CAS), Beijing 100101, China. <sup>16</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton CB10 1SD, UK. <sup>17</sup>Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>18</sup>National Institute of Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Headley Way, Oxford, UK. <sup>19</sup>Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, UK. <sup>20</sup>School of Medicine, University of Galway, Galway H91 TK33, Ireland. <sup>21</sup>Department of Veterinary Medicine, Zhejiang University College of Animal Sciences, Hangzhou 310058, China. <sup>22</sup>School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China. <sup>23</sup>Milner Centre for Evolution, University of Bath, Claverton Down, Bath, UK. <sup>24</sup>Present address: Bioinformatic Core Facility, UMS AMMICA, Gustave Roussy, Villejuif F-94800, France. <sup>25</sup>These authors contributed equally: Jane Hawkey, Lise Frézal.

✉ e-mail: [francois-xavier.weill@pasteur.fr](mailto:francois-xavier.weill@pasteur.fr)