



HAL
open science

Genomic perspective on the bacillus causing paratyphoid B fever

François-Xavier Weill, Lise Frézal, Alicia Tran-Dien, Anna Zhukova, Derek Brown, Marie Chattaway, Sandra Simon, Hidemasa Izumiya, Patricia Fields, Niall de Lappe, et al.

► **To cite this version:**

François-Xavier Weill, Lise Frézal, Alicia Tran-Dien, Anna Zhukova, Derek Brown, et al.. Genomic perspective on the bacillus causing paratyphoid B fever. 2024. pasteur-04638433

HAL Id: pasteur-04638433

<https://pasteur.hal.science/pasteur-04638433>

Preprint submitted on 8 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Genomic perspective on the bacillus causing paratyphoid B fever

François-Xavier Weill

francois-xavier.weill@pasteur.fr

Institut Pasteur <https://orcid.org/0000-0001-9941-5799>

Lise Frézal

Institut Pasteur <https://orcid.org/0000-0002-6518-0423>

Alicia Tran-Dien

Institut Pasteur

Anna Zhukova

Institut Pasteur <https://orcid.org/0000-0003-2200-7935>

Derek Brown

Glasgow Royal Infirmary <https://orcid.org/0000-0003-3742-8257>

Marie Chattaway

Gastrointestinal Bacteria Reference Unit, Salmonella Reference Service, Public Health England

Sandra Simon

Robert Koch-Institute

Hidemasa izumiya

National Institute of Infectious Diseases <https://orcid.org/0000-0002-9797-7402>

Patricia Fields

Centers for Disease Control and Prevention

Niall de Lappe

Galway University Hospitals

Lidia Kaftyreva

Pasteur Institute of St Petersburg

Xuebin Xu

Department of Microbiology, Shanghai Municipal Center for Disease Control and Prevention

Junko Isobe

Toyama Institute of Health

Dominique Clermont

Institut Pasteur <https://orcid.org/0000-0002-6018-2462>

Elisabeth Njamkepo

Institut Pasteur <https://orcid.org/0000-0001-6791-6003>

Yukihiro Akeda

Department of Bacteriology I, National Institute of Infectious Diseases

Sylvie Issenhuth-Jeanjean

Institut Pasteur

Mariia Makarova

Pasteur Institute of St Petersburg <https://orcid.org/0000-0003-3600-2377>

Yanan Wang

Henan Agricultural University <https://orcid.org/0000-0002-7461-2195>

Martin Hunt

Wellcome Sanger Institute

Brent Jenkins

Centers for Disease Control and Prevention <https://orcid.org/0000-0003-2907-660X>

Magali Ravel

Institut Pasteur

Véronique Guibert

Institut Pasteur

Estelle Serre

Institut Pasteur

Zoya Matveeva

Pasteur Institute of St Petersburg

Laetitia Fabre

Institut Pasteur

Martin Cormican

University of Galway

Min Yue

University of Chinese Academy of Sciences

Masatomo Morita

National Institute of Infectious Diseases <https://orcid.org/0000-0002-2850-0053>

Zamin Iqbal

European Bioinformatics Institute <https://orcid.org/0000-0001-8466-7547>

Carolina Silva Nodari

Institut Pasteur

Maria Pardos de la Gandara

Institut Pasteur <https://orcid.org/0000-0002-3273-6392>

Jane Hawkey

Monash University <https://orcid.org/0000-0001-9661-5293>

Article

Keywords:

Posted Date: June 18th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4502330/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

1 Genomic perspective on the bacillus causing paratyphoid B fever

2
3 Jane Hawkey^{1#}, Lise Frézal^{2#}, Alicia Tran Dien^{2§}, Anna Zhukova³, Derek Brown⁴, Marie
4 Anne Chattaway⁵, Sandra Simon⁶, Hidemasa Izumiya⁷, Patricia I Fields⁸, Niall de Lappe⁹,
5 Lidia Kaftyreva¹⁰, Xuebin Xu¹¹, Junko Isobe¹², Dominique Clermont¹³, Elisabeth Njamkepo²,
6 Yukihiro Akeda⁷, Sylvie Issenhuth-Jeanjean², Mariia Makarova¹⁰, Yanan Wang^{14,15}, Martin
7 Hunt^{16,17,18,19}, Brent M. Jenkins⁸, Magali Ravel², Véronique Guibert², Estelle Serre², Zoya
8 Matveeva¹⁰, Laëtitia Fabre², Martin Cormican^{9,20}, Min Yue^{21,22}, Baoli Zhu¹⁵, Masatomo
9 Morita⁷, Zamin Iqbal^{16,23}, Carolina Silva Nodari², Maria Pardos de la Gandara², François-
10 Xavier Weill^{2*}

11
12 ¹Department of Infectious Diseases, School of Translational Medicine, Monash University,
13 Melbourne, Victoria 3004, Australia.

14
15 ²Institut Pasteur, Université Paris Cité, Unité des Bactéries pathogènes entériques, Paris, F-
16 75015, France.

17
18 ³Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris, F-75015,
19 France.

20
21 ⁴Scottish Microbiology Reference Laboratories (SMiRL), Glasgow, G31 2ER, United
22 Kingdom.

23
24 ⁵Gastrointestinal Bacteria Reference Unit (GBRU), United Kingdom Health Security Agency,
25 London, NW9 5EQ, United Kingdom.

26
27 ⁶Unit of Enteropathogenic Bacteria and Legionella (FG11)/National Reference Centre for
28 Salmonella and Other Bacterial Enteric Pathogens, Robert Koch-Institute, Wernigerode,
29 38855, Germany.

30
31 ⁷Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, 162-8640,
32 Japan.

33
34 ⁸Division of Foodborne, Waterborne and Environmental Diseases, Centers for Disease
35 Control and Prevention, Atlanta, Georgia, USA

36
37 ⁹National Salmonella, Shigella and Listeria Reference Laboratory, Galway University
38 Hospitals, Galway, SW4 671, Ireland.

39
40 ¹⁰Pasteur Institute of St Petersburg, St Petersburg, 197101, Russia.

41
42 ¹¹Department of Microbiology, Shanghai Municipal Centre for Disease Control and
43 Prevention, Shanghai 200336, China.

44
45 ¹²Department of Bacteriology, Toyama Institute of Health, Toyama, 939-0363, Japan.

46
47 ¹³ Institut Pasteur, Université Paris Cité, Collection of Institut Pasteur (CIP), Paris, F-75015,
48 France.

49
50 ¹⁴International Joint Research Centre for National Animal Immunology, College of
51 Veterinary Medicine, Henan Agricultural University, Zhengzhou, Henan 450046, China.

52
53 ¹⁵CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of
54 Microbiology, Chinese Academy of Sciences (CAS), Beijing 100101, China.

55
56 ¹⁶European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, CB10
57 1SD, United Kingdom.

58
59 ¹⁷Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.
60

61 ¹⁸National Institute of Health Research Oxford Biomedical Research Centre, John Radcliffe
62 Hospital, Headley Way, Oxford, United Kingdom.

64 ¹⁹Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial
65 Resistance, University of Oxford, Oxford, United Kingdom.

67 ²⁰School of Medicine, University of Galway, Galway, H91 TK33, Ireland.

69 ²¹Department of Veterinary Medicine, Zhejiang University College of Animal Sciences,
70 Hangzhou, 310058, China.

72 ²²School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese
73 Academy of Sciences, Hangzhou 310024, China.

75 ²³Milner Centre for Evolution, University of Bath, Claverton Down, Bath, United Kingdom.

76
77 [§]New affiliation: Bioinformatic Core Facility, UMS AMMICA, Gustave Roussy, Villejuif, F-
78 94800, France.

79
80 [#]These authors contributed equally to this work

81
82 ^{*}Corresponding author

83
84

85 **ABSTRACT**

86 Paratyphoid B fever (PTB) is caused by an invasive lineage (phylogroup 1, PG1) of
87 *Salmonella enterica* serotype Paratyphi B (SPB). Here, we provide a genomic overview of the
88 population structure, geographic distribution, and evolution of SPB PG1 by analysing
89 genomes from 568 historical and contemporary isolates, obtained globally, between 1898 and
90 2021. We show that this pathogen existed in the 13th century, subsequently diversifying into
91 11 lineages and 38 genotypes with strong phylogeographic patterns. Following its discovery
92 in 1896, it circulated across Europe until the 1970s, after which it was mostly reimported into
93 Europe from South America and other parts of the world. Antimicrobial resistance recently
94 emerged in various genotypes of SPB PG1, mostly through mutations of the quinolone-
95 resistance-determining regions of *gyrA* and *gyrB*. This study provides an unprecedented
96 insight into SPB PG1 and an essential genomic tool for the future global genomic surveillance
97 of PTB.

98

99

100 INTRODUCTION

101 At the turn of the 20th century, investigators in Europe and North America showed that
102 Eberth's bacillus (now known as *Salmonella enterica* serotype Typhi) was not the only
103 organism causing enteric fever, a severe systemic infection causing prolonged high fever,
104 fatigue, headache, and abdominal pain, exclusively in humans^{1,2,3}. The other causal bacteria
105 identified were paratyphoid bacilli of three different types: A, B, and C. The first two cases
106 involving the paratyphoid B bacillus (now known as *S. enterica* serotype Paratyphi B and
107 referred to hereafter as SPB) were described by Achard and Bensaude in Paris, France, in
108 1896 (refs^{1,4,5}). Following the introduction of appropriate laboratory tools — initially based
109 on O- and H-antigen serotyping⁶ (recognising the antigenic formula 1,4,[5],12:b:1,2) and then
110 on an absence of *d*-tartrate (*d*-Tar⁻) fermentation by the cultured bacteria (SPB⁻ strains)⁷ —
111 SPB was more frequently detected in Europe (**Supplementary Text**). PTB disease was
112 milder than typhoid fever, with a lower incidence of complications and a lower mortality^{8,9}. It
113 occurred as sporadic cases with occasional outbreaks, rarely due to case-to-case transmission
114 in institutions (such as garrisons, hospitals, and children's homes), but was mostly foodborne,
115 particularly in natural or synthetic cream, unpasteurised milk, and bakery products⁹. PTB
116 cases remained frequent across Europe in the 1950s and 1960s^{10,11,12}, but the end of the 1970s
117 saw a progressive decline in SPB⁻ isolation accompanied by an increase in the isolation of
118 zoonotic *d*-tartrate-fermenting (SPB⁺) strains, also known as variant Java strains¹³.
119 Epidemiologically, PTB also shifted from being locally acquired to being an imported
120 disease. In the UK, 43.2% (152/352) of PTB cases were considered to have been contracted
121 locally between 1973 and 1977, whereas 56.8% (200/352) were considered to have been
122 contracted abroad, particularly in Mediterranean and Middle Eastern countries¹⁴.

123

124 A phage typing scheme was developed for the surveillance of PTB as soon as World War II
125 (WWII)^{15,16}. This scheme subtyped SPB⁻ isolates into 10 different phage types (PTs) – 1, 2,
126 3a, 3aI, 3b, BAOR (British Army of the Rhine), Jersey, Beccles, Dundee, and Taunton – and
127 reports based on tens of thousands of isolates phage-typed across the world were regularly
128 published from the 1940s to the 1990s^{10,11,12,17,18}. A particular geographic distribution was
129 observed, with PTs 1 and 2 reported to be autochthonous to the UK (whereas Taunton was
130 found in imported cases); BAOR was prevalent in Central Europe; Dundee was prevalent in
131 France, and the PT 3 series (3a, 3aI, and 3b) was more common in North America than
132 elsewhere. However, despite their different epidemiological characteristics and pathogenicity,
133 SPB⁻ and SPB⁺ strains were grouped together as a single serotype, SPB^{6,19}, and their PTs were
134 ultimately combined without distinction in world surveys, rendering these surveys less
135 informative.

136

137 From the 1990s onwards, population genetics tools were used to distinguish between SPB⁺
138 and invasive SPB⁻ strains, initially by multilocus enzyme electrophoresis²⁰ (Pb1 group for
139 SPB⁻) and then by multilocus sequence typing (MLST)²¹ (sequence type ST86 and five single-
140 locus variants of ST86 for SPB⁻). In 2003, the molecular basis of the *d*-Tar⁻ character of SPB⁻
141 strains was elucidated: a single-nucleotide variant (SNV) leading to the loss of the start codon
142 of a gene involved in the *d*-tartrate fermentation pathway²².

143

144 In 2017, an estimated 10·9 million cases of typhoid fever and 3·4 million cases of paratyphoid
145 fever (all types), resulted in 116·8 and 19·1 thousand deaths worldwide, respectively²³.
146 However, by contrast to *S. enterica* serotypes Typhi (STY)²⁴, Paratyphi A (SPA)²⁵, and
147 Paratyphi C²⁶, little is known about the population structure of SPB⁻. Connor and coworkers²⁷
148 were the first to try to elucidate the structure of this population, in a study of 191 isolates of

149 SPB (SPB⁺ and SPB⁻) from around the world. Their phylogenomic analysis identified 10
150 distinct lineages, named phylogroups (PGs 1-10). All 34 SPB⁻ isolates were grouped into the
151 invasive PG1 lineage, which was derived from the closely related lineages PG2 to -5,
152 containing SPB⁺ gastrointestinal isolates. Microbial genomic surveillance conducted in the
153 UK since 2014 led to the identification of phylogenetic clades of SPB⁻ strains associated with
154 travel to South America, Iraq, and Pakistan^{28,29}. However, the use of a limited number of
155 global SPB⁻ isolates or of contemporary routine SPB⁻ isolates ruled out any deeper global
156 phylogenomic analysis of this pathogen.

157

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

164

165 **RESULTS**

166

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

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

174 ensured that this diversity dataset comprised exclusively of genomes (i) with the correct *in*
175 *silico* serotype, (ii) containing the specific SNV associated with the loss of *d*-Tar fermentation
176 in SPB⁻ (ref.²³), (iii) and belonging to the invasive lineage, PG1 (ref.²⁶). This was achieved in
177 a straightforward manner using the HC400_1620 cluster of the EnteroBase *Salmonella* core-
178 genome MLST scheme (<https://enterobase.warwick.ac.uk/species/index/senterica>) as a proxy
179 **(Supplementary Text, Supplementary Fig. 1)**.

180

181 A maximum likelihood (ML) phylogeny of these 568 SPB⁻ PG1 genomes was constructed
182 from 15,995 single-nucleotide variants (SNVs) distributed over the non-repetitive, non-
183 recombinant core genome (**Fig. 1a, Supplementary Fig. 2, Supplementary Data 2**). Eleven
184 lineages were identified (L1 to L11), one of which (L10) predominated, accounting for 62%
185 (352/568) of the isolates. The frequency of L10 increased sharply over the study period, from
186 6.1% (2/33) between 1898 and 1950 to 81.6% (182/223) for the 2001 - 2021 period (**Fig. 1b**).
187 L10 was found worldwide, accounting for 38.6% (78/202) of the European isolates (i.e.,
188 isolates recovered from infections acquired in Europe) to 88.7% (86/97) of the Middle Eastern
189 isolates (**Fig. 1c**). Lineage L7 was also well distributed and was detected in isolates from all
190 the geographic regions other than the Middle East (**Fig. 1c**). Lineage L5 was more frequent in
191 Asia (except for Middle Eastern countries), whereas L2 and L9 were more frequently
192 identified in Europe.

193

194 Lineages were further subdivided into well-supported monophyletic groups at various
195 hierarchical levels, including clades, subclades, and an additional higher-resolution group (see
196 Methods “Defining lineages and genotypes”) (**Fig. 2a**). In total, we defined 38 hierarchical
197 genotypes with a phylogenetically informative nomenclature of the form
198 [lineage].[clade].[subclade].[higher-resolution group] (**Fig. 2a,b**). Eighteen genotypes had

199 strong geographic patterns (from country to continent level) (**Supplementary Data 1**). This
200 geographic information was added to the genotype nomenclature as an alias to make it more
201 informative.

202

203 The great majority ($n = 27$, 71%) of the 38 genotypes comprised European isolates. These
204 European isolates were found in the following genotype groups (each containing three or
205 more isolates): 2.0, 2.1, 4, 7.3, 9.0, 9.1_France, 10.1.1_Europe1, 10.2, and
206 10.3.8.5_EuropeNorthAfrica (**Table 1**). Our oldest strain (CIP A214) — probably isolated by
207 H. Conradi in Germany in 1898 — belonged to genotype 4. Middle Eastern isolates were
208 found in 50% (19/38) of the genotypes, with particularly high frequencies in the genotype 1,
209 2.1.1_Turkey1, 10.1, 10.3.2_MiddleEast1, 10.3.3_Turkey2, 10.3.5_MiddleEast2,
210 10.3.8.2_Turkey3, 10.3.8.3_MiddleEast3, and 10.3.8.4_MiddleEast4 groups. American
211 isolates belonged to 34.2% (13/38) of the genotypes and were particularly frequent in the
212 genotype 6, 7.1.1_Chile1, 10.3.4_Chile2, and 10.3.6_SouthAmerica groups. African
213 (excluding Egypt) and Asian (excluding Middle Eastern countries; see the footnotes of **Table**
214 **1**) isolates were each present in 23.7% (9/38) of the genotype groups. The African isolates,
215 which were almost exclusively from North Africa (only three were from East Africa), were
216 most frequently of genotypes 7.3.1_NorthAfrica1, 7.3.2_BAOR, 10.3.7_NorthAfrica2, and
217 10.3.9_NorthAfrica3. Two East African isolates from Madagascar collected in 1962 and 2001
218 were of genotype 10.2; whereas a third isolate collected from Ethiopia in 2018 was of
219 genotype 10.0 (**Supplementary Data 1**). The South Asian isolates belonged exclusively to
220 the genotypes of lineage L10 (in particular 10.3.1_SouthAsia1 and 10.3.8_SouthAsia2),
221 whereas almost all the East Asian isolates belonged to genotypes from older lineages (2.1, 5,
222 and 7.2_EuropeEasternAsia). The genotype distribution for animal isolates was not
223 significantly different from that for human isolates (**Supplementary Text**).

224

225 After confirming the presence of a strong temporal signal in our dataset (**Supplementary Fig.**
226 **3**), we applied a Bayesian phylogenetic approach to a spatially and temporally representative
227 subset of 256 isolates to estimate the nucleotide substitution rates and divergence times of the
228 different lineages (**Table 2**) and to construct a dated phylogeny (**Fig. 3 and Supplementary**
229 **Fig. 4**). We estimated the genome-wide substitution rate at 1.2×10^{-7} substitutions
230 site⁻¹ year⁻¹ (95% credible interval (CI) = 9.6×10^{-8} to 1.5×10^{-7}), giving a most recent
231 common ancestor (MRCA) for all SPB⁻ PG1 in our collection dating back to 1274 CE
232 (common era) (95% CI, 915 – 1583) (**Supplementary Text**). The MRCAs of the different
233 lineages were estimated to have existed in the 18th century or first half of the 19th century
234 (**Table 2**).

235

236 **Comparison of SNV-based phylogenetic data with phage-typing data**

237 World surveys of SPB PTs were regularly reported over several decades following World
238 War II (**Supplementary Fig. 5**). However, the phylogenetic value of the typing scheme used
239 at the time has never been assessed. We therefore analysed the correlation between phage-
240 typing and genomic data for the 254 SPB⁻ isolates from the diversity dataset for which phage-
241 typing results were available (**Fig. 4**). Good concordance was observed for the isolates of four
242 PTs: 92.8% (13/14) of PT 1 isolates belonged to lineage L2, 95.8% (91/95) of those typed as
243 Taunton belonged to L10, and 100% of BAOR (11/11) and Jersey (6/6) isolates belonged to
244 lineage L7 (**Supplementary Data 1**). The isolates of the BAOR and Jersey PTs even
245 belonged to single genotypes, 7.3.2_BAOR and 7.3, respectively. By contrast, isolates typed
246 as PT Taunton belonged to 10 different genotypes within L10 (**Supplementary Data 1**). The
247 only PT 2 isolate belonged to L1. The isolates of the remaining PTs were not assigned to a
248 predominant lineage and were instead considered to belong to four lineages. For example,

249 50.1% (32/63) of the isolates typed as Dundee belonged to L9 (all isolates were from Europe,
250 and all but one belonged to genotype 9.1_France), 46% (29/63) belonged to L10 (from
251 multiple genotypes, mostly corresponding to isolates from the Middle East and North Africa),
252 and 1.6% (2/63) belonged to L2.

253

254 **Evolution of antimicrobial resistance**

255 The emergence of antimicrobial resistance (AMR) in SPB⁻ PG1 is recent (**Fig. 5** and
256 **Supplementary Data 1**). Between 1898 and 2000, only one isolate (0.3%, 1/345) had
257 antibiotic resistance genes (ARGs). This human isolate (B73-1117), collected in France in
258 1973, displayed resistance to ampicillin (*bla*_{TEM-1D}), streptomycin (*strAB*, *aadA1*, and
259 *aadA2b*), sulfonamides (*sulI*), chloramphenicol (*cmlA1*), and tetracycline (*tetA*). Between
260 2001 and 2021, 23.1% (52/225) of isolates had ARGs. One isolate acquired in Turkey in 2001
261 (01-7995) produced a CTX-M-3 extended-spectrum beta-lactamase³⁰, whereas another isolate
262 (P7704) acquired in South America in 2019 produced an OXA-48 carbapenemase³¹. The
263 prevalent mechanisms of resistance (21.3%, 48/225) during this period involved mutations in
264 the quinolone resistance-determining regions of the *gyrA* and *gyrB* DNA gyrase genes,
265 leading to resistance to nalidixic acid and/or decreased susceptibility or resistance to
266 ciprofloxacin (minimum inhibitory concentration between 0.125 and 0.5 mg/L). The first
267 isolate harbouring such a mutation was isolated in 2004. Diverse mutations were observed,
268 with *gyrA*_S83F in 20 isolates, *gyrA*_D87Y in 12 isolates, *gyrA*_D87N in six isolates,
269 *gyrA*_D82N in three isolates, *gyrB*_S464F in three isolates, *gyrA*_D87G in two isolates, a
270 combination of *gyrA*_D87N and *gyrB*_S464F in one isolate, and a combination of *gyrA*_S83F
271 and *gyrA*_D87G in one isolate. These *gyrA* and *gyrB* mutations occurred over the entire
272 phylogenetic tree, in lineages L2 (*n* = 4), L3 (*n* = 1), L5 (*n* = 2), L7 (*n* = 3), L9 (*n* = 5), and
273 L10 (*n* = 33) (**Fig. 5**). The lineage L10 isolates containing such *gyrA* and *gyrB* mutations were

274 further classified into eight distinct genotypes. One cluster contained seven *gyrA*_D87Y
275 isolates acquired in Turkey between 2009 and 2017 (genotype 10.3.8.2_Turkey3)
276 **(Supplementary Data 1)**.

277

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

279 A pan-genome analysis of the 568 SPB⁻ PG1 genomes studied **(Supplementary Fig. 6)**
280 identified a closed pan-genome ($\alpha = 1.3$) containing a total of 5,681 genes, including a
281 core genome of 4,044 genes **(Supplementary Data 3)**. Most ($n = 696$, 46.2%) of the 1,506
282 accessory genes present in < 95% of the genomes were assigned to prophages
283 **(Supplementary Text, Supplementary Fig. 7, Supplementary Data 4)**. Two types of
284 prophage — a *Salmonella* Brunovirus SEN34-like and a Xuanwuvirus P88-like prophage —
285 contained *sopE*, a virulence gene that encodes an effector translocated into eukaryotic cells,
286 which promotes bacterial invasion through cytoskeleton rearrangement. All but two of the
287 568 genomes had at least one *sopE* prophage: most genomes (75.1%, 425/566) had one, some
288 (17.8%, 101/566) had two and a few (7.1%, 40/566) had three **(Fig. 6)**. All genomes
289 containing three *sopE* prophages belonged to genotype 9.1_France (associated with the
290 Dundee PT).

291

292 **Development of a new SNV-based genotyping tool and analysis of recent isolates**

293 We identified marker SNVs unique to each genotype (38 SNVs in total) **(Supplementary**
294 **Data 5)** and implemented this genotyping scheme in the open-source Mykrobe platform
295 (<https://www.mykrobe.com/>) **(Supplementary Text)**. After validation of the Mykrobe-
296 implemented version of the scheme on the 568 genomes from the diversity dataset, this
297 genotyping tool was used on the 336 genomes from the surveillance dataset (see Methods “*S.*
298 *enterica* serotype Paratyphi B sequence data collection”) **(Supplementary Data 6)** to identify

299 genotypes recently isolated in France, the UK, the US, and Canada (**Fig. 2c**). During the
300 2014-2023 period, 25 genotypes were observed. Genotype diversity was highest among
301 French isolates (17 genotypes for 84 isolates) and lowest in the UK (13 genotypes for 200
302 isolates). The most frequent genotype was 10.3.6_SouthAmerica, found in 49.1% (164/336)
303 of the isolates from all four countries. Its frequency ranged from 44% (125/284) for the
304 isolates collected in Europe to 76.9% (40/52) for the isolates collected in North America. The
305 second most frequent genotype, 10.3.2_MiddleEast1 (18.8%, 63/336), was found mostly in
306 the UK. The other genotypes were found in less than 10% of the isolates. There was an
307 overrepresentation of genotypes linked to South Asia in the UK (e.g., 10.3.1_SouthAsia1 and
308 10.3.8.1_SouthAsia2, together accounting for 11% (22/200) of the isolates), and of genotypes
309 linked to North Africa in France (e.g., 7.3.1_NorthAfrica1, 7.3.2_BAOR,
310 10.3.7_NorthAfrica2 and 10.3.9_NorthAfrica3, together accounting for 13.1% (11/84) of the
311 isolates). Interestingly, older genotypes — such as 2.1, accounting for 3.5% (7/200) of UK
312 isolates, and 9.1_France, accounting for 9.5% (8/84) of French isolates — are still being
313 isolated.

314

315 **DISCUSSION**

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

324
325 SPB⁻ strains, first identified in Europe in 1896, became a common cause of enteric fever
326 across Europe thereafter, suggesting that they were circulating in this particular region of the
327 world. Indeed, SPB⁻ genetic diversity was greatest among European isolates and those
328 collected in Turkey. Turkey, part of which lies in Europe, has a long shared history with the
329 rest of Europe through the rule of the Ottoman Empire over most of the countries in south-
330 eastern Europe for several centuries. SPB⁻ isolates were, however, less frequently reported in
331 other parts of the world (the Americas, East Asia, North Africa) early in the 20th century. In
332 Morocco, North Africa, SPB⁻ strains were reported to have been introduced around the city of
333 Fes by the French troops during the Rif war in 1925 (ref.³²). Two genotypes may have been
334 introduced into Morocco at this time: 7.3.1_NorthAfrica1 or 7.3.2_BAOR. Other
335 introductions from Europe may have occurred for older, pre-L10 lineages, including the
336 introduction of genotypes 2, 5 and 7.2_EuropeEasternAsia in East Asia, and of genotype
337 7.1.1_Chile1 in Chile.

338
339 Our data for historical isolates and a careful exploitation of the world surveys of PTs enabled
340 us to determine which genotypes were present during the first half of the 20th century. In the
341 UK, the genotypes associated with domestic infections¹⁶ were probably 2.0 and 2.1
342 (representing PTs 1 and 2) and 6 and 7.2_EuropeEasternAsia (representing PT 3). However,
343 many rare PTs were introduced into the UK after WWII following the return of millions of
344 demobilized soldiers and the reopening of tourist traffic¹⁶. Taunton predominated among
345 these new PTs, and included various genotypes of lineage L10, mostly found in non-European
346 isolates (from North Africa, South Asia, the Middle East and South America). In France, one
347 epidemic strain of PT Jersey emerged in Western France in 1951 (only 0.13%, 1/752 of the
348 isolates were phage-typed as Jersey in 1950), peaked in 1952 (60.1%, 252/419 of the isolates

349 in this year), and then became very rare in the 1970s^{12,18,33}. This strain could be assigned to
350 genotype 7.3 (monophasic SPB⁻, see **Supplementary Text**). Widespread outbreaks caused by
351 a strain phage-typed as Dundee also occurred in France during the spring and summer of 1949
352 (ref.¹⁶). This strain, assigned to genotype 9.1_France, remained prevalent in France until the
353 mid-1980s^{10,11,12,16,17,18,33}.

354

355 In Europe, improvements in hygiene following WWII, including food hygiene (control
356 measures for foods at risk implemented in bakeries in the UK), sanitation, and access to safe
357 water and antibiotics, likely prevented the spread of SPB⁻. Ultimately, this led to a steady
358 decrease in the prevalence of local PTB, with these infections increasingly confined to
359 travellers or migrants returning from the Middle East, North Africa, South America, and
360 South Asia. Our analysis of recent routine SPB⁻ PG1 isolates sequenced over the last 10 years
361 in France, the UK, the US, and Canada indicated that the most frequent genotype (~50% of all
362 these isolates) is currently 10.3.6_SouthAmerica. This genotype is mostly associated with
363 Andean countries from western South America such as Peru, Bolivia, and more recently
364 Argentina. We estimated that this genotype was introduced into South America, probably
365 from Europe, between 1899 (95%CI: 1871-1927) and 1918 (95%CI: 1888-1942). Between
366 1973 and 1977, only 1% (2/200) of the imported PTB cases in the UK were acquired in South
367 America¹⁴, and in 2023, genotype 10.3.6_SouthAmerica accounted for 78.1% (25/32) of all
368 PTB cases detected in France. In Argentina, an increasing number of PTB cases (total of
369 ~5,500 cases) have been reported in the Salta province (bordering Bolivia) since 2018 (ref.³⁴),
370 but no such epidemiological pattern was found in Peru and Bolivia. It is therefore important to
371 conduct local studies aiming to identify the areas of current PTB transmission (probably in
372 remote regions with tourist sites) and associated risk factors, to facilitate mitigation.
373 Interestingly, old genotypes are still being isolated in Europe. For example, 4% of the

374 surveillance isolates from the UK belonged to genotypes 2.1 and 5, and 9.5% of those in
375 France belonged to the 9.1_France genotype. The eight French cases, for which isolates were
376 not recovered from blood samples, were patients between 81 and 98 years of age, suggesting
377 that they may be long-term carriers infected several decades ago. We can therefore conclude
378 that the distribution of SPB⁻ PG1 genotypes within these four high-income countries reflects
379 (i) the destinations of their holidaymakers, (ii) movements of people linked to their colonial
380 and/or immigration histories, and finally (iii) remnants of past infections in long-term carriers.

381

382 One feature particular to SPB⁻ strains is their sensitivity to antimicrobial drugs, which is
383 greater than that of other agents of enteric fevers, such as STY²⁴ and, to a lesser extent, SPA
384 strains²⁵. AMR mostly concerned quinolones and their resistance determinants, emerging over
385 the last 20 years, without horizontal transmission. Different mutations of the chromosomal
386 *gyrA* gene occurred in many genotypes across the entire phylogeny, in isolates from Europe,
387 South America, East and South Asia, and the Middle East, suggesting a high level of
388 fluoroquinolone exposure worldwide. However, no successful AMR SPB⁻ strains have
389 emerged contrasting with STY, the H58 clone of which emerged in South Asia in the 1980s
390 before spreading globally^{35,36}. SPB⁻ strains are not very frequent and have only recently been
391 detected in South Asia — where selection pressures exerted by quinolones and, later,
392 fluoroquinolones, began early — may account for this lower level of AMR than in STY and
393 SPA strains.

394

395 We have also provided a new robust framework for identifying and tracking the causal agent
396 of PTB. Firstly, SPB isolates can be assigned to the 10 known lineages (PGs)²⁷ with the
397 EnteroBase cgMLST scheme, with the HC400_1620 cluster considered a signature of SPB⁻
398 PG1. This cluster has been shown to be more informative than the traditional MLST7 ST86

399 criterion. Secondly, we have developed a hierarchical SNV-based genotyping scheme for
400 tracking the different SPB⁻ PG1 strains, an approach previously successfully used for the
401 surveillance of the two main agents of enteric fever, STY³⁷ and SPA³⁸. Our scheme, now
402 implemented in the open-source Mykrobe software ([https://github.com//mykrobe-
404 tools/mykrobe](https://github.com//mykrobe-
403 tools/mykrobe)), can identify 38 different genotypes with a phylogenetically — and sometimes
405 phylogeographically — informative nomenclature. The use of this scheme and its universal
406 nomenclature will undoubtedly improve the laboratory surveillance of PTB globally. For
407 example, genomic surveillance in the UK identified an imported SPB⁻ outbreak in travellers
408 coinciding with a mass gathering in Iraq in 2021 (ref.²⁹). The isolates from these patients
409 clustered in one of the two clades labelled “travel to Iraq”. According to our genotyping
410 scheme, this clade corresponds to genotype 10.3.2_MiddleEast1. The oldest isolates
411 belonging to this genotype were collected in Iran in 1965 and Iraq in 1975, suggesting that
412 this strain has been endemic in the region for many decades. The second clade labelled “travel
413 to Iraq” identified by UKHSA corresponded to genotype 10.3.8.3_MiddleEast3.

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

421

422

423

424 **ACKNOWLEDGEMENTS**

425 This research was funded by the *Fondation Le Roch-Les Mousquetaires*; *Institut*
426 *Pasteur*; *Santé publique France*; and by the French Government's Investissement d'Avenir
427 programme, Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases'
428 (grant no. ANR-10-LABX-62-IBEID). We thank Prof. Jacques Ravel, Prof. David A. Rasko,
429 Luke Tallon, Kranthi Vavikolanu, and Michael Pietsch for submitting archived or new short
430 reads to a public repository, Susan Van Duyne for ensuring safe shipping of strains, Anthony
431 M. Smith and Chien-Shun Chiou for reviewing their data, Paul O'Dette for his support, and
432 the sequencing team at the Institut Pasteur (PF1 & P2M-Plateforme de Microbiologie
433 Mutualisée) for sequencing the samples. We also thank all the corresponding laboratories of
434 the French National Reference Centre for *Escherichia coli*, *Shigella*, and *Salmonella*. Martin
435 Hunt was funded by the National Institute for Health Research Health Protection Research
436 Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at
437 Oxford University in partnership with the UK Health Security Agency (NIHR200915), and
438 the NIHR Biomedical Research Centre, Oxford. Marie Anne Chattaway is affiliated to the
439 National Institute for Health Research Health Protection Research Unit (NIHR HPRU)
440 (NIHR200892) in Genomics and Enabling Data at University of Warwick in partnership with
441 the UK Health Security Agency (UKHSA), in collaboration with University of Cambridge
442 and Oxford. Marie Anne Chattaway is based at UKHSA. The views expressed are those of the
443 authors and not necessarily those of the Centers for Disease Control and Prevention, the
444 NIHR, the Department of Health and Social Care or the UK Health Security Agency. The
445 funders had no role in study design, data collection and analysis, decision to publish, or
446 preparation of the manuscript.

447

448

449 **CONTRIBUTIONS**

450 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.,
451 J.I., D.C., Y.A., Y.W., B.J., M.C., M.Y., B.Z., M.M., C.S.N., M.P.G., and F.-X.W. selected
452 and provided isolates or genomes with their basic metadata. L.FR., A.T.D., S.I.J., M.R., V.G.,
453 E.S., and L.FA. did the subculturing, phenotypic experiments, and DNA extractions. J.H.,
454 L.FR., A.T.D., A.Z., E.N., L.FA., and F.-X.W. analysed and/or interpreted the data. J.H.,
455 M.H. and Z.I. implemented the genotyping scheme in Mykrobe. J.H., L.FR., and F.-X.W.
456 wrote the manuscript. All the authors contributed to manuscript editing.

457

458 **ONLINE METHODS**

459 **Ethics statement**

460 This study was based exclusively on bacterial isolates (including historical and reference
461 strains) and associated metadata. The great majority of these isolates ($n = 362$) originated
462 from Institut Pasteur (reference laboratories or *Collection de l'Institut Pasteur* (CIP)); 115
463 isolates corresponded to historical or reference strains collected between 1898 and 1971 and
464 247 were bacterial isolates collected under the mandate for laboratory-based surveillance
465 awarded by the French Ministry of Health to the National Reference Centre for *Escherichia*
466 *coli*, *Shigella* and *Salmonella* (NRC-ESS) since 1972. Data collection and storage by the
467 NRC-ESS was approved by the French National Commission for Data Protection and
468 Liberties (“*Commission Nationale Informatique et Libertés* (CNIL)”; approval number:
469 1474659). For other human bacterial isolates collected (or the genome sequences derived
470 from them) by participating reference laboratories under local mandates for laboratory-based
471 surveillance of salmonellosis in line with local laws and regulations, the associated metadata
472 did not contain any personal identifiable information and were restricted to year and country

473 of isolation, and international travel information. As a result, neither informed consent nor
474 approval from an ethics committee was required.

475

476 ***S. enterica* serotype Paratyphi B sequence data collections**

477 *Diversity dataset*

478 We first studied a diversity dataset of 568 SPB⁻ genomes, 446 of which were generated
479 specifically for this study, 109 had already been published^{127,28,29,31,39,40,41}, and the other 13
480 were unpublished but deposited in EnteroBase

481 (<https://enterobase.warwick.ac.uk/species/index/senterica>) or GenBank

482 (<https://www.ncbi.nlm.nih.gov/genbank/>) (**Supplementary Data 1**). This diversity dataset
483 also contained the reference strain (116K) of *S. enterica* serotype Onarimon (1,9,12:b:1,2)
484 considered an O-antigen variant of SPB⁻ (**Supplementary Text**).

485

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

494

495 The 568 genomes from the diversity dataset were obtained from isolates collected between
496 1898 and 2021 from humans (472/568, 83.1%), environmental samples (42/568, 7.4%),
497 animals (14/568, 2.5%), food items (2/568, 0.3%) or from unknown sources (38/568, 6.7%)

498 **(Supplementary Data 1)**. For the human isolates, 43% (203/472) were obtained from blood,
499 39.2% (185/472) from stools, 3.8% (18/472) from other sources (urine, cerebrospinal fluid,
500 pus, cysts, wounds, gallbladders, bile) and 14% (66/472) were of unknown origin. The 568
501 isolates and strains were isolated locally or from travellers and originated from 41 countries in
502 Europe (202/568, 35.6%), Asia (144/568, 25.3%), the Americas (127/568, 22.3%), and Africa
503 (81/568, 13.2%). Some historical or laboratory strains were of unknown geographic origin
504 (14/568, 2.5%). The European isolates accounted for 57.9% (135/233) of the total isolates
505 collected in the 1898-1980 period, decreasing to 20% (67/335) in the 1981-2022 period.

506

507 *Surveillance dataset*

508 We assembled a surveillance dataset of 336 SPB⁻ genomes routinely obtained by four public
509 health microbiology laboratories (CDC, UKHSA, IP and Canada's National Microbiology
510 Laboratory) and submitted to EnteroBase between August 27th 2015 and May 2nd 2023. These
511 336 genomes included 111 already present in the “diversity dataset” (**Surveillance Data 6**).

512

513 **Antimicrobial drug susceptibility testing**

514 Antimicrobial drug susceptibility was determined at IP for 273 SPB⁻ isolates from the
515 diversity dataset (**Supplementary Data 1**) by disk diffusion on Mueller-Hinton (MH) agar in
516 accordance with the guidelines of the Antibiogram Committee of the French Society for
517 Microbiology (CA-SFM) / European Committee on Antimicrobial Susceptibility Testing
518 (EUCAST)⁴². The following antimicrobial drugs (Bio-Rad, Marnes-la-Coquette, France) were
519 tested: amoxicillin, ceftriaxone, ceftazidime, streptomycin, kanamycin, amikacin, gentamicin,
520 nalidixic acid, ofloxacin, ciprofloxacin, sulfonamides, trimethoprim, sulfamethoxazole-
521 trimethoprim, chloramphenicol, tetracycline, and azithromycin. *Escherichia coli* CIP 76.24
522 (ATCC 25922) was used as a control. The minimal inhibitory concentrations (MICs) of

523 nalidixic acid and ciprofloxacin were determined for 84 isolates (all 24 isolates resistant to
524 nalidixic acid in disk diffusion tests and 60 susceptible isolates chosen on the basis of year of
525 isolation and antibiotic resistance gene content) with Etest strips (bioMérieux, Marcy L'Etoile,
526 France). As a means of distinguishing *Salmonella* isolates susceptible to ciprofloxacin
527 (minimum inhibitory concentration [MIC] ≤ 0.25 mg/L), which are wild-type (WT), from
528 those that are non-WT, we defined two categories based on the epidemiological cutoffs used
529 by the CLSI: decreased susceptibility to ciprofloxacin (MIC between 0.12 and 0.5 mg/L) and
530 true susceptibility to ciprofloxacin (MIC ≤ 0.06 mg/L)⁴³. Please note that due to clinical
531 evidence of poor response to ciprofloxacin in systemic infections caused by *Salmonella* spp.
532 isolates displaying such decreased susceptibility to ciprofloxacin, the clinical breakpoint to
533 define ciprofloxacin resistance in non-enteric isolates of this species is MIC > 0.06 mg/L⁴².

534

535 **Short-read sequencing**

536 Illumina short-read sequencing was performed by the IP genomics platforms (PF1 and the
537 Mutualised Platform for Microbiology (P2M)) for 389 isolates (362 from IP plus 12 from
538 PISP and 15 from the CDC). At PF1 (150 isolates sequenced), total DNA was extracted with
539 the Wizard Genomic DNA Kit (Promega, Madison, WI, USA) and fragmented with a Covaris
540 E220 ultrasonicator. Sequencing libraries were then prepared with the NEXTflex PCR-free
541 DNA-Seq Kit (Bioo Scientific Corporation, Austin, TX, USA) and sequencing was performed
542 with the HiSeq2500 system (Illumina, San Diego, CA, USA). At P2M (239 isolates
543 sequenced), total DNA was extracted with the MagNA Pure DNA isolation kit (Roche
544 Molecular Systems, Indianapolis, IN, USA). Sequencing libraries were prepared with the
545 Nextera XT kit (Illumina), and sequencing was performed with the NextSeq 500 system
546 (Illumina). The other 57 SPB⁻ isolates were sequenced by the participating laboratories, in

547 accordance with their usual practices (Supplementary Methods section “Short-read
548 sequencing”).
549
550 Paired-end reads varied in length according to the sequencing platform/site, from 95 to 300
551 bp, yielding a mean coverage of 124-fold for each isolate (minimum 25-fold, maximum 350-
552 fold) (**Supplementary Data 1**).

553
554 Taxonomic read classification with Kraken⁴⁴ v.2.1.1 was used to confirm that sequencing
555 reads originated from *Salmonella* and not from a contaminant. All short-read data were made
556 publicly available. Their genome accession numbers are provided in **Supplementary Data 1**.

557

558 **Long-read sequencing and complete genome circularisation**

559 At the time of the study, there was no complete SPB- PG1 genome that could be used as a
560 reference genome. We therefore sent total DNA extracted from strain CIP 54.115 (ref.²⁷) to
561 GATC Biotech (now Eurofins Genomics) for long-read sequencing by the Pacific
562 BioSciences RSII platform. The high-quality *de novo* assembly was performed by GATC
563 Biotech after the hierarchical genome-assembly process (HGAP) workflow⁴⁵. We performed
564 an additional step of assembly polishing with short reads and the Pilon⁴⁶ v.1.23 tool.

565

566 Long-read sequences for another 11 isolates — selected to provide a maximal representation
567 of diversity in terms of lineages and phage types — were generated with Oxford Nanopore
568 Technology (ONT) (**Supplementary Data 1**). Bacterial DNA was extracted from cultures
569 grown Trypto-casein soy broth (Bio-Rad) at 37°C with shaking at 200 rpm. We used
570 Genomic-tip 100/G columns (Qiagen) according to the manufacturer’s instructions for DNA
571 extraction. DNA integrity and the absence of RNA were checked by agarose gel

572 electrophoresis and by the determination of A_{260}/A_{230} and A_{260}/A_{280} ratios with a NanoDrop™
573 2000 spectrophotometer. DNA concentrations were measured with the Qubit system and the
574 dsDNA *BR* Assay Kit (Invitrogen). Libraries were prepared from total DNA with the SQK-
575 LSK109 ligation sequencing kit and the EXP-NBD104/114 barcoding kit according to the
576 ONT procedure (Native Barcoding Amplicons protocol version
577 ACDE_9064_v109_revP_14Aug2019, [dx.doi.org/10.17504/protocols.io.bgzxjx7n](https://doi.org/10.17504/protocols.io.bgzxjx7n)). Libraries
578 were sequenced with R9.4.1 flow cells and a Mk1C MinION sequencer. Base calling was
579 performed with Guppy⁴⁷ v.4.3.4 or v.5.0.13. The filtlong tool
580 (<https://github.com/rrwick/Filtlong/>) v.0.2.1 was used to filter reads according to their length
581 (min_length 800 bp) and quality (keep_percent 90). Read lengths ranged from 3,827 to
582 20,347 bp (mean of 9,923 bp), with a mean of 262-fold coverage per genome (minimum 54-
583 fold, maximum 676-fold). Complete *de novo* genome assemblies were generated with the
584 Trycycler⁴⁸ v.0.5.0 pipeline using default parameters. For each isolate, long reads were
585 subsampled into 12 sets, which were subsequently used to generate 12 independent
586 assemblies (four sets per assembler) with Flye⁴⁹ v.2.9, raven⁵⁰ v.1.6.0 or miniasm⁵¹ v.0.3-&-
587 Minipolish⁵¹ v.0.1.3. The consensus assembly was first long read-polished with medaka
588 v.1.4.4 (<https://github.com/nanoporetech/medaka>) and then short read-polished four times
589 with pilon⁴⁶ v.1.23. The final assemblies were annotated with bakta⁵² v.1.5.0.

590

591 **Other genomes**

592 SPB⁺ PG2 strain 201906085 (ENA accession no. ERR12749341) and SPB⁺ PG3 strain SPB7
593 (GenBank accession no. NC_010102.1) were used as outgroups to identify the ancestral
594 lineage of SPB⁻ PG1 genomes (**Supplementary Data 1**).

595

596

597 **Genomic typing methods**

598 A two-step *in silico* serotype confirmation procedure was used: O-antigen determination by a
599 fast kmer-alignment method from KMA⁵⁵ v.1.4.14 for the alignment of raw paired-end reads
600 against SPB-specific sequences within the *rfb* cluster (**Supplementary Table 1**) and *fliC*
601 (encoding the H1 flagellin) and *fliB* (encoding the H2 flagellin), with calling by the
602 NCBI BLAST+ (ref.⁵⁴) v.2.14.1 blastn command line tool on assemblies against SPB
603 reference sequences (**Supplementary Table 1**). The specific STM 3356 SNV present in *d*-
604 Tar⁻ isolates²³ was sought with the NCBI BLAST+ (ref.⁵⁴) v.2.14.1 blastn command line
605 tool on assemblies against SPB⁻ and SPB⁺ reference sequences (**Supplementary Table 1**).
606
607 Multilocus sequence typing (MLST)²², and core genome MLST (cgMLST) were performed
608 with various tools integrated into EnteroBase⁵⁵
609 (<https://enterobase.warwick.ac.uk/species/index/senterica>). The
610 EnteroBase *Salmonella* cgMLST scheme (“cgMLST V2 + HierCC V1”) —based on 3002
611 core genes — assigns bacterial genomes to single-linkage hierarchical clusters (HCs) at 13
612 fixed levels of resolution, ranging from HC0 (high-resolution clusters consisting of identical
613 genomes with no allelic differences) to HC2850 (low-resolution clusters consisting of
614 genomes with up to 2850 allelic differences). Evaluations by Zhou and coworkers⁵⁶ found
615 that, in the genus *Salmonella*, cluster assignments at the HC2850, HC2000, and HC900 levels
616 could be used to distinguish subspecies, super-lineages and eBurst groups²², respectively, and
617 that epidemic outbreaks could be distinguished with HC2, HC5, or HC10. Sequence type and
618 cgMLST clustering at the HC2, HC5, HC10, HC20, HC50, HC100, HC200, HC400, HC900,
619 HC2000, HC2600 and HC2850 levels are shown in **Supplementary Data 1** for the genomes
620 from the diversity dataset and in **Supplementary Data 6** for the genomes from the
621 surveillance dataset.

622 We also constructed a cgMLST tree — based on core genome allelic distances —inferred
623 with the NINJA neighbour-joining algorithm (present in the “cgMLST V1+HierCC V1”
624 scheme) and visualised with GrapeTree⁵⁷ for two different datasets. The first dataset
625 corresponded to the 166 genomes from Connor and coworkers²⁷ present in EnteroBase
626 (https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86472),
627 as 12 genomes did not pass the quality control of EnteroBase and one was discarded due to a
628 probable inversion (**Supplementary Data 7**). The second corresponded to the 567 genomes
629 from our diversity dataset that were present in EnteroBase
630 (https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86468);
631 one unpublished draft genome (ATCC 10719, SRR955214) sequenced with 454 technology
632 could not be uploaded by EnteroBase, which accepts only complete genomes or Illumina
633 short reads. The resulting cgMLST trees used for **Supplementary Figures 1 and 7** are
634 publicly available from https://enterobase.warwick.ac.uk/ms_tree?tree_id=92077 and
635 https://enterobase.warwick.ac.uk/ms_tree?tree_id=92095, respectively. We found that some
636 of the strains described in the Supplementary Table 1 of the article by Connor *et al.*²⁷ were
637 assigned to incorrect STs, PGs (**Supplementary Data 7**). We have corrected the assignments
638 for this study.

639

640 EnteroBase was searched on April 9th 2021, using the HC400_1620 criterion (see
641 **Supplementary Text**) to identify additional genomes, thereby extending the geographic and
642 time coverage of our diversity dataset. EnteroBase was also searched on April 10th 2023 with
643 the same HC400_1620 criterion, to assemble the surveillance dataset of 336 genomes
644 (https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:91222).

645

646

647 **Detection of antimicrobial resistance determinants**

648 Antimicrobial resistance (AMR) genes were detected with SRST2 (ref.⁵⁸) v.0.2.0 using
649 default parameters and the CARD⁵⁹ v.3.0.8 AMR gene database. We also determined the
650 presence or absence of mutations in the quinolone resistance-determining regions (QRDRs)
651 by extracting the relevant SNV calls for codons 83 and 87 of *gyrA*, codon 464 of *gyrB*, and
652 codon 80 of *parC*, from the SPB⁻ PG1 reference genome CIP 54.115 (GenBank accession no.
653 CP147898) (**Supplementary Data 2**).

654

655 **Phylogenetic analyses**

656 The paired-end reads and simulated paired-end reads were mapped onto the complete SPB⁻
657 PG1 reference genome CIP 54.115 (GenBank accession no. CP147898) with RedDog
658 (<https://github.com/katholt/reddog-nf>). Genomes with a depth of $\leq 10x$ across the reference
659 genome and a ratio of heterozygous SNVs/homozygous SNVs > 0.7 were excluded from the
660 analysis, leaving 568 genomes for downstream analysis. We excluded SNVs present in fewer
661 than 99% of genomes, and SNVs in repeat regions, such as insertion sequences or phages.
662 Recombination was masked with Gubbins⁶⁰ v.2.3.2 using the `weighted_robinson_foulds`
663 parameter. The resulting SNV alignment of 15,995 SNVs (see **Supplementary Data 2** for a
664 full list of locations) was used to create a maximum likelihood phylogeny with RAxML⁶¹
665 v.8.2.9, with a GTR+G base substitution model and 1000 bootstraps. The final tree was rooted
666 on the lineage L1 SPB⁻ PG1 genome NCTC 8299 (GenBank accession no. AE014073) and
667 visualised with iTOL⁶² v.6 (<https://itol.embl.de>).

668

669 We first assessed the temporal signal in the diversity dataset (with and without the five
670 outliers) by “root-to-tip” regression (linear regression of the number of substitutions
671 accumulated from the root to the tips of the ML phylogenetic tree as a function of sampling

672 times) with TempEst⁶³ v.1.5.3 (**Supplementary Fig. 3 and Supplementary Data 1**). We
673 confirmed the temporal signal in a subset of 256 genomes (see below) by performing eight
674 date randomisations⁶⁴ in BEAST⁶⁵ v.2.7.1. The eight runs with randomised dates gave
675 significantly different estimates for substitution rate compared to the run with real dates
676 (**Supplementary Fig. 3**), indicating a strong temporal signal in the real data.

677
678 For the generation of a timed phylogeny, we selected a subset of 256 genomes representing
679 the diversity of the phylogeny and the full range of isolation dates and geographies
680 (**Supplementary Data 1**). We used all genomes in lineages L1 to L6 and L8, and lineages L7,
681 L9 and L10 were subsampled by selecting the oldest and newest genomes in each lineage,
682 together with a random selection of the remaining genomes in the lineage, keeping the
683 numbers of genomes sampled proportional to the full dataset (**Supplementary Data 1**). We
684 used BEAST⁶⁵ v.2.7.1 with a GTR substitution model, the optimised relaxed clock model
685 (setting the uclMean prior to a lognormal distribution with a mean of 0.0001 and a standard
686 deviation of 2, an initial value of 0.0001, and ensuring the ‘mean in real space’ option was
687 checked) and the Bayesian Skyline population model. We performed three independent runs
688 of 600 million states. We removed a 10% burn-in for each run and combined the runs together
689 to create the consensus file. A consensus tree was drawn with the ‘majority rule’ option from
690 sumtrees.py in dendropy⁶⁶ v.4.5.2, setting branch lengths to the median length. The skyline
691 plot was drawn with Tracer⁶⁷ v.1.7.2.

692

693 **Defining lineages and genotypes**

694 We used fastbaps⁶⁸ v.1.0.8 and visual inspection to define lineages in the maximum
695 likelihood tree, using the best_baps_partition function with the phylogeny as a prior.

696

697 Genotypes within lineages were defined following visual inspection. For the selection of
698 marker SNVs for each genotype, all SNVs were mapped onto the phylogeny with SNPPar⁶⁹
699 v.1.1. On branches leading to genotypes, we selected marker SNVs, prioritising SNVs that
700 were synonymous mutations within genes for which the ratio of non-synonymous SNVs to
701 synonymous SNVs was as close to zero as possible. The final selection of marker SNVs, their
702 coding consequences and genome location relative to the reference can be found in

703 **Supplementary Data 5.**

704

705 **Implementation and validation of genotyping scheme**

706 We implemented the genotyping scheme in Mykrobe⁷⁰ v.0.12.2, meaning that the scheme is
707 supported by this version onwards and is installed when running the mykrobe commands
708 “panels update_metadata” and “panels update_species all”. We used the pre-existing *invA*
709 probe⁷¹ to detect genomes belonging to *S. enterica*. An additional probe was created to
710 distinguish between SPB PG1 genomes and other serotypes, including non-PG1 Paratyphi B
711 genomes — this probe detects the *d*-Tar specific SNV (i.e., a G->A change in the start codon
712 of STM 3356)²³ (**Supplementary Table 2**). We tested the Mykrobe implementation on all
713 568 genomes from the diversity dataset (**Supplementary Data 1**), using Illumina reads as
714 input, to confirm genotypes were assigned correctly. We then tested the scheme on the 336
715 genomes of the surveillance dataset (**Supplementary Data 6**). All these genomes were
716 mapped onto the reference genome as described above and included in the ML phylogeny
717 (which was constructed with IQTree⁷² v.2, with a GTR substitution model, on 17,338 SNVs),
718 resulting in 793 genomes (**Supplementary Fig. 8**). We assigned a genotype to each genome
719 with Mykrobe, and genotypes were validated against the phylogeny to ensure they were
720 correct.

721

722 ***De novo assembly***

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

725

726 Assemblies were generated from 454-GS-FLX reads with SPAdes⁷³ v.3.15.5 (k-values of 21,
727 37, 53, 69, 77, default settings for other criteria).

728

729 **Pan-genome analysis**

730 The pan-genome analysis was performed on the diversity dataset, which included 566 Illumina
731 paired-end reads, one published complete genome (P7704), and one publicly available draft
732 genome from 454-GS-FLX reads (CFSAN024725) (**Supplementary Data 1**). We used
733 panaroo⁷⁴ v.1.3.0, with default parameters and genome assemblies annotated with bakta⁵²
734 v.1.5.0. We estimated the pan-genome openness level (open if $\alpha \leq 1$ and closed if α is
735 > 1) for the 568 genomes, using Heaps' law with the R micropan⁷⁵ v.2.1 package. The gene
736 presence/absence matrix of the pan-genome created is shown in **Supplementary Data 3**. We
737 increased the accuracy of accessory gene assignment to prophages or plasmids by performing
738 a second pan-genome analysis on 581 genomes constructed from 567 short-read assemblies and
739 14 complete genomes (see next paragraph). After manual curation of the prophage content in
740 the complete genomes (see next paragraph), we selected the 1,506 accessory gene IDs of the
741 568-pan-genome in the 581-pan-genome output, and we were able to assign the 1,506 accessory
742 CDS to prophage and plasmid regions. An additional manual curation of short-read assemblies
743 was performed for plasmid contigs potentially absent from the 14 complete genomes
744 (**Supplementary Data 4**).

745

746

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

748 Unlike short-read assemblies, complete genome assemblies provide complete information
749 about prophage content and location. We therefore used 14 complete SPB⁻ genomes: the
750 genome included in our diversity dataset (P7704), and 13 additional genomes from isolates with
751 short reads included in the “diversity dataset”: SARA41_FB_1, which was publicly available
752 (GenBank accession no. CP074225.1) and 12 genomes newly generated for this study
753 (Materials section “Long-read sequencing and complete genome circularisation”)
754 (**Supplementary Data 1**). The prophage regions were detected in these genomes and
755 taxonomically classified with the PHASTER⁷⁶ tool in June 2023. Prophages annotated as intact
756 or absent in at least one of the 14 complete genomes were retained for further analysis. Ten
757 prophage regions were then precisely delineated from the alignment of the 14 genomes around
758 each insertion site (**Supplementary Data 8**). Finally, we screened for occupancy of the 10
759 insertion sites in the 568 genomes of the diversity dataset with the blastn algorithm (BLAST⁺⁵⁴
760 v.2.14.1). As a means of obtaining a good estimate of the *sopE* prophage copy number per
761 genome (*sopE*-CN), we combined insertion site occupancy and short-read coverage data. First,
762 the short reads were aligned with the complete genome of the B62 isolate (GenBank accession
763 no. CP147902), with the very-sensitive-local option of bowtie2 (ref.⁷⁷) v.2.3.5.1, and coverage
764 was estimated across the entire genome (cov_g), at the *sopE* locus (cov_s) and the *Salmonella*
765 Pathogenicity Island 1 locus (GenBank accession no. CP147902; coordinates 981997 –
766 1036581) (cov_pi) with samtools⁷⁸ v.1.13. The mean value of the cov_s/cov_pi and cov_s/
767 cov_g ratios was chosen as an estimate of the *sopE*-CN. The sequencing platform has a strong
768 and significant effect on *sopE*-CN values (Kruskal-Wallis test $p = 4.8E-50$, large magnitude
769 effect $\eta^2[H] = 0.403$). We therefore split the dataset according to the Illumina platform used
770 (HiSeq, $n = 274$; MiSeq, $n = 45$; or NextSeq, $n = 242$) before searching for outliers in the *sopE*-
771 CN distribution with Dunn's pairwise comparison test with Bonferroni correction

772 **(Supplementary Fig. 9).** When the *sopE*-CN and the number of insertion sites occupied did
773 not correlate, we visually inspected both long- and short-read mappings onto the genomes of
774 isolates B62 and B2590, with Tablet⁷⁹ v.1.21.02.08. The complete data are summarised in
775 **Supplementary Data 9.**

776

777 Prophages inserted in the vicinity of the *hin-fljB-fljA* cluster are described in **Supplementary**
778 **Data 8.** The assemblies from the 568 diversity dataset genomes were screened for the
779 presence and structure of the *hin-fljB-fljA* gene cluster with the blastn⁸⁰ algorithm.

780

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

782 We used the gggenes⁸¹ v.0.5.0 and ggplot2 (ref.⁸²) v.3.4.2 packages of R⁸³ v.4.1.2 to visualise
783 prophage gene content.

784

785 **Statistics**

786 Statistical analysis was performed with R⁸³ v.4.1.2, R packages rstatix⁸⁴ v.0.7.2, and ggpubr⁸⁵
787 v.0.6.0.

788

789 **Data availability**

790 The publicly available sequences used in this study are available from GenBank under
791 accession numbers CP147895-CP147907. The short-read sequence data generated in this study
792 were submitted to EnteroBase (<https://enterobase.warwick.ac.uk/>) and to the European
793 Nucleotide Archive (ENA, <https://www.ebi.ac.uk/ena/>) under study numbers PRJDB11608,
794 PRJEB18998, PRJEB28356, PRJEB28356, PRJEB30317, PRJEB67705, PRJNA248792,
795 PRJEB68323, PRJEB49424, PRJEB71958. All the accession numbers of the short-read

796 sequences produced and/or used in this study are listed in **Supplementary Data 1** and
797 **Supplementary Data 6**.

798

799 The list of genomes studied (and their assembled short-read data) can be obtained from
800 EnteroBase at:

801 https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86468

802 (diversity dataset, **Supplementary Data 1**) and:

803 https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:91222

804 (surveillance data set, **Supplementary Data 4**). The list of 166 genomes (and their assembled

805 short-read data) published by Connor *et al.*³⁵ and present in EnteroBase can be obtained from

806 https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:86472.

807 The cgMLST GrapeTrees shown in **Supplementary Figures 1 and 7** can be visualised from

808 EnteroBase at: https://enterobase.warwick.ac.uk/ms_tree?tree_id=92077 and

809 https://enterobase.warwick.ac.uk/ms_tree?tree_id=92095, respectively.

810

811 **Code availability**

812 Mykrobe is available for download at <https://github.com//mykrobe-tools/mykrobe>.

813 Instructions for running Mykrobe for SPB- PG1 are available in the Mykrobe documentation

814 at <https://github.com/Mykrobe-tools/mykrobe/wiki/AMR-Prediction>. Briefly, this means

815 using the option --species paratyphiB when running the mykrobe predict command. The

816 genotyping panel developed here is available at

817 https://figshare.com/articles/dataset/Mykrobe_panel_paratyphi_B_version_20230627/249255

818 [06](#)

819 The fq2dna script (genome *de novo* assembly from raw paired-end FASTQ files) can be found

820 at <https://gitlab.pasteur.fr/GIPhy/fq2dna>.

821

822 **Data collection**

823 The data were entered into an Excel (Microsoft) version 16.76 spreadsheet or tabulation-

824 separated text files (tsv).

825

826 **REFERENCES**

827

- 828 1. Achard, C. & Bensaude, R. Infections paratyphoïdiques. *Bull. Mem. Soc. Hop. Paris*
829 **13**, 820-833 (1896).
- 830 2. Gwyn, N.B. On infection with a para-colon bacillus in a case with all the clinical
831 features of typhoid fever. *Bull. Johns Hopkins Hosp.* **9**, 54-56 (1898).
- 832 3. Hirschfeld, L. A new germ of paratyphoid. *Lancet* **193**, 296-297 (1919).
- 833 4. Pratt, J.H. On paratyphoid fever and its complications. *Boston Med. Surg.* **148**, 137-
834 142 (1903).
- 835 5. Proescher, F. & Roddy, J.A. Bacteriological studies on paratyphoid A and paratyphoid
836 B. *Arch. Intern. Med.* **3**, 263-312 (1910).
- 837 6. Grimont, P.A.D & Weill F.X. Antigenic formulae of the *Salmonella* serovars, 9th ed.
838 World Health Organization Collaborating Center for reference and research on
839 *Salmonella*, Institut Pasteur (2007). Available online at:
840 https://www.pasteur.fr/sites/default/files/veng_0.pdf (accessed March 25, 2024).
- 841 7. Kristensen, M. & Bojlen, K. Vergärungsmässig definierte Typen des Paratyphus B-
842 Bazillus. *Zentralbl. Bakteriol. Parasitenk. Infektionskr. I. Abt. Orig.* **114**, 86-108
843 (1929).
- 844 8. Vogelsang, T.M. Paratyphoid B in Western Norway: bacteriology, symptomatology,
845 morbid anatomy and epidemiology of infection with bacillus paratyphosus B. *J. Infect*
846 *Dis.* **55**, 276-298 (1934).
- 847 9. Savage, W. Paratyphoid fever: an epidemiological study. *J. Hyg.* **42**, 393-410 (1942).
- 848 10. Felix A. World survey of typhoid and paratyphoid-B phages types. *Bull. World Health*
849 *Organ.* **13**, 109-170 (1955).
- 850 11. Nicolle P. Rapport sur la distribution des lysotypes de *Salmonella typhi* et de *S.*
851 *paratyphi B* dans le monde, d'après les résultats fournis par les centres nationaux
852 membres du comité international de la lysotypie entérique à l'occasion du congrès
853 international de microbiologie, Stockholm, 1958. *Ann. Inst. Pasteur* **102**, 389-409
854 (1962).
- 855 12. International Committee for Enteric Phage-Typing (ICEPT). The geographical
856 distribution of *Salmonella typhi* and *Salmonella paratyphi A* and *B* phage types during
857 the period 1 January 1966 to 31 December 1969. *J. Hyg.* **71**, 59-84 (1973).
- 858 13. Braddick, M.R. & Sharp, J.C. Enteric fever in Scotland 1975-1990. *Public Health*
859 **107**, 193-198 (1993).
- 860 14. World Health Organization. Surveillance of *Salmonella paratyphi A* and *B*. *Wkly*
861 *Epidemiol. Rec.* **53**, 257 (1978). Available online at:
862 [https://iris.who.int/bitstream/handle/10665/222048/WER5334_257-](https://iris.who.int/bitstream/handle/10665/222048/WER5334_257-257.PDF?sequence=1&isAllowed=y)
863 [257.PDF?sequence=1&isAllowed=y](https://iris.who.int/bitstream/handle/10665/222048/WER5334_257-257.PDF?sequence=1&isAllowed=y) (accessed March 25, 2024).
- 864 15. Felix, A. & Callow, B.R. Typing of paratyphoid B bacilli by Vi bacteriophage. *Br.*
865 *Med. J.* **2**, 127-130 (1943).

- 866 16. Felix, A. & Callow, B.R. Paratyphoid-B Vi-phage typing. *Lancet* **2**, 10-14 (1951).
- 867 17. Nicolle P. Rapport sur la distribution des lysotypes de *Salmonella typhi* et de *S.*
- 868 *paratyphi B* dans le monde, d'après les résultats fournis par les centres nationaux
- 869 membres du comité international de la lysotypie entérique à l'occasion du congrès
- 870 international de microbiologie, Stockholm, 1958 Deuxième partie. *Ann. Inst. Pasteur*
- 871 **102**, 580-595 (1962).
- 872 18. International Federation for Enteric Phage-Typing (IFEPT). The geographical
- 873 distribution of *Salmonella typhi* and *Salmonella paratyphi A* and *B* phage types during
- 874 the period 1 January 1970 to 31 December 1973. *J. Hyg.* **88**, 231-254 (1982).
- 875 19. Le Minor, L., Veron, M. & and Popoff, M. Proposition pour une nomenclature des
- 876 *Salmonella*. *Ann. Microbiol.* **133B**, 245-254 (1982).
- 877 20. Selander, R.K. *et al.* Genetic population structure, clonal phylogeny, and
- 878 pathogenicity of *Salmonella paratyphi B*. *Infect. Immun.* **58**, 1891-1901 (1990).
- 879 21. Achtman M. *et al.* Multilocus sequence typing as a replacement for serotyping in
- 880 *Salmonella enterica*. *PLoS Pathog.* **8**, e1002776 (2012).
- 881 22. Malorny, B., Bunge, C. & Helmuth, R. Discrimination of d-tartrate-fermenting and -
- 882 nonfermenting *Salmonella enterica* subsp. *enterica* isolates by genotypic and
- 883 phenotypic methods. *J. Clin. Microbiol.* **41**, 4292-4297 (2003).
- 884 23. GBD 2017 Typhoid and Paratyphoid Collaborators. The global burden of typhoid and
- 885 paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study
- 886 2017. *Lancet Infect. Dis.* **19**, 369-381 (2019).
- 887 24. Carey, M.E. *et al.* Global diversity and antimicrobial resistance of typhoid fever
- 888 pathogens: Insights from a meta-analysis of 13,000 *Salmonella Typhi* genomes. *Elife*
- 889 **12**, e85867 (2023).
- 890 25. Zhou, Z. *et al.* Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi
- 891 A during 450 years of global spread of enteric fever. *Proc. Natl Acad. Sci. USA* **111**,
- 892 12199–12204 (2014).
- 893 26. Zhou, Z. *et al.* Pan-genome analysis of ancient and modern *Salmonella enterica*
- 894 demonstrates genomic stability of the invasive Para C lineage for millennia. *Curr.*
- 895 *Biol.* **28**, 2420-2428.e10 (2018).
- 896 27. Connor, T.R. *et al.* What's in a name? Species-wide whole-genome sequencing
- 897 resolves invasive and noninvasive lineages of *Salmonella enterica* serotype Paratyphi
- 898 B. *mBio* **7**, e00527-16 (2016).
- 899 28. Chattaway, M.A. *et al.* Phylogenomics and antimicrobial resistance of *Salmonella*
- 900 Typhi and Paratyphi A, B and C in England, 2016-2019. *Microb. Genom.* **7**, 000633
- 901 (2021).
- 902 29. Chattaway, M.A. *et al.* Genomic sentinel surveillance: *Salmonella* Paratyphi B
- 903 outbreak in travellers coinciding with a mass gathering in Iraq. *Microb. Genom.* **9**,
- 904 mgen000940 (2023).
- 905 30. Weill, F.X., Fabre, L., Grandry, B., Grimont, P.A. & Casin, I. Multiple-antibiotic
- 906 resistance in *Salmonella enterica* serotype Paratyphi B isolates collected in France
- 907 between 2000 and 2003 is due mainly to strains harboring *Salmonella* genomic islands
- 908 1, 1-B, and 1-C. *Antimicrob. Agents Chemother.* **49**, 2793-2801 (2005).
- 909 31. Balandraud, A. *et al.* Sepsis caused by *Salmonella* Paratyphi B producing an OXA-48
- 910 carbapenemase in a traveller. *J. Glob. Antimicrob. Resist.* **26**, 219-221 (2021).
- 911 32. Melnotte, P. Les affections typhoïdes au Maroc. *Bull. Soc. Path. Ex.* **25**, 447-460
- 912 (1932).
- 913 33. Nicolle, P. & Hamon, Y. Distribution des lysotypes du bacille typhique et du bacille
- 914 paratyphique B en France, dans les territoires d'Outre-Mer et dans quelques autres
- 915 pays. *Rev. Hyg. Med. Soc.* **2**, 424-463 (1954).

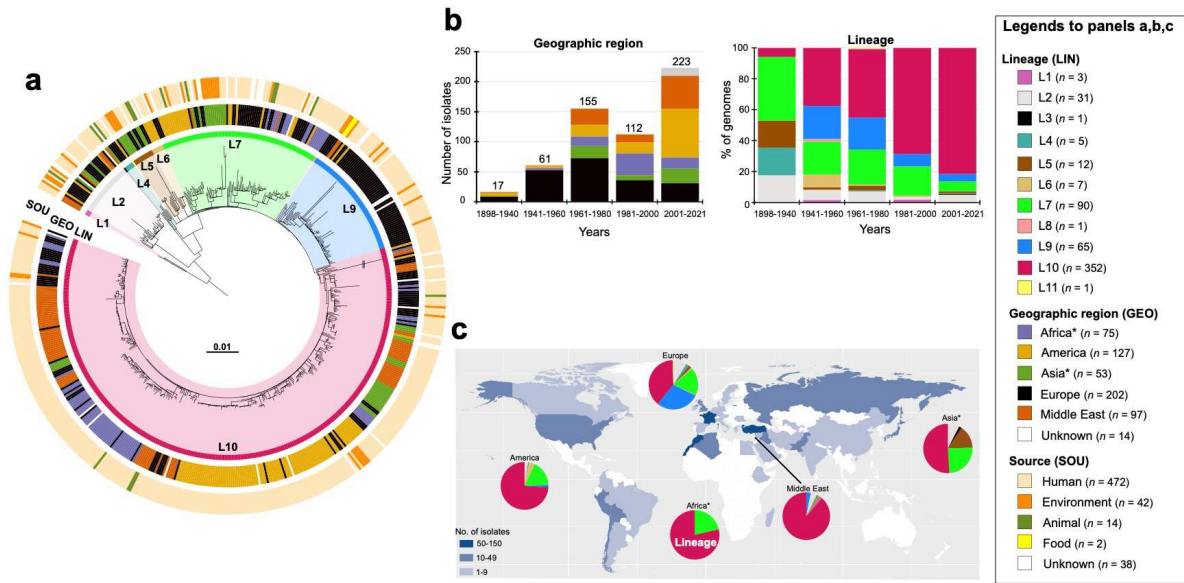
- 916 34. Ministerio de Salud Argentina. Alerta epidemiologica. Situación de fiebre paratifoidea
917 en la provincia de Salta. Semana epidemiológica 51/2022 (2022). Available on line at:
918 [https://bancos.salud.gob.ar/sites/default/files/2021-11/Alerta-fiebre-paratifoidea-](https://bancos.salud.gob.ar/sites/default/files/2021-11/Alerta-fiebre-paratifoidea-VFF.pdf)
919 [VFF.pdf](https://bancos.salud.gob.ar/sites/default/files/2021-11/Alerta-fiebre-paratifoidea-VFF.pdf) (accessed March 25, 2024).
- 920 35. Roumagnac, P. *et al.* Evolutionary history of *Salmonella* Typhi. *Science* **314**, 1301-
921 1304 (2006).
- 922 36. Wong, V.K. *et al.* Phylogeographical analysis of the dominant multidrug-resistant
923 H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission
924 events. *Nat. Genet.* **47**, 632-639 (2015).
- 925 37. Wong, V.K. *et al.* An extended genotyping framework for *Salmonella enterica* serovar
926 Typhi, the cause of human typhoid. *Nat. Commun.* **7**, 12827 (2016).
- 927 38. Tanmoy, A.M. *et al.* Paratype: a genotyping tool for *Salmonella* Paratyphi A reveals
928 its global genomic diversity. *Nat. Commun.* **13**, 7912 (2022).
- 929 39. Wang, Y. *et al.* The temporal dynamics of antimicrobial-resistant *Salmonella enterica*
930 and predominant serovars in China. *Natl Sci. Rev.* **10**, nwac269 (2022).
- 931 40. Baker, K.S. *et al.* The Murray collection of pre-antibiotic era Enterobacteriaceae: a
932 unique research resource. *Genome Med.* **7**, 97 (2015).
- 933 41. Higginson, E.E. *et al.* Improving our understanding of *Salmonella enterica* serovar
934 Paratyphi B through the engineering and testing of a live attenuated vaccine strain.
935 *mSphere* **3**, e00474-18 (2018).
- 936 42. CA-SFM., EUCAST. Comité de l'Antibiogramme de la Société Française de
937 Microbiologie Recommandations 2018. Available on line at: [https://www.sfm-](https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/CASFMV2_SEPTEMBRE2018.pdf)
938 [microbiologie.org/wp-content/uploads/2019/02/CASFMV2_SEPTEMBRE2018.pdf](https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/CASFMV2_SEPTEMBRE2018.pdf)
939 (accessed March 25, 2024).
- 940 43. CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 30th edn.
941 Supplement M100 (Clinical and Laboratory Standards Institute, 2020).
- 942 44. Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2.
943 *Genome Biol.* **20**, 257 (2019).
- 944 45. Chin, C.S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read
945 SMRT sequencing data. *Nat. Methods* **10**, 563-569 (2013).
- 946 46. Walker, B.J. *et al.* Pilon: an integrated tool for comprehensive microbial variant
947 detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
- 948 47. Wick, R.R., Judd, L.M. & Holt, K.E. Performance of neural network basecalling tools
949 for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
- 950 48. Wick, R.R. *et al.* Tricycler: consensus long-read assemblies for bacterial genomes.
951 *Genome Biol.* **22**, 266 (2021).
- 952 49. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P.A. Assembly of long, error-prone
953 reads using repeat graphs. *Nat. Biotechnol.* **37**, 540-546 (2019).
- 954 50. Vaser, R. & Šikić, M. Time- and memory-efficient genome assembly with Raven.
955 *Nat. Comput Sci.* **1**, 332–336 (2021).
- 956 51. Wick, R.R. & Holt, K.E. Benchmarking of long-read assemblers for prokaryote whole
957 genome sequencing. *F1000Res.* **8**, 2138 (2021).
- 958 52. Schwengers, O. *et al.* Bakta: rapid and standardized annotation of bacterial genomes
959 via alignment-free sequence identification. *Microb. Genom.* **7**, 000685 (2021).
- 960 53. Clausen, P.T.L.C., Aarestrup, F.M. & Lund, O. Rapid and precise alignment of raw
961 reads against redundant databases with KMA. *BMC Bioinformatics* **19**, 307 (2018).
- 962 54. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**,
963 421 (2009).
- 964 55. Zhou, Z., Alikhan, N.F., Mohamed, K., Fan, Y.; Agama Study Group & Achtman, M.
965 The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia*

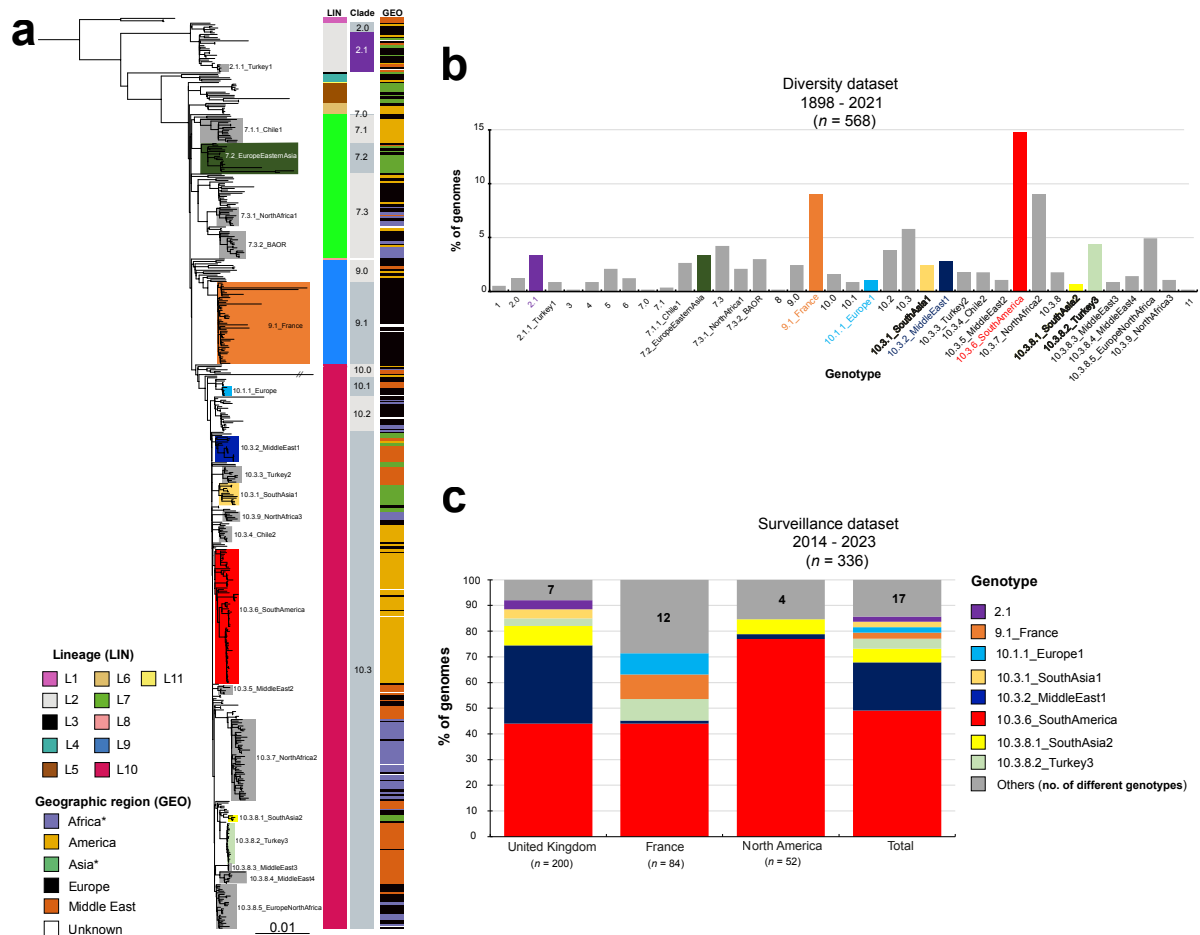
- 966 *pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* **30**, 138-152
967 (2020).
- 968 56. Zhou, Z., Charlesworth, J. & Achtman, M. HierCC: a multi-level clustering scheme
969 for population assignments based on core genome MLST. *Bioinformatics* **37**, 3645-
970 3646 (2021).
- 971 57. Zhou, Z. *et al.* GrapeTree: visualization of core genomic relationships among 100,000
972 bacterial pathogens. *Genome Res.* **28**, 1395-1404 (2018).
- 973 58. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital
974 microbiology labs. *Genome Med.* **6**, 90 (2014).
- 975 59. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive
976 antibiotic resistance database. *Nucleic Acids Res.* **45**, D566-D573 (2017).
- 977 60. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant
978 bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
- 979 61. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a
980 fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.
981 *Bioinformatics* **35**, 4453–4455 (2019).
- 982 62. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for
983 phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
- 984 63. Rambaut, A., Lam, T.T., Max Carvalho, L. & Pybus, O.G. Exploring the temporal
985 structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus*
986 *Evol.* **2**, vew007 (2016).
- 987 64. Duchêne, S., Duchêne, D., Holmes, E.C. & Ho, S.Y. The performance of the date-
988 randomization test in phylogenetic analyses of time-structured virus data. *Mol. Biol.*
989 *Evol.* **32**, 1895-906 (2015).
- 990 65. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis.
991 *PLoS Comput. Biol.* **10**, e1003537 (2014).
- 992 66. Sukumaran, J. & Holder, M.T. DendroPy: a Python library for phylogenetic
993 computing. *Bioinformatics* **26**, 1569-1571 (2010).
- 994 67. Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. Posterior
995 summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901-904
996 (2018).
- 997 68. Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W. & Corander, J. Fast
998 hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **47**, 5539-
999 5549 (2019).
- 1000 69. Edwards, D.J., Duchene, S., Pope, B. & Holt, K.E. SNPPar: identifying convergent
1001 evolution and other homoplasies from microbial whole-genome alignments. *Microb.*
1002 *Genom.* **7**, 000694 (2021).
- 1003 70. Hunt, M. *et al.* Antibiotic resistance prediction for *Mycobacterium tuberculosis* from
1004 genome sequence data with Mykrobe. *Wellcome Open Res.* **4**, 191 (2019).
- 1005 71. Ingle, D.J., Hawkey, J., Dyson, Z.A. & Holt, K.E. Genotyphi implementation in
1006 Mykrobe - preliminary technical report. Zenodo (2022). Available online at
1007 <https://zenodo.org/record/4785179> (accessed March 25, 2024).
- 1008 72. Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: a fast and
1009 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*
1010 *Biol. Evol.* **32**, 268-274 (2015).
- 1011 73. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications
1012 to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
- 1013 74. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo
1014 pipeline. *Genome Biol.* **21**, 180 (2020).

- 1015 75. Snipen, L. & Liland, K.H. micropan: an R-package for microbial pan-genomics. *BMC*
1016 *Bioinformatics*. **16**, 79 (2015).
- 1017 76. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool.
1018 *Nucleic Acids Res.* **44**, W16–W21 (2016).
- 1019 77. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat.*
1020 *Methods* **9**, 357–359 (2012).
- 1021 78. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008
1022 (2021).
- 1023 79. Milne, I. *et al.* 2013. Using Tablet for visual exploration of second-generation
1024 sequencing data. *Briefings in Bioinformatics* **14**, 193–202 (2013).
- 1025 80. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman D.J. Basic local
1026 alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- 1027 81. Wilkins, D. gggenes: Draw Gene Arrow Maps in 'ggplot2'. R package version 0.5.0
1028 (2023). Available on line at: <https://wilcox.org/gggenes/> (accessed March 28, 2024).
- 1029 82. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New
1030 York. ISBN 978-3-319-24277-4 (2016). Available on line
1031 at: <https://ggplot2.tidyverse.org> (accessed March 28, 2024).
- 1032 83. R Core Team. R: A language and environment for statistical computing. R Foundation
1033 for Statistical Computing, Vienna, Austria. 2021. Available on line at: [https://www.R-](https://www.R-project.org/)
1034 [project.org/](https://www.R-project.org/) (accessed March 28, 2024).
- 1035 84. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R
1036 package version 0.7.2 (2023). Available online
1037 at: <https://rpkgs.datanovia.com/rstatix/> (accessed March 25, 2024).
- 1038 85. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots (2023). Available
1039 online at: <https://CRAN.R-project.org/package=ggpubr> (accessed March 29, 2024).
- 1040

1041
1042

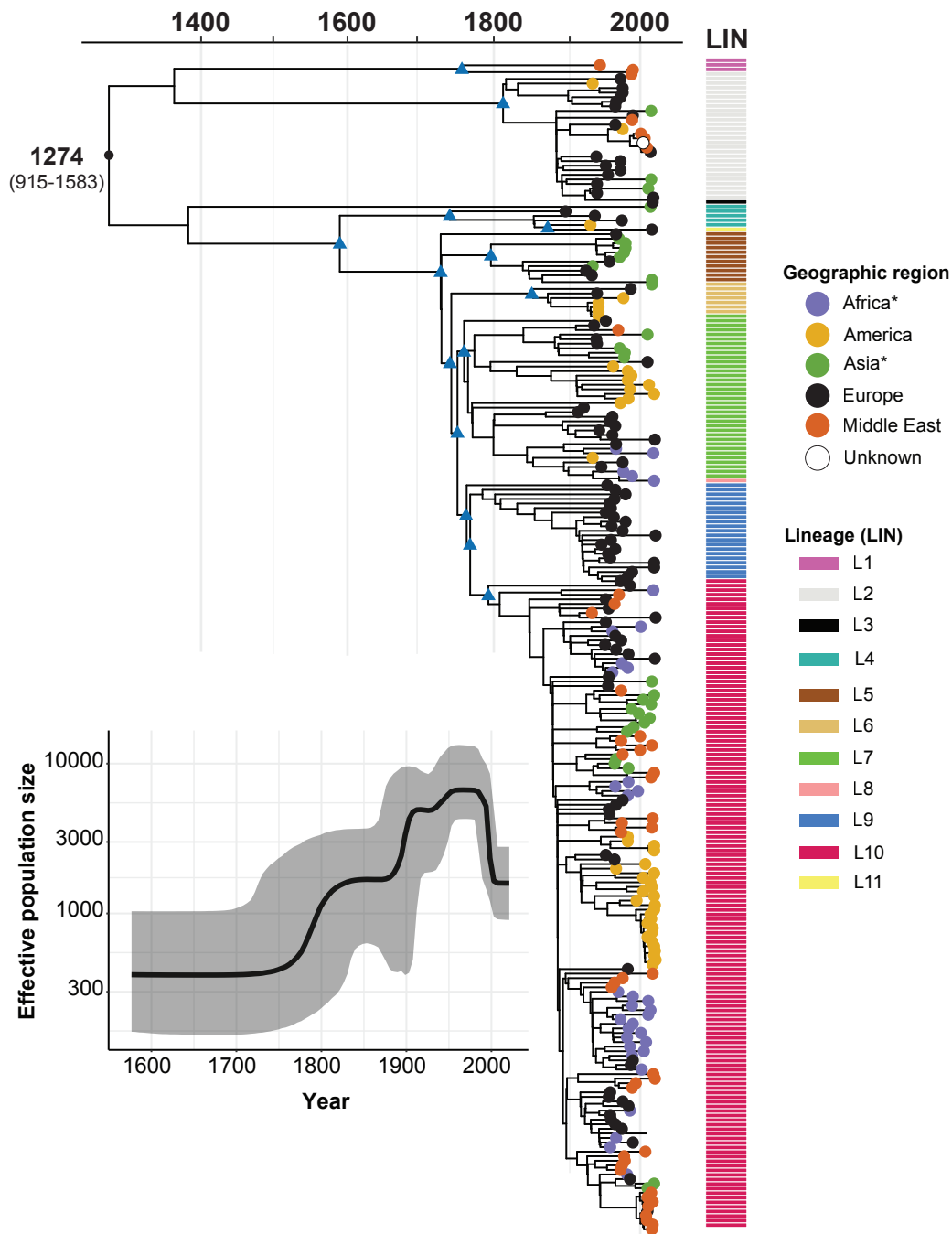
1043 **Figure 1. Phylogeny, temporal, geographic, and source distribution of the 568 SPB-**
1044 **PG1 isolates from the diversity set (1898-2021).** **a**, Circular maximum likelihood phylogeny
1045 (rooted on ancestral lineage L1 genome NCTC 8299) for the 568 SPB- PG1 isolates from the
1046 diversity dataset. The double slash (//) indicates an artificial shortening of this branch for
1047 visualisation. The rings show the associated information for each isolate, according to its
1048 position in the phylogeny, from the innermost to outermost, in the following order: (1) lineage
1049 (LIN); (2) geographic region (GEO); (3) and source (SOU). Lineages are labelled LX, where
1050 X is the lineage number. Lineages L3, L8, and L11, which contain only singletons are not
1051 labelled. The tips of the tree were highlighted according to lineage with a lighter hue of the
1052 colour used in the innermost ring (LIN). The scale bar indicates the number of substitutions
1053 per variable site (SNVs). **b**, The stacked bar chart on the left shows the distribution of the 568
1054 isolates by geographic region and time period, and the stacked bar chart on the right shows
1055 the frequencies of the lineages by the same time periods. **c**, Number of isolates per country
1056 (map) and frequencies of the lineages by world region (pie charts). An asterisk indicates that
1057 some African and Asian countries were reassigned to the Middle East (see **Table 1**). The map
1058 was drawn in R with the “ggplot2” package world map data from Wickham H (2016).
1059 ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978- 3-319-
1060 24277-4, <https://ggplot2.tidyverse.org>.
1061





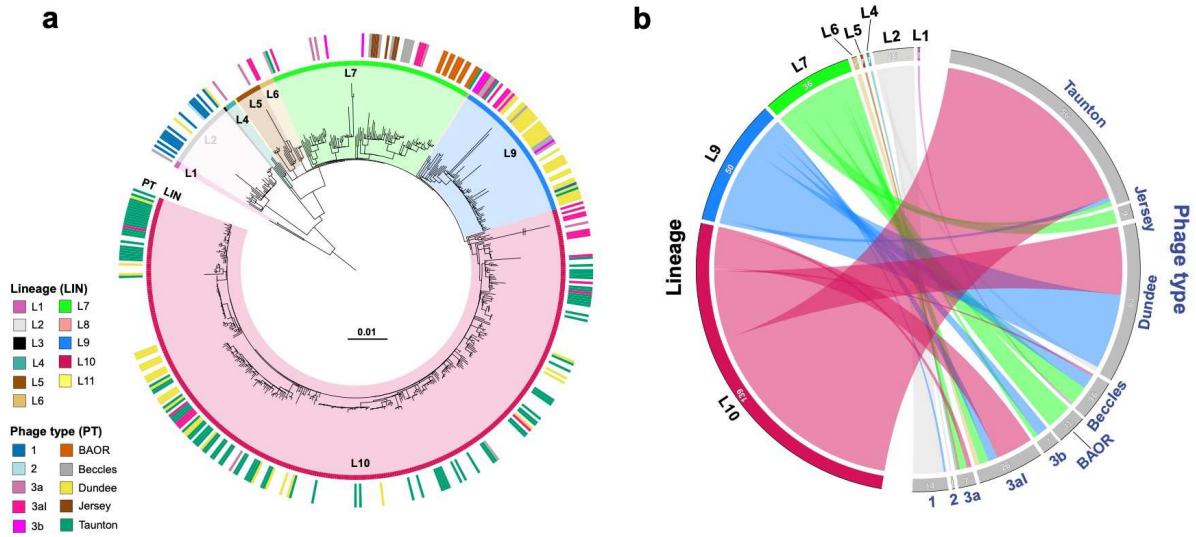
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074

Figure 2. Identification of the 38 hierarchical genotypes of SPB PG1 and their distribution between the diversity and surveillance datasets. **a**, Maximum likelihood phylogeny for the 568 SPB PG1 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 that some African and Asian countries were reassigned 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).



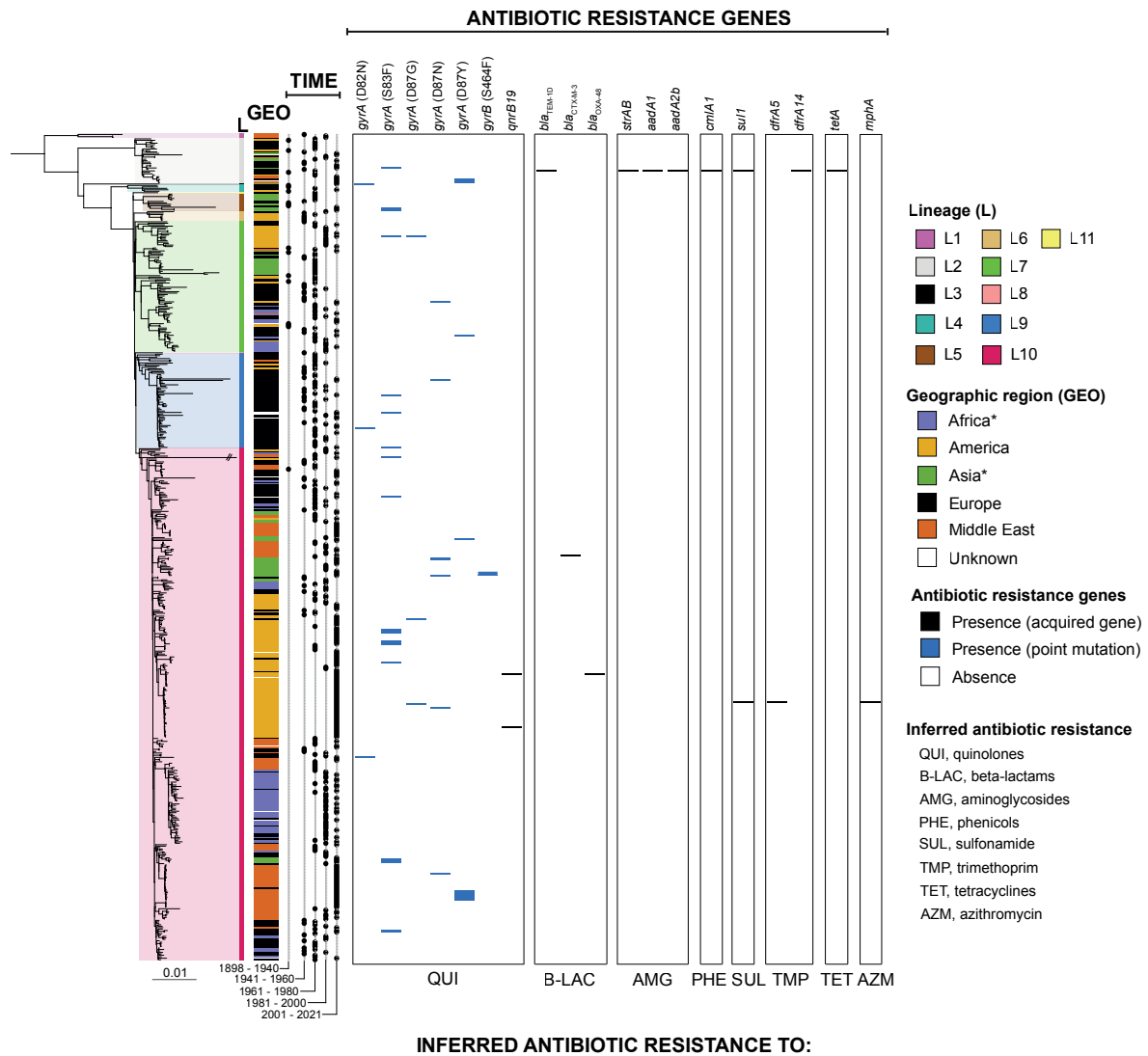
1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087

Figure 3. Timed phylogeny of a representative subsample of 256 SPB- 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 (LIN, see inset) for 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 that some African and Asian countries were reassigned to the Middle East (see Table 2). **b,** Bayesian skyline plot showing temporal changes in effective population size (black curve) with 95% confidence intervals (grey shading).



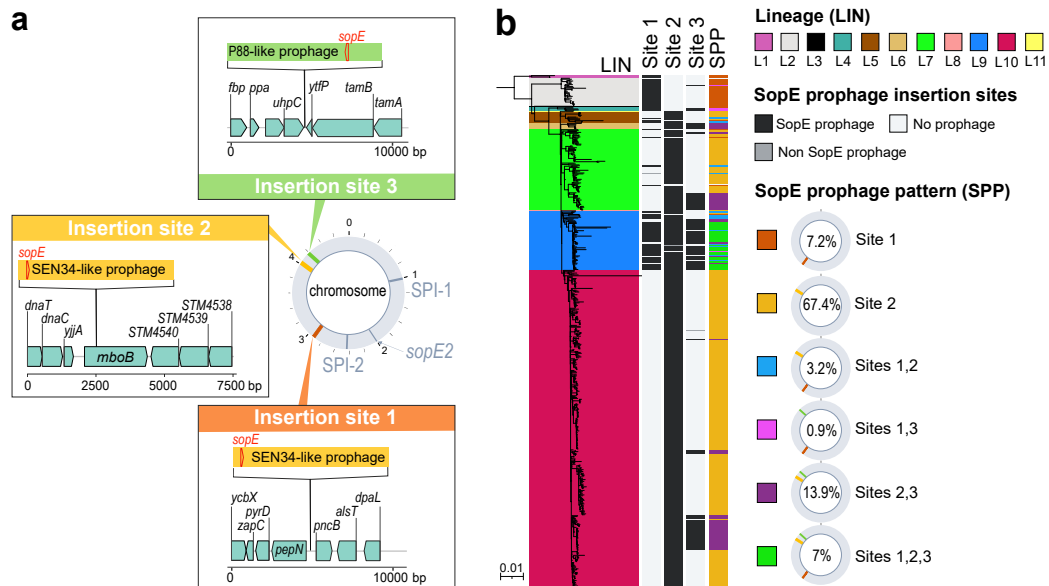
1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096

Figure 4. Correlation between genome and phage-typing data for SPB- PG1. a, Maximum likelihood phylogeny (as in Fig.1a) showing the 254 isolates from the diversity dataset that were phage-typed and their phage types (PT, see legend, inset). **b,** Circular plot illustrating the correspondence between phage type and lineage for each of these 254 isolates. The flow bars are coloured according to the lineage (see the legend in the inset of panel “a”). The number of isolates is also indicated for each phage type and lineage.



1097
1098
1099
1100
1101
1102
1103
1104

Figure 5. Genomic characterization of antibiotic resistance genes in SPB PG1. 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 that some African and Asian countries were reassigned to the Middle East (**Table 1**).



1105
1106

1107

Figure 6. Lineage-specific accumulation of *sopE* prophages in SPB⁻ PG1.

1108 **a.** Three insertion sites occupied by *sopE* prophages were identified in the 14 complete
1109 genomes of SPB⁻ PG1 isolates. Prophages of the SEN34 family [40.89-44.3 kb] were found at
1110 insertion sites #1 and/or #2. The P88-family prophage [34.3 kb] was found at insertion
1111 site #3. Further details on the insertion sites are available in **Supplementary Data 9**. Gene
1112 arrow maps were generated with the gggenes v.0.5.0 and ggplot2 v.3.4.2 packages of R
1113 v.4.1.2 software. **b.** The 568 genomes from the diversity dataset were screened for an absence
1114 of insertions at sites #1, #2 and #3. Presence/absence is colour-coded in black and light grey,
1115 respectively; dark grey indicates the presence of a potential *sopE*-free prophage. Six types of
1116 prophage insertion were recorded across the 11 lineages (**Supplementary Data 9**).

1117

1118

1119 **Table 1. Geographic distribution of the 38 genotypes found in the diversity dataset**
 1120

Genotype	n	Africa* n (%)	America n (%)	Asia** n (%)	Europe n (%)	Middle East n (%)	Unknown n (%)
1	3					3 (100%)	
2.0	7		1 (14.3%)		6 (85.7%)		
2.1	19		2 (10.5%)	4 (21.1%)	11 (57.9%)	1 (5.3%)	1 (5.3%)
2.1.1 Turkey1	5				1 (20.0%)	3 (60.0%)	1 (20.0%)
3	1			1 (100%)			
4	5		1 (20.0%)		4 (80.0%)		
5	12			8 (66.7%)	4 (33.3%)		
6	7		5 (71.4%)		2 (28.6%)		
7.0	1				1 (100%)		
7.1	2				2 (100%)		
7.1.1 Chile1	15		15 (100%)				
7.2 EuropeEasternAsia	19			13 (68.4%)	5 (26.3%)	1 (5.3%)	
7.3	24		6 (25.0%)		17 (70.8%)		1 (4.2%)
7.3.1 NorthAfrica1	12	7 (58.3%)			4 (33.3%)	1 (8.3%)	
7.3.2 BAOR	17	9 (52.9%)	1 (5.9%)		7 (41.2%)		
8	1				1 (100%)		
9.0	14		2 (14.3%)		10 (71.4%)	2 (14.3%)	
9.1 France	51				48 (94.1%)		3 (5.9%)
10.0	9	1 (11.1%)	2 (22.0%)		4 (44.4%)	2 (22.2%)	
10.1	5				2 (40.0%)	3 (60.0%)	
10.1.1 Europe1	6				4 (66.7%)	1 (16.7%)	1 (16.7%)
10.2	22	5 (22.7%)			16 (72.7%)		1 (4.5%)
10.3	33		2 (6.1%)	7 (21.2%)	15 (45.5%)	9 (27.3%)	
10.3.1 SouthAsia1	14			13 (92.9%)		1 (7.1%)	
10.3.2 MiddleEast1	16		1 (6.3%)	3 (18.8%)		12 (75.0%)	
10.3.3 Turkey2	10					10 (100%)	
10.3.4 Chile2	10		10 (100%)				
10.3.5 MiddleEast2	6					5 (83.3%)	1 (16.7%)
10.3.6 SouthAmerica	84		79 (94.0%)		3 (3.6%)		2 (2.4%)
10.3.7 NorthAfrica2	51	41 (80.4%)			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			4 (100%)			
10.3.8.2 Turkey3	25				1 (4.0%)	24 (96.0%)	
10.3.8.3 MiddleEast3	5					5 (100%)	
10.3.8.4 MiddleEast4	8					8 (100%)	
10.3.8.5 EuropeNorthAfrica	28	6 (21.4%)			20 (71.4%)	1 (3.6%)	1 (3.6%)
10.3.9 NorthAfrica3	6	5 (83.3%)			1 (16.7%)		
11	1				1 (100%)		
Total	568	75	127	53	202	97	14

1121 n, number of isolates. *the isolates from Egypt were assigned to the Middle East. **the isolates from Turkey,
 1122 Iraq, Lebanon, Syria, Mandatory Palestine, and Saudi Arabia were assigned to the Middle East. If a genotype is
 1123 found at a percentage > 50% in a particular geographic region, the data are indicated in bold.
 1124
 1125
 1126
 1127
 1128
 1129
 1130
 1131
 1132
 1133
 1134
 1135

1136
1137

Table 2. Dating of the main lineages and genotypes of SPB⁻ PG1 with BEAST2

Main lineages and genotypes	MRCA (95% HPD)
All 256 SPB ⁻ 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)

1138
1139
1140
1141
1142

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

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NMICROBIOL24051349ASOM.pdf](#)
- [NMICROBIOL24051349ASupplementaryData1.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData2.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData3.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData4.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData5.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData6.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData7.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData8.xlsx](#)
- [NMICROBIOL24051349ASupplementaryData9.xlsx](#)