



HAL
open science

Prediction of Salmonella serovars isolated from clinical and food matrices in Lebanon and genomic-based investigation focusing on Enteritidis serovar

Marie Noel Mansour, Joseph Yaghi, André El Khoury, Arnaud Felten, Michel-Yves Mistou, Ali Atoui, Nicolas Radomski

► To cite this version:

Marie Noel Mansour, Joseph Yaghi, André El Khoury, Arnaud Felten, Michel-Yves Mistou, et al.. Prediction of Salmonella serovars isolated from clinical and food matrices in Lebanon and genomic-based investigation focusing on Enteritidis serovar. International Journal of Food Microbiology, 2020, 333, pp.108831. 10.1016/j.ijfoodmicro.2020.108831 . hal-03491405

HAL Id: hal-03491405

<https://hal.science/hal-03491405>

Submitted on 25 Aug 2022

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 - NonCommercial - NoDerivatives 4.0 International License

1 **Prediction of *Salmonella* serovars isolated from clinical and food matrices in Lebanon**
2 **and genomic-based investigation focusing on Enteritidis serovar**

3 Marie Noel Mansour ^{a#}, Joseph Yaghi ^{a#}, André El Khoury ^a, Arnaud Felten ^b, Michel-Yves
4 Mistou ^c, Ali Atoui ^d, Nicolas Radomski ^{b*}

5 **Affiliation of authors**

6 ^a Centre d'Analyses et de Recherche (CAR), Unité de Recherche « Technologies et
7 Valorisation Agro-alimentaire » (UR-TVA), Faculty of Sciences, Saint-Joseph University of
8 Beirut, Campus of sciences and technologies, Mar Roukos, Lebanon.

9 ^b Paris-Est University, French Agency for Food, Environmental and Occupational Health and
10 Safety (Anses), Laboratory for Food Safety (LSAL), Maisons-Alfort, France.

11 ^c Applied Mathematics and Computer Science, from Genomes to the Environment (MaAGE),
12 National Institute for Agricultural, Food and Environmental Research (INRAE), Université Paris-
13 Saclay, Jouy-en-Josas, France.

14 ^d Laboratory of Microbiology, Department of Life and Earth Sciences, Faculty of Sciences I,
15 Lebanese University, Hadat Campus, Beirut, Lebanon.

16 # Equal contribution

17 * Corresponding author: Nicolas Radomski. Genome Analysis Modelling Risk (GAMeR),
18 Laboratory for Food Safety (LSAL), ANSES, 14 rue Pierre et Marie Curie, 94701 Maisons-
19 Alfort, France. E-mail address: nicolas.radomski@anses.fr.

20 **Authors e-mail addresses and ORCID**

21 marienoel.mansour@net.usj.edu.lb, joseph.yaghi@usj.edu.lb, andre.khoury@usj.edu.lb
22 (<https://orcid.org/0000-0003-1687-1989>), arnaud.felten@anses.fr, michel-yves.mistou@inrae.fr
23 (<https://orcid.org/0000-0001-9732-3970>), aatoui@ul.edu.lb (<https://orcid.org/0000-0002-7027-6067>), nicolas.radomski@anses.fr (<https://orcid.org/0000-0002-7480-4197>)

25

26 **ABSTRACT**

27 *Salmonella enterica* subsp. *enterica* serovars are considered major causes of food poisoning
28 and we performed this study because *Salmonella* is a burden in Lebanon. The present study
29 investigated the ability of genomic information to predict serovar using a collection of
30 *Salmonella* isolates from infected humans (n=24) and contaminated food (n=63) in Lebanon.
31 Further, the phylogenomic relationships of the serovar the predominated in Lebanon (i.e.,
32 *S. Enteritidis*; n=25) were investigated in comparison with isolates from other countries (n=130)
33 based on coregenome single nucleotide polymorphisms (SNPs). Genetic elements, specifically
34 *Salmonella* pathogenicity islands (SPI), plasmid replicons, and antibiotic-resistance genes were
35 screened in *S. Enteritidis* genomes (n=155). Our results revealed that the *Salmonella* serovars
36 identification by seroagglutination from the samples isolated in Lebanon (n=87) was highly
37 correlated with the genomic-based prediction of serovars (80.4-85.0% with SeqSero1 and 93.1-
38 94.2% with SeqSero2). The *Salmonella* serovars isolated from human and food samples in
39 Lebanon were mainly Enteritidis (28.7%) and Infantis (26%). To a rare extent, other serovars
40 included Amager, Anatum, Bredeney, Chincol, Heidelberg, Hofit, Kentucky, Montevideo,
41 Muenster, Newport, Schwarzengrund, Senftenberg and Typhimurium. In comparison with other
42 countries, *S. Enteritidis* samples isolated in Lebanon (56±27 intra-group pairwise SNP
43 differences) presented a strong phylogenomic relativeness at the coregenome level with
44 samples, as for example with samples isolated from Syria (65±31 inter-group pairwise SNP
45 differences). Most of the studied *S. Enteritidis* genomes encoded 10 SPIs involved in survival in
46 immune cells (i.e. SPIs 1, 2, 3, 4, 5, 12, 13, 14, 16 and 17). The plasmid replicons IncFIB (S)_1
47 and IncFII (S)_1 encoding elements involved in virulence were identified in the majority of the
48 *S. Enteritidis* genomes (94% and 96%, respectively), the majority exhibiting aminoglycosides
49 (gene *aac(6')-laa_1*). The IncI_1_Alpha replicon responsible for ampicillin-resistance was only
50 detected in 2 of 155 Lebanese human clinical strains. Genomic-based risk assessment of

51 *Salmonella* serovars in Lebanon showed that food imported from Syria might be an origin of
52 the *S. Enteritidis* human cases in Lebanon. The detection of several SPIs involved in the
53 survival, plasmid replicons involved in virulence, and aminoglycoside-resistance genes,
54 emphasizes that *S. Enteritidis* is of paramount importance for public health in Lebanon and
55 other countries.

56 **Keywords**

57 *Salmonella* serovars, Lebanon; *Salmonella* Enteritidis; phylogenomic relatedness; *Salmonella*
58 pathogenicity islands; plasmid replicons; aminoglycoside-resistance genes

59 **1. Introduction**

60 *Salmonella enterica* subsp. *enterica* causes infectious diseases in humans and animals
61 worldwide and is considered as a major foodborne pathogen (Kirk et al., 2015). In fact,
62 nontyphoid salmonellosis is one of the major foodborne diseases affecting the human
63 population (Acheson and Hohmann, 2001). It presents as gastroenteritis of sudden fever onset,
64 abdominal pain, diarrhea, nausea and sometimes vomiting (Acheson and Hohmann, 2001).
65 More than 2600 serovars of *S. enterica* subsp. *enterica* are described (Guibourdenche et al.,
66 2010) that can be used for source tracking. Over the past 100 to 160 million years, the
67 diversification of *Salmonella* genus from a common ancestor with *Escherichia coli* was
68 accompanied by the acquisition of genes that characterized its virulence and allowed for the
69 infection of different hosts (Meysman et al., 2013). The clusters of genes related to *Salmonella*
70 virulence have been acquired through several horizontal transfer events and are called
71 *Salmonella* pathogenicity islands (SPIs) (Rychlik et al., 2009). Twenty-three (23) SPIs have
72 been described (Hayward et al., 2014; Ilyas et al., 2017; Sabbagh et al., 2010; Shah et al.,
73 2005; Vernikos and Parkhill, 2006) and their related DNA sequences can be retrieved from
74 reference genomes (Blondel et al., 2009; Chiu, 2005; Fookes et al., 2011; Kérouanton et al.,
75 2015; McClelland et al., 2001; Parkhill et al., 2001; Thomson et al., 2008). These SPIs encode

76 genes responsible for adhesion to epithelial cells (SPI-3, 4 and 5), invasion of their *Salmonella*
77 containing vacuoles (SCV) (SPI-1 and 14), overcoming the resistance of the intestinal mucus
78 layer for proper colonization (SPI-6), induction of inflammation and neutrophil recruitment (SPI-
79 1), as well as macrophage survival (SPI-11, 12 and 16) and outer membrane remodeling (SPI-
80 2, 5 and 13) in macrophages (Fookes et al., 2011; Ilyas et al., 2017; K  rouanton et al., 2015).
81 While SPIs are common in all *S. enterica* (i.e. SPIs-1, 2, 3, 4 and 5 (Espinoza et al., 2017)),
82 others are variable between serovars (i.e. SPI-11 (Fookes et al., 2011), 16, 17 (Vernikos and
83 Parkhill, 2006), 19 (Blondel et al., 2009; Fookes et al., 2011)) or constituted of slightly different
84 nucleotide sequences (i.e. SPI-6, 9 and 10 (Thomson et al., 2008)). Other SPIs are expected to
85 be associated with *S. Typhi* (i.e. SPI-7, 8, 15 and 18 (Fookes et al., 2011)), *S. Derby* (i.e. SPI-
86 23 (Hayward et al., 2014; K  rouanton et al., 2015)), *S. arizonae* (i.e. SPI-20 and 21 (Blondel et
87 al., 2009; Fookes et al., 2011)) or *S. bongori* (i.e. SPI-22 (Fookes et al., 2011)). Concerning the
88 common SPIs, the SPI-1 and SPI-2 code for a type III secretion system (T3SS) of different
89 structure and function (Jennings et al., 2017; Lou et al., 2019). While the T3SS encoded by
90 SPI-1 is involved in non-phagocytic cell invasion and pro-inflammatory responses, that
91 encoded by SPI-2 is required for intracellular survival (Tanner and Kingsley, 2018). The SPI-3
92 is also necessary for intracellular survival and promotes bacterial growth during the systemic
93 phase of the disease as it encodes a magnesium transporter under low Mg²⁺ environments
94 (Soto et al., 2006). The SPI-4, on the other hand, is involved in epithelial cell adhesion, survival
95 in the macrophage and secretion of toxins, while the SPI-1 and SPI-5 are required for
96 enteropathogenesis because they are involved in inflammation and chloride secretion (Soto et
97 al., 2006). This SPI-5 also encodes the stimulation of SPI-1 and SPI-2 (Sabbagh et al., 2010).

98 *Salmonella* is one of the most prevalent Gram-negative pathogens and its presence is
99 correlated with the use of antibiotics (Olesen et al., 2018). The acquisition of antimicrobial
100 resistance (AMR) genes by *Salmonella* and other bacteria is due to acquisition of mobile

101 genetic elements, especially transmissible plasmids, enabling rapid adaptation to antibiotic-
102 resistance (Douarre et al., 2020). Classification of plasmids can be based on genomic-based,
103 replication-gene-typing (Carattoli et al., 2014), plasmid-mobilization (MOB) genes (Orlek et al.,
104 2017; Robertson and Nash, 2018), and by serovar-specific-virulence genes and plasmid size
105 (Rychlik et al., 2006).

106 Given the importance of *Salmonella* for public health, the purpose of this study was to: (i)
107 investigate the genomic-based prediction of serovars isolated in Lebanon, (ii) estimate the
108 phylogenomic relationships of the serovars predominating in Lebanon in comparison with other
109 countries, and (iii) screen genetic elements of interest (i.e. SPIs, plasmid replicons, AMR
110 genes).

111 **2. Materials and methods**

112 2.1. Analytical strategy

113 Overall, 87 strains of *Salmonella* were isolated from food and clinical specimens in Lebanon
114 between 2017 and 2018 (Supplementary data 1). In parallel to the next generation sequencing
115 (NGS) and *de novo* assembly, the serovars of these strains were identified by
116 seroagglutination based on Kauffmann-Le Minor scheme (Grimont, 2007) in comparison with
117 genomic-based prediction tools (i.e. SeqSero1 and SeqSero 2). Given the serovar mostly
118 isolated in Lebanon (i.e. *S. Enteritidis*), the collection of NGS data from strains isolated in
119 Lebanon was completed with NGS data from other countries based a curated version of
120 Enterobase (Alikhan et al., 2018). This *S. Enteritidis* curated version (16 928 records) of
121 Enterobase (34 714 records in January 2019) allowed standardization of complete records (i.e.
122 BioProject, ENA ID, Host, sample Matrix, serovar, source niche, source origin, source details,
123 country, and collection years). Respecting a high diversity of countries and a balanced data set
124 in terms of sources and collection years, 134 paired-end reads labelled as *Salmonella* serovar
125 *Enteritidis* were downloaded from the European Nucleotide Archive (ENA) (Harrison et al.,

126 2018). The NGS data from ENA were also *de novo* assembled to check the serovar
127 identification and used to perform a phylogenomic reconstruction of *S. Enteritidis* isolated in
128 Lebanon and other countries. The scaffolds from *S. Enteritidis de novo* assemblies were used
129 to detect genetic elements of interest including SPIs, plasmid replicons, and AMR genes.

130 2.2. Quality assessment of genomic DNA and NGS/WGS data

131 *Salmonella* strains were isolated and grown on *Salmonella*-selective media (XLD or BHI).
132 Then, the genomic DNA was extracted using the "KingFisher™ Duo Prime" protocol. The
133 quantity, purity, and integrity of the extracted DNA were assessed using a Qubit, a Nanodrop,
134 and agarose gels, respectively. NGS data were although produced by the "Institut du Cerveau
135 et de la Moelle Epinière" (ICM - Hôpital Pitié Salpêtrière, Paris). The NGS libraries were
136 prepared using the Nextera XT DNA Library Prep Kit and paired-end sequenced (2 x 150
137 bases) with an Illumina Nextseq500 sequencer. The quality assessment of NGS and WGS
138 data was supported by detecting the intra-species contamination with ConFindr (Low et al.,
139 2019), checking the assembly sizes and fragmentation with MultiQC (Ewels et al., 2016) and
140 ggplot2 (Wickham, 2009, p. 2), as well as controlling the depth and breadth of mapping
141 coverage during variant-calling analysis performed with iVarcall2 (Felten et al., 2017). The
142 non-parametric tests related to WGS data quality assessment (i.e. Kolmogorov-Smirnov (KS)
143 and Wilcoxon rank-sum (WS)) were performed with the R Stats package (R Development Core
144 Team, n.d.).

145 2.3. *de novo* assembly, scaffolding and assembly metrics

146 The in-house workflow ARTwork (Vila Nova et al., 2019) performed coverage control of reads
147 (i.e. > 20X) with Bbmap (Bushnell, 2014), read normalization (i.e. 100X) with Bbnorm (Xu et al.,
148 2015), control of read quality with FastQC (Andrews, 2018), read trimming (i.e. > 20 of Quality
149 Check) with Trimmomatic (Bolger et al., 2014), *de novo* assembly of contigs with SPAdes
150 (Bankevich et al., 2012), selection of the closely related reference genomes with MinHash

151 among an in-house collection of reference circular genomes (Ondov et al., 2016), scaffolding
152 with MeDuSa (Bosi et al., 2015), gap filling with GMcloser (Kosugi et al., 2015), trimming of
153 small scaffolds (i.e. < 200 bases) with Biopython (Cock et al., 2009), annotation of genes with
154 Prokka (Seemann, 2014) and synthesis of assembly metrics with QUAST (Gurevich et al.,
155 2013).

156 2.4. Genomic-based prediction of serovars

157 The identified *Salmonella* serovars by seroagglutination based on Kauffmann-Le Minor scheme
158 (Grimont, 2007) were compared with genomic-based predictions of serovars. To achieve this
159 goal, the genomic-based prediction tools SeqSero1 (Zhang et al., 2015) and SeqSero2 (S.
160 Zhang et al., 2019) were applied on assembled contigs and scaffolds produced by the workflow
161 ARTwork (Vila Nova et al., 2019). SeqSero1 and SeqSero2 were used to predict the
162 *Salmonella* serovars based on databases of *Salmonella* determinants (i.e. *rfb* gene clusters,
163 *fliC* and *fliB* alleles). In order to support the genomic-based prediction of serovars, the
164 sequence types (ST) were also identified from the scaffolds using the workflow mlst (version
165 2.16.1: <https://github.com/tseemann/mlst>) and the PubMLST database (Jolley and Maiden,
166 2010). The comparisons of serovars identified by seroagglutination and predicted by genomic
167 tools were performed with Venn diagrams produced by Venny
168 (<http://jvenn.toulouse.inra.fr/app/index.html>) (Bardou et al., 2014).

169 2.5. Mapping-based variant calling analysis and phylogenomic reconstruction

170 Following the Genome Analysis ToolKit (GATK) (McKenna et al., 2010), the in-house workflow
171 iVARcall2 (Felten et al., 2017; Vila Nova et al., 2019) performed trimming of reads (i.e. > 30 of
172 Quality Check) with Trimmomatic (Bolger et al., 2014), reads alignment with BWA (Li, H. and
173 Durbin, R., 2010) against the reference genome of the *S. Enteritidis* strain EC20090193
174 (CP007321-1) (Rehman et al., 2014), read sorting with Samtools (Li et al., 2009), as well as
175 duplication removal, realignment around insertion/deletions (InDels) and HaplotypeCaller-

176 based variant calling (SNPs and small InDels) with GATK (McKenna et al., 2010). The
177 pseudogenomes were produced with the script “VCFtoPseudoGenome” and correspond to the
178 reference genome where the genotypes of detected coregenome variants were replaced in
179 each genome (Felten et al., 2017). Based on these pseudogenomes, the phygenomic
180 reconstruction was performed with IQ-Tree by selecting the most appropriated model of
181 evolution based on Bayesian Information Criterion (BIC) and Akaike (AIC) (Kalyaanamoorthy et
182 al., 2017), producing consensus trees considering all possible Nearest-Neighbor-Interchanges
183 (NNIs) (Nguyen et al., 2015), and searching in the tree space from a BIONJ tree (Nguyen et al.,
184 2015) with an improved version (Hoang et al., 2018) of the ultrafast bootstrap (UFBoot) (Minh
185 et al., 2013) (1000 iterations). UFBoot convergences were checked after the IQ-Tree
186 computation (Nguyen et al., 2015). Samples of the closely related serovar Dublin (1SD0102,
187 1SD0107, SRR5583067) were used as an outgroup to root the tree of the serovars Enteritidis
188 (Felten et al., 2017). The pseudogenomes were also used as input to analyse the pairwise
189 SNP differences intra- and inter-countries based on a modified version of the R package
190 harrietr (<https://github.com/andersgs/harrietr>) (Palma et al., 2020) and ggplot2 graphical
191 representations (Wickham, 2009, p. 2).

192 2.6. Detection of genetic elements of interest (i.e. SPIs, plasmid replicons and AMR genes)
193 The SPIs, plasmid replicons, and AMR genes were detected from the assembled scaffolds with
194 Abricate (version 0.8.10: <https://github.com/tseemann/abricate>). While the SPIs were detected
195 (80% identity and 80% coverage for all related genes) based on in-house Abricate databases
196 of each SPI genes detected with Prokka (Seemann, 2014) (Table 1), the plasmid replicons and
197 AMR genes were detected following recommended settings of PlasmidFinder (Carattoli et al.,
198 2014) and ResFinder (Zankari et al., 2012), respectively. The presence of complete SPIs was
199 considered when 90% of SPI genes were detected. The presence/absence patterns of SPIs,

200 plasmid replicons, and AMR genes were presented with the phylogenomic reconstruction using
201 the version 5.5.1 of iTOL (Letunic and Bork, 2019, 2016, 2011, 2007)

202 2.7. Data availability

203 The previously developed script called iVARCall2 (Felten et al., 2017) and ARTwork (Vila Nova
204 et al., 2019) can be downloaded from the repository [https://github.com/afelten-](https://github.com/afelten-Anses/VARtools/tree/master/iVARCall2)
205 [Anses/VARtools/tree/master/iVARCall2](https://github.com/afelten-Anses/VARtools/tree/master/iVARCall2) and the corresponding open source package is now
206 available in Conda (<https://anaconda.org/itsmeludo/repo> and <https://conda.io/docs/>).
207 Sequencing data is available in the BioProject PRJEB37595 of the European Nucleotide
208 Archive (ENA) (<https://www.ebi.ac.uk/ena/browser/view/PRJEB37595>).

209 3. Results

210 3.1. Quality of genomic data

211 Out of 92 strains of *Salmonella* from Lebanon, five strains were excluded from the present
212 study because of exogenous DNA contamination and low quality of the assembly in terms of
213 scaffold number and total assembly length (data are not shown). Considering all the remaining
214 NGS data (n=221), the Confindr-based exogenous DNA contamination (Low et al., 2019)
215 ranged from 0 to 8 single nucleotide variants (SNVs) and was estimated at 0.40 ± 0.59 SNVs
216 (Supplementary data 1). Similar extremely low levels of exogenous DNA contamination were
217 observed comparing NGS data from the present study (0.25 ± 0.39 SNVs) to ENA
218 (0.50 ± 0.71 SNVs). As recently demonstrated, these levels of exogenous DNA contamination
219 lower than 10 SNVs are appropriate for phylogenomic reconstruction based on coregenome
220 SNPs (Pightling et al., 2019). The *de novo* assemblies performed in the present study present
221 a low fragmentation (from 1 to 19: 4.59 ± 3.48 scaffolds higher than 1000 bp), as well as
222 expected total assembly length (from 4.27 to 5.06 Mb: 4.76 ± 0.07 Mb) and GC% (from 51.8 to
223 52.2: 52.1 ± 0.1) (Supplementary data 1). Published *Salmonella* genome sizes range from 3.39
224 to 5.59 Mbp (Nava M. and Hernandez, 2016) with a GC% estimated at 52.12 ± 0.09 (Vila Nova

225 et al., 2019). Concerning the serovars predominating in Lebanon (25 *S. Enteritidis* among 87
226 *Salmonella*), there were no statistical differences in scaffold number (KS=0.494 and
227 WS=0.417), total length (KS=0.083 and WS=0.587) and deep coverage (KS=0.240 and
228 WS=0.715) between *S. Enteritidis* samples from the present study and ENA (n=130 excluding
229 4 genomes with ambiguous ST and serovars), emphasizing that assembly and mapping of both
230 provided comparable outcomes (Supplementary data 1 and 2). Small statistical differences in
231 breadth coverage (KS= 1.4×10^{-10} and WS= 7.3×10^{-8}) were observed between samples from the
232 present study (98.56±0.10%) and ENA (99.09±0.39%), showing that the genomes of the
233 serovar predominating in Lebanon diverged slightly from the mapped reference genome
234 (*S. Enteritidis* strain EC20090193) compared to genomes from the other considered countries
235 (Supplementary data 2). Otherwise, the total length of genomes (4.73±0.03Mb) corresponding
236 to the serovar predominating in Lebanon was quite similar to the predicted genome size from
237 the literature; namely, ranging from 4.6 to 4.8Mb due to the presence or absence of several
238 mobile elements such as phages and plasmids (Allard et al., 2013). Overall, the quality metrics
239 of the assembly (number of scaffolds: 2.67±1.54; total length: 4.73±0.03Mb) and mapping
240 (Breadth coverage: 99.01±0.42%; deep coverage: 80.05±26.12X) are suitable for downstream
241 analyses related to the serovar mostly isolated in Lebanon (Felten et al., 2017; Vila Nova et al.,
242 2019) (Supplementary data 2).

243 3.2. Genomic-based serovar prediction

244 Based on results using the Kauffmann-Le Minor scheme (Grimont, 2007), the *Salmonella*
245 strains isolated from Lebanon (n=87) were mainly represented by the serovars Enteritidis
246 (28.7%) and Infantis (26.4%). Other serotypes, namely Amager, Anatum, Bredeney, Chincol,
247 Heidelberg, Hofit, Kentucky, Montevideo, Muenster, Newport, Schwarzengrund, Senftenberg
248 and Typhimurium were also identified by at lower percentages (Supplementary data 1 and
249 Figure 1A). The serovars identified by seroagglutination of *Salmonella* samples isolated in

250 Lebanon (Grimont, 2007) were highly correlated with the genomic-based prediction of serovars
251 from SeqSero1 (Zhang et al., 2015) (80.4-85.0%) and SeqSero2 (S. Zhang et al., 2019) (93.1-
252 94.2%) applied on assembled contigs or scaffolds produced by the workflow ARTwork (Vila
253 Nova et al., 2019).

254 3.3. Coregenome phylogenomic reconstruction and deep analysis of pairwise SNP 255 distances

256 Except five divergent genomes at the root (SRR6844254, SRR7049988, SRR7408396,
257 SRR4418787, ERR1815494), the robust phylogenomic reconstruction (TVM+F+I converged at
258 102) divided *S. Enteritidis* strains (mainly ST 11: Supplementary data 1) into two main clusters.
259 A first cluster that encompasses the Lebanese strains with other ones mostly isolated in Asia
260 and Brazil (cluster 1), and a second one that comprises mainly European, North American and
261 Brazilian strains (cluster 2) (Figure 2). Considering the Lebanese strains (Supplementary data
262 3), human and avian strains had smaller intra-country pairwise SNP differences (46 ± 12 ranging
263 from 23 to 67) than human and bovine strains (96 ± 3 ranging from 88 to 101). The intra-country
264 pairwise SNP distances (Supplementary data 3) were highly variable (Figure 3A) and range
265 between 48 ± 29 for strains from Taiwan and 358 ± 225 for strain from Ireland with a maximum of
266 3 220 for strains from United-States (Table 2A). Comparing Lebanon to all the other countries
267 (Figure 3B), the Lebanese and Syria strains present the lowest inter-country pairwise SNP
268 distances ranging from 38 to 93 and estimated at 65 ± 31 (Table 3B). More precisely, Lebanese
269 strains from the *S. Enteritidis* cluster 1 present closely related pairwise SNP differences with
270 strains from Syria (SF11 vs SRR1300592: 38 SNPs), Taiwan (SF21 vs SRR7741644: 62 SNPs
271 and SF21 vs SRR7741644 vs 65 SNPs), Belgium (SF11 vs SRR1763807: 74 SNPs), United
272 Kingdom (SF21 vs SRR6900948: 78 SNPs) and Ireland (SH4 vs ERR1815489, ERR1815490
273 and ERR1815488: 80 SNPs).

274 3.4. Distributions of SPIs, plasmid replicons and AMR genes
275 SPIs (Supplementary data 4), plasmid replicons (Supplementary data 5), and AMR genes
276 (Supplementary data 6) were detected in *S. Enteritidis* genomes. Most of the *S. Enteritidis*
277 genomes harbored 10 SPIs: SPI-1, SPI-2, SPI-3, SPI-4, SPI-5, SPI-12, SPI-13, SPI-14, SPI-16
278 and SPI-17 (Supplementary data 4). The majority of the genomes (94%) harbored 2 to 3
279 plasmid replicons such as IncFIB(S)_1 and IncFII(S)_1 (Supplementary data 5). In addition,
280 two Lebanese human genomes (SH17 and SH25) harbored a plasmid replicon IncII_1_Alpha
281 which was not detected in other Lebanese *S. Enteritidis* genomes (Supplementary data 5).
282 While a strain isolated from food in Lebanon (SF20) was devoid of plasmid replicons, others
283 harbored plasmid replicons such as Col(Ye4449)_1, ColRNAI_1, and IncX1_1 and others
284 (Supplementary data 5). Elsewhere, AMR genes were detected for six antibiotic-resistance
285 gene families: aminoglycoside (*aac(3)-Vla_1*, *aac(6')-Iaa_1*, *ant(3)-Ia_1*, *aph(3)-Ib_5*, *aph(6)-*
286 *Id_1*), beta-lactamase (*blaTEM-116_1*), phenicol (*cat(pC194)_1*), trimethoprim (*dfrA1_10*,
287 *dfrA25_1*), quinolone (*qnrB2_1*) and sulphonamide (*sul1_5*, *sul2_2*). Almost all studied
288 *S. Enteritidis* genomes harbored a single AMR gene (Supplementary data 6). The most
289 common AMR gene detected was *aac(6')-Iaa_1* responsible for aminoglycoside-resistance.
290 Two Lebanese strains recovered from human infections (SH17 and SH25) carried an additional
291 plasmid replicon (IncII_1_Alpha), harbor also an additional AMR gene (*blaTEM-1B_1*). Five
292 human clinical and avian strains isolated from other countries (SRR5160664, SRR7741600,
293 SRR7741613, SRR7741630; SRR1258572) exhibited resistance to more than three antibiotic
294 families (Supplementary data 6) and are consequently considered as multiple drug resistance
295 (MDR) strains (Magiorakos et al., 2012).

296 **4. Discussion**

297 4.1. Avian origin of *S. Enteritidis* isolated from human in Lebanon

298 *S. Enteritidis* is the most prevalent serovar in Lebanon as it had already been reported by the
299 Lebanese laboratories participating in PulseNet, the network for the monitoring of foodborne
300 diseases (Fadlallah et al., 2019). This observation was confirmed in our study since
301 *S. Enteritidis* was the major serovar identified and was prevalent in 25 samples (28.7%) among
302 87 ones isolated in Lebanon between 2017 and 2018 (Figure 1A). The estimated correlations
303 between seroagglutination and genomic-based serovar prediction (80.4-85.0% for SeqSero1
304 (Zhang et al., 2015) and 93.1-94.2% for SeqSero2 (S. Zhang et al., 2019) (Figure 1B) were in
305 agreement with other authors estimating that Seqsero1 and Seqsero2 are correlated at 95%
306 and 98%, respectively (Diep et al., 2019). In agreement with our observations (Figure 2) and as
307 previously observed in comparison with the clonal serovars phylogenetically associated with
308 cattle (Dublin) and avian (Pullorum and Gallinarum), the serovar *Enteritidis* is considered as a
309 multi-host (Felten et al., 2017; Langridge et al., 2015) and presents a panmictic population
310 regarding its isolated animal source (Sheppard et al., 2018). Supporting the theory that humans
311 (Retamal et al., 2015) and poultry (Barrow, 1991) are especially susceptible to *S. Enteritidis*,
312 we observed that genetic differences between human and avian strains isolated in Lebanon
313 (Supplementary data 3) (46 ± 12 SNPs) are smaller than those between human and bovine
314 strains (96 ± 3 SNPs). Indeed, *S. Enteritidis* is known for its strong association with poultry and
315 its ability to infect chicken ovaries asymptotically (Y. Zhang et al., 2019). Based on our
316 results (Figure 2 and Supplementary data 3), the strong phylogenetic relatedness between
317 human and avian *S. Enteritidis* strains shows that salmonellosis in Lebanon is likely due to
318 poultry sources. This has been demonstrated for 20 years, since the first outbreak of
319 *S. Enteritidis* in chicken meat in Lebanon (Barbour et al., 1999). And indeed, the disease
320 caused by the serovar *Enteritidis* is one of the 25 major poultry diseases in Lebanon such as

321 colibacillosis, infectious bursal disease, necrosis of the femoral head, malabsorption, and the
322 syndrome of egg drop (Barbour et al., 1997).

323 4.2. Geographical origin of *S. Enteritidis* isolated in Lebanon

324 This phylogenomic study of *S. Enteritidis* strains from different geographic sources shows that
325 Lebanese strains are closely related with strains from Syria (38 SNPs), Taiwan (62-65 SNPs),
326 Belgium (74 SNPs), United Kingdom (78 SNPs) and Ireland (80 SNPs) (Figure 3B and
327 Supplementary data 3). This result is consistent with the fact that the meat sector in Lebanon
328 contains products that originate from various sources (e.g. Australia, Brazil, Paraguay, France,
329 New Zealand, China, Canada, Uruguay, Argentina, United States, Poland, Denmark, Porto
330 Rico, Netherlands, South Africa) (Darwish, 2013). The Syrian strain from plants presenting the
331 most closely related coregenome to Lebanese avian strains could highlight inter-country
332 contamination, which remains hypothetical at this stage of the investigation. It is estimated that
333 Lebanon imports annually about 25,000 tons of chicken breast meat from Syria and that in turn
334 imports meat, thighs, and chicken drumsticks from Lebanon (Darwish, 2013). Nevertheless, the
335 limited amount of *S. Enteritidis* genomes from Lebanon (n=25) and other 14 countries (n=130)
336 compared in the present study does not exclude the possibility that genomes from non-
337 sequenced samples and/or other countries may be genetically more related to those isolated in
338 Lebanon. The present study that is performed with limited computing and storage capacities,
339 will contribute to the international systems of foodborne pathogens' control like Pathogen
340 Detection from NCBI integrating today genomic data for 286 015 *Salmonella* isolates.

341 4.3. SPIs of *S. Enteritidis*

342 To our knowledge, the present study proposes for the first time a DNA sequence database of
343 the 23 known SPIs (Table 1). The studied *S. Enteritidis* genomes harbor 10 SPIs among the 23
344 known (i.e. SPIs 1, 2, 3, 4, 5, 12, 13, 14, 16 and 17). As expected, the studied *S. Enteritidis*
345 genomes do not harbor the SPIs supposed to be associated with *S. Typhi* (i.e. SPI-7, 8, 15 and

346 18 (Fookes et al., 2011)), *S. Derby* (i.e. SPI-23 (Hayward et al., 2014; K  rouanton et al.,
347 2015)), *S. arizonae* (i.e. SPI-20 and 21 (Blondel et al., 2009; Fookes et al., 2011)) or *S. bongori*
348 (i.e. SPI-22 (Fookes et al., 2011)). In addition, the studied *S. Enteritidis* were not shown to
349 harbor SPIs known to be variable between *S. enterica* serovars (i.e. SPI-11 and 19 (Blondel et
350 al., 2009; Fookes et al., 2011) or constituted of different nucleotide forms (i.e. SPI-6, 9 and 10
351 (Fookes et al., 2011; Thomson et al., 2008)). Among the SPIs detected in the studied
352 *S. Enteritidis*, the first 5 SPIs (SPI-1 to SPI-5) were expected because they are common in all
353 *Salmonella enterica* (Espinoza et al., 2017). Among them, the SPI-2 is considered to be the
354 most important for virulence because it possesses all necessary genes for the survival of the
355 bacterium in macrophages (Ilyas et al., 2017). The detected SPI-12, SPI-13, SPI-14, SPI-16,
356 and SPI-17 have a role during infection in various serovars. For instance, the SPI-12 has been
357 detected in various serovars such as Choleraesuis, Typhimurium (Ilyas et al., 2017), Typhi
358 (Deng et al., 2003), Enteritidis, Gallinarum, and Choleraesuis (Fookes et al., 2011). The SPI-12
359 encodes a remnant phage including genes involved in the survival of *Salmonella* in
360 macrophages and is controlled by the SsrB regulator through the *sspH2* promoter
361 (Tomljenovic-Berube et al., 2013). This *sspH2* virulence factor secretes effector proteins of
362 T2SS increasing the rate of actin polymerization in infected cells (Satyanarayana et al., 2019).
363 Among these effector proteins of the SPI-12, the *oafA* gene encodes an acetylase of the LPS
364 O antigen that confers beneficial antigenic variation for pathogenesis and immune evasion
365 (Ilyas et al., 2017). A previous study demonstrated by a screening of *S. Gallinarum* mutants
366 that the detected SPI-13 and SPI-14 promoted the colonization of chicken spleen (Shah et al.,
367 2005). Besides, the SPI-13 in *S. Enteritidis* genomes impaired infection of murine
368 macrophages, without modified infection of avian macrophages (Elder et al., 2016). The *in vitro*
369 specificity of these two SPI-13 and SPI-14 during avian infection would explain their detection
370 in the serovar *S. Enteritidis* that is considered to have an affinity for avian diseases. Otherwise,

371 the detected SPI-16 and SPI-17 are supposed to be present in all *Salmonella* serovars except
372 *S. bongori* and *S. arizonae*, as well as *S. bongori*, *S. arizonae* and *S. Typhimurium*,
373 respectively (Vernikos and Parkhill, 2006). The SPI-16 and SPI-17 identified in the studied
374 *S. Enteritidis* genomes, appear to belong to the same phage family, similar to bacteriophage
375 P22, and are involved in seroconversion because of the presence of the *gtrA* and *gtrB* genes
376 that mediate the glycosylation of the O antigen and the variation of the cell surface (Sabbagh et
377 al., 2010).

378 4.4. Plasmid replicons of *S. Enteritidis*

379 IncFIB(S)_1 and IncFII(S)_1 are the plasmids replicons predominantly detected in the studied
380 *S. Enteritidis* genomes. They belong to the family of IncF which is widely distributed and
381 restricted to the Enterobacteriaceae family, in particular *Salmonella enterica*, *Shigella* and
382 *Escherichia coli* (Villa et al., 2010). These IncF plasmid replicons contributed to the bacterial
383 ability by providing determinants of virulence including AMR, bacteriocins, siderophores,
384 cytotoxins and adhesion factors (Villa et al., 2010). For instance, IncF plasmids have been
385 associated with the sudden worldwide emergence of clinically relevant extended spectrum
386 beta-lactamases (ESBLs), such as the *CTX-M-15* gene, but also with the spread of *ampC*
387 genes (*blaCMY* and *blaDHA*) and the resistance to quinolone and aminoglycosides encoded by
388 the *aac(6')-Ib-cr*, *qnr*, *qepA*, *armA* and *rmtB* genes (Carattoli et al., 2014; Villa et al., 2010). An
389 aminoglycoside resistance that is encoded by the *aac(6')-Iaa_1* gene was detected in all the
390 *S. Enteritidis* genomes, which differs from the resistance genes generally present on IncF
391 plasmid replicons (Villa et al., 2010). The lack of plasmid replicons in some *S. Enteritidis*
392 genomes from aminoglycoside-resistant strains, could be due to the fact that the *aac(6')-Iaa_1*
393 gene can be carried by the chromosome. Furthermore, the plasmid replicons IncI_1_Alpha
394 detected in two strains isolated from Lebanese human (SH17 and SH25), are probably
395 responsible for the resistance to ampicillin, because these are the only strains among the

396 Lebanese ones carrying the ampicillin resistance gene *BlaTEM-1B_1*. As previously described,
397 plasmids belonging to the incompatibility group I1 (Incl1) detected in *Escherichia coli* and
398 *Salmonella* carry ampicillin resistance genes (Carattoli, 2009; García-Fernández et al., 2008).
399 On the other hand, the observation of ampicillin-resistant strains without Incl plasmid replicons
400 implies that this resistance can be carried by other plasmid families such as small-sized Col
401 plasmids and IncX plasmids which encode resistance genes *blaTEM* (Ahmed et al., 2017;
402 Johnson et al., 2012).

403 4.5. AMR genes of *S. Enteritidis*

404 *S. Enteritidis* has become increasingly resistant to multiple antibiotics by accumulating
405 resistance genes to different types of antibiotics. In the present study focusing on the
406 *Salmonella* serovar mostly isolated in Lebanon (i.e. Enteritidis), we observed a high frequency
407 of aminoglycoside resistance coded by the *aac(6')-Iaa* gene comprising the enzyme aac(6')
408 that inactivates, by acetylation, tobramycin, kanamycin and amikacin (Salipante, 2003). Several
409 studies have been conducted on the AMR of *S. Enteritidis* from different countries. Concerning
410 *S. Enteritidis* isolated from eggs in India, resistance to ampicillin, kanamycin, nalidixic acid,
411 neomycin, polymixin, and tetracyclin had been observed (Suresh et al., 2006). Two studies
412 were carried out with *S. Enteritidis* isolated in Italy (2004): the first with strains isolated from
413 humans, animals and food showed resistance to treptomycin, sulphonamide, nalidixic acid and
414 tetracycline (Busani et al., 2004), and the second with strains isolated from poultry, showed
415 resistance to tetracycline, ampicillin, trimethoprim-sulfamethoxazole, gentamicin and
416 kanamycin (Pezzella et al., 2004). On the other hand, a study carried out with *S. Enteritidis*
417 isolated from poultry, eggs and humans in Turkey, showed no resistance to antibiotics from the
418 family of aminoglycoside, quinolone and carbapenem, while resistance was evident from all
419 isolated strains to Class II cephem, tetracycline, and beta-lactamase (İçgen, 2002). The low
420 frequency of aminoglycoside-resistance of *S. Enteritidis* strains, suggests that the recent

421 increase in aminoglycoside use in human medicine has not yet had an impact. The *aac(6')-Iaa*
422 gene codes for resistance to different antibiotics (tobramycin, kanamycin, amikacin), but this
423 does not allow us to consider the Lebanese strains as MDR because the latter possess a
424 resistance to at least three antibiotic families (Magiorakos et al., 2012). *S. Enteritidis* strains
425 isolated in Lebanon, therefore do not show resistance to the majority of the dedicated
426 antibiotics for this serovar and their resistance involves only certain antibiotics of the
427 aminoglycoside family, as well as resistance to ampicillin. Finally, few *S. Enteritidis* genomes
428 from other countries are MDR with AMR genes corresponding to different antibiotic families
429 (aminoglycoside, beta-lactamase, phenicol, trimethoprim, quinolone and/or sulphonamide).

430 **5. Conclusions**

431 Foodborne bacterial diseases are a growing challenge for human health worldwide. Among
432 them, salmonellosis can cause high morbidity in the developed and industrialized countries.
433 The causative agent *Salmonella* can induce simple infections such as gastroenteritis or
434 invasive infections such as bacteremia and systemic disease. *Salmonella enterica* subsp.
435 *enterica* serovar Enteritidis, the main serovar isolated in Lebanon between 2017 and 2018, has
436 become the second leading cause of *Salmonella* infection in humans, after the serovar
437 Typhimurium. *S. Enteritidis* has a wide range of hosts but seems strongly related to poultry and
438 egg products. In the present study, we observed that the classical seroagglutination procedure
439 for identifying *Salmonella* serovars was highly correlated with genomic-based predictions. The
440 phylogenomic study of *S. Enteritidis* performed in the present study showed a strong similarity
441 between Lebanese strains and strains from several countries, especially Syria, which may
442 potentially highlight inter-countries contamination of imported food. The observation of close
443 phylogenomic relationships between human and avian strains in Lebanon showed that the
444 major cause of infection by this serovar is probably the avian products. *S. Enteritidis* genomes
445 harbor 10 SPIs: SPI-1, SPI-2, SPI-3, SPI-4 SPI-5, SPI-12, SPI-13, SPI -14, SPI-16 and SPI-17,

446 involved in colonization and infection of host cells. Concerning the plasmid replicons, the
447 predominantly detected replicons (IncFIB(S)_1 and IncFII(S)_1) affect the virulence of this
448 serovar but are not involved in the aminoglycosides-antibiotic resistance because of the
449 absence of the gene *aac (6')-laa_1* that appears to be localized on the chromosome. On the
450 other hand, we found that the resistance of strains (SH17 and SH25) to ampicillin is probably
451 harbored by the plasmid replicon IncI_1_Alpha. Finally, the studied *S. Enteritidis* showed
452 resistance to antibiotics belonging to the family of aminoglycosides, especially tobramycin,
453 amikacin and kanamycin, with resistance to ampicillin detected in few samples. Most of the
454 studied *S. Enteritidis* are not MDR because most of the samples showed resistance to a single
455 family of antibiotics. In summary, *S. Enteritidis* is a serovar possessing several virulence
456 factors, as well as AMR genes, that improve its survival in the host organism. To reduce the
457 infection rate caused by *S. Enteritidis*, it is necessary to analyze imported products to
458 determine their microbiological qualities, while carefully handling antibiotic usage in a way that
459 prevents excess application, especially in the livestock sector to avoid the emergence of new
460 AMR.

461 **Declaration of competing interest**

462 The authors declare that they have no conflicts of interest.

463 **Acknowledgements**

464 This study was funded and supported by the research council of Saint-Joseph University of
465 Beirut (USJ) (project FS149). The sequencing and genomic analyses were supported by the
466 project “Collaborative Management Platform for detection and Analyses of (Re-) emerging and
467 foodborne outbreaks in Europe” (COMPARE) in the framework of the European Union's
468 Horizon 2020 research and innovation programme (grant agreement No. 643476). The present
469 study represents a first stone for the ongoing European Joint Programme (EJP) “Assessing
470 Determinants of the Non-Decreasing Incidence of *Salmonella*” (ADONIS) (grant agreement No.

77380). We thank Pierre-Yves Letournel and Thomas Texier (Anses) for providing high-performance computing resources. We thank also Joseph Oliver Falkinham III and Rouaa Daou for helping to improve the grammar of the manuscript.

Author contributions

MNM, JY, AEK and AA performed the bacteriological analyses. AEK, MYM and NR designed experimental plan. MNM performed the DNA extractions and genomic analyses with the support of AF in charge of the bioinformatics infrastructure and under the analytical supervision of NR who designed the analytical strategy. All authors commented and approved the final manuscript. All of them agree to be accountable for all aspects of the work in terms of accuracy or integrity.

References

- Acheson, D., Hohmann, E.L., 2001. Nontyphoidal salmonellosis. *Clin. Infect. Dis.* 32, 263–269. <https://doi.org/10.1086/318457>
- Ahmed, S., Olsen, J.E., Herrero-Fresno, A., 2017. The genetic diversity of commensal *Escherichia coli* strains isolated from nonantimicrobial treated pigs varies according to age group. *PLoS ONE* 12, 1–18. <https://doi.org/10.1371/journal.pone.0178623>
- Alikhan, N.-F., Zhou, Z., Sergeant, M.J., Achtman, M., 2018. A genomic overview of the population structure of *Salmonella*. *PLOS Genet.* 14, e1007261. <https://doi.org/10.1371/journal.pgen.1007261>
- Allard, M.W., Luo, Y., Strain, E., Pettengill, J., Timme, R., Wang, C., Li, C., Keys, C.E., Zheng, J., Stones, R., Wilson, M.R., Musser, S.M., Brown, E.W., 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. *PLoS ONE* 8, e55254. <https://doi.org/10.1371/journal.pone.0055254>
- Andrews, S., 2018. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinforma.* <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Barbour, E.K., Hilan, S.K.H.C., Kallas, M.A.E., Sakr, W., 1997. National surveillance of poultry diseases in Lebanon 16, 770–775.
- Barbour, E.K., Jurdi, L.H., Talhouk, R., Qatanani, M., Eid, A., Sakr, W., Bouljihad, M., Spasojevic, R., 1999. Emergence of *Salmonella* Enteritidis outbreaks in broiler chickens in the Lebanon: epidemiological markers and competitive exclusion control (EN, FR, ES). *Rev. Sci. Tech. OIE* 18, 710–718. <https://doi.org/10.20506/rst.18.3.1184>
- Bardou, P., Mariette, J., Escudié, F., Djemiel, C., Klopp, C., 2014. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15, 293. <https://doi.org/10.1186/1471-2105-15-293>

- 510 Barrow, P.A., 1991. Experimental infection of chickens with *Salmonella* Enteritidis. *Avian*
511 *Pathol.* 20, 145–153. <https://doi.org/10.1080/03079459108418749>
- 512 Blondel, C.J., Jiménez, J.C., Contreras, I., Santiviago, C.A., 2009. Comparative genomic
513 analysis uncovers 3 novel loci encoding type six secretion systems differentially
514 distributed in *Salmonella* serotypes. *BMC Genomics* 10, 354.
515 <https://doi.org/10.1186/1471-2164-10-354>
- 516 Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina
517 sequence data. *Bioinforma. Oxf. Engl.* 30, 2114–2120.
518 <https://doi.org/10.1093/bioinformatics/btu170>
- 519 Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R.,
520 Fondi, M., 2015. MeDuSa: a multi-draft based scaffolder. *Bioinformatics* 31, 2443–2451.
521 <https://doi.org/10.1093/bioinformatics/btv171>
- 522 Busani, L., Graziani, C., Battisti, A., Franco, A., Ricci, A., Vio, D., Digiannatale, E., Paterlini, F.,
523 D’Incau, M., Owczarek, S., Caprioli, A., Luzzi, I., 2004. Antibiotic resistance in
524 *Salmonella enterica* serotypes Typhimurium, Enteritidis and Infantis from human
525 infections, foodstuffs and farm animals in Italy. *Epidemiol. Infect.* 132, 245–251.
526 <https://doi.org/10.1017/S0950268803001936>
- 527 Bushnell, B., 2014. BBMap: A fast, accurate, splice-aware aligner. Berkeley Lab Report
528 Number: LBNL-7065E.
- 529 Carattoli, A., 2009. Resistance Plasmid Families in *Enterobacteriaceae*. *Antimicrob. Agents*
530 *Chemother.* 53, 2227–2238. <https://doi.org/10.1128/AAC.01707-08>
- 531 Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller
532 Aarestrup, F., Hasman, H., 2014. *In Silico* detection and typing of plasmids using
533 PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.*
534 58, 3895–3903. <https://doi.org/10.1128/AAC.02412-14>
- 535 Chiu, C.-H., 2005. The genome sequence of *Salmonella enterica* serovar Choleraesuis, a
536 highly invasive and resistant zoonotic pathogen. *Nucleic Acids Res.* 33, 1690–1698.
537 <https://doi.org/10.1093/nar/gki297>
- 538 Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I.,
539 Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J.L., 2009. Biopython: freely
540 available Python tools for computational molecular biology and bioinformatics.
541 *Bioinformatics* 25, 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- 542 Darwish, A.E., 2013. Agricultural census project analysis and assessment of the poultry sector
543 in Lebanon. *Minist. Agric. Agric.* 1, 1–20.
- 544 Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J., Burland, V., Kodoyianni, V.,
545 Schwartz, D.C., Blattner, F.R., 2003. Comparative genomics of *Salmonella enterica*
546 serovar Typhi strains Ty2 and CT18. *J. Bacteriol.* 185, 2330–2337.
547 <https://doi.org/10.1128/JB.185.7.2330-2337.2003>
- 548 Diep, B., Barretto, C., Portmann, A.-C., Fournier, C., Karczmarek, A., Voets, G., Li, S., Deng,
549 X., Klijn, A., 2019. Salmonella serotyping; Comparison of the traditional method to a
550 microarray-based method and an *in silico* platform using whole genome sequencing
551 data. *Front. Microbiol.* 10, 2554. <https://doi.org/10.3389/fmicb.2019.02554>
- 552 Douarre, P.-E., Mallet, L., Radomski, N., Felten, A., Mistou, M.-Y., 2020. Analysis of
553 COMPASS, a new comprehensive plasmid database revealed prevalence of
554 multireplicon and extensive diversity of IncF plasmids. *Front. Microbiol.* 11, 1–15.
555 <https://doi.org/10.3389/fmicb.2020.00483>
- 556 Elder, J.R., Chiok, K.L., Paul, N.C., Haldorson, G., Guard, J., Shah, D.H., 2016. The
557 *Salmonella* pathogenicity island 13 contributes to pathogenesis in streptomycin pre-
558 treated mice but not in day-old chickens. *Gut Pathog.* 8, 16.
559 <https://doi.org/10.1186/s13099-016-0098-0>

560 Espinoza, R.A., Silva-Valenzuela, C.A., Amaya, F.A., Urrutia, Í.M., Contreras, I., Santiviago,
561 C.A., 2017. Differential roles for pathogenicity islands SPI-13 and SPI-8 in the
562 interaction of *Salmonella* Enteritidis and *Salmonella* Typhi with murine and human
563 macrophages. *Biol. Res.* 50, 5. <https://doi.org/10.1186/s40659-017-0109-8>
564 Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results
565 for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.
566 <https://doi.org/10.1093/bioinformatics/btw354>
567 Fadlallah, S.M., Shehab, M., Cheaito, K., Haidar-Ahmad, N., El Hafi, B., Saleh, M., Nasser, Z.,
568 El Hajj, R., Ghosn, N., Ammar, W., Matar, G.M., 2019. PulseNet Lebanon: An overview
569 of its activities, outbreak investigations, and challenges. *Foodborne Pathog. Dis.* 16,
570 498–503. <https://doi.org/10.1089/fpd.2018.2581>
571 Felten, A., Vila Nova, M., Durimel, K., Guillier, L., Mistou, M.-Y., Radomski, N., 2017. First
572 gene-ontology enrichment analysis based on bacterial coregenome variants: insights
573 into adaptations of *Salmonella* serovars to mammalian- and avian-hosts. *BMC Microbiol.*
574 17, 1–22. <https://doi.org/10.1186/s12866-017-1132-1>
575 Fookes, M., Schroeder, G.N., Langridge, G.C., Blondel, C.J., Mammina, C., Connor, T.R.,
576 Seth-Smith, H., Vernikos, G.S., Robinson, K.S., Sanders, M., Petty, N.K., Kingsley, R.A.,
577 Bäumlér, A.J., Nuccio, S.-P., Contreras, I., Santiviago, C.A., Maskell, D., Barrow, P.,
578 Humphrey, T., Nastasi, A., Roberts, M., Frankel, G., Parkhill, J., Dougan, G., Thomson,
579 N.R., 2011. *Salmonella bongori* provides insights into the evolution of the *Salmonellae*.
580 *PLoS Pathog.* 7, e1002191. <https://doi.org/10.1371/journal.ppat.1002191>
581 García-Fernández, A., Chiaretto, G., Bertini, A., Villa, L., Fortini, D., Ricci, A., Carattoli, A.,
582 2008. Multilocus sequence typing of IncI1 plasmids carrying extended-spectrum β -
583 lactamases in *Escherichia coli* and *Salmonella* of human and animal origin. *J.*
584 *Antimicrob. Chemother.* 61, 1229–1233. <https://doi.org/10.1093/jac/dkn131>
585 Grimont, P., 2007. Antigenic formulae of the *Salmonella* serovars. WHO Collab. Cent. Ref.
586 Res. *Salmonella* 9th edition, 1–166.
587 Guibourdenche, M., Roggentin, P., Mikoleit, M., Fields, P.I., Bockemühl, J., Grimont, P.A.D.,
588 Weill, F.-X., 2010. Supplement 2003–2007 (No. 47) to the White-Kauffmann-Le Minor
589 scheme. *Res. Microbiol.* 161, 26–29. <https://doi.org/10.1016/j.resmic.2009.10.002>
590 Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QAST: quality assessment tool for
591 genome assemblies. *Bioinformatics* 29, 1072–1075.
592 <https://doi.org/10.1093/bioinformatics/btt086>
593 Harrison, P.W., Alako, B., Amid, C., Cerdeño-Tárraga, A., Cleland, I., Holt, S., Hussein, A.,
594 Jayathilaka, S., Kay, S., Keane, T., Leinonen, R., Liu, X., Martínez-Villacorta, J., Milano,
595 A., Pakseresht, N., Rajan, J., Reddy, K., Richards, E., Rosello, M., Silvester, N.,
596 Smirnov, D., Toribio, A.-L., Vijayaraja, S., Cochrane, G., 2018. The European Nucleotide
597 Archive in 2018. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gky1078>
598 Hayward, M.R., AbuOun, M., La Ragione, R.M., Tchórzewska, M.A., Cooley, W.A., Everest,
599 D.J., Petrovska, L., Jansen, V.A.A., Woodward, M.J., 2014. SPI-23 of *S. Derby*: Role in
600 adherence and invasion of porcine tissues. *PLoS ONE* 9, e107857.
601 <https://doi.org/10.1371/journal.pone.0107857>
602 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2:
603 Improving the Ultrafast Bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522.
604 <https://doi.org/10.1093/molbev/msx281>
605 İçgen, B., 2002. Characterization of *Salmonella* Enteritidis isolates of chicken, egg and human
606 origin from Turkey. *Food Microbiol.* 19, 375–382. <https://doi.org/10.1006/fmic.2002.0479>
607 Ilyas, B., Tsai, C.N., Coombes, B.K., 2017. Evolution of *Salmonella*-host cell interactions
608 through a dynamic bacterial genome. *Front. Cell. Infect. Microbiol.* 7.
609 <https://doi.org/10.3389/fcimb.2017.00428>

- 610 Jennings, E., Thurston, T.L.M., Holden, D.W., 2017. *Salmonella* SPI-2 type III secretion system
611 effectors: Molecular mechanisms and physiological consequences. *Cell Host Microbe*
612 22, 217–231. <https://doi.org/10.1016/j.chom.2017.07.009>
- 613 Johnson, T.J., Bielak, E.M., Fortini, D., Hansen, L.H., Hasman, H., Debroy, C., Nolan, L.K.,
614 Carattoli, A., 2012. Expansion of the IncX plasmid family for improved identification and
615 typing of novel plasmids in drug-resistant *Enterobacteriaceae*. *Plasmid* 68, 43–50.
616 <https://doi.org/10.1016/j.plasmid.2012.03.001>
- 617 Jolley, K.A., Maiden, M.C., 2010. BIGSdb: Scalable analysis of bacterial genome variation at
618 the population level. *BMC Bioinformatics* 11, 595. <https://doi.org/10.1186/1471-2105-11-595>
- 620 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermini, L.S., 2017.
621 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14,
622 587–589. <https://doi.org/10.1038/nmeth.4285>
- 623 K  rouanton, A., Hirchaud, E., Rose, V., Esnault, E., Naquin, D., Denis, M., 2015. First
624 complete genome sequence of a *Salmonella enterica* subsp. *enterica* serovar Derby
625 strain associated with pork in France. *Genome Announc.* 3, e00853-15.
626 <https://doi.org/10.1128/genomeA.00853-15>
- 627 Kirk, M.D., Pires, S.M., Black, R.E., Caipo, M., Crump, J.A., Devleeschauwer, B., D  pfer, D.,
628 Fazil, A., Fischer-Walker, C.L., Hald, T., Hall, A.J., Keddy, K.H., Lake, R.J., Lanata,
629 C.F., Torgerson, P.R., Havelaar, A.H., Angulo, F.J., 2015. World Health Organization
630 estimates of the global and regional disease burden of 22 foodborne bacterial,
631 protozoal, and viral diseases, 2010: A data synthesis. *PLOS Med.* 12, e1001921.
632 <https://doi.org/10.1371/journal.pmed.1001921>
- 633 Kosugi, S., Hirakawa, H., Tabata, S., 2015. GMcloser: closing gaps in assemblies accurately
634 with a likelihood-based selection of contig or long-read alignments. *Bioinformatics*
635 *btv465*. <https://doi.org/10.1093/bioinformatics/btv465>
- 636 Langridge, G.C., Fookes, M., Connor, T.R., Feltwell, T., Feasey, N., Parsons, B.N., Seth-Smith,
637 H.M.B., Barquist, L., Stedman, A., Humphrey, T., Wigley, P., Peters, S.E., Maskell, D.J.,
638 Corander, J., Chabalgoity, J.A., Barrow, P., Parkhill, J., Dougan, G., Thomson, N.R.,
639 2015. Patterns of genome evolution that have accompanied host adaptation in
640 *Salmonella*. *Proc. Natl. Acad. Sci.* 112, 863–868.
641 <https://doi.org/10.1073/pnas.1416707112>
- 642 Letunic, I., Bork, P., 2019. Interactive tree of life (iTOL) v4: recent updates and new
643 developments. *Nucleic Acids Res.* 47, W256–W259. <https://doi.org/10.1093/nar/gkz239>
- 644 Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and
645 annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245.
646 <https://doi.org/10.1093/nar/gkw290>
- 647 Letunic, I., Bork, P., 2011. Interactive tree of life v2: online annotation and display of
648 phylogenetic trees made easy. *Nucleic Acids Res.* 39, W475–W478.
649 <https://doi.org/10.1093/nar/gkr201>
- 650 Letunic, I., Bork, P., 2007. Interactive tree of life (iTOL): an online tool for phylogenetic tree
651 display and annotation. *Bioinformatics* 23, 127–128.
652 <https://doi.org/10.1093/bioinformatics/btl529>
- 653 Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler
654 transform. *Bioinformatics* 26, 589–595.
- 655 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
656 Durbin, R., Subgroup, 1000 Genome Project Data Processing, 2009. The Sequence
657 Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

- 658 Lou, L., Zhang, P., Piao, R., Wang, Y., 2019. *Salmonella* pathogenicity island 1 (SPI-1) and its
659 complex regulatory network. *Front. Cell. Infect. Microbiol.* 9, 270.
660 <https://doi.org/10.3389/fcimb.2019.00270>
- 661 Low, A.J., Koziol, A.G., Manninger, P.A., Blais, B.W., Carrillo, C.D., 2019. ConFindr: Rapid
662 detection of intraspecies and cross-species contamination in bacterial whole-genome
663 sequence data. <https://doi.org/10.7287/peerj.preprints.27499v1>
- 664 Magiorakos, A.-P., Srinivasan, A., Carey, R.B., Carmeli, Y., Falagas, M.E., Giske, C.G.,
665 Harbarth, S., Hindler, J.F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D.L., Rice,
666 L.B., Stelling, J., Struelens, M.J., Vatopoulos, A., Weber, J.T., Monnet, D.L., 2012.
667 Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an
668 international expert proposal for interim standard definitions for acquired resistance.
669 *Clin. Microbiol. Infect.* 18, 268–281. <https://doi.org/10.1111/j.1469-0691.2011.03570.x>
- 670 McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik,
671 S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K.,
672 Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
673 Stoneking, T., Nhan, M., Waterston, R., Wilson, R.K., 2001. Complete genome
674 sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413, 852–856.
675 <https://doi.org/10.1038/35101614>
- 676 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.,
677 Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit:
678 a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
679 *Res.* 20, 1297–303.
- 680 Meysman, P., Sánchez-Rodríguez, A., Fu, Q., Marchal, K., Engelen, K., 2013. Expression
681 divergence between *Escherichia coli* and *Salmonella enterica* serovar Typhimurium
682 reflects their lifestyles. *Mol. Biol. Evol.* 30, 1302–1314.
683 <https://doi.org/10.1093/molbev/mst029>
- 684 Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic
685 bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- 686 Nava M., G., Hernandez, Y.E., 2016. Comparative genomics of *Salmonella* could reveal key
687 features of adaptation. *J. Data Min. Genomics Proteomics* 07.
688 <https://doi.org/10.4172/2153-0602.1000e121>
- 689 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and
690 effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol. Biol.*
691 *Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>
- 692 Olesen, S.W., Barnett, M.L., MacFadden, D.R., Brownstein, J.S., Hernández-Díaz, S., Lipsitch,
693 M., Grad, Y.H., 2018. The distribution of antibiotic use and its association with antibiotic
694 resistance. *eLife* 7, e39435. <https://doi.org/10.7554/eLife.39435>
- 695 Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy,
696 A.M., 2016. Mash: fast genome and metagenome distance estimation using MinHash.
697 *Genome Biol.* 17, 132. <https://doi.org/10.1186/s13059-016-0997-x>
- 698 Orlek, A., Stoesser, N., Anjum, M.F., Doumith, M., Ellington, M.J., Peto, T., Crook, D.,
699 Woodford, N., Walker, A.S., Phan, H., Sheppard, A.E., 2017. Plasmid classification in an
700 era of whole-genome sequencing: Application in studies of antibiotic resistance
701 epidemiology. *Front. Microbiol.* 8, 182. <https://doi.org/10.3389/fmicb.2017.00182>
- 702 Palma, F., Brauge, T., Radomski, N., Mallet, L., Felten, A., Mistou, M.-Y., Brisabois, A., Guillier,
703 L., Midelet-Bourdin, G., 2020. Dynamics of mobile genetic elements of *Listeria*
704 *monocytogenes* persisting in ready-to-eat seafood processing plants in France. *BMC*
705 *Genomics* 21, 130.
- 706 Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C.,
707 Mungall, K.L., Bentley, S.D., Holden, M.T.G., Sebahia, M., Baker, S., Basham, D.,

708 Brooks, K., Chillingworth, T., Connerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd,
709 L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
710 Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., Ó'Gaora, P., Parry, C., Quail,
711 M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G.,
712 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica*
713 serovar Typhi CT18. *Nature* 413, 848–852. <https://doi.org/10.1038/35101607>

714 Pezzella, C., Ricci, A., DiGiannatale, E., Luzzi, I., Carattoli, A., 2004. Tetracycline and
715 streptomycin resistance genes, transposons, and plasmids in *Salmonella enterica*
716 Isolates from animals in Italy. *Antimicrob. Agents Chemother.* 48, 903–908.
717 <https://doi.org/10.1128/AAC.48.3.903-908.2004>

718 Pightling, A.W., Pettengill, J.B., Wang, Y., Rand, H., Strain, E., 2019. Within-species
719 contamination of bacterial whole-genome sequence data has a greater influence on
720 clustering analyses than between-species contamination. *Genome Biol.* 20, 286.
721 <https://doi.org/10.1186/s13059-019-1914-x>

722 R Development Core Team, (2015), n.d. R: A Language and Environment for Statistical
723 Computing. R Foundation for Statistical Computing, Vienna, Austria.

724 Rehman, M.A., Ziebell, K., Nash, J.H.E., Kropinski, A.M., Zong, Z., Nafziger, E., Boerlin, P.,
725 Chui, L., Devenish, J., Bekal, S., Graham, M., Johnson, R.P., 2014. High-quality draft
726 whole-genome sequences of 162 *Salmonella enterica* subsp. *enterica* serovar Enteritidis
727 strains isolated from diverse sources in Canada. *Genome Announc.* 2, e00348-14,
728 2/2/e00348-14. <https://doi.org/10.1128/genomeA.00348-14>

729 Retamal, P., Fresno, M., Dougnac, C., Gutierrez, S., Gornall, V., Vidal, R., Vernal, R., Pujol,
730 M., Barreto, M., González-Acuña, D., Abalos, P., 2015. Genetic and phenotypic
731 evidence of the *Salmonella enterica* serotype Enteritidis human-animal interface in
732 Chile. *Front. Microbiol.* 6, 1–10. <https://doi.org/10.3389/fmicb.2015.00464>

733 Robertson, J., Nash, J.H.E., 2018. MOB-suite: software tools for clustering, reconstruction and
734 typing of plasmids from draft assemblies. *Microb. Genomics* 4.
735 <https://doi.org/10.1099/mgen.0.000206>

736 Rychlik, I., Gregorova, D., Hradecka, H., 2006. Distribution and function of plasmids in
737 *Salmonella enterica*. *Vet. Microbiol.* 112, 1–10.
738 <https://doi.org/10.1016/j.vetmic.2005.10.030>

739 Rychlik, I., Karasova, D., Sebkova, A., Volf, J., Sisak, F., Havlickova, H., Kummer, V., Imre, A.,
740 Szmolka, A., Nagy, B., 2009. Virulence potential of five major pathogenicity islands (SPI-
741 1 to SPI-5) of *Salmonella enterica* serovar Enteritidis for chickens. *BMC Microbiol.* 9,
742 268. <https://doi.org/10.1186/1471-2180-9-268>

743 Sabbagh, S.C., Forest, C.G., Lepage, C., Leclerc, J.-M., Daigle, F., 2010. So similar, yet so
744 different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars
745 Typhimurium and Typhi: Genomic comparison of *S. Typhimurium* and *S. Typhi*. *FEMS*
746 *Microbiol. Lett.* 305, 1–13. <https://doi.org/10.1111/j.1574-6968.2010.01904.x>

747 Salipante, S.J., 2003. Determining the limits of the evolutionary potential of an antibiotic
748 resistance gene. *Mol. Biol. Evol.* 20, 653–659. <https://doi.org/10.1093/molbev/msg074>

749 Satyanarayana, T., Johri, B.N., Das, S.K., 2019. Microbial diversity in ecosystem sustainability
750 and biotechnological applications: Volume 1. Microbial diversity in normal & extreme
751 environments.

752 Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–
753 2069. <https://doi.org/10.1093/bioinformatics/btu153>

754 Shah, D.H., Lee, M., Park, J., Lee, J., Eo, S., Kwon, J., Chae, J., 2005. Identification of
755 *Salmonella Gallinarum* virulence genes in a chicken infection model using PCR-based
756 signature-tagged mutagenesis. *Microbiology* 151, 3957–3968.
757 <https://doi.org/10.1099/mic.0.28126-0>

758 Sheppard, S.K., Guttman, D.S., Fitzgerald, J.R., 2018. Population genomics of bacterial host
759 adaptation. *Nat. Rev. Genet.* 19, 549–565. <https://doi.org/10.1038/s41576-018-0032-z>

760 Soto, S.M., Rodríguez, I., Rodicio, M.R., Vila, J., Mendoza, M.C., 2006. Detection of virulence
761 determinants in clinical strains of *Salmonella enterica* serovar Enteritidis and mapping
762 on macrorestriction profiles. *J. Med. Microbiol.* 55, 365–373.
763 <https://doi.org/10.1099/jmm.0.46257-0>

764 Suresh, T., Hatha, A.A.M., Sreenivasan, D., Sangeetha, N., Lashmanaperumalsamy, P., 2006.
765 Prevalence and antimicrobial resistance of *Salmonella* Enteritidis and other salmonellas
766 in the eggs and egg-storing trays from retails markets of Coimbatore, South India. *Food*
767 *Microbiol.* 23, 294–299. <https://doi.org/10.1016/j.fm.2005.04.001>

768 Tanner, J.R., Kingsley, R.A., 2018. Evolution of *Salmonella* within hosts. *Trends Microbiol.* 26,
769 986–998. <https://doi.org/10.1016/j.tim.2018.06.001>

770 Thomson, N.R., Clayton, D.J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., Quail,
771 M.A., Stevens, M., Jones, M.A., Watson, M., Barron, A., Layton, A., Pickard, D.,
772 Kingsley, R.A., Bignell, A., Clark, L., Harris, B., Ormond, D., Abdellah, Z., Brooks, K.,
773 Cherevach, I., Chillingworth, T., Woodward, J., Norberczak, H., Lord, A., Arrowsmith, C.,
774 Jagels, K., Moule, S., Mungall, K., Sanders, M., Whitehead, S., Chabalgoity, J.A.,
775 Maskell, D., Humphrey, T., Roberts, M., Barrow, P.A., Dougan, G., Parkhill, J., 2008.
776 Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum
777 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res.*
778 18, 1624–1637. <https://doi.org/10.1101/gr.077404.108>

779 Tomljenovic-Berube, A.M., Henriksbo, B., Porwollik, S., Cooper, C.A., Tuinema, B.R.,
780 McClelland, M., Coombes, B.K., 2013. Mapping and regulation of genes within
781 *Salmonella* Pathogenicity Island 12 that contribute to *In Vivo* fitness of *Salmonella*
782 *enterica* serovar Typhimurium. *Infect. Immun.* 81, 2394–2404.
783 <https://doi.org/10.1128/IAI.00067-13>

784 Vernikos, G.S., Parkhill, J., 2006. Interpolated variable order motifs for identification of
785 horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands.
786 *Bioinformatics* 22, 2196–2203. <https://doi.org/10.1093/bioinformatics/btl369>

787 Vila Nova, M., Durimel, K., La, K., Felten, A., Bessières, P., Mistou, M.-Y., Mariadassou, M.,
788 Radomski, N., 2019. Genetic and metabolic signatures of *Salmonella enterica* subsp.
789 *enterica* associated with animal sources at the pangenomic scale. *BMC Genomics* 20,
790 1–21. <https://doi.org/doi:10.1186/s12864-019-6188-x>

791 Villa, L., García-Fernández, A., Fortini, D., Carattoli, A., 2010. Replicon sequence typing of
792 IncF plasmids carrying virulence and resistance determinants. *J. Antimicrob.*
793 *Chemother.* 65, 2518–2529. <https://doi.org/10.1093/jac/dkq347>

794 Wickham, H., 2009. ggplot2. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-98141-3>

795

796 Xu, S., Ackerman, M.S., Long, H., Bright, L., Spitze, K., Ramsdell, J.S., Thomas, W.K., Lynch,
797 M., 2015. A male-specific genetic map of the *microcrustacean daphnia pulex* based on
798 single-sperm whole-genome sequencing. *Genetics* 201, 31–38.
799 <https://doi.org/10.1534/genetics.115.179028>

800 Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup,
801 F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. *J.*
802 *Antimicrob. Chemother.* 67, 2640–2644. <https://doi.org/10.1093/jac/dks261>

803 Zhang, S., den Bakker, H.C., Li, S., Chen, J., Dinsmore, B.A., Lane, C., Lauer, A.C., Fields,
804 P.I., Deng, X., 2019. SeqSero2: Rapid and improved *Salmonella* serotype determination
805 using whole-genome sequencing data. *Appl. Environ. Microbiol.* 85, e01746-19,
806 /aem/85/23/AEM.01746-19.atom. <https://doi.org/10.1128/AEM.01746-19>

- 807 Zhang, S., Yin, Y., Jones, M.B., Zhang, Z., Deatherage Kaiser, B.L., Dinsmore, B.A.,
808 Fitzgerald, C., Fields, P.I., Deng, X., 2015. *Salmonella* serotype determination utilizing
809 high-throughput genome sequencing data. J. Clin. Microbiol. 53, 1685–1692.
810 <https://doi.org/10.1128/JCM.00323-15>
- 811 Zhang, Y., Chen, Y., Gu, T., Xu, Q., Zhu, G., Chen, G., 2019. Effects of *Salmonella enterica*
812 serovar Enteritidis infection on egg production and the immune response of the laying
813 duck *Anas platyrhynchos*. PeerJ 7, e6359. <https://doi.org/10.7717/peerj.6359>
814

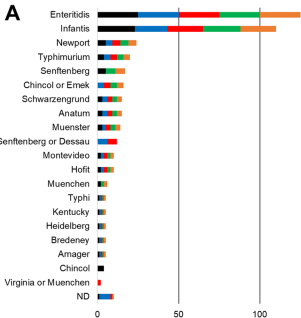
815 **FIGURE LEGENDS**

816 **Figure 1: Genomic-based predictions of *Salmonella* serovars (A) isolated from human**
817 **(n=24) and food (n=63) in Lebanon between 2017 and 2018 (A) and corresponding**
818 **prediction consistencies in comparison with seroagglutination based on Kauffmann-Le**
819 **Minor scheme (B).** The genomic-based predictions 1, 2, 3 and 4 were performed using
820 contigs (i.e. Spades) or scaffolds (i.e. Spades+Medusa+GapCloser from the workflow
821 ARTwork) as input of SeqSero 1 or 2.

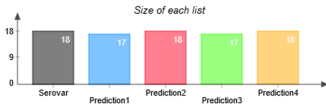
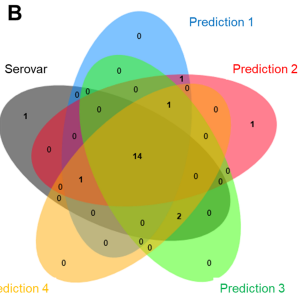
822 **Figure 2: Coregenome SNP-based phylogenomic tree by maximum likelihood (TVM+F+I**
823 **converged at 102) of *Salmonella* Enteritidis isolated from human and food in Lebanon**
824 **between 2017 and 2018 (n=25), as well as other countries (n=130).** The variant calling
825 analysis and maximum likelihood were performed with the workflows iVarCall2 and IQ-Tree,
826 respectively. The source-type, country and isolation dates are represented. Bootstrap values
827 higher than 90 are marked by black circles and the tree was rooted with Dublin serovar as an
828 external branch (sample identifiers 1SD0102, 1SD0107, SRR5583067).

829 **Figure 3: Pairwise SNP differences between intra- (A) and inter-groups of countries (B)**
830 **at the coregenome scale of *Salmonella* Enteritidis isolated in Lebanon between 2017 and**
831 **2018 (n=25), as well as reads from international Archives (n=130).** Mapping-based variant
832 calling was performed with the workflow iVarCall2. The paired-end reads corresponding to the
833 most prevalence serovar in Lebanon (i.e. *S. Enteritidis*) were mapped with the variant calling
834 workflow iVarCall2 against the reference genome of the *S. Enteritidis* strain EC20090193
835 (CP007321-1). The intra-groups of countries include at least two samples same the same
836 countries. The inter-groups of countries are compared against the country Lebanon.

837



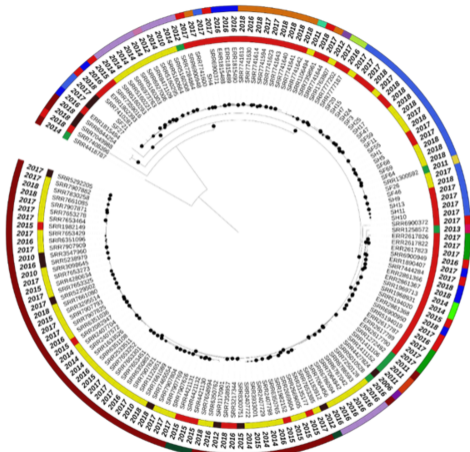
- Serovar (seroagglutination)
- Prediction 1: SeqSero1 prediction from contigs
- Prediction 2: SeqSero1 prediction from scaffolds
- Prediction 3: SeqSero2 prediction from contigs
- Prediction 4: SeqSero2 prediction from scaffolds



source-type



country



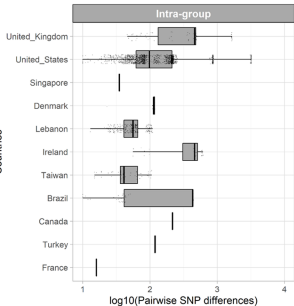
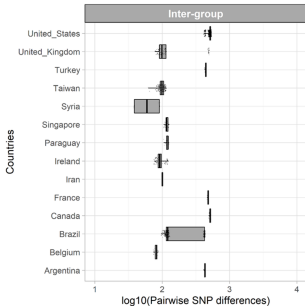
A**B**

Table 1: Database of all known *Salmonella* Pathogenicity Islands (SPIs). The reference SPI sequences were retrieved from the PAtHogenesisity Island DataBase (PAIdb) and National Center for Biotechnology Information (NCBI).

SPI	Sequence accession	Genome origin	Sequence origin	DNA sequence size (bases)	Gene number	Reference
SPI-1	NC_003197_P3	Typhimurium LT2	PAIdb	44279	47	(McClelland et al., 2001)
SPI-2	NC_003197_P2	Typhimurium LT2	PAIdb	40071	46	(McClelland et al., 2001)
SPI-3	NC_003197_P4	Typhimurium LT2	PAIdb	16616	15	(McClelland et al., 2001)
SPI-4	NC_003197_P5	Typhimurium LT2	PAIdb	23411	6	(McClelland et al., 2001)
SPI-5	NC_003197_P1	Typhimurium LT2	PAIdb	9069	9	(McClelland et al., 2001)
SPI-6	NC_003198_P1	Typhi CT18	PAIdb	58666	59	(Parkhill et al., 2001)
SPI-7	NC_003198_P9	Typhi CT18	PAIdb	133638	152	(Parkhill et al., 2001)
SPI-8	NC_003198_P6	Typhi CT18	PAIdb	6885	10	(Parkhill et al., 2001)
SPI-9	NC_003198_P4	Typhi CT18	PAIdb	15696	4	(Parkhill et al., 2001)
SPI-10	NC_003198_P10	Typhi CT18	PAIdb	32934	33	(Parkhill et al., 2001)
SPI-11	NC_006905_P2	Choleraesuis SC-B67	PAIdb	15686	15	(Chiu, 2005)
SPI-12	NC_006905_P4	Choleraesuis SC-B67	PAIdb	11075	12	(Chiu, 2005)
SPI-13	NC_003197.2:3276387-3280545	Typhimurium LT2	NCBI	4159	5	(McClelland et al., 2001)
SPI-14	NC_003197.2:926613-932916	Typhimurium LT2	NCBI	6304	6	(McClelland et al., 2001)
SPI-15	AL513382.1:3054094-3059809	Typhi CT18	NCBI	5716	7	(Parkhill et al., 2001)
SPI-16	AE006468.2:613596-617725	Typhimurium LT2	NCBI	4130	7	(McClelland et al., 2001)
SPI-17	AL513382.1:2460793-2465770	Typhi CT18	NCBI	4978	4	(Parkhill et al., 2001)
SPI-18	AL513382.1:1455055-1456801	Typhi CT18	NCBI	1747	2	(Parkhill et al., 2001)
SPI-19	AM933173.1:1110982-1150641	Gallinarum 287/91	NCBI	39660	37	(Thomson et al., 2008)
SPI-20	NC_010067.1:2617346-2651626	arizonae 62:z4,z23	NCBI	34281	31	(Blondel et al., 2009)
SPI-21	NC_010067.1:2504768-2560524	arizonae 62:z4,z23	NCBI	55757	52	(Blondel et al., 2009)
SPI-22	NC_015761.1:1349885-1369393	bongori NCTC 12419	NCBI	19509	16	(Fookes et al., 2011)
SPI-23	LAZB00000000	Derby 07CR553	NCBI	36540	36	(K�erouanton et al., 2015)

Table 2: Statistical metrics of pairwise SNP differences between intra- (A) and inter-groups of countries (B) at the coregenome scale of *Salmonella* Enteritidis isolated in Lebanon between 2017 and 2018 (n=25), as well as reads from international Archives (n=130). Mapping-based variant calling was performed with the workflow iVarCall2. The paired-end reads corresponding to the most prevalence serovar in Lebanon (i.e. *S. Enteritidis*) were mapped with the variant calling workflow iVarCall2 against the reference genome of the *S. Enteritidis* strain EC20090193 (CP007321-1). The intra-groups of countries include at least two samples same the same countries. The inter-groups of countries are compared against the country Lebanon. N/A stands for not applicable in case of less than three samples. The present comparison does not include all publicly available data at NCBI Pathogen detection.

	Countries	Mean	Standard deviation	Quantile (50%)	Quantile (25%)	Quantile (75%)	Minimum	Maximum
A	France	16	N/A	16	16	16	16	16
	Turkey	119	N/A	119	119	119	119	119
	Canada	110	116	110	3	216	3	216
	Brazil	235	208	430	32	434	0	445
	Taiwan	48	29	41	28	64	0	106
	Ireland	358	225	461	94	486	0	615
	Lebanon	56	27	54	37	65	0	110
	Denmark	87	49	114	66	116	1	119
	Singapore	35	NA	35	35	35	35	35
	United-States	268	578	91	59	213	0	3220
	United-Kingdom	527	508	462	136	483	46	1668
B	Iran	101	2	101	100	101	99	102
	Syria	65	31	65	39	91	38	93
	France	481	6	482	478	485	464	490
	Turkey	443	6	444	440	449	425	454
	Canada	517	9	520	510	524	490	530
	Brazil	240	153	121	115	426	96	438
	Taiwan	99	8	99	94	105	62	116
	Ireland	95	13	92	87	98	76	124
	Belgium	82	4	83	79	84	74	88
	Paraguay	119	5	122	116	123	105	127
	Argentina	431	6	433	428	434	415	439
	Singapore	117	5	116	114	122	105	127
	United-States	513	22	518	508	525	409	542
	United-Kingdom	193	171	99	91	115	78	501