



**HAL**  
open science

# Génomique épidémiologique de Salmonella

Alicia Tran Dien

► **To cite this version:**

Alicia Tran Dien. Génomique épidémiologique de Salmonella. Médecine humaine et pathologie. Institut agronomique, vétérinaire et forestier de France, 2018. Français. NNT : 2018IAVF0001 . tel-02006644

**HAL Id: tel-02006644**

**<https://theses.hal.science/tel-02006644>**

Submitted on 4 Feb 2019

**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.

# THESE DE DOCTORAT

préparée à l'Institut des sciences et industries du vivant et de l'environnement

(AgroParisTech)

pour obtenir le grade de

**Docteur de l'Institut agronomique vétérinaire et forestier de France**

**Spécialité : Sciences de la vie et de la santé**

École doctorale n°581

Agriculture, alimentation, biologie, environnement et santé (ABIES)

*par*

**Alicia Tu My Linh TRAN DIEN**

**Génomique épidémiologique de *Salmonella***

Directeur de thèse : **François-Xavier WEILL**

**Thèse présentée et soutenue à l'Institut Pasteur, le 11 Janvier 2018.**

**Composition du jury :**

M. Christophe SOLA, Professeur, Université Paris-Saclay	Président
M. Axel CLOECKAERT, Directeur de recherche, INRA- Tours-Nouzilly	Rapporteur
Mme Marianne CHEMALY, Ingénieur de recherche, ANSES Ploufragan	Rapporteur
Mme Anne BRISABOIS, Directeur de recherche, ANSES Maisons-Alfort	Examinatrice
M. François-Xavier WEILL, Directeur de recherche, Institut Pasteur	Directeur de thèse

**Unité des Bactéries Pathogènes Entériques**

28 rue du Dr Roux, 75724 Paris cedex 15, France



## Remerciements

En premier lieu, j'aimerais remercier mon directeur de thèse, le Docteur François-Xavier Weill sans qui, le projet n'aurait jamais vu le jour et n'aurait pu aboutir. Là pour moi en toutes circonstances, il trouvait toujours les bons mots pour me rassurer, m'encourager et sa rigueur de travail est un exemple pour moi à suivre. J'ai encore beaucoup à apprendre de vous. Mes sincères remerciements.

La réalisation de cette thèse a été rendue possible grâce au financement octroyé par La Fondation Le Roch-Les Mousquetaires. Merci pour la confiance que vous nous avez accordée.

Je voudrais également remercier les membres du jury, le Professeur Christophe Sola (président du jury), Docteur Axel Cloeckert (rapporteur), Docteur Marianne Chemaly (rapporteur) et Docteur Anne Brisabois (examinatrice) de me donner l'opportunité de présenter le fruit d'un travail de quatre années.

Je tiens à remercier de tout cœur mes parents et toute ma famille Bâ Tinti, Pon, Belle-sœur et la liste continue, ainsi que toutes les personnes parties beaucoup trop tôt à qui je dédie cette thèse... Les mots ne suffisent pas pour exprimer ma gratitude et ma reconnaissance.

Merci à Jessica Wermuth, ma chou, malgré le temps et la distance, tu as toujours été à mes côtés depuis le début. Si patiente avec moi, tu as su me dompter, partager les plus beaux moments de ma vie comme les plus tristes.

Merci à Fiona Maillard et Amélie Bhuiyan. Poussin tu as été LA première à m'accueillir à Pasteur, qui aurait cru que tu finirais dans les remerciements cinq ans plus tard, figurant parmi les premières ! Pakou, ma ministre, je te dois tout à ce niveau-là !

Merci à Laetitia Fabre, Maxime, my QOD and dearest friend, we've been through so much! You raised me, taught me everything and even sacrificed yourself for me. I owe you everything. You are the spring to my roll, the yin to my yang, the stars to my ninja and the

corn to my horse. I'm so grateful to share this special relationship with you Mate. I'll finish by saying, turtle Loves panda a lot... Like... A lot.

Merci à Isabelle Carle, Topsy ! Tous les matins j'attendais avec impatience ton arrivée pour ma séance psychologie et notre petit moment musical pour lancer et égayer mes journées ! Plus qu'une collègue, tu es une amie sur qui je peux toujours compter.

Merci à Nizar Fawal, my bro, you spent so many hours teaching me bio-informatics, always pushing me and I'm glad you did. All the apple breaks learning life with you. Thanks to you I'm now able to do things I didn't think I could. With you, everything is awesome !

Merci à Sophia Myriam Achaïbou de m'avoir fait rire tous les jours même quand tu n'étais pas là. Au moment où j'écris ces remerciements, cela fera un an que nous nous sommes rencontrées, un an où tu as été un vrai rayon de soleil, un an d'une belle amitié.

Je remercie tous les stagiaires que j'ai pu rencontrer durant ces quatre années et plus particulièrement : Alexa Saliou (mon tout petit Cupcake, raison pour laquelle je me levais tous les matins durant trois mois), Docteur Camille Gomart (on en aura mangé des petits pois au wasabi !) et Pierre Lafoy (héros de l'Univers).

Je voulais remercier toute l'équipe BPE pour m'avoir accompagnée tout au long de cette aventure, notamment Elisabeth Njamkepo Nguemkam pour son aide et son écoute, Sophie Lefevre pour ses conseils et sa bonne humeur, ainsi que Jean Rauzier et Marie-Laure Quilici pour qui j'ai une affection toute particulière.

Enfin, je tiens à remercier Catherine Dauga et Karin Eiglmeier pour leurs précieux conseils et encouragements.

A tous, un grand merci.

## Liste des abréviations

**ADN** : Acide DésoxyriboNucléique

**BLAST**: Basic Local Alignment Search Tool

**BWA**: Burrows-Wheeler-Alignment

**CG**: CRISPOL Groupe

**cgMLST**: Core Genome MultiLocus Sequence Typing

**CMI** : Concentration Minimale Inhibitrice

**CNR** : Centre National de Référence

**CRISPOL**: CRISpr POLymorphisms

**CRISPR**: Clustered Regularly Interspaced Short Palindromic Repeats

**CT** : CRISPOL Type

**CTent** : CRISPOL Type ENTeritidis

**DR** : Direct Repeat

**eBGs** : eBurstGroups

**ECP** : Electrophorèse en Champ Pulsé.

**IS** : Séquence d'Insertion

**MLST** : MultiLocus Sequence Typing

**MLVA** : MultiLocus VNTR Analysis

**MST**: Minimum Spanning Tree

**PCR**: Polymerase chain reaction

**pMLST**: Plasmid MultiLocus Sequence Typing

**rMLST**: Ribosomal MultiLocus Sequence Typing

**S.** : *Salmonella*

**SEG** : Séquençage Entier du Génome

**SNP**: (Single Nucleotide Polymorphism

**SNP**: Single Nucleotide Polymorphism

**ST**: Sequence Type

**TBW**: Transformation de Burrows Wheeler

**TCS**: Trypto-Caséine-Soja

**Tn**: Transposon

**VNTR**: Variable Number of Tandem Repeat

**wgMLST**: Whole Genome MultiLocus Sequence Typing

**WKL** : White-Kauffmann-Le Minor

## Liste des figures

Figure 1. Classification de <i>Salmonella</i> .....	17
Figure 2. Composition antigénique de <i>Salmonella</i> . ....	25
Figure 3. Illustration d'une lyse bactérienne .....	26
Figure 4. Protocole du ribotypage.....	27
Figure 5. Electrophorèse en champ pulsé avec voltage alternatif.....	28
Figure 6. Profils de dix souches de <i>Salmonella enterica</i> sérotype Typhi obtenus à partir de l'ECP.....	28
Figure 7. Protocole de la MLVA.....	30
Figure 8. Arbre couvrant de poids minimum ou MST (Minimum Spanning Tree, voir chapitre IV) des profils MLST de 4,257 isolats de <i>S. enterica</i> sous-espèce <i>enterica</i> .....	32
Figure 9. Différentes structures du système CRISPR/cas chez <i>S. enterica</i> et <i>S. bongori</i> . ....	34
Figure 10. Mécanisme immunitaire du système CRISPR-cas. ....	36
Figure 11. Protocole du CRISPOL Luminex® développé pour <i>S. enterica</i> sérotype Typhimurium. ....	38
Figure 12. Principales cibles des grandes familles d'antibiotiques.....	40
Figure 13. Représentation des séquençages single-end et paired-end.....	52
Figure 14. Schéma de reconstruction et d'analyse génomique à partir de l'ADN au séquençage entier du génome.....	56
Figure 15. Organigramme à suivre pour l'obtention d'un résultat phylogénétique fiable à partir des séquences. ....	58
Figure 16. Carte génétique de la souche de <i>S. enterica</i> sérotype Typhimurium LT2. ....	59
Figure 17. Schéma du pipeline utilisé par SeqSero pour la détermination du sérotype .....	60
Figure 18. Pipeline du CRISPOL <i>in silico</i> . ....	78
Figure 19. Relation phylogénétique de 368 génomes de <i>S. enterica</i> sérotype Typhimurium et leur profil CRISPOL correspondant.....	85
Figure 20. Relation phylogénétique de 405 génomes de <i>S. enterica</i> sérotype Enteritidis et leur profil CRISPOL correspondant.....	94
Figure 21. Les sites d'intégration du transposon Tn2 au sein du plasmide de virulence IncF pSLT chez cinq isolats de <i>S. enterica</i> sérotype Typhimurium résistants à l'ampicilline.	108
Figure 22. Arbre phylogénétique de maximum de vraisemblance des isolats de <i>S. enterica</i> sérotype Typhimurium.....	111

Figure 23. Caractéristiques géographiques, archéologiques et métagénomiques du squelette Ragna.....	120
Figure 24. Phylogénie génomique des <i>Salmonella enterica</i> et de la lignée Para C. ....	124
Figure 25. Représentation de l'évolution de la lignée Para C au cours du temps. ....	125
Figure 26. Répartition des ilots de pathogénicité de <i>Salmonella</i> au sein de la lignée C. ....	126
Figure 27. Schéma représentatif de l'histoire évolutive de la lignée Para C au cours du temps. ....	126

## Liste des tableaux

Tableau 1. Distribution des génomes séquencés par les plateformes MiSeq, HiSeq 2500 et NextSeq 500. ....	74
Tableau 2. Liste des versions et paramètres utilisés pour l'assemblage <i>de novo</i> et l'alignement.....	77
Tableau 3. Caractéristiques des génomes générés par l'assembleur.....	80
Tableau 4. Liste des espaceurs utilisés pour le CRISPOL <i>in silico</i> Enteritidis.....	93
Tableau 5. Caractéristiques des premiers isolats de <i>S. enterica</i> (et <i>E. coli</i> TEM) porteurs de plasmides de résistance à l'ampicilline .....	101
Tableau 6. Caractéristiques des 11 isolats de <i>S. enterica</i> sérotype Typhimurium résistants à l'ampicilline identifiés dans cette étude. ....	109

## Liste des graphiques

Graphique 1. Prix du séquençage d'un génome de 2001 à 2014. <sup>94</sup> .....	50
Graphique 2. Nombre de génomes soumis sur NCBI ces vingt dernières années. <sup>163</sup> .....	66
Graphique 3. Pourcentage des résultats obtenus par chaque outil avec le CRISPOL <i>in silico</i> . 81	
Graphique 4. Application du CRISPOL <i>in silico</i> pour une étude épidémiologique de 868 isolats de <i>S. enterica</i> sérotype Typhimurium isolés entre Juin et Novembre.....	82
Graphique 5. Nombre d'isolats de <i>S. enterica</i> sérotype Typhimurium résistants à l'ampicilline et producteurs de bêta-lactamase, par année d'isolation.....	105
Graphique 6. Distribution de la CMI de l'ampicilline et de la pénicilline G, sur une sélection de 67 isolats de <i>S. enterica</i> sérotype Typhimurium.....	106



## Liste des annexes

<b>Annexe 1 Chapitre V</b> .....	151
Annexe 1 a. Caractéristiques des génomes et isolats de <i>Salmonella enterica</i> étudiés .	151
Annexe 1 b. Arbre phylogénétique de maximum de vraisemblance avec les bootstraps et les noms des souches de <i>S. enterica</i> .....	158
<b>Annexe 2 Chapitre VI</b> .....	160
Annexe 2 a. Distribution des souches dans cette étude et pour le séquençage entier du génome.....	160
Annexe 2 b. Distribution phylogénétique des CMI pour l'ampicilline et la penicilline G, sur une sélection de 67 isolats de <i>S. enterica</i> sérotype Typhimurium. ....	160
Annexe 2 c. Fréquence de transfert des différents plasmides de <i>S. enterica</i> sérotype Typhimurium. ....	161
Annexe 2 d. Arbre phylogénétique de maximum de vraisemblance des souches de <i>S. enterica</i> sérotype Typhimurium.....	162
Annexe 2 e. Caractéristiques des génomes et isolats de <i>Salmonella enterica</i> étudiés.	163
Annexe 2 f. Article soumis et accepté par Lancet ID, version anglaise.....	169
<b>Annexe 3. Chapitre VII</b> .....	186
Annexe 3 a. Article “ Millennia of genomic stability within the invasive Para C lineage of <i>Salmonella enterica</i> ”.	186

## Sommaire

<b>Etude bibliographique.....</b>	<b>13</b>
<b>Chapitre I.....</b>	<b>15</b>
<b><i>Salmonella</i>.....</b>	<b>15</b>
I. Historique.....	16
II. Taxonomie et nomenclature.....	16
III. Habitat, pathogénie et infectiologie.....	17
1. <i>Habitat</i> .....	17
2. <i>Pathogénie et infectiologie</i> .....	18
IV. Surveillance, épidémiologie et prophylaxie.....	20
1. <i>Surveillance</i> .....	20
2. <i>Epidémiologie</i> .....	22
3. <i>Prophylaxie</i> .....	23
V. Identification des espèces et sous-espèces.....	23
1. Typage des <i>Salmonella</i> .....	23
2. Sous-typage des <i>Salmonella</i> .....	25
<b>Chapitre II.....</b>	<b>29</b>
<b>Typage et sous-typage moléculaire des <i>Salmonella</i> .....</b>	<b>29</b>
<b>à haut-débit.....</b>	<b>29</b>
I. Multilocus VNTR analysis.....	30
II. Multilocus Sequence Typing.....	31
III. Polymorphisme des régions CRISPR.....	33
1. Composition du CRISPR.....	33
2. Mécanisme immunitaire du système CRISPR-Cas.....	34
3. CRISPOL Luminex®.....	37
<b>Chapitre III.....</b>	<b>39</b>
<b>Génétique de la résistance aux antibiotiques chez <i>Salmonella</i>.....</b>	<b>39</b>
I. Cibles des antibiotiques.....	40
II. La résistance bactérienne.....	40

1. Généralités .....	40
2. Les mécanismes de résistance .....	41
3. Support génétique de la résistance® .....	42
4. Mécanisme de transfert de la résistance .....	43
5. Détection au laboratoire .....	44
III. Résistance chez les <i>Salmonella</i> .....	45
1. Gènes .....	45
2. Cas des $\beta$ lactam .....	45
3. Mutations .....	46
IV. Epidémies de <i>Salmonella</i> liées à la résistance aux antibiotiques .....	47
<b>Chapitre IV .....</b>	<b>49</b>
<b>Génomique et séquençage .....</b>	<b>49</b>
I. Introduction au séquençage entier du génome (SEG) .....	50
1. ADN et génome .....	50
2. Technologies de séquençage .....	51
3. Protocole de séquençage .....	52
4. Analyse génomique .....	53
5. Analyse phylogénétique .....	56
II. Génomique de <i>Salmonella</i> .....	58
III. Outils dédiés à <i>S. enterica</i> sérotype Typhimurium basés sur le SEG .....	59
1. SeqSero .....	60
2. Salmonella in silico Typing Resource (SISTR) .....	61
3. Center for Genomic Epidemiology .....	61
4. Enterobase .....	63
5. Panseq .....	64
6. PHAST .....	64
<b>Etude expérimentale .....</b>	<b>65</b>
<b>Objectifs de la thèse .....</b>	<b>66</b>

<b>Chapitre V</b> .....	<b>69</b>
<b>SEG : Typage à visée épidémiologique</b> .....	<b>69</b>
<b>(Article1)</b> .....	<b>69</b>
Abstract .....	70
CRISPOL <i>in silico</i> , méthode de typage de <i>Salmonella enterica</i> sérotype Typhimurium .....	71
I. Matériels et méthodes .....	73
II. Résultats.....	79
III. Discussion.....	86
Conclusion .....	89
CRISPOL <i>in silico</i> , méthode de sous-typage de <i>Salmonella enterica</i> sérotype Enteritidis.....	91
Introduction.....	91
Etude expérimentale .....	91
Conclusions.....	95
<b>Chapitre VI</b> .....	<b>97</b>
<b>SEG : résistance aux antibiotiques, étude retrospective</b> .....	<b>97</b>
<b>(Article 2)</b> .....	<b>97</b>
Abstract .....	98
Transmission de la résistance à l’ampicilline chez <i>Salmonella enterica</i> sérotype Typhimurium à la fin des années 1950 : une étude rétrospective par séquençage entier du génome .....	99
I. Matériels et méthodes .....	101
II. Résultats.....	105
III. Discussion.....	111
Conclusion .....	114
<b>Chapitre VII</b> .....	<b>117</b>
<b>SEG : étude évolutive d’un sérotype rare</b> .....	<b>117</b>
Abstract .....	118
<i>Salmonella enterica</i> sérotype Paratyphi C : une stabilité génomique qui s’étend sur un millénaire.....	119
I. Introduction.....	119
II. Matériel et méthode .....	120
III. Résultat et discussion.....	124

IV. Conclusion.....	127
Conclusion générale.....	129
Bibliographies.....	133
Annexes.....	150

## **Etude bibliographique**



## **Chapitre I**

### ***Salmonella***



## I. Historique

Observé pour la première fois par Karl Joseph Eberth en 1880, dans les ganglions lymphatiques abdominaux et la rate, le bacille de la typhoïde ne fut cultivé qu'en 1884 par Georg Theodor Gaffky. Les salmonelles ont été nommées en l'honneur du vétérinaire Daniel Elmer Salmon qui découvrit en 1885 avec Theobald Smith le bacille du « Hog-choléra de Salmon ».

## II. Taxonomie et nomenclature

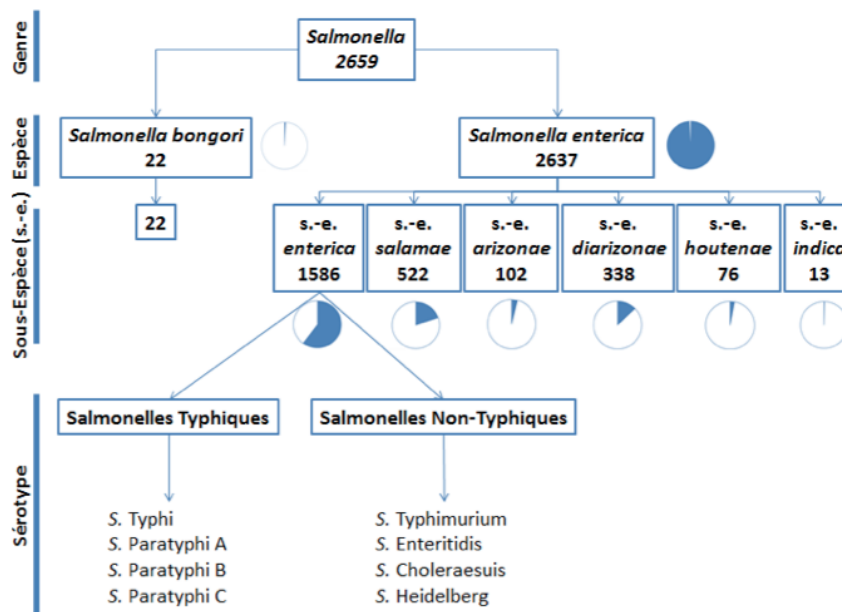
Appartenant à la famille des *Enterobacteriaceae*, le genre *Salmonella* est divisé en deux espèces, *Salmonella bongori*, espèce rare et *Salmonella enterica*, espèce habituelle.<sup>1,2</sup> Cette dernière est elle-même divisée en six sous-espèces puis sérotypes :

- *S. enterica* sous-espèce *enterica* (1586 sérotypes)
- *S. enterica* sous-espèce *salamae* (522 sérotypes)
- *S. enterica* sous-espèce *arizonae* (102 sérotypes)
- *S. enterica* sous-espèce *diarizonae* (338 sérotypes)
- *S. enterica* sous-espèce *houtenae* (76 sérotypes)
- *S. enterica* sous-espèce *indica* (13 sérotypes)

Parmi les *S. enterica* sous-espèce *enterica*, nous pouvons distinguer deux grands groupes :

- les salmonelles typhiques avec *S. enterica* sérotype Typhi et *S. enterica* sérotype Paratyphi A, B et C, les principaux responsables des fièvres typhoïdes et paratyphoïdes respectivement.
- les salmonelles non-typhiques avec *S. enterica* sérotype Typhimurium et *S. enterica* sérotype Enteritidis représentant 70% des salmonelles isolées chez l'homme en France<sup>3</sup>.

L'ensemble des salmonelles comprend actuellement plus de 2600 sérotypes (figure 1).<sup>1,2</sup>



**Figure 1. Classification de *Salmonella*.**

Les cercles indiquent en bleu le pourcentage représenté par chaque espèce et sous-espèce.

### III. Habitat, pathogénie et infectiologie

#### 1. Habitat

Le tube digestif des vertébrés est le réservoir principal des bactéries du genre *Salmonella*. Les salmonelles touchent autant l'homme que les animaux. Cependant, certains sérotypes peuvent être strictement humain ou animal. La sous-espèce *enterica* touche principalement l'homme et les animaux à sang chaud par exemple, mais certains sérotypes comme *S. enterica* sérotype Panama infectent les animaux à sang froid notamment les reptiles<sup>4</sup>. Les sérotypes *S. enterica* sérotype Typhi et Paratyphi sont strictes à l'homme, la contamination se faisant par ingestion d'eau ou d'aliments contaminés par des matières fécales d'origine humaine. Ces sérotypes se retrouvent le plus souvent dans des pays où l'hygiène reste précaire.<sup>5</sup> D'autres sérotypes sont impliqués dans des épidémies touchant la volaille (*S. enterica* sérotype Heidelberg)<sup>6</sup>, les bovins et l'homme (*S. enterica* sérotype Dublin)<sup>7</sup>, les produits laitiers (*S. enterica* sérotype Montevideo)<sup>8</sup>, les œufs (*S. enterica* sérotype Enteritidis)<sup>9</sup>, les porcs et la charcuterie (*S. enterica* sérotype Derby)<sup>10</sup>, ils sont dits ubiquistes.

## 2. Pathogénie et infectiologie

Les caractéristiques essentielles de la pathogénie des salmonelles sont leur capacité à entrer dans les cellules-hôtes et à y demeurer comme parasite intracellulaire facultatif. Une ingestion de  $10^5$  à  $10^{10}$  bactéries est requise pour déclencher l'infection.<sup>11</sup> Avant de pouvoir infecter l'hôte, les cellules de *Salmonella* ingérées doivent pouvoir survivre à l'action bactéricide des lactoperoxydases dans la salive et le pH acide de l'environnement gastrique, afin de pénétrer dans l'intestin grêle où se trouve également des agents enzymatiques. Pour les bacilles ayant résistés à l'acidité gastrique, l'infection commence par une phase d'adhésion aux entérocytes et cellules M des plaques de Peyer. Les salmonelles franchissent ensuite la muqueuse intestinale et envahissent les ganglions mésentériques. S'ensuivent une phase de multiplication dans les macrophages et cellules réticulo-endothéliales, induisant une réponse inflammatoire et la sécrétion d'une entérotoxine diarrhéique.<sup>12</sup>

Les salmonelles possèdent de nombreux facteurs de virulence (adhésines, toxines, invasines, protéines de sécrétion, système de captation du fer) leur permettant à chaque étape de la pathogénie, de s'adapter aux conditions de l'environnement et à la réponse de l'hôte.

### *Îlot de pathogénicité de Salmonella*

Les principaux facteurs de virulence intervenant dans la pathogénicité des *Salmonella* sont localisés sur des régions du chromosome appelées îlots de pathogénicité de *Salmonella* (SPI pour *Salmonella pathogenicity island*). A ce jour, une vingtaine d'îlots ont été décrites.<sup>13</sup> Ces SPI (de 10 à 100 kb) dont les variations structurales, fonctionnelles et la prévalence peuvent varier d'un sérotype à l'autre, jouent un rôle dans la spécificité de l'hôte et leur degré d'invasivité. Parmi ceux-ci, les îlots SPI-1 et SPI-2 codent pour des systèmes de sécrétion de type III appelés T3SS (permet la translocation de protéines de virulence de la bactérie dans une cellule eucaryote) et jouent un rôle majeur durant l'infection. L'îlot SPI-4 code pour un système de sécrétion de type I impliqué dans l'adhésion aux cellules épithéliales. D'autres sont exclusifs de sérotypes, comme SPI-6 à SPI-10 propres au sérotype Typhi. Les sérotypes Typhimurium et Choleraesuis possèdent neuf SPI alors que le sérotype Typhi en possède trois de plus. Les SPI sont généralement porteurs d'éléments mobiles (intégrase, transposase, séquence d'insertion, qui seront développés par la suite), les rendant instable résultant en une excision du chromosome.<sup>13</sup>

### Plasmides de virulence

Certains sérotypes de *Salmonella* appartenant à la sous-espèce I portent un plasmide contenant des gènes de virulence. Ces plasmides (50-90 kb), possèdent tous une région commune de 7,8 kb, spv (plasmide *Salmonella* virulence), nécessaire à la multiplication de la bactérie.<sup>14</sup>

Ce plasmide de virulence n'est pas présent chez tous les sérotypes de *Salmonella*. *Salmonella enterica* sérotype Typhi, Paratyphi A et Paratyphi B par exemple sont dépourvus du plasmide de virulence et notamment de la région spv. Les plasmides sont plus fréquemment retrouvés chez les sérotypes *Salmonella enterica* Typhimurium et Enteritidis. La présence de plasmides de virulence dans des sérotypes adaptés à l'hôte suggère que l'acquisition du plasmide de virulence a pu élargir la gamme d'hôtes de *Salmonella*.<sup>14</sup> Le plasmide de virulence de *Salmonella enterica* sérotype Typhimurium LT2 nommé pSLT, a été décrit comme auto-transmissible.<sup>15</sup>

### Système de captation du fer

Certaines salmonelles sont capables de synthétiser l'entérochéline, ou entérobactine, un sidérophore sécrété dans des conditions limitantes en fer. La corrélation entre la capacité à produire l'entérochéline et la pathogénicité des salmonelles a été peu étudiée. Toutefois le système entérochéline semble jouer un rôle dans la fièvre typhoïde.<sup>16</sup>

### Rôle des toxines

Les salmonelles sont capables de sécréter des toxines responsables de lésions ou symptômes observés chez l'hôte. Le lipide A du lipopolysaccharide est un constituant lipidique d'une endotoxine responsable de la toxicité des bactéries à Gram négatif.<sup>17</sup> L'endotoxine est ainsi responsable de l'activation de phénomènes inflammatoires, la plupart des symptômes de la fièvre typhoïde et du choc septicémique consécutif à une bactériémie.<sup>17</sup> Les entérotoxines jouent probablement un rôle dans les diarrhées des formes gastro-entéritiques.<sup>18</sup>

### Symptômes

Cliniquement, les *Salmonella* non typhiques sont généralement responsables de gastro-entérites, dites salmonelloses. Elles se traduisent par de la fièvre, des diarrhées, des

vomissements et des douleurs abdominales. La période d'incubation est de 12 à 24 heures après ingestion de l'aliment contaminé. Chez la majorité des personnes infectées, la guérison se fait en quelques jours. Cependant pour les nourrissons, les personnes âgées et les personnes immunodéprimées, chez qui l'infection peut être grave voire mortelle, une antibiothérapie est nécessaire.

Les fièvres typhoïdes et paratyphoïdes se manifestent par une fièvre continue, une perte d'appétit, des maux de tête, des douleurs abdominales, ainsi qu'une diarrhée ou constipation, ceci après une période d'incubation pouvant aller de une à trois semaines.

## **IV. Surveillance, épidémiologie et prophylaxie**

### *1. Surveillance*

Le réseau de surveillance de *Salmonella* centralise les résultats de sérotypage des salmonelles. Aujourd'hui, ce réseau a pour objectif principal de confirmer et détecter l'émergence de souches potentiellement problématiques en santé publique notamment au niveau européen mais également à l'échelle internationale. Toutes origines confondues, *S. enterica* sérotype Typhimurium et ses variants monophasiques ainsi que *S. enterica* sérotype Enteritidis demeurent majoritairement isolées. *Salmonella* est depuis de nombreuses années un contaminant microbiologique majeur à l'origine d'épidémie d'origine alimentaire en France et en Europe. L'optimisation de l'évaluation et de la gestion du risque de salmonellose chez l'homme et l'animal implique la collecte de données de qualité, dans un pas de temps adapté.

Le Centre National de Référence des *Escherichia coli*, *Shigella* et *Salmonella* (CNR-ESS, Institut Pasteur) durant laquelle cette thèse s'est déroulée est en charge de la surveillance microbiologique des salmonelloses humaines et reçoit des souches et des comptes rendus d'un réseau de laboratoires de centres hospitaliers (40%) et privés. Au cours de l'année 2016, le CNR-ESS a reçu et typé plus de 10,000 isollements humains en France.<sup>3</sup> En plus de l'activité d'identification des sérotypes, le CNR a également pour mission de suivre l'émergence de souches résistantes aux antibiotiques. L'étude de la résistance aux antibiotiques est réalisée sur un échantillon représentatif des principaux sérotypes de *Salmonella*, tous les ans ou les deux ans suivant les tendances. Les souches étudiées au CNR-

ESS dont les formules antigéniques ne figurent pas dans le schéma WKL sont transmises pour validation au Centre collaborateur OMS (Organisation Mondiale de la Santé) de Référence et de Recherche pour les *Salmonella* (CCOMS) dont les responsables sont ceux du CNR-ESS.

Les principaux collaborateurs du CNR-ESS et participants actifs du réseau de surveillance des *Salmonella* sont:

- l'ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail), en particulier l'unité caractérisation et épidémiologie bactérienne de Maisons-Alfort et du LNR *Salmonella* à Ploufragan. Ils sont spécialisés dans la collecte des souches de *Salmonella* d'origine non humaine. Les souches collectées par le réseau proviennent de trois secteurs de la chaîne agro-alimentaire: (1) "Santé et production animales", (2) "Hygiène des aliments", (3) "Ecosystème naturel".

- Santé publique France (SPF): le CNR-ESS lui envoie des relevés hebdomadaires et des déclarations obligatoire. Des relevés ponctuels sont également effectués en fonction des demandes d'information émanant de SPF, ou lors d'une épidémie.

- l'AFSSA-LERQAP (Agence française de Sécurité Sanitaire des Aliments - Laboratoire d'Etude et de Recherches sur la Qualité des Aliments et sur les Procédés agroalimentaires) qui collecte par l'intermédiaire de laboratoires vétérinaires, d'hygiène alimentaire, ou d'analyses de l'environnement, des souches de *Salmonella* adressées pour sérotypage et des informations sur les souches déjà sérotypées. Le CNR-Salm reçoit chaque année de l'AFSSA-LERQAP le relevé des souches identifiées avec les informations essentielles (origine du prélèvement et département de provenance).

- Le CNR-Salm fait également partie du réseau européen de surveillance Enter-Net devenu ECDC-FWD (European Food and Waterborne Diseases and Zoonoses Network) chargé de la surveillance internationale des infections gastro-intestinales humaines. Il regroupe les différents pays de l'Union européenne ainsi que d'autres pays non membres. Par ailleurs, le CNR-ESS et les laboratoires associés transmettent leurs données trimestriellement via Santé publique France au réseau européen et répond aux demandes d'informations qui leur sont adressées par l'ECDC-FWD ou par d'autres réseaux comme l'EFSA (Autorité Européenne de Sécurité des Aliments).

Chaque année le CNR-ESS adresse également à deux instances internationales, l'OMS et l'OIE (World Organisation for Animal Health), des données sur le nombre de *S. enterica* sérotypes Enteritidis et Typhimurium isolées en France.

## 2. Epidémiologie

### Salmonelles Non Typhiques

Les salmonelles non typhiques ou NTS (*None Typhoidal Salmonella*) sont les infections à *Salmonella* les plus courantes et se produisent dans le monde entier.<sup>19</sup>

En Europe, le nombre de NTS est estimé à 690 cas pour 100 000 habitants. Dans les pays sous-développés, les NTS ont un caractère endémique, plus particulièrement en Afrique subsaharienne, avec des taux élevés chez les enfants de moins de trois ans et les patients infectés par le VIH.<sup>20</sup> En revanche, les infections liées aux NTS sont moins fréquentes en Asie. Aux États-Unis, les données fournies par le FoodNet montrent que les infections liées aux NTS étaient les plus fréquentes dans cette région, avec une incidence de 17,6 cas par 100 000 habitants par an.<sup>20</sup> L'une des plus récente épidémie aux États-Unis était liée à la contamination d'œufs par *S. enterica* sérotype Enteritidis.<sup>21</sup> En France, l'épidémie nationale liée aux laits pour nourrissons distribués à l'échelon International en décembre 2017 était dû au sérotype Agona.<sup>22</sup>

### Salmonelles Typhiques

Les salmonelles typhiques à l'origine des fièvres typhoïdes sont quant à elles associées à un taux élevé de mortalité survenant principalement dans les pays sous-développés.<sup>23</sup> En 2000, la fièvre typhoïde était estimée à 22 millions de cas, entraînant 200 000 décès dans le monde, principalement dans les pays sous-développés.<sup>24</sup> Le nombre de cas liés à la fièvre typhoïde aux États-Unis et dans certains pays européens est faible. La plupart des cas signalés dans ces pays sont liés aux voyages, notamment avec des voyageurs revenant d'Afrique, d'Inde ou du Pakistan.<sup>23</sup> Depuis 2003, 100 à 250 cas d'infections à *S. Enterica* sérotype Typhi, isolées en France proviennent de cas importés d'Afrique, d'Amérique du Sud et du sous-continent Indien et sont répertoriés chaque année au CNR des *Salmonella*. L'incidence annuelle des cas en France est environ de 0,2 pour 100 000 habitants. Elle est en constante diminution depuis le dernier pic épidémique observé en 1949.<sup>25</sup>

### 3. Prophylaxie

#### Traitement

Une antibiothérapie est prescrite pour les personnes âgées, les nourrissons ou les personnes immuno-déprimées dans le cas des gastro-entérites. Pour les salmonelloses liées à la fièvre typhoïde une antibiothérapie appropriée abaisse le risque de mortalité à moins de 1%, mais de nombreuses souches deviennent de plus en plus résistantes aux antibiotiques (cf chapitre III). Un vaccin anti typhoïdique, ne nécessitant qu'une seule injection, peut être administré aux voyageurs se rendant dans des régions à risque. L'effet protecteur dure trois ans et le taux de protection en zone d'endémie est de 60%.

#### Prevention

Les principales préconisations pour lutter contre le risque de salmonellose sont une bonne cuisson des aliments (65°C pendant 5 à 6 minutes pour les viandes). Il est conseillé de conserver les œufs et les préparations à base d'œufs sans cuisson au réfrigérateur. Une bonne hygiène est également de rigueur (lavage des mains régulier). La dissémination des bactéries peut être enrayée par : une distribution d'eau de qualité, le traitement des eaux usées, le contrôle des zones de récolte des coquillages, la pasteurisation des aliments.

## V. Identification des espèces et sous-espèces

Les méthodes de typages et sous-typages sont nombreuses et se distinguent par leur pouvoir discriminant et leur reproductibilité. Le coût et le temps nécessaire à la réalisation du typage sont également deux critères importants. L'identification de l'espèce et de la sous-espèce de *Salmonella* est basée sur des tests phénotypiques et si besoin moléculaires. Les techniques présentées ci-dessous en sont une liste non exhaustive.

### 1. Typage des *Salmonella*

#### *Caractéristiques phénotypiques et biochimiques*

Les *Salmonella* sont des bacilles à coloration de Gram négative, aéro-anaérobie facultatifs, pour la plupart mobiles grâce à une ciliature péritriche. Généralement prototrophes, elles



poussent sur milieu ordinaire, fermentent le glucose en produisant du gaz, réduisent les nitrates en nitrites. Parmi les principales caractéristiques biochimiques permettant leur identification au laboratoire, se trouve également le caractère oxydase négatif, lactose négatif et sulfure d'hydrogène (H<sub>2</sub>S) positif.

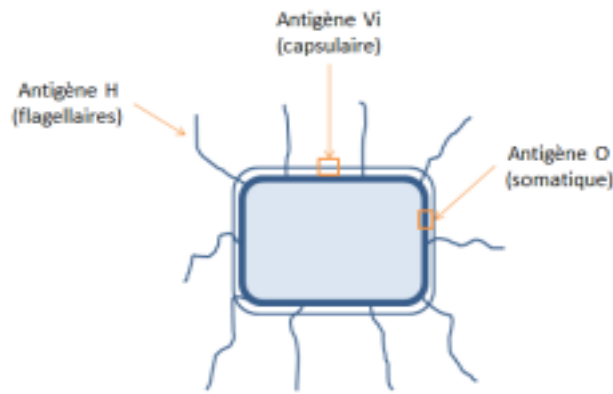
### Sérotypage

C'est en mélangeant des bacilles typhiques avec un sérum d'individus infectés par la fièvre typhoïde que Georges-Fernand Widal mit en évidence en 1896 leur diversité antigénique. Ce sont ces mêmes antigènes qui vont permettre le sérotypage des *Salmonella*. En effet, cette technique est basée sur la recherche de trois antigènes selon le schéma de White (1926), reprise par Kauffman en 1934, puis L. Le Minor en 1965, M.Y. Popoff en 1989 et enfin P.A.D. Grimont et F.-X. Weill en 2003.<sup>1,2</sup> Les trois antigènes sont présentés ci-après et dans la figure 2 :

- l'antigène O ou somatique est spécifique du lipopolysaccharide (membrane externe de la bactérie). Aujourd'hui, 46 antigènes O ont été décrits, dont certains sont communs à plusieurs sérotypes. Une région *rfa* contenant les gènes de structure des glycosyltransférases est impliquée dans la synthèse du « core » de l'antigène O. La région *rfb* contient les gènes nécessaires à la synthèse des monomères oligosaccharidiques, eux même polymérisés par le gène *wzy*. Des mutations au niveau de ces régions induisent des phénotypes dits « rough » dont le sérotype n'est pas sérotypable.<sup>26</sup>

- les antigènes H1 et H2 ou flagellaires, au nombre de 85, sont habituellement diphasiques chez les *Salmonella*. Les gènes codant pour ces deux flagellines s'expriment selon le système d'inversion de phase : aléatoirement et toutes les 1,000 à 10,000 générations, le gène qui n'est pas exprimé s'exprime et l'autre cesse de s'exprimer. L'expression de l'opéron *flj* composé des gènes *fljA* et *fljB* réprime l'expression du gène *fliC*.<sup>27</sup>

- l'antigène Vi ou capsulaire contrôlé par les gènes *viA* et *viB* n'est connu que chez trois sérotypes : *S. enterica* sérotype Typhi, *S. enterica* sérotype Paratyphi C et *S. enterica* sérotype Dublin.<sup>28</sup>



**Figure 2. Composition antigénique de *Salmonella*.**

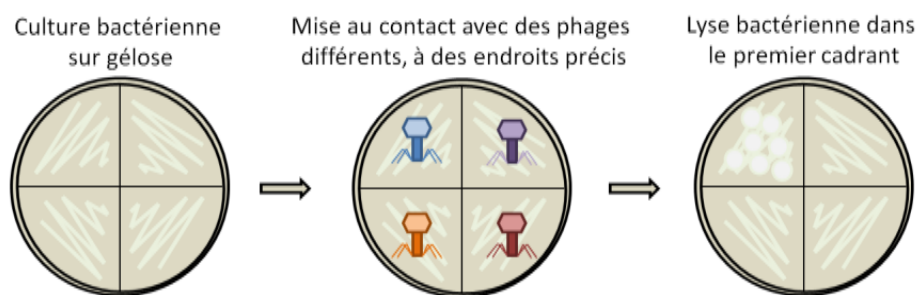
Les trois antigènes ciblés lors du sérotypage sont indiqués en orange.

La combinaison de tous ces antigènes permet de définir et d'identifier les différents sérotypes. *S. enterica* sérotype Typhimurium est par exemple défini par la formule antigénique [1],4[5],12:i:1,2 ou [1],4[5],12:i:- pour ses variants monophasiques. Plus de 150 sérums ciblant les antigènes O, H et Vi sont actuellement commercialisés. Cette méthode, longue, nécessite entre 3 à 6 jours pour un rendu de résultat mais également beaucoup de main d'œuvre. De plus, certaines souches dites « rough » sont dépourvues de la chaîne somatique O et ne sont pas sérotypables. Le faible pouvoir discriminant de cette technique ne permet pas de réaliser un sous-typage des *Salmonella*.

## 2. Sous-typage des *Salmonella*

### Lysotypie

Terme inventé par P. Nicolle en 1950, la lysotypie est la plus ancienne méthode de sous-typage permettant la caractérisation de différents bacilles, notamment ceux non agglutinables par la méthode du sérotypage, grâce à la lyse bactériophagique.<sup>29</sup> La bactérie cultivée est mise au contact avec le bactériophage. En fonction de sa sensibilité vis-à-vis de ce dernier, un halo indiquant la lyse des bactéries va se former (figure 3). Chaque bactériophage possédant un spectre d'activité particulier, ceci permet un regroupement des bactéries entre elles. L'un des premiers sérotypes typé par lysotypie est le sérotype Typhi, le subdivisant en 106 lysotypes.<sup>28,30</sup> Bien que standardisée à l'échelle internationale, 70% des souches se répartissent en trois lysotypes seulement (A, E1 et C1).<sup>31</sup>



**Figure 3. Illustration d'une lyse bactérienne**

### *Ribotypage*

Profitant de l'essor des études portant sur les ARNs ribosomiques (ARNr), le ribotypage est une technique moléculaire utilisant le polymorphisme des fragments de restriction de l'ADN codant pour les ARNr. Cette séquence d'ADN est unique à chaque bactérie, stable et donc reproductible.<sup>32,33</sup> L'ADN bactérien est extrait et digéré par des enzymes de restriction spécifiques, coupant l'ADN et générant des fragments de longueurs différentes. Les fragments obtenus sont séparés en fonction de leur taille sur un gel d'électrophorèse et transférés sur une membrane en nylon. L'ADN est hybridé à des sondes nucléiques d'ARNr 16 S ou 23 S. Le ribotypage consiste donc en la localisation de séquences d'ADN codant pour les ARNr 16 S et 23 S (figure 4). Après révélation du marquage, des profils appelés ribotypes sont obtenus. Contrairement à la lysotypie, cette méthode plus discriminante montre une plus grande diversité pour le sérotype Typhi, au sein des lysotypes A et E1.<sup>34</sup> Cependant, pour d'autres sérotypes (Typhimurium, Paratyphi B...) les résultats obtenus ne permettent pas de distinguer le sérotype.<sup>35,36</sup> De plus, le nombre de bandes parfois trop nombreuses rend son interprétation difficile.<sup>37,38</sup>

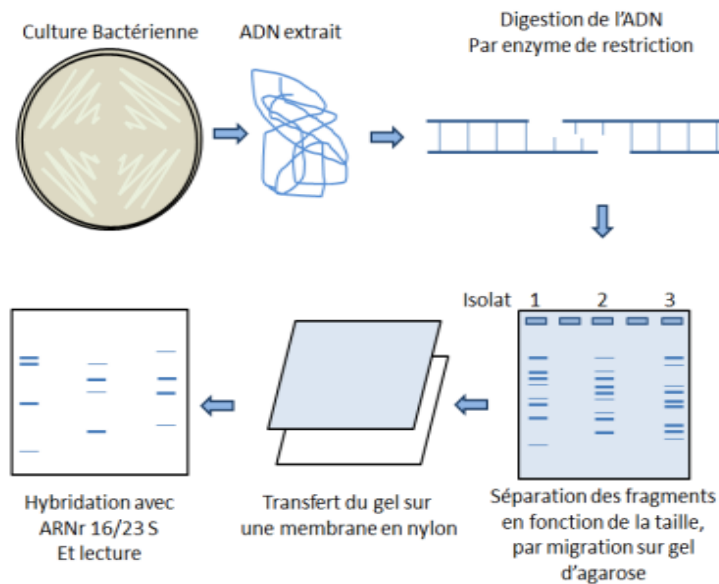
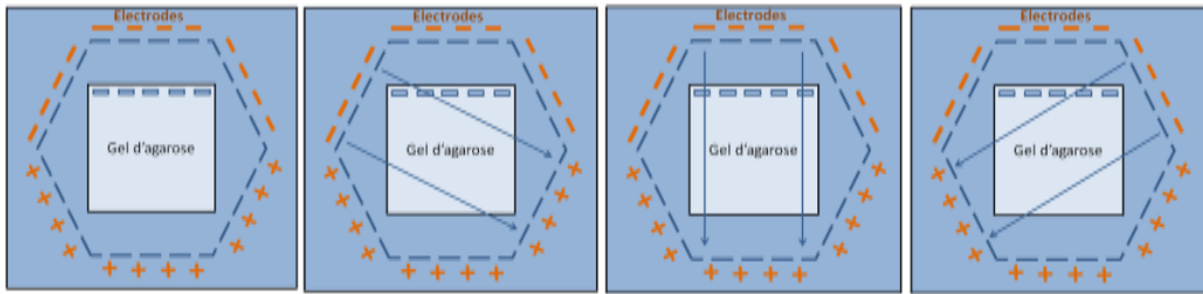


Figure 4. Protocole du ribotypage.

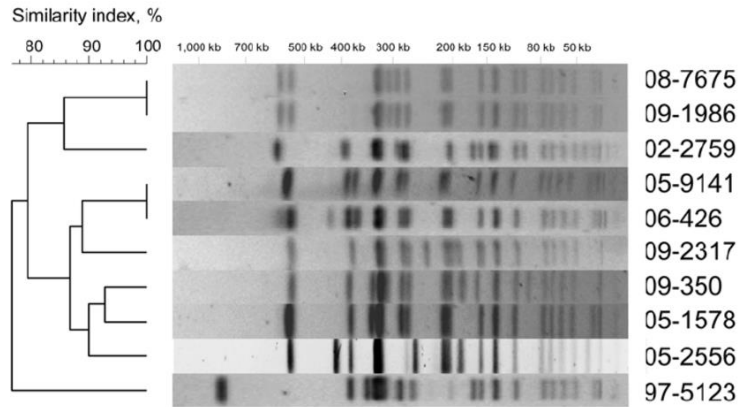
### Electrophorèse en champs pulsé (ECP)

Développée par David C. Schwartz et Charles Cantor en 1984<sup>39</sup>, l'électrophorèse en champ pulsé (ECP) utilise une endonucléase de restriction reconnaissant des sites rares (généralement *Xba*I, *Spe*I ou *Bln*I), générant des fragments d'ADN de grande taille (plus de 50 kb). La séparation de ces molécules d'ADN se fait par migration sur gel, déposé dans une cuve contenant des électrodes placées de façon hexagonale. Ainsi, l'ECP utilise un voltage alternatif pouvant aller dans trois directions différentes (figure 5), permettant une meilleure résolution et interprétation des résultats.<sup>39</sup> L'analyse des différents profils se fait grâce à des logiciels tels que BioNumerics (Applied Maths, Sint-Martens-Latem, Belgique) (figure 6). Méthode standardisée<sup>40</sup> (<http://www.cdc.gov/pulsenet>), son pouvoir discriminant est très sollicité lors d'épidémies à l'échelle nationale et internationale.<sup>41,42</sup> Néanmoins, cette technique est longue (de 4 à 5 jours) et lourde tant au niveau de l'appareillage qu'au niveau de sa réalisation. De plus, elle ne permet pas la différenciation de certains clones prédominants tel que le clone DT104 (voir chapitre III), dont 73% des souches sont représentés par un même profil en ECP.<sup>43</sup> Il est donc nécessaire de développer des techniques à haut débit plus discriminantes.



**Figure 5. Electrophorèse en champ pulsé avec voltage alternatif.**

Le sens du courant est indiqué par les flèches bleues, allant de l'électrode négative (-) à l'électrode positive (+).



**Figure 6. Profils de dix souches de *Salmonella enterica* sérotype Typhi obtenus à partir de l'ECP.**

Le dendrogramme généré avec BioNumerics version 3.5 montre le regroupement des souches sur la base des profils ECP. Accou-Demartin *et al.* 2011.

## **Chapitre II**

### **Typage et sous-typage moléculaire des *Salmonella***

#### **à haut-débit**

Dans le chapitre précédent, nous avons vu que le sérotypage traditionnel selon White-Kauffmann-Le Minor était la méthode de référence pour le typage de *Salmonella*. Cependant, lors d'épidémies il est nécessaire de pouvoir aller plus loin et différencier deux souches d'un même sérotype. Malgré le développement de méthodes plus discriminantes basées sur l'analyse génétique des bactéries telle que l'ECP, cette méthode reste lourde et ne permet pas une analyse à plus haut-débit.

## I. Multilocus VNTR analysis

Cette technique vise à déterminer la longueur de séquences courtes répétées : les répétitions en tandem polymorphes (ou VNTR pour *Variable Number of Tandem Repeat*). Le nombre de répétitions d'une souche à l'autre étant extrêmement variable, la discrimination entre deux bactéries proches peut être établie.<sup>44</sup> Le Multi Locus VNTR Analysis (MLVA) repose sur l'amplification par PCR des VNTR à l'aide d'amorces spécifiques de ces régions, suivi d'une vérification de la taille des amplicons et du nombre de répétitions par migration sur gel électrophorèse et électrophorèse par capillarité. La lecture se fait sur un électrophérogramme, chaque pic représentant un allèle précis (figure 7). Cette technique est discriminante,<sup>44,45</sup> notamment pour le sérotype Typhimurium,<sup>46,47</sup> peu coûteuse, rapide à réaliser et adaptable à haut débit. Cependant, la MLVA n'est pas universelle et un schéma de typage propre à chaque sérotype doit être mis en place et validé.

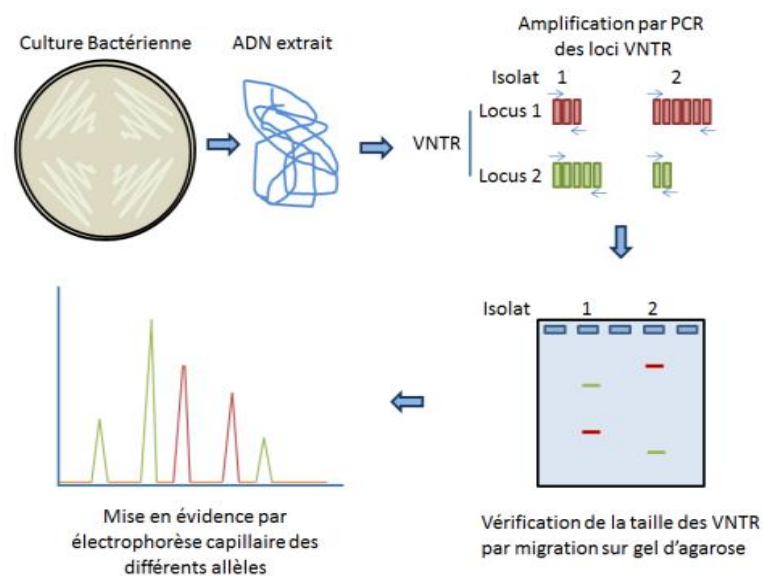


Figure 7. Protocole de la MLVA.

## II. Multilocus Sequence Typing

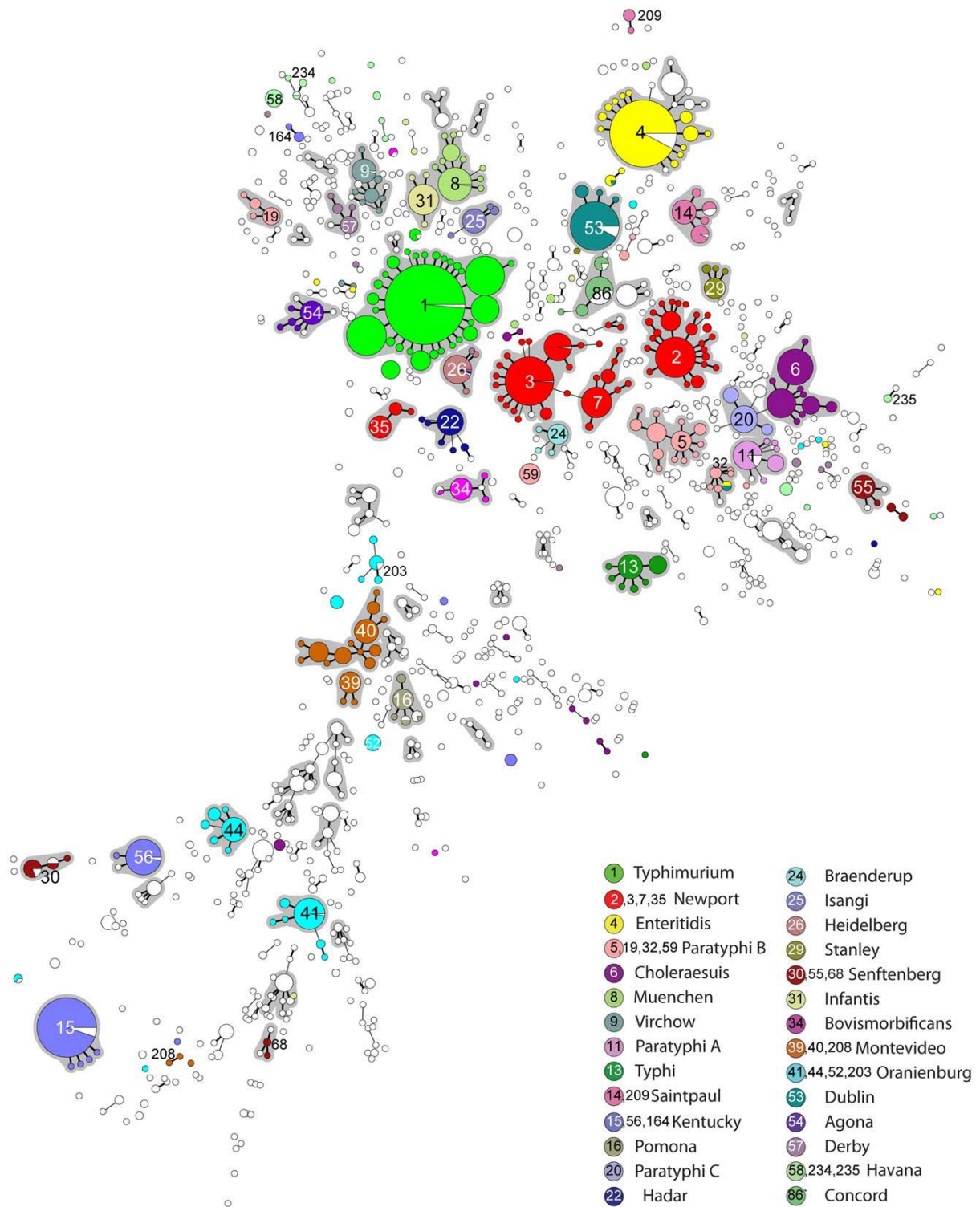
La méthode du typage par séquençage de gènes multiple ou Multi Locus Sequence Typing (MLST) dérive de l'électrophorèse des isoenzymes ou *multi-locus enzyme electrophoresis* (MLEE) et a été développée dans le but de mener des études phylogénétiques et épidémiologiques chez les microorganismes.<sup>48</sup> Elle est basée sur la détection de sept gènes de ménage (*aroC, dnaN, hemD, hisD, purE, sucA et thrA*), choisis pour leur faible taux de mutations (gènes conservés, peu soumis aux pressions de sélection) et leur stabilité dans le temps.<sup>48</sup> L'ADN bactérien est extrait et une amplification des sept gènes d'environ 500 pb est effectuée. Les amplicons obtenus sont séquencés puis comparés à la base de données MLST de Warwick. La combinaison des différents profils alléliques de chaque gène permet de définir un séquençotype (ST pour *Sequence Type*). Il existe quatre schémas MLST pour *Salmonella*<sup>49,50,51,52</sup> dont le plus connu est celui décrit par Kidgell *et al.*, utilisé par l'ancien site internet de Warwick.

Tous les ST génétiquement proches sont regroupés en « eBurstGroups » (eBGs). Un eBG est défini comme un groupe comprenant deux ou plusieurs ST qui partagent six des sept allèles définissant le ST et peut contenir plus d'un sérotype.<sup>53</sup> Il existe à ce jour (octobre 2017) plus de 1000 ST et 400 eBG différents

(<http://enterobase.warwick.ac.uk/species/index/senterica>). Chez *Salmonella*, les deux plus grands sérotypes *S. enterica* sérotype Typhimurium et *S. enterica* sérotype Enteritidis sont largement représentés par le ST19 (ST majoritaire de l'eBG1) et ST11 (présent dans l'eBG4) respectivement (figure 8).<sup>53</sup>

Standardisé, le MLST permet une reproductibilité au sein et entre les laboratoires. Validée sur 4257 souches appartenant à 554 sérotypes<sup>53</sup>, cette méthode corrèle très bien aux données de sérotypage tout en étant plus discriminante que celle-ci grâce à la nature polyphylétique de la majorité des sérotypes. Cependant pour les deux principaux sérotypes cités ci-dessus, la technique MLST ne permet pas dans la grande majorité des cas d'effectuer un sous-typage du fait d'une grande homogénéité génétique au sein de ces sérotypes.<sup>53,49</sup>





**Figure 8. Arbre couvrant de poids minimum ou MST (Minimum Spanning Tree, voir chapitre IV) des profils MLST de 4,257 isolats de *S. enterica* sous-espèce *enterica*.**

Chaque cercle correspond à un des 1,095 STs trouvés en 2012, dont la taille est proportionnelle au nombre d'isolats. Les ST isolés sont dispersés partout au sein de l'arbre.

Achtman *et al.*, 2012.

### III. Polymorphisme des régions CRISPR

En 1987, le japonais Ishino mit en évidence chez *Escherichia coli*<sup>54</sup>, l'existence de séquences d'ADN répétées pouvant se lire dans les deux sens, appelées palindromes. Rebaptisées vingt ans plus tard CRISPR pour *Clustered Regularly Interspaced Short Palindromic Repeats*, ces séquences d'ADN sont généralement associées aux gènes *cas*, le tout formant un système d'immunité adaptative, retrouvé également chez d'autres bactéries.<sup>55</sup>

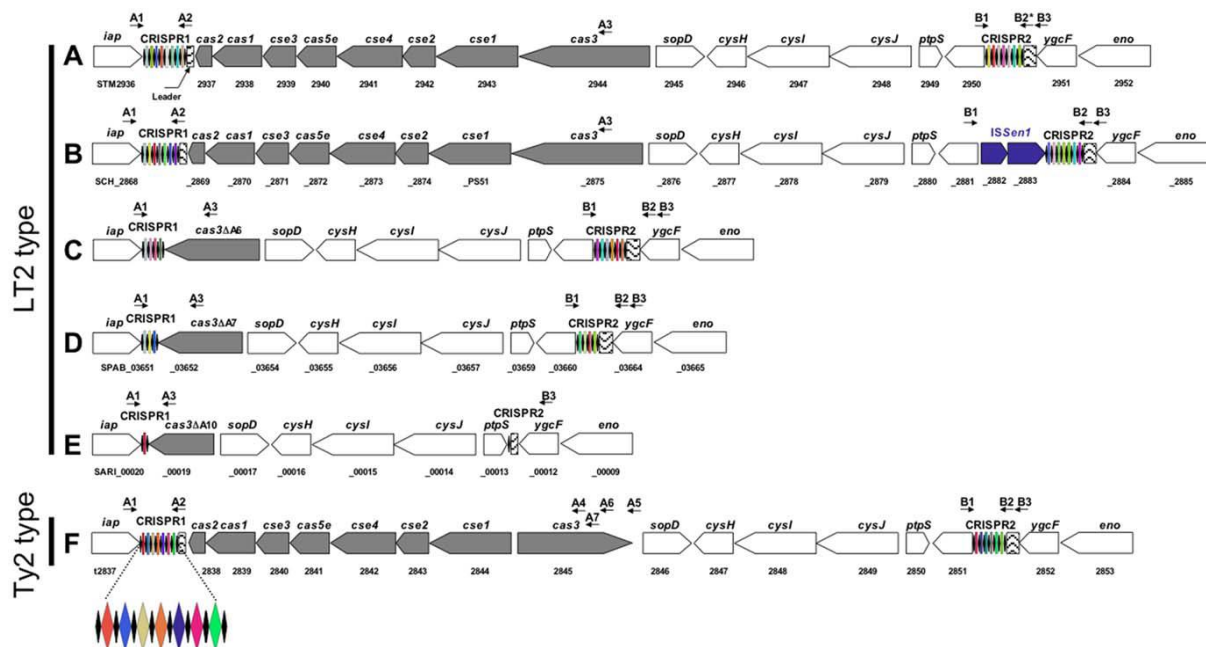
#### 1. Composition du CRISPR

Le système d'immunité CRISPR-Cas (CRISPR-associated genes) est formé :

- des régions CRISPR, au nombre de deux chez *Salmonella* : CRISPR1 et CRISPR2. Ces deux loci sont séparés par moins de 20 kb. Chacun d'eux est caractérisé par la succession de séquences directes répétées ou *direct repeats* (DR) d'une longueur moyenne de 29 pb et régulièrement intercalées par des séquences non codantes de 32 pb appelées espaceurs (27 en moyenne par souche). Le CRISPR1 se trouve en aval du gène *iap* et le CRISPR2 en amont du gène *ygcF*. Chaque région CRISPR possède une séquence riche en A-T dite « *leader* » qui lui est adjacente (figure 9).

- entre les deux loci CRISPR se trouve la machinerie cas et ses gènes : *cas1*, *cas2* (présents dans tous les mécanismes CRISPR-cas), *cse3*, *cas5e*, *cse4*, *cse2* et *cas3*. Ces gènes *cas* permettent soit la séparation des deux brins d'ADN ou d'ARN (les hélicases), soit la coupure des acides nucléiques en petits fragments (les endonucléases). S'ensuit les gènes *sopD* (codant pour une protéine de sécrétion), *cysH* (codant pour une phosphoadenosine phosphosulfate reductase), *cysI* et *cysJ* (codant pour une sous-unité de sulfite reductase), *ptpS* (codant pour une pyruvyl tetrahydrobioptérine synthase) et une ORF codant pour une hydrolase métallo-dépendante (figure 9).<sup>56</sup>

Plusieurs structures de la région CRISPR sont observables chez *Salmonella*. Cependant la structure A est la plus présente, notamment chez *S. enterica* sérotype Typhimurium (figure 9).<sup>57</sup>



**Figure 9. Différentes structures du système CRISPR/cas chez *S. enterica* et *S. bongori*.**

Les deux régions CRISPR (CRISPR1 et CRISPR2) sont présentes chez les deux espèces. Le système CRISPR-cas, et ses gènes *cas2*, *cas1*, *cse3*, *cas5e*, *cse4*, *cse2* et *cas3* sont localisés entre les deux régions CRISPR. La structure la plus fréquente est la A, représentée par la souche *S. enterica* sérotype Typhimurium LT2. Les structures B à E sont représentées par les souches *S. enterica* sérotype Choleraesuis SC-B67, Javiana GA\_MM04042433, Paratyphi B SPB7 et *S. enterica* sous-espèce *arizonae* sérotype 62:z4,z33:- souche CDC346-86, respectivement. La structure F est représentée par la souche *S. enterica* sérotype Typhi Ty2. Les losanges noirs représentent les séquences répétées (DR), ceux en couleur les espaceurs.

Fabre *et al.*, 2012.

## 2. Mécanisme immunitaire du système CRISPR-Cas

Lorsqu'un corps étranger envahit la bactérie, il lui injecte son ADN. Le système CRISPR-Cas permet de lutter contre des éléments génétiques invasifs tels que les bactériophages et les plasmides. Pour une bactérie dépourvue du système CRISPR-cas, l'ADN phagique se multiplie et lyse la bactérie. Le mécanisme immunitaire se déroule en trois étapes (figure 10) :

- l'acquisition ou adaptation : insertion de nouveaux espaceurs dans la région CRISPR.
- l'expression : transcription du locus CRISPR.
- l'interférence : détection et dégradation de l'ADN invasif par l'ARN CRISPR et les protéines Cas.

### Phase d'acquisition

La première phase est celle de l'acquisition ou adaptation. Lors de cette étape, la bactérie peut être envahie par un bactériophage qui va lui injecter son ADN ou un plasmide. L'ADN

étranger contient une séquence cible appelée proto-espaceur avec à ses côtés, une courte séquence appelée PAM pour *proto-spacer adjacent motif* (ou motif adjacent au proto-espaceur). Ce motif PAM est spécifique à l'ADN envahisseur et absent de la bactérie envahie. Une mutation de ce motif peut empêcher l'immunité bactérienne et favoriser l'invasion. Le proto-espaceur est reconnu puis clivé par le complexe protéique Cas1 et Cas2. Les gènes codant pour les protéines Cas1 et Cas2 sont extrêmement conservés, d'où leur présence dans tous les types de CRISPR-cas. Le complexe Cas1-Cas2 permet également l'intégration du nouvel espaceur dans la région CRISPR, entre la séquence leader et la première séquence répétée (DR). Dans un même temps, la séquence répétée adjacente au nouvel espaceur est dupliquée. Les CRISPR sont donc constitués d'espaceurs correspondant à l'accumulation d'ADN étranger, entraînant une résistance spécifique de la bactérie vis-à-vis de l'être envahisseur.

### *Phase d'expression*

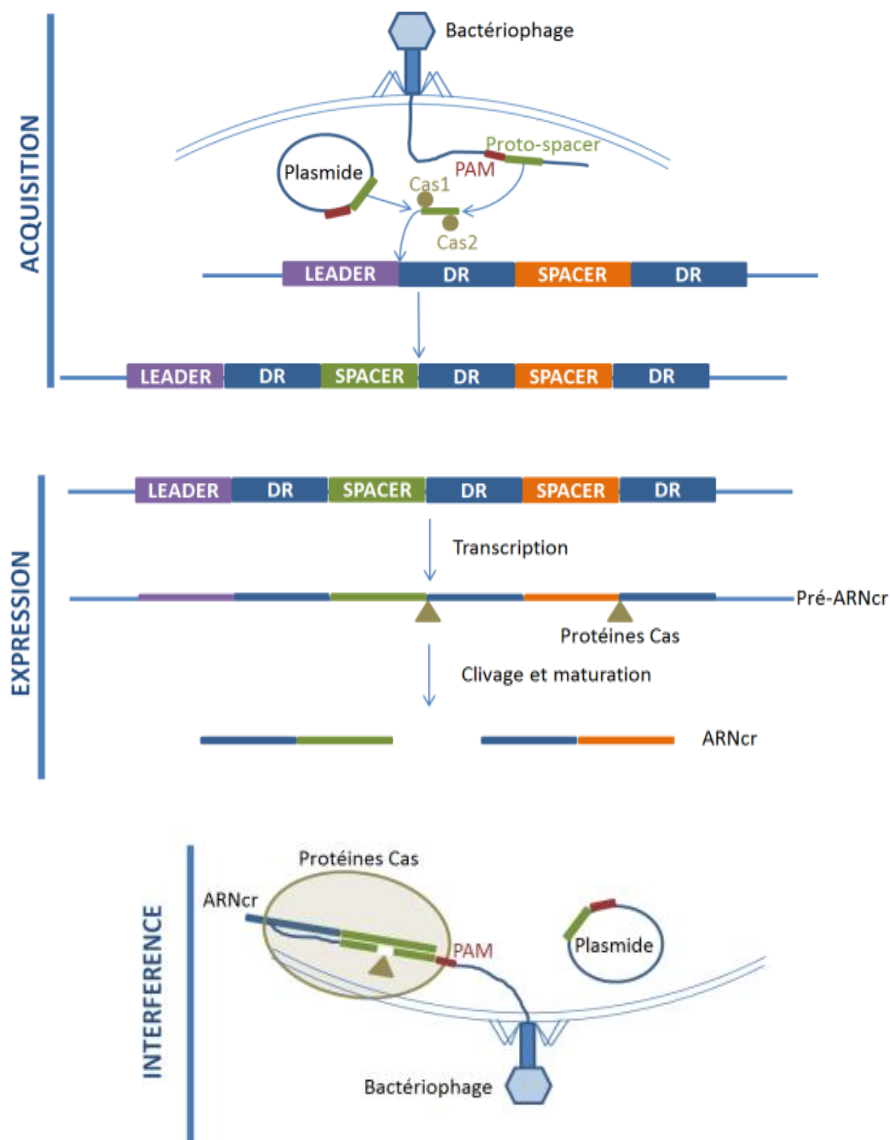
La deuxième étape est la phase d'expression. La séquence CRISPR obtenue lors de la phase d'acquisition est transcrite en une longue séquence primaire appelée pré-ARNcr (pour ARN CRISPR). Les protéines Cas viennent cliver le pré-ARNcr en petits fragments qui vont maturer en ARNcr. Chacun d'entre eux est spécifique et complémentaire d'un ADN invasif. Les protéines Cas impliquées dans le clivage et la maturation des ARNcr diffèrent d'un système CRISPR-cas à un autre. Il existe trois grandes classes de système CRISPR-Cas :

- Type I : utilise un ensemble de protéines Cas (Cas3, Cas4, Cas5, Cas6, Cas7) appelé cascade.
- Type II : seule la protéine Cas9 est impliquée.
- Type III : seule la protéine Cas6 est impliquée.

### *Phase d'interférence*

La dernière étape est celle de l'interférence. Les ARNcr sont couplés aux protéines Cas les guidant vers l'ADN étranger. Afin d'éviter une réaction auto-immune, c'est-à-dire la destruction de l'ADN de la bactérie hôte, celle-ci va reconnaître grâce aux protéines Cas, le motif PAM. Tant que le motif PAM est absent, les protéines Cas n'hydrolysent pas le brin d'ADN, évitant ainsi un clivage de l'ADN bactérien et permettant également de limiter le type et le nombre d'espaceurs pouvant être acquis par la bactérie. Une fois reconnu, le

complexe ARNcr/protéines Cas vient s'hybrider au brin d'ADN phagique/plasmidique, puis l'hydrolysent.<sup>58,55</sup>



**Figure 10. Mécanisme immunitaire du système CRISPR-cas.**

Le schéma montre les trois étapes, acquisition, expression et interférence, du système immunitaire CRISPR-cas.

Au final, les CRISPR sont composés de séquences répétées, intercalées par des séquences dérivant d'ADN étrangers.<sup>59</sup>

Le contenu en espaceurs d'une collection globale de 1000 souches représentatives de plus de 200 sérotypes dont les plus fréquemment impliqués en pathologie humaine, comme les sérotypes Typhimurium et Enteritidis, a ainsi été répertorié. Plus de 7000 espaceurs

distribués sur les deux régions CRISPR ont été mis en évidence (de 500 pb à 4 kb par région). Le contenu en espaceurs de chaque souche était parfaitement corrélé au sérotype et même à la structure génétique des populations obtenue par MLST.<sup>57</sup> Des micro-variations de ce contenu en espaceurs (duplication, SNP [Single Nucleotide Polymorphism substitution d'un nucléotide par un autre], variant VNTR, perte et acquisition d'espaceurs) permet de sous-typer les souches au sein du sérotype monophylétique le plus prévalent, *S. enterica* sérotype Typhimurium avec un pouvoir discriminant équivalent voire meilleur que l'ECP.<sup>57</sup>

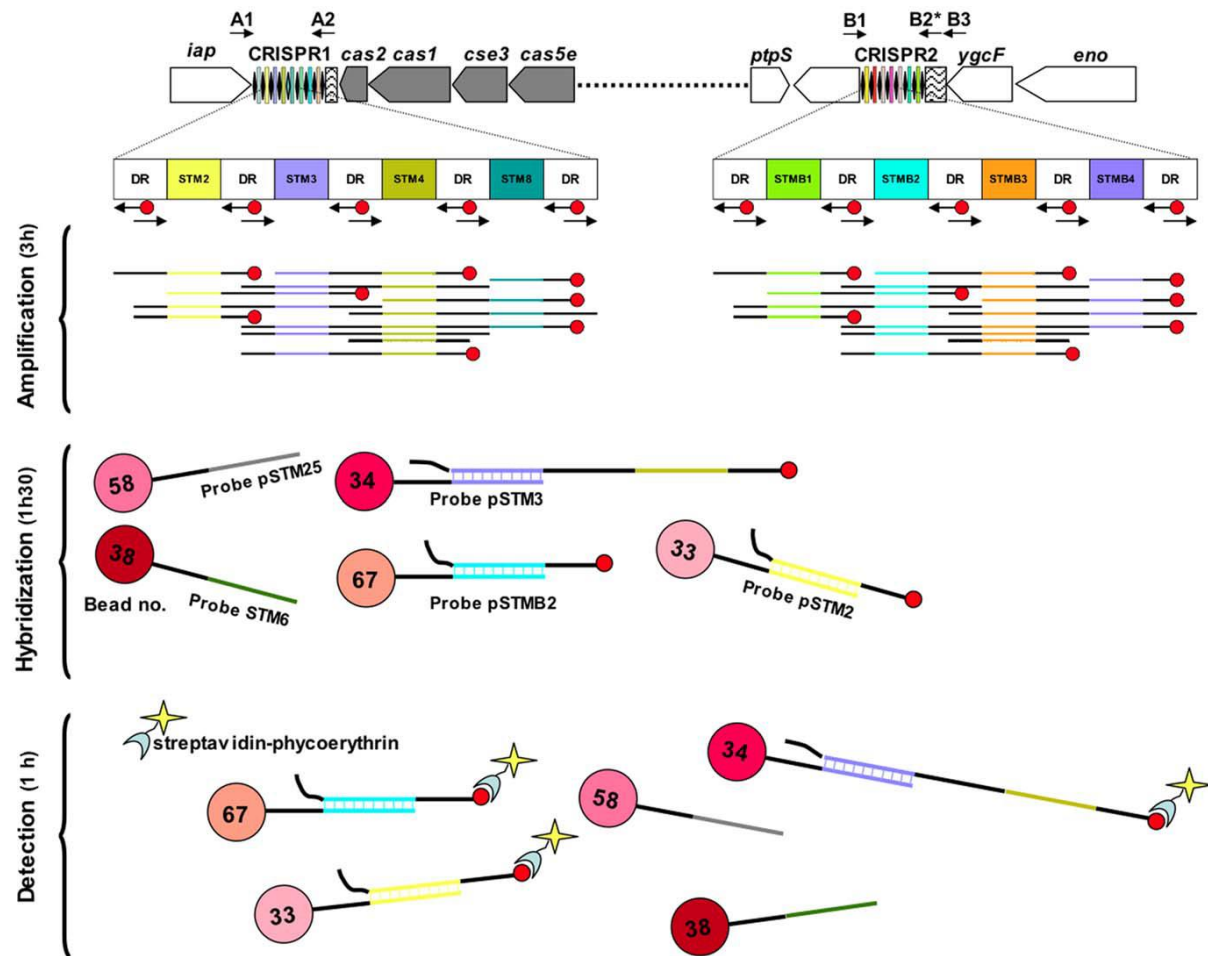
### 3. CRISPOL Luminex®

Développée pour *S. enterica* sérotype Typhimurium par l'Unité des Bactéries Pathogènes Entériques, en collaboration avec l'équipe Infection Génétique Evolution des Pathogènes Emergents (IGEPE), le CRISPOL (CRISPr POLymorphisms) repose sur la technologie Luminex® et sur 72 sondes (de 25 à 32 nucléotides) spécifiques des espaceurs de ce sérotype.<sup>57</sup> Chacune de ces sondes est couplées à une bille dont l'adresse spectrale est unique. La technique se déroule en plusieurs étapes (figure 11) :

- **l'amplification** : l'ADN est extrait par choc thermique puis les deux régions CRISPR sont amplifiées par PCR, en utilisant un couple d'amorces (dont un est biotinylé) ciblant les régions répétées DR.
- **l'hybridation** : les produits PCR sont ensuite hybridés aux 72 sondes.
- **la détection** : le couple produit PCR/sondes est incubé avec de la streptavidine-phycoerythrine venant se fixer à la biotine. Ces derniers émettent un signal lumineux et sont détectés par la plateforme Luminex® (appareil BioPlex® 100, Bio-Rad, Marnes-la-Coquette, France).

Méthode rapide (environ 6h) et peu coûteuse, ses résultats corrélaient parfaitement avec ceux du MLST. Chaque schéma CRISPR obtenu est appelé CRISPOL Type (CT). Il en existe aujourd'hui près de 990 pour 13,024 souches. La perte ou duplication des espaceurs, ainsi que les variations nucléotidiques pour certains espaceurs (STM03 var1, variant de l'espaceur STM03, diffère d'un nucléotide par exemple), sont les principaux déterminants permettant une discrimination des souches au sein d'un même sérotype. Elle est utilisée en routine au Centre National de Référence mais également au laboratoire de l'ANSES Maisons-Alfort, responsable de la surveillance des salmonelloses non-humaines, pour la détection des épidémies liées à *S. enterica* sérotype Typhimurium.

Cependant, seulement adaptée pour le sérotype Typhimurium, il est nécessaire de développer une méthode universelle pour l'ensemble des 2,600 sérotypes de *Salmonella*. En effet, la création d'une base de données nécessite à ce jour une vérification des CT par amplification puis séquençage Sanger (voir Chapitre IV) des deux régions CRISPR, ainsi qu'une implémentation des nouveaux espaceurs à la base de données. Toutes ces étapes restent manuelles et sont risquées d'erreurs. Cette limite technologique doit être remplacée par une méthode rapide, automatisable, fiable et utilisable en routine au CNR.



**Figure 11. Protocole du CRISPOL Luminex® développé pour *S. enterica* sérotype Typhimurium.**

Le temps estimé pour chaque étape est basé sur 65 souches d'une plaque 96 puits. L'amplification comprend la préparation des thermolysats à la migration sur gel d'électrophorèse des produits PCR.

Fabre *et al.*, 2012.

## **Chapitre III**

### **Génétique de la résistance aux antibiotiques chez**

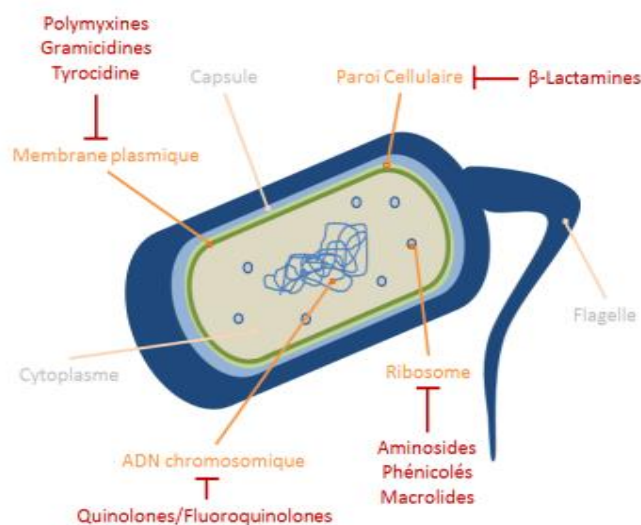
### ***Salmonella***



L'antibiothérapie est généralement préconisée pour les personnes âgées, immunodéprimées ou les nourrissons atteints de salmonellose. Cependant, une mauvaise utilisation de ces derniers peut entraîner une résistance de la bactérie vis-à-vis du traitement.

## I. Cibles des antibiotiques

Les antibiotiques sont des substances ayant une action sur la bactérie, d'ordre bactéricide (qui tue la bactérie) ou bactériostatique (qui empêche les bactéries de se multiplier). Afin d'agir sur les bactéries, l'antibiotique doit posséder et pouvoir accéder à une cible spécifique (paroi bactérienne, membrane plasmidique, ribosome ou ADN). La figure 12 montre les principales cibles des grandes familles d'antibiotiques.<sup>60</sup>



**Figure 12. Principales cibles des grandes familles d'antibiotiques.**

En rouge sont indiqués les grandes familles d'antibiotiques et leur activité inhibitrice (barres rouges), ainsi que leurs cibles bactérienne (en orange). En gris clair, les autres structures bactérienne.

## II. La résistance bactérienne

La résistance bactérienne peut être définie comme la capacité d'une bactérie à continuer à croître et survivre en présence de l'antibiotique.

### 1. Généralités

Face aux antibiotiques, la bactérie possède deux types de résistance : la résistance naturelle et la résistance acquise.

### *La résistance naturelle ou dite intrinsèque*

Cette résistance présente chez toutes les souches d'une même espèce ou d'un même genre est stable et constante à l'intérieur d'un taxon.<sup>61</sup> La résistance peut être transmise à la descendance (transfert verticale).<sup>61</sup>

### *La résistance acquise*

D'autres bactéries, soumises à des pressions de sélection acquièrent une résistance aux antibiotiques pour lesquels elles étaient sensibles. Contrairement à la résistance naturelle, celle-ci ne concerne que quelques souches.<sup>61</sup> L'acquisition peut être due à une mutation affectant un gène de structure ou de régulation ou à l'acquisition de matériel génétique exogène (transfert horizontal, plus répandue en clinique).<sup>61</sup>

## *2. Les mécanismes de résistance*

Pour se défendre contre les antibiotiques, les bactéries emploient trois grands mécanismes qui ont pour but d'inactiver l'antibiotique, d'empêcher la reconnaissance de leur cible ou de rendre la membrane imperméable à leur pénétration.

### *Production d'enzymes*

La production d'enzymes altère les antibiotiques les rendant inactifs. C'est le cas notamment des  $\beta$ -lactamases qui hydrolysent l'anneau  $\beta$ -lactame des pénicillines et des céphalosporines.<sup>62</sup> La phosphorylation ou l'acétylation peuvent également inhiber l'activité des aminosides.<sup>63</sup>

### *Altération de la cible*

Les modifications structurales des cibles des antibiotiques empêchent l'antibiotique de s'y fixer. L'altération de la protéine liant la pénicilline (PLP) fait partie des mécanismes de résistance.<sup>64</sup>

### *Expulsion des antibiotiques*

L'expulsion des antibiotiques se fait par la production de protéines membranaires agissant comme pompes moléculaires. C'est le mécanisme principal de la résistance à la tétracycline, chez les bactéries Gram positives et Gram négatives.<sup>65</sup>

### 3. Support génétique de la résistance®

#### *Le chromosome*

Principal support de la résistance naturelle, le chromosome bactérien contient des gènes de résistance codant pour des enzymes d'inactivation. La résistance chromosomique acquise résulte quant à elle d'une mutation, dont la faible fréquence est difficile à définir et déterminer.<sup>66</sup> En effet, l'émergence de mutants résistants aux antibiotiques est un phénomène complexe dont la physiologie, la génétique de la bactérie, le milieu et l'environnement dans lequel elle vit ont un rôle prépondérant.<sup>67</sup>

#### *Les éléments exogènes et mobiles*

La résistance acquise est principalement due à la présence de matériel génétique exogène : les plasmides et les éléments transposables. Non essentiels à la survie de la bactérie, ils lui permettent de mieux s'adapter à un environnement donné.

**Les plasmides** : ce sont des éléments extra-chromosomiques d'ADN double brin circulaire, de taille variable (entre 0,5 et 500 kb). Comme les chromosomes, ils sont capables de se répliquer durant la croissance bactérienne.<sup>68</sup> Il en existe de plusieurs types et leur classification se fait selon leur groupe d'incompatibilité (Inc).<sup>69</sup> En effet, deux plasmides appartenant au même groupe d'incompatibilité ne peuvent pas coexister. Ainsi, les plasmides de types IncF, IncI1, IncX, IncN, IncH, IncB/O, IncK, IncK/M, IncP, IncQ, IncT, IncU, IncW et IncY sont parmi les plus connus.<sup>69,70</sup> Une bactérie peut contenir plusieurs plasmides porteurs d'un ou plusieurs gènes de résistances, conférant une résistance multiple aux antibiotiques.<sup>68,71</sup>

**Les transposons** (Tn) : sont bordés par des séquences d'insertion (transposon composite) ou par des répétitions inversées (IR, transposon non composite). Les séquences d'insertion (IS) sont des séquences d'ADN comprises entre 0,8 et 2,5 kb. Ces séquences ne codent que pour des protéines impliquées dans des activités de transposition. Elles sont bordées par des IR de 15 à 25 pb, indispensables à leur transposition.<sup>72</sup> Certains transposons cumulent des propriétés d'intégration (avec un gène d'intégrase) et des propriétés de transfert (gènes *tra*) (transposon conjugatif), leur conférant la capacité de transposer d'un site donneur vers un site accepteur. Ils sont également capables de s'exciser et former un élément circulaire

extra-chromosomique.<sup>73</sup> Le transposon Tn3, l'un des plus classiques confère la résistance à l'ampicilline et est largement distribué parmi les bactéries à coloration de Gram négative.<sup>72</sup> Le transposon Tn1 appartenant à la même famille, dérivant du plasmide RP4 est le premier transposon à avoir été découvert.<sup>74</sup>

**Les intégrons** : contrairement aux plasmides et aux transposons, les intégrons ne sont pas transposables par eux-mêmes mais résident sur les éléments mobiles cités ci-dessus<sup>72</sup> ou sur le chromosome. Décrit depuis les années 1980<sup>75</sup>, les intégrons sont caractérisés par la présence d'une intégrase, d'un site de recombinaison et d'un promoteur.<sup>72</sup> La majorité des intégrons appartiennent à la classe 1 et sont fréquemment associés à la famille des transposons Tn21.<sup>76</sup>

**Les îlots génomiques de résistance** : formés d'assemblages d'éléments génétiques de résistance, ils sont localisés sur le chromosome. L'îlot génomique 1 de *Salmonella* (ou SGI1 pour *Salmonella genomic island 1*) a été décrit pour la première fois chez le clone DT104 de *S. enterica* sérotype Typhimurium. Cet îlot de 43 kb contient un intégron de class 1 appartenant au groupe In4, lui conférant une résistance multiple à l'ampicilline, le chloramphénicol, le florfenicol, la streptomycine, la spectinomycine, les sulfamides et la tétracycline. SGI1 est également capable de s'exciser et passer d'une souche à l'autre.<sup>77</sup>

#### 4. Mécanisme de transfert de la résistance

Les plasmides et les éléments mobiles peuvent être transférés d'un endroit à un autre (plasmidique ou chromosomique), d'une bactérie à l'autre. L'acquisition d'éléments mobiles peut se faire selon trois mécanismes :

- la **transformation** : premier modèle d'échange connu<sup>78</sup> de petits fragments d'ADN entre une souche donatrice et une souche réceptrice, dite en état de compétence.<sup>79</sup>
- la **transduction** : transfert de l'ADN bactérien d'une bactérie donneuse à une bactérie receveuse, par l'intermédiaire de bactériophages.<sup>80</sup>
- la **conjugaison** : processus de transfert du matériel génétique d'une souche donatrice à une souche réceptrice, grâce à un pont formé par le pilus. L'ADN plasmidique est découpé au niveau de son origine de transfert (*oriT*) et se déroule grâce à un ensemble de protéines appelées relaxosome. L'ADN est transféré dans le sens 5' -> 3' puis le brin complémentaire

est resynthétisé afin de garder un plasmide double brin. Le transfert se fait notamment à l'aide de gènes de transfert *tra*.<sup>81</sup>

Certains plasmides sont dits conjugatifs ou autotransférables. Ils sont classés selon leur structure génétique:

- Plasmides conjugatifs de type IncF : IncF, IncS, IncC, IncD et IncI
- Plasmides conjugatifs de type IncP : IncP, IncU, IncM, IncQ et IncW
- Plasmides conjugatifs de type Ti : IncX, IncH, IncN et IncT
- Plasmides conjugatifs de type IncI : IncI, IncB et IncK.<sup>81</sup>

Les deux premiers groupes sont connus pour participer dans la dissémination des gènes de résistance aux antibiotiques, le plasmide IncF de 100 kb<sup>82</sup> et le plasmide RK2 de 60 kb (IncP)<sup>83</sup> étant les premiers à avoir été séquencés. Le premier transfert de gènes de résistances médié par un plasmide a été démontré en 1959 au Japon, avec un transfert entre une souche de *Shigella* et *Escherichia coli*.<sup>84</sup>

### 5. Détection au laboratoire

L'étude de souches résistantes au laboratoire permet d'une part de trouver le traitement approprié mais sert plus particulièrement à la surveillance épidémiologique et à la détection de résistances émergentes. Les tests peuvent être phénotypiques, c'est le cas de la méthode des disques de diffusion sur milieu gélosé ou antibiogramme,<sup>85,86</sup> qui permet dans un premier temps d'observer les différents profils : sensible, intermédiaire ou résistant (déterminés par le Comité de l'antibiogramme de la Société Française de Microbiologie [CASFM] et *European Committee on Antimicrobial Susceptibility Testing* [EUCAST])<sup>87</sup>, mais également de détecter les mécanismes enzymatiques. La concentration minimale inhibitrice (CMI) permet quant à elle de déterminer *in vitro* la concentration d'antibiotique la plus faible pouvant inhiber la croissance de la bactérie. Pour une étude génétique plus approfondie, une PCR pouvant détecter les gènes responsables de la résistance peut venir compléter les tests phénotypiques.

La classification des plasmides et leur identification moléculaire se fait grâce au typage des réplicons (*Replicon Typing*) qui consiste en la détection et l'amplification de gènes impliqués dans la réplication tels que les gènes *repA*, spécifiques à chaque plasmide.<sup>88</sup>

### III. Résistance chez les *Salmonella*

#### 1. Gènes

De nombreux gènes impliqués dans la résistance aux antibiotiques ont été identifiés chez *Salmonella*. Parmi eux:

- **Tétracycline:** *tet(A)*, *tet(B)*, *tet(C)*, *tet(D)* et *tet(G)* codant pour des pompes à efflux. Les gènes *tet(A)* et *tet(B)* sont les plus courants et le gène *tet(G)* n'est localisé qu'au niveau de l'îlot génomique de *Salmonella* 1 (SGI1) ou 2 (SGI2).<sup>89</sup>

- **Phénicolés** (incluant le chloramphenicol et le florfenicol) longtemps prescrits pour la fièvre typhoïde: *catA1/catA2*, *catB2/catB3/catB8*, *cmlA1/cmlA4/cmlA9* et *floR* codant notamment pour une acétyltransférase.<sup>89</sup>

- **Aminoglycosides:**

\**aadA* codant pour l'aminoglycoside-3'-O-adenyltransferase. A ce jour, une dizaine de gènes *aadA* impliqués dans cette résistance ont été décrites. Tous ces gènes sont localisés sur des cassettes au niveau des intégrons 1 et 2 (CFXXXXXXXXXX) ou sont localisés sur SGI1 et SGI2.<sup>89</sup>

\**aadB* codant pour une aminoglycoside-2'-O-adenyltransferase (gentamicine, kanamycine et tobramycine)

\**aacC* et ses variants [*aac(6'')-Ib-cr*] codant pour une aminoglycoside-N-acetyltransferase. Ces gènes ont une localisation plasmidique ou au niveau du SGI1 généralement.

\**strA* et *strB* codant pour une phosphotransferase (streptomycine). Ces gènes ont une localisation plasmidique ou chromosomique

\**aphA1* et *aphA2* codant pour un aminoglycoside 3'-phosphotransferases

\**armA* et *rmtC* codant pour une 16S ARNr methylases

- **Triméthoprime:** *dfrA* dont plus d'une dizaine ont été décrites et *dfrB*. Ils sont généralement localisés sur des cassettes au niveau des intégrons de classe 1 ou 2, des plasmides, ou les SGI1 and SGI2. Les gènes *sul1* à 3 confèrent également une résistance aux sulfamides.<sup>89</sup>

- **Macrolides:** *ere(A)*, *mph(A)*, *mph(E)*, et *msr(E)*<sup>89</sup>

#### 2. Cas des β lactam

La grande famille des β-lactamines agit sur la paroi bactérienne et comprend la pénicilline et ses dérivées (pénicilline G, ampicilline etc.), les céphalosporines, les carbapénèmes et les

monobactames. Les mécanismes de résistance des bactéries face aux  $\beta$ -lactamines sont nombreux mais le plus intéressant pour notre étude est la production de  $\beta$ -lactamase. Mécanisme de résistance majeur des bactéries à coloration de Gram négative, les  $\beta$ -lactamases sont des enzymes capables d'hydrolyser le cycle  $\beta$ -lactame rendant l'antibiotique inactif.<sup>90</sup> Les gènes codant pour les  $\beta$ -lactamases peuvent être localisés sur le chromosome mais sont le plus souvent situés sur les éléments mobiles et exogènes (transposons et plasmides) augmentant les risques de dissémination.<sup>91</sup> Un inhibiteur des  $\beta$ -lactamases est l'acide clavulanique.

Les  $\beta$ -lactamases sont classées selon leur fonction et leur affinité pour le cycle  $\beta$ -lactame comme suit :

- les pénicillinases à spectre restreint d'origine chromosomique, ayant pour principaux substrats les pénicillines, les carboxypénicillines et certains antibiotiques de type céphalosporines de première génération (C1G)
- les pénicillinases à large spectre et leurs variantes TEM (du nom du malade chez qui la première souche porteuse de ce type d'enzyme a été isolée, Temoneria<sup>92</sup>), OXA pour oxacillinase, SHV pour sulfhydrile-variable, PSE pour *Pseudomonas specific enzyme*. Souvent d'origine plasmidique, elles ont une activité enzymatique à l'oxacilline en plus
- les  $\beta$ -lactamases à spectre étendu d'origine plasmidique. Parmi elles se trouvent les pénicillinases à large spectre mais aussi celles à spectre étendu ou BLSE (PER, CTX, VEB, GES, BES). Elles hydrolysent toutes les bétalactamines jusqu'aux céphalosporines troisième génération (C3G) mais pas les céphamycines ni l'imipénème
- les céphalosporinases d'origine chromosomique ont pour substrats les céphalosporines C1G et certaines C2G, ainsi que les pénicillines G et les aminopénicillines
- les carbapénémases : d'origine chromosomique ou plasmidique. Elles peuvent être de type pénicillinase et inactive sur les C3G (OXA-48) ou de type métallo-enzyme, hydrolysant les C3G et les carbapénèmes (NDM, VIM). Parmi les carbapénémases se trouvent également les KPC pour *Klebsiella pneumoniae* carbapénémase et IMP pour imipénémase
- les céphamycinases : ayant pour substrats les céphems (céfoxitine, cefmétazole) CMY<sup>90</sup>

### 3. Mutations

Parmi les résistances dites chromosomiques se trouve la résistance aux quinolones (acide nalidixique) et fluoroquinolones (ciprofloxacine), antibiotiques largement utilisés contre la fièvre typhoïde à la fin des années 80 dans le Sud-est asiatique.<sup>93,94,95</sup> L'ADN gyrase et la

topoisomérase IV sont les principales cibles de ces antibiotiques chez les bactéries à coloration de Gram-négative. Une mutation de ces deux structures induit une altération de la perméabilité cellulaire et des pompes à efflux.<sup>96</sup> Ces mutations surviennent dans des régions spécifiques aux quinolones, appelées QRDR (pour *quinolone resistance determining regions*), telles que les gènes *gyrA*, *gyrB*, *parC* et *parE*, codant pour l'ADN gyrase et la topoisomérase IV. Pour le gène *gyrA*, il s'agit généralement d'une mutation de l'asparagine en position 87 en glycine, de la sérine en position 83 en phénylalanine et pour *gyrB* de la sérine en position 463 en phénylalanine.<sup>97</sup>

#### **IV. Epidémies de *Salmonella* liées à la résistance aux antibiotiques**

Naturellement sensibles aux antibiotiques, les premiers cas liés à la résistance aux antibiotiques chez les *Salmonella* ont été identifiés chez le sérotype Typhi. D'abord sporadiques avec la résistance au chloramphénicol dans les années 50, ils sont devenus épidémiques au début des années 70 en Inde, avec la première résistance lié à un plasmide.<sup>94</sup> Ainsi, en l'espace de dix ans (1992-2002), plus de 80% des souches au Vietnam sont devenues multi-résistants (MR).<sup>98</sup> Pour les sérotypes de *Salmonella* non-typhiques, des souches résistantes à la tétracycline ont été observées à partir des années 50, en même temps que la commercialisation de ces derniers<sup>99</sup>. Au Royaume-Uni, en 1989, une épidémie liée à la résistance aux antibiotiques est due à un clone de *S. enterica* sérotype Typhimurium, au lysotype particulier : le clone DT104 ayant intégré dans son chromosome un îlot génomique de 43 kb (SGI1). Ce clone est penta-résistant avec un profil de résistance « ASCSuT » (voir plus haut). En France, les premières épidémies étaient essentiellement hospitalières : 1963-1964 épidémie à *S. enterica* sérotype Derby au profil phénotypique ACSKSuTe<sup>100</sup>, 1965 épidémie causée par *S. enterica* sérotype Panama (AKTe)<sup>101</sup> porteur de plasmides de résistances autotransférables. Depuis le début des années 80, les données de surveillance indiquent une augmentation de la résistance aux antibiotiques passant de 20% dans les années 1990, à près de 70% dans certains pays dans les années 2000.<sup>102</sup> Plus récemment, depuis 2002, des souches résistantes à la ciprofloxacine ont émergé chez *S. enterica* sérotype Kentucky, en Egypte et se sont propagées à travers l'Afrique et le Moyen-Orient.<sup>103</sup> Ces souches appartenaient à un séquençotype particulier, le ST198 et portaient un variant de l'îlot génomique SGI1, SGI1-K.<sup>103</sup> Les années 2000 ont été également marquées par l'émergence du sérotype 1,4,[5],12,i:-, un variant monophasique du sérotype



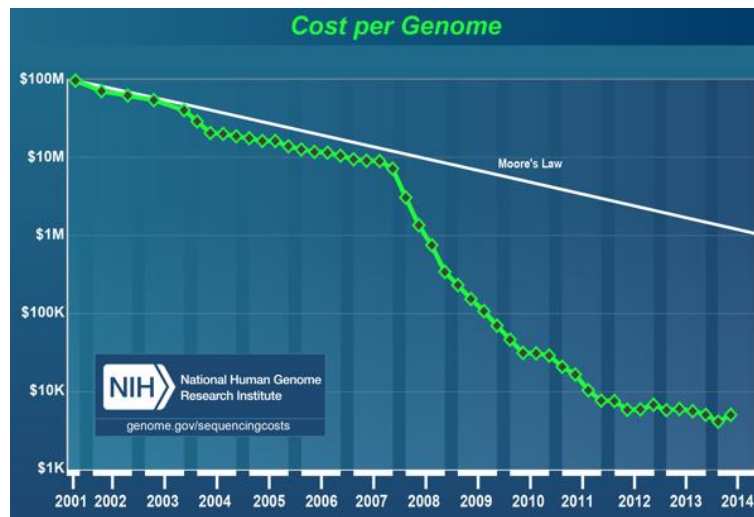
Typhimurium caractérisé par un profil de résistance de type ASSuTe (ampicilline, streptomycine, spectinomycine et tétracycline).<sup>104</sup>

## **Chapitre IV**

### **Génomique et séquençage**

Longtemps réservé à la recherche scientifique, le séquençage a vu son coût diminuer au fil des années (graphique 1), permettant ainsi son application dans le domaine de la santé publique et de la surveillance épidémiologique. Les technologies de séquençage nouvelles générations et les outils d'analyse qui en découlent ne cessent de progresser, bouleversant ainsi les pratiques traditionnelles de plus d'un siècle.

Dans ce chapitre, nous allons passer en revue ces outils d'analyse adaptés à la surveillance de *Salmonella*. Ce sera également l'occasion de définir les notions de génomique et de séquençage.



Graphique 1. Prix du séquençage d'un génome de 2001 à 2014.<sup>105</sup>

## I. Introduction au séquençage entier du génome (SEG)

### 1. ADN et génome

Découvert en 1869 par le biochimiste suisse F. Miescher<sup>106</sup>, l'ADN ou acide désoxyribonucléique est la base du patrimoine génétique. Il est formé de quatre petites molécules appelées nucléotides : A (adénine), T (thymine), C (cytosine) et G (guanine). Au sein de cette double hélice, les bases s'apparient par paires : A avec T et C avec G, reliées par des liaisons covalentes, formant deux brins antiparallèles. Il existe deux types d'ADN : l'ADN chromosomique et l'ADN extra-chromosomique (phagique ou plasmidique par exemple).

L'ADN peut être découpé en petits morceaux appelés gène correspondant à une information génétique particulière. Le génome est donc l'ensemble du matériel génétique d'un

organisme. Il peut contenir des séquences dites codantes pouvant coder pour une protéine ou non codantes. Le génome humain représente environ trois milliards de paires de bases et celui de *Salmonella* un peu moins de cinq millions de paires de bases.

L'ensemble des gènes d'une même espèce représente le pan-génome. Cela comprend :

- les gènes présents chez toutes les souches d'une même espèce, appelé le core-génome. Celui-ci inclut les gènes nécessaires à la survie et à la croissance de la bactérie comme les gènes codant pour des enzymes intervenant dans des mécanismes de synthèse. Tous ces gènes sont hautement conservés, tant au niveau de leur séquence qu'au niveau de la syntonie (ordre d'apparition des gènes dans le génome), évoluent lentement au cours du temps et sont généralement utilisés pour déterminer les relations existant entre les isolats bactériens.<sup>107</sup>
- les gènes qui ne sont pas présents chez toutes les bactéries d'une même espèce, sont appelés le génome accessoire. Ces gènes spécifiques à une ou plusieurs souches sont souvent situés sur des îlots génomiques ou des prophages.

## 2. Technologies de séquençage

Le séquençage consiste en la lecture des nucléotides qui composent l'ADN, une à une grâce à un dispositif appelé séquenceur. En 1977 fut séquencé le premier génome, celui du virus bactérien phi X174<sup>108</sup> (numéro d'accèsion GenBank NC\_001422.1), par l'un des pionniers du séquençage, Frederick Sanger, qui donna également son nom au séquençage Sanger, dit séquençage de première génération. En pratique, l'ADN est dénaturé, puis l'ajout des nucléotides par l'ADN polymérase va permettre la synthèse du brin complémentaire. La synthèse se termine par l'incorporation d'un didésoxynucléotide (nucléotide privé du groupement OH en position 3'), marqué par un fluorochrome (A vert, T rouge, G jaune et C bleu). Les fragments d'ADN (entre 20 pb à ~1000 pb) obtenus sont soumis à une électrophorèse capillaire durant laquelle les fragments migrent en fonction de leur taille. Le signal fluorescent émis par chaque bande permet la reconstitution de la séquence. Malgré des séquences longues, cette technique coûteuse en matériel et en temps, génère peu de séquences. A titre de comparaison, pour séquencer le génome humain, il aura fallu générer environ 30 million de petites séquences de 800 pb, correspondant à 24 billion de bases (Gb), c'est-à-dire huit fois la taille du génome humain (3 Gb).<sup>109,110</sup> Cette technique n'est adaptée qu'à des études de séquençage de gènes simples ou à une région précise du génome.

Ce n'est que 28 ans plus tard, en 2005, qu'est apparu le séquençage dit de nouvelle génération ou « *Next Generation Sequencing* » (NGS), en réponse à un prix élevé et un faible débit du séquençage de première génération.<sup>111</sup> Contrairement à la méthode Sanger, des dizaines de milliers de séquences sont générées parallèlement, d'où le nom de séquençage à haut débit. Un suivi en temps réel de la réaction enzymatique permet d'observer les variations structurales. Actuellement, trois plates-formes dominent le marché du séquençage : Illumina<sup>112</sup>, Ion Torrent<sup>113</sup> et Pacific Biosciences<sup>114</sup>. Illumina et Ion Torrent peuvent être considérés comme des séquenceurs de paille, générant de nombreuses séquences courtes allant d'une centaine de nucléotides à 300 paire de bases (pb) pour Illumina<sup>115</sup> et 400 pb pour Ion Torrent<sup>116</sup>. Ces deux séquenceurs utilisent des technologies bien distinctes : Illumina utilise la technologie de séquençage par synthèse avec une mesure de la fluorescence pour chaque nucléotide incorporé. La technologie Ion Torrent utilise des puces semi-conductrices remplies de puits et détecte le changement de pH par la mesure des ions. Le séquençage par Pacific Biosciences se base sur une toute autre méthode, le « *Single Molecule Real Time* » (SMRT). Contrairement aux deux autres séquençages, où l'ADN est lu base après base, la technologie SMRT lit un seul brin d'ADN dans son ensemble. Les séquences obtenues sont beaucoup plus longues, allant de 3,000 à 15,000 pb.<sup>117</sup> Le signal émis est plus faible induisant un bruit de fond et un taux d'erreurs plus élevés.<sup>118</sup>

### 3. Protocole de séquençage

Il existe deux types de séquençage : « **single-end** » avec le séquençage de l'ADN que d'un côté et le « **paired-end** », séquençage des deux extrémités 5' et 3' (*forward* et *reverse*), permettant d'obtenir des séquences beaucoup plus longues (de l'ordre de 50 pb pour le single-end et 100 pb pour le paired-end) (figure 13). Le séquençage single-end est plus rapide et moins coûteux. Le séquençage paired-end apporte des informations supplémentaires sur le génome et permet également une meilleure résolution.<sup>119</sup>



**Figure 13. Représentation des séquençages single-end et paired-end.**  
Les barres bleues représentent les séquences nucléotidiques.

L'une des premières étapes du séquençage est l'extraction de l'ADN génomique à étudier. Cet ADN fait ensuite l'objet d'une fragmentation enzymatique ou par ultrasons afin de former une **librairie** d'ADN. Les fragments obtenus sont de taille spécifique (en fonction de la demande de l'utilisateur) et marqués avec des adaptateurs (par ligation au niveau des extrémités 5' et 3' de l'ADN). Les fragments d'ADN sont ensuite déposés sur un support ou « *flow cell* » contenant elles-mêmes des adaptateurs complémentaires et spécifiques à chaque brin d'ADN. Ceci permet d'identifier les différents pathogènes présents sur une même *flow cell*. Les brins d'ADN sont amplifiés par PCR, générant des amas de fragments d'ADN identiques et démultipliés. Après cette étape, le séquençage peut alors débuter. Les désoxyribonucléotides (dNTP) A, T, G et C sont ajoutés et à chaque nucléotide incorporé, complémentaire au brin d'ADN, un signal fluorescent ou une émission de protons est mesuré. L'ensemble de ces signaux correspond à un enchaînement nucléotidique.

A la fin du séquençage, des fragments d'ADN d'une centaine de nucléotides sont synthétisés. Ces séquences sont appelées **reads** (figure 14). Chaque plateforme de séquençage produit un nombre différent de reads. Selon l'instrument utilisé, Illumina peut produire de 1,000,000 de reads (Illumina MiSeq v2 Nano) à des centaines de million (Illumina HiSeq X Ten), Ion Torrent génère une moyenne de 2,000,000 de reads et PacBio une moyenne de 35,000 reads.<sup>120</sup>

Les fichiers contenant les reads ont un format bien particulier appelé **fastq**. Ces fichiers contiennent les séquences nucléotidiques mais également, l'identification, la localisation de la séquence sur la *flow cell* et les scores de qualités associés à chaque nucléotide. Le format de fichier **fasta** ne contient que la séquence nucléotidique.

#### 4. Analyse génomique

Avant d'être analysés, les reads générés sont nettoyés et filtrés. AlienTimmer<sup>121</sup> ou CUTADAP<sup>122</sup> sont des outils permettant la détection de contaminants tels que les amorces ou les adaptateurs utilisés lors du séquençage. Sachant que les erreurs de séquençage se traduisent par des scores de faible qualité, ces logiciels éliminent les bases concernées.<sup>123</sup>

Une fois nettoyés, deux approches sont utilisées pour l'analyse des reads : l'**assemblage de novo** et l'**alignement** ou **mapping**.

## L'assemblage de novo

Les séquences de référence ou étroitement liées au génome étudié n'étant pas toujours disponibles, l'assemblage *de novo* permet la reconstruction du génome sans séquence de référence. Avec l'avancé du séquençage, de nombreux outils d'assemblage *de novo* se sont développés. Ces logiciels vont rechercher les reads chevauchants et les concaténer pour former des fragments d'ADN plus longs appelés **contig** (figure 14). Pour cela, ils utilisent deux principaux algorithmes : la superposition des graphiques (ou OLC pour *Overlap-Layout-Consus*) et le graphique de De Bruijn.

**OLC** : C'est un assemblage intuitif inventé par Staden en 1980.<sup>124</sup> Comme l'indique son nom, la méthode OLC va chercher les zones chevauchantes parmi tous les reads (overlaps, O) puis répertorie ces résultats sous forme d'un graphique (layout, L) et enfin, les concatène pour former une séquence consensus (consensus, C). L'utilisation de cet algorithme est appropriée pour le séquençage Sanger produisant peu mais de longues séquences (entre 100 et 1000 pb).<sup>125</sup> Les assembleurs tels que Celera Assembler<sup>126</sup> ou CAP3<sup>127</sup> utilisent cette approche.

**Graphique de De Bruijn** : introduit par Idury et Waterman en 1995<sup>128</sup>, les graphes de De Bruijn vont limiter les efforts de calcul, en coupant les séquences d'ADN en petites séquences appelées k-mer, avec k la taille de la séquence fixée par l'utilisateur. Pour éviter les séquences palindromiques (même séquence lue dans les deux sens), la valeur du k-mer est souvent un nombre impair. Le choix de sa valeur est crucial et dépend du génome étudié. En coupant les séquences en k-mer, le graph de De Bruijn, réduit les séquences redondantes. Ainsi, pour un génome comportant de nombreuses séquences répétées, il est plus judicieux de choisir une taille de k-mer supérieure à la séquence répétée pour éviter leur concaténation. Cet algorithme est plus adapté pour les séquenceurs actuels qui génèrent des millions de séquences courtes (entre 25 et  $\geq 100$  pb). La plupart des assembleurs tels que Ray<sup>129</sup>, Velvet<sup>130,131</sup>, SPAdes<sup>132</sup> et CLC assembler (outil payant et nécessitant une licence, CLC Genomics Workbench, Finlandsgade, Danemark) utilisent cette approche.

A la fin, le génome *de novo* est formé de plusieurs contigs de taille différentes. Plusieurs paramètres permettent d'évaluer la qualité du génome construit :

- le **N50** : lorsque les contigs sont triés par taille décroissante, le N50 est la longueur du dernier contig obtenu lorsque la somme des contigs couvre la moitié de la taille du génome. Plus le N50 est élevé, moins le génome est fragmenté, plus les contigs générés sont longs, évitant ainsi la présence de gènes tronqués situés sur plusieurs contigs différents.

- la **profondeur de couverture** : le séquençage génère des reads sensés recouvrir tout le génome. Cependant, produits de manière aléatoire et indépendante, ils ne sont pas répartis de manière égale sur tout le génome. La profondeur de couverture représente le nombre de fois moyen où une base nucléotidique est lue durant le séquençage. Par conséquent, une couverture de 50 X signifie que chaque base a été séquencée 50 fois. Plusieurs facteurs peuvent influencer sur la profondeur de couverture. Par exemple, plus il y a de régions avec un pourcentage de GC élevé dans le génome, plus la profondeur de couverture sera élevée.

133,134

### L'alignement

Contrairement à l'approche *de novo*, l'alignement compare des milliers de reads à une référence (génome entier ou séquences partielles) (figure 14). Le principal avantage d'une telle technique est la possibilité de détecter les gènes et séquences communs entre la référence et le génome étudié, ainsi que les variations génomiques telles que les insertions (ajout d'un nucléotide dans la séquence d'ADN), les délétions (suppression d'un nucléotide dans la séquence d'ADN) et les SNPs (*Single Nucleotide Polymorphism*, substitution d'un nucléotide par un autre).

Les outils d'alignement disponibles peuvent être divisés en deux catégories en fonction de l'algorithme utilisé : la table de hachage ou la transformation de Burrows Wheeler (TBW). Le premier algorithme coupe la séquence de référence en petits fragments et note leur position au sein du génome dans une table. Il coupe ensuite les reads en petits fragments puis les comparent à la référence. Lorsqu'une correspondance est trouvée, l'algorithme reconstitue les fragments grâce à leur position dans le génome. Cependant, cette méthode trouve ses limites lorsque le génome contient de nombreuses séquences répétées.<sup>135</sup> Pour résoudre ce problème, la transformation de Burrows Wheeler coupe également la séquence en fragments, mais tient compte des nucléotides en amont et en aval (préfixe et suffixe). A la différence de la table de hachage, ces fragments sont réordonnés puis transformés, évitant ainsi les séquences de répétitions.<sup>135</sup> En pratique, ces algorithmes sont beaucoup plus complexes car ils prennent en considération d'autres paramètres, y compris le score de chaque nucléotide. Les logiciels d'alignement TBW les plus utilisés sont Bowtie 2<sup>136</sup> et Burrows-Wheeler-Alignment (BWA)<sup>137</sup>.

Un autre outil d'alignement largement connu et utilisé est le Basic Local Alignment Search Tool (BLAST).<sup>138,139</sup> Comme la table de hachage, BLAST fragmente la séquence d'ADN puis



construit une matrice de similarité (BLOSUM 62 pour *BLOCKS SUBstitution Matrix*) qui permet d'attribuer une valeur positive pour chaque similarité ou une valeur négative pour chaque substitution. Contrairement aux deux algorithmes évoqués précédemment, BLAST ne tient pas compte de la qualité des nucléotides (fichier fastq) mais prend seulement la séquence nucléotidique (fichier fasta).

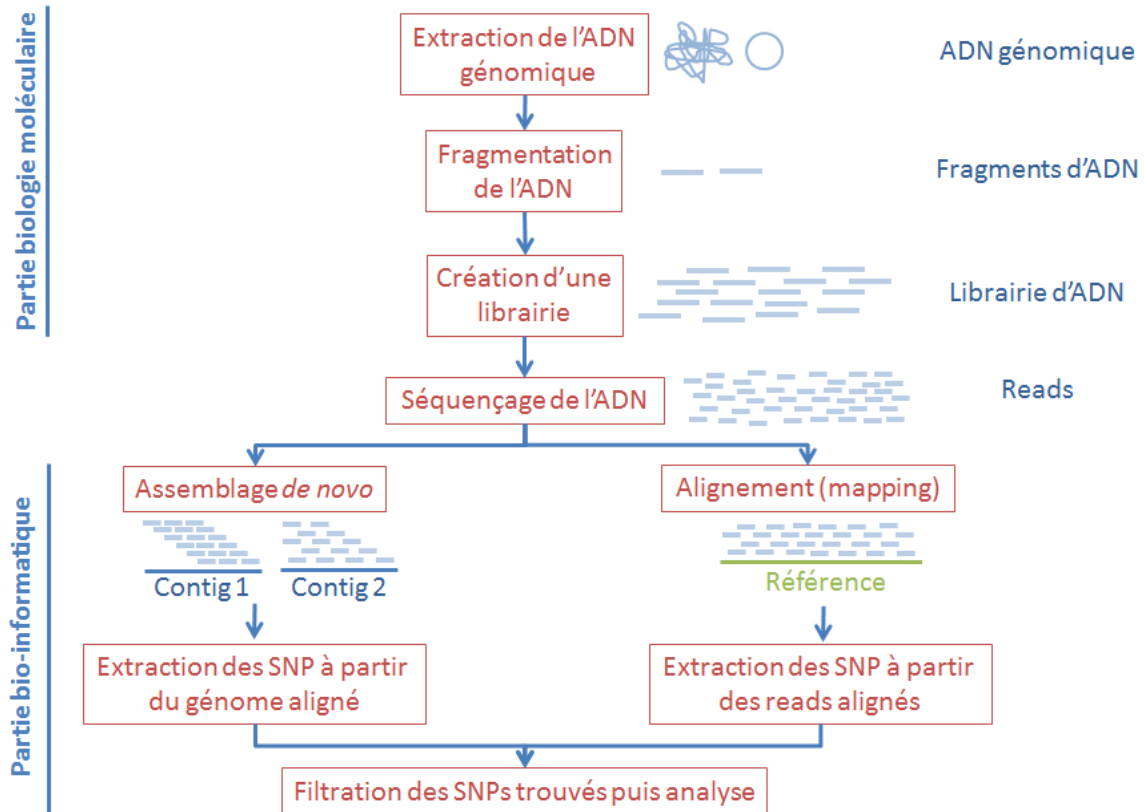


Figure 14. Schéma de reconstruction et d'analyse génomique à partir de l'ADN au séquençage entier du génome.

### 5. Analyse phylogénétique

L'étude phylogénétique permet de caractériser les liens de parentés entre les populations et les espèces au cours de leur évolution. Historiquement, les premières comparaisons bactériennes étaient basées sur une classification morphologique puis biochimique.<sup>140</sup> L'introduction de l'ARN 16s a permis une classification plus discriminante et a ouvert la porte à la phylogénie moléculaire.<sup>141</sup> Le point de départ de l'analyse phylogénétique est un ensemble de séquences homologues alignées à une référence pour aboutir à la fin à un arbre décrivant leurs relations évolutives. L'arbre phylogénétique est formé de branches

regroupant les souches proche entre elles à leurs extrémités, appelé clade. Chaque branche se rejoint en un point appelé nœud.

Les méthodes utilisées pour la construction d'arbres phylogénétiques sont fondées sur deux approches : les méthodes de distances et les méthodes basées sur les caractères (figure 15).

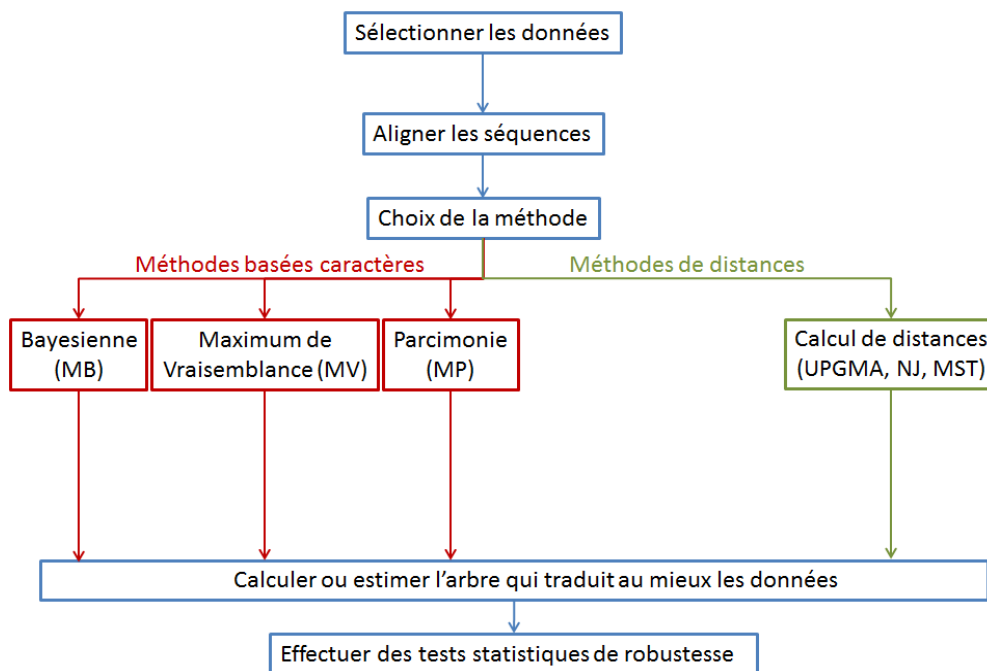
- **méthodes de distances** : elles consistent à calculer à partir d'un alignement multiple les distances entre les séquences prises deux à deux. A partir de cette matrice, les séquences sont regroupées pour dessiner un arbre. C'est une analyse quantitative adaptée à de grands jeux de données de par sa rapidité (moins de cinq minutes pour une centaine de séquences). Réduisant la séquence à un nombre, elle est moins précise pour l'analyse des séquences éloignées.<sup>142</sup> Les méthodes de classification ascendante hiérarchique au lien moyen (*Unweighted Pair-Group Method with Arithmetic means*, UPGMA), de Neighbor-Joining (NJ) ou d'arbre couvrant de poids minimal (*Minimum Spanning Tree*, MST) sont les plus utilisées. Contrairement à l'UPGMA et au NJ, le MST est un arbre où les clades sont représentés par des cercles et les branches remplacées par des liens dont la longueur est proportionnelle à la distance au clade adjacent. Cette méthode permet une meilleure visualisation lors d'une étude populationnelle.<sup>143,48,53</sup>

- **méthodes basées sur les caractères** : parmi elles, les méthodes de Parcimonie (MP), de Maximum de Vraisemblance (MV) et Bayésienne (MB). Contrairement à la méthode précédente, ces analyses sont qualitatives et donnent une information sur les séquences ancestrales. Beaucoup plus fiables (notamment la méthode MV la plus utilisée)<sup>144</sup>, elles sont coûteuses en calculs (en fonction du nombre de souches, cela peut aller jusqu'à 2 jours). La méthode MP gère également mal les substitutions multiples au sein du génome.<sup>144</sup> De nombreux logiciels permettent ces constructions d'arbres, RAxML(MV)<sup>145,146</sup> et BEAST (MB)<sup>147</sup> étant les plus connus. L'avantage de BEAST est la prise en compte des dates d'isolement, permettant une datation et une visualisation évolutive au cours du temps.

Cependant, il n'existe pas de bases de données ou de convention internationale pour l'analyse phylogénétique par les SNPs, laissant ainsi chaque laboratoire définir leur propre paramètre. Néanmoins, contrairement au MLST qui compare des gènes entre eux, l'approche SNPs va au-delà du gène et compare le génome dans son ensemble, permettant une meilleure discrimination.

A la fin, la robustesse et la fiabilité de l'arbre sont évaluées par le « *bootstrap* ». Cette technique consiste en la construction de plusieurs arbres et la comparaison des résultats obtenus. L'arbre final aura une valeur de bootstrap à chacun de ses nœuds, indiquant le

nombre de fois où cette branche a été retrouvée au fil des répétitions. Cette valeur peut aller de 0 à 100 (100 étant la meilleure valeur).



**Figure 15. Organigramme à suivre pour l'obtention d'un résultat phylogénétique fiable à partir des séquences.**

Les méthodes basées sur les caractères sont représentées en rouge et celles sur les distances en vert.

Adapté de Hillis *et al.* (1993)<sup>142</sup>

## II. Génomique de *Salmonella*

En génomique des *Salmonella*, le génome de *S. enterica* sérotype Typhimurium LT2 est généralement utilisé en tant que référence. Et pour cause, cette souche a été la base des premières cartes génétiques publiées en 1965 par Sanderson et Demerec<sup>148</sup> (figure 16) et du premier séquençage génomique de *Salmonella* par McClelland en 2001.<sup>149</sup> Au départ, les études génomiques consistaient en la construction de cartes génétiques, par l'utilisation d'enzyme de restriction. Ces cartes représentaient une succession de gènes, sans aucune indication sur la taille ni la séquence nucléotidique du génome.

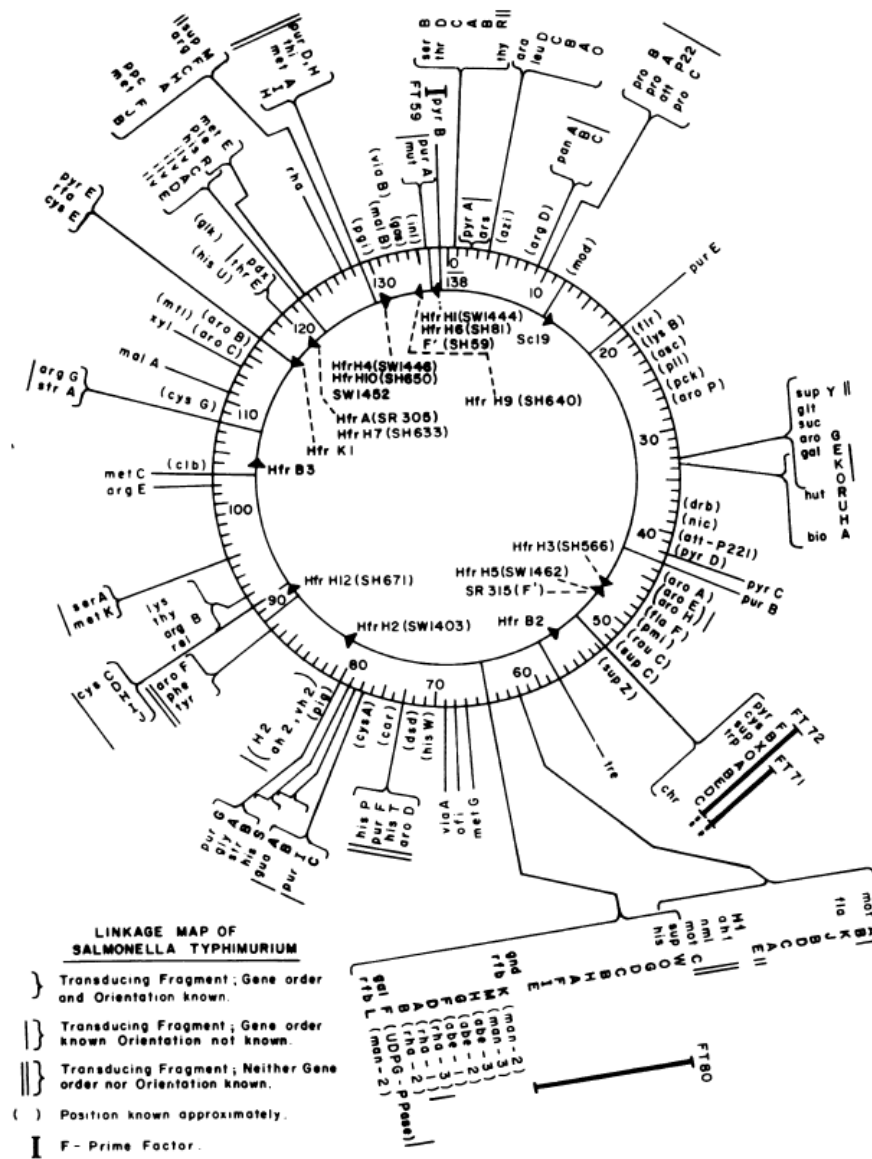


Figure 16. Carte génétique de la souche de *S. enterica* sérotype Typhimurium LT2. Sanderson et Demerec, 1965. <sup>148</sup>

Depuis, d'autres génomes tels que *S. enterica* sérotype Typhi CT18 (STY)<sup>150</sup> ont été séquencés par la méthode de séquençage traditionnelle Sanger. On compte aujourd'hui près de 8,000 génomes de *Salmonella* séquencés.<sup>151</sup>

### III. Outils dédiés à *S. enterica* sérotype Typhimurium basés sur le SEG

Les données biologiques ne cessent de croître de manière exponentielle, tant en nombre qu'en type. Développer des outils permettant d'extraire, de recouper les données et

permettre aux utilisateurs de comparer leurs résultats à l'existant est devenu une nécessité. Mêlant les connaissances de plus d'un siècle aux nouvelles technologies, nombreux sont les outils mis en place pour l'identification et la caractérisation de *Salmonella*. Ils sont disponibles et ouverts à tous. En voici une liste non-exhaustive.

### 1. SeqSero

SeqSero (*Salmonella Serotyping by Whole Genome Sequencing*) a été développé par l'Université de Géorgie en 2015 et permet le sérotypage de *Salmonella*.<sup>152</sup> (<http://www.denglab.info/SeqSero>). Il est basé sur la recherche des gènes codant pour les antigènes O et H. Parmi eux se trouve le cluster de gènes *rfb* synthétisant l'antigène somatique O, comprenant d'autres gènes spécifiques à la caractérisation de cet antigène et les gènes *fliC* et *fliB* codant pour les antigènes flagellaires chez *Salmonella*. Une base de données répertoriant les principaux déterminants de chaque sérotype a été construite. La recherche de ces gènes se fait à partir d'assemblages *de novo* ou de reads (single-end ou paired-end). Une prédiction du sérotype est donnée au bout de dix minutes et ressemble donc à quelques exceptions près, au schéma Kauffmann-White établi il y a près d'un siècle. La figure 17 est un exemple de pipeline utilisé par le site pour la détermination du sérotype.

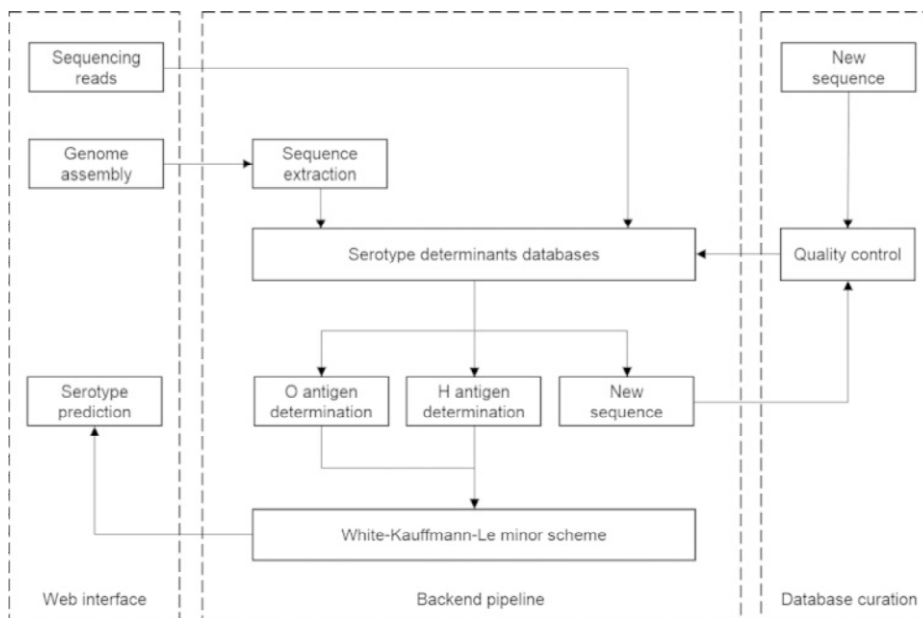


Figure 17. Schéma du pipeline utilisé par SeqSero pour la détermination du sérotype

Malgré une base de données préexistante, SeqSero n'arrive pas à différencier les sérotypes ayant la même formule antigénique mais qui diffèrent d'un seul facteur au niveau de

l'antigène O qui, contrairement aux antigènes flagellaires, résultent de voies métaboliques complexes.<sup>153</sup> C'est le cas du sérotype Carrau (6,14,[24]:y:1,7) et du sérotype Madelia (1,6,14,25:y:1,7). Il en est de même pour certains sérotypes monophasiques (Typhimurium et son variant monophasique) ou rough (n'exprimant pas l'antigène O). Les nombreuses mutations affectant les gènes somatiques et flagellaires rendent leur identification plus difficile. De plus, du fait de la grande taille de la région *rfb*, pouvant aller jusqu'à 20 kb selon les sérotypes, c'est une région longue et riche en A-T, induisant une faible profondeur de couverture et un assemblage difficile.<sup>154</sup> Le gène se retrouve généralement sur deux contigs différents engendrant une perte de l'information, voire l'absence totale de l'antigène O. La base de données de SeqSero ne comprend pas tous les sérotypes et seulement 100 sérotypes sont identifiés.

## 2. *Salmonella in silico Typing Resource (SISTR)*

Développé par le laboratoire National de microbiologie au Canada, SISTR est un outil qui, en plus de la détermination du sérotype par génosérotypage, il est implémenté de trois autres recherches : le MLST, le rMLST (*ribosomal multilocus sequence typing*) et le cgMLST (core génome MLST), permettant ainsi une identification affinée du sérotype (<https://lfz.corefacility.ca/sistr-app/>).<sup>155</sup> L'approche rMLST permet une détection de 53 gènes codant pour des protéines ribosomales, les gènes *rps*. Ces gènes ont été choisis pour leur présence chez toutes les bactéries, leur distribution sur tout le génome et pour leur conservation au cours du temps.<sup>156</sup> Cette technique permet une meilleure discrimination et résolution que le typage par les sous-unités 16S de l'ARNr.<sup>156</sup> Pour le cgMLST, la base de données de SISTR s'appuie sur 330 core gènes de *Salmonella* (cgMLST330), déterminés à partir de 361 sérotypes différents.<sup>155</sup>

Avec l'approche MLST, la discrimination du sérotype Typhimurium et de son variant monophasique 1,4,[5],12:i:-, n'est pas possible, les deux sérotypes étant regroupés au sein du séquençotype ST19. De plus, SISTR ne permet pas une soumission des reads mais seulement du génome assemblé.

## 3. *Center for Genomic Epidemiology*

Le Centre de Génomique Epidémiologique (CGE) développé à l'Université danoise en 2010 regorge d'outils pour l'étude de *Salmonella* (<http://www.genomicepidemiology.org/>). Les démarches d'identification et de typage ressemblent à celles développées dans le premier

chapitre mais ont pour fichier d'entrée une séquence génomique. Selon le nombre d'utilisateurs sur le site, les résultats peuvent varier entre dix minutes et plus d'une heure pour la recherche des marqueurs et entre un et trois jours pour l'analyse phylogénétique.

### Génotypage

Une base de données regroupant les principales familles d'antibiotiques a été mise en place afin de permettre la recherche de gènes de résistance aux antibiotiques (ResFinder<sup>157</sup>, <https://cge.cbs.dtu.dk/services/ResFinder/>), grâce à un alignement par BLAST. Basée seulement sur la présence de ces gènes, ResFinder ne détecte pas les mutations chromosomiques pouvant induire une résistance telles que les mutations des gènes *gyrA*, *gyrB*, *parC* et *parE*, responsables de la résistance à l'acide nalidixique. Il est donc important de réaliser un antibiogramme en parallèle afin de comparer les profils phénotypiques et génotypiques. Ceci permettra également de mieux comprendre les mécanismes de résistance.

D'autres outils de génotypages sont également à disposition tels que la recherche des îlots de pathogénicité de *Salmonella* (SPIfinder).

### Typage

L'outil SeqSero décrit précédemment a été implémenté sur ce site.

Tout comme le ribotypage, SpeciesFinder<sup>158</sup> (<https://cge.cbs.dtu.dk/services/SpeciesFinder/>) et Reads2Type<sup>159</sup> (<https://cge.cbs.dtu.dk/services/Reads2Type/>) sont basés sur la recherche de l'ARNr 16S. L'un en cherchant des gènes dans leur totalité et l'autre en cherchant des séquences spécifiques d'une espèce d'une longueur de 50 nucléotides. Les gènes ribosomiaux essentiels à la survie des cellules, sont hautement conservés au sein d'une même espèce. En raison de leur faible variabilité, l'approche de l'ARNr 16S est rarement utilisée lors d'études épidémiologiques.<sup>35,36</sup>

Pour ne pas limiter la recherche à un seul gène, l'extraction des sept gènes de ménage (MLST) est également disponible sur le site (<https://cge.cbs.dtu.dk/services/MLST/>).<sup>160</sup> Cette approche s'appuie sur une base de données préexistante.

Nous avons vu dans le chapitre III l'importance des plasmides dans la dissémination des gènes de résistance aux antibiotiques. Il est donc important de pouvoir les typer grâce à PlasmidFinder<sup>161</sup> (<https://cge.cbs.dtu.dk/services/PlasmidFinder/>) qui utilise la même

technique que le réplicon typing, en cherchant les séquences des réplicons spécifiques à chaque groupe d'incompatibilité.

De la même manière que le MLST type le génome bactérien, le typage plasmidique multilocus (pMLST) permet de typer les plasmides en se basant sur la diversité allélique de chaque groupe d'incompatibilité plasmidique (<https://cge.cbs.dtu.dk/services/pMLST/>).<sup>161</sup>

### Phylogénétiques

CGE a mis à disposition trois outils d'analyse phylogénétique basés principalement sur la recherche de SNPs (snpTree<sup>162</sup> [<https://cge.cbs.dtu.dk/services/snpTree/>], CSI Phylogeny<sup>163</sup> [<https://cge.cbs.dtu.dk/services/CSIPhylogeny/>] et NDtree [<https://cge.cbs.dtu.dk/services/NDtree/>]).<sup>164,165</sup>

### Reconstruction génomique

Pour les novices en bio-informatique ne sachant pas utiliser les logiciels d'assemblage en ligne de commande, CGE a implémenté son site de deux outils d'assemblage *de novo* : Velvet (<https://cge.cbs.dtu.dk/services/Assembler/>) et SPAdes<sup>166</sup> (<https://cge.cbs.dtu.dk/services/SPAdes/>).

## 4. Enterobase

Enterobase est une base de données développée pour *Salmonella*, *Escherichia/Shigella*, *Yersinia* et *Moraxella* par l'équipe de Mark Achtman en 2016 (<https://enterobase.warwick.ac.uk/>) et remplace l'ancienne base MLST de Warwick ([mlst.warwick.ac.uk](http://mlst.warwick.ac.uk)). La base *Salmonella* comprend à ce jour (octobre 2017) 123,626 souches. L'utilisateur soumet ses reads bruts (.fastq) sous un format compressé (.gz), accompagnés des métadatas pour chaque génome. A la fin de l'analyse, l'utilisateur peut télécharger les résultats sous forme de tableau. Ces résultats comprennent, en plus de l'analyse MLST (sept gènes), une extraction rMLST (53 gènes)<sup>156</sup>, cgMLST (contrairement à SISTR, celui-ci comprend 3,002 gènes), wgMLST (en incluant le cgMLST et le génome accessoire, celui-ci représente le pan-génome avec 21,065 gènes)<sup>167</sup>, une analyse CRISPR réalisée à partir de la base de données Fabre *et al.*, une prédiction du sérotype en utilisant SISTR, une annotation du génome en utilisant l'outil d'annotation génomique Prokka<sup>168</sup> (cet outil recherche notamment les séquences codantes et les gènes ribosomiaux) et une analyse phylogénétique (MST ou Maximum de vraisemblance). L'utilisateur peut également



télécharger les génomes assemblés et créer des espaces communs de travail facilitant les échanges de résultats. Malheureusement, Enterobase ne permet pas une soumission de fichiers .fastq ni de fichiers .fasta, obligeant l'utilisateur à effectuer une étape en plus qui demande une certaine connaissance en informatique. En plus de ces inconvénients, Enterobase est une base de données implémentée par tous les laboratoires et devient public au bout de six mois minimum (l'utilisateur pouvant prolonger le mode privé). Outre les incohérences rencontrées au niveau des métadatas entrées par les utilisateurs, Enterobase n'est pas adaptée pour une étude rapide, en routine et personnelle.

### 5. Panseq

L'agence de santé publique du Canada a développé en 2017 un site internet permettant de comparer les génomes soumis (fichiers .fasta) en extrayant le pan-génome, core-génome et génome accessoire, ainsi que le core-SNP (<https://lfz.corefacility.ca/panseq/>).<sup>169</sup> Le logiciel aligne les génomes entre eux ou à une référence et crée des fichiers contenant les séquences communes ou uniques à chaque génome. N'étant pas spécifique à *Salmonella*, il permet tout de même une analyse rapide du génome dans sa totalité pour un novice en bio-informatique.

### 6. PHAST

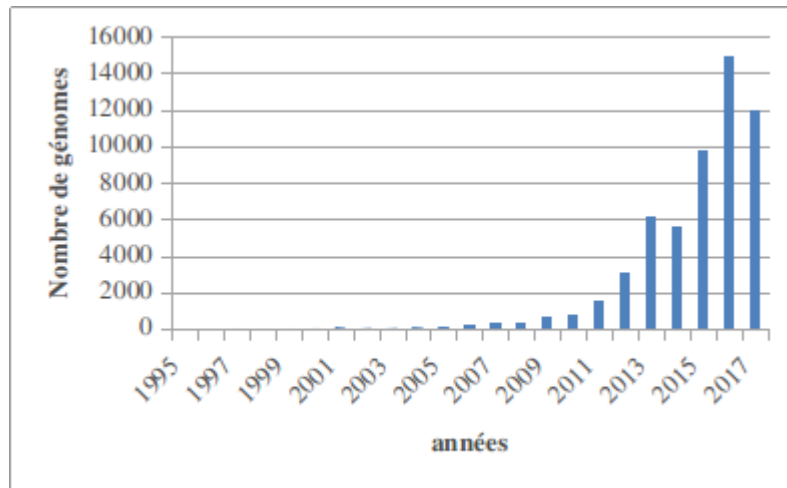
PHAST est un site internet Canadien, mis en place par Arndt *et al.* en 2016 permettant une analyse et une visualisation des phages présents au sein du génome, ainsi que les gènes présents dans ces phages<sup>170,171</sup> (<http://phast.wishartlab.com/>, <http://phaster.ca/>). Tout comme Panseq, PHAST n'est pas entièrement dédié à *Salmonella* mais sachant qu'un génome bactérien est composé de plus de 20% de gènes phagiques<sup>172</sup>, il est important de pouvoir les identifier. De plus, nous avons vu précédemment le rôle prépondérant des phages dans la formation des régions CRISPR.

Le SEG est donc la technique privilégiée pour la détection et l'analyse des gènes caractérisant une souche bactérienne. La plupart de ces outils permettent l'analyse de plusieurs souches en même temps. Toutefois, l'utilisateur devra soumettre le même génome pour chaque outil qu'il voudra utiliser séparément. Certains sites comme CGE imposent une restriction d'une vingtaine de soumissions par utilisateur. La vitesse d'analyse des génomes dépend également du nombre d'utilisateur connecté sur le serveur.

## **Etude expérimentale**

## Objectifs de la thèse

Près d'un siècle s'est écoulé entre la découverte de l'ADN (1869)<sup>106</sup> et le séquençage du premier génome bactérien (1977)<sup>108</sup>. Depuis, le séquençage entier du génome a révolutionné le monde de la microbiologie et les technologies utilisées ne cessent d'évoluer à une vitesse sans précédent. L'émergence du séquençage dit à haut débit puis des séquenceurs de pailleuse a contribué à la réduction du coût<sup>173</sup> du séquençage du génome bactérien, le rendant abordable à un grand nombre de laboratoires et conduisant à sa démocratisation.<sup>111</sup> Ainsi, en l'espace d'une dizaine d'années, 112,292 génomes ont été soumis sur le site de NCBI (2017) (graphique 2) et plus d'une dizaine de milliers de projets de métagénomique sur GOLD (2017). Les champs d'application du séquençage sont innombrables (microbiologie, phylogénie, diagnostic, épidémiologie etc)



Graphique 2. Nombre de génomes soumis sur NCBI ces vingt dernières années.<sup>174</sup>

Avec l'usage accéléré et massif du SEG, le nombre colossal et la complexité des séquences générées peuvent devenir un frein à ces nouvelles technologies. Face à cette quantité de données, le biologiste doit être suffisamment qualifié et en mesure de vérifier, d'extraire les informations, de comparer, d'interpréter et de visualiser l'ensemble des résultats. Il existe déjà de nombreux outils de traitements génomiques (assemblage *de novo*, alignement) ainsi que des sites dédiés à l'analyse des génomes (SeqSero, SISTR, CEG, Enterobase) mais la demande ne cesse d'augmenter.

Néanmoins, grâce au travail collaboratif entre bioinformaticiens et biologistes, le SEG est en passe de devenir la méthode privilégiée pour l'étude génomique des pathogènes. L'analyse *in silico* des séquences permet une exploitation complète de l'information génétique et la compréhension des mécanismes jusqu'alors inconnus, de manière rapide et en une seule étape.

Le CNR-ESS durant laquelle cette thèse s'est déroulée a pour mission (1) de détecter précocement les épisodes épidémiques, par la caractérisation et le typage rapide des souches de *Salmonella*, (2) de suivre l'évolution et les mécanismes de résistance aux antibiotiques de *Salmonella* et enfin, (3) d'établir un lien phylogénétique entre les différentes souches afin d'observer leur évolution au cours du temps.

Dans ce cadre, **l'objectif principal de ce travail est d'utiliser le séquençage entier du génome pour étudier un pathogène touchant des millions de personnes par an : *Salmonella***. Les salmonelles sont l'une des causes principales de maladies diarrhéiques dans le monde.<sup>175</sup> Dépister les foyers épidémiques le plus rapidement possible est un élément clé pour lutter contre ces bactéries et éviter leur propagation. De nombreux outils d'analyses *in silico* spécifiques à ce pathogène ont été mis en place afin d'**identifier** ces bactéries. Peu discriminants, ils ne sont pas adaptés pour une **étude épidémiologique** sur un grand nombre de génomes. Il est donc important de développer un outil rapide et efficace qui puisse détecter et établir des liens unissant les bactéries.

De plus, depuis une vingtaine d'années, nous avons été témoins d'une **explosion de la résistance** des bactéries face aux antibiotiques. L'émergence inquiétante des souches résistantes aux antibiotiques, dont le plus connu est le clone DT104 de *S. enterica* sérotype Typhimurium, touchent autant l'homme que l'animal et plusieurs pays industrialisés en Amérique du Nord et en Europe. Certains antibiotiques sont devenus inefficaces et bientôt, les marchés pharmaceutiques ne seront plus en mesure de fournir la quantité et les antibiotiques nécessaires au traitement des patients. Il est essentiel à ce stade de comprendre les mécanismes d'émergence des populations résistantes aux antibiotiques. Encore une fois, le SEG est l'outil idéal pour ce type d'étude.

Tandis que certains sérotypes ne cessent d'évoluer et de s'adapter pour se répandre dans le monde entier, d'autres sérotypes semblent subir peu de changements. Suivre l'**évolution génomique** d'un sérotype au cours du temps, depuis sa découverte à nos jours est un pas de plus dans la **compréhension de l'adaptation bactérienne face à son environnement**.

Aujourd'hui, le SEG se montre comme le seul outil rapide capable de retracer l'histoire d'un pathogène, du passé jusqu'au présent et de permettre l'identification et la détection d'épidémies en cours. Le SEG est également un outil d'avenir qui remplacera sûrement la plupart des méthodes devenues certes archaïques mais qui resteront les techniques de référence pour le développement et la validation de l'analyse génomique.

Avec une collection de souches récoltées partout dans le monde depuis des décennies, l'Unité des Bactéries Pathogènes Entériques renferme une ressource inestimable d'informations qui ne demandent qu'à être exploitées. C'est avec cette collection que nous allons aborder les trois points cités ci-dessus : **(1)** le typage à visée épidémiologique, grâce au développement d'un outil de sous-typage *in silico* d'un sérotype prédominant, *S. enterica* sérotype Typhimurium, basé sur le polymorphisme des régions CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*), **(2)** l'étude rétrospective de la transmission de la résistance à l'ampicilline chez *S. enterica* sérotype Typhimurium et enfin, **(3)** l'évolution d'un sérotype rare, *S. enterica* sérotype Paratyphi C.

L'ensemble de ces études sera présenté sous forme d'articles publiés ou en cours de publication. .

## **Chapitre V**

### **SEG : Typage à visée épidémiologique**

#### **(Article1)**

## Abstract

Les salmonelloses sont une des premières causes de diarrhées bactériennes d'origine alimentaire dans le monde. Une technique de sous-typage à haut-débit de *Salmonella enterica* sérotype Typhimurium a précédemment été développée au sein de l'Unité des Bactéries Pathogènes Entériques, nommée CRISPOL et basée sur le polymorphisme des régions CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*). Cette méthode utilise la technologie Luminex pour cibler la présence de 72 séquences variables CRISPR appelées espaceurs. Cependant, le séquençage entier du génome (SEG) plus discriminant, est devenu la méthode privilégiée pour l'analyse des relations génétiques d'un grand ensemble de souches bactériennes. Ainsi, nous avons développé l'outil CRISPOL *in silico* à partir des séquences de SEG et en le comparant à la méthode Luminex CRISPOL, méthode de référence.

Pour cela, les génomes de 360 isolats de *S. enterica* sérotype Typhimurium ont été obtenus par séquençage (technologie Illumina). Ces isolats ont été choisis pour être représentatifs de la diversité du sérotype tant au niveau temporel (1911-2013), géographique, que génétique (MLST *Multilocus Sequence Typing* et CRISPOL [169 CRISPOL Type ou CT ont été identifiés par le CRISPOL Luminex]). De nombreux outils d'assemblage *de novo* et d'alignement ont été comparés afin de définir l'outil le plus adapté pour l'extraction des régions CRISPR. L'efficacité du CRISPOL *in silico* a été déterminée par une étude rétrospective sur l'année 2016. Grâce à ce nouvel outil, une épidémie de *S. enterica* sérotype Typhimurium liée à la consommation de poulets contaminés a pu être mise en évidence. L'approche CRISPOL a également été comparée à la phylogénie basée sur le core génome (SNPs, *Single Nucleotide Polymorphisms*), afin d'établir une nouvelle nomenclature.

Utilisé en routine, CRISPOL *in silico* permettra le sous-typage des *S. enterica* sérotype Typhimurium, tout en gardant et implémentant une base de donnée déjà bien établie. Un site internet dédié, ouvert au public est en cours de développement.

# CRISPOL *in silico*, méthode de typage de *Salmonella enterica*

## sérotype Typhimurium

Alicia Tran Dien, Laetitia Fabre, Nizar Fawal, Simon Le Hello, François-Xavier Weill

### ARTICLE EN COURS DE PUBLICATION

## Introduction

*Salmonella* est l'une des causes principales de maladies infectieuses d'origine alimentaire dans le monde.<sup>176</sup> En 2016, le Centre National de référence des *Escherichia coli*, *Shigella* et *Salmonella* (CNR-ESS) a enregistré 10,545 isollements humains de *Salmonella* en France, (contre 12,357 en 2015).<sup>3</sup> Les sérotypes prédominants sont Typhimurium (ubiquitaire) et son variant monophasique 1,4,[5],12:i- (filière porcine) émergé durant cette dernière décennie et le sérotype Enteritidis (filière œuf). Ces trois sérotypes représentent 70% des isollements de *Salmonella* en France chez l'homme.<sup>3</sup> Pathogène à caractère zoonotique, il nécessite une surveillance à tous les niveaux, tant chez l'homme que chez l'animal, mais également dans le secteur de l'environnement. Cette surveillance repose sur l'identification précoce de la bactérie ainsi que son typage au laboratoire. L'information générée à travers les réseaux de surveillance peut alors permettre l'identification de cas groupés liés à un même type bactérien et de suspecter une épidémie causée par un aliment contaminé. Plus cette information est générée et analysée rapidement, plus vite sera identifié et détruit l'aliment contaminé. La morbidité et la mortalité en lien avec cette épidémie en seront amoindries.

Depuis près d'un siècle, le sérotypage est la méthode de référence pour typer les *Salmonella* grâce à l'agglutination des antigènes somatiques O et flagellaires H (H1 et H2), basée sur le schéma White-Kauffmann-Le Minor (WKL).<sup>1</sup> Avec une nomenclature universelle, cette méthode utilisée dans tous les laboratoires reste longue, lourde et coûteuse. Le développement des techniques telles que l'ECP et la détection des sept gènes de ménage par le MLST ont amélioré le dépistage épidémiologique mais ne sont pas assez discriminants<sup>40,53</sup> pour sous-typer les deux sérotypes majoritaires, Typhimurium et Enteritidis.

Depuis 2010, de nombreux laboratoires dont celui du CNR-ESS utilisent une technique de sous-typage à haut débit de *S. enterica* sérotype Typhimurium et de ses variants monophasiques ou immobiles basée sur le polymorphisme des régions CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) à l'aide de la technologie xMAP de



Luminex®.<sup>57</sup> Cette méthode tend à disparaître pour laisser place au séquençage entier du génome.

L'arrivée des technologies de séquençage haut-débit adaptées aux génomes bactériens (Illumina, Ion Torrent et Pacific Biosciences) est en train de bouleverser le monde épidémiologique. Au cours des dernières années, des applications dédiées à l'analyse génomique des *Salmonella* ont vu le jour. Ces outils allient des techniques sérologiques établies il y a plus d'un siècle aux nouvelles technologies. Des sites tels que SeqSero<sup>152</sup> sont à la recherche de la région *rfb* et des gènes *fliC/fliB* codant pour les antigènes somatiques O et flagellaires H respectivement, à partir des séquences génomiques. La formule antigénique déduite est ensuite comparée au schéma classique WKL pour fournir le sérotype correspondant. Outre la détermination des gènes codant pour les antigènes, d'autres sites comme Enterobase sont implémentés d'outils tels que le cgMLST, le wgMLST ou encore l'étude du CRISPR en utilisant la base de données Fabre *et al.* Malgré ses performances, Enterobase n'est pas un outil adapté pour un dépistage en routine ou une étude rapide. En effet, l'utilisation du site nécessite des connaissances en bio-informatique et les génomes soumis deviennent publics. Enterobase est donc plus une base de données génomique qu'un outil d'analyse.

Durant ce chapitre, nous avons voulu développer une méthode unique, universelle et à haut débit pour le sous-typage des salmonelles, en utilisant le polymorphisme des régions CRISPR basée sur les technologies de séquençage, le CRISPOL *in silico*. Cet outil se veut d'être facile d'utilisation, que l'on soit novice ou non en bio-informatique et surtout qu'il puisse être utilisé en routine, notamment en cas d'épidémies. Une approche phylogénétique permettra à l'utilisateur de voir les liens entre les différentes populations. En plus de la mise en place de cet outil, nous voulions accompagner l'utilisateur dans sa démarche avant la soumission des séquences au CRISPOL *in silico*. En effet, avant que les données de SEG puissent être analysées, il faut s'assurer que les séquences ou les assemblages soumis soient les plus adaptés possibles pour un meilleur résultat.

Avec une collection historique mais également récente de plus de 300 souches de *S. enterica* sérotype Typhimurium, nous avons testé et comparé différentes approches (assemblage *de novo* et alignement) et différents logiciels de traitement de séquences (CLC assembler, Velvet, Ray et SPAdes pour l'assemblage et Bowtie2, BWA et Blastall pour l'alignement), à la méthode CRISPOL Luminex®, méthode de référence lors de cette étude. Après une mise en

place du CRISPOL *in silico*, une application directe sur plus de 800 souches de *S. enterica* sérotype Typhimurium isolées durant le pic saisonnier de ce sérotype, permettra sa validation en routine.

Ce sera l'occasion de redéfinir une nouvelle nomenclature des souches de *Salmonella* tout en restant corrélée aux anciennes méthodes de typage de façon à ne pas perdre les données accumulées au cours de ces 80 dernières années. L'étude de la coévolution entre le chromosome (SNPs du core génome) et les deux régions CRISPR permettront d'affiner cette nomenclature.

## I. Matériels et méthodes

### *Souches bactériennes*

Pour cette étude, 364 isolats de *Salmonella enterica* sérotype Typhimurium dont 35 proviennent de l'étude Fabre *et al.*<sup>57</sup> ont été analysés. Ces isolats ont été sélectionnés pour être représentatifs de la diversité de ce sérotype tant au niveau temporel (isolés entre 1911 et 2012), géographique (comprenant l'Europe,  $n = 199$  ; l'Afrique,  $n = 83$  ; l'Asie,  $n = 26$  et l'Amérique,  $n = 16$ ), mais également au niveau génétique par le séquençotype (ST) MLST (29 ST différents) et le CRISPOL Type Luminex® (229 CT différents). Les caractéristiques des génomes sont présentées en Annexe 1a.

Pour la surveillance épidémiologique, 868 isolats de *S. enterica* sérotype Typhimurium ont été sélectionnés entre Juin 2016 et Novembre 2016 (pique saisonnier de *Salmonella*). Parmi ces souches, 102 ont vu leur CT déterminés par la technologie Luminex® en guise de référence.

### *Extraction de l'ADN*

L'ADN total a été extrait avec le kit Instagene (Biorad) ou le kit Maxwell 16 cell DNA purification kit (Promega, Madison WI), pour le typage par le CRISPOL et le séquençage entier du génome respectivement, selon les recommandations du fabricant.

Pour l'étude épidémiologique, l'ADN total a été extrait en utilisant le MagNAPure 96 system (Roche).

### Typage CRISPOL Luminex®

Le CRISPOL a permis le sous-typage à haut débit de *S. enterica* sérotype Typhimurium et de son variant monophasique, en utilisant la technologie du Luminex®, comme précédemment décrit.<sup>57</sup>

### Séquençage entier du génome

Pour le séquençage entier du génome, 364 isolats de *S. enterica* sérotype Typhimurium ont été séquencés. Le séquençage a été réalisé à la plateforme génomique de l'Institut Pasteur sur deux séquenceurs différents, Illumina MiSeq ( $n = 7$ ) et Illumina HiSeq 2500 ( $n = 357$ ), (tableau 1) générant des reads de 118 à 300 pb paired-end.

Pour la partie étude épidémiologique, le séquençage a été réalisé avec le NextSeq 500 à la plateforme de microbiologie mutualisée (P2M) à l'Institut Pasteur. Les reads générés avaient une taille de 150 pb paired-end.

Les reads ont été soumis à Enterobase (<http://enterobase.warwick.ac.uk/>).

### Autres génomes étudiés

Deux génomes publiés de *S. enterica* Typhimurium, DT104, souche représentative du clone multi-résistant DT104 (numéro d'accèsion GenBank NC\_002569), et SO4698-09, souche représentative du clone Européen multi-résistant monophasique (numéro d'accèsion GenBank NZ\_LN999997), ont été téléchargés sur GenBank et des reads de 100 pb ont été générés. Les reads de la souche *S. enterica* sérotype Weltevreden 840K, utilisée pour enraciner l'arbre, ont été téléchargés.

Séquenceur	Taille de l'insert (pb)	Nombre d'isolats séquencés	Nombre de CRISPOL Luminex®
MiSeq	150	7	7
	300	7	
HiSeq	108	69	357
	138	42	
	146	246	
	Total	357	
NextSeq 500	150	868	102

Tableau 1. Distribution des génomes séquencés par les plateformes MiSeq, HiSeq 2500 et NextSeq 500.

### Détection des SNPs

Pour la détection des SNPs, tous les reads ont été alignés avec le génome de référence *S. enterica* sérotype Typhimurium LT2, comprenant le chromosome (numéro d'accèsion GenBank AE006468) et le plasmide pSLT (numéro d'accèsion GenBank AE006471), avec Bowtie2 version 2.1.0. Les SNPs ont été identifiés avec SAMtools version 0.1.19 comme décrit précédemment.<sup>177</sup> Les SNPs présents dans les séquences d'insertion identifiés avec ISfinder (<https://www-is.biotoul.fr/>) ou dans les régions recombinantes identifiés par Gubbins ont été enlevés, comme décrit précédemment.<sup>177</sup>

### Analyses phylogénétiques

L'arbre phylogénétique de maximum de vraisemblance a été construit avec 27,721 SNPs issus de l'alignement de tous les génomes de *S. enterica* au génome de référence de *S. enterica* sérotype Typhimurium LT2. RAxML version 7.8.6 (modèle de substitution GTRGAMMA) a été utilisé pour la construction de l'arbre. Afin de vérifier la robustesse de l'arbre, 100 bootstrap ont été réalisés. L'arbre a été enraciné sur le génome *S. enterica* sérotype Weltevreden 840K. La version finale de l'arbre a été visualisée avec R version 3.3.3.

### Analyses MLST

La méthode MLST *in silico* a été réalisée avec l'outil « Achtman 7 Genes » sur Enterobase, de même pour les groupes MLST eBURST.

### CRISPOL *in silico* à partir des reads

Une stratégie basée sur l'alignement des reads a été développée pour l'extraction du CRISPOL Type (CT) à partir du séquençage entier du génome. D'une manière générale, les reads bruts sans aucun contrôle qualité ni filtrations ont été alignés à 79 couples de séquences répétées (DR) et espaceurs d'une soixantaine de nucléotides (41 DR/espaceurs comprenant dix variants pour le CRISPR1 et 38 DR/espaceurs comprenant un variant pour le CRISPR2), avec Bowtie2 ou CLC mapper. Les paramètres utilisés sont indiqués dans le tableau 2. Ces paramètres ont été adaptés pour un alignement de séquences paired-end, en étant le plus strict possible. De cette manière, les variants des espaceurs STM03, STM07B, STM12, STMB8, qui diffèrent d'un seul SNP peuvent être détectés. Les variants de l'espaceur STM18, peu discriminants n'étaient pas pris en compte pour la détermination du CT lors du CRISPOL Luminex®. Cependant, avec le séquençage entier du génome, les variants de taille

variable ont dû être ajoutés aux séquences de référence pour une meilleure détection de cet espaceur. Les reads doivent couvrir en moyenne 87% (90/103 nucléotides) de la séquence DR/STM07A-B d'une taille plus longue et 73% (45/62 nucléotides) de la séquence de tous les autres couples DR/espaceur. Cette taille a été choisie pour être supérieure à la longueur des séquences répétées et éviter toutes ambiguïtés. Tous les DR/espaceur ayant une profondeur de couverture de trois reads sont considérés comme étant potentiellement présents dans le génome testé, les autres étant considérés comme des artefacts. Le schéma du CRISPOL est ensuite reconstitué afin d'être comparé à la base de données CT contenant les 990 profils. Lorsqu'un nouveau CT est trouvé, l'alignement est relancé mais avec des paramètres moins stricts, notamment au niveau de la longueur et de la couverture des reads. (60% [37/62 nucléotides] au lieu de 73%).

Après cette étape d'alignement, si un nouveau CT est trouvé, les reads dont le format d'origine est un fichier fastq indiquant les qualités de séquençage pour chaque base, sont convertis en fichier fasta (séquence nucléotidique seule). Un alignement utilisant Blastall est effectué entre le fichier multi-fasta et la base de données. Avec les paramètres choisis (tableau 2), tous les DR/espaceurs trouvés sont gardés pour former le schéma CRISPOL.

#### *CRISPOL in silico à partir des assemblages de novo*

Une approche consistant à extraire le CRISPOL à partir des assemblages *de novo* a également abouti à l'élaboration d'un script. Plus complexe que l'approche précédente, le script réalise un alignement avec Blastall entre le génome *de novo* et la base de données. Il recherche dans un premier temps le gène *iap*, gène qui initie le début du CRISPR1 et donc du CRISPOL. La fin de la région CRISPR1 est déterminée par l'identification des gènes *cas* séparant les deux régions CRISPR. Enfin, le gène *ycgF* indique la fin du CRISPR2. Ces gènes permettent non seulement la détection des deux régions CRISPR mais également leur orientation au sein du génome. Après avoir repéré le CRISPR1 et le CRISPR2, le script va détecter les DR et les espaceurs séparément. Les nouveaux espaceurs peuvent être identifiés et implémentés à la base de données.

	Outil	Paramètres	
		Défaut	Utilisé pour l'étude
<b>Assemblage <i>de novo</i></b>	Ray version 2.0.0	k-mer = 21	k-mer = 41
	Velvet version 1.2.10 + Velvet optimizer version 2.2.5	k-mer screening	k-mer = 41
	Spades version 3.6.0	k-mer = 21, 33, 55, 77	k-mer = 41
	CLC assembler (CLCbio) version 4.4.0	k-mer = 21-22	k-mer = 41
<b>Alignement</b>	Blastall version 2.2.26	-e (Expectation value) = 10 -F (filter query sequence) = T	-e 1e-25 -F F
	CLC mapper (CLCbio) version 4.4.0	-l 0.5 -s 0.8 -r random -p fb ss 180 250	-l 0.21 -s 1 -r ignore -p fb ss 100 750
	Bowtie2 version 2.1.0	-sensitive-local mode	-very sensitive-local mode -L 32

**Tableau 2. Liste des versions et paramètres utilisés pour l'assemblage *de novo* et l'alignement.**

### *CRISPOL in silico nouveau CT*

En cas de nouveau CT, une matrice de distance est réalisée : le nouveau schéma CRISPOL trouvé est retranscrit en mode binaire (0 pour absence de l'espaceur et 1 pour sa présence) et comparé à la base de données des différents profils CRISPOL. Cette matrice va être utilisée pour la construction d'un arbre phylogénétique avec la méthode UPGMA, permettant ainsi de voir quel CT connu est génétiquement proche du nouveau CT. Enfin, une comparaison de ces deux CT (connu et nouveau), permet d'identifier les différences en espaceurs qui composent le CRISPOL.

Toutes ces étapes sont résumées dans la figure 18.

## CRISPOL *in silico* pipeline

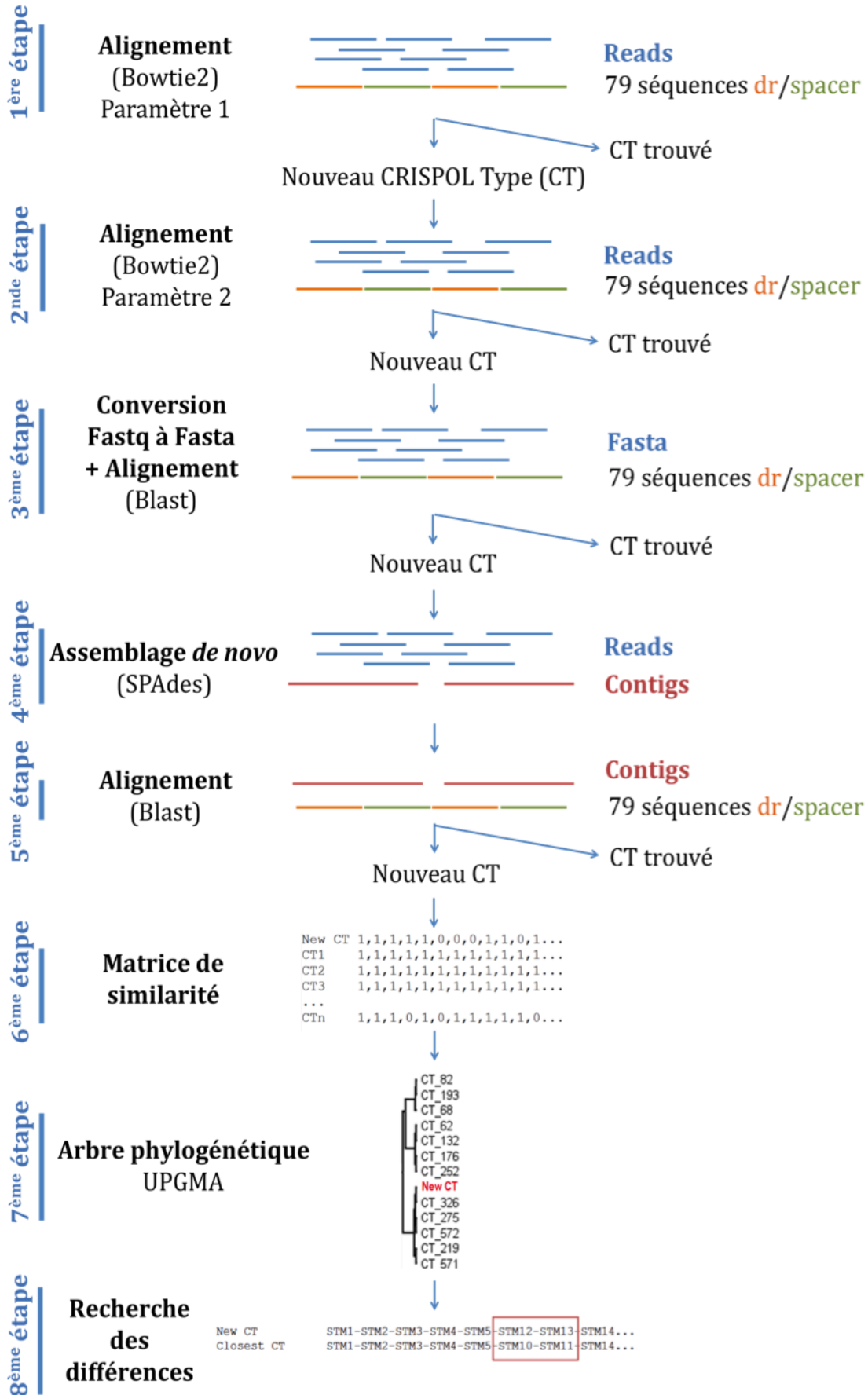


Figure 18. Pipeline du CRISPOL *in silico*.

## II. Résultats

### *CRISPOL in silico*

Les différentes étapes du script de sous-typage des *S. enterica* sérotype Typhimurium par extraction du CRISPOL *in silico*, sont représentées dans la figure 18 et les détails pour chaque étape décrits dans le matériel et méthode.

### *CRISPOL in silico à partir des reads : méthode d'alignement*

Pour tester l'approche de l'alignement, deux méthodes ont été évaluées : un alignement des séquences qualitées (fastq) et un alignement des séquences brutes (fasta) contre la base de données. Pour la première méthode, une comparaison a été faite entre CLC mapper et Bowtie2. Seulement 1% (3/364) des profils CRISPOL n'ont pas pu être extraits (graphique 3). Les erreurs rencontrées sont essentiellement dues à une absence des espaceurs localisés à la fin du CRISPR1 ou au début du CRISPR2.

Avec l'approche des séquences multi-fasta, tous les profils CRISPOL ont pu être extraits correctement. Ces deux méthodes permettent une extraction du CRISPOL en moins de 0.25 h pour CLC mapper et Bowtie2 et 0.15 h pour Blastall.

### *Analyse statistique des assembleurs de novo*

Nombreux sont les logiciels d'assemblage *de novo* qui ont été développés avec l'avancée du séquençage entier du génome. Afin de déterminer l'outil le plus adapté à notre étude, quatre logiciels d'assemblage ont été comparés : Ray, Velvet, SPAdes et CLC assembler. Pour cela, nous avons évalué le N50, le nombre de contigs générés et le temps mis par chaque logiciel pour assembler un génome. Chaque génome a été assemblé au moins deux fois afin de tester la répétabilité de chaque logiciel. Ray et SPAdes génèrent des génomes dont le N50 a une valeur moyenne de 128 kpb et 150 kpb respectivement. Celle de Velvet est 2.7 fois plus grande avec une valeur de N50 de 347 kpb. CLC assembler se situe au milieu avec un N50 de 237 kpb. Concernant le nombre de contigs, Ray et CLC assembler produisent le moins de contigs (88 et 68 respectivement). Viennent ensuite Velvet et SPAdes avec 105 et 110 contigs à la fin de l'assemblage. Au niveau du temps généré, le plus long est Ray avec une valeur moyenne de 3 h pour reconstruire un génome. Les trois autres sont presque équivalents, SPAdes étant le plus rapide (0.3 h), vient ensuite CLC assembler (0.4 h) et enfin Velvet (0.6 h) (tableau 3).



Outil	Insert (pb)	N50 (kpb)	Nombres de contigs	Temps d'assemblage (h)
RAY	108	110	80	3
	138	100	110	
	146	175	75	
VELVET	108	390	95	0,5
	138	250	140	
	146	400	80	
CLC	108	220	60	0,3
	138	210	90	
	146	280	55	
SPAdes	108	150	110	0,2
	138	120	120	
	146	180	100	

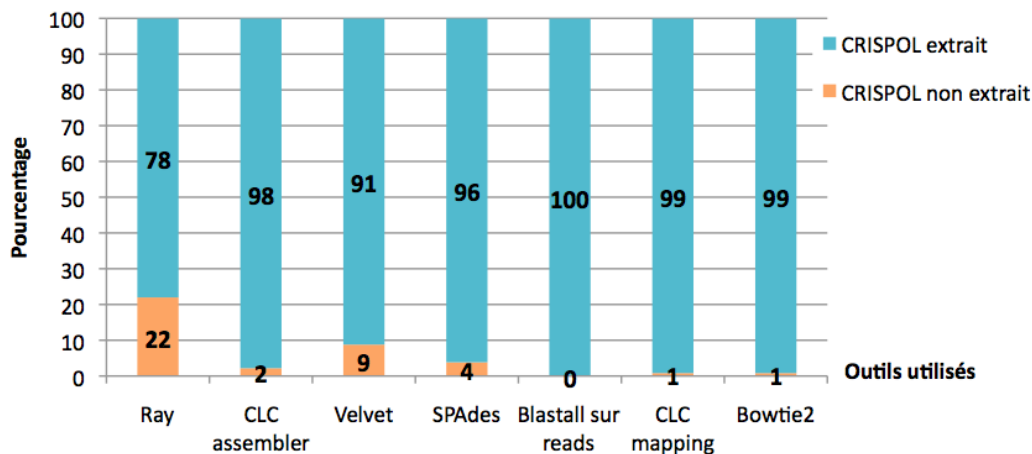
Tableau 3. Caractéristiques des génomes générés par l'assembleur.

### CRISPOL *in silico* à partir d'assemblages de novo

Une analyse préliminaire du sous-typage *in silico* à partir des génomes assemblés montre que les résultats diffèrent d'un outil à un autre. Le script n'a réussi à extraire que 78% (284/364) des profils CRISPOL à partir des génomes assemblés avec Ray. Les génomes reconstruits avec CLC assembler et SPAdes ont présenté de meilleurs résultats avec 98% (356/364) et 96% (350/364) de réussite respectivement. Velvet a quant à lui assemblé des génomes dont 91% (332/364) des CRISPOL ont pu être extraits (graphique 3).

Parmi les erreurs observées lors de l'extraction des profils CRISPOL, toutes présentaient des espaceurs manquants comparé aux résultats du CRISPOL Luminex®, méthode de référence. Sur les 22% (80/364) de CRISPOL mal assemblés avec Ray, 44% (35/80) était dû à une absence de l'espaceur STMB13. Il en est de même pour Velvet, sur les 9% de profil CRISPOL mal assemblés, 44% (14/32) d'entre eux n'avaient pas l'espaceur STMB13. Pour CLC assembler et SPAdes, aucune erreur concernant le STMB13 n'a été observée. Les autres erreurs fréquemment rencontrées sont généralement localisées en début ou fin des régions CRISPR1 et CRISPR2. De plus, 21% (3/14) des profils CRISPOL assemblés avec SPAdes ne contenaient pas les espaceurs STMB10 et STMB11.

En regardant le génome dans son ensemble et les différents contigs qui le compose, pour chacune de ces erreurs, notamment l'absence des espaceurs STMB13, STMB10 et STMB11 ; le CRISPR2 est toujours assemblé sur deux contigs différents, chacun comprenant les espaceurs adjacents à ceux manquants.



**Graphique 3. Pourcentage des résultats obtenus par chaque outil avec le CRISPOL *in silico*.**

Le pourcentage de profil CRISPOL ayant pu être extrait est indiqué en bleu et en orange le pourcentage de profil CRISPOL mal assemblé sur une totalité de 364 génomes.

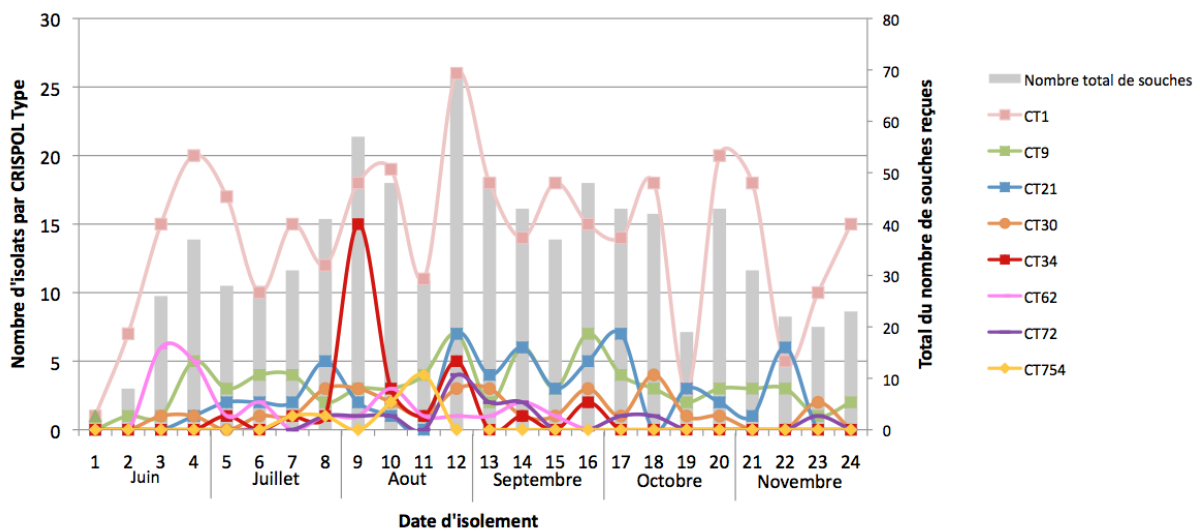
### *Impact du séquenceur et de la taille de l'insert sur le CRISPOL in silico*

Sachant que la qualité des données générées par le SEG peut avoir une influence sur l'extraction *in silico* du CRISPOL, nous avons regardé parmi les génomes comportant des erreurs si le séquenceur et la taille des inserts utilisés avaient un impact sur le résultat final. Ici, deux séquenceurs ont été utilisés : MiSeq et HiSeq 2500. La taille des inserts variait de 108 pb à 300 pb. De manière générale, aucune différence significative n'a pu être observée. Le séquenceur et la taille de l'insert n'ont ici aucune influence sur l'assemblage du CRISPR.

### *Application épidémiologique*

Afin de vérifier la l'efficacité du CRISPOL *in silico*, nous l'avons testé en routine durant une période de six mois correspondant au pic saisonnier (de Juin à Novembre) du pathogène *S. enterica* sérotype Typhimurium. Au total, 868 souches ont été séquencées puis analysées, parmi lesquelles 12% (102/868) ont été passées au CRISPOL Luminex®, en guise de contrôle. Au total, 69 CT connus différents ont été trouvés, avec une large majorité de CT1 (39%, 339/868), souvent associé à l'émergence du variant monophasique 1,4,[5],12 :i :- en Europe, suivi du CT9 (9%, 76/868) et du CT21 (7%, 58/868) connu pour être associé au clone DT104. Dans le graphique 4 sont représentés les CT majoritaires trouvés (CT1, CT9, CT21 et CT30) et d'autres CT plus rares (CT62, CT72 et CT754), servant de contrôle négatif à la détection d'épidémies. Durant la première semaine d'Août, il est observé que le CT34 resté jusqu'alors rare, voit son nombre de souches augmenter : de 2% (3/128) au mois de Juillet, il passe à 12% (24/203) pour le mois d'Août, indiquant la présence d'une possible épidémie. Après

investigation, il s'agissait d'une épidémie de *Salmonella enterica* sérotype Typhimurium, liée à des poulets rôtis contaminés.



**Graphique 4. Application du CRISPOL *in silico* pour une étude épidémiologique de 868 isolats de *S. enterica* sérotype Typhimurium isolés entre Juin et Novembre.**

Les statistiques sont représentées par semaine. Les barres grises indiquent le nombre total de souches reçues par semaine. Les lignes de couleurs indiquent le nombre de souches représentées pour chaque CT par semaine.

#### Coévolution CRISPR et SNPs.

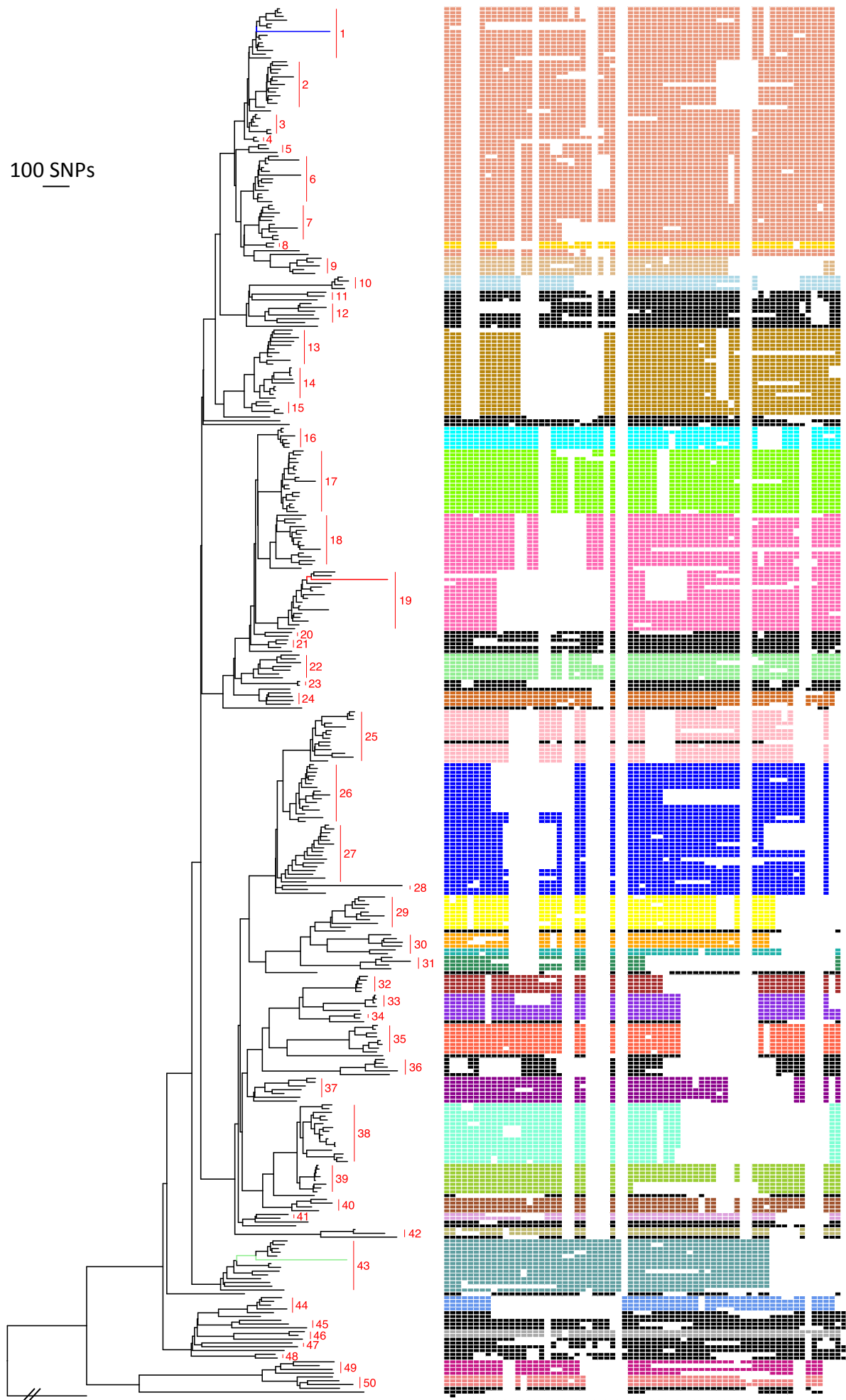
L'analyse CRISPOL des 364 souches a montré leur appartenance à 229 CT différents avec les CT117 et CT46 les plus souvent représentés (3.02%,  $n = 11/364$  ; 2.75%,  $n = 10/364$  respectivement). Ces 229 CT ont été regroupés dans 30 CRISPOL Groupe (CG) différents. Un CG est défini comme un groupe comprenant deux ou plusieurs CT qui diffèrent de cinq espaceurs au plus. La délétion est définie comme un seul évènement contigu. A partir de cette définition, les CG1, CG3 et CG9 sont les plus importants (comprenant 16 CT, 37 CT et 17 CT respectivement). Parmi les CT du CG1, nous retrouvons le clone de multi-résistance DT104.

La majorité des souches de *S. enterica* sérotype Typhimurium appartiennent au groupe MLST eBG 1 (91.5%, 333/364), suivi du groupe MLST eBG 138 (4.8%, 17/364) et du groupe MLST eBG 364 (2.7%, 10/364). Au total, 20 ST différents ont été trouvés *in silico*. Le ST19 (eBG1) étant largement majoritaire (81%, 295/364), suivi du ST36 (eBG138, 4.4%, 16/364).

Une phylogénie des 364 souches a été construite sur la base de 27,721 SNPs du core génome trouvés après un alignement avec le génome de référence *S. enterica* sérotype Typhimurium LT2. L'arbre phylogénétique montre une répartition des souches en 50 clades (chaque clade doit comporter au moins deux souches) différents (figure 19). Les deux CT

principaux CT117 et CT46 appartiennent aux clades 17 et 13 respectivement. Le CG1 comprend les clades 18 et 19 et CG3 de 1 à 7. Le groupe variant monophasique de *S. enterica* sérotype Typhimurium (ayant ici pour référence la souche S04698-09), largement représenté par le CT1, appartient au CG13, correspondant au clade 43. Ce clade est caractérisé par la présence de l'espaceur STM31 et contient huit autres CT tous porteur de cet espaceur. Plus distant des autres clades, les clades 44 à 48 regroupent tous les ST36, caractérisés par la présence de l'espaceur BraB14.

100 SNPs



CG1 CG2 CG3 CG4 CG5 CG6 CG7 CG8 CG9 CG10 CG11 CG12 CG13 CG14 CG15 CG16 CG17 CG18 CG19 CG20 CG21 CG22 CG23 CG24 CG25 CG26 CG27 CG28 CG29 CG30

**Figure 19. Relation phylogénétique de 368 génomes de *S. enterica* sérotype Typhimurium et leur profil CRISPOL correspondant.**

L'arbre phylogénétique basé sur le maximum de vraisemblance a été construit à partir de 27,721 SNPs appartenant au core génome grâce à RAxML. La séquence de référence *S. enterica* sérotype Typhimurium LT2 est indiquée par une branche bleue, celle du clone DT104 en rose et celle de la souche monophasique SO4698-09 en vert. La souche *S. enterica* sérotype Weltevreden 840K a été utilisée pour enraciner l'arbre et la longueur de sa branche a été divisée par 100, comme indiqué par les doubles barres (//), afin d'améliorer la visualisation. A gauche et en rouge sont indiqués par des numéros les différents clades. A droite sont indiqués les profils CRISPOL obtenus avec la méthode du CRISPOL Luminex®. Les carrés de couleurs et les carrés blancs indiquent la présence et l'absence des espaceurs respectivement. Chaque couleur correspond à un CRISPOL groupe (CG) auquel appartient le CT. Les profils CT de couleur noire indique qu'ils sont trop éloignés génétiquement pour appartenir à un CG. Le même arbre avec les bootstraps et le nom des isolats est présenté en Annexe 1 b.

### III. Discussion

Durant cette étude, nous avons mis en place un nouvel outil de sous-typage de *S. enterica* sérotype Typhimurium, basé sur la méthode du CRISPOL, à partir du SEG. Deux approches d'extraction du profil CRISPOL ont été comparées, celle de l'alignement à partir des reads et celle à partir d'un assemblage *de novo*.

L'utilisation de séquenceurs et de taille d'inserts différents indique que ces deux critères n'avaient pas d'influence sur l'extraction du CRISPOL *in silico*. Cependant, aucun génome n'a pu être séquencé selon ces dix paramètres (deux séquenceurs et cinq tailles d'insert), la comparaison s'est faite sur des génomes différents. La raison du coût étant ici évidente, nous avons préféré nous concentrer par la suite sur l'analyse post-séquençage.

Avec une moyenne de deux millions de reads par génome, la méthode d'alignement permet une analyse beaucoup plus rapide (en moins de dix minutes), sans passer par une étape d'assemblage du génome (12 minutes pour le plus rapide et trois heures pour le plus long), suivi d'une extraction ultérieure du CRISPOL. Outre le problème de rapidité, l'approche d'assemblage *de novo* ne semble pas être la plus adaptée pour l'assemblage du CRISPOL (98% pour le meilleur assembleur et 78% pour le moins performant, contre 100% pour l'approche de l'alignement). Ces assembleurs sont basés sur la recherche de séquences ou parties de séquence chevauchantes afin de les concaténer. Or les régions CRISPR sont composées essentiellement de séquences répétées engendrant des confusions et des ambiguïtés pour l'assembleur. Pour Ray et Velvet, la majorité des erreurs rencontrées étaient dues à l'absence de l'espaceur STMB13. Ce dernier, mais également les espaceurs STM5, STM8, STM22 et STM29 sont fréquemment dupliqués<sup>57</sup> au sein des régions CRISPR, induisant en erreur les algorithmes. Pour éviter le moins d'erreurs possible, l'assembleur est programmé pour omettre les régions ambiguës.<sup>178</sup> CLC assembler et SPAdes n'ont quant à eux pas de difficultés pour réassembler un ou deux espaceurs dupliqués. Ces quatre assembleurs utilisent le même algorithme, celui du graph de De Bruijn, mais contrairement à Velvet, SPAdes est doté d'une correction des reads en utilisant un outil appelé BayesHammer<sup>179</sup> et d'une conversion du graph de De Bruijn en graphes dit rectangulaires, permettant une meilleure résolution des zones répétées.<sup>180</sup> A la fin, SPAdes corrige les erreurs des contigs générés en réalignant grâce à BWA les reads initiaux aux contigs.

Cependant, SPAdes n'arrive pas dans 21% des cas à assembler les espaceurs STMB10 et STMB11. Le CRISPOL Luminex® à lui seul ne permet que de voir l'absence et la présence des espaceurs et non les phénomènes de duplications souvent source d'erreurs comme nous venons de le voir. Pour aller plus loin, nous avons comparé 105 profils CRISPR des génomes *de novo* à leurs séquences Sanger. Les 105 séquences étaient réparties sur 96 profils CRISPR différents, 52.4% ( $n = 55/105$ ) des profils contenaient des espaceurs dupliqués et 3.8% ( $n = 4/105$ ) présentaient des blocs entiers d'espaceurs dupliqués. Les séquences Sanger montrent que pour les génomes dont les espaceurs STMB10 et STMB11 étaient manquants, des blocs entiers d'espaceurs (notamment STMB3-STMB4 et STMB5-STMB6-STMB7-STMB8-STMB9) sont répétés en amont et en aval du couple STMB10-STMB11. SPAdes n'est pas en mesure d'assembler autant de duplications.

Concernant les erreurs rencontrées en début et en fin de région CRISPR, il est à noter que l'espaceur STMB1, localisé au début du CRISPR2, est riche en homopolymères (répétitions identiques de la même base, dans notre cas « AAAAA » 5 A) et possède un pourcentage de GC (25% de GC) inférieur à la moyenne (environ 46% pour les autres espaceurs). Les espaceurs STM02, STM24 (situés en début et fin du CRISPR1 respectivement) et l'espaceur STMB35 (situé à la fin du CRISPR2), sont les trois autres espaceurs également riches en homopolymères (« AAAAA », 5 A). L'influence du pourcentage de GC au sein du génome est connue pour jouer un rôle dans la qualité de séquençage et d'assemblage du génome. Cependant, les raisons pour lesquelles le pourcentage GC affecte l'assemblage ne sont pas encore claires.<sup>181</sup> Les homopolymères tout comme les régions répétées, sont des problèmes récurrents durant le séquençage, entraînant des erreurs d'assemblage.<sup>182</sup>

En conclusion, pour une utilisation optimale du CRISPOL *in silico*, l'utilisateur doit privilégier une soumission des reads et éviter les assemblages *de novo*. Cependant, si l'utilisateur préfère soumettre un assemblage, il est conseillé d'utiliser SPAdes ou s'ils sont détenteurs d'une licence, CLC assembler, pour un meilleur assemblage des deux régions CRISPR.

La comparaison coévolutive entre le CRISPOL et les variations génomiques des SNPs montre une corrélation entre les 30 CRISPOL Groupe et les 50 clades trouvés. L'approche phylogénétique basé sur les SNPs du core génome est plus discriminante et permet une subdivision des CG en sous groupes. Ces résultats confirment ceux trouvés par Fu *et al.*



L'ensemble de leurs 105 génomes étudiés montre une répartition des souches en 14 groupes répétés (RG) équivalents à nos CG et pour le core génome, cinq clades et 19 lignées. Ils en avaient conclu que leur étude était représentative de la diversité des RG disponibles à ce jour. Cependant, avec une collection composée principalement de souches récentes (années 2000) et Australiennes, un premier biais au niveau de la diversité spatiotemporelle est à noter. Avec un grand échantillonnage de plus de 360 isolats, isolés entre 1911 à 2012, répartis sur sept continents, 229 CT et 20 ST différents, nous avons montré une plus grande diversité génomique de *S. enterica* sérotype Typhimurium. Sur les 105 génomes, seulement 66 étaient disponibles sur le site NCBI. En appliquant notre script CRISPOL *in silico* sur ces génomes, nous avons trouvé des discordances avec les profils CRISPOL présentés dans leur article. C'est le cas notamment de la souche DT135 (numéro d'accèsion GenBank ERS007567) où l'espaceur STMB13 est absent pour Fu *et al.* et présent avec notre script mais également sur Enterobase. Pour assembler leur génome, Fu *et al.* ont utilisé Velvet. Or, durant cette étude, nous avons démontré l'incapacité de Velvet à assembler les espaceurs dupliqués dont STMB13. Une discordance au niveau de la souche SARA8 a également été observée avec l'absence de l'espaceur STMB23 pour Fu *et al.* Or SARA8 a déjà été passée au CRISPOL Luminex® mais également au séquençage Sanger lors d'études précédentes<sup>57</sup> et affichait un profil CT38 avec présence de l'espaceur STMB23. Au final, avec 36% (24/66 génomes) de discordances trouvées, la comparaison des groupes RG avec nos CG n'a pu être réalisée.

Leur approche reste toutefois intéressante, notamment l'analyse comparative et évolutive entre le CRISPOL avec les prophages, ces derniers étant impliqués dans l'acquisition de nouveaux espaceurs par la bactérie. Cependant, l'acquisition d'un nouvel espaceur est un phénomène rare.<sup>57</sup> La variabilité du CRISPOL se fait principalement sur la base de délétions ou de duplications d'un ou plusieurs espaceurs contigus et enfin des variations d'un ou deux nucléotides au sein d'un espaceur.<sup>57</sup>

La distribution des CT en groupes CG a permis de voir que certains de ces groupes avaient des marqueurs spécifiques, comme l'espaceur STM31 pour le CG13 ou BraB14 pour ST36. Par conséquent, une détermination du core SNPs pour chaque clade permettra l'établissement d'une nouvelle nomenclature universelle, jusque-là inexistante, pour le sous-typage de *S. enterica* sérotype Typhimurium, basée sur le génotype.

## Conclusion

Les infections à *Salmonella* non typhiques restent l'une des causes principales de maladies diarrhéiques dans le monde. *S. enterica* sérotype Typhimurium est l'un des sérotypes prédominants de salmonellose, transmise de l'animal à l'homme. Les méthodes de typage s'appuient encore largement sur une méthode établie il y a près d'un siècle, celle du sérotypage basée sur le schéma White-Kauffman-LeMinor. Cette technique et toutes celles développées par la suite visant à réaliser un sous-typage de *Salmonella* ne sont pas assez spécifiques et discriminantes pour un dépistage épidémiologique.

Avec l'avancée récente des nouvelles générations de séquençage dite de paillasse, le SEG est devenu un outil accessible à tous, permettant ainsi une utilisation au quotidien et en routine dans les laboratoires de surveillance. Cette approche permet à elle seule de remplacer la plupart des méthodes de typage et sous-typage actuelles et ouvre la voie à de nouvelles techniques de génotypage.

Ici nous avons développé le CRISPOL *in silico*, un outil de sous-typage de *S. enterica* sérotype Typhimurium, basé sur le séquençage du génome. Facile d'utilisation, cet outil sera disponible et ouvert à tous. Le script pouvant être parallélisé, il permettra une analyse rapide et à haut débit de plusieurs génomes en même temps. En plus de fournir un résultat de sous-typage, nous accompagnons l'utilisateur dans son étude, en réalisant une analyse phylogénétique incluse dans le script, leur permettant une visualisation populationnelle de leurs souches.

D'autres outils disponibles en ligne, permettent également une analyse du CRISPR. Néanmoins, en plus d'être pour certains peu fiables<sup>57</sup> ou difficiles d'utilisation, la plupart de ces outils ne possèdent pas une base de données complète et universelle des différents espaceurs, séquences répétées et surtout d'une base regroupant plus de 900 profils CRISPOL. Sans ces informations, une approche phylogénétique ne serait pas possible. En soumettant leurs génomes, l'utilisateur contribuera au développement mais également à l'implémentation de nouveaux espaceurs pour la base de données.

En parallèle, la coévolution CRISPOL et variations génomiques basées sur les SNPs, a montré une corrélation entre les deux approches indiquant que le polymorphisme des régions CRISPR serait un moyen fiable d'apprécier la distance génétique entre deux isolats.

Ayant prouvé son utilité en routine et en période de pic saisonnier de *S. enterica* sérotype Typhimurium, le script va pouvoir s'étendre à d'autres sérotypes tels que *S. enterica* sérotype Enteritidis, le deuxième sérotype prédominant de *Salmonella*.

L'étude concernant ce dernier sérotype étant toujours en cours, l'ébauche et les premiers résultats sont brièvement présentés dans la partie suivante.

# CRISPOL *in silico*, méthode de sous-typage de *Salmonella* *enterica* sérotype Enteritidis

## ETUDE EN COURS

### Introduction

Au début des années 2000, *Salmonella enterica* sérotype Enteritidis était la principale cause d'infection bactérienne d'origine alimentaire, laissant place à partir de 2005 à *S. enterica* sérotype Typhimurium et son variant monophasique 1,4,[5],12:i:-.<sup>3</sup> Le sérotype Enteritidis occupait depuis 2011 le rang de 3<sup>ème</sup> place mais en 2016, il reprend la 1<sup>ère</sup> place des sérotypes responsables des salmonelloses humaines et reste activement impliqué dans des épidémies d'amplitude européenne, comme nous avons pu le voir en 2016.<sup>183</sup> Les œufs contaminés sont le principal réservoir de *S. enterica* sérotype Enteritidis, représentant 80% des épidémies et approximativement 50,000 à 110,000 des cas associés aux œufs contaminés chaque année aux Etats-Unis.<sup>184</sup> La volaille et plus particulièrement le poulet représentent également un réservoir important de ce sérotype.<sup>185</sup>

Les méthodes de sous-typage disponibles à ce jour pour ce sérotype restent classiques, MLVA et ECP utilisées en routine, malgré leur faible pouvoir discriminant.<sup>186,44</sup>

En gardant la même approche que *S. enterica* sérotype Typhimurium, nous voulons développer un outil de sous-typage de *S. enterica* sérotype Enteritidis, basé sur le polymorphisme des régions CRISPR. Contrairement au sérotype Typhimurium où le CRISPOL Luminex® était déjà appliqué en routine, il faudra mettre en place une base de données des différents espaceurs et répertorier tous les profils CRISPOL pour le sérotype Enteritidis, en nous appuyant sur les travaux préliminaires de Fabre *et al.*

Cette étude étant actuellement en cours, nous voulions exposer ici les premiers résultats obtenus et les travaux restant à faire.

### Etude expérimentale

Pour cette étude, nous avons récupéré les profils CRISPR de 180 souches de *S. enterica* sérotype Enteritidis de l'étude Fabre *et al.*, 407 génomes du même sérotype séquencés en

routine au CNR-ESS à l'Institut Pasteur en 2016 et 8,440 génomes de *S. enterica* sérotype Enteritidis provenant du site Enterobase (<https://enterobase.warwick.ac.uk>) ont été analysés. Les caractéristiques des génomes et les données entrées sur Enterobase pouvant être erronées, la recherche des génomes s'est réduite aux principaux séquençotypes du sérotype Enteritidis, les ST11, ST74 et ST180.

Pour les 407 génomes provenant du CNR, le CRISPOL type Enteritidis (CTent) a été extrait grâce au script développé pour Typhimurium en utilisant une nouvelle base de données DR/espaceurs adaptée au sérotype Enteritidis. La liste des espaceurs est indiquée dans le tableau 4.

CRISPR1	Espaceur	Séquence
(1)	Ent1	TATTTATAAGCGTGTCATCTATGCAACCCAAC
(2)	Dub1	GGATTTTTCTGAACGTGATTCGGGATACGC
(3)	Ent2	ACCTGCCCCGACCCAATAAGGGGGCCCTCGTGA
(3)	Ent2var1	ACCTGCCCCGACCCAATAAGGAGGCCCTCGTGA
(4)	Ent3	GGCCGCTGGTCAAATTCCTCAATCTGAGCAATC
(5)	Ent4	ATAGCCCCGGCAGCGATAGCTAAACCAGTTCC
(6)	Ent5	GCCTCAAATCTCTCGGTGAGATGTAAGCGTC
(6)	Ent5var1	TCCTCAAATCTCTCGGTGAGATGTAAGCGTC
(7)	Ent6	ACCAGTGGTCAGCGGCGGATGAATTTGCCCTG
(8)	Ent10	CGGCGCAATTGCCACATGCTTTATTGTGCTTC
(9)	Ent11	AGCTAATACCACGCCAGCATTAGAAAAACCGC
(10)	Ent7	GAGAATGCTCATGCGCGTGAGCGCCATATATT
(10)	Ent7var1	GAGAATGCTCATGCGTGAGCGCCATATATT
(11)	Ent9	CATGGCAATTTTACGGCGGACGTGCTCGCTCT
(12)	Ent8	AGGCGGACCGAAAAACCGTTTTTCAGCCAACGT
(13)	Ent12	TCCAAATCCGGTGAGGCTGATATCAAACCCGT

CRISPR2	Espaceur	Séquence
(1)	EntB0	GGCTACACGCAAAAATTCCAGTCGTTGGCGCA
(2)	EntB1	CCGATTAAGATCCGCGAGTCTGCATCAGTAACT
(3)	EntB2	CGATTCTACGGCAACAGGCCAGGCTGCGACCG
(4)	IndB1	ATGACATGACCGTGACTAGCGAGAATGTCCCA
(5)	EntB24	GAGGCCATGATGCTAATCGCTGAGCAGCAGGA
(6)	EntB11	ATTACTTAACAAAGAATCATATCGTTGCTGTA
(7)	EntB12	CACCACGACGCATAAGAGGCAACGCAATGCTC
(8)	EntB13	TTCAACCTGGCGCAGATCGCTATCGAGGAATT
(9)	EntB14	TTAATGAAGTTGGAGACTGTGGGTGTGATCTC
(10)	EntB15	TTTATTTAAAAAAGGCATACGGATTATGAATC
(11)	EntB16	CCGACAACGAAACCGCGCAAAGTGATGCTCT
(12)	EntB17	TCAAGCTGGCTAATCTGCGTGTGATGTTGTC
(13)	EntB18	ACCACGGCAGACCTGTGTACGTCTAACCCAC
(14)	EntB19	CGAAAATCAGCCCGCATATTCGACACATCGC
(15)	EntB21	GTGCAGGCGACTGGAAAAGACGATATATACGC

(16)	EntB20	CCGACACATCCACGGCCAGCGACATGGGGGCC
(17)	EntB10	GGTTCAGGATTCAGCCCCACCGAAACCGGCG
(18)	EntB22	CTGCAATATGGGAAATGATTATTTAATTATTT
(19)	EntB23	GACGGAAATATCAGGCCATATCCGCCTATAGT
(20)	EntB3	ATCAAACATGGAAACCCCTTTAATGAGAGCAA
(21)	EntB4	TCAGGAACGCGCGGCGGAAGAGCTTGGTGTTTG
(22)	EntB5	GCTGCCTTTCCCGGAGTTCCGGCCCTAAATT
(23)	EntB6	TCATGCGCTATAAAAATCAGACTGTCCACATGC
(24)	EntB7	TGATTATTGACGACAACAGCACAGACCGGCAG
(25)	EntB8	AATAATCGGCAATTTGTCCTGGACAGGCACGG
(26)	EntB9	GAATCTGGAGGCCAACAGCGCGGCGAAATCCT
(26)	EntB9var1	GAATCTGGAGGCCAACAGCGCGGCGAAATACT

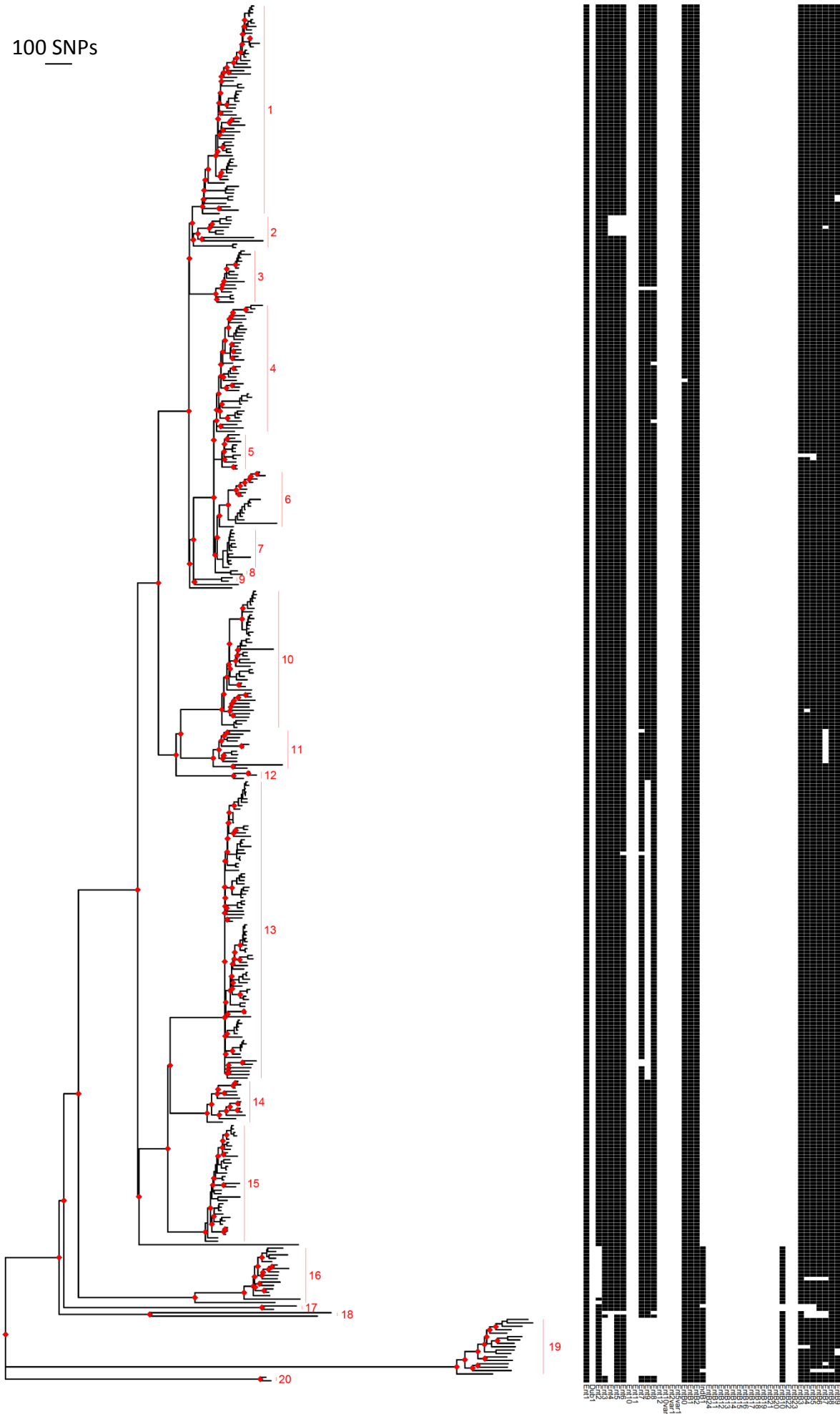
**Tableau 4. Liste des espaceurs utilisés pour le CRISPOL *in silico* Enteritidis**

Au total, sur 9,027 CRISPOL extraits, plus de 220 CTent différents ont été trouvés, contre 990 CT pour 13,024 souches de *S. enterica* sérotype Typhimurium. Cependant, deux profils CTent que nous nommerons CTent1 et CTent2 sont largement prédominants. CTent1 représente 27% (2,465/9,027 CTent) et CTent2 53% (4,783/9,027) des CTent, correspondant à 80% de la totalité des CTent. Il est donc nécessaire de trouver une autre méthode pouvant les discriminer.

Comme pour le sérotype Typhimurium, nous avons comparé les profils CTent obtenus avec un arbre phylogénétique basé sur 16,357 SNPs du core génome. La séquence P125109 (numéro d'accèsion GenBank NC\_011294.1) a été utilisée comme génome de référence. Pour des raisons de visibilité, les 8,847 génomes n'ont pas pu être inclus et seuls les 405 génomes séquencés au CNR-ESS sont représentés (figure 20). Malgré un manque de diversité (souches isolées en France en 2016 essentiellement), le choix s'est porté sur ces génomes en vue d'une application en routine au CNR-ESS.

Sur les 405 génomes, les profils CRISPOL étaient distribués sur 25 CTent différents, avec 21 % (85/405 génomes) de CTent1 et 61% (246/405 génomes) de CTent2. L'arbre phylogénétique basé sur les SNPs du core génome montre la division des profils CTent en 20 clades différents. Le CTent1 se trouve principalement dans le clade 13 et le CTent2 est réparti sur 14 clades différents (clade 1 à 12, puis 14 et 15).

100 SNPs



**Figure 20. Relation phylogénétique de 405 génomes de *S. enterica* sérotype Enteritidis et leur profil CRISPOL correspondant.**

L'arbre phylogénétique de maximum de vraisemblance a été construit à partir de 16,357 SNPs du core génome grâce à RAXML. A droite sont indiqués les profils CRISPOL obtenus avec la méthode du CRISPOL *in silico*. Les carrés noirs et blancs indiquent la présence et l'absence d'espaces respectivement.

Les noeuds dont les bootstraps ont une valeur  $\geq 0.9$  sont indiqués par des losanges rouges. A côté de l'arbre sont indiqués les différents clades.

## Conclusions

Les premiers résultats obtenus indiquent que l'homogénéité génétique du sérotype Enteritidis avec la prédominance de deux CTent, ne permet pas son sous-typage par la méthode CRISPOL. Une comparaison avec les SNPs du core-génome montre qu'il est possible de différencier ces CTent en plusieurs sous-ensembles. Pour continuer cette étude, la prochaine étape serait d'étudier plus en détails les SNPs de chaque clade, afin d'établir une nouvelle nomenclature et base de données.



L'outil présenté dans le premier article montre tout l'intérêt du SEG dans le sous-typage à haut débit et en routine de *Salmonella*. Son pouvoir discriminant et sa base de données déjà établit faciliteront les échanges entre laboratoires.

La deuxième mission du CNR-ESS est de suivre l'évolution des souches émergentes résistantes aux antibiotiques. Connaître le mécanisme d'apparition de ces dernières permettra de mieux comprendre ce phénomène qui reste un réel problème de santé publique de nos jours.

L'article qui suit publié dans Lancet Infectious Diseases en 2017 est une étude rétrospective sur la transmission de l'ampicilline chez *S. enterica* sérotype Typhimurium dans les années 1950.

## **Chapitre VI**

### **SEG : résistance aux antibiotiques, étude rétrospective**

#### **(Article 2)**

## Abstract

L'ampicilline, première pénicilline semi-synthétique active contre les *Enterobacteriaceae*, a été mise sur le marché en 1961. Les premiers foyers épidémiques impliquant des souches résistantes à l'ampicilline de *Salmonella enterica* sérotype Typhimurium ont été identifiés au Royaume-Uni en 1962 et 1964. Le but de cette étude rétrospective était d'étudier l'émergence de cette résistance dans des isolats historiques de *S. enterica* sérotype Typhimurium.

Pour cela, 288 isolats de *S. enterica* sérotype Typhimurium prélevés entre 1911 et 1969 dans 31 pays et provenant de diverses sources ont été analysés. Un test de sensibilité aux antibiotiques par la méthode des disques par diffusion a été réalisé sur l'ensemble des isolats. Des expériences de transfert de la résistance ont été menées sur les isolats résistants à l'ampicilline. Afin de connaître la structure des populations résistantes à l'ampicilline, nous avons séquencé le génome entier de 225 isolats, sélectionnés pour être représentatif de la diversité (sources, spatiotemporelles et génétiques).

11 (4%) des 288 isolats étaient résistants à l'ampicilline par acquisition de gènes de  $\beta$ -lactamase, dont blaTEM-1, portés par différents plasmides, y compris le plasmide de virulence de *S. enterica* sérotype Typhimurium. Un isolat produisant la  $\beta$ -lactamase TEM-1 a été isolé en France en 1959 et deux isolats producteurs de  $\beta$ -lactamase TEM-1 ont été isolés en Tunisie en 1960, avant la commercialisation de l'ampicilline. Les vecteurs de résistance à l'ampicilline étaient différents de ceux rapportés au Royaume-Uni dans les années 1960.

L'association entre l'utilisation d'antibiotiques et la sélection des déterminants de la résistance n'est pas toujours évidente. Nos résultats suggèrent que l'utilisation de pénicillines à spectre étroit en tant que promoteur de croissance dans le bétail pourrait avoir favorisé la propagation des plasmides porteurs du gène blaTEM-1 chez *S. enterica* sérotype Typhimurium à la fin des années 1950.

# Transmission de la résistance à l'ampicilline chez *Salmonella enterica* sérotype Typhimurium à la fin des années 1950 : une étude rétrospective par séquençage entier du génome

Alicia Tran Dien, Simon Le Hello, Christiane Bouchier, François-Xavier Weill

**Article soumis à Lancet Infectious Diseases le 28 Août 2017 et accepté le 9 octobre 2017.**

**(Voir Annexe 2f pour la version anglaise).**

## Introduction

En mai 2015, la 68<sup>ème</sup> Assemblée Mondiale de la Santé a approuvé un plan d'action mondial pour lutter contre la résistance bactérienne aux antibiotiques devenue un problème majeur de santé publique pour le 21<sup>ème</sup>.<sup>187</sup>

La découverte de la pénicilline par Alexander Fleming en 1928 a sans nul doute ouvert une nouvelle ère dans la lutte contre les infections bactériennes, éclipsant celle de la sulfamidothérapie.<sup>188</sup> La pénicilline est une substance bactéricide naturelle issue du *Penicillium rubrum*, dont la synthétisation et la production de masse ont débuté dans les années quarante, durant la Seconde Guerre mondiale.<sup>188</sup> Cependant, dès le début, A. Fleming observa que la pénicilline, contrairement aux bactéries à coloration de Gram-positive telles que les staphylocoques, pneumocoques et autres streptocoques, était inefficace contre les bactéries à coloration de Gram-négative de type « coli-typhoïde ». <sup>189</sup> La première pénicilline pouvant inhiber la croissance des entérobactéries, dite pénicilline à large spectre, fut l'ampicilline, un dérivé semi-synthétique de l'acide-6-aminopénicillanique<sup>190</sup>, commercialisé par Beecham Research Laboratories sous le nom de « Penbritin » en 1961.<sup>191</sup>

En 1965 furent publiées les premières observations décrivant des souches d'Enterobacteriaceae productrices de bêta-lactamase et pouvant transmettre leur résistance à l'ampicilline par un facteur nommé « R » (plasmide).<sup>192</sup> Dans cet article, il est rapporté que deux des trois souches d'Enterobacteriaceae étudiées étaient productrices de pénicillinase de type TEM. Parmi elles, une souche de *E. coli* isolée en Grèce en 1963 et une *S. enterica* sérotype Paratyphi B isolée au Royaume-Uni également en 1963, tous deux porteur du gène *bla*<sub>TEM-1A</sub>, localisé sur un plasmide de type IncX2 et un plasmide IncF d'environ 40 kb et 100

kb respectivement (tableau 5).<sup>193,194</sup> La troisième souche, une *S. enterica* sérotype Typhimurium ayant un phage de type 1a (reclassé plus tard en phage-type DT2), portait une bêta-lactamase au profil différent de TEM (tableau 5). Isolée en 1962 à Brighton, au Royaume-Uni, elle coïncide avec la première épidémie causée par une souche du même sérotype, résistante à l'ampicilline. A cette époque, des dizaines de souches de *S. enterica* sérotype Typhimurium DT2 résistantes à l'ampicilline, la streptomycine, aux sulfamides et à la tétracycline ont été retrouvées chez l'homme et la filière porcine.<sup>195</sup> Le plasmide d'une de ces souches, R<sub>1818</sub>, renommé plus tard en R46,<sup>193,194</sup> a été séquencé (numéro d'accèsion GenBank AY046276) et s'est avéré être un plasmide de type IncN d'environ 50 kb, portant le gène *bla*<sub>OXA-2</sub>.

En 1963 et 1964, plusieurs souches épidémiques de *Salmonella enterica* capables de transmettre la résistance à l'ampicilline ont été identifiées en Europe (table 1).<sup>196,197</sup> En particulier *S. enterica* sérotype Typhimurium DT29 ayant émergé suite aux nouvelles pratiques agricoles intensives pour l'élevage des veaux au Royaume-Uni.<sup>198</sup> En 1965, 1,297 et 576 isolats de ce pathogène ont été prélevés sur du bétail et chez l'homme respectivement. Plus de 96% de ces isolats étaient multi-résistants aux antibiotiques. Parmi eux, la souche RT1 qui possède un plasmide, NTP1, résistant à l'ampicilline<sup>199</sup>, de petite taille (8,3 kb) non-conjugatif et porteur du gène *bla*<sub>TEM-1</sub>.<sup>197</sup>

Cette chronologie suggère que l'utilisation de l'ampicilline pourrait être la cause principale de la sélection de souches résistantes de cet agent pathogène d'origine alimentaire.

Afin d'améliorer notre compréhension sur l'émergence de la résistance à l'ampicilline, un antibiotique encore largement utilisé à ce jour, nous avons étudié une grande collection de souches historiques de *S. enterica* sérotype Typhimurium, un agent zoonotique resté l'un des deux sérotypes prédominants, causant des maladies infectieuses en Europe depuis la Seconde Guerre mondiale.<sup>200,201</sup> Ces 288 isolats ont été collectés entre 1911 et 1969, dont 107 ont été isolés avant la mise sur le marché de l'ampicilline.

Agent pathogène	Origine	Profil de résistance	Plasmide	Type	gène <i>bla</i>	Autres gènes de résistance	Référence
<i>S. Typhimurium</i> DT2	1962, RU, Humain et porc	ApStrSulTe	R <sub>1818</sub> = R46*	51kb, IncN	OXA-2 <sup>Ω</sup>	<i>aadA1,sul1,tet</i> (C)	6,10
<i>S. Paratyphi</i> B	1963, RU, Humain	ApKCSul	R <sub>7268</sub> = R1 <sup>†</sup>	97.5kb, IncF	TEM-1A	<i>aadA1, aph(3')-Ic, catA1,sul1</i>	6,8
<i>E. coli</i>	1963, Grèce, Humain	ApStr	R <sub>TEM</sub> = R6K <sup>#</sup>	39kb, IncX2	TEM-1A	<i>strA/B</i>	6,9
<i>S. Derby</i>	1963, France, Humain	ApStrKCSulTe	R <sub>4-5</sub>	-	-	-	11
<i>S. Oranienburg</i>	1964, France, Humain	ApStrKCSulTe	R <sub>7</sub>	-	-	-	11
<i>S. Typhimurium</i> DT29	1964, RU, Humain et bétail	ApStrSulTe	NTP1 <sup>NC</sup>	8.3kb	TEM-1	None	12-14

**Tableau 5. Caractéristiques des premiers isolats de *S. enterica* (et *E. coli* TEM) porteurs de plasmides de résistance à l'ampicilline**

Ap, ampicilline; Str, streptomycine; K, kanamycine; C, chloramphenicol; Sul, sulfamides; Tmp, triméthoprime; Te, tétracycline; -, indéterminé; DT, phage-type définitif; disponible sur GenBank avec le numéro d'accèsion AY046276\* ou KY749247<sup>†</sup> ou depuis le site du Wellcome Trust Sanger Institute<sup>#</sup>, <sup>NC</sup> non-conjugatif; <sup>Ω</sup> gène dupliqué.

## I. Matériels et méthodes

### *Souches bactériennes*

Pour cette étude, les 288 isolats de *S. enterica* sérotype Typhimurium analysés proviennent de sources différentes (comprenant l'homme,  $n = 149$  ; les animaux et l'alimentation,  $n = 113$ ) et de zones géographiques diverses (comprenant l'Europe,  $n = 161$  ; l'Afrique,  $n = 90$  ; l'Asie,  $n = 19$  ; l'Amérique,  $n = 15$ ), entre 1911 et 1969 (Annexe 2a et Annexe 2e). Ils appartiennent aux collections du Centre National de Référence des *E. coli*, *Shigella* et *Salmonella* (CNR-ESS) ( $n = 277$ ), de la Collection de l'Institut Pasteur (CIP) ( $n = 9$ ), et du Centre collaborateur de l'Organisation Mondiale de la Santé de Référence et de Recherche pour les Salmonella ( $n = 2$ ), Institut Pasteur, Paris, France. Tous les isolats du CNR-ESS ont été conservés dans des tubes de conservation à température ambiante. Ils n'ont pas été manipulés depuis leur mise en conserve (généralement une à deux semaines après leur isolement à partir des échantillons biologiques). Tous les autres isolats ont été conservés - 80°C.

### Tests de sensibilités aux antibiotiques

La sensibilité aux antibiotiques a été déterminée par la méthode des disques de diffusion sur milieu gélosé Mueller-Hinton, avec un panel de 32 antibiotiques (Bio-Rad, Marnes-la-Coquette, France), comme décrit précédemment.<sup>202</sup> La concentration minimale inhibitrice (CMI) de l'ampicilline et de la pénicilline G (également connu sous le nom de benzylpénicilline) de 67 isolats a été déterminée à l'aide de bandelettes Etest (BioMerieux, Marcy L'Etoile, France) (Annexe 2e).

### Détermination du transfert de la résistance

Une expérience de transfert de la résistance a été réalisée sur 11 souches de *S. enterica* sérotype Typhimurium résistantes à l'ampicilline sur milieu solide. Les souches *E. coli* K-12 J5 Azi<sup>R</sup> et *S. enterica* sérotype Typhimurium STM 36-57 Nal<sup>R</sup>, un mutant résistant à l'acide nalidixique généré au laboratoire à partir d'une souche sauvage STM 36-57 (IncF-), ont été utilisées en tant que souches réceptrices pour le test de transférabilité plasmidique par conjugaison et pour quantifier la fréquence de transférabilité respectivement. La conjugaison a été réalisée sur un filtre placé soit sur du milieu gélosé Trypto-Caséine-Soja (TCS) (lorsque la souche réceptrice *S. enterica* sérotype Typhimurium STM 36-57 Nal<sup>R</sup> était utilisée) soit sur Drigalski (lorsque la souche réceptrice *E. coli* K-12 J5 Azi<sup>R</sup> était utilisée). Les transconjugants ont été sélectionnés sur le même type de milieu gélosé utilisé pour la conjugaison, tout en ajoutant soit de l'ampicilline (75 mg/L) ou de la pénicilline G (15 mg/L) et de l'acide nalidixique (64 mg/L) ou de l'azide de sodium (500 mg/L). Pour les conjugaisons où aucuns transconjugants n'a été obtenus sur milieu gélosé TCS, une autre conjugaison a été réalisée en milieu minimum M9.<sup>15</sup> Trois transconjugants ont été aléatoirement sélectionnés pour chaque expérience. Le typage plasmidique s'est basé sur la méthode du replicon typing par PCR.<sup>202</sup> Pour les conjugaisons faites entre deux souches de *S. enterica* sérotype Typhimurium, un typage CRISPOL (voir ci-dessous) a été réalisé afin de distinguer les souches réceptrices des donatrices. Chaque expérience de conjugaison a été faite trois fois. La fréquence de conjugaison a été calculée en divisant le nombre de transconjugants (dénombrés sur milieu gélosé contenant de l'acide nalidixique et de l'ampicilline ou de la pénicilline G) par le nombre total de souches réceptrices (dénombré sur milieu gélosé contenant seulement de l'acide nalidixique). Les valeurs indiquées sont une moyenne des trois expériences.

### Extraction de l'ADN

L'ADN total a été extrait avec le kit Instagene (Biorad) ou le kit Maxwell 16 cell DNA purification kit (Promega, Madison WI), pour le typage par le CRISPOL et le séquençage entier du génome respectivement, selon les recommandations du fabricant.

### Typage CRISPOL Luminex®

Le CRISPOL a permis le sous-typage à haut débit de *S. enterica* sérotype Typhimurium et de son variant monophasique, en utilisant la technologie du Luminex®, comme précédemment décrit.<sup>57</sup>

### Séquençage entier du génome

La sélection de 225 génomes (provenant de l'étude CRISPOL *in silico*) sur 288 souches de *S. enterica* sérotype Typhimurium s'est faite afin d'obtenir une large couverture en terme de source, spatiotemporelle et diversité génétique (basée sur le typage CRISPOL) (Annexe 2a et Annexe 2e). Le séquençage s'est déroulé à la plateforme génomique de l'Institut Pasteur avec le séquenceur Illumina HiSeq 2500 générant des reads de 118 à 300 pb paired-end, avec une couverture moyenne de 233X (valeur minimale 45X, maximale 960X). Les séquences brutes ont été soumises à Enterobase (<http://enterobase.warwick.ac.uk/>).

La plus vieille souche de *S. enterica* sérotype Typhimurium résistante à l'ampicilline, ExPB 5-59 a également été séquencée par GATC Biotech (Konstanz, Allemagne), par la plateforme PacBio RS II (Pacific Biosciences, CA, Etats-Unies). Les erreurs de séquençage au niveau des homopolymères générées par PacBio ont été corrigées sur la base du séquençage Illumina. La séquence plasmidique a été annotée avec le serveur RAST (<http://rast.nmpdr.org/>), avec une inspection et correction manuelle. La séquence a été déposée sur GenBank sous le numéro d'accèsion MF428416.

### Autres génomes étudiés

Deux génomes publiés de *S. enterica* Typhimurium, DT104, souche représentative du clone multi-résistant DT104<sup>203</sup> (numéro d'accèsion GenBank NC\_002569), et SO4698-09, souche représentative du clone Européen multi-résistant monophasique (numéro d'accèsion GenBank NZ\_LN999997)<sup>204</sup>, ont été téléchargés sur GenBank, et des reads de 100 pb ont été générés. Les reads de la souche *S. enterica* sérotype Weltevreden 840K, utilisée pour enraciner l'arbre, ont été téléchargés sur EnteroBase.



### Alignement de séquences et détection des SNPs

Pour la détection des SNPs, tous les reads ont été alignés avec le génome de référence *S. enterica* sérotype Typhimurium LT2, en prenant le chromosome (numéro d'accèsion GenBank AE006468) et le plasmide pSLT (numéro d'accèsion GenBank AE006471), avec Bowtie2 version 2.1.0. Les SNPs ont été identifiés avec SAMtools version 0.1.19 comme décrit précédemment. Les SNPs présents dans les séquences d'insertion ont été identifiés avec ISfinder (<https://www-is.biotoul.fr/>) ou dans les régions recombinantes identifiés par Gubbins ont été enlevés, comme décrit précédemment.<sup>177</sup>

### Analyses phylogénétiques

L'arbre phylogénétique de maximum de vraisemblance a été construit avec 32,741 SNPs chromosomaux issus de l'alignement de tous les génomes de *S. enterica*. RAxML version 7.8.6 (modèle de substitution GTRGAMMA) a été utilisé pour la construction de l'arbre. Afin de vérifier la robustesse de l'arbre, 100 bootstraps ont été réalisés. L'arbre a été enraciné sur le génome *S. enterica* sérotype Weltevreden 840K. La version finale de l'arbre a été visualisée avec FigTree version 1.4.2.

### Assemblage de novo

Tous les reads pour chaque souche ont été assemblés *de novo* avec SPAdes version 3.6.0.<sup>132</sup>

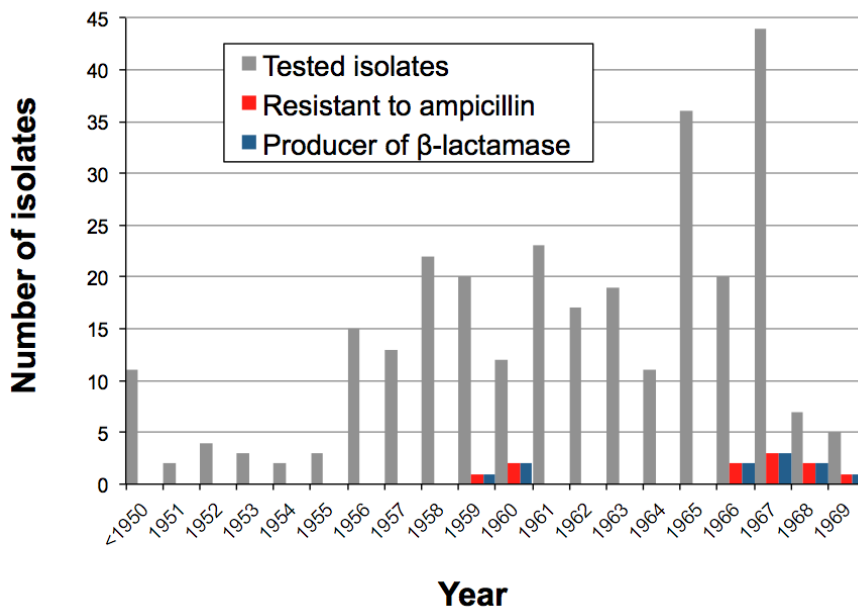
### Analyses génétiques

La méthode MLST et le sérotypage *in silico* ont été réalisés avec les outils d'Enterobase « Achtman 7 Genes » et « Serotype Prediction (SISTR) » respectivement. La présence des gènes de résistance aux antibiotiques (GRAs) ou des structures contenant des GRA ont été déterminées avec ResFinder version 2.1 (<https://cge.cbs.dtu.dk/services/ResFinder/>), PlasmidFinder version 1.3 (<https://cge.cbs.dtu.dk/services/PlasmidFinder/>), pMLST version 1.2 (<https://cge.cbs.dtu.dk/services/pMLST/>) à partir des assemblages de SPAdes. Les résultats de CRISPOL Luminex® (présence ou absence des 72 séquences ADN de 25 à 32 pb) ont été confirmés pour chaque isolat, par la recherche des 72 séquences au sein de la séquence du génome entier (soit à partir des reads, soit par l'assemblage SPAdes).

## II. Résultats

### Test de sensibilité aux antibiotiques des souches historiques

Des tests de sensibilité aux antibiotiques ont été effectués sur 288 souches historiques de *S. enterica* sérotype Typhimurium, provenant de diverses sources et zones géographiques, isolées entre 1911 et 1969 (Annexe 2e). La plupart de ces isolats sont sensibles à tous les antibiotiques testés (253/288, 87.8%). Cependant, onze (11/288, 3.8%) d'entre eux étaient résistants à l'ampicilline, dont six (6/288, 2.1%) résistants à d'autres antibiotiques tels que les sulfamides, le triméthoprime, la tétracycline et les aminoglycosides (tableau 6, Annexe 2e et graphique 5). Ces onze isolats ont été prélevés chez l'homme ( $n = 10$ ), en France ( $n = 8$ ) et en Tunisie ( $n = 3$ ). Trois (3/107, 2.8%) des 107 isolats prélevés entre 1911 et 1960 (c'est-à-dire avant la mise sur le marché de l'ampicilline) étaient résistants à l'ampicilline. Ces souches ont été prélevées chez l'homme, en France ( $n = 1$ ) et en Tunisie ( $n = 2$ ), en 1959 et 1960 respectivement.



**Graphique 5. Nombre d'isolats de *S. enterica* sérotype Typhimurium résistants à l'ampicilline et producteurs de bêta-lactamase, par année d'isolation.**

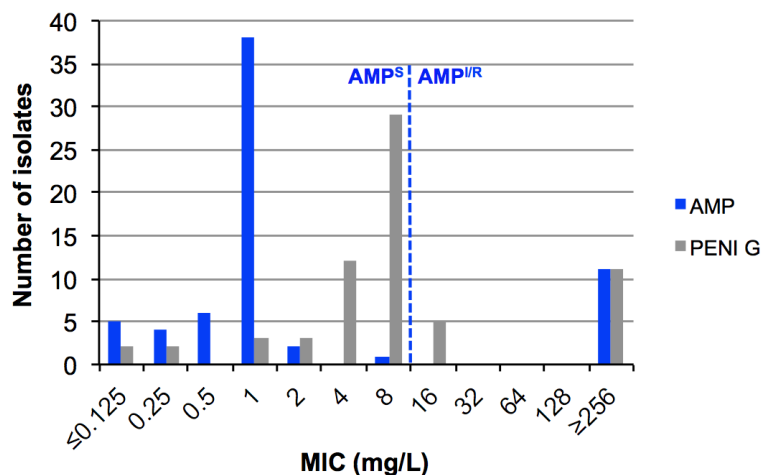
En gris est indiqué le nombre total d'isolats de *S. enterica* sérotype Typhimurium étudiés, en rouge ceux résistants à l'ampicilline et en bleu, les producteurs de bêta-lactamase.

### Composition de la population des souches historiques

Les 225 souches historiques de *S. enterica* sérotype Typhimurium dont le génome entier a été séquencé, appartiennent au groupe MLST eBurst 1 (98.7%, 222/225), 138 (0.4%, 1/225) et 243 (0.9%, 2/225). Les séquences types (STs) ST19 (89.3 %, 201/225), suivi du ST313 et ST302 (1.3%, 3/225 chacun) sont les plus fréquents. Au total, 123 CRISPOL types ont été observés, avec une prédominance du CT117 (4.9%, 11/225).

### Caractérisation des souches résistantes à l'ampicilline

Les onze souches de *S. enterica* sérotype Typhimurium résistantes à l'ampicilline détectées par la méthode des disques de diffusion et listées dans le tableau 6 ont une CMI à la pénicilline (ampicilline ou pénicilline G)  $\geq 256$  mg/L (tableau 6 et graphique 6). Pour tous ces isolats, l'analyse génomique a identifié un gène codant pour la bêta-lactamase correspondant aux gènes *bla*<sub>TEM</sub> ( $n = 9$ ) ou *bla*<sub>OXA</sub> ( $n = 2$ ). Cinquante-six isolats de *S. enterica* sérotype Typhimurium dépourvus de gènes codant pour la bêta-lactamase ont été sélectionnés pour être représentatifs de chaque clade génomique de la collection (Annexe 2b). Leur CMI à l'ampicilline varie entre 0.06 et 8 mg/L (avec une médiane de 1 mg/L), tandis que leur CMI à la pénicilline G varie entre 0.125 à 6 mg/L (avec une médiane de 8 mg/L) (graphique 6 et Annexe 2e).



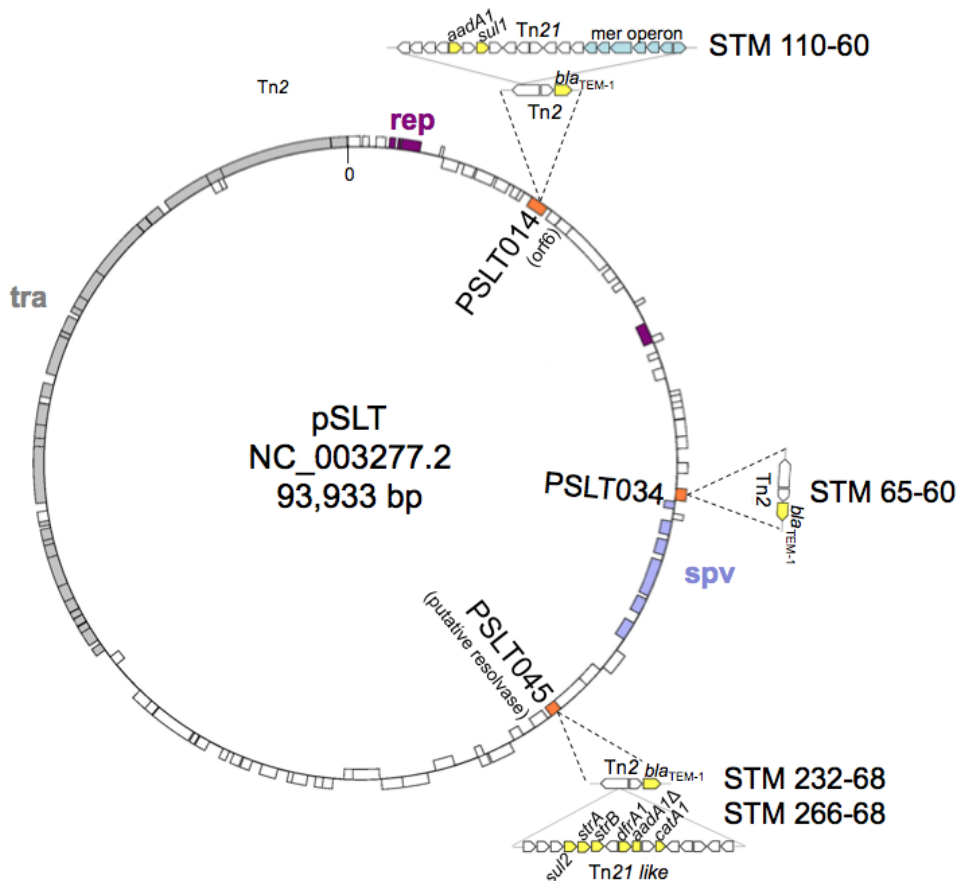
**Graphique 6. Distribution de la CMI de l'ampicilline et de la pénicilline G, sur une sélection de 67 isolats de *S. enterica* sérotype Typhimurium.**

La limite distinguant une souche sensible d'une souche résistante à l'ampicilline selon l'EUCAST ((S  $\leq 8$  mg / L ; R > 8 mg / L) et le CLSI (S  $\leq 8$  mg / L ; I = 16 mg / L ; R  $\geq 32$  mg / L) est représentée par une barre bleue en pointillée.

Parmi les gènes de bêta-lactamase, *bla*<sub>TEM-1B</sub>, porté par le transposon Tn2<sup>205</sup>, est le plus fréquent et a été trouvé chez huit isolats (8/11, 72.7%) de *S. enterica* sérotype Typhimurium

résistants à l'ampicilline. Le transposon Tn2 a été localisé sur des plasmides du groupe d'incompatibilité IncF ( $n = 5$ ) et IncX1 ( $n = 3$ ).

Pour quatre isolats, une analyse des jonctions du transposon Tn2 montre que ce dernier est localisé sur un plasmide de 93 kb, de type IncF, le plasmide pSLT, connu comme étant le plasmide de virulence de la souche de *S. enterica* sérotype Typhimurium LT2.<sup>15</sup> Le Tn2 s'est intégré à trois endroits différents sur ce plasmide pSLT (figure 21). Pour les isolats STM 232-68 et STM 266-68, au séquençotype ST313, isolés en France et Tunisie en 1968, le site d'intégration de Tn2 était identique à celui de pSLT-BT. Ce plasmide de multi-résistance aux antibiotiques, appartient à la lignée II du clone ST313, une population du sérotype Typhimurium généralement localisée en Afrique.<sup>206</sup> De plus, comme observé pour pSLT-BT, un élément mobile similaire au transposon Tn21 (appelé Tn21-like), localisé dans les plasmides IncF des souches 232-68 et STM 266-68, s'est également intégré dans le Tn2. Cependant, ces deux plasmides diffèrent légèrement du pSLT-BT à cause de la présence d'une séquence d'insertion, IS1 de 6.8 kb, englobant l'intégron (In2) du Tn21-like, induisant une délétion (région localisée entre les nucléotides 41,666 et 48,543, numéro d'accèsion GenBank FN432031). Le plasmide IncF du cinquième isolat (STM 366-66) dérive du plasmide NR1/R100 (sans le transposon Tn21) (GenBank numéro DQ364638). Cet isolat ne contenait pas le plasmide de virulence pSLT.



**Figure 21. Les sites d'intégration du transposon Tn2 au sein du plasmide de virulence IncF pSLT chez cinq isolats de *S. enterica* sérotype Typhimurium résistants à l'ampicilline.**

Les gènes dans lesquels le Tn2 s'est intégré sont représentés en orange avec leur nom et leur fonction (entre parenthèses), selon le numéro d'accèsion GenBank NC\_003277.2. Les gènes de résistance aux antibiotiques sont indiqués en jaune. Les gènes du plasmide impliqués dans le transfert (*tra*), la réplication (*rep*) et la virulence (*spv*) sont indiqués en gris, violet et lavande respectivement.

Le plasmide de 45 kb pExPB5-59-1 de type IncX1 contenant le transposon Tn2 a également été séquencé avec la technologie PacBio. Ce plasmide est présent chez une population clonale (ST19-CT160) de bactéries isolées en France sur une période de plus de dix ans (1959 à 1969). Le corps principal du plasmide pExPB5-59-1 dépourvu du Tn2 est similaire à un plasmide conjugatif de 42.2 kb (pYU39\_IncX ; numéro d'accèsion GenBank CP011431) retrouvé chez une population de souches envahissantes de type ST213 de *S. enterica* sérotype Typhimurium, isolées au Mexique dans les années 2000. Dans notre étude, cinq autres isolats de *S. enterica* sérotype Typhimurium portent également un plasmide IncX1 mais sont dépourvus de gènes de résistance aux antibiotiques (Annexe 2e). Ces plasmides ont seulement 45 à 77% de leur séquence d'ADN (avec 98 à 99% d'identité nucléotidique) en commun avec pExPB5-59-1. Il est intéressant à noter que le plus vieux de ces isolats a été collecté durant une épidémie d'origine alimentaire à Newcastle, au Royaume-Uni en 1911.

Les trois isolats de *S. enterica* sérotype Typhimurium collectés en France et en Tunisie avant la mise sur le marché de l'ampicilline contenaient le transposon Tn2, porté par le plasmide IncX1 ( $n = 1$ ) et le plasmide IncF dérivé du pSLT ( $n = 2$ ). Ce transposon s'est inséré à deux endroits différents au sein du plasmide IncF.

Isolat	Année	Pays	Hôte	Profil de résistance	Gène bla	Autres résistances	ST	CT	Plasmide
ExPB 5-59	1959	France	Humain	Ap	TEM-1B	-	19	160	IncX1 <sup>†‡</sup>
STM 110-60*	1960	Tunisie	Humain	ApStrSpSul	TEM-1B	<i>aadA1</i>	19	896	IncF <sup>†‡</sup>
STM 65-60*	1960	Tunisie	Humain	Ap	TEM-1B	-	19	896	IncF <sup>†‡</sup>
STM 345-66	1966	France	Humain	ApStrSpSulTe	OXA-2	<i>aadA1,sul1,tet(C)</i>	19	46	Incl1, IncN <sup>Ω</sup>
STM 366-66	1966	France	Humain	ApSul	TEM-1B	<i>sul1</i>	19	46	Incl1, Col,IncF <sup>†‡</sup> #
ExPB 63-67	1967	France	Humain	Ap	TEM-1B	-	19	160	IncX1 <sup>†‡</sup>
STM 84-67	1967	France	Humain	ApSpKCSulTe	OXA-1	<i>aph(3')-Ia,</i> <i>aadA1,catA1,sul1,</i> <i>tet(B)</i>	19	46	Col, IncF
STM 127-67	1967	France	Humain	ApK	TEM-1A	<i>aph(3')-Ia</i>	19	46	Incl1 <sup>†‡</sup>
STM 232-68	1968	France	Humain	ApStrCSulTmp	TEM-1B	<i>strA/B,catA1,sul2,</i> <i>dfrA1</i>	313	810	IncF <sup>†‡</sup>
TM 266-68	1968	Tunisie	?	ApStrCSulTmp	TEM-1B	<i>strA/B,catA1,sul2,</i> <i>dfrA1</i>	313	810	IncF <sup>†‡</sup>
ExPB 37-69	1969	France	Humain	Ap	TEM-1B	-	19	160	IncX1 <sup>†‡</sup>

**Tableau 6. Caractéristiques des 11 isolats de *S. enterica* sérotype Typhimurium résistants à l'ampicilline identifiés dans cette étude.**

Ap, ampicilline; Str, streptomycine; Sp, spectinomycine; K, kanamycine; C, chloramphénicol; Sul, sulfamides; Tmp, triméthoprime; Te, tétracycline; ?, indéterminé; \*, les deux isolats ont été reçus sans aucune informations épidémiologique, de l'Institut Pasteur de Tunis en Aout (STM 65-60) et Octobre (STM 110-60) 1960, en même temps que d'autres isolats de *Salmonella* spp., *Shigella* spp. et entéropathogénique *E. coli*; localisation du gène de bêta-lactamase sur le plasmide démontrée par conjugaison<sup>†</sup> et/ou analyses génomiques<sup>‡</sup>; <sup>Ω</sup> plasmide incomplet de IncN (fragment de 8.8 kb fragment similaire à R46, comprenant la région allant de *arda* à *repA*); # plasmide IncF dérivé de NR1/R100 sans Tn21, les autres plasmides IncF dérivent du plasmide de virulence pSLT.

### Transfert de la résistance à l'ampicilline

Une conjugaison a été réalisée entre les onze isolats de *S. enterica* sérotype Typhimurium résistants à l'ampicilline avec la souche réceptrice STM 36-57 NaI<sup>R</sup> en présence d'ampicilline (75 mg/L), pour déterminer la transférabilité des différents gènes *bla* portés par les plasmides. Les neuf isolats ayant le gène *bla*<sub>TEM</sub> porté par le plasmide ont transféré la résistance à l'ampicilline par conjugaison. La fréquence de transfert allait de  $2.4 \times 10^{-4}$  (pour le plasmide Incl1 de la souche STM 127-67) à  $1.6 \times 10^{-8}$  (plasmide IncF dérivant du pSLT de la souche STM 266-68) transconjugants/réceptrice (Annexe 2c). L'utilisation de milieu minimum M9 a été requise pour obtenir des transconjugants avec la souche STM 110-60. Les

deux isolats portant le gène *bla*<sub>OXA</sub> (STM 84-67 et STM 345-66) n'ont pas transféré leur résistance à l'ampicilline par conjugaison.

Les quatre plasmides porteurs du transposons Tn2 (le plasmide IncX1 de l'isolat ExPB 5-59, les plasmides IncF dérivant du pSLT des isolats STM 65-60 et STM 110-60 et le plasmide IncF dérivant du NR1/R100 de l'isolat STM 366-66) ont été transférés entre les deux souches de *S. enterica* sérotype Typhimurium en présence de pénicilline G (15 mg/L). La fréquence de transfert variait de  $1.2 \times 10^{-3}$  (plasmide IncX1 de la souche ExPB 5-59) à  $3.0 \times 10^{-6}$  (plasmide IncF dérivant du pSLT de la souche STM 110-60) transconjugants/réceptrice.

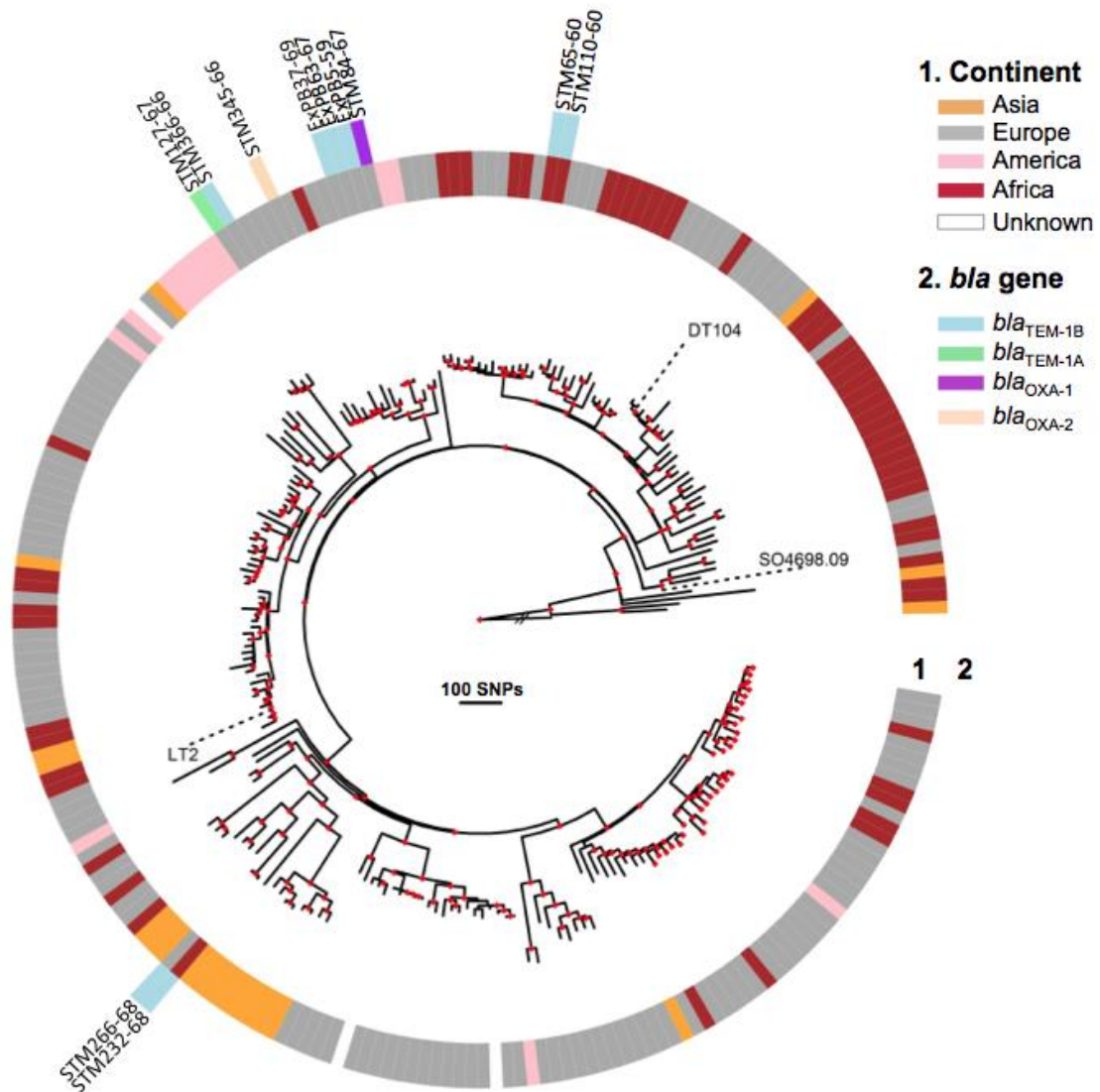
### *Relation génétique entre les souches résistantes à l'ampicilline*

Une étude phylogénétique du génome basée sur 32,741 SNPs, répartis uniformément sur le core génome sans les parties répétitives et recombinantes a montré le regroupement des onze isolats de *S. enterica* sérotype Typhimurium résistants à l'ampicilline en trois groupes (figure 22 et Annexe 2d).

Le groupe principal comprenait sept isolats résistants à l'ampicilline collectés en France entre 1959 et 1969, avec quatre types de gènes bêta-lactamase (*bla*<sub>TEM-1A</sub>, *bla*<sub>TEM-1B</sub>, *bla*<sub>OXA-1</sub>, et *bla*<sub>OXA-2</sub>). Ces sept isolats appartenaient à un clade comprenant 13 isolats d'Europe de l'Est et du Sud, avec un isolat du Nord de l'Afrique, collectés d'humains, d'aliments et d'animaux (poussins), appartenant au ST19 et CT46 ou à des variants du CT46 (Annexe 2e).

Le second groupe de souches résistantes à l'ampicilline comprenait deux isolats humains ST19-CT896 collectés en Tunisie en 1960. Ils appartenaient à un clade composé de 14 souches ST19-CT117 (ou des dérivés de CT117) provenant principalement d'humains (à l'exception d'un isolat provenant d'une saucisse) en France, Belgique et Tunisie entre 1953 et 1963.

Le troisième groupe contenait les deux isolats ST313-CT810 résistants à l'ampicilline collectés en France et Tunisie en 1968. Ils appartenaient à un clade comprenant six isolats aux séquençotypes ST302 ( $n = 3$ ) et ST313 ( $n = 3$ ). Les quatre isolats non résistants à l'ampicilline étaient originaires de Turquie (deux ST302 et un ST313) et Tunisie (un ST302) et étaient collectés entre 1956 et 1965.



**Figure 22. Arbre phylogénétique de maximum de vraisemblance des isolats de *S. enterica* sérotype Typhimurium.**

L'origine géographique (continent) des isolats est indiquée dans l'anneau 1. Les gènes de bêta-lactamase des onze isolats résistants à l'ampicilline ainsi que leur nom sont indiqués dans l'anneau 2. Les gènes de bêta-lactamase des génomes de comparaison DT104 ( $bla_{PSE-1}=bla_{CARB-2}$ ) et de la monophasique SO4698-09 ( $bla_{TEM-1-B}$ ) ne sont pas indiqués. Le nom des génomes de comparaisons LT2, DT104 et SO4698-09 sont indiqués par des lignes en pointillées noirs. Les noeuds dont les bootstraps ont une valeur  $\geq 0.9$  sont indiqués par des losanges rouges. La souche *S. enterica* sérotype Weltevreden 840K a été utilisée pour enracer l'arbre et la longueur de sa branche a été divisée par 100, comme indiqué par les doubles barres (//), afin d'améliorer la visualisation.

### III. Discussion

Nous avons montré ici que la résistance à l'ampicilline a émergé avant sa commercialisation, chez *S. enterica* sérotype Typhimurium, avec la présence du gène bêta-lactamase  $bla_{TEM-1B}$  trouvé sur deux plasmides différents (IncX1 et IncF), chez trois isolats provenant d'Europe (France) et d'Afrique (Tunisie) collectés en 1959 et 1960. Cependant, nous n'avons pas



cherché ici à déterminer l'étendue réelle de la distribution de la résistance à l'ampicilline chez *S. enterica* sérotype Typhimurium, en raison d'un biais d'échantillonnage évident (les isolats historiques étudiés ici provenaient principalement de la France et des anciennes colonies françaises en Afrique et en Asie). Une autre limitation à notre étude était le mode de stockage des souches collectées. Presque tous les isolats ont été stockés dans des tubes de conserve et l'absence de pression de sélection par des antibiotiques n'a sûrement pas favorisé à long terme, le maintien des plasmides dans ces souches, comme indiqué précédemment pour le pathogène *Bacillus anthracis*.<sup>207</sup> La proportion d'isolats résistants à l'ampicilline dans notre sélection de *Salmonella* a sûrement été sous-estimée.

Néanmoins, l'utilisation de l'ampicilline, la première pénicilline à large spectre préconisée en 1961<sup>191</sup> pour le traitement des infections dues à des Enterobacteriaceae, n'a probablement pas abouti à la sélection de ces premiers isolats résistants à l'ampicilline. Suite à l'émergence d'une résistance multiple aux antibiotiques chez le sérotype Typhimurium DT29, E.S. Anderson du Laboratoire central de santé publique au Royaume-Uni a averti que l'utilisation de la pénicilline G et des tétracyclines comme additifs alimentaires pour le bétail (chez les volailles et le porc) et comme prophylactique chez les bovins était responsable de l'émergence des isolats multirésistants de *S. enterica* sérotype Typhimurium.<sup>199</sup> L'exposition à long terme et à des concentrations sub-thérapeutiques d'antibiotiques chez le bétail a sûrement été un élément critique pour la sélection de la résistance à l'ampicilline d'un agent zoonotique largement répandu chez des animaux dédiés à la consommation. Nous avons constaté que la médiane de la CMI pour la pénicilline G chez les isolats de *S. enterica* sérotype Typhimurium de type sauvage n'était pas particulièrement élevée, 8 mg / L, tandis que Fleming, durant ses essais rudimentaires basés sur des bouillons de *Penicillium*, a signalé que les membres du groupe "coli-typhoïde" n'ont pas été affectés par le pouvoir inhibiteur de la pénicilline.<sup>189</sup> Les faibles valeurs de CMI obtenues pour la pénicilline G contre *S. enterica* sérotype Typhimurium peuvent expliquer l'acquisition de la résistance à l'ampicilline dans un environnement où les résidus de pénicilline G atteignent des concentrations juste au-dessus des CMI de la bactérie. Par exemple, en France dans les années 1960, la pénicilline G a été autorisée en tant qu'additif alimentaire et a été utilisée à des concentrations allant de 5 à 200 mg / kg de nourriture, chez tous les animaux destinés à la consommation humaine.<sup>208</sup> La pénicilline G est sensible à l'acide gastrique, mais sa faible absorption par le système digestif peut entraîner des concentrations non négligeables dans

le tube digestif et dans les excréments. Une étude aux États-Unis en 1975 a révélé que les excréments de poulets à chair nourris aux antibiotiques contenaient de la pénicilline entre 0 et 25 mg / kg (avec une moyenne de 12,5 mg / kg).<sup>209</sup> En utilisant des concentrations similaires (15 mg / L de pénicilline G), nous avons montré que le gène *bla* porté par les plasmides IncX1 ou IncF pouvait être transféré entre des souches de *S. enterica* sérotype Typhimurium. Nos résultats ne fournissent aucune preuve d'un lien de causalité entre l'utilisation de la pénicilline G et l'émergence d'une résistance transmissible à l'ampicilline chez le bétail, mais ils suggèrent qu'un environnement dans lequel les résidus de pénicilline G étaient abondants, comme l'environnement agricole décrit par Anderson<sup>198</sup> ont favorisé à la propagation des plasmides porteurs de gène *bla* chez le sérotype de *Salmonella* le plus répandu à l'époque. Nous n'avons identifié aucun isolat résistant à l'ampicilline provenant du bétail. Cependant, cela peut refléter le fait que seulement ~ 50% des isolats d'animaux testés proviennent du bétail, avec ~ 40% des isolats de bétail originaires d'Afrique et d'Amérique du Sud. Après 1969, au Royaume-Uni, conformément aux recommandations du comité sur l'utilisation d'antibiotiques dans l'élevage et la médecine vétérinaire (connu sous le nom de Comité "Swann" en l'honneur de son président), l'utilisation de promoteurs de croissance chez le bétail a finalement été limitée aux médicaments avec peu ou pas d'applications thérapeutiques chez l'homme ou les animaux et les pénicillines en particulier, ont été identifiées comme inappropriées pour une utilisation en tant que promoteurs de croissance.

Il est intéressant de noter que les vecteurs de la résistance à l'ampicilline chez nos isolats de *S. enterica* sérotype Typhimurium, principalement en provenance de France, diffèrent de ceux impliqués dans les premiers foyers au Royaume-Uni, malgré la proximité géographique de la France et du Royaume-Uni. Ainsi, l'émergence précoce de la résistance à l'ampicilline chez *S. enterica* sérotype Typhimurium était due à de multiples transferts horizontaux de gènes *bla*, suivi d'une expansion clonale à travers différents pays, en allant de l'isolement très rare de CT160 durant une période de dix ans en France, aux centaines d'isolats de DT29 par an au Royaume-Uni. Cette situation contraste avec la dissémination rapide observée par la suite pour deux clones de *S. enterica* sérotype Typhimurium multirésistants aux antibiotiques : DT104 dans les années 80 et le variant monophasique 1,4,[5],12:i:- dans les années 2000.<sup>210,104</sup> La différence peut être due aux changements du commerce alimentaire, moins mondialisé à la fin des années 1950/début des années 1960, que dans les années 80.

## Conclusion

Les antibiotiques sont généralement prescrits pour traiter et prévenir les infections bactériennes. Cependant, la résistance aux antibiotiques touche à présent toutes les régions du monde et a atteint un niveau dangereusement élevé. Leur usage abusif et excessif accélère le phénomène de la résistance, certaines maladies bactériennes deviennent difficiles à traiter et la mise sur le marché d'une nouvelle molécule se fait de plus en plus rare.

Cette étude rétrospective nous a permis, grâce au SEG, de mettre en lumière les premières étapes de l'émergence de la résistance à l'ampicilline, première pénicilline active contre les *Enterobacteriaceae*, encore largement utilisée.

Nous avons constaté que la relation entre l'utilisation d'antibiotiques et la sélection des plasmides codant pour les déterminants de la résistance n'était pas aussi directe. Depuis le début de l'utilisation des antibiotiques, l'agriculture aux côtés de la médecine, a joué un rôle dans la sélection des agents pathogènes résistants. Ces résultats mettent en évidence la nécessité d'une approche «unique» afin d'éradiquer ce phénomène qui reste aujourd'hui un véritable problème de santé publique.

Dans les chapitres précédents, nous avons montré la contribution du SEG pour le typage en routine et le suivi évolutif des souches résistantes aux antibiotiques.

Dans ce dernier article, nous allons aborder la troisième mission du CNR : le suivi et l'évolution des sérotypes au cours du temps. Il s'agit ici d'une collaboration pour étude de métagénomique sur un sérotype rare: *Salmonella enterica* sérotype Paratyphi C.



## **Chapitre VII**

### **SEG : étude évolutive d'un sérotype rare**

## Abstract

La plupart des cas de salmonellose chez les mammifères sont causés par la sous-espèce I de *Salmonella enterica*. Certains sérotypes sont à l'origine des maladies invasives chez l'homme (Paratyphi C), le porc (Typhisuis) ou les deux (Choleraesuis). L'analyse métagénomique des os et dents d'une femme de 25 ans enterrée à Trondheim en Norvège en l'an 1200 ± 25CE, montre un génome lié au sérotype Paratyphi C.

Le séquençage de 219 génomes de *Salmonella enterica* sérotype Paratyphi C, Choleraesuis et Typhisuis montre que ces sérotypes sont très proches (avec seulement 1000 SNP dispersés dans tous leurs génomes) et partagent un même ancêtre commun. Le pan-génome est quant à lui assez diversifié avec des gains et des pertes d'ilots génomiques, de bactériophages et de plasmides.

L'étude phylogénétique datée montre des changements qui ont eu lieu au niveau du génome accessoire au cours de ces 4,000 dernières années. La plupart semblent refléter des événements aléatoires et non liés à la sélection naturelle.

Cette analyse illustre les avantages obtenus en combinant la métagénomique à l'étude populationnelle des différents sérotypes.

# Salmonella enterica sérotype Paratyphi C : une stabilité génomique qui s'étend sur un millénaire

Zhemin Zhou, Inge Lundstrøm, **Alicia Tran Dien**, Sebastián Duchêne, Nabil-Fareed Alikhan, Martin J. Sergeant, Gemma Langridge, Anna K. Fotakis, Satheesh Nair, Hans K. Stenøien, Stian S. Hamre, Sherwood Casjens, Axel Christophersen, Christopher Quince, Nicholas R. Thomson, François-Xavier Weill, Simon Y. W. Ho, M. Thomas P. Gilbert, Mark Achtman.

**Article soumis à Current Biology, en révision. Cet article est issu d'une collaboration. Les résultats présentés ci-dessus sont un récapitulatif des points essentiels. La version complète se trouve en annexe 3**

## **I. Introduction**

La sous-espèce I de *Salmonella enterica*, *S. enterica* sous-espèce *enterica* est connue pour être la première cause de salmonellose.<sup>211</sup> Cette sous-espèce est elle-même divisée en deux groupes, les salmonelles dites non typhiques et typhiques, parmi lesquelles, certaines peuvent causer des maladies invasives chez l'homme (*S. enterica* sérotype Paratyphi C), le porc (*S. enterica* sérotype Typhisuis) ou les deux (*S. enterica* sérotype Choleraesuis).<sup>53,212</sup> *S. enterica* sérotype Paratyphi C a été trouvée pour la première fois durant la Première Guerre mondiale en 1916, isolée de soldats d'Europe de l'Est souffrant de fièvre entérique (également appelée fièvre paratyphoïde).<sup>213</sup> La fièvre entérique est sévère et occasionnellement mortelle. A cette époque, le pathogène était également localisé en Afrique du Nord et en Amérique. Elle a depuis disparu d'Europe, à l'exception de voyageurs occasionnels en provenance d'Asie de l'Est ou d'Afrique.<sup>53</sup> En dehors de la communauté microbiologique, ce pathogène était très rare et peu connu en Europe et en Amérique.<sup>53</sup> En effet, il était difficile à l'époque de distinguer la fièvre typhoïde causée par *S. enterica* sérotype Typhi de la fièvre entérique. Par conséquent, aucune *S. enterica* sérotype Paratyphi C n'a été décrite ou très peu.<sup>211</sup> Les sérotypes Typhisuis et Choleraesuis, sont toujours présents dans le sud et l'est de l'Asie, *S. enterica* sérotype Choleraesuis est aujourd'hui rare en Europe, sauf chez le sanglier<sup>214</sup> et *S. enterica* sérotype Typhisuis a été éradiquée des porcs européens après un abattage de troupeaux infectés.<sup>53,212</sup>



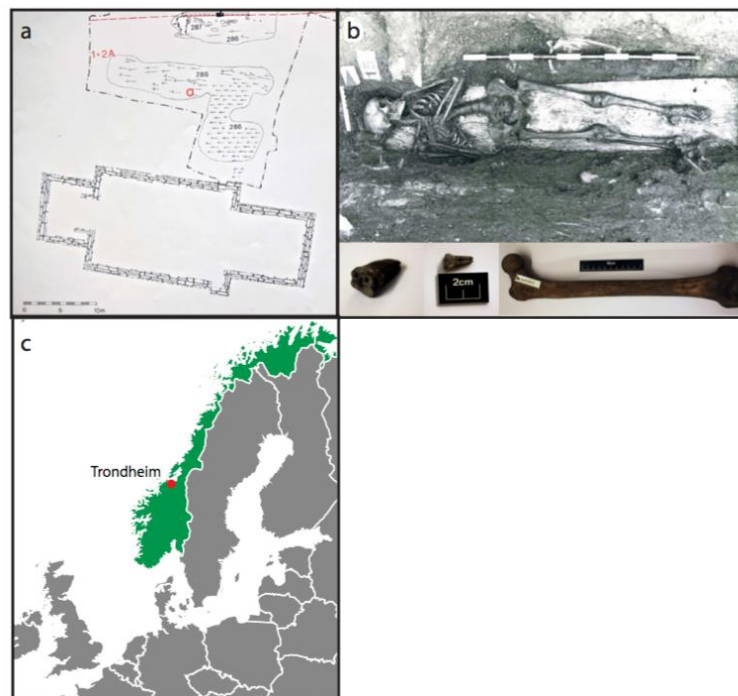
Le squelette d'une jeune femme a récemment été découvert à Trondheim en Norvège. Les analyses de datation montrent qu'elle était âgée entre 19 et 24 ans et serait décédée il y a plus de 800 ans.

Cette étude de métagénomique va permettre de mieux comprendre l'évolution du sérotype Paratyphi C, sérotype rare et de mettre en lumière les causes du décès de la jeune femme trouvée.

## II. Matériel et méthode

### Matériel

Les données de métagénomique ont été extraites à partir des dents et des os d'un squelette (SK152) d'une jeune femme (molaire supérieure droite [200 mg], pulpe radiculaire intérieure [200 mg], os fémoral [300 mg], et mélange de plaque dentaire minéralisée [30 mg] prélevé sur plusieurs dents) découvert à Trondheim en Norvège. Cette jeune femme était âgée d'une vingtaine d'années et serait décédée il y a plus de 800 ans selon les datations (figure 23).



**Figure 23. Caractéristiques géographiques, archéologiques et métagénomiques du squelette Ragna.**

a) Site d'excavation (Folkebibilotekstomten, 1973-1985) du cimetière de l'église de St. Olav à Trondheim, en Norvège. b) Squelette entier (en haut) et os long fémoral plus deux dents à partir desquels l'ADN de *Salmonella* était extrait (en bas). c) Carte de l'Europe entourant la Norvège (en vert).

### *Séquençage métagénomique du squelette*

Toutes les manipulations moléculaires, y compris la préparation des librairies ont été menées dans des laboratoires propres dédiés au GeoGenetics Center du Musée d'Histoire Naturelle de l'Université de Copenhague. Tous les échantillons ont été collectés et traités en appliquant des protocoles strictes pour les ADN anciens (ADNa). Les banques ont été séquencées avec le Hiseq (Illumina) avec 15 passages.

### *Assemblage génomique des séquences de métagénomique*

L'étude taxonomique des séquences de métagénomique par les méthodes traditionnelles actuellement disponibles peut conduire à des erreurs quant à la détection de pathogènes bactériens. En effet, les pathogènes sont sur-représentés dans les bases de données publiques par rapport aux organismes environnementaux et les génomes complets de nombreuses bactéries environnementales ne sont pas encore disponibles. Par conséquent, l'approche de l'assemblage de novo à partir des données brutes est souvent privilégiée. Cette approche a été utilisée afin de mettre en évidence la présence de pathogènes bactériens, à partir de neuf banques de métagénomique de SK152.

Tous les reads s'alignant avec le génome humain selon BWA<sup>137</sup> ont été exclus. Les reads non-humains ont été assemblés en séquences consensus (contigs) avec MEGAHIT<sup>215</sup>, en utilisant les paramètres par défaut. Tous les contigs dont la taille était supérieure à 20 kb ont été divisés en fragments de 10 kb. Les 39,016 contigs (>1 kb) ont été groupés par CONCOCT<sup>216</sup> en 76 groupes selon la composition de la séquence et leur couverture. Les protéines codant pour ces séquences ont été définis par PRODIGAL<sup>217</sup> et leur fonction classée selon le Clusters of Orthologous Groups of proteins (COGs)<sup>218</sup>.

### *Genotyping par Enterobase*

Enterobase assemble automatiquement tous les génomes dont les séquences Illumina ont été rendues publics. Pour cela, il utilise un pipeline utilisant SPAdes<sup>132</sup> pour l'assemblage et le post-assemblage (détection des SNPs par alignement des reads aux contigs formés). Les assemblages ayant passé les critères de qualités : les couvertures  $\geq 20 \times$  sont automatiquement génotypés par la méthode du MLST (sept gènes de ménage ; 51 protéines ribosomales [rMLST] ; 21,065 gènes du pangéome [wgMLST] et 3,002 core-gènes [cgMLST]).

### *Arbre phylogénétique de S. enterica sous-espèce I*

Une collection de 50,000 génomes représentant toute la diversité génétique de *S. enterica* sous-espèce I ont été choisis aléatoirement parmi les 2,964 rMLST STs dans EnteroBase (Mai 2016). Ces génomes sont disponibles sur EnteroBase. Des arbres phylogénétiques ont été réalisés à partir des séquences des 3,002 core-gènes (EnteroBase schéma cgMLST v2). Les séquences de chaque core-gène ont été alignées avec MAFFT<sup>219</sup>. Deux algorithmes ont été utilisés pour ces alignements : i) un arbre maximum de vraisemblance basé sur un multifasta d'alignement de toutes les séquences du core-génome (2.8 Mb) et le modèle GTRCAT a été généré en utilisant un RAXML v8.2.442. ii) Un autre arbre de maximum de vraisemblance a été généré en utilisant RAXML pour chacun des 3,002 alignements des core-gènes eux-même analysés par ASTRID.<sup>220</sup>

### *Origine des génomes de la lignée Para C*

Les 100 génomes d'EnteroBase appartenant à la lignée Para C proviennent de souches épidémiologiquement récentes (PHE, R.-U. et FDA Etats-Unis), mais également de souches séquencées par le Sanger Center (MLST legacy) ou de la collection historique de Murray. Afin d'obtenir des isolats ayant une diversité temporelle, 119 isolats anciens provenant de la collection de l'Unité des Bactéries Pathogènes Entériques (Institut Pasteur, Paris) ont été séquencés et implémentés à la base de données d'EnteroBase.

### *Identification de l'ADNa de Paratyphi C au sein des données de métagénomique*

Une analyse initiale de métagénomique sur des séquences de 33 squelettes humains (Trondheim, Norvège) a été réalisée en utilisant KRAKEN<sup>221</sup> et ses paramètres par défaut. KRAKEN a identifié 304 séquences métagénomiques de Salmonella pour le squelette SK152. Ces séquences ont été alignées à 91,000 génomes (bactérien, archée, viral) de la base de données de GenBank avec BLASTn.

### *Identification des séquences spécifiques à Ragna*

Les séquences ont été analysées dans un premier temps avec BBMERGe et BBDUK2 du module BBMAP.<sup>222</sup> Les séquences spécifiques à Ragna ont ensuite été identifiées en les alignant à deux génomes de référence dont RKS4594 (génome de référence de Paratyphi C) en utilisant BOWTIE2 et SAMtools/BCFtools 1.2.

### *Détection des SNPs pour les génomes récents*

Les assemblages de la lignée Para C et de Birkenhead ont été alignés contre RKS4594 en utilisant LAST<sup>223</sup> et les SNPs filtrés (base ayant une qualité < 10 et les alignements ambigus > 0.1). Les régions dont l'alignement correspond aux séquences répétées (taille > 100 bp) (BLASTn) ou aux loci CRISPR (PILER-CR<sup>224</sup>), à plus de 95% ont été supprimées. Avec ces paramètres, 61,451 SNPs ont été retenus.

### *Reconstruction phylogénétique de la lignée Para C*

A partir du core-génome SNPs, un arbre phylogénétique a été généré en utilisant RAxML v8.2.4, en utilisant un modèle GTRCAT et Stamatakis pour la correction des sites invariants.

### *Datation de Paratyphi c et de la lignée ParaC*

La méthode bayésienne BEAST v1.8.3 (modèle GTR+G) a été utilisée pour générer des arbres phylogénétiques datés.

### *Analyses génétiques*

Tous les gènes présentant un codon stop ou un changement de cadre de lecture au niveau des régions codantes ont été définis comme pseudogènes. Les prophages de la lignée Para C ont été identifiés par PHASTER. BLASTn a permis l'identification des îlots génomiques au sein du pan-génome. Les séquences îlots génomiques de référence (SPI-1 à SPI-12) ont été téléchargées sur PAIDB<sup>225</sup>. Tous les gènes du pan-génome sont passés par ISfinder (séquences d'insertion) PlasmidFinder 1.3 (groupes d'incompatibilité plasmidique) et CONJscanT4SSscan (relaxases et système de sécrétion de type IV).

### *Accessibilité des données*

Les séquences des 119 souches de la collection de l'Institut Pasteur ont été déposées sur le site de NCBI sous le numéro d'accès PRJEB19916. Tous les génomes mentionnés dans cette étude sont disponibles sur EnteroBase au nom de « rST representatives » (2,964 génomes) et « Para C Lineage ».

### III. Résultat et discussion

Une analyse phylogénétique faite à partir du core-génome (3,002 gènes) de Ragna et de 2,964 autres génomes récents et représentatifs de *S. enterica* sous-espèce *enterica* montre l'appartenance du génome de Ragna à un clade appelé « la lignée Para C », composé de *S. enterica* sérotype Paratyphi C, *S. enterica* sérotype Choleraesuis, *S. enterica* sérotype Typhisuis et un sérotype extrêmement rare, *S. enterica* sérotype Lomita, (figure 24). Pour aller plus loin, un alignement du génome de Ragna avec 219 génomes dont 119 historiques (Institut Pasteur, Paris) de *S. enterica* sérotype Paratyphi C et ses variants (PC-1, PC-2), Choleraesuis et ses variants (CS Kunzendorf, CS *sensu stricto*, CS-3) et Typhisuis et ses variants (TS-1, TS-2), montre que 0.05-0.18% de l'ensemble des reads provenant des dents de Ragna et 0.01% des reads du fémur étaient spécifiques à *S. enterica* sérotype Paratyphi C. Ces reads couvrent également 98.4% du génome de référence de *S. enterica* sérotype Paratyphi C (RKS4594). La comparaison entre les 219 souches historiques et récentes indique que 78% des  $4,388 \pm 99$  gènes de la lignée Para C appartenaient principalement au core-génome et pour certains, à des éléments génétiques mobiles (plasmides de virulence, bactériophages, îlots de pathogénicité), présents de manière aléatoire chez d'autres *Salmonella*.

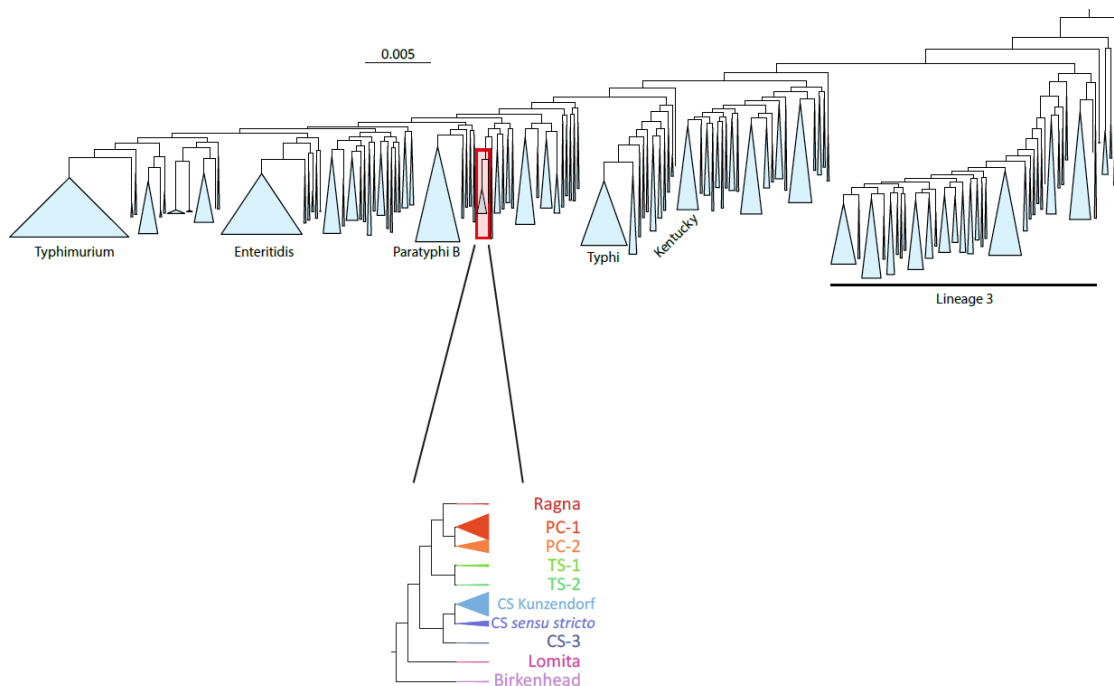
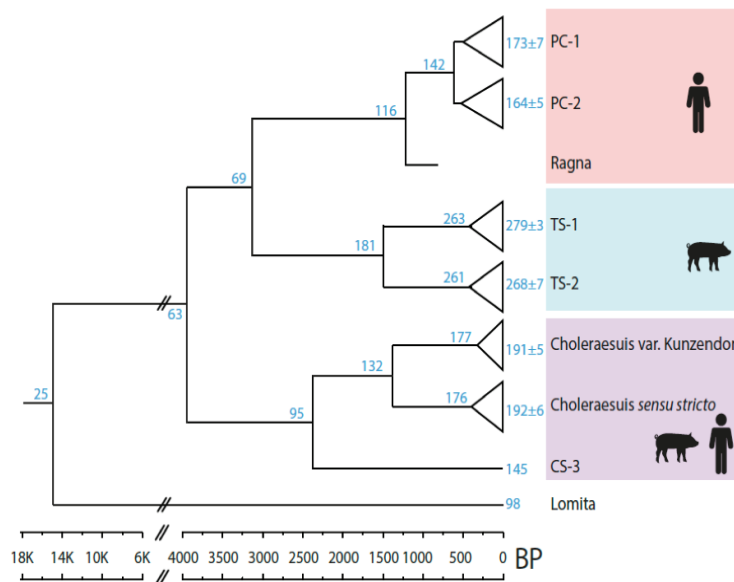


Figure 24. Phylogénie génomique des *Salmonella enterica* et de la lignée Para C.

Phylogénie basée sur la méthode de maximum de vraisemblance à partir de 2,964 génomes représentatifs de *S. enterica* sous-espèce *enterica*. Chaque génome est aléatoirement représentatif d'un séquençotype ribosomal multi-locus (rMLST). La lignée ParaC et d'autres lignées contenant de nombreux génomes ont été regroupés en triangles bleus et les sérotypes associés avec ces lignées sont indiqués sous l'arbre. L'encadré rouge indique la lignée Para C plus le groupe extérieur, Birkenhead.

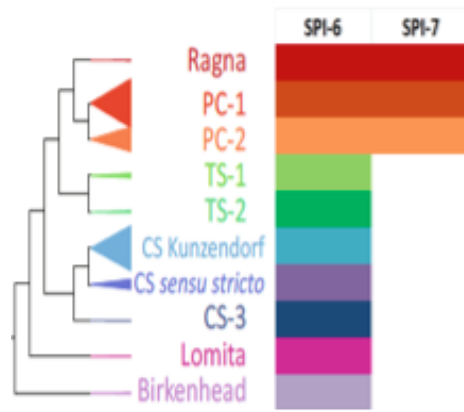
L'étude des pseudogènes (gènes inactifs au sein d'un génome, n'étant plus nécessaire pour l'infection de l'hôte, mais qui traduisent l'adaptation microbienne) montre une accumulation de pseudogènes au sein des sous-groupes de chaque sérotypes. Cependant, au niveau de l'ancêtre commun de la lignée Para C, peu de changements est observable. Les pseudogènes accumulés ont sûrement participé à la différenciation et à la spécificité de l'hôte pour chaque sérotype (figure 25).



**Figure 25. Représentation de l'évolution de la lignée Para C au cours du temps.**

En bleu est représenté le nombre de pseudogènes accumulés. Les encadrés représentent les différents sérotypes de la lignée Para C et leur spécificité d'hôte

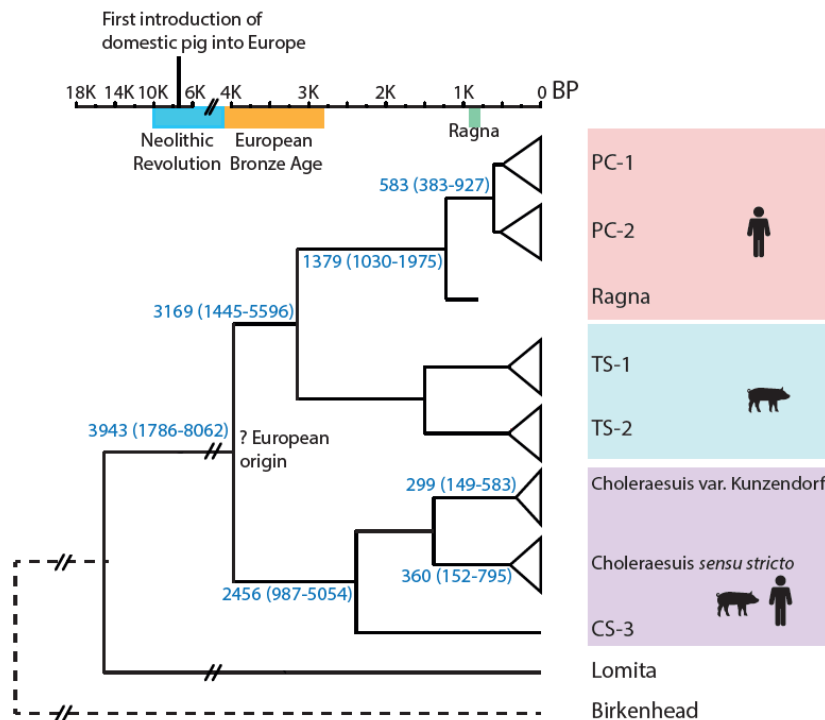
L'extraction des îlots de pathogénicité de *Salmonella* montre que l'îlot SPI-6 qui participe à la colonisation gastro-intestinale est présent chez Ragna et tous les sérotypes (PC-1/2, TS-1/2, CS Kunzendorf et *sensu stricto*, CS-3, Lomita et Birkenhead). Cet îlot code pour des gènes spécifiques à l'adhésion chez l'homme (*tcf*), la colonisation des porcs (*saf*) et sont présents de manière variable au sein des génomes. Cependant, l'îlot de pathogénicité SPI-7 n'est présent que chez Ragna et Paratyphi C (PC-1 et PC2). Cet îlot contient des gènes de synthèse de la capsule Vi, participe à l'activité inhibitrice de l'opsonisation et favorise la fièvre entérique (figure 26).



**Figure 26. Répartition des îlots de pathogénicité de *Salmonella* au sein de la lignée C.**

Sont représentés ici les îlots SPI-6 et SPI-7. Les couleurs indiquent les différents sérotypes.

Enfin, l'analyse évolutive de la lignée Para C par une approche phylogénétique Bayésienne, indique que le sérotype Paratyphi C s'est différencié des autres sérotypes Choleraesuis et Typhisuis il y a plus de 3,000 ans. Ces trois sérotypes auraient pour ancêtre commun une souche Européenne, ce qui coïnciderait avec l'existence de *S. enterica* sérotype Paratyphi C au nord de la Norvège, 800 ans auparavant (figure 27).



**Figure 27. Schéma représentatif de l'histoire évolutive de la lignée Para C au cours du temps.**

L'échelle en haut à gauche représente les différentes périodes historiques (BP : *Before Present*). Les chiffres en bleus sur le phylogramme indiquent les estimations de datation avec les marges d'erreurs entre parenthèses, selon une approche phylogénétique bayésienne. Les spécificités de l'hôte pour chaque sérotype sont indiquées dans les encadrés de couleurs.

## IV. Conclusion

L'ensemble des résultats obtenus à partir des os et des dents d'une jeune femme enterrée en Norvège il y a plus de 800 ans indiquait que *S. enterica* sérotype Paratyphi C était à l'origine d'une infection en Norvège, provoquant la mort de Ragna par septicémie associée à la fièvre entérique. Originnaire d'Europe, *S. enterica* sérotype Paratyphi C aurait disparu pour aller sévir en Asie et en Afrique.

Toutes les bactéries étant soumises à des pressions de sélections dues à l'environnement dans lequel elles vivent, subissent des variations génomiques au cours de leur existence.<sup>226</sup> Avec une période de plus de mille ans séparant Ragna, le génome le plus ancien au génome le plus récent, nous nous attendions à ce que les différences génomiques relevées soit nombreuses et encore plus grandes entre *S. enterica* sérotype Paratyphi C et les autres sérotype Choleraesuis et Typhisuis. Les 219 génomes historiques et récents de ces trois sérotypes ont montré qu'ils étaient étroitement liés, avec seulement 1000 SNPs dispersés au sein de leur core-génome. L'arbre évolutif daté de la lignée Paratyphi C montre que *S. enterica* sérotype Paratyphi C aurait pour ancêtre les sérotypes Choleraesuis et Typhisuis et que les rares changements dans leur génome accessoire au cours des 4,000 dernières années, seraient, pour la plupart, des événements aléatoires plutôt qu'une sélection naturelle. Son évolution stable au cours du temps est sûrement due à un transfert zoonotique provenant du porc durant la période du Néolithique.<sup>227</sup>

Ainsi, cette étude collaborative de métagénomique illustre les avantages pouvant être obtenus en combinant l'étude génomique de souches historiques et récentes à celle d'autres domaines tels que la datation et l'archéologie.





## **Conclusion générale**

Ce travail s'inscrit dans le cadre de l'étude épidémiologique de *Salmonella*, pathogène à l'origine des infections alimentaires les plus courantes et répandues, à l'ère du séquençage massif. Outil pluridisciplinaire, le séquençage entier du génome permet de répondre à de nombreux questionnements biologiques.

Les travaux de cette thèse ont été réalisés dans l'Unité des Bactéries Pathogènes Entériques, qui renferme le Centre National de Référence des *Escherichia coli*, *Shigella* et *Salmonella* (CNR-ESS). Le CNR-ESS a pour principales missions : (1) de réaliser une identification et une caractérisation rapide des souches de *Salmonella* afin de détecter précocement les épisodes épidémiques, (2) de suivre l'évolution et les mécanismes de résistance des souches émergentes résistantes aux antibiotiques et enfin (3) d'établir un lien phylogénétique entre les différentes souches afin d'observer leur évolution au cours du temps. Dans ce manuscrit, trois articles publiés ou en cours de publication ont été présentés et répondent pour chacun d'entre eux aux problématiques et missions du CNR-ESS.

(1) Le développement exponentiel du séquençage a changé le monde de la microbiologie et nous avons vu naître au cours de ces dernières années, des outils dédiés à l'analyse de *Salmonella*. Cependant, rare sont ceux capables de réaliser un sous-typage en vue d'un dépistage épidémiologique en routine. Le premier article intitulé "CRISPOL *in silico*, méthode de typage de *Salmonella enterica* sérotype Typhimurium", introduit un nouvel outil de sous-typage basé sur le SEG et le polymorphisme des régions CRISPR. Rapide (temps d'analyse post séquençage moins de 15 minutes), allie surveillance épidémiologique et liens de parenté entre les différentes populations grâce à l'implémentation d'une approche phylogénétique. La comparaison entre l'approche assemblage *de novo* et celle de l'alignement a montré que cette dernière était plus rapide (15 minutes d'analyse contre trois heures d'assemblage et 10 minutes d'analyse pour le plus long assembleur) et plus efficace (extraction du CRISPOL réussie pour 100% des génomes contre 98% pour l'assemblage). Son utilisation en routine au CNR-ESS pour le sérotype Typhimurium, a déjà prouvé son efficacité avec l'étude rétrospective des génomes de 2016 ayant détecté une épidémie liée au poulet, non détectée par agglutination. Cet outil, sera bientôt déployé sur le cluster de calculs de l'Institut Pasteur et un site internet sera mis en place afin que chaque laboratoire puisse disposer d'un outil d'analyse facile d'utilisation. Pour le sérotype Enteritidis, les tests sont

encore en cours et une nouvelle nomenclature basée cette fois sur les SNPs (l'approche CRISPOL n'étant pas assez discriminante) devra être mise en place.

A la suite de cette thèse, nous souhaitons étendre et adapter l'approche CRISPOL *in silico* à tous les sérotypes. Un grand travail de création de base de données a déjà été amorcé en amont par le Docteur Laetitia Fabre. La mise en place d'une nouvelle nomenclature basée sur les SNPs du core génome permettra d'affiner le sous-typage de sérotypes génétiquement homogènes tels que *S. enterica* sérotype Enteritidis. Connaissant l'implication des phages dans l'acquisition des nouveaux espaceurs (bien que l'évènement reste rare), une étude similaire à celle de Fu *et al.* pour une comparaison alliant CRISPOL, SNPs et phages serait intéressante avec une collection de souches aussi diversifiée et importante que le CNR-ESS.

(2) Le deuxième article intitulé "Transmission de la résistance à l'ampicilline chez *Salmonella enterica* sérotype Typhimurium à la fin des années 1950 : une étude rétrospective par séquençage entier du génome", a permis de mieux comprendre l'émergence au sein de ce sérotype de souches résistantes à l'ampicilline avant même sa commercialisation en 1961. Connaissant déjà les méfaits d'une utilisation massive des antibiotiques, nous avons pu confirmer grâce au SEG sa probable implication et notamment celle de la pénicilline G (utilisée dans la filière du bétail en tant que promoteur de croissance), dans l'acquisition de déterminants à la résistance (bla-TEM1, plasmides et transposons) à l'ampicilline.

Durant cette étude, nous avons pu collecter deux souches isolées en 1968, au séquençotype particulier, le ST313. Ce séquençotype est connu pour être un clone invasif (voir chapitre VI), résistant aux antibiotiques et responsable des maladies infectieuses principalement en Afrique. De nombreux articles<sup>206,228</sup> ont déjà été publiés, mais leur collection comprenant que des souches récentes ne pouvait pas retracer correctement l'histoire évolutive de ce clone Africain. Débuté en 1947 et associé à de nombreux laboratoires dont ceux du réseau mondial de l'Institut Pasteur, la collection de souches de l'unité des Bactéries Pathogènes Entériques, permettra d'en apprendre plus sur ce clone, l'origine de son émergence et les raisons de son expansion dans les pays d'Afrique.

(3) Enfin, le troisième article intitulé "*Salmonella enterica* sérotype Paratyphi C : une stabilité génomique qui s'étend sur un millénaire", répond à la dernière mission du CNR-ESS qui est le suivi évolutif ou l'émergence de nouveaux sérotypes. Cette étude de métagénomique a été

effectuée dans le cadre d'une collaboration et montre que le sérotype Paratyphi C, sérotype rare a très peu évolué au cours de ces 1,300 ans. Encore une fois, l'analyse du génome entier a permis de comparer des sérotypes entre eux et de voir les liens de parentés qui les unissent mais également les caractéristiques propres à chacun d'eux (ilots de pathogénicité participant à la spécificité de l'hôte des sérotypes).

L'utilisation du séquençage génomique tant à l'échelle du diagnostic qu'à l'échelle de l'évolution microbienne est sans nul doute l'approche à privilégier à l'avenir, de par sa rapidité et son efficacité, avec les outils qui lui sont dédiés et ceux qui restent encore à développer. Le séquençage réunit à lui seul toutes les informations et données pouvant caractériser une bactérie (identification du sérotype, détection des gènes de résistance, caractérisation plasmidique etc).

Cependant, le manque de (1) normalisation lors du traitement des séquences (paramètres de filtration des reads par exemple) mais également des (2) bases de données (mise place d'une nomenclature basée sur les SNPs sans base de données commune) entre les laboratoires pourraient être un frein au SEG.

Avec l'avancée des nouvelles technologies, le typage et sous-typage des *Salmonella* devient de plus en plus discriminants: des caractères phénotypiques et biochimiques nous sommes passés à l'analyse nucléotidique et génomique, en passant par l'étude des déterminants moléculaires. Néanmoins, il est important de fixer des limites et de savoir jusqu'où la discrimination est utile, notamment lors du sous-typage. En effet, l'analyse du polymorphisme nucléotidique est-elle nécessaire pour détecter des souches d'une même épidémie ou l'analyse moléculaire par la méthode du CRISPOL *in silico* est suffisante? Encore une fois, ces critères restent spécifiques à chaque étude et chaque laboratoire et aucune règle n'a été définie jusqu'à présent. La démocratisation du séquençage récente du séquençage mais également le nombre important d'outils qui lui sont dédiés ne permettent pas à ce jour d'avoir assez de recul pour répondre à cette question.

Garder un lien avec les techniques et nomenclatures établies il y a près de 100 ans, devenues maintenant des méthodes de référence permettra d'améliorer et normaliser celles du SEG.

## **Bibliographies**

1. Grimont, P. A., Weill, F.-X. & others. Antigenic formulae of the Salmonella serovars. *WHO Collab. Cent. Ref. Res. Salmonella* **9**, 1–161 (2007).
2. Issenhuth-Jeanjean, S. *et al.* Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme. *Res. Microbiol.* **165**, 526–530 (2014).
3. Centre National de Référence des Escherichia coli, Shigella, Salmonella. Rapport d'activité annuel d'exercice 2016. Available at: [https://www.pasteur.fr/sites/default/files/rubrique\\_pro\\_sante\\_publicue/les\\_cnr/escherichia\\_coli\\_shigella\\_salmonella/rapport-cnr\\_escherichia-coli-shigella-salmonella-2016\\_pdf\\_final\\_.pdf](https://www.pasteur.fr/sites/default/files/rubrique_pro_sante_publicue/les_cnr/escherichia_coli_shigella_salmonella/rapport-cnr_escherichia-coli-shigella-salmonella-2016_pdf_final_.pdf).
4. Gay, N., Le Hello, S., Weill, F.-X., de Thoisy, B. & Berger, F. Salmonella serotypes in reptiles and humans, French Guiana. *Vet. Microbiol.* **170**, 167–171 (2014).
5. Spanò, S. Mechanisms of Salmonella Typhi Host Restriction. *Adv. Exp. Med. Biol.* **915**, 283–294 (2016).
6. Foley, S. L., Johnson, T. J., Ricke, S. C., Nayak, R. & Danzeisen, J. Salmonella Pathogenicity and Host Adaptation in Chicken-Associated Serovars. *Microbiol. Mol. Biol. Rev. MMBR* **77**, 582–607 (2013).
7. McDonough, P. L., Fogelman, D., Shin, S. J., Brunner, M. A. & Lein, D. H. Salmonella enterica serotype Dublin infection: an emerging infectious disease for the northeastern United States. *J. Clin. Microbiol.* **37**, 2418–2427 (1999).
8. Dominguez, M. *et al.* Outbreak of Salmonella enterica serotype Montevideo infections in France linked to consumption of cheese made from raw milk. *Foodborne Pathog. Dis.* **6**, 121–128 (2009).
9. Golden, N. J. *et al.* Review of induced molting by feed removal and contamination of eggs with Salmonella enterica serovar Enteritidis. *Vet. Microbiol.* **131**, 215–228 (2008).
10. Boyen, F. *et al.* Non-typhoidal Salmonella infections in pigs: A closer look at epidemiology, pathogenesis and control. *Vet. Microbiol.* **130**, 1–19 (2008).
11. Darwin, K. H. & Miller, V. L. Molecular Basis of the Interaction of Salmonella with the Intestinal Mucosa. *Clin. Microbiol. Rev.* **12**, 405–428 (1999).
12. Finlay, B. B. & Falkow, S. Comparison of the invasion strategies used by Salmonella cholerae-suis, Shigella flexneri and Yersinia enterocolitica to enter cultured animal cells: endosome acidification is not required for bacterial invasion or intracellular replication. *Biochimie* **70**, 1089–1099 (1988).
13. Hensel, M. Evolution of pathogenicity islands of Salmonella enterica. *Int. J. Med. Microbiol. IJMM* **294**, 95–102 (2004).
14. Rotger, R. & Casadesús, J. The virulence plasmids of Salmonella. *Int. Microbiol.* **2**, 177–184 (1999).

15. Ahmer, B. M., Tran, M. & Heffron, F. The virulence plasmid of *Salmonella typhimurium* is self-transmissible. *J. Bacteriol.* **181**, 1364–1368 (1999).
16. Furman, M., Fica, A., Saxena, M., Di Fabio, J. L. & Cabello, F. C. *Salmonella typhi* iron uptake mutants are attenuated in mice. *Infect. Immun.* **62**, 4091–4094 (1994).
17. Danner, R. L. *et al.* Endotoxemia in human septic shock. *Chest* **99**, 169–175 (1991).
18. Finlay, B. B. & Falkow, S. Virulence factors associated with *Salmonella* species. *Microbiol. Sci.* **5**, 324–328 (1988).
19. Eng, S.-K. *et al.* *Salmonella*: A review on pathogenesis, epidemiology and antibiotic resistance. *Front. Life Sci.* **8**, 284–293 (2015).
20. Gordon, M. A. Invasive Non-typhoidal *Salmonella* Disease – epidemiology, pathogenesis and diagnosis. *Curr. Opin. Infect. Dis.* **24**, 484–489 (2011).
21. Outbreaks Involving *Salmonella* | CDC. (2018). Available at: <https://www.cdc.gov/salmonella/outbreaks.html>. (Accessed: 8th April 2018)
22. Santé publique France - Epidémie nationale de salmonelloses à *Salmonella Agona* en cours associée à des laits pour nourrissons distribués à l'échelon international, France, décembre 2017. Available at: <https://www.santepubliquefrance.fr/Sante-publique-France/L-article-du-mois/2018/Epidemie-nationale-de-salmonelloses-a-Salmonella-Agona-en-cours-associee-a-des-laits-pour-nourrissons-distribues-a-l-echelon-international-France-decembre-2017>. (Accessed: 8th April 2018)
23. Hardy, A. *Salmonella*: a continuing problem. *Postgrad. Med. J.* **80**, 541–545 (2004).
24. Crump, J. A., Luby, S. P. & Mintz, E. D. The global burden of typhoid fever. *Bull. World Health Organ.* **82**, 346–353 (2004).
25. Weill, F.-X. Les fièvres typhoïdes et paratyphoïdes en France de 2001 à 2003.
26. Reeves, P. R. *et al.* Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol.* **4**, 495–503 (1996).
27. Jones, C. J. & Aizawa, S. The bacterial flagellum and flagellar motor: structure, assembly and function. *Adv. Microb. Physiol.* **32**, 109–172 (1991).
28. Craigie, J. & Yen, C. H. The Demonstration of Types of *B. typhosus* by Means of Preparations of Type II Vi Phage. I. Principles and Technique. *Can. Public Health J.* **29**, 448–63 (1938).
29. Dragas, Z. [Lysotyping of *Salmonella typhimurium*. Study of strains from hospital epidemics]. *Arch. Roum. Pathol. Exp. Microbiol.* **28**, 991–996 (1969).
30. Anderson, E. S., Ward, L. R., De Saxe, M. J. & De Sa, J. D. H. Bacteriophage-typing designations of *Salmonella typhimurium*. *J Hyg* 297–300 (1977).
31. (Icept), T. I. C. F. E. P.-T. The Geographical Distribution of *Salmonella typhi* and



Salmonella paratyphi A and B Phage Types during the Period 1 January 1966 to 31 December 1969. *J. Hyg. (Lond.)* **71**, 59–84 (1973).

32. Grimont, F. & Grimont, P. A. Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. *Ann. Inst. Pasteur Microbiol.* **137B**, 165–175 (1986).
33. Regnault, B., Grimont, F. & Grimont, P. A. Universal ribotyping method using a chemically labelled oligonucleotide probe mixture. *Res. Microbiol.* **148**, 649–659 (1997).
34. Altwegg, M., Hickman-Brenner, F. W. & Farmer, J. J. Ribosomal RNA gene restriction patterns provide increased sensitivity for typing Salmonella typhi strains. *J. Infect. Dis.* **160**, 145–149 (1989).
35. Selander, R. K. *et al.* Genetic population structure, clonal phylogeny, and pathogenicity of Salmonella paratyphi B. *Infect. Immun.* **58**, 1891–1901 (1990).
36. Esteban, E., Snipes, K., Hird, D., Kasten, R. & Kinde, H. Use of ribotyping for characterization of Salmonella serotypes. *J. Clin. Microbiol.* **31**, 233–237 (1993).
37. Renvoisé, A., Brossier, F., Sougakoff, W., Jarlier, V. & Aubry, A. Broad-range PCR: past, present, or future of bacteriology? *Med. Mal. Infect.* **43**, 322–330 (2013).
38. Echeita, M. A. & Usera, M. A. Chromosomal rearrangements in Salmonella enterica serotype typhi affecting molecular typing in outbreak investigations. *J. Clin. Microbiol.* **36**, 2123–2126 (1998).
39. Schwartz, D. C. & Cantor, C. R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75 (1984).
40. Hunter, S. B. *et al.* Establishment of a universal size standard strain for use with the PulseNet standardized pulsed-field gel electrophoresis protocols: converting the national databases to the new size standard. *J. Clin. Microbiol.* **43**, 1045–1050 (2005).
41. Murase, T., Nakamura, A., Matsushima, A. & Yamai, S. An Epidemiological Study of Salmonella enteritidis by Pulsed-Field Gel Electrophoresis (PFGE): Several PFGE Patterns Observed in Isolates from a Food Poisoning Outbreak. *Microbiol. Immunol.* **40**, 873–875 (1996).
42. Sandt, C. H. *et al.* The Key Role of Pulsed-Field Gel Electrophoresis in Investigation of a Large Multiserotype and Multistate Food-Borne Outbreak of Salmonella Infections Centered in Pennsylvania. *J. Clin. Microbiol.* **44**, 3208–3212 (2006).
43. Weill, F.-X. *et al.* Multidrug resistance in Salmonella enterica serotype Typhimurium from humans in France (1993 to 2003). *J. Clin. Microbiol.* **44**, 700–708 (2006).
44. Liu, Y. *et al.* The evaluation and application of multilocus variable number tandem repeat analysis (MLVA) for the molecular epidemiological study of Salmonella enterica subsp. enterica serovar Enteritidis infection. *Ann. Clin. Microbiol. Antimicrob.* **15**, 4 (2016).

45. Lindstedt, B.-A., Vardund, T., Aas, L. & Kapperud, G. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. *J. Microbiol. Methods* **59**, 163–172 (2004).
46. Larsson, J. T. *et al.* Development of a new nomenclature for *Salmonella typhimurium* multilocus variable number of tandem repeats analysis (MLVA). *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **14**, (2009).
47. Lindstedt, B. A. *et al.* Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **18**, 20385 (2013).
48. Maiden, M. C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3140–3145 (1998).
49. Fakhr, M. K., Nolan, L. K. & Logue, C. M. Multilocus sequence typing lacks the discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. *J. Clin. Microbiol.* **43**, 2215–2219 (2005).
50. Kidgell, C. *et al.* *Salmonella typhi*, the causative agent of typhoid fever, is approximately 50,000 years old. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2**, 39–45 (2002).
51. Kotetishvili, M., Stine, O. C., Kreger, A., Morris, J. G. & Sulakvelidze, A. Multilocus sequence typing for characterization of clinical and environmental salmonella strains. *J. Clin. Microbiol.* **40**, 1626–1635 (2002).
52. Sukhnanand, S. *et al.* DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella enterica* serotypes. *J. Clin. Microbiol.* **43**, 3688–3698 (2005).
53. Achtman, M. *et al.* Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* **8**, e1002776 (2012).
54. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
55. Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* **117**, 119–128 (2015).
56. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60 (2005).
57. Fabre, L. *et al.* CRISPR typing and subtyping for improved laboratory surveillance of

- Salmonella infections. *PloS One* **7**, e36995 (2012).
58. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
  59. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
  60. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol.* **8**, 423–435 (2010).
  61. Munita, J. M. & Arias, C. A. Mechanisms of Antibiotic Resistance. *Microbiol. Spectr.* **4**, (2016).
  62. Jacoby, G. A. Extrachromosomal resistance in gram-negative organisms: the evolution of  $\beta$ -lactamase. *Trends Microbiol.* **2**, 357–360 (1994).
  63. Shaw, K. J., Rather, P. N., Hare, R. S. & Miller, G. H. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol. Rev.* **57**, 138–163 (1993).
  64. Spratt, B. G. Resistance to antibiotics mediated by target alterations. *Science* **264**, 388–393 (1994).
  65. Nikaido, H. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* **264**, 382–388 (1994).
  66. Mass, W. K. Mutations to antibiotic resistance. in *Antibiotics in laboratory and medicine* 669–682 (Lorian V).
  67. Martinez, J. L. & Baquero, F. Mutation Frequencies and Antibiotic Resistance. *Antimicrob. Agents Chemother.* **44**, 1771–1777 (2000).
  68. Actis, L. A., Tolmasky, M. E. & Crosa, J. H. Bacterial plasmids: replication of extrachromosomal genetic elements encoding resistance to antimicrobial compounds. *Front. Biosci. J. Virtual Libr.* **4**, D43-62 (1999).
  69. Novick, R. P. Plasmid incompatibility. *Microbiol. Rev.* **51**, 381 (1987).
  70. Couturier, M., Bex, F., Bergquist, P. L. & Maas, W. K. Identification and classification of bacterial plasmids. *Microbiol. Rev.* **52**, 375–395 (1988).
  71. Davies, J. E. Origins, acquisition and dissemination of antibiotic resistance determinants. *Ciba Found. Symp.* **207**, 15-27; discussion 27-35 (1997).
  72. Bennett, P. M. Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods Mol. Biol. Clifton NJ* **266**, 71–113 (2004).
  73. Burrus, V. & Waldor, M. K. Shaping bacterial genomes with integrative and conjugative elements. *Res. Microbiol.* **155**, 376–386 (2004).
  74. Kleckner, N., Chan, R. K., Tye, B. K. & Botstein, D. Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. *J. Mol. Biol.* **97**, 561–575 (1975).

75. Stokes, H. W. & Hall, R. M. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol. Microbiol.* **3**, 1669–1683 (1989).
76. Grinstead, J., de la Cruz, F. & Schmitt, R. The Tn21 subgroup of bacterial transposable elements. *Plasmid* **24**, 163–189 (1990).
77. Doublet, B., Boyd, D., Mulvey, M. R. & Cloeckaert, A. The Salmonella genomic island 1 is an integrative mobilizable element. *Mol. Microbiol.* **55**, 1911–1924 (2005).
78. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond.)* **27**, 113–159 (1928).
79. Lorenz, M. G. & Wackernagel, W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* **58**, 563–602 (1994).
80. Davison, J. Genetic Exchange between Bacteria in the Environment. *Plasmid* **42**, 73–91 (1999).
81. Waters, V. L. Conjugative transfer in the dissemination of  $\beta$ -lactam and aminoglycoside resistance. *Front Biosci* **4**, 10–2741 (1999).
82. Frost, L. S., Ippen-Ihler, K. & Skurray, R. A. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **58**, 162–210 (1994).
83. Pansegrau, W. *et al.* Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J. Mol. Biol.* **239**, 623–663 (1994).
84. Ochiai, K., Yamanaka, T., Kimura, K. & Sawada, O. Studies on inheritance of drug resistance between Shigella strains and Escherichia coli strains. *Nippon Iji Shimpo* **1861**, 34–46 (1959).
85. Bauer, A. W., Perry, D. M. & Kirby, W. M. Single-disk antibiotic-sensitivity testing of staphylococci; an analysis of technique and results. *AMA Arch. Intern. Med.* **104**, 208–216 (1959).
86. Bauer, A. W., Kirby, W. M., Sherris, J. C. & Turck, M. Antibiotic susceptibility testing by a standardized single disk method. *Am. J. Clin. Pathol.* **45**, 493–496 (1966).
87. Comité de l'antibiogramme de la Société Française de Microbiologie (CASFM) & European committee on Antimicrobial Susceptibility Testing (EUCAST). Recommendations 2017 V.1.0 MARS. (2017). Available at: [http://www.sfm-microbiologie.org/UserFiles/files/casfm/CASFMV1\\_0\\_MARS\\_2017.pdf](http://www.sfm-microbiologie.org/UserFiles/files/casfm/CASFMV1_0_MARS_2017.pdf).
88. Carattoli, A. *et al.* Identification of plasmids by PCR-based replicon typing. *J. Microbiol. Methods* **63**, 219–228 (2005).
89. Michael, G. B., Butaye, P., Cloeckaert, A. & Schwarz, S. Genes and mutations conferring antimicrobial resistance in Salmonella: an update. *Microbes Infect.* **8**, 1898–1914

(2006).

90. Bush, K. & Jacoby, G. A. Updated Functional Classification of  $\beta$ -Lactamases. *Antimicrob. Agents Chemother.* **54**, 969–976 (2010).
91. Sirot, D. Extended-spectrum plasmid-mediated  $\beta$ -lactamases. *J. Antimicrob. Chemother.* **36 Suppl A**, 19–34 (1995).
92. Medeiros, A. A. B-lactamases. *Br. Med. Bull.* **40**, 18–27 (1984).
93. Accou-Demartin, M. *et al.* *Salmonella enterica* Serotype Typhi with Nonclassical Quinolone Resistance Phenotype. *Emerg. Infect. Dis.* **17**, 1091–1094 (2011).
94. Wain, J. & Kidgell, C. The emergence of multidrug resistance to antimicrobial agents for the treatment of typhoid fever. *Trans. R. Soc. Trop. Med. Hyg.* **98**, 423–430 (2004).
95. Wain, J. *et al.* Quinolone-resistant *Salmonella typhi* in Viet Nam: molecular basis of resistance and clinical response to treatment. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **25**, 1404–1410 (1997).
96. Eaves, D. J. *et al.* Prevalence of Mutations within the Quinolone Resistance-Determining Region of *gyrA*, *gyrB*, *parC*, and *parE* and Association with Antibiotic Resistance in Quinolone-Resistant *Salmonella enterica*. *Antimicrob. Agents Chemother.* **48**, 4012–4015 (2004).
97. Miró, E. *et al.* [Resistance to quinolones and  $\beta$ -lactams in *Salmonella enterica* due to mutations in topoisomerase-encoding genes, altered cell permeability and expression of an active efflux system]. *Enferm. Infecc. Microbiol. Clin.* **22**, 204–211 (2004).
98. Le, T. A. H. *et al.* Endemic, epidemic clone of *Salmonella enterica* serovar typhi harboring a single multidrug-resistant plasmid in Vietnam between 1995 and 2002. *J. Clin. Microbiol.* **42**, 3094–3099 (2004).
99. Ramsey, C. H. & Edwards, P. R. Resistance of *Salmonellae* isolated in 1959 and 1960 to tetracyclines and chloramphenicol. *Appl. Microbiol.* **9**, 389–391 (1961).
100. Chabbert, Y. A. & Le Minor, L. [Transfer of the resistance to several antibiotics in Enterobacteriaceae. I. Definition--general bacteriology of the transfer]. *Presse Med.* **74**, 2407–2410 (1966).
101. Raffi, A., Véron, M., Kazmierczak, A., Colin, P. & Schirrer, J. [124 cases of *Salmonella panama* infections in infants]. *Pediatric* **25**, 721–733 (1970).
102. Su, L.-H., Chiu, C.-H., Chu, C. & Ou, J. T. Antimicrobial resistance in nontyphoid *Salmonella* serotypes: a global challenge. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **39**, 546–551 (2004).
103. Le Hello, S. *et al.* International spread of an epidemic population of *Salmonella enterica* serotype Kentucky ST198 resistant to ciprofloxacin. *J. Infect. Dis.* **204**, 675–684

(2011).

104. Hopkins, K. L. *et al.* Multiresistant *Salmonella enterica* serovar 4,[5],12:i:- in Europe: a new pandemic strain? *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **15**, 19580 (2010).

105. DNA Sequencing Costs: Data. *National Human Genome Research Institute (NHGRI)* Available at: <https://www.genome.gov/27541954/DNA-Sequencing-Costs-Data>. (Accessed: 10th October 2017)

106. Miescher F. Letter I; to Wilhelm His; Tübingen, February 26th, 1869. *W His Al Eds Histochem. Physiol. Arb. Von Friedrich Miescher--Ausdem Wiss. Briefwechsel Von F Miescher* **1**, 33–38 (1869).

107. Kohara, Y., Akiyama, K. & Isono, K. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495–508 (1987).

108. Sanger, F. *et al.* Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **24;265**, 687–95 (1977).

109. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).

110. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

111. Jarvie, T. Next generation sequencing technologies. *Drug Discov. Today Technol.* **2**, 255–260 (2005).

112. Illumina | Sequencing and array-based solutions for genetic research. Available at: <https://www.illumina.com/>. (Accessed: 9th October 2017)

113. Ion Torrent. Available at: <https://www.thermofisher.com/fr/fr/home/brands/ion-torrent.html>. (Accessed: 9th October 2017)

114. Home - Pacific Biosciences. Available at: <http://www.pacb.com/>. (Accessed: 9th October 2017)

115. Sequencing Platforms | Compare NGS platforms (benchtop, production-scale). Available at: <https://www.illumina.com/systems/sequencing-platforms.html>. (Accessed: 24th October 2017)

116. Ion Torrent Life technologies. The Ion PGM™ System, with 400-base read length chemistry, enables routine high-quality de novo assembly of small genomes. (2013). Available at: <https://tools.thermofisher.com/content/sfs/brochures/Small-Genome-Ecoli-De-Novo-App-Note.pdf>. (Accessed: 24th October 2017)

117. SMRT Sequencing: Read Lengths. *Pacific Biosciences*

118. Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).
119. Chhangawala, S., Rudy, G., Mason, C. E. & Rosenfeld, J. A. The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biol.* **16**, (2015).
120. Designing Next-Generation Sequencing Runs. Available at: <https://genohub.com/next-generation-sequencing-guide/>. (Accessed: 24th October 2017)
121. Criscuolo, A. & Brisse, S. AlienTrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **102**, 500–506 (2013).
122. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
123. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
124. Staden, R. A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* **8**, 3673–3694 (1980).
125. Miller, J. R., Koren, S. & Sutton, G. Assembly Algorithms for Next-Generation Sequencing Data. *Genomics* **95**, 315–327 (2010).
126. Denisov, G. *et al.* Consensus generation and variant detection by Celera Assembler. *Bioinforma. Oxf. Engl.* **24**, 1035–1040 (2008).
127. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).
128. Idury, R. M. & Waterman, M. S. A New Algorithm for DNA Sequence Assembly. *J. Comput. Biol.* **2**, 291–306 (1995).
129. Boisvert, S., Laviolette, F. & Corbeil, J. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **17**, 1519–1533 (2010).
130. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
131. Zerbino, D. R. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **CHAPTER**, Unit-11.5 (2010).
132. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

133. Sims, D., Sudbery, I., Illott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–132 (2014).
134. Escalona, M., Rocha, S. & Posada, D. A comparison of tools for the simulation of genomic next-generation sequencing data. *Nat. Rev. Genet.* **17**, 459–469 (2016).
135. Schbath, S. *et al.* Mapping Reads on a Genomic Sequence: An Algorithmic Overview and a Practical Comparative Analysis. *J. Comput. Biol.* **19**, 796–813 (2012).
136. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
137. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
138. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
139. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
140. PRINCIPLES of Microbial classification; a report of the discussion meeting of the Society for General Microbiology, September 1954. *J. Gen. Microbiol.* **12**, 314–386 (1955).
141. Chang, H. R., Loo, L. H., Jeyaseelan, K., Earnest, L. & Stackebrandt, E. Phylogenetic relationships of *Salmonella typhi* and *Salmonella typhimurium* based on 16S rRNA sequence analysis. *Int. J. Syst. Bacteriol.* **47**, 1253–1254 (1997).
142. Hillis, D. M., Allard, M. W. & Miyamoto, M. M. Analysis of DNA sequence data: phylogenetic inference. *Methods Enzymol.* **224**, 456–487 (1993).
143. Excoffier, L. & Smouse, P. E. Using Allele Frequencies and Geographic Subdivision to Reconstruct Gene Trees within a Species: Molecular Variance Parsimony. *Genetics* **136**, 343–359 (1994).
144. Brooks, D. R. *et al.* Quantitative Phylogenetic Analysis in the 21st Century. *Rev. Mex. Biodivers.* (2007).
145. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
146. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* **57**, 758–771 (2008).
147. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
148. Sanderson, K. E. & Demerec, M. THE LINKAGE MAP OF *SALMONELLA TYPHIMURIUM*. *Genetics* **51**, 897–913 (1965).
149. McClelland, M. *et al.* Complete genome sequence of *Salmonella enterica* serovar



- Typhimurium LT2. *Nature* **413**, 852–856 (2001).
150. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848–852 (2001).
  151. Mukherjee, S. *et al.* Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* **45**, D446–D456 (2017).
  152. Zhang, S. *et al.* *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J. Clin. Microbiol.* **53**, 1685–1692 (2015).
  153. Reyes, R. E., González, C. R., Jiménez, R. C., Herrera, M. O. & Andrade, A. A. Mechanisms of O-Antigen Structural Variation of Bacterial Lipopolysaccharide (LPS). (2012). doi:10.5772/48147
  154. Schnaitman, C. A. & Klena, J. D. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**, 655–682 (1993).
  155. Yoshida, C. E. *et al.* The *Salmonella* In Silico Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft *Salmonella* Genome Assemblies. *PloS One* **11**, e0147101 (2016).
  156. Jolley, K. A. *et al.* Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiol. Read. Engl.* **158**, 1005–1015 (2012).
  157. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
  158. Larsen, M. V. *et al.* Benchmarking of methods for genomic taxonomy. *J. Clin. Microbiol.* **52**, 1529–1539 (2014).
  159. Saputra, D. *et al.* Reads2Type: a web application for rapid microbial taxonomy identification. *BMC Bioinformatics* **16**, 398 (2015).
  160. Larsen, M. V. *et al.* Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *J. Clin. Microbiol.* **50**, 1355–1361 (2012).
  161. Carattoli, A. *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
  162. Leekitcharoenphon, P. *et al.* snpTree - a web-server to identify and construct SNP trees from whole genome sequence data. *BMC Genomics* **13**, S6 (2012).
  163. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the Problem of Comparing Whole Bacterial Genomes across Different Sequencing Platforms. *PLOS ONE* **9**, e104984 (2014).
  164. Leekitcharoenphon, P., Nielsen, E. M., Kaas, R. S., Lund, O. & Aarestrup, F. M. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. *PloS One* **9**, e87991 (2014).

165. Joensen, K. G. *et al.* Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **52**, 1501–1510 (2014).
166. Nurk, S. *et al.* Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. in *Research in Computational Molecular Biology* 158–170 (Springer, Berlin, Heidelberg, 2013). doi:10.1007/978-3-642-37195-0\_13
167. enterobase / enterobase-web / wiki / Salmonella Statistics — Bitbucket. Available at: <https://bitbucket.org/enterobase/enterobase-web/wiki/Salmonella%20Statistics>. (Accessed: 12th October 2017)
168. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30**, 2068–2069 (2014).
169. Laing, C. *et al.* Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics* **11**, 461 (2010).
170. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* **39**, W347–W352 (2011).
171. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
172. Casjens, S. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* **49**, 277–300 (2003).
173. Land, M. *et al.* Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* **15**, 141–161 (2015).
174. NCBI. Genome Information by organism - Prokaryotes. Available at: <https://www.ncbi.nlm.nih.gov/genome/browse/>. (Accessed: 17th October 2017)
175. Todd, E. C. Epidemiology of foodborne diseases: a worldwide review. *World Health Stat. Q. Rapp. Trimest. Stat. Sanit. Mond.* **50**, 30–50 (1997).
176. World Health Organization & Foodborne Disease Burden Epidemiology Reference Group. *WHO estimates of the global burden of foodborne diseases.* (2015).
177. Kuijpers, L. M. F. *et al.* Genomic analysis of *Salmonella enterica* serotype Paratyphi A during an outbreak in Cambodia, 2013-2015. *Microb. Genomics* **2**, e000092 (2016).
178. Bishop, O. T. *Bioinformatics and Data Analysis in Microbiology.* (Caister Academic Press, 2014).
179. Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14 Suppl 1**, S7 (2013).
180. Vyahhi, N., Pyshkin, A., Pham, S. & Pevzner, P. A. From de Bruijn Graphs to

Rectangle Graphs for Genome Assembly. in *Algorithms in Bioinformatics* 249–261 (Springer, Berlin, Heidelberg, 2012). doi:10.1007/978-3-642-33122-0\_20

181. Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y. & Hwang, C.-C. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *PLoS ONE* **8**, (2013).

182. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).

183. EFSA (European Food Safety Authority & ECDC (European Centre For Disease Prevention and Control. Multi-country outbreak of Salmonella Enteritidis phage type 8, MLVA type 2-9-7-3-2 and 2-9-6-3-2 infections. <https://ecdc.europa.eu> (2016). Available at: <https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/ROA-ECDC-EFSA-Salmonella%20Enteritidis.pdf>. (Accessed: 15th October 2017)

184. Braden, C. R. Salmonella enterica Serotype Enteritidis and Eggs: A National Epidemic in the United States. *Clin. Infect. Dis.* **43**, 512–517 (2006).

185. Henzler, D. J., Ebel, E., Sanders, J., Kradel, D. & Mason, J. Salmonella enteritidis in eggs from commercial chicken layer flocks implicated in human outbreaks. *Avian Dis.* **38**, 37–43 (1994).

186. Ziebell, K. *et al.* Subtyping of Canadian isolates of Salmonella Enteritidis using Multiple Locus Variable Number Tandem Repeat Analysis (MLVA) alone and in combination with Pulsed-Field Gel Electrophoresis (PFGE) and phage typing. *J. Microbiol. Methods* **139**, 29–36 (2017).

187. World Health Organization (WHO). Global action plan on antimicrobial resistance. Geneva: World Health Organization. Available at: [http://www.wpro.who.int/entity/drug\\_resistance/resources/global\\_action\\_plan\\_eng.pdf](http://www.wpro.who.int/entity/drug_resistance/resources/global_action_plan_eng.pdf). (Accessed: 26th June 2017)

188. Bud, R. *Penicillin: Triumph and Tragedy*. (Oxford University Press, 2007).

189. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzae. *Br. J. Exp. Pathol.* **10**, 226–236 (1929).

190. Batchelor, F. R., Doyle, F. P., Nayler, J. H. & Rolinson, G. N. Synthesis of penicillin: 6-aminopenicillanic acid in penicillin fermentations. *Nature* **183**, 257–258 (1959).

191. Knox, R. A survey of new penicillins. *Nature* **192**, 492–496 (1961).

192. Datta, N. & Kontomichalou, P. Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. *Nature* **208**, 239–241 (1965).

193. Hedges, R. W., Datta, N., Kontomichalou, P. & Smith, J. T. Molecular specificities of R factor-determined  $\beta$ -lactamases: correlation with plasmid compatibility. *J. Bacteriol.* **117**,

56–62 (1974).

194. Cox, K. E. L. & Schildbach, J. F. Sequence of the R1 plasmid and comparison to F and R100. *Plasmid* **91**, 53–60 (2017).
195. Anderson, E. S. & Datta, N. RESISTANCE TO PENICILLINS AND ITS TRANSFER IN ENTEROBACTERIACEAE. *Lancet Lond. Engl.* **1**, 407–409 (1965).
196. Chabbert, Y. A. & Baudens, J. G. Transmissible resistance to six groups of antibiotics in Salmonella infections. *Antimicrob. Agents Chemother.* **5**, 380–383 (1965).
197. Grindley, J. N. & Nakada, D. The nucleotide sequence of the replication origin of plasmid NTP1. *Nucleic Acids Res.* **9**, 4355–4366 (1981).
198. Anderson, E. S. Drug resistance in Salmonella typhimurium and its implications. *Br. Med. J.* **3**, 333–339 (1968).
199. Anderson, E. S., Threlfall, E. J., Carr, J. M., McConnell, M. M. & Smith, H. R. Clonal distribution of resistance plasmid-carrying Salmonella typhimurium, mainly in the Middle East. *J. Hyg. (Lond.)* **79**, 425–448 (1977).
200. Kelterborn, E. & M, D. V. Salmonella species. First isolations, names and occurrence. *Salmonella Species First Isol. Names Occur.* (1967).
201. European Food Safety Authority & European Centre for Disease Prevention and Control. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015. *EFSA J.* **14**, n/a-n/a (2016).
202. Le Hello, S. *et al.* Highly drug-resistant Salmonella enterica serotype Kentucky ST198-X1: a microbiological study. *Lancet Infect. Dis.* **13**, 672–679 (2013).
203. Mather, A. E. *et al.* Distinguishable epidemics of multidrug-resistant Salmonella Typhimurium DT104 in different hosts. *Science* **341**, 1514–1517 (2013).
204. Petrovska, L. *et al.* Microevolution of Monophasic Salmonella Typhimurium during Epidemic, United Kingdom, 2005-2010. *Emerg. Infect. Dis.* **22**, 617–624 (2016).
205. Partridge, S. R. & Hall, R. M. Evolution of transposons containing blaTEM genes. *Antimicrob. Agents Chemother.* **49**, 1267–1268 (2005).
206. Kingsley, R. A. *et al.* Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res.* **19**, 2279–2287 (2009).
207. Marston, C. K. *et al.* Effects of long-term storage on plasmid stability in Bacillus anthracis. *Appl. Environ. Microbiol.* **71**, 7778–7780 (2005).
208. Gounelle, H. & Szakvary, A. [Antibiotics and food. IV. Use of antibiotics outside medical prescriptions]. *Bull. Acad. Natl. Med.* **150**, 149–156 (1966).
209. Webb, K. E. & Fontenot, J. P. Medicinal drug residues in broiler litter and tissues from

- cattle fed litter. *J. Anim. Sci.* **41**, 1212–1217 (1975).
210. Threlfall, E. J. Epidemic salmonella typhimurium DT 104--a truly international multiresistant clone. *J. Antimicrob. Chemother.* **46**, 7–10 (2000).
211. Franco-Paredes, C. *et al.* Enteric Fever: A Slow Response to an Old Plague. *PLoS Negl. Trop. Dis.* **10**, (2016).
212. Uzzau, S. *et al.* Host adapted serotypes of Salmonella enterica. *Epidemiol. Infect.* **125**, 229–255 (2000).
213. Hirschfeld, L. A NEW GERM OF PARATYPHOID. *The Lancet* **193**, 296–297 (1919).
214. Methner, U., Heller, M. & Bocklisch, H. Salmonella enterica subspecies enterica serovar Choleraesuis in a wild boar population in Germany. *Eur. J. Wildl. Res.* **56**, 493–502 (2010).
215. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinforma. Oxf. Engl.* **31**, 1674–1676 (2015).
216. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
217. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
218. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
219. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
220. Vachaspati, P. & Warnow, T. ASTRID: Accurate Species TREes from Internode Distances. *BMC Genomics* **16**, S3 (2015).
221. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
222. BBTools. *DOE Joint Genome Institute* Available at: <https://jgi.doe.gov/data-and-tools/bbtools/>. (Accessed: 11th April 2018)
223. LAST: genome-scale sequence comparison. Available at: <http://last.cbrc.jp/>. (Accessed: 11th April 2018)
224. Edgar, R. C. PILER-CR: Fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18 (2007).

225. PAthogenisity Island DataBase (PAI DB). Available at: [http://www.paidb.re.kr/view\\_cnpai.php?g\\_acn=NC\\_006086&gp\\_id=1&cn\\_type=c&m=g](http://www.paidb.re.kr/view_cnpai.php?g_acn=NC_006086&gp_id=1&cn_type=c&m=g). (Accessed: 11th April 2018)
226. Cordero, O. X. & Polz, M. F. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* **12**, 263–273 (2014).
227. Krause-Kyora, B. *et al.* Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. *Nat. Commun.* **4**, ncomms3348 (2013).
228. Okoro, C. K. *et al.* Intra-continental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nat. Genet.* **44**, 1215–1221 (2012).

## **Annexes**

# Annexe 1 Chapitre V

## Annexe 1 a. Caractéristiques des génomes et isolats de *Salmonella enterica* étudiés

Isolate name	Alternative name	Species	Recorded serotype	In silico Serotyping	Phage type	Sequence Type	MLST eBURst group	CRISPOL type	Alleles CRISPR1	Alleles CRISPR2
CIP 58-58	Mutton	Salmonella enterica	Typimurium	B1:1,2	2066	1	162	A78	B165	
CIP 106162	Murray/21	Salmonella enterica	Typimurium	B1:1,2	19	1	203	A06	B178	
CIP 106233	Murray/383=Binns=NCTC73	Salmonella enterica	Typimurium	B1:1,2	19	1	876	A12	New	
Felix 8391	Glasgow	Salmonella enterica	Typimurium	B1:1,2	19	1	41	A25	B19	
11K	W5604	Salmonella enterica	Typimurium	B1:1,2	19	1	894prov	New	New	
CIP 106186	Murray/40	Salmonella enterica	Typimurium	B1:1,2	19	1	13			
CIP 106243	Murray/452	Salmonella enterica	Typimurium	B1:1,2	19	1	876		B07	
CIP 106245	Murray/458	Salmonella enterica	Typimurium	B1:1,2	19	1	876			
CIP 106190	Murray/155	Salmonella enterica	Typimurium	B1:1,2	19	1	876			
CIP 106244	Murray/453	Salmonella enterica	Typimurium	B1:1,2	19	1	886prov			
STM 10-51		Salmonella enterica	Typimurium	B1:1,2	19	1	320	A09	B178	
STM 7-51		Salmonella enterica	Typimurium	B1:1,2	19	1	885			
CIP 52-10		Salmonella enterica	Typimurium	B1:1,2	19	1	869prov			
STM 4-52		Salmonella enterica	Typimurium	B1:1,2	19	1	738prov			
STM 46-52		Salmonella enterica	Typimurium	B1:1,2	19	1	142			
STM 88-52		Salmonella enterica	Typimurium	B1:1,2	19	1	117			
STM 11-53		Salmonella enterica	Typimurium	B1:1,2	19	1	26	A06	B07	
STM 27-53		Salmonella enterica	Typimurium	B1:1,2	19	1	874prov			
STM 29-53		Salmonella enterica	Typimurium	B1:1,2	19	1	203			
STM 1-54		Salmonella enterica	Typimurium	B1:1,2	19	1	890prov	New derived from A94 + STM19	New	
STM 15-54		Salmonella enterica	Typimurium	B1:1,2	19	1	216	A200	B169	
STM 18-55		Salmonella enterica	Typimurium	B1:1,2	19	1	884prov			
STM 29-55		Salmonella enterica	Typimurium	B1:1,2	19	1	76			
STM 30-55		Salmonella enterica	Typimurium	B1:1,2	19	1	140	A12	B198	
STM 12-56		Salmonella enterica	Typimurium	B1:1,2	19	1	117			
STM 14-56		Salmonella enterica	Typimurium	B1:1,2	19	1	117			
STM 17-56		Salmonella enterica	Typimurium	B1:1,2	19	1	696	A33	B36	
STM 27-56		Salmonella enterica	Typimurium	B1:1,2	302	1	302	A140	B87	
STM 28-56		Salmonella enterica	Typimurium	B1:1,2	1805	1	696			
STM 29-56		Salmonella enterica	Typimurium	B1:1,2	243	1	895prov			
STM 31-56		Salmonella enterica	Typimurium	B1:1,2	140	1	140			
STM 32-56		Salmonella enterica	Typimurium	B1:1,2	1649	1	882prov			
STM 38-56		Salmonella enterica	Typimurium	B1:1,2	117	1	117			
STM 38-56		Salmonella enterica	Typimurium	B1:1,2	19	1	786	New	B14	
STM 46-56		Salmonella enterica	Typimurium	B1:1,2	98	1	878prov			
STM 47-56		Salmonella enterica	Typimurium	B1:1,2	19	1	415			
STM 50-56		Salmonella enterica	Typimurium	B1:1,2	19	1	864prov			
STM 52-56		Salmonella enterica	Typimurium	B1:1,2	19	1	865prov			
STM 56-72		Salmonella enterica	Typimurium	B1:1,2	19	1	868prov			
STM 57-56		Salmonella enterica	Typimurium	B1:1,2	19	1	875prov			
STM 11-57		Salmonella enterica	Typimurium	B1:1,2	15	1	15			
STM 14-57		Salmonella enterica	Typimurium	B1:1,2	19	1	877prov			
STM 15-57		Salmonella enterica	Typimurium	B1:1,2	19	1	320			
STM 25-57		Salmonella enterica	Typimurium	B1:1,2	19	1	203			
STM 28-57		Salmonella enterica	Typimurium	B1:1,2	19	1	140			
STM 35-57		Salmonella enterica	Typimurium	B1:1,2	19	1	91			
STM 36-57		Salmonella enterica	Typimurium	B1:1,2	19	1	880prov			
STM 37-57		Salmonella enterica	Typimurium	B1:1,2	19	1	13			
STM 38-57		Salmonella enterica	Typimurium	B1:1,2	19	1	872prov			
STM 4-57		Salmonella enterica	Typimurium	B1:1,2	19	1	930prov			
STM 5-57		Salmonella enterica	Typimurium	B1:1,2	3014	1	20			
STM 8-57		Salmonella enterica	Typimurium	B1:1,2	19	1	13			
STM 10-58		Salmonella enterica	Typimurium	B1:1,2	19	1	117			
STM 12-58		Salmonella enterica	Typimurium	B1:1,2	19	1	912prov			
STM 17-58		Salmonella enterica	Typimurium	B1:1,2	19	1	415			
STM 18-58		Salmonella enterica	Typimurium	B1:1,2	19	1	185			
STM 19-58		Salmonella enterica	Typimurium	B1:1,2	19	1	892prov	A25	B20	
STM 2-58		Salmonella enterica	Typimurium	B1:1,2	19	1	911prov			
STM 31-58		Salmonella enterica	Typimurium	B1:1,2	19	1	185			
STM 32-58		Salmonella enterica	Typimurium	B1:1,2	19	1	869prov			
STM 33-58		Salmonella enterica	Typimurium	B1:1,2	19	1	869prov			
STM 42-58		Salmonella enterica	Typimurium	B1:1,2	19	1	867prov			
STM 43-58		Salmonella enterica	Typimurium	B1:1,2	19	1	867prov			
STM 50-58		Salmonella enterica	Typimurium	B1:1,2	19	1	738prov			
STM 51-58		Salmonella enterica	Typimurium	B1:1,2	19	1	541			
STM 54-58		Salmonella enterica	Typimurium	B1:1,2	19	1	76			
STM 58-58		Salmonella enterica	Typimurium	B1:1,2	19	1	907prov			
STM 61-58		Salmonella enterica	Typimurium	B1:1,2	19	1	203			
STM 70-58		Salmonella enterica	Typimurium	B1:1,2	19	1	20			
STM 77-58		Salmonella enterica	Typimurium	B1:1,2	19	1	323	A33	B146	
STM 89-58		Salmonella enterica	Typimurium	B1:1,2	19	1	314	A205	B20	
STM 90-58		Salmonella enterica	Typimurium	B1:1,2	19	1	140			
EXPB 5-59		Salmonella enterica	Typimurium	B1:1,2	19	1	160			
STM 15-59		Salmonella enterica	Typimurium	B1:1,2	19	1	203			
STM 1-59		Salmonella enterica	Typimurium	B1:1,2	19	1	117			
STM 17-59		Salmonella enterica	Typimurium	B1:1,2	19	1	235			
STM 19-59		Salmonella enterica	Typimurium	B1:1,2	302	1	866prov	A06	B178	
STM 33-59		Salmonella enterica	Typimurium	B1:1,2	19	1	203			
STM 3-59		Salmonella enterica	Typimurium	B1:1,2	19	1	880prov			
STM 38-59		Salmonella enterica	Typimurium	B1:1,2	19	1	893prov			
STM 44-59		Salmonella enterica	Typimurium	B1:1,2	19	1	889prov			
STM 45-59		Salmonella enterica	Typimurium	B1:1,2	19	1	929prov			
STM 47-59		Salmonella enterica	Typimurium	B1:1,2	98	1	878prov			
STM 51-59		Salmonella enterica	Typimurium	B1:1,2	19	1	879prov			
STM 53-59		Salmonella enterica	Typimurium	B1:1,2	19	1	260			
STM 55-59		Salmonella enterica	Typimurium	B1:1,2	19	1	20			
STM 60-59		Salmonella enterica	Typimurium	B1:1,2	19	1	203			
STM 61-59		Salmonella enterica	Typimurium	B1:1,2	19	1	873prov			



Isolate name	Continent	Country	Host	Host details	Year	Reference	WGS	Fabre et al Yes/No	Sanger	Antimicrobial susceptibility testing
CIP 58-58	Europe	United Kingdom	Food	Unknown	1911	Lancet ID	Yes	No	Yes	Pansusceptible
CIP 106162	Europe	United Kingdom	Unknown	Unknown	1917	Lancet ID	Yes	No	Yes	Pansusceptible
CIP 106233	Europe	France	Human	Unknown	1917	Lancet ID	Yes	No	Yes	Pansusceptible
Felix 8391	Europe	United Kingdom	Unknown	Unknown	1930	Lancet ID	Yes	No	Yes	Pansusceptible
11K	Europe	Denmark	Unknown	Unknown	1934	This study	Yes	No	No	Pansusceptible
CIP 106186	Europe	United Kingdom	Unknown	Unknown	1935	Lancet ID	Yes	No	No	Pansusceptible
CIP 106243	Europe	United Kingdom	Unknown	Unknown	1936	Lancet ID	Yes	No	No	Pansusceptible
CIP 106246	Europe	United Kingdom	Unknown	Unknown	1937	Lancet ID	Yes	No	No	Pansusceptible
CIP 106190	Europe	United Kingdom	Unknown	Unknown	1938	Lancet ID	Yes	No	No	Pansusceptible
CIP 106244	Europe	United Kingdom	Unknown	Unknown	1940	Lancet ID	Yes	No	No	Pansusceptible
STM 10-51	Europe	France	Human	Unknown	1951	Lancet ID	Yes	No	No	Pansusceptible
STM 7-51	Europe	France	Animal	Pig	1951	Lancet ID	Yes	No	Yes	Pansusceptible
CIP 52-10	Europe	France	Unknown	Unknown	1952	Lancet ID	Yes	No	Yes	Pansusceptible
STM 4-52	Europe	France	Human	Unknown	1952	Lancet ID	Yes	No	No	Pansusceptible
STM 46-52	Europe	France	Human	Stool	1952	Lancet ID	Yes	No	No	Pansusceptible
STM 88-52	Europe	Portugal	Human	Stool	1952	Lancet ID	Yes	No	No	Pansusceptible
STM 11-53	Europe	France	Human	Unknown	1952	Lancet ID	Yes	No	No	Pansusceptible
STM 27-53	Europe	France	Human	Unknown	1953	Lancet ID	Yes	No	No	Pansusceptible
STM 29-53	Europe	France	Food	Horse meat	1953	Lancet ID	Yes	No	Yes	Pansusceptible
STM 1-54	Europe	France	Human	Pig	1954	Lancet ID	Yes	No	No	Pansusceptible
STM 15-54	Europe	France	Animal	Stool	1954	Lancet ID	Yes	No	Yes	Pansusceptible
STM 18-55	Europe	France	Human	Unknown	1955	Lancet ID	Yes	No	No	Pansusceptible
STM 29-55	Europe	France	Animal	Duck	1955	Lancet ID	Yes	No	No	Pansusceptible
STM 30-55	Europe	France	Human	Unknown	1955	Lancet ID	Yes	No	No	Pansusceptible
STM 12-56	Europe	Belgium	Human	Unknown	1956	Lancet ID	Yes	No	Yes	Pansusceptible
STM 14-56	Europe	South America	Human	Unknown	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 17-56	Europe	Venezuela	Human	Unknown	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 27-56	Europe	Switzerland	Animal	Bird	1956	Lancet ID	Yes	No	Yes	Pansusceptible
STM 28-56	Asia	Turkey	Human	Unknown	1956	Lancet ID	Yes	No	Yes	Pansusceptible
STM 29-56	Africa	Turkey	Human	Unknown	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 31-56	Africa	Sub Saharan Africa	Human	Pus	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 32-56	Asia	France	Animal	Cow	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 38-56	Asia	Vietnam	Human	Access	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 46-56	Africa	Tunisia	Human	Cerebrospinal Fluid	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 47-56	Africa	Democratic Republic of the Congo	Human	Stool	1956	Lancet ID	Yes	No	Yes	Pansusceptible
STM 50-56	Africa	Democratic Republic of the Congo	Human	Stool	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 52-56	Africa	Senegal	Human	Pus	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 56-72	Asia	Vietnam	Human	Stool	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 57-56	Europe	Vietnam	Animal	Turtle intestine	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 11-57	Africa	France	Human	Unknown	1956	Lancet ID	Yes	No	No	Pansusceptible
STM 14-57	Europe	Rwanda	Unknown	Unknown	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 15-57	Europe	France	Food	Bovine liver	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 25-57	Europe	France	Human	Stool	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 28-57	Europe	France	Human	Stool	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 35-57	Europe	France	Human	Unknown	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 36-57	Europe	France	Human	Stool	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 37-57	Europe	Belgium	Human	Stool	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 38-57	Europe	France	Human	Stool	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 4-57	Africa	France	Animal	Sheep	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 5-57	Africa	Algeria	Animal	Rabbit	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 8-57	Europe	France	Human	Stool	1957	Lancet ID	Yes	No	No	Pansusceptible
STM 10-58	Europe	France	Animal	Pig stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 12-58	Europe	France	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 17-58	Africa	Tunisia	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 18-58	Europe	France	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 19-58	Africa	Senegal	Animal	Calf	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 2-58	Africa	Morocco	Animal	Guinea pig	1958	Lancet ID	Yes	No	Yes	Pansusceptible
STM 31-58	Africa	Democratic Republic of the Congo	Animal	Monkey Blood	1958	Lancet ID	Yes	No	Yes	Pansusceptible
STM 32-58	Europe	France	Animal	Calif	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 3-58	Africa	Morocco	Human	Pus	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 42-58	Europe	France	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 43-58	Europe	France	Animal	Dog	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 50-58	Europe	Italy	Food	Duck meat	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 51-58	Europe	Italy	Food	Duck meat	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 54-58	Europe	France	Animal	Rabbit	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 5-58	Europe	France	Food	Horse meat	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 58-58	Europe	Italy	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 61-58	Europe	France	Human	Calf	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 70-58	Europe	France	Human	Blood	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 77-58	Europe	France	Animal	Pig stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 80-58	Europe	Belgium	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 89-58	Europe	France	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
STM 90-58	Europe	France	Human	Stool	1958	Lancet ID	Yes	No	No	Pansusceptible
EF93-5-59	Europe	France	Animal	Cat	1958	Lancet ID	Yes	Yes	Yes	Pansusceptible
STM 15-59	Africa	Morocco	Human	Blood	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 1-59	Africa	Madagascar	Animal	Guinea pig	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 17-59	Africa	Ethiopia	Human	Rabbit brain	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 33-59	Europe	Tunisia	Human	Unknown	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 3-59	Europe	Italy	Human	Blood	1959	Lancet ID	Yes	No	Yes	Pansusceptible
STM 36-59	Africa	France	Animal	Cattle	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 44-59	Africa	Democratic Republic of the Congo	Human	Stool	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 45-59	Europe	France	Animal	Pig	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 47-59	Africa	France	Animal	Duck	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 51-59	Africa	Cameroun	Human	Blood	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 53-59	Africa	Madagascar	Food	Mayonnaise	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 5-59	Europe	France	Human	Urine	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 60-59	Africa	Madagascar	Animal	Duckling	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 61-59	Africa	Algeria	Animal	Dog	1959	Lancet ID	Yes	No	No	Pansusceptible
STM 79-59	Africa	Democratic Republic of the Congo	Animal	Duckling	1959	Lancet ID	Yes	No	No	Pansusceptible

Isolate name	Alternative name	Species	Recorded serotype	In silico Serotyping	Phage type	Sequence Type	MLST eBURST group	CRISPOL type	Alleles CRISPR1	Alleles CRISPR2
STM 92-59		Salmonella enterica	Typhimurium	B1:1,2		19	1	CRISPOL		
STM 110-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	896prov		
STM 122-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	15		
STM 48-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	871prov		
STM 50-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	738prov		
STM 57-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	157	A34	B56
STM 60-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	314		
STM 65-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	896prov		
STM 92-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	117		
STM 98-60		Salmonella enterica	Typhimurium	B1:1,2		19	1	15		
STM 105-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	05		
STM 111-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	738prov		
STM 116-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	76		
STM 134-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	907prov		
STM 151-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	752prov		
STM 154-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	281		
STM 41-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	15	A54	B147
STM 73-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	41		
STM 75-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	898prov		
STM 80-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	46	A27	B25
STM 82-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	203	A10	B213
STM 92-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	124		
STM 97-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	05	A12	B04
STM 99-61		Salmonella enterica	Typhimurium	B1:1,2		19	1	05		
STM 114-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	05		
STM 115-62		Salmonella enterica	Typhimurium	B1:1,2		3013	1	867prov	A84 (derived from A42 without the d1 B12)	
STM 132-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	01		
STM 142-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	53		
STM 19-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	293		
STM 40-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	850prov		
STM 55-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	924prov		
STM 61-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	46		
STM 65-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	881prov		
STM 71-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	905prov		
STM 72-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	90		
STM 73-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	20		
STM 79-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	898prov		
STM 8-62		Salmonella enterica	Typhimurium	B1:1,2		19	1	117		
STM 117-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	915prov	A1	B08
STM 136-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	59		
STM 147-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	818prov		
STM 156-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	140		
STM 166-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	26		
STM 186-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	85		
STM 189-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	149prov		
STM 191-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	923prov		
STM 222-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	117		
STM 232-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	216		
STM 34-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	85	A10	B37
STM 38-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	05		
STM 39-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	897prov		
STM 57-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	863prov		
STM 6-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	927prov		
STM 92-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	91	A28	B20
STM 9-63		Salmonella enterica	Typhimurium	B1:1,2		19	1	823prov		
STM 124-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	117		
STM 132-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	913prov		
STM 151-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	26		
STM 163-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	906prov		
STM 209-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	313	A175	B20
STM 219-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	117		
STM 264-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	90		
STM 268-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	906prov		
STM 275-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	313		
STM 296-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	738prov		
STM 950-64		Salmonella enterica	Typhimurium	B1:1,2		19	1	20		
STM 105-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	90		
STM 108-65		Salmonella enterica	Typhimurium	B1:1,2	DT1	19	1	919prov		
STM 111-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	90		
STM 113-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	323		
STM 127-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	461prov		
STM 133-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	917prov		
STM 139-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	877prov		
STM 157-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	933prov		
STM 162-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	415	A16	B14
STM 165-65		Salmonella enterica	Typhimurium	B1:1,2		3021	1	149	A27	B107
STM 169-65		Salmonella enterica	Typhimurium	B1:1,2		3021	1	149		
STM 170-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	909prov		
STM 176-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	850prov		
STM 181-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	323		
STM 201-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	59	A30	B21
STM 210-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	48		
STM 212-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	46		
STM 215-65		Salmonella enterica	Typhimurium	B1:1,2		313	1	910prov		
STM 216-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	686		
STM 219-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	842prov		
STM 225-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	20		
STM 23-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	931prov		
STM 23-65		Salmonella enterica	Typhimurium	B1:1,2		19	1	46		

Isolate name	Continent	Country	Host	Host details	Year	Reference	WGS	Fabre et al Yes/No	Sanger	Antimicrobial susceptibility testing
STM 92-59	Europe	Belgium	Human	Stool	1959	Lancet ID	Yes	No	No	Panusceptible
STM 110-60	Africa	Tunisia	Human	Unknown	1960	Lancet ID	Yes	No	No	ApSifsu
STM 122-60	Europe	France	Human	Stool	1960	Lancet ID	Yes	No	No	Panusceptible
STM 48-60	Europe	France	Human	Stool	1960	Lancet ID	Yes	No	No	Panusceptible
STM 50-60	Europe	France	Animal	Mouse	1960	Lancet ID	Yes	No	No	Panusceptible
STM 57-60	Europe	Italy	Animal	Colt	1960	Lancet ID	Yes	No	Yes	Panusceptible
STM 60-60	Europe	Belgium	Human	Stool	1960	Lancet ID	Yes	No	No	Ap
STM 65-60	Africa	Tunisia	Human	Unknown	1960	Lancet ID	Yes	No	No	Panusceptible
STM 92-60	Europe	France	Human	Stool	1960	Lancet ID	Yes	No	No	Panusceptible
STM 98-60	Europe	Belgium	Human	Unknown	1960	Lancet ID	Yes	No	No	Panusceptible
STM 105-61	Asia	Vietnam	Human	Gecko viscera	1961	Lancet ID	Yes	No	No	Panusceptible
STM 111-61	Europe	France	Human	Stool	1961	Lancet ID	Yes	No	No	Panusceptible
STM 116-61	Europe	France	Food	Unknown	1961	Lancet ID	Yes	No	No	Panusceptible
STM 134-61	Europe	France	Animal	Guinea pig	1961	Lancet ID	Yes	No	No	Panusceptible
STM 151-61	Africa	Mozambique	Animal	Pig	1961	Lancet ID	Yes	No	No	Panusceptible
STM 154-61	Europe	Belgium	Human	Stool	1961	Lancet ID	Yes	No	No	Panusceptible
STM 73-61	Africa	Morocco	Animal	Hedgehog Stool	1961	Lancet ID	Yes	No	Yes	Panusceptible
STM 75-61	Africa	Morocco	Animal	Hedgehog Stool	1961	Lancet ID	Yes	No	No	Panusceptible
STM 80-61	Europe	Portugal	Food	Meat	1961	Lancet ID	Yes	No	No	Panusceptible
STM 82-61	Europe	France	Human	Stool	1961	Lancet ID	Yes	No	No	Panusceptible
STM 92-61	Europe	France	Animal	Bird in cage	1961	Lancet ID	Yes	No	Yes	Panusceptible
STM 97-61	Asia	Vietnam	Animal	Gecko viscera	1961	Lancet ID	Yes	No	No	Panusceptible
STM 99-61	Asia	Vietnam	Unknown	Unknown	1961	Lancet ID	Yes	No	No	Panusceptible
STM 114-62	Asia	Vietnam	Human	Stool	1962	Lancet ID	Yes	No	No	Panusceptible
STM 115-62	Asia	Vietnam	Human	Stool	1962	Lancet ID	Yes	No	No	Panusceptible
STM 133-62	Europe	Italy	Food	Duck soup	1962	Lancet ID	Yes	No	No	Panusceptible
STM 142-62	Europe	Portugal	Food	Chicken	1962	Lancet ID	Yes	No	Yes	Panusceptible
STM 19-62	Europe	Portugal	Food	Blood	1962	Lancet ID	Yes	No	No	Panusceptible
STM 40-62	Africa	Senegal	Human	Unknown	1962	Lancet ID	Yes	No	No	Panusceptible
STM 55-62	Europe	France	Human	Unknown	1962	Lancet ID	Yes	No	No	Panusceptible
STM 61-62	Europe	France	Animal	Porcine	1962	Lancet ID	Yes	No	No	Panusceptible
STM 65-62	Europe	Belgium	Human	Stool	1962	Lancet ID	Yes	No	No	Panusceptible
STM 71-62	Africa	Morocco	Animal	Rat	1962	Lancet ID	Yes	No	No	Panusceptible
STM 72-62	Europe	France	Animal	Canary	1962	Lancet ID	Yes	No	No	Panusceptible
STM 73-62	Europe	Belgium	Human	Stool	1962	Lancet ID	Yes	No	No	Panusceptible
STM 79-62	Europe	France	Human	Unknown	1962	Lancet ID	Yes	No	No	Panusceptible
STM 8-62	Europe	France	Human	Pigeon	1962	Lancet ID	Yes	No	No	Panusceptible
STM 117-63	Europe	France	Human	Unknown	1963	Lancet ID	Yes	No	No	Panusceptible
STM 136-63	Europe	Romania	Human	Unknown	1963	Lancet ID	Yes	No	No	Panusceptible
STM 147-63	Europe	France	Human	Pigeon	1963	Lancet ID	Yes	No	No	Panusceptible
STM 15-63	Europe	France	Human	Guinea pig	1963	Lancet ID	Yes	No	Yes	Panusceptible
STM 156-63	South America	French Guiana	Human	Stool	1963	Lancet ID	Yes	No	No	Panusceptible
STM 166-63	Europe	France	Animal	Chicken	1963	Lancet ID	Yes	No	No	Panusceptible
STM 184-63	Europe	France	Animal	Dog	1963	Lancet ID	Yes	No	No	Panusceptible
STM 185-63	Europe	France	Animal	Chick	1963	Lancet ID	Yes	No	No	Panusceptible
STM 194-63	Africa	Cameroon	Unknown	Unknown	1963	Lancet ID	Yes	No	No	Panusceptible
STM 222-63	Africa	Morocco	Animal	Cow	1963	Lancet ID	Yes	No	No	Panusceptible
STM 3-63	Africa	Morocco	Unknown	Unknown	1963	Lancet ID	Yes	No	No	Panusceptible
STM 3-63	Asia	Vietnam	Unknown	Unknown	1963	Lancet ID	Yes	No	No	Panusceptible
STM 36-63	Europe	Belgium	Human	Unknown	1963	Lancet ID	Yes	No	No	Panusceptible
STM 38-63	Europe	France	Animal	Bird	1963	Lancet ID	Yes	No	No	Panusceptible
STM 57-63	Europe	France	Human	Stool	1963	Lancet ID	Yes	No	No	Panusceptible
STM 8-63	Europe	France	Human	Stool	1963	Lancet ID	Yes	No	No	Panusceptible
STM 9-63	Europe	Italy	Animal	Pig	1963	Lancet ID	Yes	No	No	Panusceptible
STM 92-63	Europe	France	Animal	Porcine	1963	Lancet ID	Yes	No	Yes	Panusceptible
STM 9-63	Asia	Vietnam	Human	Stool	1963	Lancet ID	Yes	No	No	Panusceptible
STM 124-64	Europe	France	Animal	Rabbit	1964	Lancet ID	Yes	No	No	Panusceptible
STM 132-64	Europe	Belgium	Human	Unknown	1964	Lancet ID	Yes	No	No	Panusceptible
STM 151-64	Asia	Vietnam	Human	Stool	1964	Lancet ID	Yes	No	Yes	Panusceptible
STM 163-64	Europe	France	Animal	Guinea pig	1964	Lancet ID	Yes	No	No	Panusceptible
STM 209-64	Europe	France	Food	Sausage	1964	Lancet ID	Yes	No	No	Panusceptible
STM 219-64	Europe	Romania	Animal	Chicken	1964	Lancet ID	Yes	No	No	Panusceptible
STM 264-64	Africa	Morocco	Human	Unknown	1964	Lancet ID	Yes	No	No	Panusceptible
STM 269-64	Africa	Senegal	Human	Unknown	1964	Lancet ID	Yes	No	No	Panusceptible
STM 275-64	Europe	France	Human	Stool	1964	Lancet ID	Yes	No	No	Panusceptible
STM 296-64	Asia	India	Human	Unknown	1964	Lancet ID	Yes	No	Yes	Panusceptible
STM 658-64	Africa	Mali	Human	Stool	1964	Lancet ID	Yes	No	No	Panusceptible
STM 105-65	Europe	France	Animal	Guinea pig spleen	1965	Lancet ID	Yes	No	No	Panusceptible
STM 109-65	Africa	Senegal	Human	Unknown	1965	Lancet ID	Yes	No	No	Panusceptible
STM 114-65	Europe	France	Human	Unknown	1965	Lancet ID	Yes	No	No	Panusceptible
STM 113-65	Europe	France	Human	Stool	1965	Lancet ID	Yes	No	No	Panusceptible
STM 127-65	Africa	Senegal	Human	Stool	1965	Lancet ID	Yes	No	No	Panusceptible
STM 132-65	Asia	Vietnam	Human	Blood	1965	Lancet ID	Yes	No	No	Panusceptible
STM 139-65	Europe	France	Animal	Guinea pig	1965	Lancet ID	Yes	No	No	Panusceptible
STM 158-65	Europe	France	Human	Stool	1965	Lancet ID	Yes	No	No	Panusceptible
STM 152-65	Central America	Sanctia	Human	Unknown	1965	Lancet ID	Yes	No	No	Panusceptible
STM 167-65	South America	Argentina	Human	Unknown	1965	Lancet ID	Yes	No	No	Panusceptible
STM 162-65	South America	Venezuela	Animal	Chicken	1965	Lancet ID	Yes	No	Yes	Panusceptible
STM 165-65	South America	Venezuela	Animal	Chicken	1965	Lancet ID	Yes	No	Yes	Panusceptible
STM 169-65	South America	Peru	Human	Unknown	1965	Lancet ID	Yes	No	No	Panusceptible
STM 170-65	South America	Peru	Human	Meat industry	1965	Lancet ID	Yes	No	No	Panusceptible
STM 176-65	Europe	Italy	Food	Meat industry	1965	Lancet ID	Yes	No	No	Panusceptible
STM 184-65	Europe	France	Animal	Pigeon	1965	Lancet ID	Yes	No	No	Panusceptible
STM 204-65	Europe	France	Animal	Horse	1965	Lancet ID	Yes	No	Yes	Panusceptible
STM 210-65	Europe	France	Human	Stool	1965	Lancet ID	Yes	No	No	Panusceptible
STM 212-65	Africa	Cameroon	Human	Stool	1965	Lancet ID	Yes	No	No	Panusceptible
STM 215-65	Africa	Turkey	Human	Abscess	1965	Lancet ID	Yes	No	No	Panusceptible
STM 216-65	Asia	Chad	Human	Chicken	1965	Lancet ID	Yes	No	No	Panusceptible
STM 216-65	Asia	Iran	Human	Chicken	1965	Lancet ID	Yes	No	No	Panusceptible
STM 225-65	Asia	Portugal	Human	Pig	1965	Lancet ID	Yes	No	No	Panusceptible
STM 235-65	Europe	France	Human	Stool	1965	Lancet ID	Yes	No	No	Panusceptible

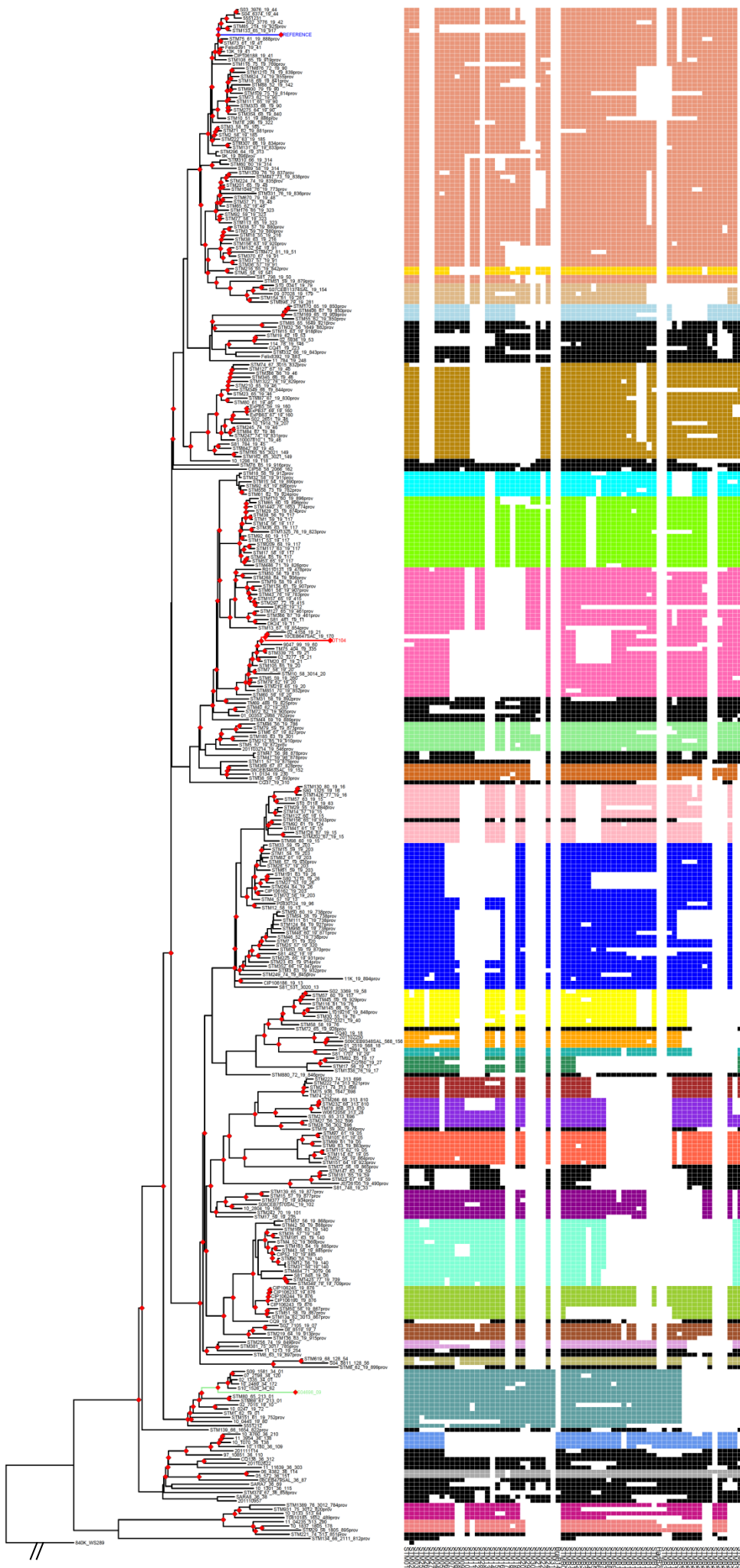
Isolate name	Alternative name	Species	Recorded serotype	In silico Serotyping	Phage type	Sequence Type	MLST eBURST group	CRISPOL type	Alleles CRISPR1	Alleles CRISPR2
STM 53-65		Salmonella enterica	Typimurium	B:1,2		19	1	117		
STM 54-65		Salmonella enterica	Typimurium	B:1,2		19	1	117		
STM 65-214		Salmonella enterica	Typimurium	B:1,2		19	1	925prov		
STM 72-65		Salmonella enterica	Typimurium	B:1,2		19	1	928prov		
STM 78-65		Salmonella enterica	Typimurium	B:1,2		19	1	916prov		
STM 80-65		Salmonella enterica	Typimurium	B:1,2		213	1	01		
STM 85-65		Salmonella enterica	Typimurium	B:1,2		1649	1	921prov		
STM 92-65		Salmonella enterica	Typimurium	B:1,2		19	1	17		
STM 134-66	TM 66-139	Salmonella enterica	Typimurium	B:1,2		2111	243	812prov		New
STM 139-66		Salmonella enterica	Typimurium	B:1,2		1654	1	822prov		
STM 145-66		Salmonella enterica	Typimurium	B:1,2		19	1	76	A177	
STM 307-66		Salmonella enterica	Typimurium	B:1,2		19	1	834prov		
STM 310-66		Salmonella enterica	Typimurium	B:1,2		19	1	314		
STM 332-66		Salmonella enterica	Typimurium	B:1,2		19	1	843prov		
STM 333-66		Salmonella enterica	Typimurium	B:1,2		19	1	90	A25	B24
STM 345-66		Salmonella enterica	Typimurium	B:1,2		19	1	46		
STM 349-66		Salmonella enterica	Typimurium	B:1,2		19	1	844prov		
STM 352-66		Salmonella enterica	Typimurium	B:1,2		19	1	847prov		
STM 366-66		Salmonella enterica	Typimurium	B:1,2		19	1	46		
ExpB 63-67		Salmonella enterica	Typimurium	B:1,2		19	1	160		
STM 126-67		Salmonella enterica	Typimurium	B:1,2		19	1	15		
STM 127-67		Salmonella enterica	Typimurium	B:1,2		19	1	46		
STM 131-67	TM 67-13	Salmonella enterica	Typimurium	B:1,2		19	1	833prov		
STM 13-67		Salmonella enterica	Typimurium	B:1,2		19	1	854prov		
STM 202-67		Salmonella enterica	Typimurium	B:1,2		19	1	15		
STM 20-67	TM 67-20	Salmonella enterica	Typimurium	B:1,2		19	1	21		
STM 23-67	TM 67-23	Salmonella enterica	Typimurium	B:1,2		19	1	59		
STM 366-67		Salmonella enterica	Typimurium	B:1,2		19	1	461prov		
STM 369-67		Salmonella enterica	Typimurium	B:1,2		19	1	828prov		
STM 370-67		Salmonella enterica	Typimurium	B:1,2		19	1	91		
STM 379-67		Salmonella enterica	Typimurium	B:1,2		36	138	858prov		
STM 406-67		Salmonella enterica	Typimurium	B:1,2		19	1	850prov		
STM 6-67	TM 67-6	Salmonella enterica	Typimurium	B:1,2		2015	1	827prov		
STM 69-67		Salmonella enterica	Typimurium	B:1,2		313	1	01		
STM 84-67		Salmonella enterica	Typimurium	B:1,2		19	1	46		
STM 87-67		Salmonella enterica	Typimurium	B:1,2		19	1	830prov		
STM 232-68		Salmonella enterica	Typimurium	B:1,2		313	1	810		
STM 266-68		Salmonella enterica	Typimurium	B:1,2		313	1	810		
STM 354-68		Salmonella enterica	Typimurium	B:1,2		19	1	840		
STM 619-68		Salmonella enterica	Typimurium	B:1,2		128	1	54		
ExpB 37-69		Salmonella enterica	Typimurium	B:1,2	DT1	19	1	160		
STM 18-69		Salmonella enterica	Typimurium	B:1,2		19	1	160		
TM 69-488		Salmonella enterica	Typimurium	B:1,2		19	1	841prov		
STM 242-70		Salmonella enterica	Typimurium	B:1,2		19	1	825prov		
STM 377-70		Salmonella enterica	Typimurium	B:1,2		19	1	101		
STM 851-70		Salmonella enterica	Typimurium	B:1,2		19	1	934prov		
STM 37-71		Salmonella enterica	Typimurium	B:1,2		19	1	852prov		
STM 446-71		Salmonella enterica	Typimurium	B:1,2		19	1	48		
STM 484-71		Salmonella enterica	Typimurium	B:1,2		19	1	826prov		
STM 297-72		Salmonella enterica	Typimurium	B:1,2		3019	1	06		
STM 876-72		Salmonella enterica	Typimurium	B:1,2		19	1	415		
STM 880-72		Salmonella enterica	Typimurium	B:1,2		19	1	90		
STM 447-73		Salmonella enterica	Typimurium	B:1,2		19	1	846prov		
STM 558-73		Salmonella enterica	Typimurium	B:1,2		19	1	838prov		
STM 1210-74	TM 74-1210	Salmonella enterica	Typimurium	B:1,2		19	1	782prov		
STM 211-74		Salmonella enterica	Typimurium	B:1,2		313	1	688		
STM 212-74		Salmonella enterica	Typimurium	B:1,2		313	1	688	A76	
STM 221-74		Salmonella enterica	Typimurium	B:1,2		513	243	851prov		New
STM 221-74		Salmonella enterica	Typimurium	B:1,2		313	1	821prov		
STM 222-74		Salmonella enterica	Typimurium	B:1,2		313	1	688		
STM 223-74		Salmonella enterica	Typimurium	B:1,2		19	1	835prov		
STM 224-74	TM 74-224	Salmonella enterica	Typimurium	B:1,2		19	1	46		
STM 245-74		Salmonella enterica	Typimurium	B:1,2		19	1	831prov		
STM 247-74		Salmonella enterica	Typimurium	B:1,2		19	1	845prov		
STM 249-74		Salmonella enterica	Typimurium	B:1,2		19	1	849prov		
STM 256-74	TM 74-256	Salmonella enterica	Typimurium	B:1,2		19	1	855prov		
STM 824-74		Salmonella enterica	Typimurium	B:1,2		19	1	814prov		
STM 109-75	TM 75-109	Salmonella enterica	Typimurium	B:1,2		19	1	769prov		
STM 116-75		Salmonella enterica	Typimurium	B:1,2		19	1	21		
STM 339-75		Salmonella enterica	Typimurium	B:1,2	104	3017	1	785prov		
STM 381-75		Salmonella enterica	Typimurium	B:1,2		3012	243	820prov		
STM 931-75	TM 75-404	Salmonella enterica	Typimurium	B:1,2	DT104	19	1	335		
TM 75-404		Salmonella enterica	Typimurium	B:1,2		1647	243	688		
TM 75-936		Salmonella enterica	Typimurium	B:1,2		19	1	773prov		
STM 1048-76		Salmonella enterica	Typimurium	B:1,2		19	1	829prov		
STM 1322-76		Salmonella enterica	Typimurium	B:1,2		19	1	823prov		
STM 1325-76		Salmonella enterica	Typimurium	B:1,2		19	1	17		
STM 1338-76		Salmonella enterica	Typimurium	B:1,2		19	1	837prov		
STM 1339-76		Salmonella enterica	Typimurium	B:1,2		3012	243	784prov		
STM 1389-76		Salmonella enterica	Typimurium	B:1,2		1653	1	774prov		
STM 1440-76		Salmonella enterica	Typimurium	B:1,2		19	1	836prov		
STM 331-76		Salmonella enterica	Typimurium	B:1,2		19	1	709		
STM 43-76		Salmonella enterica	Typimurium	B:1,2		19	1	783prov		
STM 1425-77		Salmonella enterica	Typimurium	B:1,2		19	1	16		
STM 1426-77		Salmonella enterica	Typimurium	B:1,2		19	1	146	A61	B245
114-78		Salmonella enterica	Typimurium	B:1,2		19	1	322		
TM 78-296	TM78-296	Salmonella enterica	Typimurium	B:1,2	35	19	1	810		
TM 78-858		Salmonella enterica	Typimurium	B:1,2		313	1	709prov		
STM 349-79		Salmonella enterica	Typimurium	B:1,2		19	1	48		
STM 670-79		Salmonella enterica	Typimurium	B:1,2		19	1			

Isolate name	Continent	Country	Host	Host details	Year	Reference	WGS	Fabre et al	Yes/No	Sanger	Antimicrobial susceptibility testing
STM 53-85	Africa	Tunisia	Human	Unknown	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 54-85	Africa	Tunisia	Unknown	Unknown	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 65-214	Asia	Vietnam	Animal	Guinea pig blood	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 72-85	Europe	France	Food	Unknown	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 78-85	Africa	Democratic Republic of the Congo	Animal	Pig	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 80-85	Europe	Portugal	Food	Fish meal	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 85-85	Europe	South America	Animal	Pig	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 92-85	Europe	Italy	Animal	Canary	1965	Lancet ID	Yes	No	No	No	Pansusceptible
STM 134-86	Africa	Cote d'Ivoire	Human	Stool	1966	Lancet ID	Yes	No	No	No	St/Sul
STM 139-86	Africa	Madagascar	Human	Stool	1966	Lancet ID	Yes	No	No	No	Pansusceptible
STM 145-86	Europe	France	Food	Frozen egg	1966	Lancet ID	Yes	No	No	No	Pansusceptible
STM 307-86	Africa	Algeria	Animal	Dog	1966	Lancet ID	Yes	No	Yes	Yes	Pansusceptible
STM 310-86	South America	Argentina	Animal	Hare	1966	Lancet ID	Yes	No	No	No	Pansusceptible
STM 332-86	South America	Brazil	Animal	Rabbit	1966	Lancet ID	Yes	No	No	Yes	St/Sul
STM 333-86	Europe	France	Human	Stool	1966	Lancet ID	Yes	No	No	No	Pansusceptible
STM 345-86	Europe	France	Human	Stool	1966	Lancet ID	Yes	No	No	No	ApSt/Spc/Sul/Tc
STM 349-86	Europe	France	Human	Stool	1966	Lancet ID	Yes	No	No	No	Pansusceptible
STM 352-86	Europe	France	Human	Stool	1966	Lancet ID	Yes	No	No	No	Ap
STM 366-86	Europe	France	Human	Stool	1966	Lancet ID	Yes	No	No	No	Pansusceptible
ExpB 63-67	Europe	France	Human	Blood	1967	Lancet ID	Yes	No	No	No	ApK
STM 126-67	Europe	France	Human	Blood	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 127-67	Europe	France	Human	Stool	1967	Lancet ID	Yes	No	No	No	ApK
STM 131-67	Europe	France	Human	Stool	1967	Lancet ID	Yes	No	No	No	St/Sul
STM 13-67	Africa	Senegal	Animal	Chicken	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 202-67	Europe	France	Animal	Chicken	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 20-67	Africa	Morocco	Human	Blood	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 23-67	Africa	Madagascar	Human	Blood	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 366-67	Africa	Senegal	Human	Blood	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 369-67	Africa	Senegal	Human	Blood	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 370-67	Europe	France	Animal	Monkey (from Hambur	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 379-67	Asia	India	Food	Frog legs	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 406-67	South America	Peru	Animal	Guinea pig	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 6-67	Africa	Democratic Republic of the Congo	Animal	Guinea pig	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 69-67	Europe	Portugal	Feed	Fishmeal	1967	Lancet ID	Yes	No	No	No	St/Sul
STM 74-67	Europe	France	Human	Stool	1967	Lancet ID	Yes	No	No	No	Pansusceptible
STM 84-67	Europe	France	Human	Stool	1967	Lancet ID	Yes	No	No	No	ApSpK/Sul/Tc
STM 87-67	Europe	Italy	Animal	Chick	1967	Lancet ID	Yes	No	No	No	St/Spc/Sul
STM 232-68	Europe	France	Human	Unknown	1968	Lancet ID	Yes	No	No	No	ApSt/Sul/TmpC
STM 266-68	Africa	Tunisia	Human	Unknown	1968	Lancet ID	Yes	No	No	No	Pansusceptible
STM 354-68	Europe	France	Human	Pigeon	1968	Lancet ID	Yes	No	No	No	Pansusceptible
STM 619-68	South America	Brazil	Animal	Human	1968	Lancet ID	Yes	No	No	No	Pansusceptible
ExpB 37-69	Europe	France	Human	Blood	1969	Lancet ID	Yes	No	No	No	Ap
STM 18-69	Europe	France	Human	Stool	1969	Lancet ID	Yes	No	No	No	Pansusceptible
TM 69-488	Africa	Cameroon	Animal	Meuse	1969	Lancet ID	Yes	No	No	No	Pansusceptible
STM 242-70	Africa	Mozambique	Animal	Chick	1970	This study	Yes	No	No	No	Pansusceptible
STM 377-70	Africa	Djibouti	Unknown	Unknown	1970	This study	Yes	No	No	No	Pansusceptible
STM 851-70	Europe	France	Human	Pus	1970	This study	Yes	No	No	No	Pansusceptible
STM 37-71	Africa	Senegal	Human	Blood	1971	This study	Yes	No	No	No	Pansusceptible
STM 446-71	Africa	Tunisia	Animal	Unknown	1971	This study	Yes	No	No	No	ApCa/Sul
STM 484-71	Europe	France	Human	Bird liver	1971	This study	Yes	No	No	No	Pansusceptible
STM 297-72	Africa	Senegal	Human	Stool	1972	This study	Yes	No	No	No	Pansusceptible
STM 876-72	Europe	France	Human	Stool	1972	This study	Yes	No	No	No	Pansusceptible
STM 880-72	Europe	France	Human	Stool	1972	This study	Yes	No	No	No	Pansusceptible
STM 447-73	Europe	France	Human	Stool	1973	This study	Yes	No	No	No	Pansusceptible
STM 658-73	Europe	France	Human	Stool	1973	This study	Yes	No	No	No	Pansusceptible
STM 1210-74	Europe	France	Human	Stool	1974	This study	Yes	No	No	No	Ap
STM 211-74	Africa	Burkina Faso	Unknown	Unknown	1974	This study	Yes	No	No	No	Pansusceptible
STM 212-74	Africa	Burkina Faso	Unknown	Unknown	1974	This study	Yes	No	No	No	Pansusceptible
STM 221-74	Africa	Burkina Faso	Unknown	Unknown	1974	This study	Yes	No	No	Yes	Pansusceptible
STM 222-74	Africa	Burkina Faso	Unknown	Unknown	1974	This study	Yes	No	No	No	Pansusceptible
STM 223-74	Africa	Burkina Faso	Unknown	Unknown	1974	This study	Yes	No	No	No	Pansusceptible
STM 224-74	Europe	France	Human	Stool	1974	This study	Yes	No	No	No	ApSt/Spc/Sul/Tc
STM 245-74	Europe	France	Human	Stool	1974	This study	Yes	No	No	No	Pansusceptible
STM 247-74	Asia	Iraq	Unknown	Unknown	1974	This study	Yes	No	No	No	Pansusceptible
STM 249-74	Asia	Iraq	Unknown	Unknown	1974	This study	Yes	No	No	No	Pansusceptible
STM 256-74	Europe	France	Human	Stool	1974	This study	Yes	No	No	No	Pansusceptible
STM 824-74	Europe	France	Human	Stool	1974	This study	Yes	No	No	No	Pansusceptible
STM 109-75	Europe	France	Human	Stool	1974	This study	Yes	No	No	No	Pansusceptible
STM 116-75	Europe	France	Human	Stool	1974	This study	Yes	No	No	No	Pansusceptible
STM 339-75	Europe	France	Human	Stool	1975	This study	Yes	No	No	No	Pansusceptible
STM 381-75	Europe	France	Food	Milk Powder	1975	This study	Yes	No	No	No	Pansusceptible
STM 931-75	Europe	France	Environment	Water	1975	This study	Yes	No	No	No	Streptomycine
TM 75-404	Europe	France	Environment	Water	1975	This study	Yes	No	No	No	Streptomycine
TM 75-536	Europe	France	Human	Stool	1975	This study	Yes	No	No	No	Pansusceptible
STM 1048-76	Africa	Burkina Faso	Human	Blood	1975	This study	Yes	No	No	No	Pansusceptible
STM 1322-76	Europe	France	Human	Blood	1975	This study	Yes	No	No	No	Pansusceptible
STM 1325-76	Europe	France	Human	Stool	1976	This study	Yes	No	No	No	Pansusceptible
STM 1338-76	Europe	France	Human	Stool	1976	This study	Yes	No	No	No	Pansusceptible
STM 1339-76	Europe	France	Human	Stool	1976	This study	Yes	No	No	No	Pansusceptible
STM 1389-76	Europe	France	Human	Stool	1976	This study	Yes	No	No	No	Pansusceptible
STM 1440-76	Africa	Cote d'Ivoire	Avian	Chicken	1976	This study	Yes	No	No	No	KTe
STM 331-76	Africa	Morocco	Unknown	Unknown	1976	This study	Yes	No	No	No	Pansusceptible
STM 43-76	Africa	Senegal	Food	Pork	1976	This study	Yes	No	No	No	ApSt/Spc/Sul
STM 1425-77	Europe	France	Human	Stool	1976	This study	Yes	No	No	No	Pansusceptible
STM 1426-77	Europe	France	Human	Stool	1977	This study	Yes	No	No	No	Pansusceptible
STM 114-78	Europe	France	Human	Stool	1978	This study	Yes	No	No	No	Pansusceptible
TM 78-296	Europe	France	Human	Stool	1978	This study	Yes	No	No	Yes	Pansusceptible
TM 78-858	Africa	Burkina Faso	Human	Stool	1978	This study	Yes	No	No	No	ApSt/Sul/TmpC
STM 349-79	Europe	France	Human	Stool	1979	This study	Yes	No	No	No	Pansusceptible
STM 670-79	Europe	France	Human	Stool	1979	This study	Yes	No	No	No	Pansusceptible

Isolate name	Alternative name	Species	Recorded serotype	In silico Serotyping	Phage type	Sequence Type	MLST eBurst group	CRISPOL type	Alleles CRISPR1	Alleles CRISPR2
STM 898-79		Salmonella enterica	Typidiumium	B1:12		19	1	281		
STM 900-79		Salmonella enterica	Typidiumium	B1:12	Not typable	19	1	90	A06	B07
STM 900-79		Salmonella enterica	Typidiumium	B1:12	Not typable	19	1	26	A10	B38
S80-1210		Salmonella enterica	Typidiumium	B1:12		19	1	16		
S80-1325		Salmonella enterica	Typidiumium	B1:12		19	1	16		
STM 130-80		Salmonella enterica	Typidiumium	B1:12		19	1	45		
STM 842-80		Salmonella enterica	Typidiumium	B1:12		19	1	29	A36	B31
S81-1707		Salmonella enterica	Typidiumium	B1:12	40	3020	1	11	A16	B16
S81-481		Salmonella enterica	Typidiumium	B1:12	3	19	1	19	A09	B30
S81-481		Salmonella enterica	Typidiumium	B1:12	3	19	1	19	A10	B178
STM 81-531		Salmonella enterica	Typidiumium	B1:12	RDNC	19	1	33	A10	B09
STM 81-531		Salmonella enterica	Typidiumium	B1:12	RDNC	3020	1	13	A10	B09
STM 81-748		Salmonella enterica	Typidiumium	B1:12		19	1	45	A27	B22
STM 81-784		Salmonella enterica	Typidiumium	B1:12	15a	19	1	50	A31	B20
STM 81-798		Salmonella enterica	Typidiumium	B1:12	8	19	1	06	A13	B34
STM 81-848		Salmonella enterica	Typidiumium	B1:12		19	1	51		
STM 472-81		Salmonella enterica	Typidiumium	B1:12		36	138	110	A71	B116
97-10851		Salmonella enterica	Typidiumium	B1:IND		19	1	60		
9047/99		Salmonella enterica	Typidiumium monophasic	B1:IND		19	1	490prov		
J0726105		Salmonella enterica	Typidiumium	B1:12		19	1	46	A27	B25
S1000/7810-1		Salmonella enterica	Typidiumium	B1:12	42	19	1	18	A24	B27
01-2519		Salmonella enterica	Typidiumium monophasic	B1:12		568	1	762prov		
01-00353		Salmonella enterica	Typidiumium	B1:12	RDNC	2988	1	12	A16	B05
DK 28		Salmonella enterica	Typidiumium	B1:12		19	1	848prov		
L1019216		Salmonella enterica	Typidiumium	B1:12		19	1	21	A08	B39
02-3277		Salmonella enterica	Typidiumium	B1:12		19	1	21	A08	B39
02-4158		Salmonella enterica	Typidiumium	B1:12		19	1	21	A08	B39
02-5936		Salmonella enterica	Typidiumium	B1:12		19	1	53		B03
02-7015		Salmonella enterica	Typidiumium	B1:12	104var3	19	1	01	A15	
02-1335		Salmonella enterica	Typidiumium	B1:12	193	34	1	01	A16	B16
DK 24		Salmonella enterica	Typidiumium	B1:12	DT74	19	1	11	A16	B16
SD2-0321		Salmonella enterica	Typidiumium	B1:12	41	19	1	40	A22	B35
SD2-2651		Salmonella enterica	Typidiumium	B1:12	12	19	1	46	A27	B25
SD2-3369		Salmonella enterica	Typidiumium	B1:12	68	19	1	58	A34	B02
SD2-3776		Salmonella enterica	Typidiumium	B1:12	120	19	1	42	A25	B23
SD2-7105		Salmonella enterica	Typidiumium	B1:12		19	1	07	A11	B06
S03-3976		Salmonella enterica	Typidiumium	B1:12	12var	19	1	44	A26	B19
S04-6374		Salmonella enterica	Typidiumium	B1:12	12var	19	1	44	A26	B19
S04-6811		Salmonella enterica	Typidiumium	B1:12	2	128	1	56	A32	B28
5591212		Salmonella enterica	Typidiumium	B1:12		19	1	175	A138	B95
5591231		Salmonella enterica	Typidiumium	B1:15		19	151	41		
S05-2664		Salmonella enterica	Typidiumium	B1:12		36	138	151	A81	B06
06-1392		Salmonella enterica	Typidiumium	B1:12	40	19	1	41	A10	B31
07-2198		Salmonella enterica	Typidiumium	B1:12		36	138	114	A81	B06
R0110131		Salmonella enterica	Typidiumium monophasic	B1:IND	120	34	1	120	A42	E202
S07CEB1137ASAL		Salmonella enterica	Typidiumium	B1:12		19	1	479prov		
06-9519		Salmonella enterica	Typidiumium	B1:12		19	1	154	A54	B59
08CEB3463SAL		Salmonella enterica	Typidiumium	B1:12		19	1	7		
08CEB479SAL		Salmonella enterica	Typidiumium	B1:12		36	138	152	A83	B109
S08CEB7570SAL		Salmonella enterica	Typidiumium	B1:12		19	1	87	A132	B86
09-07028		Salmonella enterica	Typidiumium	B1:12		19	1	109	A52	B57
S09-1581		Salmonella enterica	Typidiumium	B1:12	193	19	1	179	A28	B33
S09CEB9348SAL		Salmonella enterica	Typidiumium monophasic	B1:IND	Not typable	34	1	01	A42	B62
T0810185		Salmonella enterica	Typidiumium	B1:12		568	1	156	A57	
10-1914		Salmonella enterica	Typidiumium	B1:12		1652	243	489prov		
10-2489		Salmonella enterica	Typidiumium	B1:12		34	1	207	A27	B159
10-2804		Salmonella enterica	Typidiumium monophasic	B1:IND		19	1	172	A19	New
10-4760		Salmonella enterica	Typidiumium	B1:12		19	1	166	A12	B186
10-0133		Salmonella enterica	Typidiumium	B1:12		36	138	210	A71	B162
10-0247		Salmonella enterica	Typidiumium	B1:12		513	243	64	A75	B130
10-0445		Salmonella enterica	Typidiumium	B1:12		19	1	72	A15	B12
10-1070		Salmonella enterica	Typidiumium	B1:12		19	1	80	A62	B12
10-1180		Salmonella enterica	Typidiumium	B1:12		36	138	138	A71	B161
10-1288		Salmonella enterica	Typidiumium	B1:12		36	138	109	A71	B162
10-1301		Salmonella enterica	Typidiumium	B1:12		19	1	118	A190	B119
10-1837		Salmonella enterica	Typidiumium	B1:12		36	138	115	A92	B129
10CEB647SAL		Salmonella enterica	Typidiumium	B1:12		1805	243	178	A77	B98
S10-0118		Salmonella enterica	Typidiumium	B1:12	Not typable	19	1	170	A227	New
S10-0341		Salmonella enterica	Typidiumium	B1:12	Not typable	19	1	63	A10	New
S10-1526		Salmonella enterica	Typidiumium	B1:12	Not typable	34	1	79	A54	B60
11-3954		Salmonella enterica	Typidiumium	B1:12	Not typable	36	138	62	A50	B12
201102250		Salmonella enterica	Typidiumium	B1:12		568	138	330	A157	B27
201102622		Salmonella enterica	Typidiumium monophasic	B1:12		36	138	341	A90	B27
201103214		Salmonella enterica	Typidiumium	B1:12		19	1	341	A90	B236
20110957		Salmonella enterica	Typidiumium	B1:12		3137	1	546prov		
201111114		Salmonella enterica	Typidiumium	B1:12		36	138	345	A158	B125
11-0134		Salmonella enterica	Typidiumium	B1:12	Not typable	19	1	353	A71	B125
11-04235		Salmonella enterica	Typidiumium	B1:12		19	1	230	A67	B68
11-11639		Salmonella enterica	Typidiumium	B1:12		513	243	290	A76	B109
11-1213		Salmonella enterica	Typidiumium monophasic	B1:12		36	138	303	A20	B118
11-784		Salmonella enterica	Typidiumium monophasic	B1:IND		19	1	254	A146	B196
W0612058		Salmonella enterica	Typidiumium	B1:12		19	1	248	A211	B45
W0616352		Salmonella enterica	Typidiumium	B1:12		313	1	28		B239
W0616352		Salmonella enterica	Typidiumium	B1:12		19	1	883	derived from A192 (STM18var3)	New
W0616352		Salmonella enterica	Typidiumium	B1:12		19	1	41		
W0616352		Salmonella enterica	Typidiumium	B1:12		19	1	41		
CIP 106188		Salmonella enterica	Typidiumium	B1:12	Not typable	19	1	60	A45	B48
SAR93		Salmonella enterica	Typidiumium	B1:12	Not typable	36	138	60	A37	B10
CO 136		Salmonella enterica	Typidiumium	B1:12		36	138	38		
CO 160		Salmonella enterica	Typidiumium	B1:12	DT136	36	138	312		
CO 37		Salmonella enterica	Typidiumium	B1:12	DT160	19	1	27	A38	B36
CO 40		Salmonella enterica	Typidiumium	B1:12	DT37	19	1	310	A123	B151
CO 41		Salmonella enterica	Typidiumium	B1:12	DT40	19	1	18	A24	B27
CO 9		Salmonella enterica	Typidiumium	B1:12	DT41	19	1	223	A56	B61
P0830124		Salmonella enterica	Typidiumium	B1:12	DT9	19	1	57	A2	B174

Isolate name	Continent	Country	Host	Host details	Year	Reference	WGS	Fabre et al	Sanger	Antimicrobial susceptibility testing
STM 696-79	Europe	France	Human	Stool	1979	This study	Yes	No	No	
STM 905-79	Europe	France	Human	Stool	1979	This study	Yes	No	No	Pansusceptible
S90-1210	Africa	Morocco	Human		1980	This study	Yes	Yes	Yes	Pansusceptible
S90-1325	Europe	France	Bovine		1980	This study	Yes	Yes	Yes	ApSKCTe
STM 136-80	Europe	France	Human	Stool	1980	This study	Yes	No	No	ApSISuCTe
STM 842-80	Europe	France	Human	Stool	1980	This study	Yes	No	No	Pansusceptible
S91-1707	Africa	Egypt	Human		1981	This study	Yes	Yes	Yes	
S91-481	Africa	Senegal	Unknown		1981	This study	Yes	Yes	Yes	
S91-482	Oceania	New Caledonia	Human		1981	This study	Yes	Yes	Yes	
S91-531	Europe	France	Poultry		1981	This study	Yes	Yes	Yes	
S91-748	Oceania	French Polynesia	Human		1981	This study	Yes	Yes	Yes	Pansusceptible
S91-784	Oceania	France	Human		1981	This study	Yes	Yes	Yes	
S91-798	Africa	Madagascar	Human		1981	This study	Yes	Yes	Yes	
S91-948	Europe	France	Avian	Chick	1981	This study	Yes	Yes	Yes	
97-10851	Europe	France	Human	Stool	1981	This study	Yes	Yes	Yes	Pansusceptible
9047690	Asia	India	Human	Stool	1997	This study	Yes	No	No	
J0726105	Europe	Spain	Swine	Unknown	1999	This study	Yes	No	No	
S1000 7810-1	Africa	Norway	Unknown	Pig	1999	This study	Yes	No	No	
01-2519	Europe	Senegal	Hedgehog	Unknown	2000	This study	Yes	No	No	SISuTe
01-00353	Europe	France	Human	Hedgehog	2000	This study	Yes	Yes	Yes	Pansusceptible
DK 28	Africa	Morocco	Unknown	Unknown	2001	This study	Yes	No	No	Pansusceptible
L1019216	Africa	Senegal	Human	Unknown	2001	This study	Yes	Yes	Yes	Pansusceptible
02-3277	Africa	Senegal	Unknown	Unknown	2001	This study	Yes	No	No	SI
02-4158	Europe	France	Human	Stool	2002	This study	Yes	Yes	Yes	ApSISuTe
02-5936	Europe	France	Human	Unknown	2002	This study	Yes	Yes	Yes	
02-7015	Europe	France	Human	Unknown	2002	This study	Yes	No	No	SI
02-1335	Europe	France	Human	Unknown	2002	This study	Yes	Yes	Yes	
DK 24	Africa	Senegal	Human	Unknown	2002	This study	Yes	Yes	Yes	
S02-0321	Europe	France	Human	Unknown	2002	This study	Yes	Yes	Yes	
S02-2651	Europe	France	Human	Unknown	2002	This study	Yes	No	No	SI
S02-3369	Europe	France	Human	Unknown	2002	This study	Yes	Yes	Yes	ApSISuTe
S02-3776	Europe	France	Human	Unknown	2002	This study	Yes	No	No	Pansusceptible
S03-3976	Europe	France	Human	Unknown	2002	This study	Yes	Yes	Yes	Pansusceptible
S04-6374	Europe	France	Human	Unknown	2002	This study	Yes	Yes	Yes	Pansusceptible
S04-8811	Europe	France	Human	Unknown	2002	This study	Yes	Yes	Yes	Pansusceptible
5551212	Asia	France	Animal	Snake	2004	This study	Yes	Yes	Yes	ApSISpToGCSuTmPteNaICip
05-572	Asia	Vietnam	Animal	Pigeon	2004	This study	Yes	Yes	Yes	ApSISpToGCSuTmPteNaICip
S05-2864	Europe	France	Human	Unknown	2005	This study	Yes	No	No	PanPansusceptible
06-4382	Europe	France	Human	Unknown	2005	This study	Yes	Yes	Yes	
07-2198	Europe	France	Human	Unknown	2005	This study	Yes	Yes	Yes	
R0110131	Europe	France	Human	Human	2006	This study	Yes	No	No	
S07CEB11374SAL	Africa	Senegal	Human	Stool	2007	This study	Yes	No	No	Pansusceptible
08-8519	Europe	Senegal	Swine	Pig	2007	This study	Yes	No	No	
08CEB3463SAL	Africa	Algeria	Human	Pig	2008	This study	Yes	No	No	
08CEB479SAL	Europe	France	Food	Chicken	2008	This study	Yes	No	No	
S08CEB7570SAL	Europe	France	Food	Seafood	2008	This study	Yes	No	No	
09-07028	Oceania	France	Avian	Pigeon	2008	This study	Yes	No	No	
S09-1581	Europe	Germany	Unknown	Unknown	2009	This study	Yes	Yes	Yes	ApSISpKToGCSuTmPteNaICip
T0810185	Europe	China	Human	Goose	2009	This study	Yes	No	No	Pansusceptible
10-1914	Africa	Senegal	Animal	Unknown	2009	This study	Yes	No	No	
10-2489	Oceania	French Polynesia	Human	Stool	2010	This study	Yes	No	No	
10-2804	Europe	France	Human	Unknown	2010	This study	Yes	No	No	
10-4760	Europe	France	Human	Unknown	2010	This study	Yes	No	No	ApSISuTmPte
10-0133	Europe	France	Human	Unknown	2010	This study	Yes	No	No	Pansusceptible
10-0247	Europe	France	Human	Unknown	2010	This study	Yes	No	No	
10-0445	Africa	Mauritius	Unknown	Unknown	2010	This study	Yes	No	No	
10-1070	Europe	France	Human	Unknown	2010	This study	Yes	No	No	
10-1180	Europe	France	Human	Unknown	2010	This study	Yes	No	No	
10-1298	America	French Guiana	Human	Stool	2010	This study	Yes	No	No	
10-1301	Europe	France	Human	Unknown	2010	This study	Yes	No	No	
10-1837	Europe	France	Human	Unknown	2010	This study	Yes	No	No	
10CEB647SAL	Europe	France	Unknown	Unknown	2010	This study	Yes	No	No	Pansusceptible
S10-0341	Europe	France	Human	Unknown	2010	This study	Yes	No	No	
S10-1526	Europe	France	Human	Unknown	2010	This study	Yes	Yes	Yes	
11-3954	Europe	France	Food	Milkery	2010	This study	Yes	Yes	Yes	
201102250	Europe	France	Human	Stool	2011	This study	Yes	No	No	ApFox CroCazSKTGSuTmPCTe
201102622	Europe	France	Human	Stool	2011	This study	Yes	No	No	
201103214	Africa	Niger	Human	Unknown	2011	This study	Yes	No	No	
20110957	Europe	France	Human	Unknown	2011	This study	Yes	No	No	
11-0134	Europe	France	Human	Unknown	2011	This study	Yes	No	No	Pansusceptible
11-04235	Europe	France	Human	Unknown	2011	This study	Yes	No	No	
11-11639	Europe	France	Human	Duck	2011	This study	Yes	No	No	
11-1213	Europe	France	Human	Unknown	2011	This study	Yes	No	No	
11-784	America	Mexico	Human	Stool	2011	This study	Yes	No	No	
W0612058	Africa	Martinique (French West Indies)	Human	Stool	2012	This study	Yes	No	No	ApSISuTmPc
Felix 8392	Unknown	Unknown	Unknown	Unknown	<1934	Lancet ID	Yes	No	No	Pansusceptible
13K	Europe	United Kingdom	Unknown	Unknown	<1935	This study	Yes	No	No	Pansusceptible
9K	Europe	Denmark	Unknown	Unknown	<1935	This study	Yes	No	No	Pansusceptible
CIP-106188	Europe	United Kingdom	Unknown	Unknown	<1935	Lancet ID	Yes	No	No	Pansusceptible
SAR7	Europe	Norway	Unknown	Unknown	<1981	This study	Yes	Yes	Yes	Pansusceptible
SAR8	Europe	Finland	Unknown	Unknown	<1981	This study	Yes	Yes	Yes	Pansusceptible
CO 136	Europe	United Kingdom	Unknown	Unknown	<1981	This study	Yes	No	No	
CO 160	Europe	784	Unknown	Unknown	<1981	This study	Yes	No	No	
CO 37	Europe	United Kingdom	Unknown	Unknown	<1981	This study	Yes	No	No	
CO 40	Europe	United Kingdom	Unknown	Unknown	<1981	This study	Yes	No	No	
CO 41	Europe	United Kingdom	Unknown	Unknown	<1981	This study	Yes	No	No	
CO 9	Europe	United Kingdom	Unknown	Unknown	<1981	This study	Yes	No	No	
P0830124	Africa	Central African Republic	Unknown	Unknown	<1981	This study	Yes	No	No	ApSuTmPte

Annexe 1 b. Arbre phylogénétique de maximum de vraisemblance avec les bootstraps et les noms des souches de *S. enterica*



Au bout de chaque branche est indiqué le nom de chaque souche, le ST et le CT, séparés par un « \_ ».  
 Les bootstraps dont la valeur est  $\geq 90$  sont indiqués par un carré rouge.  
 Les génomes de référence sont indiqués en bleu (LT2), rouge (DT104) et vert (S04698-09).

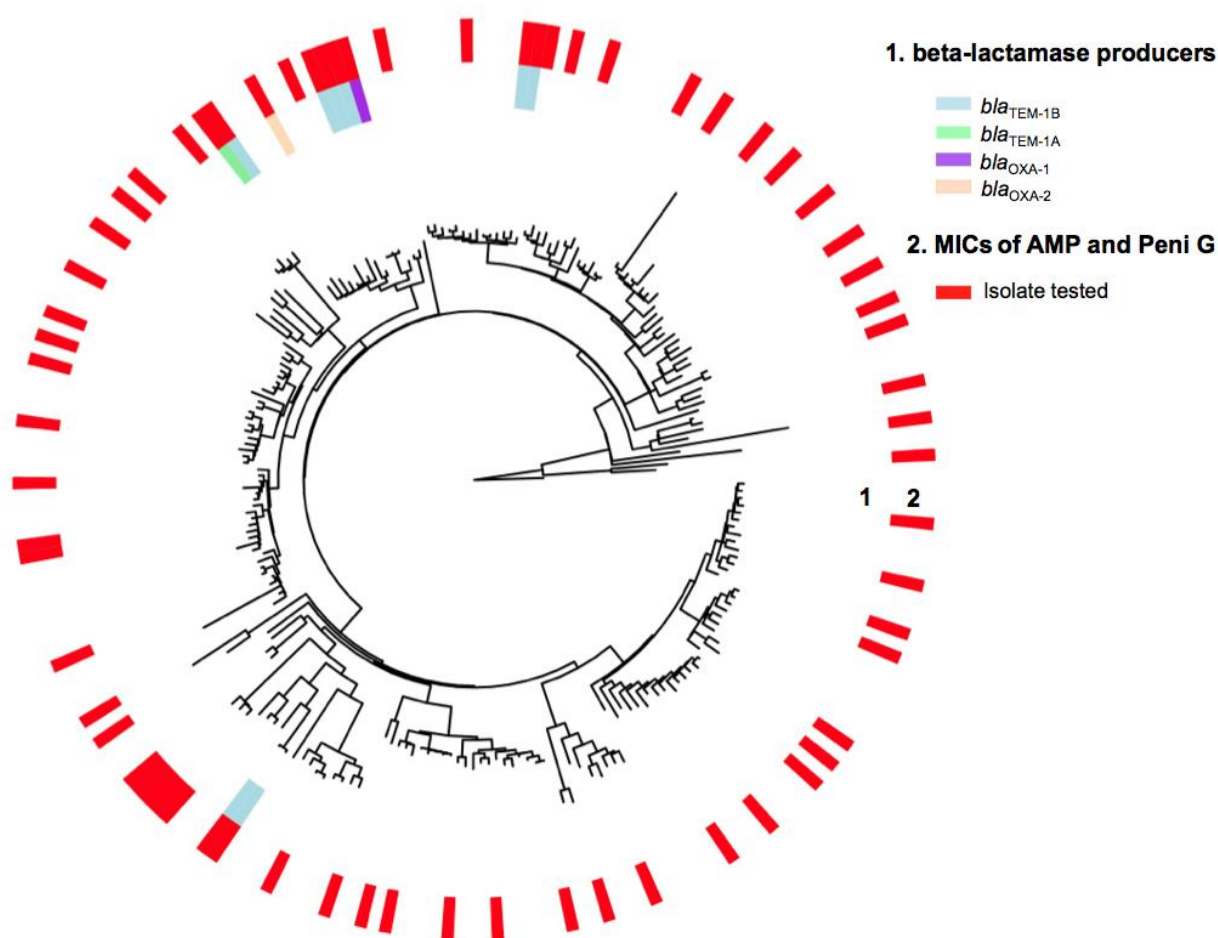


Annexe 2 a. Distribution des souches dans cette étude et pour le séquençage entier du génome

Sources and geographic areas	Global collection isolates	Sequenced isolates
Humans	149	113
Animals, food, feed	113	93
Europe	161	132
Africa	90	58
Asia	19	19
America	15	13
<b>Total</b>	<b>288</b>	<b>225</b>

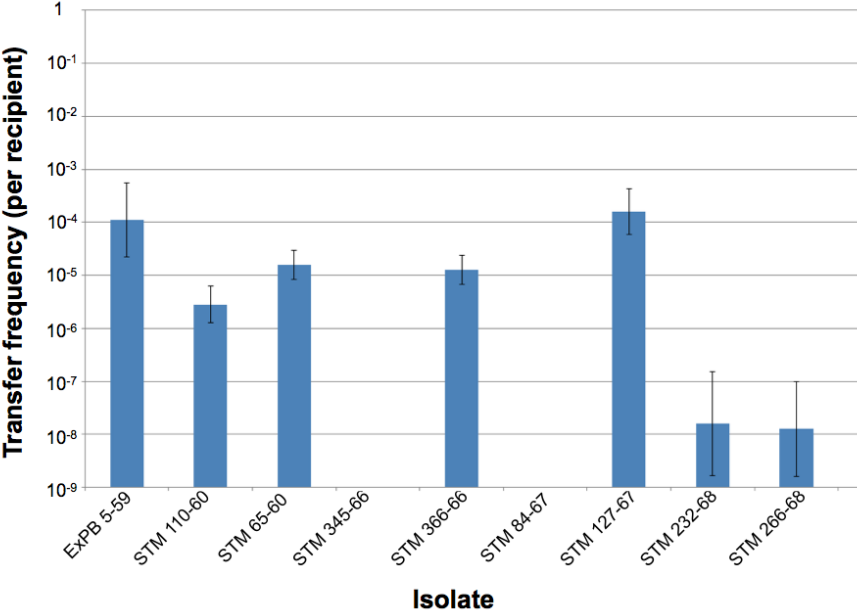
Annexe 2 b. Distribution phylogénétique des CMI pour l'ampicilline et la penicilline G, sur une sélection de 67 isolats de *S. enterica* sérotype Typhimurium.

Le premier cercle indique les souches productrices de  $\beta$ -lactamase, avec quatre gènes différents en bleu, vert, violet et rose représentant *bla*<sub>TEM-1B</sub>, *bla*<sub>TEM-1A</sub>, *bla*<sub>OXA-1</sub> et *bla*<sub>OXA-2</sub> respectivement. Le second cercle indique les isolats testés pour la CMI pour l'ampicilline et la penicilline G.



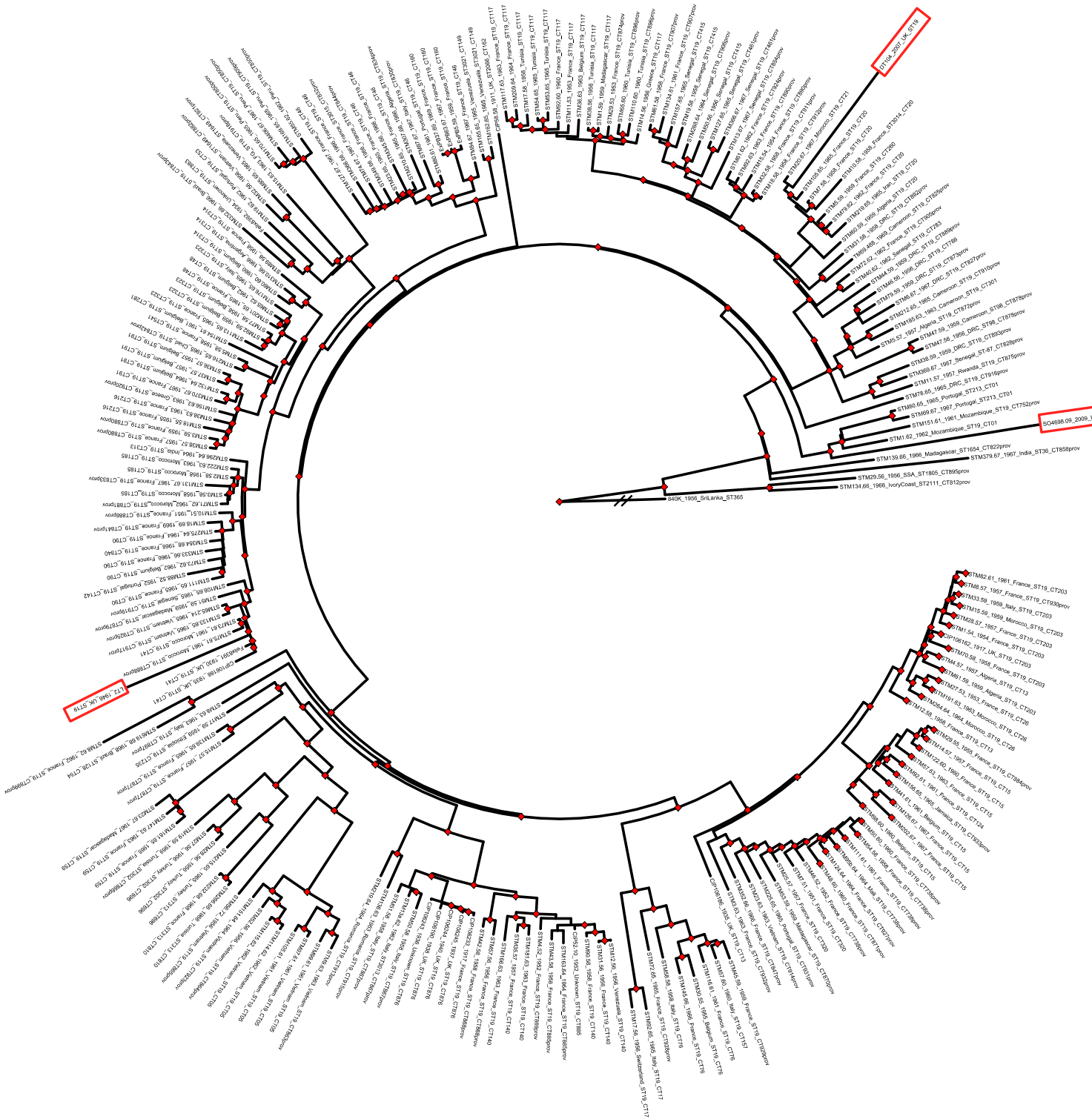
**Annexe 2 c. Fréquence de transfert des différents plasmides de *S. enterica* sérotype Typhimurium.**

Les statistiques ont été obtenues en faisant l'expérience trois fois.



## Annexe 2 d. Arbre phylogénétique de maximum de vraisemblance des souches de *S. enterica* sérotype Typhimurium

Le nom des souches est indiqué au bout de chaque branche. Les bootstraps dont la valeur est  $\geq 90$  sont indiqués par un carré rouge. Les génomes de référence sont encadrés en rouge.



## Annexe 2 e. Caractéristiques des génomes et isolats de *Salmonella enterica* étudiés.

Isolate name	Alternative name	Continent	Country	Host	Host details	Year	Species	Recorded serotype	In silico Serotyping	Antimicrobial susceptibility testing	Resistance gene	MIC AMP (mg/L)	MIC Peni G (mg/L)
840K	Sri Lanka	Asia	Sri Lanka	Human	Unknown	1956	Salmonella enterica	Wakaraseden Typhimurium	B:1,2	Unknown	Unknown		
CIP 106162	Murray/21	Europe	United Kingdom	Unknown	Unknown	1917	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.125	
CIP 106186	Murray/40	Europe	United Kingdom	Unknown	Unknown	1935	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
CIP 106188	Murray/49=Gasgow	Europe	United Kingdom	Unknown	Unknown	~1935	Salmonella enterica	Typhimurium	B:1,2	Str			
CIP 106190	Murray/55	Europe	United Kingdom	Unknown	Unknown	1938	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
CIP 106233	Murray/483=Bnms=NCTC73	Europe	France	Human	Unknown	1917	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.5	3
CIP 106243	Murray/452	Europe	Unknown	Unknown	Unknown	1936	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		1	8
CIP 106244	Murray/453	Europe	United Kingdom	Unknown	Unknown	1940	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
CIP 52.10	Murray/458	Europe	United Kingdom	Unknown	Unknown	1937	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.125	0.75
CIP 52.10	Murray/458	Europe	United Kingdom	Unknown	Unknown	1952	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		≥256	≥256
DT104	Mutton	Europe	United Kingdom	Human	Unknown	2007	Salmonella enterica	Typhimurium	B:1,2	Unknown		≥256	≥256
EXPB 37-69		Europe	France	Human	Blood	1969	Salmonella enterica	Typhimurium	B:1,2	Ad	aadA2, blaCARB-2, floR, sul1, tet(G)	0.125	0.75
EXPB 5-59		Europe	France	Human	Blood	1959	Salmonella enterica	Typhimurium	B:1,2	Ad	blaTEM-1B	≥256	≥256
EXPB 63-67		Europe	France	Human	Blood	1967	Salmonella enterica	Typhimurium	B:1,2	Ad	blaTEM-1B	≥256	≥256
Folk 8391		Europe	United Kingdom	Unknown	Unknown	1930	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
Folk 8392		Europe	Unknown	Unknown	Unknown	~1934	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
LT2	Gasgow NCTC 3040	Europe	United Kingdom	Human	Cattle	1946	Salmonella enterica	Typhimurium (monophasic) Typhimurium	B:1,2	Unknown			
SO4698-09		Africa	Senegal	Human	Stool	2009	Salmonella enterica	Typhimurium	B:1,2	Unknown	strfAB, blaTEM-1B, sul2, tet(B)		
STM 76-67		Africa	Senegal	Human	Stool	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 1-54		Africa	Madagascar	Animal	Pig bit brain	1954	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 1-59		Africa	Madagascar	Animal	Chicken	1954	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 1-62		Africa	Madagascar	Human	Chicken	1962	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 10-51		Europe	France	Human	Unknown	1951	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 10-58		Europe	France	Human	Pig stool	1958	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	6
STM 05-61		Asia	Vietnam	Animal	Gecko viscera	1961	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 05-65		Europe	France	Animal	Guinea pig spleen	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 08-65		Africa	Senegal	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 11-53		Europe	France	Human	Unknown	1953	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		1	6
STM 11-57		Africa	Rwanda	Unknown	Unknown	1957	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 110-60		Africa	Tunisia	Human	Unknown	1960	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 111-61		Europe	France	Human	Stool	1961	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 111-61		Europe	France	Human	Stool	1961	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 113-65		Europe	France	Human	Stool	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 113-65		Europe	France	Human	Stool	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 114-60		Africa	Madagascar	Human	Stool	1960	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 114-62		Asia	Vietnam	Human	Stool	1962	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 115-62		Asia	Vietnam	Human	Stool	1962	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	6
STM 116-61		Europe	France	Food	Unknown	1961	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 117-60		Africa	Morocco	Unknown	Unknown	1960	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 117-63		Europe	France	Human	Stool	1963	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 12-56		South America	Venezuela	Human	Unknown	1956	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 12-56		Europe	France	Human	Stool	1958	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 12-58		Europe	France	Human	Stool	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 121-67		Europe	France	Human	Stool	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 122-60		Africa	Morocco	Animal	Redgenny	1960	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 124-62		Europe	France	Human	Unknown	1964	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 124-62		Europe	France	Human	Unknown	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 125-67		Europe	France	Human	Stool	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 126-67		Europe	Senegal	Human	Blood	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 127-65		Africa	Senegal	Human	Stool	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 127-67		Europe	France	Human	Stool	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 131-67	TM 67-13	Europe	France	Human	Stool	1964	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible	aph(3)-Ia, blaTEM-1A	≥256	≥256
STM 132-64		Europe	Belgium	Human	Unknown	1964	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 133-65		Europe	France	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 134-61		Europe	Vietnam	Animal	Guinea pig	1961	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 134-61		Africa	Cote d'Ivoire	Human	Stool	1960	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	3
STM 134-65		Europe	France	Human	Stool	1960	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	8
STM 134-65		Europe	France	Human	Stool	1960	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.125	0.25
STM 139-65		Europe	France	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 139-66		Africa	Madagascar	Human	Stool	1966	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 139-66	TM 66-139	Europe	Italy	Food	Duck soup	1962	Salmonella enterica	Typhimurium	B:1,2	StrSul	strfAB, sul2	1	8
STM 13a-62		Europe	France	Human	Unknown	1966	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	8
STM 14-56		Europe	Greece	Food	Unknown	1956	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	6
STM 14-57		Europe	France	Food	Bovine liver	1957	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 14-59		Africa	Morocco	Animal	Rat	1959	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 14-59		Europe	France	Food	Frozen egg	1966	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 146-67		Europe	France	Animal	Pigeon	1967	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 147-63		Europe	France	Animal	Pigeon	1963	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	8
STM 147-63		Europe	France	Human	Stool	1954	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	4
STM 15-64		Europe	France	Human	Stool	1954	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	9
STM 15-64		Europe	France	Human	Stool	1954	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		1	4
STM 15-69		Africa	Madagascar	Human	Guinea pig	1959	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 15-63		South America	Mozambique	Human	Guinea pig	1963	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 15-63		Europe	France	Human	Chicken	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 151-61		Africa	Senegal	Human	Stool	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		1	8
STM 151-61		Europe	Vietnam	Human	Stool	1964	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	6
STM 154-61		Europe	Belgium	Human	Stool	1961	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.75	8
STM 156-61		Africa	Madagascar	Animal	Fly	1961	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 156-63		Europe	Greece	Human	Stool	1963	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 156-63		Central America	Jamaica	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 157-65		Africa	Senegal	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 162-65		South America	Venezuela	Animal	Chicken	1965	Salmonella enterica	Typhimurium	B:1,2	Str			
STM 162-65		Europe	France	Human	Guinea pig	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 165-64		Europe	Venezuela	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 165-65		Europe	Venezuela	Human	Chicken	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 166-63		Europe	France	Human	Chicken	1963	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 169-65		South America	Peru	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 169-65		Europe	Switzerland	Animal	Bird	1956	Salmonella enterica	Typhimurium	B:1,2	Str		1	8
STM 17-58		Africa	Tunisia	Human	Stool	1958	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 17-59		Africa	Ethiopia	Human	Stool	1959	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 170-65		South America	Peru	Human	Unknown	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 176-65		Europe	Italy	Food	Meat industry	1965	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 179-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 18-55		Europe	France	Human	Unknown	1955	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		0.25	1
STM 18-58		Europe	France	Animal	Calf	1958	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible			
STM 18-58		Europe	France	Human	Stool	1958	Salmonella enterica	Typhimurium	B:1,2	Pansusceptible		1	12
STM 181-63		Europe	France	Human	Stool	1963	Salmonella enterica	Typhimurium	B:1,2</				

Isolate name	Phage type	Sequence Type	MLST eBurst group	CRISPOL type	Plasmid	pMLST InCf	pMLST Other Inc	Reference	Enterobase Überstrain number	WGS
8406		385						This study	SAL_EA4356A	Yes
CP 06162		19	1	203	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 06186		19	1	13	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 06188		19	1	41	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 06190		19	1	876	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 06233		19	1	876	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 06243		19	1	876	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 06244		19	1	876	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 06245		19	1	885	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 52_10		19	1	162	IncX1 (95.5%)	S1A-B17		This study	SAL_EA4356A	Yes
CP 58_58		2066	1		-			This study	SAL_EA4356A	Yes
DT104	DT104	19	1	160	IncX1			Mather et al. Science 2013	SAL_EA4509A	Yes
ExpB 37-69		19	1	160	IncX1			This study	SAL_EA4509A	Yes
ExpB 5-59		19	1	160	IncX1			This study	SAL_EA4509A	Yes
ExpB 63-67		19	1	41	IncFB(S), IncF(S), IncI1	S1A-B17	Inc11 andA_3_pil_1_rep1_1_sogs_1_tbrA_16	This study	SAL_EA4370AA	Yes
Felix 8391		19	1	883	IncFB(S), IncF(S), IncI1	S1A-B17		This study	SAL_EA456AA	Yes
Felix 8392		19	1		IncFB(S), IncF(S)			McClanland et al. Nature 2001		Reference genome
L72		34	1		IncX1			Petrovska et al. Emerg Infect Dis 2016		Comparison genome
SO4698-09	DT193	19	1	283	Unknown			This study		No
STM 76-67		19	1	203	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4496AA	Yes
STM 1-54		19	1	117	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA8192AA	Yes
STM 1-59		19	1	01	IncF(S) (99.5%)	S1A-B17		This study	SAL_EA4430AA	Yes
STM 1-62		19	1	886p	IncFB(S), IncF(S), IncI2 (98.4%)	S1A-B17		This study	SAL_EA4460AA	Yes
STM 10-51		3014	1	20	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4331AA	Yes
STM 105-51		19	1	05	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4520AA	Yes
STM 105-51		19	1	20	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4564AA	Yes
STM 108-55		19	1	919p	IncFB(S), IncF(S), IncI2 (98.1%), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4528AA	Yes
STM 11-53		19	1	117	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA439AA	Yes
STM 11-57		19	1	875p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA8189AA	Yes
STM 110-50		19	1	895p	IncF(S), IncFV(S)	S1A-B17		This study	SAL_EA8195AA	Yes
STM 111-61		19	1	738p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4812AA	Yes
STM 111-65		19	1	90	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4472AA	Yes
STM 113-65		19	1	323	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4385AA	Yes
STM 114-60		Unknown	1		Unknown			This study		No
STM 114-62		19	1	05	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4361AA	Yes
STM 115-62		19	1	76	IncFB(S), IncF(S), IncI1	S1A-B17		This study	SAL_EA4562AA	Yes
STM 116-61		19	1	203	Unknown	S1A-B17, FZA-B-		This study		No
STM 117-60		Unknown	1		Unknown			This study		No
STM 117-63		19	1	117	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4463AA	Yes
STM 12-61		19	1	140	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA629AA	Yes
STM 12-66		19	1	15	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4529AA	Yes
STM 12-67		Unknown	1		None			This study		No
STM 124-60		19	1	15	Unknown			This study	SAL_EA4578AA	Yes
STM 123-62		Unknown	1	185	Unknown			This study		No
STM 124-64		19	1	927p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_LA0704AA	Yes
STM 125-64		Unknown	1	140	Unknown			This study		No
STM 126-67		19	1	15	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4560AA	Yes
STM 127-65		19	1	461p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4373AA	Yes
STM 127-67		19	1	46	IncI1		Inc11 andA_3_pil_3_rep1_4_sogs_6_tbrA_3_	This study	SAL_LA0577AA	Yes
STM 13-67		19	1	854p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4470AA	Yes
STM 131-67		19	1	833p	IncI1, IncF(S)	S1A-B17		This study	SAL_EA4381AA	Yes
STM 132-64		19	1	91	IncFB(S) (99.8%), IncF(S)	S1A-B35		This study	SAL_EA4384AA	Yes
STM 133-65		19	1	917p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4325AA	Yes
STM 134-61		19	1	907p	IncFB(S), IncF(S), IncI1	S1A-B17		This study	SAL_EA4394AA	Yes
STM 134-66		2111	243		None		Inc11 andA_2_pil_2_rep1_1_sogs_10_tbrA_4	This study	SAL_EA4411AA	Yes
STM 136-63		19	1	915p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4523AA	Yes
STM 139-65		19	1	877p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4577AA	Yes
STM 139-66		1654	1	822p	IncF(S) (97.7%), IncI1 (99.3%)	S3A-B-		This study	SAL_EA4314AA	Yes
STM 138-62		3013	1	867p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4308AA	Yes
STM 14-56		19	1	117	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA8431AA	Yes
STM 14-57		19	1	15	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA8438AA	Yes
STM 14-59		Unknown	1	203	Unknown			This study		No
STM 145-66		19	1	76	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4503AA	Yes
STM 146-67		Unknown	1		Unknown			This study		No
STM 147-63		19	1	54	IncFB(S), IncF(S), IncX1 (98.9%)	S1A-B17		This study	SAL_EA4420AA	Yes
STM 15-54		19	1	890p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA8428AA	Yes
STM 15-57		19	1	877p	IncFB(S), IncF(S), IncI1 (98.3%)	S1A-B17		This study	SAL_EA8190AA	Yes
STM 15-59		19	1	203	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA8422AA	Yes
STM 15-63		19	1	918p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4471AA	Yes
STM 151-51		19	1	752p	IncF(S) (97.7%), IncI2 (97.8%)	S1A-B17		This study	SAL_EA849AA	Yes
STM 151-54		19	1	923p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA456AA	Yes
STM 154-61		19	1	281	IncFB(S), IncF(S), IncX1			This study	SAL_EA4531AA	Yes
STM 156-61		Unknown	1		Unknown			This study		No
STM 156-63		19	1	920p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4509AA	Yes
STM 156-65		19	1	933p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA452AA	Yes
STM 157-65		19	1	415	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA458AA	Yes
STM 162-85		3021	1	149	None			This study	SAL_EA4497AA	Yes
STM 163-64		19	1	885p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4473AA	Yes
STM 166-65		3021	1		None			This study		No
STM 166-63		19	1	140	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4509AA	Yes
STM 17-65		19	1	939p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4479AA	Yes
STM 17-66		19	1	117	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA430AA	Yes
STM 17-69		19	1	235	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA435AA	Yes
STM 170-55		19	1	650p	IncFB(S), IncF(S), IncI1 (99.3%)	S1A-B17		This study	SAL_EA4579AA	Yes
STM 176-65		19	1	323	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA4578AA	Yes
STM 179-66		Unknown	1		Unknown			This study		No
STM 18-55		19	1	216	IncFB(S), IncF(S), IncFII, IncI1	S1A-B35, FZA-B-	Inc11 andA_2_pil_1_rep1_1_sogs_3_tbrA_21	This study	SAL_EA8184AA	Yes
STM 18-59		19	1	912p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA437AA	Yes
STM 18-69		19	1	841p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA440AA	Yes
STM 181-63		19	1	140	None			This study	SAL_EA4301AA	Yes
STM 181-65		19	1	59	IncFB(S), IncF(S), IncX1 (98.9%)	S1A-B17		This study	SAL_EA454AA	Yes
STM 185-63		19	1	401	IncFB(S), IncF(S), ColI56 (96%)	S1A-B17		This study	SAL_EA429AA	Yes
STM 19-58		19	1	315	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA840AA	Yes
STM 19-59		302	1	866p	IncFB(S), IncF(S)	S1A-B17		This study	SAL_EA843AA	Yes

Isolate name	Alternative name	Continent	Country	Host	Host details	Year	Species	Recorded serotype	In silico Serotyping	Antimicrobial susceptibility testing	Resistance gene	MIC AMP (mg/L)	MIC Peni G (mg/L)
STM 19-62		Europe	Portugal	Food	Ovine	1962	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 191-63		Africa	Morocco	Animal	Cow	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 192-67		Africa	Morocco	Human	Unknown	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 2-58		Africa	Morocco	Human	Guinea pig	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 20-67	TM 67-20	Africa	Morocco	Human	Stool	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	8
STM 201-65		Europe	France	Animal	Horse	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	3
STM 202-67		Europe	France	Animal	Chicken	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	1	12
STM 209-64		Europe	France	Animal	Chicken	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.38	4
STM 209-67		Europe	France	Food	Sausage	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 210-65	TM 67-21	Africa	Morocco	Human	Stool	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 210-65		Europe	France	Human	Stool	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 211-65		Africa	Democratic Republic of the Congo	Human	Parrot	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 212-65		Africa	Cameroun	Human	Absecess	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 214-65		Asia	Vietnam	Human	Guinea pig blood	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 215-65		Africa	Turkey	Animal	Cerebrospinal Fluid	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 216-65		Africa	Chad	Animal	Chicken	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	6
STM 219-65		Europe	Portugal	Human	Unknown	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	6
STM 219-65		Asia	Iran	Human	Unknown	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 222-63		Africa	Morocco	Unknown	Unknown	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 225-65		Europe	Portugal	Animal	Pig	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 23-63		Asia	Vietnam	Unknown	Unknown	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.5	4
STM 23-65		Europe	France	Human	Stool	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 23-67	TM 67-23	Africa	Madagascar	Human	Blood	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.125	0.125
STM 232-68		Europe	France	Human	Unknown	1968	Salmonella enterica	Typhimurium	B1:1,2	ApStiSulTmPC	Unknown	>256	>256
STM 24-67	TM 67-24	Africa	Madagascar	Animal	Unknown	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 25-57		Europe	France	Human	Stool	1957	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 256-68		Europe	France	Human	Stool	1968	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 26-62		Africa	Madagascar	Human	Stool	1962	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 26-65		South America	French Guiana	Animal	Rat	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 264-64		Africa	Morocco	Human	Unknown	1964	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	8
STM 265-68		Europe	France	Human	Stool	1968	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	>256	6
STM 266-68		Africa	Tunisia	Human	Unknown	1968	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	1	
STM 268-64		Africa	Senegal	Human	Unknown	1964	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 270-64		Europe	France	Food	Horse meat	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 271-66		Europe	Turkey	Human	Unknown	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 275-64		Europe	Turkey	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	6
STM 28-56		Asia	France	Human	Stool	1956	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 28-57		Europe	Turkey	Human	Unknown	1957	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 29-53		Europe	France	Human	Unknown	1953	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	6
STM 29-55		Europe	France	Animal	Duck	1953	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 29-56		Europe	France	Animal	Pus	1956	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 296-64		Asia	Sub Saharan Africa	Human	Unknown	1964	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 3-58		Africa	India	Human	Unknown	1964	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 3-59		Europe	Morocco	Human	Pus	1958	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 4-59		Europe	France	Animal	Cattle	1959	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 3-63		Europe	France	Animal	Bird	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 30-55		Europe	Belgium	Human	Unknown	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 30-55		Africa	Senegal	Human	Unknown	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.5	4
STM 30-59		Europe	France	Human	Stool	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 303-67		Europe	France	Human	Stool	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 304-67		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 305-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 305-66		Africa	Algeria	Animal	Dog	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 307-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 31-66		Europe	France	Human	Cow	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 31-67		Europe	Italy	Human	Unknown	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	6
STM 31-68		Africa	Democratic Republic of the Congo	Animal	Monkey Blood	1958	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 310-66		South America	Argentina	Animal	Hare	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 32-56		Asia	Vietnam	Human	Unknown	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 32-58		Europe	France	Animal	Calf	1958	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.38	2
STM 33-59		Europe	Italy	Animal	Cow	1958	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.5	2
STM 331-66		South America	Brazil	Human	Blood	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 332-66		South America	Brazil	Animal	Guinea pig	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 333-66		South America	Brazil	Animal	Rabbit	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 34-61		Africa	Morocco	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 345-66		Europe	France	Animal	Small mammal	1962	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	>256	>256
STM 346-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 347-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 348-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 351-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 352-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 352-68		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 36-57		Europe	Belgium	Human	Unknown	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 36-61		Africa	Morocco	Animal	Small mammal	1961	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 36-63		Europe	Belgium	Human	Stool	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 365-67		Africa	Senegal	Human	Blood	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	>256	>256
STM 366-66		Europe	France	Human	Stool	1966	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 366-67		Africa	Senegal	Human	Blood	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 367-67		Africa	Senegal	Human	Blood	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 368-67		Africa	Senegal	Human	Blood	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 368-67		Africa	Senegal	Human	Stool	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.25	1.5
STM 37-57		Europe	Belgium	Animal	Monkey (from Hamburg)	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	1.5	16
STM 37-61		Africa	Morocco	Human	Stool	1961	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 370-67		Europe	France	Animal	Small mammal	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 375-67		Asia	India	Food	Frog legs	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 38-56		Africa	Tunisia	Human	Cerebrospinal Fluid	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 38-57		Europe	France	Human	Stool	1967	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 38-59		Africa	Democratic Republic of the Congo	Human	Stool	1959	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 39-59		Europe	France	Human	Stool	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 39-68		Europe	France	Human	Stool	1963	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 4-57		Europe	France	Human	Unknown	1965	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 4-57		Africa	Algeria	Animal	Sheep	1957	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	6
STM 40-62		Africa	Senegal	Human	Blood	1962	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 406-67		South America	Peru	Human	Unknown	1961	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 41-61		Europe	Belgium	Human	Stool	1961	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		
STM 42-58		Europe	France	Human	Stool	1958	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown	0.75	6
STM 43-58		Europe	France	Animal	Dog	1958	Salmonella enterica	Typhimurium	B1:1,2	Pansusceptible	Unknown		

Isolate name	Phage type	Sequence Type	MLST eBURST group	CRISPR type	Plasmid	pMLST IncF	pMLST Other Inc	Reference	Enterobase UBERstrain number	WGS
STM 19-52	19	19	1	53	IncF(S), IncF(II), IncF(S), IncF(II), IncI1	ST1A-B17	IncI1 ST-23	This study	SAL_EA40695AA	Yes
STM 191-63	19	19	1	26	Unknown	ST1A-B17		This study	SAL_EA4484AA	Yes
STM 192-67	Unknown	Unknown	1	46	IncF(S), IncF(II)	ST1A-B17		This study		No
STM 2-58	19	19	1	185	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4375AA	Yes
STM 20-87	19	19	1	21	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4324AA	Yes
STM 201-65	19	19	1	48	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4424AA	Yes
STM 202-67	19	19	1	15	IncF(S), IncF(II), IncX1 (98.9%)	ST1A-B17		This study	SAL_EA4530AA	Yes
STM 209-64	19	19	1	117	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA431AA	Yes
STM 209-67	Unknown	Unknown	1	46	Unknown			This study		No
STM 210-66	Unknown	Unknown	1	46	Unknown			This study		No
STM 211-65	Unknown	Unknown	1	17	Unknown			This study	SAL_EA415AA	Yes
STM 212-65	19	19	1	910prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA456AA	Yes
STM 214-65	19	19	1	925prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA457AA	Yes
STM 215-65	313	313	1	696	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA0579AA	Yes
STM 216-65	19	19	1	842prov	IncF(S), IncF(II), Col156 (98%)	ST1A-B17		This study	SAL_EA452AA	Yes
STM 219-64	19	19	1	913prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA433AA	Yes
STM 219-65	19	19	1	20	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4423AA	Yes
STM 222-63	19	19	1	185	IncF(S), IncF(II), IncI1 (99.5%)	ST1A-B17	IncI1 ST-7	This study	SAL_EA401AA	Yes
STM 225-65	19	19	1	931prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4530AA	Yes
STM 22-63	19	19	1	914prov	IncF(S), IncF(II), Col(BGS12)	ST1A-B17		This study	SAL_EA6451AA	Yes
STM 22-65	19	19	1	46	IncF1	FZA-B-		This study	SAL_EA4380AA	Yes
STM 22-67	19	19	1	59	IncX1 (98.9%)			This study	SAL_EA432AA	Yes
STM 232-68	313	313	1	810	IncF(S), IncF(II), IncQ1	ST1A-B17		This study	SAL_EA417AA	Yes
STM 24-67	Unknown	Unknown	1	320	IncF(S), IncF(II)	ST1A-B17		This study		No
STM 25-57	19	19	1	91	IncF(S), IncF(II)	ST1A-B17		This study		No
STM 256-68	Unknown	Unknown	1	770prov	Unknown			This study		No
STM 26-65	Unknown	Unknown	1	738prov	Unknown			This study		No
STM 264-64	19	19	1	26	IncF(S), IncF(II), IncI1	ST1A-B17	IncI1 arda_10.pill_1, rep1_1, sogS_6, tba_8 (99.8%)	This study	SAL_EA4300AA	Yes
STM 265-68	Unknown	Unknown	1	90	Unknown			This study		No
STM 268-68	313	313	1	810	IncF(S), IncF(II), IncQ1	ST1A-B17		This study	SAL_EA439AA	Yes
STM 269-64	19	19	1	506prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4296AA	Yes
STM 27-53	302	302	1	68	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA435AA	Yes
STM 27-54	302	302	1	96	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA0432AA	Yes
STM 28-56	302	302	1	696	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA418AA	Yes
STM 28-57	19	19	1	203	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA432AA	Yes
STM 29-53	19	19	1	874prov	IncF(S), IncF(II), IncI1 (99.3%), IncI2 (98.1%)	ST1A-B17	IncI1 arda_32.pill_8, rep1_2, sogS_4, tba_8	This study	SAL_EA8425AA	Yes
STM 29-55	19	19	1	884prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA8427AA	Yes
STM 29-56	1805	1805	243	895prov	None			This study	SAL_EA0580AA	Yes
STM 296-64	19	19	1	313	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4527AA	Yes
STM 3-58	19	19	1	185	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA457AA	Yes
STM 3-59	19	19	1	880prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA8441AA	Yes
STM 3-63	19	19	1	932prov	None	ST1A-B35		This study	SAL_EA4500AA	Yes
STM 30-55	19	19	1	76	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4580AA	Yes
STM 30-59	Unknown	Unknown	1	415	Unknown			This study	SAL_EA8428AA	Yes
STM 303-67	Unknown	Unknown	1	90	Unknown			This study		No
STM 304-67	Unknown	Unknown	1	16	Unknown			This study		No
STM 306-66	Unknown	Unknown	1	90	Unknown			This study		No
STM 309-66	Unknown	Unknown	1	323	Unknown			This study		No
STM 31-56	Unknown	Unknown	1	46	IncF(S), IncF(II)	ST1A-B17		This study	SAL_LA0697AA	Yes
STM 31-57	Unknown	Unknown	1	140	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA8186AA	Yes
STM 31-58	19	19	1	320	Unknown			This study		No
STM 310-66	19	19	1	892prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_LA0578AA	Yes
STM 32-56	1649	1649	1	314	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4379AA	Yes
STM 33-58	19	19	1	882prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_LA0696AA	Yes
STM 33-59	19	19	1	911prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4469AA	Yes
STM 34-56	Unknown	Unknown	1	303	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4380AA	Yes
STM 32-66	19	19	1	843prov	Unknown			This study	SAL_EA4466AA	Yes
STM 33-66	19	19	1	203	IncF(S), IncF(II), IncI1	ST1A-B17	IncI1 arda_2.pill_2, rep1_2, sogS_4, tba_5	This study	SAL_EA4486AA	Yes
STM 34-61	Unknown	Unknown	1	90	Unknown			This study	SAL_EA4489AA	No
STM 34-66	19	19	1	46	IncN, IncI1	ST1A-B17	IncI1 ST12; IncI1 arda_3.pill_3 (99.61%), rep1_4, sogS_6, tba_3	This study	SAL_EA4540AA	Yes
STM 346-66	Unknown	Unknown	1	46	Unknown			This study		No
STM 347-66	Unknown	Unknown	1	90	Unknown			This study		No
STM 349-66	19	19	1	844prov	IncF(II)	ST1A-B17		This study	SAL_LA0699AA	Yes
STM 35-57	19	19	1	140	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4477AA	Yes
STM 351-66	Unknown	Unknown	1	847prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_LA0700AA	Yes
STM 352-66	19	19	1	847prov	IncF(S), IncF(II), ColE10	ST1A-B17		This study	SAL_EA4533AA	Yes
STM 354-68	19	19	1	840	None			This study	SAL_EA4478AA	Yes
STM 36-57	19	19	1	91	None			This study	SAL_EA4478AA	No
STM 36-61	Unknown	Unknown	1	203	Unknown			This study	SAL_EA4542AA	Yes
STM 36-63	19	19	1	117	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4542AA	Yes
STM 385-67	Unknown	Unknown	1	461prov	IncF1, Col156 (96.1%), IncI1 (99.3%)			This study	SAL_EA4551AA	Yes
STM 386-66	19	19	1	46	IncF(S), IncF(II)	FZA-B-		This study	SAL_EA4551AA	Yes
STM 386-67	19	19	1	461prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4297AA	Yes
STM 387-67	Unknown	Unknown	1	461prov	Unknown			This study		No
STM 388-67	Unknown	Unknown	1	461prov	Unknown			This study		No
STM 389-67	-87	-87	1	828prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA437AA	Yes
STM 37-57	19	19	1	91	IncI2 (97.8%)			This study	SAL_EA4476AA	Yes
STM 37-61	Unknown	Unknown	1	203	Unknown			This study	SAL_EA437AA	Yes
STM 370-67	19	19	1	91	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4476AA	Yes
STM 379-67	36	36	138	859prov	None			This study	SAL_EA4395AA	Yes
STM 38-59	19	19	1	860prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4320AA	Yes
STM 38-59	19	19	1	860prov	IncF(S) (89.8%), IncF(II), IncI1 (87.9%)	ST1A-B35	IncI1 arda_14.pill_8, rep1_3, sogS_3, tba_6	This study	SAL_EA0433AA	Yes
STM 38-59	19	19	1	893prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA439AA	Yes
STM 38-63	19	19	1	216	IncF(S) (89.8%), IncF(II)	ST1A-B35		This study	SAL_LA0694AA	Yes
STM 39-66	Unknown	Unknown	1	16	Unknown			This study		No
STM 4-57	19	19	1	869prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4355AA	Yes
STM 4-57	19	19	1	13	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA8181AA	Yes
STM 40-62	19	19	1	283	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA8186AA	Yes
STM 406-67	19	19	1	850prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4322AA	Yes
STM 41-61	19	19	1	15	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4400AA	Yes
STM 42-58	19	19	1	868prov	IncF(S), IncF(II), IncP (99.4%)	ST1A-B17		This study	SAL_EA4437AA	Yes
STM 43-58	19	19	1	885prov	IncF(S), IncF(II)	ST1A-B17		This study	SAL_EA4436AA	Yes

Isolate name	Alternative name	Continent	Country	Host	Host details	Year	Species	Recorded serotype	In silico Serotyping	Antimicrobial susceptibility testing	Resistance gene	MIC AMP (mg/L)	MIC Peni G (mg/L)
STM 437-67		Europe	France	Human	Stool	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 444-59		Africa	Democratic Republic of the Congo	Animal	Pig	1959	<i>Salmonella enterica</i>	Typhimurium		Pansusceptible	Unknown		
STM 440-67		Europe	France	Human	Stool	1967	<i>Salmonella enterica</i>	Typhimurium		Pansusceptible	Unknown		
STM 441-67		Europe	France	Human	Stool	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 451-59		Europe	France	Animal	Duck	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 451-59		Europe	France	Human	Stool	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 46-52		Africa	Democratic Republic of the Congo	Human	Stool	1956	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible		1.5	8
STM 47-56		Africa	Democratic Republic of the Congo	Human	Stool	1956	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 48-60		Africa	Cameroon	Human	Blood	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 51-57		Europe	France	Human	Stool	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 51-57		Africa	Algeria	Animal	Rabbit	1957	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 51-58		Europe	France	Animal	Rabbit	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 51-58		Europe	France	Animal	Horse meat	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 51-58		Europe	France	Animal	Duckling	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 50-56		Africa	Senegal	Human	Pus	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 50-58		Europe	France	Food	Duck meat	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 50-60		Europe	France	Animal	Mouse	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown	1	12
STM 51-58		Europe	France	Human	Stool	1968	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 51-59		Africa	Madagascar	Food	Duck meat	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 52-56		Asia	Vietnam	Food	Mayonnaise	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 53-59		Africa	Madagascar	Human	Stool	1956	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 53-65		Africa	Tunisia	Human	Urine	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 54-58		Europe	France	Human	Unknown	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 54-60		Africa	Senegal	Animal	Rabbit	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 54-65		Africa	Tunisia	Human	Stool	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 55-62		Europe	France	Unknown	Unknown	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 57-60		Europe	France	Unknown	Unknown	1962	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 57-60		Europe	France	Animal	Unknown	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 59-53		Europe	Italy	Animal	Colt	1963	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 60-60		Africa	Democratic Republic of the Congo	Human	Stool	1963	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 60-60		Africa	Democratic Republic of the Congo	Animal	Pig	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 61-58		Europe	France	Animal	Dog	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 61-58		Africa	Algeria	Animal	Stool	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 61-62		Europe	France	Animal	Dog	1962	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 619-68		South America	Brazil	Animal	Pig	1968	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 65-60		Africa	Tunisia	Human	Unknown	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 65-62		Europe	Belgium	Human	Stool	1962	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 66-61		Africa	Morocco	Unknown	Unknown	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	blaTEM-1B	0.125 ≥256	0.25 ≥256
STM 68-61		Africa	Morocco	Unknown	Unknown	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 69-67		Europe	Portugal	Feed	Fishmeal	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 71-51		Europe	France	Animal	Pig	1951	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible		0.75	4
STM 71-58		Africa	France	Animal	Pig stool	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 70-58		Europe	France	Animal	Guinea pig	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 71-62		Europe	France	Human	Blood	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 71-62		Africa	Morocco	Animal	Stool	1962	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 72-66		Europe	France	Animal	Turkey intestine	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 72-66		Europe	France	Animal	Canary	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 73-61		Africa	Morocco	Food	Unknown	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible		0.75	8
STM 73-62		Europe	Belgium	Animal	Hedgehog Stool	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 73-67		Africa	Chad	Human	Stool	1962	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 74-63		Africa	Morocco	Animal	Mouse	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 74-67		Europe	France	Unknown	Unknown	1963	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 75-61		Africa	Morocco	Human	Stool	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 75-67		Europe	France	Animal	Hedgehog Stool	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 76-61		Africa	Morocco	Animal	Hedgehog Stool	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 77-58		Europe	Belgium	Human	Stool	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 79-65		Africa	Democratic Republic of the Congo	Animal	Pig	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 79-59		Africa	Democratic Republic of the Congo	Animal	Duckling	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 79-62		Europe	France	Human	Unknown	1962	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible		0.75	4
STM 8-57		Europe	France	Human	Stool	1957	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 8-62		Europe	France	Animal	Pigeon	1962	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 8-63		Europe	France	Animal	Pigeon	1963	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 80-61		Europe	Portugal	Animal	Carn	1963	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 82-61		Europe	Portugal	Food	Meat	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 84-67		Europe	France	Food	Fish meal	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 85-65		Europe	France	Human	Stool	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 85-65		South America	Venezuela	Human	Stool	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 85-67		Europe	France	Animal	Pig	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 87-61		Africa	Madagascar	Human	Stool	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 87-67		Europe	Italy	Animal	Guinea pig blood	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 88-52		Europe	Portugal	Animal	Chick	1952	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 88-65		Africa	Cameroon	Human	Unknown	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown	0.25	3
STM 89-58		Europe	France	Animal	Rabbit	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 9-63		Asia	Vietnam	Human	Stool	1963	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 90-58		Europe	France	Animal	Cat	1958	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 92-59		Europe	Belgium	Human	Stool	1959	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 92-60		Europe	France	Human	Stool	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 92-61		Europe	France	Animal	Bird in cage	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 92-63		Europe	France	Animal	Pig	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 92-65		Europe	France	Animal	Canary	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 95-67		Europe	France	Human	Stool	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 95-67		Europe	France	Human	Dog stool	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 95b-54		Africa	Tunisia	Human	Stool	1964	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 96-65		Europe	France	Animal	Dog stool	1965	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible			
STM 96-67		Europe	France	Animal	Stool	1967	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 97-61		Asia	Vietnam	Animal	Gecko viscera	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 98-60		Europe	Belgium	Human	Unknown	1960	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
STM 99-61		Asia	Vietnam	Unknown	Unknown	1961	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
TM 69-186		Africa	Cameroon	Unknown	Unknown	1969	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
TM 69-488		Africa	Cameroon	Animal	Mouse	1969	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		
TM 69-692		Africa	Tunisia	Animal	Dog	1969	<i>Salmonella enterica</i>	Typhimurium	B:1,1,2	Pansusceptible	Unknown		



Isolate name	Phage type	Sequence Type	MLST eBurst group	CRISPOL type	Plasmid	pMLST IncF	pMLST Other Inc	Reference	Enterobase	Uberstrain number	WGS
STM 437-67	Unknown	19	1	889prov	None			This study	SAL_EA4543AA		No
STM 445-59	Unknown	90	1	Unknown	None			This study			No
STM 446-67	Unknown	90	1	Unknown	None			This study			No
STM 45-69	Unknown	920prov	1	920prov	None			This study			No
STM 46-52	Unknown	738prov	1	738prov	None			This study			No
STM 46-56	Unknown	878prov	1	878prov	None			This study			No
STM 47-56	Unknown	98	1	98	None			This study			No
STM 47-59	Unknown	88	1	878prov	None			This study			No
STM 48-60	Unknown	19	1	871prov	None			This study			No
STM 493-67	Unknown	19	1	117	None			This study			No
STM 5-57	Unknown	19	1	872prov	None			This study			No
STM 5-58	Unknown	19	1	541	None			This study			No
STM 5-59	Unknown	19	1	260	None			This study			No
STM 50-56	Unknown	19	1	415	None			This study			No
STM 50-58	Unknown	19	1	867prov	None			This study			No
STM 50-60	Unknown	19	1	738prov	None			This study			No
STM 508-68	Unknown	19	1	90	None			This study			No
STM 51-58	Unknown	19	1	867prov	None			This study			No
STM 51-59	Unknown	19	1	879prov	None			This study			No
STM 52-56	Unknown	19	1	864prov	None			This study			No
STM 53-58	Unknown	19	1	970prov	None			This study			No
STM 54-58	Unknown	19	1	738prov	None			This study			No
STM 54-60	Unknown	19	1	415	None			This study			No
STM 54-65	Unknown	19	1	117	None			This study			No
STM 55-62	Unknown	19	1	850prov	None			This study			No
STM 55-62	Unknown	19	1	868prov	None			This study			No
STM 57-56	Unknown	19	1	157	None			This study			No
STM 57-60	Unknown	19	1	15	None			This study			No
STM 57-63	Unknown	19	1	827prov	None			This study			No
STM 58-58	Unknown	19	1	76	None			This study			No
STM 6-67	Unknown	19	1	827prov	None			This study			No
STM 60-59	Unknown	19	1	20	None			This study			No
STM 60-60	Unknown	19	1	314	None			This study			No
STM 61-58	Unknown	19	1	907prov	None			This study			No
STM 61-59	Unknown	19	1	203	None			This study			No
STM 61-62	Unknown	128	1	54	None			This study			No
STM 619-68	Unknown	19	1	896prov	None			This study			No
STM 65-60	Unknown	19	1	203	None			This study			No
STM 65-62	Unknown	19	1	203	None			This study			No
STM 66-61	Unknown	203	1	203	None			This study			No
STM 68-61	Unknown	213	1	01	None			This study			No
STM 69-67	Unknown	19	1	320	None			This study			No
STM 7-51	Unknown	19	1	20	None			This study			No
STM 7-58	Unknown	19	1	827prov	None			This study			No
STM 7-67	Unknown	19	1	203	None			This study			No
STM 70-58	Unknown	19	1	46	None			This study			No
STM 70-67	Unknown	19	1	881prov	None			This study			No
STM 71-62	Unknown	19	1	865prov	None			This study			No
STM 72-56	Unknown	19	1	905prov	None			This study			No
STM 72-62	Unknown	19	1	928prov	None			This study			No
STM 72-85	Unknown	19	1	41	None			This study			No
STM 73-61	Unknown	19	1	90	None			This study			No
STM 73-62	Unknown	19	1	842prov	None			This study			No
STM 73-67	Unknown	19	1	203	None			This study			No
STM 74-63	Unknown	3015	1	832prov	None			This study			No
STM 74-67	Unknown	19	1	889prov	None			This study			No
STM 75-61	Unknown	19	1	323	None			This study			No
STM 75-67	Unknown	19	1	4	None			This study			No
STM 76-61	Unknown	19	1	323	None			This study			No
STM 77-58	Unknown	19	1	916prov	None			This study			No
STM 78-65	Unknown	19	1	873prov	None			This study			No
STM 79-59	Unknown	19	1	20	None			This study			No
STM 79-62	Unknown	19	1	930prov	None			This study			No
STM 8-57	Unknown	19	1	899prov	None			This study			No
STM 8-62	Unknown	19	1	897prov	None			This study			No
STM 8-63	Unknown	19	1	827prov	None			This study			No
STM 8-67	Unknown	19	1	46	None			This study			No
STM 80-61	Unknown	19	1	01	None			This study			No
STM 80-65	Unknown	213	1	203	None			This study			No
STM 82-61	Unknown	19	1	46	None			This study			No
STM 84-67	Unknown	1649	1	921prov	None			This study			No
STM 85-65	Unknown	19	1	323	None			This study			No
STM 85-67	Unknown	19	1	Unknown	None			This study			No
STM 87-61	Unknown	19	1	870prov	None			This study			No
STM 87-67	Unknown	19	1	830prov	None			This study			No
STM 88-52	Unknown	19	1	142	None			This study			No
STM 88-56	Unknown	19	1	863prov	None			This study			No
STM 89-58	Unknown	19	1	2	None			This study			No
STM 90-58	Unknown	19	1	140	None			This study			No
STM 92-59	Unknown	19	1	323	None			This study			No
STM 92-60	Unknown	19	1	117	None			This study			No
STM 92-61	Unknown	19	1	124	None			This study			No
STM 92-63	Unknown	19	1	890prov	None			This study			No
STM 92-65	Unknown	19	1	17	None			This study			No
STM 94-67	Unknown	19	1	90	None			This study			No
STM 95-65	Unknown	19	1	117	None			This study			No
STM 95-67	Unknown	19	1	46	None			This study			No
STM 95b-64	Unknown	19	1	738prov	None			This study			No
STM 96-65	Unknown	19	1	117	None			This study			No
STM 96-67	Unknown	19	1	Unknown	None			This study			No
STM 97-61	Unknown	19	1	46	None			This study			No
STM 98-60	Unknown	19	1	05	None			This study			No
STM 99-61	Unknown	19	1	15	None			This study			No
STM 99-66	Unknown	19	1	05	None			This study			No
TM 69-168	Unknown	325	1	825prov	None			This study			No
TM 69-468	Unknown	19	1	825prov	None			This study			No
TM 69-692	Unknown	19	1	15	None			This study			No

DT2

**Title: Early transmissible ampicillin resistance in zoonotic *Salmonella enterica* serotype Typhimurium in the late 1950s: a retrospective whole-genome sequencing study**

**Authors :** Alicia Tran Dien, M.Sc<sup>1</sup>, Simon Le Hello, Pharm.D<sup>1</sup>, Christiane Bouchier, Ph.D<sup>2</sup>, and François-Xavier Weill, M.D <sup>1‡</sup>

<sup>1</sup> Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Centre National de Référence des *Escherichia coli*, *Shigella* et *Salmonella*, Paris 75015, France

<sup>2</sup> Institut Pasteur, Plate-forme Génomique (PF1), Paris, 75015, France

‡ Corresponding author

François-Xavier Weill: Centre National de Référence des *Escherichia coli*, *Shigella* et *Salmonella*, Unité des Bactéries Pathogènes Entériques, Institut Pasteur, 28 rue du docteur Roux, 75724 Paris cedex 15. Tel: +33-1 45 68 83 45, Fax: +33-1 45 68 88 37. E-mail: francois-xavier.weill@pasteur.fr

Running title: Early ampicillin resistance in *Salmonella* Typhimurium

Keywords: *Salmonella* Typhimurium, multidrug resistance, ampicillin, penicillins,  $\beta$ -lactamase, antimicrobial resistance, emergence

Summary (251 words); main text (3590 words); four figures, two tables, one box (248 words), 30 references. Appendix with four pages.

## SUMMARY

### Background

Ampicillin, the first semisynthetic penicillin active against Enterobacteriaceae, was released onto the market in 1961. The first outbreaks of disease caused by ampicillin-resistant *Salmonella enterica* serotype Typhimurium strains were identified in the United Kingdom in 1962 and 1964.

### Methods

We studied a collection of 288 historical isolates of *S. enterica* serotype Typhimurium collected between 1911 and 1969 from 31 countries on four continents. Antimicrobial drug susceptibility was tested for all isolates and whole-genome sequencing was carried out on a subset of 225 isolates. The mechanisms of resistance to ampicillin were identified.

### Findings

Eleven isolates (3.8%) from three phylogenomic groups were resistant to ampicillin due to the acquisition of various  $\beta$ -lactamase genes carried by different plasmids, including the

virulence plasmid of *S. enterica* serotype Typhimurium. Three isolates producing TEM-1 were isolated in France ( $n = 1$ ) and Tunisia ( $n = 2$ ), in 1959 and 1960, respectively, before ampicillin went on sale. The vectors for ampicillin resistance were different from those reported in the strains responsible for the outbreaks in the United Kingdom in the 1960s.

## **Interpretation**

The relationship between antibiotic use and the selection of plasmids encoding resistance determinants is not as direct as often presumed. The presence of transmissible resistance to ampicillin in a prominent zoonotic bacterium before the release of ampicillin for clinical use is consistent with a role for the narrow-spectrum penicillins (e.g., benzylpenicillin) used as growth promoters in livestock since the 1950s. These findings highlight the need for a “one-health” approach to tackling the issue of antibiotic resistance.

## **Funding**

Institut Pasteur, Santé publique France, the French government’s Investissement d’Avenir programme, the Fondation Le Roch-Les Mousquetaires.

## **Panel: Research in context**

### **Evidence before this study**

We searched PubMed for articles published between January 1, 1900 and December 31, 1965, with the following search terms: “Salmonella and ampicillin and resistance”, “Salmonella and penicillin and resistance”, “Salmonella and penicillinase”, and “Salmonella and  $\beta$ -lactamase”. We also searched for articles published between January 1, 1900 and August 31, 2017, with the following search terms “Salmonella and TEM”, and “Salmonella and OXA”. No language restrictions were used. We found no reports describing transmissible ampicillin resistance or  $\beta$ -lactamase in *Salmonella* spp. isolates collected before 1962.

### **Added value of this study**

We identified TEM-1  $\beta$ -lactamase-producing *S. enterica* serotype Typhimurium isolates, dating from 1959-1960, several years before the release of ampicillin onto the market. The genetic diversity of these ampicillin-resistant isolates, their resistance mechanisms and their geographic distribution suggest that ampicillin resistance had already diffused in this zoonotic pathogen at that time. We also found relative susceptibility of wild-type *S. enterica* serotype Typhimurium isolates to penicillin G (8 mg/L), a narrow-spectrum penicillin widely used as a feed additive (subtherapeutic doses of antimicrobial agents over long periods of time) in livestock from the 1950s to the end of the 1960s.

### **Implications of all the available evidence**

Our data shed light on the origin of resistance to a major antimicrobial agent in a prominent zoonotic bacterium. They show that environments, in which penicillin G residues are abundant, such as farming environments in the late 1950s, can foster transmissible ampicillin resistance in *S. enterica* serotype Typhimurium.

## INTRODUCTION

In May 2015, the 68th World Health Assembly endorsed a global action plan to tackle antimicrobial resistance,<sup>1</sup> a global public health issue for the 21st century. Eighty years ago, the discovery and commercialisation of “sulfa drugs” (sulfonamidochrysoidine and sulfanilamide) opened up a new era in the fight against bacterial infections, that of antibiotics.<sup>2</sup> These synthetic bacteriostatic antibiotics were hailed as “magic bullets” or “wonder drugs”, but were rapidly eclipsed by penicillin, a natural bactericidal substance from *Penicillium rubrum* that had been discovered earlier (1928) but did not go into mass production until World War 2 (WW2).<sup>2</sup> In his princeps paper, Alexander Fleming noted that, unlike Gram-positive bacteria, the bacteria of the “coli-typhoid” group were not sensitive to penicillin.<sup>3</sup> The first penicillin drug active against enterobacteria (broad-spectrum penicillin) was ampicillin, a semisynthetic derivative of 6-aminopenicillanic acid,<sup>4</sup> commercialised by Beecham Research Laboratories under the name “Penbritin” in 1961.<sup>5</sup> The first report describing  $\beta$ -lactamases, encoded by “R factors” (plasmids), in Enterobacteriaceae strains with transmissible ampicillin-resistance, was published in 1965.<sup>6</sup> In this report, two of the three Enterobacteriaceae strains studied were found to produce the TEM penicillinase. The *E. coli* isolate from Greece obtained in 1963 (the strain and its  $\beta$ -lactamase were named TEM after the patient’s name, Temoneria)<sup>7</sup> and the *S. enterica* serotype Paratyphi B isolate obtained in the United Kingdom in 1963 were subsequently found to contain the *bla*<sub>TEM-1A</sub> gene carried by a ~ 40 kb IncX2 plasmid and a ~ 100 kb IncF plasmid, respectively (Table 1).<sup>8,9</sup> The third strain, a *S. enterica* serotype Typhimurium phage-type 1a (later reclassified as definitive phage-type [DT] 2), isolated in 1962 in Brighton, United Kingdom, produced a  $\beta$ -lactamase with a substrate profile different from that of TEM (Table 1). This *S. enterica* serotype Typhimurium strain was isolated during the first reported outbreak caused by an ampicillin-resistant *S. enterica* serotype Typhimurium strain. It occurred in the second half of 1962, in the United Kingdom, at a time at which dozens of *S. enterica* serotype Typhimurium DT2 isolates resistant to ampicillin, streptomycin, sulfonamides and tetracyclines were found in humans and pigs.<sup>10</sup> The plasmid of a representative strain, R<sub>1818</sub>, later renamed R46,<sup>9</sup> was sequenced (GenBank accession no. AY046276) and found to be a ~ 50 kb IncN plasmid carrying the *bla*<sub>OXA-2</sub> gene.

In 1963 and 1964, epidemic isolates of *Salmonella enterica* with transmissible ampicillin-resistance were identified several times in Europe (Table 1).<sup>11-14</sup> In particular, *S. enterica* serotype Typhimurium DT29 emerged following the adoption of intensive farming methods for the rearing of calves in the United Kingdom.<sup>12</sup> In 1965, 1,297 and 576 isolates of this pathogen were obtained from cattle and humans, respectively. More than 96% of these isolates were multidrug-resistant. NTP1, the ampicillin resistance plasmid of representative strain RT1,<sup>13</sup> was found to be a small (8.3 kb), multicopy, non-conjugative plasmid carrying the *bla*<sub>TEM-1</sub> gene.<sup>14</sup>

This timeline suggests that the use of ampicillin may have been the primary cause of the selection of resistant isolates of this major foodborne pathogen.

With a view to improving our understanding of the emergence of resistance to ampicillin, one of the most widely used antibiotics to date, we studied a large collection of historical isolates of *S. enterica* serotype Typhimurium, a zoonotic agent that has remained one of the two predominant serotypes causing human disease in Europe since WW2.<sup>15,16</sup> These 288

isolates were collected between 1911 and 1969, and 107 were obtained before ampicillin went on sale.

## **MATERIALS & METHODS**

### **Bacterial isolates**

The 288 *S. enterica* serotype Typhimurium isolates, analysed here were obtained from various sources (including humans,  $n = 149$ ; animals, food and feed,  $n = 113$ ) and geographic areas (including Europe,  $n = 161$ ; Africa,  $n = 90$ ; Asia,  $n = 19$ ; America,  $n = 15$ ), between 1911 and 1969 (table S1). They originated from the collections of the French National Reference Centre for *E. coli*, *Shigella* and *Salmonella* (FNRC-ESS) ( $n = 277$ ), the Collection de l' Institut Pasteur (CIP) ( $n = 9$ ), and the World Health Organization Collaborative Centre for Reference and Research on *Salmonella* ( $n = 2$ ), Institut Pasteur, Paris, France. All isolates from the FNRC-ESS were maintained as stab cultures at room temperature. They have not been manipulated since being placed in storage (generally one to two weeks after their isolation from biological samples). All other isolates were stored as freeze-dried cultures.

### **Antimicrobial drug susceptibility testing**

Antimicrobial drug susceptibility was determined by the disk diffusion method on Mueller-Hinton agar, with a panel of 32 antimicrobial drugs (Bio-Rad, Marnes-la-Coquette, France), as previously described.<sup>17</sup> The minimal inhibitory concentrations (MICs) of ampicillin and penicillin G (also known as benzylpenicillin) of 67 isolates were determined with Etest strips (bioMérieux, Marcy L'Etoile, France) (table S1).

### **Resistance transfer determination**

A resistance transfer experiment was carried out with the 11 ampicillin-resistant *S. enterica* serotype Typhimurium isolates on solid medium. *E. coli* K-12 J5 Azi<sup>R</sup> and *S. enterica* serotype Typhimurium STM 36-57 NaI<sup>R</sup>, a nalidixic acid-resistant mutant generated in the laboratory from the wild-type isolate STM 36-57 (IncF-), were used as recipient strains for plasmid conjugative transfer screening, and for estimating transfer efficiency, respectively. Strains were mated on a filter placed on either Trypto-Casein-Soy (TCS) (when using *S. enterica* serotype Typhimurium STM 36-57 NaI<sup>R</sup> recipient strain) or Drigalski (when using *E. coli* K-12 J5 Azi<sup>R</sup> recipient strain) agar. Transconjugants were selected on the same type of agar medium as used for mating, but supplemented with either ampicillin (75 mg/L) or penicillin G (15 mg/L) and nalidixic acid (64 mg/L) or sodium azide (500 mg/L). For isolates yielding no transconjugants on TCS agar, another mating experiment was performed on M9 minimal medium rather than TCS.<sup>18</sup> Three transconjugants were selected at random for each experiment. Plasmid classification was based on PCR-based replicon typing.<sup>17</sup> For matings between *S. enterica* serotype Typhimurium strains, CRISPOL typing (see below) was performed to distinguish between recipient and donor strains. Each conjugation experiment was carried out in triplicate. Conjugation frequencies were calculated by dividing the number of transconjugants (counted on agar medium containing nalidixic acid and ampicillin or penicillin G) by the total number of recipient cells (counted on agar medium containing only nalidixic acid). The values shown are means from triplicate experiments.

## DNA extraction

Total DNA was extracted with the Instagene kit (Biorad) or the Maxwell 16 cell DNA purification kit (Promega, Madison WI) for CRISPOL typing and whole-genome sequencing, respectively, in accordance with the manufacturer's recommendations.

## CRISPOL typing

The CRISPOL (CRISPr POLymorphisms) high-throughput method for subtyping *S. enterica* serotype Typhimurium and its monophasic variant was carried out on the Luminex platform, as previously described.<sup>19</sup>

## Whole-genome sequencing

Whole-genome sequencing was performed on a selection of 225 of the 288 *S. enterica* serotype Typhimurium isolates. The selection was designed to obtain wide coverage in terms of source, spatiotemporal and genetic diversity (based on CRISPOL typing) (table S1, Appendix, page 1). Whole-genome sequencing was carried out at the genomics platform of Institut Pasteur on an Illumina HiSeq 2500 platform generating 118 to 146 bp paired-end reads, yielding a mean of 233-fold coverage (minimum 41-fold, maximum 960-fold). Short-read sequence data were submitted to EnteroBase (<http://enterobase.warwick.ac.uk/>).

The oldest ampicillin-resistant *S. enterica* serotype Typhimurium isolate, ExPB 5-59 was also sequenced by GATC Biotech (Konstanz, Germany), on the PacBio RS II platform (Pacific Biosciences, CA, USA). PacBio sequencing errors for homopolymers were corrected on the basis of Illumina sequences. The plasmid sequence was annotated with the RAST server (<http://rast.nmpdr.org/>), with manual inspection and correction. The sequence was deposited in GenBank under the accession code MF428416.

## Other studied genomes

Two published *S. enterica* serotype Typhimurium genome sequences, those for DT104,<sup>20</sup> a representative strain of the multidrug-resistant DT104 clone (GenBank accession no. NC\_022569), and SO4698-09,<sup>21</sup> a representative strain of the multidrug-resistant European monophasic clone (GenBank accession no. NZ\_LN999997), were downloaded from GenBank, and 100 bp overlapping simulated reads were generated. Short-read sequences from the *S. enterica* serotype Weltevreden 840K strain, used as the outgroup, were downloaded from EnteroBase.

## Read alignment and SNP detection

Illumina-generated paired-end reads and simulated paired-end reads from publicly available assembled genomes were mapped onto the reference genome of *S. enterica* serotype Typhimurium strain LT2, including the chromosome (GenBank accession no. AE006468) and plasmid pSLT (GenBank accession no. AE006471), with Bowtie 2 version 2.1.0. Single-nucleotide polymorphisms (SNPs) were identified with SAMtools version 0.1.19, as previously described.<sup>22</sup> SNPs present in insertion sequences identified by ISfinder (<https://www-is.biotoul.fr/>) or recombinogenic regions identified by Gubbins were removed, as previously described.<sup>22</sup>

## Phylogenetic analyses

The maximum-likelihood (ML) phylogenetic tree was built from a 32,741-chromosomal SNP alignment of all *S. enterica* genomes. RAxML version 7.8.6 (GTRGAMMA substitution model) was used to construct the tree. We performed 100 bootstrap pseudoreplicate analyses to assess support for the ML phylogeny. The tree was rooted on *S. enterica* serotype Weltevreden 840K. The final tree was visualised with FigTree version 1.4.2.

## De novo assembly

The reads for each strain were assembled *de novo* with SPAdes version 3.6.0.<sup>23</sup>

## Genetic analyses

*In silico* MLST and serotyping were performed with the “Achtman 7 Gene” and “Serotype Prediction (SISTR)” tools of Enterobase, respectively. The presence and type of antimicrobial resistance genes (ARGs) or ARG-containing structures were determined with ResFinder version 2.1 (<https://cge.cbs.dtu.dk/services/ResFinder/>), PlasmidFinder version 1.3 (<https://cge.cbs.dtu.dk/services/PlasmidFinder/>), pMLST version 1.2 (<https://cge.cbs.dtu.dk/services/pMLST/>) on SPAdes assemblies. The CRISPOL Luminex data (presence or absence of 72 defined DNA sequences of 25-32 bp in length) were confirmed for each isolate, by searches for the 72 sequences within the whole-genome sequence (either from reads or the SPAdes assembly).

## ROLE OF THE FUNDING SOURCE

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## RESULTS

### Antimicrobial drug susceptibility testing of historical isolates

We carried out antimicrobial drug susceptibility testing on 288 historical *S. enterica* serotype Typhimurium isolates from various sources and geographic areas, between 1911 and 1969 (table S1). Most (253/288, 87.8%) of these isolates were susceptible to all the antibiotics tested (table S1). However, 11 (3.8%) were resistant to ampicillin, including 6 (2.1%) also resistant to other antibiotics, such as sulfonamides, trimethoprim, tetracycline, and aminoglycosides (tables 2, S1, figure 1). These 11 isolates were collected from humans ( $n = 10$ ) in France ( $n = 8$ ) and Tunisia ( $n = 3$ ). Three (2.8%) of the 107 isolates collected between 1911 and 1960 (i.e., before ampicillin went on sale) were resistant to ampicillin. These ampicillin-resistant isolates were collected from humans in France ( $n = 1$ ) and Tunisia ( $n = 2$ ), in 1959 and 1960, respectively.

### Population structure of the historical isolates

The 225 historical *S. enterica* serotype Typhimurium isolates for which whole-genome sequences were obtained belonged to MLST *eBurst groups* 1 (98.7%, 222/225), 138 (0.4%, 1/225) and 243 (0.9%, 2/225). The most frequent sequence types (STs) were ST19 (89.3 %,

201/225), followed by ST313 and ST302 (1.3%, 3/225 each). In total, 123 CRISPOL types were observed, with CT117 the most frequently represented (4.9%, 11/225).

### Characterisation of the ampicillin-resistant isolates

The 11 ampicillin-resistant *S. enterica* serotype Typhimurium isolates detected by the disk diffusion method and shown in Table 2 had penicillin MICs (for ampicillin or penicillin G)  $\geq$  256 mg/L (table S1, figure 2). The genomic analysis identified a  $\beta$ -lactamase gene, either *bla*<sub>TEM</sub> ( $n = 9$ ) or *bla*<sub>OXA</sub> ( $n = 2$ ) in all these isolates. In a selection of 56 *S. enterica* serotype Typhimurium isolates devoid of  $\beta$ -lactamase genes and belonging to the different genomic clades of the collection (see below and Appendix, page 2), the MIC of ampicillin ranged from 0.06 to 8 mg/L (median 1 mg/L), whereas the MIC of penicillin G ranged from 0.125 to 16 mg/L (median 8 mg/L) (table S1, figure 2).

The most frequent  $\beta$ -lactamase gene, *bla*<sub>TEM-1B</sub>, carried by transposon Tn2,<sup>26</sup> was found in eight (72.7%) ampicillin-resistant *S. enterica* serotype Typhimurium isolates. The Tn2 transposon was located on plasmids from incompatibility groups IncF ( $n = 5$ ) and IncX1 ( $n = 3$ ).

An analysis of Tn2 transposon junctions revealed that, in four isolates, Tn2 was located on pSLT, a 93 kb IncF plasmid known as the *S. enterica* serotype Typhimurium type strain LT2 virulence plasmid.<sup>18</sup> Tn2 was found integrated into three different sites of this pSLT (figure 3). In two ST313 isolates, STM 232-68 and STM 266-68, collected in France and Tunisia in 1968, the Tn2 integration site was identical to that of pSLT-BT, a multidrug resistance plasmid found in lineage II of the ST313 clone, a serotype Typhimurium population associated with invasive infections in Africa.<sup>25</sup> Furthermore, as observed in pSLT-BT, a Tn21-like element was also found integrated into the Tn2 of the STM 232-68 and STM 266-68 IncF plasmids. However, these two plasmids differed slightly from pSLT-BT in having a 6.8 kb IS1-mediated deletion (corresponding to region 41,666-48,543 of GenBank accession number FN432031) encompassing the integron (In2) region of the Tn21-like element. The IncF plasmid of the fifth isolate (STM 366-66) was derived from the NR1/R100 plasmid (lacking Tn21) (GenBank accession no. DQ364638). This isolate did not contain the pSLT virulence plasmid.

The 45 kb Tn2-containing IncX1 plasmid pExpB5-59-1 was also sequenced with the PacBio platform. This plasmid was present in a clonal bacterial population (ST19-CT160) isolated in France over a 10-year period (1959 to 1969). The non-Tn2 backbone of plasmid pExpB5-59-1 resembled a 42.2 kb conjugative plasmid (pYU39\_IncX; GenBank accession no. CP011431) found in an emerging invasive ST213 *S. enterica* serotype Typhimurium population from Mexico in the 2000s. In our study, five other *S. enterica* serotype Typhimurium isolates also carried an IncX1 plasmid but were devoid of antibiotic resistance genes (table S1). These plasmids had only 45-77% of their DNA sequences (98-99% nucleotide identity) in common with pExpB5-59-1. Interestingly, the oldest of these isolates was collected during a foodborne outbreak of disease in Newcastle, United Kingdom in 1911.

The three *S. enterica* serotype Typhimurium isolates collected in France and Tunisia before the release of ampicillin contained the Tn2 transposon, in the IncX1 ( $n = 1$ ) and pSLT-derived IncF ( $n = 2$ ) plasmids. Tn2 was found inserted at two different sites within IncF.

### Transfer of ampicillin resistance



The 11 ampicillin-resistant *S. enterica* serotype Typhimurium isolates were crossed with *S. enterica* serotype Typhimurium STM 36-57 Nal<sup>R</sup> recipient strains in the presence of ampicillin (75 mg/L), to determine the transferability of the different *bla* gene-carrying plasmids. The nine isolates containing a *bla*<sub>TEM</sub> gene-carrying plasmid were found to transfer ampicillin resistance by conjugation. Transfer frequencies ranged from  $2.4 \times 10^{-4}$  (Inc11 plasmid from isolate STM 127-67) to  $1.6 \times 10^{-8}$  (pSLT-derived IncF plasmid from isolate STM 266-68) transconjugants/recipient (Appendix, page 3). The use of M9 minimal medium was required to obtain transconjugants for STM 110-60. The two isolates containing a *bla*<sub>OXA</sub> gene (STM 84-67 and STM 345-66) did not transfer their ampicillin resistance by conjugation.

The four Tn2-carrying plasmids tested (the IncX1 plasmid from isolate ExPB 5-59, pSLT-derived IncF plasmids from isolates STM 65-60 and STM 110-60, and NR1/R100-derived IncF plasmids from isolate STM 366-66) were transferred between the two strains of *S. enterica* serotype Typhimurium in the presence of penicillin G (15 mg/L). Transfer frequencies ranged from  $1.2 \times 10^{-3}$  (IncX1 plasmid from isolate ExPB 5-59) to  $3.0 \times 10^{-6}$  (pSLT-derived IncF plasmid from isolate STM 110-60) transconjugants/recipient.

### Genetic relationships between the ampicillin-resistant isolates

A whole-genome phylogeny based on 32,741 SNPs evenly distributed over the non-repetitive non-recombinant core genome showed the clustering of the 11 ampicillin-resistant *S. enterica* serotype Typhimurium isolates into three groups (figure 4, Appendix, page 4).

The main group comprised seven ampicillin-resistant isolates collected in France between 1959 and 1969, with four types of  $\beta$ -lactamase genes (*bla*<sub>TEM-1A</sub>, *bla*<sub>TEM-1B</sub>, *bla*<sub>OXA-1</sub>, and *bla*<sub>OXA-2</sub>). These seven ampicillin-resistant isolates belonged to a clade consisting of 13 West and South European isolates (plus one North African isolate) collected from humans, food and animals (chicks) and belonging to ST19 and CT46 or variants of CT46 (table S1).

The second group of ampicillin-resistant isolates comprised two human ST19-CT896 isolates collected in Tunisia in 1960. They belonged to a clade consisting of 14 ST19-CT117 (or CT117-derived) isolates mostly obtained from humans (with the exception of one isolate from a sausage) in France, Belgium and Tunisia between 1953 and 1963.

The third group comprised the two ampicillin-resistant ST313-CT810 isolates collected in France and Tunisia in 1968. They belonged to a clade consisting of six isolates belonging to either ST302 ( $n = 3$ ) or ST313 ( $n = 3$ ). The four non ampicillin-resistant isolates originated from Turkey (two ST302 and one ST313) and Tunisia (one ST302) and were collected between 1956 and 1965.

## DISCUSSION

We show here that transmissible ampicillin resistance actually emerged several years earlier than previously thought in *S. enterica* serotype Typhimurium, as the  $\beta$ -lactamase *bla*<sub>TEM-1B</sub> gene was found on two different plasmids (IncX1 and IncF) in three isolates from Europe (France) and Africa (Tunisia) collected in 1959-1960. However, we did not aim here to determine the true extent of the global distribution of ampicillin resistance in *S. enterica* serotype Typhimurium, due to an obvious sampling bias (i.e., the available historical isolates studied here were predominantly from France and former French colonies in Africa and

Asia). Another limitation of our study was the mode of storage of the collection strains. Almost all the isolates were stored as stab cultures and the absence of antibiotic selection pressure may not have favored the maintenance of acquired plasmids in the long term in these strains, as previously reported for *Bacillus anthracis*.<sup>26</sup> The proportion of ampicillin-resistant isolates in our selection of *Salmonella* isolates may therefore be underestimated.

Nevertheless, the use of ampicillin, the first broad-spectrum penicillin indicated for the treatment of infections due to Enterobacteriaceae, which began in 1961,<sup>5</sup> is unlikely to have led to the selection of these early ampicillin-resistant isolates. Following the emergence of multidrug resistance in zoonotic *S. enterica* serotype Typhimurium DT29, E.S. Anderson from the Central Public Health Laboratory in the United Kingdom warned that the use of penicillin G and tetracyclines as feed additives for livestock (in poultry and pigs) and as prophylactics in cattle was responsible for the emergence of these multidrug-resistant *S. enterica* serotype Typhimurium isolates.<sup>13</sup> Long-term exposure to subtherapeutic levels of antibiotics in livestock may have been critical for the selection of ampicillin resistance in a prevalent zoonotic agent found in a wide range of food-producing animals. We found that the median MIC of penicillin G in wild-type *S. enterica* serotype Typhimurium isolates was not particularly high, at 8 mg/L, whereas Fleming, in his rudimentary assays based on *Penicillium* broth filtrates, reported that members of the “coli-typhoid group” were unaffected by the inhibitory power of penicillin.<sup>3</sup> The relatively low MIC values obtained for penicillin G against *S. enterica* serotype Typhimurium may account for the acquisition of ampicillin resistance in an environment in which penicillin G residues reached concentrations just above the MICs of the bacteria. For example, in France during the 1960s, penicillin G was permitted as a food additive, and was used at concentrations of 5-200 mg/kg of feed, in all animals for human consumption.<sup>27</sup> Penicillin G is sensitive to gastric acid, but its poor absorption by the digestive system can result in non-negligible concentrations in the digestive tract and excretion onto the litter. A study in the US in 1975 found that the litter of broiler chickens fed with antibiotics contained 0 to 25 mg/kg (mean 12.5 mg/kg) penicillin.<sup>28</sup> Using similar concentrations (15 mg/L of penicillin G), we showed that a *bla* gene carried by IncX1 or IncF plasmids could be successfully transferred between *S. enterica* serotype Typhimurium strains. Our results provide no proof of a causal link between the use of penicillin G and the emergence of transmissible ampicillin-resistance in livestock, but they suggest that an environment in which penicillin G residues were abundant, such as the farming environment described by Anderson<sup>22</sup>, may have favored the diffusion of *bla* gene-carrying plasmids in the most prevalent *Salmonella* serotype at the time. We identified no ampicillin-resistant isolates originating from livestock. However, this may reflect the fact that only ~50% of the animal isolates tested were from livestock, with ~40% of the livestock isolates originating from Africa and South America. After 1969, in the United Kingdom, in accordance with the recommendations of the joint committee on the use of antibiotics in animal husbandry and veterinary medicine (known as the “Swann” Committee in honor of its chairman), the use of growth promoters in livestock was eventually restricted to drugs with little or no application as therapeutic agents in humans or animals, and penicillins, in particular, were identified as unsuitable for use as growth promoters.

Interestingly, the vectors of ampicillin resistance in our *S. enterica* serotype Typhimurium isolates, mostly from France, differed from those implicated in the first outbreaks in the United Kingdom, despite the geographic proximity of France and the United Kingdom. Thus, the early emergence of ampicillin resistance in *S. enterica* serotype Typhimurium was due to multiple horizontal gene transfers of *bla* genes followed by clonal expansion of various extents within individual countries, from the very rare isolation of CT160 during a 10-year

period in France to the hundreds of isolations of DT29 per year in the United Kingdom. This situation contrasts with the rapid global dissemination subsequently observed for two multidrug-resistant *S. enterica* serotype Typhimurium clones: DT104 in the 1980s and the monophasic 1,4,[5],12:i:- in the 2000s.<sup>29,30</sup> This difference may be due to changes in the food trade, which was much less globalised in the late 1950s-early 1960s than it has been since in the 1980s.

In conclusion, our study sheds light on the early steps in the emergence of resistance to ampicillin, the first penicillin active against Enterobacteriaceae, still in widespread use. We found that the relationship between antibiotic use and the selection of plasmids encoding resistance determinants was not as direct as often presumed. Since the early days of antibiotic use, agriculture has played a role in the selection of resistant pathogens alongside medicine, particularly during periods of intensification. These findings highlight the need for a “one-health” approach to tackling the issue of antibiotic resistance.

Table 1: Characteristics of the first *S. enterica* (and *E. coli* TEM) isolates with plasmid-mediated ampicillin resistance to be reported.

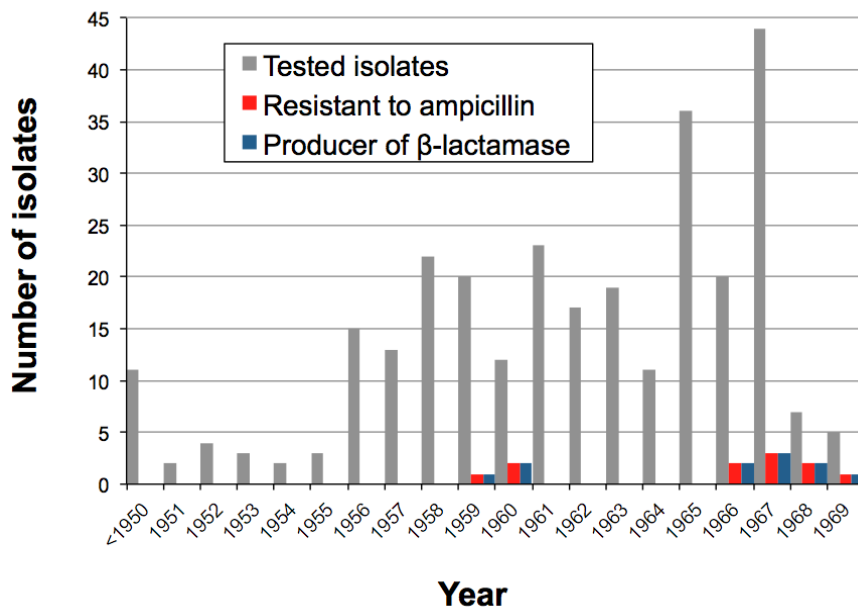
Bacterial host	Origin	AST	Plasmid	Type	<i>bla</i> gene	Other resistance genes	Reference
<i>S. Typhimurium</i> DT2	1962, UK, Humans and pigs	ApStrSulTe	R <sub>1818</sub> = R46*	51kb, IncN	OXA-2 <sup>Ω</sup>	<i>aadA1,sul1,tet(C)</i>	6,10
<i>S. Paratyphi</i> B	1963, UK, Human	ApKCSul	R <sub>7268</sub> = R1 <sup>†</sup>	97.5kb, IncF	TEM-1A	<i>aadA1, aph(3')-Ic, catA1,sul1</i>	6,8
<i>E. coli</i>	1963, Greece, Human	ApStr	R <sub>TEM</sub> = R6K <sup>#</sup>	39kb, IncX2	TEM-1A	<i>strA/B</i>	6,9
<i>S. Derby</i>	1963, France, Humans	ApStrKCSulTe	R <sub>4-5</sub>	-	-	-	11
<i>S. Oranienburg</i>	1964, France, Humans	ApStrKCSulTe	R <sub>7</sub>	-	-	-	11
<i>S. Typhimurium</i> DT29	1964, UK, Humans and cattle	ApStrSulTe	NTP1 <sup>NC</sup>	8.3kb	TEM-1	None	12-14

AST, antimicrobial drug susceptibility type; Ap, ampicillin; Str, streptomycin; K, kanamycin; C, chloramphenicol; Sul, sulfonamides; Tmp, trimethoprim; Te, tetracycline; -, unknown; DT, definitive phage-type; available from GenBank under accession nos. AY046276\* or KY749247<sup>†</sup> or from the Wellcome Trust Sanger Institute<sup>#</sup>, <sup>NC</sup> non-conjugative; <sup>Ω</sup> duplicated gene.

Table 2: Characteristics of the 11 ampicillin-resistant *S. enterica* serotype Typhimurium isolates identified in this study.

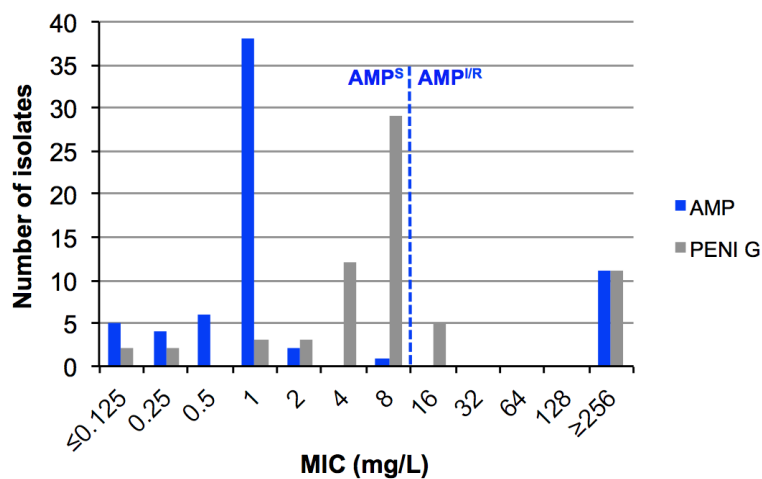
Isolate	Year	Country	Host	AST	<i>bla</i> gene	Other resistance genes	ST	CT	Plasmid
ExPB 5-59	1959	France	Human	Ap	TEM-1B	-	19	160	IncX1 <sup>+Ω</sup>
STM 110-60*	1960	Tunisia	Human	ApStrSpSul	TEM-1B	<i>aadA1</i>	19	896	IncF <sup>+Ω</sup>
STM 65-60*	1960	Tunisia	Human	Ap	TEM-1B	-	19	896	IncF <sup>+Ω</sup>
STM 345-66	1966	France	Human	ApStrSpSulTe	OXA-2	<i>aadA1,sul1,tet(C)</i>	19	46	IncI1, IncN <sup>Ω</sup>
STM 366-66	1966	France	Human	ApSul	TEM-1B	<i>sul1</i>	19	46	IncI1, Col, IncF <sup>+Ω#</sup>
ExPB 63-67	1967	France	Human	Ap	TEM-1B	-	19	160	IncX1 <sup>+Ω</sup>
STM 84-67	1967	France	Human	ApSpKCSulTe	OXA-1	<i>aph(3')-Ia, aadA1,catA1,sul1, tet(B)</i>	19	46	Col, IncF
STM 127-67	1967	France	Human	ApK	TEM-1A	<i>aph(3')-Ia</i>	19	46	IncI1 <sup>+Ω</sup>
STM 232-68	1968	France	Human	ApStrCSulTmp	TEM-1B	<i>strA/B,catA1,sul2, dfrA1</i>	313	810	IncF <sup>+Ω</sup>
TM 266-68	1968	Tunisia	?	ApStrCSulTmp	TEM-1B	<i>strA/B,catA1,sul2, dfrA1</i>	313	810	IncF <sup>+Ω</sup>
ExPB 37-69	1969	France	Human	Ap	TEM-1B	-	19	160	IncX1 <sup>+Ω</sup>

AST, antimicrobial drug susceptibility type; Ap, ampicillin; Str, streptomycin; Sp, spectinomycin; K, kanamycin; C, chloramphenicol; Sul, sulfonamides; Tmp, trimethoprim; Te, tetracycline; ?, unknown; \*, both isolates were received, without associated epidemiological information, from the Pasteur Institute of Tunis in August (STM 65-60) and October (STM 110-60) 1960, together with other isolates of *Salmonella* spp., *Shigella* spp. and enteropathogenic *E. coli*; plasmid location of the β-lactamase gene demonstrated by conjugation<sup>†</sup> and/or genomic analysis<sup>Ω</sup>; <sup>Ω</sup> incomplete IncN plasmid (8.8 kb fragment similar to R46, comprising the region from *ardA* to *repA*); <sup>#</sup>IncF plasmid derived from NR1/R100 lacking Tn21, the other IncF plasmids are derived from virulence plasmid pSLT.



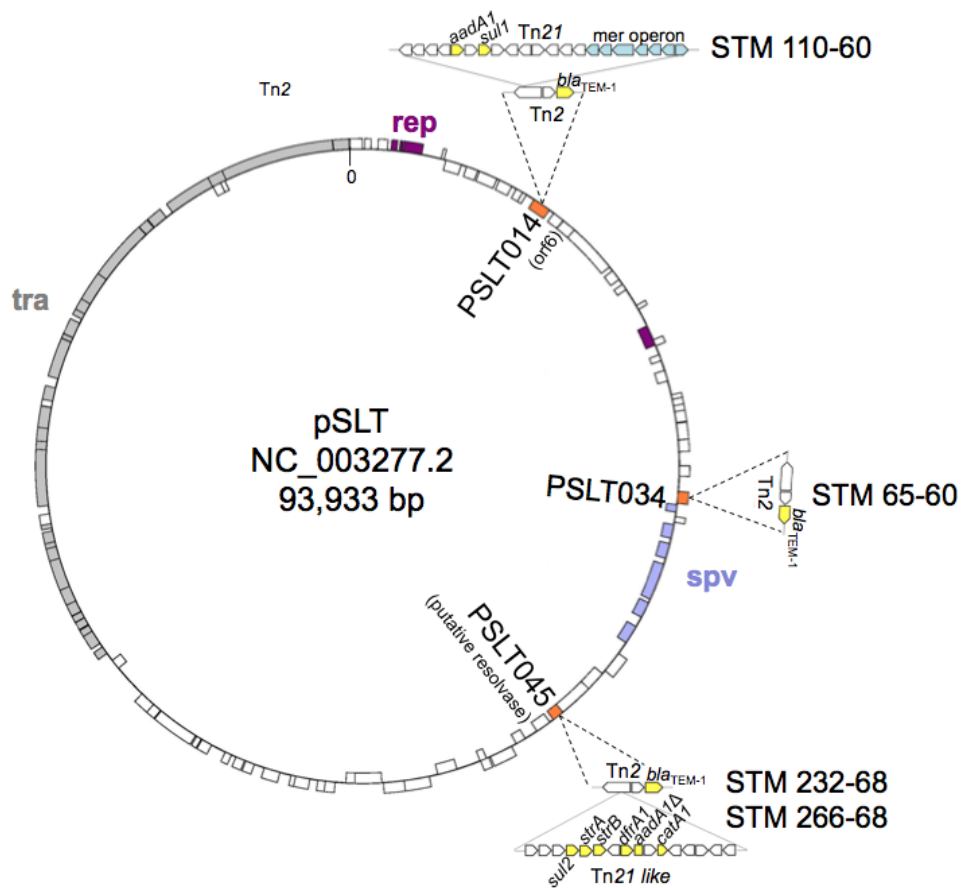
**Figure 1. Number of ampicillin-resistant and  $\beta$ -lactamase-producing *S. enterica* serotype Typhimurium isolates, by year of isolation.**

The number of *S. enterica* serotype Typhimurium isolates studied is shown in gray, that of isolates resistant to ampicillin is shown in red, and that of isolates producing a  $\beta$ -lactamase is shown in blue.



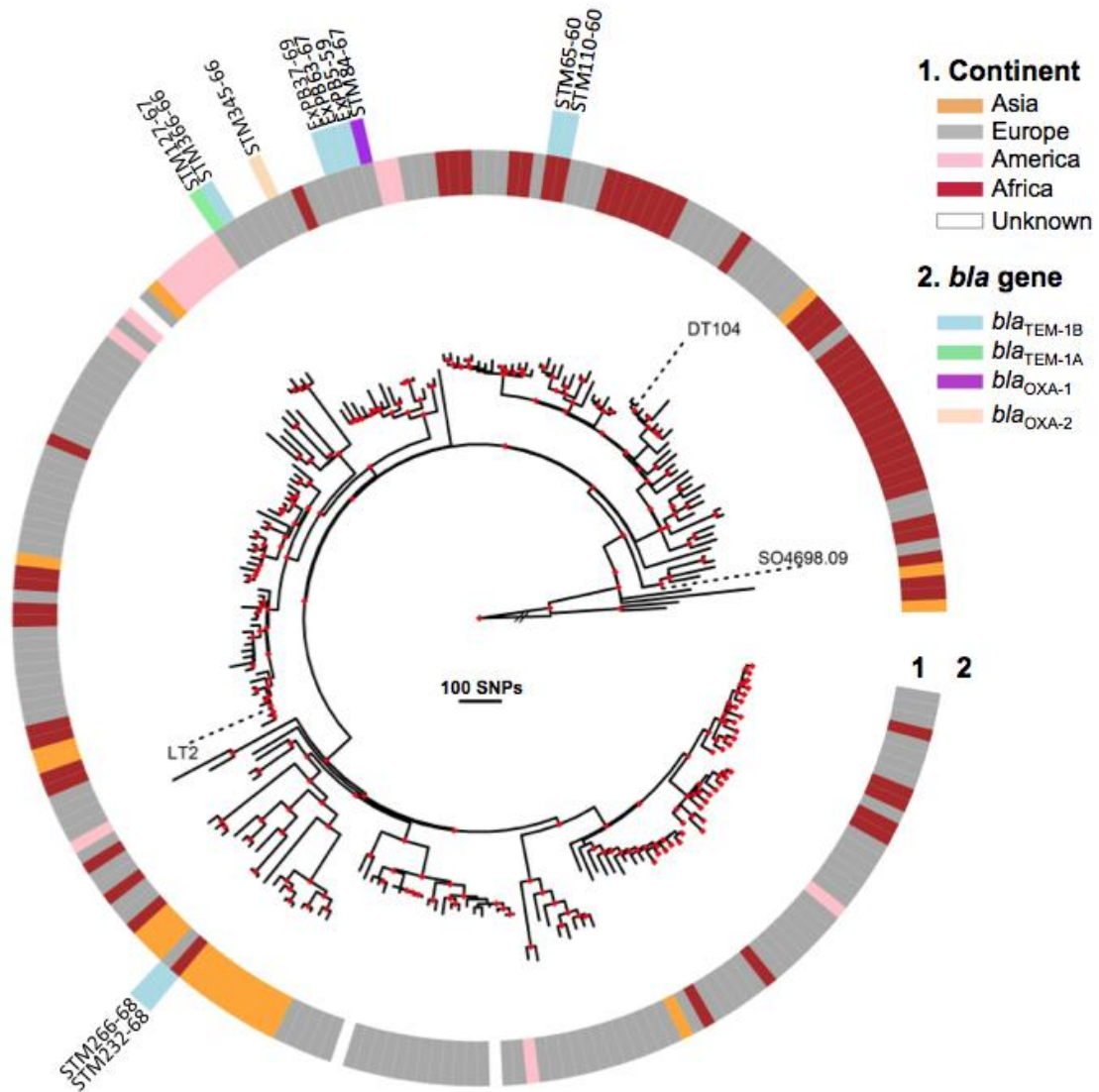
**Figure 2: The distribution of ampicillin and penicillin G MICs in a selection of 67 *S. enterica* serotype Typhimurium isolates.**

The clinical breakpoint distinguishing ampicillin susceptibility from ampicillin non-susceptibility according to EUCAST ( $S \leq 8$  mg/L;  $R > 8$  mg/L) and CLSI ( $S \leq 8$  mg/L;  $I = 16$  mg/L;  $R \geq 32$  mg/L) is shown by a dashed blue bar.



**Figure 3: Integration sites of the Tn2 transposon within the IncF virulence plasmid pSLT in five ampicillin-resistant *S. enterica* serotype Typhimurium isolates.**

The genes into which Tn2 is integrated are boxed in red and their names and function (in parentheses), according to GenBank accession no. NC\_003277.2, are shown. The antibiotic resistance genes are boxed in yellow. The plasmid genes involved in transfer (tra), replication (rep) and virulence (spv) are shown in gray, purple and light blue, respectively.



**Figure 4: Maximum-likelihood phylogeny of the *S. enterica* serotype Typhimurium isolates.**

The geographic origin (continent) of the isolates is shown in ring 1. The  $\beta$ -lactamase genes of the 11 ampicillin-resistant isolates are indicated in ring 2. The  $\beta$ -lactamase genes of comparison genomes DT104 ( $bla_{PSE-1}=bla_{CARB-2}$ ) and monophasic ( $bla_{TEM1-B}$ ) are not shown. The names of the 11 ampicillin-resistant isolates are indicated above ring 2. The LT2, DT104 and monophasic (SO4698-09) comparison genomes are indicated by black dashed lines. The nodes supported by bootstrap values  $\geq 0.9$  are indicated by red diamonds. The *S. enterica* serotype Weltevreden 840K strain was used as the outgroup, and its branch length has been shortened by a factor of 100, as indicated by the double slash (//), to improve visualisation.

## **AUTHOR CONTRIBUTIONS**

Conceived and designed the experiments: FXW. Performed the experiments: ATD. Contributed reagents/materials/analysis tools: SLH, CB. Analysed the data: ATD and FXW. Wrote the paper: ATD and FXW. Reviewed, critiqued and offered comments on the text: SLH, CB.

## **ACKNOWLEDGMENTS**

We thank G. Costilhes, L. Ma, and A. Rivasseau for expert technical assistance. We thank all corresponding laboratories of the French National Reference Centre for *E. coli*, *Shigella*, and *Salmonella*.

The French National Reference Centre for *E. coli*, *Shigella*, and *Salmonella* is funded by the Institut Pasteur and *Santé Publique France*. A. Tran Dien was supported by a grant from the *Fondation Le Roch-Les Mousquetaires*. The “Unité des Bactéries Pathogènes Entériques” belongs to the “Integrative Biology of Emerging Infectious Diseases” Laboratory of Excellence funded by the French Government “Investissement d’Avenir” programme (grant no. ANR-10-LABX-62-IBEID). The Genomics Platform is member of the “France Génomique” consortium (ANR10-INBS-09-08).

## **CONFLICTS OF INTEREST**

None of the authors has any competing interests or conflicts of interest to declare.



## REFERENCES

1. WHO. Global action plan on antimicrobial resistance. Geneva: World Health Organization, 2015. Available from: [http://www.wpro.who.int/entity/drug\\_resistance/resources/global\\_action\\_plan\\_eng.pdf](http://www.wpro.who.int/entity/drug_resistance/resources/global_action_plan_eng.pdf) (accessed June 26th 2017).
2. Bud R. Penicillin: Triumph and Tragedy. *Oxford University Press*, Oxford 2007.
3. Fleming A. On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzae*. *Br J Exp Pathol* 1929; **10**: 226–236.
4. Batchelor FR, Doyle FP, Nayler JH, Rolinson GN. Synthesis of penicillin: 6-aminopenicillanic acid in penicillin fermentations. *Nature* 1959; **183**:257–8.
5. Knox R. A survey of new penicillins. *Nature* 1961; **192**:492–6.
6. Datta N, Kontomichalou P. Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. *Nature* 1965; **208**:239–41.
7. Medeiros AA. B-lactamases. *Br Med Bull* 1984; **40**:18–27.
8. Hedges RW, Datta N, Kontomichalou P, Smith JT. Molecular specificities of R factor-determined  $\beta$ -lactamases: correlation with plasmid compatibility. *J Bacteriol* 1974; **117**:56–62.
9. Cox KEL, Schildbach JF. Sequence of the R1 plasmid and comparison to F and R100. *Plasmid* 2017; **91**:53–60.
10. Anderson ES, Datta N. Resistance to penicillins and its transfer in Enterobacteriaceae. *Lancet* 1965; **1**:407–9.
11. Chabbert YA, Baudens JG. Transmissible resistance to six groups of antibiotics in *Salmonella* infections. *Antimicrob Agents Chemother (Bethesda)* 1965; **5**:380–3.
12. Anderson ES. Drug resistance in *Salmonella* Typhimurium and its implications. *Br Med J* 1968; **3**:333–9.
13. Anderson ES, Threlfall EJ, Carr JM, McConnell MM, Smith HR. Clonal distribution of resistance plasmid-carrying *Salmonella* Typhimurium, mainly in the Middle East. *J Hyg (Lond)* 1977; **79**:425–48.
14. Grindley JN, Nakada D. The nucleotide sequence of the replication origin of plasmid NTP1. *Nucleic Acids Res* 1981; **9**:4355–66.
15. Kelterborn E. *Salmonella*-species. First isolations, names and occurrence. *Springer-Verlag*, New York 1967.
16. European Food Safety Authority. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015 <http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2016.4634/epdf> (accessed June 26, 2017). *EFSA Journal* 2016; **14**:4634.
17. Le Hello S, Harrois D, Bouchrif B, et al. Highly drug-resistant *Salmonella enterica* serotype Kentucky ST198-X1: a microbiological study. *Lancet Infect Dis*. 2013; **13**:672–9.
18. Ahmer BM, Tran M, Heffron F. The virulence plasmid of *Salmonella typhimurium* is self-transmissible. *J Bacteriol* 1999; **181**:1364–8.
19. Fabre L, Zhang J, Guigon G, et al. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PLoS One* 2012; **7**:e36995.
20. Mather AE, Reid SW, Maskell DJ, et al. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science* 2013; **341**:1514–7.
21. Petrovska L, Mather AE, AbuOun M, et al. Microevolution of Monophasic *Salmonella* Typhimurium during Epidemic, United Kingdom, 2005–2010. *Emerg Infect Dis* 2016; **22**:617–24.

22. Kuijpers LM, Le Hello S, Fawal N, et al. Genomic analysis of *Salmonella enterica* serotype Paratyphi A during an outbreak in Cambodia, 2013-2015. *Microb Genom* 2016; **2**:e000092.
23. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**:455–77.
24. Partridge SR, Hall RM. Evolution of transposons containing blaTEM genes. *Antimicrob Agents Chemother* 2005; **49**:1267–8.
25. Kingsley RA, Msefula CL, Thomson NR, et al. Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* 2009 Dec; **19**:2279-87.
26. Marston CK, Hoffmaster AR, Wilson KE, et al. Effects of long-term storage on plasmid stability in *Bacillus anthracis*. *Appl Environ Microbiol* 2005; **71**:7778–80.
27. Gounelle H, Szakvary A. [Antibiotics and food. IV. Use of antibiotics outside medical prescriptions]. *Bull Acad Natl Med* 1966; **150**:149-56. In French.
28. Webb KE Jr, Fontenot JP. Medicinal drug residues in broiler litter and tissues from cattle fed litter. *J Anim Sci* 1975; **41**:1212–7.
29. Threlfall EJ. Epidemic *Salmonella* Typhimurium DT 104—a truly international multiresistant clone. *J Antimicrob Chemother* 2000; **46**:7–10.
30. Hopkins KL, Kirchner M, Mevius DJ. Multiresistant *Salmonella enterica* serovar 4,[5],12:i:- in Europe: a new pandemic strain? *Euro Surveill* 2010; **15**:19580.

**Annexe 3 a. Article “ Millennia of genomic stability within the invasive Para C lineage of *Salmonella enterica*”.**

**Title: Millennia of genomic stability within the invasive Para C Lineage of  
*Salmonella enterica***

**Authors:** Zhemin Zhou<sup>1\*</sup>, Inge Lundstrøm<sup>2</sup>, **Alicia Tran Dien**<sup>3</sup>, Sebastián Duchêne<sup>4</sup>, Nabil-Fareed Alikhan<sup>1</sup>, Martin J. Sergeant<sup>1</sup>, Gemma Langridge<sup>5†</sup>, Anna K. Fotakis<sup>2</sup>, Satheesh Nair<sup>5†</sup>, Hans K. Stenøien<sup>6</sup>, Stian S. Hamre<sup>7</sup>, Sherwood Casjens<sup>8</sup>, Axel Christophersen<sup>6</sup>, Christopher Quince<sup>1</sup>, Nicholas R. Thomson<sup>5</sup>, François-Xavier Weill<sup>3</sup>, Simon Y. W. Ho<sup>9</sup>, M. Thomas P. Gilbert<sup>2,6\*</sup>, Mark Achtman<sup>1\*</sup>

**Affiliations:**

<sup>1</sup>Warwick Medical School, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK.

<sup>2</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark.

<sup>3</sup>Unité des Bactéries Pathogènes Entériques, Institut Pasteur, Paris, France.

<sup>4</sup>Centre for Systems Genomics, University of Melbourne, Parkville, Victoria 3010, Australia.

<sup>5</sup>Wellcome Trust Sanger Institute, Cambridge, UK

<sup>6</sup>NTNU University Museum, N-7491 Trondheim, Norway

<sup>7</sup>Department of Archaeology, History, Cultural Studies and Religion, University of Bergen, Post Box 7805, 5020 Bergen, Norway

<sup>8</sup>Pathology Department, University of Utah School of Medicine, Salt Lake City, Utah 84112, U.S.A.

<sup>9</sup>School of Life and Environmental Sciences; University of Sydney; Sydney, NSW 2006, Australia

\*Correspondence to: M.A.: [m.achtman@warwick.ac.uk](mailto:m.achtman@warwick.ac.uk); M.T.P.G.: [tgilbert@snm.ku.dk](mailto:tgilbert@snm.ku.dk); Z.Z.: [zhemin.zhou@warwick.ac.uk](mailto:zhemin.zhou@warwick.ac.uk)

†Present addresses: Molecular Microbiology, Norwich Medical School, University of East Anglia, Norwich, UK (G.L.); Gastrointestinal Bacteria Reference Unit, Public Health England, London, UK (S.N.)

*Salmonella enterica* serovar Paratyphi C is the causative agent in Africa and Asia of enteric (paratyphoid) fever<sup>1,2</sup>, a potentially lethal infection of humans. Here we compare a draft Paratyphi C genome from the 800-year-old skeleton of a young woman in Trondheim, Norway, against a new database of related modern genomes. We demonstrate that Paratyphi C is descended from the ancestors of swine pathogens, serovars Choleraesuis and Typhisuis, and may have evolved after zoonotic transfer from swine during the Neolithic

period.. Our results also indicate that only minor changes have occurred in the genome of Paratyphi C over 1,300 years.

We generated metagenomic data from teeth and long bones of 33 Trondheim skeletons (Fig. 1A,C). Initial scanning identified 266 *Salmonella enterica* reads in a tooth from skeleton SK152, a 19–24 year old woman who was buried in 1200±25 CE<sup>3</sup>. Additional sequences were generated from nine further libraries from SK152 tooth, bone and dental calculus (Extended Data 1-2; Supplementary Table 1; Supplementary Data 1). CONCOCT<sup>4</sup> reconstituted 11 near-complete microbial genomes from the data (Fig. 1D; Extended Data 3; Supplementary Tables 2-6; Supplementary Data 2-3), but only few, short (<1 kb) *S. enterica* contigs were recovered, consistent with the low number of reads.

We therefore maximize the genomic data recovered used by mapping reads to the closest relatives of the SK152 *S. enterica* (henceforth designated Ragna). To this end, we inferred phylogenetic trees of 2,964 genomes from [Enterobase](#) that represent the diversity of 50,000 modern *S. enterica* subspecies *enterica* genomes (Supplementary Data 4). These contained 711,009 SNPs (single-nucleotide polymorphisms) in 3,002 core genes (2.8 Mb). The concatenated core genes were used to infer a maximum-likelihood tree (Fig. 2A; Supplementary Data 5-6), which included the “Para C Lineage”, composed of monophyletic clades of serovars Paratyphi C, Choleraesuis, Typhisuis<sup>2</sup> plus one genome of the extremely rare serovar Lomita (Fig. 2). All SK152 metagenomic reads specific to *S. enterica* were assigned to that Lineage.

Reliable inference of the evolutionary timescale and phylogeographic history of the Para C Lineage depends on having bacterial strains collected from a broad temporal and spatial range, so we also sequenced 119 genomes from historical strains at the Institut Pasteur, Paris. Our assembled data set comprised 219 modern Para C Lineage genomes, isolated between 1914 and 2015 from multiple continents (Supplementary Table 7-8; Supplementary Data 7-9). These genomes defined sub-lineages within each serovar (Paratyphi C: PC-1, PC-2; Choleraesuis: CS Kunzendorf, CS *sensu stricto*, CS-3; Typhisuis: TS-1, TS-2) (Fig. 2B). Mapping the SK152 metagenomic reads against these 219 genomes revealed 1,030,108 unique Paratyphi C-specific reads in teeth (0.05–0.18% of all reads) or the femur (0.01%), but not in dental calculus (Supplementary Table 1). These covered 98.4% of a reference Paratyphi C genome (RKS4594) with a mean read depth of 7.3-fold (Extended Data 4).

Paratyphi C was first recognized in 1916 when it was isolated from Eastern European soldiers suffering from enteric fever<sup>5</sup>. It was common early in the 20<sup>th</sup> century in eastern Europe, northern Africa, and the Americas but has since largely disappeared from Europe<sup>2</sup>. Our data demonstrate that Paratyphi C bacteria were causing human infections in Norway 800 years ago, and their presence in both teeth and bones suggests that SK152 died of septicaemia associated with enteric fever. Paratyphi C has also recently been described from mass graves in Mexico in 1,545 CE<sup>6</sup>. Paratyphi C only infects humans, but serovar Choleraesuis is associated with septicaemia in swine (and occasionally humans) and Typhisuis with epidemic, chronic paratyphoid/caseous lymphadenitis in domestic pigs<sup>1,2</sup>. Although these other serovars continue to cause disease in southern and eastern Asia, Choleraesuis is rare in Europe today, except in wild boar<sup>7</sup>, and Typhisuis has been eradicated from European pigs.

The selective pressures associated with local ecological interactions are thought to cause variation of gene content in microbes<sup>8</sup>. We therefore anticipated dramatic differences in gene content between Ragna and modern Paratyphi C genomes, and expected even greater differences from the other serovars within the Para C Lineage. Surprisingly, 78% of the 4,388±99 genes (total length 4.8±0.08 Mb) in a Para C Lineage genome were intact core

genes even though some of them belong to mobile genetic elements that are variably present in other *Salmonella* (Fig. 2B, Supplementary Table 9; Supplementary Data 10-11). For example, the virulence plasmid was present throughout the Para C Lineage except for Typhisuis sub-lineage TS-2.

Other studies have indicated that microbial host adaptation is accompanied by the accumulation of pseudogenes<sup>9</sup>, driven by the streamlining of genes that are no longer necessary for the infection of multiple hosts<sup>10</sup>, or by rewiring of transcriptional regulation<sup>11</sup>. A median of 40-60 pseudogenes was observed within the 2,964 representative *Salmonella* genomes. We inferred 25 pseudogenes for the most recent common ancestor (MRCA) of the Para C Lineage and 69 for the MRCA of Paratyphi C, Choleraesuis plus Typhisuis (Extended Data 5A; Supplementary Data 12), suggesting that the MRCA had not yet adapted to a particular host. However, the splits into the individual serovars may well mark the beginnings of host adaptation because higher numbers of pseudogenes were inferred for the MRCAs of each of the serovars (Choleraesuis: 95; Paratyphi C: 116; Typhisuis: 181), with increasing numbers towards the present (Extended Data 5A, 6). It is possible that some of these pseudogenes contribute to host specificity.

We attempted to identify mobile genetic elements among the 3,901 genes in the accessory genome of the Para C Lineage that could account for the different host specificities of the serovars. These accessory genes clustered together within 227 GIs (genomic islands and other mobile elements), including 37 plasmids, 32 prophages, 16 IMEs (Integrative and Mobilizable Elements), SPI-5 to SPI-7, and two ICEs (Integrative and Conjugative Elements) (Supplementary Table 10; Supplementary Data 13). Parts or all of these GIs were acquired or lost on 311 independent occasions. However, at least 60% of gains or losses are unlikely to be important for host specificity because they were restricted to a single genome (Supplementary Table 10). Most of the other gains or losses are also unlikely to represent successful evolutionary changes in virulence or host specificity, because, as in other *Salmonella*<sup>12-14</sup>, they were restricted to individual sub-lineages, and sister sub-lineages differing in the possession of those genes are also prevalent in invasive disease (Fig. 2B, Extended Data 5B & 7, Supplementary Table 11). For example, Paratyphi C is more virulent for mice after lysogenization by SCP-P1 (GI076) upstream of *pgtE*<sup>15</sup> whose gene product prevents opsonization<sup>16</sup>. However, GI076 is absent from half of the Paratyphi C genomes, including Ragna (Fig. 2B; Supplementary Data 13). Similarly, the *fimH102* allele of a type 1 fimbrial adhesin facilitates specific adhesion to porcine cells by serovar Choleraesuis<sup>17</sup> but *fimH* is totally lacking in CS *sensu stricto*. Indeed, none of the virulence factors and GIs seemed likely to be consistently related to differential virulence or host specificity (Supplementary Table 11), with the notable exceptions of SPI-7 and SPI-6.

SPI-7 (GI107) is a pathogenicity island in serovars Typhi, Paratyphi C (including Ragna) and Dublin which encodes the Vi capsular polysaccharide. Vi prevents the opsonization and clearance that is triggered by binding of the C3 component of complement to lipopolysaccharide<sup>18</sup>, and might thereby promote enteric fever in humans. The absence of SPI-7 from serovars Typhisuis and Choleraesuis suggests that it was acquired prior to the expansion of serovar Paratyphi C, and might be associated with its human specificity.

SPI-6 (GI008) is present throughout the Para C Lineage. It encodes a Type Six Secretion System (T6SS) as well as *Salmonella* atypical fimbriae (*saf*) and Typhi colonization factor (*tcf*). T6SS systems encode intracellular, inverted bacteriophage-tail-like structures that can inject lethal effectors (TaeX) into neighbouring eukaryotic and bacterial cells<sup>19</sup>. SPI-6 contributes to gastrointestinal colonization and pathogenesis of mice<sup>20</sup> and chickens<sup>21</sup> by serovar Typhimurium. Expression of *tcf* resulted in specific adhesion of *Escherichia coli* to human

epithelial cells<sup>22</sup> and *saf* mutations reduced gastrointestinal colonization of pigs by serovar Choleraesuis<sup>23</sup>. The genes encoding *tcf* and *saf* are variably present within SPI-6 in the multiple serovars within the Para C Lineage (Extended Data 8), which might account for their differential host specificity. A parsimonious interpretation of this diversity is that an intact SPI-6 was initially acquired after Lomita branched off by horizontal transfer, followed by successive, internal deletions prior to the MRCAs of Paratyphi C and Typhisuis. Alternatively, each SPI-6 variant might represent an independent horizontal transfer event.

We dated key stages in the evolution of the Para C Lineage by a Bayesian phylogenetic approach (Fig. 3, Extended Data 9-10, Supplementary Tables 12-13; Supplementary Data 14). The substitution rate for Paratyphi C ( $9.4 \times 10^{-8}$  substitutions/site/y; CI:  $6.9 \times 10^{-8}$ - $1.2 \times 10^{-7}$ ) was not sensitive to the inclusion or exclusion of Ragna. The age of the crown group encompassing modern isolates was dated at 583 y (CI: 383-927) and its split from Ragna at 1,379 y (CI: 1,030-1,975) (Supplementary Table 12). Thus, the addition of the Ragna genome sheds light on an additional 800 years of the evolution of Paratyphi C.

The maximum age of Paratyphi C is 3,169 y, the date when it split from serovar Typhisuis. In turn, the MRCA for Paratyphi C, Typhisuis plus Choleraesuis evolved in Europe approximately 3,943 years ago (Extended Data 9, Supplementary Table 14; Supplementary Data 14). A European ancestry is consistent with the existence of Paratyphi C in northern Norway 800 years ago, and of the possible import by Europeans to Mexico before 1,545 CE of other Paratyphi C bacteria<sup>6</sup>. We also note that serovars Choleraesuis and Typhisuis infect swine, and that age of their MRCAs overlap with the Neolithic domestication of pigs from wild boar in Europe<sup>24</sup> (Fig. 3, Extended Data 9). It is therefore possible that this invasive sub-lineage evolved in swine or in boar, and that Paratyphi C represents a subsequent host adaptation to humans by a zoonotic pathogen of domesticated animals.

Our analyses indicate that our understanding of the evolutionary histories of bacterial pathogens can be substantially expanded by combining genomes from both extant and ancient bacteria. Our results also indicate that both the core and accessory genome of bacterial pathogens are remarkably stable over millennia, and that much of the dramatic variation between extant genomes represents transient genetic fluctuation whose evolutionary relevance to ecological fitness is uncertain.

## Methods

Additional details of methods used in this study can be found in the Supplementary Methods within Supplementary Information.

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized except for Bayesian analyses and the investigators were not blinded to allocation during experiments and outcome assessment.

**Dating of SK152.** SK152 was dated by two alternative methods. Archaeological evidence indicated that SK 152 was part of sub-phase FN7/level II within the extension of the cemetery, which was dated to 1175-1225 CE (Fig. 23 in citation 25). Accelerator Mass Spectrometry (AMS) dating was performed at DirectAMS and Oxford Radiocarbon Accelerator Unit, yielding dates in radiocarbon years BP (Before 1950 CE) using the  $^{14}\text{C}$  half-life of 5,568 years. Isotopic fractionation was corrected using the  $\delta^{13}\text{C}$  values measured on the AMS. The quoted  $\delta^{13}\text{C}$  values were measured independently on a stable isotope mass spectrometer (to  $\pm 0.3$  per mil relative to VPDB). For details of the chemical pre-treatment, target preparation and AMS measurement see 26,27. Extended Data 2 shows calculated calendar age ranges generated by the OxCAL computer program (v4.2) of C. Bronk Ramsey, using the 'IntCal13' dataset. These calculations support dates of 994-1052 CE with 53% likelihood and 1081-1152 CE with 43% likelihood. We summarized the two periods as  $1073 \pm 79$  years, which is the mean  $\pm$  range of the lowest and highest extremes.

**Metagenomic sequencing of samples from SK152.** All molecular work including pre-library amplification was conducted in dedicated aDNA clean laboratory facilities at the Centre for GeoGenetics, Natural History Museum, University of Copenhagen. All samples were collected and processed using strict aDNA guidelines. Nine sequencing libraries were prepared from the upper 3<sup>rd</sup> left molar root including dental pulp, the upper 2<sup>nd</sup> right molar dentine/cementum (200 mg) and pulp (interior root canal, 200 mg), femoral long bone (300 mg), and a mixture of mineralized dental plaque (calculus, 30 mg) taken from multiple teeth. Libraries were sequenced on an Illumina HiSeq in 15 runs (Supplementary Table 1).

**Genomic assemblies from metagenomic reads.** Taxonomic profiling of metagenomic reads with currently available traditional methods can yield false indications of the presence of bacterial pathogens because pathogens are over-represented in public databases relative to environmental organisms, and completed genomes of many environmental bacteria are not yet available. As a result, ecological metagenomic analyses have used *de novo* assemblies of raw reads<sup>28</sup> but this has not yet been applied in aDNA studies. We therefore investigated whether *de novo* assemblies could reveal the presence of bacterial pathogens within the nine metagenomic libraries from SK152, and used subsequent analysis of deamination rates to determine whether those assemblies were ancient or modern.

**Assembly and binning.** We excluded reads within the metagenomic libraries that mapped to the human genome according to BWA<sup>29</sup>. The non-human reads were co-assembled into contiguous sequences (contigs) with MEGAHIT<sup>30</sup>, using default parameters, and producing the assembly statistics in Supplementary Table 2 (All contigs). Contigs which exceeded 20 kb were split into 10 kb fragments. All 39,016 contigs greater than 1 kb in length were clustered by CONCOCT<sup>4</sup> into 76 bins using both sequence composition and coverage across all samples. All of the contigs in each bin were potentially derived from a single bin-specific genomic assembly. Protein encoding sequences on these contigs were called using PRODIGAL<sup>31</sup>, and then categorised in terms of function according to Clusters of Orthologous Groups of

proteins (COGs)<sup>32</sup>. We previously identified 36 single copy core COGs (SCGs) that are found in a single copy in all bacterial genomes<sup>4</sup>. We therefore tested each of the bins for the number of copies of these 36 SCGs, and found that 11 bins largely represented MAGs (unique metagenome assembled genomes) because they contained at least 27 SCGs (75% of 36) in a single copy (Extended Data 3). The contigs that belong to the 11 MAGs are available in Supplementary Data 2.

**Taxonomic assignments.** We constructed a phylogenetic tree of the 36 SCGs from 1,755 reference genomes plus the 11 MAGs. The reference genomes consisted of one representative from each bacterial genus and each archaeal species with complete genome sequences in NCBI. Each SCG was aligned separately using MAFFT<sup>33</sup>, and overhangs were trimmed with TRIMAL<sup>34</sup>. Where exceptional MAGs contained multiple SCGs we chose one of the sequence variants at random. The concatenated SCG alignments were used to construct a single tree with FASTTREE<sup>35</sup>. For taxonomic assignments, we identified the ancestral node containing each MAG plus one or more reference genomes, and the most frequent taxonomic designation among the neighbouring reference genomes was assigned to that MAG (Supplementary Table 4). C72 was initially assigned to *Mogibacterium timidum*, but only few related reference genomes existed in NCBI. We therefore compared C72 against the rMLST database<sup>36</sup> and re-assigned C72 to *Eubacterium*.

**Calculation of deamination rates.** BWA was used to map the metagenomic reads onto all contigs within each MAG, as well as to the human genome hg38 and Paratyphi C RKS4594. We then used MAPDAMAGE 2.0<sup>37</sup> to separately characterise the DNA damage associated with each of the 13 organisms. The posterior mean estimates and standard deviations for  $\delta S$ , the single-stranded deamination rate for all genomes and MAGs, are listed in Supplementary Table 5 and depicted graphically in Fig. 1D.

**Interpretation of sources of MAGs.** Nine of the MAGs likely reflect recent soil contamination because their 5'-single-stranded deamination rates were low (DNA damage <0.22) versus values of 0.47 for the human DNA and 0.9 for Ragna (Supplementary Table 5)<sup>38</sup>. The two other assembled genomes exhibited high levels of DNA damage. C72, a novel species of *Eubacterium*, was found almost exclusively in dental calculus (Supplementary Table 5), and may have been a major component of a biofilm associated with periodontal disease<sup>39</sup>. C18 belongs to *Acidovorax*, which is associated with plant pathogens<sup>40</sup>, and may have been introduced with the wood chips that covered the skeleton.

**Genotyping by Enterobase.** [Enterobase](#) automatically assembles genomes from all publicly available Illumina short reads from certain genera as well as from short reads that are uploaded by users. To this end, it uses a pipeline which depends on SPAdes<sup>41</sup> for assembly and post-assembly verification of SNP calls by remapping reads onto contigs. Assemblies that pass internal quality criteria, including a mean coverage of  $\geq 20$  fold are automatically genotyped by multiple multi-locus sequence typing (MLST) schemes into Sequence Types (STs) consisting of unique allele numbers for each genetic locus. These MLST schemes for *Salmonella* currently include a 7 housekeeping gene legacy scheme<sup>2</sup>, rMLST based on 51 ribosomal proteins<sup>36</sup>, wgMLST based on a pan-genome of 21,065 genes and cgMLST based on 3,002 core genes. Details of these schemes are available on the [Help pages](#) at Enterobase, which also describe the detailed stages involved in defining the *Salmonella* schemes.

**Species trees of *S. enterica* subspecies I.** We assembled a collection of genomes that represent the entire genetic diversity of 50,000 genomes of *S. enterica* subsp. I by choosing a



random representative of each of the 2,964 rMLST STs in Enterobase (May, 2016) (Supplementary Data 4). These genomes can be accessed at [Enterobase Workspace of Representative Strains](#). We performed super-tree analyses on the sequences of 3,002 core genes (Enterobase cgMLST v2 scheme) from these 2,964 representatives. The sequences of each core gene was aligned using MAFFT<sup>33</sup>. These alignments were used to calculate species trees by two different algorithms. i) A Maximum Likelihood tree of the concatenated, aligned sequences of all core gene sequences (2.8 Mb)(Supplementary Data 7) was generated using a GTRCAT model in RAXML v8.2.4<sup>42</sup> (Fig. 2A). ii) A Maximum Likelihood tree was also generated by RAXML for each of the 3,002 core gene alignments, and these 3,002 gene trees were processed by ASTRID<sup>43</sup>. ASTRID is a recently developed coalescent-based method for species tree estimation that is statistically consistent under the multi-species coalescent, can account for incomplete lineage sorting, is one of the most accurate currently available methods, and can handle large amounts of data. Because ASTRID only infers topology, we derived the branch lengths from sequence distances within the concatenated alignment using RAXML with a GTRGAMMA model (Supplementary Data 6). The topologies of the two trees were identical near the root and towards the tips, and at 75% of the intermediate branches. These trees were used for phylogenetic placement (MGPLACER<sup>44</sup>) of the metagenomic reads from SK152.

**Sources of genomes from the Para C Lineage (Supplementary Data 7).** The species trees identified 100 genomes in Enterobase that were in the Para C Lineage, or closely related. Most of these genomes were very recent, epidemiological isolates from the UK (PHE) and USA (FDA), but a number of them from strains sequenced by the Sanger Centre that had been tested by legacy MLST<sup>2</sup> or were from the historical Murray collection<sup>45</sup>. In order to obtain additional isolates of diverse sources and dates, we cultivated 119 old strains from the historical, global collection at the Institut Pasteur, Paris, sequenced their genomes and uploaded the short reads to Enterobase.

**Initial identification of Paratyphi C aDNA within metagenomic data.** In initial analyses, metagenomic reads from 33 human skeletons from Trondheim, Norway were screened using KRAKEN<sup>46</sup> with its default genomic database. KRAKEN scored 304 metagenomic reads from skeleton SK152 as *Salmonella*, but did not score any reads as *Salmonella* or other potentially invasive bacterial pathogens from any other skeleton. However, the taxon assignments provided by KRAKEN for reads present at low concentrations can be problematical<sup>47</sup>. In our hands, the KRAKEN analyses of SK152 failed to identify 7/11 taxa that yielded assembled MAGs with CONCOCT (Supplementary Data 1-2). We therefore tested the reads from SK152 identified as *S. enterica* by *ad hoc* BLASTN alignments to a custom database of 91,000 bacterial, archaeal and viral genomes from the non-redundant nucleotide database in GenBank (RefSeq), and confirmed that 266 (87%) scored as *S. enterica*. MGPlacer<sup>44</sup> was then used to iteratively map those *Salmonella*-specific metagenomic reads onto a core SNP phylogeny from one genome from each of the 20 *Salmonella* serovars represented in RefSeq (May, 2016). All *Salmonella*-specific reads mapped on the branch leading to serovar Paratyphi C strain RKS4594.

**Identification of Ragna-specific reads in metagenomic libraries.** Reads were pre-processed using BBMERGE and BBDUK2 from the BBMAP package<sup>48</sup>. Sequences specific to the Ragna genome were then identified by alignment against two sets of reference genomes using BOWTIE2<sup>49</sup>: i) The “ParaC” set, consisting of all 219 modern genomes in the Para C Lineage and ii) the “outgroup” set, consisting of representative of all (4,441) non-*Salmonella* bacterial genomes in RefSeq, plus the human genome hg38. We assigned sequencing reads

to Ragna if they yielded equal or better alignment scores to the ParaC set than to the outgroup, and differed from the most similar genome in the ParaC lineage by no more than 4%. Potentially duplicated reads were removed by retaining only one representative, unique Ragna-specific read sequence when multiple identical reads were identified.

**Reference based mapping of aDNA.** The Ragna-specific, unique reads were aligned against Paratyphi C reference genome RKS4594 using BOWTIE2<sup>49</sup> with the end-to-end option and analysed using SAMTOOLS/BCFTOOLS 1.2<sup>50</sup>. In order to exclude potential spurious SNPs due to deamination, the consensus base was only called on sites that were covered by at least two reads with a consensus base quality  $\geq 10$  and which were located at least 5 bases from either read end. Exceptionally, SNP calls with single coverage matching these criteria were included if the same call was provided by previously excluded 5' and 3' ends of reads.

**Reference based SNP calling in modern genomes.** Assemblies within the Para C lineage plus Birkenhead (Supplementary Data 7) were aligned against RKS4594 using LAST<sup>51</sup>, and SNPs from these alignments were filtered to remove regions with low base qualities ( $Q < 10$ ) or ambiguous alignment (ambiguity  $\geq 0.1$ ). Sites were also removed if they aligned with  $\geq 95\%$  identity to disperse repetitive regions that were longer than 100 bp (BLASTN), overlapped with tandem repeats (TRF<sup>52</sup>), or overlapped with CRISPR regions (PILER-CR<sup>53</sup>). Of the remaining SNPs, 61,451 were called in  $\geq 95\%$  of the genomes from the ParaC Lineage plus Birkenhead, and were retained as core genomic SNPs (Supplementary Data 8).

**Reconstruction of the phylogeny of the Para C Lineage.** Stretches of clustered SNPs, which are the hallmarks of homologous recombination, were identified and removed from the core genomic SNPs. Briefly, an initial phylogeny was calculated on all core SNPs using RAXML v8.2.4<sup>42</sup> under a GTRCAT model with Stamatakis ascertainment correction for invariant sites<sup>54</sup>. SNPs were assigned onto branches in that phylogeny using a maximum likelihood method with a symmetric transition model<sup>55</sup>. Clustered SNPs on each branch were identified *via* a hidden Markov model using RECHMM<sup>12</sup> and excluded from further analysis. A final RAXML phylogeny (Figs. 2B and 3; Extended Data 7 & 9; Supplementary Data 14) was then calculated as above from the remaining 49,610 core mutational SNPs.

**Date estimates for Paratyphi C and the Para C Lineage.** We performed molecular-clock analyses using three different methods, with the results summarised in Supplementary Table 12.

Bayesian phylogenetic approach (BEAST v1.8.3<sup>56</sup>). Sequence diversity was analysed using the GTR+G model of nucleotide substitution on the core mutational SNPs, with a discrete gamma distribution and six rate categories to account for rate heterogeneity across sites. To account for SNP ascertainment bias, we applied a correction that incorporated the nucleotide frequencies across all of the constant sites. For computational tractability, we analysed two subsamples each from the Para C Lineage; serovar Paratyphi C serovar (including Ragna); and modern crown serovar Paratyphi C strains (excluding Ragna). Each subsample comprised 50 random genomes, except that the Ragna sequence and the sole genome in the CS-3 cluster were deliberately included when present in the test set. Sub-samples including the Ragna genomes were run with either of two dates for Ragna (1200 CE from archaeological dates and 1073 CE from AMS calibrated ages). Thus, our Bayesian phylogenetic analyses involved 12 data sets (Supplementary Table 12 and Supplementary Data 14).

Posterior distributions of parameters, including the tree, were estimated using Markov chain Monte Carlo sampling. Samples were drawn every 5000 steps over a total of  $5 \times 10^7$  steps, with the first 10% of samples discarded as burn-in. Marginal likelihoods were used to compare three demographic models for the tree prior: constant population size, Bayesian Skyride coalescent<sup>57</sup>, and birth-death process with serial sampling<sup>58</sup>. Marginal likelihoods were estimated using stepping-stone sampling<sup>59</sup>, with 20 path steps and a chain length of  $10^6$  per path step.

Temporal structure. The date randomization test<sup>60</sup> checks for temporal structure within the data set by analysing multiple, date-randomized replicate data sets after randomly reassigning the ages of the sequences. The temporal structure is considered assured when the 95% credibility interval of the rate estimate from the original data does not overlap with any of those multiple rate estimates from the date-randomized replicates<sup>61</sup>. We used 10 date randomizations (Extended Data 10), and found strong temporal signals for all sub-samples of modern genomes analysed with BEAST.

Root-to-tip regression and least-squares dating. Both methods were used to analyse all the genomes that had been sub-sampled for Beast analyses: the Para C Lineage and the Paratyphi C serovar, each with two possible ages of Ragna, in addition to the modern Paratyphi C strains without Ragna. A root-to-tip regression analysis was performed in TEMPEST<sup>62</sup> and the significance of the slope was evaluated in R v3.0.1<sup>63</sup>. A least-squares dating was performed in LSD<sup>64</sup>. To test for the presence of temporal structure in the data set, substitution rates were also estimated from 1,000 date-randomized replicates. We verified that the rate estimate from the original data set was not contained within the range of values obtained from the date-randomized replicates.

**Inferring ancestral geographic sources.** A subtree of genomes with geographic source information was extracted from the Para C Lineage phylogeny plus the outgroup Birkenhead with the ETE3 Python package<sup>65</sup>. (Lomita is lacking from this subtree because its geographic source is uncertain). The ancestral geographic states of internal nodes in the subtree were inferred by three independent algorithms: i) the Maximum Likelihood comparison<sup>66</sup> in BayesTraits<sup>67</sup>; ii) the Markov chain Monte Carlo (MCMC) approach<sup>68</sup> in BayesTraits; and iii) Bayesian Binary MCMC in RASP<sup>69</sup>. All three algorithms indicated a European origin for the Para C Lineage (Extended Data 9, Supplementary Table 13), whether Ragna and serovar Birkenhead were included or not. In addition, very strong support for a European origin of the Para C Lineage was provided by BayesTraits MCMC and RASP, and moderate support was provided by BayesTraits ML.

**Reconstruction of the pan genome of the Para C Lineage (Supplementary Data 10).** A wide collection of annotated reference genes was collected from three sources: 1) all 21,065 unique sets of homologs in the wgMLST scheme in Enterobase; 2) published annotations for the complete genomes of Choleraesuis strains C500<sup>70</sup> and SC-B67<sup>71</sup> and Paratyphi C strain RKS4594<sup>72</sup>; 3) PROKKA<sup>73</sup> annotations of all draft genomes in EnteroBase. All these reference genes were grouped into 29,436 gene clusters using UCLUST<sup>74</sup>. In order to obtain sets of homologous regions that cover  $\geq 50\%$  of the length of the centroid sequences with  $\geq 70\%$  nucleotide identity, the centroid sequence from each cluster was aligned with all the modern genomes in the Para C Lineage using BLASTN. Overlapping paralogs between the homolog sets were identified through the same iterative methodology used to construct the entire *Salmonella* wgMLST scheme. After removal of paralogs, the remaining 6,665 homolog sets were treated as the pan genes of the Para C Lineage (Supplementary Data 10-11). The

synteny of the pan genome was reconstructed as described in Supplementary Materials of Zhou *et al.*<sup>13</sup> and can be examined in detail in a publicly accessible, interactive local version of [Anvi'o server Pan-genome](#).

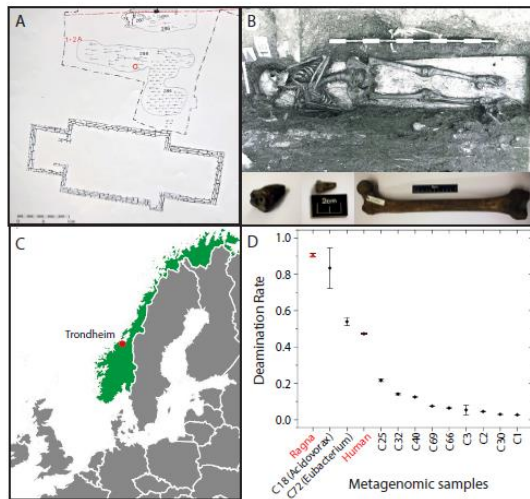
**Pseudogenes and genes gained or lost.** The gene alignments used to reconstruct the pan genome of the Para C Lineage were screened for potential gene disruptions. Genes with stop codons or frameshifts anywhere in the coding regions were scored as pseudogenes. Pseudogenes and genes that were gained or lost were assigned to ancestral states in the core SNP phylogeny (Extended Data 7) by the Maximum Likelihood algorithm<sup>55</sup> as implemented in the R function ACE (Ancestral Character Estimation) within the APE (Analyses of Phylogenetics and Evolution) package (Supplementary Data 12-13). These ancestral states were used to infer the phylogenetic branches on the SNP tree of the Para C Lineage on which 3,125 genes or parts of genes and 1,251 pseudogenes were independently gained or lost ((Extended Data 5 & 7, Supplementary Table 9B). The relative frequencies of gene gain/loss and pseudogene versus time were estimated as described<sup>12</sup>: time (years) was calculated from the core SNP phylogeny by dividing branch lengths by the SNP substitution rate estimated by BEAST ( $9.4 \times 10^{-8}$  substitutions per site per year; Supplementary Table 12). The 95% confidence intervals for these frequencies were inferred from 1,000 bootstrap resamplings of genomic islands or pseudogenes. The data and script used for these purposes can be found in Supplementary Data 15.

**Identification of bacteriophages.** Prophages in modern genomes of the Para C Lineage were identified using the online API in PHASTER<sup>75</sup>. The identified prophages were taxonomically labelled according to similarities of their major capsid protein (MCP) to the clusters of bacteriophages described by Casjens *et al.*<sup>76</sup>. Further manual taxonomic refinements were based on comparisons of all genes in each prophage against those in the published phage clusters. Prophage names, including SPC-P1, are the same as those of previously described phages which they most closely resembled except for GI28, which has no known close relatives.

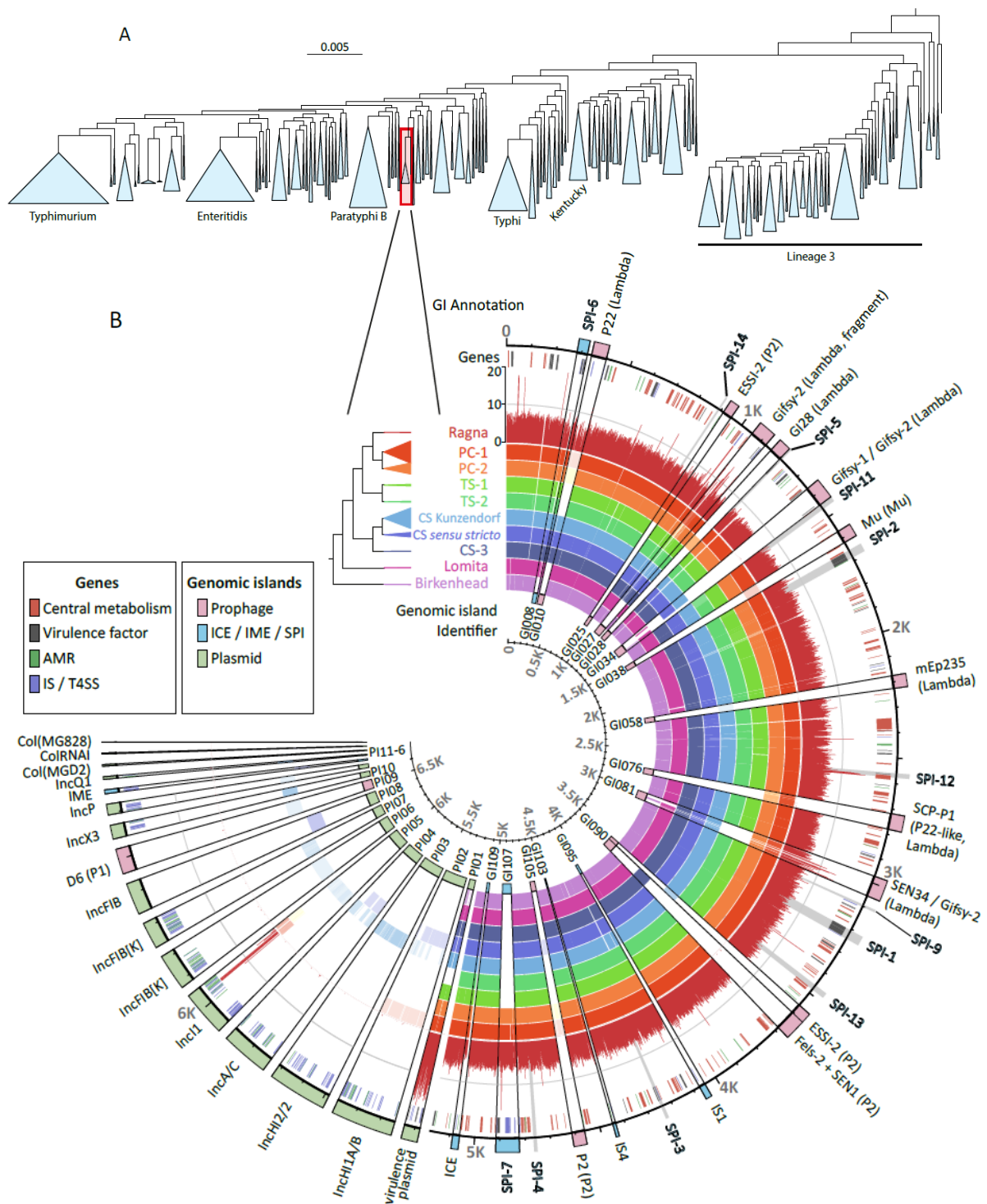
**Annotation of genomic islands.** BLASTN was used to identify known genomic islands within the pan genome as tight clusters of genes in which  $\geq 60\%$  of the sequences aligned with a previously described SPI/SGI sequence with  $\geq 80\%$  nucleotide identity. The genomic sequences used as references for SPI-1 through SPI-12 were downloaded from PAIDB<sup>77</sup>, as were SGI-1 and SGI-2. SPI-13 through SPI-21 were obtained from citations 78-82. Genes in the pan genome were also screened for an association with mobile elements against ISFINDER (IS elements)<sup>83</sup>, PLASMIDFINDER 1.3 (incompatibility groups of plasmids)<sup>84</sup> and CONJSCAN-T4SSCAN (relaxases and key components of type IV secretion systems)<sup>85</sup>. Strong matches were annotated according to those resources (Fig. 2B; Supplementary Data 10 & 13). These analyses revealed that genes gained or lost were associated with at least 227 genetic islands belonging to 127 distinct categories (Supplementary Table 9B).

**Data availability.** Genomic reads for 119 strains from the Institut Pasteur collection have been deposited in the NCBI short read archive under project accession PRJEB19916. All genomes referred to here are available from the publicly available workspaces entitled "rST representatives" (2,964 genomes) and "Para C Lineage" in the *Salmonella* database of Enterobase (<http://Enterobase.warwick.ac.uk>). Interactive versions of Fig. 2 and Extended Data 8 can be found at [https://enterobase.warwick.ac.uk/anvi\\_public/ParaC\\_pangenome](https://enterobase.warwick.ac.uk/anvi_public/ParaC_pangenome) and [https://enterobase.warwick.ac.uk/anvi\\_public/ParaC\\_SPI6](https://enterobase.warwick.ac.uk/anvi_public/ParaC_SPI6). Data supplements are

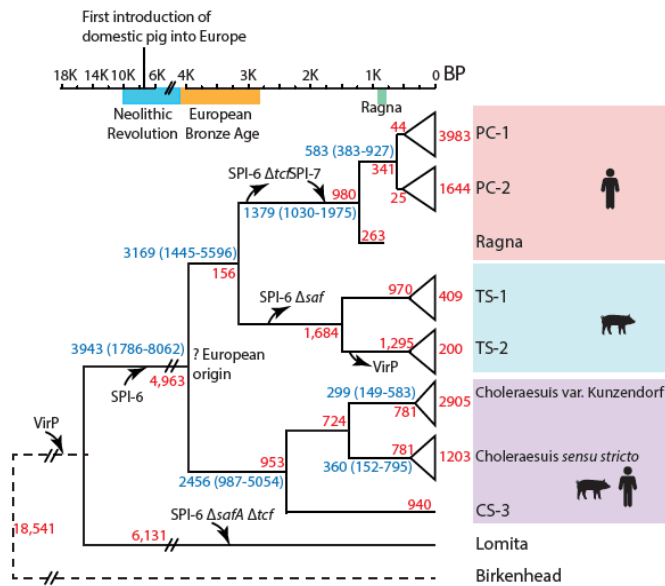
permanently stored at the University of Warwick (<http://wrap.warwick.ac.uk/85593>) and non-human metagenomic data at [https://sid.erda.dk/wsgi-bin/lis.py?share\\_id=E56xgi8CEI](https://sid.erda.dk/wsgi-bin/lis.py?share_id=E56xgi8CEI).



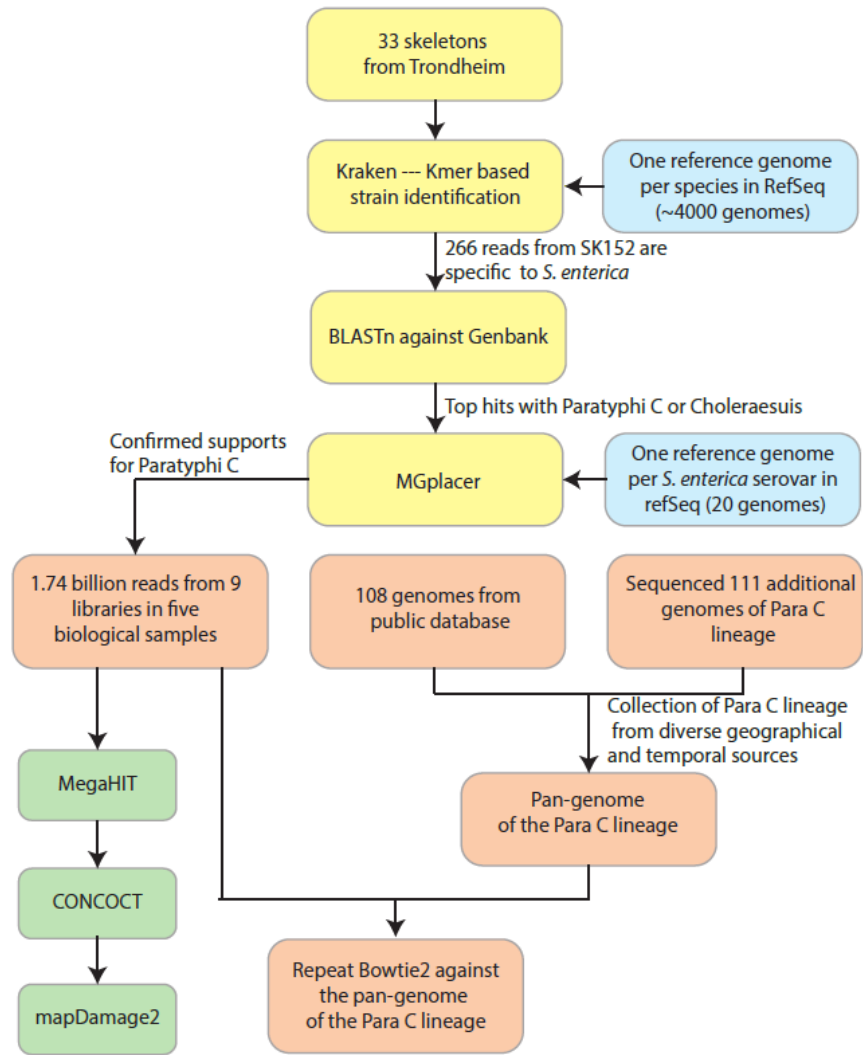
**Figure 1.** Geographic, archaeological and metagenomic features of skeleton SK152. A) Excavation site (Folkebibilotekestomten, 1973-1985) of the church cemetery of St. Olav in Trondheim, Norway. The burial location of SK152 (red circle) belongs to a building phase which has been dated archaeologically<sup>3</sup> to 1200 CE (range 1175-1225). B) Entire skeleton (top) and femoral long bone plus two teeth from which *Salmonella* DNA was extracted (bottom). The skeleton rested on a wooden plank (a symbolic coffin) in a grave filled and covered with anoxic, acidic and waterlogged wood chips and soil. C) Map of Europe surrounding Norway (green). D) Deamination rate for metagenomic reads in the *Salmonella* Paratyphi C Ragna genome, human DNA and 11 single genome assemblies (Cxx) identified by CONCOCT<sup>4</sup>. C18 (*Acidovorax*) and C72 (*Eubacterium*) show high levels of deamination rates as do reads from humans or Ragna, while the other assemblies likely represent modern environmental bacteria.



**Figure 2.** Genomic phylogenies of *S. enterica* and the Para C Lineage. A) Maximum-likelihood phylogeny of 2,964 representative genomes of *S. enterica* subspecies *enterica*. Each genome is a random representative of one ribosomal multi-locus sequence typing (rMLST) sequence type. The Para C Lineage and other lineages containing numerous genomes are collapsed into blue triangles, and several common serovars that are associated with those lineages are indicated under the tree. Red rectangle: Para C Lineage plus the outgroup Birkenhead. B) Pan-genomic contents for 222 genomes of the Para C Lineage, including Ragna, plus Birkenhead. Genes are ordered as in Supplementary Data 10-11 (primary ticks every 500 genes). Circles (inner to outer): Circle 1. Sixteen major, variably present chromosomal genomic islands (GI008-GI109; Supplementary Data 10) followed by sixteen cytoplasmic elements (PI01-PI16), colour-coded as in the right key. Circles 2-10: The frequency of presence or absence of each gene per sub-lineage within the phylogram is indicated by colour opacity. Circle 11: Coverage of aDNA reads per gene within Ragna (Scale 0-20 at 12:00). Circle 12: Genes color-coded as in the left key. Circle 13: Traditional designations of GIs, PIs and other variably present genomic elements, such as SPIs (grey wedges).



**Figure 3.** Cartoon of the evolutionary history of the Para C Lineage on the historical timeframe (BP: before present) for human history in Europe (top). The phylogram is annotated with the acquisitions of the virulence plasmid (VirP), SPI-6 and SPI-7 by inward arrows and deletions of parts of SPI-6 by outwards arrows (see Extended Data 8). Numbers in blue indicate date estimates and their 95% credible intervals (parentheses) according to a Bayesian phylogenetic approach (Supplementary Tables 12-13). Red numbers indicate numbers of non-recombinant core single nucleotide variants. The host specificities of the individual serovars are indicated by cartoons at the right.

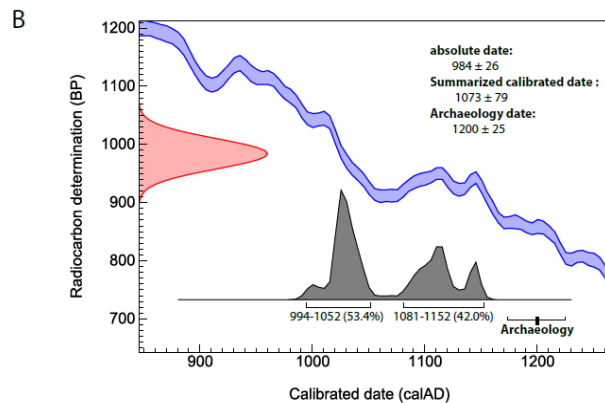


**Extended Data 1.** Overview of the workflow used in this study.

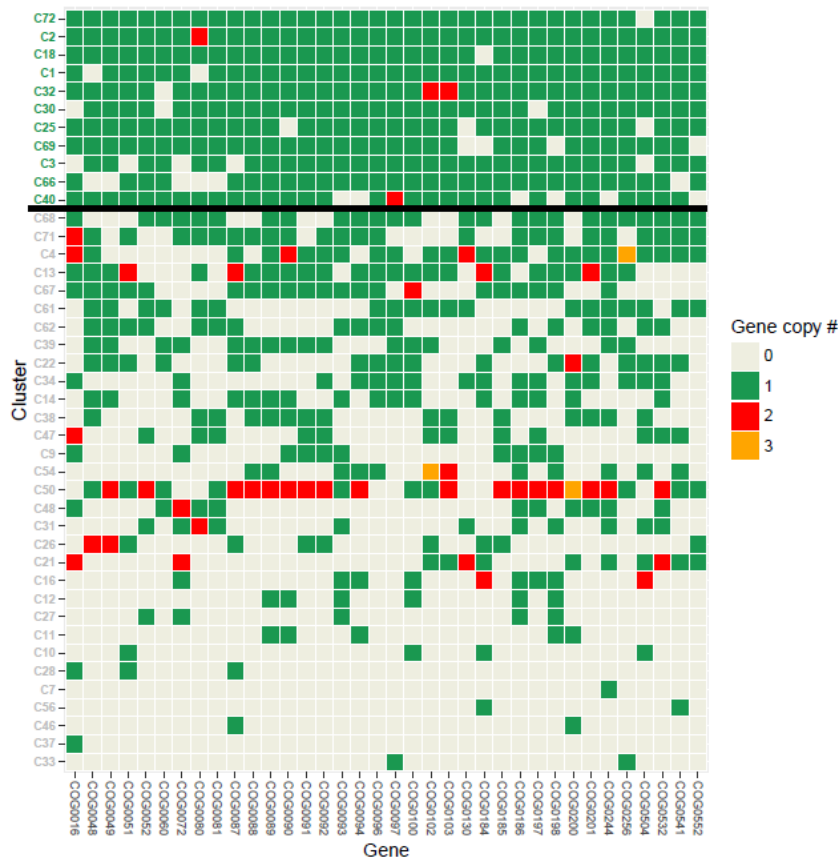


A

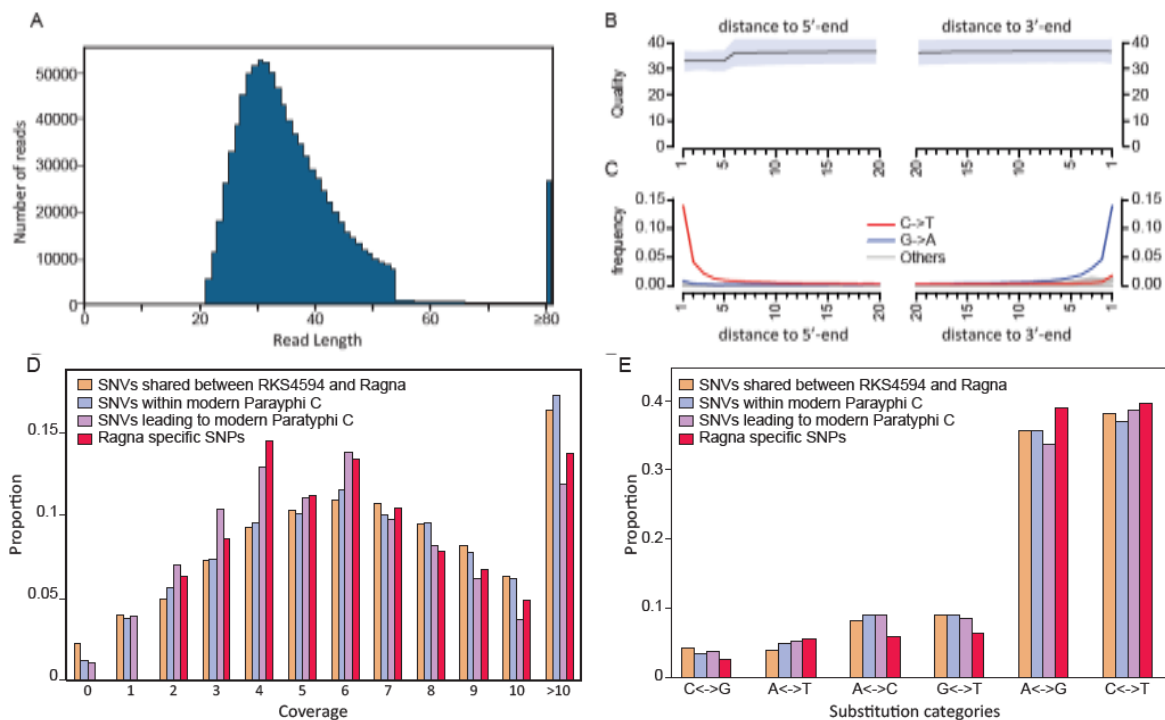
Skeleton number	SK152
Sex	Female
Age (y)	19-24
Height (cm)	155.4 ± 4.5
Date (Archaeology)	1200 ± 25 CE
AMS result (Oxford)	1073 ± 79 CE
AMS result (DirectAMS)	911 ± 30 CE
possible birth place ( $\delta^{18}\text{O}_{\text{Carbon}}$ from first molar)	northernmost inland areas of Scandinavia / Northwest Russia
travel history ( $\delta^{18}\text{O}_{\text{Carbon}}$ from third molar)	travelled south at age of 3-14 and arrived Trondheim in early teens



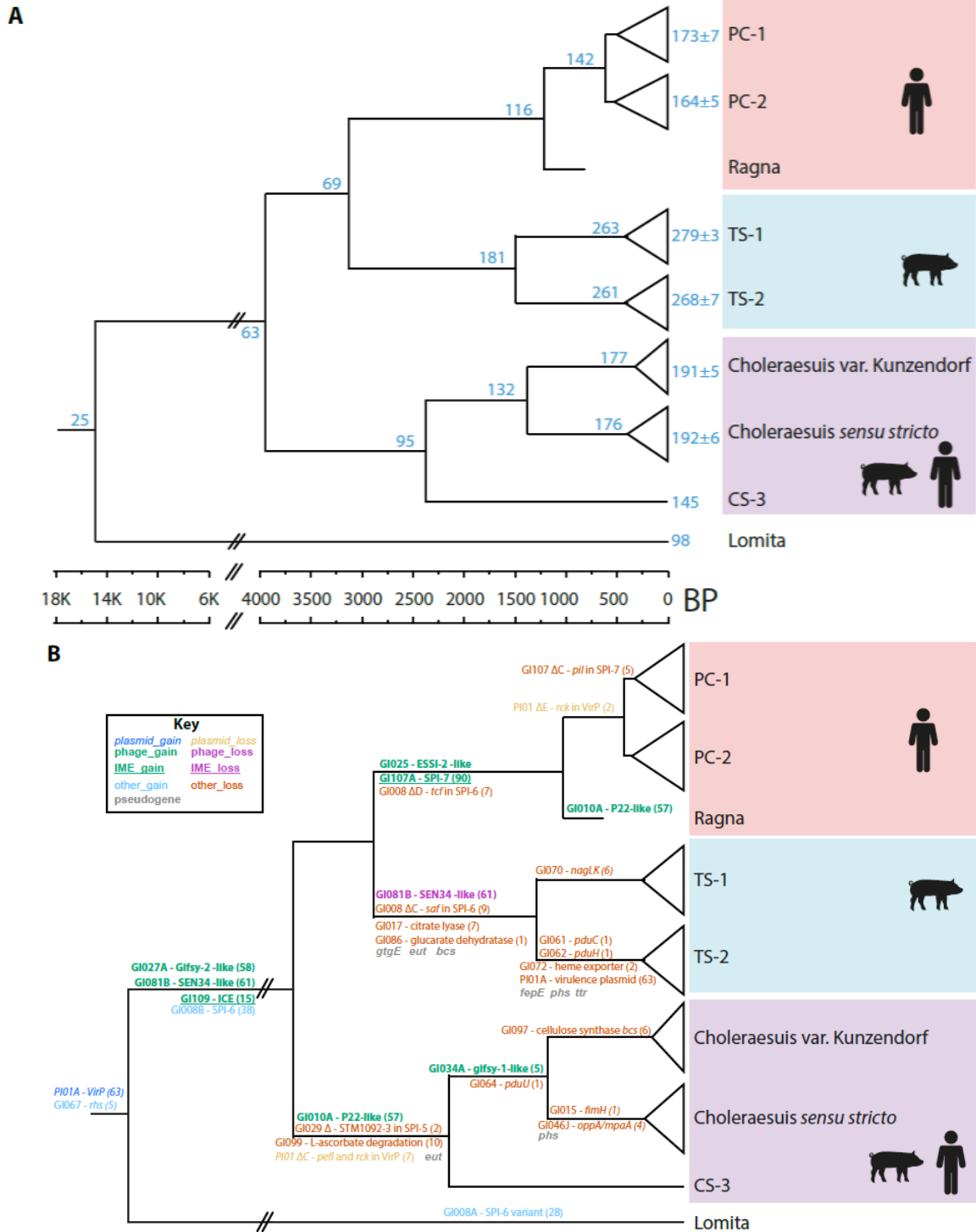
**Extended Data 2.** Archaeological findings for SK152. A) Summary of detailed information for SK152. B) Graphic illustration of the difference between Radiocarbon ( $^{14}\text{C}$ ) dating of tooth (from Oxford) and archaeological dating. The red curve on the Y axis shows the uncalibrated  $^{14}\text{C}$  estimate of years before present. The blue line is the calibration curve, and the grey curves on the X-axis are the calibrated dates according to that calibration curve. The summarized  $^{14}\text{C}$  date is the mean value of two extremes of the calibrated dates with a wide confidence interval of 79 years. The age of SK152 inferred from archaeological interpretations of the building phase of the cemetery are shown at the lower right (Archaeology).



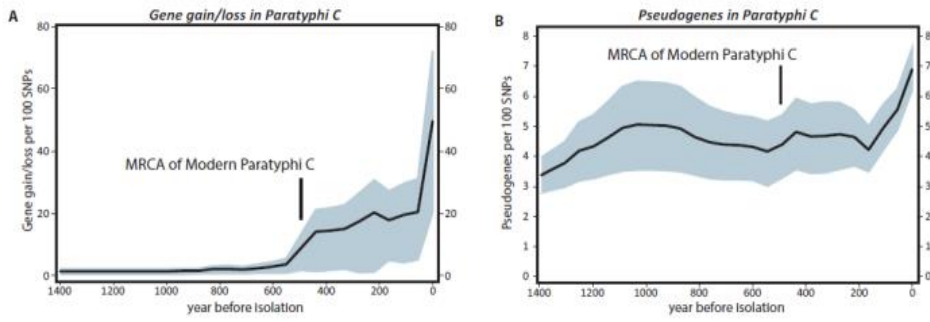
**Extended Data 3.** Single core gene (SCG) frequencies in the 76 CONCOCT clusters generated after binning contigs  $\geq 1$  kb. Each row corresponds to a cluster and columns are SCGs. The top 11 clusters, delineated by a thick black line, included  $\geq 75\%$  of the 36 SCGs. Gene copy number is indicated by the key.



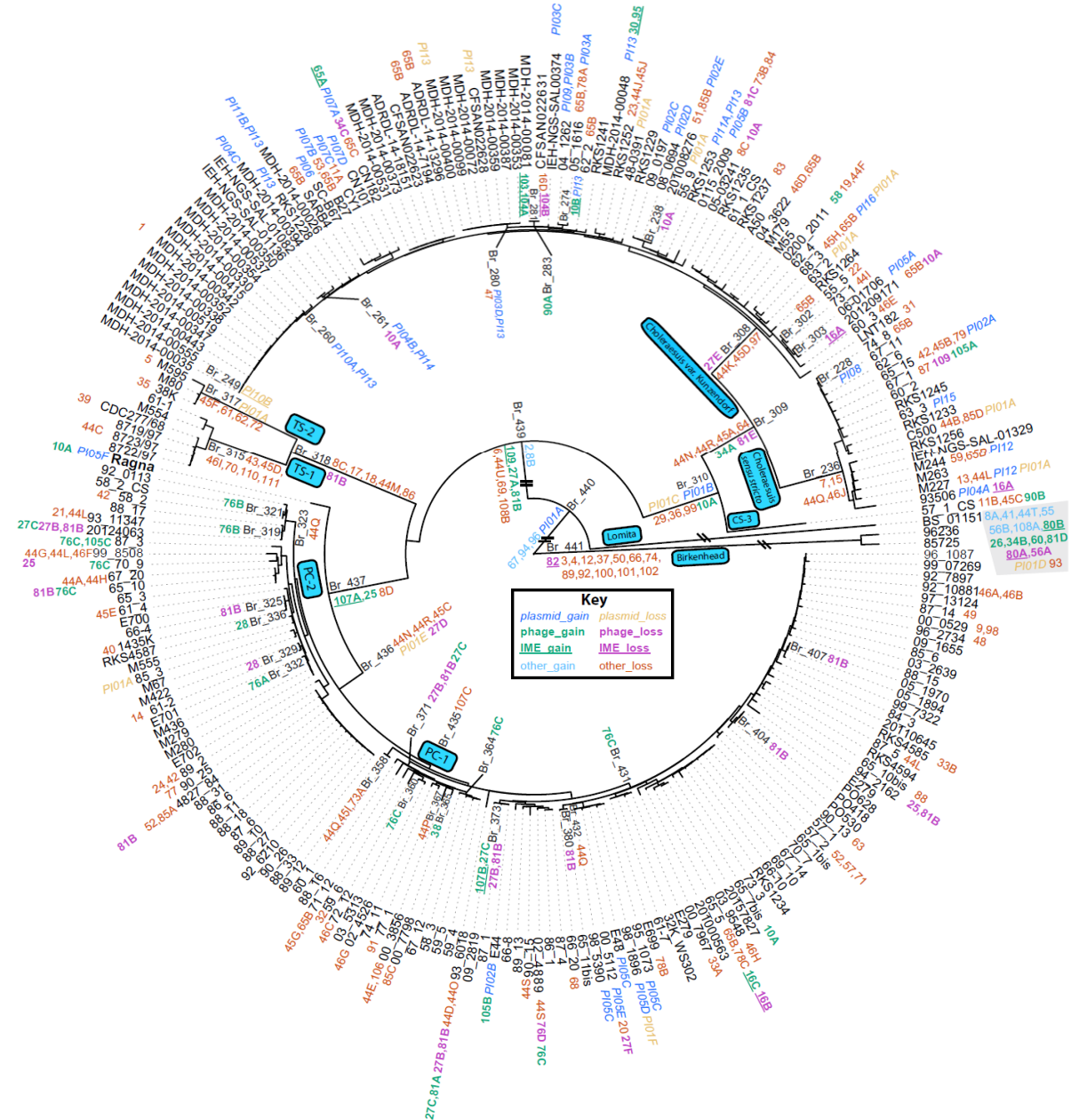
**Extended Data 4.** Summary of aDNA reads specifically aligned onto Paratyphi C (RKS4594) (A-C) and SNPs called in Ragna (D, E). A) Distribution of the reads lengths. The spike at 81 bps reflects reads that originated from single-end sequencing libraries. B) The average (black) and standard deviation (gray shadow) base qualities at both ends of the reads. C) The deamination rate at both ends of the mapped reads. D) Distribution of read coverages in different categories of SNPs (see key). E) The distribution of nucleotide composition in different categories of SNPs (see key).



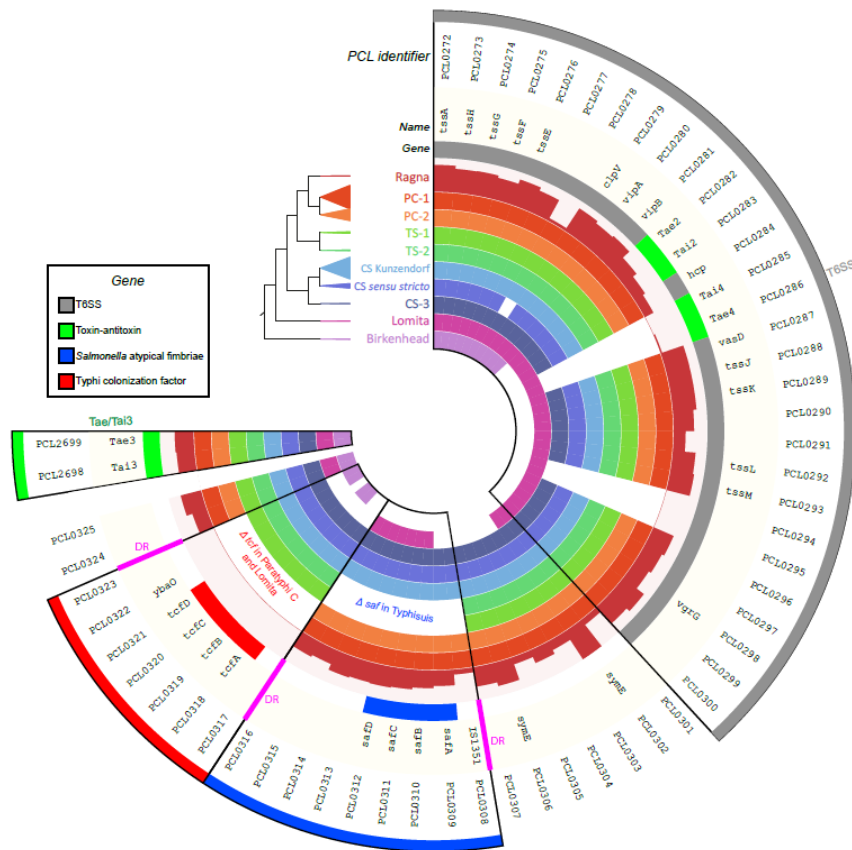
**Extended Data 5.** Cartoon trees of the changes in pseudogenes (A) and genomic islands (B) within deep branches of the Para C Lineage scaled with dates before present (BP) as estimated with BEAST. Sub-lineage designations are at the right and the existence of multiple genomes within a sub-lineage is indicated by a collapsed triangle at the tip of the tree. Host specificities of the serovars are indicated by cartoon characters at the far right. A) The numbers in light blue above branches indicated the total number $\pm$ Std. deviation of pseudogenes accumulated since the root of the tree (details in Supplementary Data 12). B) Numbers in parentheses indicate the numbers of genes affected by each change (details in Supplementary Data 13). Partial deletions in the PI01 (virulence plasmid VirP) and GI008 (SPI-6) are indicated by "Δ" followed by letters for distinct deletions. Colours in the Key indicate the types of GIs and gain/loss.



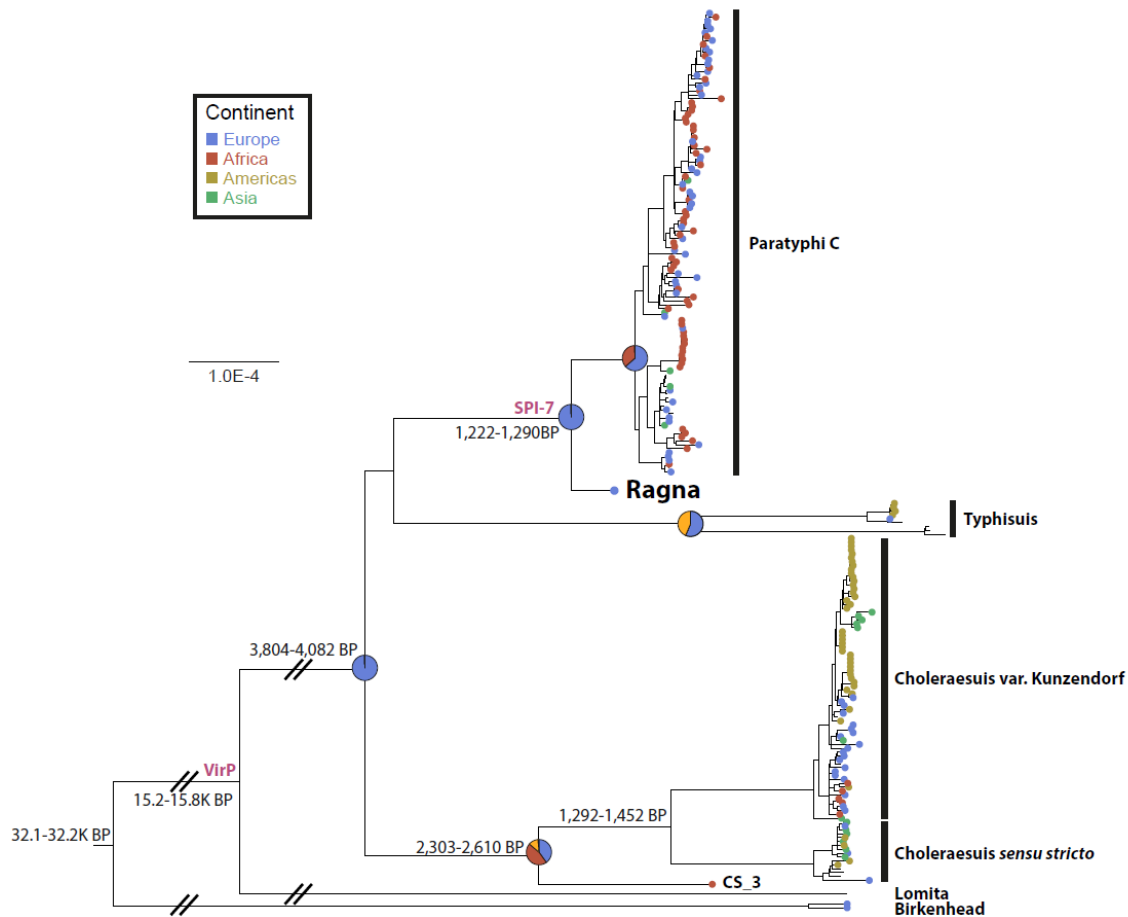
**Extended Data 6.** Frequencies of gene gain/loss (A) and pseudogenes (B) over time. Details are as described in Methods (Pseudogenes and genes gained or lost) and discussed under Supplementary Discussion (Long term stability of *Salmonella* genomes). The black lines show mean values and are flanked by CI 95% (blue).



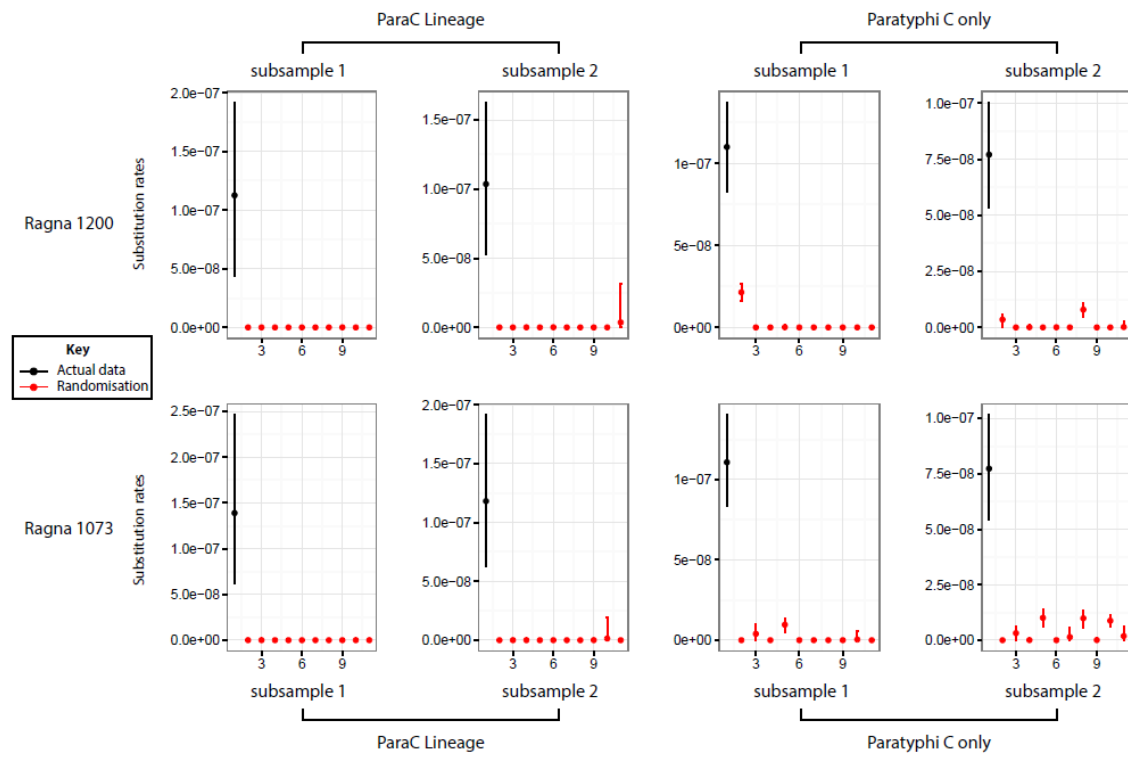
**Extended Data 7.** Maximum-likelihood radial phylogeny derived from 49,610 non-recombinant, non-repetitive core SNPs within the Para C Lineage plus Birkenhead. Internal branch numbers are indicated in black (BR\_XXX) while internal nodes for major sub-lineages are indicated by cyan rounded rectangles. Other numbers refer to gain or loss of genomic islands (42 for GI042, etc.) or plasmids (P1xx) in the accessory genome coloured according to the key (see Supplementary Data 10).



**Extended Data 8.** SPI-6 gene within the Para C Lineage plus Birkenhead. All T6SS genes within SPI-6 were present throughout the Para C Lineage but *tssL*, *tssM* and *vgrG* are absent in serovar Birkenhead, suggesting that a new version of SPI-6 was acquired by horizontal gene transfer (HGT) during the evolution of the ancestor of the Para C Lineage. In that event, the entire *tcf* and *saf* clusters have been deleted in Paratyphi C plus Lomita and in Typhisuis, respectively. These deletions are likely due to excision via bounding direct-repeats (DR, purple) that are associated with IS1351. SPI-6 in Lomita may also represent an independent acquisition by HGT: the termini of the deletion of the *saf* genes in Lomita differed from the termini in Typhisuis, and a *Tae4/Tai4* toxin/antitoxin T6SS effector is found exclusively in Lomita's SPI-6. Furthermore, unlike other members of the Para C Lineage, the SNP density in Lomita's SPI-6 was fourfold higher (0.015) than in the core genome. An interactive version of this figure is also available at Anvi'o server SPI-6



**Extended Data 9.** Maximum likelihood phylogeny and geographical origins derived from 49,610 SNPs in the non-recombinant, non-repetitive core genome of the Para C Lineage plus Birkenhead. Filled circles at the tips are colour-coded according to the source continent (see Continent key). Pie-charts on internal nodes show the geographical inference estimates of the likelihoods of continental sources according to standard MCMC in BayesTraits (Supplementary Table 13). The position of acquisition of SPI-7 by Paratyphi C is indicated at an arbitrary position on the basal branch for that serovar. The numbers on branches indicate the estimated dating estimates by BEAST from independent subsamples of the Para C Lineage (Supplementary Table 14), except that the dates for the two deepest nodes were extrapolated from deeper trees (Fig. 2A) that included representatives of all *S. enterica* subspecies I. Note that the basal branches are discontinuous, and include breaks. Scale bar: substitution frequency.



**Extended Data 10.** Date randomization tests for the temporal signal in multiple replicate data sets that were analysed by BEAST (Supplementary Tables 10-11). Solid circles represent mean estimates of substitution rates, and errors bars show their 95% credible intervals. Black symbols were estimated with real data and red symbols were obtained from 10 randomizations of dates per data set.

## Résumé vulgarisé

Découverte il y a plus d'un siècle, *Salmonella*, bactérie responsable de diarrhées d'origine alimentaire appelées salmonelloses, n'a cessé d'intriguer les chercheurs. Sa capacité à résister à de nombreux antibiotiques est de plus en plus préoccupante. Sa surveillance repose sur un dépistage rapide et spécifique de façon à identifier le plus précocement possible les sources alimentaires contaminées. Les méthodes classiques sont longues, lourdes et non automatisables. Comprendre l'émergence et l'évolution des *Salmonella* est la clé pour éradiquer ce pathogène. Au cours des dernières décennies, des progrès spectaculaires ont été menés dans le monde de la biologie, avec notamment, l'arrivée de nouvelles machines permettant de scanner et d'analyser dans les moindres détails les bactéries, passant du traitement d'une dizaine à des centaines de millions de données. L'accès facilité à ces données et aux outils qui leurs sont dédiés sont devenus une nécessité. Nous avons voulu lors de ce travail, montrer l'intérêt de ces scanners, pour l'étude de *Salmonella*. **(1)** Après avoir analysé plus de 300 bactéries de *Salmonella*, nous avons mis au point une technique de dépistage des épidémies à l'aide d'outils informatiques (et non plus chimiques) d'un type prédominant de *Salmonella* appelé Typhimurium, basé sur des structures répétées de la bactérie. La surveillance biologique à haut débit des salmonelloses a été validée en routine sur plus de 800 bactéries. L'étude évolutive entre les structures répétées et les mutations bactériennes ont permis d'établir une nomenclature définissant les différentes populations de Typhimurium. **(2)** L'analyse de 280 bactéries historiques a montré que les bactéries résistantes à un antibiotique appelé ampicilline étaient déjà répandues chez Typhimurium à la fin des années 1950, bien avant sa commercialisation et son utilisation. La présence d'antibiotiques dans le milieu agricole où ces composés ont été utilisés en tant que promoteurs de croissance a pu conduire à l'émergence des premières bactéries résistantes aux antibiotiques. **(3)** L'étude des liens de parentés entre une bactérie prélevée sur un cadavre d'une jeune femme décédée il y a plus de 800 ans en Norvège avec plusieurs bactéries récentes a montré que ces bactéries étaient très similaires malgré leurs années d'écart. Ainsi, l'étude combinée entre la biologie et l'informatique a accru nos connaissances sur l'évolution de ce pathogène.

**Mots clés : analyse bactérienne par un scanner, surveillance épidémiologique, structure répétée, mutation, résistance antibiotique, lien de parenté, évolution**





**Title : Genomic and epidemiology of *Salmonella***

**Keywords : Whole genome sequencing, epidemiological monitoring, CRISPR, SNP, antibiotic resistance, phylogeny, evolution**

**Abstract :** Over a century has passed since the discovery of *Salmonella* and yet, this pathogen still intrigues researchers. Its ability to withstand many antibiotics is of increasing concern. The monitoring of this pathogen is based on a rapid and discriminatory typing to identify the sources of contaminated food as early as possible. The conventional methods are long, heavy and non-automatable. Understanding the emergence and evolution of *Salmonella* is the key to eradicate this pathogen, which has remained one of the leading causes of foodborne bacterial diarrhea in the world. During the last decades, spectacular progress has been made in the world of microbiology with the arrival of workbench sequencers, passing from a dozen to hundreds of millions of sequences processed. Facilitated access to numerous genome sequences and dedicated tools are mandatory. Tools currently available are not sufficiently discriminating for the subtype of *S. enterica* serotype Typhimurium, a predominant serotype of *Salmonella*. Throughout this study, we showed the interest of whole genome sequencing, a multidisciplinary tool, for the genomic study of *Salmonella*. **(1)** After sequencing over 300 *S. enterica* serotype Typhimurium genomes, we have developed an *in silico* subtyping tool for this serotype, based on the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) polymorphism. High-throughput microbiological monitoring of salmonellosis has been routinely validated on over 800 genomes. The study of coevolution between the chromosome (SNPs of the core genome) and the two CRISPR regions made it possible to establish a nomenclature defining the different populations of this serotype. **(2)** Genomic analysis of 280 historical strains of *S. enterica* serotype Typhimurium showed that plasmids carrying  $\beta$ -lactamase genes, which confer resistance to ampicillin, were widespread within this serotype in the late 1950s, years before ampicillin was first used for clinical purposes. The presence of penicillin G in the farming environment where these compounds were used as growth promoters, may have led to the selection of the first ampicillin-resistant strains. **(3)** The phylogenetic study of a genome from the corpse of a young woman who died over 800 years ago, probably due to enteric fever, and 219 historical and recent genomes of the serotypes Paratyphi C, Choleraesuis and Typhisuis have shown, despite the differences in host specificity, that their genomes were very similar over the past 4000 years. Thus, the combination of genotypic and phylogenetic approaches has increased our knowledge of the evolution of this pathogen.

**Titre : Génomique épidémiologique de *Salmonella***

**Mots-clés : Mots clés : Séquençage entier du génome, surveillance épidémiologique, CRISPR, SNP, résistance antibiotique, phylogénie, évolution**

**Résumé :** Découverte il y a plus d'un siècle, *Salmonella* n'a cessé d'intriguer les chercheurs. Sa capacité à résister à de nombreux antibiotiques est de plus en plus préoccupante. La surveillance de ce pathogène repose sur un typage rapide et discriminant de façon à identifier le plus précocement possible les sources alimentaires contaminées. Les méthodes classiques sont longues, lourdes et non automatisables. Comprendre l'émergence et l'évolution des *Salmonella* est la clé pour éradiquer ce pathogène resté l'une des premières causes de diarrhées bactériennes d'origine alimentaire dans le monde. Au cours des dernières décennies, des progrès spectaculaires ont été menés dans le monde de la microbiologie avec l'arrivée des séquenceurs de paillasse, passant du traitement d'une dizaine à des centaines de millions de séquences. L'accès facilité aux séquences génomiques et aux outils qui leurs sont dédiés sont devenus une nécessité. Les outils actuellement disponibles ne sont pas assez discriminants pour sous-typier *S. enterica* sérotype Typhimurium (STM), sérotype prédominant de *Salmonella*. Nous avons voulu lors de ce travail, montrer l'intérêt du séquençage entier du génome, pour l'étude génomique de *Salmonella*. **(1)** Après avoir séquencé plus de 300 génomes de STM, nous avons mis au point un outil de sous-typage *in silico* de ce sérotype, basé sur le polymorphisme des CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*). La surveillance à haut débit des salmonelloses a été validée en routine sur plus de 800 génomes. L'étude de la coévolution entre le chromosome (SNPs) et les régions CRISPR ont permis d'établir une nomenclature définissant les différentes populations de STM. **(2)** L'analyse génomique de 280 souches historiques de STM a montré que les gènes de bêta-lactamase conférant une résistance à l'ampicilline et portés par des plasmides étaient répandus chez STM à la fin des années 1950, bien avant l'utilisation de cet antibiotique. La présence de la pénicilline G dans le milieu agricole où ces composés ont été utilisés en tant que promoteurs de croissance a pu conduire à la sélection des premières souches résistantes à l'ampicilline. **(3)** L'étude phylogénétique d'un génome issu du cadavre d'une femme décédée il y a plus de 800 ans, probablement à cause de la fièvre entérique et de 219 génomes historiques et récents des sérotypes Paratyphi C, Choleraesuis et Typhisuis ont montré que leurs génomes étaient très similaires au cours des 4000 dernières années. Ainsi, la combinaison des approches génotypique et phylogénétique a accru nos connaissances sur l'évolution de ce pathogène.