Typage et sous-typage du genre Salmonella basé sur le polymorphisme des loci CRISPR

Laetitia Fabre

NNT: 2017IAVF0002

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 : Microbiologie

École doctorale n°581
Agriculture, alimentation, biologie, environnement et santé (ABIES)

par

Laetitia FABRE

Typage et sous-typage du genre *Salmonella* basé sur le polymorphisme des loci CRISPR

Directeur de thèse : François-Xavier WEILL

Thèse présentée et soutenue à Paris, le 19 janvier 2017

Composition du jury :

M. Guillaume ARLET, Professeur des universités-praticien hospitalier, Université Paris 6	Président
Mme Alexandra AUBRY, Professeur des universités-praticien hospitalier, Université Paris 6	Rapporteur
Mme Catherine BRANGER, Maître de conférences des universités-praticien hospitalier Université Paris 7	Rapporteur
Mme Anne BRISABOIS, Directrice de recherche, ANSES	Examinatrice
M. François-Xavier WEILL, Directeur de recherche, Institut Pasteur	Directeur de thèse

Institut Pasteur
28 rue du Dr Roux 75724 PARIS cedex 15

A mes parents

A ma grand−mère Suzanne

A mon grand−père Maurice

Aux familles Fabre et Lefèvre

REMERCIEMENTS

Je tiens tout d'abord à remercier le Docteur François−Xavier Weill, qui a su me faire confiance pour réaliser ce travail, mais surtout pour m'avoir permis d'évoluer tout au longde ces années. Le chemin a été long et parfois éprouvant mais c'est ainsi que l'on gagne en rigueur et en ténacité. Je vous suis sincèrement reconnaissante pour tout ce que j'ai appris en travaillant avec vous.

Je tiens à exprimer mes plus sincères remerciements aux membres du jury qui me font l'honneur d'évaluer ce travail : le Professeur Guillaume Arlet en tant que président du jury, le Docteur Catherine Branger et le Professeur Alexandra Aubry en tant que rapporteurs et le Docteur Anne Brisabois en tant qu'examinatrice.

Je tiens aussi à remercier Eduardo Rocha, tuteur de cette thèse, pour son soutien et ses conseils.

Merci au Professeur Patrick Grimont, ce fut un honneur de travailler avec vous.

Docteur Simon Le Hello, co−chef, co−thésard et parrain de thèse. Merci pour ton soutien, ta patience, tes encouragements, ta patience, ton amitié, ta patience, ton humour, ta patience.

Il n'y a pas que le travail dans la vie ; et pourtant … Il prend une assez grande place dans ma vie pour avoir à remercier les personnes suivantes :

Isabelle «Isa» Carle. Comment te dire… Le mot merci n'est pas assez fort pour exprimer tout ce que je te dois et à quel point je te suis reconnaissante. Tes encouragements, ton optimisme sans limite m'ont aidée à ne pas baisser les bras et à y croire jusqu'au bout.

J’aurais presque pu te dédier cette thèse ; il faudra continuer à te contenter de mon amitié désolée. Merci. Vraiment.

Marie « Madaaaaaame » Accou. Merci pour ton soutien, tes encouragements et ton amitié, sans oublier mon premier fan Clément !

Alicia « minime » Tran Dien. The best minion ever. The little voice in my head thanks to which I never feel alone. And for so much more, thank you sooooooooooooooooo much.

Un grand merci à tous les amis du yéti en particulier Alexis, Thierry, Georges, Viguinie, Laure.

Renaud. Je te suis vraiment reconnaissante pour tout ce que tu as fait pour moi. Pour avoir toujours cru en moi. Et pour tant d’autres choses, merci.

Cela n’a pas été toujours facile de jongler entre les études et le travail alors un grand merci à mes collègues Véronique−Princesse Véro, Chrystelle−Kiki, Sylvie−Dame Snuth, Corinne−Dame Cruckly, Monique, Martine L. et les « mamies » Françoise−Doudouce et Martine G. Merci à toute l’équipe BPE de m’avoir rendu la tâche plus facile et de m’avoir soutenue, aussi bien sur le plan pratique que sur le plan personnel.

Et parce qu’il n’y a effectivement pas que le travail dans la vie,

Merci à Loïc−Lolo, Catherine−Cat, Gildas, Nicolas−Kitch, Eurydice−Didice pour leur amitié et leur soutien depuis tant d’années.

Titre : **Typage et sous-typage du genre *Salmonella* basé sur le polymorphisme des loci CRISPR**

Mots clés : *Salmonella*, Typage moléculaire, CRISPR

RESUME

Les salmonelloses sont la première cause de diarrhées bactériennes d'origine alimentaire dans les pays industrialisés, demeurant un problème de santé publique majeur. Depuis près d'un siècle, le sérotypage reste la méthode de référence pour l'identification des salmonelles. Basée sur la recherche d'antigènes (un antigène de paroi et un ou deux antigènes flagellaires), cette technique manuelle est longue (3 à 6 jours) et très coûteuse (salaire et utilisation de >150 sérums commercialisés de typage). Près de 2 600 sérotypes ontété décrits chez le genre *Salmonella*. Leur distribution étant la base des investigations épidémiologiques, une identification rapide est primordiale. Par ailleurs, du fait de la prédominance de certains sérotypes, une étape supplémentaire de sous−typage (électrophorèse en champs pulsé−PFGE) est nécessaire pour une meilleure discrimination etla mise en évidence de clones épidémiques (5 jours en plus du délai du sérotypage). Une nouvelle méthode plus rapide, robuste avec des résultats facilement interchangeables entre laboratoires est nécessaire. C'est dans cette optique que nous avons commencé l'étude du polymorphisme de la région CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) chez *Salmonella*. Notre étude portant sur 134 sérotypes a mis en évidence la corrélation entre sérotype et dans certains cas sous−types et profil CRISPR. Ce travail nous a également permis de développer une nouvelle méthode de sous−typage du sérotype Typhimurium (CRISPOL) ainsi qu'une PCR spécifique pour la détection rapide des sérotypes Typhi et Paratyphi A et une autre pour l'identification du clone ST198 du sérotype Kentucky.

Le pouvoir discriminant du typage CRISPR offre une nouvelle perspective pour la surveillance des salmonelloses avec une méthode rapide et robuste et qui pourrait parfaitement se substituer aux méthodes de références que sont le sérotypage et le PFGE.

Title: Development of a novel molecular method for *Salmonella* typing and subtyping based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)polymorphism

Keywords: *Salmonella*, molecular typing, CRISPR

ABSTRACT

Salmonella is the first causative agent of bacterial diarrhea due to food poisoning in developed countries and is still considered as a major public health problem. For almost a century, serotyping has been the gold standard method for *Salmonella* identification. Based on the identification of somatic and flagellar antigens, this manual technique remains time consuming (3−6 days) and expensive (requires the use of > 150 manufactured antisera). About 2 600 serotypes have been described within the *Salmonella* genus. Their distribution being the basis for epidemiological investigations, a fast and accurate identification is essential. Moreover, due to the predominance of particular serotypes, an additional step of subtyping (Pulse−Field Gel Electrophoresis−PFGE) is necessary for a better discrimination and the identification of epidemic clones (5 more days after serotyping). In order to optimize epidemiological investigations, we need to develop a new high throughput method for fast and accurate typing/subtyping, which led us to study the polymorphism of the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) region in *Salmonella*. Our study on 134 serotypes highlighted the strong correlation between spacer content and serotype/subtype. We also developed a high−throughput method to subtype serotype Typhimurium (CRISPOL), a new PCR method for fast identification of serotypes Typhi and

Paratyphi A and a novel PCR method for fast detection of ST198 population of serotype Kentucky.

The discriminatory power of CRISPR typing offers a new prospective for salmonellosis monitoring with a fast and accurate method that could easily replace gold standards serotyping and PFGE.

TABLE DES MATIERES

LISTE DES ILLUSTRATIONS

LISTE DES TABLEAUX

LE GENRE *SALMONELLA*

I. UN PEU D'HISTOIRE

I.1. DECOUVERTE

Bien avant l'ère bactériologique, la typhoïde, première salmonellose reconnue, fut définie sur la base de ses caractères cliniques et des changements anatomiques (1). Bien que les pathologistes Petit et Serres remarquèrent en 1813 des ulcérations intestinales prédominantes au niveau caecal, la fièvre entérique ne fut distinguée de la tuberculose qu'en 1822 par Bretonneau qui mis en évidence sa contagiosité. C'est en 1829 que Louis décrivit certains aspects de la fièvre entérique sous le terme typhoïde. Ainsi, son mode de transmission, sa pathogénèse pouvaient être étudiés et un traitement et un mode de prévention pouvaient être envisagés. En 1856, Budd émit l'hypothèse selon laquelle chaque cas de typhoïde est lié épidémiologiquement à un cas antérieur et qu'une toxine spécifique est disséminée via les selles des patients infectés. Douze ans plus tard, Pettenkofer mis en évidence le rôle des eaux usées dans sa dissémination. Ce n'est qu'en 1880 que le bacille typhique fut observé pour la première fois par Eberth dans des coupes de rate et de ganglion lymphatique d'un patient décédé de typhoïde. Koch confirma cette observation et il a fallu attendre quatre ans pour une culture du germe par Gaffky (1884). D'autres bactéries semblables, dites paratyphiques, furent isolées de cadavres d'animaux. Salmon et Smith (1885) isolèrent *Bacillus cholerae−suis* de porcs atteints du « hog cholera » pensant qu'ils avaient identifié la bactérie responsable du « hog cholera » mais des bactéries similaires furent identifiées à la suite de toxi−infections alimentaires par Gaffky et Paak en 1885 et de zoonoses par Loeffler en 1892). En 1900, Lignières définit un nouveau genre bactérien incluant ces isolats qu'il nomma *Salmonella* (2).

A cette époque, en raison du manque de caractères différentiels, la distinction avec d'autres bactéries entériques déjà décrites fut difficile, ce qui remit quelque peu en question ces découvertes. Le doute fut levé en 1896 par Pfeiffer et Kolle puis Gruber et Durham qui mirent en évidence que le sérum d'un animal immunisé avec le bacille typhique agglutinait avec le dit bacille. De leur côté, Widal à Paris et Grunbaum à Londres observèrent que le sérum de patients atteints de typhoïde agglutinait avec le bacille, le sérodiagnostic devint alors possible. Cependant, la même année, deux souches furent isolées de patients présentant les symptômes de la fièvre typhoïde mais avec un sérodiagnostic de Widal négatif. Achard et Bensaude les nommèrent « bacilles paratyphiques » et la pathologie associée « maladie paratyphique ». Un cas similaire fut publié deux ans plus tard par Gwyn qui nomma la bactérie « paracoli bacillus ». En 1901, Schottmüller confirma que « bacille paratyphique » et « paracoli bacillus » étaient distincts du bacille typhique, aussi bien au niveau culture qu'au niveau sérologie et les nomma « Paratyphus B » et « Paratyphus A » respectivement. Un nouveau concept de groupe de maladies causées par un groupe de bactéries venait d'émerger.

L'analyse antigénique suivit en 1902 grâce aux travaux de Castellani décrivant une méthode d'absorption des antisérums. Les antigènes somatiques et flagellaires furent différenciés par Smith et Reagh l'année suivante et ça n'est qu'en 1920 que Weil et Felix les nommèrent O, de l'Allemand Ohne Hauch (« sans film ») et H pour Hauch (« film »), respectivement (3). Andrewes (1922,1925) mit en évidence que les antigènes H pouvaient être diphasiques. Plus tard, Felix et Pitt (1934) montrèrent qu'un antigène de surface (Vi) pouvait empêcher l'agglutination du bacille typhique. Ainsi, White (1926) définit le premier schéma antigénique pour *Salmonella* complété régulièrement par Kauffmann (1934−1965) puis Le Minor (1965−

1989). En 1941, le schéma Kauffmann−White recensait 100 sérotypes ; il en comptait 2 000 en 1978 et 2 600 aujourd'hui.

I.2. TAXONOMIE

Les bactéries du genre *Salmonella* appartiennent à la famille des *Enterobacteriaceae*. Ce sont des bacilles à coloration de Gram négative, aéro−anaérobies facultatifs, souvent mobiles grâce à une ciliature péritriche, prototrophes, capables de produire de l'H_2S à partir du thiosulfate et réduisant les nitrates en nitrites. Ce genre rassemble des bactéries reliées entre elles selon des caractéristiques phénotypiques et génotypiques. Le contenu en G+C de l'ADN de *Salmonella* est compris entre 50 et 52 % comme *Escherichia*, *Shigella* et *Citrobacter* et le genre est défini comme un groupe d'hybridation ADN−ADN dont les membres ont des ADN hybridant à plus de 80 %. En 1963, Kauffmann divise le genre en quatre sous−genres : I, II, III et IV. La plupart des souches de salmonelles isolées de l'Homme et des vertébrés à sang chaud appartiennent au sous genre I. Les sous genres II et III (appelés aussi Arizona) sont très proches et leur différence mise en doute. Avec le sous genre IV, ils sont le plus souvent associés aux reptiles chez qui ils sont pour la plupart commensaux. En 1972, Edwards et Ewing proposent une autre définition taxonomique dans laquelle le genre *Salmonella* est limité au sous genre I de Kauffmann, le sous genre III devenant *Arizona* et les souches des sous genres II et IV étant considérées comme des atypies des deux autres.

La nomenclature du genre *Salmonella* a souvent été sujette à controverses puisqu'elle ne suit pas les règles du système international de nomenclature des bactéries. Au départ, les bactéries identifiées étaient nommées selon la pathologie associée : *S. typhi* (typhoïde), *S. paratyphi* (bacille paratyphique), *S. abortus−ovis* (agent causal de l'avortement chez la brebis), *S. typhimurium* (pathogène murin), *etc*. Lorsque le caractère ubiquiste des différents

sérotypes fut établi (*S. typhimurium*, *S. bovis−morbificans*), le choix du nom des nouveaux sérotypes identifiés fut basé sur le lieu de primo identification de la bactérie (*S. london*, *S. panama, S. stanleyville*) et ce pour toutes les souches du sous genre I. En 1970, Le Minor, Rohde et Taylor proposent de diviser le genre en quatre espèces correspondant aux sous genres de Kauffmann : *kauffmannii* (I), *salamae* (II), *arizonae* (III) et *houtenae* (IV). Depuis l'hybridation ADN−ADN a mis en évidence que le genre était en fait divisé en deux espèces distinctes : *enterica* et *bongori*, l'espèce *enterica* étant elle−même sous divisée en six sous−espèces : *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV) et *indica* (VI).

La définition actuelle du genre *Salmonella* est basée sur ses caractères biochimiques (**tableau 1**). Les noms des nouveaux sérotypes de la sous−espèce *enterica* correspondent à laville où ils ont été isolés, doivent être écrits avec une majuscule et précédés du genre de l'espèce et de la sous espèce en italique ; *S. typhi* devient ainsi *Salmonella enterica* sous espèce *enterica* sérotype Typhi, qui peut s'abréger en *S*. *enterica* sérotype Typhi ou *S.* Typhi. Pour les autres sous espèces et l'espèce *bongori*, le sérotype est désigné par la formule antigénique. Cette nomenclature a définitivement été acceptée en 2005 (4).

TABLEAU 1 : CARACTERES DIFFERENTIELS DES ESPECES ET SOUS ESPECES DU GENRE SALMONELLA D'APRES WHITE-KAUFFMANN-LE MINOR (5)

Species	*S. enterica*						*S. bongori*
Subspecies	*enterica*	*salamae*	*arizonae*	*diarizonae*	*houtenae*	*indica*	
Characters							
Dulcitol	+	+	–	–	–	d	+
ONPG (2 h)	–	–	+	+	–	d	+
Malonate	–	+	+	+	–	–	–
Gelatinase	–	+	+	+	+	+	–
Sorbitol	+	+	+	+	+	–	+
Growth with KCN	–	–	–	–	+	–	+
L(+)-tartrate[a]	+	–	–	–	–	–	–
Galacturonate	–	+	–	+	+	+	+
γ-glutamyltransferase	+(*)	+	–	+	+	+	+
ß-glucuronidase	d	d	–	+	–	d	–
Mucate	+	+	+	– (70%)	–	+	+
Salicine	–	–	–	–	+	–	–
Lactose	–	–	– (75%)	+ (75%)	–	d	–
Lysed by phage O1	+	+	–	+	–	+	d
Usual habitat	Warm-blooded animals		Cold-blooded animals and environment				

(a) = *d*-tartrate.
(*) = Typhimurium d, Dublin –.
\+ = 90 % or more positive reactions.
– = 90 % or more negative reactions.
d = different reactions given by different serovars.

II. DES SALMONELLES ET DES HOMMES

II.1. HABITAT

Les salmonelles se retrouvent naturellement et indifféremment chez les animaux homéothermes (homme, volaille, mammifères) ou poïkilothermes (reptiles, insectes). D'une manière générale, la volaille d'élevage (poulets, dindes, canards) constitue le réservoir le plus important et la plus grande source de contamination pour l'Homme. Les salmonelles sont disséminées dans l'environnement (eau, sol, plantes) par les excrétions animales ou humaines. Même si elles ne se multiplient pas de manière significative dans la nature, elles

peuvent survivre des semaines dans l'eau et jusqu'à plusieurs années dans le sol si les conditions de température, d'humidité et de pH le permettent.

II.2. PATHOLOGIE

L'infection à *Salmonella* est appelée salmonellose et se manifeste sous deux formes : la gastroentérite (la plus commune) et la fièvre typhoïde ou paratyphoïde. Elle survient en général après la consommation d'eau ou d'aliments (viande peu cuite, ovoproduits, produits laitiers), ou plus rarement par contact direct avec des animaux.

Les gastroentérites sont provoquées par des *Salmonella* ubiquistes présentes chez l'Homme et l'animal. La durée d'incubation est généralement d'un à deux jours et dépend de la dose ingérée, de la santé de l'hôte et des caractéristiques de la souche de *Salmonella* mise en cause. Les salmonelloses provoquent une forte fièvre accompagnée de diarrhées, vomissements et douleurs abdominales. Chez des adultes de condition physique normale, une gastroentérite disparaît sans traitement après trois à cinq jours en moyenne. En revanche, une antibiothérapie doit être prescrite chez les personnes âgées, les nourrissons, ou les personnes immunodéprimées chez lesquels l'infection peut être plus sévère, voire mortelle.

Les fièvres typhoïdes et paratyphoïdes sont causées par des *Salmonella enterica* strictement adaptées à l'Homme, *S. enterica* sérotype Typhi (fièvre typhoïde), Paratyphi A, Paratyphi C et certaines souches de Paratyphi B (fièvres paratyphoïdes). Après une période d'incubation d'une à deux semaines survient une fièvre continue accompagnée de maux de tête, d'anorexie, d'abattement ("tuphos" : torpeur en grec), de douleurs abdominales avec diarrhée ou constipation. Dans les formes bénignes, l'état reste stationnaire pendant une quinzaine de jours puis la convalescence dure plusieurs semaines. Dans les formes plus

graves où des complications peuvent survenir au niveau de l'intestin, du cœur ou de la vésicule, la fièvre typhoïde peut être fatale en l'absence de traitement. Le taux de mortalité est de 10% comparé à moins de 1% pour les autres formes de salmonellose. Il existe de plus un portage chronique des *S. enterica* sérotype Typhi : après guérison d'une fièvre typhoïde, 2 à 5% des individus continuent à excréter ces bactéries. Les symptômes des fièvres paratyphoïdes sont similaires, mais le plus souvent moins sévères, le taux de mortalité de cette salmonellose étant par ailleurs bien plus bas que celui de la fièvre typhoïde.

II.3. EPIDEMIOLOGIE

En dépit d'une connaissance approfondie du pouvoir pathogène et des moyens de préventions de contamination, les salmonelles restent la principale cause de gastro entérite d'origine alimentaire dans le monde. L'industrialisation des méthodes d'élevage et de conditionnement des viandes a largement contribué à l'augmentation de l'incidence des cas d'infection à *Salmonella*. De même, les œufs et les ovoproduits sont des sources courantes de contamination.

Les salmonelloses d'origine alimentaire peuvent donner lieu à des foyers épidémiques très importants, qui peuvent atteindre une échelle nationale (et même internationale) si un aliment commercialisé à large diffusion se trouve contaminé. En 1994, aux Etats−Unis, par exemple, une épidémie provoquée par une crème glacée a touché 224 000 personnes (6). En France, une des plus importantes épidémies, dont la source n'a pu être identifiée, survenue fin 1985, aurait touché 25 000 personnes d'après l'estimation la plus faible.

La contamination par les *Salmonella* responsables des fièvres typhoïdes et paratyphoïdes se fait par ingestion d'eau ou d'aliments ayant subi une contamination fécale d'origine humaine. Comme toutes les maladies à transmission oro−fécale, ces fièvres surviennent le

plus souvent dans des zones où l'hygiène est précaire, et frappent principalement les pays en développement en Asie, en Afrique ou en Amérique Latine. Les données mondiales les plus récentes font état de 21 millions de cas annuels de fièvre typhoïde et de 200 000 morts (7). La maladie n'a cependant pas totalement disparu des pays industrialisés. En 2003, un foyer de sept cas groupés liés à un lieu de restauration, a été détecté dans le XVI^{e} arrondissement de Paris., la source de la contamination ayant été imputée à un porteur sain travaillant en cuisine (données InVS). Le nombre de cas annuels en France est inférieur à 0,3 pour 100 000 habitants. Depuis 1999, cent à cent−cinquante souches de *S. enterica* sérotype Typhi, isolées en France, ces souches provenant quasi−exclusivement de cas importés.

La surveillance épidémiologique est assurée par le Centre National de Référence des *Escherichia coli−Shigella−Salmonella* (CNR−ESS, Institut Pasteur, Paris) dont les principales missions sont (8):

- suivre les tendances évolutives temporelles des différents sérotypes de *Salmonella* en s'appuyant sur un réseau national de laboratoires d'analyses de biologie médicale,
- contribuer à l'alerte en signalant tout évènement inhabituel (augmentation du nombre de cas, survenue de cas groupés, modification des profils de résistance, apparition de souches inhabituelles…),
- contribuer à la surveillance et à l'investigation des toxi−infections alimentaires collectives en signalant à l'Institut de Veille Sanitaire (InVS) les foyers de cas groupés notifiés au CNR,
- collaborer avec les réseaux nationaux de surveillance des salmonelles chez l'animal, dans les aliments et l'environnement,

- participer avec l'InVS au réseau européen des surveillances des *Salmonella* de l'European Center of Disease Control (ECDC).

III. REVUE DES DIFFERENTES METHODES DE SOUS-TYPAGE DU GENRE *SALMONELLA*

Depuis maintenant 100 ans, le sérotypage reste la méthode de référence d'identification des salmonelles, l'agglutination des antigènes somatiques et flagellaires se faisant en suivant le schéma de White−Kauffmann−Le Minor (5, 9).

De nombreuses alternatives moléculaires ont été développées par identification des gènes codant les antigènes recherchés : l'opéron *rfb* pour les antigènes O, *fliC* et *fljB* codant pour les antigènes flagellaires (10, 11, 12, 13). Chaque gène peut être amplifié et séquencé pour identification (14, 15, 16). Cette approche demande cependant une bonne connaissance de ces gènes et des bases de données complètes afin de couvrir tout le spectre des sérotypes. La complexité de l'opéron *rfb* (**Figure 1**) rend difficile la mise au point d'une méthodecouvrant la totalité des sérogroupes. En effet, les différences antigéniques entre les 46 différents groupes ne résident pas dans la variation génétique d'un seul gène du cluster (17) mais dans les variations de l'ensemble de l'opéron (18). Dans certains cas, les facteurs antigéniques peuvent même être codées par des gènes extérieurs à l'opéron, comme les antigènes O:24 et O:25 (19). A ce jour, les séquences de 28 des 46 sérogroupes de *Salmonella* ont été complètement annotées.

FIGURE 1 : ORGANISATION DE L'OPERON *RFB* POUR LES NEUF SEROGROUPES MAJORITAIRES. LES CIBLES PCR SEROGROUPES-SPECIFIQUES SONT ENTOUREES, D'APRES LES REFERENCES 12,18 ET 19.

Malgré tout, de nombreux systèmes PCR multiplex ont été mis au point pour l'identification rapide des sérotypes les plus fréquents (19, 20, 21). L'étude la plus complète portait sur 500 échantillons représentatifs de 30 sérotypes avec pour cibles cinq antigènes O (O:4, O:7, O:8, O:9, O:3,10), huit H1 (i ; r ; l,v ; e,h ; z10 ; b ; d : G) et sept H2 (1,2 ; 1,5 ; 1,6 ; 1,7 ; l,w ; e,n,x ; e,n,z15) ; pour 423 des 500 échantillons testés, le sérotypage moléculaire a donné le même résultat que le sérotypage classique, soit 85 % de fiabilité (21).

Le sérotypage n'est cependant pas suffisant du fait de la prévalence de certains comme Typhimurium et sont variant monophasique qui représentent plus de 50 % des salmonelles isolées. Afin de contenir l'émergence de souches virulentes et de prévenir de nouveaux cas d'infection, il est important de définir des marqueurs épidémiologiques fiables. Les investigations épidémiologiques des salmonelloses nécessitent de plus en plus souvent une analyse moléculaire des souches isolées de façon à relier les différents cas décrits entre eux et identifier avec plus de certitude la ou les souches responsables de ces épidémies. De

nombreuses méthodes de sous−typage ont été mises au point pour différencier certains isolats au sein même d'un sérotype.

III.1. LYSOTYPIE

Au sein d'un même sérotype de *Salmonella*, la sensibilité à certains bactériophages peut varier, ce qui a permis la mise au point de différents schémas de lysotypie (1). En 1938, Craigie utilise une collection de variants du phage Vi II pour typer les souches exprimant l'antigène capsulaire Vi (*S.* Typhi, *S.* Paratyphi C et certaines *S.* Dublin). D'autres schémas ont suivi pour les sérotypes Paratyphi B (22) et Typhimurium (Anderson, 1964). Jusqu'au début des années 2 000, la lysotypie fut considérée comme, un marqueur épidémiologique majeur pour les sérotypes Typhi (96 lysotypes) et Typhimurium (200 lysotypes). Malgré l'abandon de la méthode au profit des méthodes moléculaires, les lysotypes prévalents restent ancrés dans le langage commun, le plus connu étant le DT104 de Typhimurium, clone majoritaire connu pour sa pentarésistance aux antibiotiques.

III.2. RIBOTYPIE

La ribotypie est basée sur l'étude de l'ADN codant les ARNr 16S et 23S. Les gènes codant ces ARNr sont organisés en opérons (*rrn*...) présent en une ou plusieurs copies sur le chromosome bactérien. Ainsi, un profil caractéristique est obtenu par restriction enzymatique de l'ADN total suivi d'une hybridation selon Southern à l'aide d'une sonde universelle ciblant l'opéron *rrn* (23a). Le pouvoir discriminant de cette méthode repose sur lefait que les gènes ciblés ici sont extrêmement stables. Celui−ci peut être augmenté par l'utilisation de différentes enzymes de restriction. L'analyse des profils permet de distinguer certaines souches parmi les sérotypes et lysotypes communs. Elle n'est cependant pas

substituable au sérotypage et malgré son automatisation (Riboprinter®), cette méthode a été abandonnée au profit de nouvelles méthodes plus discriminantes.

III.3. IS*200*

La séquence d'insertion IS*200* est un autre marqueur spécifique chez le genre *Salmonella* (23b). Cette séquence d'insertion de 708 paires de bases (pb) est présente et répétée plusieurs fois le long du génome des salmonelles en fonction du sérotype. Le nombre de copies et leur localisation sur le chromosome sont très peu variables donccaractéristiques d'une souche de *Salmonella*. Le profil IS*200* est obtenu par hybridation à l'aide d'une sonde ciblant la séquence d'insertion après restriction de l'ADN et transfert sur membrane selon Southern. Jusqu'au début des années 2 000, cette méthode est souvent combinée à l'électrophorèse en champ pulsé et la ribotypie pour les caractérisations moléculaires de certaines populations comme les sérotypes Paratyphi B ou Typhi (24, 25).

III.4. ELECTROPHORÈSE EN CHAMPS PULSÉ (PULSED FIELD GEL ELECTROPHORESIS-PFGE)

Ce type d'électrophorèse a été développé par Schwartz et Cantor en 1984 afin de séparer les grandes molécules d'ADN (> 50 kb) que l'électrophorèse classique en gel d'agarose ne permet pas de résoudre. La macrorestriction est basée sur l'utilisation d'enzymes à sites de coupures rares afin d'obtenir après migration des profils permettant dediscriminer les souches d'un même sérotype entre elles. Pour augmenter le pouvoir discriminant de la méthode, il est possible d'utiliser plusieurs enzymes de restrictions et de comparer les profils combinés pour étudier une population particulièrement homogène. Cette technique est actuellement la méthode de référence en matière de sous typage du genre *Salmonella* et est devenu un outil de surveillance en temps réel des épidémies de salmonelloses via le réseau PulseNet réunissant les agences de santé publique et les

laboratoires de contrôle alimentaire qui soumettent leurs profils dans une base de données internationale grâce à un protocole standardisé (26).

Cette méthode connaît quand même certaines limites à commencer par son temps d'exécution et l'impossibilité de l'automatiser, ce qui allonge le délai de mise en place de mesures de contrôle d'une épidémie (retrait/destruction de la source alimentaire en cas de TIAC). L'interprétation des profils peut aussi varier d'un laboratoire à un autre et certains sérotypes ne sont pas typables par cette méthode (lyse de l'ADN).

III.5 Analyse Multi-locus Variable Number Tandem Repeats (MLVA)

Face à la progression rapide du nombre d'épidémies à *S.* Typhimurium DT104 et au faible pouvoir discriminant du PFGE sur cette population, une nouvelle méthode de sous−typage a été développée en 2003 avec l'étude des loci VNTR (Variable Number Tandem Repeats) (27). Ces loci sont amplifiés par PCR et leur taille est mesurée par électrophorèse en capillaire. Chaque taille de locus correspond à un allèle et l'ensemble des allèles d'un échantillon donne le sous−type. Deux schémas standards pour les sérotypes Typhimurium (28) et Enteritidis (29) ont été ainsi mis au point et utilisés en routine pour les investigations d'épidémiologies imputées à ces sérotypes (30).

III.6. Multi-Locus Sequence Typing (MLST)

Le typage MLST consiste à cibler directement la variation nucléotidique de différents gènes de ménage. En règle générale, les schémas en comptent sept car augmenter le nombre de gènes ne permet pas d'augmenter le nombre de génotypes par le séquençage direct de partie interne de ces gènes (31). Cette méthode est dérivée de l'analyse des isoenzymes ou MLEE (multilocus enzyme electrophoresis) qui se base sur les variations de la mobilité électrophorétique de certaines enzymes bactériennes (isoenzymes) (32). L'avantage

majeur du séquençage est que c'est une technologie générant des données non ambigües pouvant être stockées et partagées électroniquement et donc échangeables entre les différents laboratoires du monde entier (http://pubmlst.org; http://mlst.warwick.ac.uk/mlst/). L'analyse des données MLST n'utilise pas les séquences directement mais un code allélique. En effet, pour un locus donné, chaque séquence unique se voit assigné un numéro d'allèle. L'ensemble des numéros d'allèle de tous les loci sont combinés pour former un profil allélique (séquençotype−ST). Deux souches proches auront le même profil ou des profils ne variant que sur un ou deux loci alors que des souches éloignées n'auront pas d'allèle commun ou très peu.

Un schéma MLST basé sur sept gènes de ménage (*aroC*, *dnaN*, *hisD*, *hemD*, *purE*, *sucA*, *thrA*) a d'abord été développé pour l'étude évolutive du sérotype Typhi (33) puis testé sur 110 isolats représentant 25 sérotypes de la sous espèce *enterica* (34). S'en sont suivi plusieurs études sur les sérotypes Newport (35), Typhimurium (36) et d'autres sérotypes qui ont mis en évidence une corrélation en le séquençotype et le sérotype. Cette méthode a même été proposée comme substitut possible au sérotypage grâce à une étude d'Achtman *et al* portant sur 4 257 souches de 554 sérotypes représentant une diversité de 1 092 séquençotypes (37, **Figure 2**).

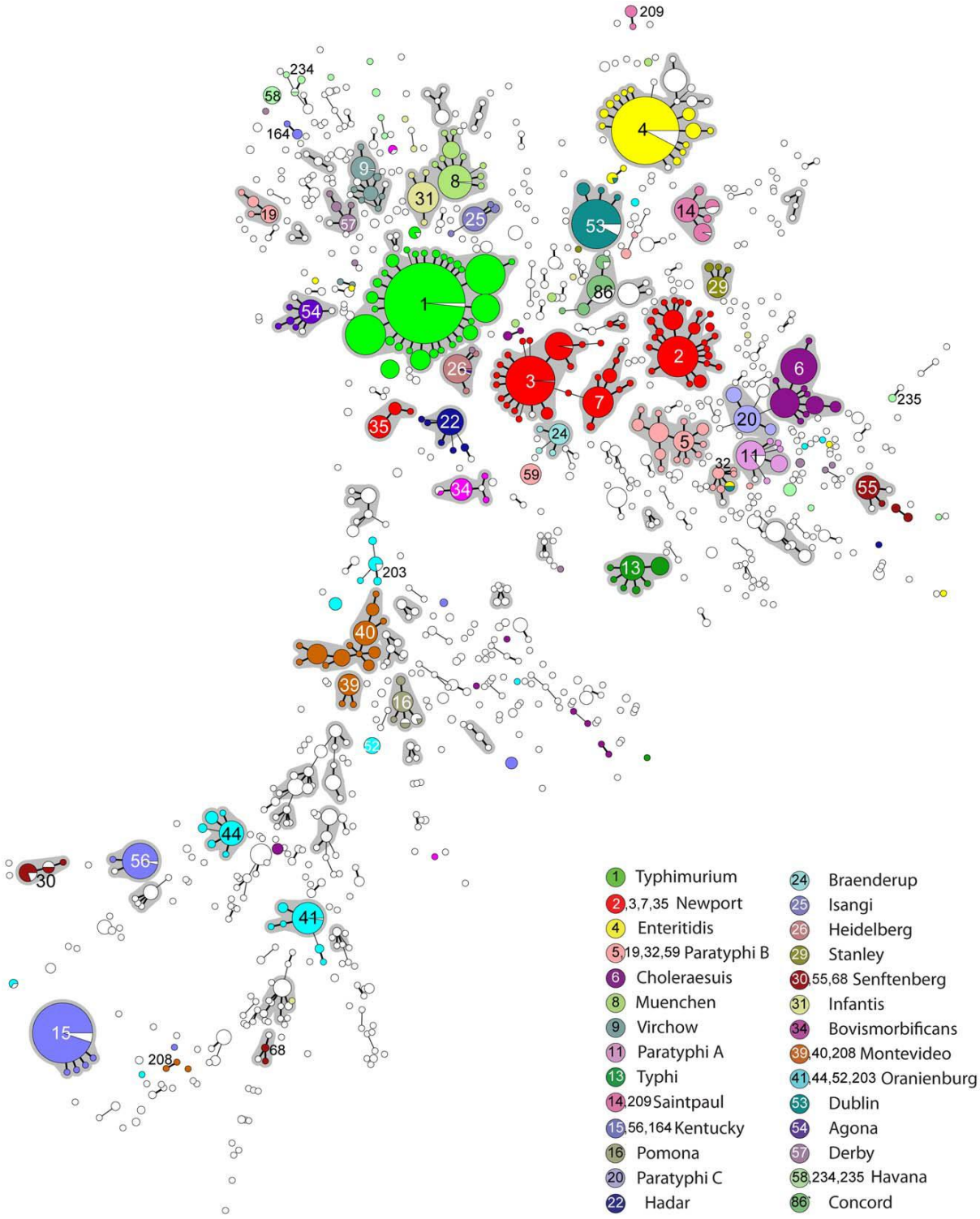


FIGURE 2 : MINIMUM SPANNING TREE (MST) DE 4 257 SOUCHES DE *SALMONELLA ENTERICA* SUBSP *ENTERICA*. CHAQUE CERCLE CORRESPOND A UN ST ; SA TAILLE EST PROPORTIONNELLE AU NOMBRE DE SOUCHES, D'APRES ACHTMAN ET AL. (37).

Cette étude a mis en évidence une corrélation entre ST et sérotype. Dans le cas de profils polyphilétiques, la méthode est même plus discriminante que le sérotype. En revanche, elle ne permet pas de sous−typer les sérotypes prévalents Typhimurium (ST19 majoritaire) et Enteritidis (ST11 majoritaire).

CLUSTERED REGULARLY INTERSPACED SHORT PALINDROMIC REPEATS (CRISPR)

En 1987, une nouvelle famille de courtes séquences répétées flanquant une région de 24 kb est décrite chez *Escherichia coli :* l'une située 24 pb en amont du gène *iap* et en aval du gène *ygcF*, chacun contenant des séquences répétées de 29 pb (38, 39).

I. ORGANISATION DES LOCI CRISPR

Un locus CRISPR est caractérisé par deux types de séquences non codantes : (i) les répétitions (Direct Repeats−DR, 21−48 pb) séparées par des séquences uniques appelées (ii) spacers (21−72 pb) **(Figure 3)**. Au sein de chaque région CRISPR, la longueur des DR et des spacers est conservée, excepté la dernière répétition à l'extrémité −3' qui est tronquée dans 30 % des cas (40, 41, 42) ou dégénérée. L'identification de la dernière répétition est ainsi un facteur clé dans l'orientation du locus CRISPR et permet son annotation correcte.

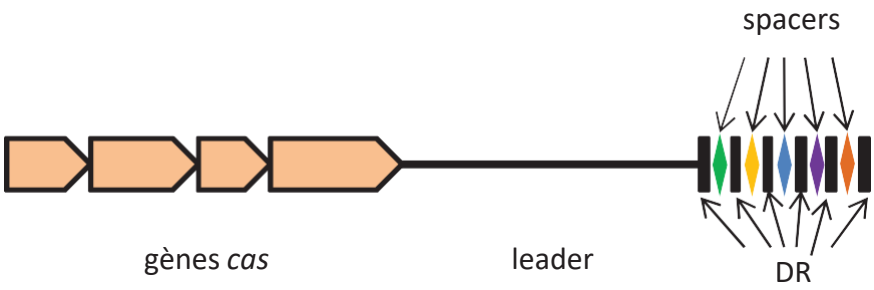


FIGURE 3 : REPRESENTATION SCHEMATIQUE D'UN LOCUS CRISPR

I.1. REPETIONS-DR

Les séquences répétées sont conservées au sein d'un locus et semblent espèce−spécifiques (43, 44). Des similarités entre les séquences de certaines DR permettent de les

diviser en 12 groupes (45) dont certains contiennent des motifs palindromiques, habituellement GTTTg/c et GAAAC, aux extrémités qui joueraient un rôle dans latranscription du locus (43, 45).

I.2. SPACERS

A l'opposé du caractère conservé des DR, les séquences de spacers sont, elles, distinctes et variables au sein d'un même locus. De plus, il apparaît que toutes les séquences de spacers au sein d'un même génome sont uniques, à quelques rares exceptions près (46, 47). De la même manière, le nombre de motifs DR−spacers au sein d'un locus CRISPR est différent entre deux espèces distinctes mais également au sein d'une même espèce, ce nombre pouvant aller de 2 à 375 (44, 46).

Les spacers identifiés chez les bactéries et les archaea présentent certaines homologies avec des séquences d'éléments mobiles, phages et plasmides conjugatifs essentiellement, indiquant leur provenance (48, 49, 50, 51). Une étude chez *Streptococcus thermophilus* a en effet mis en évidence que parmi 40 % des séquences de spacers inventoriées, 75 % étaient homologues à des séquences de phages et 20 % à des séquences de plasmides (48).

I.3. TRANSMISSION PAR TRANSFERT HORIZONTAL

Comme défini précédemment, les loci CRISPR sont différents d'une espèce à l'autre. Cependant, certaines espèces éloignées comme *Escherichia coli* et *Mycobacterium avium* possèdent des séquences CRISPR similaires (41), ce qui suggère qu'elles ont pu évoluer indépendamment du reste du chromosome. La prédominance des régions CRISPR chez les archaea (90 %) et leur fréquence limitée chez les bactéries (40−70%) laisse à penser que ces régions se sont développées au sein de souches ancestrales d'archaea et ont été transmises par transfert horizontal chez les bactéries. Cette hypothèse est renforcée par le contenu en

G−C des loci CRISPR distinct du G−C % du reste du chromosome. Des motifs CRISPR ont été identifiés sur des mégaplasmides (52) et sur des prophages (53), ce qui en fait des vecteurs probables de cette dissémination entre espèces éloignées.

I.4. SEQUENCE LEADER

Cette région non codante, riche en A−T, est située en région −5' d'un locus CRISPR et varie selon les espèces (20−534 pb) (41). Au sein d'un même génome, les séquences leader sont quasiment identiques (80 %) d'un locus à l'autre (54, 55, 56).

La séquence leader est supposée jouer un rôle dans l'acquisition de nouveaux spacers au sein d'un locus CRISPR, cette région contenant un site de fixation pour certaines protéines spécifiques (sûrement codées par les gènes *cas*) (51, 57). Une étude métagénomique de deux populations naturelles de *Leptospirillum* sp. a mis en évidence un pattern spécifique selon lequel les spacers étaient intégrés (58). De nouveaux spacers à certaines souches ont en effet été identifiés du côté de la séquence leader, plus précisément entre le leader et le premier motif DR−spacer, alors que les spacers les plus anciens étaient localisés à l'autre extrémité du locus. La même observation a été faite en laboratoire à partir de cultures de *S. thermophilus* (42, 44, 57).

Il a aussi été suggéré que la séquence leader agissait comme un promoteur de la transcription des loci CRISPR (49, 59, 60, 61). Cette hypothèse fut confortée par l'absence d'acquisition de nouveaux spacers observée au niveau de loci CRISPR non précédés de séquences leader, conséquemment considérés comme des reliquats inactifs (49, 62). Cependant, l'activité d'un locus CRISPR fonctionnel est mise en évidence par la transcription en ARNm initiée à partir de la séquence leader. Le transcrit résultant est alors pris en charge par des protéines Cas spécifiques, générant ainsi les ARNcr (CRISPR ARN). Ces molécules sont

considérées comme la base des systèmes CRISPR/*cas* qui fonctionne selon un mécanisme d'interférence (59) décrit plus loin.

II. LES GENES *CAS*

La majorité des loci CRISPR identifiés est associée à un groupe de gènes conservés nommés *cas* (CRISPR−associated genes) (41, 63, 64). Ces gènes ne sont présents qu'au sein de génomes contenant des loci CRISPR. Leur nombre varie de 4 à 20 (57, 63, 65) et sont situés aussi bien en amont qu'en aval des motifs DR−spacers mais toujours du même côté pour un locus donné.

A l'image des répétions, les gènes *cas* sont locus−spécifiques et les deux semblent fonctionner en tandem (63, 42, 45). Leur orientation est souvent la même que les DR. Ilexiste également une corrélation entre le nombre de motifs au sein d'un locus CRISPR et l'organisation des gènes cas associés : une étude de la région des genres *Salmonella* et *Escherichia* a mis en évidence que lorsque la région cas est intacte, le nombre de répétitions des loci CRISPR est élevé (66). Il est à noter que lorsqu'au moins deux loci CRISPR ayant la même séquence répétée sont présents au sein d'un génome, un seul d'entre eux est associé aux gènes *cas* (46). Selon la classification proposée par Makarova définissant 12 groupes de répétitions, quatre sont clairement associés à des sous types CRISPR/cas spécifiques (67).

Les gènes *cas* sont couramment distribués parmi les archaea et les bactéries. Certains sont conservés et limités à certaines espèces, d'autres peuvent se retrouver chez des espèces phylogénétiquement éloignées, suggérant une dissémination par transfert horizontal. Une

analyse bioinformatique à large échelle réalisée sur 200 génomes a permis d'identifier 45 familles de protéines Cas (63). Makarova reclassa cette famille en 23 groupes incluant des sous−groupes phylum−spécifiques (64). Plus de 65 gènes *cas* distincts ont été identifié depuis (43).

II.1. GENES « CORE »

En dépit d'une grande variabilité, quatre « core » gènes *cas* ont été définis (*cas*1− *cas*4) (41) puis deux autres (*cas5* et *cas6*) (48). Actuellement, 10 gènes ont été décrits dans lafamille *cas* (**Tableau 2**). Cette famille de gènes est largement répandue au sein d'espèces bactériennes souvent éloignées et se situe à proximité d'un locus CRISPR. En général, tousles gènes ne sont pas présents mais *cas1* et *cas2* restent les plus fréquents (63, 64). Les protéines pour lesquelles ils codent, Cas1 et Cas2 respectivement, sont les plus conservées au sein de tous les systèmes CRISPR/cas contrairement aux autres protéines Cas (67). En raison de sa fréquence, Cas1 est considérée comme un marqueur pour détecter les systèmesCRISPR/cas (65).

TABLEAU2 : « CORE » GENES *CAS* ET LEUR FONCTION PUTATIVE

Gène	Protéine	Prédiction
cas1	Endonucléase/intégrase métallo−dépendante	Impliquée dans l'intégration des nouveaux spacers
cas2	Endonucléase métallo−dépendante	Facilite l'intégration des nouveaux spacers Coupure des ADNss
cas3	Nucléase ADN simple−brin ; hélicaseATP−dépendante	Coupe l'ADN lors de la phase d'interférence
cas4	Nucléase type RecB	Digestion de l'ADN étranger
cas5, cas6, cas7	Endonucléases métallo−indépendantes/protéines RAMP	Interférence ; fixe les ARNcr
cas8	Sous−unité du complexe Cascade	Liaison ADN / intéraction avec protéines RAMP
cas9	Endonucléase avec domaines type RuvC et HNH	Liaison avec le complexe ARNcr−ARNtracr pour cliver l'ADN étranger
cas10	Polymérase avec domaine type PALM	Séparation des brins et clivage de l'ADN simple brin

II.2. GENES « NON-CORE »

Ces gènes sont distribués de manière plus restreinte au sein des systèmes CRISPR/*cas*. Leur nom provient de l'organisme chez lequel ils ont été identifiés (ex : *cse*, CRISPR System *E. coli*, i.e. *casA−casB−casE−casC−casD*). Ils sont organisés en clusters de deux àsix gènes au sein d'un organisme (63). En général, un ou plusieurs de ces gènes accompagnent les gènes «core ».

II.3. PROTÉINES RAMP (REPAIR-ASSOCIATED MYSTERIOUS PROTEINS)

D'abord décrites comme impliquées dans la réparation de l'ADN (« Repair »), elles ont été redéfinies en Repeat Associated Mysterious Proteins (RAMPs) (46). Leur position au sein de la région CRISPR est assez aléatoire (63). De toutes les protéines Cas, ce sont les plus diverses et les moins conservées. Sur la base de l'analyse des séquences et des structuresdes protéines produites, la famille RAMP a été divisée en trois groupes : cas5, cas6 et cas7

(68). Leur analyse structurale par cristallographie a montré qu'elles possédaient un ou deux domaines de liaison ARN, appelés RNA recognition Motif (RRM). Ces protéines sont impliquées dans la reconnaissance des cibles spécifiques pendant la phase d'interférence (59).

III. CLASSIFICATION DES SYSTEMES CRISPR/*CAS*

La première classification de ces systèmes a été basée sur la phylogénie de Cas1 et le clustering des gènes a mis en évidence huit groupes distincts : Ecoli, Ypest, Apern, Nmeni, Hmari, et Dvulg, identifiés respectivement chez les espèces suivantes : *E. coli*, *Yersinia pestis*, *Aeropyrum pernix*, *Neisseria meningitidis*, *Mycobacterium tuberculosis*, *Thermotoga neapolitana*, *Haloarcula marismortui* et *Desulfovibrio vulgaris* (63, 64). Au fur et à mesure de nouvelles études plus fines sur cette famille, la classification a évolué en trois groupes (type I, II, III) en combinant les phylogénies des gènes *cas1* et *cas2*, chaque groupe ayant un gène particulier associé : type I−*cas3*, type II−*cas9*, type III−*cas10* (68, **Figure 4**).

III.1. TYPE I

Outre le gène *cas3*, ses principaux éléments sont le gène *cas4* et les gènes codants les protéines RAMPs. Ces systèmes sont divisés en sous−types : I−A (Apern/CASS5), I−B (Tnep−Hmari/CASS7), I−C (Dvulg/CASS1) , I−D, I−E (Ecoli/CASS2) et I−F (Ypest/CASS3).

III.2. TYPE II

A ce jour, ce type de système n'a été identifié que chez les bactéries. Il comprend deux sous−types : II−A (Nmenni/CASS4) et II−B (Nmenni/CASS4a). En plus de *cas1* et *cas2*, lesgènes spécifiques de ce type sont *cas4* et *cas9*.

III.3. TYPE III

Ce système est surtout retrouvé chez les archaea. Le gène associé est *cas10*, codant pour l CRISPR polymerase et les protéines RAMPs, (Cas6 et plusieurs types de Cas7). Deux sous−types ont été identifiées : III−A (Mtube.CASS6) et III−B (module polymérase−RAMP/ système Cmr).

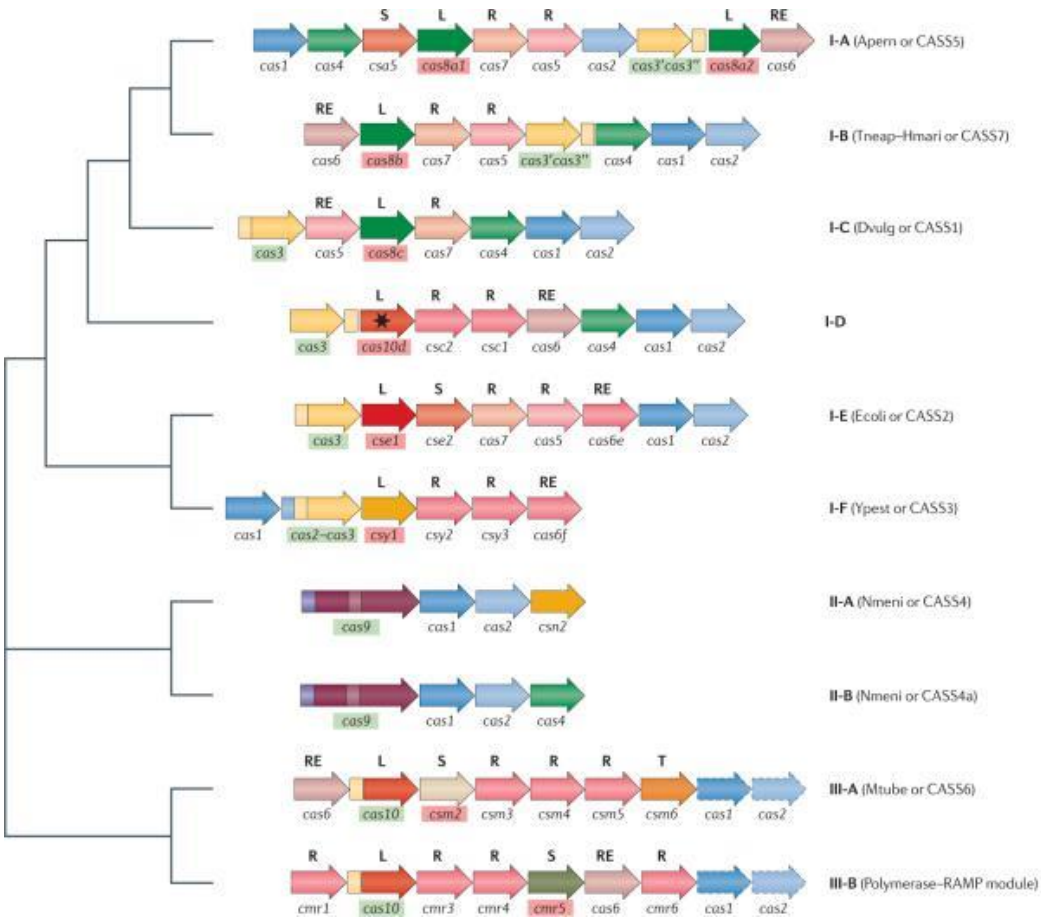


FIGURE 4 : RELATION ENTRE LES TROIS TYPES ET LEURS SOUS-TYPES DES SYSTEMES CRISPR/CAS D'APRES MAKAROVA ET AL (67)

IV. IMMUNITE ADAPTATIVE

Depuis l'identification de cette nouvelle famille de séquences répétées par Ishino, de nombreux travaux ont été menés pour définir le rôle du système CRISPR/Cas. De nombreuses hypothèses ont été émises incluant son implication dans les mécanismes de réparation de l'ADN ou la stabilité du génome (50), mais la majorité des expériences menéesont mis en évidence une relation forte entre la présence de certaines séquences de spacers et la résistance à certains phages (57). Lors d'une infection d'une culture de *S. thermophilus* par un phage lytique, certaines cellules ont résisté à la lyse (Bacteriophages insensitive Mutants−BIMs). L'analyse comparative des séquences des BIMs avec la souche sauvage a mis en évidence la présence de quatre nouveaux spacers. Leurs séquences étaient identiques à celle du phage lytique utilisé pour l'infection et appelées « proto−spacers ». Une autre étude a montré que CRISPR pouvait empêcher la conjugaison plasmidique chez *Staphylococcus epidermidis* (69) et même l'acquisition de plasmides par d'autres mécanismes de transformation et électroporation. Dans tous les cas, le mécanisme de protection contre l'invasion d'ADN plasmidique était basé sur l'incorporation d'un spacer homologue aux séquences des plasmides.

Il est alors établi que le système CRISPR/cas confère une immunité adaptative contre certains éléments génétiques étrangers où les spacers jouent un rôle déterminant dans la reconnaissance des ADNs étrangers (57, 59, 68). Ce mécanisme de défense est décrit en trois phases (70, **Figure 5**) :

- immunisation (ou adaptation) durant laquelle le(s) nouveau(x) spacer(s) dérivant de l'ADN étranger est/sont intégré(s) dans le locus CRISPR,

- <u>expression</u> du CRISPR qui implique la transcription du locus et le process des ARNcr,
- <u>interférence</u> ARNcr complexés aux protéines Cas qui éliminent l'ADN étranger par fixation et/ou dégradation de l'ADN étranger.

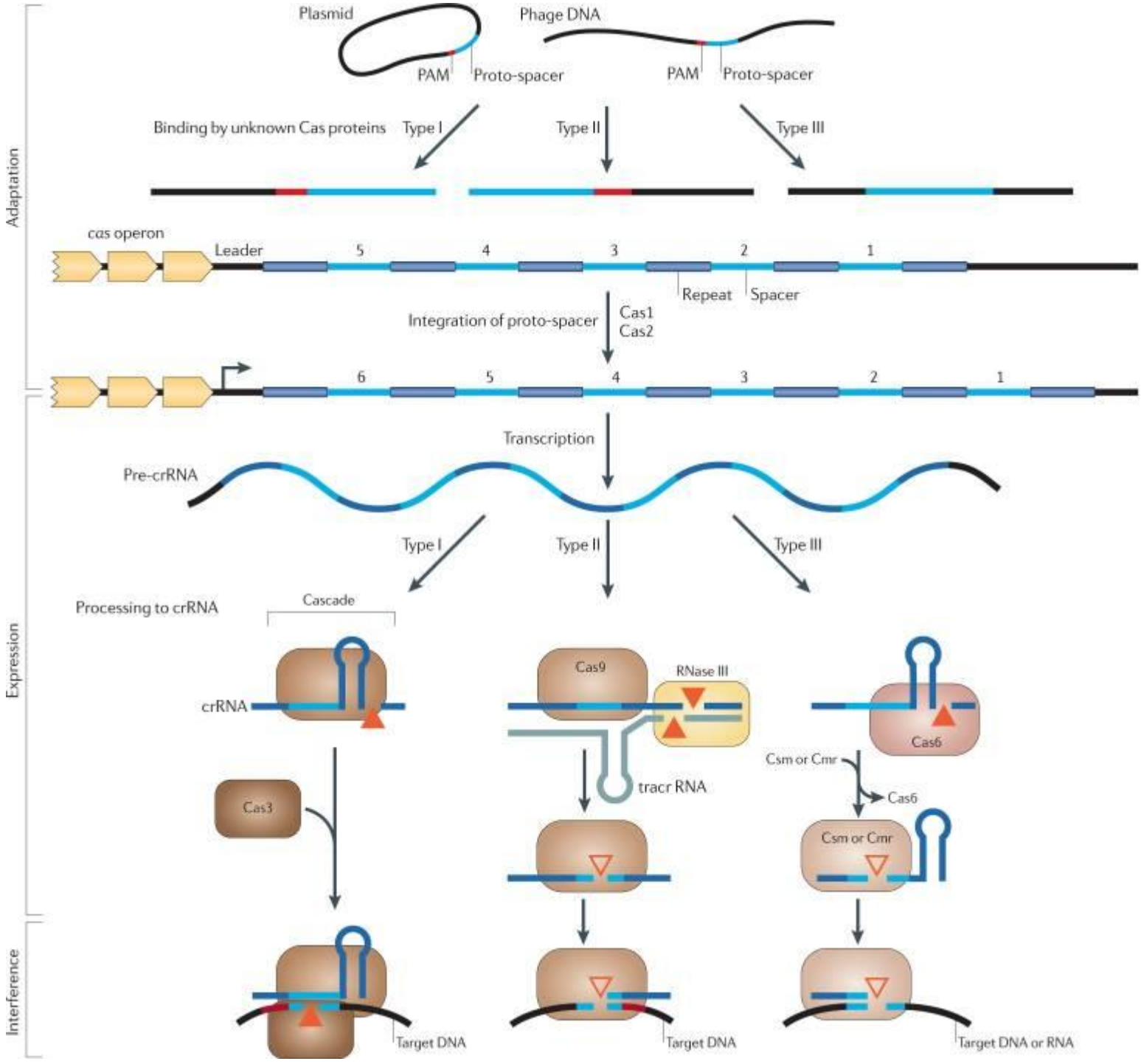


FIGURE 5 : MODE D'ACTION DU SYSTEME CRISPR/CAS ILLUSTRE PAR MAKAROVA ET AL.(67)

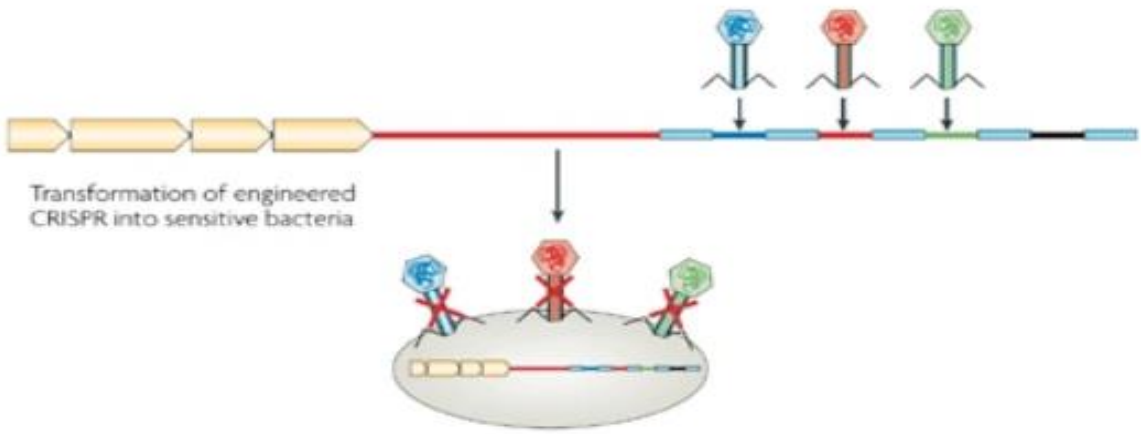
V. APPLICATIONS

La grande diversité des séquences CRISPR offre de nombreuses opportunités d'applications comme l'analyse phylogénétique de certaines populations bactériennes (42, 66, 71, 72) et plus récemment, le genome editing.

L'identification et la caractérisation des séquences CRISPR (CRISR/cas et motifs spacers−DR) a été facilitée par le développement d'outils bioinformatiques tels que CRISPRFinder (73), CRISPI (74) et CRISPRdb (46).

V.1. MODIFICATION DES DEFENSES CONTRE LES VIRUS (FIGURE 6)

De par leur propriété de défense immunitaire contre les infections phagiques, les systèmes CRISPR/cas représentent un intérêt particulier pour l'industrie comme les produits laitiers où la production microbienne joue un rôle important. En effet, la présence de phages dirigés contre les bactéries lactiques peut empêcher la fermentation et ainsi dégrader fortement la qualité des produits, avec un fort impact économique. Comme décrit précédemment, la sélection de cellules BIM peut contrer ce phénomène (42, 57).



:

FIGURE 6 : MODIFICATION ARTIFICIELLE DE LA SENSIBILITE D'UNE BACTERIE A UN PHAGE, D'APRES SOREK ET AL. (65)

V.2. INACTIVATION D'UN GENE (FIGURE 7)

La propriété RNAi du système CRISPR offre la possibilité d'inactiver tout gène d'intérêt sans modification génétique en transformant dans l'organisme d'intérêt un plasmide portant un système CRISPR−cas9 avec un spacer identique à la séquence du gène à inactiver.

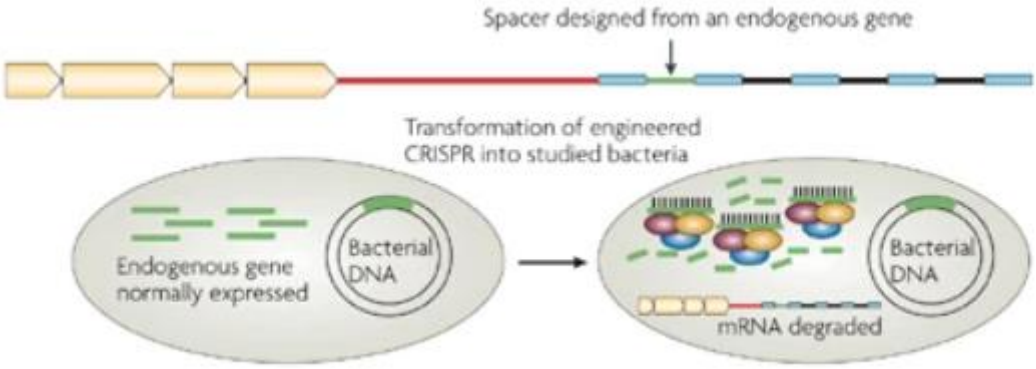


FIGURE 7 : INACTIVATION D'UN GENE PAR LA METHODE CRISPR, D'APRES SOREK ET AL. (65)

V.3. TYPAGE (FIGURE 8)

Le typage basé sur l'analyse comparative des séquences de spacers a été décrit pour la première fois en 1993 par Groenen *et al.* dans le but de différencier des souches de *Mycobacterium tuberculosis* (75). A partir de ces observations, Kemerbeek mit au point le spoligotyping, consistant à hybrider les spacers amplifiés par PCR à partir d'amorces marquées dessinées sur le DR sur une membrane contenant des sondes complémentaires aux séquences de spacers précédemment décrites chez *Mycobacterium tuberculosis* (76). La technique fut ensuite adaptée à d'autres espèces comme *Corynebacterium diphteriae* (77) et*Yersinia pestis* (78). D'autres espèces sont également sous−typées par comparaison du contenu en spacers des loci CRISPR telles que *Campylobacter jejuni* (79) et *Lactobacillus* sp. (65).

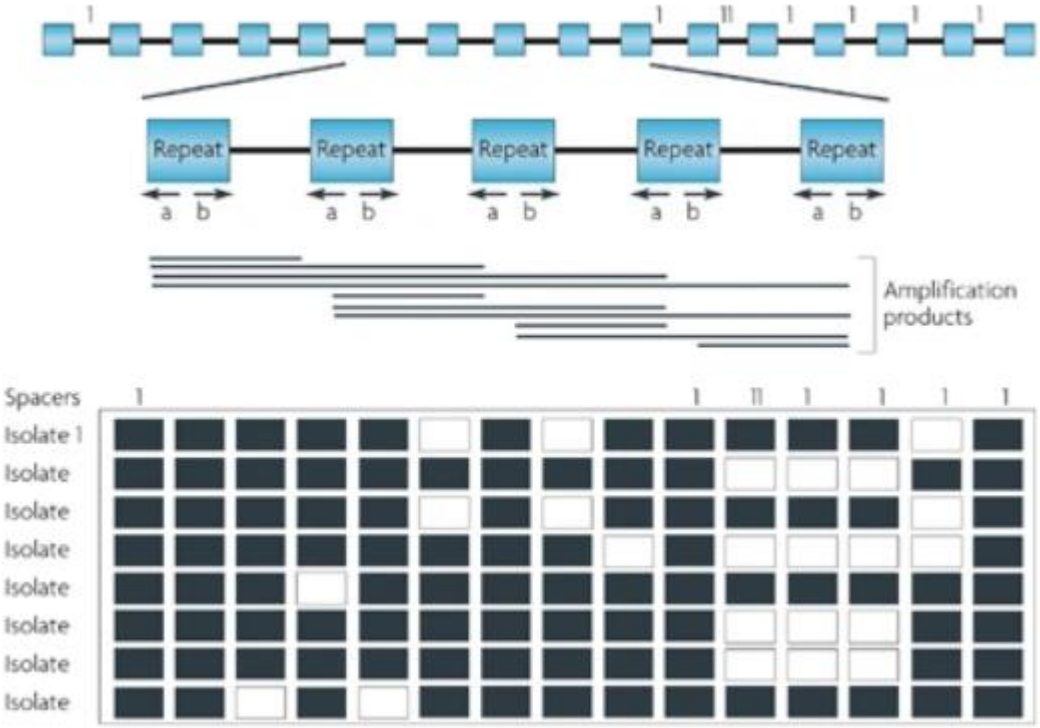


FIGURE 8 : SPOLIGOTYPING, D'APRES SOREK ET AL (65)

Aux vues des possibilités de typage offertes par le polymorphisme des loci CRISPR, l'objet de ce travail de thèse est d'étudier la région CRISPR du genre *Salmonella*, son organisation, son rôle et son potentiel dans l'avancée du typage moléculaire du genre.

TRAVAUX PERSONNELS

PUBLICATION 1

CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections.

Optimisation de la surveillance des infections à *Salmonella* par typage et sous-typage basés sur le polymorphisme des loci CRISPR

Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou−Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth−Jeanjean S, Achtman M, Brisse S, Sola C, Weill FX.

PLoS One. 2012;7(5):e36995. doi: 10.1371/journal.pone.0036995
78 citations, 11 440 vues (nov. 2016)

CONTEXTE ET OBJECTIFS

Les salmonelloses sont la première cause de diarrhées bactériennes d'origine alimentaire dans les pays industrialisés, demeurant un problème de santé publique majeur. Depuis près d'un siècle, le sérotypage reste la méthode de référence pour l'identification des salmonelles. Basée sur la recherche d'antigènes (un antigène de paroi et un ou deux antigènes flagellaires), cette technique manuelle est longue (3 à 6 jours) et très coûteuse (salaire et utilisation de >150 sérums commercialisés de typage). Près de 2 600 sérotypes ontété décrits chez le genre *Salmonella*. Leur distribution étant la base des investigations épidémiologiques, une identification rapide est primordiale. Par ailleurs, du fait de la prédominance de certains sérotypes, une étape supplémentaire de sous−typage (pulsed−field gel electrophoresis ou PFGE) est nécessaire pour une meilleure discrimination et la mise en

évidence de clones épidémiques (5 jours en plus du délai du sérotypage). Afin d'améliorer et accélérer la surveillance des salmonelloses, une nouvelle méthode de typage/sous−typage moléculaire rapide, discriminante et robuste est nécessaire. C'est dans cette optique que nous avons commencé l'étude du polymorphisme de la région CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) chez *Salmonella*. Notre étude préliminaire portant sur 400 souches représentatives de chaque espèce et sous−espèce appartenant à 56 sérotypes a mis en évidence une forte corrélation entre le contenu en spacers et le sérotype voire le sous−type d'une souche (80).

MATERIELS ET METHODES

Inventaire du contenu en spacers

L'analyse *in silico* de la région CRISPR a été réalisée sur 39 génomes publiés. L'inventaire du contenu en spacers d'une collection de 744 souches isolées entre 1885 et 2010 de sources humaines, animales ou alimentaires d'origines géographiques diverses appartenant à 130 sérotypes différents représentatifs de chaque espèce et sous espèce a été réalisé par amplification par PCR et séquençage Sanger des loci CRISPR (CRISPR1 et CRISPR2) comme décrit précédemment (80).

Développement d'une méthode de sous-typage haut débit du sérotype Typhimurium

De nouvelles amorces ont été dessinées à partir des séquences DR afin d'amplifier les spacers, l'une d'entre elles étant biotinylée. Les 72 sondes ont été dessinées à partir des séquences des spacers spécifiques du sérotype Typhimurium et greffées sur microbilles avec lesquelles les produits PCR seront hybridés. L'analyse a été réalisée sur plateforme Luminex.

RESULTATS

Organisation et structure des loci CRISPR de Salmonella

L'analyse *in silico* des 39 génomes publics a mis en évidence deux loci CRISPR distants de 20 kb. Le locus CRISPR1 était situé en aval du gène *iap* et CRISPR2 en amont du gène *ygcF*. Six différentes structures (A−F) ont été définies selon l'organisation des gènes *cas*. La séquence DR (29 pb) était conservée entre les deux loci à quelques variants SNPs près. Surles 39 génomes, 705 séquences de spacers ont été extraites. Au sein du locus CRISPR1, le nombre de spacers variait de 1 à 55 et celui de CRISPR2 variait de 0 à 32. A de rares exceptions près, la taille des spacers était de 32 pb.

Corrélation entre contenu en spacers et sérotype

Le locus CRISPR1 a pu être amplifié par PCR chez 639 des 744 souches testées ; sa taille variait de 400 pb à 3 kb. En revanche, le locus CRISPR2 était bien présent chez tous les échantillons (500 pb−3 kb), exceptés pour les isolats de la sous espèce *arizonae*. L'absence d'amplification pour les deux loci s'explique par des délétions en aval de CRISPR1 et en amont de CRISPR2. Plus de 3 800 spacers d'une longueur moyenne de 32 pb ont été identifiés chez les 744 isolats testés et les 39 génomes publiés. Le nombre de spacers variait de 1 à 124 pour CRISPR1 et était compris entre 0 (sous espèce *arizonae*) et 50 pour CRISPR2. Seulement deux populations rares ont montré une faible corrélation entre le contenu en spacers et le sérotype ou le ST (profile MLST−Séquençotype). La première rassemblait des souches de la sous−espèce *arizonae* associées à des reptiles et arborait le même spacer sur CRISPR1 en dépit de leurs sérotypes et ST différents, la seconde des sérotypes de la sous− espèce *enterica* eux aussi associés à des reptiles. Les gènes *cas* de ces deux groupes étaient

délétés. En dehors de ces deux cas particuliers, le contenu en spacers était clairement corrélé au MLST et au sérotype pour 730 des 744 souches testées (98.1%).

La majorité des spacers étaient uniques pour certains sérotypes. Cependant, les groupes de sérotypes partageant des spacers communs étaient liés génétiquement par comparaison de leur ST. Ce phénomène a été observé pour les sérotypes Typhimurium (4:i:1,2), Heidelberg (4:r:1,2) et Kisangani (4:a:1,2) : tous les spacers communs de CRISPR1 se situaient à l'extrémité *iap* alors que ceux présents du côté de la séquence leader semblaient sérotype− spécifique.

Microvariations du contenu en spacers et pouvoir discriminant

Des microvariations (duplication, triplicat, perte ou gain de spacer, variants SNPs, variant VNTR) ont été observées au sein des sérotypes monophylétiques. Ce fut le cas pour le sérotype majoritaire Typhimurium et son variant monophasique pour lesquels nous avons analysé huit génomes publics et 150 souches d'une collection bien caractérisée et isolée de 1947 à 2010.

Quarante spacers ont été identifiés pour le locus CRISPR1 (6 à 31 spacers/souche). L'ordre des spacers était strictement conservé. La plupart de ces séquences avait une longueur de32 pb (31/40) ou 33 pb (2/40). Quatre spacers avaient des variants SNPs, un autre (STM18) avait quatre variants VNTR (26 à 52 pb) et deux spacers étaient particulièrement longs, leur séquence étant le résultat de la fusion entre un spacer de 28 pb avec un DR de 14 pb suivi d'un spacer normal de 32 pb. Le nombre de spacers identifié pour le locus CRISPR2 (4 à 40 spacers/souche) était de 39. Leur taille était de 32 pb (38/39) ou 33 pb (1/39). Seulement deux d'entre eux avaient des variants SNPs.

L'ordre des spacers était strictement conservé pour tous excepté quatre allèles de CRISPR2 (8 échantillons). La variabilité du contenu en spacers des deux loci résidait dans la duplication d'un motif spacer−DR (STM5, STM8, STM22, STM28 et STMB13) et/ou la délétion d'un ou plusieurs motifs. Ainsi, nous avons obtenu 57 allèles CRISPR1 et 62 allèles CRISPR2, soit 83 allèles combinés CRISPR1−CRISPR2, offrant une meilleure discrimination que d'autres méthodes de sous typage comme le PFGE, certains profils étant population−spécifique comme observé pour les clones d'intérêt DT104 et ST313 multi résistant aux antibiotiques.

Applications

Ces travaux ont permis la mise au point de trois applications d'intérêt en microbiologie clinique :

Le « sizing » par PCR pour une comparaison rapide des isolats de salmonelle : la variabilité du nombre de spacers peut être utilisée pour tracer certaines souches d'intérêt et différencier deux souches par simple visualisation sur gel après amplification des loci CRISPR, même si deux échantillons ayant des loci CRISPR1 et 2 de même taille ne sont pas forcément identiques.

La méthode CRISPOL pour un sous typage haut débit du sérotype Typhimurium et son variant monophasique par hybridation sur bille en milieu liquide (technologie Luminex®), les sondes étant dessinées à partir de 72 des 79 spacers spécifiques de Typhimurium, ce qui permet d'obtenir le profil CRISPR des échantillons en six heures.

Les protocoles PCR spécifiques de certains sérotypes ou certains clones d'intérêt : la présence de certains spacers spécifiques de certaines souches peut servir de base pour la mise au point de PCR ciblant ces spacers, comme illustré dans la publication suivante.

CONCLUSIONS

Le pouvoir discriminant de la méthode CRISPR offre une nouvelle perspective pour la surveillance des salmonelloses avec une méthode rapide et robuste et qui pourrait parfaitement se substituer aux méthodes de références que sont le sérotypage et le PFGE, le contenu en spacers étant clairement corrélé au sérotype qui sert de base aux investigations épidémiologiques.

OPEN ACCESS Freely available online

PLoS one

CRISPR Typing and Subtyping for Improved Laboratory Surveillance of *Salmonella* Infections

Laëtitia Fabre[1], Jian Zhang[2], Ghislaine Guigon[3], Simon Le Hello[1], Véronique Guibert[1], Marie Accou-Demartin[1], Saïana de Romans[1], Catherine Lim[1], Chrystelle Roux[1], Virginie Passet[3], Laure Diancourt[3], Martine Guibourdenche[1], Sylvie Issenhuth-Jeanjean[1], Mark Achtman[4], Sylvain Brisse[3], Christophe Sola[2], François-Xavier Weill[1]*

1 Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Paris, France, **2** Institute of Genetics and Microbiology, UMR8621, IGEPE Team, Universud, CNRS Université Paris-Sud 11, Orsay, France, **3** Institut Pasteur, Genotyping of Pathogens and Public Health, Paris, France, **4** Environmental Research Institute, University College Cork, Cork, Ireland

Abstract

Laboratory surveillance systems for salmonellosis should ideally be based on the rapid serotyping and subtyping of isolates. However, current typing methods are limited in both speed and precision. Using 783 strains and isolates belonging to 130 serotypes, we show here that a new family of DNA repeats named CRISPR (clustered regularly interspaced short palindromic repeats) is highly polymorphic in *Salmonella*. We found that CRISPR polymorphism was strongly correlated with both serotype and multilocus sequence type. Furthermore, spacer microevolution discriminated between subtypes within prevalent serotypes, making it possible to carry out typing and subtyping in a single step. We developed a high-throughput subtyping assay for the most prevalent serotype, Typhimurium. An open web-accessible database was set up, providing a serotype/spacer dictionary and an international tool for strain tracking based on this innovative, powerful typing and subtyping tool.

Citation: Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, et al. (2012) CRISPR Typing and Subtyping for Improved Laboratory Surveillance of *Salmonella* Infections. PLoS ONE 7(5): e36995. doi:10.1371/journal.pone.0036995

Editor: Igor Mokrousov, St. Petersburg Pasteur Institute, Russian Federation

Received February 23, 2012; **Accepted** April 11, 2012; **Published** May 18, 2012

Funding: This work was supported by the Institut Pasteur (permanent support for Research and Programme Transversal de Recherche 354) and by the Institut de Veille Sanitaire. The MLST data are publicly available from http://mlst.ucc.ie, which is currently funded by a grant from the Science Foundation of Ireland (05/FE1/B882). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have read the journal's policy and have the following conflicts: Some of the authors (FXW, LF, VG, LD, and SB) have filed an international patent application (no. PCT/IB2008/004004) for the molecular typing and subtyping of Salmonella by identification of the variable nucleotide sequences of the CRISPR loci. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: fxweill@pasteur.fr

Introduction

Salmonellosis is one of the most common causes of food-borne diarrheal disease worldwide. Most infections are zoonotic and are transmitted from food animals to humans through the ingestion of contaminated food. In the United States, 1.4 million nontyphoidal *Salmonella* infections are thought to occur in humans annually, resulting in approximately 15,000 hospitalizations and 400 deaths [1]. An efficient surveillance system for salmonellosis is therefore crucial. Various non exclusive strategies have been developed, including sentinel surveillance, periodic population-based surveys, and laboratory-based surveillance. Laboratory-based approaches are a key component of monitoring strategies in developed countries. They require a network of clinical laboratories covering the population and referring isolates or information to a central public health reference laboratory. The speed with which public health laboratories obtain information after the onset of symptoms and the regular sharing of information between public health laboratories and epidemiologists are critical for the successful use of information to detect outbreaks early and to identify their source. The basic information currently provided by laboratories is the serotype of the isolates. Hence, each year, more than 200,000 human isolates of *Salmonella* are serotyped in the United States and Europe [2,3]. Serotyping, the reference method for *Salmonella* typing since the 1930s, is based on the determination of two surface antigens– O-polysaccharide and flagellin proteins – by agglutination with a large set of polyclonal rabbit antisera. This technique can recognize more than 2,500 serotypes [4], but its discriminatory capacity is limited, because two serotypes, Typhimurium and Enteritidis, are highly prevalent worldwide and account for most outbreaks. The sensitivity of serotyping for the detection of outbreaks involving these common serotypes, even with the use of cluster-detection algorithms, is therefore unsatisfactory [5].

Differentiation between isolates within the most common serotypes requires the use of subtyping methods, which were initially based on determination of the sensitivity of certain *Salmonella* serotypes to several bacteriophage suspensions (phage typing) [6]. DNA-based subtyping methods were subsequently developed, including, in particular, pulsed-field gel electrophoresis (PFGE) [7], which is based on analysis of the restriction pattern of high-molecular weight DNA digested with a rare-cutting restriction enzyme. Real-time subtyping methods have increased the power of laboratory-based surveillance to detect outbreaks, distinguishing them from the background of sporadic cases by identifying the phage type or molecular "fingerprint" of an

outbreak strain. PFGE is currently the gold standard method for this purpose. This real-time subtype surveillance has been implemented in the US through PulseNet, an internet-based network of public health and food regulatory agency laboratories that perform real-time standardized PFGE and submit normalized PFGE patterns or raw TIFF gel images electronically to a national database. Regular searches of this database are made, with a view to identifying clusters of identical patterns. However, PFGE has several limitations: it is a technically demanding, non automated method. This may explain why, in a study of outbreaks of food-infection occurring in the US in 2002, the median interval from the onset of symptoms to PFGE results was 18, with a period of 10 days elapsing between the submission of isolates to public health laboratories and PFGE results [8]. Furthermore, the interpretation and comparison of banding profiles is not straightforward, even with standard protocols and analysis software. The discovery of short DNA sequence repeats in the genomes of prokaryotic organisms has recently led to the development of new subtyping methods. Multilocus variable number of tandem repeats (VNTR) analysis (MLVA) is based on the number of contiguous DNA repeats present at several loci. Following a repeat-spanning PCR for each locus, the number of repeats can be determined by sequencing or inferred from electrophoresis (molecular weight being correlated with the number of repeats). An MLVA scheme for serotype Typhimurium based on the analysis of five loci (with repeat units of 6 to 33 bp) has been established and evaluated [9]. Unlike PFGE, MLVA is rapid, technically simple and suitable for the processing of large numbers of isolates. It can also distinguish between clonal isolates indistinguishable by PFGE, such as those belonging to the multidrug-resistant DT104 strain. However, this method has several drawbacks. MLVA schemes have been validated (i.e., shown to meet performance and convenience criteria, including the epidemiological concordance required for typing methods for use in bacterial epidemiology [10]) for only two *Salmonella* serotypes, Typhimurium and Enteritidis [11–13]. It requires a capillary electrophoresis system and it is difficult to size fragments accurately, as observed in multicenter studies. Finally, these repetitive DNA sequences may evolve too rapidly, leading to changes in repeat numbers during the course of an outbreak [12,13]. MLVA is therefore often used in addition to existing subtyping methods, such as PFGE or phage typing.

Jansen *et al.* identified a new family of repeated DNA sequences, named CRISPR (clustered regularly interspaced short palindromic repeats) in many prokaryotes [14]. This family is characterized by 24–47 bp DNA direct repeats (DRs), separated by variable 21–72 bp sequences called "spacers" [15,16]. A "leader sequence" and *cas* (CRISPR-associated sequence) genes are often identified adjacent to the CRISPR locus. Since the middle of the 1990s, the CRISPR locus of *Mycobacterium tuberculosis* has been extensively studied and the high degree of polymorphism of its spacer content has led to the development of a subtyping method known as spoligotyping [17]. Subtyping methods based on analyses of the spacers of CRISPR loci have since been developed for bacteria of medical interest, such as *Yersinia pestis* [18], *Corynebacterium diphtheriae* [19] and *Campylobacter* [20]. CRISPR seem to confer resistance to foreign DNA, such as plasmids and phages, and the newly integrated spacers are derived from the invading DNA [21,22]. Interestingly, these spacers are integrated into the CRISPR locus in a polarized manner [18,21]. The spacer content of a strain therefore reflects previous DNA introductions and can provide evolutionary information.

Several studies have reported the presence of two CRISPR loci in *Salmonella* [14,23,24]. We previously showed, in a preliminary study of 400 *Salmonella enterica* and *Salmonella bongori* reference strains and isolates from 56 serotypes, that CRISPR polymorphisms (i.e., spacer content) were strongly correlated with serotype and subtype [25]. Two studies recently suggested that CRISPR loci might provide information useful for typing [26,27]. However, these studies considered only a limited number of serotypes from a single geographic area.

We aimed to demonstrate that CRISPR polymorphism analysis is an efficient and powerful alternative to both serotyping and PFGE methods. We first analyzed the spacer content of the two *Salmonella* CRISPR loci in a large global collection of reference strains and well documented isolates belonging to 130 serotypes of all species and subspecies, focusing particularly on the serotypes most frequently involved in human infections. Analysis of the distribution of the >3,800 unique spacers identified showed that spacer content was strongly correlated with both serotype and multilocus sequence typing (MLST) type. Furthermore, the microevolution of spacer content facilitated the robust discrimination of subtypes within most serotypes, including the most prevalent serotypes, Typhimurium and Enteritidis.

We also present here three applications of CRISPR polymorphisms for *Salmonella* surveillance. In particular, we describe a novel high-throughput subtyping assay for serotype Typhimurium (and its emerging monophasic 1,4,[5],12:i:- variant). This bead-based liquid hybridization assay is both rapid and easy to carry out, and is therefore highly suitable for use in public health laboratories.

Results

In silico Analysis of the Organization and Structure of CRISPR Loci in *Salmonella*

Two CRISPR loci, CRISPR1 and CRISPR2, were separated by less than 20 kb in all 39 complete genomes of *S. enterica* and *S. bongori* analyzed (Figure 1, Table 1). The CRISPR1 locus was located downstream from the *iap* gene, whereas CRISPR2 was located upstream from the *ygcF* gene. The ordered CRISPR-associated (*cas*) genes belonging to the Ecoli subtype defined by Haft *et al.* [28] were located between the CRISPR loci: *cas2*, *cas1*, *cse3*, *cas5e*, *cse4*, *cse2*, and *cas3*. Following the *cas* genes were *sopD* (encoding a secreted effector protein), *cysH* (encoding a phosphoadenosine phosphosulfate reductase), *cysI*, *cysJ* (both encoding sulfite reductase subunits), *ptpS* (encoding pyruvyl tetrahydrobiopterin synthase) and an ORF encoding a putative metal-dependent hydrolase. Structure A was the most frequent, and was observed in 26 (67%) genomes of *S. enterica* subsp. *enterica* (including representative serotype Typhimurium strain LT2), *S. enterica* subsp. *diarizonae* and *S. bongori*. Structure B, which was found only in serotype Choleraesuis SC-B67, differed from structure A by an insertion sequence, IS*Sen1*, immediately upstream from CRISPR2. Structure F, which was found in nine genomes of *S. enterica* subsp. *enterica* (including representative serotype Typhi strain Ty2) differed from structures A and B in having a different orientation of the *cas3* gene and in terms of the degree of similarity of Cas proteins (40 to 85%, depending on the Cas proteins considered; data not shown) [23]. In structures C, D and E (found in *S. enterica* serotype Paratyphi B strain SPB7, *S. enterica* subsp. *arizonae* serotype 62:z4,z23:- strain CDC346-86, and *S. enterica* serotype Javiana strain GA_MM04042433, respectively), there was a deletion beginning at the end of the last DR of CRISPR1 and encompassing all *cas* genes with the exception of a 5′ remnant of *cas3*, which was in the same orientation as that in structure A.

The DRs of both CRISPR loci were conserved. They were 29 bp long and had the consensus sequence 5′-CGGTTTATCCCCGCTGGCGCGGGGAACAC-3′. However,

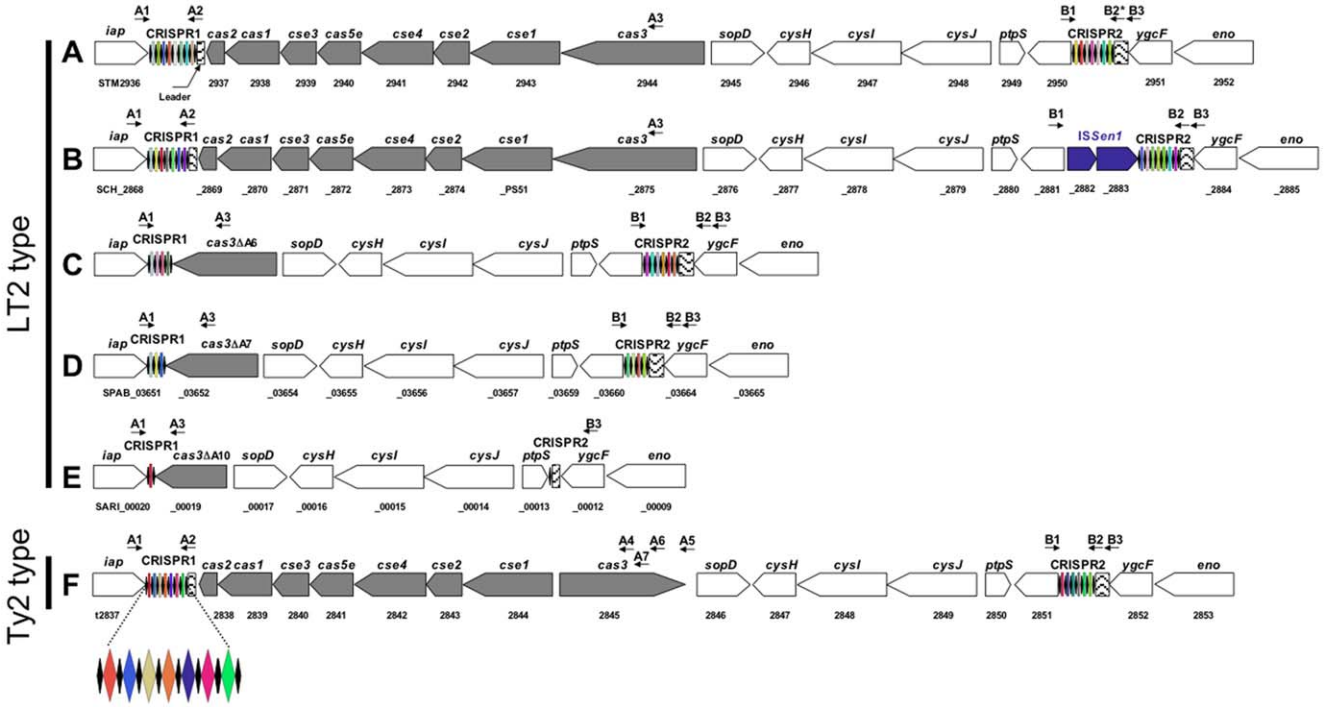


Figure 1. CRISPR/Cas system structures from 39 available genome sequences for *S. enterica* and *S. bongori*. Two CRISPR loci (CRISPR1 and CRISPR2) are present in all genomes. The CRISPR-associated (cas) genes *cas2*, *cas1*, *cse3*, *cas5e*, *cse4*, *cse2*, and *cas3 genes* of the "Ecoli" subtype [28] are located between the CRISPR loci. The most frequent structure, A, is represented by *S. enterica* serotype Typhimurium strain LT2. Structures B to E are represented by *S. enterica* serotypes Choleraesuis strain SC-B67, Javiana strain GA_MM04042433, Paratyphi B strain SPB7, and *S. enterica* subsp. *arizonae* serotype 62:z4,z23:- strain CDC346-86, respectively. Structure F is represented by *S. enterica* serotype Typhi strain Ty2. Black diamonds represent direct repeats, with colored diamonds indicating spacers. The CRISPR1 locus of serotype Typhi strain Ty2 is enlarged. The primers used to amplify and sequence the CRISPR loci for the spacer inventory are indicated by horizontal arrows.
doi:10.1371/journal.pone.0036995.g001

some DR variants carrying single-nucleotide polymorphisms (SNPs) with respect to the consensus sequence were observed (Table S1).

There were 705 unique spacers between the DRs in the two loci from the 39 available genomes (Table S2). Depending on the genome, the number of CRISPR1 spacers varied from 1 to 55 (mean 18.6±standard deviation 13.6) and of the number of CRISPR2 spacers varied from 0 (subsp. *arizonae*) to 32 (15.0±9.8). Spacers were typically 32 bp long (681/705). One was 29 bp long, two were 31 bp long, sixteen were 33 bp long, one was 38 bp long (spacer STM18var2, which contained a VNTR) one was 50 bp long, one was 72 bp long and two were 74 bp long (spacers STM7A/7B and STM7A/7Bvar2 of serotype Typhimurium) (see below). Some spacers were common to different serotypes.

While our study was underway, two CRISPR databases (CRISPRdb and CRISPI) went online [15,29]. These generalist databases containing >1500 prokaryote genomes incorporate various bioinformatics tools, including one for identifying CRISPR sequences in a selected genome. The application of this tool to *Salmonella* genomes resulted in incorrect results for four to six of the 17 genomes present in both databases. CRISPRfinder did not detect the short CRISPR2 locus of serotype Typhi strains Ty2 and CT18, which have a unique spacer (EntB0var1) between two DRs (DR27 and DR), one of which is degenerate (identity of 20/29 bp). CRISPRfinder detected three CRISPR in serotype Typhimurium strain LT2 and serotype Heidelberg strain SL476. The CRISPR1 locus was actually artificially split into two CRISPR, due to the presence of an unusual fused spacer-DR unit (STM7A/7B, see below). The CRISPI tool detected no CRISPR in four genomes and only one CRISPR in two others. The CRISPR loci identified in all six genomes were short (1 to 6 spacer-DR units) in our study, confirming that the CRISPI tool is not suitable for detecting short CRISPR loci. Thus, although bioinformatics tools are undoubtedly useful for screening for CRISPR within genomes, careful manual inspection is required to complete the analysis for a given species.

Spacer Content is Strongly Associated with Serotype and MLST

PCR amplification of CRISPR1 with primers A1 and A2 generated a product of between 400 bp and 3 kb in size, in 639 of 744 strains and isolates. By contrast PCR amplification of CRISPR2 with primers B1, B2 and B3 generated a product of between 500 bp and 3 kb in size in all but subsp. *arizonae* strains and isolates (Figure 1, Tables 2 and S2). Various deletions downstream from CRISPR1 or upstream from CRISPR2 were responsible for amplification failure (see below and Table 3). In one reference strain of serotype Mbandaka, PCR was unsuccessful because the CRISPR1 locus was very large (>6 kb) and contained 124 spacer-DR units.

More than 3,800 different spacers (mean length of 32 nucleotides) were identified in the 39 available genomes and 744 strains and isolates tested (Tables S2, S3, S4). The number of spacers present in a given strain ranged from 1 to 124 for CRISPR1, and from 0 (subsp. *arizonae*) to 50 for CRISPR2 (Table S3). Two rare groups of strains displayed low levels of correlation between spacer content and serotype or MLST type. First, all the reptile-associated subsp. *arizonae* strains had the same single

Table 1. CRISPR loci detected in the 39 available genomes of *Salmonella*.

Strain	Source (accession no.)	CRISPR structure	CRISPR1 coordinates	CRISPR2 coordinates
***S. enterica* subsp. *enterica* serotype:**				
Agona strain SL483	GenBank (CP001138)	A	2988105-2989231 (18)	3005517-3006033 (8)
Choleraesuis strain SC-B67	GenBank (AE017220)	B	3031533-3031805 (4)	3049243-3049698 (7)
Dublin strain CT_02021853	GenBank (CP001144)	A	3121101-3121251 (2)	3137409-3137742 (4)
Enteritidis strain P125109	GenBank (AM933172)	A	2961370-2961886 (8)	2978038-2978677 (10)
Gallinarum strain 287/91	GenBank (AM933173)	A	2952175-2952325 (2)	2968478-2969117 (10)
Hadar strain RI_05P0661	GenBank (ABFG00000000)	F	NA (28)	NA (29)
Hadar strain ''Sanger''1	Sanger Institute2	F	NA (28)	NA (30)
Heidelberg strain SL476	GenBank (CP001120)	A	3051217-3052879 (28)	3069137-3070263 (18)
Heidelberg strain SL4861	GenBank (ABEL00000000)	A	NA (26)	NA (18)
Infantis ''Sanger'' 1	Sanger Institute2	A	NA (31)	NA (14)
Javiana strain GA_MM040424331	GenBank (ABEH00000000)	C	NA (6)	NA (12)
Kentucky strain CDC1911	GenBank (ABEI00000000)	A	NA (19)	NA (18)
Kentucky strain CVM291881	GenBank (ABAK00000000)	A	NA (18)	NA (17)
Newport strain SL254	GenBank (CP001113)	F	3054859-3056473 (26)	3073142-3074328 (19)
Newport strain SL3171	GenBank (ABEW00000000)	A	NA (12)	NA (18)
Paratyphi A strain ATCC 9150	GenBank (CP000026)	F	2889569-2889902 (5)	2906453-2906664 (3)
Paratyphi A strain AKU_12601	GenBank (FM200053)	F	2885105-2885560 (7)	2902111-2902322 (3)
Paratyphi B strain SPB7	GenBank (CP000886)	D	3041329-3041479 (2)	3050804-3051137 (5)
Paratyphi C strain RKS4594	GenBank (CP000857)	A	3010604-3011242 (10)	3028681-3029258 (9)
Saintpaul strain SARA231	GenBank (ABAM00000000)	A	NA (13)	NA (26)
Saintpaul strain SARA291	GenBank (ABAN00000000)	A	NA (19)	NA (7)
Schwarzengrund strain CVM19633	GenBank (CP001127)	F	2981949-2982709 (12)	2999469-3000534 (17)
Schwarzengrund strain SL4801	GenBank (ABEJ00000000)	F	NA (12)	NA (17)
Tennessee strain CDC07-01911	GenBank (ACBF00000000)	A	NA (41)	NA (21)
Typhi strain CT18	GenBank (AL627276)	F	2926182-2926567 (5)	2943123-2943212 (1)
Typhi strain Ty2	GenBank (AE014613)	F	2912041-2912461 (6)	2929017-2929106 (1)
Typhimurium strain LT2	GenBank (AE006468)	A	3076611-3078147 (23)	3094279-3096260 (32)
Typhimurium strain SL1344	GenBank (FQ312003)	A	3099172-3100159 (15)	3116291-3117723 (22)
Typhimurium strain D23580	GenBank (FN424405)	A	3069598-3071012 (22)	3087144-3088271 (18)
Typhimurium strain 14028S	GenBank (CP001363)	A	3096848-3098323 (23)	3114455-3116070 (25)
Typhimurium strain T000240	GenBank (AP011957)	A	3100041-3101393 (21)	3117525-3119506 (32)
Typhimurium strain DT21	Sanger Institute2	A	NA (21)	NA (26)
Typhimurium strain NCTC 133481	Sanger Institute2	A	NA (10)	NA (26)
Virchow strain SL4911	GenBank (ABFH00000000)	A	NA (55)	NA (16)
Weltevreden strain HI_N05-5371	GenBank (ABFF00000000)	A	NA (40)	NA (26)
4,5,12:i:- strain CVM237011	GenBank (ABAO00000000)	A	NA (23)	NA (26)
***S. enterica* subsp. *arizonae* serotype:**				
62:z4,z23:- strain CDC346-86	GenBank (CP000880)	E	25560-25471 (1)	17801-17773 (0)
***S. enterica* subsp. *diarizonae* serotype:**				
61:l,v:1,5,7 strain CDC01-00051	Washington State University3	A	NA (30)	NA (1)
***S. bongori* serotype:**				
66:z41:- strain 12419	Sanger Institute2	A	2791744 -2792992 (20)	2808974 -2810039 (20)

[1]Genomes not finished or annotated.
[2]These data were provided by Dougan's group at the Wellcome Trust Sanger Institute and could be obtained from http://www.sanger.ac.uk/resources/downloads/bacteria/salmonella.html.
[3]Available from http://genome.wustl.edu/genomes/.
[4]NA, not applicable; the number of spacers per locus is indicated in brackets.
doi:10.1371/journal.pone.0036995.t001

Table 2. Primers used for the spacer content inventory.

Primer	Sequence 5′-3′[1]	Coordinates in *Salmonella* genomes[2]	Function[3]
A1	GTRGTRCGGATAATGCTGCC	AE006468 (3076537-3076556) AE014613 (2911967-2911986)	Forward primer for amplification of CRISPR1 or for combined amplification of both CRISPR1 and CRISPR2 loci
A2	CGTATTCCGGTAGATBTDGATGG	AE006468 (3078306-3078284) AE014613 (2912608-2912586)	Reverse primer for amplification of CRISPR1 in 640 (86%) of the isolates
A3	CTATTTTGGRCTRCCGACRATG	AE006468 (3085738-3085717)	Reverse primer for amplification of CRISPR1 in 60 (8.1%) isolates of type A structure
A4	GCAATCGGAGCGATTGATGGC	AE014613 (2920120-2920100)	Reverse primer for amplification of CRISPR1 in 29 (3.9%) isolates of type F structure
A5	TCAACACTCTCTTCACCCAG	AE014613 (2921235-2921216)	Reverse primer for amplification of CRISPR1 in 7 (0.9%) isolates of type F structure
A6	TAACCAGCCCTCTTCTGCCTG	AE014613 (2920910-2920892)	Reverse primer for amplification of CRISPR1 in 2 (0.3%) isolates of type F structure
A7	CGCATCATCAACCGTGTTGCG	AE014613 (2920524-2920504)	Reverse primer for amplification of CRISPR1 in 6 (0.8%) isolates of type F structure
B1	GAGCAATACYYTRATCGTTAACGCC	AE006468 (3094155-3094179) AE014613 (2928893-2928917)	Forward primer for amplification of CRISPR2
B2	GTTGCDATAKGTYGRTRGRATGTRG	AE006468 (3096328-3096303) AE014613 (2929174-2929150)	Reverse primer for amplification of CRISPR2 for the isolates belonging to subspecies other than arizonae and diarizonae
B3	CTGGCGGCTGTCTATGCAAAC	AE006468 (3096602-3096582) AE014613 (2929448-2929428)	Reverse primer for single amplification of CRISPR2 for the isolates belonging to all subspecies or reverse primer for combined amplification of both CRISPR1 and CRISPR2 loci

[1]Degenerate positions: R = G or A; Y = T or C; M = A or C; K = G or T; D = G or A or T; B = G or T or C.
[2]AE006468, serotype Typhimurium LT2 strain; AE014613, serotype Typhi Ty2 strain.
[3]The primer pairs used for CRISPR1 amplification for each of the 744 strains are indicated in Table S2.
doi:10.1371/journal.pone.0036995.t002

CRISPR1 spacer, despite the diversity of their serotypes and STs. Second, some reptile-associated subsp. *enterica* serotypes, such as Urbana, Johannesburg, Reading, Pomona, Gueuletapee, Rubislaw, Goettingen and Sandiego had a limited set of spacers shared between these serotypes that belonged to a highly recombinogenic group known as clade B [30] or lineage 3 [31]. Both groups of strains displayed deletions (ΔA10 for subsp. *arizonae* serotypes and ΔF4 and ΔF5 for clade B subsp. *enterica* serotypes) of the *cas* genes (see below). However, with the exception of these two groups, spacer content was strongly correlated with serotype and/or MLST type for 730 of 744 (98.1%) strains. Moreover, for polyphyletic serotypes comprising unrelated MLST groups, spacer content was strongly correlated with the population structure defined by MLST (Table S2). For example, the CRISPR data for three recently described genetic lineages of serotype Newport [32], gave correct serotype recognition and genetic lineage assignment (Tables 4 and S5).

Most of the spacers were unique to particular serotypes. The degree of spacer sharing varied among groups of serotypes identified as closely related on the basis of MLST. For example, in serotypes such as Typhimurium (4:i:1,2) Heidelberg (4:r:1,2) and Kisangani (4:a:1,2), all the spacers on the *iap* gene side tended to be the same whereas the spacers present on the leader side tended to be more serotype-specific. For isolates of the ST11 group of serotype Enteritidis (9,12:g,m:-) and the closely related Gallinarum (9,12:-:-) and Dublin (9,12:g,p:-) serotypes, only a subset of common spacers was identified (Tables 5 and S6). This was also the case for the complex group of "bioserotypes" Paratyphi C, Choleraesuis (*sensu stricto* and variant Kunzendorf), and Typhisuis, which have the antigenic formula 6,7:c:1,5 in common, indicating descent from a common ancestor, consistent with MLST data (Figure 2). The presence of IS*Sen1* at the same position upstream from CRISPR2 in these bioserotypes is also consistent with the hypothesis of a common ancestor. By contrast, the polyphyletic serotype Decatur (formerly known as serotype Choleraesuis variant Decatur which also has an antigenic formula of 6,7:c:1,5) did not have an IS*Sen1* element upstream from CRISPR2 and included various spacers not found in Paratyphi C, Choleraesuis and Typhisuis (Table S2).

Deletions Downstream from CRISPR1 and Upstream from CRISPR2 in some *Salmonella* Populations

For representative strains or isolates for which no PCR was obtained from CRISPR1 or CRISPR2, we carried out a long-range PCR encompassing both CRISPR loci, with primers A1 and B3. Amplicons were obtained from all isolates and were between 10 kb and 20 kb in size. DNA sequencing of the ends of the PCR products showed that amplification failure resulted principally from large deletions affecting the *cas* genes, ending at the *cas3* gene and preventing the annealing of the A2 primer (Table 3). For subsp. *arizonae* and serotypes Paratyphi B and Javiana, these deletions were consistent with data for the three corresponding available genome sequences and from Fricke *et al.* [24]. Such deletions were also observed in other serotypes or populations within a single serotype, but they were of different types. We therefore designed and validated new CRISPR1 reverse primers binding to the residual region of the *cas3* gene from structure A (A3 and A4) or F (A5 to A7) (Table 2). Due to a deletion upstream from CRISPR2, the B1 primer did not bind to DNA from subsp. *arizonae*. However, the available genome sequence and sequencing of the long PCR fragment generated with the A1 and B3 primers

Table 3. Serotypes with deletions of the Cas machinery.

Name	Deleted cas2-cas3 region[1]	*cas3* remnant size in bp[2]	Serotypes with such deletion (no. of isolates)
Type A CRISPR structure			
ΔA1	3078148-3084080	2622	Stourbridge (7)
ΔA2	3078148-3084337	2365	Kundunchi (1)
ΔA3	3078148-3084606	2096	Choleraesuis (2)
ΔA4	3078148-3084649	2053	Napoli (1)
ΔA5	3078148-3084656	2046	Mbandaka (2)
ΔA6	3078148-3084763	1939	Javiana (4)
ΔA7	3078148-3084995	1707	Abony (1), Paratyphi B (25)
ΔA8	3078148-3085040	1662	Enteritidis (1)
ΔA9	3078148-3085289	1413	subsp. indica 6,7:z41:1,7 (1)
ΔA10	3078148-3085385	1317	All subsp. arizonae (5)
ΔA11	3078148-3085559	1143	Enteritidis (1)
ΔA12	3078148-3085681	1021	Worthington (12)
Type F CRISPR structure			
ΔF1	2912462-2919593	1406 (1899)	subsp. houtenae 48:g,z51:- (1)
ΔF2	2912462-2919670	1329 (1822)	Portedeslilas (1)
ΔF3	2912462-2919881	1118 (1611)	Newport (5)
ΔF4	2912462-2919890	1109 (1602)	Johannesburg (1), Urbana (2)
ΔF5	2912462-2919901	1098 (1591)	9,12:l,v:- (1), Arechavaleta (1), Brandenburg (3), Chester (1), Glostrup (1), Goettingen (1), Gueuletapee (1), Maracaibo (1), Miami (4), Panama (3), Pomona (2), Reading (1), Rubislaw (1), Sandiego (1)
ΔF6	2912462-2920094	905 (1398)	Albany (1), Duesseldorf (1)
ΔF7	2912462-2920154	845 (1338)	Choleraesuis (1)
ΔF8	2912462-2920810	189 (682)	subsp. houtenae 1,40:z4,z24:- and 44:a:- (2)
ΔF9	2912462-2920937	62 (555)	Bardo (1), Newport (1)
ΔF11	2912462-2921077	0 (415)	Carrau (1), Madelia (1)
ΔF12	2912462-2921188	0 (303)	Newport (3)

[1]The coordinates of the deleted regions of isolates with type A CRISPR structure and those of isolates with type F CRISPR structure are based on *S. enterica* serotype Typhimurium strain LT2 (GenBank AE006468) and serotype Typhi strain Ty2 (GenBank AE014613) genomes, respectively. The reverse primers used for these isolates are indicated in Table 2.

[2]The *cas3* gene of serotype Typhimurium LT2 strain is 2663 bp in size, whereas that of serotype Typhi strain Ty2 is 2207 bp in size, due to a frameshift leading to a premature stop codon. The sizes of the *cas3* gene remnant are shown in brackets, not taking into account the serotype Typhi-specific frameshift.

doi:10.1371/journal.pone.0036995.t003

from four other isolates showed that there was only one DR (DR68) and no CRISPR2 spacer in this subspecies.

Microvariations of the Spacer Content Discriminate below the Serotype Level

We observed stable microvariation (duplication, triplication, loss or gain of spacers, presence of SNP variant spacers or VNTR variant spacers) within the strains of monophyletic serotypes. This was the case for the most prevalent serotype worldwide, Typhimurium and its monophasic 1,4,[5],12:i:- variant, for which we analyzed eight genomes and 150 well characterized isolates collected between 1947 and 2010 (Table S7).

In silico analysis of genome sequences identified 28 unique spacers within CRISPR1. This number increased to 40 after analysis of the additional 150 isolates (between 6 and 31 spacers per isolate). The order of spacers was strictly conserved. Most were 32 bp (31/40) or 33 bp (2/40) long. Four of the 40 spacers had SNP variants, one (STM18) had four VNTR variants (26 to 50 bp), and the 74 bp STM7A/7B and STM7A/7Bvar2 spacers contained a 28 bp spacer fused to 14 bp from the end of a DR followed by a classical 32 bp spacer (Figure S1). This fusion spacer-DR unit may have been generated accidentally during the process of spacer acquisition.

In silico analysis identified 36 unique spacers within CRISPR2. This number increased to 39 upon sequencing of the additional 150 isolates (between 4 and 40 spacers per isolate). All spacers were 32 bp (38/39) or 33 bp (1/39) long. The 39 spacers included only two variant spacers (SNP variants, all the other spacers being unrelated). As for CRISPR1, the variability of CRISPR2 was due to duplication of a single spacer (STMB13)-DR unit and/or to deletion of single or contiguous spacer-DR units.

The order of the spacers was strictly conserved in all but four alleles of CRISPR2 (8 isolates). The variability of CRISPR1 and CRISPR2 spacer content resulted from the duplication of single spacer (STM5, STM8, STM22, STM28, and STMB13)-DR units and/or the deletion of single or contiguous spacer-DR units. This microvariation resulted in 57 CRISPR1 and 62 CRISPR2 alleles or into 83 CRISPR1-CRISPR2 combined alleles, thus providing a higher resolution than other subtyping methods, such as PFGE. Particular populations, such as multidrug-resistant (MDR) DT104

Table 4. Comparison of CRISPR1 spacer content with the population structure of *S. enterica* serotype Newport, as assessed by MLST.

Lineage	Strain	MLST[1]	CAS type(deletion)[2]	CRISPR1 spacer content[3]
Newport-I	00-4093	ST156	Ty2 (ΔF3)	Ind1var1-H1-H2-H3-H4-H5-H7-H8-H9-N32-H14-N33-N51-N52-N53-Bovis3-H15-N54-N55-DueB1-N56-N60-N61-N57-N58-N59
	01-2174	ST156	Ty2 (ΔF3)	Ind1var1-H1-H2-H3-H4-N55-DueB1-N56-N60-N61-N57-N58-N59
	00-973	ST166	Ty2 (ΔF3)	Ind1var1-H1-H2-H3-H4-H5-H7-H8-H9-N32-H14-N33-N51-N52-N53-Bovis3-H15-N54-N55-DueB1-N56-N62-N60-N61-N57-N58-N59
	04-2487	ST166	Ty2 (ΔF3)	Ind1var1-H1-H2-H3-H4-H5-H7-H8-H9-N32-H14-N33-N51-N52-N53-Bovis3-H15-N54-N55-DueB1-N56-N62-N60-N61-N57-N59
	39/64	ST166	Ty2 (ΔF3)	Ind1var1-H1-H2-H3-H4-H5-H7-H8-H9-N32-H14-N33-N51-N52-N53-Bovis3-H15-N54-N55-DueB1-N56-N60-N61-N57-N58-N59
Newport-II	10/66	ST45	Ty2	H7-N1-H8-N2-N3-N4-N5-N31-N6-N7-N8-N9-N10-N22-N12-N14-N15-N16-N17-N18-N21
	00-4165	ST45	Ty2	H7-N1-H8-N2-N3-N4-N5-N6-N7-N8-N9-N10-N22-N23-N24-N11-N12-N13-N14-N15-N16-N17-N18-N19-N20-N21
	02-7891	ST45	Ty2	H7-N1-H8-N2-N3-N4-N5-N6-N7-N8-N9-N10-N22-N23-N24-N11-N12-N13-N14-N15-N16-N17-N18-N19-N20-N21
	04-9597	ST45	Ty2	H7-N1-H8-N2-N3-N4-N5-N6-N7-N8-N9-N10-N22-N23-N24-N11-N12-N13-N14-N15-N16-N17-N18-N19-N20-N21
	SL254	ST45	Ty2	H7-N1-H8-N2-N3-N4-N5-N6-N7-N8-N9-N10-N22-N23-N24-N11-N12-N13-N14-N15-N16-N17-N18-N19-N20-N21
	01-2010	ND	Ty2	H7-N1-H8-N2-N3-N4-N5-N6-N7-N8-N9-N10-N22-N23-N24-N11-N12-N13-N14-N15-N16-N17-N18-N19-N20-N21
	03-3224	ND	Ty2	H7-N1-H8-N2-N3-N4-N5-N6-N7-N8-N9-N10-N22-N23-N24-N11-N12-N13-N14-N15-N16-N17-N18-N19-N20-N21
	10/56	ST46	Ty2	H7-N1-H8-NB25var1-N26-N27-N2-N28-N3-N4-N5-N31-N6-N7-N8-N24-N11-N12-N13-N29-N30-N16-N17-N18-N19-N21
	50K	ST31	Ty2 (ΔF12)	H7-N1-H8-NB25var1-N26-N27-N2-N28-N3-N4-N5-N31-N6-N7-N8-N9-N23-N24-N11-N12-N13-N29-N30-N14-N15-N16-N17-N18-N19-N20
	04-1198	ST31	Ty2 (ΔF12)	H7-N1-N17-N18-N19-N20
	50/3	ST31	Ty2 (ΔF12)	H7-N1-H8-NB25var1-N26-N27-N2-N28-N3-N4-N5-N31-N6-N7-N8-N9-N23-N24-N29-N30-N14-N15-N16-N17-N18-N19-N20
	2/58	ST211	Ty2 (ΔF9)	H7-N1-H8-NB25var1-N26-N27-N2-N28-N3-N19-N20-N21
Newport-III	05-0815	ST118	LT2	STM1var1-N45-N34-N35-N36-N37-N38-N39-N40-N46-N47-N48-N49-N42-N43-N44
	4/51	ST118	LT2	STM1var1-N45-N35-N36-N37-N38-N39-N40-N46-N47-N48-N49-N50-N42-N43
	03-8748	ST118	LT2	STM1var1-N45-N34-N35-N36-N46-N47-N48-N42-N43-N44
	SL317	ST5	LT2	STM1var1-N34-N35-N36-N37-N38-N39-N40-N41-N42-N43-N44

[1]ND, Not done.
[2]The deletions are named according to Table 3.
[3]Due to space constraints, the spacer names Newp and Had are abbreviated to N and H, respectively.
doi:10.1371/journal.pone.0036995.t004

isolates [33], African MDR ST313 isolates [34], and ciprofloxacin-resistant isolates [35], each had typical CRISPR alleles.

In vitro stability experiments showed no difference in spacer content for either of the CRISPR loci between the five original serotype Enteritidis isolates and their derived cultures after one month or two months with daily passages. Furthermore, the genome sequences of widely used laboratory strains of serotype Typhimurium, LT2 (isolated in 1947) and SL1344 (isolated in the 1970s), showed these strains to have the same spacer content (15 to 32 per locus) as strains LT2 and SL1344 available in our laboratory. The stability of this marker was also assessed by performing the microbead-based CRISPOL assay (see below) one year later on fresh cultures (grown from single colonies) of the 150 serotype Typhimurium and monophasic 1,4,[5],12:i:- isolates (mean of 41±9 spacers per isolate). One isolate had a discordant CT with respect to the initial spacer content determined by sequencing. We resequenced both CRISPR loci in this subcultured isolate and found that a single CRISPR1 spacer-DR unit had been lost. During the systematic CRISPOL testing of all serotype Typhimurium and monophasic isolates obtained between January 1 2010 and July 7 2010, 43 duplicate and four triplicate isolates obtained from the same patients on different days were analyzed (see below). All but one had concordant CTs. The final isolate had lost a CRISPR1 spacer-DR unit that was present in the other two isolates from the same patient. Thus, spacer content is, at least in serotypes Enteritidis and Typhimurium, stable enough for use in surveillance and outbreak investigation. However, although rare and often minor (single spacer variant of the original CT in both cases detected), CRISPR variation may occur and, whatever its origin, occurring before or during carriage in the patient or subculture in the laboratory, should be taken into account when defining outbreak-related CTs.

We assessed the discriminatory power of the method, an important parameter for surveillance purposes, by comparing CRISPR spacer diversity with classical first-line (PFGE, phage typing) and second-line (MLVA) subtyping methods for a subset of

Table 5. CRISPR1 spacer content in various O:9 and O:2 serotypes.

Serotype	Antigenic formula	Biotype	MLST	No. of isolates	CRISPR1 spacer content
Enteritidis	9,12:g,m:-		ST11 group1		
				1	Ent1
				2	Ent1-Dub1-Ent3-Ent8
				2	Ent1-Ent2-Ent3-Ent4-Ent4-Ent5-Ent6-Ent7-Ent8
				1	Ent1-Ent2-Ent3-Ent4-Ent5
				76	Ent1-Ent2-Ent3-Ent4-Ent5-Ent6-Ent7-Ent8
				7	Ent1-Ent2-Ent3-Ent4-Ent5-Ent6-Ent7-Ent9-Ent8
				10	Ent1-Ent2-Ent3-Ent4-Ent5-Ent7-Ent8
				7	Ent1-Ent2-Ent3-Ent4-Ent5-Ent7-Ent9-Ent8
				1	Ent1-Ent2-Ent3-Ent4-Ent5var1-Ent6-Ent7-Ent9-Ent8
				92	Ent1-Ent2-Ent3-Ent5-Ent6-Ent7-Ent9-Ent8
				1	Ent1-Ent2-Ent5-Ent6
				1	Ent1-Ent2-Ent5-Ent6-Ent7-Ent8
				1	Ent1-Ent2var1-Ent3-Ent4-Ent5-Ent6-Ent7-Ent8
				51	Ent1-Ent2var1-Ent3-Ent4-Ent5-Ent6-Ent7-Ent9-Ent8
				2	Ent1-Ent3-Ent5-Ent6-Ent7-Ent9-Ent8
			Other STs		
			ST180	1	Ent1-Ent5-Ent6-Ent10-Ent11-Ent7-Ent9-Ent12
			ST180	1	Ent1-Ent5-Ent6-Ent10-Ent11-Ent7var1-Ent12
			ST180	1	Ent1-Ent5-Ent6-Ent10-Ent7-Ent12
			ST6	1	Ent16-Ent17-Ent18-Ent19-Ent20?//?Ent353
			ST77	1	STM1-Ent13-EmeB14-Ent14-CholB19-Ent15
	9,12:g,m,p:-		ST74	1	Ent1-Ent5-Ent6-Ent10-Ent11-Ent7-Ent12
	9,12:g,m,p:-		ST74	1	Ent1-Ent5-Ent6-Ent11-Ent7-Ent12
	9,12:g,m:1,7		ST746	1	Ent36-Mba9-Ent37-Ent38
Blegdam	9,12:g,m,q:-		ST739 (ST11 SLV)	1	Ent1-Ent2-Ent3-Ent5-Ent6-Ent7-Ent9-Ent8
Rosenberg	9,12:g,z85:-		ST11	2	Ent1-Ent2-Ent3-Ent4-Ent5-Ent6-Ent7-Ent8
			ST11	1	Ent1-Ent2-Ent3-Ent4-Ent4-Ent5-Ent6-Ent7-Ent8
Dublin	9,12:g,p:-		ST10	4	Ent1-Dub1
			ST73	2	Ent1-Dub1
Gallinarum	9,12:-:-				
		Gallinarum	ST78	7	Ent5-Ent6var1
		Pullorum	ST92	3	Ent3-Ent4
		Pullorum	ST747 (ST92 SLV)	1	Ent3-Ent4
		Pullorum	ST92	1	Ent1-Ent3-Ent4
		Duisburg4	ST762	2	Ent1-Ent3-Ent4
Nitra	2,12:g,m:-		ST11	2	Ent1-Ent2-Ent3-Ent4-Ent5-Ent6-Ent7-Ent8
			ST11	1	Ent1-Ent2-Ent3-Ent5-Ent6-Ent7-Ent9-Ent8
Kiel	2,12:g,p:-		ST10	3	Ent1-Dub1

[1]ST (sequence type) 11 group consists of ST11 and its single-locus variants (SLV).
[2]Includes the 5 ST136 "Danysz"» strains used as rodenticides.
[3]Ent20−//−Ent35, 15 unique spacers are located between Ent20 and Ent35 (see Table S2).
[4]Serotype Gallinarum biovar Duisburg is different from serotype Duisburg.
doi:10.1371/journal.pone.0036995.t005

50 randomly selected clinical isolates collected in 2002 [33]. We found 17 different alleles for CRISPR1 (Simpson's discrimination index (DI) = 0.84), 23 for CRISPR2 (DI = 0.84), and 26 if a combination of the two loci was considered (DI = 0.88). These isolates gave 26 *Xba*I-PFGE profiles (DI = 0.87) and 14 phage types (DI = 0.74). For prevalent MDR DT104 isolates, the discriminatory power was higher for combined CRISPR analysis (5 profiles, DI = 0.64) than for PFGE (also 5 profiles, but DI = 0.38). The best discrimination was that achieved with the five-locus loci MLVA method, for which all the DT104 isolates had different profiles (DI = 1). However, it was not possible to amplify some MLVA loci from some isolates (null alleles were seen

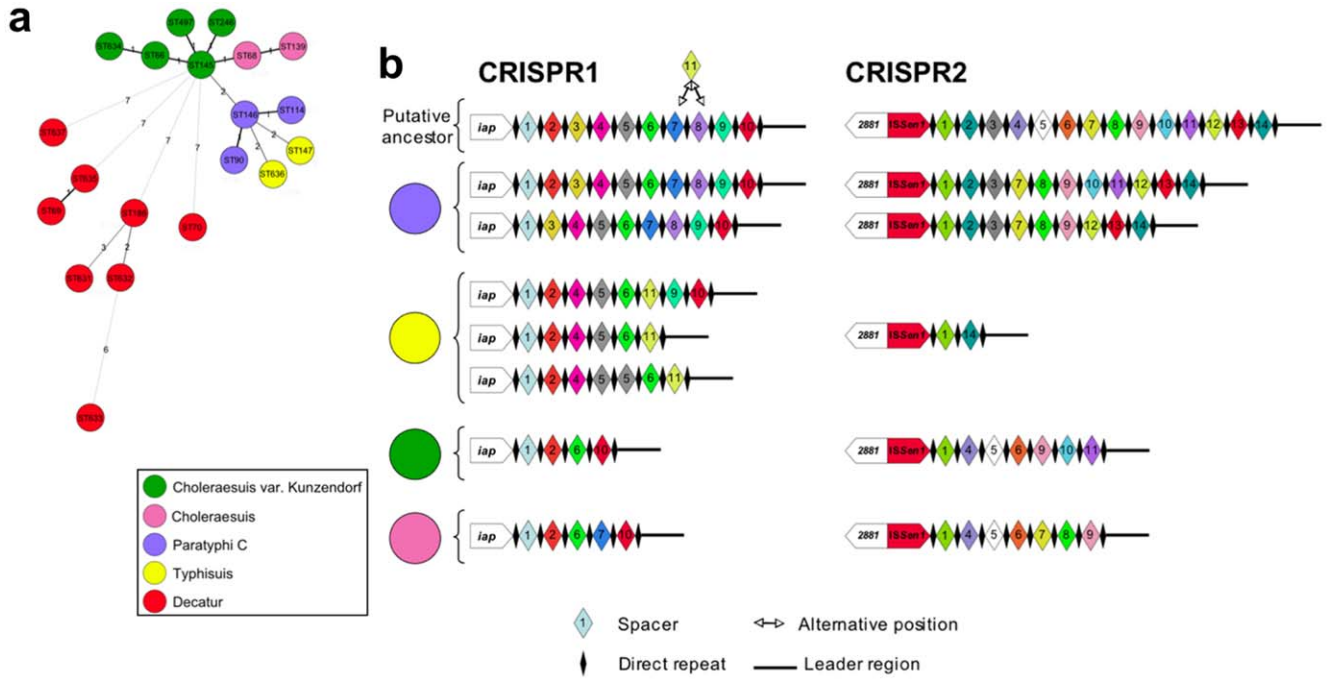


Figure 2. Multilocus sequence typing and CRISPR spacer content of 34 *S. enterica* strains and isolates with the antigenic formula 6,7:c:1,5. Based on additional biochemical characters, we can identify five subserotypes: Choleraesuis *sensu stricto*, Choleraesuis variant Kunzendorf, Paratyphi C (human-restricted), Typhisuis (pig-restricted) and Decatur. MLST (a) and CRISPR data (b) show that Choleraesuis, Paratyphi C and Typhisuis share a common ancestor, whereas Decatur is made up of at least five unrelated populations. The numbers in panel "a" correspond to the allelic difference between STs. The size of the circle is not correlated with the number of strains with the corresponding ST. The exact name of the spacers and the spacer content of Decatur strains from panel b can be found in Table S2.
doi:10.1371/journal.pone.0036995.g002

for STTR3 in 4 isolates, STTR5 in 1 isolate, STTR6 in 3 isolates and STTR10 in 3 isolates) and variations in the number of repeats of some loci were observed in outbreak-related isolates, indicating lower levels of epidemiological concordance, possibly due to the very rapid evolution of these markers or to outbreaks being caused by more than two MLVA types (Table S7).

The spacer content of serotypes Typhimurium and Enteritidis isolates from 10 documented outbreaks was studied by sequencing (Tables S7 and S8) or with the microbead-based CRISPOL assay (see below). In all cases, the outbreak isolates had the same CRISPR type. Epidemiological concordance was thus complete.

Four strains (SARA8, 81-784, 02-7015 and 07-1777) with a known spacer content covering all the spacers identified were tested in every microbead-based CRISPOL experiment. Their CRISPR types were identical in all cases.

In addition to 100% typeability, the other performance criteria [10] for CRISPR analysis, such as stability, discriminatory power, epidemiological concordance and reproducibility, indicated that this was a very powerful method for use in the molecular epidemiology of *Salmonella*.

Applications of CRISPR Polymorphisms

There are at least three applications of CRISPR polymorphisms of potential interest in clinical microbiology or public health laboratories.

Application 1: CRISPR sizing by PCR for the rapid comparison of Salmonella spp isolates. The first application is a double-locus PCR assay for the rapid comparison of *Salmonella* isolates. We demonstrate above that variation in the number and type of spacers can be used to track strains, given the discrimination between the most prevalent *Salmonella* serotypes. Remarkably, simple PCR amplification of CRISPR1 and CRISPR2 loci, followed by agarose gel electrophoresis and sizing of the PCR products, differentiated outbreak isolates from non outbreak isolates and was therefore found to be a useful screening approach. For example, for the eight isolates of serotype Typhimurium isolated from the same city during a single week in 2005 (cluster E in Table S7), it was possible to discriminate between four isolates from the same food poisoning cluster and four other isolates unrelated to this cluster (Figure 3). This size variation results from variation in the number of spacer-DR units (total 60 bp), which thus provides some discrimination even in the absence of qualitative information (i.e., the spacer type). Another advantage of this approach is that it does not require prior serotype identification, as sequences from isolates of all serotypes can be amplified by at least one of the two primer pairs used for CRISPR amplification. A different amplicon size for one or both loci demonstrates that the analyzed isolates are unrelated, but it should be borne in mind that a similar amplicon size for both loci does not necessarily imply that the two isolates belong to the same strain. Two unrelated isolates could have the same number of spacers but of different types, and low-level variation in the number of spacers might not be detected on agarose electrophoresis of large PCR products. However, this simple screening approach is suitable for low-capacity public health or hospital laboratories, including those in developing countries, which need to be able to compare several *Salmonella* isolates rapidly in a single experiment and that cannot afford complete serotyping or subtyping by PFGE.

Application 2: High-throughput method for subtyping serotype Typhimurium or its monophasic variant in real time: the CRISPOL assay. We present below a second

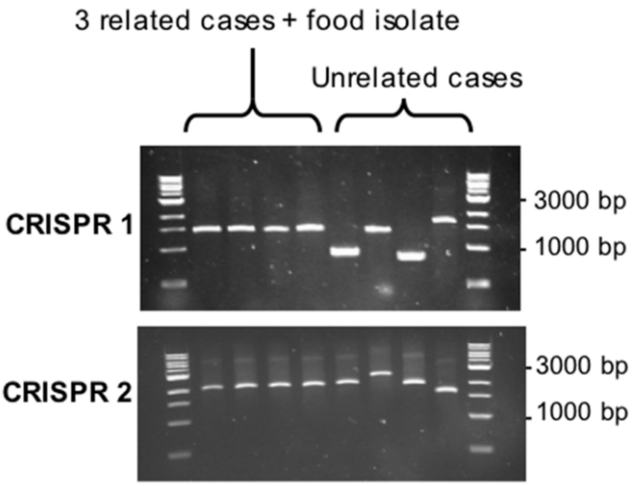


Figure 3. CRISPR sizing by PCR for the rapid comparison of *Salmonella* spp isolates. Results of PCR amplification for 8 *S. enterica* serotype Typhimurium isolates collected from the same city during a single week (cluster E in Table S7). Three cases were from the same food poisoning cluster (the food isolate was also tested), whereas the other cases were unrelated.
doi:10.1371/journal.pone.0036995.g003

application resulting from the development of a high-throughput method for the real-time subtyping of serotype Typhimurium and its monophasic variant. Based on the 83 CRISPR1-CRISPR2 combined alleles identified above, we developed a bead-based liquid hybridization assay (Luminex® technology), CRISPOL (for CRISPR polymorphism; Figure 4). A 25 to 32 bp capture probe was designed for each of 72 of the 79 spacers identified (Table 6; it was not possible to distinguish between some of the remaining seven spacers by this approach. For example, STMB8var1 has a single SNP located in position 1 of the spacer). Each capture probe was coupled to a defined xMAP bead. We used thermolysates as the DNA template and a single primer pair (including a biotinylated primer) hybridizing to DR sequences to amplify the spacer content of the two CRISPR loci rapidly. The PCR mixture was hybridized with the 72 probe-coupled beads and incubated with streptavidin-phycoerythrin for detection. The Luminex® platform was then used to measure the fluorescence associated with each bead (corresponding to a unique probe/spacer). This method gave a highly robust readout, with mean fluorescence signals of 709 to 5,707 in the presence of the spacer, and of 52 to 193 in the absence of the spacer (Table 7). The positive/negative ratios were between 13 (for a bead for which coupling was not optimal) and 92 (mean of 50.8). It was also easy to identify the four SNP-variant spacers (Table 8). One probe, pSTMB26 had a trimodal distribution, due to an intermediate population (MFI between 300 and 1,200), whereas the positive population had an MFI of more than 3,500. This intermediate population (consisting mostly of emerging European monophasic isolates) contained spacer STMB34, which is partly complementary to pSTMB26. We designed a new probe targeting the other side of spacer STMB26, but this probe was also partly complementary to another spacer, STM28. We resolved this problem by subtracting the value for the control strain 02-7015 (STMB34 positive) from that for probe pSTMB26 in each experiment.

The repeatability of the CRISPOL assay was assessed by running 30 isolates in triplicate and was high (data for five strains provided in Table S9), with low standard deviations. The assay was also highly reproducible, based on the results for the four control strains (which, together, contained all the known spacers) analyzed in each experiment. A cutoff of five times the value for the background sample (consisting of all reaction components except template DNA, which was replaced with water), which had an MFI of approximately 300, was used to determine whether the result was positive or negative for a given spacer. The distribution of crude MFI values in a typical experiment with 65 isolates showed the clear-cut distinction between negative and positive results for spacers (Figure S2).

The concordance of the CRISPOL assay and sequencing results was 100% for the spacer content of the 150 sequenced isolates. Despite the presence of some genetic variation, such as duplications of spacers, VNTR variation of STM18 and the appearance of SNP variants, which might result in spacers being missed by the CRISPOL assay, no such effect was observed in practice as we found no isolates with identical CTs but with alleles with different sequences.

This method has two major advantages: its rapidity and low cost. It requires 5.5 hours in total, with 2.75 hours of hands-on time, to test 65 isolates from bacterial colonies, at an estimated cost of €4 per sample for reagents and consumables.

The application of this method to almost 2,000 serotype Typhimurium and monophasic variant isolates led to the identification of 245 different CRISPOL types (CTs; Figure 5). CT21 and its variants were strongly associated with the MDR DT104 serotype Typhimurium clone (MLST type ST19). Similarly, CT1 and its variants were strongly associated with the emerging European monophasic 1,4,[5],12:i:- variant (MLST type ST34). In total, 1,084 isolates (one per patient) were received at the French National Reference Center for *Salmonella* (FNRC-Salm) between January 1and July 4 2010; 89 CTs were observed among the serotype Typhimurium isolates (n = 677). The two most prevalent types were CT21 (33% of isolates) and CT30 (11.5%); both were associated with the DT104 clone. For monophasic isolates (n = 407), we identified 39 CTs, of which CT1 (50%) and CT9 (14.7%) were the most prevalent. During this period a steady increase in the prevalence of CT1 and three peaks of CT21, CT136 and CT62 (two of which corresponded to documented outbreaks) [36] were observed (Figure 6).

Application 3: Development of PCR assays targeting specific serotypes or particular strains. The presence of unique, constant spacers in certain serotypes, such as Typhi and Paratyphi A, should make it possible to develop PCR assays specific for these serotypes. As a proof-of-principle, we have successfully developed and validated such PCRs for serotypes Typhi and Paratyphi A (manuscript in preparation). Moreover, it would be possible to develop a PCR assay for the detection of any strain of interest with a particular spacer content provided that a culture of this strain was available. For strains with no specific spacers (common spacers only), we can use other stable characteristics of the strain, such as the absence of a spacer-DR unit between STM06 and STM24 (e.g., a MDR serotype Typhimurium DT104 strain), to design primers yielding a PCR product of known size.

Discussion

We demonstrate here that the assessment of CRISPR spacer content is a robust, highly discriminatory and practical method for typing *Salmonella* isolates. Serotyping has been the reference method for *Salmonella* typing for almost 80 years. However, this technique has a number of drawbacks, including low throughput, high costs due to the need for highly trained staff and expensive antisera, and accreditation problems. It is also of limited value for

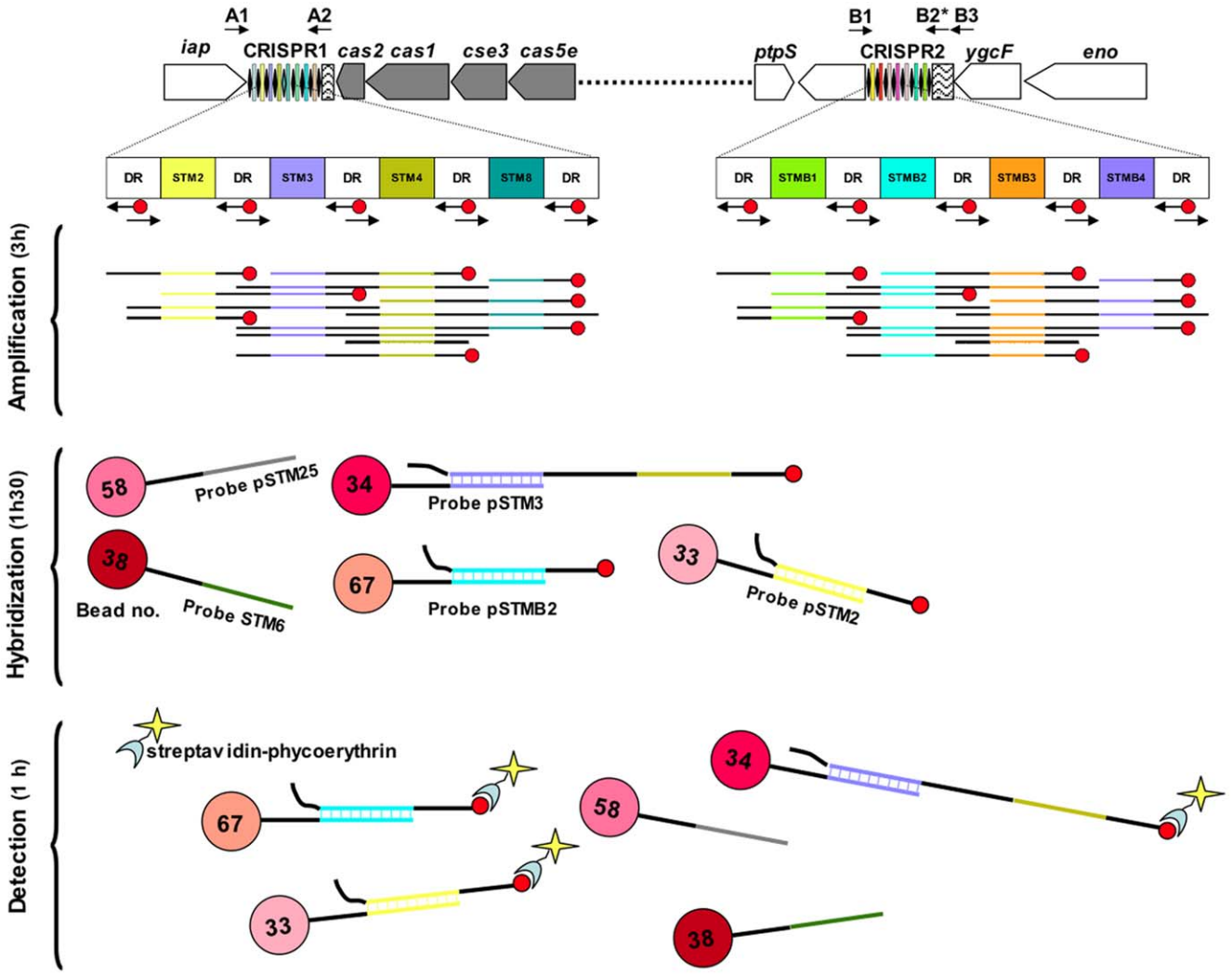


Figure 4. Overview of the bead-based CRISPOL assay for *S. enterica* serotype Typhimurium developed here. The estimated time for each step is based on the testing of 65 isolates in a 96-well plate. The amplification step spans from the preparation of thermolysates to the gel electrophoresis of PCR products.
doi:10.1371/journal.pone.0036995.g004

strain discrimination, given the overdominance of a small number of serotypes. There is therefore a clear need for improved methods [37].

Recently developed "molecular serotyping" methods have been proposed as an alternative. These methods mimic serotyping in that they target the genes involved in biosynthesis of the flagellar (*fliC* and *fljB*) [38] and/or the O-polysaccharide (encoded by the *rfb* locus) antigens [2]. However, due to the complexity of the *rfb* locus (8 to 23 kb, including more than 10 open reading frames), it is currently possible to identify only a minority of the 46 O serogroups of *Salmonella* by PCR. These approaches are also subject to all the limitations inherent to serotyping in terms of a lack of discrimination and a lack of polyphyletic group recognition.

MLST is a promising method for defining evolutionarily and epidemiologically meaningful groups of *Salmonella*. A publicly accessible database (http://mlst.ucc.ie/mlst/dbs/Senterica) includes data for more than 4,250 isolates from more than 500 serotypes [39]. Analyses of these data have revealed that most serotypes are probably polyphyletic and therefore do not correspond to natural groups descended from a single ancestor and sharing important host association or virulence features. This recent study highlights the importance of using phylogenetically informative methods recognizing natural groups rather than serotypes. However, MLST has a low discriminatory power and is not suitable for the detection or investigation of outbreaks due to highly prevalent monophyletic serotypes.

The data presented here for a global collection of 783 reference strains and isolates from 130 serotypes of *Salmonella*, including the most common serotypes involved in human infections, show a high degree of CRISPR polymorphism. This polymorphism makes it possible to distinguish between most serotypes and between MLST groups within polyphyletic serotypes. Furthermore, microvariations, such as the loss, acquisition, duplication of spacers or point mutations within spacers, have a strain discrimination capacity similar to that of current gold standard methods, such as PFGE. The CRISPR method can therefore be used for simultaneous typing – defined as the determination of serotype or MLST group – and subtyping. It therefore represents a single alternative to several widely used reference methods: serotyping, PFGE and phage typing. This genetic marker is based on polymorphic DNA sequences of limited length, 0.5 to 3 kb. It therefore has major advantages, in terms of analysis, throughput, standardization,

Table 6. Serotype Typhimurium spacers and corresponding probes for the CRISPOL assay.

CRISPR locus	Spacer name (position) *	Spacer DNA sequence (5′–3′)‡	Probe name (position) **
1	STM01 (1)	TTTTCAGCCCTTGTCGACTGCGGAACGCCCCT	pSTM01 (1)
1	STM02 (2)	GCGAAATAGTGGGGAAAAACCCCTGGTTAACC	pSTM02 (2)
1	STM03 (3)	TAGGCCTTGATACCATCGCTCGCACCTCGTCA	pSTM03 (3)
1	STM03var1 (3)	TAGGCCTTGATACCATC**A**CTCGCACCTCGTCA	pSTM03var1 (69)
1	STM03var2 (3)	**C**AGGCCTTGATACCATCGCTCGCACCTCGTCA	
1	STM25 (4)	GTTTATTACTGCTTAGTTAATTAATGGGTTGC	pSTM25 (4)
1	STM26 (5)	AGGCGAATAATCTCTAATAGTCTCAATTCGTT	pSTM26 (5)
1	STM27 (6)	TAAATCTGGCGTCGAGACATTCGAAATAGTGC	pSTM27 (6)
1	STM04 (7)	TCTTTTGATTTTGCTGCGATGTTATAACCAGA	pSTM04 (7)
1	STM05 (8)	TATCCACATATACCCGCAATCATATTCAAGAA	pSTM05 (8)
1	STM06 (9)	AATCACTGCGGGGGTATTTAGCGGAAACGGCT	pSTM06 (9)
1	STM07A (10)	GATCGAGTAACGTGCGCTGGAACGCGTCGGCGCGGGGAACAC	
1	STM07B (10)	AATTAAAGCCGAGGGTGGCACCGCGCCTTATT	pSTM07 (10)
1	STM07Bvar2 (10)	AATTAAAGCCGAGGGTGG**T**ACCGCGCCTTATT	pSTM07var2 (70)
1	STM08 (11)	GCACCTCGAAACGGTTTTAAAACACTACCGTTT	pSTM08 (11)
1	STM09 (12)	TGGACCGATGGGGCCAACATCGCCGAACGTGG	pSTM09 (12)
1	STM10 (13)	GTTACGTTCGGTAAATGGAAAGCGGCGAATAT	pSTM10 (13)
1	STM11 (14)	CCAGAAAGTGCCGGTAGTGCCTGATGAACGAC	pSTM11 (14)
1	STM12 (15)	CGCGCCCACTTCCGTAAAATACAGATAATCCA	pSTM12 (15)
1	STM12var1 (15)	CGCGCCCACTTCCGTAA**G**ATACAGATAATCCA	pSTM12var1 (71)
1	STM28 (16)	GGCAGCGGGCGAGGCAAACACATTCGGGGCGT	pSTM28 (16)
1	STM13 (17)	GGTAATTTCTCATCTAACAGCCTGTACGCCTC	pSTM13 (17)
1	STM14 (18)	GAATCTAATGCAACAGATGAATAAACACGTAA	pSTM14 (18)
1	STM15 (19)	TCTTTATCGTCAATGCGAAATTTTCCGCGACG	pSTM15 (19)
1	STM16 (20)	TCCCATTCACCAACAACAATATCGCCCTGCAA	pSTM16 (20)
1	STM17 (21)	CGTTGCGTCAGGTTGATCCAGTGCGTCAGCGG	pSTM17 (21)
1	STM18 (22)	TCTCGGTCTCGGTCTCGGTCTCGGTAGTGACG	pSTM18 (22)
1	STM18var1 (22)	TCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTAGTGACG	
1	STM18var2 (22)	TCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTAGTGACG	
1	STM18var3 (22)	TCTCGGTCTCGGTCTCGGTAGTGACG	
1	STM18var5 (22)	TCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTCTCGGTAGTGACG	
1	STM19 (23)	ACTTCCTTCAGTCTTAACGATAATCCGCAACG	pSTM19 (23)
1	STM20 (24)	GCAAAATAGCGATGAGCTGGCTACGCCCACTGG	pSTM20 (24)
1	STM29 (25)	AGCCGGCGCGAGCCTGGAGGGTTGAATAATGG	pSTM29 (25)
1	STM30 (26)	CAATCTCGCATTCGTTACCCCACCTGCATTTT	pSTM30 (26)
1	STM21 (27)	GAGGGGATAGGAGTTACGATCCAGCCTGGTTG	pSTM21 (27)
1	STM22 (28)	GTGGTTGCAGACCAATCAGCCCGCCAGCGGTT	pSTM22 (28)
1	STM24 (29)	CAGCACGAAAAATTATTTACTGTCGTTGCTCA	pSTM24 (29)
1	STM31 (30)	TGTAACAGTCCGTCGTTAATCAGCGCGGTGGG	pSTM31 (30)
1	BraB14 (31)	GAAGGTACGGGGAAAACAAAGATTACTCGTTC	pBraB14 (31)
2	STMB0 (1)	ATCTTCATATTGCGTGACGCTGCCGATGAACG	pSTMB0 (32)
2	STMB32 (2)	TCTTTATCAGCTAACCATTTCCAGAACTCGTC	pSTMB32 (33)
2	STMB01 (3)	TATAATATGAATTAATTTTTGCGCATAACCTG	pSTMB01 (34)
2	STMB01var1 (3)	TATAATATGAATTAATTTTTGCG**T**ATAACCTG	
2	STMB02 (4)	TGCCCGTTCTGCCTCTTCGCACTCTCGATCAA	pSTMB02 (35)
2	STMB03 (5)	TGCGTAATGGGCTACCTGAACTTCACATATCC	pSTMB03 (36)
2	STMB04 (6)	ATTAAGCGCGCAAAGTTTGGGTTAATTGGACA	pSTMB04 (37)
2	STMB05 (7)	CGTATTCGTCACACAGCCCCGTCCAGAAATGA	pSTMB05 (38)
2	STMB06 (8)	TAACGAACTGAATAAAATGTCAGAAAGTGACG	pSTMB06 (39)

Table 6. Cont.

CRISPR locus	Spacer name (position) *	Spacer DNA sequence (5′–3′)‡	Probe name (position) **
2	STMB07 (9)	GCAGCTTAGCGACGAAATTAAAACCGAACTCAC	pSTMB07 (40)
2	STMB08 (10)	TGCCAGTGACTACAGAAGCGTCGCTATCGGGG	pSTMB08 (41)
2	STMB08var1 (10)	TGCCAGTGACTACAGAAGCGTC**T**CTATCGGGG	pSTMB08var1 (72)
2	STMB09 (11)	ACCGATAAACAACCGCATAGCCTCTTTCGTTT	pSTMB09 (42)
2	STMB10 (12)	TGCTCAATAACGTCGTAAATAGCGTAAGCTGG	pSTMB10 (43)
2	STMB11 (13)	TATTTCGCCTTCGGCACTGACGTCACCGTCAA	pSTMB11 (44)
2	STMB12 (14)	GTCGCGTTCGTTGCCGGTATAGACCAGCGTCA	pSTMB12 (45)
2	STMB13 (15)	ATCGAATCGAAACCCCAGCCACAGAAATAATT	pSTMB13 (46)
2	STMB14 (16)	GCTCATGTCAAACGCCATCAGCGTTCCGGCAT	pSTMB14 (47)
2	STMB15 (17)	AATCGCCAGCCTCGGAAATATTCCATCCTCCG	pSTMB15 (48)
2	STMB16 (18)	AGGAACTAAACAGCCTGACCGTTGAGGATCTG	pSTMB16 (49)
2	STMB17 (19)	ACCGGACAAATCTTTTTTTTCCTGTTCCTGTT	pSTMB17 (50)
2	HadB20 (20)	GGGCGGTCCCCGGCCTCAATACCGCGCTGACG	pHadB20 (51)
2	STMB34 (21)	TTGAGGTGCCGCTTGCCGTTCTTCTGTTTTTT	pSTMB34 (52)
2	STMB18 (22)	GGGCACTATGAACGGATCGGCGCTGATGCCGG	pSTMB18 (53)
2	STMB19 (23)	GGTAAAGCCACACCATTTTTTATTGACCTCGC	pSTMB19 (54)
2	STMB33 (24)	CTAGGAGGCGTAATGAATACTACGTATCAAAA	pSTMB33 (55)
2	STMB20 (25)	GTGGTGGCCTCAAATAAATTCGAGCGCTGGAG	pSTMB20 (56)
2	STMB21 (26)	TCGACGTGGACGAGGAGTTACTCAACCGCTGC	pSTMB21 (57)
2	STMB22 (27)	AGCGCCACATGGCCCACCGGCACCACCCGATC	pSTMB22 (58)
2	STMB23 (28)	AAATGACCAAATCAGAAATCTTCACCAAAGCC	pSTMB23 (59)
2	STMB24 (29)	TAATGGCCACAGTAAGTCAAACGGTTCTGGAA	pSTMB24 (60)
2	STMB25 (30)	GAGTCCGGGGGTTATATAGTTATTTAATGAGC	pSTMB25 (61)
2	STMB26 (31)	TTGGGCGTCGGTTTTTTCAGGTTTAGGTCCGG	pSTMB26 (62)
2	STMB27 (32)	TCAACTGTCAGTTCGTCGTTAGCCAGTAATTC	pSTMB27 (63)
2	STMB28 (33)	CTGAAAACGCATGGAATCCGGTATAAACAGTC	pSTMB28 (64)
2	STMB29 (34)	GATGTAACTGATAGCGAAATATATTGGGATAA	pSTMB29 (65)
2	STMB30 (35)	GAAACGTAAACAGGGTAAGATACAACTCTGCA	pSTMB30 (66)
2	STMB31 (36)	TGTAAAGGGTGGTCTGGAAGGGGATCGGCAAA	pSTMB31 (67)
2	STMB35 (37)	TCGTGTGAGGTCGCTGAGAAAAACGGGGCGTA	pSTMB35 (68)

*position of the spacer within the CRISPR1 or CRISPR2 locus.
‡single polymorphic nucleotides that define spacer variants are shown in bold typeface; probe sequences are underlined.
**position of the probe in the CRISPOL assay.
doi:10.1371/journal.pone.0036995.t006

interpretation and data exchange, over current typing and subtyping methods. We believe that this novel approach will constitute a real improvement in the monitoring of *Salmonella* infections, by making it possible to obtain results more rapidly, thereby optimizing surveillance and outbreak detection.

We propose several different strategies for CRISPR genotyping in *Salmonella*. First, determination of the sizes of the two CRISPR loci by PCR can be used as an initial screen that is easy to implement, even in low-capacity laboratories. This approach requires no preliminary serotyping. Second, when more precise discrimination is required, the spacer content can then be investigated by Sanger sequencing of the PCR products, which has the additional advantage of facilitating the detection of new putative spacers. However, once the spacer diversity within a serotype is known (i.e., after the analysis of a representative collection of isolates of this serotype), higher throughput is required for daily surveillance. We have developed a Luminex®-based approach that is suitable for serotype Typhimurium, which accounts for 50% of all cases of human salmonellosis in France and is one of the two major serotypes worldwide. PFGE is currently recommended for the real-time surveillance of this serotype, but is technically demanding and poorly standardized in many laboratories. It is therefore difficult to use PFGE in many countries in which a single reference laboratory processes a large number of isolates. The CRISPOL assay developed here covers both serotype Typhimurium and its emerging monophasic variant. It provides an excellent alternative to PFGE, being cheaper, less technically demanding and yielding data that are easy to interpret and exchange. An approach based on the initial use of the CRISPOL assay, followed by MLVA for genetically homogeneous populations, such as the DT104 clone (CT21) or the emerging monophasic strain (CT1) would be highly effective. However, due to limitations in the number of beads that can be mixed (500 for the latest Luminex® platform), the universal use of a Luminex®-based approach is not possible (the *Salmonella* serotypes analyzed to date include 3,800 different spacers). Whole-genome sequencing

Table 7. Probe responses in the CRISPOL assay (data from 25 sequenced isolates).

Probe name (bead no.)	Spacer absent Median (MFI)±SD[1]	Spacer present Median (MFI)±SD	Ratio[2] positive/negative
pBraB14 (29)	65±10	3866±570	59
pSTMB35 (30)	70±14	3540±445	51
pHadB20 (31)	62±8	1544±200	25
pSTM01 (32)	53±9	2547±332	48
pSTM02 (33)	81±12	2586±308	32
pSTM03 (34)	56±4	3808±5394	68
pSTM03var1 (35)	52±3	2409±5324	48
pSTM04 (36)	57±8	4187±390	73
pSTM05 (37)	173±13	5216±605	30
pSTM06 (38)	64±4	3019±337	47
pSTM07 (39)	66±11	4580±4944	69
pSTM07var2 (40)	65±13	3283±4714	54
pSTM08 (41)	55±7	709±154	13
pSTM09 (42)	68±11	4065±397	60
pSTM10 (43)	94±17	4036±402	43
pSTM11 (44)	62±9	4564±518	74
pSTM12 (45)	69±11	3343±5154	46
pSTM12var1 (46)	61±8	1173±2804	20
pSTM13 (47)	58±7	3812±539	66
pSTM14 (48)	59±8	4183±381	71
pSTM15 (49)	72±14	3458±452	48
pSTM16 (50)	71±13	4990±1061	70
pSTM17 (51)	89±15	2092±333	24
pSTM18 (52)	79±14	3863±610	49
pSTM19 (53)	68±15	3359±289	49
pSTM20 (54)	65±9	3113±375	48
pSTM21 (55)	74±14	4100±573	56
pSTM22 (56)	71±9	2158±304	30
pSTM24 (57)	50±6	4603±440	92
pSTM25 (58)	60±11	4178±382	70
pSTM26 (59)	62±7	4772±521	78
pSTM27 (60)	59±10	4061±364	69
pSTM28 (61)	61±9	4186±328	69
pSTM29 (62)	61±9	3508±418	58
pSTM30 (63)	73±11	4532±670	62
pSTM31 (64)	68±10	3209±409	47
pSTMB0 (65)	55±12	3952±398	73
pSTMB01 (66)	64±16	5409±674	85
pSTMB02 (67)	65±11	3987±500	62
pSTMB03 (68)	58±9	2508±713	43
pSTMB04 (69)	68±13	4913±558	72
pSTMB05 (70)	60±8	5196±574	87
pSTMB06 (71)	61±8	1802±343	30
pSTMB07 (72)	62±10	4447±477	71
pSTMB08 (73)	69±10	1210±4535	53
pSTMB08var1 (74)	69±10	1266±2515	27
pSTMB09 (75)	66±9	1266±251	19
pSTMB10 (76)	64±7	2629±480	41
pSTMB11 (77)	59±9	2036±279	35

Table 7. Cont.

Probe name (bead no.)	Spacer absent Median (MFI)±SD[1]	Spacer present Median (MFI)±SD	Ratio[2] positive/negative
pSTMB12 (78)	73±11	2147±276	30
pSTMB13 (79)	69±14	5142±805	75
pSTMB14 (80)	64±8	2034±289	32
pSTMB15 (81)	67±10	3524±466	53
pSTMB16 (82)	78±12	5062±547	65
pSTMB17 (83)	94±16	2947±437	32
pSTMB18 (84)	80±11	3745±532	47
pSTMB19 (85)	101±13	4570±733	45
pSTMB20 (86)	75±14	4573±483	61
pSTMB21 (87)	73±12	3556±319	49
pSTMB22 (88)	84±16	5707±636	68
pSTMB23 (89)	69±9	5434±642	78
pSTMB24 (90)	71±8	3525±450	50
pSTMB25 (91)	72±10	2039±395	28
pSTMB26 (92)	193±2443	3599±664	19
pSTMB27 (93)	101±15	3464±505	34
pSTMB28 (94)	87±12	3218±391	37
pSTMB29 (95)	87±13	3542±519	41
pSTMB30 (96)	79±10	3805±442	48
pSTMB31 (97)	83±15	2195±304	27
pSTMB32 (98)	83±12	4832±551	58
pSTMB33 (99)	70±8	2143±249	30
pSTMB34 (100)	78±11	3109±513	40

[1]MFI, median fluorescence intensity (raw data); SD, standard deviation.
[2]The ratio is the positive average median fluorescence intensity (MFI) divided by the negative average MFI.
[3]Due to partial identity with spacer STMB34, the signal of pSTMB26 is stronger in isolates containing spacer STMB34 (such as the emergent monophasic population). This has been taken into account by subtracting the MFI of control strain #02-7015 from that of pSTMB26 in each experiment. The corrected median is 183±46 with a ratio of 20.
[4]Median calculated for 24 isolates.
[5]Median calculated for 20 isolates.
doi:10.1371/journal.pone.0036995.t007

(WGS) is a possible alternative, provided costs and analysis times can be decreased. WGS could be customized to focus exclusively on the two CRISPR sequences. The known spacer sequences would be extracted and compared with the contents of a CRISPR/serotype database. Another alternative would involve the use of a microarray approach based on DNA oligonucleotides

Table 8. Probe responses in the CRISPOL assay for SNP variants (individual isolates).

Probe name (bead no.)	Isolate #81-299 (STM03 variant 1) Median (MFI)±SD	Isolate #DK4 (STM07 variant 2) Median (MFI)±SD	Isolate #02-277 (STM12 variant 1) Median (MFI)±SD	Isolates #02-3369, #02-7105, #01-1639, #81-482, #81-831 (STMB08 variant 1) Median (MFI)±SD
pSTM03 (34)	3807±89	NA	NA	NA
pSTM03var1 (35)	4573±132	NA	NA	NA
pSTM07 (39)	NA	1944±41	NA	NA
pSTM07var2 (40)	NA	5143±140	NA	NA
pSTM12 (45)	NA	NA	1493±61	NA
pSTM12var1 (46)	NA	NA	4146±85	NA
pSTMB08 (73)	NA	NA	NA	2204±328
pSTMB08var1 (74)	NA	NA	NA	4277±420

Results in triplicate; MFI, median fluorescence intensity (raw data); SD, standard deviation; NA: not applicable.
doi:10.1371/journal.pone.0036995.t008

Figure 5. Dendrogram presentation of the 245 distinct CRISPOL types detected among 2,200 isolates of *S. enterica* serotype Typhimurium or its monophasic variant of antigenic formula 1,4,[5],12:i:-. Black squares indicate presence of the spacer, as detected by the corresponding probe, whereas white indicates an absence of the spacer. For the determination of CRISPOL types (CTs), each of the 68 spacers was

treated as a numerical character indicating absence (0) or presence (1 for all spacers except BraB14, for which an arbitrary value of 10 was assigned) in BioNumerics 6.5 software (Applied Maths, Sint-Martens-Latem, Belgium). Similarities between CTs were assessed by calculating the Pearson product-moment, and a dendrogram was constructed by the unweighted pair group method with arithmetic mean (UPGMA). The four SNP-variant spacers targeted by probes 69 to 72 are shown but were excluded from the phylogenetic analysis, as they were not independent. A indicates a group of profiles derived from CT1, the main type of emerging monophasic isolates. B indicates a group of profiles derived from CT21, which is associated with multidrug-resistant DT104 serotype Typhimurium isolates. C indicates a group of serotype Typhimurium isolates of ST36 that may have one or two specific spacers on the leader side of CRISPR1 (BraB14) and CRISPR2 (STMB35).
doi:10.1371/journal.pone.0036995.g005

corresponding to highly informative spacers. In the meantime, subtyping applications remain to be developed for the most epidemiologically important serotypes or MLST groups, such as Enteritidis and Newport. For this purpose, it should be straightforward to apply the strategy presented here for serotype Typhimurium.

Clearly, there is also a need to extend the serotype coverage of the spacer content inventory, as only the 130 most important serotypes have been investigated so far. We hope to capture all the diversity of *Salmonella* (>2,500 serotypes) in the next 10 years, and the CRISPR/serotype dictionary available from our open-access website will be updated accordingly. This web tool can be used to extract and identify spacers from a submitted DNA sequence and for comparisons with a well curated database (i.e., containing accurately serotyped isolates). The application of this tool to CRISPR sequences identified as corresponding to Enteritidis isolates by Liu *et al.* [27], showed that 10 of the 27 considered actually corresponded to Typhimurium (EST21, EST22), Infantis (EST17), Kentucky (EST23), and Heidelberg (EST15), rather than Enteritidis. All these discrepancies related to isolates obtained from a local diagnostic laboratory and not from the reference laboratory participating in the study, which suggests that serotyping errors were the cause.

In conclusion, we have demonstrated that CRISPR is a powerful method suitable for use in the molecular typing and subtyping of *Salmonella* isolates. We believe that, given its combined advantages, CRISPR strain characterization is an excellent potential alternative to both serotyping and PFGE, the current gold standard methods. Given the rapidity of this method, in particular, it should have a major impact on surveillance and outbreak investigation and is likely to be of benefit to public health.

Materials and Methods

Salmonella Strains and Isolates

In the first part of this study (spacer content inventory and comparison with current typing and subtyping methods), we used 744 *Salmonella* reference strains or isolates belonging to 130 serotypes (including those most frequently identified in human and food products). These serotypes belonged to two species of the *Salmonella* genus: *S. enterica* and *S. bongori* and the six subspecies of *S. enterica*: *enterica*, *salamae*, *arizonae*, *diarizonae*, *indica* and *houtenae* (Tables S2 and S4). The *Salmonella* serotype reference strains were obtained from the World Health Organization Collaborative Center for Reference and Research on *Salmonella* (WHOCC-Salm). Most of the isolates were from the French National Reference Center for *Salmonella* (FNRC-Salm). Both these centers are located at the Institut Pasteur, Paris. Other strain and isolate providers are acknowledged at the end of the manuscript. The strains and isolates studied were obtained from around the world, between 1885 and 2010. Larger subsets of isolates from prevalent serotypes were assembled to reflect as accurately as possible the diversity of these populations: Typhimurium or its monophasic variant with antigenic formula 1,4,[5],12:i:- (n = 150), Enteritidis (n = 187), polyphyletic serotypes such as Newport (n = 21) and Paratyphi B (n = 36), and serotypes with the antigenic formula 6,7:c:1,5 (n = 34), or clinically important serotypes, such as Typhi (n = 20) and Paratyphi A (n = 14). This test population was generally well defined in terms of its epidemiological context,

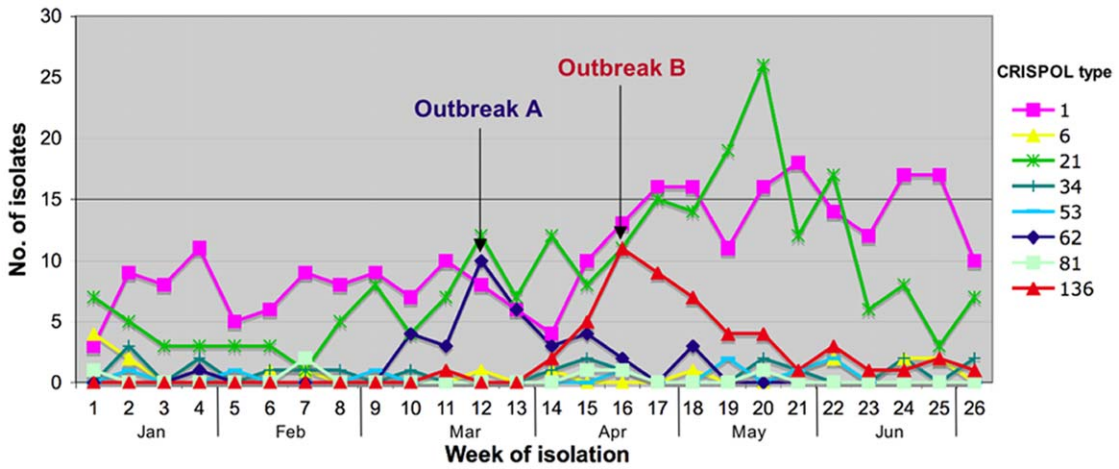


Figure 6. Distribution of selected CRISPOL types of *S. enterica* serotype Typhimurium and *S. enterica* with antigenic formula 1,4,[5],12:i :- isolated from humans in France between January 1 and July 4 2010. Over this period, all 1,084 isolates (one per patient) were CRISPOL-typed and two outbreaks were investigated. Outbreak A (≈40 cases) was due to the consumption of a raw milk cheese contaminated with a CT62 highly multidrug-resistant *S. enterica* serotype Typhimurium strain, whereas outbreak B (≈50 cases) was caused by the consumption of a dried pork sausage contaminated with a CT136 *S. enterica* 4,12:i:- strain. The third peak, corresponding to CT21, in May was neither detected nor investigated, as CRISPOL typing was carried out retrospectively.
doi:10.1371/journal.pone.0036995.g006

antimicrobial resistance phenotype, phage type, PFGE type, MLVA type, haplotype and MLST type, as determined by methods described elsewhere [9,32,33,40,41].

In the second part of the study, we validated the CRISPOL method on 150 serotype Typhimurium or 1,4,[5],12:i:- isolates for which both CRISPR loci were sequenced. The method was then applied to a collection of 1,900 isolates from the WHOCC-Salm and from the FNRC-Salm, including all isolates received by the FNRC-Salm from January 1 2010 to July 4 2010 (n = 1,131 isolates from 1,084 patients).

Inventory of the Spacer Content of 744 *Salmonella* Strains and Isolates and 39 Genomes of 130 Serotypes

In silico analysis. We analyzed CRISPR spacer content in 39 full genome sequences of *S. enterica* and *S. bongori* (Table 1). Regions containing CRISPR sequences were identified by a blast (ncbi) search of the 29 bp DR consensus sequence of *S. enterica* serotype Typhi strain Ty2 (5′-CGGTTTATCCCCGCTGGCGCGGGGAACAC-3′) [14]. Regions downstream from the *iap* gene and upstream from the *ygcF* gene were downloaded and the spacer-DR units of each CRISPR locus were extracted manually.

DNA extraction. Total DNA was extracted with the InstaGene matrix (BioRad, Marnes la Coquette, France) or the Wizard kit (Promega, Madison, WI, USA) in accordance with the manufacturer's recommendations.

PCR and sequencing of the CRISPR1 and CRISPR2 loci. Oligonucleotide primers for amplification of the CRISPR1 and CRISPR2 loci from all *Salmonella* spp. were designed on the basis of consensus alignments of the available *Salmonella* genomes (Tables 1 and 2). The CRISPR1 locus was amplified with the forward A1 primer (binding 74 bp upstream from the CRISPR1 of serotype Typhimurium strain LT2) and the reverse primer A2 (binding 130 bp downstream). Alternative reverse primers, such as A3 to A7, were required for some isolates. The CRISPR2 locus was amplified with the forward primer B1 (binding 110 bp upstream from CRISPR2) and the reverse primers B2 and B3 (binding 45 bp and 324 bp downstream from the CRISPR2 locus of strain LT2, respectively). The B3 primer was designed because no region homologous to B2 was found in subsp. *diarizonae*. The primers were synthesized by MWG-Biotech (Ebersberg, Germany). Single-locus amplifications were performed in a total volume of 50 µl containing DNA (2.5 µl from InstaGene matrix or 2 µl diluted 10-fold from Wizard), primers (10 pmol each), deoxynucleoside triphosphate (100 µM), *Taq* DNA polymerase (0.85 U of GoTaq Flexi DNA polymerase; Promega) and its buffer, $MgCl_2$ (1.5 mM) and dimethylsulfoxide (5%). The cycling conditions were as follows: 2 min for denaturation at 94°C (1 cycle), followed by 35 cycles of 1 min at 94°C for denaturation, 1 min at 59°C (61°C when using the A1-A4 pair) for annealing, and 90 s at 72°C for polymerization, followed by an additional 10 min at 72°C for extension.

The entire region spanning both CRISPR loci was amplified with primers A1 and B3. For this purpose, DNA was extracted with the Promega Wizard kit and PCR was carried out with the Expand Long Template PCR System kit (Roche).

Both strands of purified amplicons were sequenced with Big Dye Terminator version 3.1 (Applied Biosystems, Foster City, CA) on an ABI 3730XL apparatus (Applied Biosystems).

BioNumerics 6.5 software (Applied Maths, Sint-Martens-Latem, Belgium) was used to analyze nucleotide sequences.

Development of Web-accessible Tools and CRISPR/serotype Dictionary

We have developed a web tool for the creation and storage of catalogues of spacers and DR variants. This "Institut Pasteur CRISPR database for *Salmonella*" can be queried online at http://www.pasteur.fr/recherche/genopole/PF8/crispr/CRISPRDB.html. The content of the catalogue is used to identify known spacers and DRs in a submitted DNA sequence, which is coded into a succession of DR and spacer identifiers by the query "**Search spacers composition for query**". If part of the sequence has no exact matches in the DR and spacers dictionary, a blast query can then be used to obtain the nearest match ("**Blast unknown spacer sequence against dictionary**"), to identify new spacers or new DR variants. Isolates analyzed at the FNRC-Salm and coded as spacer-DR arrays within the CRISPR/serotype dictionary can be downloaded with the "**Browse spacers composition for the published strains**" query. .

Spacer nomenclature. The spacer names start with a three- to four-letter prefix indicating the serotype from which the spacer was extracted for the first time. The suffix B indicates spacers found in the CRISPR2 locus. Spacers were numbered consecutively in order of discovery. The start of spacer arrays is described, starting downstream from the *iap* gene. SNP or VNTR variants are denoted as "var" (e.g., EntB0var1, STM18var2).

Calculation of discrimination indices. The discriminatory abilities of the CRISPR, PFGE and phage typing methods were assessed by calculating Simpson's index of diversity (D value), as previously described [42].

Statistic analysis. The mean number and standard deviation of the spacers in the CRISPR loci were calculated with Excel (Microsoft).

Nucleotide sequence accession numbers. The nucleotide sequences of the CRISPR loci have been assigned GenBank accession numbers JF724159 to JF725640.

Development of a High-throughput Subtyping Method (CRISPOL) for Serotype Typhimurium

DNA extraction. We increased the throughput of this method by using thermolysates as the DNA template. Briefly, we suspended a 10 µl loop of bacteria in 200 µl of molecular biology-grade water. The suspension was vortexed for 10s, incubated at 95°C for 10 min and then centrifuged for 5 minutes at 13000 rpm in a Jouan A14 centrifuge. The supernatant was transferred to a 1.5 ml microtube and stored at −20°C until use.

PCR amplification. We followed the strategy used for *Mycobacterium* spoligotyping, based on the use of two primers hybridizing to the DRs in opposite directions, one of which was 5′-biotinylated [17]. The primers were DRSTMA (5′-CCGCTGGCGCGGGGAACA-3′) and DRSTMB (5′Biot-CGCCAGCGGGGATAAACC-3′). Amplifications were carried out in a volume of 50 µl containing 1 µl of thermolysate (or 1 µl of molecular biology-grade water for the blank), primers (50 pmol each), deoxynucleoside triphosphate (200 µM), *Taq* DNA polymerase (0.85 U of Go *Taq*; Promega) and its buffer, $MgCl_2$ (1.5 mM). The cycling conditions were as follows: 2 min at 95°C for initial denaturation, followed by 20 cycles of 1 min at 95°C for denaturation, 30 s at 59°C for annealing and 15 s at 72°C for polymerization. The PCR products were checked by electrophoresis in 1.2% agarose gels and were stored at −20°C for no more than three days before use in the Luminex® assay.

Probes and microbead coupling. We designed 72 spacer-derived oligonucleotide probes (including four SNP-variant probes) of between 25 and 32 nucleotides in size (Table 6). These probes were synthesized by Eurogentec (Angers, France), with a 5′

terminal amino group modification, using a 12-carbon spacer linker. The 72 Luminex xMap microbeads (L100-C129 to L100-C200) were coupled to the 72 probes, as previously described [43]. Each type of coupled microbead was resuspended in 1 × TE (10 mM Tris, 1 mM EDTA, pH 8) at a final concentration of approximately 50,000 microbeads per µl. We then combined equal volumes of each type of coupled microbead in Protein LoBind tubes (Eppendorf, Hamburg, Germany). The mixture was stored in the dark at 4°C before and after use.

Hybridization. Hybridization was performed in a polycarbonate plate with 96 conical wells (Corning, Corning, NY, USA), to which we added 10 µl of PCR product, 7 µl of 1 × TE and 33 µl of probe-coupled microbeads diluted in 1.5 × TMAC buffer (5 M tetramethyl ammonium chloride [Sigma, St. Louis, MO, USA], 20% Sarkosyl, 1 M Tris-HCl [pH 8.0], 0.5 M EDTA [pH 8.0]) to a final concentration of approximately 75 microbeads per µl for each type of coupled microbead. The plates were sealed with adhesive PCR film (ABgene, Epsom, UK) and heated to 94°C for 3 min for initial denaturation, followed by hybridization at 59°C for 20 min. The plate was centrifuged at 4,000 g for 3 min and the supernatant was carefully discarded. A reporter mix consisting of 90 µl of streptavidin-R-phycoerythrin (1.25 µg/ml in 1 × TMAC; Invitrogen, Carlsbad, CA, USA) was added to each well and the microbead pellet was resuspended. The microplate was then incubated for 5 minutes and analyzed on the Luminex® platform.

Analysis on the Luminex platform. The microplate was analyzed in a Luminex® 200 system at a temperature of 50°C. Analyses were based on counts for 50 beads per set.

Data analysis. Four strains (SARA8, 81-784, 02-7015 and 07-1777) with a known spacer content covering all the spacers in the assay and a blank (in which DNA was replaced with water) were analyzed in each experiment. For all probes except pSTMB26, relative fluorescence unit values were corrected by subtracting the value for the blank. In the case of negative corrected values, an arbitrary value of 25 was attributed. For STMB26, values were corrected by subtracting those for 07-1777, due to a weak cross-reaction observed in emerging monophasic variant strains (07-1777 being a monophasic variant isolate). A cutoff value five times higher than the corrected value was defined. For pSTM03, pSTM07, pSTM12 and pSTMB9 probes and their SNP variant probes pSTM03var1, pSTM07var2, pSTM12var1, and pSTMB9var1, assignment to the wild-type spacer or its SNP variant was based on the ratio of crude values for each probe. If the wild-type probe/SNP variant probe ratio was >1.1, then the wild-type spacer was attributed to the isolate. If the wild-type probe/SNP variant probe ratio was <0.9, then the SNP variant spacer was attributed to the isolate. An R tool was developed to automate the analysis. For each strain an allelic pattern, referred to as the CRISPOL type (CT), consisting of the presence or absence of the 68 ordered CRISPR1, then CRISPR2 probes, followed by the four SNP-variant probes, was generated. Data were incorporated into a dedicated CRISPOL database with BioNumerics® software. A provisional number was assigned to each strain with a new CT, (e.g., CT62prov) until both CRISPR regions had been sequenced, to check for consistency with the CT and for an absence of new spacers.

Acknowledgments

We thank K. Sanderson (Salmonella Genetic Stock Centre, University of Calgary, Canada), A. Brisabois (Anses, Maisons-Alfort, France), M. Mikoleit (Centers for Disease Control and Prevention, Atlanta, USA), C. Bizet (Collection de l'Institut Pasteur, Paris, France), L. Bossi (Centre National de la Recherche Scientifique, Gif-sur-Yvette, France), L. Ward (Health Protection Agency, Colindale, United Kingdom), J.D. Perrier-Gros-Claude (Institut Pasteur de Dakar, Dakar, Senegal), J. Lassen (Norwegian Institute of Public Health, Oslo, Norway), A. Miko (Federal Institute for Risk Assessment (BfR), Berlin, Germany), C.H. Chiu (Chang Gung Children's Hospital, Taoyuan, Taiwan), and R. Hendriksen (Technical University of Denmark, Lyngby, Denmark) for providing some of the strains studied. We thank F. Topin (Luminex B.V., Oosterhout, The Netherlands) for his support.

Supporting Information

Figure S1 *S. enterica* serotype Typhimurium spacers of non classical length.
(DOC)

Figure S2 Distribution of median fluorescence intensity (MFI) values for each probe in a typical CRISPOL experiment with 65 isolates.
(DOC)

Table S1 Direct repeats in the 39 available genomes of *Salmonella* spp.
(DOC)

Table S2 Characteristics, CRISPR alleles, and spacer sequences of the 783 *Salmonella* strains, isolates and genomes studied.
(XLS)

Table S3 DNA sequences of the inventoried spacers.
(XLS)

Table S4 Number of spacers per *Salmonella* serotype.
(DOC)

Table S5 Comparison of CRISPR2 spacer content with the population structure of *S. enterica* serotype Newport as assessed by MLST.
(DOC)

Table S6 CRISPR2 spacer content in various O:9 and O:2 serotypes.
(DOC)

Table S7 Characteristics, PFGE types, MLVA types, phage types, CRISPR alleles, and CRISPOL types of the 158 *S. enterica* serotype Typhimurium strains and isolates (including the monophasic variant) studied.
(XLS)

Table S8 Epidemiological concordance of CRISPR spacer content for separate outbreaks due to serotype Enteritidis.
(DOC)

Table S9 Probe responses in the CRISPOL assay for triplicates of 5 individual strains or isolates.
(DOC)

Author Contributions

Conceived and designed the experiments: FXW LF CS SB. Performed the experiments: LF JZ VG MAD SDR CL CR VP LD SIJ. Analyzed the data: FXW LF. Contributed reagents/materials/analysis tools: GG SLH MG MA. Wrote the paper: FXW. Reviewed, critiqued and offered comments on the text: SB CS LF.

References

1. Voetsch AC, Van Gilder TJ, Angulo FJ, Farley MM, Shallow S, et al. (2004) FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin Infect Dis Suppl 3: S127–134.
2. Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, et al. (2007) Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol 45: 3323–3334.
3. Fisher IS, Enter-net participants (2004) International trends in *Salmonella* serotypes 1998–2003–a surveillance report from the Enter-net international surveillance network. Euro Surveill 9: 45–47.
4. Grimont PAD, Weill FX (2007) Antigenic formulae of the *Salmonella* serovars. 9th ed. Paris, France: WHO Collaborating Center for Reference and Research on *Salmonella*, Institut Pasteur website. Available : http://www.pasteur.fr/ip/portal/action/WebdriveActionEvent/oid/01s-000036-089. Accessed 2012 Apr 23.
5. Bender JB, Hedberg CW, Boxrud DJ, Besser JM, Wicklund JH, et al. (2001) Use of molecular subtyping in surveillance for *Salmonella enterica* serotype Typhimurium. N Engl J Med 344: 189–195.
6. Anderson ES, Ward LR, Saxe MJ, de Sa JD (1977) Bacteriophage-typing designations of *Salmonella typhimurium*. J Hyg 78: 297–300.
7. Olsen JE, Skov MN, Threlfall EJ, Brown DJ (1994) Clonal lines of *Salmonella enterica* serotype Enteritidis documented by IS200-, ribo-, pulsed-field gel electrophoresis and RFLP typing. J Med Microbiol 40: 15–22.
8. Hedberg CW, Greenblatt JF, Matyas BT, Lemmings J, Sharp DJ, et al. (2008) Timeliness of enteric disease surveillance in 6 US states. Emerg Infect Dis 14: 311–313.
9. Lindstedt BA, Heir E, Gjernes E, Kapperud G (2003) DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. J Clin Microbiol 41: 1469–1479.
10. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, et al. (2007) Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect Suppl 3: 1–46.
11. Torpdahl M, Sørensen G, Lindstedt BA, Nielsen EM (2007) Tandem repeat analysis for surveillance of human *Salmonella* Typhimurium infections. Emerg Infect Dis 13: 388–395.
12. Hopkins KL, Maguire C, Best E, Liebana E, Threlfall EJ (2007) Stability of multiple-locus variable-number tandem repeats in *Salmonella enterica* serovar Typhimurium. J Clin Microbiol 45: 3058–3061.
13. Hopkins KL, Peters TM, de Pinna E, Wain J (2011) Standardisation of multilocus variable-number tandem-repeat analysis (MLVA) for subtyping of *Salmonella enterica* serovar Enteritidis. Euro Surveill 16: pii. 19942 p.
14. Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43: 1565–1575.
15. Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8: 172.
16. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327: 167–170.
17. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 35: 907–914.
18. Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151: 653–663.
19. Mokrousov I, Limeschenko E, Vyazovaya A, Narvskaya O (2007) *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. Biotechnol J 2: 901–906.
20. Schouls LM, Reulen S, Duim B, Wagenaar JA, Willems RJ, et al. (2003) Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. J Clin Microbiol 41: 15–26.
21. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709–1712.
22. Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468: 67–71.
23. Touchon M, Rocha EP (2010) The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One 5: e11126.
24. Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, et al. (2011) Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. J Bacteriol 193: 3556–3568.
25. Weill FX, Fabre-Berland L, Guibert V, Diancourt L, Brisse S (2007) Molecular typing and subtyping of *Salmonella* by identification of the variable nucleotide sequences of the CRISPR loci. French Patent Application (no. FR07/09188) filed on December 28, 2007. International Patent Application (no. PCT/IB2008/004004) filed on Dec. 29, 2008.
26. Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, et al. (2011) Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. Appl Environ Microbiol 77: 1946–1956.
27. Liu F, Kariyawasam S, Jayarao BM, Barrangou R, Gerner-Smidt P, et al. (2011) Subtyping *Salmonella enterica* serovar Enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). Appl Environ Microbiol 77: 4520–4526.
28. Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 1: e60.
29. Rousseau C, Gonnet M, Le Romancer M, Nicolas J (2009) CRISPI: a CRISPR interactive database. Bioinformatics 25: 3317–3318.
30. den Bakker HC, Moreno Switt AI, Govoni G, Cummings CA, Ranieri ML, et al. (2011) Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*. BMC Genomics 12: 425.
31. Didelot X, Bowden R, Street T, Golubchik T, Spencer C, et al. (2011) Recombination and population structure in *Salmonella enterica*. PLoS Genet 7: e1002191.
32. Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, et al. (2010) Evolution and population structure of *Salmonella enterica* serovar Newport. J. Bacteriol 192: 6465–6476.
33. Weill FX, Guesnier F, Guibert V, Timinouni M, Demartin M, et al. (2006) Multidrug resistance in *Salmonella enterica* serotype Typhimurium from humans in France (1993 to 2003). J Clin Microbiol 44: 700–708.
34. Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, et al. (2009) Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. Genome Res 19: 2279–2287.
35. Xia S, Hendriksen RS, Xie Z, Huang L, Zhang J, et al. (2009) Molecular characterization and antimicrobial susceptibility of *Salmonella* isolates from infections in humans in Henan Province, China. J Clin Microbiol 47: 401–409.
36. Bone A, Noel H, Le Hello S, Pihier N, Danan C, et al. (2010) Nationwide outbreak of *Salmonella enterica* serotype 4,12:i:- infections in France, linked to dried pork sausage, March-May 2010. Euro Surveill 15: 19592.
37. Wattiau P, Boland C, Bertrand S (2011) Methodologies for *Salmonella enterica* subsp. *enterica* subtyping: gold standards and alternatives. Appl Environ Microbiol 77: 7877–7885.
38. McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI (2011) Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol 49: 565–573.
39. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, et al. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. Plos Pathogens. In revision).
40. Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S, et al. (2006) Evolutionary history of *Salmonella* Typhi. Science 314: 1301–1304.
41. Larsson JT, Torpdahl M, Petersen RF, Sorensen G, Lindstedt BA, et al. (2009) Development of a new nomenclature for *Salmonella typhimurium* multilocus variable number of tandem repeats analysis (MLVA). Euro Surveill 14: 19174.
42. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26: 2465–2466.
43. Zhang J, Abadia E, Refregier G, Tafaj S, Boschiroli ML, et al. (2010) *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. J Med Microbiol 59: 285–294.

PUBLICATION 2

CRISPR is an optimal target for the design of specific PCR assays for *Salmonella enterica* serotypes Typhi and Paratyphi A.

CRISPR est une cible optimale pour la mise au point d'une PCR pour la détection de *Salmonella enterica* sérotypes Typhi et Paratyphi A.

Fabre L, Le Hello S, Roux C, Issenhuth−Jeanjean S, Weill FX.

PLoS Negl Trop Dis. 2014 Jan 30;8(1):e2671. doi: 10.1371/journal.pntd.0002671
6 citations, 5 086 vues (nov. 2016)

CONTEXTE ET OBJECTIFS

Les fièvres typhoïdes et paratyphoïdes restent un problème de santé publique majeur dans les pays en voie de développement, en particulier en Asie du Sud−Est et en Afrique. Leur diagnostic repose principalement sur des méthodes sérologiques peu sensibles et spécifiques (test Widal). Nos précédents travaux sur la région CRISPR ont démontré la possibilité de mise au point de protocoles PCR ciblant des spacers spécifiques de certains sérotypes. Nous avons ainsi développé et validé un protocole PCR simplex et multiplex ainsi qu'un protocole en PCR en temps réel pour l'identification des sérotypes Typhi et Paratyphi A, agents causatifs des fièvres typhoïdes et paratyphoïdes respectivement.

MATERIELS ET METHODES

La sensibilité de la méthode a été testée sur une large collection de souches le plus représentative possible de la diversité génétique de ces pathogènes. Cette collection comprenait un total de 188 souches du sérotype Typhi : 79 souches de la collection, isolées entre 1918 et 2006 représentatives de 65 différents haplotypes, ainsi que la totalité des 109 souches reçues au CNR−ESS en 2009. Pour le sérotype Paratyphi A, un total de 74 souches a été testé : 32 souches de la collection isolées entre 1926 et 2006, ainsi que les 42 souches reçues en 2009 au CNR.

La spécificité de la méthode a été testée sur 71 souches contrôles incluant certaines espèces fréquemment impliquées dans des septicémies : 29 souches de salmonelles non−typhiques et 42 souches de 37 espèces bactériennes différentes.

RESULTATS ET DISCUSSION

La sélection des cibles PCR s'est faite sur l'étude du polymorphisme des loci CRISPR de 18 souches et deux génomes pour le sérotype Typhi et 12 souches et 2 génomes pour le sérotype Paratyphi A réalisée lors de notre précédente étude. Sept profils combinés CRISPR1−CRISPR2 ont été identifiés chez Typhi contre trois pour le sérotype Paratyphi A. Pour l'identification du sérotype Typhi, l'amorce sens se plaçait 206 pb en amont du locus CRISPR2 et l'amorce anti−sens sur la jonction entre le spacer EntB0var1 et sa DR en amont afin de générer un produit PCR de 244 pb. Pour le sérotype Paratyphi A, l'amorce sens se plaçait 293 pb en amont du locus CRISPR1 et l'anti−sens sur le spacer ParA2 pour un produit PCR de 409 pb. Les amorces ont été utilisées en simplex et en multiplex. La sensibilité et la spécificité des deux simplex et du multiplex était de 100 %.

Les mêmes amorces ont été utilisées pour la mise au point d'un protocole PCR en temps réel en SYBRgreen sur 45 ADNs du sérotype Typhi et 36 ADNs du sérotype Paratyphi A sélectionnés aléatoirement contre les 70 ADNs contrôles. La spécificité de la méthode était de 100 %. La limite de détection du protocole Typhi était de 1.7 pg (60 unités formant colonie−UFC) et de 3 pg (90 UFC) pour le sérotype Paratyphi A.

CONCLUSION

Ces travaux confirment que CRISPR est une cible optimale pour la mise au point de protocoles PCR spécifiques de certaines souches d'intérêt. Dans le cas des sérotypes Typhi et Paratyphi A, le protocole est à présent utilisé en routine au CNR−ESS et pourrait être adapté à la détection de ces bactéries dans des échantillons sanguins.

OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

CRISPR Is an Optimal Target for the Design of Specific PCR Assays for *Salmonella enterica* Serotypes Typhi and Paratyphi A

Laetitia Fabre, Simon Le Hello, Chrystelle Roux, Sylvie Issenhuth-Jeanjean, François-Xavier Weill*

Unité des Bactéries Pathogènes Entériques, Centre National de Référence des *Escherichia coli*, *Shigella* et *Salmonella*, WHO Collaborative Center for Reference and Research on *Salmonella*, Institut Pasteur, Paris, France

Abstract

Background: Serotype-specific PCR assays targeting *Salmonella enterica* serotypes Typhi and Paratyphi A, the causal agents of typhoid and paratyphoid fevers, are required to accelerate formal diagnosis and to overcome the lack of typing sera and, in some situations, the need for culture. However, the sensitivity and specificity of such assays must be demonstrated on large collections of strains representative of the targeted serotypes and all other bacterial populations producing similar clinical symptoms.

Methodology: Using a new family of repeated DNA sequences, CRISPR (clustered regularly interspaced short palindromic repeats), as a serotype-specific target, we developed a conventional multiplex PCR assay for the detection and differentiation of serotypes Typhi and Paratyphi A from cultured isolates. We also developed EvaGreen-based real-time singleplex PCR assays with the same two sets of primers.

Principal findings: We achieved 100% sensitivity and specificity for each protocol after validation of the assays on 188 serotype Typhi and 74 serotype Paratyphi A strains from diverse genetic groups, geographic origins and time periods and on 70 strains of bacteria frequently encountered in bloodstream infections, including 29 other *Salmonella* serotypes and 42 strains from 38 other bacterial species.

Conclusions: The performance and convenience of our serotype-specific PCR assays should facilitate the rapid and accurate identification of these two major serotypes in a large range of clinical and public health laboratories with access to PCR technology. These assays were developed for use with DNA from cultured isolates, but with modifications to the assay, the CRISPR targets could be used in the development of assays for use with clinical and other samples.

Citation: Fabre L, Le Hello S, Roux C, Issenhuth-Jeanjean S, Weill F-X (2014) CRISPR Is an Optimal Target for the Design of Specific PCR Assays for *Salmonella enterica* Serotypes Typhi and Paratyphi A. PLoS Negl Trop Dis 8(1): e2671. doi:10.1371/journal.pntd.0002671

Editor: Edward T. Ryan, Massachusetts General Hospital, United States of America

Received September 10, 2013; **Accepted** December 13, 2013; **Published** January 30, 2014

Funding: This work was supported by the Institut Pasteur, the Institut de Veille Sanitaire and by the French Government ''Investissement d'Avenir'' programme (Integrative Biology of Emerging Infectious Diseases'' Laboratory of Excellence, grant no. ANR-10-LABX-62-IBEID). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: francois-xavier.weill@pasteur.fr

Introduction

Typhoid and paratyphoid fevers remain a major health problem in the developing world, particularly in South East Asia and Africa, with estimates of more than 20 million illnesses and over 200,000 deaths due to typhoid fever and more than 5 million illnesses due to paratyphoid fever in 2000 [1]. The causal agents of these diseases, *Salmonella enterica* serotypes Typhi and Paratyphi A, respectively, must be identified early and accurately, to ensure appropriate antimicrobial therapy for the individual concerned, particularly in endemic areas in which the main differential diagnosis is malaria. Correct identification is also fundamental for prevalence and monitoring studies of typhoid and paratyphoid fevers, for the implementation, strengthening and evaluation of health policies (vaccination campaigns, sanitation, water supply treatment, empirical antimicrobial therapy). In high-income countries, in which typhoid fever mostly affects travelers, laboratory confirmation is based on culture, mostly semi-automated blood culture, and further serotyping and antimicrobial drug susceptibility testing of the bacterial isolates. The cost of typing sera and the generalization of accreditation for private laboratories have resulted in the restriction of serotyping to public health laboratories or national reference centers, delaying diagnosis by several days. Furthermore, some isolates of serotypes Typhi and Paratyphi A may not be serotypable due to rough phenotypes (autoagglutination due to the complete or partial loss of lipopolysaccharide) and non motility, and some isolates of serotype Typhi may be classified as not serotypable if overexpression of the Vi polysaccharide masks the O antigens and the Vi antigen is not destroyed by heating before O agglutination. In the regions in which these diseases are endemic, both culture and serotyping are considered expensive and are rarely used. Thus, despite their low sensitivity and/or specificity, serological tests, such as the Widal test, are the predominant method for typhoid diagnosis in these regions [2].

Author Summary

Typhoid and paratyphoid fevers are a major health problem in developing countries, where they are still endemic. Accurate diagnosis remains a challenge, as identification of the causal agents, *Salmonella enterica* serotypes Typhi and Paratyphi A, respectively, requires appropriate facilities (for bacterial culture) and specific reagents (typing antisera). Several *Salmonella* serotype-specific PCR assays have been published and even used directly on biological samples, but none has been validated on a large collection of strains to check for reliability and robustness. We report here the successful development of conventional and real-time PCR assays for the identification of serotypes Typhi and Paratyphi A from cultured isolates. These assays, based on an original target (CRISPR), had a sensitivity and specificity of 100% in tests on a collection of more than 330 strains. The conventional multiplex PCR assay provides an alternative to the serotyping method for laboratories in developing countries unable to obtain a regular supply of typing antisera but with access to PCR technology.

Molecular PCR-based methods have been developed [3–14] (Table 1), to overcome these problems of the need for blood culture (requirement for specialist microbiological facilities, low sensitivity). Given the small numbers of circulating bacteria (<1/ml during infection), several strategies have been applied, involving a broth pre-enrichment step [11,12], nested PCR [3] or real-time PCR [13]. Various molecular targets have been used (Table 1). Initially, the targets chosen were based on genetic structures encoding somatic (O) and/or flagellar (H) antigens and/or Vi antigen [3–5,10–12]. More recently, *in silico* comparative genomics has been used to identify serotype-specific regions [6–9,13,14].

We recently described a new molecular method for *Salmonella* typing, based on polymorphism of the CRISPR (clustered regularly interspaced short palindromic repeats) region [15]. This region contains a new family of repeated DNA sequences [16] characterized by 24–47 bp (29 bp for *Salmonella*) direct repeats separated by variable 21–72 bp (32 bp for *Salmonella*) sequences known as "spacers". In *Salmonella*, this region consists of two CRISPR loci (CRISPR1 and CRISPR2) flanking the Cas machinery (Figure 1). The analysis of 783 strains and genomes from 130 serotypes, including 18 strains and two genomes of serotype Typhi and 12 strains and two genomes of serotype Paratyphi A, led to the identification of more than 3,800 different spacers and showed that the spacer content of a strain was strongly correlated with its serotype. Furthermore, the presence of unique, constant spacers in certain serotypes, such as Typhi and Paratyphi A, was thought to constitute an optimal molecular target for the development of serotype-specific PCR assays. This approach has since been used by Delannoy *et al.* [17,18] for the identification of eight major serotypes of enterohemorrhagic *Escherichia coli*.

The aim of this work was to provide proof-of-principle that CRISPR polymorphism can be used for the development of specific PCRs targeting serotypes of interest. We report here the successful development and validation of singleplex and multiplex, conventional and real-time PCRs for the rapid identification of serotypes Typhi and Paratyphi A from cultured isolates.

Methods

In silico analysis

We used CRISPR sequencing data from a previous study, which included 18 strains and two genomes (Ty2 and CT18, GenBank accession nos. AE014613 and AL627276, respectively) of *S. enterica* serotype Typhi and 12 strains and two genomes (ATCC 9150 and AKU_12601, GenBank accession nos. CP000026 and FM200053, respectively) of *S. enterica* serotype Paratyphi A [15]. A new genome (P-stx-12, GenBank accession no. CP003278) of serotype Typhi was also analyzed.

Bacterial strains and isolates

For assessment of the sensitivity of the serotype-specific PCRs, the *S. enterica* serotype Typhi and Paratyphi A strains and isolates used in this study were selected so as to be representative of the broadest possible genetic diversity for these pathogens. We used collection strains (including the 18 strains of serotype Typhi and 12 of serotype Paratyphi A for which the CRISPR regions had been sequenced) previously shown to be representative of the principal genetic lineages of the two serotypes, together with recent clinical isolates. For serotype Typhi, we used a total of 188 strains and isolates: 79 collection strains (isolated between 1918 and 2006), including the reference strain Ty2, representative of 65 different haplotypes [19] and 109 isolates recovered in 2009. For serotype Paratyphi A, we used a total of 74 strains and isolates: 32 collection strains (isolated between 1926 and 2006), including the reference strain 1K, part of an ongoing whole-genome sequencing project, and 42 isolates recovered in 2009. All the serotype Typhi and Paratyphi A strains and isolates were from the French National Reference Center for *Escherichia coli-Shigella-Salmonella* (FNRC-ESS, Institut Pasteur, Paris) or the WHO Collaborative Center for Reference and Research on *Salmonella* (WHO- Salm, Institut Pasteur, Paris) (Table S1).

For assessment of the specificity of the two serotype-specific PCRs, we used 70 strains of bacteria, some of which are frequently encountered in bloodstream infections. These included 29 non typhoidal *Salmonella* strains (NTS) and 42 strains from 37 species from other genera. These strains were from the FNRC-ESS, the WHO-Salm, the *Collection de l'Institut Pasteur* (CIP, Institut Pasteur, Paris) or the French National Reference Center for *Neisseria* (Institut Pasteur, Paris) (Table 2).

Genomic DNA extraction

Total DNA was extracted from reference strains with the Wizard Genomic DNA purification Kit (Promega, Wisconsin, United States). For clinical isolates, we increased the throughput of the extraction step with the InstaGene matrix kit (Bio-Rad, Marnes-la-Coquette, France). Both DNA extraction methods were performed according to the manufacturer's instructions. For PCR, we used DNA resuspended in a rehydration buffer provided with the Wizard kit or the final supernatant of the InstaGene matrix extraction process.

Serotype-specific PCRs

The PCR specific for *S. enterica* serotype Typhi was carried out with forward primer TY-F (5′-ACGTAGACTCATCCTC-GACC-3′ corresponding to region 2928811–2928830 of *S. enterica* serotype Typhi strain Ty2; GenBank accession no. NC_004631), binding 206 bp upstream from CRISPR2, and the reverse primer TY-R (5′-GCGTGTAGCAGTATTCACCA-3′ corresponding to region 2929055–2929036 of strain Ty2), binding to the DR-EntB0var1 spacer junction of CRISPR2.

The PCR specific for *S. enterica* serotype Paratyphi A was carried out with forward primer PA-F (5′-ACGGGTTGCTGTAAT-GATGC-3′ corresponding to region 2889276–2889295 of *S. enterica* serotype Paratyphi A strain ATCC 9150; GenBank accession no. NC_006511), binding 293 bp upstream from CRISPR1, and the reverse primer PA-R (5′-GCATCATCGCG-

Table 1. Published PCR assays targeting *S. enterica* serotypes Typhi and/or Paratyphi A.

Targeted serotypes	Target genes	Target selection by comparative genomics	Type of PCR assay	Bacterial strains tested (no./origin)	Clinical samples tested (no./origin)	Detection limits[d]	Ref.
Typhi	*fliC*-d[‡]	No	Nested	Typhi (2), other *Salmonella* serotypes (8), non–*Salmonella* (9)	Whole blood (12/Korea)	10 cfu/ml	3
Typhi, Paratyphi A	*tyv**, *prt**, *viaB*[§], *fliC*-d[‡], *fliC*-a[‡]	No	Conventional	Typhi (15/Japan), Paratyphi A (11/Japan), other *Salmonella* serotypes (19), non–*Salmonella* (26)	No	NT	4
Typhi[†], Paratyphi A	*rfbJ**, *tyv**, *prt**, *viaB*[§], *fliC*-d[‡], *fliC*-a[‡]	No	Conventional	Typhi (136/Mali, Chile), Paratyphi A (33/Chile), other *Salmonella* serotypes (541)	No	NT	5
Paratyphi A	spa0180 (*stkF*)[a], spa2473[a], spa2539[a], spa4289 (*hsdM*)[a]	Yes	Conventional	Paratyphi A (52/Malaysia), other *Salmonella* serotypes (75), non–*Salmonella* (14)	No	1×10^5 cfu/ml	6, 7
Typhi[††]	STY1599[b]	Yes	Conventional	Typhi (31/Korea), other *Salmonella* serotypes (43), non–*Salmonella* (37)	No	NT	8
Typhi, Paratyphi A	STY0312[b] (=SPA2476[c]), STY0313–STY0316[b]	Yes	Conventional	Typhi and Paratyphi A (35, India), other *Salmonella* serotypes (14), non–*Salmonella* (11)	Whole blood (78/India)	4 cfu/ml	9
Typhi	*tyv**, *viaB*[§], *ratA*, *fliC*-d[‡],	No	Conventional	Typhi (1), other *Salmonella* serotypes (1)	Sera (18, India)	5 fg	10
Typhi	*fliC*-d[‡]	No	Conventional	Typhi (1)	No	3×10^5 cfu/ml	11, 12
Typhi, Paratyphi A	STY0201[b], SSPA2308[c]	Yes	Real-time	Typhi (81/Nepal), Paratyphi A (61/Nepal), other *Salmonella* serotypes (10), non–*Salmonella* (13)	Whole blood (100/Nepal), bone marrow (28/Vietnam)	2.5×10^2 cfu/ml (Typhi)	13
Typhi[†††], Paratyphi A	*stg*A, SPA1723a-SSPA1724[a] intergenic region, STY4220[b]	Yes	Conventional	Typhi (14/Singapore), Paratyphi A (7/Singapore), other *Salmonella* serotypes (103), non–*Salmonella* (8)	No	1×10^5 cfu/ml (Typhi), 2×10^5 cfu/ml (Paratyphi A)	14
Typhi, Paratyphi A	CRISPR	Yes	Conventional and real-time	Typhi (188/global), Paratyphi A (74/global), other *Salmonella* serotypes (29), non–*Salmonella* (42)	No	1.7 pg (Typhi), 3 pg (Paratyphi A)	This study

[†]This multiplex assay also aims to detect serotype Paratyphi B,
[††]This multiplex assay also aims to detect serotypes Typhimurium, Enteritidis, *S. enterica* subsp. *enterica* and *Salmonella* spp.,
[†††]This multiplex assay also aims to detect *Salmonella* spp.,
*O antigen,
[§]Vi antigen,
[‡]H antigen,
[a]from the serotype Paratyphi A ATCC 9150 genome,
[b]from the serotype Typhi CT18 genome,
[c]from the serotype Paratyphi A AKU_12601 genome.
[d]Results for detection limits are presented in colony-forming units (cfu)/ml when determined with artificially inoculated blood samples (without broth enrichment) or in amount of DNA per PCR when determined by 10-fold dilutions of the bacterial culture; NT, not tested.
doi:10.1371/journal.pntd.0002671.t001

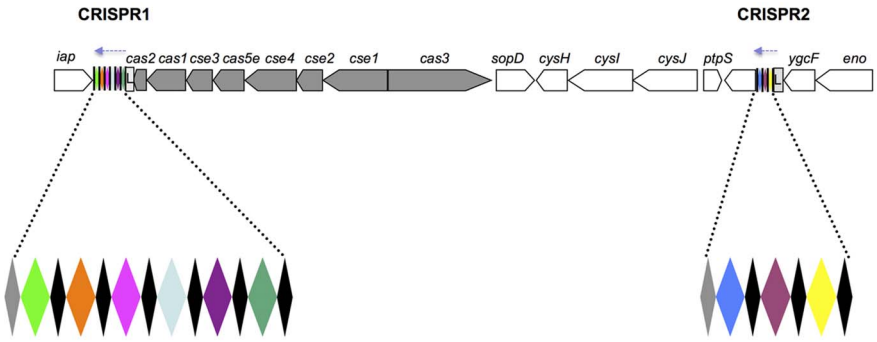


Figure 1. Structure of the CRISPR/Cas system from *S. enterica* serotypes Typhi and Paratyphi A. Two CRISPR loci (CRISPR1 and CRISPR2) flank the CRISPR-associated sequences (cas). The *cas* genes, spacers, direct repeats, leader sequences are represented by gray arrows, colored diamonds, black diamonds, and light gray boxes marked L, respectively. The genomic orientation from *iap* to *ygcF* has been maintained to ensure consistency with previous studies [15,28,29]. The probable transcriptional orientation of the CRISPR loci, as extrapolated from studies in *E. coli* [30,31], is indicated by light blue dashed arrows.
doi:10.1371/journal.pntd.0002671.g001

CATAGTGTC-3′ corresponding to region 2889685–2889666 of strain ATCC 9150), binding to the ParA2 spacer of CRISPR1.

The primers were synthesized by MWG-Biotech (Ebersberg, Germany). The TY-R primer was purified by denaturing polyacrylamide gel electrophoresis (HYPUR grade), whereas the remaining primers were purified by either high-pressure liquid chromatography (HPLC grade) or reverse-phase chromatography (HPSF grade).

Singleplex PCR was performed in a reaction volume of 50 µl containing DNA (2 µl for the InstaGene matrix or 1 µl diluted 10-fold for Wizard), primers (TY-F and TY-R or PA-F and PA-R, 10 pmol each), GoTaq Flexi DNA polymerase (0.85 U) and its buffer (Promega), deoxyribonucleotide triphosphate (0.1 mM each), $MgCl_2$ (1.5 mM), dimethyl sulfoxide (5%), and sterile water (Aqua B. Braun, Melsungen, Germany). The cycling conditions were as follows: 2 min of denaturation at 94°C, followed by 25 cycles of denaturation for 15 s at 94°C, annealing for 15 s at 56°C and polymerization for 30 s at 72°C and a final extension phase at 72°C for 5 min. Amplification products were visualized by electrophoresis in a 2.2% agarose gel. The sizes of the expected PCR products for serotypes Typhi (TY-F and TY-R primers) and Paratyphi A (PA-F and PA-R) strains were 244 bp and 409 bp, respectively.

Multiplex PCR was performed in a reaction volume of 50 µl containing DNA (2 µl for the InstaGene matrix or 1 µl diluted 10-fold for Wizard), TY-F and TY-R primers (10 pmol each), PA-F and PA-R (5 pmol each), GoTaq Flexi DNA polymerase (0.85 U) and its buffer (Promega), deoxyribonucleotide triphosphate (0.1 mM each), $MgCl_2$ (1.5 mM), dimethyl sulfoxide (5%), and sterile water (Aqua B. Braun). The cycling conditions were as follows: 2 min of denaturation at 94°C, followed by 25 cycles of denaturation for 15 s at 94°C, annealing for 15 s at 56°C and polymerization for 30 s at 72°C and a final phase of extension, for 5 min at 72°C. Amplification products were visualized by electrophoresis in a 2.2% agarose gel.

The DNA of the reference strains of serotypes Typhi (Ty2) and Paratyphi A (1K) was used as a positive control for all these PCRs. Negative controls were obtained by replacing the DNA with sterile water (Aqua B. Braun).

EvaGreen-based real-time PCR assay

The same primers as for standard PCR were used in singleplex reactions. For each set of primers, the optimized EvaGreen protocol was carried out in a final volume of 20 µl containing DNA (2 µl diluted 10-fold for the InstaGene matrix or 2 µl diluted 100-fold for Wizard), 10 µl of SsoFast^{TM} EvaGreen Supermix (Bio-Rad), 100 nM PA-F and PA-R primers or 300 nM TY-F and TY-R primers and sterile water (Aqua B. Braun). PCR was carried out on a CFX96^{TM} Real-Time PCR detection system (Bio-Rad) with the following cycling conditions: 98°C for 2 min, followed by 40 cycles of 98°C for 5 s, 62°C for 2 s and 72°C for 10 s). A melting curve analysis was performed for the amplicons obtained, immediately after the completion of PCR cycling, with a ramp of 0.5°C s^{-1} over the interval 60 to 95°C. We calculated Ct values from automatic baseline settings and the melting curve analyses were performed with CFX Manager version 2.1 (Bio-Rad).

The detection limits were determined by two methodologies. First, we amplified, in duplicate, serial dilutions of genomic Wizard DNA from the reference strains Ty2 and 1K of serotypes Typhi and Paratyphi A, respectively. The concentrations of the Wizard DNA extracts were determined with both a NanoDrop 2000 and a NanoDrop 8000 spectrophotometer (Thermo Scientific, France) and these extracts were then diluted in sterile water (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 2×10^{-5}, 4×10^{-5}, 6×10^{-5}, 8×10^{-5}, 10^{-6} dilutions tested). Second, serial dilutions of cultures of *S. enterica* serotype Typhi Ty2 and Paratyphi A 1K cells were used as a template for PCR. An overnight culture with aeration at 37°C in trypticase soy broth (Bio-Rad) was adjusted to McFarland 0.5 then serially diluted 10-fold (from 10^{-1} to 10^{-7}) with sterile water. We took a 1 ml aliquot of each dilution, in duplicate, centrifuged it for 5 min at 13.800×*g* then resuspended it in 200 µl of InstaGene Matrix (Bio-Rad). For each DNA extract, a 2 µl undiluted aliquot from InstaGene matrix and one aliquot diluted 10-fold were used as templates for PCR. Viable cell counts were obtained by plating 10 or 100 µl of each dilution of bacterial culture in duplicate on trypticase soy agar plates and incubating overnight at 37°C. DNA samples were considered positive if the Ct values were <40 and there was a clear melting curve peak at the expected temperature.

The sensitivity of each real-time PCR assay was assessed by testing either 46 isolates of serotype Typhi or 37 isolates of serotype Paratyphi A. Specificity was assessed by testing negative controls, including the 43 serotype Typhi isolates (for the Paratyphi A-specific real-time PCR assay), 37 serotype Paratyphi A isolates (for the Typhi-specific real-time PCR assay), 29 other NTS strains, 18 non-*Salmonella* strains, and sterile water (Aqua B. Braun). A 10^{-2} dilution of the Wizard DNA extracts of the Ty2

Table 2. Results of conventional PCR assays targeting *S. enterica* serotypes Typhi and Paratyphi A.

S. enterica **subsp.** *enterica* **serotype**	**No. of strains tested**[1]	**PCR Typhi**[2]	**PCR Paratyphi A**[2]
Typhi	188	+	−
Paratyphi A	74	−	+
Typhimurium	1	−	−
Enteritidis	1	−	−
Dublin	1	−	−
Newport	1	−	−
Paratyphi B biovar Java	1	−	−
Paratyphi B	1	−	−
Paratyphi C	1	−	−
Choleraesuis *sensu stricto*	1	−	−
Choleraesuis var. Decatur	1	−	−
Muenchen	1	−	−
Manhattan	1	−	−
Mikawasima	1	−	−
Nitra	1	−	−
Blockley	1	−	−
Blegdam	1	−	−
Mbandaka	1	−	−
Canada	1	−	−
Indiana	1	−	−
Emek	1	−	−
Urbana	1	−	−
Muenster	1	−	−
Virchow	1	−	−
Brandenburg	1	−	−
Panama	1	−	−
Schwarzengrund	1	−	−
Poona	1	−	−
Heidelberg	1	−	−
Kentucky	1	−	−
Oranienburg	1	−	−
Other species			
Escherichia coli	5	−	−
Escherichia blattae	1	−	−
Escherichia vulneris	1	−	−
Shigella dysenteriae 1	1	−	−
Shigella boydii 1	1	−	−
Shigella dysenteriae 2	1	−	−
Shigella sonnei g	1	−	−
Shigella flexneri 2a	1	−	−
Enterobacter aerogenes	1	−	−
Enterobacter kobei	1	−	−
Enterobacter gergoviae	1	−	−
Enterobacter cloacae	1	−	−
Citrobacter farmerii	1	−	−
Citrobacter rodentium	1	−	−
Citrobacter murliniae	1	−	−
Citrobacter werkmanii	1	−	−
Citrobacter sedlakii	1	−	−

Table 2. Cont.

S. enterica subsp. *enterica* serotype	No. of strains tested[1]	PCR Typhi[2]	PCR Paratyphi A[2]
Citrobacter braakii	1	–	–
Citrobacter amalonaticus	1	–	–
Citrobacter diversus	1	–	–
Edwardsiella tarda	1	–	–
Kluyvera cryocrescens	1	–	–
Pantoea agglomerans	1	–	–
Serratia marcescens	1	–	–
Leminorella grimontii	1	–	–
Buttauxiella agrestis	1	–	–
Yersinia pseudotuberculosis	1	–	
Providencia rettgeri	1	–	–
Dickeya chrysanthemi	1	–	–
Klebsiella pneumoniae	1	–	–
Morganella morganii	1	–	–
Proteus vulgaris	1	–	–
Streptococcus pyogenes	1	–	–
Staphylococcus aureus	1	–	–
Staphylococcus epidermidis	1	–	–
Streptococcus agalactiae	1	–	–
Streptococcus pneumoniae	1	–	–
Neisseria meningitidis	1	–	–

[1]The details of the strains are provided in Supplemental Table 1.
[2]+, amplicon of the expected size; −, no amplicon of the expected size.
doi:10.1371/journal.pntd.0002671.t002

and 1K reference strains was used as a positive control for the tests.

Statistical analysis

The mean and standard deviation of the Ct values were calculated with Excel (Microsoft).

Results

Design of CRISPR-based primers for serotype Typhi- and Paratyphi A-specific PCRs

The selection of targets for serotype Typhi- and Paratyphi A-specific PCRs was based on the analysis of CRISPR region polymorphism in 18 strains and two genomes (Ty2 and CT18) of serotype Typhi and 12 strains and two genomes (ATCC 9150 and AKU_12601) of serotype Paratyphi A, as previously reported [15]. Seven combined CRISPR1-CRISPR2 profiles were identified for serotype Typhi (Figure 2), and three were identified for serotype Paratyphi A (Figure 3).

We selected CRISPR2 as the target for serotype Typhi-specific PCR, as no CRISPR1 spacer was common to all 20 strains studied or to all the genomes of this serotype. The CRISPR2 of serotype Typhi contains a unique, constant spacer, EntB0var1. However, this spacer is also known to be present as the first spacer (probably corresponding to the last transcribed) of CRISPR2 in a rare serotype, Emek [15]. It is also a single nucleotide polymorphism (SNP) variant of the EntB0 spacer, located in the first position of CRISPR2 in at least 20 serotypes, including the prevalent serotype Enteritidis [15]. The forward primer (TY-F) was designed to bind 206 bp upstream from the CRISPR2 locus (TY-F), whereas the reverse primer (TY-R) was designed to bind at the junction of EntB0var1 and its upstream DR (Figures 3 and 4), to prevent amplification in serotype Emek, which also contains EntB0var1, and in serotypes containing EntB0. The DR upstream from EntB0var1 in serotype Typhi (DR27) contains a SNP with respect to the sequences of Emek (DR1G) and Enteritidis (DR1C), and this SNP was used for annealing to the 3′ end of the HPLC-purified TY-R primer. We also deliberately incorporated a destabilizing mismatch at position −4 from the 3′-end of TY-R, to increase the specificity of the PCR [20].

For *S. enterica* serotype Paratyphi A detection, the forward primer (PA-F) was designed to bind 293 bp upstream from the CRISPR1 locus, whereas the ParA2 spacer of CRISPR1 was used as a template for the reverse primer (PA-R), as it was encountered in all the Paratyphi A strains and genomes studied and was found only in this serotype.

PCR identification of *S. enterica* serotypes Typhi and Paratyphi A from cultured isolates

The PCR described here was developed for DNAs obtained from bacterial isolates. This PCR is based on the use of one pair of primers targeting serotype Typhi (TY-F and TY-R) and one pair of primers targeting serotype Paratyphi A (PA-F and PA-R). These primers can be used in singleplex or multiplex PCR assays.

All of the 188 strains of serotype Typhi tested yielded only the 244 bp amplicon expected for this serotype, whereas the 74 strains of serotype Paratyphi A yielded only the 409 bp amplicon expected for this serotype (Figure 5). All NTS tested gave negative results, as did the other bacterial species tested

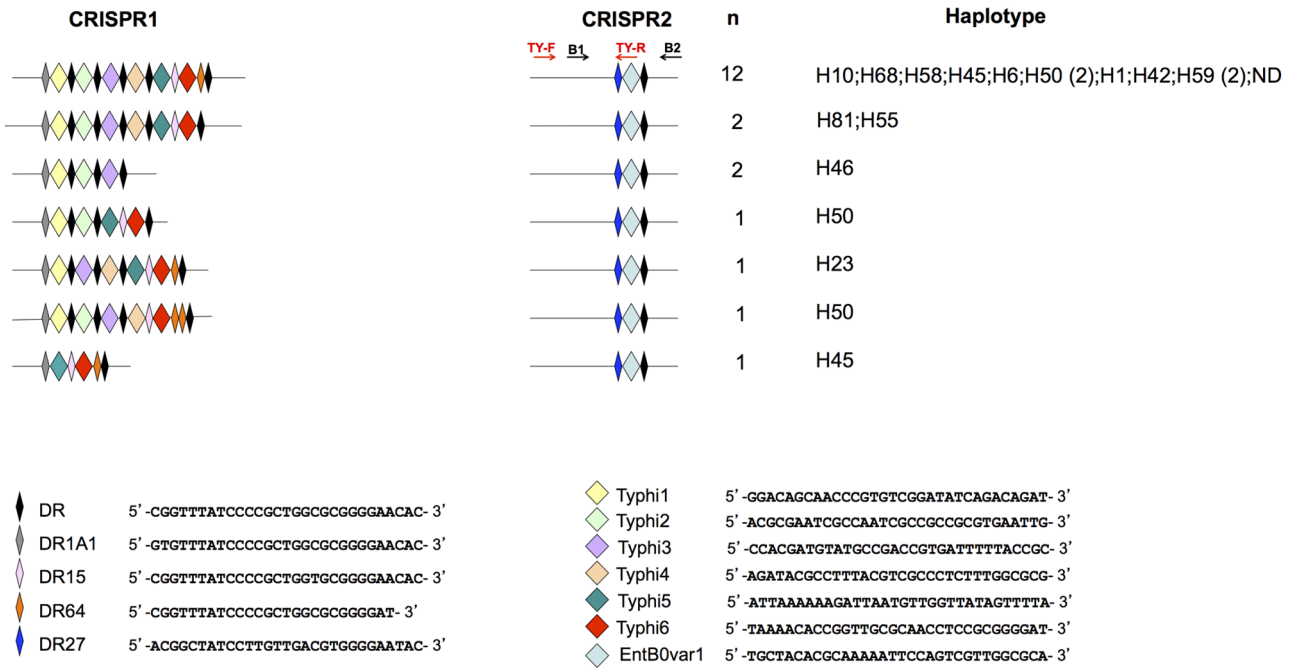


Figure 2. CRISPR profiles obtained for the 18 strains and two genomes of *S. enterica* serotype Typhi. The spacer sequences and direct repeat (DR) sequences are indicated. The set of strains and genomes has been reported elsewhere [15]. "n" is the number of strains harboring each profile. Haplotypes [19] are indicated, to illustrate the genetic diversity of the strains studied. ND, not determined. The primers used for serotype Typhi-specific amplification (TY-F and TY-R) and those for amplification of the entire CRISPR2 sequence (B1 and B2) [15] are indicated in different colors.
doi:10.1371/journal.pntd.0002671.g002

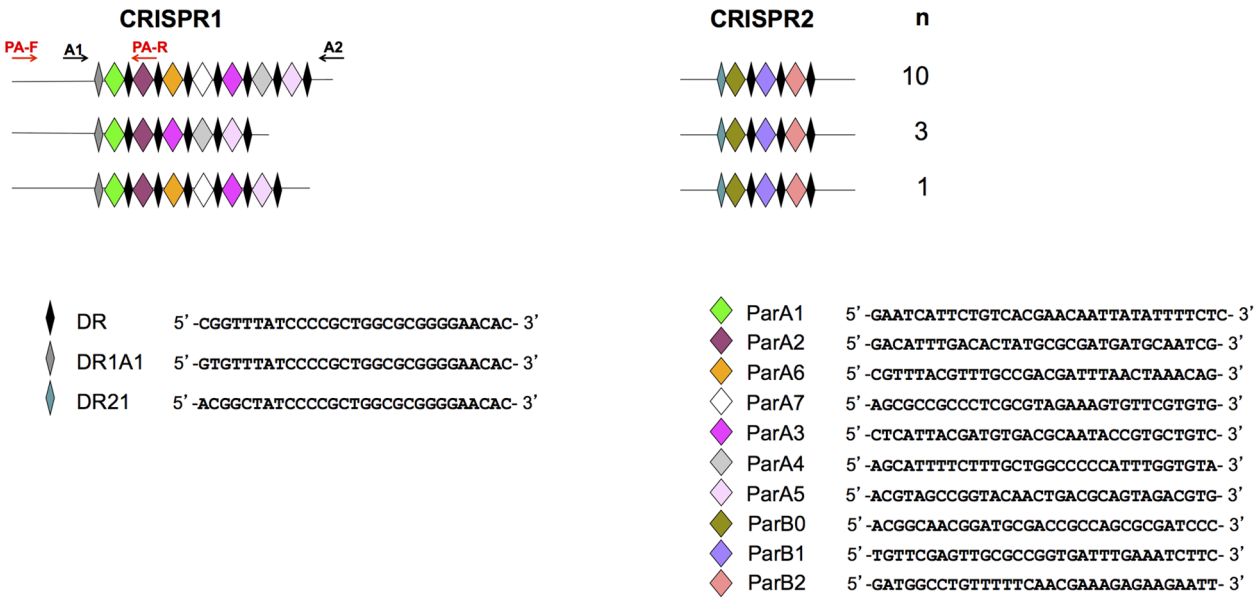


Figure 3. CRISPR profiles obtained for 12 strains and two genomes of *S. enterica* serotype Paratyphi A. The spacer sequences and direct repeat (DR) sequences are indicated. The set of strains and genomes has been reported elsewhere [15]. n" is the number of strains harboring each profile. The primers used for serotype Paratyphi A-specific amplification (PA-F and PA-R) and for amplification of the entire CRISPR1 sequence (A1 and A2) [15] are indicated in different colors.
doi:10.1371/journal.pntd.0002671.g003

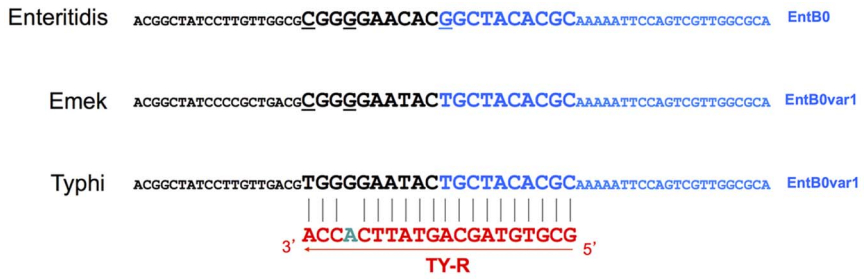


Figure 4. Strategy used to design TY-R, the reverse primer for the serotype Typhi-specific PCR assay. The DR sequence upstream from EntB0/EntB0var1 is indicated in black letters. The spacers EntB0 and EntB0var1 are indicated in blue letters. The nucleotides belonging to the template for primer TY-R are represented by letters of larger size. The sequence of TY-R is indicated in red, with the deliberate mismatch in green. Mismatch positions between TY-R and the templates of serotypes Emek (EntB0) and Enteritidis (EntB0var1) are underlined.
doi:10.1371/journal.pntd.0002671.g004

(Table 2). Similar results were obtained in singleplex and multiplex assays.

EvaGreen-based real-time PCR assay

Each pair of primers (TY-F and TY-R; PA-F and PA-R) was tested separately on a subset of DNAs, in duplicate.

In the serotype Typhi-specific real-time PCR assay, all 45 DNAs from the serotype Typhi isolates randomly selected from those isolated at the FNRC-ESS in 2009, together with the reference strain Ty2, gave Ct values between 20.5 and 26.2 (mean value 22.8±1.7) and a single melting curve peak at 82.5–83°C in fluorescence melting curve analysis. No positive results were obtained for the 37 serotype Paratyphi A strains, the 29 NTS strains (including serotype Emek, Dublin and Enteritidis strains) or the other bacterial species (18 tested) (Figure 6). Only one strain (of serotype Emek) had a Ct value below 40 (38.6) but no melting curve peak.

In the serotype Paratyphi A-specific assay, all 36 DNAs from the serotype Paratyphi A isolates randomly selected from those isolated at the FNRC-ESS in 2009, together with the reference strain 1K, gave Ct values between 21.2 and 26.5 (mean value 23.1±1.1) and a single melting curve peak at 85–85.5°C. No positive results (Ct values always >40 and no melting curve peak) were obtained for the 43 serotype Typhi strains, the other NTS strains or the other bacterial species (Figure 6).

The limit of detection of the EvaGreen-based real-time PCR assay was 1.7 pg for Typhi and 3 pg for Paratyphi A, for the testing of series of purified DNA extract dilutions for the reference strain of each serotype. A sensitivity of ≈60 viable bacterial cells per PCR for Typhi and ≈90 for Paratyphi A was obtained when serial dilutions of bacterial cell cultures were used as the PCR template.

Discussion

We show here that CRISPR regions could be used as an original target for the development of PCR assays specific for particular *Salmonella* serotypes containing constant and unique spacers. As a proof-of-principle, we have successfully developed such PCRs for serotypes Typhi and Paratyphi A, the causal agents of typhoid and paratyphoid fevers.

Traditionally, the identification of these two serotypes has been based on serotyping, which requires the culture of the isolate and an adequate set of polyclonal rabbit antisera. However, these antisera are costly and are therefore not always available for use in the clinical or public health laboratories of the regions in which these diseases are endemic. Furthermore, some isolates cannot be serotyped, due to changes in the structure of the lipopolysaccharide or its accessibility to typing O antisera, or to the synthesis or assembly of the flagellum structure. PCR is increasingly being used, but the main challenge in such approaches is the

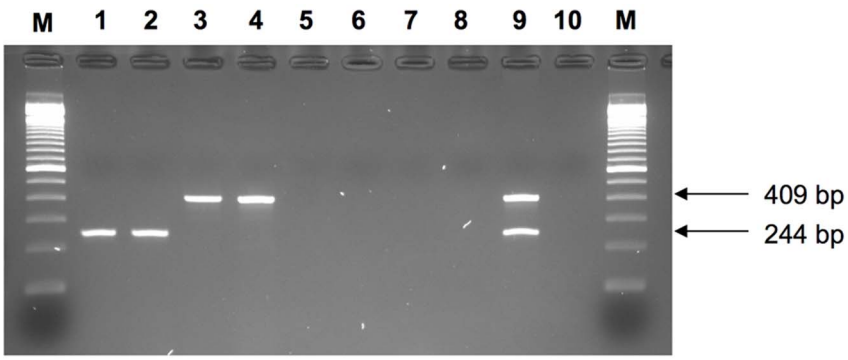


Figure 5. Gel electrophoresis image of the multiplex PCR used for the screening of *S. enterica* serotypes Typhi and Paratyphi A. M: 100-bp DNA ladder (Invitrogen, Carlsbad, CA, USA); lane 1, *S. enterica* serotype Typhi isolate 09-2213, lane 2, *S. enterica* serotype Typhi isolate 09-6791; lane 3, *S. enterica* serotype Paratyphi A isolate 09-2344; lane 4, *S. enterica* serotype Paratyphi A isolate 09-9426; lane 5, *S. enterica* serotype Typhimurium strain LT2; lane 6: *S. enterica* serotype Emek strain 297K; lane 7, *Escherichia coli* isolate 10-7043; lane 8: *Shigella flexneri* isolate 11-1445; lane 9, *S. enterica* serotypes Typhi and Paratyphi A reference strains Ty2 and 1K, respectively; lane 10, negative control (sterile water). We checked that the template DNA was of sufficiently high quality, by amplifying the housekeeping genes *aroC* for *Salmonella* and *adk* for *E. coli*/*Shigella* (data not shown), as previously described [32,33].
doi:10.1371/journal.pntd.0002671.g005

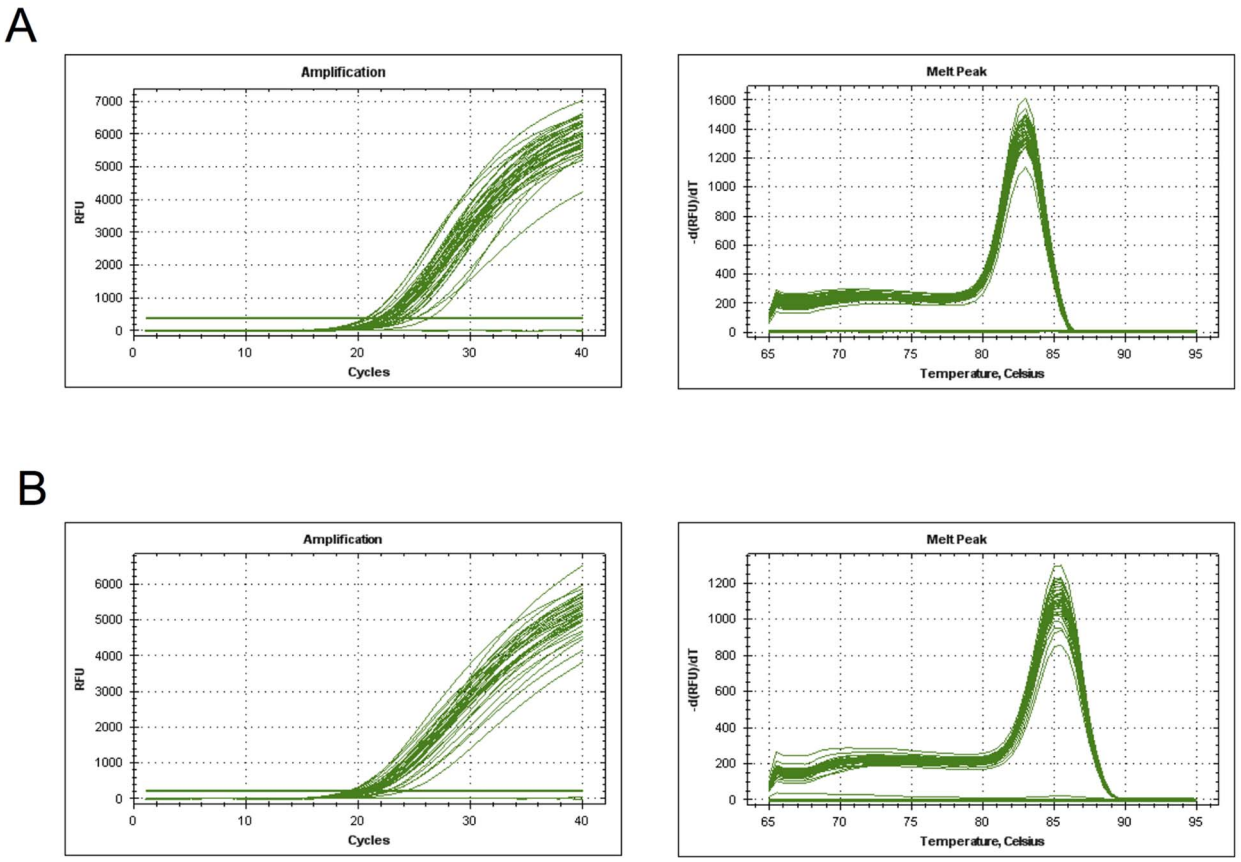


Figure 6. EvaGreen real-time PCR assay. (A) Amplification and melting curves for 43 serotype Typhi isolates tested in duplicate with the TY-F and TY-R primers, yielding a mean Ct value of 22.6±1.1 and a single melting curve peak at 82.5–83°C. The two negative controls (*S. enterica* serotype Paratyphi A 1K and sterile water), also tested in duplicate, appear as flat lines. (B) Amplification and melting curves for 37 serotype Paratyphi A isolates tested in duplicate with the PA-F and PA-R primers, yielding a mean Ct value of 23.1±1.1 and a melting curve peak at 85–85.5°C. The two negative controls (*S. enterica* serotype Typhi Ty2 and sterile water), also tested in duplicate, appear as flat lines.
doi:10.1371/journal.pntd.0002671.g006

identification of a stable molecular target specific to the serotype concerned. The molecular targets initially selected were related to serotype: genes involved in the biosynthesis of the flagellar (*fliC* and *fljB*) and/or O-polysaccharide (encoded by the *rfb* locus) Vi antigens. However, none of these genes, in isolation, can be considered specific for the identification of serotypes Typhi and Paratyphi A. Indeed, the Vi locus is not always present in serotype Typhi [21,22], and *fliC-d* is not sufficient to affirm an identification of serotype Typhi, as it is also common to 145 NTS serotypes [23]. Thus, for 100% specificity to be achieved with this approach, several targets must be incorporated into the PCR assay. A new approach for identifying targets on the basis of genome comparison analysis has recently been described. However, the complete *Salmonella* genomes available cannot really be considered representative of the entire genus (39 complete genomes online, covering only 25 serotypes from the >2,500 described when this work was begun in 2011). Moreover, only three assembled genomes of serotypes Typhi (Ty2, CT18, and P-stx-12) and two of Paratyphi A (ATCC 9150 and AKU_1260) are available online. Seventeen 454 shotgun-sequenced serotype Typhi strains representing various genetic lineages [24] have been deposited in whole-genome sequence databases since 2008, but have rarely been used for the selection of serotype-specific targets [13]. This ideal approach for identifying good targets may therefore currently be biased by the limited number of genome sequences available. Furthermore, the CRISPR loci comprise only arrays of DR-spacer units in non coding regions, and this interesting target has therefore generally been ignored in comparative genomics studies, in which targets have generally been selected from open reading frames (ORFs).

Another key limitation of many published studies on serotype-specific PCR assays is the application of the validation step to a limited number of strains and/or blood samples, collected locally (Table 1). In our study, a sensitivity of 100% (i.e., the stability of the DNA target) was obtained for a large collection of 262 serotype Typhi and Paratyphi A strains chosen to be representative of the genetic diversity of these pathogens. This collection included old and recently isolated strains from around the world and strains from various genetic lineages, as previously demonstrated by haplotyping [19] and confirmed [24] or currently in the process of being confirmed by whole-genome sequencing (unpublished). The 100% specificity of the assays was also confirmed with 70 other strains, including NTS from 29 serotypes frequently found associated with bloodstream infections [25], and 42 non-*Salmonella* strains. We are therefore very confident that these CRISPR targets are suitable for the specific detection of all populations of serotypes Typhi and Paratyphi A.

The CRISPR-based PCR assays we have developed were validated on cultured bacterial isolates and may facilitate the identification of these serotypes in addition to (gain of time) or as an alternative to serotyping (lack of availability of typing sera, rough or non motile strains) in clinical or public health

laboratories. The next step will be to validate a CRISPR-based PCR assay suitable for use on blood samples in the field. For this purpose, in addition to conventional PCR, we have developed an EvaGreen-based real-time PCR assay with a reasonably good sensitivity when applied to purified DNA. As this real-time PCR was initially developed for well-equipped laboratories (blood culture and real-time PCR technology), we used the same primers as designed for the conventional multiplex PCR, generating amplicons of 244 bp for serotype Typhi and 409 bp for serotype Paratyphi A. These sizes are convenient for rapid differential identification by the agarose gel electrophoresis of PCR products, but it will be necessary to decrease the size of the amplified regions for both serotypes to about 70 to 100 bp to optimize the sensitivity of the real-time PCR assay, an essential requirement when dealing with biological samples, particularly blood. This could be done by modifying the forward primers, TY-F and PA-F, to make them bind further downstream, and the reverse primer, PA-R, to make it bind further upstream, within the ParA1 spacer. An internal control amplification target should also be identified and included in the assay. Of course, other factors such as blood volume, the nature of the anticoagulant used and the presence of a pre-enrichment step are important for the optimization of real-time PCR assays of this type. For laboratories not equipped with real-time PCR technology located in countries in which these diseases are endemic, a nested PCR could be designed to improve the detection limits. This CRISPR-based nested PCR could use TY-F and B2 as the first-round primers and B1 and TY-R as the second-round primers for serotype Typhi (Figure 2). For serotype Paratyphi A, PA-F and A2 could be used as the first-round primers and A1 and PA-R could be used as the second-round primers (Figure 3). The A1, A2, B1 and B2 primers are those previously used for amplification of the entire CRISPR loci [15]. All these primers could potentially be combined, for the development of a multiplex nested PCR targeting both serotypes, Typhi and Paratyphi A. It would also be possible to use a recently developed technique: the loop-mediated isothermal amplification (LAMP) of DNA [26]. This rapid and sensitive amplification method makes use of four to six primers, which recognize six to eight regions of the target DNA (here, CRISPR). This method is suitable for use in resource-limited settings, as a water-bath can be used in place of a thermocycler and visual colorimetric detection by staining with a dye, such as SYBR Green I, can replace gel electrophoresis with ethidium bromide staining [27].

In conclusion, whereas most other published PCR protocols use combinations of different markers for each targeted serotype, with a non optimal validation step, we were able to identify a unique target in a conserved region of the *Salmonella* genome for each targeted serotype and we demonstrated that the two targets were highly stable and specific, in tests of a large, representative collection of strains of serotypes Typhi and Paratyphi A. The multiplex conventional and EvaGreen-based real-time PCR assays described here can be used for the rapid confirmation of identification for these two major serotypes and for differentiation between them, using DNA from cultured isolates. These assays, which are now routinely used at the FNRC-ESS, would be easy to implement in a large range of clinical and public health laboratories with access to PCR technology. Finally, we confirmed that CRISPR was an optimal target for the development of a DNA amplification assay for the detection of any strain of interest with a particular spacer content, provided that a culture of this strain is available.

Supporting Information

Table S1 Characteristics, CRISPR alleles, sequence types, haplotypes of the 338 *Salmonella* strains, isolates and genomes studied. Excel file.
(XLS)

Acknowledgments

We would like to thank all corresponding laboratories of the French National Reference Center for *E. coli-Shigella-Salmonella*. We thank M.K. Taha and S. Brisse for providing DNA for bacteria other than *Salmonella*, *E. coli* and *Shigella*.

Author Contributions

Conceived and designed the experiments: FXW LF. Performed the experiments: LF CR. Analyzed the data: FXW LF. Contributed reagents/materials/analysis tools: SLH SIJ. Wrote the paper: FXW SLH LF.

References

1. Crump JA, Luby SP, Mintz ED (2004) The global burden of typhoid fever. Bull World Health Organ 82: 346–353.
2. Wain J, Hosoglu S (2008) The laboratory diagnosis of enteric fever. J Infect Dev Ctries 2: 421–425.
3. Song JH, Cho H, Park MY, Na DS, Moon HB, et al. (1993) Detection of *Salmonella* Typhi in the blood of patients with typhoid fever by polymerase chain reaction. J Clin Microbiol 31: 1439–1443.
4. Hirose K, Itoh K, Nakajima H, Kurazono T, Yamaguchi M, et al. (2002) Selective amplification of *tyv* (*rfbE*), *prt* (*rfbS*), *viaB*, and *fliC* genes by multiplex PCR for identification of *Salmonella enterica* serovars Typhi and Paratyphi A. J Clin Microbiol 40: 633–636.
5. Levy H, Diallo S, Tennant SM, Livio S, Sow SO, et al. (2008) PCR method to identify *Salmonella enterica* serovars Typhi, Paratyphi A, and Paratyphi B among *Salmonella* isolates from the blood of patients with clinical enteric fever. J Clin Microbiol 46: 1861–1866.
6. Ou HY, Ju CT, Thong KL, Ahmad N, Deng Z, et al. (2007) Translational genomics to develop a *Salmonella enterica* serovar Paratyphi A multiplex polymerase chain reaction assay. J Mol Diagn 9: 624–630.
7. Teh CS, Chua KH, Puthucheary SD, Thong KL (2008) Further evaluation of a multiplex PCR for differentiation of *Salmonella* Paratyphi A from other salmonellae. Jpn J Infect Dis 61: 313–314.
8. Park SH, Kim HJ, Cho WH, Kim JH, Oh MH, et al. (2009) Identification of *Salmonella enterica* subspecies I, *Salmonella enterica* serovars Typhimurium, Enteritidis and Typhi using multiplex PCR. FEMS Microbiol Lett 301: 137–146.
9. Nagarajan AG, Karnam G, Lahiri A, Allam US, Chakravortty D (2009) Reliable means of diagnosis and serovar determination of blood-borne *Salmonella* strains: quick PCR amplification of unique genomic loci by novel primer sets. J Clin Microbiol 47: 2435–2441.
10. Kumar A, Balachandran Y, Gupta S, Khare S, Suman (2010) Quick PCR based diagnosis of typhoid using specific genetic markers. Biotechnol Lett 32: 707–712.
11. Zhou L, Pollard AJ (2010) A fast and highly sensitive blood culture PCR method for clinical detection of *Salmonella enterica* serovar Typhi. Ann Clin Microbiol Antimicrob 9: 14. doi: 10.1186/1476-0711-9-14.
12. Zhou L, Pollard AJ (2012) A novel method of selective removal of human DNA improves PCR sensitivity for detection of *Salmonella* Typhi in blood samples. BMC Infect Dis 12: 164. doi: 10.1186/1471-2334-12-164.
13. Nga TV, Karkey A, Dongol S, Thuy HN, Dunstan S, et al. (2010) The sensitivity of real-time PCR amplification targeting invasive *Salmonella* serovars in biological specimens. BMC Infect Dis 10: 125. doi: 10.1186/1471-2334-10-125.
14. Ngan GJ, Ng LM, Lin RT, Teo JW (2010) Development of a novel multiplex PCR for the detection and differentiation of *Salmonella enterica* serovars Typhi and Paratyphi A. Res Microbiol 161: 243–248.
15. Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V et al. (2012) CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One 7: e36995. doi: 10.1371/journal.pone.0036995.
16. Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43: 1565–1575.
17. Delannoy S, Beutin L, Burgos Y, Fach P. (2012) Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains by use of the CRISPR locus as a target for a diagnostic real-time PCR. J Clin Microbiol 50: 3485–3492.
18. Delannoy S, Beutin L, Fach P. (2012) Use of clustered regularly interspaced short palindromic repeat sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26:H11, O45:H2,

O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR. J Clin Microbiol 50: 4035–4040.
19. Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S, et al. (2006) Evolutionary history of *Salmonella* Typhi. Science 314: 1301–1304.
20. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, et al. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17: 2503–2516.
21. Hashimoto Y, Li N, Yokoyama H, Ezaki T. (1993) Complete nucleotide sequence and molecular characterization of *ViaB* region encoding Vi antigen in *Salmonella* Typhi. J Bacteriol 175: 4456–4465.
22. Nair S, Alokam S, Kothapalli S, Porwollik S, Proctor E, et al. (2004) *Salmonella enterica* serovar Typhi strains from which SPI7, a 134-kilobase island with genes for Vi exopolysaccharide and other functions, has been deleted. J Bacteriol 186: 3214–3223.
23. Grimont PAD, Weill FX. (2007) Antigenic formulae of the *Salmonella* serovars. 9th ed. Paris, France: WHO Collaborating Center for Reference and Research on *Salmonella*, Institut Pasteur website. 23: Available : http://www.pasteur.fr/ip/portal/action/WebdriveActionEvent/oid/01s-000036-089. Accessed 2013 December 3.
24. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, et al. (2008) High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. Nat Genet 40: 987–993.
25. Jones TF, Ingram LA, Cieslak PR, Vugia DJ, Tobin-D'Angelo M, et al. (2008) Salmonellosis outcomes differ substantially by serotype. J Infect Dis 198: 109–114.
26. Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, et al. (2000) Loop-mediated isothermal amplification of DNA. Nucleic Acids Res 28: E63.
27. Njiru ZK, Mikosza AS, Armstrong T, Enyaru JC, Ndung'u JM, et al. (2008) Loop-mediated isothermal amplification (LAMP) method for rapid detection of *Trypanosoma brucei rhodesiense*. PLoS Negl Trop Dis 2:e147. doi: 10.1371/journal.pntd.0000147.
28. Touchon M, Rocha EP. (2010) The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One 5:e11126. doi: 10.1371/journal.pone.0011126.
29. Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, et al. (2011) Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. Appl Environ Microbiol 77: 1946–1956.
30. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, et al. (2010) Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol 75: 1495–1512.
31. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, et al. (2010). Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. Mol Microbiol 77: 1367–1379.
32. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, et al. (2012) Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog 8: e1002776. doi: 10.1371/journal.ppat.1002776.
33. Wirth T, Falush D, Lan R, Colles F, Mensa P, et al. (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60: 1136–51.

PUBLICATION 3

CRISPR is a target of choice for PCR detection of-fluoroquinolone resistant *Salmonella enterica* serotype Kentucky ST198

Fabre L., Weill F.X., Le Hello S.

Article en cours de préparation pour la revue Journal of Clinical Microbiology.

CONTEXTE ET OBJECTIFS

Identifiée en Egypte au début des années 1990, la population *Salmonella enterica* sérotype Kentucky ST198−X1 a rapidement acquis au fil des années de nombreux marqueurs chromosomiques de la résistance aux antibiotiques, via le *Salmonella* Genomic Island 1 (SGI1) d'une part et, d'autre part, l'accumulation de mutations au niveau des gènes *gyrA* et *parC* lui conférant une résistance à l'acide nalidixique (Nal) puis à la ciprofloxacine (CIP) à partir de 2002. Depuis, cette souche s'est rapidement propagée en Afrique et au Moyen− Orient (81). En 2009, certains cas ont été décrits chez des voyageurs de retour d'Inde et dansdes élevages de volailles au Bengladesh (82, 83). Autre fait marquant, la concentration minimale inhibitrice (CMI) de la ciprofloxacine n'a cessé d'augmenter au cours du temps chez *S.* Kentucky (84). Enfin, le réservoir de ce clone ST198 est de plus en plus vaste : il fut d'abord identifié de manière autochtone chez la volaille en Afrique puis s'est répandu chez plusieurs animaux et dans certains produits alimentaires (81, 82, 83, 85, 86) dont la volaille d'élevage qui a largement contribué à sa dissémination dans les pays en voie de développement depuis 2005 (81) pour finalement atteindre les pays développés en 2010 (Pologne (86), Allemagne (87) et France (88)).

L'émergence d'un tel clone pose un problème majeur de santé publique et de nouveaux outils sont nécessaires pour contenir cette propagation. C'est dans cette optique que nous avons étudié le polymorphisme de la région CRISPR du sérotype Kentucky afin de développerune méthode de détection rapide par PCR de la population ST198 à l'intention des laboratoires d'hygiène et de microbiologie alimentaire et vétérinaire.

MATERIELS ET METHODES

Comme décrit précédemment, nous avons inventorié le contenu en spacers d'une collection de deux génomes publics et 54 souches isolées de 1937 à 2013 de *Salmonella enterica* sérotype Kentucky sélectionnées sur la base de la diversité des profils PFGE, MLST. La sensibilité aux antibiotiques, la CMI ciprofloxacine, la recherche de mutations *gyrA* et *parC* et la recherche du SGI1 ont également été caractérisés (**Annexes 1 et 2**). La sensibilité de la méthode PCR a été testée sur une nouvelle collection de 64 souches d'origines diversesisolées de 1959 à 2013 (**Tableau 3**). La spécificité de la méthode a été testée sur 20 souches de *Salmonella enterica* de sérotype autre que Kentucky (**Tableau 4**) et sur 36 souches représentatives de 32 espèces différentes (**Tableau 5**).

RESULTATS ET DISCUSSION

L'analyse du contenu en spacers des 56 souches et 2 génomes a mis en évidence 18 profils CRISPR1 (A1−A18), 15 profils CRISPR2 (B1−B15) et 22 profils combinés CRISPR1−CRISPR2. Aucune cible potentielle n'est apparue évidente sur le locus CRISPR1, nous avons donc analysé plus attentivement le locus CRISPR2. Les souches ST198 présentaient 10 profils CRISPR2 différents (B1, B3, B4, B5, B6, B7, B8, B9, B10, B11). La comparaison de ces profils a

mis en évidence la région située entre les spacers KenB55 et KenB38 comme cible de choix pour les différencier, ces deux spacers n'étant pas présents chez les profils B12, B13 et B14. Parmi ces souches, les ST198−X1−SGI1−CIP−R présentaient les profils B5, B6, B9, B10 et B11. Les souches de profil B1, B4, B8, B2, B3 et B7 étaient toutes sensibles à la ciprofloxacine et de profil PFGE différent de X1. L'amorce sens a ainsi été dessinée sur le spacer KenB55 et l'anti−sens sur le spacer KenB38 (**Figure 9**).

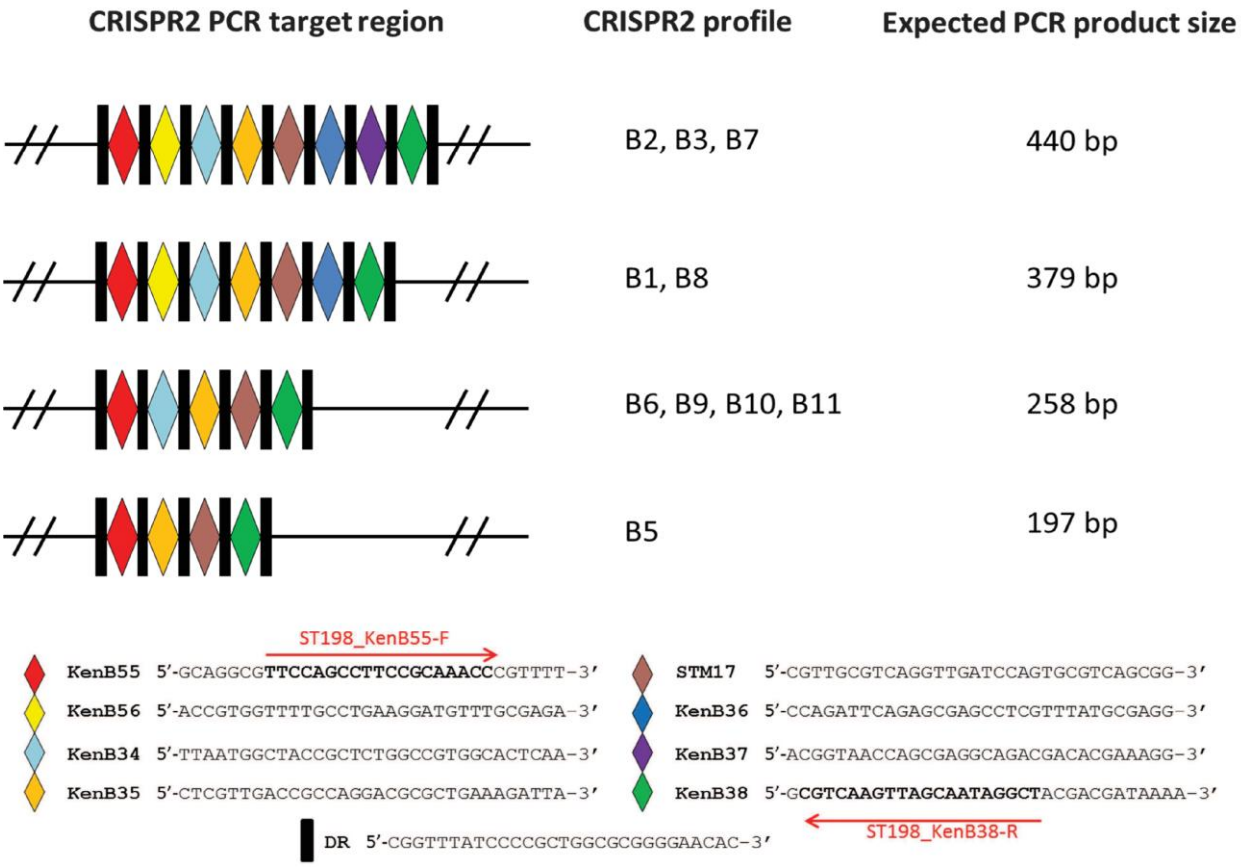


FIGURE 9 : REGION CIBLEE PAR PCR POUR LA DETECTION DU CLONE ST198. LES SPACERS SONT INDIQUES PAR DES LOSANGES DE COULEUR, LA REPETITION ASSOCIEE PAR UN RECTANGLE NOIR. LES SEQUENCES DES AMORCES ST198_KENB55-F ET ST198_KENB38-R SONT INDIQUEES EN GRAS DANS LA SEQUENCE DU SPACER CONCERNE.

Sur les 64 souches testées en PCR, 54 étaient positives dont 35 ont donné un amplicon de 258 pb, 11 de 440 pb, 8 de 379 pb, 2 de 200 pb. Les deux produits d'amplification de taille 200 pb ont été générés à partir d'échantillons ST198. De même, ceux de taille 258 pb et ceuxde 379 pb appartenaient tous au ST198. Ceux de 440 pb appartenaient tous au ST198 saufun de ST727. Ce dernier est un SLV (Single Locus Variant) du ST198, ce qui signifie qu'il appartient au même complexe clonal que les souches ST198. Les échantillons négatifs appartenaient à des séquençotypes plus éloignés du ST198 (ST152, ST314, ST393, ST696, ST723, ST728, ST724) (**Tableau 3**).

Toutes les souches dont le produit PCR était de 200 ou 258 pb possédaient le SGI−1, étaient de profil PFGE X1 et résistantes à la ciprofloxacine. Les échantillons ayant donné un produit PCR de 379 pb et 440 pb ne possédaient pas de SGI1 (sauf quatre qui arboraient le variant −J, associé au profil PFGE X2, (89)), de profil PFGE différent de X1 et sensibles à la ciprofloxacine. Les échantillons négatifs étaient sensibles aux antibiotiques, ne possédaient pas de SGI1 et étaient de profil PFGE différent de X1.

TABLEAU 3 : CARACTERISTIQUES DES SOUCHES DE *SALMONELLA ENTERICA* SEROTYPE KENTUCKY TESTEES PAR PCR

Sample	source	year	country	AST (*bla* genes) (1)	PFGE	MLST	SGI-1	PCR product size (bp)
2–59	human	1959	Tunisia	susceptible		ST728	-	-
1–63	human	1963	Vietnam	susceptible	X19	ST198	-	440
2–64	fish meal	1964	Portugal	susceptible		ST152	-	-
1–68	human	1968	Senegal	SSul	X1b	ST727	-	440
61–69	turkey meat	1969	France	Te	X2a	ST198	-	440
14–72	human	1972	Martinique	susceptible	X16	ST198	-	440
3–75	human	1975	Djibouti	susceptible	X11	ST723	-	-
8–76	superficial water	1976	Indonesia	Str		ST198	-	440
9–76	meat flour	1976	France	susceptible	X9	ST198	-	440
4–77	poultry meal	1977	France	Str		ST198	-	440
12–78	fish meal	1978	Turkey	Str	X15	ST198	-	440
18–78	human	1978	Iraq	Str	X17	ST198	-	440
2–79	human	1979	Guinea	susceptible	X9	ST198	-	440
6–79	human	1979	Burkina Faso	susceptible	X9	ST724	-	-
4–82	human	1982	Morocco	susceptible	X20	ST393	-	-
1–83	human	1983	Gabon	susceptible	X9	ST198	-	379
17–85	human	1985	Egypt	SSul	X14	ST198	-	379
1–86	human	1986	Vietnam	SSpSulCTe	X2b	ST198	Js	379
12–89	human	1989	Thaïland	SSpSulCTe	X2b	ST198	Js	379
10–90	human	1990	Indonesia	Sul	X2b	ST198	Js	379
92–6276	human	1992	Indonesia	SSpSulCTe	X2b	ST198	Js	379
93–6429	human	1993	Indonesia	Sul	X2c	ST198	J4	379
94–3543	human	1994	East Africa	susceptible	X1c	ST198	-	379
200102959	human	2001	Kenya	susceptible	X10	ST314	-	-
200208051	human	2002	Egypt	ASSpGSulTeNal	X1a	ST198	Ks	258
200208141	human	2002	Egypt	ASSpGSulCTeNal	X1m	ST198	K1	258
200208629	human	2002	Morocco	susceptible	X10	ST314	-	-
200309270	human	2003	India	Nal	X2d	ST198	-	440
200402049	human	2004	Egypt	NalCip	X1b	ST198	Qs	258
200407734	human	2004	Egypt	ASSpKGSulTmpCTeNal	X1h	ST198	K1	258
200408262	human	2004	Egypt	SSpGSulNalCip	X1a	ST198	K5	258
200500236	human	2005	Egypt	ANalCip	X1b	ST198	P	258
200500520	human	2005	Egypt	ANalCip	X1a	ST198	P2	258
200501199	human	2005	Egypt	SSpGSulTeNalCip	X1a	ST198	Q3	258
200502131	human	2005	Egypt	ANalCip	X1a	ST198	Q1	258
200503290	human	2005	Egypt	AKNalCip	X1a	ST198	K	258
200601060	human	2006	Kenya/Tanzania	ASSpGSulTeNalCip	X1c	ST198	K	258
200605737	human	2006	Tunisia	sensible	X5	ST696	-	-
200700033	turkey	2007	Morocco	ASSpSulTeNalCip	X1a	ST198	Ks	258
200800015	seafood	2008	Morocco	ASSpTGSulTeNalCip	X1d	ST198	Ks	258
200804979	human	2008	Cameroon	ANalCip	X1a	ST198	P	258
200805707	human	2008	Tanzania	ASSpGSulTeNalCip (TEM)	X1c	ST198	K	258
K_26	layer poultry farms	2009	Bangladesh	ASSpGSulTeNalCip (TEM)	X1e	ST198	Ks	258
K_50	layer poultry farms	2009	Bangladesh	ASSpGSulTeNalCip (TEM)	X1e	ST198	K	258
201002378	human	2010	Egypt	ASSpGSulCTeNalCip		ST198	K	200
201008911	human	2010	Egypt	ASSpGSulTeNalCip		ST198	K	200
201205586	poultry meat	2010	Côte d'Ivoire	ASSpGSulTeNalCip (TEM)		ST198	+	258
201205587	horse placenta	2010	France	ASSpGSulTmpTeNalCip (TEM)		ST198	+	258
201205588	river	2010	France	SSpGSulTeNalCip		ST198	+	258
201205585	compost	2010	France	ANalCip (TEM)		ST198	+	258
201007469	human	2010	Egypt	ATeNalCip		ST198	Q	258
201008929	human	2010	Djibouti	ASSpGSulTeNalCip		ST198	K	258
B_11	broiler poultry farms	2010	Bangladesh	ASSpTGSulTmpTeNalCip (TEM)	X1e	ST198	+	258
B_81	broiler poultry farms	2010	Bangladesh	ASSpGSulTeNalCip (TEM)	X1a	ST198	+	258
K_78	layer poultry farms	2010	Bangladesh	ASSpGSulTeNalCip (TEM)	X1i	ST198	+	258
11CEB3342	spice	2011	France	ANalCip (TEM–1)		ST198	+	258
11CEB4816	marinated turkey meat	2011	France	ASSpGSulTeNalCip (TEM–1)		ST198	+	258
12CEB4452	turkey farm	2012	France	ASSpGSulTeNalCip	X1f	ST198	Ks	258
12CEB716	dog	2012	France	ANalCip (TEM)		ST198	+	258
201203105	human	2012	Indonesia	ASSpGSulTeNalCip (TEM)	X1a	ST198	Ks	258
201205363	human	2012	Kuwait	ASSpGTeNalCip (TEM)	X1l	ST198	Ks	258
201207374	human	2012	Vietnam	ASulTeNalCip (TEM)	X1c	ST198	Ks	258
13CEB2160	turkey meat	2013	Poland	ASSpGSulTeNalCip (TEM)	X1x	ST198	Ks	258
201301062	human	2013	Algeria	A–ImpSSpGSulTmpTeNalCip (TEM–1 + OXA–48)	X1b	ST198	Ks	258

(1) A: ampicillin; S: streptomycin; Sp: spectinomycin; G: sulfametoxazole; Sul: sulfonamides; Te: tetracyclin; Nal: nalidixic acid; Cip: ciprofloxacin

Les 20 échantillons appartenant à d'autres sérotypes que Kentucky étaient tous négatifs, à l'exception de la souche 4/75 de sérotype Poeseldorf (8,20,54 :i :z6) (ST198) pour laquelle nous avons obtenu un amplicon de 440 pb (**Tableau 4**). Ce résultat n'est cependant pas surprenant et ne remet pas en cause la spécificité du protocole puisque ce sérotype est un variant O:54 de Kentucky, l'opéron codant l'antigène O:54 étant porté par un plasmide (90). Les 36 échantillons appartenant à d'autres espèces bactériennes étaient tous négatifs (**Tableau 5**).

TABLEAU 4 : RESULTATS DES TESTS PCR SUR SEROTYPES AUTRES QUE *SALMONELLA ENTERICA* KENTUCKY

STRAIN	SUBSPECIES	SEROTYPE	ST	PCR (bp)
LT2	*enterica*	Typhimurium	ST19	-
2−74	*enterica*	Hadar	ST33	-
50K	*enterica*	Newport	ST31	-
1K	*enterica*	Paratyphi A	ST85	-
53K	*enterica*	Bovismorbificans	ST377	-
158K	*enterica*	Infantis	ST493	-
173K	*enterica*	Napoli	ST765	-
65K	*enterica*	Dublin	ST10	-
343K	*enterica*	Albany	ST292	-
46K	*enterica*	Montevideo	ST138	-
20K	*enterica*	Derby	ST40	-
1137/72	*enterica*	Agona	ST13	-
41K	*enterica*	Virchow	ST755	-
Ty2	*enterica*	Typhi	ST1	-
17K	*enterica*	Chester	ST411	-
4/75	*enterica*	Poeseldorf	ST198	440
553/69	*enterica*	Istanbul	ST770	-
891K	*enterica*	Altona	ST731	-
263K	*enterica*	Corvallis	ST1357	-
747/70	*enterica*	Kedougou	ST1543	-
201308345	*enterica*	Enteritidis	ST11	-

TABLEAU 5 : AUTRES ESPECES BACTERIENNES TESTEES PAR PCR

Other species	Number of strains tested	Strain	Source (1)
Escherichia coli	5	CNR2011	CNR-ESS
Escherichia blattae	1	CDC 9005-74	CIP
Shigella dysenteriae 1	1	NCDC 1007-71	CIP
Shigella boydii 1	1	11-1445	CNR-ESS
Shigella dysenteriae 2	1	11-1663	CNR-ESS
Shigella sonnei g	1	11-2511	CNR-ESS
Shigella flexneri 2a	1	11-2500	CNR-ESS
Enterobacter aerogenes	1	CIP 60-85	CIP
Enterobacter kobei	1	CIP 105 566	CIP
Enterobacter gergoviae	1	CIP 76.1	CIP
Enterobacter cloacae subsp. *cloacae*	1	CIP 60-85	CIP
Citrobacter murliniae	1	CIP 104556	CIP
Citrobacter sedlakii	1	CDC 4696-86	CIP
Citrobacter braakii	1	CDC 80-58	CIP
Edwardsiella tarda	1	NCTC 10396	CIP
Kluyvera cryocrescens	1	CDC 2065-78	CIP
Pantoea agglomerans	1	CIP 57-51	CIP
Serratia marcescens subsp. *marcescens*	1	CIP 103235	CIP
Leminorella grimontii	1	CDC 1944-81	CIP
Buttauxiella agrestis	1	CIP 80.31	CIP
Yersinia pseudotuberculosis subsp. *pseudotuberculosis*	1	ATCC 29833	CIP
Providencia rettgeri	1	ATCC 29944	CIP
Dickeya chrysanthemi	1	CIP 82.99	CIP
Klebsiella pneumoniae subsp. *ozenae*	1	CIP 52.211	CIP
Morganella morganii subsp. *morganii*	1	ATCC 25830	CIP
Proteus vulgaris	1	ATCC 29833	CIP
Streptococcus pyogenes	1	CIP 56.41	CIP
Staphylococcus aureus subsp. *Aureus*	1	CIP 65.8	CIP
Staphylococcus epidermidis	1	CIP 81.55	CIP
Streptococcus agalactiae	1	CIP 103227	CIP
Streptococcus pneumoniae	1	CIP 103566	CIP
Nesseria meningitidis	1		CNR-*Nesseria*

(1) CNR-ESS: Centre National de référence *Escherichio coli-Shigella-Salmonella* ; CIP : Collection de l'Institut Pasteur; CNR-Nesseria: Centre National de Référence des *Nesserio*

CONCLUSIONS

Cette étude a permis la mise au point d'une PCR spécifique du clone émergent S. Kentucky ST198 basée sur le polymorphisme des régions CRISPR, renforçant son utilité dans la mise au point d'outils de détection de populations particulières de *Salmonella*. La taille de l'amplicon obtenu permet en outre de différencier la population X1−SGI1−CIP−R. La surveillance de ce clone n'est pas règlementée, sauf en France depuis 2015 (arrêté du 17 février 2015 relatif à la définition des dangers sanitaires de première et deuxième catégorie

pour les volailles, https://www.legifrance.gouv.fr/eli/arrete/2015/2/17/AGRG1504715A/jo/texte), réglementation basée uniquement sur le sérotype. La méthode de détection décrite ici serait un argument de poids pour faire évoluer la règlementation vers la détection du clone ST198 en particulier.

CONCLUSIONS–PERSPECTIVES

Notre étude préliminaire (80) a mis en évidence l'intérêt de l'étude du polymorphisme de la région CRISPR pour le typage et le sous−typage rapide du genre *Salmonella*. L'étude plus approfondie de la région a permis de mettre en relief le potentiel de la méthode à se substituer aux méthodes de référence que sont le sérotypage et le PFGE. Les travaux décrit dans cette thèse ont montré l'impact de cette méthode sur la surveillance épidémiologique des salmonelloses par l'identification de certaines particularités de clones épidémiques (Publication 3, 91), les rendant ainsi plus facilement traçables mais aussi par le développement de nouveaux outils d'identification rapide des certains sérotypes majeurs (Typhi et Paratyphi A) et de sous−typage haut−débit du sérotype prévalent Typhimurium et son variant monophasique. D'autres méthodes basées sur l'analyse du polymorphisme des loci CRISPR ont été proposées, comme la combinaison du profil CRISPR associé à certains gènes de virulence (*fimH* et *sseL*) (92, 93) mais n'offrent pas de meilleure discrimination ou de possibilité d'automatisation.

Ces résultats nous ont permis d'envisager une automatisation de la méthode. Le problème cependant est le nombre important de spacers identifiés. Au moment de ces travaux, plus de 3 800 séquences ont été répertoriées parmi 134 sérotypes. Depuis, nous avons complété ces résultats à 280 sérotypes et ajouté plus de 3 000 spacers à notre inventaire. La technologie Luminex® utilisée pour CRISPOL Typhimurium n'est pas envisageable car limitée en nombre de sondes multiplexables. Seule une puce à ADN aurait pu être adaptée. Cependant, avec la démocratisation rapide des nouvelles techniques de séquençage (NGS−Next Generation Sequencing), le séquençage complet (WGS−Whole Genome Sequencing) est de plus en plus utilisé pour l'étude des salmonelles. De nouvelles bases de données ont ainsi été développées et associées à des outils bioinformatiques permettant d'extraire toutes les informations nécessaires au typage des salmonelles comme l'outil en ligne fourni par le

Center of Genomic Epidemiology (Danemark) (http://genomicepidemiology.org/) pour l'extraction des gènes de résistance aux antibiotiques, MLST, plasmides, pMLST, ou encore la migration de la base de données publique MLST vers Enterobase (http://enterobase.warwick.ac.uk/) qui permet en outre de soumettre un génome complet et de le comparer à d'autres présents dans la base de données et relier ainsi certaines souches épidémiques. Notre inventaire de spacers a été implémenté à ce dernier outil pour l'extraction des profils CRISPR à partir de données NGS. De nouveaux schémas sont également développés sur la base de l'analyse du core genome (cgMLST) comme définis dans l'outil *Salmonella In Silico* Typing Ressource (SISTR) (94) ou encore Enterobase. L'approche WGS comporte cependant certaines limites concernant le typage CRISPR. En effet, pour un séquençage short reads (2x150 pb), l'assemblage des loci CRISPR reste difficileen raison des répétitions, ce qui peut fausser le résultat avec par exemple une duplication artificielle de spacers. Une approche par mapping des reads bruts sur un dictionnaire de séquences de spacers est alors plus appropriée. Celle−ci est en cours de développement au sein de l'unité et sera implémentée dans notre nouvelle version de l'outil web pour la recherche de spacers dans une séquence de *Salmonella*.

BIBLIOGRAPHIE

Ces références sont citées par leur ordre d'apparition dans le texte.

1 Le Minor L. The Genus *Salmonella* in The Prokaryotes Vol., II. Chp. 92. 1981 (Starr M.P., Stolp H., Trüper H.G., Balows A., Schlegel H.G., Eds)

2 Widal, F., Lemierre, A. & Abrami, P. (1921): Fièvres typhoïdes et para−typhoïdes. In: Roger, G. H., Widal, F. & Teissier, P. J.: Nouveau Traité de Médecine, fasc. 3, Paris.

3 Weil, E. & Felix, A. (1917) Wien. Klin. Wschr. 30, 1509, cited in Smith, R.W. & Koffler, H., Bacterial Flagella, In Advances in Microbial Physiology, Vol. 6 (A.H. Rose & J.F. Wilkinson, Eds.), p. 251, Academic Press, 1971

4 Judicial Commission of the International Committee on Systematics of Prokaryotes. The type species of the genus *Salmonella* Lignères 1900 is *Salmonella enterica* (ex Kauffmann and Edwards 1952) Le Minor and Popoff 1987, with the type strain LT2T, and conservation of the epithet enterica in *Salmonella enterica* over all epithets that may be applied to this species. Opinion 80. *Int J Syst Evol Microbiol 2005;*55:519−20.

5 Grimont P.A. D, Weill F.−X. Antigenic formulae of the *Salmonella* serovars, 9th ed. Paris, France: WHO Collaborating Center for Reference and Research on *Salmonella*, Institut Pasteur, 2007. Available at : http://www.pasteur.fr/fr/sante/centrescollaborateurs− l−oms−ccoms/salmonella

6 Vought, K. J. and S. R. Tatini (1998). "*Salmonella* Enteritidis contamination of ice cream associated with a 1994 multistate outbreak." J Food Prot 61(1): 5−10

7 Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. *Bull World Health Organ 2004*;82:346−53

8 Weill FX, Le Hello S. Centre National de Référence des *Salmonella*. Rapports d'activités annuels 2011−2014. Paris: Institut Pasteur; Disponible à http://www.pasteur.fr/cnr/salmonella

9 Issenhuth−Jeanjean S, Roggentin P, Mikoleit M, Guibourdenche M, de Pinna E, Nair S, Fields PI, Weill F.−X. Supplement 2008−2010 (no. 48) to the White−Kauffmann−Le Minor scheme. *Res Microbiol 2014*;165(7):526−30

10 Winokur PL. Molecular epidemiological techniques for *Salmonella* strain discrimination. Front Biosci. 2003 Jan 1;8:c14−24. Review

11 Shi C, Singh P, Ranieri ML, Wiedmann M, Moreno Switt AI. Molecular methods for serovar determination of *Salmonella*. Crit Rev Microbiol. 2015;41(3):309−25. doi: 10.3109/1040841X.2013.837862. Review

12 Luk JM, Kongmuang U, Reeves PR, Lindberg AA. Selective amplification of abequose and paratose synthase genes (*rfb*) by polymerase chain reaction for identification of *Salmonella* major serogroups (A, B, C2, and D). J ClinMicrobiol. 1993 Aug;31(8):2118−23

13 Mortimer CK, Peters TM, Gharbia SE, Logan JM, Arnold C. Towards the development of a DNA−sequence based approach to serotyping of *Salmonella enterica*. BMC Microbiol. 2004 Aug 6;4:31

14 Jiang XM, Neal B, Santiago F, Lee SJ, Romana LK, Reeves PR. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium(strain LT2). Mol Microbiol. 1991 Mar;5(3):695−713

15 Smith NH, Selander RK. Sequence invariance of the antigen−coding central region of the phase 1 flagellar filament gene (*fliC*) among strains of *Salmonella* typhimurium. J Bacteriol. 1990 Feb;172(2):603−9

16 Vanegas RA, Joys TM. Molecular analyses of the phase−2 antigen complex 1,2 of *Salmonella* spp. J Bacteriol. 1995 Jul;177(13):3863−4

17 Verma N, Reeves P. Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D salmonellae. J Bacteriol. 1989 Oct;171(10):5694−701

18 Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P. Multiplex, bead−based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol. 2007 Oct;45(10):3323−34

19 Fitzgerald C, Sherwood R, Gheesling LL, Brenner FW, Fields PI. Molecular analysis of the *rfb* O antigen gene cluster of *Salmonella enterica* serogroup O:6,14 and development of a serogroup−specific PCR assay. Appl Environ Microbiol.2003 Oct;69(10):6099−105

20 Garaizar J, Porwollik S, Echeita A, Rementeria A, Herrera S, Wong RM, Frye J, Usera MA, McClelland M. DNA microarray−based typing of an atypical monophasic *Salmonella enterica* serovar. J Clin Microbiol. 2002 Jun;40(6):2074−8

21 Herrera−León S, Ramiro R, Arroyo M, Díez R, Usera MA, Echeita MA. Blind comparison of traditional serotyping with three multiplex PCRs for the identification of *Salmonella* serotypes. Res Microbiol. 2007 Mar;158(2):122−7

22 Felix A, Callow BR. Typing of Paratyphoid B Bacilli by Vi Bacteriophage. Br Med J. 1943 Jul 31;2(4308):127−30

23a Esteban E, Snipes K, Hird D, Kasten R, Kinde H. Use of ribotyping for characterization of *Salmonella* serotypes. J Clin Microbiol. 1993 Feb;31(2):233−7

23b Lam S, Roth JR. IS*200*: a *Salmonella*−specific insertion sequence. Cell. 1983 Oct;34(3):951−960

24 Weill FX, Fabre L, Grandry B, Grimont PA, Casin I. Multiple−antibiotic resistance in *Salmonella enterica* serotype Paratyphi B isolates collected in France between 2000 and 2003 is due mainly to strains harboring *Salmonella* genomic islands 1, 1−B, and 1−C. Antimicrob Agents Chemother. 2005 Jul;49(7):2793−801

25 Weill FX, Tran HH, Roumagnac P, Fabre L, Minh NB, Stavnes TL, Lassen J, Bjune G, Grimont PA, Guerin PJ. Clonal reconquest of antibiotic−susceptible *Salmonella enterica* serotype Typhi in Son La Province, Vietnam. Am J Trop Med Hyg. 2007 Jun;76(6):1174−81

26 Hedberg CW, Greenblatt JF, Matyas BT, Lemmings J, Sharp DJ, et al. Timeliness of enteric disease surveillance in 6 US states. Emerg Infect Dis. 2008;14:311−313

27 Lindstedt BA, 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. 2004;59(2):163−72

28 Larsson JT, Torpdahl M, Petersen RF, Sorensen G, Lindstedt BA, Nielsen EM. Development of a new nomenclature for *Salmonella* typhimurium multilocus variable number of tandem repeats analysis (MLVA). Euro Surveill. 2009;14(15):pii=19174

29 Hopkins KL, Peters TM, de Pinna E, Wain J. Standardisation of multilocus variable−number tandem−repeat analysis (MLVA) for subtyping of *Salmonella enterica* serovar Enteritidis. Euro Surveill. 2011;16(32). pii=19942

30 Lindstedt BA, Torpdahl M, Vergnaud G, Le Hello S, Weill FX, Tietze E, Malorny B, Prendergast DM, Ní Ghallchóir E, Lista RF, Schouls LM, Söderlund R, Börjesson S, Åkerström S. Use of multilocus variable−number tandem repeat analysis (MLVA) in eight European countries, 2012. Euro Surveill. 2013;18(4):pii=2038

31 Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998 Mar 17;95(6):3140−5

32 Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl Environ Microbiol. 1986 May;51(5):873−84. Review

33 Kidgell C, Reichard U, Wain J, Linz B, Torpdahl M, et al. *Salmonella* typhi, the causative agent of typhoid fever, is approximately 50,000 years old. Infect Genet Evol. 2002;2:39−45

34 Torpdahl M, Skov MN, Sandvang D, Baggesen DL. Genotypic characterization of *Salmonella* by multilocus sequence typing, pulsed−field gel electrophoresis and amplified fragment length polymorphism. J Microbiol Methods. 2005

35 Harbottle H, White DG, McDermott PF, Walker RD, Zhao S. Comparison of multilocus sequence typing, pulsed−field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J Clin Microbiol. 2006;44:2449−2457

36 Wiesner M, Zaidi MB, Calva E, Fernandez−Mora M, Calva JJ, et al. Association of virulence plasmid and antibiotic resistance determinants with chromosomal multilocus genotypes in Mexican *Salmonella enterica* serovar Typhimurium strains. BMC Microbiol. 2009;9:131

37 Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH, Brisse S; *S. enterica* MLST Study Group.. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog. 2012;8(6):e1002776. doi: 10.1371/journal.ppat.1002776

38 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. 1987 Dec;169(12):5429−33

39 Nakata A, Amemura M, Makino K. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K−12 chromosome. J Bacteriol. 1989 Jun;171(6):3553−6

40 Jansen R, van Embden JD, Gaastra W, Schouls LM. Identification of a novel family of sequence repeats among prokaryotes. OMICS. 2002;6(1):23−33

41 Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol. 2002 Mar;43(6):1565−75

42 Horvath P, Romero DA, Coûté−Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol. 2008 Feb;190(4):1401−12

43 Godde JS, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol. 2006 Jun;62(6):718− 29

44 Deveau H, Garneau JE, Moineau S. CRISPR/Cas system and its role in phage−bacteria interactions. Annu Rev Microbiol. 2010;64:475−93. doi: 10.1146/annurev.micro.112408.134123

45 Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. 2007;8(4):R61

46 Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics. 2007 May 23;8:172.

47 Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010 Jan 8;327(5962):167−70. doi: 10.1126/science.1179555

48 Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005 Aug;151(Pt 8):2551−61

49 Lillestøl RK, Redder P, Garrett RA, Brügger K. A putative viral defence mechanism in archaeal cells. Archaea. 2006 Aug;2(1):59−72

50 Mojica FJ, Díez−Villaseñor C, García−Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol. 2005 Feb;60(2):174−82

51 Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire newrepeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 2005 Mar;151(Pt 3):653−63

52 Greve B, Jensen S, Brügger K, Zillig W, Garrett RA. Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. Archaea. 2004 Oct;1(4):231−9

53 Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño−Tárraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabbinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. The multidrug−resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet. 2006 Jul;38(7):779−86

54 Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science. 1996 Aug 23;273(5278):1058−73

55 Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Peterson S, Reich CI, McNeil LK, Badger JH, Glodek A, Zhou L, Overbeek R, Gocayne JD, Weidman JF, McDonald L, Utterback T, Cotton MD, Spriggs T, Artiach P, Kaine BP, Sykes SM, Sadow PW, D'Andrea KP, Bowman C, Fujii C, Garland SA, Mason TM, Olsen GJ, Fraser CM, Smith HO, Woese CR, Venter JC. The complete genome sequence of the hyperthermophilic, sulphate−reducing archaeon *Archaeoglobus fulgidus*. Nature. 1997 Nov 27;390(6658):364−70

56 Smith DR, Doucette−Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, Harrison D, Hoang L, Keagle P, Lumm W, Pothier B, Qiu D,

Spadafora R, Vicaire R, Wang Y, Wierzbowski J, Gibson R, Jiwani N, Caruso A, Bush D, Reeve JN, et al. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. J Bacteriol. 1997 Nov;179(22):7135−55

57 Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007 Mar 23;315(5819):1709−12

58 Tyson GW, Banfield JF. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol. 2008 Jan;10(1):200−7

59 Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008 Aug 15;321(5891):960−4. doi: 10.1126/science.1159689

60 Hale C, Kleppe K, Terns RM, Terns MP. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. RNA. 2008 Dec;14(12):2572−9. doi: 10.1261/rna.1246808

61 Marraffini LA, Sontheimer EJ. CRISPR interference: RNA−directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 2010 Mar;11(3):181−90. doi:10.1038/nrg2749. Review

62 Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, GarrettRA. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Mol Microbiol. 2009 Apr;72(1):259−72. doi: 10.1111/j.1365−2958.2009.06641.x

63 Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR−associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol. 2005 Nov;1(6):e60

64 Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA−interference−based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct. 2006 Mar 16;1:7

65 Sorek R, Kunin V, Hugenholtz P. CRISPR−−a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol.2008 Mar;6(3):181−6.

66 Touchon M, Rocha EP. The small, slow and specialized CRISPR and anti−CRISPR of *Escherichia* and *Salmonella*. PLoS One. 2010 Jun 15;5(6):e11126. doi:10.1371/journal.pone.0011126

67 Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P,Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. Evolution and classification of the CRISPR−Cas systems. Nat Rev Microbiol. 2011Jun;9(6):467−77. doi: 10.1038/nrmicro2577

68 Makarova KS, Aravind L, Wolf YI, Koonin EV. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR−Cas systems. Biol Direct. 2011 Jul 14;6:38. doi: 10.1186/1745−6150−6−38

69 Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008 Dec19;322(5909):1843−5. doi: 10.1126/science.1165771

70 van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. CRISPR−based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci. 2009Aug;34(8):401−7. doi: 10.1016/j.tibs.2009.05.002. Review

71 Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D. Germ warfare in a microbial mat community: CRISPRs provide insights into the co−evolution of host and viralgenomes. PLoS One. 2009;4(1):e4169. doi: 10.1371/journal.pone.0004169

72 Held NL, Herrera A, Cadillo−Quiroz H, Whitaker RJ. CRISPR associated diversity within a population of *Sulfolobus islandicus*. PLoS One. 2010 Sep 28;5(9). pii:e12988. doi: 10.1371/journal.pone.0012988

73 Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res.2007 Jul;35 (Web Server issue):W52−7

74 Rousseau C, Gonnet M, Le Romancer M, Nicolas J. CRISPI: a CRISPR interactive database. Bioinformatics. 2009 Dec 15;25(24):3317−8. doi:10.1093/bioinformatics/btp586

75 Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol Microbiol.1993 Dec;10(5):1057−65

76 Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol. 1997 Apr;35(4):907−14

77 Mokrousov I, Limeschenko E, Vyazovaya A, Narvskaya O. *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. Biotechnol J.2007 Jul;2(7):901−6

78 Vergnaud G, Li Y, Gorgé O, Cui Y, Song Y, Zhou D, Grissa I, Dentovskaya SV, Platonov ME, Rakin A, Balakhonov SV, Neubauer H, Pourcel C, Anisimov AP, Yang R. Analysis of the three

Yersinia pestis CRISPR loci provides new tools for phylogenetic studies and possibly for the investigation of ancient DNA. Adv ExpMed Biol. 2007;603:327−38

79 Schouls LM, Reulen S, Duim B, Wagenaar JA, Willems RJ, Dingle KE, Colles FM, Van Embden JD. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeats sequencing: strain diversity, host range, and recombination. J Clin Microbiol.2003 Jan;41(1):15−26

80 Weill FX, Fabre−Berland L, Guibert V, Diancourt L, Brisse S. French Patent Application (no. FR07/09188) filed on December 28, 2007. International Patent Application (no. PCT/IB2008/004004) filed on Dec. 29, 2008; 2007. Molecular typing and subtyping of *Salmonella* by identification of the variable nucleotide sequences of the CRISPR loci

81 Le Hello S, Hendriksen RS, Doublet B, Fisher I, Nielsen EM, Whichard JM, Bouchrif B, Fashae K, Granier SA, Jourdan−Da Silva N, Cloeckaert A, Threlfall EJ, Angulo FJ, Aarestrup FM,Wain J, Weill FX. International spread of an epidemic population of *Salmonella enterica* serotype Kentucky ST198 resistant to ciprofloxacin. J Infect Dis. 2011 Sep 1;204(5):675−84. doi: 10.1093/infdis/jir409

82 Barua H, Biswas PK, Olsen KE, Christensen JP. Prevalence and characterization of motile *Salmonella* in commercial layer poultry farms in Bangladesh. PLoS One. 2012;7(4):e35914. doi: 10.1371/journal.pone.0035914

83 Barua H, Biswas PK, Olsen KE, Shil SK, Christensen JP. Molecular characterization of motile serovars of *Salmonella enterica* from breeder and commercial broiler poultry farms in Bangladesh. PLoS One. 2013;8(3):e57811. doi: 10.1371/journal.pone.0057811

84 Le Hello S, Bekhit A, Granier SA, Barua H, Beutlich J, Zając M, Münch S, Sintchenko V, Bouchrif B, Fashae K, Pinsard JL, Sontag L, Fabre L, Garnier M, Guibert V, Howard P,

Hendriksen RS, Christensen JP, Biswas PK, Cloeckaert A, Rabsch W, Wasyl D, Doublet B, Weill FX. The global establishment of a highly−fluoroquinolone resistant *Salmonella enterica* serotype Kentucky ST198 strain. Front Microbiol. 2013 Dec 18;4:395. doi: 10.3389/fmicb.2013.00395

85 Münch S., Braun P., Wernery U., Kinne J., Pees M., Flieger A., et al. (2012). Prevalence, serovars, phage types, and antibiotic susceptibilities of *Salmonella* strains isolated from animals in the United Arab Emirates from 1996 to 2009. Trop. Anim. Health Prod. 44, 1725−1738 10.1007/s11250−012−0130−4

86 Wasyl D., Hoszowski A. (2012). First isolation of ESBL−producing *Salmonella* and emergence of multiresistant *Salmonella* Kentucky in turkey in Poland. Food Res. Int. 45, 958−961 10.1016/j.foodres.2011.07.024

87 Beutlich J., Guerra B., Schroeter A., Arvand M., Szabo I., Helmuth R. (2012). [Highly ciprofloxacin resistant *Salmonella enterica* serovar Kentucky isolates in turkey meat and a human patient]. Berl. Munch. Tierarztl. Wochenschr. 125, 89−95 10.2376/0005−9366−125−89

88 Guillon F., Chasset P., Le Hello S., Granier S. A. (2013). Epidemiological investigation of highly ciprofloxacin resistant *Salmonella* Kentucky detected for the first time in French avian production. Bull. Épidémiol. 57, 22−23 Available online at: http://www.ansespro.fr/bulletin−epidemiologique/Documents/BEP−mg−BE57.pdf

89 Le Hello S, Weill FX, Guibert V, Praud K, Cloeckaert A, Doublet B. Earlystrains of multidrug−resistant *Salmonella enterica* serovar Kentucky sequence type198 from Southeast Asia harbor *Salmonella* genomic island 1−J variants with a novel insertion sequence. Antimicrob Agents Chemother. 2012 Oct;56(10):5096−102. doi: 10.1128/AAC.00732−12

90 Rohde R. [On the possibility of conversion in relation to S. poeseldorf=(8),20,54:i:z6, a new species of *Salmonella* of the subgenus I]. ArchHyg Bakteriol. 1968 Jan;152(1):79−81

91 Le Hello S, Maillard F, Mallet HP, Daudens E, Levy M, Roy V, Branaa P, Bertrand S, Fabre L, Weill FX. *Salmonella enterica* serotype Enteritidis in French Polynesia, South Pacific, 2008−2013. Emerg Infect Dis. 2015 Jun; 21(6):1045−8. doi: 10.3201/eid2106.141103

92 Liu F, Barrangou R, Gerner−Smidt P, Ribot EM, Knabel SJ, Dudley EG. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. Appl Environ Microbiol. 2011 Mar;77(6):1946−56. doi: 10.1128/AEM.02625−10

93 Li H, Li P, Xie J, Yi S, Yang C, Wang J, Sun J, Liu N, Wang X, Wu Z, Wang L, Hao R, Wang Y, Jia L, Li K, Qiu S, Song H. New clustered regularly interspaced short palindromic repeat locus spacer pair typing method based on the newly incorporated spacer for *Salmonella enterica*. J Clin Microbiol. 2014 Aug;52(8):2955−62. doi: 10.1128/JCM.00696−14

94 Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VPJ, Nash JHE, Taboada EN (2016) The *Salmonella In Silico* Typing Resource (SISTR): an open web−accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. PLoS ONE 11(1): e0147101. http://dx.doi.org/10.1371/journal.pone.0147101

ANNEXES

ANNEXE 1: PROFILS CRISPR1 DE DE *SALMONELLA ENTERICA* SEROTYPE KENTUCKY (PUBLICATION 3)

Isolate	Serotype	Profil CRISPR 1 - A	SPACER CONTENT CRISPR 1
00-1059	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
01-2100	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
02-9866	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
04-4567	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
05-1016	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
05-4680	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
06-5737	Kentucky	A16	Ken27-Ent16-Ent17-Ent18-Ken28-Ken29-Ken30-Ken31-Ken32-Ken33-Ken34-Ken35-Ken36
07-1511	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
07-6574	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
07-7991	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
08-4705	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
08-5707	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
08-KS6	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
09-8381	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
09-8745	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
09-9322	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201000305	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201000720	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
1090-10	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201005456	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201007071	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201007503	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201008553	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201009778	Kentucky	A8	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22
201100664	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201101683	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201101973	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201106348	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
201007297	Kentucky	A10	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken44-Ken45-Ken46-Ken47-Ken48-Ken49-Ken21-Ken22-Ken23-Ken24-Ken25-Ken26
1-61	Kentucky	A3	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken43-Ken44-Ken45-Ken46-Ken47-Ken48-Ken49-Ken56-Ken57-Ken21-Ken22-Ken23-Ken24-Ken25-Ken26
1-66	Kentucky	A2	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
1-68	Kentucky	A7	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken47-Ken48-Ken49-Ken56-Ken57-Ken21-Ken22-Ken23-Ken24-Ken25-Ken26
19-85	Kentucky	A6	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken47-Ken48-Ken49-Ken21-Ken22-Ken23-Ken24-Ken25-Ken26
2-75	Kentucky	A12	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken47-Ken48-Ken49-Ken56-Ken57-Ken21-Ken22-Ken23-Ken24-Ken25-Ken26
3-59	Kentucky	A14	Ken1-Ken2-Had4-Had7-Ken3-Ken4-Ken5-Ken6-Ken7-Ken8-Ken9-Ken10-Ken11-Ken12-Ken58-Ken59-Ken13-Ken14-Stp12-Ken15-Ken16
4-90	Kentucky	A5	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44
5-76	Kentucky	A11	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38var1-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50
07F4403	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
80-11-227-5139	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
80-11-252-4482	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
80-11-309-2385	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
93-6429	Kentucky	A6	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken47-Ken48-Ken49-Ken21-Ken22-Ken23-Ken24-Ken25-Ken26
96-11313	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
97-11473	Kentucky	A4	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55
97-6819	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
98K	Kentucky	A6	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken47-Ken48-Ken49-Ken21-Ken22-Ken23-Ken24-Ken25-Ken26
99-2998	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
BfR05-4625	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
BfR10-2164	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
BfR10-2979	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
CDC191*	Kentucky	A13	Ken1-Ken2-Had4-Had7-Ken3-Ken4-Ken5-Ken6-Ken7-Ken8-Ken9-Ken10-Ken11-Ken12-Ken13-Ken14-Stp12-Ken15-Ken16
CVM29188*	Kentucky	A15	Ken1-Ken2-Had4-Had7-Ken3-Ken5-Ken6-Ken7-Ken8-Ken9-Ken10-Ken11-Ken12-Ken13-Ken14-Stp12-Ken15-Ken16
Em06-2339	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
Em07-4654	Kentucky	A9	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken25-Ken26
10-2378	Kentucky	A17	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Newp8-Ken40-Ken41-Chin1var1-Ken42-Ken50-Ken51-Ken52-Ken53-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken26
10-8911	Kentucky	A18	Alt1-Conc15-Ana3-Ber1-Ken17-Ken18-Ken19-Ken20-Ken37-Ken38-Ken39-Ken54-Ken55-Ken43-Ken44-Ken45-Ken46-Ken56-Ken57-Ken21-Ken22-Ken26
(*) génomes publiques			

ANNEXE 2 : PROFILS CRISPR2 DE *SALMONELLA ENTERICA* SEROTYPE KENTUCKY (PUBLICATION 3)

Isolate	Serotype	Profil CRISPR 2 - B	SPACER CONTENT CRISPR 2
00-1059	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
01-2100	Kentucky	B9	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB47-KenB48-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
02-9866	Kentucky	B10	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
04-4567	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
05-1016	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
05-4680	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
06-5737	Kentucky	B12	EntB39-IndB1-KenB1-KenB2-KenB3-KenB4-KenB5-KenB6-KenB7-KenB16-KenB17-KenB18-KenB19-KenB57-KenB58-KenB59-KenB20-KenB21-KenB22-KenB23-KenB24-KenB25-KenB10-KenB26-KenB27-KenB28
07-1511	Kentucky	B9	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB47-KenB48-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
07-6574	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
07-7991	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
08-4705	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1KenB41
08-5707	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1KenB41
08-KS6	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
09-8381	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
09-8745	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
09-9322	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201000305	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1KenB41
201000720	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
1090-10	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201005456	Kentucky	B11	EntB0-MonB22-MonB24-NiaB2-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201007071	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201007503	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201008553	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201009778	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201100664	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201101683	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201101973	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201106348	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
201007297	Kentucky	B4	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
1-61	Kentucky	B4	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
1-66	Kentucky	B3	EntB0-MonB22-MonB24-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB56-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
1-68	Kentucky	B2	EntB0-MonB22-MonB24-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB56-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
19-85	Kentucky	B1	EntB0-MonB22-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB56-KenB34-KenB35-STM17-KenB36-KenB38-KenB39-Ago27var1-KenB41
2-75	Kentucky	B4	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
3-59	Kentucky	B13	EntB39-IndB1-KenB1-KenB2-KenB3-KenB4-KenB5-KenB6-KenB7-KenB8-KenB62-KenB9-KenB9-KenB10-KenB11-NewpB2var1-KenB12-KenB13-Ent5-KenB14-KenB15
4-90	Kentucky	B7	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB56-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
5-76	Kentucky	B7	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB56-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
07F4403	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
80-11-227-5139	Kentucky	B10	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
80-11-252-4482	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
80-11-309-2385	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
93-6429	Kentucky	B8	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB56-KenB34-KenB35-STM17-KenB36-KenB38-KenB39-Ago27var1-KenB41
96-11313	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
97-11473	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
97-6819	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
98K	Kentucky	B7	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB56-KenB34-KenB35-STM17-KenB36-KenB37-KenB38-KenB39-Ago27var1-KenB41
99-2998	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
BfR05-4625	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
BfR10-2164	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
BfR10-2979	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
CDC191*	Kentucky	B14	IndB1-KenB1-KenB2-KenB3-KenB4-KenB5-KenB6-KenB7-KenB8-KenB9-KenB10-KenB11-NewpB2var1-KenB12-KenB13-Ent5-KenB14-KenB15
CVM29188*	Kentucky	B15	IndB1-KenB1-KenB3-KenB4-KenB5-KenB6-KenB7-KenB8-KenB9-KenB10-KenB11-NewpB2var1-KenB12-KenB13-Ent5-KenB14-KenB15
Em06-2339	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
Em07-4654	Kentucky	B6	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB34-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
10-2378	Kentucky	B5	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41
10-8911	Kentucky	B5	EntB0-MonB22-MonB24-NiaB2-KenB29-ParCB2-KenB30-KenB31-KenB32-CholB4-BloB2-KenB33-Tho6-KenB42-KenB43-KenB44-KenB45-KenB46-KenB47-KenB48-KenB49-KenB50-KenB51-KenB52-KenB53-KenB54-KenB55-KenB35-STM17-KenB38-KenB39-Ago27var1-KenB41

(*) génomes publiques

Title : Development of a novel molecular method for Salmonella typing and subtyping based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) polymorphism

Keywords : *Salmonella*, molecular typing, CRISPR

Abstract :

Salmonella is the first causative agent of bacterial diarrhea due to food poisoning in developed countries and is still considered as a major public health problem. For almost a century, serotyping has been the gold standard method for *Salmonella* identification. Based on the identification of somatic and flagellar antigens, this manual technique remains time consuming (3–6 days) and expensive (requires the use of > 150 manufactured antisera). About 2 600 serotypes have been described within the *Salmonella* genus. Their distribution being the basis for epidemiological investigations, a fast and accurate identification is essential. Moreover, due to the predominance of particular serotypes, an additional step of subtyping (Pulse–Field Gel Electrophoresis–PFGE) is necessary for a better discrimination and the identification of epidemic clones (5 more days after serotyping). In order to optimize epidemiological investigations, we need to develop a new high throughput method for fast and accurate typing/subtyping, which led us to study the polymorphism of the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) region in Salmonella. Our study on 134 serotypes highlighted the strong correlation between spacer content and serotype/subtype. We also developed a high–throughput method to subtype serotype Typhimurium (CRISPOL), a new PCR method for fast identification of serotypes Typhi and Paratyphi A and a novel PCR method for fast detection of ST198 population of serotype Kentucky.

The discriminatory power of CRISPR typing offers a new prospective for salmonellosis monitoring with a fast and accurate method that could easily replace gold standards serotyping and PFGE.

Titre : Typage et sous-typage du genre *Salmonella* basé sur le polymorphisme des loci CRISPR

Mots-clés : *Salmonella*, Typage moléculaire, CRISPR

Résumé :

Les salmonelloses sont la première cause de diarrhées bactériennes d'origine alimentaire dans les pays industrialisés, demeurant un problème de santé publique majeur. Depuis près d'un siècle, le sérotypage reste la méthode de référence pour l'identification des salmonelles. Basée sur la recherche d'antigènes (un antigène de paroi et un ou deux antigènes flagellaires), cette technique manuelle est longue (3 à 6 jours) et très coûteuse (salaire et utilisation de >150 sérums commercialisés de typage). Près de 2 600 sérotypes ont été décrits chez le genre *Salmonella*. Leur distribution étant la base des investigations épidémiologiques, une identification rapide est primordiale. Par ailleurs, du fait de la prédominance de certains sérotypes, une étape supplémentaire de sous–typage (électrophorèse en champs pulsé–PFGE) est nécessaire pour une meilleure discrimination et la mise en évidence de clones épidémiques (5 jours en plus du délai du sérotypage). Une nouvelle méthode plus rapide, robuste avec des résultats facilement interchangeables entre laboratoires est nécessaire. C'est dans cette optique que nous avons commencé l'étude du polymorphisme de la région CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) chez *Salmonella*. Notre étude portant sur 134 sérotypes a mis en évidence la corrélation entre sérotype et dans certains cas sous–types et profil CRISPR. Ce travail nous a également permis de développer une nouvelle méthode de sous–typage du sérotype Typhimurium (CRISPOL) ainsi qu'une PCR spécifique pour la détection rapide des sérotypes Typhi et Paratyphi A et une autre pour l'identification du clone ST198 du sérotype Kentucky.

Le pouvoir discriminant du typage CRISPR offre une nouvelle perspective pour la surveillance des salmonelloses avec une méthode rapide et robuste et qui pourrait parfaitement se substituer aux méthodes de références que sont le sérotypage et le PFGE.