Source-attribution study of *S.* Derby sporadic cases in France (En préparation)

5.1 Résumé

Suite à la production des génomes de référence pour *S.* Derby, il a été possible de construire une phylogénie plus fine à partir d'une référence adaptée à la lignée majoritaire (ST40). Nous avons inclus l'ensemble des 302 souches humaines collectées en France en 2014-2015 afin de déterminer la source-attribution des souches humaines. Nos résultats montrent la présence, chez l'humain, des 4 lignées identifiées (ST39, ST40, ST71, ST682) dans les aliments. Cependant les lignées identifiées chez le porc (les ST40, ST39 et ST682) représentent 93% des cas humains, là où les souches issues du ST71 n'en représentent que 2% (6/302). 13 profils sporadiques propres aux souches humaines sont également identifés correspondant essentiellement à des profils ST encore non définis.

5.2 Publication

Source-attribution study of S. Derby sporadic cases in France

Yann Sévellec[1], Sophie A. Granier[1], Simon Le Hello[2], Michel-Yves Mistou[1] and Sabrina Cadel-Six[1]*

1. Université PARIS-EST, ANSES, Laboratory for Food Safety, F-94701 Maisons-Alfort, France.

2. Institut Pasteur, Centre National de Référence des *Salmonella*, Unité des Bactéries Pathogènes Entériques, Paris, France.

Running title: Source-attribution study of *S.* Derby.

Corresponding author: Sabrina Cadel-Six, sabrina.cadelsix@anses.fr

ANSES, Laboratory for Food Safety, 14 Rue Pierre et Marie Curie, 94701 Maisons-Alfort

Keywords: *Salmonella* Derby, source-attributionstudy, human infection, antimicrobial resistance.

5.2.1 Introduction

Non-typhoid *Salmonella enterica* subsp. *enterica* is a major cause of food illness causing self-limiting gastroenteritis with 550 million people getting sick each year, including 220 children under 5 year old. *Salmonella* is one of the 4 main causes of gastroenteritis (WHO, 2018). Salmonellosis is usually characterized by acute onset of fever, abdominal pain, diarrhea, nausea and sometimes vomiting. The onset of disease symptoms occurs 6–72 hours (usually 12–36 hours) after ingestion of *Salmonella*, and illness lasts 2–7 days (WHO, 2018). Symptoms of salmonellosis are relatively mild and patients will make a recovery without specific treatment in most cases. However, in some cases, particularly in children and elderly patients, the associated dehydration can become severe and life-threatening.

Salmonella enterica subsp. *enterica* serotype Derby (*S.* Derby) is one of the serovars whose food poisoning mainly affects the population at risk (elderly or children). Appart from these cases, most of the *S.* Derby cases are considered sporadic. In 1946, *S.* Derby caused an epidemic in Australia in 68 infants and young children causing the death of 10 of them (Mushin, 1948). In 1963, an outbreak caused by egg contamination by *S.* Derby caused 822 cases in 53 hospitals in the United States (Sanders et al., 1963). More recently, between the end of 2013 and the beginning of 2014, *S.* Derby is responsible for an outbreak in Germany involving 145 patients, mostly elderly people in Berlin and the Brandenburg region in Germany (Simon et al., 2017). In China, *S.* Derby is the third most reported serovar in clinical cases (Ran et al., 2011) and the most reported serovar in infants and toddlers (Cui et al., 2009). Otherwise these outbreaks affecting immune depressed people, this serovar is mainly associated with sporadic cases: in the United States, from 2005, *S.* Derby causes an average of 120 cases each year (CDC, 2017). In Europe *S.* Derby is the 5th most isolated serovar in human with 570 cases in 2016 (EFSA, 2017). In France, *S*. Derby varied between the 8th to the 5th position (n =164 to 178 clinical isolates) of the most frequently isolated serotypes in humans since 2000 (Weill, 2014).

The principal reservoirs of this pathogen are pigs and poultry worldwide. *S.* Derby was recorded as the most abundant serovar isolated from slaughter pigs with a presence in 28.5% of the production holding in whole European Union in 2008 (European Food Safety, 2008). It is also the 4th most frequently isolated serovar in non-human sector in USA (Schmidt et al., 2012) and the main serovar isolated from slaughter pigs in China (Cai et al., 2016). From 2014, this serovar is also the most frequently isolated in turkey flocks in Europe (EFSA, 2015; 2016; 2017; Sévellec et al., 2018a).

This serovar adapted to pigs and poultry animals is frequently isolated from pork meat. Several studies have associated pork meat and *S.* Derby (Valdezate et al., 2005; Hauser et al., 2011; Kerouanton et al., 2013; Simon et al., 2017; Zheng et al., 2017). In Europe, *S.* Derby accounted for 22.9% of the isolates from pork meat, followed by monophasic strains of *S.* Typhimurium (22.3%) and *S.* Typhimurium (20.6%) (EFSA, 2016). Data from the ANSES *Salmonella* Network (jointly with the National Reference Laboratory) show that this serovar is mainly isolated from

pork and poultry meat in France with 242/598 (40%) and 63/598 (11%) isolates respectively between 2014 and 2015 (Leclerc et al., 2016).

According to a whole genomic analysis realized on a collection of 140 genomes, we identified three different genomic lineages in France for this serovar, ST40-ST39, ST682 and ST71; two of them associated to pork (ST40-ST39 and ST682) and one (ST71) to poultry sector (Sévellec et al., 2018a).

On the other hand, the high prevalence of *S.* Derby in pork and poultry sectors demonstrates the necessity to evaluate the genetic diversity of isolates to identify the source of human clinical cases and to evaluate the pathogenic potential of each lineage to infect humans. Previous attempts of subtyping of the *S.* Derby serovar and to link animal and food strains to human cases where either limited in scope and considered only the pork sector (Hauser et al., 2011; Kerouanton et al., 2013; Zheng et al., 2017) or lacked a representative collection of strains to encompass the real genetic diversity of *S.* Derby (Hayward et al., 2016). We undertook a source-attribution study by variant calling and MLST analyses on the whole *S.* Derby sporadic cases that occurred in France in 2014 and 2015 to identify the sources of contamination and the way of transmission of this serovar within the French territory. We analyzed also the antibiotic resistance potential of the whole collection, 442 *S.* Derby strains isolated from humans, pork and poultry sectors.

5.2.2 Material and methods

Genome Selection.

The 140 *S.* Derby genomes issues from strains isolated from pork and poultry sectors in France in 2014 and 2015 and described in previous publication (Sévellec et al., 2018a) were selected. Human French genomes were from the 302 strains recorded by the CNR (*Institut Pasteur*) and corresponding to the totality of the *S.* Derby clinical cases occurred in metropolitan France in 2014 and 2015. These 302 strains were identified as serotype Derby by glass slide agglutination, according to the White-Kauffmann-Le Minor scheme (Grimont and Weill, 2007) and their genomes were extracted and sequenced as previously described in Sévellec et al. 2018 (Sévellec et al., 2018a). The final human and non-human French collection used for the source-attributionand genome comparative studies comprised 442 genomes.

The metadata of the whole genomes used in this study are reported in **Annexe 2 of this memoire.**

Multilocus sequence typing (MLST).

All genomes were characterized by MLST using seven housekeeping genes *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA* and *thrA* (Achtman et al., 2012). The seven housekeeping gene sequences for each strain were uploaded to the MLST service of the Center for Genomic Epidemiology (CGE) (https://cge.cbs.dtu.dk/services/MLST/), which allowed us to determine the sequence type (ST) directly from the read files.

Single Nucleotide Polymorphism (SNP) and phylogenetic analyses.

The SNP analysis of the 442 French strains was conducted using the VARCall workflow (Felten et al., 2017). The 2014LSAL02547 *S.* Derby reference sequence (NCBI accession number CP029486) (Sévellec et al., 2018b) was used to generate the VCF files. Phylogenetic analyses on the dataset were computed using RaXml software (Stamatakis, 2014). The phylogenetic trees were constructed under the maximum likelihood criterion using the GTR-gamma model of nucleotide evolution. The phylogenetic analyses were based on the pseudogenome obtained using the GATK. The phylogenetic data were visualised using interactive Tree Of Life (iTOL https://itol.embl.de/) (Letunic and Bork, 2016).

Statistical analyses.

The non-normality of the data (number of paired SNP differences) for analyses of both French and international collections, was assessed using Shapiro test (Royston, 1995) on R from the pairwise matrix obtained. The comparison between the paired SNP differences was tested by a Kolmogorov-Smirnov test (KS-test) (Huang et al., 2016) being the variance of the distribution by paired SNP proven significantly unequal by the Fisher test (Markowski and Markowski, 1990).

Identification of acquired resistance genes.

The French collection of genomes was analyzed by ResFinder 2.1 application (Zankari et al., 2012) on the CGE server as described in previous publication (Sévellec et al., 2018a). The localization of resistance gene inside the assembled genomes was tested by running a blast on Bionumerics V. 7.6.1. To verify the presence of a class 1 integron we used the INTEGRALL (Moura et al., 2009) database to identify a class 1 integron for *S.* Derby (NCBI: HG314953.2) to test with BLAST.

Antimicrobial susceptibility tests.

Antibiotic susceptibility of the 140 French non-human strains was described in Sévellec et al. in 2018 (Sévellec et al., 2018a). Antibiotic susceptibility of the 302 French human strains were determined using the disc diffusion method as recommended by the Clinical and laboratory standards institute (CLSI) (CLSI, 2015; 2016). Fifteen antimicrobials (Bio-Rad, Marne-la-coquette, France) were tested: amoxicillin/clavulanic acid (AMC; 30 µg), ampicillin (AMP; 10 µg), cephalothin (CEF; 30 µg), cefotaxime (CTX; 30 µg), ceftazidime (CAZ; 30 µg), chloramphenicol (CHL, 30 µg), sulfonamides (SSS; 300 µg), trimethoprim-sulfamethoxazole (SXT; 1.25+23.75 µg), streptomycin (STR, 10 U), gentamicin (GEN; 10 µg), kanamycin (KAN; 30 UI), tetracycline (TET; 30 UI), nalidixic acid (NAL; 30 µg), ciprofloxacin (CIP; 5 µg), pefloxacin (PEF, 5 µg). Colistin disk (CST; 10 µg) was used on each plate on quality management purposes to ensure the absence of contamination and assess the bacterial identification. Automatic readings were performed using the BIOMIC® V3 system (Giles Scientific Inc., Santa Barbara, California). Isolates were classified as “susceptible”, “intermediate” or “resistant” according to the clinical interpretative criteria recommended by the CLSI (CLSI, 2016).

5.2.3 Results

Genetic diversity and source-attribution study of the French collection.

As shown in **figure 1** the SNPs analysis clustered the French human and non-human genomes in four main groups. These four groups are consistent with the ones already identify by Sévellec et al. (Sévellec et al., 2018a) and identified by MLST analysis (100% identity), indeed, to simplify the comprehension of the results, we have decided to call these 4 SNPs analysis's groups ST39, 40, 71 and 682.

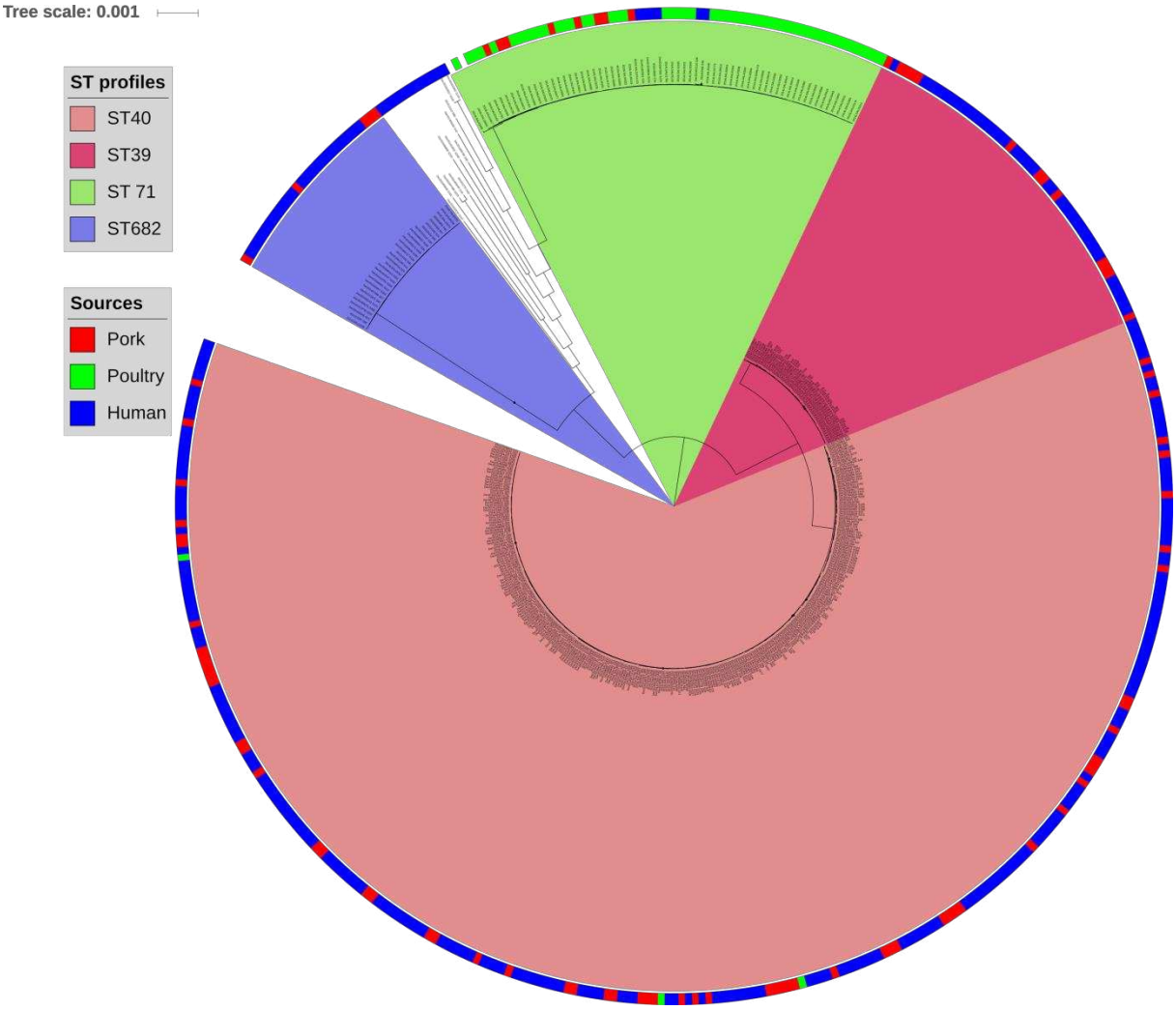


Figure 1: Phylogenetic tree of the French *Salmonella* Derby human and non-human isolates.

Within each of these groups, genomes were found to differ by less than 300 SNPs. The genomes belonging to ST39 were most closely related to ST40 genomes with an average of 3962 SNPs and a standard deviation (SD) of 20 SNPs. The strains belonging to the ST71 cluster were distant from the ones belonging to ST40 by 26,957 SNPs, with an SD of 1,583. The ST682 was the most genetically distant from ST39, ST40 and ST71 with an average of 33,961 (SD of 4,102 SNP). Considering epidemiological information, there was no evidence of a relationship

between the genomic proximity and geographical localization for the different lineages. Closely related strains could have very different geographical origins and *vice versa*. It is also the case for the ST39 and ST682 lineages, which grouped together only 13 and 5 genomes respectively. These lineages presented a very wide geographic distribution, clustering together strains from Pays-de-la-Loire, Bourgogne, Languedoc-Roussillon and Centre. Region Pays-de-la-Loire, however, presented considerable diversity in terms of ST profiles with strains from the four different lineages. Consistent with the distribution of the pork and poultry food production chains in France, ST40 was geographically more widespread than ST71. As previously described by Sévellec et al. (Sévellec et al., 2018a), we can clearly associate three genetic groups with the pork sector (ST39, ST40, and ST682) and one with the poultry sector (ST71).

94% of the French human strains clustered in the genetic groups associated to pork ST39 (54/442), ST40 (284/442) and ST682 (30/442) with 71% of the total human strains belonging to the ST40 alone. Only 6 strains (2% of the human strains) belong to group ST71 related to poultry. The remaining human strains 4% (n=12) belong to various ST profiles already described for *S.* Derby (ST683, ST2639, ST3135) or new ST profiles (10 strains) identified and validated by the CNR for *Salmonella*.

Antimicrobial resistance.

Complete and partial resistance genes identified in the 442 genomes included in this study is reported in the supplementary table 2. No resistance gene were detected for 42% of the strains (188/442) with most of the susceptible strains belonging to the ST71 and ST682 with respectively 97% (62/64) and 93% (28/30) of the strains found to be carrying no resistance gene. Only 17% of the ST40 strains (48/275) presented no resistance genes. In the ST40 64% the strains were carrying simultaneously the *aadA2*, *sul1* and *tetA* genes corresponding to the majoritarian resistance profiles to aminoglycosides, sulfonamides and tetracyclines (STR-SSS-TET) which correspond to 62% of the human strains and 82% of the strains isolated from pork. Twenty strains presented resistance for 4 antibiotic families or more, 16 from the ST40, 2 from ST71 and 2 from ST39. All ST39 are carrying the *fosA7* gene conferring resistance to fosfomycin, this gene is present in the ST40 with a sequence identity of 97%.

5.2.4 Discussion

Genetic diversity of *Salmonella* Derby in France.

In this study, two collections of genomes were compared to investigate the genetic diversity of *S.* Derby in France. The polyphyletic nature of the Derby serovar in food sources have been established in previous studies (Hayward et al., 2016; Zheng et al., 2017; Sévellec et al., 2018a) but this is the first attempts to our knowledge to investigate the diversity of the whole Derby serovar at this scale.

In France, 4 STs are presents in non-human strains forming 3 genomic lineages, the ST39-40, ST71 and ST682 (Sévellec et al., 2018a). The human stains were characterized by a higher diversity in ST profiles. Thirteen other, profiles were identified within human strains, corresponding to 1 or 2 strains each. For these STs no food sources were identified.

This can be due either to the panel of non-human strains analyzed not including bovine and ovine sources, either to the specificity of this ST for human contamination and invasion. More analysis would be needed to confirm these hypotheses. On the 302 strains isolated from human clinical cases, 93% are associated to STs associated with the pork sector. Among them indeed the ST40 is responsible for 71% of the human clinical cases and both clades (1 and 2) of the ST40 are associated with human clinical cases. The ST71 associated with the poultry sector accounts only for 2% of the human cases, two of the clinical cases associated with ST71 were caused by strains of the "Asian clade" of the ST71 one of those cases is associated with travel in Thailand.

Surprisingly, there are few chronological or geographical links between human isolates and the most closely related food isolates. This can be explained by the structure of the pork sector in France where most of the slaughtering is done in a few slaughterhouses that handle industrial transformation and provide retail sale nationwide. Slaughterhouses represent a vulnerable spot for microbiological contamination as described by multiple publications (Rostagno et al., 2003; Botteldoorn et al., 2004; Sofos, 2008; Fois et al., 2017). The industrialization and the globalization of the food sector promotes a geographical dispersion of contaminated products or animal which facilitates the dissemination of pathogen, thus makes the localization of the origin of an outbreak more difficult (Newell et al., 2010; Ercsey-Ravasz et al., 2012). The congelation of the stock may also to take in account as *Salmonella* survive congelation (Chaves et al., 2011). Nonetheless, we were able to link cluster of human strains to a specific food strain. As an example, In the ST71, the human strains 201506934 and 201507632 are linked to strains associated with *Gallus Gallus* isolated in Britanny. In the ST40, 5 clinical strains are linked to the food strain 2014LSAL04065 collected from pepper sausage.

Studying the genetic diversity of the population of a given pathogen of interest provides valuable information about the potential sources of contamination and help to direct future investigations of outbreaks. The WGS technology have been proven in this study a valuable tool for the source-attributionof human cases as reported in previous publications (Gilchrist et al., 2015; Dunn, 2016; Ferrari et al., 2017).

Antimicrobial susceptibility of the French strains.

Antimicrobial use within animal sectors constitutes an evident evolutionary pressure for this pathogen as 67% of the strains analyzed present at least one resistance gene in their genome. As reported in previous study (Sévellec et al., 2018a), most of the susceptible strains belong to the ST682 (28/30 strains) and the ST71 (62/64 strains). The strains of the ST40 are in majority resistant to at least on class of antibiotic with a majoritarian resistance profiles (STR SSS TET) is present in 64% of the isolates. This resistance profile is specific of the clade 2 of the

ST40 and is carried by an SGI-1 specific of *S.* Derby described as SGI-1C (Beutlich et al., 2011; Sévellec et al., 2018a). Aminoglycosides, sulfonamides and tetracycline being the most used antibiotics in the pork sector (Méheust et al., 2016), this resistance profile might explain the success of this particular clade (ST40) (Newell et al., 2010). This resistance profile was already reported in Spain in 2005 (Valdezate et al., 2005). In France, in 2013, this resistance profiles was reported in pigs (Kerouanton et al., 2013) and is widespread in Europe in *S.* Derby strains from the pork sector or human cases (Bonardi, 2017).

The SGI-1 present in the genome of the dominant clade of ST40 in the pork sector carries also a mercuric resistance operon *mer*. Resistance to heavy metals such as mercury have been demonstrated as a co-selection factor for antimicrobial resistance as they are often carried by the same transposons (Baker-Austin et al., 2006; Pal et al., 2015; Deng et al., 2017b). The presence of resistance to heavy metals was already observed in the major clone in Europe of monophasic variant of *Salmonella* Typhimurium (Petrovska et al., 2016).

The presence of a resistance gene for fosfomycin: *fosA7* (Rehman et al., 2017) was detected in all isolates from the ST39 and a variant with 97% of identity is present in all the strains of the ST40. Those genes were not reported in previous prediction (Sévellec et al., 2018a) as it was included recently in the Resfinder database. This demonstrates the importance of keeping track of the updates in public databases since some resistance gene can go unnoticed in previous predictions. The resistance to fosfomycin was not tested nor in our antimicrobial susceptibility test or in previous publications at our knowledge should further antimicrobial susceptibility tests be performed to verify the functionality of the *fosA7* gene both in ST39 and in ST40.

Conclusion.
We were able to distinguish, in France, four major genetically distant STs in the *S.* Derby serovar (ST39, ST40, ST71 and ST682) in pork and poultry sectors and in human. The ST39, ST40 and ST682 are mainly associated to pork strains and represent 93% of the total human strains collected by *Pasteur Institute* in 2014-2015. The pork source is the main source for human contamination. The structure of the pork sector in France and the consumption habit in France were able to explain the spread and the sporadic pattern of the human cases associated with the food sources.

6 Publication V: Investigation of the host specificity of *S.* Derby. (En préparation)

6.1 Résumé

Suite au constat de la très faible proportion des souches isolées depuis le secteur volailles parmi les cas cliniques humains, nous avons voulu étudier la pathogénicité des différentes lignées de *S.* Derby identifiées lors de cette étude afin de pouvoir comprendre les différences

d'hôtes entre les lignées de ce sérovar. Nos questions de recherche étaient : quelle est la gamme d'hôtes des différentes lignées de *Salmonella* Derby ? Les souches du ST71 ont-elles une capacité atténuée à infecter l'humain ?

Pour y répondre nous avons étudié les variations génomiques concernant des facteurs de virulence identifiées de *Salmonella* en les comparant à *S.* Typhimurium -le sérovar généraliste par excellence- et l'un des mieux étudiés. Nous avons également fait une étude du pan génome de *S.* Derby afin d'identifier les principales différences génétiques entre les lignées. Enfin nous avons confirmé nos résultats par une étude de la capacité d'invasion cellulaire des lignées ST40, ST71 et ST682 chez l'humain, le porc et la volaille. Nos résultats montrent une capacité d'invasion cellulaire réduite des souches du ST71 chez l'ensemble des souches du ST71.

6.2 Publication

Investigation of the host specificity of S. Derby.

Yann Sévellec[1], Sylvie M. Roche[2], Sophie A. Granier[1], Isabelle Virlogeux-Payant[2],Simon Le Hello[3], Laurent Guilliet[1], Michel-Yves Mistou[1] and Sabrina Cadel-Six[1]*

1. Université PARIS-EST, ANSES, Laboratory for Food Safety, F-94701 Maisons-Alfort, France.
2. ISP, Institut National de la Recherche Agronomique (INRA), UMR 1282, Université de Tours, Paris, France
3. Institut Pasteur, Centre National de Référence des *Salmonella*, Unité des Bactéries Pathogènes Entériques, Paris, France.

Running title: Host specificity of *S.* Derby.

Corresponding author: Sabrina Cadel-Six, sabrina.cadelsix@anses.fr

ANSES, Laboratory for Food Safety, 14 Rue Pierre et Marie Curie, 94701 Maisons-Alfort

Keywords: *Salmonella* Derby, host specificity, core genome, pan genome.

6.2.1 Introduction

Salmonella displays an important diversity in host range dependent of the serovar. The different serovar can be distinguished between serovar that are host restricted (e.g. *S.* Typhi in humans), host adapted (e.g. *S.* Choleraesuis and the pork) or broad range as *S.* Typhimurium (Yue et al., 2012; Yue et al., 2015). These serovars possess very diverse pathogenicity patterns determined by variation in virulence factors especially in the *Salmonella* pathogenicity island

(Sabbagh et al., 2010), in adhesion factor (Yue et al., 2012) or even in the allelic variation of key virulence factor (Eswarappa et al., 2008; Yue et al., 2015).

Several studies have reported differences in host association inside the *S.* Derby serovar with a lineage (ST71) associated with poultry (Hayward et al., 2016; Sévellec et al., 2018a; Sévellec et al., 2018c). Zheng et al. reported the ST71 in China but this serovar was not associated with human cases (Zheng et al., 2017). In Uruguay Betancor et al. reported the presence of a lineages of *S.* Derby in eggs which was not found in human, those strains were however not associated with an ST profiles (Betancor et al., 2010). A previous study demonstrated differences in virulence factors of the different lineages of *S.* Derby identified in France, with the presence in the ST39-ST40 lineage of the SPI-23 associated with pork invasion capacity (Sévellec et al., 2018a). The result of our source-attribution study described in the part 5 of the results suggested also a lack in pathogenicity for the strains of the ST71 associated with poultry.

Several studies (Betancor et al., 2010; Hauser et al., 2011; Hayward et al., 2013) have investigated the set of virulence factors in *S.* Derby's genomes but those studies were limited to specific lineages of this serovar and did not provide a comprehensive view of the pathogenicity of *S.* Derby. We reported in previous publication (Sévellec et al., 2018a) differences in allelic variations and in the structure of the SPI-1 to -5 and the presence of the SPI-23 in specific lineages of *S.* Derby but this investigation was incomplete and did not revealed critical differences that could explain the difference in host association between the different *S.* Derby lineages.

The aim of this study was to identify the genomic signatures able to explain the divergence of behavior between these pathogens (Derby and Typhimurium) and between the different lineages of *S.* Derby. We made a detailed comparison study of the genome sequences associated to the virulence factors between the three different *S.* Derby lineages (ST39-ST40, ST682 and ST71) and compared to *S.* Typhimurium serovar which is the most frequently associated with outbreaks worldwide. The results of this genomic analysis were confirmed by *in vitro* adhesion, invasion assay on epithelial cells from human, pork and *Gallus Gallus*.

6.2.2 Material and Methods

Genome Selection.

We used the collection of 442 *S.* Derby described in 4.5 to constitute a collection of assembled genomes using Spades V 3.9.1. (Bankevich et al., 2012). Due to poor quality of the assembly of several genomes of human strains, we limited the collection to 288 human clinical cases for a total of 428 strains (see annexe 3).

Pan-genome investigation of the French lineages.

Pan genome analysis was performed on the genomes belonging to the well described lineages ST39-ST40, ST682 and ST71 (Sévellec et al., 2018a; Sévellec et al., 2018b; Sévellec et al.,

2018c). The 428 genomes were processed using Prokka software (V. 1.11) (Seemann, 2014) to produce annotation of the CDS. The pan genome content was analyzed using ROARY (V. 3.8.0) (Page et al., 2015). Specific traits were then tested using Scoary software (V. 1.6.16) (Brynildsrud et al., 2016). Pan genome comparisons were performed using different traits: the sector the strain was isolate from, the lineage the strains belongs to and the cluster of each ST40 strain.

Virulence factors identification.

SPI-1 to -19; SPI-23 and CS54 were researched within the 442 genomes of *S*. Derby French collection using blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with a cut-off of 90% of identity. The references for the SPI-1 to -5 and the SPI-23 were extracted as described in previous work (Sévellec et al., 2018a). SPI 6 to 19 were extracted using the pathogenicity island database (PAI DB) (Yoon et al., 2015). CDS were extracted from reference genomes. The references used to extract the SPI are reported in the Annexe 4.

Additional virulence related genes were extracted from the virulence factor database (VFDB) (Chen et al., 2012).

Cell lines and culture conditions.

Human adenocarcinoma HT-29 cell line (ECACC no. 85061109, Salisbury, UK) was grown in Dulbecco's Modified Eagle's Medium (DMEM with 4.5 g/L glucose (Gibco) supplemented with 2 mM L-glutamine (Gibco) and 10% fetal calf serum (SVF) (Gibco). Chicken hepatoma LMH cell line (ATCC CRL-2117) was grown in Williams' Medium E (Lonza) supplemented with 10% SVF. New-born piglet intestinal IPEC-1 cell line (Gonzalez-Vallina et al., 1996) was cultured in DMEM-F12 (Lonza) supplemented with 5% SVF, 2 mM L-glutamine, 10 µg/mL insulin - 5 µg/mL transferrin - 5 ng/mL sodium selenate (ITS – Sigma), 5 ng/mL epidermal growth factor. Cells were routinely grown in 75 cm2 plastic tissue culture flasks at 37 °C under 5% CO2. Throughout this study, cells were grown without antimicrobial compounds.

Adhesion-invasion assays.

Adhesion-invasion assays were performed as previously described (Roche et al., 2005) on 4 *S.* Derby strains: two from the ST40 (2014LSAL02547 and 07CR553 one from the ST71 2014LSAL01779 and one from the ST682 2014LSAL03787). Briefly, 24-well tissue culture plates were seeded five days before infection in order to obtain confluent monolayers for the day of infection. Bacteria were grown overnight in TSB medium at 37 °C without shaking. Each experiment used three plates: one to determine the total number of adhered and invaded bacteria (total-cell associated (TCA) assay), the second the invaded bacteria only (invasion assay) and the third one for the multiplication assay. Cell monolayers were incubated in medium for 24 h and then infected for 1.5 h, 3 h or 24h at 37 °C with 10^7 CFU in 300 µL (multiplicity of infection 10). For TCA assays, the cell monolayers were gently washed six times with phosphate buffered saline (PBS) (pH 7.3) and then disrupted with 1 mL cold distilled water (4 °C). Viable bacteria (intra- and extracellular) were counted after plating serial dilutions on TSA (Difco). For the invasion and multiplication assays, the plates were washed

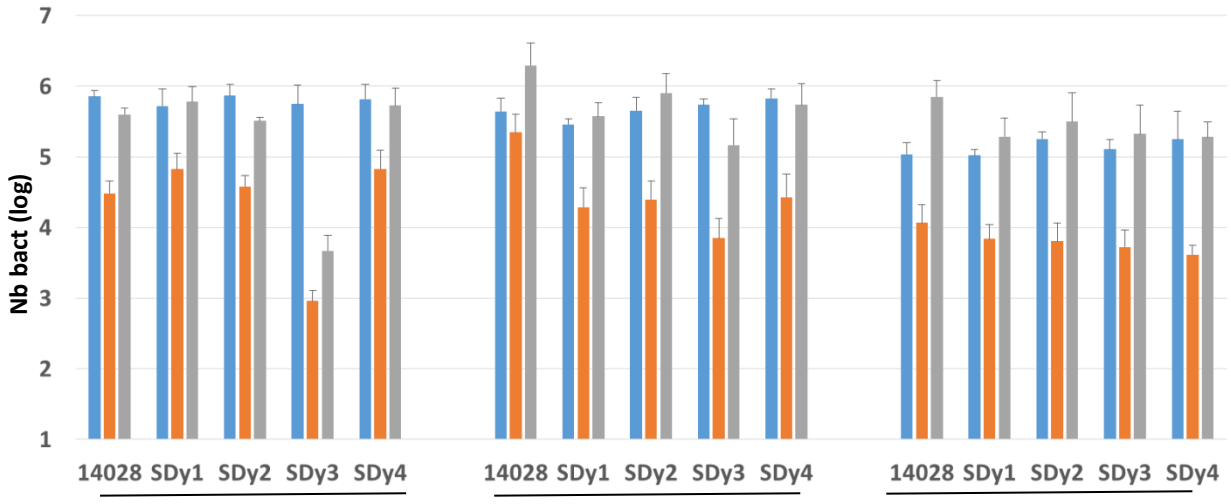
once with appropriate medium and incubated in culture medium containing 100 µg/mL gentamicin. After 1.5 h at 37 °C, the plate for the invasion assay was washed once with PBS and lysed with 1 mL cold distilled water (4 °C). Viable intracellular bacteria were assessed by serial dilutions plated on TSA. For the multiplication assay, the plate was washed once with appropriate medium and incubated in culture medium containing 10 µg/mL gentamicin. After 21 h at 37 °C, the plate was washed once with PBS and lysed with 1 mL cold distilled water (4 °C). Viable intracellular bacteria were assessed by serial dilutions plated on TSA. Results were expressed as mean ± SD of the number of bacteria for 10^7 CFU. Experiments were carried out in duplicate and repeated third for each strain.

6.2.3 Results

Adhesion-invasion assays.

The strains 2014LSAL01779 belonging to the ST71, the 2014LSAL03785 belonging to the ST682 and the strains 2014LSAL02547 and 07CR553 (Kerouanton et al., 2015) belonging to the ST40 were tested against *S.* Typhimurium 14028 to evaluate their adhesion and invasion capability on epithelial cells HT29 (human), IPEC-1 (Porc) and LMH (*Gallus gallus* epithelial cell). The results are shown in **figure 1**.

A.



B.

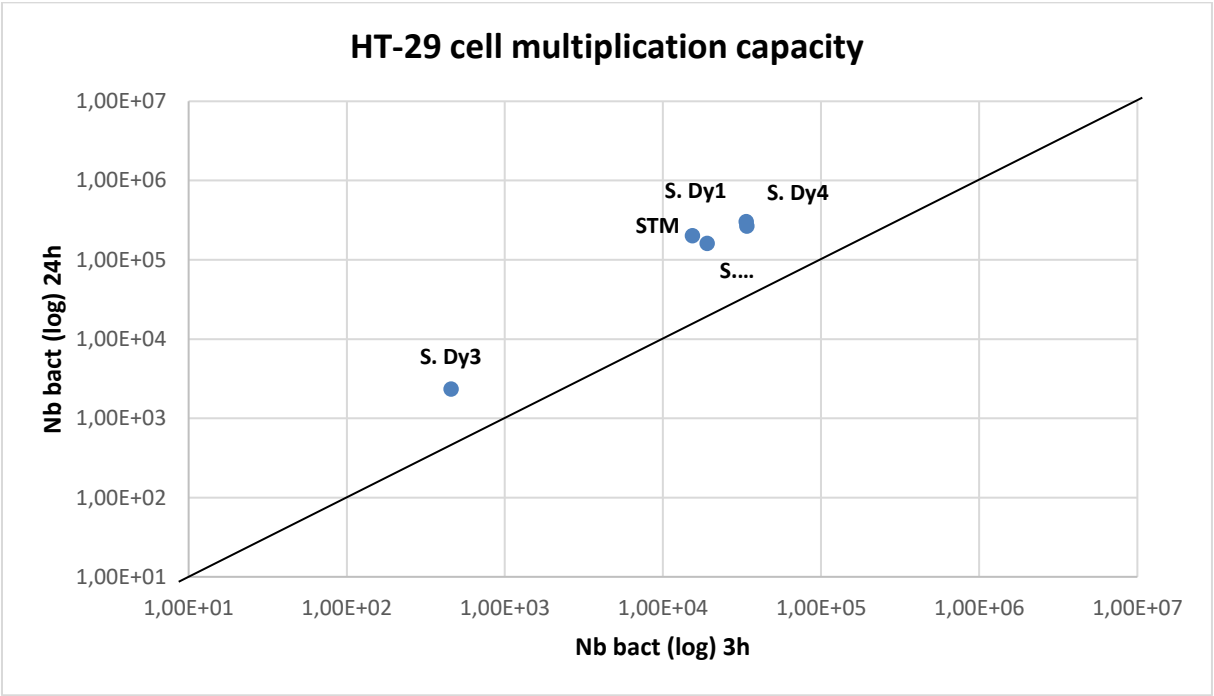


Figure 1: Differences in adhesion, invasion and multiplication capacity of the ST40, ST71 and ST682 lineages of *S.* Derby on human (HT 29), *Gallus gallus* (LMH) and pig (IPEC-1) epithelial cells.

14028: STM14028; SDy1: 2014LSAL02547; SDY2: 2014LSAL03785; SDy3 2014LSAL01779; S. Dy407CR553.

A: capacity on adhesion, invasion and multiplication of the different lineages of *S.* Derby. *In blue: adhesion; in orange: invasion; in gray: multiplication.*

B: Correlation between the invasion and multiplication capaticy of the different S. Derby lineages.The strain S. Dy3 (ST71) present a significantly lower invasion capacity leading to decreased multiplication in human HT 29 epithelial cells. The diagonal correspond to the X=Y function as a comparison point.

All the *S*. Derby strains have a similar adhesion capacity on all the three epithelial cell type. Those adhesion capacities were statistically similar to *S.* Typhimurium 14028.

At the invasion stage, the three *S.* Derby STs showed the same invasion capacity for pork epithelial cells IPEC-1 (similar to STM14028) and *Gallus Gallus* epithelial cells LMH (which was 1 log lower than STM14028). The strains from ST71 presented 1 log decrease in CFU at the invasion stage on HT 29 cells compared to STM14028 and the other *S.* Derby of the ST40 and ST682. The multiplication capacity was similar for all the epithelial cell type and all the tested strains.

A second experiment was performed to assess the in vitro invasion capacity on 6 strains of *S.* Derby ST71 lineages from poultry (strains 2014LSAL01779 and 2014LSAL05133) and for human strains from the European (strains 201402501 and 201510930) and Asia clade (strains 201402459 and 201407239) of the ST71. The results were consistent with the experiment presented in the figure 1 for all the tested strains expect the strain 201402501 which showed impaired invasion capacity compared to the other ST71 strains.

The results of this experiment are presented in the annexe 5.

Virulence factors.

As demonstrated in previous study (Sévellec et al., 2018a), all *S.* Derby STs shared the SPI-1 to 5 with a high identity (superior to 98%) for most of the genes. All the *S.* Derby analyzed also possessed then SPI-9 and fragments of the SPI-6, -11, -13, -16 and -19 as reported in the table 2. As for the SPI-1 to -5, the genes present in the different ST groups of *S.* Derby revealed an identity of 98% or more to the reference. Some key virulence factor present nonetheless some more important variation between the 4 STs of *S.* Derby as exposed in **Table 1**.

Table 1: Presence of known SPI in the different *S.* Derby lineages.

**Absent or present at an identity cutoff of 90%. [a] presence of fimbriae operon; [b]absence of sugR gene.*

***Salmonella* pathogenicity island**	**length (kpb)**	**genes nb**	**ST40-ST39***	**ST71***	**ST682***
SPI-1	38	39	39/39	39/39	39/39
SPI-2 common region	12	10	10/10	10/10	10/10
SPI-2 specific region	25	31	31/31	31/31	31/31
SPI-3	16	10	8/10[a]	8/10[a]	9/10[b]
SPI-4	27	10	10/10	10/10	10/10
SPI-5	9	8	8/8	8/8	8/8
SPI-6	47	52	34/52	34/52	37/52
SPI-7	134	75	2/75	0/75	1/75
SPI-8	8	12	2/14	14/14	8/14
SPI-9	16	4	4/4	4/4	4/4
SPI-10	33	24	0/24	0/24	0/24
SPI-11	14	19	12/19	12/19	12/19
SPI-12	16	12	9/12	9/12	10/12
SPI-13	25	19	13/19	13/19	13/19
SPI-14	9	6	0/6	0/6	0/6
SPI-15	6,5	5	0/5	0/5	0/5
SPI-16	4,5	6	0/6	0/6	0/6
SPI-17	5	7	1/7	6/7	0/7
SPI-18	2	2	0/2	0/2	0/2
SPI-19	42	30	3/30	4/30	5/30
SPI-23	36	41	41/41	0/41	0/41
CS54	25	8	5/8	6/8	8/8

In the SPI-6 all ST groups lacked the *safA* gene from the *saf* fimbrial operon, the whole *tcr* operon from *S.* Typhi and the *tinR* revolvase but possessed a functional T6SS operon.

The SPI-17 was only complete for the ST71 with the 6 CDSs present in the reference and an additional hypothetical protein between *gtrA* and the IS256 transposase corresponding to STY2626. Only 1 CDSs was present in the ST40 and ST39 (STY2631, corresponding to an IS256 transposase). The SPI-17 was absent in the ST682.

The SPI-13 was present in all the genomes with the genes *SG3012* to *SG3017* missing. In the ST71 and ST682 this locus contained instead the SPI-8 from *S.* Typhi. The ST71 presented a complete SPI-8 with 14 CDS (12 from the reference and 2 additional CDS coding for a partial T4SS system). The ST682 presented a SPI-8 more closely related to *S.* Typhi CT18 except the missing genes *STY3289* to *STY3292*. In the ST40 and ST39 2 genes only (*STY3291* and *STY3292* coding for hypothetical protein) were present in the genomes but the two lineages presented a new insertion at this locus containing 5 genes included an iron reductase as presented in the Annexe 6. The SPI-9, -10, -14, -15 and -18 are totally absent in all lineages.

The locus CS54 was complete for the ST682. The ST71 present an 802 bp deletion in *shdA* and lack the *ratB* and *ratC* gene. The ST40 and ST39 lack the *ratB*, *ratC* and the *sinI* genes is not detected at a cutoff of 90% of identity.

If we considerate the comparison between *S.* Derby and Typhimurium serovars, the main diversities are in the SPI-6, -11, -12, -13 partial in the different lineages of *S.* Derby analyzed and the absence of the SPI-14 and-16 in *S.* Derby lineages.

The **table 2** summarizes the presence or absence of several *Salmonella* virulence factors.

Table 2: Presence of known virulence factors in the different *S.* Derby lineages.

**Absent or present based on an identity cutoff of 90%; partial[a] : corresponds to at least 20% of the genes missing for gifsy1 and gyfsy2; LT2: Salmonella Typhimurium LT2 (NC_003197.2). In red: virulence factor which present differencies between the lineages.*

Virulence factor	Role	ST40-ST39*	ST71*	ST682*	Reference used
virulence plasmid	Plasmid	absent	absent	absent	plasmid (AB040415.1)
gifsy1	Phage	absent	absent	partial[a]	425..48,265 (NC_010392.1)
gifsy2	Phage	absent	absent	partial[a]	42..42,753 (NC_010393.1)
misL	Adhesion factor	present	present	present	3,954,765..3,957,640 (LT2)
ratB	Adhesion factor (CS54)	absent	absent	present	2634402..2641709 (LT2)
shdA	Adhesion factor (CS54)	present	present	present	2627589..2633708 (LT2)
sinH	Adhesion factor (CS54)	present	present	present	2648609..2650801 (LT2)
bcf	Fimbriae (adhesion)	present	present	present	24,469..32,119 (LT2)
csg	Fimbriae (adhesion)	present	present	present	1,228,014..1,232,462 (LT2)
fim	Fimbriae (adhesion)	present	present	present	604,118..613,305 (LT2)
lpf	Fimbriae (adhesion)	absent	present	present	3,823,024..3,827,461 (LT2)
pef	Fimbriae (adhesion)	absent	absent	absent	7,904..14,362 (LT2)
saf* (safA absent)	Fimbriae (adhesion)	present	present	present	342,210..346,569 (LT2)
stb	Fimbriae (adhesion)	present	present	present	378,982..384,940 (LT2)
stc	Fimbriae (adhesion)	ST39 only	absent	absent	2,242,830..2,247,625 (LT2)
stf	Fimbriae (adhesion)	present	present	present	230,658..236,265 (LT2)
sti	Fimbriae (adhesion)	present	present	present	203,991..208,922 (LT2)
stj	Fimbriae (adhesion)	absent	present	absent	4,826,485..4,829,614 (LT2)
tiv (typhi)	immune evasion (antigen)	absent	absent	absent	4,516,537..4,524,140 (typhi CT18)
vex (typhi)	immune evasion (antigen)	absent	absent	absent	4,510,583..4,515,425 (typhi CT18)
rck	Adhesion/invasion factor	absent	absent	absent	plasmid (AB040415.1)
pagN	Invasion factor	present	present	present	STM0306 (LT2)
sinR	transcription regulator	present	present	present	348,611..349,558 (LT2)
sseI	Cellular survival	present	present	present	gifsy2 - 42..42,753 (NC_010393.1)
sseJ	T3SS effector	present	present	present	gifsy2 - 42..42,753 (NC_010393.1)
sopE	T3SS effector	ST40 only	absent	absent	4,481,801..4,482,523 (typhi CT18)
sopE2	T3SS effector	present	present	present	1,952,009..1,952,738 (LT2)
pipB2	T3SS effector	present	present	present	2,926,806..2,927,863 (LT2)
sifA	T3SS effector	present	present	present	1,309,834..1,310,844 (LT2)
sifB	T3SS effector	present	present	present	1,692,171..1,693,125 (LT2)
sopD	T3SS effector	present	present	present	1,054,059..1,055,023 (LT2)
sopD2	T3SS effector	present	present	absent	3,087,148..3,088,108 (LT2)

Most virulence factors are either absent (at a 90% identity cutoff) in all lineages or present with an identity to the reference higher than 98%. With the exception of several adhesion factors: the *ratB* gene form the genomic island CS54 is only present in ST682. The fimbriae

operon *csg* corresponding to the *agf* aggregative fimbriae is absent in the ST40 and ST39. The *stc* fimbriae operon is only present in the ST39 and the *stj* fimbriae operon is only present in the ST71. The *sopE* effector of the SPI-1 is only present in the ST40.

The virulence factors with an identity to the reference inferior to 98% to at least one lineage are: *sseC* (95%), *sseD* (96%), *sseF* (97%), *sseJ* (96%), *sopD2* (between 78% and 95%), *pipB2* (93%) and *sopE2* (97%).

According to the comparison with *S.* Typhimurium the main differences are the absence in all the ST groups of *S.* Derby analyzed of the *pef* (fimbriae-adhesion) gene. Other differences observed according to the ST groups are within the *ipf*, *stc* and *stj* genes.

Several virulence-associated genes from SPI or outside of SPI presented significant allelic variations as presented in the **table 3**. Those genes are adhesion factor (*safB*) or T3SS effector (*sseJ*, *sifA*, *pipB2*, *sopD2*) *sseD* is part of the translocon of the T3SS-2.

Table 3: Identity Variation between the lineages of *S.* Derby in several virulence related gene.

Virulence factor	Function	ST40 vs ST39	ST40 vs ST71	ST40 vs ST682	ST71 vs ST682
sseD	T3SS-2 translocon	100%	98%	98%	96%
sifA	T3SS-2 effector	96%	95%	98%	94%
safB	Fimbriae (adhesion)	99%	93%	92%	99%
sseJ	T3SS-2 effector	100%	98%	96%	97%
pipB2	T3SS effector	98%	96%	94%	94%
sopD2	T3SS-2 effector	100%	93%	78%	78%

Pan genome analysis.

From the ROARY analysis, the *core* genome of the Derby serovar was constituted of 3 501 genes (3 214 genes plus 287 genes present in more than 95% of the strains). 1 464 additional genes were part of the shell genome (present in more than 15% of the strains) and 13 184 genes constituted the cloud genome, present in less than 15% of the strains.

Figure 2 summarizes the genes that are specific to the different *S.* Derby lineages. The main differences in the accessory genome concern hypothetical proteins. Other important group of protein function was: phage associated genes, defense mechanisms, membrane protein, included fimbriae related genes, DNA recombination proteins and signal transduction mechanisms. The ST39 and ST40 hypothetical protein include the genes associated with SPI-23. The details are presented in annexe 7.

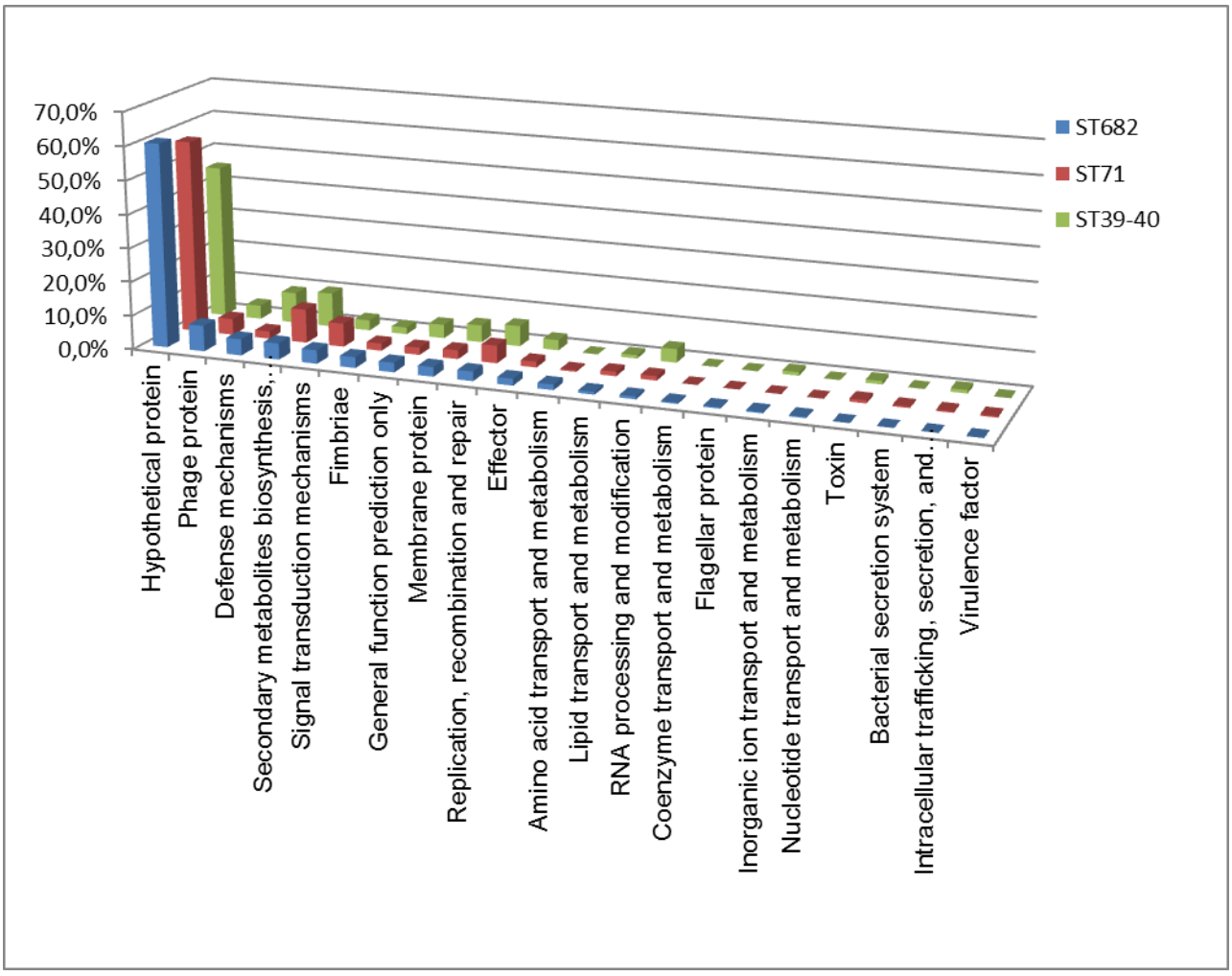


Figure 2: general function of lineages specifics genes.

The Scoary analysis demonstrates that the operon *yehABCD* can be used to discriminate the ST39 from the other ST groups.

ST682 can be characterized by the *tfaE* gene associated with phage tail fiber assembly protein and by several genes (*yteP* and *garK*) associated with sugar metabolism. The *yafA* gene is also specific in the ST682. This gene was previously reported as absent of the Derby lineage (Anjum et al., 2005).

ST40 is characterized by the presence of several integrases associated with the SPI-23: *tnpA*. And by the presence of a second copy of the *lexA* gene associated to the repression of DNA repair proteins. Between the clade 1 and 2 of ST40, all characteristic differences are genes from the SGI-1, characteristic of the clade 2.

ST71 is characterized by the present of the *rrrD* gene coding for a lysozyme associated with the DLP12 prophage and by the *ddrA* gene associated with phage SJ46.

We tested the characteristic differences between the strains isolated from the poultry sector and the pork sector. The *sopD2* variant of the ST71 was associated to the poultry sector with a P-value >$1*10^{-26}$. As well as the beta lactamase precursor hcp and the gene *ugpA* coding for a sn-glycerol-3-phosphate transport system permease protein. Pork strains were characterized by the *rayT* gene coding for a REP-associated tyrosine transposase and by hypothetical proteins.

6.2.4 Discussion

Pathogenicity potential of the different *S.* Derby lineages.

The source-attribution studies on the different lineages of *S.* Derby demonstrated that the poultry related ST71 lineage is marginally present in human. As the poultry meat represents 28 % of the French meat consumption comparable with the 33 % of the pork sector, we may suppose that the relative burden of strains isolated from the poultry and pork sector would better be explained by a difference in pathogenicity between the strains.

The results of the invasion assays seem to confirm this hypothesis. Significant differences were found in the invasion capacity for human epithelial cells between the ST71 and the other *S.* Derby lineages. All the lineages presented the same adhesion and multiplication capacities. Those results are consistant with population structure detailed in this study as well as the results reported by Zheng et al. which does not reported ST71 strains in human (Zheng et al., 2017).

In order to understand the genomics factor that can explain the differences in invasion assay we investigated key virulence factors and performed pan genome analysis. We found that all *S.* Derby lineages lacked the *S.* Typhi associated SPI-7, -10, -15 and -18 as reported by Hayward et al. in 2013 (Hayward et al., 2013). The SPI-14, and -19 (except a few hypotheticals proteins) present in *S.* Typhimurium a ubiquitous serovar, were absent in all *S.* Derby lineages as well as the *Salmonella* virulence plasmid and presented a fragmented SPI-16 which in associated with immune evasion through the modification of the O-antigen (Ilyas et al., 2017; Singh et al., 2018). The absence of the SPI-16 in the ST71 could lead to a decrease in long term survival in the eukaryotic cells and to an impaired systemic invasion capacity (Ilyas et al., 2017). The fragmentation of the SPI-16 could explain the presumed low pathogenicity of S. Derby compared to *S.* Typhimurium in our invasion assays but this hypothesis should be investigated further to assess the long-term survivability of *S.* Derby and understand the differences in pathogenicity between *S.* Derby and *S.* Typhimurium.

The SPI-1 to -5 and -9 were present in all lineages as well with few variations in the sequences of tested CDS. The SPI-6, -11, -12 and -13 were found incomplete in all lineages. The know virulence related gene of the SPI-11 were present in all lineages. The SPI-12 lacked the *sspH2* gene coding for an effector of the T3SS-2 except for the ST71. All lineages lacked the *safA* gene of the *saf* fimbriae operon and the *tcf* fimbriae operon associated to host specificity in *S.* Typhi. Those results were reported in previous studies (Betancor et al., 2010).

The CS54 was incomplete in the ST39, ST40 and ST71, as well with the genes *ratB* (involved in *Salmonella* colonization of the intestine) and *ratC* incomplete in those 3 lineages as reported by Hayward el al. (Hayward et al., 2013). The *sinI* gene coding for an outer membrane protein is only present in ST682 and ST71 but the role of this gene in virulence has not been demonstrated (Kingsley et al., 2003).

All the lineages present a deletion of the genes *SG3012-SG3017* of *S.* Gallinarum 287/91 in the SPI-13, this locus is known to have undergone multiple recombination events in *Salmonella* (Espinoza et al., 2017). The main difference in the SPI structure of the different Derby lineages was the presence of 3 different insertions in this locus. The ST71 presented the SPI-8 at the location of those genes, similarly to *S.* Typhi. The SPI-8 present in the ST71 contained 14 CDS including 2 genes coding for an incomplete T4SS and 2 bacterioncin immunity genes previously identified in *S.* Typhi (Sabbagh et al., 2010). The ST682 presented a partial SPI-8 with 10 CDS including the 2 bacteriocin immunity protein. The ST40 and ST39 possess, at this locus, an insertion of 7 CDS including the *lysR* regulator and a gene coding for an iron reductase. Similar results were reported by Betancor et al. (Betancor et al., 2010) with missing elements of the SPI-13.

ST71 also possesses the SPI-17 from *S.* Typhi with the *gtrA* and *gtrB2* gene involved in LPS O-antigen modification (Singh et al., 2018).

All the lineages present differences in the set of fimbriae operon but those differences seem not been translated in difference of the adhesion capacity for human, porcine and *Gallus Gallu*s epithelial cells, during the *in vitro* test. The variation of fimbriae was reported a marker of adaptation to different host ranges we cans infer that the differences lineages of S. Derby is an evolution to a different set of hosts (Yue et al., 2012) but this hypothesis must be further investigated by different adhesion assays to understand the host range of each S. Derby lineages. the complete fimbriome of S. Derby remain to be systematically investigated as well as this study focused only of fimbrial adhesion factors present in *S.* Typhimurium or SPI and cannot be considered exhaustive.

As for virulence factors, the most important differences between the lineages of *S.* Derby concern the effector of the T3SS that are located outside of the SPIs. With the *sopE* gene only present in ST40. This gene is involved in the T3SS-1 as the SopE protein mediate invasion by stimulating the membrane ruffling of epithelial cells (Friebel et al., 2001). SopE is also a promoter of intracellular and extracellular inflammation (Ramos-Morales, 2012; Vonaesch et al., 2014) and plays an important role in the early intracellular multiplication of *Salmonella*. The presence of *sopE* could explain the greater invasion capacity of ST40 over ST71 but the similarities in invasion capacity between ST40, ST39 and ST682 have yet to be explained.

The ST682 lacked also the *sopD2* gene and the ST40-ST39 allele of this gene presented important allelic variations compared to the ST71 allele, with an identity of 93% between the two alleles. The variants of the gene *sifA*, *sseJ* and *pipB2* were also important (table 4).

The SifA and SseJ effector synergize together into a complex with the mammalian protein SKIP/RHoA to promote the stabilization of the *Salmonella* Containing Vacuole (SCV) and intracellular multiplication of *Salmonella*. SifA is also known to interact with pipB2 and sopD2 as antagonist (Ramos-Morales, 2012). The allelic variations of those 4 genes could have an important impact on the multiplication of *S.* Derby in macrophages and epithelial cells and thus explain the difference in pathogenicity of the different *S.* Derby lineages.

The impact of the variations of those effectors is beyond the scope of this study but should provide valuable information about the pathogenicity mechanism of *S.* Derby.

Pan genome analysis.

Most of the genetic differences between the lineages correspond to hypothetical proteins, to mobile elements (as phage proteins or genes coding for DNA binding proteins), to membrane proteins or to secondary metabolism genes mostly associated to carbohydrate metabolism. The ST682 and ST39-ST40 also possess an important proportion of specific defense mechanisms. The ST40 and ST39 shared most of their genes in common.

Specific genomic component associated with the different lineages were identified. Most of the genes specific to a lineage was either a hypothetical protein or phage associated genes. The investigation of the phage contents of the different lineages of S. Derby could provide valuables information and should be investigate more deeply.

The ST39 is characterized by the presence of the *yehABCD* operon coding for the stc fimbriae. The SGI-1 was also confirmed as the distinctive genomic feature between the clade 1 and 2 of ST40. More interestingly the *sopD2* gene was confirmed to be a distinctive feature between the pork and human associated lineage and the poultry associated lineage ST71.

No significant feature was reported between the pork and clinical *S.* Derby strains apart for gene linked to assembly gaps, confirming the proximity of the human strains with the pork strains.

Conclusion.

The *in vitro* epithelial cells invasion assays demonstrated a hindered invasion capacity for ST71 lineage associated with poultry strains. Genomic analysis demonstrated differences in pan genome and key virulence factors between the different lineages of *Salmonella* Derby and in allelic variations.