GHENT UNIVERSITY

FACULTY OF SCIENCES

Development of a genoserotyping system for the identification of *Salmonella* serotypes

Mathieu Gand

Supervisors:
Prof. Dr. Kathleen Marchal (UGhent)
Dr. Sigrid De Keersmaecker (Sciensano)
Dr. Sophie Bertrand (FOD Volksgezondheid)

Members of the Examination Committee:
Prof. Bart Devreese, chair (UGhent)
Dr. Filip Boyen, secretary (UGhent)
Prof. Marc Heyndrickx (UGhent)
Dr. Nancy Roosens (Sciensano)
Dr. François-Xavier Weill (Institut Pasteur)

A dissertation submitted to Ghent University
in partial fulfilment of the requirements for the degree
of Doctor in Science: Biochemistry and Biotechnology
Academic year: 2019 - 2020

DEVELOPMENT OF A GENOSEROTYPING SYSTEM

FOR THE IDENTIFICATION OF *SALMONELLA* SEROTYPES

Mathieu Gand

Supervisors:
Prof. Dr. Kathleen Marchal (UGhent)
Dr. Sigrid De Keersmaecker (Sciensano)
Dr. Sophie Bertrand (FOD Volksgezondheid)

Members of the Examination Committee:
Prof. Bart Devreese, chair (UGhent)
Dr. Filip Boyen, secretary (UGhent)
Prof. Marc Heyndrickx (UGhent)
Dr. Nancy Roosens (Sciensano)
Dr. François-Xavier Weill (Institut Pasteur)

A dissertation submitted to Ghent University
in partial fulfilment of the requirements for the degree
of Doctor in Science: Biochemistry and Biotechnology
Academic year: 2019 – 2020

June 2020

SUMMARY

Salmonella is a major pathogen, widely spread and responsible of salmonellosis. It can cause different symptoms, from simple gastroenteritis to the more dangerous typhoid fever. S*almonella* can infect humans, but also animals which are its main reservoir. Humans are mostly infected through the consumption of animal-derived food products. Besides its impact on public health, another major concern of *Salmonella* is economic loss for the professional of the food sectors due to contaminated animals or food products, and economic inactivity due to sickness leave. The genus *Salmonella* is divided by a complex classification system into 2 species, 6 subspecies and more than 2500 serotypes. The severity of salmonellosis is highly conditioned by, amongst other factors, the infected host species and the serotype of the infecting *Salmonella*. Therefore, the determination of the serotype is a first key diagnostic for *Salmonella* control. Moreover, 6 serotypes and their variants, i.e. *S.* Enteritidis, *S.* Hadar, *S.* Infantis, *S.* Paratyphi B var. Java, *S.* Typhimurium including its monophasic variant 1,4,[5],12:i:- and *S.* Virchow are particularly targeted by the legislation as to be excluded from the food chain (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8), with the aim to limit their transmission to humans. Unfortunately, the classical methods for *Salmonella* serotyping, i.e. slide-agglutination and biochemical tests, are expensive, time-consuming and subjective. Therefore, highly trained and experienced technicians are required to perform these techniques which are usually only fully mastered at National Reference Centers (NRCs). Despite the fact that these techniques are implemented worldwide since more than 80 years, they are not fully adapted to the need of the field, especially for the professionals of the food sector who need to rapidly, accurately and cost-efficiently detect the serotypes targeted by the legislation. Fortunately, during these last years, molecular techniques have shown their potential as replacement method for *Salmonella* serotype identification. A wide range of different molecular technologies, based on the detection of molecular markers or on the analysis of sequencing data, are described in the scientific literature. Based on a critical review of some of these techniques, the Multiplex Oligonucleotide Reaction-PCR (MOL-PCR) & Luminex method was selected as the principle in this PhD study to develop a fast, cost-effective and accurate *Salmonella* genoserotyping system.

The first step of the new method development was to choose the serotypes to be targeted. Eighteen serotypes and their variants were selected based on their occurrence in the legislation, their clinical relevance (invasive serotypes) and their prevalence in the poultry and pork sectors in Belgium, i.e. *S.* Agona, *S.* Anatum, *S.* Brandenburg, *S.* Choleraesuis, *S.* Derby, *S.* Enteritidis including its vaccine variants AviPro SALMONELLA VAC E and Salmovac SE, *S.* Gallinarum including its variants Gallinarum and Pullorum, *S.* Hadar, *S.* Infantis, *S.* Livingstone, *S.* Mbandaka, *S.* Minnesota, *S.* Ohio, *S.* Paratyphi B var. Java, *S.* Rissen, *S.* Senftenberg, *S.* Typhimurium including its monophasic variant 1,4,[5],12:i:- and *S.* Virchow. Secondly, molecular markers specific to the targeted serotypes were

selected from *Salmonella* EnteroBase (a database with the MLST sequences of more than 230 000 *Salmonella* isolates), from the scientific literature and from genomic studies using publicly available and in-house produced Whole Genome Sequencing (WGS) data (achieving a number of 100 genomes used for comparisons) and bioinformatics tools such as Gegenees and BioNumerics. This marker selection was particularly complex for the detection of the heterogeneous population of *S.* Paratyphi B var. Java where only one suitable marker could be retrieved among more than 3 million Single Nucleotide Polymorphism (SNP) positions obtained from a genomic comparison. This valuable SNP marker was used to develop in addition a real-time PCR as an alternative method for the rapid identification of *S.* Paratyphi B and the determination of its variant Java, replacing a complex and subjective biochemical test. From this molecular markers' selection, 4 MOL-PCR assays were developed, i.e. the molecular markers were recognized by probes through a ligation-amplification reaction (MOL-PCR), followed by a capturing of the created amplicons by specific oligonucleotides coated on color-coded microspheres, which are themselves detected by a device through a fluorescence reaction (Luminex technology). Additionally, a Decision Support System (DSS), hosted by a web-application, was created for an automatic interpretation of the Luminex results with recommendations provided to the users, and for a centralization of the results in a database to improve the *Salmonella* surveillance in Belgium. The 4 modules and the DSS were validated by comparison with the classical method, including more than 1300 strains and reaching an accuracy above 99%. Finally, the complete genoserotyping system was evaluated for its ability to completely identify auto-agglutinable isolates which cannot be typed by the slide-agglutination technique.

This PhD work showed that a targeted molecular method such as the MOL-PCR & Luminex technology, even though not the most complete technique as compared to WGS, has the potential to improve the accuracy, cost- and time-effectiveness of *Salmonella* serotype identification in a routine setting. The 4 MOL-PCR assays developed here are up to 7.5 less expensive than the classical methods and they are able to completely identify, in 1 to 2 days, more than 75% of the serotypes usually encountered in Belgium. The developed genoserotyping system is complementary to WGS and an ideal workflow including both techniques was proposed for global *Salmonella* surveillance and control at a national level.

SAMENVATTING

Salmonella is een belangrijke ziekteverwekker, wijdverspreid en verantwoordelijk voor salmonellose. De bacterie kan verschillende symptomen veroorzaken, gaande van eenvoudige gastro-enteritis tot de gevaarlijkere tyfus. *Salmonella* kan mensen infecteren, maar ook dieren die het belangrijkste reservoir vormen. Mensen worden meestal besmet door de consumptie van dierlijke producten. Naast de impact op de volksgezondheid, zijn het economisch verlies voor de professional in de voedingssector als gevolg van besmette dieren of gecontamineerd voedsel, en economische inactiviteit als gevolg van ziekteverlof andere belangrijke bezorgdheden gerelateerd aan *Salmonella*. Het geslacht *Salmonella* is via een complex classificatiesysteem verdeeld in 2 soorten, 6 ondersoorten en meer dan 2 500 serotypes. De ernst van de salmonellose wordt, onder andere, sterk bepaald door de geïnfecteerde gastheersoort en het serotype van de infecterende bacterie *Salmonella*. Daarom is de bepaling van het serotype een eerste belangrijke diagnose voor de bestrijding van *Salmonella*. Bovendien zijn 6 serotypes en hun varianten, i.e. *S.* Enteritidis, *S.* Hadar, *S.* Infantis, *S.* Paratyphi B var. Java, *S.* Typhimurium met inbegrip van de monofasische variant 1,4, [5], 12: i: - en *S.* Virchow in het bijzonder vermeld in de wetgeving om van de voedselketen te worden uitgesloten (EU-verordening nr. 2160/2003, Belgisch koninklijk besluit 27/04/2007 en Belgisch FAVV nota BP-MN-FDS / LABO / 1557457 v8), met als doel hun overdracht op mensen te beperken. Helaas zijn de klassieke methoden voor serotypering van *Salmonella*, d.w.z. de glasplaat-agglutinatie en biochemische tests, duur, tijdrovend en subjectief. Daarom zijn hoogopgeleide en geëxperimenteerde technici vereist om deze technieken, die meestal alleen volledig in Nationale Referentie Centers (NRC's) beheersd zijn, uit te voeren. Ondanks het feit dat deze technieken sinds meer dan 80 jaar wereldwijd geïmplementeerd zijn, zijn ze niet volledig aangepast aan de behoeften van het werkveld, vooral voor de professionals in de voedingssector die de serotypen vermeld in de wetgeving, snel, nauwkeurig en kostenefficiënt moeten kunnen detecteren. Gelukkig hebben moleculaire technieken de afgelopen jaren hun potentieel als vervangingsmethode voor de identificatie van *Salmonella*-serotype laten zien. Een breed scala aan verschillende moleculaire technologieën, gebaseerd op de detectie van moleculaire markers of op de analyse van sequentiegegevens, zijn in de wetenschappelijke literatuur beschreven. Op basis van een kritische beoordeling van enkele van deze technieken, werd de Multiplex Oligonucleotide Reaction-PCR (MOL-PCR) & Luminex-methode als principe om in dit doctoraatswerk een snel, kosteneffectief en nauwkeurig *Salmonella*-genoserotyping-systeem te ontwikkelen, gekozen.

De eerste stap bij de ontwikkeling van de nieuwe methode was het bepalen van de serotypen die geïdentificeerd moesten kunnen worden. Achttien serotypes en hun varianten werden geselecteerd op basis van hun vermelding in de wetgeving, hun klinische relevantie (invasieve serotypes) en hun prevalentie in de pluimvee- en varkenssector in België, i.e. *S.* Agona, *S.* Anatum, *S.* Brandenburg, *S.* Choleraesuis , *S.* Derby, *S.* Enteritidis inclusief de vaccinvarianten AviPro SALMONELLA VAC E

en Salmovac SE, *S.* Gallinarum inclusief de varianten Gallinarum en Pullorum, *S.* Hadar, *S.* Infantis, *S.* Livingstone, *S.* Mbandaka, *S.* Minnesota, *S.* Ohio, *S.* Paratyphi B var. Java, *S.* Rissen, *S.* Senftenberg, *S.* Typhimurium inclusief zijn monofasische variant 1,4, [5], 12: i: - en *S.* Virchow. Vervolgens werden moleculaire merkers die specifiek zijn voor de beoogde serotypes geselecteerd uit *Salmonella* EnteroBase (een databank met de MLST-sequenties van meer dan 230.000 *Salmonella*-isolaten), uit de wetenschappelijke literatuur en via genomische studies met behulp van publiek beschikbare en eigen gegenereerde Whole Genome Sequencing (WGS) gegevens (waarbij 100 genomen werden gebruikt voor vergelijkingen) en bioinformatica-tools zoals Gegenees en BioNumerics. Deze merkerselectie was bijzonder complex voor de detectie van de heterogene populatie van *S.* Paratyphi B var. Java waar slechts één geschikte merker uit meer dan 3 miljoen posities van Single Nucleotide Polymorphisms (SNP) verkregen uit een genomische vergelijking, kon worden gevonden. Deze waardevolle SNP-merker werd gebruikt om bijkomend een real-time PCR assay te ontwikkelen als een alternatieve methode voor de snelle identificatie van *S.* Paratyphi B en de bepaling van zijn variant Java, ter vervanging van een complexe en subjectieve biochemische test. Uit de selectie van deze moleculaire merkers werden 4 MOL-PCR-assays ontwikkeld. Dit wil zeggen dat de moleculaire merkers dienden herkend te worden door sondes via een ligatie-amplificatiereactie (MOL-PCR), gevolgd door het vangen van de bekomen amplicons door specifieke oligonucleotiden gecoat op kleurgecodeerde microsferen, die zelf door een apparaat via een fluorescentiereactie (Luminex-technologie) werden gedetecteerd. Bovendien werd een beslissingsondersteunend syteem (Decision Support System, DSS), gehost door een web-applicatie, gecreëerd voor een automatische interpretatie van de Luminex-resultaten met aanbevelingen aan de gebruikers, en voor een centralisatie van de resultaten in een databank om de surveillance van *Salmonella* in België te verbeteren. De 4 modules en het beslissingsondersteunend systeem werden in vergelijking met de klassieke methode gevalideerd, waarbij meer dan 1 300 stammen gebruikt werden, resulterend in een nauwkeurigheid van meer dan 99%. Uiteindelijk werd het complete genoserotyping-systeem op zijn vermogen om auto-agglutineerbare isolaten die niet kunnen worden getypeerd door de glasplaat-agglutinatie-techniek volledig te identificeren, geëvalueerd.

Dit promotieonderzoek heeft aangetoond dat een gerichte moleculaire methode zoals de MOL-PCR & Luminex-technologie, hoewel niet de meest exhaustieve techniek in vergelijking met WGS, het potentieel heeft om de nauwkeurigheid, kosten- en tijdeffectiviteit van *Salmonella*-serotype-identificatie in een routine omgeving te verbeteren. De 4 MOL-PCR-testen die hier ontwikkeld werden, zijn tot 7,5 minder duur dan de klassieke methoden en ze zijn in staat om binnen 1 tot 2 dagen meer dan 75% van de serotypes meestal in België aangetroffen, te identificeren. Het ontwikkelde genoserotyping-systeem is complementair aan WGS en een ideale workflow waarin beide technieken gecombineerd worden, werd voor globale surveillance en controle van S*almonella* op nationaal niveau voorgesteld.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AC:	Allele call
Arsia:	Association Régionale de Santé et d'Identification Animales
CRISPR :	Clustered Regularly Interspaced Short Palindromic Repeats
CSV:	Comma Separated Value
CTRL_+ and CTRL_-:	positive and negative controls
DGZ:	DierenGezondheidszorg Vlaanderen
DNA:	DeoxyriboNucleotide Acid
dNTP:	DeoxyriboNucleotide Triphosphate
DSS:	Decision Support System
dT-:	d-tartrate non-fermenting isolates
dT+:	d-tartrate fermenting isolates
d-tartrate:	dextrorotatory [L(+)]-tartrate
eBG:	eBurst group
ECDC:	European Center for Disease prevention and Control
EFSA:	European Food Safety Authority
EU:	European Union
FASFC:	Federal Agency for the Security of the Food Chain
GPP:	Gödel Prime Product
WKL:	White-Kauffmann- Le Minor
MFI:	Medium Fluorescence Intensity
MLST:	MultiLocus Sequence Typing
MLVA:	MultiLocus Variable-number tandem Analysis
MOL-PCR:	Multiplex Oligonucleotide Ligation-Polymerase Chain Reaction
NGS:	Next Generation Sequencing
NRC:	National Reference Center
NRL:	National Reference Laboratory
NTS:	Non-Typhoidal *Salmonella*
PCR:	Polymerase Chain Reaction
PFGE:	Pulsed Field Gel Electrophoresis
PG:	Phylogenetic Group
rep-PCR:	repetitive element-Polymerase Chain Reaction
RNA:	RiboNucleic Acid

SAPE:	Streptavidin, R-Phycoerythrin Conjugate
SCV:	*Salmonella* Containing Vacuole
SISTR:	*Salmonella In Silico* Typing Resource
SNP:	Single Nucleotide Polymorphism
SNR:	Signal to Noise Ratio
SPI:	*Salmonella* Pathogenic Island
SRA:	Sequence Read Archive
ST:	Sequence Type
TAT:	Turn-Around Time
TS:	Typhoidal *Salmonella*
TTSS:	Type Three Secretion System
WGS:	Whole Genome Sequencing
WHO:	World Health Organization
XLD:	Xylose Lysine Deoxycholate

CHAPTER 1
Introduction

1.1. General context

Salmonella is a major pathogen responsible of salmonellosis, a disease which manifests as gastroenteritis, typhoid fever or even silent symptoms. This pathogen can infect both animals and humans, and is transmitted to the latter mainly through the consumption of food. At the beginning of the 20th century, *Salmonella* was responsible of huge outbreaks and infected about 3 000 people in New York between 1906 and 1907. At that time, the knowledge about transmission was not complete and it was not easy to find the origin of these outbreaks. But one famous case was reported by Georges Soper, a sanitary engineer specialized in typhoid cases, who was charged by the rich Warren family to investigate why 6 of the 11 family members living in New York suffered from typhoid fever between 27 August and 3 September 1906. He noticed that Mary Mallon, the cook of the family, had previously served in 8 families among which 7 had experienced cases of typhoid fever. Nevertheless, Mary Mallon never showed strong signs of illness due to *Salmonella typhi* (the main pathogen described as responsible for typhoid fever at that time) and refused to be considered as the cause of outbreaks. Finally, even if she was not really cooperating, Georges Soper, helped by the police, succeeded to obtain stool samples from Mary Mallon. *Salmonella typhi* was isolated from these samples and confirmed the cook as an healthy, i.e. asymptomatic carrier, spreading the dangerous invasive bacteria responsible of typhoid

fever, everywhere she worked. After 2 years of quarantining, Mary Mallon was released in exchange of her will to not work anymore with food related jobs, something that she did not respect in the following years and she continued to contaminate people. Later, she was caught again by the sanitary authorities who put her in quarantining until her death in 1938. An autopsy revealed that *Salmonella typhi* was still present in her gallbladder. At that time, no vaccine nor antibiotic treatment were available against *Salmonella* and the mortality rate was high, approaching 15%. Mary Mallon was proven responsible for the contamination of at least 122 people, including 5 deaths, and suspected of even more (Marineli et al. 2013; Soper 1939).

This famous story, known as "Typhoid Mary" **(Figure 1)**, shows how outbreak investigations, including accurate identification and characterization of the causative agent, are important key steps to trace the source of the contamination in order to confine the outbreak. More generally, the monitoring and surveillance of *Salmonella* is of major importance for public health, with the aim to limit the spreading of the bacteria and their transmission to and between humans.

Figure 1: Mary Mallon as "Typhoid Mary" in the local newspaper of that time
(Marineli et al. 2013)

Since its official report in 1884 by Dr. Daniel Elmer Salmon, a veterinary pathologist who isolated the rod-shaped bacterium from the intestines of pigs showing signs of swine fever, the classification and nomenclature of *Salmonella* has considerably evolved and its genus has been subdivided into a great number of different types and variants (Eng et al. 2015; Ryan, O'Dwyer, and Adley 2017). These different types result in different clinical manifestations and affect the hosts differently, making their identification mandatory when isolating *Salmonella* from the field. Consequently, there is a need for routine laboratories to use the best and most efficient method for the identification of the *Salmonella* types. Ideally, this method must be rapid, robust, inexpensive, easy to implement in the laboratories and resulting in objective data that can be easily shared worldwide.

1.2. The pathogen *Salmonella*

1.2.1. The complex classification structure

The genus *Salmonella* is currently partitioned in a complex taxonomic structure including species, subspecies, serotypes and variants defined by the White-Kauffmann-Le Minor (WKL) scheme (Grimont and Weill 2007; Ryan, O'Dwyer, and Adley 2017). However, not such a long time ago, this nomenclature was even more complex with all *Salmonella* serotypes believed to be different *Salmonella* species (such as *Salmonella typhi* mentioned in the Typhoid Mary story, section 1.1) until that a new nomenclature was proposed in 1987 (Le Minor and Popoff 1987) and updated the following years until arriving at the actual WKL classification which will be described in this section.

First, *Salmonella* is divided into 2 species: *Salmonella enterica* and *Salmonella bongori*. Secondly, the species *Salmonella enterica* is itself subdivided into 6 subspecies, historically numbered with roman numbers from I to VI but also named: *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb) *houtenae* (IV) and *indica* (VI). The subspecies I is more encountered in warm-blooded animals while subspecies II to VI are isolated from cold-blooded animals and the environment, even though some human cases have been reported. These species and subspecies are discriminated based on their biochemical characteristics, presented later in section 1.4.2.

Then, the subspecies are divided into serotypes counting for a total of 2 659, among which the most common 1 586 belong to the subspecies *enterica* and represent 99.5% of the isolated *Salmonella*. The serotypes (also named serovar) are determined through the characterization of 3 antigenic sites at the surface of the bacteria, i.e. the somatic antigen O and the 2 flagellar antigens (also named phase) H1 and H2, by an antigen-antibody agglutination reaction following the WKL scheme (Grimont and Weill 2007; Issenhuth-Jeanjean et al. 2014). The alphabetic or numerical codes corresponding to each serum positive for the agglutination reactions (further explained in section 1.4) are allocated to the serotype of the bacteria. These serotypes are thus defined by their somatic and flagellar antigenic formula, each separated by a colon (":"), i.e. O:H1:H2 like for instance: 1,4,[5],12:i:1,2 or 3,{10}{15}{15,34}:e,h:l,w (**Figure 2**).

To simplify these formulas, a name was attributed to all the serotypes of the subspecies *enterica*. When this antigenic classification system was first used 80 years ago, names were given to certain serotypes in relation to the disease they triggered (serovar Typhi) or the host they infected (serovar Abortus-ovis (sheep), serovar Typhi-murium (mouse) or serovar Cholerae-suis (pig)), although these clinical relationships were not always correct. Later, to avoid any confusion, serotype names were given according to the location where they were isolated for the first time. For example, the serotype 1,42:c:e,n,z15 is named *Salmonella enterica* subsp. *enterica* serovar Antwerp, in reference to the Flemish city where it was first isolated, and is abbreviated *Salmonella* Antwerp. However, the abbreviation *S.* Antwerp is also commonly found in the scientific literature (Ryan, O'Dwyer, and Adley

2017) and this type of abbreviated nomenclature will be used throughout this manuscript. Some *Salmonella* serotypes express only one flagellar antigen (e.g. *S.* Enteritidis with formula 1,9,12:g,m:–) or are non-motile (e.g. *S.* Gallinarum with formula 1,9,12:–:–).

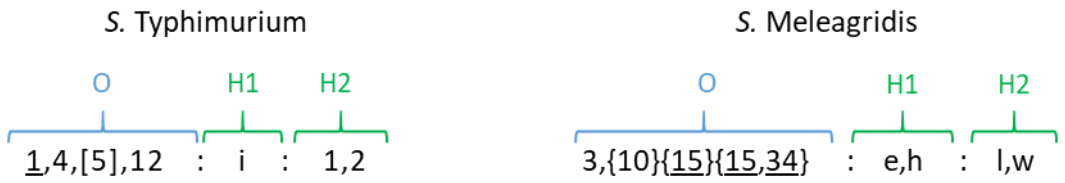

Figure 2: Antigenic nomenclature of 2 *Salmonella* serotypes.
O: somatic antigens O; H1: flagellar antigens H1; H2: flagellar antigens H2; __ : underlined antigens are determined by phage conversion, which means that they are present only if the culture is lysogenized by the corresponding converting phage; [ ] : antigens in square brackets are present or absent in the formula but without relation to phage conversion; { } : antigens in curly brackets are exclusive, which means that they cannot coexist together in the formula. For *S.* Meleagridis, the O formula can be O:3,10, O:3,15 or O:3,15,34 but never all these numbers at the same time.

Finally, some biochemical or motility characteristics are tested (further detailed in section 1.4.2) to discriminate variants inside a serotype. This is for example the case for the frequently isolated monophasic variant of *S.* Typhimurium (1,4,[5],12:i: –), which does not express the H2 antigen 1,2, or the variant Java of the serotype Paratyphi B (i.e. *S.* Paratyphi B var. Java) which has the ability to ferment the dextrorotatory L(+)-tartrate (dT) (Malorny, Bunge, and Helmuth 2003). For the latter, dT fermenting and non-fermenting *S.* Paratyphi B isolates are sometimes designated *S.* Paratyphi B dT+ and *S.* Paratyphi B dT-, respectively, in the scientific literature, including in this manuscript. Biochemical tests are also used for the discrimination between the 2 variants Gallinarum and Pullorum of the serotype Gallinarum (Christensen et al. 1992; Shivaprasad 2000).

The classification of *Salmonella* is a topic constantly in discussion in the scientific community and still evolving. The nomenclature presented above, based on phenotypic characteristics, is worldwide used since years by the public health authorities and recommended by the World Health Organization (WHO) Collaborating Centre. But with the advent of the genetic era and the use of molecular tools for the study of the *Salmonella* genome, this classification is put into question (Achtman et al. 2012). Indeed, some *Salmonella* isolates were basically clustered together based on their shared antigenic formula while the molecular analyses of their genome showed a high variability between strains of a same serotype. This is for example the case for *S.* Paratyphi B isolates clustered in the same serotype because of their shared antigenic formula 1,4,[5],12:b:1,2, while recent studies reported a high genetic diversity in this group and divided 191 *S.* Paratyphi B strains into 10 different PhyloGroups (PGs) based on the analysis of their genome (Connor et al. 2016). Similarly, for the most common serotypes such as *S.* Typhimurium and *S.* Enteritidis, a discrimination below the serotype level is sometimes required. Indeed, some

serotypes and their variants can be discriminated into subtypes using molecular subtyping methods such as Pulsed Field Gel Electrophoresis (PFGE), MultiLocus Variable-number tandem Analysis (MLVA) or Single Nucleotide Polymorphism (SNP) typing based on Whole Genome Sequencing (WGS).

1.2.2. Pathogenicity and virulence

Salmonella can cause salmonellosis by the ingestion of contaminated food or water. Consequently, this pathogen has to survive the digestive system before infecting the host intestines. For doing that, one of the remarkable characteristics of *Salmonella* is its ability to quickly adapt to harsh conditions. Indeed, these bacteria can secrete a set of protective proteins to handle the acid environment of the stomach or to survive in the intestinal mucosa composed of digestive enzymes, bile salts and anti-microbial peptides (Audia, Webb, and Foster 2001; Rychlik and Barrow 2005). Once arrived in the intestines, *Salmonella* uses fimbrial adhesins to adhere to the epithelium and 2 Type Three Secretion Systems (TTSSs), TTSS-1 and TTSS-2, to invade and survive in the host cells. The TTSS is a needle-like structure which creates a channel in the host cell membranes allowing the pathogen to inject virulence proteins, also called effector proteins, in the cytoplasm. The TTSS-1 and the associated effector proteins are encoded by genes, e.g. *invA* involved in the cytoplasmic export machinery, on the *Salmonella* Pathogenic Island 1 (SPI-1). The effector proteins injected by the TTSS-1 proceed to a rearrangement of the enterocyte actin cytoskeleton, leading to its ruffling and finally to the uptake of the bacteria. Basically, *Salmonella* regulates its own entry by hijacking host functions. Another action of these effector proteins is the induction of a pro-inflammatory response, causing the typical symptoms of gastroenteritis such as diarrhea. Once internalized by the host cell into a vacuole named the *Salmonella* Containing Vacuole (SCV), the bacteria stay safe and hidden from the immune system and can replicate. To survive in this SCV, other virulence proteins are injected in the host cell though the TTSS-2, encoded by genes of the SPI-2. This time, the role of these effector proteins is to block the fusion between the SCV and the host cell lysosome, avoiding the killing of the bacteria, and helping the latter to survive to the harsh conditions of the SCV environment, i.e. few nutrients, low pH and low Mg^{2+} and Ca^{2+} concentration. These immune escape mechanisms are also employed by *Salmonella* when this one is phagocyted by macrophages (Eng et al. 2015; Foley et al. 2013).

1.2.3. Clinical manifestation and host specificity depending on the serotype

From a clinical point of view, *Salmonella* can be divided into 2 forms depending on its pathogenicity profile in humans: typhoidal and non-typhoidal (**Figure 3**). The Typhoidal *Salmonella* (TS) are strictly specific to humans, who are their unique reservoir, and include the serotypes Typhi and Paratyphi (A, B or C). These serotypes are responsible of the dangerous typhoid and paratyphoid fevers, both known as enteric fever, that are potentially life threatening with a mortality rate estimated at 10%

without antibiotic treatment. A low dose of infecting cells is sufficient to trigger the symptoms of this disease including headache, abdominal pain and diarrhea, followed by the onset of fever which can reach 41.5°C. The TS are transmitted between humans through the ingestion of water and food contaminated by the waste of infected people. Consequently, despite the fact that TS are present worldwide, the enteric fever is mainly a problem in developing countries which do no dispose of an efficient wastewater treatment (Eng et al. 2015; WHO 2019).

All the *Salmonella* serotypes other than Typhi and Paratyphi are considered as Non-Typhoidal *Salmonella* (NTS) and are commonly found in the intestinal tract of wild and domestic animals, including livestock, that are their major reservoir. All the NTS are potentially pathogenic for humans but the infection dose must generally be higher than for TS infections. They are responsible of the less dangerous gastroenteritis, causing an inflammatory condition in the gastrointestinal tract, accompanied by symptoms such as diarrhea, vomiting, nausea, headache and abdominal cramps. NTS infections have a shorter incubation period (6-72h) than TS infections (7 days or more) and they are generally cleared by infected people in 10 days or less. But if the patient belongs to a risk population such as infants, elderly, pregnant women or immunocompromised people, they can develop more severe symptoms. Additionally, if the NTS succeed to break the intestinal barrier and enter the bloodstream, the simple gastroenteritis can evolve in bacteremia and other intestinal complications. When it happens, high fever similar to the enteric fever is observed and the immune response triggered by the bacteremia can lead to a septic shock with a high mortality rate (Eng et al. 2015; Heredia and García 2018; Jessica and Beau 2019). Almost all the *Salmonella* serotypes can trigger bacteremia, but some invasive serotypes such as *S.* Dublin and *S.* Choleraesuis are more known to be responsible of this outcome. Some studies made the hypothesis that the presence of *spv* genes, encoded on a virulence plasmid and involved in mechanisms for persistence in the host, could help these serotypes to survive in the SCV for a longer period before host cell apoptosis and thus, this explains their ability to be more invasive (Guiney and Fierer 2011; Andino and Hanning 2015).

Although *Salmonella* can sometimes colonize the intestinal tract of animals without provoking any symptoms, thus making them a healthy, asymptomatic carrier, this bacterium is also a pathogen for a wide range of animal species and is not only restricted to humans. Indeed, wild, domestic and livestock animals can develop salmonellosis with similar symptoms as for humans including gastroenteritis and enteric fever. Actually, the *Salmonella* serotypes can be clustered in 3 groups depending on their host specificity. Some of them are host-specific, which means they are strictly limited to a small number of related host species in which they will cause systemic disease. This is for instance the case for *S.* Typhi and *S.* Gallinarum which are highly invasive serotypes adapted to human and bird species, respectively.

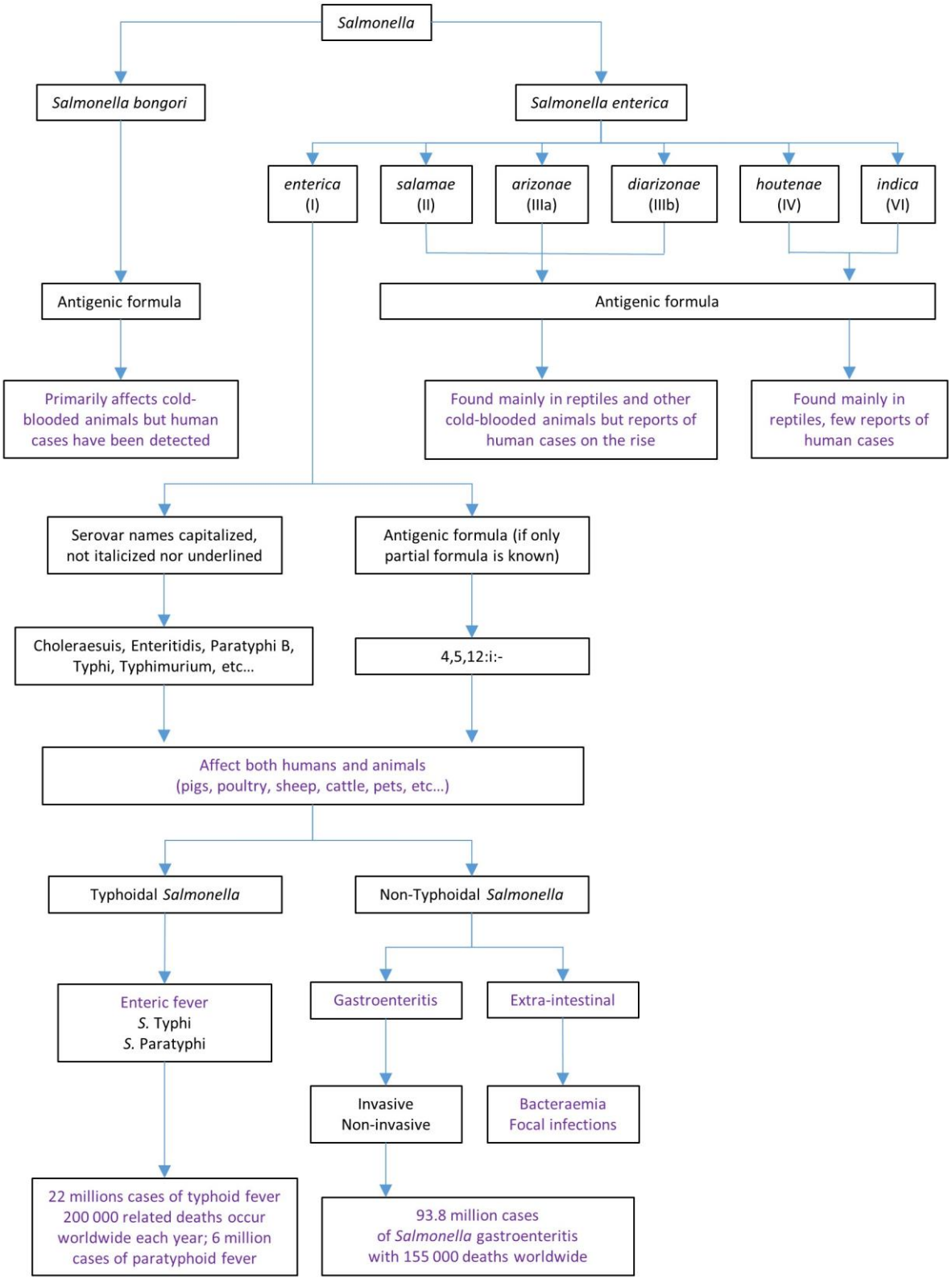

Figure 3: *Salmonella* genus nomenclature breakdown (adapted from Ryan et al. 2017)
In purple are indicated the host specificity, the resulting symptoms or disease and the related epidemiology

Additionally, the 2 variants *S.* Gallinarum var. Gallinarum and *S.* Gallinarum var. Pullorum, affect differently their hosts depending on the age of the birds. Indeed, similarly as for the risk populations mentioned earlier for humans, the variant Gallinarum triggering the fowl typhoid is more associated with adult birds while the variant Pullorum causing the Pullorum disease is more found in young birds and is even responsible of dead-in-shell chicks.

The second group of serotypes are host-restricted, which means they are usually associated with 1 or 2 closely related host species but are also sometimes able to infect other hosts. For example, *S.* Choleraesuis and *S.* Dublin are known to cause systemic disease in pigs and ruminants but can also infect humans and other species. Finally, the third and largest group is composed of serotypes which can infect a broader range of host species without a strong species specificity. Belonging to this group, *S.* Enteritidis, even though more associated with poultry, can infect humans and other species, similarly as *S.* Typhimurium which can contaminate birds, pork, cattle and humans. In poultry, these 2 non-invasive serotypes lead to low or undetectable symptoms (Demirbilek 2016; Andino and Hanning 2015).

In conclusion, the severity and the type of disease triggered by *Salmonella* is highly related to its serotype but also to the infecting dose, the strain virulence, the host species, its age and its immune status (**Figure 3**).

1.3. The importance of monitoring *Salmonella* for public health and for its economic impact

1.3.1. The global burden of Salmonella

Salmonella is a pathogen of major concern that is responsible of enteric fever and gastroenteritis worldwide. In the developed countries like those of the European Union (EU), the TS responsible of enteric fever are under control as only few cases are diagnosed each year. Indeed, only 1 161 confirmed typhoid/paratyphoid fever cases were reported in Europe in 2016 (0.33 cases per 100 000 population) mostly due to people traveling back from endemic regions and accounting for 82.5% of the cases (ECDC 2018). In Belgium, less than 1.46% of the *Salmonella* isolates serotyped by the National Reference Center (NRC) were identified in 2018 as *S.* Typhi (0.67%), *S.* Paratyphi A (0.23%) or *S.* Paratyphi B (0.56%) (NRC data). Consequently, the current study focuses more on the NTS which are one of the most common foodborne pathogens, accounting for around 93.8 million foodborne illnesses and 155 000 deaths per year worldwide (Eng et al. 2015). Moreover, it is feared that the number of deaths due to salmonellosis will increase in the future because of the emergence of multi-drug-resistant (MDR) serotypes (Nair, Venkitanarayanan, and Johny 2018). In Europe, *Salmonella* is the second cause of foodborne infections due to zoonotic agents, after *Campylobacter* infections, with 91 662 cases reported in 2017, resulting in a notification rate of 19.7 cases per 100 000 population **(Figure 4)** (EFSA 2018). This foodborne pathogen is highly associated with large outbreaks, especially during the summer season

where an increase of the cases is observed each year at this period. According to the European Food Safety Authority (EFSA), the overall economic impact of human salmonellosis has been estimated at more than 3 billion of euros per year (EFSA 2019a). This includes the cost of investigations for diagnostics, the treatment of the patients but also the decreased worker's productivity amongst others. (Oxford Analytica 2012).

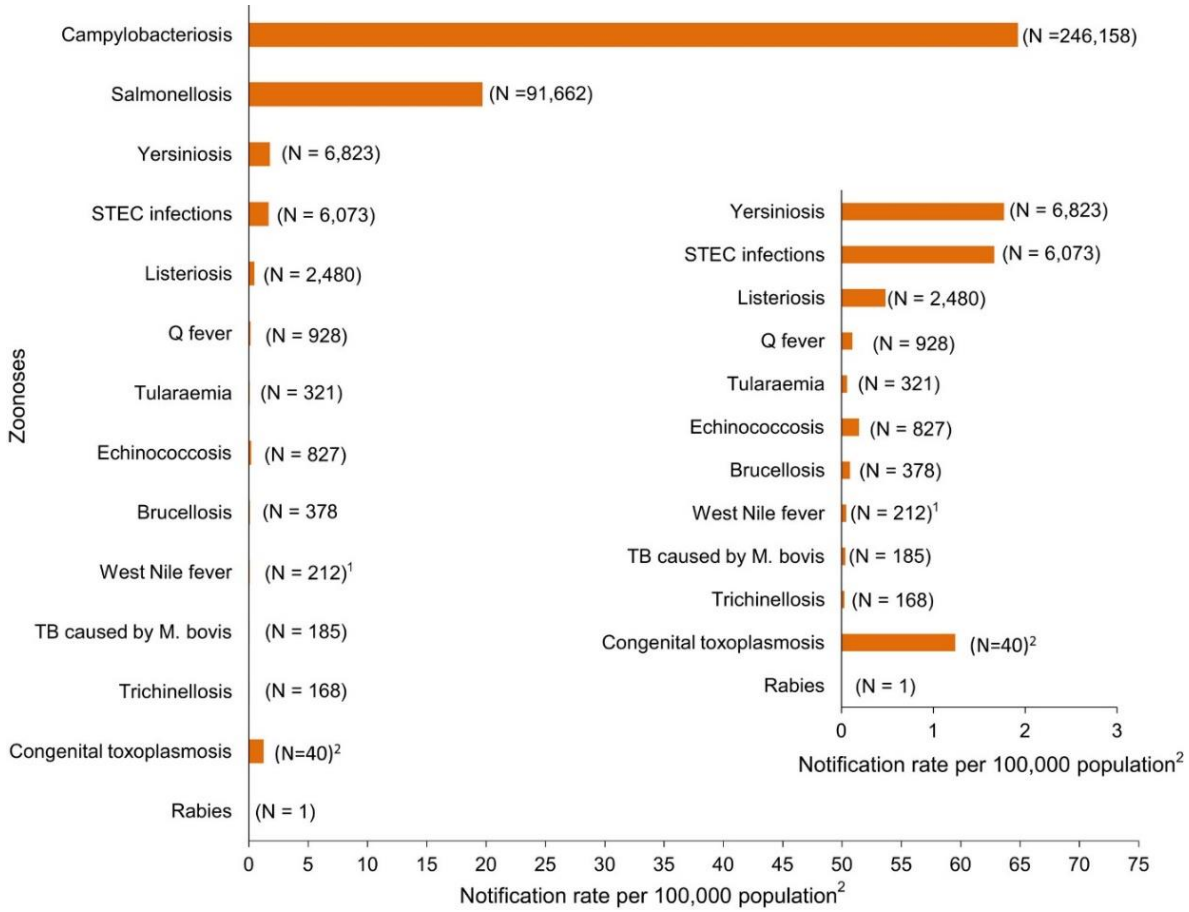

Figure 4: Reported numbers and notification rates of confirmed human zoonoses in the EU, 2017 (EFSA 2018). Note: Total number of confirmed cases is indicated in parenthesis at the end of each bar; [1]Exception: West Nile fever where total number of cases were used; [2]Exception: congenital toxoplasmosis notification rate per 100 000 live births.

The estimation of salmonellosis is largely underestimated as a lot of cases are not diagnosed nor reported (Havelaar et al. 2013; Mellou et al. 2013; Mølbak et al. 2014). Indeed, a fraction of the infected people develop mild symptoms (or even silent symptoms), they have the knowledge that the body can cure the disease by itself or they cannot afford the cost of a medical consultation, and thus they do not seek healthcare and are hence not captured by the surveillance system. Some other cases are unreported because the infected people attend healthcare but the infection is not diagnosed or misdiagnosed and the pathogen is not isolated. Indeed, to be captured by a laboratory-based surveillance system, a sick individual must (1) seek for a medical consultation, (2) have a sample (stool, urine or blood) requested,

(3) submit a sample for testing, (4) the causative agents must be properly isolated and identified by a valid laboratory method and finally (5) the positive result must be reported to the surveillance system (Thomas et al. 2013). The surveillance system is managed by the NRCs which also master the reference techniques for identification and characterization. The multiplication factor, which is the ratio between the symptomatic cases and the number of cases reported to the surveillance system, was determined to be 20 for non-typhoidal salmonellosis in France (**Figure 5**) (Van Cauteren et al. 2015). This means that for every case reported to the NRC, 20 other cases occurred and are not diagnosed nor reported. But it must be mentioned that salmonellosis has not the obligation to be reported to the competent authorities in all the EU countries. For instance, in Belgium, France, Luxembourg and Spain the reporting of human salmonellosis is made on a voluntary basis, and in the UK the isolated pathogen is reported rather than the disease (Gibbons et al. 2014).

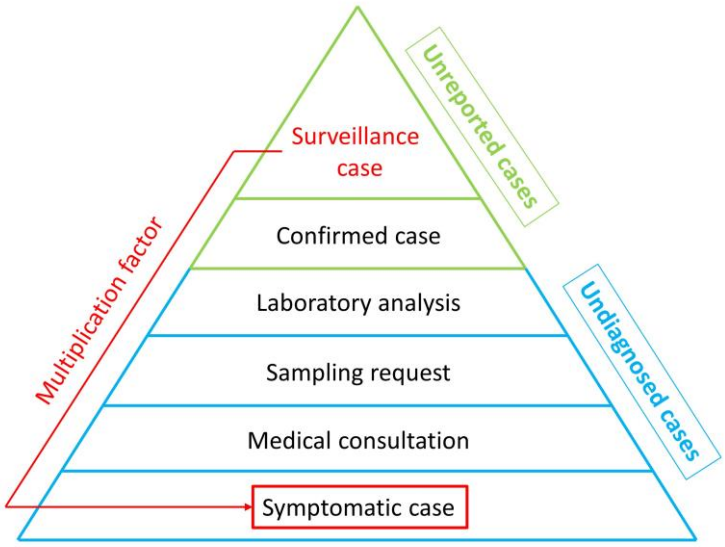

Figure 5: Under-reporting and under-diagnostic factors (adapted from Van Cauteren, 2016)

1.3.2. Salmonella *in food-producing animals*

Humans are mostly infected by *Salmonella* through the consumption of contaminated food products as diversified as animal meat, animal products (including eggs), dairy products (including cheese), sweets, chocolate, vegetable or seafood products. In 2017, the most incriminated food vehicles responsible of *Salmonella* outbreaks were: eggs & egg products (36.8%), bakery products (16.7%) and meat & meat products (8.2%) (EFSA 2018). Indeed, dishes prepared with contaminated raw eggs, egg products and insufficiently heated poultry meat and pork are concerned. Consequently, it is highly important to monitor *Salmonella* in the food-producing animal sector, first for the protection of the consumers but also for the maintenance of animal productivity. Indeed, when speaking about *Salmonella* infections in farm animals composed for instance of poultry, pork and cattle, 2 patterns can be described.

At the one hand the livestock can be infected by host-related invasive serotypes like *S.* Choleraesuis in pork or *S.* Gallinarum in poultry, leading to the manifestation of symptoms such as diarrhea or fever, and resulting in a high mortality rate. In these cases, the livestock contamination will be more easily noticed and one of the concerns of the breeder will be the economic loss linked to the decrease in productivity (weight loss, abortions, milk production, treatment of contaminated eggs, etc…) in addition to the quarantining and medical treatment of diseased animals (Evangelopoulou et al. 2015; Oxford Analytica 2012). At the other hand, the livestock can be contaminated by other unrestricted host serotypes, such as *S.* Enteritidis or *S.* Typhimurium, which will trigger limited or silent symptoms not easily noticeable by the farmers without specific control. If salmonellosis symptoms can be detected and the disease is properly diagnosed, this will also have an economic impact for the breeders as elaborated above. But if the symptoms are not noticed, the infected animals risk to not be excluded from the food chain and they will be able to potentially transmit *Salmonella* and cause salmonellosis to humans.

Despite a constant improving of hygiene and disease control in the food sector, pork, poultry and associated products such as eggs, are still the main source of *Salmonella* transmission to humans through food consumption (Pires, de Knegt, and Hald 2011). Many factors could be involved in livestock contamination. The presence of wild animals, rodents or insects in the environment of the farm can be vectors of transmission as well as the handling by humans in slaughtering, food processing and storage procedures, if the hygiene standards are not carefully respected (Sofos 2008). The world globalization has oriented the consumer habits into more protein in the diet and thus the increase of animal products consumption. Some studies project that this kind of consumption can achieve 376 millions of tons in 2030 (Dhama et al. 2013). Such a demand conduces to an intensive animal production, with a more complex management of the food safety, increasing the risk of defective processing practices and contamination by foodborne pathogens at multiple points from the farm to the fork (Heredia and García 2018).

1.3.3. The surveillance and monitoring of Salmonella *serotypes*

For a better understanding of the epidemiology of *Salmonella* and the limitation of its transmission to humans, the determination of the species and subspecies is not sufficient. Indeed, there is a need to go under the subspecies level and for that the serotyping provides helpful information and this is the basis of all *Salmonella* surveillance programs. First, when isolating *Salmonella*, it is important to know whether it belongs to an invasive serotype potentially dangerous for the host and requiring a medical treatment or a quarantining for farm animals. Secondly, the determination of the serotype is the initial point to start outbreak investigations and trace the source of the contamination, even though the subtyping of the strains (i.e., typing below the serotype level) is usually additionally required (Jourdan-da Silva et al. 2018; Leekitcharoenphon et al. 2019; Pijnacker et al. 2019). The outbreak traceability

allows the professional of the food sector to conduct efficient disinfection procedures where it is effectively needed but also to perform corrective actions and identify new outbreak vehicles, with the aim to avoid further contaminations (Angelo et al. 2015; Gambino-Shirley et al. 2018; Mba-Jonas et al. 2018). With the globalization and the world exchanges, the *Salmonella* contaminations are most of the time not restricted to a state or a country. The serotyping of *Salmonella* strains provides an international language used since many years by laboratories to easily share data during multi-state outbreaks. Finally, the surveillance of the circulating serotypes and the evaluation of those that are the most involved in outbreaks help the food authorities to establish what are the serotypes to combat in priority. Indeed, even if more than 1 500 serotypes exist in the *Salmonella enterica* subspecies *enterica*, less than 100 account for most human infections (CDC 2019) and only 5 represented more than 70% of the European human cases in 2017, including *S.* Enteritidis (49.1%), *S.* Typhimurium (13.4%), monophasic *S.* Typhimurium (1,4,[5],12:i:-) (8%), *S.* Infantis (2.3%) and *S.* Newport (1.2%). The surveillance of the prevalent serotypes gives also information about their preferred sources. When isolated from food or animal sources, *Salmonella* is preferably found in broiler flocks (67.2%), broiler meat (11.3%), turkey flocks (6.5%), laying hens flocks (6.4%), pig meat (4.0%), cattle (1.6%) and pigs (1.4%) (**Figure 6**) (EFSA 2018).

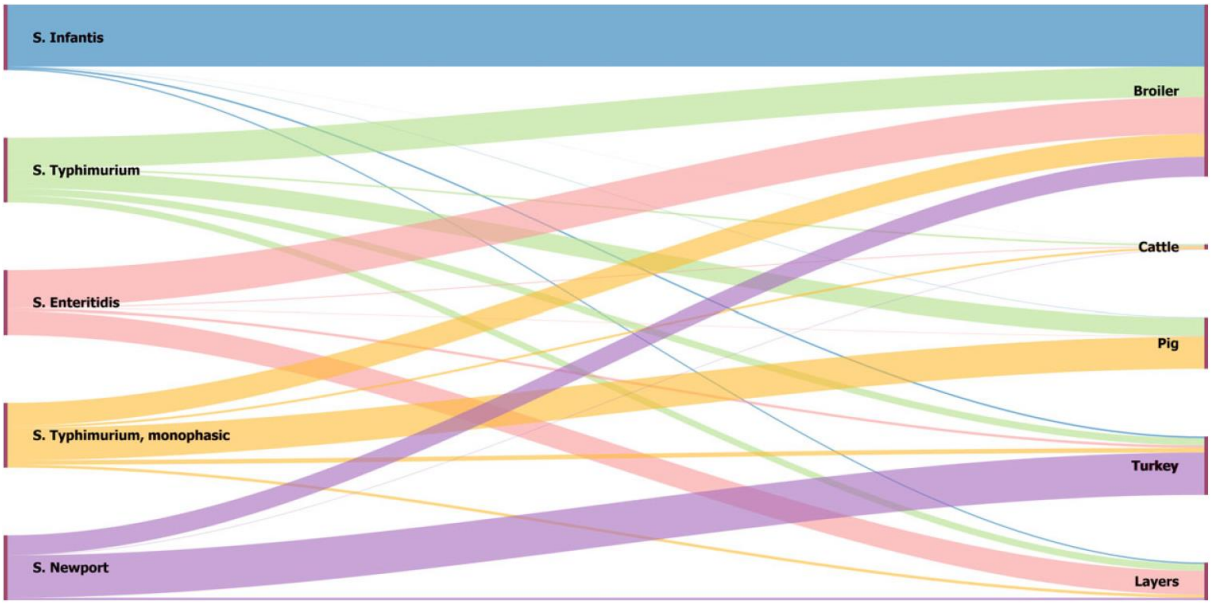

Figure 6: Sankey diagram of the distribution of the EU top-five *Salmonella* serotypes in human salmonellosis acquired in the EU, across different food and animal sources (broiler, cattle pig, turkey and layers), by source, EU, 2017 (EFSA 2018)

But it must be noticed that the composition of the top 5 most prevalent serotypes can slightly vary between member states of the EU. In Belgium, the surveillance of *Salmonella* is performed by the National Reference Laboratory (NRL) and the NRC, which are part of an international network through the European Center for Disease prevention and Control (ECDC) and the WHO. Briefly, the Belgian

NRC performs the serotyping of *Salmonella* samples isolated from human, food or animal sources by first-line laboratories and sent on a voluntary basis. The samples coming from food and animals are collected by the Federal Agency for the Safety of the Food Chain (FASFC) and the veterinary laboratories (DGZ for Flanders and Arsia for Wallonia), respectively. Samples can also be sent to the NRL by private companies. *Salmonella* isolation is made directly by these laboratories or by the NRL after transport of the samples. The isolates are then transferred to the NRC for complete identification. Finally, isolates from human samples are sent by hospitals or clinical laboratories directly to the NRC (**Figure 7**).

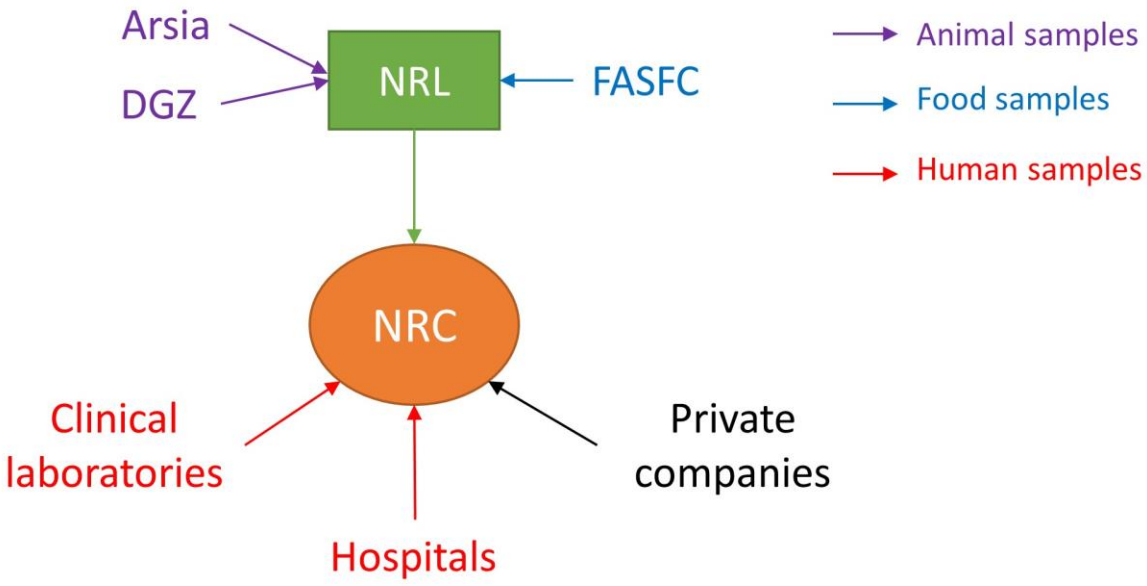

Figure 7: Origins of *Salmonella* isolates sent to the NRC for serotyping and surveillance.

The Belgian NRC reported that the 5 most prevalent serotypes isolated from human cases were in 2017-2018: *S.* Typhimurium (24%), monophasic *S.* Typhimurium (1,4,[5],12:i:-) (19%), *S.* Enteritidis (17%), *S.* Infantis (2%) and S. Derby (2%). As for the *Salmonella* isolated from animal and food sources in the same period, they belonged mainly to *S.* Infantis (31%), *S.* Paratyphi B var. Java (9%), *S.* Enteritidis (8%), monophasic *S.* Typhimurium (1,4,[5],12:i:-) (7%) and *S.* Typhimurium (6%) (**Figure 8**). *S.* Enteritidis was historically known to be the most prevalent serotype in Belgium in the beginning of the 2000s with a high number of reported cases. But the introduction at a national level of an obligation to vaccinate layer flocks with live attenuated strains of *S.* Enteritidis drastically decreased the incidence of human salmonellosis and foodborne outbreaks due to this serotype (Collard et al. 2008). The inoculated vaccine strain of *S.* Enteritidis triggers an immunisation of the animals but without persisting in the flocks because of drifting mutations in its metabolic genes.

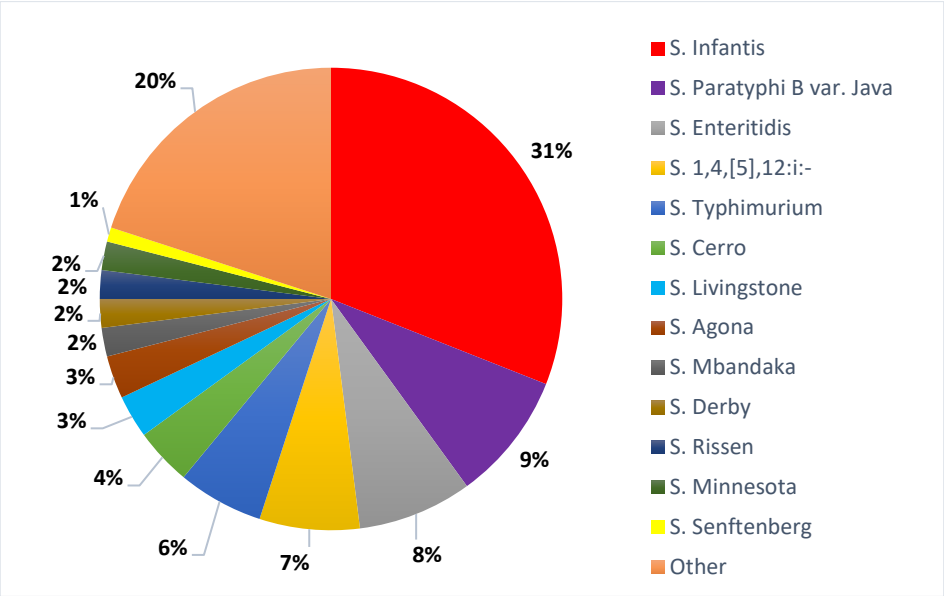

Figure 8: Prevalence of the *Salmonella* serotypes isolated from food and animal samples in Belgium between 2017 and 2018 (Data NRC)

1.3.4. The Salmonella *serotypes subjected to an official control in the food sector*

The high prevalence of NTS infections in Europe conducted the European and member state authorities to set in place a strong monitoring of this pathogen to manage the risks from farm to retail, especially in the poultry sector, and to protect the consumer health (Antunes et al. 2016; Martínez-Avilés et al. 2019; Pires, De Knegt, and Hald 2011). Indeed, to prevent *Salmonella* infections, surveillance program for the timely detection of this pathogen along the food chain (animal feed, living animals, slaughterhouses, retail sector, and restaurants) and appropriate sanitary measures are required (Bertrand et al. 2016). Consequently, the EU regulation N°2160/2003 was created in 2003, and updated in the following years, with the aim to reduce the maximum percentage of adult breeding flocks of *Gallus gallus*, laying hens and broilers, positive for some target *Salmonella* serotypes, to 1% or less. These target serotypes were defined based on their public health significance taking into account the criteria described in Annex III of Regulation (EC) No 2160/2003: (a) the most frequent *Salmonella* serotypes associated with human salmonellosis; (b) the route of infection; (c) whether any serotype shows a rapid and recent ability to spread and cause disease in humans and/or animals; and (d) whether any serotype shows increased virulence, e.g. regarding invasiveness or resistance to relevant therapies for human infections. These criteria, included in the Belgian regulation (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8), are rather complex as different serotypes are targeted and different corrective measures are applied depending on the breeding type. For food

producing animals (i.e. broilers and laying hens), the serotypes subjected to an official control through this regulation are *S.* Enteritidis and *S.* Typhimurium including its monophasic variant 1,4,[5],12:i:-. These 3 targets are also subjected to an official control in adult breeding flocks of *Gallus gallus*, in addition to *S.* Hadar, *S.* Infantis, *S.* Virchow and *S.* Paratyphi B var. Java. Although the latter was not initially listed in the EU regulation N°2160/2003, it was included in the Belgian law because particularly prevalent in Belgium. When these targeted serotypes are isolated in poultry farms, some corrective measures must be taken. For instance, for laying hens and breeding flocks, it is stated in the regulation that if one of their respective targeted serotypes is detected, the entire flocks must be eliminated by logistic slaughtering (at the end of egg-laying for laying hens), strict disinfection of the farm must be performed and absence of any serotype of *Salmonella* has to be proven before the installation of a new animal batch. Additionally, eggs coming from laying hens and intended for human consumption are restricted to food applications including a thermic treatment. Concerning hatching eggs from adult breeding, they must be eliminated when the eggs are already incubated; if they were not already incubated they can, besides elimination, also be transferred to food processing for human consumption, when including a heat treatment (FASFC circulars PCCB/S2/418588 and PCCB/S2/409035). For broilers, logistic slaughtering is based on the presence of *Salmonella* spp. but serotyping is performed to instruct hygiene measures at the farm (FASFC circular PCCB/S2/589616). As it is mandatory to vaccinate adult breeding flocks and laying hens of *Gallus gallus* with a live attenuated strain of *S.* Enteritidis (FASFC circulars PCCB/S2/418588 and PCCB/S2/409035), it is sometimes required to make the discrimination between the wild-type and vaccine strains, when isolating this serotype, if the vaccination is too close to the sampling period.

Salmonella official controls in the pork sector were in place in Belgium until 2015, but they were discontinued afterwards because too constraining for the breeders and with no real impact on the *Salmonella* prevalence in pork breeding. Therefore, these last years the *Salmonella* control measures in the pork sector were more focused on global hygiene improvement along the food chain and evaluation of *Salmonella* vaccine efficiency. However, the surveillance and monitoring of the *Salmonella* serotypes circulating in pork farms is still strongly recommended by the competent authorities.

1.4. The classical methods for *Salmonella* serotyping

As elaborated above, in Belgium, *Salmonella* is isolated from human, food and animal sources by the clinical laboratories, the NRL or the veterinary laboratories (DGZ for Flanders and Arsia for Wallonia), respectively. Then, when required, the isolates are sent to the NRC for further characterization including serotyping (**Figure 7**).

1.4.1. Detection and isolation of Salmonella

Salmonella can be isolated from various samples such as stool, blood, urine, pus, sputum or other body fluids for humans, feces and meat for animals, portion of food for food products, and from the environment. A part of the sample can be directly cultured on differential and selective media but an enrichment step is usually performed prior to selective isolation. For instance, for animal and food samples, a pre-enrichment is performed by homogenizing 25 g or 25 ml in 225 ml of buffered peptone water and incubating the suspension for 16-18 hours at 37°C. Then, selective enrichment is operated by inoculating 0.1 ml and 1 ml of the pre-enrichment to 10 ml of Rappaport-Vassiliadis with soja broth and Muller-Kauffmann tetrathionate broth, and incubating for 24h at 41.5°C and 37°C, respectively. A full loop of each enrichment is subsequently inoculated on 2 selective media, Xylose Lysine Deoxycholate (XLD) agar and *Salmonella* ID2 agar, and incubated at 37°C for another 24 h. One colony of each plate with a characteristic aspect is subcultured on Kligler Iron agar for further confirmation analyses including biochemical tests or Maldi-TOF mass spectrometry. If the picked colony cultured on Kligler Iron agar is negative, 4 other colonies are tested from the XLD and *Salmonella* ID2 agar plates (ISO 6579).

1.4.2. Salmonella *serotype identification by slide-agglutination and biochemical tests*

The classical method for *Salmonella* serotyping is based on the determination of its somatic (O) and flagellar (H1 and H2) antigens by slide-agglutination and biochemical tests, following the WKL scheme (ISO 6579). For the slide-agglutination, the test consists of putting into contact on a slide a drop of serum with a small loop of bacteria from a culture on Kligler Iron agar. After mixing the reagents, the presence of an agglutination reaction is observed on a black background with a magnifying glass **(Figure 9)**. The agglutination must be visible in less than 60s, if not, it is determined as negative.

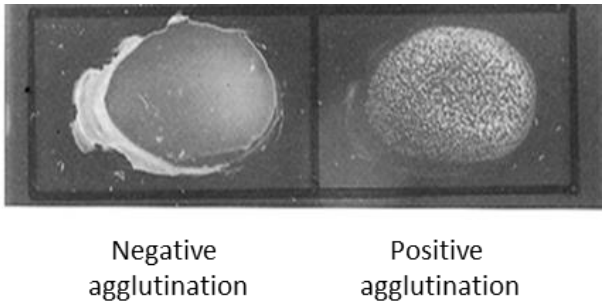

Figure 9: Serotyping by slide-agglutination

Prior to the reaction with sera, the isolate is tested for its auto-agglutinable potential. To do this, a drop of distilled water is used instead of the serum and no agglutination reaction must be observed. If an agglutination is seen with water, the *Salmonella* strain is determined as auto-agglutinable and cannot be typed by slide-agglutination. Once the sample has been confirmed as typable, the operator first tests polyvalent sera composed of a mix of antibodies targeting several O antigens: OMA, OMB, OMC, OMD, OME, OMF and OMG. For instance, the serum OMA targets the following O antigens: 1,2,12 + 4,5,12 + 9,12 + 9,46 + 3,10 + 3,15 + 1,3,19 + 21. If the isolate is positive for one of these polyvalent sera, the technician tests the corresponding monovalent sera individually. When an agglutination is obtained, the numeric code of the serum used for this positive reaction is reported in the antigenic formula of the sample. It is not mandatory to find all the O antigens included in the antigenic formula of a serotype for its determination. Indeed, only some major O antigens of each antigenic formula are needed. For example, in the formula 1,4,[5],12:i:1,2, the determination of O:12 is not mandatory because this antigen is present in the formula of a large number of serotypes and is thus not discriminative. Secondary O antigens are sometimes required for the determination of variants. In the previous example, the absence of the antigen O:5 is used for the discrimination of the variant Copenhagen from the serotype Typhimurium, i.e. 1,4,12:i:1,2 gives *S.* Typhimurium var. Copenhagen.

The next step is the determination of the H antigens. The same protocol as described above is used. Again, the laboratory operator uses first polyvalent sera, targeting this time several H antigens: HMA, HMB, HMC, HMD and HM3. Then, monovalent sera, corresponding to the positive polyvalent serum, are tested. In some cases, both H1 and H2 antigens can be determined by this way, but usually, only one of the 2 flagellar antigens is determined and another culture on a soft solid media supplemented with sera targeting the obtained H antigens must be performed. The role of this supplement is to block the expression of the previously determined H antigen and force the bacteria to express the second H flagellar phase.

The combination of the O and H formulas gives the antigenic characteristics of the *Salmonella* isolate. If the strain belongs to the subspecies I, a name is linked to this result in the WKL scheme (Grimont and Weill 2007). Sometimes a common antigenic formula is shared by several subspecies of *Salmonella enterica*. In this case, additional tests are used for the discrimination between these subspecies, based on their biochemical characters (**Table 1**). Identically, biochemical tests are also required for the determination of some variants. This is for instance the case for the variant Java of the serotype Paratyphi B for which the ability to ferment the d-tartrate can be determined by 2 cultural methods. One is the historical lead-acetate test which consists of incubating the *S.* Paratyphi B isolate in a minimum broth supplemented with d-tartrate as source of energy. After 7 days at 37°C, the use of d-tartrate is measured by addition of lead-acetate to the culture. The lead-acetate has the ability to form a precipitate in the presence of d-tartrate: if a small precipitate is formed, the bacteria used all the d-tartrate present in the broth (dT+); if a fluffy fine precipitate is created, the initial amount of d-tartrate

is still present and was not used by the bacteria (dT-). The second more recent cultural method is the commercial Jordan's Tartrate test which consists of inoculating the bacteria on a solid medium containing d-tartrate and observing after 24h at 37°C, a color change of the culture, from red to yellow, if fermentation of d-tartrate has occurred (dT+).

Table 1: Differential characters of *Salmonella* species and subspecies (adapted from Grimont and Weill 2007)

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

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

A collection of more than 120 sera is needed for being able to identify all the antigenic formulas composing the WKL scheme. The choice of sera to test and the ability to properly detect an agglutination reaction without mistake, is highly correlated with the training and the experience of the laboratory technicians. Therefore, the serotyping by slide-agglutination is expensive, labor-intensive and not fully accurate as it is linked to the subjectivity of the technician’s interpretation. In Belgium, this method is only fully mastered in the NRC.

1.5. Alternative molecular methods for *Salmonella* serotype identification

With the advent of the genetic era, a plethora of molecular methods were developed for the genoserotyping of *Salmonella*, as valid alternatives to the classical serotyping using slide-agglutination and biochemical tests. Basically, 2 types of approaches are generally used: sequencing and non-sequencing based techniques, which are often targeted methods. The latter use the detection of a single or a combination of molecular markers, specific to a given serotype or variant, by PCR or hybridization based methods. These molecular markers are sometimes linked to the genes which code for the somatic and flagellar antigens, allowing the deduction of the antigenic formula and thus the serotype, or to specific DNA regions only present in the genomes of the targeted serotype. The sequencing methods are based on the analysis of a part (e.g. MLST genes or CRISPR loci) or the totality of the bacterial genome which allows the deduction of the serotypes and associated variants.

It needs to be mentioned that some molecular subtyping (i.e. below the serovar level) methods can sometimes be used for serotyping purposes. This is for instance the case for the repetitive element-PCR (rep-PCR) techniques using primers to amplify non-coding repetitive sequences interspersed throughout bacterial genomes, which are subsequently separated by electrophoresis. This technique showed its high potential for serotype prediction from the generated genomic fingerprints (Rasschaert et al. 2005; Wise et al. 2009). Another example is the PFGE method which consists of the cutting of the bacterial genomic DNA into large fragments using a restriction enzyme and their subsequent separation by gel electrophoresis with a pulsed electrical field. The result is a pattern of bands (PFGE profile or pulsotype), similar to a barcode and relatively conserved between bacterial clones. Some parts of the band pattern showed to be specific to some serotypes. The deduction of the serotype is made after comparison with other PFGE profiles of the laboratory's proper database (or the international standardized database PulseNet) composed of strains for which the serotype identification is known. Although initially used for subtyping, this method proved to be reliable for the identification of *Salmonella* serotypes (Bopp et al. 2016; Kérouanton et al. 2007; Zou et al. 2010).

Some efforts were also made to develop alternative serotyping methods based on mass spectrometry, such as Matrix-Assisted Laser Desorption Ionization-Time Of Flight (MALDI-TOF) (Bell et al. 2016). Based on the measurement of proteins' mass from intact bacterial cells or cell extracts, MALDI-TOF is able to generate reproducible patterns from the obtained spectra. Being rapid, cost-effective, user-friendly and using standardized protocols as well as reference spectra libraries, MALDI-TOF is the method of choice for *Salmonella* spp. identification by first-line laboratories. Even though recent advances have shown that this method might be used for *Salmonella* identification below the genus and species level (Dieckmann and Malorny 2011), this requires the extension of the upper mass range of detection from 20 000 Da up to 40 000 Da, which is not supported by all the commercial

methods. Additionally, while common MALDI-TOF tests are based on the detection of ribosomal proteins, referenced in standardized databases, serovar-specific combinations of several non-ribosomal proteins were used as markers in the protocol of Dieckmann and Malonry (2011), which would require the establishment of new standardized reference libraries for potential future use. Finally, Kang and colleagues failed to reproduce these results and concluded, based on their own data, that MALDI-TOF was very limited for *Salmonella* identification at the serotype level (Kang et al. 2017).

In conclusion, the molecular methods are the most reliable alternatives for the identification of *Salmonella* at the serotype level. Some of these methods will be reviewed in this section and their advantages and drawbacks will be discussed (**Table 2**).

1.5.1. PCR-based serotyping methods

PCR and detection by electrophoresis on agarose gel

It is not necessary anymore to present this famous and widely used technique consisting of the specific amplification of a DNA region with a pair of forward and reverse primers, and the subsequent detection of the amplicons by electrophoresis on an agarose gel. Some singleplex PCR assays using one pair of primers were developed for the specific detection of one serotype, e.g. *S.* Enteritidis (Lampel, Keasler, and Hanes 1996) or *S.* Typhimurium (H. J. Kim et al. 2006). In this case, if a specific band is observed on the agarose gel, this means that the related serotype was detected. Other duplex or multiplex assays are composed of at least 2 pairs of primers and aim to amplify several DNA regions simultaneously. This time, the amplicons are distinguishable based on their size on the agarose gel after separation by electrophoresis. This is the case for the duplex PCR methods developed by Zhai et al. (2014) and Malorny, Bunge, and Helmuth (2003) for, respectively, the detection of *S.* Paratyphi B and the discrimination between its variant dT- and dT+. These 2 duplex methods are composed of a pair of primers to detect their respective targets, but also of a second pair of primers for the simultaneous confirmation of the *Salmonella* genus, each set of primers generating amplicons with different sizes. Of the same principle, some multiplex PCR methods were developed for the identification of genes coding for the somatic and flagellar antigens (Echeita et al. 2002; Herrera-León et al. 2004; Liu et al. 2011; Luk et al. 1993) or other genomic regions specific to the serotypes (Alvarez et al. 2004; S. Kim et al. 2006; Laetitia Fabre et al. 2014). Despite the fact that PCR is a cheap universal method commonly implemented in laboratories worldwide, a detection by electrophoresis on agarose gel is needed in most cases. Additionally, this technology does not allow a high level of multiplexing and several assays are usually needed for the identification of numerous targets, which can be time-consuming and labor-intensive.

Real-time PCR, i.e. detection using fluorescence

The principle of the real-time PCR (also called qPCR) is similar to the PCR but here the presence of molecular markers is detected in real time through the measurement of fluorescence produced at each amplification step. Two main technologies are used: SYBR green (intercalating dye) and TaqMan (probe based). The principle of the nonspecific fluorescent dye SYBR is that when SYBR dyes bind to double-stranded DNA, its fluorescence increases by 20-100 fold. As the amount of double-stranded DNA increases during the PCR process, the SYBR fluorescent signal increases correspondingly. The TaqMan assay is more widely used because it is more sensitive and specific. It is based on the Taq DNA polymerase 5′–3′ exonuclease activity, and the TaqMan probe which is a sequence-specific oligonucleotide with a reporter fluorescent dye at its 5′ end and a quencher dye at its 3′ end. When the probe is not hydrolyzed by the Taq DNA polymerase, the reporter dye emitted fluorescent light is absorbed by the quencher dye because of fluorescent resonance energy transfer. When the probe is hydrolyzed by the Taq DNA polymerase, the 5′ reporter dye is separated from the quencher dye. Therefore, the quenching effect is gone and the 5′ reporter dye fluorescent light is able to be detected by the qPCR instrument (Jia 2012). Singleplex and multiplex assays using this technology were developed for a fast and accurate detection of various *Salmonella* serotypes (e.g. *S.* Dublin, *S.* Enteritidis, *S.* Gallinarum, *S.* Kentucky, *S.* Paratyphi A *S.* Typhi and *S.* Typhimurium; Farrell et al. 2005; Hwa Lee et al. 2009; Laetitia Fabre et al. 2014; O'Regan et al. 2008) or discrimination between 2 variants of a same serotype (e.g. vaccinal variant of *S.* Enteritidis; Maurischat et al. 2015). Despite the fact that this detection method is commonly used, very fast and easy to perform, the number of targets per assay is very low due to the limited number of available spectra used by the fluorophores for detection.

1.5.2. Hybridization-based serotyping methods

Microarray

Microarrays have been commonly used in gene expression studies (RNA) but they also allow to easily detect in one shot a large numbers of molecular markers at the DNA level. A DNA microarray is composed of a multitude of oligonucleotides representing different alleles of target genes. Small regions of these genes are amplified by multiplex PCR and a reporter dye is incorporated into the amplicons. Depending on the allelic sequence of these amplified genes, they specifically hybridize to the oligonucleotides coupled at the microarray and are subsequently detected through fluorescence. This method is mostly used for the determination of the somatic O and flagellar H antigens based on their coding sequence (Braun et al. 2012; Guo et al. 2013; Robertson, Yoshida, Gurnik, et al. 2018; Yoshida et al. 2007). The detection by microarray has the advantage to detect a large number of molecular markers in one run but most of the time, several multiplex PCRs are needed before hybridization to the array. Additionally, the analysis of a high number of samples in high-throughput is limited because one

array is required for each sample to identify its serotype. A commercial kit based on this technology and named Check&Trace *Salmonella* claims the possible identification of 300 serotypes. However, this kit is expensive and its protocol is performed using separate tubes, each containing one array, analyzed one by one by the Check&Trace reader. Therefore, if multiple isolates need to be identified, all the reaction and reading steps have to be performed for each isolate, each in a separate tube. In these conditions, the method becomes labor-intensive and time-consuming when many isolates have to be serotyped.

Multiplex Oligonucleotide Ligation-PCR (MOL-PCR) and Luminex xTAG technology

The Luminex technology is a multiplex detection platform using a liquid bead-suspension array. The beads in this array are polystyrene microspheres with a diameter of 5.6 to 6.5 µm and which are divided into different spectral set depending on their combination of red and infra-red colors. Thanks to this unique spectral address, the Luminex device is able to categorize each microsphere individually in a mixture of microspheres of different regions. With the xTAG technology, the beads are magnetic MagPlex-TAG microspheres that are pre-coupled with a specific anti-TAG DNA sequence linked to the unique bead color. These anti-TAG sequences are composed of a 24 bp oligonucleotide with minimal cross hybridization (no C, only T, A and G) and they all hybridize with their complementary TAG sequence at 37°C. In the Multiplex Oligonucleotide Ligation-PCR (MOL-PCR) assay, specific molecular markers are detected through a multiplex ligation-dependent reaction followed by a singleplex (universal) amplification reaction. The multiplex ligation reaction is working with sets of probe pairs: upstream probes composed of a 5' universal primer site (e.g. T7), an internal anti-TAG sequence (further used for the Luminex detection) which is unique for each target and a 3' target-specific sequence; and downstream probes including a 5' target-specific sequence and a 3' universal primer site (e.g. T3) **(Figure 10)**.

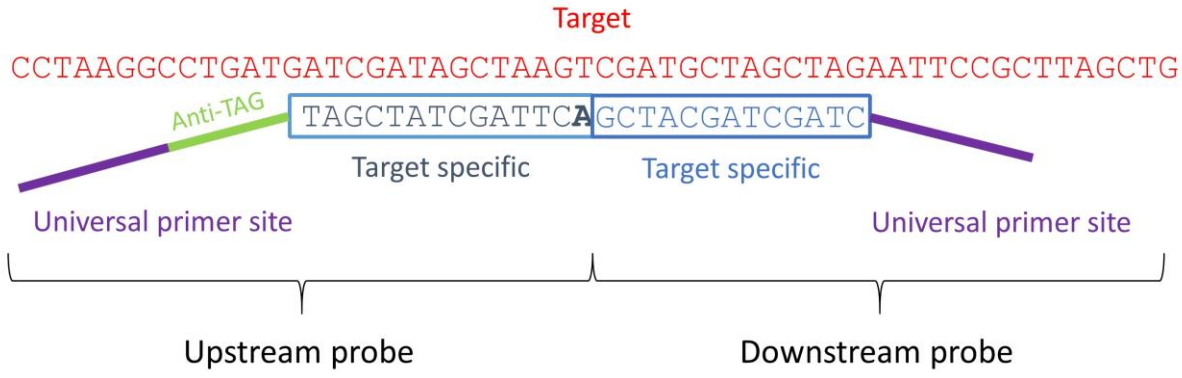

Figure 10: Architecture of the ligation probes annealing close to each other on their target site. Target specific: sequence of the probe which anneals to the target site of the molecular marker; anti-TAG: unique sequence used later for hybridization to MagPlex-TAG microspheres; universal primers site: sequences used by universal primers (e.g. T3 and T7) for the singleplex PCR step.

For a specific detection of molecular markers, upstream and downstream probes must anneal adjacent to each other on their target site (**Figure 10**), for being linked by the ligase enzyme during the first step of the method: the multiplex ligation reaction. The so created ligation fragments form a PCR template which are subsequently amplified during the singleplex PCR step with universal primer pairs. The produced ligation-amplification fragments are then able to be hybridized to MagPlex-TAG microspheres through the recognition between the anti-TAG (coupled to the beads) and the complementary TAG (present in the MOL-PCR fragment) sequences. Finally, thanks to one of the PCR primers which is biotinylated, the incubation with Streptavidin, R-Phycoerythrin Conjugate (SAPE) triggers a fluorescence reaction which is detected by the Luminex platform, attesting the presence of the related molecular marker (**Figure 11**). Basically, the read-out of a multiplex bead-suspension array by a Luminex device is the measurement of the red signal for the microsphere spectral address and the green signal for the presence of the target. The green signal detects if there was an hybridization event to the beads, and thus indicates the presence or absence of the target in the sample, and the red signal identifies the specific region of the bead and thus the identity of the detected molecular marker. Concretely, the Luminex device reports the Median Fluorescence Intensities (MFIs) for each marker and each sample. These data must be processed to determine if they are sufficiently above the fluorescence background noise to be considered as positive signals. Then, according to the combination of the molecular markers detected, the serotype identification can be determined.

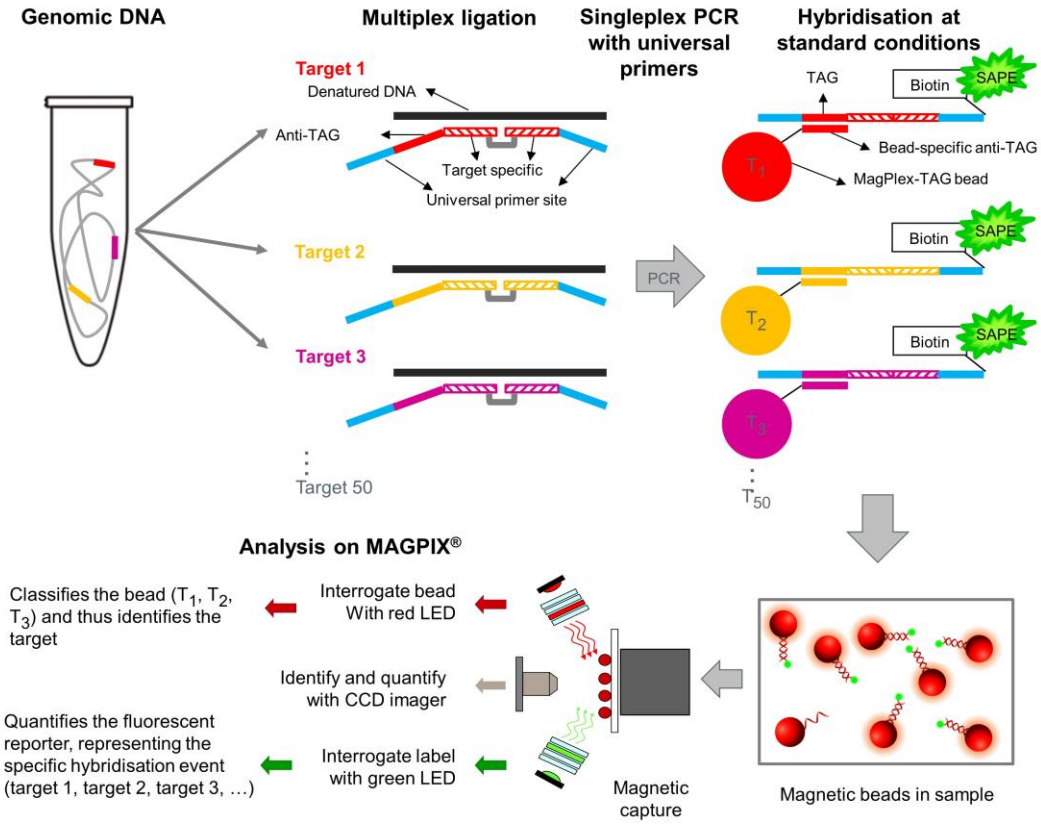

Figure 11: MOL-PCR and Luminex xTAG technology (adapted from Wuyts 2015)

The analysis of this bead-suspension array is performed with LEDs and a CCD camera after magnetic capture of the beads in the MagPix apparatus which allows multiplexing up to 50 targets (**Figure 11**). It is because of the multiplex ligation reaction, followed by a universal singleplex PCR reaction, that such a high number of targets can be simultaneously detected. Other platforms can go up to 100 (Luminex 100/200) or 500 (FlexMap 3D) targets by using lasers and flow cytometry for the beads detection. The MOL-PCR and Luminex read-out can both be performed in one day using a 96-well plate. Therefore, this technique is fast, high-throughput and allows a higher level of multiplexing compared to PCR and qPCR methods. The MOL-PCR can be performed in common thermocyclers but the detection step requires the purchase of a Luminex instrument, i.e. MagPix, Luminex 100/200 or FlexMap 3D. The Centers for Disease Control and Prevention (CDC) developed a method compatible with the Luminex 100/200 and based on a similar technology named xMAP, for the detection of more than 85% of the most common *Salmonella* serotypes, using three separate multiplex assays that simultaneously detect O and H antigens, in addition to other serotype-specific markers. This method has been commercialized by Luminex corporation. However, considering the fact that the Luminex 100/200 is more expensive than the MagPix instrument, that the method is not modular because the probe composition of the multiplex assays is not known and that the deduction of the serotype from the MFI raw data is not easy because no interpretation software is provided, this commercial kit is not really adapted to routine analysis in first-line laboratories. Indeed, these laboratories aim to detect in priority the serotypes subjected to an official control (see section 1.3.4), and this in a fast, user-friendly and cheap way.

1.5.3. Sequencing-based serotyping methods

MultiLocus Sequence Typing (MLST)

The MLST technique is based, for *Salmonella*, on the targeted sequencing of 7 housekeeping genes conserved in the bacterial genome: i.e. *aroC* (chorismate synthase gene), *dnaN (*DNA polymerase III beta subunit gene), *hemD* (uroporphyrinogen III cosynthase gene), *hisD* (histidinol dehydrogenase gene), *purE* (phosphoribosylaminoimidazole carboxylase gene), *sucA* (alpha ketoglutarate dehydrogenase gene) and *thrA* (aspartokinase + homoserine dehydrogenase gene). A region of approximatively 500 bp of each gene is amplified by separate PCR reactions and sequenced by the Sanger method after purification. After the analysis of the sequence, a number is attributed to each allele of the housekeeping genes. This means that every time that the DNA sequence of these genes differs by only one nucleotide from the previously obtained sequences (available in a database), a new number is allocated to the new allele. The combination of the 7 allele numbers forms a barcode system which is the base of the MLST scheme. All the *Salmonella* isolates having the same MLST profile are clustered

together into Sequence Types (STs). Additionally, when STs share all but one or 2 alleles, they are clustered together into ST-based clonal complex, also called eBurst Group (eBG).

In 2012, based on previous studies and their own experiments analyzing MLST data coming from more than 4 000 S*almonella* isolates belonging to 554 different serotypes, Achtman and his collaborators (2012) demonstrated the high correlation between MLST and classical serotyping. Indeed, most of the *Salmonella* isolates which were clustered together into eBG shared the same serotype, as it is illustrated for *S.* Enteritidis in **Figure 12**. Some exceptions were observed for the serotypes Newport, Oranienburg and Paratyphi B which encompassed multiple and distinct eBGs, demonstrating the genomic variety inside these populations (Achtman et al. 2012).

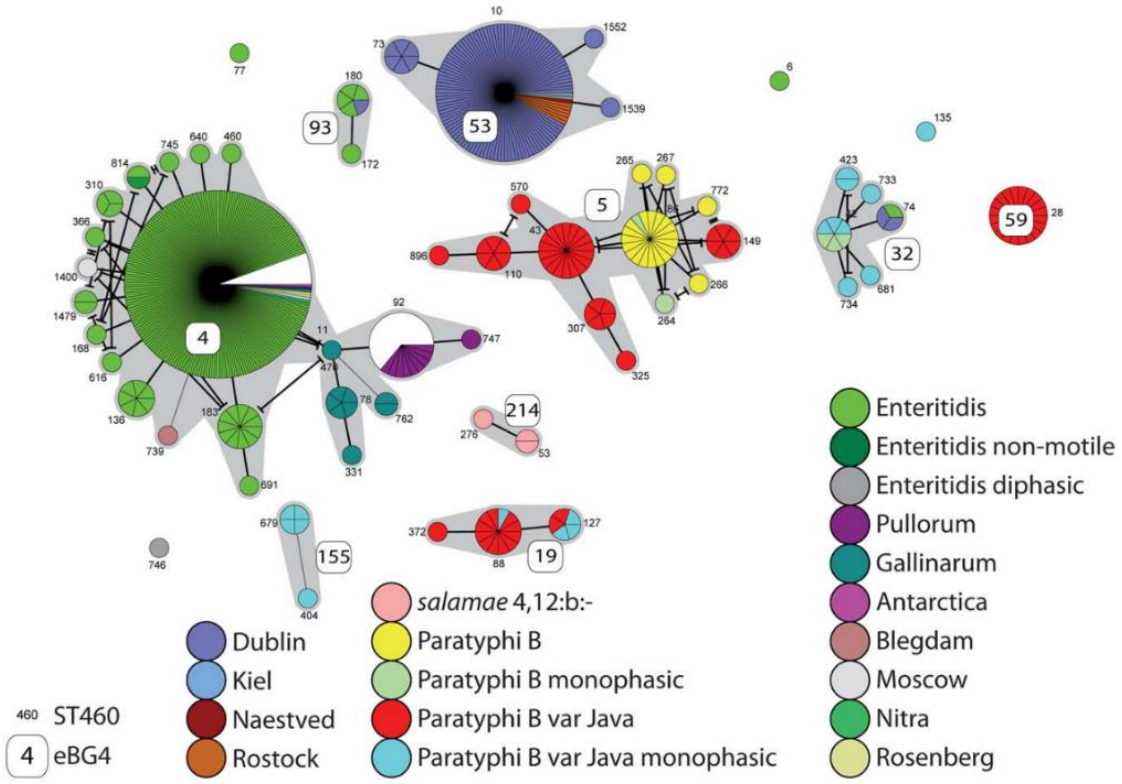

Figure 12: MSTree of *S.* Enteritidis, *S.* Dublin, *S.* Paratyphi B and their serological variants (Achtman et al. 2012)

EnteroBase is the MLST database containing all the MLST data submitted worldwide. It is is based on the analysis of sequencing data generated by Sanger sequencing or WGS, submitted by the users or automatically imported from the European Nucleotide Archive (ENA). This database contained in early 2016 more than 30 000 *Salmonella* entries (after in-house curation) and is now composed of more than 230 000 *Salmonella* entries (https://enterobase.warwick.ac.uk/; accessed in October 2019) thanks to a

huge supply of WGS data (**Figure 13**) (Alikhan et al. 2018). Most entries contain at least the following information: Sample ID, serotype determined by slide-agglutination, MLST clustering into ST and eBG when possible, predicted serotype based on the MLST analysis and other metadata if provided by the users.

Clustering of MLST types into STs and eBGs allows accurate serotype predictions. Consequently, this technique is highly efficient for the identification of almost all the *Salmonella* serotypes. Nevertheless, the amplification of the housekeeping genes requiring 7 PCRs, followed by the Sanger sequencing step, is expensive, labor intensive and thus limiting for cost-effective analysis in high-throughput in routine laboratories.

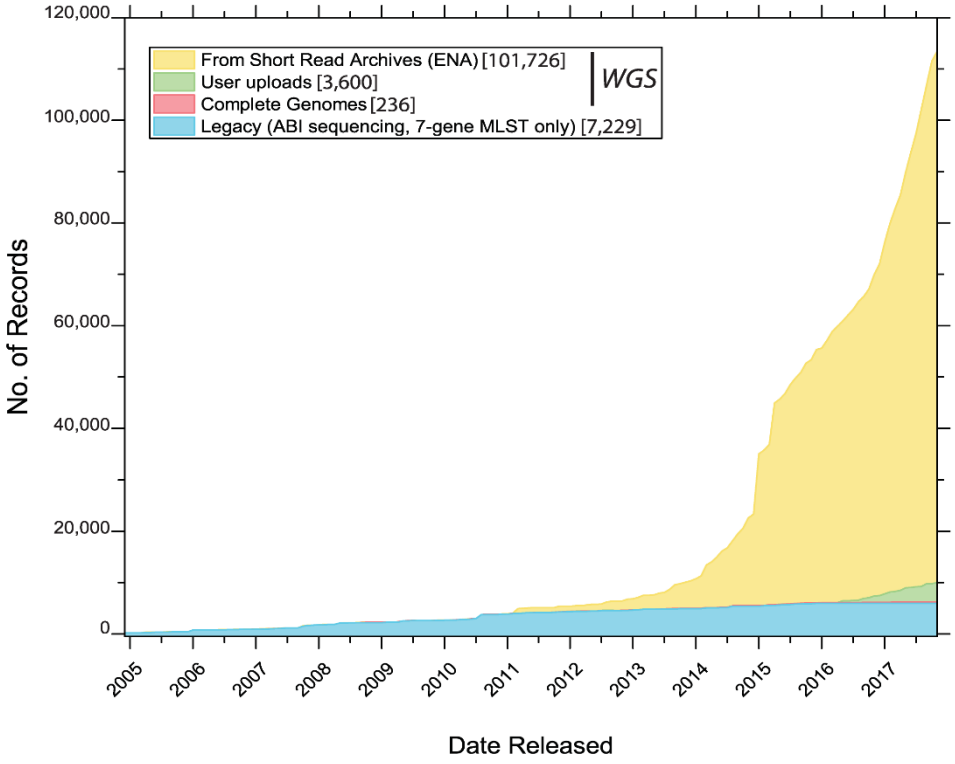

Figure 13: Evolution of number of entries in EnteroBase between 2005 and 2018. EnteroBase has performed genomic assemblies from sequence reads that were originally submitted to ENA short-read archives or directly uploaded to EnteroBase by users. EnteroBase also contains all entries with legacy MLST genotypes based on Sanger sequencing that were originally submitted to the former legacy MLST website. Historical release dates in EnteroBase for assembled *Salmonella* genomes and strains subjected to legacy sanger-based MLST. The curves indicate an exponential increase in the numbers of publicly released short reads over time versus only a linear increase in legacy entries (adapted from Alikhan et al., 2018).

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) typing

CRISPR is a family of repeated DNA sequences present in many prokaryotes. This family is composed of 24-27 bp DNA direct repeats, separated by variable 21-72 bp sequence spacers. Two CRISPR loci, CRISPR1 and CRISPR2, have been identified in *Salmonella* and the determination of their sequence allows the identification of isolates at the serotype level in addition to their subtyping. Indeed, a reliable link was demonstrated between the presence or absence of certain variable sequences (i.e. spacers), as well as their arrangement in the CRISPR loci, and the *Salmonella* serotypes, which led to the establishment of a CRISPR/serotype dictionary (Laëtitia Fabre et al. 2012; Weill et al. 2007). In practice, similarly as for MLST, 3 set of primers are used to amplify either CRISPR1, CRISPR2 or both, and the composition of the amplified sequences is determined by Sanger sequencing. After comparison of the retrieved sequence(s) with the CRISPR/serotype dictionary, the serotype of the *Salmonella* isolate can be determined. The authors who developed this method created in 2012 a dictionary containing the association of 130 serotypes, the most frequently involved in human infections, with the inventory of their related spacer content, and planned to expand it in the following years to capture the diversity of all *Salmonella* serotypes (Laëtitia Fabre et al. 2012). Even though a complete and informative database, ideally publicly available and allowing an easy interpretation of the results, is still awaited, this method has a strong potential as it can perform *Salmonella* serotyping and subtyping in one analysis. Nevertheless, CRISPR typing has the same drawbacks than MLST as several PCR amplifications are needed, followed by Sanger-sequencing, and hence this method is not the most adapted to the needs of the food sector for a fast, easy and cost-effective *Salmonella* serotyping.

Whole Genome Sequencing (WGS)

WGS is the determination of the complete DNA sequence of the genome of a bacterial isolate in high-throughput, using Next Generation Sequencing (NGS) technologies, of which the Illumina technology is one of the most commonly used. The principle of Illumina is the sequencing by synthesis. First, the genome is fragmented while simultaneously integrating amplification primers and adapters at each end allowing their capture by a flow cell. Then the DNA polymerase catalyzes the incorporation of fluorescently labeled deoxyribonucleotide triphosphates (dNTPs) into the DNA template strand hybridized to the flow cell, during sequential cycles of DNA synthesis. During each cycle, at the point of incorporation, the nucleotides are identified by fluorophore excitation. With the Illumina sequencing, millions of DNA fragments are sequenced as such in parallel. The produced WGS data are fragmented raw sequencing reads which must be processed and assembled using bioinformatics tools to reconstruct the genome. Then, the assembled genome is imported into automatic pipelines for serotype identification. One of these web-based pipelines is the *Salmonellas In Silico* Typing Resource (SISTR) which performs serotype prediction based on the sequence of the somatic O and flagellar H genes, the MLST profile of the isolate and matching with existing *Salmonella* genomes publicly available.

During the last years, WGS was mostly used by public health authorities for retrospective studies of outbreaks or to evaluate the genomic diversity among pathogen populations. But with the decrease of the analysis cost and time linked to the NGS technologies, WGS has been slowly implemented in big laboratories in Canada, France, United Kingdom and U.S.A., for the routine serotyping of *Salmonella* (Allard 2016; Ashton et al. 2016; Jain, Mukhopadhyay, and Thomassin 2019). In addition to identify the serotype, WGS analysis provides valuable information on the subtype, the antimicrobial resistance and virulence profile and the potential link with past or current outbreaks. Nevertheless, the complete analysis from genomic DNA extraction to bioinformatics analysis takes at least 4 days to be performed with the currently in routine used sequencing technologies and is cost-effective only when batching with a consistent number of samples is possible (Jain, Mukhopadhyay, and Thomassin 2019; Ibrahim and Morin 2018). Consequently, this method is not adapted for small laboratories with limited resources.

1.5.4. What is the best method for Salmonella *genoserotyping in a routine setting?*

The fast identification of *Salmonella* serotypes is of major importance when first-line laboratories isolate this pathogen from the field, especially for the serotypes subjected to an official control in the food sector. The perfect genoserotyping method, for implementation in routine laboratories, must be fast, accurate, cost-effective and easy to perform. Additionally, the alternative molecular method must be compatible with the WKL scheme commonly used by laboratories worldwide and included in the regulations. Moreover, the newly developed test must be validated by comparison with the classical method (i.e. the slide-agglutination and the biochemical tests) and must match the criteria of the ISO 15189 and ISO 17025 to be used in routine laboratories under accreditation.

The advantages and drawbacks of the alternative molecular methods for *Salmonella* serotype identification described in this chapter have been summarized in **Table 2**. The PCR-based methods are easy to perform and have the advantage to be commonly used by a lot of laboratories worldwide, but the number of targeted molecular makers, and thus the number of serotypes which can be simultaneously detected, is limited. Additionally, the detection by gel electrophoresis can be labor-intensive and not user-friendly. The number of possible targets is much higher with the microarray or the sequencing methods. But with the microarray the number of analyzed samples is limited, and the sequencing methods are time-consuming and labor-intensive, although highly accurate. Additionally, WGS is only cost-effective when used in large sequencing platforms, where several samples can be analyzed together and this at regular time points. The MOL-PCR and Luminex technology offer a good compromise because the analysis of 96 samples by one multiplex assay is performed in one day with a high level of multiplexing (Wuyts, Mattheus, et al. 2015). But in the existing commercial kit using this technology for *Salmonella* genoserotyping, 3 MOL-PCR assays are needed for the complete identification of the samples, so only 30 samples can be analyzed in 1 day. Additionally, no interpretation software is

provided with the kit and the raw data must be manually converted into identification results. Also, this commercial kit is only compatible with the Luminex 100/200 which represents a non-negligible investment cost. A new Luminex test, using less multiplex assays, compatible with the less expensive MagPix apparatus and provided with a Decision Support System (DSS) for an automatic interpretation of the raw data, would be perfectly adapted to the need of the first-line laboratories. Another need of this new test, in comparison with the commercial kit which is a black box, would be to be modular, and thus easily adaptable following the evolution of the regulation or the prevalence of the *Salmonella* serotypes to detect in priority.

Table 2: Advantages and drawbacks of alternative molecular methods for *Salmonella* serotype identification

	PCR	qPCR	Microarray[1]	Luminex[2]	MLST	CRISPR typing	WGS
Analysis time	4-5h	3h	7h	7h	≥ 3 days	≥ 3 days	≥ 4 days
Analysis cost [3]	Cheap	Cheap	Average	Average	Expensive	Expensive	Expensive
User friendly	Yes	Yes	Relatively	Relatively	No	No	No
Number of samples in 1 analysis	40-96 [4]	96	≤ 8	30	≤ 10	30	35 [5]
Number of targeted serotypes in 1 assay	Limited	Very limited	Very high[6]	High[7]	Very High[8]	High[9]	Unlimited

The different parameters presented in this table are estimated for the execution of one assay (e.g. one multiplex for PCR, qPCR and MOL-PCR, in 1 instrument) performed by one technician. If more than one assay is used for PCR, qPCR and MOL-PCR, the number of samples possible in 1 analysis must be divided by the number of assay needed.

[1]: the parameters are estimated based on the commercial Check & Trace *Salmonella* microarray.

[2]: the parameters are estimated based on the commercial Luminex xMAP® *Salmonella* Serotyping Assay (SSA). Despite the Luminex technology is working with a 96-well plate, 3 multiplex assays are needed for *Salmonella* serotyping using this SSA. Consequently, only 30 samples can be analyzed in 7h with this commercial kit.

[3]: the analysis cost includes the availability or the purchasing of the necessary instruments.

[4]: depending on the capacity of the gel electrophoresis system.

[5]: for a coverage of 40-65x using an Illumina MiSeq.

[6]: according to the manufacturer, the Check & Trace *Salmonella* commercial kit can discriminate over 300 serotypes.

[7]: according to the manufacturer, the Luminex xMAP® SSA can identify more than 85% of the most commonly encountered *Salmonella* isolates.

[8]: the MLST technique can theoretically identify all the serotypes as long as they are well conserved into STs and eBGs, and sufficiently represented in EnteroBase, i.e. not rare serotypes

[9]: the CRISPR typing technique was described as potentially able to identify all the *Salmonella* serotypes. However, only a dictionary of 130 serotypes and their related CRISPR spacer content was made available until now.

CHAPTER 2
Scientific problem addressed and aims of the project

Salmonella is a major pathogen commonly involved in foodborne diseases worldwide. As these bacteria are transmitted to humans mainly through the consumption of food, and more specifically food products coming from animals which are their main reservoir, a strict monitoring and control of *Salmonella* serotypes is required along the food chain. A special focus is put on the poultry and pork sector as chickens and pigs, and their associated products such as eggs, have been identified as the most common sources of *Salmonella* infections in humans. Consequently, following the EU Regulation (EC) No 2160/2003, Belgium adopted measures for the reduction of *Salmonella* in breeding flocks of *Gallus gallus*, laying hens and broilers, with the aim to limit its spreading along the food chain by controlling it directly at the initial level, i.e. the farm (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). In this regulation, specific serotypes are subjected to an official control. This means that when the first-line laboratories isolate *Salmonella* from the field, the determination of its serotype is a key diagnostic to match the criteria of the regulation. Additionally, the serotyping of *Salmonella* isolates is an essential tool to quickly detect the emergence of outbreaks and trace their source. Nevertheless, the classical methods for *Salmonella* serotyping and variant determination (i.e. slide-agglutination and biochemical tests), following the WKL scheme, are expensive, labor-intensive, potentially time-consuming and require carefully trained technicians for the interpretation of the results, which is subjective. Despite this, the serotyping by slide-agglutination is widely used since 80 years and the WKL scheme is included in the regulations of a lot of countries. Consequently, most of the *Salmonella* strains isolated from the field by the first-line laboratories, must be sent to the NRC which fully masters the classical methods for serotyping. Unfortunately, this situation adds extra cost and delays to the analysis demanded from the professionals of the food sectors to comply with the regulation. Additionally, a non-negligible number of *Salmonella* isolates, coming mostly from the food sector, cannot be serotyped by slide-agglutination because of their auto-agglutinable character. If such an isolate belongs to one of the serotypes subjected to an official control, it cannot be identified as such and it will not be excluded from the food chain because no corrective measures will be taken.

To solve these issues, the Belgian public federal service for public health, security of the food chain and environment raised the need for a new, fast and accurate method for the identification of zoonotic *Salmonella* serotypes which must be monitored in the poultry and pork sectors. The former Scientific Institute for Public Health, now Sciensano, was chosen to develop this new method through this PhD research. Therefore, the aim of the present work was to design an alternative serotyping method based on the molecular techniques available and to validate it by comparison with the classical method. For doing this, the following scientific research questions were investigated (**Figure 1**):

- What is the best approach for *Salmonella* serotyping, taking into account the regulatory aspects?
- How to develop a valid and accurate alternative method?
- What is the method's applicability as replacement of the classical methods?

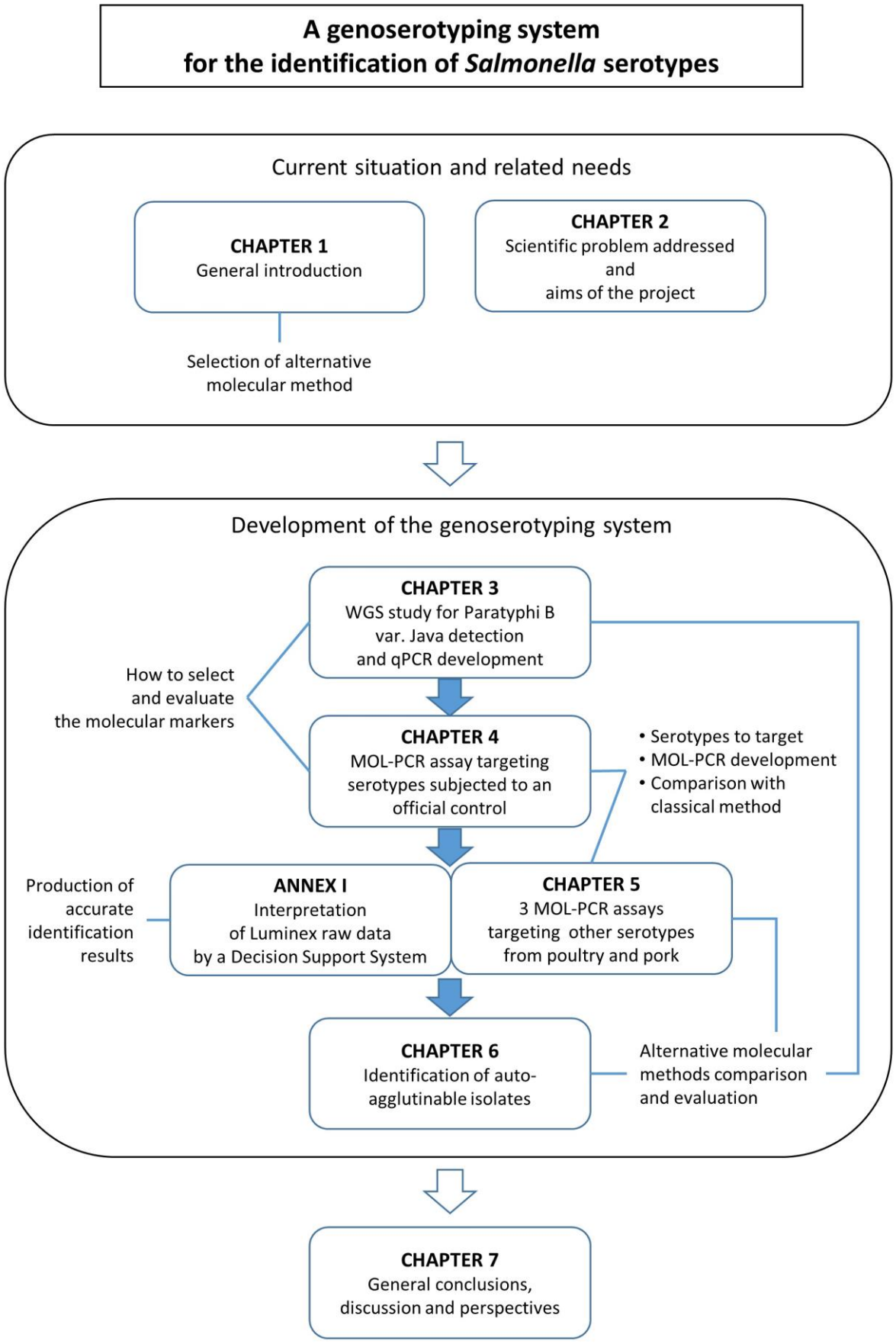
A genoserotyping system
for the identification of Salmonella serotypes
Current situation and related needs
CHAPTER 1
General introduction
CHAPTER 2
Scientific problem addressed
and
aims of the project
Selection of alternative
molecular method
Development of the genoserotyping system
CHAPTER 3
WGS study for Paratyphi B
var. Java detection
and qPCR development
How to select
and evaluate
the molecular markers
CHAPTER 4
MOL-PCR assay targeting
serotypes subjected to an
official control
• Serotypes to target
• MOL-PCR development
• Comparison with
classical method
Production of
accurate
identification
results
ANNEX I
Interpretation
of Luminex raw data
by a Decision Support System
CHAPTER 5
3 MOL-PCR assays
targeting other serotypes
from poultry and pork
CHAPTER 6
Identification of auto-
agglutinable isolates
Alternative molecular
methods comparison
and evaluation
CHAPTER 7
General conclusions,
discussion and perspectives

Figure 1: Outline of the thesis

The following detailed research questions were investigated in this PhD work throughout 7 chapters:

- What are the most important *Salmonella* serotypes which must be targeted by the method? (Chapters 4 and 5)
- What is the best molecular technique as an alternative for *Salmonella* serotype identification? (chapters 3, 5 and 6)
- How to select the molecular markers and evaluate their specificity? (Chapters 3 and 4)
- How to produce accurate identification results easily obtainable by the laboratory technicians? (Chapters 4, 5 and annex I)
- Is the method valid after comparison with the classical methods? (Chapters 4 and 5)
- Is the developed method able to replace complex biochemical tests for serotype variants determination? (Chapters 3 and 5)
- Is the method able to clearly identify auto-agglutinable *Salmonella* isolates not typable by slide-agglutination? (Chapter 6)

CHAPTER 3
Development of a real-time PCR method for the genoserotyping of *Salmonella* Paratyphi B variant Java

Context of this chapter:

The strategy of MLST marker selection using EnteroBase, fully described in Chapter 4, was not successful for the specific detection of *S.* Paratyphi B var. Java. Therefore, in the current chapter, the specificity of a publicly available molecular marker, targeting this serotype, was checked, and an additional SNP marker was selected from genomic comparison using WGS data. From this, a real-time PCR test was developed for the fast and accurate identification of *S.* Paratyphi B and the determination of its variant Java, as an alternative to a complex and time-consuming biochemical test.

This chapter was adapted from the previously published manuscript:
Gand, M., W. Mattheus, A. Saltykova, N. Roosens, K. Dierick, K. Marchal, S.C.J. De Keersmaecker* and S. Bertrand*. 2019. "Development of a real-time PCR method for the genoserotyping of *Salmonella* Paratyphi B variant Java." *Applied Microbiology and Biotechnology* 103 (12): 4987–96. https://doi.org/10.1007/s00253-019-09854-4. (* equal contribution)

Authors' contributions:

M. Gand designed the study, performed the experiments, analyzed the data and drafted the manuscript. A. Saltykova created the in-house script for the filtering of the mutation list. S.C.J. De Keersmaecker and S. Bertrand conceived and supervised the study, helped to design the study, to interpret the results and to draft the manuscript. K Marchal, N. Roosens and W. Mattheus provided specialist feed-back. All authors read and approved the final manuscript.

Abstract:

Discriminating between d-tartrate fermenting and non-fermenting strains of *Salmonella enterica* subsp. *enterica* serotype Paratyphi B is of major importance as these 2 variants have different pathogenic profiles. While d-tartrate non-fermenting *S.* Paratyphi B isolates are the causative agent of typhoid-like fever, d-tartrate fermenting isolates (also called variant Java) of the same serotype trigger the less dangerous gastroenteritis. The determination of *S.* Paratyphi B variants requires a time-consuming process and complex biochemical tests. Therefore, a quadruplex real-time PCR method, based on the allelic discrimination of molecular markers selected from the scientific literature and from whole genome sequencing data produced in-house, was developed in this study, to be applied to *Salmonella* isolates. This method was validated with the analysis of 178 *S.* Paratyphi B (d-tartrate fermenting and non-fermenting) and other serotypes reaching an accuracy, compared to the classical methods, of 98% for serotyping by slide-agglutination and 100% for replacement of the biochemical test. The developed real-time PCR permits to save time and to obtain an accurate identification of *S.* Paratyphi B serotype and its d-tartrate fermenting profile, which is needed in routine laboratories for fast and efficient diagnostics.

3.1. Introduction

Salmonella is one of the major causes of food poisoning all over the world. These bacteria can contaminate a large variety of food products including those of animal origin such as eggs, milk products or meat. This is why the combat against zoonotic *Salmonella* (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8) is crucial to rapidly identify serotypes that may contaminate the food chain like Paratyphi B variant Java in poultry products. Additionally, *Salmonella* can cause diseases in poultry and pork farming. One of the major concerns of *Salmonella* is economic loss due to contaminated food destruction and economic inactivity due to sickness leave.

The *Salmonella* genus is composed of more than 2 500 serotypes divided in 2 species, i.e. *Salmonella enterica* and *Salmonella bongori*. *Salmonella enterica* is itself subdivided in 6 subspecies among which the 1 500 serotypes of the subspecies 1 (also called *Salmonella enterica* subsp. *enterica*) are the main cause of *Salmonella* infections in human (Ryan, O'Dwyer, and Adley 2017). The gold standard technique for the characterization of *Salmonella*, widely used since 60 years, is the serotyping by slide agglutination following the Kauffmann- White-Le Minor (WKL) scheme, consisting of the identification of three antigenic sites (somatic O and two flagellar H antigens) by specific antisera. In spite of its worldwide use, this technique is time-consuming, not always objective and it requires carefully trained personnel. Moreover, for the differentiation between two variants of a same serotype, additional biochemical tests are needed. This is amongst others important for *Salmonella enterica* subsp. *enterica* serotype Paratyphi B (*S.* Paratyphi B) as it can be discriminated into two variants depending on its ability to ferment dextrorotatory [L(+)]-tartrate (d-tartrate). The pathogenicity of these two variants is totally different: whereas the rare d-tartrate-non-fermenting (dT-) variant causes typhoid-like fever, the more-spread d-tartrate-fermenting (dT+) variant, called var. Java, leads to a less dangerous food poisoning (Malorny, Bunge, and Helmuth 2003). The ability of strains to ferment d-tartrate is tested by culture-based biochemical methods, i.e. the lead-acetate or the commercial Remel™ Jordan's Tartrate Agar tests. These methods are however poorly reproducible, time-consuming (2 to 7 days) and can lead to false negative results (Alfredsson et al. 1972; Barker 1985; Malorny, Bunge, and Helmuth 2003).

Since a few years, molecular techniques have proven to be suitable tools for the genoserotyping of *Salmonella*, including for the determination of variants. Indeed, the Multi Locus Sequence Typing (MLST) technique showed how genotype clusters defined by molecular typing method correspond (for most of the serotypes) to serotype clusters determined by slide-agglutination (following the WKL scheme) and was therefore proposed as replacement for classical serotyping (Achtman et al. 2012). The MLST technique was, however, not able to cluster all the Paratyphi B strains in one close related group, as the Paratyphi B population is polyphyletic and a large heterogeneity of genotypes exists inside this serotype. Later, Connor and his collaborators (2016) described the genomic variation in the Paratyphi B

group after analysis of a large amount of Whole Genome Sequencing (WGS) data (191 strains sequenced), giving the first high-resolution view of this serotype. They were able to cluster the analyzed strains into Phylogenetic Groups (PGs) numbered from 1 to 10.

Other genoserotyping methods are based on molecular markers specific for some serotypes which are detected by PCR-based technologies (Franklin et al. 2011; Maurischat et al. 2015; Rajtak et al. 2011; Yoshida, Simone, et al. 2016). For example, Malorny, Bunge, and Helmuth (2003) developed a PCR method for the differentiation between dT- and dT+ *Salmonella* strains as an alternative to the biochemical tests mentioned above. They discovered that the non-fermenting characteristic of dT- strains was due to SNP in the start codon (ATA instead of ATG) of a gene (*STM 3356*) encoding a putative cation transporter involved in the d-tartrate fermentation pathway. Based on this SNP, they designed PCR primers specific to dT+ strains. The amplified fragments are detected through agarose gel electrophoresis. Similarly, Zhai and his collaborators (2014) developed a PCR test, based on the *SPAB_01124* gene (a specific marker resulting from a genomic study) for the detection of the serotype *S.* Paratyphi B in food. For the determination of the serotype and its variant, the disadvantages, however, are that 2 separate PCR tests are required followed by a detection using agarose gel electrophoresis.

As asked by the legislation, it is important to clearly and rapidly identify *S.* Paratyphi B var. Java (dT+) isolates entering in the food chain. Therefore, there is a need to develop a fast and accurate technique, especially for the discrimination between dT- and dT+ variants. In this study we developed a multiplex real-time PCR (qPCR) method, based on markers found in the scientific literature and on in-house produced WGS results, in order to replace the dT variant biochemical test and simultaneously confirm the Paratyphi B serotype identification, once *Salmonella* is isolated from its matrix.

3.2. Materials and methods

3.2.1. Bacterial strains

All the strains used (**Supplemental Table S1**) are reference isolates from the Belgian National Reference Center (NRC) for *Salmonella* and *Shigella*. These strains have been sent to the NRC for further characterization after the isolation from human, food or animal matrices by the first-line laboratories and *Salmonella* spp. identification. All the analyses were performed from a characteristic colony on XLD agar and confirmation of *Salmonella* genus identification was done by MALDI-TOF method if needed. All isolates are available upon request. The serotype of these isolates was confirmed, prior to use, by slide-agglutination (Grimont and Weill 2007). To avoid confusion, the name *S.* Paratyphi B will be used in this study for isolates belonging to the serotype Paratyphi B *stricto sensu* with no information on the d-tartrate fermentation ability. *S.* Paratyphi B var. Java isolates which can ferment the d-tartrate will be named *S.* Paratyphi B (dT+) in contrast to isolates which cannot, named *S.* Paratyphi B (dT-).

3.2.2. Biochemical tests for the d-tartrate fermentation ability

The lead-acetate test was performed as described by Alfredsson et al. (1972) but with the modified inoculation step (a loopful of bacteria from an overnight (14 to 20 hours) culture at 37°C on Nutrient agar (Neogen® Culture Media, Lansing, USA)) as recommended by Malorny, Bunge, and Helmuth (2003).

The commercial Remel Jordan's tartrate test (ThermoFisher Scientific, Waltham, USA) was used according to the manufacturer's instructions.

3.2.3. DNA extraction

For qPCR and Sanger sequencing, the DNA template was prepared by heat lysis. To perform this, a single colony from an overnight (14 to 20 hours) culture at 37°C on Nutrient agar was dissolved in 60 µl sterile de-ionized water and incubated at 95°C in a heating block for 10 minutes. After cooling for minimum 10 minutes at 4°C (in the fridge) and centrifugation for 10 minutes at 11 000 × g using Centrifuge 5417C (Eppendorf, Hamburg, Germany), the supernatant was stored at −20°C and used for further analysis.

For WGS and parts of the qPCR analysis, genomic DNA was extracted with the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich, Saint-Louis, USA) according to the manufacturer's instructions.

3.2.4. PCR tests for the identification of S. *Paratyphi B dT+ isolates*

The PCR test of Zhai et al. (2014) (mentioned in the present study as "PCR Zhai") and the PCR of Malorny, Bunge, and Helmuth (2003) (mentioned in the present study as "PCR Malorny") were performed according to the author's instructions. Nuclease free distilled water was used as a no template control (NTC).

3.2.5. WGS and genome comparison study

Genomic DNA of 13 *S.* Paratyphi B isolates (5 dT- and 8 dT+) was sequenced with an Illumina MiSeq instrument (2 x 300 bp, Nextera XT libraries). The serotype Paratyphi B was confirmed for each of the isolates using SeqSero (Zhang et al. 2015) with raw reads as input. FASTQ reads from all sequences were deposited at the SALMSTID BioProject on NCBI (PRJNA509747).

Table 1: Target genomes for *S.* Paratyphi B

Serotype	Type of sequence	Reference[a]
Paratyphi B	Complete genome	NC_010102.1
Paratyphi B (PG1)	Raw reads	ERR023396
Paratyphi B (PG1)	Raw reads	ERR460132
Paratyphi B (PG2)	Raw reads	ERR129870
Paratyphi B (PG2)	Raw reads	ERR460150
Paratyphi B (PG3)	Raw reads	ERR278708
Paratyphi B (PG3)	Raw reads	ERR460145
Paratyphi B (PG4)	Raw reads	ERR278698
Paratyphi B (PG4)	Raw reads	ERR278712
Paratyphi B (PG5)	Raw reads	ERR023399
Paratyphi B (PG6)	Raw reads	ERR460141
Paratyphi B (PG6)	Raw reads	ERR460153
Paratyphi B (PG7)	Raw reads	SRR1965575
Paratyphi B (PG8)	Raw reads	ERR278705
Paratyphi B (PG9)	Raw reads	ERR129875
Paratyphi B (PG10)	Raw reads	ERR403703

[a]**:** references of complete genomes, contig lists and raw reads are accession numbers, sequenced strain references and Sequence Read Archive (SRA) respectively.

Table 2: Background genomes

Serotype	Type of sequences	Reference[a]
Agona	Complete genome	CP006876.1
Anatum	Complete genome	CP013222.1
Blockley	Contig list	CRJJGF_00147
Bovismorbificans	Complete genome	HF969015.2
Braenderup	Contig list	CFSAN044976
Brandenburg	Contig list	CVM N45949
Bredeney	Complete genome	CP007533.1
Cerro	Complete genome	CP012833.1
Chester	Complete genome	CP019178.1
Choleraesuis	Complete genome	CP007639.1
Derby	Contig list	07CR553
Dublin	Complete genome	CP019179.1
Enteritidis	Complete genome	CP007434.2
Gallinarum var. Pullorum	Complete genome	LK931482.1
Gallinarum var. Gallinarum	Complete genome	CP019035.1
Gaminara	Contig list	SA20063285
Hadar	Contig list	SA20026260
Hvittingfoss	Contig list	SA20014981
Indiana	Contig list	ATCC 51959
Infantis	Complete genome	LN649235.1
Javiana	Contig list	CVM N42337
Litchfield	Contig list	CVM N32042
Livingstone	Contig list	CKY-S4
Manhattan	Contig list	SA20034532
Mbandaka	Complete genome	CP019183.1
Minnesota	Complete genome	CP019184.1
Montevideo	Complete genome	CP007222.1
Muenchen	Contig list	CVM N42480
Muenster	Complete genome	CP019198.1
Newport	Complete genome	CP016014.1
Ohio	Contig list	CVM N29382
Oranienburg	Contig list	CFSAN039514
Panama	Complete genome	CP012346.1
Paratyphi A	Complete genome	CP019185.1
Pomona	Contig list	ATCC 10729
Poona	Contig list	2010K-2244
Rissen	Contig list	150_SEER
Saintpaul	Complete genome	CP017727.1
Senftenberg	Complete genome	LN868943.1
Stanley	Contig list	06-0538
Tennessee	Contig list	SALC_70
Typhimurium	Complete genome	NC_003197.2
Virchow	Contig list	SVQ1
Weltevreden	Complete genome	LN890524.1

In CLC Genomics Workbench 8.0 (Qiagen, Hilden, Germany), the raw FASTQ reads were first trimmed to quality score limit 0.05 with maximum 2 ambiguous nucleotides and reads with length below 30 nucleotides were discarded. These trimmed reads were then *de novo* assembled with automatic bubble and word size, in mapping mode "map reads back to contigs" with scaffolding and a minimum contig length of 1 000 nucleotides. The WGS data were subsequently analyzed with Gegenees which is a software for comparative analysis of microbial WGS data, allowing to define genomic signatures unique for specified target groups. The contigs were exported to Gegenees (version 2.2.1; downloaded from http://www.gegenees.org; Ågren et al. 2012) on a Linux platform with 16 *S.* Paratyphi B genomes (including 15 from Connor et al. 2016) belonging to different PGs (2 of each PG when possible) (**Table 1**) and 44 other genomes belonging to other frequent serotypes (**Table 2**), all publicly available on NCBI. The complete genomes mentioned in the **Tables 1 and 2** are annotated genomes which are preferably used as reference genomes. The downloaded raw reads were first trimmed and assembled as described for the in-house sequenced data. A fragment all-against-all comparison was made with all the genomes. The genomes belonging to serotype Paratyphi B were labeled as TARGET in the software (and the genome NC_010102.1 as REFERENCE additionally) and the other genomes as BACKGROUND. For each comparison, the Biomarker score was used to find sequences specific of the TARGET group and absent in the BACKGROUND group.

Multiple alignments of all the genomes were performed with the BioNumerics software (Applied Maths, Sint-Martens-Latem, Belgium; version 7.6) and a mutation list containing SNP differences and their position in the genomes was created. This list was filtered using command line tools on a Linux platform, i.e. retrieving SNP markers present in the TARGET group and absent in the BACKGROUND group.

3.2.6. qPCR for detection of S. *Paratyphi B var. Java*

The TaqMan probes ParaB and Java, for the identification of the Paratyphi B serotype and the dT variant, were inspired from the marker *SPAB_04460* found in our genomic study and from the primer 166 (gene *STM3356*) of the study of Malorny, Bunge, and Helmuth (2003), respectively. For each marker, a SNP probe and a WT probe were designed by putting the specific nucleotide locus in the middle of the TaqMan probe. The probes were synthetized with Locked Nucleic Acids (LNAs) in order to achieve the targeted Tm of 66°C with a probe length lower than 25 bp, corresponding to the qPCR guidelines given by IDT (Designing PCR primers and probes; https://eu.idtdna.com). Corresponding primers were designed in order to amplify a region of ~ 100 bp flanking the ParaB probes and the Java probes, respectively. All the probes and primers were ordered at IDT (Leuven, Belgium) (**Table 3**).

Real-time PCR reactions were performed in one single quadruplex reaction in a final volume of 25 µL composed of 1x Takyon™ Rox Probe MasterMix UNG (Eurogentec, Liège, Belgium), 0.25 µM

of corresponding TaqMan probes (except for the ParaB_SNP probe for which 0.05 µM was used as asymmetric concentrations gave better results for the pair of probes ParaB), 0.4 µM of corresponding primers and 5 µL of DNA (extracted by heat lysis or GenElute extraction kit (Sigma-Aldrich, Saint-Louis, USA) at 5ng/µL). Nuclease free distilled water was used as a no template control (NTC).

Table 3: Sequences of TaqMan probes, qPCR primers and sequencing primers

Target	Type	Name	Sequence (5' – 3')
Paratyphi B	TaqMan probes	ParaB_SNP	/FAM/TCGGCATAG{T}{**T**}AGATCTTTGCC/BHQ_1/
		ParaB_WT	/Tex615/TCGGCATAGT{**C**}AGATCTTTGCC/BHQ_2/
	Primers	ParaB_Fw	AACATGCCGAGCGTAAAC
		ParaB_Rv	ACTGGCAGCGATTTACAC
		ParaBSeq_FwT7	TAATACGACTCACTATAGGGTGCTAAAGACGCCGGTATAA
		ParaBSeq_Rv	ATTAACCCTCACTAAAGGGA
dT-/dT+	TaqMan probes	Java_SNP(dT-)	/HEX/ATTATAAATA{T}{**A**}{G}{A}ACCCATTACCC/BHQ_1/
		Java_WT(dT+)	/Cy5/ATTATAAATA{T}{**G**}G{A}ACCCATTACCC/BHQ_2/
	Primers	Java_FW	TTCTCCCTGTCAACATTGG
		Java_Rv	TTCCCATACAAACATGACGA
		JavaSeq_FWT7	TAATACGACTCACTATAGGGGAGAATATGCTGACCCGCTA
		JavaSeq_Rv	ATTAACCCTCACTAAAGGGA

/FAM/: 6-carboxyfluorescein
/Cy5/: cyanin 5
/HEX/: Phosphoramidite
/Tex615/: TexasRed615
Nucleotides between { } are LNA base
Nucleotides in **bold** are specific for the SNP or WT marker

The concentration of the DNA extracted with the GenElute kit was measured with Nanodrop (ThermoFisher Scientific, Waltham, USA). Extraction by heat lysis was selected as extraction method as it gave the same results than with the GenElute extraction kit and because it is cheaper. Other master mixes were tested at the same concentration of probes and primers (RT-PCR Mastermix (Diagenode, Liège, Belgium) and TaqMan Genotyping MasterMix (Applied Biosystem, Foster City, USA)) but the Takyon™ Rox Probe MasterMix UNG (Eurogentec, Liège, Belgium) was kept as it gave a good discrimination between the 2 alleles. The PCR conditions for the qPCR reaction were: 10 min at 95 °C, 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Fluorescence intensity was collected at the end of the annealing step. The reaction was performed on a CFX96 (Bio-Rad, Hercules, CA, USA). The

S. Paratyphi B isolate II-37-NH was used as a positive control for ParaB_SNP and Java_SNP(dT-) and a negative control for the WT version of the same probes. Identically, the *S.* Enteritidis isolate S15BD02868 was used as a positive control for ParaB_WT and Java_WT(dT+) and a negative control for the SNP version of the same probes.

Real-time PCR fluorescence results were analyzed using the Allelic Discrimination tab of the Bio-Rad CFX Manager (version 3.1; Bio-Rad). For each isolate, the relative fluorescence (RFU) of SNP probes was divided by the relative fluorescence of their respective WT probes. For both markers, if this ratio was greater than 1.0, the SNP version of the marker is present in the genome of the isolate. If it was below 1.0, the WT version of the marker is present in the genome of the isolate. Isolates which have the SNP allele or the WT allele of the markers *SPAB_04460* are identified as *S.* Paratyphi B or belonging to another serotype than Paratyphi B, respectively. Isolates which have the SNP allele or the WT allele of the marker *STM 3356* are discriminated as dT- or dT+ strains, respectively.

To assess the selectivity of the developed method, the sensitivity and specificity were determined by inclusivity and exclusivity tests, respectively, as described previously by Barbau-Piednoir, Bertrand, et al. (2013) and Barbau-Piednoir, Botteldoorn, et al. (2013). Sensitivity is the ability of the developed method to identify correctly true positive samples whereas specificity is the ability of the same method to identify correctly true negative samples. True negative and positive samples are determined by the reference method (here: slide-agglutination and biochemical test). The accuracy is determined by the closeness of agreement between a test result and the accepted reference value (Banoo et al. 2010; Berwouts, Morris, and Dequeker 2010; Burd 2010). The parameters were calculated with the following formulas:

$$Sensitivity\ (inclusivity) = \frac{a}{(a+d)}$$

$$Specificity\ (exclusivity) = \frac{b}{(b+c)}$$

$$Accuracy = \frac{a+b}{(a+b+c+d)}$$

where a = number of true positive samples, b = number of true negative samples, c = number of false positive samples and d = number of false negative samples.

3.2.7. Sanger sequencing

The marker sequences targeted by the TaqMan probes ParaB and Java were determined on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster city, USA) according to the manufacturer's instructions. Sequencing primers were designed with Primer3 (http://primer3.ut.ee; Untergasser et al. 2012) with the aim to amplify a region between 500 and 600 bp flanking the TaqMan probes annealing sites. Forward primers were extended with a T7 primer binding site at their 5' end for the sequencing

step **(Table 3)**. The PCR to prepare the sequencing templates was performed in a final reaction volume of 48 µl including 1× FastStart PCRMaster (Roche, Bâle, Switzerland), and 2 µl of the DNA (extracted by heat lysis) used for the qPCR assay. The following protocol was run in a thermal cycler: 4 minutes at 95°C, 30 cycles of 30 s at 94°C, 1 min at 55°C and 1 minute at 72°C, 10 minutes at 72°C. PCR products were visualised by agarose gel electrophoresis with ethidium bromide staining, and cleaned up before sequencing with ExoSAP-IT (Affymetrix, Santa Clara, USA) according to the manufacturer's protocol. Sequence alignments were made with Muscle in MEGA7 (version 7.0.18; MEGA software; Kumar, Stecher, and Tamura 2016).

3.3. Results

3.3.1. *Specificity of the markers* SPAB_01124 *and* STM 3356

The aim of this study was to develop a multiplex qPCR test, to rapidly identify *S.* Paratyphi B (dT-/dT+) based on the previously reported markers *SPAB_01124* (Zhai et al. 2014) and *STM 3356* (Malorny, Bunge, and Helmuth 2003). Prior to the development of this test, the specificity of these 2 markers was tested with their respective PCR tests. The 2 PCRs were performed on 2 *S.* Paratyphi B (dT-), 4 *S.* Paratyphi B (dT+) and 3 other common serotypes (Typhimurium, Enteritidis and Livingstone). Unexpectedly, while all the dT+ isolates were correctly identified by the PCR Malorny, no 384 bp fragments were detected for 3 (2012-45, S16BD08024 and S16BD08272) of the 6 *S.* Paratyphi B isolates analysed with the PCR Zhai (**Table 4**).

Therefore, WGS was performed on these 6 *S.* Paratyphi B isolates in order to investigate why no amplification was detected for 3 of them. The 6 genomes were *de novo* assembled and multiple aligned with 16 publicly available *S.* Paratyphi B genomes (**Table 1**). The *SPAB_01124* gene locus was screened on this multiple alignment and it appeared that this gene was present in all the *S.* Paratyphi B genomes except for the genomes of the 2012-45, S16BD08024 and S16BD08272 isolates as well as the publicly available genome ERR403703 (**Figure 1**).

3.3.2. Genomic study for a marker specific of S. *Paratyphi B*

As the *SPAB_01124* gene appeared not to be a suitable marker for the detection of the Paratyphi B serotype, a comparative genome study was performed to find a specific genetic marker for this serotype. In addition to the 6 genomes already sequenced, WGS was performed on 7 additional *S.* Paratyphi B genomes, achieving a total of 13 WGS datasets (5 dT- and 8 dT+). None of the genetic markers retrieved with Gegenees were specific for all *S.* Paratyphi B or suitable for the design of qPCR probes and primers, after checking the candidate sequences in the multiple alignments of the respective genomes.

Table 4: PCRs Zhai and Malorny tested on Paratyphi B dT-, Paratyphi B dT+ and other serotypes isolates.

Bacterial isolates		PCR Zhai[a] (*SPAB_01124*)		PCR Malorny[a] (*STM 3356*)	
		Expected	Obtained	Expected	Obtained
***S.* Paratyphi B (dT-)**	**II-37-NH**	384 bp	~ 380 bp	No fragment	No fragment
***S.* Paratyphi B (dT-)**	**2012/2966**	384 bp	~ 380 bp	No fragment	No fragment
***S.* Paratyphi B (dT+)**	**2012/45**	384 bp	No fragment	290 bp	~ 290 bp
***S.* Paratyphi B (dT+)**	**2012/60**	384 bp	~ 380 bp	290 bp	~ 290 bp
***S.* Paratyphi B (dT+)**	**S16BD08024**	384 bp	No fragment	290 bp	~ 290 bp
***S.* Paratyphi B (dT+)**	**S16BD08272**	384 bp	No fragment	290 bp	~ 290 bp
***S.* Typhimurium**	**S15BD01386**	No fragment	No fragment	NA	NA
***S.* Enteritidis**	**S15BD02868**	No fragment	No fragment	NA	NA
***S.* Livingstone**	**S15BD01242**	No fragment	No fragment	NA	NA

[a]: fragments expected or obtained after electrophoresis on agarose gel and estimation of their size by comparison with a molecular weight marker. Performed in duplicate in independent assays.
NA: Not Analysed, the d-tartrate fermentation ability test is only performed on *S.* Paratyphi B confirmed isolates.

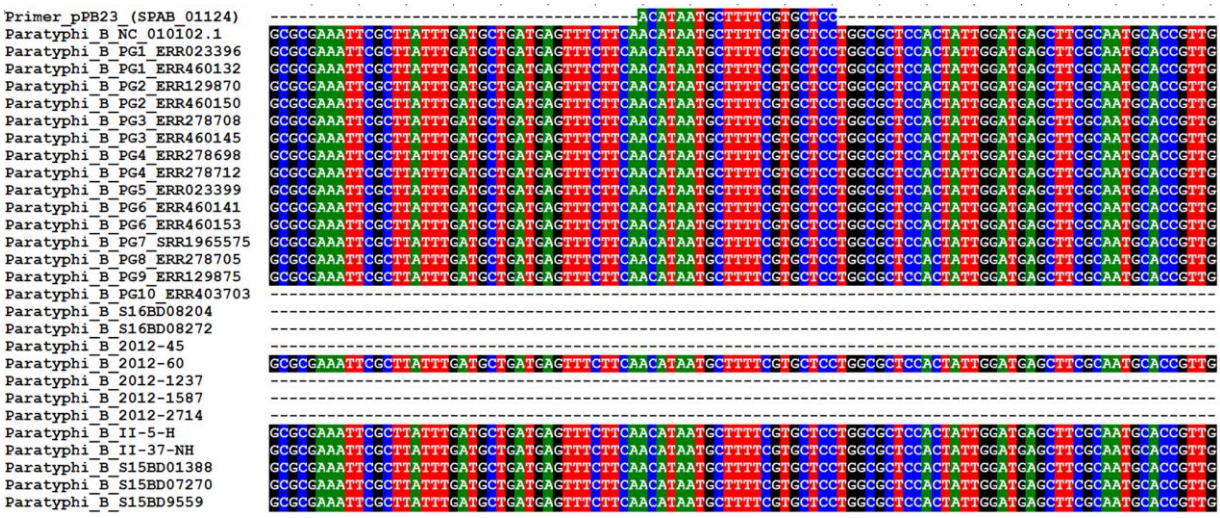

Figure 1: Presence of the marker *SPAB_01124* in the *S.* Paratyphi B genomes.
Alignment of the primer pPB23 (used in PCR Zhai and based on the marker *SPAB_01124*) against the multiple alignment of the *de novo* assembled in-house sequenced *S.* Paratyphi B genomes and the 16 publicly available *S.* Paratyphi B genomes using BioNumerics.

Consequently, a second strategy was applied. A mutation list containing more than 3 million of SNP positions in the genomes was generated from the multiple alignment. The filtering of this list retrieved only one position for which a SNP was present in all genomes of the TARGET group (the Paratyphi B genomes) and absent in those 44 of the BACKGROUND group (genomes belonging to other serotypes, **Table 2**). This position, located in a transporter gene (*SPAB_04460*), was selected as a genetic marker for *S.* Paratyphi B (**Figure 2**).

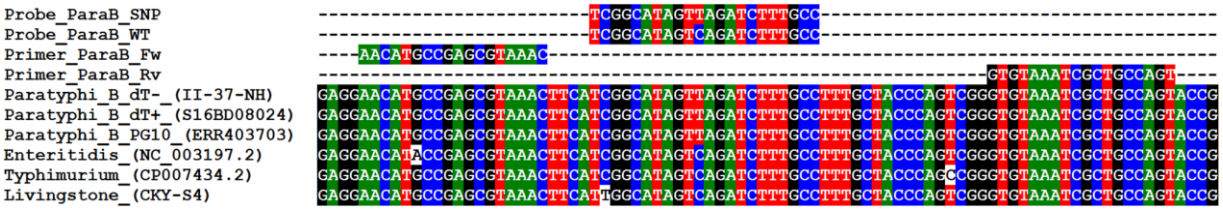

Figure 2: Alignment of primers and probes designed for marker *SPAB_04460* against sequences of the serotypes mentioned in Table 4.
The designed primers (ParaB Fw and Rv) are amplifying a fragment of 79 bp. The probe ParaB contains in the middle of its sequence the SNP specific for *S.* Paratyphi B. The SNPs located in the annealing sites of some serotypes did not affect the efficiency of the qPCR assay, as they were not in the 3' end of the primer nor in the middle of the TaqMan probes

3.3.3. qPCR development

The development of the multiplex qPCR assay for the specific identification of the Paratyphi B serotype and the discrimination between dT- and dT+ variants was based on the marker *SPAB_04460* selected in the present study and the marker *STM 3356* from the PCR Malorny (Malorny, Bunge, and Helmuth 2003). Primers and TaqMan probes were designed, amplifying and targeting these markers respectively (**Figure 2**, primers and probe illustrated with marker *SPAB_04460*). The resulting method is a genoserotyping test using allelic discrimination. The multiplex qPCR assay was successfully tested on the 9 isolates already used previously for the specificity tests (**Table 4** and **Supplemental Table S1**).

3.3.4. Comparison between qPCR and classical method for the detection of S. Paratyphi B var. Java

A total of 17 *S.* Paratyphi B (dT-) (i.e. all the strains available in the NRC collection), 53 *S.* Paratyphi B (dT+) and 108 isolates belonging to other serotypes, species or genus were analyzed by the qPCR method achieving a total of 178 strains. The results were compared to those found with the classical methods: the slide-agglutination serotyping technique and the d-tartrate fermenting biochemical tests (only performed on *S.* Paratyphi B isolates) (**Supplemental Table S1**). All the tests have been repeated 3 times in independent assays.

All the tested strains (178) were correctly identified by the qPCR method except for 4 isolates: *S.* Berta, *S.* Meleagridis, *S.* Singapore and *S.* Stanleyville which were wrongly serotyped as *S.* Paratyphi B. The biochemical tests failed to discriminate 1 *S.* Paratyphi B (S15BD06384) isolate (in bold in the **Supplemental Table S1**) in dT- or dT+ whereas the qPCR method identified it as a *S.* Paratyphi B dT+. For this strain, 4 analyses with the lead-acetate test were performed and gave 2 dT+ results and 2 dT- results, while the Remel Jordan's tartrate test gave negative results after 24h of incubation and positive results after 48h of incubation, both at 37°C. For all these problematic strains (4 discordances at the serotype determination level and one unclear dT fermenting status), the qPCR results were confirmed by Sanger-sequencing.

According to these results, the sensitivity (inclusivity) and specificity (exclusivity) of the developed method were determined to be 100% and 96% for the identification of *S.* Paratyphi B serotype respectively and both 100% for the differentiation between *S.* Paratyphi B dT- and *S.* Paratyphi B dT+ variants (see **Supplemental Table S1**). Therefore, the accuracy of this assay was calculated to be 98% for the *S.* Paratyphi B identification and 100% for dT fermenting discrimination profile.

3.4. Discussion

The aim of this study was to develop a fast and accurate method for the discrimination between the dT- and the dT+ (also called Java) variants, and the confirmation of Paratyphi B serotype identification of *Salmonella* isolates. Consequently, the development of a qPCR method, based on the previously reported markers *SPAB_01124* (Zhai et al. 2014) and *STM 3356* (Malorny, Bunge, and Helmuth 2003) was chosen. Unfortunately, preliminary tests showed that the *SPAB_01124* marker was not able to specifically identify all the *S.* Paratyphi B tested. This result was not surprising regarding to the heterogeneous genomic background of this serotype, illustrated by the 10 PGs described by Connor et al. (2016). Our investigations on the *SPAB_01124* marker showed that it was absent in some of our *S.* Paratyphi B genomes and in the *S.* Paratyphi B genome ERR403703 belonging to the PG10. This might suggest that the marker *SPAB_01124* is specific of *S.* Paratyphi B PGs 1 to 9 but not to PG10[1]. The genomic variation among the *S.* Paratyphi B population can also explain why no adequate genetic marker was found with the Gegenees software. Fortunately, whereas the search of specific sequences (genetic markers) was not successful, the study of specific mutations retrieved one SNP (located in the *SPAB_04460* gene) present in all the *S.* Paratyphi B PGs and absent in the genomes belonging to other serotypes taken as BACKGROUND during the study. This valuable marker was used instead of *SPAB_01124* for the detection of *S.* Paratyphi B in the qPCR development.

In this study, the developed qPCR method correctly identified all the *S.* Paratyphi B dT+ (53) and *S.* Paratyphi B dT- (17) tested (100% accuracy). The marker *STM 3356* was even able to resolve an unknown dT fermenting profile, unable to be clearly determined by the biochemical tests, demonstrating

[1]: Additional analyses made afterwards showed that the *S.* Paratyphi B var. Java genomes lacking *SPAB_01124* in the current study belong to ST28 and are multi-drug resistant which is typical for PG10 isolates (Doublet et al. 2013; Connor et al. 2016; Castellanos et al. 2020).

the efficiency of molecular methods vs. classical methods. Indeed, after 4 analyses, no clear results were obtained with the lead-acetate test whereas the commercial Remel Jordan's tartrate test orientated towards dT+ ability after 48h of incubation at 37°C. This illustrates the limits and the poor repeatability of the lead acetate test which were already pointed out by previous studies (Alfredsson et al. 1972; Barker 1985; Malorny, Bunge, and Helmuth 2003). These kind of untypable strains are a major problem in diagnostic laboratories as they cannot be clustered in one of the 2 different pathogenic profiles, i.e. simple gastroenteritis or the more severe typhoid fever. As a consequence, the laboratory is unable to comply with the legislation. By using the qPCR method developed in this study, this issue will be solved. Moreover, as this is a qPCR method, it is easier and faster to perform in the laboratory compared to the PCR combined with gel electrophoresis detection.

Among the 108 other different serotypes tested with the qPCR method, all were correctly identified as non *S.* Paratyphi B except 4 (*S.* Berta, *S.* Meleagridis, *S.* Singapore and *S.* Stanleyville), achieving 2% of false positives (98% accuracy). However, these serotypes were not reported as frequently encountered in Europe in 2016 (EFSA 2017). Indeed, they are not very common as they represent less than 0.1% of the isolated *Salmonella* in Europe between 2002 and 2017 (data extracted from the TESSy database, ECDC). Additionally, *S.* Berta (O:9), *S.* Meleagridis (O:3,10) and *S.* Singapore (O:7) differ from *S.* Paratyphi B (O:4) at their serogroup level and all (including *S.* Stanleyville H1:z_4,z_{23}) differ from *S.* Paratyphi B (H1:b) at their first flagellar phase level. These false positives are therefore not a major issue as the developed qPCR test will be used, in routine laboratories, mainly for isolates already serotyped as 4:b:? by slide-agglutination. For these samples, the qPCR method will confirm the *S.* Paratyphi B identification (which is the most common serotype with formula O:4 and H1:b) and perform the dT variant discrimination on the same day (Day 1). For rare cases in which the second flagellar phase (H2:1,2) is detected by slide-agglutination in first at Day 1, a confirmation of H1:b will be needed the day after (Day 2), using H2 blocking phase culture for excluding *S.* Stanleyville. In this situation, in case of *S.* Paratyphi B confirmation by the classical method, the qPCR test will be used for a fast and accurate dT variant discrimination instead of using the biochemical tests (**Figure 3**).

As such, the qPCR method developed in this study will be highly valuable in National Reference Centers and Laboratories as well as in first-line laboratories. In most cases, the complete identification of *S.* Paratyphi B dT-/dT+ will be obtained accurately in 1 day instead of 3 to 9 days with risks of no clear results. Consequently, this method saves time, money and helps to obtain a clear and accurate dT variant identification. Thanks to this, *S.* Paratyphi B dT+ can be detected in time and corrective measures can allow its rapid exclusion from the food chain as required by the regulation in Belgium (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8).

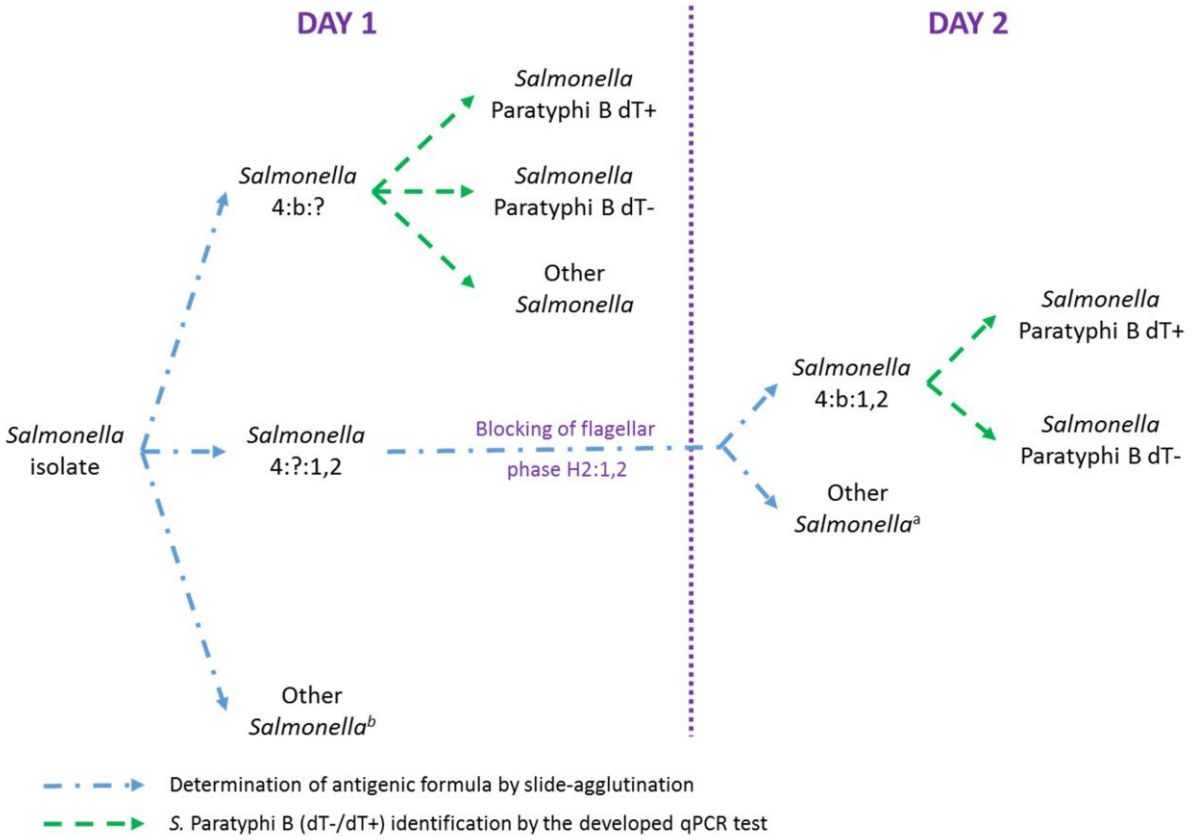

Figure 3: Proposed analysis process for *S.* Paratyphi B dT-/dT+ identification in routine laboratories.

In case of « Other *Salmonella* » the full antigenic formula is determined by classical method

[a] : e.g. *S.* Stanleyville

[b] : e.g. *S.* Berta, *S.* Meleagridis and *S.* Singapore

Acknowledgements

The research that yielded these results, was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 14/5 SALMSTID 1. Sanger and whole genome sequencing were performed at the service Transversal activities in Applied Genomics at Sciensano. The authors want to thank the technicians of the National Reference Center for *Salmonella* and *Shigella*, for their expertise in serotyping by slide-agglutination.

Supplementary data

Supplemental Table S1 can be found online at: https://doi.org/10.1007/s00253-019-09854-4

CHAPTER 4
A multiplex oligonucleotide ligation-PCR method for the genoserotyping of common *Salmonella* using a liquid bead suspension assay

Context of this chapter:

Based on the review of the alternative *Salmonella* serotyping molecular methods (Chapter 1), the Multiplex Oligonucleotide Ligation – PCR (MOL-PCR) combined with the Luminex technology was selected to develop a genoserotyping assay targeting the 6 *Salmonella* serotypes and their variants subjected to an official control in Belgium. An innovative strategy was applied to select highly specific molecular markers from the MLST database EnteroBase. The SNP marker targeting *S.* Paratyphi B, selected from the genomic study described in Chapter 3, was included in the development of the MOL-PCR assay described in the current chapter. The new genoserotyping technique, which includes a barcode system for easy interpretation of the data, was validated by comparison with the classical methods.

This chapter was adapted from the previously published manuscript:

Gand, M., W. Mattheus, N.H.C. Roosens, K. Dierick, K. Marchal, S.C.J. De Keersmaecker* and S. Bertrand*. 2020. "A multiplex oligonucleotide ligation-PCR method for the genoserotyping of common *Salmonella* using a liquid bead suspension assay." *Food Microbiology* 87 (May). https://doi.org/10.1016/j.fm.2019.103394. (* equal contribution)

Authors' contributions:

M. Gand designed the study, performed the experiments, analyzed the data and drafted the manuscript. S.C.J. De Keersmaecker and S. Bertrand conceived and supervised the study, helped to design the study, to interpret the results and to draft the manuscript. K Marchal, W. Mattheus and N. Roosens provided specialist feed-back. All authors read and approved the final manuscript.

Abstract:

Salmonella is a major pathogen having a public health and economic impact in both humans and animals. Six serotypes of the *Salmonella* genus are mentioned in the Belgian and European regulation as to be rapidly excluded from the food chain (EU regulation N°2160/2003, Belgian royal decree 27/04/2017). The reference method for *Salmonella* serotyping, including slide-agglutination and biochemical tests, is time-consuming, expensive, not always objective, and therefore does not match the fast identification criteria required by the legislation. In this study, a molecular method, using genetic markers detected by Multiplex Oligonucleotide Ligation – PCR and Luminex technology, was developed for the identification of the 6 *Salmonella* serotypes and their variants subjected to an official control. The resulting method was validated with the analysis of 971 *Salmonella* isolated from different matrixes (human, animal, food or environment) and 33 non-*Salmonella* strains. The results were compared with the reference identifications, achieving an accuracy of 99.7%. The cost-effective high-throughput genoserotyping assay is performed in 1 day and generates objective results, thanks to the automatic interpretation of raw data using a barcode system. In conclusion, it is fully adapted to the implementation in first-line laboratories and meets the requirements of the regulation.

4.1. Introduction

In 2017, human salmonellosis was still the second bacterial zoonosis reported in Europe after *Campylobacter* infections (EFSA 2018; Eng et al. 2015). *Salmonella*, the causing agent of salmonellosis, can infect both humans and animals which leads to public health issues and economical loss. This pathogen can contaminate a large variety of food products from vegetables to products of animal origins including eggs, dairy products and meat. The genus *Salmonella* is divided according to a complex classification system including 2 species, 6 subspecies and more than 2500 serotypes. The species and subspecies are characterized by biochemical tests and the serotypes are determined by slide-agglutination, following the White-Kauffmann-Le Minor (WKL) scheme (Ryan, O'Dwyer, and Adley 2017; Grimont and Weill 2007). The most common serotypes belong to the 1500 *Salmonella enterica* subsp. *enterica* and include Enteritidis, Hadar, Infantis, Virchow, Typhimurium including its monophasic variant 1,4,[5],12:i:- and Paratyphi B including its variant which has the ability to ferment the dextrorotatory [(L(+)]-tartrate also named Java or dT+. These 6 serotypes and their variant must be monitored because they are the most frequently isolated in Belgium and mentioned in the European and Belgian legislations as to be rapidly excluded from the food chain (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). Indeed, in Belgium, if one of these *Salmonella* serotypes is detected in adult breeding of *Gallus gallus*, the entire flock must be eliminated by logistic slaughtering, the farm must be completely decontaminated and absence of these serotypes must be proven before arrival of a new poultry batch. The same goes for laying hens of *Gallus gallus*, but only if *S.* Enteritidis, *S.* Typhimurium or *S.* 1,4,[5],12:i:- are detected. Additionally, eggs coming from laying hens, which are positive to their corresponding targeted serotypes elaborated above, can only be used for human consumption after undergoing a thermic treatment. Regarding this situation, a rapid and accurate identification of these *Salmonella* serotypes is required to avoid economical loss for the breeders and to be in agreement with the regulation.

The classical methods for *Salmonella* serotyping consist of biochemical tests and slide-agglutination which are very complex, subjective, expensive and time-consuming. Therefore, most of the *Salmonella* isolates must be sent by the first-line laboratories to National Reference Centers (NRCs) where the reference techniques are fully mastered and the complete serotype identification can be obtained. This additional step has the effect to increase the analysis time and cost. Additionally, only these NRCs hold the complete collection of the expensive antisera needed for the slide-agglutination, contrarily to the first-line laboratories which usually are a smaller structure with limited resources. Luckily these last years, molecular techniques based on the detection of molecular markers, specific to the targeted serotypes, have proven to be a better alternative as they yield more objective and accurate results (Wattiau, Boland, and Bertrand 2011). In addition these new techniques are sometimes cheaper or faster and consequently more adapted to the rapid exclusion of serotypes demanded by the legislation

and can be done directly at the first line level. Most of the time, these genoserotyping methods are PCR based, therefore requiring a detection by electrophoresis on agarose gel, or real-time PCR based, and they target only one or few serotypes at the same time, or several assays are needed for complete identification (Fitzgerald et al. 2007; Franklin et al. 2011; Gand et al. 2019; Malorny, Bunge, and Helmuth 2003; Maurischat et al. 2015; Rajtak et al. 2011; Wattiau, Boland, and Bertrand 2011). To avoid the use of multiple assays, a better option is the Multiplex Oligonucleotide Ligation – PCR (MOL-PCR), using a liquid bead suspension assay (Luminex xTAG technology), which allows a high level of multiplexing (Wuyts, Mattheus, et al. 2015; Yoshida et al. 2014). The MOL-PCR consists of the detection of molecular markers through a ligation-dependent amplification reaction, in combination with the xTAG technology. The latter is based on color-coded microspheres, divided into distinct (color) sets. Each bead set is coated with an oligonucleotide allowing the specific capture of MOL-PCR fragments after hybridization, and detection via fluorescence measurements on a Luminex instrument. A commercial Luminex kit already exists for the identification of the 100 most common *Salmonella* serotypes (kit: xMap Salmonella Serotyping Assay Kit). Nevertheless, this kit is too expensive, not modular nor adjustable because commercial without detailed probe information, and is too labor-intensive (3 multiplex assays and using the Lx200 apparatus) for the first-line laboratories which aim to detect in priority the 6 *Salmonella* serotypes and their variants mentioned in the legislation elaborated above. Data interpretation of the results generated by this kit is not automated and can be complex. Consequently, there is a need for a simpler method, focused on the identification of mandatory *Salmonella* serotypes subjected to an official control, that can be adapted following the evolution of the legislation criteria or serotype prevalence.

In this study, a MOL-PCR assay using the Luminex technology was developed for a fast, accurate and cheap detection of *Salmonella* isolates previously isolated from human, animal, food or environmental samples and belonging to the serotypes, and their variants, to be combatted as outlined by the European and Belgian regulations. The molecular markers targeted by the method were selected from the MultiLocus Sequence Typing (MLST) database named EnteroBase or from the scientific literature. The detection of the MOL-PCR fragments is performed on a Luminex platform called the MagPix which allows the simultaneous detection of up to 50 molecular markers. Compared to the Lx200, the MagPix apparatus is smaller, less expensive and therefore more suitable and cost-effective for implementation in first-line laboratories. Moreover, the results' interpretation is facilitated by a barcode system using the Gödel Prime Product (GPP) (Van den Bulcke et al. 2008; Van Den Bulcke et al. 2010) and ensuring an objective conversion of the fluorescence data into serotype identifications. The developed method was validated with the analysis of 1004 bacterial isolates composed of 971 *Salmonella* strains from 114 different serotypes and 33 non-*Salmonella* strains.

4.2. Materials and methods

4.2.1. Bacterial strains

The strains used in this study are reference isolates coming from the collection of the Belgian NRC for *Salmonella* and *Shigella*. The *Salmonella* strains were previously isolated by the first-line laboratories between 2005 and 2018 from food, animal or human matrixes, or from environment. These isolates were sent to the NRC for further characterization, including serotyping by slide-agglutination following the WKL scheme, and confirmation of *Salmonella* genus identification was done by MALDI-TOF method if needed. All isolates are available upon request. The *Salmonella* strains and strains from other bacterial genus were cultured on Nutrient agar (Neogen® Culture Media, Lansing, USA).

4.2.2. DNA extraction

For MOL-PCR, the bacterial DNA was extracted by heat lysis (as described by Wuyts, Mattheus, et al. 2015) from an overnight (14 to 20 hours) culture at 37°C. Briefly, a single colony was sampled in 60 µl sterile de-ionized water and incubated at 95°C in a heating block for 10 minutes. After cooling for minimum 20 minutes at 4°C (in the fridge) and centrifugation for 10 minutes at 11000 × g, the supernatant was used immediately or stored at −20°C for further analysis. Positive controls were made by mixing in one tube a single colony of specific strains in function of their targets characteristics (**Table 1**). The DNA of the strain's mix is extracted by heat lysis as previously described.

For Whole Genome Sequencing (WGS), genomic DNA was extracted with the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich, Saint-Louis, USA) according to the manufacturer's instructions.

4.2.3. Selection of molecular markers from EnteroBase and NCBI database

The MLST database, EnteroBase (Achtman et al., 2012; Alikhan et al., 2018; https://enterobase.warwick.ac.uk), was screened to select conserved housekeeping alleles among the genetically close related strains of a targeted serotype which are clustered together in Sequence Types (STs) or eBurst Groups (eBGs). In order to find a Single Nucleotide Polymorphism (SNP) that can discriminate this allele, and hence the genoserotype to which it belongs, genetic alignments were made using MUSCLE (Edgar 2004) in the MEGA6 software (Tamura et al. 2013) with these conserved alleles against all the alleles present in EnteroBase for a given housekeeping gene (a total average of 500 alleles per housekeeping gene, in early 2016). The specificity of the selected MLST SNP markers were therefore confirmed *in silico* on the whole MLST Database composed of 31 848 entries, after in-house curation, in early 2016 (**Table 2**).

Table 1: Negative and positive controls composition

Mix Number	CTRL name	Targeted characteristics	Strains ID	Probes controlled for positive signal*
-	CTRL_- Vibrio	*Vibrio alginolyticus*	M/5035	NA
1	CTRL_+ H	*S.* Hadar	S17BD01821	STID3, STID334_WT, STID34_SNP, STID4_SNP, STID491_WT, STID5, STID71_WT, STID191
2	CTRL_+ PB	*S.* Paratyphi B dT-	II-37-NH	invA, SAL-73, STID16, STID334_SNP, STID34_WT, STID4_WT, STID71_SNP
3	CTRL_+ 12	*S.* Anatum *S.* Enteritidis *S.* Enteritidis ST183	S16BD07249 S17BD07653 S16BD09144	STID171, STID2, STID31, STID491_SNP
4	CTRL_+ 24	*S.* Infantis *S.* Indiana	S17BD01991 S17BD06592	rpoB, STID50, STID6
5	CTRL_+ 34	*S.* Minnesota *S.* Virchow	S17BD02503 S17BD00950	STID13, STID15, STID18, STID35

dT-: non-fermenting d-tartrate strain
CTRL_-: negative control
CTRL_+: positive control
NA: Not Applicable
*: probes for which a positive signal is expected with the associated CTRL_+

Other markers selected from the scientific literature are based on a SNP or on the presence or absence of a complete sequence (Abs/Pres) specific to a serotype. The specificity of these markers was checked *in silico* by BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the NCBI Database (**Table 3**).

4.2.4. Design of ligation probes

The first step of the MOL-PCR consists in the adjacent annealing of the upstream and downstream ligation probes, to the genomic DNA, for a specific detection of the selected molecular marker. These ligation probes were designed using Visual OMP (version 7.6.58.0; DNA Software) according to the guidelines of Wuyts, Mattheus, et al. (2015) (**Table 2 and 3**). For markers coming from the scientific literature, ligation probes are based on existing probes, primers or specific amplified sequences. For some SNP markers, for which an interpretation using allelic discrimination is needed, a probe with the wild-type (WT) allele was also designed. All probes and primers were ordered from Integrated DNA Technologies (IDT, Leuven, Belgium) with a standard desalted purification.

Table 2: Ligation probes designed from molecular marker selection using EnteroBase

Target	Probe	DNA sequence	Beads region[1]	Interpretation[2]	MLST gene	Allele number	SNP position[3]
S. Enteritidis ST183	STID491-U_SNP	*TAATACGACTCACTATAGGg*tgattgaatagtagattgtttaaCAGCTTCGCCGAAACGGCGGA**G**	46	AC	*purE*	60	132
	STID491-U_WT	*TAATACGACTCACTATAGGg*ttatgatatagtgagttgttattCAGCTTCGCCGAAACGGCGGAA	77				
	STID491-D	P-GAGAACGGATATCAAGTGATTATTGC*TCCCTTTAGTGAGGGTTAAT*					
S. Hadar	STID3-U_SNP	*TAATACGACTCACTATAGGG*agtagaaagttgaaattgattatgCTTCTCGCTGTCCACGCT**T**	12	SNR	*dnaN*	5	183
	STID3-D	P-CCTGCCGCCGATTTCCCGA*TCCCTTTAGTGAGGGTTAAT*					
	STID4-U_SNP*	*TAATACGACTCACTATAGGg*attgatatttgaatgtttgtttgCACCACCGAGATCCCGGC**A**	22	AC	*thrA*	12	423
	STID4-U_WT*	*TAATACGACTCACTATAGGg*tatgttgtaatgtattaagaaa*g*CACCACCGAGATCCCGGCG	25				
	STID4-D*	P-CGAGACATGGCGGCGAAAAC*TCCCTTTAGTGAGGGTTAAT*					
	STID5-U_SNP	*TAATACGACTCACTATAGGg*atagatttagaatgaattaagtgACGCCTGATGCTGACATTGCC**T**	28	SNR	*hisD*	7	385
	STID5-D	P-GCAAGGTGGCGGAGGCGGTAGA*TCCCTTTAGTGAGGGTTAAT*					
S. Indiana	STID50-U_SNP	*TAATACGACTCACTATAGGGa*gtgaatgtaagattatgtatttgTGAATGTTATCAACGATGTG**A**	13	SNR	*hemD*	11	351
	STID50-D	P-GAAACATTACGATGGCGCGG*TCCCTTTAGTGAGGGTTAAT*					
S. Infantis	STID6-U_SNP	*TAATACGACTCACTATAGGG*aattagaagtaagtagagtttaagGCGCCAAATTCCCGGG**G**	56	SNR	*sucA*	21	69
	STID6-D	P-GCGAAACGTTTCTCGCTCGA*TCCCTTTAGTGAGGGTTAAT*					
	STID71-U_SNP	*TAATACGACTCACTATAGGG*attgtgaaagaaagagaagaaattGTCGAAACTGGCCTGTG**G**	14	AC	*hemD*	22	99
	STID71-U_WT	*TAATACGACTCACTATAGGg*tgttatagaagttaaatgttaagTCGAAACTGGCCTGTGT	30				
	STID71-D	P-CGCCGCGCTATTTCGCG*TCCCTTTAGTGAGGGTTAAT*					
S. Virchow	STID13-U_SNP*	*TAATACGACTCACTATAGGG*aataagagaattgatatgaagatgGGCCATCGAAAACTGGGTCG**A**	35	SNR	*dnaN*	7	271
	STID13-D*	P-TTCAATCAGGCGCTTCATCGTGGC*TCCCTTTAGTGAGGGTTAAT*					
	STID15-U_SNP*	*TAATACGACTCACTATAGGg*taagattagaagttaatgaagaaCGCTAAACATCGCCATGTT**A**	52	SNR	*thrA*	14	339
	STID15-D*	P-TTAAGGTTAGAGATCCCTTTAACCGGC*TCCCTTTAGTGAGGGTTAAT*					

*: the probes anneal on the reverse complement strand of the MLST gene

[1]: correspond to the specific color and TAG sequence of the bead

[2]: the presence of the molecular marker is determined by the calculation of the Signal to Noise Ratio (SNR) or Allele Call (AC)

[3]: correspond to the location of the SNP in multiple alignments of all alleles related to the considered MLST gene

P- : Phosphate

Primer (T7 and T3), anti-TAG, target-specific sequences and SNP positions are indicated by italic, lower-case, underlined and bold sequences, respectively.

Table 3: Ligation probes designed from molecular markers selected in the scientific literature

Target	Probe	DNA sequence	Beads region[1]	Type[2]	Interpretation[3]	Marker	Based on	Source
Enteritidis	STID2-U	*TAATACGACTCACTATAGGGt*attagagtttgagaataagtagtCGGCGCATTCCTCCGTTT	33	Abs/Pres	SNR	*sdr*	primer sdr-R	Rajtak et al. 2011
	STID2-D	P-TTTCGTCGTGGGCGTCAGTA*TCCCTTTAGTGAGGGTTAAT*						
Paratyphi B (including var. Java)	STID334-U_SNP	*TAATACGACTCACTATAGGg*ttgtaaattgtagtaaagaagtaGCGTAAACTTCATCGGCATAGT**T**	15	SNP	AC	*SPAB_04460*	TaqMan probes ParaB_SNP and ParaB_WT	Gand et al. 2019
	STID334-U_WT	*TAATACGACTCACTATAGGg*ttagttatgatgaatattgtgtaGCGTAAACTTCATCGGCATAGTC	45					
	STID334-D	P-AGATCTTTGCCTTTGCTACCCA*TCCCTTTAGTGAGGGTTAAT*						
	STID34-U_SNPdT-	*TAATACGACTCACTATAGGG*tgtatatgttaatgagatgttgtaTCATATATCATTGATTGGATAATTATAAATAT**A**	29					
	STID34-U_WTdT+	*TAATACGACTCACTATAGGG*tttaagtgagttatagaagtagtaTCATATATCATTGATTGGATAATTATAAATATG	37	SNP	AC	*STM3356*	Primer #166	Malorny, Bunge, and Helmuth 2003
	STID34-D	P-GAACCCATTACCCTTACCTTATGTT*CCCTTTAGTGAGGGTTAAT*						
Salmonella	invA-U	*TAATACGACTCACTATAGGG*gataagaaagtgaaatgtaaattgATAAACTTCATCGCACCGTCA	51	Abs/Pres	SNR	*invA*	Ligation probes invA-U and invA-D	Wuyts, Mattheus, et al. 2015
	invA-D	P-AAGGAACCGTAAAGCTGGCTT*TCCCTTTAGTGAGGGTTAAT*						
Serogroup O:4	STID16-U	*TAATACGACTCACTATAGGG*tttgatttaagagtgttgaatgtaTCAAGTTGGAACTGGTGCT	26	Abs/Pres	SNR	*rfbJ*	Sequence amplified by B_rfbJ_F and B_rfbJ_R	Franklin et al. 2011
	STID16-D	P-GGGGTAAGTTTGAAAGATTTTCTGG*TCCCTTTAGTGAGGGTTAAT*						
Serogroup O:9	STID171-U_SNP	*TAATACGACTCACTATAGGG*aattgagaaagagataaatgatagCATATACTAAACAAAAAGCAAATGAA**C**	72	SNP	SNR	*prt*	Serogroup D capture probe	Fitzgerald et al. 2007
	STID171-D	P-TCGCCGCCGCCATTATAGA*TCCCTTTAGTGAGGGTTAAT*						

(continued on next page)

Table 3 (continued)

Target	Probe	DNA sequence	Beads region[1]	Type[2]	Interpretation[3]	Marker	Based on	Source
Serogroup O:6,7	STID18-U	*TAATACGACTCACTATAGGg*taagagtattgaaattagtaagaCGTTGGCAGACTGGTACTGATTG	66	Abs/Pres	SNR	*wbaA*	Primer C1_wbaA_F3	Franklin et al. 2011
	STID18-D	P-GCTCCCCTATTACGATGATTTC*TCCCTTTAGTGAGGGTTAAT*						
Serogroup O:8	STID191-U	*TAATACGACTCACTATAGGG*aaataagaatagagagagaaagttTTATAAATTTACGTTTAGAACATGTTTAC	43	Abs/Pres	SNR	*rfbJ*	Sequence amplified by C2_rfbJ_F and C2_rfbJ_R	Franklin et al. 2011
	STID191-D	P-GGTGAGAGGGATAAAGCAGGTAAAAT*CCCTTTAGTGAGGGTTAAT*						
Serogroup O:10	STID31-U	*TAATACGACTCACTATAGGG*tgatatagtagtgaagaaataagtTCTCTACGCAGACAATTATGTCA	34	Abs/Pres	SNR	*wzx*	Primer E_wzx_F	Franklin et al. 2011
	STID31-D	TGGAGTTATTATCCGGATGGG*TCCCTTTAGTGAGGGTTAAT*						
Serogroup O:21	STID35-U	*TAATACGACTCACTATAGGg*ttgagaattagaatttgataaagCCACTGTCATTGGTGGTTATGAG	73	Abs/Pres	SNR	*wzx*	Primer L_wzx_F2	Franklin et al. 2011
	STID35-D	P-TATGAATGGCTGGTATACGACATC*TCCCTTTAGTGAGGGTTAAT*						
Typhimurium (including var. monophasic)	rpoB-U	*TAATACGACTCACTATAGGG*gtaattgaattgaaagataagtgtTTTCTCAGCTGCACCGTAG**C**	18	SNP	SNR	*rpoB*	Ligation probes rpoB-U and rpoB-D	Wuyts, Mattheus, et al. 2015
	rpoB-D	P-CCTGGCGTCTTCTTTGACTCC*TCCCTTTAGTGAGGGTTAAT*						
	SAL-73-U	*TAATACGACTCACTATAGGG*tgaaatgtgtatttgtatgtttagCCAGCCGCAAGGGTTACTGTAC	62	Abs/Pres	SNR	*fljB*[4]	Ligation probes SAL-73-U and SAL-73-D	Wuyts, Mattheus, et al. 2015
	SAL-73-D	P-CGTCAGTAGCAACGTTAACTTCATAA*TCCCTTTAGTGAGGGTTAAT*						

[1]: correspond to the specific color and TAG sequence of the bead
[2]: Single Polymorphism Nucleotide (SNP) or presence or absence of a sequence (Abs/Pres)
[3]: the presence of the molecular marker is determined by the calculation of the Signal to Noise Ratio (SNR) or Allele Call (AC)
[4]: targeting fljB sequence coding for flagellar antigen H:1,2 but also positive for some *Salmonella* isolates belonging to H:1,5 H:1,6 and H:1,7
P- : Phosphate
Primer (T7 and T3), anti-TAG, target-specific sequences and SNP positions are indicated by italic, lower-case, underlined and bold sequences, respectively

4.2.5. MOL-PCR protocol and Luminex read-out

The MOL-PCR assay, the hybridization to MagPlex-TAG microspheres (Luminex, Austin, USA) and the staining reaction using streptavidin-R-phycoerythrin (SAPE) (ThermoFisher Scientific, Waltham, USA) were performed following a modified version of the protocol described by (Wuyts, Mattheus, et al. 2015). Briefly, all the probes (**Table 2** and **Table 3**), except invA-U and invA-D, were mixed together at a final concentration of 50 nM and stored at -20°C in single use aliquots. This probe mix was added to the ligation mix like a single probe at a final concentration of 2 nM. Probes invA-U and invA-D were identically diluted and stored but added separately to the ligation mix to avoid high background noise. DNA from *Vibrio alginolyticus* strain M/5035 was extracted and used in the assay, like described for other samples, as a negative control (CTRL_-) for all probes. Identically, five mixes of *Salmonella* DNA (see composition of the mix in **Table 1**) were prepared as described in section 4.2.2 and used in the assay as positive controls (CTRL_+), in a way that each marker targeted by the method was present in one of these DNA mixes. The ligation and PCR reactions were performed in a thermal cycler SimpliAmp (Applied Biosystems, Foster City, USA).

For the Luminex read-out, a volume of 100 µL of MOL-PCR amplicons hybridized to Luminex beads was analyzed by a MagPix device (Luminex, Austin, USA) at 37°C with a minimum beads count of 50 and a sample wash, according to the manufacturer's instructions.

4.2.6. Data processing and interpretation

For each sample, the MagPix gives the Median Fluorescence Intensity (MFI) corresponding to each marker. For a Signal-to-Noise Ratio (SNR) interpretation, the signal to noise ratio of the marker "a" was calculated by dividing the MFI of the sample "x" by the corresponding MFI of the CTRL_- (**Equ. 1**):

Equ. 1 $$\mathrm{SNR}_{sample\ x\ marker\ a} = \frac{\mathrm{MFI}_{Sample\ x_{marker\ a}}}{\mathrm{MFI}_{CTRL\_-marker\ a}}$$

For some SNP markers which need an allelic discrimination interpretation, a SNP Allele Call (AC) was calculated by dividing the SNR of the SNP probe by the sum of the SNR of the SNP probe and the SNR of its corresponding WT probe (**Equ. 2**):

Equ. 2 $$\mathrm{AC\_SNP}_{sample\ x\ SNP\ a} = \frac{\mathrm{SNR}_{Sample\ x\ SNP\ a}}{\mathrm{SNR}_{Sample\ x\ SNP\ a} + \mathrm{SNR}_{Sample\ x\ WT\ a}}$$

It is recommended to everyone who wants to implement this method, to perform a small validation procedure, including at least 25 *Salmonella* strains per targeted serotypes, 75 untargeted *Salmonella* strains and 25 bacterial strains belonging to another genus, in order to determine the proper cut-off

values which can vary a bit between laboratories. Indeed, Wuyts, Roosens, et al. (2015) showed that the intensity of the fluorescence signals generated by the assay can depend on some intrinsic factors like for example the heating and cooling rates of the thermocycler used during the MOL-PCR step. Here, the cut-off values were calculated for each marker from the MFI generated during the validation procedure and by taking into account the variability obtained with the background noise (MFI signal of CTRL_-). Briefly, for a SNR interpretation, the maximum MFI of the negative samples (increased by 10 % when lower than 200) and the minimum MFI of the positive samples (decreased by 10% when higher than 400) were determined, and respectively divided by the weakest MFI signal encountered in CTRL_- and the highest MFI value allowed in CTRL_- . The cut-off values were calculated as the average of these 2 values. A marker is present in the genome of the *Salmonella* isolate, when the SNR of its corresponding probe is above or equal to its determined cut-off value. For an AC interpretation, the cut-off values were determined as the average of the maximum AC value of negative samples and the minimum AC value of positive samples, encountered during the validation process. If the SNP allele call is higher than the cut-off value, then the SNP allele is assigned to the sample.

A barcode system using the GPP (Van Den Bulcke et al. 2010; Van den Bulcke et al. 2008) was used to identify if a combination of molecular markers detected in the analyzed *Salmonella* sample is specific to a serotype (**Table 4**). Similarly as previously described by Wuyts, Roosens, et al. (2015), a prime number was allocated to each marker. For each sample, the prime numbers of all detected markers were all multiplied to give a product which was subsequently compared to the expected GPPs listed in the **Table 4**, which are associated to serotype identifications. In case of an unknown GPP was obtained, the product was divided by each prime number of the corresponding serogrouping probes: i.e. STID16, STID18, STID31, STID35, STID171 and STID191. If the result of one of these divisions is an integer, the molecular marker linked to the probe was present in the genome of the *Salmonella* isolate, and consequently, the serogroup could be determined (**Table 3)**.

All the processing of the data (SNR/AC calculation, GPP attribution, CTRL check and final identification results displaying) was automatically performed using a programmed Excel workbook compatible with the Comma Separated Value (CSV) files generated by the MagPix and containing all the MFI values for each assay. This Excel template is available upon request.

4.2.7. Comparison study

A total of 1 004 bacterial isolates, including 971 *Salmonella* and 33 non-*Salmonella*, were analyzed with the developed genoserotyping test. From this total, 690 *Salmonella* isolates belonged to the serotypes targeted by the method including 134 *S.* 1,4,[5],12:i:-, 106 *S.* Enteritidis, 31 *S.* Hadar, 154 *S.* Infantis, 13 *S.* Paratyphi B, 84 *S.* Paratyphi B var. Java, 123 *S.* Typhimurium and 45 *S.* Virchow. The remaining strains, which are not targeted by the method, were composed of 281 *Salmonella* isolates

Table 4: Genotype profiles which can be detected by the method

Probe ID:	invA	rpoB	SAL-73	STID2	STID3	STID4	STID5	STID6	STID13	STID15	STID16	STID18
Targets:	*Salmonella*	Typhimurium *	*fljB*	Enteritidis	Hadar	Hadar	Hadar	Infantis	Virchow	Virchow	O:4**	O:7**
Probe prime numbers:	3	23	5	67	41	31	37	47	19	59	7	43
	+	+	-	-	-	-	-	-	+[1]	-	+	-
	+	+	-	-	-	-	-	-	+[2]	-	+	-
	+	+	-	-	-	-	-	-	-	-	+	-
	+	+	-	-	-	-	-	-	-	-	+	-
	+	-	-	+	-	-	+	-	-	-	-	-
	+	-	-	-	-	-	+	-	-	-	-	-
	+	-	-	-	+	+	+[5]	-	-	-	-	-
	+	-	-	-	+	+	-	-	-	-	-	-
	+	-	+	-	-	-	-	+	-	-	-	+
	+	-	+	-	-	-	-	-	-	-	+	-
	+	-	+	-	-	-	-	-	-	-	+	-
	+	-	+	-	-	-	-	-	-	-	+	-
	+	-	+	-	-	-	-	-	-	-	+	-
	+	+	+	-	-	-	-	-	+[1]	-	+	-
	+	+	+	-	-	-	-	-	+[2]	-	+	-
	+	+	+	-	-	-	-	-	-	-	+	-
	+	+	+	-	-	-	-	-	-	-	+	-
	+	-	+	-	-	-	-	-	+	+	-	+

(continued on next page)

According to the *in silico* analysis (using EnteroBase, 2016) performed during marker selection:
[1]: the SNP #99 of the allele hemD and the SNP #271 of the gene dnaN are present together in 94% of the Typhimurium population
[2]: the SNP #271 of the gene dnaN is present in 0.7% of the Typhimurium population
[3]: the SNP #99 of the gene hemD is present in 4% of the Typhimurium population
[4]: the ST183 represents 1.7% of the Enteritidis population
[5]: the SNP #385 of the gene hisD is present in 94% of the Hadar population
[6]: the SNP #99 of the gene hemD is present in 54% of the Paratyphi B (dT-/dT+) population

Table 4 (continued)

Probe ID:	STID31	STID34	STID35	STID50	STID71	STID171	STID191	STID334	STID491	GPP	Associated serotyping result
Targets:	O:10**	dT- var.	O:21**	Indiana	Infantis	O:9**	O:8**	Paratpyhi B*	Enteritidis		
Probe prime numbers:	73	17	79	61	11	53	29	13	71		
	-	-	-	-	+[1]	-	-	-	-	100947	Monophasic Typhimurium
	-	-	-	-	-	-	-	-	-	9177	Monophasic Typhimurium
	-	-	-	-	+[3]	-	-	-	-	5313	Monophasic Typhimurium
	-	-	-	-	-	-	-	-	-	483	Monophasic Typhimurium
	-	-	-	-	-	+	-	-	-	394161	Enteritidis
	-	-	-	-	-	+	-	-	+	417693	Enteritidis (ST183[4])
	-	-	-	-	-	-	+	-	-	4091349	Hadar
	-	-	-	-	-	-	+	-	-	110577	Probably Hadar*** (ST473)
	-	-	-	-	+	-	-	-	-	333465	Infantis
	-	+	-	-	+[6]	-	-	+	-	255255	Paratyphi B (dT-)
	-	+	-	-	-	-	-	+	-	23205	Paratyphi B (dT-)
	-	-	-	-	+[6]	-	-	+	-	15015	Paratyphi B var. Java (dT+)
	-	-	-	-	-	-	-	+	-	1365	Paratyphi B var. Java (dT+)
	-	-	-	-	+[1]	-	-	-	-	504735	Typhimurium
	-	-	-	-	-	-	-	-	-	45885	Typhimurium
	-	-	-	-	+[3]	-	-	-	-	26565	Typhimurium
	-	-	-	-	-	-	-	-	-	2415	Typhimurium
	-	-	-	-	-	-	-	-	-	723045	Virchow

*: including variants
**: serogroup tageted by the associated probe
***: to be confirmed by classical method (i.e. slide-agglutination)
Monophasic Typhimurium: 1,4,[5],12:i:-
dT-: non d-tartrate fermenting strains

from 108 other serotypes, 1 *S. bongori* and 33 non-*Salmonella* isolates (**Table 5**). Prior to use, the serotype identification of each *Salmonella* isolate was confirmed by the reference methods including slide-agglutination and biochemical tests following the WKL scheme, from the same culture used for DNA extraction. If discordant identifications were obtained between the 2 methods, 2 repetitions of the slide-agglutination analysis were performed, i.e. one blind test and one performed by a different technician, to confirm the results. The accuracy of the method, which means the closeness of agreement between the test result and the reference identification, was determined from inclusivity (sensitivity) and exclusivity (specificity) tests as previously described by (Gand et al. 2019).

4.2.8. Whole Genome Sequencing

The whole genome of 3 *S.* Virchow (S16BD00604, S17BD03634 and S17BD08736) was obtained with an Illumina MiSeq (2 x 300 bp, Nextera XT libraries) (Illumina, San Diego, United States). FASTQ reads from all sequences were deposited at the Sciensano-*Salmonella* BioProject at NCBI (PRJNA509747).

In CLC Genomics Workbench 8.0 (Qiagen, Hilden, Germany), the raw FASTQ reads were first trimmed to quality score limit 0.05 with maximum 2 ambiguous nucleotides and reads with length below 30 nucleotides were discarded. These trimmed reads were then *de novo* assembled with automatic bubble and word size, in mapping mode “map reads back to contigs” with scaffolding and a minimum contig length of 1 000 nucleotides. For each isolate, the serotype and the 7 gene MLST ST were characterized using SISTR (Yoshida, Kruczkiewicz, et al. 2016) with genome assemblies (FASTA format) as input.

Table 5: Determination of the sensitivity, specificity and accuracy of the developed method

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Targeted isolates	*S.* 1,4,[5],12:i:-	O:4	133	100947, 9177, 5313 or 483	*S.* Typhimurium var. monophasic	NA	100947	*S.* Typhimurium var. monophasic	NA	133 TP
	S. 1,4,[5],12:i:-	O:4	1	100947, 9177, 5313 or 483	*S.* Typhimurium var. monophasic	NA	9177	*S.* Typhimurium var. monophasic	NA	1 TP
	S. Enteritidis	O:9	99	394161 or 417693	*S.* Enteritidis	NA	394161	*S.* Enteritidis	NA	99 TP
	S. Enteritidis	O:9	7	394161 or 417693	S. Enteritidis (ST183)	NA	417693	S. Enteritidis (ST183)	NA	7 TP
	S. Hadar	O:8	29	110577 or 4091349	*S.* Hadar	NA	4091349	*S.* Hadar	NA	29 TP
	S. Hadar	O:8	2	110577 or 4091349	Probably *S.* Hadar (ST474)[4]	O:8	110577	Probably *S.* Hadar (ST474)[4]	O:8	2 TP
	S. Infantis	O:7	154	333465	*S.* Infantis	NA	333465	*S.* Infantis	NA	154 TP
	S. Paratyphi B (dT-)	O:4	13	23205 or 255255	*S.* Paratyphi B (dT-)	NA	255255	*S.* Paratyphi B (dT-)	NA	13 TP
	S. Paratyphi B var. Java (dT+)	O:4	83	1365 or 15015	*S.* Paratyphi B var. Java (dT+)	NA	15015	*S.* Paratyphi B var. Java (dT+)	NA	83 TP
	S. Paratyphi B Var. Java (dT+)	O:4	1	1365 or 15015	*S.* Paratyphi B var. Java (dT+)	NA	1365	*S.* Paratyphi B var. Java (dT+)	NA	1 TP
	S. Typhimurium	O:4	122	504735, 45885, 26565 or 2415	*S.* Typhimurium	NA	504735	*S.* Typhimurium	NA	122 TP
	S. Typhimurium	O:4	1	504735, 45885, 26565 or 2415	*S.* Typhimurium	NA	26565	*S.* Typhimurium	NA	1 TP
	S. Virchow	O:7	42	723045	*S.* Virchow	NA	723045	*S.* Virchow	NA	42 TP
	S. Virchow	O:7	2	723045	*S.* Virchow	NA	38055	*Salmonella*	O:7	2 FN
	S. Virchow	O:7	1	723045	*S.* Virchow	NA	645	*Salmonella*	O:7	1 FN

(continued on next page)

Table 5 (continued)

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Untargeted isolates	*Acinetobacter baumanii*	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Bacillus cereus	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Citrobacter koseri	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Enterobacter aerogenes	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Enterococcus faecium	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Escherichia coli	NA	2	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	2 TN
	Klebsiella oxytoca	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Klebsiella pneumoniae	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Listeria monocytogenes	NA	5	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	5 TN
	Morganella morganii	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Neisseria meningitidis	NA	4	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	4 TN
	S. Abony	O:4	1	Unknown GPP	*Salmonella*	O:4	1239	*Salmonella*	O:4	1 TN
	S. Agama	O:4	1	Unknown GPP	*Salmonella*	O:4	105	*Salmonella*	O:4	1 TN
	S. Ago	O:30	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Agona	O:4	11	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	11 TN
	S. Agoueve	O:13	1	Unknown GPP	*Salmonella*	Unknown	33	*Salmonella*	Unknown	1 TN
	S. Ajiobo	O:13	4	Unknown GPP	*Salmonella*	Unknown	39	*Salmonella*	Unknown	4 TN
	S. Albany	O:8	2	Unknown GPP	*Salmonella*	O:8	87	*Salmonella*	O:8	2 TN
	S. Albany	O:8	1	Unknown GPP	*Salmonella*	O:8	957	*Salmonella*	O:8	1 TN
	S. Altona	O:8	1	Unknown GPP	*Salmonella*	O:8	435	*Salmonella*	O:8	1 TN
	S. Anatum	O:3,10	3	Unknown GPP	*Salmonella*	O:10	12045	*Salmonella*	O:10	3 TN
	S. Anecho	O:35	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Bardo	O:8	2	Unknown GPP	*Salmonella*	O:8	17166405	Probably *S.* Bardo/Newport[5]	O:8	2 TN
	S. Bardo	O:8	1	Unknown GPP	*Salmonella*	O:8	148335	Probably *S.* Bardo/ Newport/Blockley[4]	O:8	1 TN

(continued on next page)

Table 5 (continued)

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Untargeted isolates	*S.* Bardo	O:8	1	Unknown GPP	*Salmonella*	O:8	435	*Salmonella*	O:8	1 TN
	S. Bareilly	O:7	2	Unknown GPP	*Salmonella*	O:7	645	*Salmonella*	O:7	2 TN
	S. Berta	O:9	1	Unknown GPP	*Salmonella*	O:9	2067	*Salmonella*	O:9	1 TN
	S. Blockley	O:8	2	Unknown GPP	*Salmonella*	O:8	148335	Probably *S.* Bardo/Newport/Blockley[4]	O:8	2 TN
	S. *bongori*	NA	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Bovismorbificans	O:8	2	Unknown GPP	*Salmonella*	O:8	2697	*Salmonella*	O:8	2 TN
	S. Braenderup	O:7	3	Unknown GPP	*Salmonella*	O:7	1419	*Salmonella*	O:7	3 TN
	S. Brancaster	O:4	1	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	1 TN
	S. Brandenburg	O:4	1	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	1 TN
	S. Bredeney	O:4	3	Unknown GPP	*Salmonella*	O:4	105	*Salmonella*	O:4	3 TN
	S. Brive	O:42	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Butantan	O:3,10	1	Unknown GPP	*Salmonella*	O:10	219	*Salmonella*	O:10	1 TN
	S. Carmel	O:17	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Cero	O:18	2	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	2 TN
	S. Cero	O:18	3	Unknown GPP	*Salmonella*	Unknown	33	*Salmonella*	Unknown	3 TN
	S. Cero	O:18	3	Unknown GPP	*Salmonella*	Unknown	15	*Salmonella*	Unknown	3 TN
	S. Chester	O:4	3	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	3 TN
	S. Choleraesuis	O:7	1	Unknown GPP	*Salmonella*	O:7	645	*Salmonella*	O:7	1 TN
	S. Coeln	O:4	2	Unknown GPP	*Salmonella*	O:4	3255	*Salmonella*	O:4	2 TN
	S. Colindale	O:7	1	Unknown GPP	*Salmonella*	O:7	939765	Probably *S.* Colindale (ST584)[4]	O:7	1 TN
	S. Corvallis	O:8	1	Unknown GPP	*Salmonella*	O:8	87	*Salmonella*	O:8	1 TN
	S. Cotham	O:28	2	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	2 TN
	S. Derby	O:4	11	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	11 TN
	S. Derby	O:4	6	Unknown GPP	*Salmonella*	O:4	231	*Salmonella*	O:4	6 TN

(continued on next page)

Table 5 (continued)

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Untargeted isolates	*S.* Dublin	O:9	5	Unknown GPP	*Salmonella*	O:9	159	*Salmonella*	O:9	5 TN
	S. Dugbe	O:45	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Durban	O:9	1	Unknown GPP	*Salmonella*	O:9	159	*Salmonella*	O:9	1 TN
	S. Durham	O:13	1	Unknown GPP	*Salmonella*	Unknown	141	*Salmonella*	Unknown	1 TN
	S. Eastbourne	O:9	1	Unknown GPP	*Salmonella*	O:9	159	*Salmonella*	O:9	1 TN
	S. Ebrie	O:35	1	Unknown GPP	*Salmonella*	Unknown	33	*Salmonella*	Unknown	1 TN
	S. Escanaba	O:7	1	Unknown GPP	*Salmonella*	O:7	1419	*Salmonella*	O:7	1 TN
	S. Gallinarum var Gallinarum	O:9	1	Unknown GPP	*Salmonella*	O:9	5883	Probably *S.* Gallinarum[4]	O:9	1 TN
	S. Gallinarum var Pullorum	O:9	1	Unknown GPP	*Salmonella*	O:9	5883	Probably *S.* Gallinarum[4]	O:9	1 TN
	S. Gaminara	O:16	1	Unknown GPP	*Salmonella*	Unknown	15	*Salmonella*	Unknown	1 TN
	S. Give	O:3,10	4	Unknown GPP	*Salmonella*	O:10	1095	*Salmonella*	O:10	4 TN
	S. Glostrup	O:8	1	Unknown GPP	*Salmonella*	O:8	94047	*Salmonella*	O:8	1 TN
	S. Goldcoast	O:8	1	Unknown GPP	*Salmonella*	O:8	87	*Salmonella*	O:8	1 TN
	S. Grumpensis	O:13	1	Unknown GPP	*Salmonella*	Unknown	15	*Salmonella*	Unknown	1 TN
	S. Havana	O:13	2	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	2 TN
	S. Heidelberg	O:4	2	Unknown GPP	*Salmonella*	O:4	61845	Probably *S.* Kisangani/Heidelberg/ Saintpaul/Stanleyville[4]	O:4	2 TN
	S. Hvittingfoss	O:16	1	Unknown GPP	*Salmonella*	Unknown	627	Probably *S.* Hvittingfoss[4]	Unknown	1 TN
	S. Ibadan	O:13	1	Unknown GPP	*Salmonella*	Unknown	17205	*Salmonella*	Unknown	1 TN
	S. Idikan	O:13	2	Unknown GPP	*Salmonella*	Unknown	15	*Salmonella*	Unknown	2 TN
	S. Indiana	O:4	5	Unknown GPP	*Salmonella*	O:4	147315	Probably *S.* Indiana[4]	O:4	5 TN
	S. Isangi	O:7	1	Unknown GPP	*Salmonella*	O:7	1419	*Salmonella*	O:7	1 TN
	S. Ituri	O:4	1	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	1 TN

(continued on next page)

Table 5 (continued)

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Untargeted isolates	*S.* Jalisco	O:11	1	Unknown GPP	*Salmonella*	Unknown	15	*Salmonella*	Unknown	1 TN
	S. Javiana	O:9	1	Unknown GPP	*Salmonella*	O:9	159	*Salmonella*	O:9	1 TN
	S. Jerusalem	O:7	1	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	1 TN
	S. Kapemba	O:9	1	Unknown GPP	*Salmonella*	O:9	795	*Salmonella*	O:9	1 TN
	S. Kasenyi	O:38	2	Unknown GPP	*Salmonella*	Unknown	42009	*Salmonella*	Unknown	2 TN
	S. Kedougou	O:13	3	Unknown GPP	*Salmonella*	Unknown	33	*Salmonella*	Unknown	3 TN
	S. Kentucky	O:8	10	Unknown GPP	*Salmonella*	O:8	87	*Salmonella*	O:8	10 TN
	S. Kisarawe	O:11	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Kottbus	O:8	2	Unknown GPP	*Salmonella*	O:8	87	*Salmonella*	O:8	2 TN
	S. Lagos	O:4	1	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	1 TN
	S. Limete	O:4	2	Unknown GPP	*Salmonella*	O:4	4935	*Salmonella*	O:4	2 TN
	S. Litchfield	O:8	3	Unknown GPP	*Salmonella*	O:8	17835	*Salmonella*	O:8	3 TN
	S. Livingstone	O:7	14	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	14 TN
	S. Livingstone	O:7	3	Unknown GPP	*Salmonella*	O:7	1677	*Salmonella*	O:7	3 TN
	S. London	O:3,10	1	Unknown GPP	*Salmonella*	O:10	1095	*Salmonella*	O:10	1 TN
	S. Manhattan	O:8	1	Unknown GPP	*Salmonella*	O:8	87	*Salmonella*	O:8	1 TN
	S. Mbandaka	O:7	8	Unknown GPP	*Salmonella*	O:7	1419	*Salmonella*	O:7	8 TN
	S. Meleagridis	O:3,10	1	Unknown GPP	*Salmonella*	O:10	2847	*Salmonella*	O:10	1 TN
	S. Mgulani	O:38	1	Unknown GPP	*Salmonella*	Unknown	6105	*Salmonella*	Unknown	1 TN
	S. Miami	O:9	1	Unknown GPP	*Salmonella*	O:9	159	*Salmonella*	O:9	1 TN
	S. Mikawasima	O:7	1	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	1 TN
	S. Minnesota	O:21	16	Unknown GPP	*Salmonella*	O:21	237	Probably *S.* Minnesota[5]	O:21	16 TN
	S. Monschaui	O:35	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Montevideo	O:7	1	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	1 TN
	S. Muenchen	O:8	3	Unknown GPP	*Salmonella*	O:8	4785	*Salmonella*	O:8	3 TN

(continued on next page)

Table 5 (continued)

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Untargeted isolates	*S.* Muenster	O:3,10	9	Unknown GPP	*Salmonella*	O:10	219	*Salmonella*	O:10	9 TN
	S. Napoli	O:9	1	Unknown GPP	*Salmonella*	O:9	159	*Salmonella*	O:9	1 TN
	S. Newport	O:8	2	Unknown GPP	*Salmonella*	O:8	256215	Probably *S.* Bardo/Newport[4]	O:8	2 TN
	S. Newport	O:8	4	Unknown GPP	*Salmonella*	O:8	148335	Probably *S.* Bardo/Newport/Blockley[4]	O:8	4 TN
	S. Newport	O:8	1	Unknown GPP	*Salmonella*	O:8	25665	*Salmonella*	O:8	1 TN
	S. Newport	O:8	5	Unknown GPP	*Salmonella*	O:8	435	*Salmonella*	O:8	5 TN
	S. Norwich	O:7	2	Unknown GPP	*Salmonella*	O:7	1419	*Salmonella*	O:7	2 TN
	S. Nyborg	O:3,10	1	Unknown GPP	*Salmonella*	O:10	1095	*Salmonella*	O:10	1 TN
	S. Ohio	O:7	2	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	2 TN
	S. Oranienburg	O:7	2	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	2 TN
	S. Panama	O:9	1	Unknown GPP	*Salmonella*	O:9	795	*Salmonella*	O:9	1 TN
	S. Paratyphi A	O:2	3	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	3 TN
	S. Pomona	O:28	1	Unknown GPP	*Salmonella*	Unknown	15	*Salmonella*	Unknown	1 TN
	S. Poona	O:13	1	Unknown GPP	*Salmonella*	Unknown	15	*Salmonella*	Unknown	1 TN
	S. Potsdam	O:7	1	Unknown GPP	*Salmonella*	O:7	4773	*Salmonella*	O:7	1 TN
	S. Rissen	O:7	6	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	6 TN
	S. Rissen	O:7	1	Unknown GPP	*Salmonella*	O:7	7869	Probably *S.* Rissen (ST1846)[4]	O:7	1 TN
	S. Rubislaw	O:11	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Saintpaul	O:4	4	Unknown GPP	*Salmonella*	O:4	105	*Salmonella*	O:4	4 TN
	S. Sandiego	O:4	1	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	1 TN
	S. Schwarzengrund	O:4	1	Unknown GPP	*Salmonella*	O:4	105	*Salmonella*	O:4	1 TN
	S. Senftenberg	O:1,3,19	6	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	6 TN
	S. Singapore	O:7	1	Unknown GPP	*Salmonella*	O:7	1677	*Salmonella*	O:7	1 TN

(continued on next page)

Table 5 (continued)

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Untargeted isolates	*S.* Soerenga	O:30	1	Unknown GPP	*Salmonella*	Unknown	141	*Salmonella*	Unknown	1 TN
	S. Stanley	O:4	1	Unknown GPP	*Salmonella*	O:4	105	*Salmonella*	O:4	1 TN
	S. Stanleyville	O:4	1	Unknown GPP	*Salmonella*	O:4	273	*Salmonella*	O:4	1 TN
	S. Stanleyville	O:4	2	Unknown GPP	*Salmonella*	O:4	21	*Salmonella*	O:4	2 TN
	S. Stanleyville	O:4	1	Unknown GPP	*Salmonella*	O:4	23541	Probably *S.* Stanleyville[4]	O:4	1 TN
	S. Takoradi	O:8	1	Unknown GPP	*Salmonella*	O:8	87	*Salmonella*	O:8	1 TN
	S. Tees	O:16	1	Unknown GPP	*Salmonella*	Unknown	20163	*Salmonella*	Unknown	1 TN
	S. Telelkebir	O:13	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Tennessee	O:7	5	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	5 TN
	S. Thompson	O:7	1	Unknown GPP	*Salmonella*	O:7	129	*Salmonella*	O:7	1 TN
	S. Typhi	O:9	1	Unknown GPP	*Salmonella*	O:9	159	*Salmonella*	O:9	1 TN
	S. Uganda	O:3,10	2	Unknown GPP	*Salmonella*	O:10	2409	*Salmonella*	O:10	2 TN
	S. Umbilo	O:28	1	Unknown GPP	*Salmonella*	Unknown	93	*Salmonella*	Unknown	1 TN
	S. Urbana	O:30	1	Unknown GPP	*Salmonella*	Unknown	3	*Salmonella*	Unknown	1 TN
	S. Wandsworth	O:39	1	Unknown GPP	*Salmonella*	Unknown	2145	*Salmonella*	Unknown	1 TN
	S. Weltevreden	O:3,10	2	Unknown GPP	*Salmonella*	O:10	219	*Salmonella*	O:10	2 TN
	Serratia odorifera	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Shigella boydii	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Shigella flexneri	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Shigella sonei	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Staphylococcus aureus	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Staphylococcus epidermidis	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Staphylococcus mileri	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN

(continued on next page)

Table 5 (continued)

	Reference identification[1]		Nbr of isolates	Expected result with the Luminex assay			Obtained results with the Luminex assay			Comparison results[3]
	Serotype	Serogroup[2]		GPP	Identification	Serogroup[2]	GPP	Identification	Serogroup[2]	
Untargeted isolates	*Staphylococcus saprophyticus*	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Streptococcus agalactiae	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Streptococcus bovis	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Streptococcus dysgalactiae	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Streptococcus pyogenes	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Vibrio alginolyticus	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Yersinia enterocolitica	NA	1	-	Non-*Salmonella*	NA	-	Non-*Salmonella*	NA	1 TN
	Total number of samples		1004						Total TP	687
									Total TN	314
									Total FP	0
									Total FN	3
									Inclusivity: sensitivity	99,6%
									Exclusivity: specificity	100%
									Accuracy	99,7%

1,4,[5],12:i:-: Typhimurium monophasic

Paratyphi B (dT-): non d-tartrate fermenting Paratyphi B isolates (dT-)

Paratyphi B var. Java (dT+): d-tartrate fermenting Paratyphi B isolates (dT+)

NA: not applicable

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

[1]: determined by classical methods i.e., slide-agglutination and biochemical tests

[2]: Determination of the serogroup (if possible), only mentioned when the sample does not belong to one of the targeted serotypes.

[3]: Expected serotype identification vs. obtained serotype identification

[4]: Probability determined according to the present results and the *in silico* analysis on EnteroBase

[5]: Probability determined according to the present results

4.3. Results

4.3.1. Selection of molecular markers

In this study, molecular markers were selected for the specific detection of the 6 *Salmonella* serotypes and their variants frequently isolated in Belgium and mentioned in the regulation as to combat (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). A total of 12 molecular markers was inspired from the literature for the specific detection of *S.* Enteritidis, *S.* Paratyphi B including the Java variant, *S.* Typhimurium including its monophasic variant and for sample serogrouping (**Table 3)**. These markers are based on a SNP or on the presence or absence of a complete sequence. Their specificity was checked successfully *in silico* using BLASTn on the NCBI Database (data not shown). The rest of the molecular markers (9) was selected from the MLST database (EnteroBase) for the specific identification of *S.* Hadar, *S.* Infantis, *S.* Virchow, a specific ST of *S.* Enteritidis (ST183) and the exclusion of *S.* Indiana, which can be mistaken with a low percentage of *S.* Typhimurium isolates (**Table 2**).

First, specific SNPs were screened among the allele sequences of the 7 housekeeping genes (which constitute the MLST scheme: *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*) that are conserved in a serotype population. After an *in silico* verification on the whole EnteroBase database, it became clear that used separately, no single SNP candidate was sufficient for the specific detection of the targeted serotypes. Consequently, the choice was made to use a combination of SNP markers selected among at least 2 conserved alleles. Used together with the serogroup markers selected from the literature (**Table 3**), these combinations of MLST markers (**Table 4**) gave a false positive rate of 0% for the concerned serotypes, after an *in silico* verification on EnteroBase. Few strains of the targeted serotypes, belonging to rare STs, were reported as false negative: i.e. less than 6% for Hadar serotype and less than 3% for the 4 other serotypes mentioned earlier in this paragraph. These false negative rates were investigated during the validation process (section 4.3.3).

4.3.2. MOL-PCR development

From the selection of molecular markers, a total of 26 upstream and 21 downstream probes were designed (**Table 2** and **Table 3**). In the programmed Excel workbook, a serotype identification is automatically assigned to a sample when a specific combination of positive probes is obtained and converted into a known GPP. If no positive signal is obtained for the marker *invA* targeting all *Salmonella* species, no GPP interpretation is performed. If the obtained GPP is unknown, the serogroup of the sample is determined, if possible, as described in materials and methods. The expected

combinations of molecular markers, the GPPs and the associated serotyping results are listed in **Table 4**. In case of unknown or partial identification, the sample must be serotyped by the classical method.

Based on preliminary results (data not shown), an allelic discrimination, i.e. AC interpretation including the use of a WT probe, was needed to properly detect some SNP markers and to improve the fluorescence signal of probes STID4, STID34, STID334, STID71 and STID491, allowing the characterization of *S.* Hadar, *S.* Paratyphi B, *S.* Infantis and *S.* Enteritidis. All other probes are interpreted using SNR calculation, as elaborated in material and methods.

The probe STID50 targeting *S.* Indiana was introduced in the multiplex to exclude this serotype from a relatively infrequent population of *S.* Typhimurium isolates identified by GPP 2415 and representing less than 1% of the *S.* Typhimurium population, according to an *in silico* analysis performed on EnteroBase in 2019. Additionally, STID491 was added to the assay for the identification of *S.* Enteritidis isolates belonging to ST183, representing only 1.6% of the *S.* Enteritidis isolates on EnteroBase (2019), but circulating in Belgium, (**Supplementary Table S1**) and not detected by STID2 which is targeting the other *S.* Enteritidis isolates. Finally, the marker targeted by STID5 is not able to detect a small part of the *S.* Hadar population belonging to ST473 with a frequency expected to be 6% like previously anticipated during the marker selection from EnteroBase. Consequently, when the GPP 110577 will be obtained, the Hadar genoserotype result will have to be confirmed by slide-agglutination because false positive results can be retrieved when obtaining a positive result for the probe combination including invA, STID191, STID3, STID4 and negative for STID5.

4.3.3. Comparison between the reference and the developed molecular method

With the aim to check the specificity of the developed genoserotyping test compared to the reference methods, i.e. serotyping by slide-agglutination and biochemical tests, 690 *Salmonella* isolates belonging to the targeted serotypes, 281 *Salmonella* isolates belonging to untargeted serotypes and 33 non-*Salmonella* isolates were analyzed with both methods. When an unknown GPP was retrieved, the presence of a serogroup marker was screened among the signals of the serogrouping probes. The obtained results are listed in **Table 5** and a detailed version including the probe combinations is available in **Supplementary Table S1**. According to the results, except for 3/45 *S.* Virchow isolates which gave false negative results (with ID numbers S16BD00604, S17BD03634 and S17BD08736), all the other 687 targeted isolates were correctly identified. For the 314 untargeted isolates, no false positive results were obtained.

To investigate why the 3 *S.* Virchow isolates (S16BD00604, S17BD03634 and S17BD08736) were not detected by STID13 and/or STID15, the identification was confirmed by 2 repetitions of the slide-agglutination and their full genome was sequenced. The upload of the corresponding genome assemblies to the SISTR tool also confirmed the *S.* Virchow identification. Additionally, the MLST function

(7 genes MLST) of SISTR clustered these 3 isolates into a rare ST (ST 2563 for S16BD00604 and S17BD08736, representing only 0.7% of *S.* Virchow strains in EnteroBase (2019)) or into an unknown ST (for S17BD03634) which are not targeted by the SNP markers of STID13 and/or STID15.

Also, of the 97 *S.* Paratyphi B (dT-/dT+) tested, only one isolate yielded the GPP 1365 and no isolate resulted in the theoretically possible GPP 23205 (both lacking marker of STID71) demonstrating that these two populations of *S.* Paratyphi B are not frequently distributed in Belgium while their expected frequency was near 46% according to the percentage obtained during the selection of molecular markers from EnteroBase (2016) (**Table 4** and **Table 5**).

From the 30 *S.* Hadar isolates that were analyzed during this comparison study, 2 (6.6%) were identified as belonging to the ST473 (combination of 4 markers leading to GPP 110577), which is close to the expected frequency determined during the *in silico* analysis using EnteroBase (**Table 4** and **Table 5**). Additionally, for all other *Salmonella* isolates tested, no other genoserotyping results linked to GPP 110577 were obtained. This could lead to the hypothesis that a similar specificity detection of *S.* Hadar can be obtained with 4 markers leading to GPP 110577 as well as with 5 markers leading to GPP 4091349.

Interestingly, some molecular markers combinations, other than those foreseen for the targeted serotypes (**Table 4**), seem to be specific for certain other serotypes (**Table 5**). When possible, for MLST markers in association with serogroup and/or *fljB* marker, the specificity of these combinations was checked *in silico* on EnteroBase (2019) and allowed the prediction of "probable serotype" identification: i.e. *S.* Bardo, *S.* Blockley, *S.* Colindale ST584, *S.* Gallinarum, *S.* Heidelberg, *S.* Hvittingfoss, *S.* Indiana, *S.* Kisangani, *S.* Newport, *S.* Rissen ST1836, *S.* Saintpaul and *S.* Stanleyville eBG79 (**Supplementary Table S1**). Additionally, considering the few serotypes present in the serogroup O:21 (Grimont and Weill 2007) and the detection frequency of *S.* Minnesota in Belgium (personal communication, NRC), it is likely that samples resulting in the GPP 237 belong to Minnesota serotype. All these "probable serotype" predictions were in agreement with the results obtained during the comparison study with the reference methods (**Table 5**). Consequently, the GPPs linked to these predictions were added to the list of possible identifications (**Supplementary Table S1**) which can be interpreted by the automated Excel file.

According to the inclusivity tests, i.e. identification results of targeted *Salmonella* strains, and exclusivity tests i.e. identification results of untargeted strains, the sensitivity and specificity were calculated to be 99.6% and 100%, respectively. Based on these results, the accuracy of the developed method was determined to be 99.7% (**Table 5**). During this comparison study, one technician was able to serotype a maximum of 25 samples per day by slide-agglutination. This did not include the determination of the second antigenic phase or the performing of biochemical tests for variant discrimination, when needed, which required at least 1 to 8 additional days. In comparison, using the developed genoserotyping method for the targeted serotypes, the complete identification could be

obtained for 90 samples in only one working day, in a single 96-well plate. The price of a *Salmonella* serotype identification by classical and new molecular methods was estimated using the data of the Belgian NRC. This estimation included the current cost of the consumables, the reagents, the technicians, the purchase of a MagPix apparatus and its maintenance (personal communication, NRC). Already with the analysis of only one sample, the developed genoserotyping assay is 1.7 times cheaper than the classical method. If 25 samples are considered, the new method is 3.5 times less expensive. Finally, if a full 96-well plate is analyzed, which can be performed in one day by one technician, the cost per sample is 7.5 times cheaper.

4.4. Discussion

In this study, a *Salmonella* genoserotyping tool, based on genetic markers (selected from EnteroBase and scientific literature) and using the Luminex technology, was developed. This tool is able to identify the most common serotypes in Belgium, i.e. Enteritidis, Hadar, Infantis, Virchow, Paratyphi B including its variant Java and Typhimurium including its monophasic variant , that are subjected to an official control (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). To check the specificity of this new method, 1 004 isolates including 971 *Salmonella* and 33 non-*Salmonella* isolates were analyzed. Resulting genoserotyping profiles were compared with serotyping identifications obtained using the reference methods, i.e. slide-agglutination and biochemical tests, and lead to an accuracy of 99.7%. Only 3 *S.* Virchow isolates of 45 tested, representing 6.8% of the tested *S.* Virchow isolates and belonging to rare STs, were not correctly identified by the molecular test. Consequently, when these genoserotypes will be analyzed by the developed method, they will be reported as *Salmonella* belonging to the serogroup O:7 and the Virchow serotype identification will have to be obtained by slide-agglutination.

Some differences were observed concerning the genoserotype distribution between the percentages estimated during the *in silico* analysis and the results obtained with the comparison study. This can be explained by the genoserotype distribution present in EnteroBase which was not always representative of the *Salmonella* population circulating in Belgium between 2005 and 2018, and used during the comparison study. This was especially the case for *S.* Paratyphi B (dT-/dT+).

Interestingly, the developed method was able to make genoserotype predictions, based on specific marker combinations other than those used for the targeted serotypes. Although only a few *Salmonella* isolates were tested to check these predictions, the specificity of the associated marker combinations was validated *in silico* in EnteroBase. In addition to the clustering made by the serogrouping probes, these predictions are an added value to the method, as they can give some clues about the identification of untargeted serotypes and contribute to the confirmation by classical methods. For example, one of

these predictions concerns the serotype Newport which is present in the top 5 of the most common serotypes in Europe, reported by EFSA (2018) for the year 2017, and the closely related serotype Bardo. These 2 serotypes differ by only one somatic antigen (O:6 vs. O:6,8) in their antigenic formula (Grimont and Weill 2007), which explains that they are hardly discriminated with molecular methods. Consequently, when the result "Probably *S.* Bardo/Newport" is retrieved by the GPP automatic interpretation, slide-agglutination of antigens H1:e,h for exclusion of other serotypes and O:6 for Bardo/Newport discrimination, can be performed in the same day to confirm the identification. Identically, the slide-agglutination of targeted antigens, following the WKL scheme, can be done for the verification of the other genoserotyping predictions obtained during the comparison study.

As the developed test uses a barcode system (GPP) and an automated Excel file to process Luminex data generated by the MagPix, the results are more accurate and objective compared to the classical method, for which trained technicians are required to properly interpret subjective slide-agglutinations and biochemical tests. However, this genoserotyping system follows the WKL classification and is, therefore, fully compatible with the regulation and the serotypes mentioned in it (EU regulation N°2160/2003 Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). Moreover, the complete genoserotyping analysis, from DNA extraction to data interpretation, takes only one working day compared to the classical method where at least 2 to 9 days are sometimes required to have a complete identification, including variant determination like for example *S.* Paratyphi B var. Java for which a complex and time-consuming biochemical test is needed (Alfredsson et al., 1972; Barker, 1985). The method is also cost-effective as it is performed in 96-well plates and only one well is required per sample, making it 7.5 times less expensive compared to the classical serotyping. To be able to perform this genoserotyping assay, commonly used equipment such as a PCR instrument and a centrifuge are needed, in addition to the purchase of a MagPix instrument.

Retrospectively considering the serotyping analyses performed by the Belgian NRC during the last 10 years using classical methods, the developed genoserotyping method could have identified more than 77% of the *Salmonella* isolates sent to the center, and could have given serogroup orientation or probable serotype prediction for even more (annual reports and personal communication, NRC). In conclusion, although the developed *Salmonella* genoserotyping method can detect less serotypes than other methods (Yoshida, Simone, et al. 2016; Wattiau, Boland, and Bertrand 2011) or the commercially available *Salmonella* identification kit, it is perfectly adapted to first-line laboratories for which a fast, accurate and cost-effective tool is needed, avoiding the sending of most *Salmonella* samples to the NRC. Like this, the results are rapidly transmitted to the professionals of the food sector who then know if they are in agreement with the criteria of the regulation and they can sell properly their food products as soon as possible. Furthermore, other MOL-PCR assays compatible with the MagPix apparatus exist for pathogen diagnostics (Wessels et al. 2014), subtyping (Ventola et al. 2019; Wuyts, Mattheus, et al. 2015) or antimicrobial resistance screening (Ceyssens et al. 2016). All these compatible tests can be performed

with the MagPix in the same run using the multi-batch function (with a limit of 96 reactions in total) thereby lowering the analysis price per sample. Moreover, the present method is modular and the target composition can easily be adapted following the evolution of the most common *Salmonella* serotypes or the required law adaptations. In the future, additional MOL-PCR assays could be developed to detect other *Salmonella* serotypes commonly encountered in the Belgian food sector.

Acknowledgements

The research that yielded these results, was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 14/5 SALMSTID 1. Whole genome sequencing were performed at the service Transversal Activities in Applied Genomics at Sciensano. The authors want to thank the technicians of the National Reference Center for *Salmonella* and *Shigella*, for their expertise in serotyping by slide agglutination and their participation to the comparison study; and the technicians of the service Transversal Activities in Applied Genomics for their technical support of the MagPix.

Supplementary data

Supplementary Table S1 can be found online at: https://doi.org/10.1016/j.fm.2019.103394

CHAPTER 5
A genoserotyping system for a fast and objective identification of *Salmonella* serotypes commonly isolated from poultry and pork food sectors in Belgium

Context of this chapter:

The molecular marker selection method using EnteroBase, successfully applied for the 6 serovars subjected to an official control in Belgium as described in Chapter 4, was used again, in addition to the software Gegenees previously described in Chapter 3, to develop 3 additional MOL-PCR assays. These 3 new multiplex assays target *Salmonella* serotypes which were chosen for their invasive character or their prevalence in the Belgian poultry and pork sectors. Moreover, the barcode system elaborated in the Chapter 4 was implemented as a Decision Support System (DSS), hosted by a web application, for a user-friendly and automatic interpretation of the results (described into more detail in Annex 1). Together with the first module developed in Chapter 4, and further improved as described here, the 3 new MOL-PCR assays and the DSS presented in this Chapter constitute a fast, accurate and cost-effective genoserotyping system which was validated by comparison with the classical methods.

This chapter was adapted from the previously published manuscript:

Gand, M., W. Mattheus, N.H.C. Roosens, K. Dierick, K. Marchal, S. Bertrand* and S.C.J. De Keersmaecker*. 2020. "A genoserotyping system for a fast and objective identification of Salmonella serotypes commonly isolated from poultry and pork food sectors in Belgium." *Food Microbiology* 91 (October) 103534. https://doi.org/10.1016/j.fm.2020.103534. (* equal contribution)

Authors' contributions:

M. Gand designed the study, performed the experiments, implemented the Decision Support System (DSS) developed by ICT, analyzed the data and drafted the manuscript. S.C.J. De Keersmaecker and S. Bertrand conceived and supervised the study, helped to design the study, to interpret the results and to draft the manuscript. N.H.C Roosens provided specialist feed-back on the design of the DSS. K Marchal and W. Mattheus provided specialist feed-back. All authors read and approved the final manuscript.

Abstract:

Humans are mostly contaminated by *Salmonella* through the consumption of pork- and poultry-derived food products. Therefore, a strict monitoring of *Salmonella* serotypes in food-producing animals is needed to limit the transmission of the pathogen to humans. Additionally, *Salmonella* can lead to economic loss in the food sector. Previously, a genoserotyping method using the MOL-PCR and Luminex technology was developed for the identification of the 6 *Salmonella* serotypes, and their variants, subjected to an official control in the Belgian food sector. In this study, 3 additional assays using the same technology were developed for the rapid and cost-effective detection of 13 highly invasive serotypes or other serotypes frequently isolated from the Belgian poultry and pork sectors, i.e. Agona, Anatum, Brandenburg, Choleraesuis, Derby, Enteritidis vaccine strains, Gallinarum var. Gallinarum/Pullorum, Livingstone, Mbandaka, Minnesota, Ohio, Rissen and Senftenberg. Moreover, the previously developed first MOL-PCR assay was improved for *S.* Paratyphi B and serogroup O:3 detection. Finally, a Decision Support System hosted by a web application was created for an automatic and objective interpretation of the Luminex raw data. The 3 new assays and the modifications of the first assay were validated with a 100% accuracy, using 553 *Salmonella* and non-*Salmonella* strains in total.

5.1. Introduction

In 2018, salmonellosis was once again considered as the second foodborne disease in Europe (EFSA 2019b). Its causing agent, *Salmonella,* can infect a large variety of food-producing animals like poultry, pigs and cattle, which are their major reservoirs. While the transmission between animals is oro-fecal, humans are infected by contact with animals and mostly by the consumption of contaminated food (estimated to account for 85% in transmissions to humans) (Heredia and García 2018; Oxford Analytica 2012). The *Salmonella* genus is divided into more than 2 500 serotypes (following the White-Kauffmann-Le Minor (WKL) scheme) among which the 1 500 of the *Salmonella enterica* subsp. *enterica* are generally responsible for food poisoning (Antunes et al. 2016; Grimont and Weill 2007; Ryan, O'Dwyer, and Adley 2017). The *Salmonella* serotypes have different host-specificities. Depending on the *Salmonella* serotype and the host infected by it, the salmonellosis can result either in non-invasive symptoms like gastroenteritis (or even silent symptoms for healthy carriers) or in more dangerous invasive symptoms like fever and bacteremia, leading potentially to the decease of the host without treatment (Heredia and García 2018). Indeed, while *Salmonella* serovar Enteritidis is more associated with poultry, *Salmonella* serovar Typhimurium can infect a broader range of animal species including poultry, pork and cattle. In poultry and pork, these 2 non-invasive serotypes will lead to low or undetectable symptoms, allowing the bacteria to infect humans through the consumption of contaminated eggs and meat (Demirbilek 2016). Contrarily, *Salmonella* serovar Gallinarum biovar. Gallinarum (*S.* Gallinarum var. Gallinarum) and *Salmonella* serovar Gallinarum biovar. Pullorum (*S.* Gallinarum var. Pullorum) are restricted to some avian species (depending on their age) and cause invasive symptoms like severe septicemia, resulting in a high mortality rate in a.o. the poultry sector (Alves Batista et al. 2018). Similarly, *Salmonella* serovar Choleraesuis is historically known to cause large outbreaks especially in pigs, leading to septicemia, enterocolitis and pneumonia. Despite the fact that these invasive *Salmonella* serotypes are rare in Europe, their surveillance must be maintained to detect possible reemergence, like the outbreak caused by *S.* Choleraesuis in 4 Danish pig farms in 2012-2013 (Pedersen et al. 2015). *S.* Enteritidis and *S.* Typhimurium are the most prevalent serotypes in developed countries, although infections by *S.* Enteritidis decreased these last years, thanks to the vaccination obligation of adult breeding of *Gallus gallus* and laying hens with live attenuated vaccines (Griffin and O'Brien, 2013; NRC personnal communication) such as Salmovac SE (IDT Biologika, Dessau, Germany) or AviPro SALMONELLA VAC E (Elanco GmbH, Cuxhaven, Germany).

While the presence of non-invasive *Salmonella* serotypes must be monitored in food-producing animals to limit transmissions to human, the rapid detection of invasive serotypes is more needed to avoid animal suffering and economical loss for the breeders, linked to the decrease in productivity (weight loss, abortions, milk production, treatment of contaminated eggs, etc…), the quarantining of diseased animals and the destruction or treatment of contaminated food products (Heredia and García

2018; Majowicz et al. 2010; Oxford Analytica 2012). For example, in poultry farms, depending on the type of breeding, the farmers must check the presence of some *Salmonella* serotypes in the flock before further food applications (see section 1.3.4). The Turn-Around Time (TAT), which means the maximum period allowed to communicate *Salmonella* serotyping results by the first-line laboratories, after collection of the samples at the farm, is established by the Belgian Federal Agency for the Security of the Food Chain (FASFC) at 14 days for poultry breeding. This time includes the sampling, the transport to the first-line laboratories, the isolation of the *Salmonella* and the serotype identification for which a referring to the National Reference Center (NRC) is generally needed. During this period, the selling of the eggs is postponed until a positive agreement is obtained from the first-line laboratories. If a *Salmonella* is isolated from the field, corrective actions must be undertaken depending on the serotype identified and the type of breeding (laying hens or adult breeding), going from logistic slaughtering to destruction of the eggs or restriction of their selling to applications that involve a thermic treatment. A complete disinfection of the farm must be performed in each case and absence of *Salmonella* must be proven on site before the arrival of a new animal batch.

Consequently, it is crucial to rapidly identify the serotype of the *Salmonella* isolated from poultry and pork sectors, with the aim to reduce human food poisoning, but also to quickly react in case of outbreaks due to invasive serotypes such as Choleraesuis and Gallinarum. The gold standard method for *Salmonella* serotyping consists of the characterization of 3 antigenic sites (somatic antigen O and flagellar antigens H1 and H2), located at the surface of the bacterium, by slide agglutination with specific antisera. More than 120 antisera are needed to be able to identify all of the 2 500 serotypes included in the WKL scheme (Grimont and Weill 2007). A positive agglutination is not always clearly obtained, making the result interpretation subjective. Additionally, for the discrimination of some specific variants (e.g. *Salmonella* serovar Paratyphi B variant Java, *S.* Gallinarum var. Gallinarum and *S.* Gallinarum var. Pullorum), biochemical tests based on culturing methods are needed. Also, when isolating *S.* Enteritidis, it can be important to make the discrimination between the vaccine and the wild-type field strain, if the vaccination campaign was too close to the sampling period. The vaccine strains included in the vaccines AviPro SALMONELLA VAC E and Salmovac SE are respectively resistant to antibiotics (streptomycin and rifampicin) and auxotrophic double-mutant (ade- and his-). Therefore, their differentiation is made by testing their growth characteristics on specific media containing antibiotics or lacking adenine and histidine. But as these biochemical and growth tests are based on culturing on specific media, they are complex, time-consuming and not always reliable (Batista et al. 2013; Gand et al. 2019; Maurischat et al. 2015). This is why all these reference methods (i.e. slide-agglutination, biochemical and growth tests) are only fully mastered in the NRCs, which dispose of the totality of the antisera collection, and where the tests are performed by experienced and carefully trained technicians. For this reason, most of the *Salmonella* isolates must be sent to the NRC for a complete identification and this causes additional delays not always compatible with the short TAT asked by the regulation. Therefore, this situation is

not suitable for the animal sectors. This is why there is a need for a faster, cheaper and more accurate identification technique, which could be used by the NRCs to reduce their TAT and analysis costs, or directly by the first-line laboratories, avoiding the need to send the sample to another laboratory.

Luckily, alternative appropriate methods exist for the identification of *Salmonella* serotypes and their variants, among the new molecular tools developed these last years. Some target-based molecular methods, such as Multiplex Oligonucleotide Ligation – PCR (MOL-PCR) linked to the Luminex technology, have proven to be suitable and cost-effective for rapid diagnostics (Jean-Gilles Beaubrun et al. 2014; Liang et al. 2016; Yoshida, Simone, et al. 2016). The MOL-PCR allows the detection of genomic molecular markers, linked to the serotype, by specific probes through a ligation-amplification reaction, at a high multiplexing level. The so created MOL-PCR fragments are then hybridized to unique color-coded MagPlex beads, subsequently detected by a MagPix apparatus, based on a fluorescence reaction (Luminex xTAG technology). An assay using this technology was previously developed by Gand et al. (2020) for a fast, objective and cost-effective genoserotyping of 6 *Salmonella* serotypes (and their variants) mentioned in the Belgian regulation (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8) and for the determination of common serogroups (O:3,10, O:4, O;7, O:8, O:9 and O:21).

In the present study, we have developed 3 new MOL-PCR assays with the aim to (i) rapidly detect highly invasive *Salmonella* serotypes like *S.* Gallinarum var. Gallinarum, *S.* Gallinarum var. Pullorum and *S.* Choleraesuis, (ii) make the discrimination between wild-type and vaccine (AviPro SALMONELLA VAC E and Salmovac SE) isolates of *S.* Enteritidis and (iii) identify common serotypes isolated in the food chain including: *Salmonella* serovar Agona, *Salmonella* serovar Anatum, *Salmonella* serovar Brandenburg, *Salmonella* serovar Derby, *Salmonella* serovar Livingstone, *Salmonella* serovar Mbandaka, *Salmonella* serovar Minnesota, *Salmonella* serovar Ohio, *Salmonella* serovar Rissen and *Salmonella* serovar Senftenberg. Additionally, the first MOL-PCR assay described by Gand et al. (2020) was improved for a more specific detection of *S.* Paratyphi B var. Java and serogroup O:3. The molecular markers, specific to the serotypes targeted by the developed assays, were selected from the MultiLocus Sequence Typing (MLST) database (EnteroBase), from the scientific literature or based on in-house genomic comparison studies using Whole Genome Sequencing (WGS). In addition, a Decision Support System (DSS) hosted by a web-application was created. The aim of this DSS is to perform an automatic interpretation of the Luminex raw data and to give recommendations to the users in case of partial identification. Through this web-application, all the final identification results are also centralized in a database for national surveillance of the *Salmonella* serotypes circulating in Belgium. The *Salmonella* genoserotyping system, including the multiplex assays and the DSS, was compared to the classical methods (slide-agglutination, growth and biochemical tests) with the analysis of 553 *Salmonella* and non-*Salmonella* strains.

5.2. Materials and methods

5.2.1. Bacterial strains and DNA preparation

Seventeen isolates of *S.* Choleraesuis and 2 isolates of *S.* Gallinarum were respectively provided by the Belgian Institute of Tropical Medicine (Antwerp, Belgium) and the company Biovac (Beaucouzé, France). The serotype of these isolates was confirmed by the NRC prior to use. All the other *Salmonella* strains used in this study have previously been isolated from food, animal or human samples (between 2005 and 2018 in Belgium) and were sent to the NRC for further characterization (including serotype identification), after *Salmonella* genus confirmation by the first-line laboratories. These isolates, in addition to the non-*Salmonella* strains used in this study, were stored in the collection of the NRC and are available upon request (**Supplementary Table S4**). All isolates were cultured on Nutrient agar (Neogen® Culture Media, Lansing, USA).

For MOL-PCR, the bacterial DNA of samples and controls were extracted by heat lysis as described by Gand et al. (2020). For WGS, genomic DNA was extracted with the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich, Saint-Louis, USA) according to the manufacturer's instructions.

5.2.2. Selection of molecular markers from WGS data

Using an Illumina MiSeq instrument (2 x 300 bp, Nextera XT libraries), genomic DNA of 11 *S.* Livingstone, 1 *S.* Gallinarum var. Gallinarum and 4 *S.* Gallinarum var. Pullorum isolates was sequenced. The FASTQ reads were deposited at the SALMSTID BioProject on NCBI (PRJNA509747).

In CLC Genomics Workbench 8.0 (Qiagen, Hilden, Germany), the raw FASTQ reads were first trimmed to a quality score limit of 0.05 with maximum 2 ambiguous nucleotides and reads with a length below 30 nucleotides were discarded. These trimmed reads were then *de novo* assembled with automatic bubble and word size, in mapping mode "map reads back to contigs" with scaffolding and a minimum contig length of 1 000 nucleotides. Identically, the downloaded Sequence Reads Archive (SRA) (**Supplementary Table S1**) were trimmed and assembled as described for the in-house sequenced data. All assemblies were uploaded to the *Salmonella In Silico* Typing Resource (SISTR), developed by Yoshida, Kruczkiewicz, et al. (2016), for serotype confirmation and MLST typing.

The downloaded and in-house produced WGS data were all exported to Gegenees (version 2.2.1; downloaded from http://www.gegenees.org; Ågren et al. 2012) on a Linux platform and analyzed for the selection of molecular markers specific of targeted serotypes, as described by Gand et al. (2020). All the genomes were labelled in the software as target, reference or background as indicated in the **Supplementary Table S1**.

5.2.3. Selection of molecular markers from EnteroBase and scientific literature

Single Nucleotide Polymorphism (SNP) markers were selected among the allele sequences of the 7 housekeeping genes (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*) of the MLST database EnteroBase (Achtman et al. 2012; Alikhan et al. 2018 ; https://enterobase.warwick.ac.uk), as previously described by Gand et al. (2020) (**Table 1**). The genetic alignments were made using MUSCLE (Edgar 2004) in the MEGA7 software (Kumar, Stecher, and Tamura 2016). The specificity of the alleles conserved in close Sequence Types (STs) or eBurst Groups (eBGs) was checked *in silico* on an in-house curated version of Enterobase downloaded in early 2019 and composed of 186 900 entries at that time. Abs/Pres markers (based on the presence or absence of a genetic sequence) and other SNP markers, presented in **Table 2**, were inspired from genetic studies or molecular methods published in the scientific literature.

5.2.4. Ligation probe design and MOL-PCR protocol

The ligation probes presented in **Tables 1 and 2** were designed using the Visual OMP (version 7.6.58.0; DNA Software) according to the guidelines of Wuyts, Mattheus, et al. (2015). When the molecular markers were selected from the literature, the ligation probes were designed based on existing probes, primers or a specific amplified sequence, if available. These probes were ordered from Integrated DNA Technologies (IDT, Leuven, Belgium) with a standard desalted purification.

The MOL-PCR reactions, the hybridization to MagPlex-TAG microspheres (Luminex, Austin, USA), the staining reaction using streptavidin-R-phycoerythrin (SAPE) (ThermoFisher Scientific, Waltham, USA) and the read-out using a Mag-Pix device (Luminex, Austin, USA) were performed following the protocol detailed by Gand et al. (2020). For the 4 MOL-PCR assays, the negative control (CTRL_-) was composed of *Vibrio alginolyticus* DNA (strain M/5035) extracted and used identically as for the other samples. For positive controls composed of DNA belonging to several *Salmonella* serotypes, one colony per serotype was mixed in a tube at the DNA extraction step (section 5.2.1) and processed like the other samples. The positive controls (CTRL_+) used for each assay, and their composition, are listed in **Supplementary Table S2**.

Table 1: Ligation probes designed from MLST markers selected in EnteroBase

Target	MOL-PCR assay	Probe	DNA sequence	Beads region[1]	Interpretation[2]	MLST gene	Allele number	SNP position[3]
S. Agona	O3-4-21	STID20-U_SNP	*TAATACGACTCACTATAGGg*ttgtaaattgtagtaaagaagtaCGGGCAAACGCGCGCT**A**	15	SNR	*hemD*	7	249
		STID20-D	P-ATTTTGCGTGGCAATGGCG*TCCCTTTAGTGAGGGTTAAT*					
		STID21-U_SNP	*TAATACGACTCACTATAGGG*aataagagaattgatatgaagatgGCAGAAATACGGCCTGCG**T**	35	SNR	*aroC*	3	53
		STID21-D	P-GATTACCGTGGCGGTGGACG*TCCCTTTAGTGAGGGTTAAT*					
S. Anatum	O3-4-21	STID22-U_SNP	*TAATACGACTCACTATAGGG*tgaaatgtgtatttgtatgtttagAAGTGGTTTCCGCTCATCG**T**	62	SNR	*purE*	25	87
		STID22-D	P-ACCCCCGATAAGCTGTTCAGCTT*TCCCTTTAGTGAGGGTTAAT*					
		STID23-U_SNP	*TAATACGACTCACTATAGGG*agtagaaagttgaaattgattatgCTGAAAGAGATGGTTCG**A**	12	SNR	*sucA*	20	132
		STID23-D	P-CATGCGGGTAACAGCGGCAC*TCCCTTTAGTGAGGGTTAAT*					
S. Brandenburg	O7	STID36-U_SNP*	*TAATACGACTCACTATAGGg*taattgaattgaaagataagtgtCTTCCGCGCCATCGTAATGTTT**T**	18	SNR	*hemD*	13	345
		STID36-D*	P-GCACATCGTTGATAACATTCACAAAAAC*TCCCTTTAGTGAGGGTTAAT*					
		STID37-U_SNP	*TAATACGACTCACTATAGGg*attgatatttgaatgtttgtttgCGGGCGGCGCGGCACA**T**	22	SNR	*purE*	10	174
		STID37-D	P-CTGCCGGGAATGATTGCGGC*TCCCTTTAGTGAGGGTTAAT*					
S. Choleraesuis	O7	STID45-U_SNP	*TAATACGACTCACTATAGGg*ttagttatgatgaatattgtgtaCCTGCGCGATTACCGTGG**T**	45	SNR	*aroC*	36	66
		STID45-D	P-GGTGGACGTTCTTCCGCGC*TCCCTTTAGTGAGGGTTAAT*					
S. Derby ST682	O3-4-21	STID48-U_SNP	*TAATACGACTCACTATAGGg*taagagtattgaaattagtaagaCTGTCTACGCTGCCTGCCG**T**	66	SNR	*dnaN*	60	191
		STID48-D	P-CGATTTCCCGAATCTTGACGAC*TCCCTTTAGTGAGGGTTAAT*					
S. Gallinarum	O9	STID42-U	*TAATACGACTCACTATAGGg*tgattgaatagtagattgtttaaCCGCCGCTGAAGGGCT**A**	46	SNR	*sucA*	41	39
		STID42-D	GAACGTTATCTGGGTGCCAAA*TCCCTTTAGTGAGGGTTAAT*					
S. Mbandaka	O7	STID10-U_SNP	*TAATACGACTCACTATAGGg*tgattgaatagtagattgttaaCAACATGGCGATGTTTAGCGT**T**	46	SNR	*thrA*	68	360
		STID10-D	P-TCCGGCCCAGGAATGAAAGGGATGA*TCCCTTTAGTGAGGGTTAAT*					
S. Mbandaka *S.* Rissen	O7	STID11-U_SNP	*TAATACGACTCACTATAGGGg*tgttatagaagttaaatgttaagATCGGTAAAGCCGGTGCGGCTAA**T**	30	SNR	*purE*	64	333
		STID11-D	P-GCCGCCCTGCTCGCCGCGCA*TCCCTTTAGTGAGGGTTAAT*					

(continued on next page)

Table 1 (continued)

Target	MOL-PCR assay	Probe	DNA sequence	Beads region[1]	Interpretation[2]	MLST gene	Allele number	SNP position[3]
S. Minnesota	O3-4-21	STID24-U_SNP	*TAATACGACTCACTATAGGg*tatgttgtaatgtattaagaaagAGATTGCCGTTCAGTTGGAAGG**T**	25	AC	*dnaN*	11	132
		STID24-U_WT	*TAATACGACTCACTATAGGG*tattagagtttgagaataagtagtGAGATTGCCGTTCAGTTGGAAGGC	33				
		STID24-D	P-GATCGGATGCTGGTGCGTTCTGG*TCCCTTTAGTGAGGGTTAAT*					
		STID25-U_SNP	*TAATACGACTCACTATAGGg*tgattgaatagtagattgtttaaCTGCCGCGCGCGGACAC**T**	46	SNR	*hisD*	102	441
		STID25-D	P-GCCCGGCAGGCCCTGAGCGC*TCCCTTTAGTGAGGGTTAAT*					
S. Ohio	O7	STID46-U_SNP	*TAATACGACTCACTATAGGg*taagattagaagttaatgaagaaCGCTGCGCGGCAGCCAG**G**	52	SNR	*aroC*	82	475
		STID46-D	P-ATCGCGATGAAATCACGGCG*TCCCTTTAGTGAGGGTTAAT*					
		STID9-U_SNP*	*TAATACGACTCACTATAGGGg*taattgaattgaaagataagtgtTACAAAGGCGTTGCCGGG**A**	18	SNR	*hisD*	12	174
		STID9-D*	P-CCAAAAATTTTATCCACTTTCGGTACGGAC*TCCCTTTAGTGAGGGTTAAT*					
S. Rissen	O7	STID12-U_SNP	*TAATACGACTCACTATAGGG*tattagagtttgagaataagtagtTCGCCATGCGGGTAACAG**T**	33	SNR	*sucA*	151	147
		STID12-D	P-GGCACTCGCGAAGTGGTGC*TCCCTTTAGTGAGGGTTAAT*					
S. Senftenberg	O3-4-21	STID7-U_SNP	*TAATACGACTCACTATAGGG*attgtgaaagaaagagaagaaattCCGCCATGTCTCGCGCCGGGATCTC**A**	14	SNR	*thrA*	19	435
		STID7-D	P-GTGGTGCTCATTACCCAGTCCTCCT*CCCTTTAGTGAGGGTTAAT*					
		STID26-U_SNP	*TAATACGACTCACTATAGGG*aattagaagtaagtagagtttaagAAACGGAAGGTAGCGAACTGCG**T**	56	SNR	*dnaN*	6	354
		STID26-D	P-ACTGTCGCGACCGACGGCCA*TCCCTTTAGTGAGGGTTAAT*					
		STID28-U_SNP	*TAATACGACTCACTATAGGg*atagatttagaatgaattaagtgATCCTGCTGACGCCTGATGCT**C**	28	SNR	*hisD*	75	376
		STID28-D	P-ACATTGCCCGCAAGGTGGCG*TCCCTTTAGTGAGGGTTAAT*					
		STID29-U_SNP	*TAATACGACTCACTATAGGGg*tgttatagaagttaaatgttaagCGTCCAGCGACGATGATAACCT**A**	30	SNR	*thrA*	64	312
		STID29-D	P-CCAGTTAAAGGGATCTCTAACCTTAAC*TCCCTTTAGTGAGGGTTAAT*					

*: the probes anneal on the reverse complement strand of the MLST gene

[1]: corresponds to the specific color and TAG sequence of the bead

[2]: the presence of the molecular marker is determined by the calculation of the Signal to Noise Ratio (SNR) or Allele Call (AC)

[3]: corresponds to the location of the SNP in multiple alignments of all alleles related to the considered MLST gene

P- : Phosphate

Primer (T7 and T3), anti-TAG, target-specific sequences and SNP positions are indicated by italic, lower-case, underlined and bold sequences, respectively.

Table 2: Ligation probes designed from molecular markers inspired from the scientific literature and WGS study

Target	MOL-PCR assay	Probe	DNA sequence	Beads region[1]	Type[2]	Interpretation[3]	Marker	Based on	Source
AviPro SALMONELLA VAC E	O9	STID401-U	*TAATACGACTCACTATAGGGttt*gttagaatgagaagatttatgCGCCGCCAAAGACCA**T**TTCG**T**	75	SNP	SNR	*kdpA*	kdpA-V_probe	Maurischat et al. 2015
		STID401-D	P-CAATTTGCATCAGCCACATCGGC*TCCCTTTAGTGAGGGTTAAT*						
S. Brandenburg	O7	STID38-U_SNP	*TAATACGACTCACTATAGGg*taagattagaagttaatgaagaaGGGTGTTGGTGGTGCAACAA**A**	52	SNP	SNR	*fliC*	Genbank accession no. AY434709	Herrera-León et al. 2004
		STID38-D	P-CGGTAC**T**GCTGCAATTAAGGA*TCCCTTTAGTGAGGGTTAAT*						
S. Derby	O3-4-21	STID1-U	*TAATACGACTCACTATAGGg*ttagttatgatgaatattgtgtaTTGGAGATCTTTCTAATGCGGAT	45	Abs/ Pres	SNR	*troN*	Genes of SPI-23 conserved in *S.* Derby	Hayward, Jansen, and Woodward 2013
		STID1-D	P-TCTACTAATACTGTCATCATGTTTGGAC*TCCCTTTAGTGAGGGTTAAT*						
S. Enteritidis	O9*	STID2-U	*TAATACGACTCACTATAGGGt*attagagtttgagaataagtagtCGGCGCATTCCTCCGTTT	33	Abs/ Pres	SNR	*sdr*	Ligation probes STID2-U and STID2-D	Gand et al. 2020a
		STID2-D	P-TTTCGTCGTGGGCGTCAGTA*TCCCTTTAGTGAGGGTTAAT*						
S. Gallinarum	O9	STID41-U_SNP	*TAATACGACTCACTATAGGg*atagatttagaatgaattaagtgCATTTATGGCCCCGGAGGCG**A**	28	SNP	SNR	hypothetical protein	Genomic study using Gegenees	This study
		STID41-D	P-ATGACGGGGACGACACCTGG*TCCCTTTAGTGAGGGTTAAT*						
S. Gallinarum var. Gallinarum	O9	STID43-U_SNP	*TAATACGACTCACTATAGGg*tatgttgtaatgtattaagaaagCCGACCTCTTCGCTGAATATTATG**A**	25	SNP	SNR	MscS	Genomic study using Gegenees	This study
		STID43-D	P-TGTATTGCTTTACCAAAACAACAGTA*T*CCCTTTAGTGAGGGTTAAT						
S. Livingstone	O7	STID47-U	*TAATACGACTCACTATAGGg*atagatttagaatgaattaagtgCTGAGAACCTGATAAATCG	28	Abs/ Pres	SNR	hypothetical protein	Genomic study using Gegenees	This study
		STID47-D	P-CGGCTATCTGGAGCGCGT*CCCTTTAGTGAGGGTTAAT*						

(continued on next page)

Table 2 (continued)

Target	MOL-PCR assay	Probe	DNA sequence	Beads region[1]	Type[2]	Interpretation[3]	Marker	Based on	Source
S. Paratyphi B (including var. Java)	BASE	STID333-U STID333-D	*TAATACGACTCACTATAGGG*tgatatagtagtgaagaaataagtAG GACGTATCTCTTTAGCAGGC P-GGAGGGGAGAAAGATATTTATGCC*TCCCTTTAGTGAGG GTTAAT*	34	Abs/ Pres	SNR	*SPAB_01124*	Reverse primer pPB23	Zhai et al. 2014
Salmonella	All	invA-U invA-D	*TAATACGACTCACTATAGGG*gataagaaagtgaaatgtaaattgAT AAACTTCATCGCACCGTCA P-AAGGAACCGTAAAGCTGGCTT*TCCCTTTAGTGAGGGTT AAT*	51	Abs/ Pres	SNR	*invA*	Ligation probes invA-U and invA-D	Wuyts, Mattheus, et al. 2015
Salmovac SE	O9	STID40-U_SNP STID40-D	*TAATACGACTCACTATAGGG*ttgtaaattgtagtaaagaagtaCGGT CTGACCTCCATGCT**A** P-CCGTTGGGAATTATTGCCG*TCCCTTTAGTGAGGGTTAAT*	15	SNP	SNR	*nhaA*	nhaA_V_probe	Maurischat et al. 2015
Serogroup O:4	O3-4-21*	STID16-U STID16-D	*TAATACGACTCACTATAGGG*tttgatttaagagtgttgaatgtaTCAA GTTGGAACTGGTGCT P-GGGGTAAGTTTGAAAGATTTTCTGG*TCCCTTTAGTGAGG GTTAAT*	26	Abs/ Pres	SNR	*rfbJ*	Ligation probes STID16-U and STID16-D	Gand et al. 2020a
Serogroup O:9	O9*	STID171-U_SNP STID171-D	*TAATACGACTCACTATAGGG*aattgagaaagagataaatgatagC ATATACTAAACAAAAAGCAAATGAA**C** P-TCGCCGCCGCCATTATAGA*TCCCTTTAGTGAGGGTTAAT*	72	SNP	SNR	*prt*	Ligation probes STID171-U and STID171-D	Gand et al. 2020a
Serogroup O:6,7	O7*	STID18-U STID18-D	*TAATACGACTCACTATAGGG*taagagtattgaaattagtaagaCGT TGGCAGACTGGTACTGATTG P-GCTCCCCTATTACGATGATTTC*TCCCTTTAGTGAGGGTT AAT*	66	Abs/ Pres	SNR	*wbaA*	Ligation probes STID18-U and STID18-D	Gand et al. 2020a
Serogroup O:3	O3-4-21*	STID301-U STID301-D	*TAATACGACTCACTATAGGG*tgatatagtagtgaagaaataagtTC TCTACGCAGACAATTATGTCA P-TGGAGTTATTATCCGGATGGG*TCCCTTTAGTGAGGGTTA AT*	39	Abs/ Pres	SNR	*wzx*	Primer E_wzx_F	Franklin et al. 2011

(continued on next page)

Table 2 (continued)

Target	MOL-PCR assay	Probe	DNA sequence	Beads region[1]	Type[2]	Interpretation[3]	Marker	Based on	Source
Serogroup O:3,10	O3-4-21	STID31-U	*TAATACGACTCACTATAGGG*aaataagaatagagagagaaagttTTATAAATTTACGTTTAGAACATGTTTAC	34	Abs/ Pres	SNR	*wzx*	O:3,10 sequence not present in O:1,3,19	Zhang et al. 2015
		STID31-D	P-GGTGAGAGGGATAAAGCAGGTAAAA*TCCCTTTAGTGAGGGTTAAT*						
Serogroup O:1,3,19	O3-4-21	STID321-U	*TAATACGACTCACTATAGGG*tgatatagtagtgaagaaataagtTCTCTACGCAGACAATTATGTCA	13	Abs/ Pres	SNR	*wzx*	O:1,3,19 sequence not present in O:3,10	Zhang et al. 2015
		STID321-D	P-TGGAGTTATTATCCGGATGGG*TCCCTTTAGTGAGGGTTAAT*						
Serogroup O:21	O3-4-21*	STID35-U	*TAATACGACTCACTATAGGg*ttgagaattagaatttgataaagCCACTGTCATTGGTGGTTATGAG	73	Abs/ Pres	SNR	*wzx*	Ligation probes STID21-U and STID21-D	Gand et al. 2020a
		STID35-D	P-TATGAATGGCTGGTATACGACATC*TCCCTTTAGTGAGGGTTAAT*						

*: Additionally present in the module BASE (Gand et al. 2020a)
[1]: corresponds to the specific color and TAG sequence of the bead
[2]: the presence of the molecular marker is determined by the calculation of the Signal to Noise Ratio (SNR) or Allele Call (AC)
[3]: corresponds to the location of the SNP in multiple alignments of all alleles related to the considered MLST gene
P- : Phosphate
Primer (T7 and T3), anti-TAG, target-specific sequences and SNP positions are indicated by italic, lower-case, underlined and bold sequences, respectively.

5.2.5. Data interpretation using the DSS

The MagPix produces Median Fluorescence Intensity (MFI) for each probe and each sample. From these raw data, the Signal to Noise Ratios (SNRs) are calculated for all the molecular markers (Abs/Pres and SNP markers). When it is needed, an Allele Call (AC) is also performed for some SNP markers. Then, these processed data are converted into a serotype identification based on a barcode system: the Gödel Prime Product (GPP) (Van den Bulcke et al. 2008; Van Den Bulcke et al. 2010). The data interpretation, including the checking of quality controls (CTRL_- and CTRL_+), the processing of raw data, the conversion into serotype identification and the serogroup clustering (using the GPP), was previously described in detail by Gand et al. (2020). In this study, these operations were automatically performed by the DSS developed as a web-application and accessible (through a login request) at: https://salmstid.wiv-isp.be/. The web-application was developed and deployed according to the DTAP principle, i.e. following the 4 phases of Development, Testing, Acceptance and Production. The web-application was first configured for each MOL-PCR assay, using a protocol setting function (included in it) in which all the cut-off values, the quality controls limits, the serogrouping probes and the correspondences between GPPs and identification results can be set. Then, the Comma Separated Value (CSV) file generated by the MagPix for each assay, containing all the MFI results for each probe and sample, was uploaded in the DSS. After the automatic processing of the data, the results were displayed on the screen and were exported as a PDF report. The performance of the DSS was tested with many scenarios simulating different case studies which can happen during routine analyses such as normal identifications, serogroup clustering, mixed (not pure) sample, failed quality controls, fluorescence detection issues or incorrect setting of the MagPix apparatus.

Briefly, all the serotype identifications are linked to a GPP, itself linked to a specific combination of positive probes which include at least the detection of the marker *invA* (for the detection of *Salmonella* spp.) and a marker targeting one of the following serogroups: O:3, O:4, O:7, O:8, O:9, O:21. When obtaining a GPP configured in the DSS, the system displays the name of the detected serotype. If the marker *invA* is not retrieved, the sample is reported as "No *Salmonella*" by the DSS. If it is present in the sample but without a serogroup marker, the *Salmonella* isolate is characterized as "Unknown serogroup". On the opposite, if more than one serogroup marker is detected (which is not possible considering the serogroup targeted by the method) in a *Salmonella* sample (positive for *invA*), the DSS will report it as "not pure sample" because a mix of serotypes is suspected. In case of a sample with an unknown GPP and not belonging to one of the cases described above, the DSS will display the serogroup to which it belongs, based on the division of the GPP by the prime number of the probe, and recommend what further analysis must be performed to complete the identification. When the interpretation of the data cannot be properly done, because of failed quality controls, detection errors or wrong settings of the MagPix, the DSS displays an error message to the user, describing the anomaly (**Figure 1**).

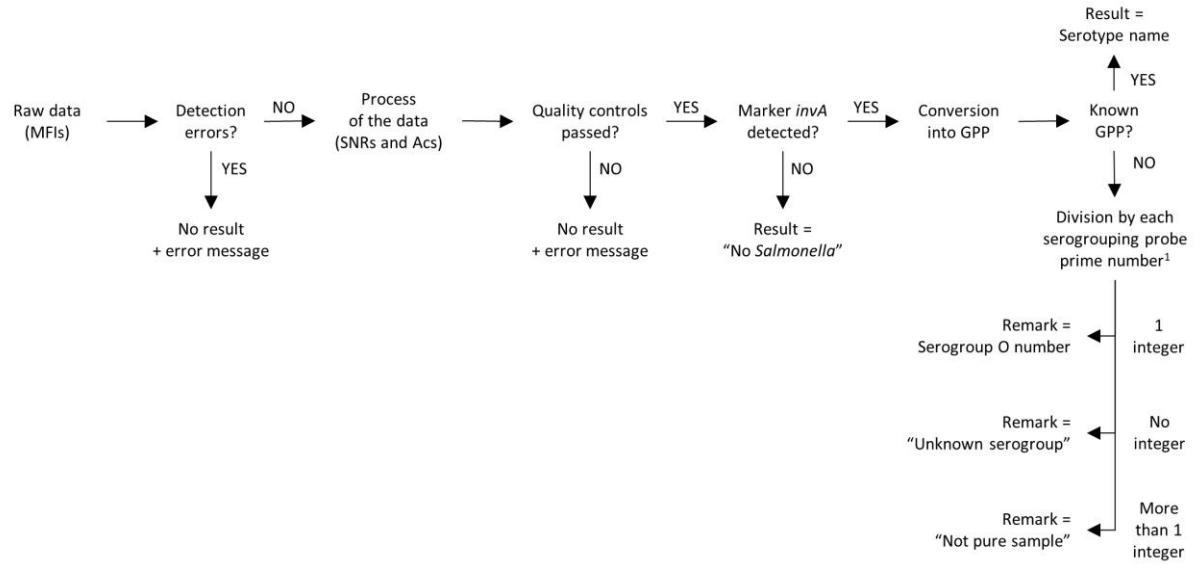

Figure 1: Decision tree for results interpretation performed by the Decision Support System (DSS)

AC: Allele Call; GPP: Gödel Prime Product; MFI: Median Fluorescence Intensity; SNR: Signal to Noise Ratio.

[1] :to know if a specific molecular marker was detected through the probe combination resulting into a GPP, this latter is divided by the prime number of the probe targeting this molecular marker. If an integer is obtained, the molecular marker was detected. If a decimal number is obtained the molecular marker was not present.

From the upload of the comma separated value (.csv) file on the web application, the DSS performs the automatic interpretation of the raw data, using the GPP, to display identification results. The processing of the data includes experimental error detection and quality control checks.

5.2.6. Evaluation of the MOL-PCR assays

A total of 464 *Salmonella* isolates and 33 non-*Salmonella* isolates were used to evaluate the sensitivity (using inclusivity tests), the specificity (using exclusivity tests) and the accuracy of the 3 new developed MOL-PCR assays following the approach used by Gand et al. (2020). From the 464 *Salmonella* isolates, 330 belonged to the serotypes targeted by the method and were used for the inclusivity tests (**Table 3**). The remaining 134 *Salmonella* isolates, belonging to 75 other serotypes not targeted by the method (**Table 4 and Table 5**), and 33 non-*Salmonella* isolates (**Table 5**), were used for the exclusivity tests.

A part of these isolates were also used for the validation of the modifications made to the MOL-PCR previously developed by Gand et al. (2020). This included all the isolates belonging to serogroup O:3 (56 to O:3,10 and 32 to O:1,3,19) and those belonging to serotype Paratyphi B (11) (**Table 3 and Table 4**). In addition, 56 other *S.* Paratyphi B isolates were also tested for the evaluation of these modifications (**Table 6**).

Table 3: Results of the inclusivity tests using *Salmonella* isolates belonging to the targeted serotypes

MOL-PCR assay	Number of tested isolates	Reference identification[1]	Serogroup	Expected identifications with the MOL-PCR assays			Obtained identifications with the MOL-PCR assays			Comparison expected vs. obtained
				GPP	Serotype	Serogroup	GPP	Serotype	Serogroup	
O3-4-21	27	*S.* Agona	O:4	2040753	*S.* Agona	O:4	2040753	*S.* Agona	O:4	27 TP
	28	*S.* Anatum	O:3,10	33915	*S.* Anatum	O:3	33915	*S.* Anatum	O:3	28 TP
	25	*S.* Brandenburg	O:4	35547369	*S.* Brandenburg	O:4	35547369	*S.* Brandenburg	O:4	25 TP
	30	*S.* Derby	O:4	31317 or 779493	*S.* Derby	O:4	31317	*S.* Derby	O:4	30 TP
	16	*S.* Derby	O:4	31317 or 779493	*S.* Derby	O:4	779493	*S.* Derby ST682	O:4	16 TP
	28	*S.* Minnesota	O:21	572241	*S.* Minnesota	O:21	572241	*S.* Minnesota	O:21	28 TP
	21	*S.* Senftenberg	O:1,3,19	110055 or 189255	*S.* Senftenberg	O:3	110055	*S.* Senftenberg eBG55 eBG69	O:3	21 TP
	7	*S.* Senftenberg	O:1,3,19	110055 or 189255	*S.* Senftenberg	O:3	189255	*S.* Senftenberg eBG30	O:3	7 TP
O7	17	*S.* Choleraesuis	O:7	435	*S.* Choleraesuis	O:7	435	*S.* Choleraesuis	O:7	17 TP
	30	*S.* Livingstone	O:7	1155	*S.* Livingstone	O:7	1155	*S.* Livingstone	O:7	30 TP
	29	*S.* Mbandaka	O:7	4845	*S.* Mbandaka	O:7	4845	*S.* Mbandaka	O:7	29 TP
	18	*S.* Ohio	O:7	2415 or 26565	*S.* Ohio	O:7	26565	*S.* Ohio	O:7	18 TP
	8	*S.* Ohio	O:7	2415 or 26565	*S.* Ohio	O:7	2415	*S.* Ohio	O:7	8 TP
	27	*S.* Rissen	O:7	3315	*S.* Rissen	O:7	3315	*S.* Rissen	O:7	27 TP
O9	1	AviPro SALMONELLA VAC E	O:9	5865	AviPro SALMONELLA VAC E	O:9	5865	AviPro SALMONELLA VAC E	O:9	1 TP
	5	Enteritidis	O:9	255	*S.* Enteritidis wild-type	O:9	255	*S.* Enteritidis wild-type	O:9	5 TP
	2	*S.* Gallinarum var. Gallinarum	O:9	1365	*S.* Gallinarum var. Gallinarum	O:9	1365	*S.* Gallinarum var. Gallinarum	O:9	2 TP
	10	*S.* Gallinarum var. Pullorum	O:9	1155	*S.* Gallinarum var. Pullorum	O:9	1155	*S.* Gallinarum var. Pullorum	O:9	10 TP
	1	Salmovac SE	O:9	4845	Salmovac SE	O:9	4845	Salmovac SE	O:9	1 TP
TOTAL:	330								Total TP:	330
									Total FN:	0
									Inclusivity (sensitivity):	100%

[1]: obtained by classical methods, i.e. slide-agglutination, cultural and biochemical tests

TP: True Positive; FN: False Negative

In case of discordances in the results obtained between the developed and the reference methods, 2 additional serotyping analyses by slide-agglutination were performed (one blind and one by another technician) in order to verify the reference identification.

5.3. Results

5.3.1. Selection of the molecular markers

For each *Salmonella* serotype, including its variant, specific SNP markers were screened in the housekeeping gene alleles of the MLST scheme (Achtman et al. 2012), which are conserved in a serotype population **(Table 1**). After having checked *in silico* the specificity of the candidates in EnteroBase, it appeared that a combination of at least 2 MLST markers (in association with the serogroup markers, **Table 2** and **Supplementary Table S3**) was needed to reach a false positive rate close to 0% and a false negative rate lower than 10%. These values were later experimentally verified during the validation of the method (see section 5.3.3).

Additional molecular markers were selected or inspired from scientific papers describing a target-based genoserotyping method or a genomic comparison study (**Table 2**). The specificity of these markers was checked *in silico* by BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the NCBI database.

As no appropriate markers could be selected in EnteroBase nor in the scientific literature for the specific detection of *S.* Livingstone and *S.* Gallinarum var. Gallinarum, genomic comparison studies were performed using publicly available and in-house generated WGS data (**Supplementary Table S1**). Fragmented alignments were performed using the software Gegenees for the determination of genomic signatures specific for each of the 3 target groups (*S.* Livingstone, *S.* Gallinarum and *S.* Gallinarum var. Gallinarum genomes) and absent in the background group (*Salmonella* genomes belonging to 43 other serotypes). For each of the 2 serotypes or variant, one molecular marker located in a coding sequence, suitable for a ligation probe design and offering a good specificity after a BLAST check in NCBI, was selected (**Table 2**).

5.3.2. Development of the MOL-PCR assays

The MOL-PCR is based on the principle that upstream and downstream probes must anneal close to each other on the target sequence to be subsequently linked by the ligase enzyme. For each of the previously selected molecular markers (in section 5.3.1), these probes were designed. Their sequence is listed in **Table 1** and **Table 2**. A wild-type version of the upstream probe was also designed when an interpretation using AC was needed to improve the fluorescence detection. The probes developed in this study were divided in 3 different MOL-PCR modules, i.e. MOL-PCR O3-4-21, MOL-PCR O7 and

MOL-PCR O9. In each module, serogrouping probes used to cluster each *Salmonella* isolate (positive for the marker *invA*) into one of the serogroups targeted by the modules, when possible, were included: STID16 (O:4), STID18 (O:7), STID35 (O:21), STID171 (O:9) and STID301 (O:3) (**Table 2**). With exception of the latter, all other serogrouping probes were already previously described in the context of the MOL-PCR assay developed by Gand et al. (2020), named here MOL-PCR BASE, allowing the detection of 6 *Salmonella* serotypes (and their variants) mentioned in the Belgian regulation, i.e. *S.* Enteritidis, *S.* Hadar, *S.* Infantis, *S.* Paratyphi B including the Java variant, *S.* Typhimurium including its monophasic variant and *S.* Virchow.

The module MOL-PCR O3-4-21 was developed to detect the *Salmonella* serotypes: Agona, Anatum, Brandenburg, Derby, Minnesota and Senftenberg. The probe STID48 was included in the assay for the identification of the particular sequence type ST682 of *S.* Derby (representing 10% of the Derby population in EnteroBase) not detected by STID1. Also in this assay, the detection of *S.* Senftenberg was obtained though the combination of STID7 and STID26, representing 86% of the Senftenberg population in Enterobase clustered in eBG55 and eBG69, or STID28 and STID29 representing 10% of the Senftenberg population in Enterobase clustered in eBG30. The probes STID31 and STID321 included in the MOL-PCR O3-4-21 were used for the discrimination between O:3,10 and O:1,3,19 respectively, but only when the *Salmonella* sample is already serogrouped as O:3 by STID301 (**Supplementary Table S3a**).

The module MOL-PCR O7 was created for the specific detection of the invasive *S.* Choleraesuis and other common *Salmonella* serotypes belonging to serogroup O:7: Livingstone, Mbandaka, Ohio and Rissen. The probe combination of invA (*Salmonella*), STID18 (O:7), STID9 and STID46 is used for the specific detection of *S.* Ohio. For the detection of *S.* Livingstone, the probes invA (*Salmonella*), STID18 (O:7) and STID47 must be positive. In the module O7, it can be noticed that the probe STID47 was also sometimes positive for *S.* Ohio but always together with STID9 and STID46 (**Supplementary Table S3b**).

The module MOL-PCR O9 was developed for a fast identification of the serotype Gallinarum, the discrimination between its 2 variants Gallinarum and Pullorum but also for the differentiation between the *S.* Enteritidis wild-type field and vaccine strains. In this module, STID40 and STID401 are respectively used for the detection of Salmovac SE and AviPro SALMONELLA VAC E when the *S.* Enteritidis identification has been confirmed by the probes invA, STID2 and STID171 also included in the module O9. The discrimination between the variants Gallinarum and Pullorum is performed in this assay by STID43 and STID42, respectively, when the sample is already positive for invA, STID41 and STID171 (**Supplementary Table S3c**).

For all modules, the expected combinations of molecular markers, the GPPs, the associated serotyping results and the corresponding remarks are listed in the **Supplementary Table S3**. These parameters were set in the DSS for the automatic interpretation of the data.

Table 4: Results of the exclusivity tests using untargeted *Salmonella* isolates for which the serogroup can be determined by the developed method

Number of tested isolates	Reference identification[1]	Serogroup	Results obtained with the MOL-PCR assays				Comparison with the reference identification	
			Using the module[2]	GPP	Identification	Serogroup	Serotype identification	Serogroup clustering
5	*S.* 1,4,[5],12:i:-	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Abony	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Agama	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Bareilly	O:7	O7	105	*Salmonella*	O:7	TN	OK
1	*S.* Berta	O:9	O9	15	*Salmonella*	O:9	TN	OK
1	*S.* Braenderup	O:7	O7	15	*Salmonella*	O:7	TN	OK
1	*S.* Brancaster	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Bredeney	O:4	O3-4-21	20163	*Salmonella*	O:4	TN	OK
1	*S.* Butantan	O:3,10	O3-4-21	6195	Probably *S.* Butantan	O:3	TN	OK
1	*S.* Chester	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Coeln	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Colindale	O:7	O7	15	*Salmonella*	O:7	TN	OK
1	*S.* Dublin	O:9	O9	15	*Salmonella*	O:9	TN	OK
1	*S.* Durban	O:9	O9	15	*Salmonella*	O:9	TN	OK
1	*S.* Eastbourne	O:9	O9	15	*Salmonella*	O:9	TN	OK
4	*S.* Give	O:3,10	O3-4-21	291165	Probably *S.* Give	O:3	TN	OK
3	*S.* Haifa	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Heidelberg	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
2	*S.* Indiana	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK

(continued on next page)

Table 4 (continued)

Number of tested isolates	Reference identification[1]	Serogroup	Results obtained with the MOL-PCR assays				Comparison with the reference identification	
			Using the module[2]	GPP	Identification	Serogroup	Serotype identification	Serogroup clustering
5	*S.* Infantis	O:7	O7	345	*Salmonella*	O:7	TN	OK
1	*S.* Ituri	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Javiana	O:9	O9	15	*Salmonella*	O:9	TN	OK
1	*S.* Jerusalem	O:7	O7	15	*Salmonella*	O:7	TN	OK
1	*S.* Kapemba	O:9	O9	15	*Salmonella*	O:9	TN	OK
1	*S.* Lagos	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
2	*S.* Liverpool	O:1,3,19	O3-4-21	165	*Salmonella*	O:3	TN	OK
2	*S.* Llandoff	O:1,3,19	O3-4-21	165	*Salmonella*	O:3	TN	OK
6	*S.* London	O:3,10	O3-4-21	389865	Probably *S.* London	O:3	TN	OK
4	*S.* Meleagridis	O:3,10	O3-4-21	4935	*Salmonella*	O:3	TN	OK
1	*S.* Mikawasima	O:7	O7	15	*Salmonella*	O:7	TN	OK
1	*S.* Montevideo	O:7	O7	15	*Salmonella*	O:7	TN	OK
4	*S.* Muenster	O:3,10	O3-4-21	105	*Salmonella*	O:4	TN	OK
1	*S.* Napoli	O:9	O9	15	*Salmonella*	O:9	TN	OK
1	*S.* Nyborg	O:3,10	O3-4-21	7455	*Salmonella*	O:3	TN	OK
3	*S.* Nyborg	O:3,10	O3-4-21	105	*Salmonella*	O:3	TN	OK
1	*S.*Oranienburg	O:7	O7	195	*Salmonella*	O:7	TN	OK
1	*S.* Panama	O:9	O9	15	*Salmonella*	O:9	TN	OK
11	*S.* Paratyphi B dT+	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Saintpaul	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK

(continued on next page)

Table 4 (continued)

Number of tested isolates	Reference identification[1]	Serogroup	Results obtained with the MOL-PCR assays				Comparison with the reference identification	
			Using the module[2]	GPP	Identification	Serogroup	Serotype identification	Serogroup clustering
1	*S.* Sandiego	O:4	O3-4-21	18447	*Salmonella*	O:4	TN	OK
1	*S.* Schwarzengrund	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
1	*S.* Singapore	O:7	O7	105	*Salmonella*	O:7	TN	OK
1	*S.* Stanley	O:4	O3-4-21	30459	*Salmonella*	O:4	TN	OK
3	*S.* Stanleyville	O:4	O3-4-21	30459	*Salmonella*	O:4	TN	OK
1	*S.* Tennessee	O:7	O7	15	*Salmonella*	O:7	TN	OK
1	*S.* Thompson	O:7	O7	15	*Salmonella*	O:7	TN	OK
5	*S.* Typhimurium	O:4	O3-4-21	429	*Salmonella*	O:4	TN	OK
2	*S.* Uganda	O:3,10	O3-4-21	7455	*Salmonella*	O:3	TN	OK
5	*S.* Virchow	O:7	O7	15	*Salmonella*	O:7	TN	OK
3	*S.* Weltevreden	O:3,10	O3-4-21	105	*Salmonella*	O:3	TN	OK
TOTAL: 102						Total TN:	102	
						Total FP:	0	
						Exclusivity (specificity):	100%	

[1]: obtained by classical methods, i.e. slide-agglutination and biochemical tests.
[2]: the results presented here are those obtained using one of the 3 modules, depending on the serogroup of the analyzed isolate.
dT+: d-Tartrate fermenting isolates; TN: True Negative, including probable results which are not complete identifications and need classical methods, i.e. slide-agglutination and biochemical test, to be confirmed; FP: False Positive

5.3.3. Validation of the 3 new MOL-PCR assays by comparison with the classical method

For the validation of the 3 new MOL-PCR assays (MOL-PCR O3-4-21, MOL-PCR O7 and MOL-PCR O9) and the DSS, at least 25 targeted *Salmonella* isolates (when available), 75 untargeted *Salmonella* isolates and 25 non-*Salmonella* isolates were analyzed per module for the inclusivity and exclusivity tests. The identification results produced with the new method were compared with those obtained using the reference techniques (**Table 3, Table 4 and Table 5**).

Using the MOL-PCR O3-4-21, 27 *S.* Agona, 28 *S.* Anatum, 25 *S.* Brandenburg, 46 *S.* Derby, 28 *S.* Minnesota and 28 *S.* Senftenberg isolates were correctly completely identified by the DSS. A part of the *S.* Derby isolates (16, representing 34%) were reported as belonging to the sequence type ST682 which is more than the proportion of this cluster estimated previously in EnteroBase (10%). Identically for the *S.* Senftenberg detection, more isolates (25%) belonging to the eBG30 were identified compared to the percentage of this population in EnteroBase (10%). With the MOL-PCR O7, the complete identification of 17 *S.* Choleraesuis, 30 *S.* Livingstone, 29 *S.* Mbandaka, 26 *S.* Ohio and 27 *S.* Rissen isolates was correctly obtained by the DSS. The ability of the MOL-PCR O9 to detect the vaccine strains of *S.* Enteritidis was validated with the correct discrimination between 5 wild-type *S.* Enteritidis coming from the field and 2 isolates coming each from one of the commercial vaccines AviPro SALMONELLA VAC E and Salmovac SE. The serotype of 12 *S.* Gallinarum isolates was confirmed by the MOL-PCR O9 and they were correctly discriminated into 2 *S.* Gallinarum var. Gallinarum and 10 *S.* Gallinarum var. Pullorum isolates. Unfortunately, not enough *S.* Choleraesuis and *S.* Gallinarum isolates were available in the NRC collection to achieve the validation criteria of at least 25 *Salmonella* isolates per targeted serotypes. As no false negative was obtained among the targeted *Salmonella* used here for the inclusivity tests, the sensitivity was determined to be 100% for each of the tested modules (**Table 3 and Supplementary Table S4**).

For the exclusivity tests, 134 untargeted *Salmonella* isolates were analyzed with the 3 modules, among which 102 belonged to one of the serogroups targeted by the method and were correctly clustered by the serogrouping probes (**Table 4**), and 32 belonged to 13 other serogroups and were correctly reported as "Unknown serogroup" by the DSS (**Table 5**). Additionally, for each module, 1 *Salmonella* isolate per serotype targeted by the 2 other assays was analyzed and correctly determined as "Unknown serogroup" (data not shown). For these cases described above, the DSS recommended to use the classical method to obtain the complete identification. Finally, 33 non-*Salmonella* isolates were analyzed with the 3 MOL-PCR assays and successfully reported as "No *Salmonella*" by the DSS (**Table 5**). From these exclusivity tests, no false positive was obtained and the specificity was determined to be 100%.

Table 5: Results of the exclusivity tests using non-*Salmonella* isolates and *Salmonella* isolates for which the serogroup cannot be determined by the developed method

Number of isolates tested	Reference Identification[1]	Luminex Identification result[2]	Comparison with the reference identification
1	*Acinetobacter baumanii*	No *Salmonella*	TN
1	*Bacillus cereus*	No *Salmonella*	TN
1	*Citrobacter koseri*	No *Salmonella*	TN
1	*Enterobacter aerogenes*	No *Salmonella*	TN
1	*Enterococcus faecium*	No *Salmonella*	TN
2	*Escherichia coli*	No *Salmonella*	TN
1	*Klebsiella oxytoca*	No *Salmonella*	TN
1	*Klebsiella pneumoniae*	No *Salmonella*	TN
5	*Listeria monocytogenes*	No *Salmonella*	TN
1	*Morganella morganii*	No *Salmonella*	TN
4	*Neisseria meningitidis*	No *Salmonella*	TN
1	*S.* Ago (O:30)	*Salmonella* - Unknown serogroup	TN
1	*S.* Agoueve (O:13,22)	*Salmonella* - Unknown serogroup	TN
1	*S.* Brive (O:42)	*Salmonella* - Unknown serogroup	TN
1	*S.* Carmel (O:17)	*Salmonella* - Unknown serogroup	TN
1	*S.* Cero (O:18)	*Salmonella* - Unknown serogroup	TN
1	*S.* Cotham (O:28)	*Salmonella* - Unknown serogroup	TN
1	*S.* Dugbe (O:45)	*Salmonella* - Unknown serogroup	TN
1	*S.* Durham (O:13,23)	*Salmonella* - Unknown serogroup	TN
1	*S.* Ebrie (O:35)	*Salmonella* - Unknown serogroup	TN
3	*S.* Gaminara (O:16)	*Salmonella* - Unknown serogroup	TN
1	*S.* Havana (O:13)	*Salmonella* - Unknown serogroup	TN
1	*S.* Hvittingfoss (O:16)	*Salmonella* - Unknown serogroup	TN
1	*S.* Idikan (O:13)	*Salmonella* - Unknown serogroup	TN
2	*S.* Kasenyi (O:38)	*Salmonella* - Unknown serogroup	TN
2	*S.* Kedougou (O:13)	*Salmonella* - Unknown serogroup	TN
1	*S.* Kisarawe (O:11)	*Salmonella* - Unknown serogroup	TN
1	*S.* Mgulani (O:38)	*Salmonella* - Unknown serogroup	TN
1	*S.* Monschaui (O:35)	*Salmonella* - Unknown serogroup	TN
3	*S.* Paratyphi A (O:2)	*Salmonella* - Unknown serogroup	TN
1	*S.* Pomona (O:28)	*Salmonella* - Unknown serogroup	TN
1	*S.* Poona (O:13,22)	*Salmonella* - Unknown serogroup	TN
1	*S.* Rubislaw (O:11)	*Salmonella* - Unknown serogroup	TN
1	*S.* Telelkebir (O:13,23)	*Salmonella* - Unknown serogroup	TN
1	*S.* Umbilo (O:28)	*Salmonella* - Unknown serogroup	TN
1	*S.* Urbana (O:30)	*Salmonella* - Unknown serogroup	TN
1	*S.* Wandsworth (O:39)	*Salmonella* - Unknown serogroup	TN
1	*Serratia odorifera*	No *Salmonella*	TN
1	*Shigella boydii*	No *Salmonella*	TN
1	*Shigella flexneri*	No *Salmonella*	TN
1	*Shigella sonei*	No *Salmonella*	TN
1	*Staphylococcus aureus*	No *Salmonella*	TN
1	*Staphylococcus epidermidis*	No *Salmonella*	TN
1	*Staphylococcus mileri*	No *Salmonella*	TN
1	*Staphylococcus saprophyticus*	No *Salmonella*	TN
1	*Streptococcus agalactiae*	No *Salmonella*	TN
1	*Streptococcus bovis*	No *Salmonella*	TN
1	*Streptococcus dysgalactiae*	No *Salmonella*	TN
1	*Streptococcus pyogenes*	No *Salmonella*	TN
1	*Vibrio alginolyticus*	No *Salmonella*	TN
1	*Yersinia enterocolitica*	No *Salmonella*	TN
Total: 65		Total TN:	65
		Total FP:	0
		Exclusivity (specificity):	100%

[1]: according to the NRC collection or obtained by classical method (i.e. slide-agglutination and biochemical tests) for the *Salmonella* isolates; [2]: obtained with each of the 3 developed assays

It could be observed that the 3 MOL-PCR assays were even able to make "probable" serotype predictions, thereby recommending the user for these cases what classical test to perform for completing the identification. These predictions concerned the serotypes: Butantan, Give and London (**Table 4 and Supplementary Table S4)**.

In conclusion, based on the data produced during this comparison study with the inclusivity (**Table 3**) and exclusivity tests (**Table 4 and Table 5**), the accuracy of the 3 new MOL-PCR assays was determined to be 100%.

5.3.4. Modification of the MOL-PCR BASE

During the development of the 3 new MOL-PCR modules, in view of combining the 4 modules into one complete genoserotyping system, the MOL-PCR BASE module (Gand et al. 2020a) was modified at two levels.

First, an adaptation was needed because when this assay was implemented for routine analyses at the Belgian NRC in 2019 and performed in parallel with the classical method, 3 *S.* Haifa were incorrectly confounded with an uncommon genotype of *S.* Paratyphi B var. Java dT+, detected with the GPP 1365. This GPP is obtained with the rare probe combination including invA (*Salmonella*), SAL-73 (H:1,2), STID16 (O:4) and STID334 (Paratyphi B) but without STID71 (*hemD*22) (**Table 6**). Indeed, this probe combination was retrieved in only 1 of the 54 *S.* Paratyphi B dT+ isolates tested, despite the fact that the SNP marker of *hemD*22 is absent in 46% of the *S.* Paratyphi B population in EnteroBase. No similar problem was observed with the *S.* Paratyphi B dT- isolates (16) as they are always positive for the probe STID34, which discriminate them from the *S.* Haifa isolates that are dT+. To avoid any confusion between *S.* Haifa and *S.* Paratyphi B dT+, the probe STID333, targeting the marker *SPAB_01124* described by Zhai et al. (2014) for the detection of *S.* Paratyphi B (**Table 2**), was added to the MOL-PCR BASE with the prime number 83. It appeared that the marker *SPAB_01124* was absent in some *S.* Paratyphi B dT+ isolates, but not at the same time than *hemD*22. Consequently, the specific detection was successfully obtained for the 54 *S.* Paratyphi B dT+ isolates tested, with the combination of STID334 with at least STID333 or STID71 (**Table 6**). Concerning the 3 *S.* Haifa isolates, the marker *SPAB_01124* was not present in these isolates and they were reported as *Salmonella* O:4 by the DSS (**Supplementary Table S4**).

Secondly, the serogrouping probe STID31 (targeting the serogroup O:3,10) was replaced by the probe STID301 to detect all the *Salmonella* isolates belonging to the serogroup O:3, including O:3,10 and O:1,3,19. All the tested *Salmonella* isolates belonging to O:3 (56 to O:3,10 and 32 to O:1,3,19) were correctly serogrouped as O:3 by STID301 (**Supplementary Table S4**).

Table 6: Probes involved in the identification of *S.* Paratyphi B (dT-/dT+) using the adapted MOL-PCR BASE module.

		Prime numbers:	3	5	7	83	13	11	17		
Reference identification[1]	Number of tested isolates	Antigenic formula[2]	invA (*Salmonella*)	SAL-73 (H:1,2)	STID16 (O:4)	STID333 (Paratyphi B)	STID334 (Paratyphi B)	STID71 (Paratyphi B)[3]	STID34 (dT- variant)	GPP	Comparison expected vs. obtained results
S. Paratyphi B dT-	13	1,4,[5],12:b:1,2	+	+	+	+	+	+	+	21186165	13 TP
S. Paratyphi B dT+	16	1,4,[5],12:b:1,2	+	+	+	+	+	+	-	1246245	16 TP
S. Paratyphi B dT+	1	1,4,[5],12:b:1,2	+	+	+	+	+	-	-	113295	1 TP
S. Paratyphi B dT+	37	1,4,[5],12:b:1,2	+	+	+	-	+	+	-	15015	37 TP
S. Haifa	3	1,4,[5],12:z_{10}:1,2	+	+	+	-	+	-	-	1365	3 TN
Total	70									Total TP:	67
										Total TN:	3
										Total FP:	0
										Total FN:	0
										Inclusivity (sensitivity):	100%
										Exclusivity (specificity):	100%
										Accuracy:	100%

[1]: Obtained with the classical method, i.e. slide-agglutination and biochemical test
[2]: according to Grimont and Weill, 2007
[3]: present in 54% of the S. Paratyphi B population in EnteroBase (Gand et al. 2020a)
TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative

To test the results' interpretation and the recommendations provided by the DSS, all the isolates (553) used in this study were analyzed with the module BASE. From this, 94 were completely identified with the module BASE only and 394 were partially identified (i.e. probable serotype or determination of the serogroup only), 32 were determined to be *Salmonella* from unknown serogroup and 33 were identified as No *Salmonella*, all in agreement with the expected results. For the 394 *Salmonella* isolates partially identified, 389 were successfully recommended by the DSS to be analyzed by one of the 3 other MOL-PCR assay. To confirm some of the probable serotypes, the DSS recommended to complete the identification using the classical method (**Supplementary Table S4**). The MOL-PCR BASE was already validated by Gand et al. (2020) with an accuracy of 99.7%, so its specificity was only evaluated here for the addition of STID333 and STID301. As no false positive nor false negative were obtained (**Table 6 and Supplementary Table S4**), the accuracy was determined to be 100% for these modifications.

5.4. Discussion

In the present study, 3 new MOL-PCR assays (MOL-PCR O3-4-21, MOL-PCR O7 and MOL-PCR O9) were developed for a fast and accurate genoserotyping of common *Salmonella* serotypes (and their variants) which are possible to be isolated in Belgium from the poultry and pork sectors. The MOL-PCR assays O7 and O9 can also be used for a fast detection of important invasive serotypes (e.g. *S.* Choleraesuis and *S.* Gallinarum) if, based on clinical symptoms, they are suspected to cause infections in animal breeding. As such, actions can be quickly taken to avoid the spread of these invasive serotypes. Moreover, when the sampling at the poultry farms is too close to the time point of vaccination of the breeding against *S.* Enteritidis, the MOL-PCR O9 can be used for a reliable discrimination between wild-type and vaccine strains. For the validation of these 3 tests, a comparison study with the classical method, using 464 bacterial isolates, was conducted and an accuracy of 100% was obtained for the detection of the serotypes included in the modules, i.e. Agona, Anatum, Brandenburg, Choleraesuis, Derby, Enteritidis (including the vaccine strains), Gallinarum var. Gallinarum/Pullorum, Livingstone, Mbandaka, Minnesota, Ohio, Rissen and Senftenberg.

Additionally, a previously developed MOL-PCR assay (Gand et al. 2020a), called here module BASE, was modified. This module BASE can identify the 6 serotypes (Enteritidis, Hadar, Infantis, Paratyphi B including its variant Java, Typhimurium including its monophasic variant and Virchow) subjected to an official control (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8) and cluster the other unknown serotypes in one of the common serogroup (O:3, O:4, O:7, O:8, O:9, O:21) if they belong to one of them. Despite the fact that the module BASE was previously validated by (Gand et al. 2020a) with the analysis of 1 004 bacterial isolates belonging to 114 of the most common serotypes, 3 rare *S.* Haifa isolates were wrongly

reported in the current study as an in Belgium uncommon genotype of Paratyphi B var. Java, thus leading to a false positive result. This shows the limitation of the tests performed for the validation of alternative molecular methods which use only isolates belonging to the most common serotypes (*S.* Haifa was isolated only 7 times in Belgium during the last 5 years, personal communication NRC) because it would be too labor intensive to analyze the 2 500 *Salmonella* serotypes of the WKL scheme. Moreover, only few laboratories in the world have the complete collection of these 2 500 serotypes. In the current study, modifications were made to the BASE module to exclude the 3 *S.* Haifa isolates detected as false positive. Additionally, improvements were made to the detection of *Salmonella* isolates belonging to serogroup O:3, including the cluster O:1,3,19. These 2 modifications were validated, identically as for the 3 new MOL-PCR assays, with an accuracy of 100%. Finally, a DSS accessible through a web-application was created for an automatic interpretation of the Luminex data, using a barcode system (GPP), and for the centralization of the results in a database improving the surveillance at a national level. Furthermore, this DSS is also able to provide recommendations to the users in case of partial identifications, i.e. probable serotype or serogroup clustering only.

For the serotypes Derby, Paratyphi B (without *hemD22* marker), and Senftenberg, the genotype distribution observed during the study, i.e. 34%, less than 1% and 25%, respectively, was different from the one estimated *in silico* with EnteroBase, i.e. 16%, 46% and 10%, respectively. This can be explained by the fact that the *Salmonella* isolates used in this study were selected from the collection of the Belgian NRC, composed of routine samples isolated in Belgium, and therefore are more representative of the genotypes circulating in this country. In contrast, the genomic data of *Salmonella* samples uploaded in EnteroBase come from all over the world but are not identical to the frequencies of prevalent serotypes obtained with a national surveillance program.

Altogether, the 4 MOL-PCR assays piloted by the DSS compose a validated *Salmonella* genoserotyping system. The MOL-PCR BASE module is recommended to be used for a first screening of new samples because if one of the serotypes targeted by this module is present at the breeding site, strict and constraining disinfection procedures must be performed at the farm, and the animals are excluded from the food chain which leads to economical loss for the farmers. In case of a partial identification result obtained with this first assay, the DSS displays a recommendation to the user on which of the 3 other MOL-PCR modules (MOL-PCR O3-4-21, MOL-PCR O7 and MOL-PCR O9) he/she can perform to complete the identification of the isolates (**Figure 2**). In some cases, the web-application will recommend to directly switch to the serotyping by slide-agglutination with targeted antisera to test.

The MOL-PCR and the Luminex technology used to develop the genoserotyping system presented in this study allow a high-throughput analysis as the method is based on experiments in a 96-well plate. Additionally, the Mag-Pix apparatus offers the possibility to perform several MOL-PCR assays (including other tests based on the Luminex xTAG technology) at the same time, thereby saving time

and money. Indeed, the MOL-PCR BASE module can be run for new samples simultaneously with the other MOL-PCR modules used to complete the identification of isolates partially identified the day before, with a limit of 96 samples in total. The addition of samples to the plate does not drastically increase the price of the analysis, but rather allows to reduce the cost per sample (Gand et al. 2020).

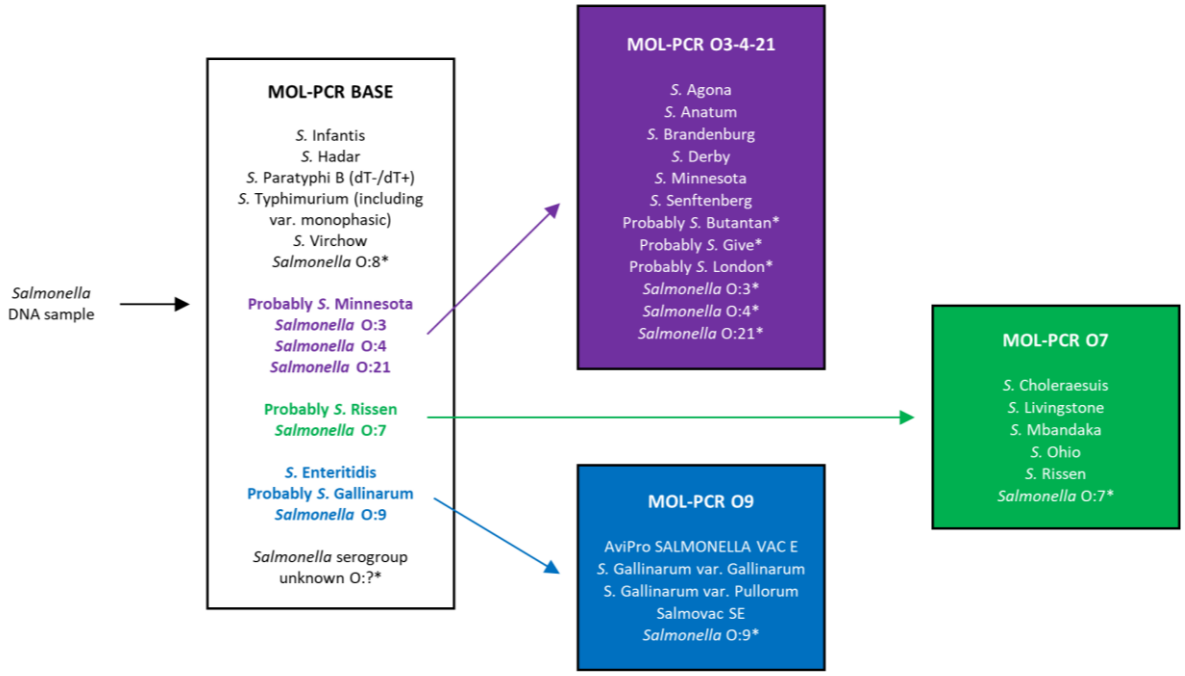

Figure 2: Recommended workflow for *Salmonella* serotype identification using the 4 MOL-PCR assays, as configured in the DSS.

The figure shows the serogroups, serotypes and their variant that can be identified with each of the 4 modules. The MOL-PCR BASE is used for a first screening of all DNA samples extracted from *Salmonella* isolates. For the partial identification obtained using this module, indicated in purple, green and blue in the figure, the DSS recommends to use the MOL-PCR O3-4-21, MOL-PCR O9 or MOL-PCR O7, respectively, to complete the identification if possible. For the partial identifications marked with a star (*), the DSS indicates this time to use the classical method, i.e. slide-agglutination and biochemical tests, to complete the identification.

Compared to the classical method (i.e. slide-agglutination and biochemical tests) for which 2 to 9 days of analyses are usually needed, performed by an experimented technician at the NRC, the developed genoserotyping system does not require particular skills, and objective results are obtained in 1 to 2 days for the targeted serotypes. Therefore, the genoserotyping test can easily be implemented in first-line laboratories as well as in NRCs and helps to reduce the analysis time, thus complying with the short TAT required by the food sector. Moreover, as the price of the antisera is constantly increasing, the use of this alternative molecular method is also cost-effective. The professionals of the food sector are not ready to pay the expensive price required to completely identify by classical method the *Salmonella* serotypes which are not subjected to an official control. However, when the new and less

expensive MOL-PCR assays developed in this study will be used, they will be more disposed to do so. Additionally, as all the identification results obtained with these assays are centralized in a database included in the DSS, this will help the transmission of the serotype identification data to the NRC and thus improve the *Salmonella* surveillance at a national level.

According to the serotyping analyses performed at the NRC between 2017 and 2018, the MOL-PCR BASE could have completely identified 59% and 54% of the *Salmonella* isolates coming from food and veterinary (including animal feed) sources, respectively. Among the remaining unidentified samples, 36% from food source and 50% from veterinary source (including animal feed) could have been genoserotyped by one of the 3 other modules. Concerning the rest of the samples, the serotype identification must be determined using the classical method. But the MOL-PCR and Luminex technologies are modular and the composition of the modules can easily be adapted if needed, like it was the case for the module BASE in this study, following the evolution of the most common serotypes circulating in Belgium according to the database of the DSS or the modifications of the law. Since 2017, *Salmonella* serovar Newport reached the top 5 of the most commonly reported cases in Europe (EFSA 2018, 2019b). A part of the genotypes composing the *S.* Newport population can already be detected as "Probable serotype" by the module BASE (Gand et al. 2020a). But the detection of this serotype could be improved by including complementary markers either in the module BASE or in the module O7 (thus becoming the module O7-8).

Another alternative method for *Salmonella* typing is based on WGS which offers a complete identification of *Salmonella* isolates (including serotype, variant and subtype), in addition to providing other information such as antibiotic resistance or phylogenetic profiles (Ibrahim and Morin 2018; Pornsukarom, van Vliet, and Thakur 2018; Yachison et al. 2017). WGS is already routinely used by big public health institutes in Canada, France, United Kingdom and U.S.A. (Allard 2016; Ashton et al. 2016; Institut Pasteur 2018; Jain, Mukhopadhyay, and Thomassin 2019). But despite the fact that this technology is more complete compared to target-based methods like MOL-PCR, it is time-consuming (1 analysis takes at least 4 days) and too expensive for small institutions with limited resources, such as first-line laboratories which have to respect short TAT and cannot wait for sample batching to reduce analysis costs (Ibrahim and Morin 2018).

In conclusion, unless WGS will become the mandatory, less expensive and more rapid gold standard method in the future, for the characterization of *Salmonella* (including the serotyping), target-based molecular methods such as MOL-PCR linked to the Luminex technology still have their utility. The genoserotyping system developed in this study is able to perform a fast and cheap identification of the most common *Salmonella* serotypes isolated from the poultry and pork sectors. With this method, objective and accurate results are obtained thanks to the automatic interpretation of the Luminex data by a DSS which can also give recommendations for further testing in case of partial identification. Consequently, the method is fully adapted to the needs of the food-producing animal sector. The

database present in the DSS will also help to improve the surveillance of *Salmonella* serotypes at the national level and orientate the future modifications of the module composition to follow the trends of the most prevalent serotypes in Belgium.

Acknowledgements

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 14/5 SALMSTID 1. Whole genome sequencing was performed at the service Transversal Activities in Applied Genomics and the DSS was developed in collaboration with the ICT service, both at Sciensano. The authors want to thank the technicians of the National Reference Center for *Salmonella* and *Shigella*, for their expertise in serotyping by slide-agglutination and their participation to the comparison study; and the technicians of the service Transversal Activities in Applied Genomics for their technical support of the MagPix.

Supplementary data

Supplementary Table S1, Supplementary Table S2, Supplementary Table S3 and Supplementary Table S4 can be found online at: https://doi.org/10.1016/j.fm.2020.103534

CHAPTER 6
Evaluation of a genoserotyping system for the serotype identification of auto-agglutinable *Salmonella* isolates

Context of this chapter:

This chapter illustrates the applicability of the developed method. The *Salmonella* genoserotyping system, composed of the 4 MOL-PCR modules and the DSS, described in chapters 4 and 5, were evaluated in the current chapter, and compared to historical identification data and Whole Genome Sequencing, for its ability to completely identify auto-agglutinable *Salmonella* isolates. These are isolates which cannot be serotyped by the slide-agglutination method, and therefore their clinical relevance and the compliance with the legislation cannot be determined by the classical method. From these data the serotype prevalence of auto-agglutinable isolates coming from the routine analyses of the Belgian National Reference Center (NRC) during the years 2016-2018 was estimated.

This chapter is ready to be submitted.

Contributions:

M. Gand designed the study, performed the experiments and analyzed the data. W. Mattheus and K. Marchal provided specialist feed-back. S.C.J. De Keersmaecker and S. Bertrand conceived and supervised the study, helped to design the study and to interpret the results.

Abstract:

Salmonella is a major pathogen divided into different serotypes, each potentially leading to different symptoms, from simple gastroenteritis to the more dangerous typhoid fever. Therefore, some of these serotypes are subjected to official controls along the food chain with the aim to limit their transmission to humans (EU regulation N°2160/2003). The *Salmonella* serotypes are determined through the characterization of 2 surface antigens (O and H) with an agglutination reaction using specific sera. In addition to be expensive, time-consuming and highly subjective, this technique can sometimes lead to untypable results due to an auto-agglutination reaction. In this case, the clinical relevance of the isolates and the compliance with the legislation cannot be determined. Fortunately, these last years, molecular methods showed to be good alternatives for *Salmonella* serotype identification. In this study, some of these techniques, including Whole Genome Sequencing (WGS) and a recently developed *Salmonella* genoserotyping system, using the Multiplex Oligonucleotide Ligation - PCR (MOL-PCR) and Luminex technology, were evaluated for the identification of auto-agglutinable isolates historically identified by PCR and Pulsed Field Gel Electrophoresis (PFGE) or coming from the routine analyses of the Belgian National Reference Center (NRC). *S.* Paratyphi B var. Java (34%), *S.* Typhimurium (22%) and its monophasic variant (11%) were the most prevalent serotypes among the auto-agglutinable *Salmonella* isolates retrieved from 2016-2018.

6.1. Introduction

Salmonella enterica subsp. *enterica* is a world-wide spread bacterium that can infect either humans, animals or both depending on its serotype of which there exist more than 1 500 different ones. Indeed, some serotypes of this pathogen are more adapted (or restricted) to a specific host while others can contaminate a wide range of species. The serotype and the host species determine also the type of the resulting salmonellosis: the less harmful, even sometimes silent, non-invasive type causing simple gastroenteritis or the more dangerous invasive type leading to fever, bacteremia and decease of the host without treatment. Consequently, the identification of *Salmonella* serotypes is needed for a proper surveillance of circulating strains that can infect humans. *Salmonella* is the second zoonotic agent responsible of large outbreaks in the European Union, mostly due to the consumption of contaminated food (EFSA 2018). To limit this way of contamination, some *Salmonell*a serotypes, such as Enteritidis, Hadar, Infantis, Paratyphi B var. Java, Typhimurium including its monophasic variant 1,4,[5],12:i:– and Virchow, are subjected to an official control (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). It means that their absence must be proven in some type of poultry breeding (see section 1.3.4). Indeed, if one of these serotypes is detected in adult breeding flock of *Gallus gallus* for example, the entire flock must be eliminated by logistic slaughtering and eggs coming from these animals are destroyed or can undergo a thermic treatment, if used in the food industry, depending on their incubation status. Finally, the farm has to be entirely disinfected and decontaminated before the installation of a new animal batch, after demonstration of the absence of *Salmonella* on site.

The determination of the *Salmonella* serotypes and their variants is historically performed since 80 years with the gold standard methods including slide-agglutination and biochemical tests following the White-Kauffmann-Le Minor (WKL) scheme (Grimont and Weill 2007). The slide-agglutination consists of the characterization of 3 antigenic sites (somatic antigen O and flagellar antigens H) located at the surface of the bacteria, by an agglutination reaction with specific antisera. A minimum of 120 antisera is needed to be able to properly identify all the antigenic formula of the WKL. But the agglutination reaction is not always easy to be determined by the human eye. Therefore, the interpretation of the results depends on the experience of the laboratory technician and is not always objective. This complex and subjective test is, most of the time, only fully mastered at the National Reference Centers (NRCs) which are in charge of the serotype identification of *Salmonella* isolates sent for characterization from human, food or veterinary sources. Sporadically, the antigenic formula cannot be determined by slide-agglutination because the sample shows an absence of agglutination reaction or a nonspecific reaction with all the sera tested. In this last case, the isolate is characterized as auto-agglutinable and reported as non-typable. Rough *Salmonella* isolates, i.e. isolates showing an altered antigen O structure and having a rough aspect when cultured on Petri dish, are usually not agglutinable but the data of some studies

suggested that they could also be responsible of auto-agglutinable characteristics due to chemical modifications of their membrane (Lalsiamthara, Kim, and Lee 2018; Herzberg and Green 1964). But the exact mechanisms leading to auto-agglutinability are not well described in the literature and stay unclear. At the Belgian NRC, 7% of the *Salmonella* isolates sent yearly from the food and animal sectors between 2014 and 2016 were auto-agglutinable (Personal communication NRC). These unidentified isolates lead to an incomplete diagnostic, not compatible with the regulation which requires the exclusion of some major serotypes as mentioned above and in section 1.3.4 (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). Indeed, if one of these serotypes is isolated from certain type of poultry breeding, but cannot be properly identified, no restrictive actions will be taken. Consequently, these major serotypes will be able to persist in the farms and potentially be transmitted to humans, where it can cause disease.

Fortunately, alternative molecular methods, including PCR and Pulsed Field Gel Electrophoresis (PFGE), are available for the determination of *Salmonella* serotypes (Wattiau, Boland, and Bertrand 2011). In the 1990s, PFGE was applied to *Salmonella* and rapidly became the following years the gold standard method, implemented in laboratories worldwide, for the subtyping of this pathogen. This method also showed a reliable link between PFGE patterns and serotype (Bopp et al. 2016; Kérouanton et al. 2007; Zou et al. 2010) and was used for the identification of auto-agglutinable isolates by Hoszowski et al. (2011). In 2007, the Belgian Federal Agency for the Safety of the Food Chain (FASFC) asked the NRC to identify by a molecular method the serotype of auto-agglutinable isolates. Following this request, PFGE was successfully used at the NRC to determine the serotype of 66 auto-agglutinable *Salmonella* isolates coming from the food sector. The results were additionally confirmed with the determination of the flagellar antigens H1 and H2 using PCR methods developed by Herrera-León et al. (2004) and Echeita et al. (2002), respectively. But PFGE is not commonly used anymore and less suitable for routine and rapid identification because it is labor-intensive and sometimes with limited reproducibility between laboratories, although efforts have been made to build an international standardized protocol and database through the PulseNet network (Wattiau, Boland, and Bertrand 2011; Keefer et al. 2019; Kozyreva et al. 2016). Nowadays, Whole Genome Sequencing (WGS) of bacterial isolates is the most accurate method as it offers a full characterization of *Salmonella* samples in one shot, including serotype identification, subtyping and antimicrobial resistance determination. But despite the fact that this method is very complete and provides a lot of data, it is still time-consuming and expensive. Indeed, an analysis by WGS takes at least 4 days of work to be completed and is cost-effective only when batching of multiple samples in one sequencing run is possible (Ibrahim and Morin 2018). Thus, this technique is not always adapted to smaller laboratories with limited resources and a low sample flow, that cannot wait for batching because this increases the time until reporting of the result to the client (also called Turn-Around Time (TAT)). This is why WGS cannot be implemented everywhere and is routinely used only by big institutes such as Public Health England, the American

Food & Drug Administration, the Public Health Agency of Canada or the French Pasteur Institute (Allard 2016; Ashton et al. 2016; Institut Pasteur 2018; Robertson, Yoshida, Kruczkiewicz, et al. 2018; Yachison et al. 2017).

In this context, other molecular techniques, target-based, like the genoserotyping method developed at the Belgian NRC for *Salmonella* and using a Multiplex Oligonucleotide Ligation-PCR (MOL-PCR) coupled to a liquid bead suspension assay (Luminex xTAG technology), have proven to be more adapted to rapid, objective and cost-effective routine identification of *Salmonella* serotypes (Gand et al. 2020a; Gand et al. 2020b). This method is based on the screening of molecular markers in the bacterial genome, that are specific to serotypes. These molecular markers are specifically detected through a ligation-dependent amplification reaction (MOL-PCR) using ligation probes. The generated MOL-PCR fragments are subsequently hybridized to unique color-coded microspheres (Mag-Plex TAG), themselves detected by a Luminex device through a fluorescence staining. The genoserotyping system developed by Gand et al. (2020b) is composed of 4 MOL-PCR assays piloted by a Decision Support System (DSS) which performs an objective and automatic interpretation of the Luminex results. The main multiplex assay, called MOL-PCR BASE, is first used for the screening of all *Salmonella* isolates because it targets the 6 serotypes mentioned above which are subjected to an official control along the food chain (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). Using this MOL-PCR BASE at the Belgian NRC, the complete identification could be obtained for 77% of the *Salmonella* samples sent for characterization in the previous year (Gand et al. 2020a). From the remaining 33%, some isolates are partially identified and their serogroup is determined when they belong to O:3, O:4, O:7, O:8, O:9 or O:21. For these partial identifications, the DSS recommends to the user what further analysis he/she has to run to complete the identification, i.e. one of the 3 other MOL-PCR assays (MOL-PCR O3-4-21, MOL-PCR O7 or MOL-PCR O9) or the classical method (slide-agglutination and biochemical tests). These 3 other MOL-PCR assays target highly invasive serotypes like Choleraesuis and Gallinarum, including the variants Gallinarum and Pullorum, or serotypes frequently isolated in the poultry and pork sector such as Agona, Anatum, Brandenburg, Derby, Livingstone, Mbandaka, Minnesota, Ohio, Rissen and Senftenberg. They are able to identify 36% and 50% of the *Salmonella* isolates sent to the Belgian NRC from food and veterinary sources, respectively, and which were not completely identified by the MOL-PCR BASE (Gand et al. 2020b).

The genoserotyping system developed by Gand et al. (2020a and 2020b) could be a good alternative method for the identification of *Salmonella* strains isolated from the field by routine laboratories and non-typable by slide-agglutination. To evaluate this, this test was used in this study for the analysis of the 66 auto-agglutinable strains, isolated from food sources in 2007, and identified at that time using PFGE and PCRs methods. Some of these identifications were confirmed by WGS and the web tool *Salmonella In Silico* Typing Resource (SISTR) developed by Yoshida, Kruczkiewicz, et al. (2016) to perform serotype identification from WGS data. Secondly, the genoserotyping system was used to

identify the serotype of 74 additional *Salmonella* strains, isolated in Belgium between 2016 and 2018 from human, food or animal sources, and determined as auto-agglutinable at that time. From these results, the serotype prevalence of these non-typable isolates was determined.

6.2. Materials and methods

6.2.1. Bacterial strains

The *Salmonella* strains used in this study are bacterial isolates sent to the Belgian NRC by first-line laboratories for routine characterization. All the 66 isolates from 2007 were isolated from food. The 74 strains isolated between 2016 and 2018, came from human (13), food (51) and veterinary (10) sources. These were the ones for which a culture could be obtained from the total of 130 strains reported by the Belgian NRC at that time as auto-agglutinable.

The *Salmonella* genus identification has been confirmed by the first-line laboratories before the transfer to the NRC where they were characterized as auto-agglutinable by experienced technicians using slide-agglutination. All these isolates were stored in the collection of the Belgian NRC and are available upon request. The isolates were cultured on Nutrient agar (Neogen® Culture Media, Lansing, USA).

6.2.2. Molecular serotyping using PFGE and PCR methods

In 2007, analyses by PFGE were performed according to the PulseNet Europe protocol (www.cdc.gov/pulsenet), and the genomic DNA was digested with restriction enzymes XbaI (New England Biolabs, Leusden, The Netherlands). *Salmonella* Braenderup H9812 digested with XbaI was used as a size marker. The PFGE profiles were compared using BioNumerics software (Applied Maths, Sint-Martens-Latem, Belgium; version 7.6). The generated bands were analyzed by using the Dice coefficient and the unweighted pair group method with averages and tolerance of 1%. Serotype identification was attributed according to comparisons with specific PFGE patterns determined using the pulsed field database of the Belgian NRC of which a part is presented in **Figure 1**.

To confirm the serotype identification obtained with PFGE method, the antigenic flagellar formula H1 and H2 were additionally determined according to the PCR methods described by Herrera-León et al. (2004) and Echeita et al. (2002), respectively.

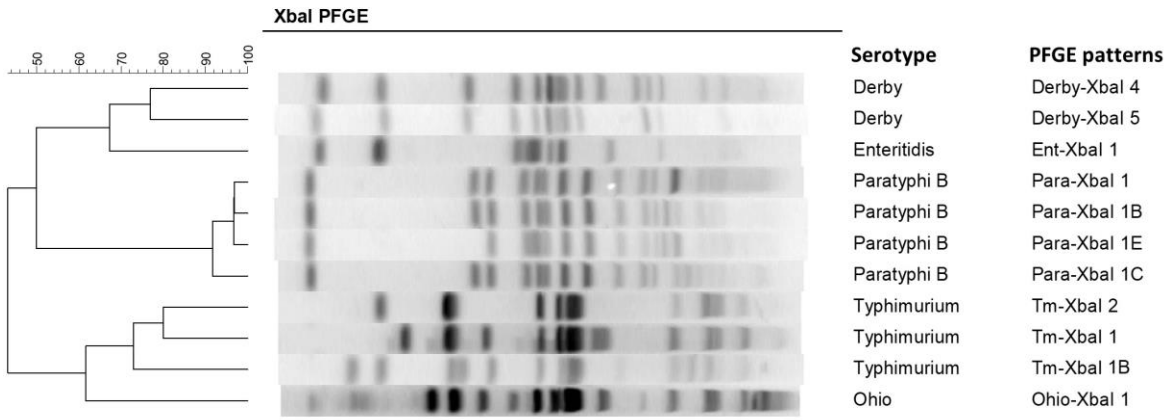

Figure 1: PFGE patterns extracted from the pulsefield database of the Belgian NRC and specific to *Salmonella* serotypes.
A phylogenetic tree was created with completely identified isolates to establish the correlation between PFGE patterns and serotypes. This correlation was later used to deduce the serotype of auto-agglutinable isolates according to their PFGE patterns.

6.2.3. Salmonella *serotype identification and genome analysis using WGS data*

The genomic DNA of 16 and 13 auto-agglutinable *Salmonella* isolates from 2007 and 2016-2018, respectively, was sequenced using an Illumina MiSeq instrument (2 x 250 bp, Nextera XT libraries). The FASTQ reads are accessible at the SALMSTID BioProject on NCBI (PRJNA509747).

Using the software CLC Genomics Workbench 8.0 (Qiagen, Hilden, Germany), the raw FASTQ reads were first trimmed to a quality score limit of 0.05 with maximum 2 ambiguous nucleotides and the reads with a length below 30 nucleotides were discarded. Then, these trimmed reads were *de novo* assembled with automatic bubble and word size, in mapping mode "map reads back to contigs", with scaffolding, and a minimum contig length of 1 000 nucleotides. All the generated assemblies were uploaded to SISTR (Yoshida, Kruczkiewicz, et al. 2016) in FASTA format for serotype identification.

Multiple alignments were performed using the BioNumerics software (Applied Maths, Sint-Martens-Latem, Belgium; version 7.6). The specific sequence of the O:7 molecular marker used for the probe STID18 in the MOL-PCR BASE and MOL-PCR O7 (CGTTGGCAGACTGGTACTGATTGGCTCCCCTATTACGATGATTTC; Gand et al. 2020a) was screened for in this multiple alignment using the "Sequence Search" function.

6.2.4. Genoserotyping by MOL-PCR and Luminex technology

The serotype identification using the 4 modules MOL-PCR BASE, MOL-PCR O-4-21, MOL-PCR O7, MOL-PCR O9 and the DSS was performed according to the protocol fully described by Gand et al. (2020b). Briefly, all the isolates were first analyzed with the module MOL-PCR BASE. When only the

serogroup could be determined using this module, the concerned *Salmonella* isolates were subsequently analyzed with one of the 3 other modules according to the recommendation provided by the DSS.

6.3. Results

6.3.1. Use of PFGE, PCR and WGS for Salmonella *serotype identification of auto-agglutinable isolates*

When the serotype identification of *Salmonella* isolates cannot be determined by slide-agglutination because they are behaving as auto-agglutinable, alternative molecular methods must be employed. In 2007, the PFGE method was used to identify 66 auto-agglutinable *Salmonella* isolates coming from the food sector. The PFGE patterns of these isolates were compared to the patterns from the pulsed field database of the Belgian NRC (**Figure 1**) and 2 *S.* Derby, 2 S. Enteritidis, 1 *S.* Ohio, 52 *S.* Paratyphi B var. Java and 9 *S.* Typhimurium were identified (**Table 1**). To confirm these identifications, the 66 *Salmonella* isolates were also analyzed in 2007 using the PCR methods of Herrera-León et al. (2004) and Echeita et al. (2002). According to the size of the amplicons obtained using electrophoresis on agarose gel, the H1 and H2 formula were determined. For all the isolates analyzed, the obtained flagellar formula were in agreement with the antigenic formula of the serotypes determined by PFGE (**Table 1**). At least one isolate per PFGE pattern was selected for WGS, i.e. 2 *S.* Derby, 2 *S.* Enteritidis, 1 *S.* Ohio, 6 *S.* Paratyphi B var. Java and 5 *S.* Typhimurium. The assembled genomes were exported to SISTR and the serotype identifications obtained by PFGE and PCR methods were confirmed except for 2 *S.* Typhimurium isolates (07-06114 and 07-00516) which were additionally discriminated as monophasic variants 1,4,[5],12:i:– (**Table 1**). *S.* Paratyphi B var. Java represented the most prevalent serotype (52) among the 66 isolates analyzed.

6.3.2. Evaluation of the MOL-PCR genoserotyping system for serotype identification of auto-agglutinable Salmonella *isolates*

To evaluate the ability of the genoserotyping system developed by Gand et al. (2020b) to identify the serotype of auto-agglutinable *Salmonella* strains, the 66 isolates from 2007, identified by PFGE, PCR and WGS, were analyzed with the MOL-PCR BASE. Using this module, the complete identification was obtained for 63 isolates: 2 *S.* Enteritidis, 52 *S.* Paratyphi B var. Java, 7 *S.* Typhimurium and 2 monophasic *S.* Typhimurium. Concerning the 3 remaining isolates, 2 (07-04529 and 07-05518) were identified as *Salmonella* belonging to serogroup O:4 and 1 (07-01263) as *Salmonella* belonging to serogroup O:7.

Table 1: Serotype identification of auto-agglutinable *Salmonella* isolates using alternative molecular methods

ID strain	Serotype identification by PFGE		Identification of flagellar antigen H1 and H2 by PCR methods				Serotype identification using WGS and SISTR	Serotype Identification using the genoserotyping system		Comparison [4]
	PFGE profile	Serotype[1]	Amplicon size H1[2]	H1 antigenic formula[2]	Amplicon size H2[3]	H2 antigenic formula[3]		First screening with the module BASE	Second analysis using a complementary module	
07-04529	Derby-XbaI 4	Derby	500 bp	g	-	-	*S.* Derby	*Salmonella* O:4	*S.* Derby[5]	OK
07-05518	Derby-XbaI 5	Derby	500 bp	g	-	-	*S.* Derby	*Salmonella* O:4	*S.* Derby[5]	OK
07-00858	Ent-XbaI 1	Enteritidis	500 bp and 300 bp	g,m	-	-	*S.* Enteritidis	*S.* Enteritidis	*S.* Enteritidis wild-type strain[6]	OK
07-04877	Ent-XbaI 1	Enteritidis	500 bp and 300 bp	g,m	-	-	*S.* Enteritidis	*S.* Enteritidis	*S.* Enteritidis wild-type strain[6]	OK
07-01263	Ohio-XbaI 1	Ohio	150 bp	b	250 bp	l,w	*S.* Ohio	*Salmonella* O:7	*S.* Ohio[7]	OK
07-00764	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	*S.* Paratyphi B var. Java	*S.* Paratyphi B var. Java	-	OK
07-00805	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	*S.* Paratyphi B var. Java	*S.* Paratyphi B var. Java	-	OK
07-01049	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	*S.* Paratyphi B var. Java	*S.* Paratyphi B var. Java	-	OK
07-01050	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01092	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01093	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01094	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01431	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01553	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01608	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01633	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01634	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01637	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK

(continued on next page)

Table 1 (continued)

ID strain	Serotype identification by PFGE		Identification of flagellar antigen H1 and H2 by PCR methods				Serotype identification using WGS and SISTR	Serotype Identification using the genoserotyping system		Comparison [4]
	PFGE profile	Serotype[1]	Amplicon size H1[2]	H1 antigenic formula[2]	Amplicon size H2[3]	H2 antigenic formula[3]		First screening with the module BASE	Second analysis using a complementary module	
07-01646	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01647	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01648	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01672	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01674	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01675	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01721	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01761	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01762	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01768	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01811	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01812	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-01966	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02116	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02117	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02136	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02234	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK

(continued on next page)

Table 1 (continued)

ID strain	Serotype identification by PFGE		Identification of flagellar antigen H1 and H2 by PCR methods				Serotype identification using WGS and SISTR	Serotype Identification using the genoserotyping system		Comparison [4]
	PFGE profile	Serotype[1]	Amplicon size H1[2]	H1 antigenic formula[2]	Amplicon size H2[3]	H2 antigenic formula[3]		First screening with the module BASE	Second analysis using a complementary module	
07-02340	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02382	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02394	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02400	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02401	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02673	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-02882	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-03951	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-03953	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-04089	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-04134	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-04163	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-04184	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-04382	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-04763	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-05350	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-05870	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-05930	Para-XbaI 1	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK

(continued on next page)

Table 1 (continued)

ID strain	Serotype identification by PFGE		Identification of flagellar antigen H1 and H2 by PCR methods				Serotype identification using WGS and SISTR	Serotype Identification using the genoserotyping system		Comparison [4]
	PFGE profile	Serotype[1]	Amplicon size H1[2]	H1 antigenic formula[2]	Amplicon size H2[3]	H2 antigenic formula[3]		First screening with the module BASE	Second analysis using a complementary module	
07-04649	Para-XbaI 1B	Paratyphi B var. Java	150 bp	b	400 bp	1,2	*S.* Paratyphi B var. Java	*S.* Paratyphi B var. Java	-	OK
07-02229	Para-XbaI 1C	Paratyphi B var. Java	150 bp	b	400 bp	1,2	-	*S.* Paratyphi B var. Java	-	OK
07-03315	Para-XbaI 1C	Paratyphi B var. Java	150 bp	b	400 bp	1,2	*S.* Paratyphi B var. Java	*S.* Paratyphi B var. Java	-	OK
07-04830	Para-XbaI 1E	Paratyphi B var. Java	150 bp	b	400 bp	1,2	*S.* Paratyphi B var. Java	*S.* Paratyphi B var. Java	-	OK
07-02082	Tm-XbaI 1	Typhimurium	250 bp	i	400 bp	1,2	-	*S.* Typhimurium	-	OK
07-04673	Tm-XbaI 1	Typhimurium	250 bp	i	400 bp	1,2	*S.* Typhimurium	*S.* Typhimurium	-	OK
07-02799	Tm-XbaI 1B	Typhimurium	250 bp	i	400 bp	1,2	-	*S.* Typhimurium	-	OK
07-04526	Tm-XbaI 1B	Typhimurium	250 bp	i	400 bp	1,2	*S.* Typhimurium	*S.* Typhimurium	-	OK
07-00575	Tm-XbaI 1B	Typhimurium	250 bp	i	400 bp	1,2	-	*S.* Typhimurium	-	OK
07-00672	Tm-XbaI 1B	Typhimurium	250 bp	i	400 bp	1,2	-	*S.* Typhimurium	-	OK
07-00972	Tm-XbaI 2	Typhimurium	250 bp	i	400 bp	1,2	*S.* Typhimurium	*S.* Typhimurium	-	OK
07-06114	Tm-XbaI 2	Typhimurium	250 bp	i	400 bp	1,2	*S.* Typhimurium var. monophasic	*S.* Typhimurium var. monophasic	-	OK
07-00516	Tm-XbaI 3	Typhimurium	250 bp	i	400 bp	1,2	*S.* Typhimurium var. monophasic	*S.* Typhimurium var. monophasic	-	OK

[1]: determined according to Figure 1.
[2]: after electrophoresis on agarose gel according to the method developed by Herrera-León et al. (2004).
[3]: after electrophoresis on agarose gel according to the method developed by Echeita et al. (2002).
[4]: the comparison is "OK" if the same serotype identification is retrieved using all the molecular methods i.e. PFGE, PCR, WGS (if available) and genoserotyping system.
[5]: Identification obtained using the MOL-PCR O3-4-21.
[6]: Identification obtained using the MOL-PCR O9.
[7]: Identification obtained using the MOL-PCR O7.
Antigenic formula of serotypes Enteritidis, Derby, Ohio, Paratyphi B var. Java and Typhimurium are respectively, 1,9,12:g,m:–, 1,4,[5],12:f,g:[1,2], 6,7,14:b:l,w, 1,4,[5],12:b:1,2 and 1,4,[5],12:i:1,2 according to the WKL scheme.

Following the recommendations of the DSS, the isolates 07-04529 and 07-05518 were analyzed with the MOL-PCR O3-4-21, and both identified as *S.* Derby. Similarly, the isolate 07-01263 was analyzed with the MOL-PCR O7 and identified as *S.* Ohio (**Table 1**). Identically as for WGS, the MOL-PCR BASE was even able to make the discrimination between *S.* Typhimurium and its monophasic variant 1,4,[5],12:i:–, which was not possible using PFGE and PCR methods. Additionally, the differentiation between the vaccine and wild-type strain was made for *S.* Enteritidis 07-00858 and 07-04877 using MOL-PCR O9 and these 2 isolates were reported as *S.* Enteritidis wild-type strains (**Table 1**). All the results obtained with the genoserotyping system were in agreement with those obtained by PFGE and PCR, or even more discriminative, demonstrating the ability of this method to clearly identify auto-agglutinable *Salmonella* isolates.

6.3.3. MOL-PCR genoserotyping of auto-agglutinable Salmonella *isolates selected from routine analyses*

A total of 74 *Salmonella* strains, taken from the routine isolates reported as auto-agglutinable between 2016 and 2018 by the Belgian NRC, were analyzed with the MOL-PCR BASE with the aim to evaluate the ability of the genoserotyping system to completely identify these isolates and to estimate the serotype distribution among them. This module succeeded to completely identify 76% of the tested isolates which belonged to the serotypes Enteritidis (3), Infantis (4), Paratyphi B var. Java (25), Typhimurium (16) and Typhimurium var. monophasic (8). Concerning the remaining isolates, some were partially identified and reported as *Salmonella* belonging to "Probably Kisangani/Heidelberg/Saintpaul/Stanleyville" (2), serogroups O:3 (1), O:4 (5) or O:8 (3), while the serogroup of the others could not be determined (7). Following the recommendations provided by the DSS, 6 partially identified isolates were subsequently analyzed by the MOL-PCR O3-4-21 and identified as *Salmonella* belonging to the serogroup O:1,3,19 (1), *S.* Brandenburg (1) and *S.* Derby (4) (**Table 2**). Using the modules MOL-PCR BASE and MOL-PCR O3-4-21, 61 of the 74 isolates (82%) were completely identified. WGS was performed on the 13 partially or unidentified identified isolates and their serotype identification was determined using SISTR. As such, the 2 probable MOL-PCR identification were confirmed to be *S.* Heidelberg, in agreement with one of the predictions provided by the DSS. The strains determined as *Salmonella* O:8 (3) by the MOL-PCR BASE were identified as *S.* Bovismorbificans (2) and *S.* Stourbridge (1), and the one determined as *Salmonella* O:1,3,19 by the MOL-PCR O3-4-21 was identified as *S.* Kouka, all in agreement with the serogroup clustering obtained from the Luminex results. Among the isolates (7) for which the serogroup could not be determined by the MOL-PCR methods, 2 were identified as *S.* Carrau and *Salmonella enterica* subsp. *houtenae* 44:z4,z23:-. These serotypes and their serogroup were indeed not targeted by any of the MOL-PCR assays composing the genoserotyping system. Unexpectedly, the remaining 5 isolates for which the serogroup could not be determined were identified by WGS as *S.* Livingstone (S18BD09301 and

S18FP02627) and *S.* Rissen (S17FP07900, S18FP05109 and S18FP09545), all belonging to the serogroup O:7. This serogroup is normally targeted by the module BASE, but the concerned 5 isolates were not identified as such, and therefore not analyzed using the MOL-PCR O7. After obtaining the WGS results, these 5 isolates were analyzed with the MOL-PCR O7, resulting into GPP 231 and GPP 663, for *S.* Livingstone and *S.* Rissen isolates, respectively, which represent the specific probe combinations targeting these 2 serotypes but lacking the prime number of the O:7 molecular marker (**Table 2**). To investigate the reason of these unexpected results, multiple alignments of the genomes of the isolates S17FP07900, S18BD09301, S18FP02627, S18FP05109 and S18FP09545 were performed with 3 publicly available (NCBI) genomes of serotypes belonging to O:7, i.e. of *S.* Choleraesuis (CP007639.1), *S.* Infantis (LN649235.1) and *S.* Mbandaka (CP019183.1). The molecular marker specific sequence, used in the probe STID18 for serogroup O:7 detection in the MOL-PCR BASE and O7 assays, was screened for in these multiple alignments. Interestingly, the sequence was found in the NCBI genomes but not in the others. In the genomes of the 5 *S.* Livingstone and *S.* Rissen isolates, a similar gap of 14 148 bp was noticed in the sequence of their *rfb* gene cluster coding for enzymes, sugars and proteins involved in the biosynthesis of subunits composing the somatic antigen O.

Table 2: Genoserotyping of auto-agglutinable *Salmonella* isolates selected from routine analyses

Number of isolates tested	First screening using the module BASE	Second analysis using a complementary module		Identification using WGS	Serotype prevalence*
	Identification	Module	Identification		
25	*S.* Paratyphi B var. Java	-	-	-	34%
16	*S.* Typhimurium	-	-	-	22%
8	*S.* Typhimurium var. monophasic	-	-	-	11%
4	*S.* Infantis	-	-	-	5%
4	*Salmonella* O:4	O3-4-21	*S.* Derby	-	5%
3	*S.* Enteritidis	O9	*S.* Enteridis wild-type strain	-	4%
3	*Salmonella* unknown serogroup	O7	GPP 663	*S.* Rissen	4%
2	Possibly *S.* Kisangani/Heidelberg/Saintpaul/ Stanleyville	-	-	*S.* Heidelberg	3%
2	*Salmonella* O:8	-		*S.* Bovismorbificans	3%
2	*Salmonella* unknown serogroup	O7	GPP 231	*S.* Livingstone	3%
1	*Salmonella* O:3	O3-4-21	*Salmonella* O:1,3,19	*S.* Kouka	1%
1	*Salmonella* O:4	O3-4-21	*S.* Brandenburg	-	1%
1	*Salmonella* O:8	-		*S.* Stourbridge	1%
1	*Salmonella* unknown serogroup	-	-	*S.* Carrau	1%
1	*Salmonella* unknown serogroup	-	-	*S.* IV 44:z4,z23:-	1%

*: among the analyzed isolates

From these results, it appeared that *S.* Paratyphi B var. Java (34%), *S.* Typhimurium (22%) and its monophasic variant (11%) were the most prevalent serotypes among the auto-agglutinable *Salmonella* isolates retrieved from 2016-2018.

6.4. Discussion

The auto-agglutinable isolates are a problem for routine laboratories in charge with the identification of *Salmonella* serotypes. Indeed, the serotype of this non-typable strains cannot be determined by the classical method, i.e. the slide-agglutination, and the professionals of the food sector do not know if they match the criteria of the regulation demanding the exclusion of 6 *Salmonella* serotypes known for their prevalence and public health impact: i.e. Enteritidis, Hadar, Infantis, Paratyphi B var. Java, Typhimurium including its monophasic variant and Virchow (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). Because of this incorrect identification, and consequently the lack of restrictive measures taken, these major *Salmonella* serotypes can persist in the farms and potentially be transmitted to humans through the consumption of food products.

The genoserotyping system developed by Gand et al. (2020b), including 4 MOL-PCR assays and a DSS, allows the detection of the 6 *Salmonella* serotypes mentioned in the legislation, in addition to highly invasive serotypes or other serotypes commonly isolated from the poultry and pork sectors. In the present study, this method was evaluated for the serotype determination of 66 auto-agglutinable isolates previously identified by PFGE and PCR, and partially complemented with WGS for 12 of them. The MOL-PCR and WGS showed a better discriminatory power, with the identification of 2 monophasic variants of *S.* Typhimurium, not determined as such by PFGE and PCR methods. All the other results obtained with the different molecular techniques were in agreement between each other, demonstrating the ability of the MOL-PCR and Luminex technology to clearly and quickly identify auto-agglutinable *Salmonella* isolates non-typable by slide-agglutination. Therefore, the MOL-PCR genoserotyping system was tested for the serotype determination of 74 auto-agglutinable *Salmonella* strains isolated in Belgium between 2016 and 2018, and completely identified 82% of them. For the remaining isolates, including some partially identified by Luminex methods (8%), WGS and SISTR were required for complete identification. Interestingly, 2 *S.* Livingstone and 3 *S.* Rissen isolates could not be serogrouped by the MOL-PCR BASE nor completely identified by the MOL-PCR O7 because the molecular marker used in these modules for O:7 detection was absent in their genomes. This deletion of a part of the DNA sequence coding for the somatic antigen O:7 is maybe responsible for the non-typable character of these strains. Indeed, rough isolates having an altered structure in their O antigen were already shown to have an increased hydrophobic membrane which was associated with a higher auto-agglutinability in saline solution (Herzberg and Green 1964; Lalsiamthara, Kim, and Lee 2018). However, the exact link between

genomic modifications, bacterial membrane alteration and auto-agglutinability needs to be further studied.

As for the isolates from 2007, *S.* Paratyphi B var. Java was the most detected serotype (34%) among the auto-agglutinable isolates analyzed from 2016-2018, followed by *S.* Typhimurium (22%) and its monophasic variant (11%). These are 3 serotypes that are amongst the 6 to combat according to the regulation. Therefore it is of utmost importance to be able to identify these auto-agglutinable strains. Two hypotheses could be made about the overrepresentation of Paratyphi B among the auto-agglutinable serotypes. First, the biosynthesis of surface antigens of this serotype could be more easily and permanently altered by food processing compared to other serotypes, leading to non-specific reactions during slide-agglutination. Secondly, this serotype could be more likely isolated from processed food products, such as eggs based products, leading to the alterations of the surface antigens. Further investigations need to be done to evaluate the impact of food processing on surface antigens biosynthesis.

PFGE was historically used as the gold standard method for *Salmonella* subtyping. Additionally, a reliable link was shown between PFGE patterns and serotype identification (Bopp et al. 2016; Kérouanton et al. 2007; Zou et al. 2010), but this technique, although inexpensive, is time-consuming, labor intensive and not user-friendly for inter-laboratory comparisons. WGS slowly replaced PFGE these last years as it provides additional data for *Salmonella* subtyping and outbreak traceability. However, this method cannot be implemented in all laboratories, especially in the smaller ones with limited resources, because cost-effectiveness is only obtained when sample batching is possible. Target-based molecular methods like the PCR tests developed by Herrera-León et al. (2004) and Echeita et al. (2002) are more adapted to first-line laboratories as an alternative to the serotyping by slide-agglutination. However, these 2 PCR methods can only be used for the identification of the most common flagellar antigens H1 and H2. For the determination of the serogroup O, other assays are needed (Cardona-Castro et al. 2009; Karns, Haley, and Van Kessel 2015; Liu et al. 2011). Therefore, using these PCR methods, at least 3 assays are needed to obtain a complete antigenic formula. The different tests cannot be combined in one big PCR because the number of targets in a multiplex PCR is limited. Additionally, a detection by electrophoresis on agarose gel or through capillary electrophoreses is often required. In this context, the genoserotyping system evaluated in the present study is a better option. Indeed, the MOL-PCR and Luminex technology allow a high level of multiplexing and the complete assay from DNA extraction to result interpretation by the DSS is performed in one day using a 96-well plate. Additionally, the simultaneous use of several MOL-PCR assays is possible in one run, with a limit of 96 reactions including the negative and positive controls, using the multi-batch option of the Luminex device. Used since early 2019 at the Belgian NRC for routine *Salmonella* serotyping, the genoserotyping system including the 4 MOL-PCR assays and the DSS were shown to be rapid, accurate, cost-effective and could identify more than 75% of the *Salmonella* samples sent for characterization (Gand et al. 2020a

and 2020b). Thanks to this method, the number of auto-agglutinable isolates reported as non-typable drastically decreased below 1% at our NRC (personal data NRC). In this study, the most important serotypes subjected to an official control, including *S.* Paratyphi B var. Java, *S.* Typhimurium and its monophasic variant (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8), were directly detected by the module MOL-PCR BASE in only one day. For laboratories who are routinely performing the serotyping by slide-agglutination, the developed test can be rapidly used as a replacement method for the exclusion of serotypes mentioned in the regulation, including the auto-agglutinable isolates. Concerning the remaining partially identified or unidentified isolates for which the identification is less mandatory, but still useful for national surveillance, they can be sent to the NRC for complete identification using WGS in batch.

In conclusion, the genoserotyping system developed by Gand et al. (2020b) demonstrated its ability to be a good alternative to the slide-agglutination method for the identification of auto-agglutinable isolates. As this method is adapted to be directly used by first-line laboratories, it reduces the TAT and objective results are rapidly transmitted to the professionals of the food sector. Consequently, it helps these latter to comply with the regulation as all the serotypes subjected to an official control can be detected, even if isolates are auto-agglutinable. As the MOL-PCR method is modular, the probe composition of the modules could easily be adapted if the trend of this non-typable isolates evolves towards serotypes not targeted by the method or to improve the detection of O:7 isolates showing a gap in the *rfb* cluster. For a better understanding of the mechanisms responsible of *Salmonella* auto-agglutinability, it could be interesting to further investigate the impact of this deletion on the antigen O structure and evaluate the effect of the food processing on its biosynthesis.

CHAPTER 7
General conclusions, discussion and perspectives

Identifying the serotype of *Salmonella* when isolated from the field is of major importance. First, because the host specificity, the clinical symptoms and the virulence of the pathogen can vary depending on its serotype, so this one must be rapidly determined to evaluate if it is life threatening and if special actions are required, i.e. medical treatment, quarantining, etc... Secondly, to reduce *Salmonella* transmission to humans, specific serotypes are subjected to an official control and must be excluded from the food chain (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8). Thirdly, the Belgian National Reference Center (NRC) performs national surveillance (including the yearly reporting of prevalent strains) and outbreak investigation by first identifying the serotype of *Salmonella* isolates before potentially making further characterization such as subtyping and antimicrobial resistance testing. As a reference center, the Belgian NRC for *Salmonella*, part of Sciensano, has the duty to master the reference techniques for *Salmonella* serotyping, i.e. slide-agglutination and biochemical tests. But the NRC has also as mission to investigate the most recent technological advances which can be more accurate and efficient for the rapid identification of *Salmonella* serotypes. This is why in 2015 the Belgian public federal service for public health, security of the food chain and environment charged the NRC with the development of an alternative molecular method for the genoserotyping of *Salmonella* and which was the starting point of this PhD research.

The detailed results of this PhD work were extensively discussed in the scientific manuscripts that compose the Chapters 3 to 6. The present chapter shows how the findings described in the different manuscripts/chapters helped to answer the research questions elaborated in Chapter 2.

7.1. The *Salmonella* serotypes to be targeted by the alternative molecular method

Although *Salmonella* can contaminate a large range of animal species (such as poultry, pork and cattle), with various health and economic impacts, only serotypes related to poultry and pork were considered in this work, as food products coming from these 2 animal species were considered as the main vehicles leading to salmonellosis in Europe (EFSA 2019b). After consultation at the beginning of this project with the first-line laboratories and the competent authorities involved in *Salmonella* detection in the Belgian food sectors, the serotypes and their variants to detect in priority were selected based on: (i) their demand by the European and Belgian regulation in the poultry sector (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8), i.e. Enteritidis, Hadar, Infantis, Paratyphi B var. Java, Typhimurium including its monophasic variant 1,4,[5],12:i:- and Virchow, (ii) their highly invasive character in pork and poultry species, i.e. Choleraesuis and Gallinarum including its variants Gallinarum and Pullorum, (iii) their prevalence in the poultry and pork sectors, i.e. Agona, Anatum, Brandenburg, Derby, Livingstone, Mbandaka, Minnesota, Ohio, Rissen and Senftenberg and (iv) the need to discriminate the wild-type and vaccine strains of *S.* Enteritidis when the mandatory vaccination campaign in poultry farms (breeding animals and laying hens) is too close to the control period, i.e. AviPro SALMONELLA VAC E and Salmovac SE. All these serotypes were covered by the 4 modules developed in this work. Even though the serotypes Hadar and Virchow were more common in the poultry sector in the 2000s, with a special concern about some *S.* Virchow isolates resistant to commonly used antibiotics (Bertrand et al. 2006), these serotypes, targeted by the module BASE, do not belong anymore to the top 5 of the most prevalent serotypes reported recently in Belgium (see Chapter 1 section 1.3.3). Therefore, their occurrence in the European and Belgian regulation can be questioned. Recently in 2019, following a request from the European Commission, the Scientific Panel on Biological Hazards (BIOHAZ) was asked to provide a scientific opinion on *Salmonella* control in poultry flocks and its public health impact (Koutsoumanis et al. 2019). In this report, the authors proposed to update the composition of the target serotypes used in the EU regulation N°2160/2003. Regarding the serotype prevalence described in chapter 1 (section 1.3), it is clear that *S.* Enteritidis, *S.* Infantis and *S.* Typhimurium including its monophasic variant 1,4,[5],12:i:- are key serotypes for which a strict and careful control is still needed. In the BIOHAZ report, *S.* Kentucky was proposed as the 4th target serotype as it has recently spread among broiler populations in several EU member states and because many strains are resistant to multiple antimicrobials, including the fluoroquinolones used in first intention in case of salmonellosis (Le Hello

et al. 2013). However, when Kentucky would be included in the EU regulation, the potential emergence in the EU of other Multi-Drug Resistant (MDR) strains, such as some *S.* Newport isolates reported in the USA, might also have to be taken into account (Crim et al. 2019; Iwamoto et al. 2017). Both serotypes, Kentucky and Newport, belong to the top 10 of prevalent *Salmonella* in Belgium and their spread must be monitored (NRC data). Considering the 5th serotype, *S.* Heidelberg or *S.* Thompson were mentioned in the report but a dynamic 5th serotype, specific to each member state, was suggested to be preferable as the serotype prevalence is not the same in all the EU. For instance, these 2 serotypes represented less than 1% of *Salmonella* isolated in Belgium these last years, while *S.* Paratyphi B, not mentioned in the BIOHAZ report, is the second most prevalent serotype coming from food and animals samples in this country (NRC data). Consequently, this serotype would be the ideal 5th 'dynamic' candidate for Belgium. This is why *S.* Paratyphi B is included in the Belgian regulation (Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8) and thus already targeted by the module BASE developed in this work. Finally, the authors of the BIOHAZ report proposed to use an "all serotypes" approach, with criteria based on the presence of the *Salmonella* species instead of specific target serotypes, to improve the control of *Salmonella* spp. in the food chain (Koutsoumanis et al. 2019).

In conclusion, the serotypes targeted by the genoserotyping assays developed during this PhD are perfectly adapted to the criteria of the regulation. Considering *S.* Paratyphi B as the 5th 'dynamic' serotype to combat, the targets of the method match also the recommendations of the BIOHAZ report, except for the detection of *S.* Kentucky, and potentially *S.* Newport, for which the spread must be monitored. Luckily, the Multiplex Oligonucleotide Ligation-PCR (MOL-PCR) & Luminex technique developed in the current work is modular and can be easily updated. The MOL-PCR BASE assay is already able to detect a part of the *S.* Newport population, thanks to the specific combination of some MultiLocus Sequence Typing (MLST) markers used for other serotypes targeted by the method. Using the huge possibilities offered by EnteroBase (see in section 7.3), demonstrated through this PhD research, other markers could be selected to detect all the *S.* Newport isolates. Identically, MLST markers could be selected for *S.* Kentucky as more than 75% of the isolates belonging to this serotype are clustered in only 3 STs in EnteroBase. Also, specific detection of the ST linked to the MDR *S.* Kentucky isolates could be considered. These new markers could be included directly in the module BASE. If an "all serotypes" detection approach is adopted, all the modules developed in the current work are already able to identify the isolates as *Salmonella* spp. through the detection of the *invA* molecular marker. Furthermore, the genoserotyping system developed in the scope of this work was strongly focused on serotypes coming from animal and food sources. But the NRC, which is now using the method in routine, deals mostly with isolates coming from humans including the highly dangerous Typhoid *Salmonella* (TS). Despite the fact that these cases are not really common in Belgium, representing less than 1% of the isolates analysed by the NRC (NRC data), they should still be detected

in priority because they are life threatening and subjected to a mandatory notification. Therefore, some validated markers published in the literature targeting the serogroup O:2 (Franklin et al. 2011), *S.* Paratyphi A (O:2) and *S.* Typhi (O:9) (Ranjbar et al. 2017; Tennant et al. 2015) are planned to be included in the module BASE as this module is recommended to be used for a first screening of all samples (Chapter 5).

7.2. Best alternative molecular technique for *Salmonella* genoserotyping

During this PhD work, different alternative molecular methods for *Salmonella* genoserotyping were used: PCR, qPCR, MOL-PCR & Luminex, Pulsed Field Gel Electrophoresis (PFGE), MLST and Whole Genome Sequencing (WGS). PCR and PFGE were helpful in 2007, for the identification of auto-agglutinable isolates (Chapter 6), when Next Generation Sequencing (NGS) was not as commonly implemented and used like today. But these 2 techniques are less suitable for routine diagnostic, because not user-friendly. Especially, with PFGE, the band pattern generated by this method is stored in large and complex files, not easy to interpret, making the method poorly reproducible. Consequently the data exchange between laboratories is difficult. The qPCR test developed in this study (Chapter 3) was efficient for the fast and easy specificity evaluation of the molecular marker selected for *S.* Paratyphi B dT-/dT+ detection, but the number of possible molecular targets remained very limited with this kind of method. Indeed, a quadruplex assay using allelic discrimination with 4 dyes was needed for the proper detection of the Paratyphi B and dT- SNP/WT markers, and this monopolized already 4 of the 5 fluorescence spectra usually available on most qPCR instruments. WGS was certainly the most accurate, discriminative and informative tool among the tested ones. Each time that there was a doubt in some identification or characterization, the whole genome sequencing of the isolate helped to resolve the problematic cases. When a MLST analysis was needed, WGS and *Salmonella In Silico* Typing Resource (SISTR), which includes MLST, were used instead of the required 7 PCRs followed by Sanger sequencing (Chapter 4 and 5). But WGS is still time-consuming (5 working days at Sciensano), labour-intensive, expensive and thus only adapted to big organisations which have a sequencing platform, such as at Sciensano.

Concerning the future of *Salmonella* genoserotyping, it is certain that one day, WGS will become the gold standard method when its time- and cost-effectiveness will be improved. But for now, this technique is not adapted to small structures and only routinely used by some big public health institutes in Canada, France, United Kingdom and U.S.A. (Allard 2016; Ashton et al. 2016; Institut Pasteur 2018; Jain, Mukhopadhyay, and Thomassin 2019). And even when this method is implemented in routine, laboratories are facing time and budget limitations, and must conserve the technical expertise as well as the sera collection required for serotyping by slide-agglutination. For instance at the French NRC, not all but 74% and 76% of the *Salmonella* isolates were serotyped by WGS only, respectively in 2017 and

2018, due to budget restrictions. The remaining samples were identified by classical methods (Institut Pasteur 2018, 2019). Additionally, in early stages of outbreaks, the serotype identification of implicated isolates must be determined as fast as possible for proper investigation, and the use of WGS in first intention is not adapted in this case. For instance, during the French outbreak concerning the contamination of infant milks by *S.* Agona in 2017, the French NRC had to step back for a time from routine WGS to serotyping by slide-agglutination, to quickly trace the origin of the outbreak and limit the spreading of the cases by identifying and recalling the infected food products (Personal communication). Moreover, a WGS protocol standardisation and an harmonization of the bioinformatics pipelines, including quality control guidelines, are required for the entire process from genomic DNA extraction to sequencing data interpretation, and this, for a proper communication and data exchange between the NRCs at an international level, especially during multistate outbreaks.

The Illumina technology, based on the sequencing of multiple short reads in high-throughput, is the method usually used for *Salmonella* serotyping based on WGS. But other sequencing solutions, faster (real-time) and without batching requirement, could be used for this purpose. For example, the Oxford Nanopore technology allows the sequencing of long reads in only few hours using small devices, such as MinION or Flongle, which are scalable for the analysis of 1 to 24 samples. The sequencing output is generated in real-time and the analysis can be stopped when enough data has been collected for result interpretation (Leggett and Clark 2017). However, until now, no automatic nor user-friendly pipeline, such as SISTR, is available for the analysis of these sequencing results. Consequently, bioinformatics skills and powerful servers with large storage capacity are required for data processing. Moreover, the technology is still constantly evolving and is mainly used for R&D purpose, but it has the potential to be used for routine genoserotyping in the future. It needs however to be properly investigated how the increased error rate (as compared to Illumina sequencing, although evolving and being optimized) affects the correct genoserotyping, or other downstream characterisation, as it would be less cost effective to use the obtained data for genoserotyping only (Leggett and Clark 2017). Another potentially promising approach is the use of shotgun metagenomics applied to pathogen diagnostic, directly on matrices without bacteria isolation. The principle of this technique is the sequencing of all DNA fragments present in a sample. Molecular markers specific to bacterial species, serotypes or variants could be looked up directly in the matrix without the time-consuming protocols required for bacterial isolation, including multiple cultures in enrichment broths and selective media. More informative, some virulence and invasive genes, such as *invA* or *spv* genes, could also be screened for to instantly evaluate the clinical importance of the positive samples (Miller et al. 2013; Oniciuc et al. 2018). But using also NGS techniques, such as Illumina, metagenomics has the same drawbacks as elaborated earlier for WGS. Additionally, the presence of the DNA markers in the samples does not attest of the viability of the pathogen(s). Moreover, as the virulence of pathogens is usually determined by a combination of specific genes, further complex analyses are needed to assess if the detected genes

are part of the same genome, i.e. same isolate, or not. The same holds true for the genoserotyping, as it is usually based on the detection of a combination of markers, like for the MOL-PCR & Luminex assays developed in this PhD research and requiring the isolation of the pathogen for a correct identification.

Retrospectively evaluated, based on the experiences using different alternative molecular methods during this PhD work, such as PCR, qPCR, MOL-PCR & Luminex, PFGE, MLST and WGS, the MOL-PCR & Luminex technology was and still remains the right option for developing a *Salmonella* genoserotyping test suitable for the current routine setting. Indeed, with this technique, 90 samples can be analysed by only one person in one working day. Thus, this technique is adapted to routine analysis but also to outbreak investigations for which a fast serotype identification is needed at early stage before further subtyping analyses. Used routinely since January 2019 at the Belgian NRC under accreditation, the test was evaluated to be cost-effective because it is up to 7.5 times less expensive than the classical methods. Furthermore, several MOL-PCR assays can be performed simultaneously on a MagPix device in multi-batch mode with a limit of 96 reactions. This means that the modules developed here, or other tests developed at the NRC and using the same technology such as *Shigella* genoserotyping (Ventola et al. 2019), *S.* Typhimurium subtyping (Wuyts, Mattheus, et al. 2015) or anti-microbial resistance determination (Ceyssens et al. 2016), can be combined in one run, thereby saving time and money. According to the NRC surveillance data of the past years, more than 77% of the *Salmonella* isolates sent yearly could have been completely identified using only the MOL-PCR BASE, and even more could have been partially identified through serogroup determination and probable serotype prediction. Concerning the 3 other modules, they could have identified more than 36 % and 50% of the remaining isolates coming from food and veterinary sources, respectively

7.3. Selection of the molecular markers and evaluation of their specificity

When developing a target-based molecular method such as MOL-PCR & Luminex, the first step is the selection of molecular markers specific to the targeted *Salmonella* serotypes. The main challenge of this process is to determine genomic signatures, which can be used for DNA-based detection, specific to bacterial clusters such as serotypes, that were initially defined by phenotypic properties, i.e. antigenic and biochemical characteristics. In other words, it is a question of identifying phenotypic clusters with molecular techniques, which are both different approaches. Consequently, a rigorous specificity evaluation of the marker candidates is required to avoid, as much as possible, false positives and negatives. Considering the complex structure of the *Salmonella* genus (Chapter 1), this means that the selected genetic markers must ideally be present in all the strains belonging to the targeted serotype and absent in the strains belonging to other serotypes. There are around 1 500 serotypes of *Salmonella enterica* subsp. *enterica*, the subspecies containing the most frequently isolated strains. The time and money needed to test the specificity of the selected marker for all these serotypes in the wet lab would

be enormous and is thus not feasible. Moreover, only few laboratories in the world hold a collection of *Salmonella* isolates representing all the 1 500 serotypes of the subspecies *enterica*. Fortunately, the use of bioinformatics tools in the present work allowed the *in silico* evaluation of this specificity. The MLST scheme was exploited in an innovative way for the selection of most of the molecular markers included in the development of the MOL-PCR assays. This solution was retained because in 2012, Achtman and his team established that a reliable link exists between the MLST clusters, i.e. Sequence Type (ST) and eBurst Group (eBG), and most of the serotypes composing the White-Kauffmann-Le Minor (WKL) scheme (Achtman et al. 2012). For all the isolates analysed with the MLST technique, the typing results are linked to serotype identifications determined by classical methods, in a database named EnteroBase. Consequently, this database was screened to select alleles of the 7 genes, composing the MLST scheme, that are conserved among isolates of the same serotype. Then, multiple alignments of all the alleles were performed to identify SNPs specific to the selected alleles, i.e. specific to the targeted serotypes. Thanks to the considerable number of *Salmonella* entries composing EnteroBase, i.e. serotype determined by classical methods and linked to MLST typing for each isolates, a strong and rigorous *in silico* specificity check of the marker candidates was performed against thousands of *Salmonella* isolates. EnteroBase demonstrated to be a powerful tool, cost and time-efficient, for markers' selection and *in silico* specificity evaluation. For 12 serotypes, specific markers could be identified using this approach (Chapters 4 and 5).

But unfortunately, no SNP(s) specific to the serotype Paratyphi B (dT-/dT+) could be retrieved using this strategy. Even the user-friendly software Gegenees, made for the determination of specific genomic signatures from WGS data sorted in target and background groups, and successfully used for *S.* Livingstone and *S.* Gallinarum markers' selection (Chapter 5), failed to retrieve good candidates for *S.* Paratyphi B detection (Chapter 3). This is not surprising regarding the genomic variety inside the Paratyphi B population clustered in a lot of disparate STs and eBGs by the MLST technique (Achtman et al. 2012). A genomic study using 191 *S.* Paratyphi B genomes demonstrated that this population could actually be divided into 10 distinct PGs (Connor et al. 2016) making the task to retrieve a specific genetic marker conserved in all these PGs difficult, as illustrated with the publicly available marker of Zhai et al. (2014) designed for Paratyphi B detection but which failed to detect isolates belonging to PG10 (Chapter 3). Consequently, another approach was followed by comparing *Salmonella* genomes, belonging to the 10 Paratyphi B PGs and 44 other serotypes, to find a specific SNP. An in-house new script had to be written to filter the 3 million SNPs retrieved this way and, surprisingly, only one mutation seemed to be conserved among the 10 PGs and absent in the 44 other serotypes. Nevertheless, even for this SNP, some rare false positives were later obtained during the MOL-PCR BASE development, if not combined with other markers (Chapter 5). The number of 44 genomes, belonging to other serotypes in the SNP comparison, was probably not enough to retrieve a strong specific marker. At the time that this genomic study was done, not so many complete and assembled genomes belonging

to various *Salmonella* serotypes, other than the most prevalent ones, were publicly available. To include a consequent variety of different serotypes in the genomic study, the time- and resource-demanding individual download of contigs or Sequence Reads Archive (SRA) was needed, requiring in-house performed genome assemblies for the latter. Nowadays, EnteroBase offers the public access to complete genomes, assembled from the sequenced reads uploaded in the database, that can be easily downloaded in batch. Therefore, more genomes could now be included in the genomic study for the *S.* Paratyphi B marker selection. But as only one specific SNP was retrieved from this analysis, if more genomes from more serotypes would have been used, probably no marker would have been obtained. As the parameters of this genomic comparison are too stringent, a better solution would now be to do the same kind of SNP comparison but by splitting the Paratyphi B genomes belonging to the 10 PGs into 2 groups or more, with the aim to select at least 1 marker specific to each group.

The difficulties encountered with the Paratyphi B marker selection illustrate perfectly the challenge of keeping a link with the historical and widely used classification system of *Salmonella*, defined by phenotypic properties, when developing a molecular method based on DNA detection for genoserotyping. In the future, the complex nomenclature of *Salmonella* should evolve by taking into account the new data provided by the molecular and genomic studies. Indeed, the new MLST based clustering tools (such as ST and eBG), using WGS data, proved their efficiency for *Salmonella* serotype identification and further subtyping analyses (Achtman et al. 2012; Alikhan et al. 2018; Yoshida, Kruczkiewicz, et al. 2016). But as discussed in section 7.2, WGS is still not ready to be implemented in all laboratories worldwide. Therefore, the WKL nomenclature should be conserved before being slowly replaced by new molecular classification systems. Moreover, a correlation between classical serotyping and genoserotyping must be maintained, like this is the case in the molecular assays developed in this PhD research, to keep a link with years and years of epidemiological data based on the WKL scheme and to be able to exchange information between countries (e.g. in case of multi-state outbreaks), where WGS is not yet implemented.

7.4. Production of accurate identification results from Luminex data

A plethora of new molecular technologies emerged these last decades and permitted the generation of a substantial amount of experimental data in one high-throughput analysis. But user-friendly tools allowing the fast and easy interpretation of these experimental data, ideally executed by the laboratory technicians themselves, are not always provided with these methods. This is clearly true for the WGS technologies, which required the development of pipelines such as SISTR, but also for the target-based molecular methods. For example, with the commercial kit for *Salmonella* detection provided by Luminex and using 3 multiplex assays (xMAP® *Salmonella* Serotyping Assay), no software is included for the interpretation of the data. This means that the user has to perform him/herself the calculation of

the raw data generated by the different assays to deduce the serotype identifications based on the combination of the detected molecular markers. In the present work, during the development of the MOL-PCR assays, this data processing was first performed using an automated Excel workbook (Chapter 4). The results produced by this Excel file were validated by comparison with the identifications obtained using the classical methods, i.e. the slide-agglutination and the biochemical tests. But this solution is not user-friendly because not modular. Indeed, the Excel workbook was specifically and gradually designed during the experimental work for the method validation and is not easily adaptable. Therefore, it is only applicable in an R&D context, and by people that know how to work properly in Excel. To improve this, a Decision Support System (DSS) was developed to be used with the 4 MOL-PCR assays developed in this PhD research (Annex 1). To be easily accessible on-line by the first-line laboratories, this DSS is hosted by a web-application which requires only a simple browser and no specific installation. For an automatic interpretation of the data, the Gödel Prime Product (GPP) already successfully used for GMO detection (Van Den Bulcke et al. 2010) was ingeniously applied to convert the fluorescence data generated by the MagPix into a unique barcode which represents the combination of molecular markers specific to the *Salmonella* serotype. Another strength of the DSS is that it is able to provide partial identification results and recommendations to the users even if the retrieved barcode is not fully known by the system. This means that this tool is not only limited to displaying the serotype names or an "unknown serotype" message, but it also takes advantage of all that can be exploited from the Luminex raw data, to give interpretations as complete as possible to the users, without loss of information (e.g. the serogrouping). Moreover, the web-application performs extensive quality control checks and detection of analysis issues, and it offers user traceability, all of which are usually required when the method is planned to be used under accreditation following ISO standards such as ISO 15189 and ISO 17025. As the DSS was created to be used in routine by laboratories, it was developed using the DTAP (Development, Testing, Acceptance and Production) principle and subsequently rigorously validated by simulating all the possible scenarios and experimental issues. During this validation phase, all the interpretation results generated by the DSS were verified by comparison with those obtained using the automated Excel workbook.

The DSS was designed to be modular, to follow the update of the MOL-PCR modules, and it can be easily configured directly by the users for each MOL-PCR assay without the intervention of an IT developer. This also means that even if the web-application was developed in this work for the genoserotyping of *Salmonella*, it can actually be used for the automatic interpretation of any Luminex results based on DNA detection and generated by a MagPix device. Consequently, the DSS will be configured to be used with other Luminex methods developed at the NRC (Ceyssens et al. 2016; Ventola et al. 2019; Wuyts, Mattheus, et al. 2015) and for further applications.

Finally, all the serotyping results analysed by the DSS are stored in a database accessible by the NRC. Thanks to this, when the genoserotyping system will be implemented in the Belgian laboratories,

the serotyping data analysed with it will be automatically transmitted to the NRC. Consequently, it will drastically improve the collection of the data and the surveillance of *Salmonella* serotypes at a national level.

7.5. Validation of the MOL-PCR genoserotyping assays by comparison with the classical methods

All the selected molecular markers were converted into upstream and downstream ligation probes and, from this, 4 MOL-PCR assays were designed: MOL-PCR BASE, MOL-PCR O3-4-21, MOL-PCR O7 and MOL-PCR O9. The MOL-PCR BASE is able to identify the serotypes and their variants subjected to an official control (EU regulation N°2160/2003, Belgian royal decree 27/04/2007 and Belgian FASFC note BP-MN-FDS/LABO/1557457 v8), and to cluster the other samples in one of the following serogroups, O:3, O:4, O:7, O:8, O:9, and O:21, if they belong to one of them (Chapter 4). The other targeted serotypes and their variants are divided amongst the 3 other multiplex assays based on their serogroups, which are also targeted in these modules (Chapter 5). As the developed *Salmonella* genoserotyping system was intended to be used for routine analyses under accreditation, a strong validation of the method was required. Consequently, the 4 MOL-PCR assays were validated with the analysis of at least 25 isolates per targeted serotype, except for the invasive *S.* Choleraesuis (17) and *S.* Gallinarum (12) which are less common in Belgium, at least 130 other *Salmonella* isolates (among which 13 other serogroups than those targeted were represented) and 33 non-*Salmonella* isolates, following the guidelines of the ISO 16140-6. The generated identification results were compared with those obtained with the classical methods, resulting in an accuracy above 99% for the 4 modules (Chapters 4 and 5). The missing 1% concerned only 3 isolates of *S.* Virchow (O:7), determined by WGS as belonging to rare STs of this serotype, that were only identified as "*Salmonella* O:7" by the MOL-PCR BASE, leading thus to false negative results (Chapter 4). However, the significance of these isolates can be questioned. They were determined as "*S.* Virchow" based on their antigenic characteristics but as they seem to be genetically different, maybe their virulence or host specificity is also not the same. This is why a shift from phenotypic to molecular detection methods is preferable for the identification of *Salmonella*, because it can be more informative and meaningful. These rare Virchow STs were so uncommon that it was decided not to adapt the method for their identification. Luckily, on the 1 004 bacterial isolates tested for the validation of the module BASE, these kind of results were scarce. It is not so common to test so many isolates for the validation of an alternative method. When consulting the scientific literature reviewed for this PhD work, on 21 scientific articles dealing with development, validation or evaluation of molecular methods (similar to those reviewed in section 7.2) for the identification of several *Salmonella* serotypes, the average number of used isolates was 350. In only one multi-laboratory study evaluating a microarray based genoserotyping assay, more than 1 000 isolates were tested (Yoshida et al. 2014). However, we can imagine that the more *Salmonella* isolates are tested,

the more rare STs can be obtained, leading potentially to incorrect results. This is due to the fact that the molecular markers selected for DNA-based alternative methods, such as the MOL-PCR assays developed in this PhD research and that target some serotypes specifically, are not always related to genes coding for the somatic and flagellar antigens. So if a mutation or a deletion happens in a molecular marker not linked to the O and H1/H2 genes, this will not automatically trigger a change in the antigenic formula of the bacterial strain and hence not a change in the serotype defined by slide-agglutination. Nevertheless, the mutation in the molecular marker can cause a failure in its molecular detection, thereby making the serotype not able to be defined by the genoserotyping method. For example, a *S.* Typhimurium isolate defined by slide-agglutination will not be detected anymore as such by the MOL-PCR method because of a mutation in a molecular marker (specific for *S.* Typhimurium) unrelated with O and H1/H2 genes. Identically, if a mutation occurs in the flagellar or somatic coding sequence, this will potentially give another antigenic formula and serotype name with the slide-agglutination technique, while the molecular marker(s) linked to the serotype will stay unchanged. For example, a *S.* Typhimurium isolate will potentially become another serotype but will still be detected as "S. Typhimurium" by the MOL-PCR assay. Fortunately, of the considerable amount of *Salmonella* isolates tested in the present validation study, one part came from the Belgian NRC collection, and the other part originated from the routine samples, when the technique was used at the NRC during the 3 last months of 2018 in parallel with the classical methods. Consequently, the specificity of the developed tests was successfully validated with *Salmonella* isolates frequently encountered in Belgium, representative of the strains circulating in the country. But we can imagine that if the method is used in another country, or continent, maybe the accuracy would be slightly different. Finally, it was also interesting to notice that some false negative or false positive cases were resolved by repeating the slide-agglutination or biochemical tests, by different laboratory technicians in blind, and this actually led to the confirmation of the results obtained by the new molecular method. This demonstrated again the limitations and subjectivity of the classical methods. The developed MOL-PCR assays were, thus, more accurate than the reference methods for *Salmonella* serotype identification.

The substantial amount of data generated during this validation process demonstrated the specificity, robustness and efficiency of the method, and allowed its accreditation at the NRC following ISO 15189 and ISO 17025. Considering this successful switch from phenotypic to molecular methods at the NRC for most of the *Salmonella* serotyping, this should incite the first-line laboratories to implement the developed MOL-PCR assays in their laboratory procedures.

7.6. Evaluation of the developed method for variant determination and auto-agglutinable isolates identification

When using the classical methods for *Salmonella* serotype and variant identification, it is sometimes difficult to obtain a clear and reliable result. For example, for the determination of some

serotype variants, complex and time-consuming biochemical tests are needed and their interpretation is highly subjective, as it was illustrated in this PhD research with the use of the lead-acetate and the Jordan's tartrate tests for the d-tartrate fermenting ability determination of *S.* Paratyphi B isolates (Chapter 3). Indeed, both tests are culture-based and their interpretation was not easy. Moreover, repetition of these tests led sometimes to contradictory results. Concerning the discrimination between *S.* Gallinarum var. Gallinarum and *S.* Gallinarum var. Pullorum, several culture-based tests are available but they are based on biochemical properties which are variable or which can show exceptions between isolates. The different variants of *S.* Gallinarum and *S.* Paratyphi B are responsible of different clinical symptoms and have different host specificities (Chapter 1). Additionally, *S.* Paratyphi B var. Java (dT+) belongs to the serotypes subjected to an official control in Belgium. Consequently, this lack of accuracy in *Salmonella* serotype and variant identification in pathogen diagnostic is not acceptable. However, the MOL-PCR Base and MOL-PCR O9 (Chapters 4 and 5) were perfectly able to replace these time-consuming and/or poorly reliable biochemical tests for *S.* Paratyphi B and *S.* Gallinarum variant determination, respectively, with the production of fast and accurate results.

Also, when performing the serotyping test by slide-agglutination, some isolates can show a reaction of auto-agglutination, and thus, their antigenic formula cannot be identified. Again, it cannot be determined if the isolates belong to serotypes subjected to an official control or to invasive serotypes with clinical importance. To obtain the complete identification of the auto-agglutinable isolates, alternative molecular methods must be used. The genoserotyping system developed in this PhD research was evaluated by comparison with other tailored molecular techniques for the detection of these non-typable isolates, and it successfully identified them when technically feasible (Chapter 6). Consequently, the MOL-PCR assays were used to retrospectively identify auto-agglutinable isolates originating from the routine analyses (Chapter 6). It appeared that a great part of these isolates belonged to the serotype Paratyphi B which is frequently found in processed food made from poultry or eggs products (Personal communication NRC; EFSA, 2018). Therefore, the hypothesis could be made that the processing of these food products may alter the surface antigens of this serotype, thereby making them auto-agglutinable. Also, the antigens of the serotype Paratyphi B are maybe more easily altered during food processing than for other serotypes. A potential link with the *Salmonella* rough isolates showing an altered antigen O structure and auto-agglutinability in saline reaction can be made, but not a lot of scientific studies deal with these questions and further investigations are needed to understand what makes a *Salmonella* to become auto-agglutinable. However, with the MOL-PCR assays developed in this PhD study, these auto-agglutinable isolates can still be identified, and hence action can be taken, if needed, based on the regulation.

7.7. General conclusion: What is the best method for *Salmonella* genoserotyping in a routine setting?

Target-based methods such as the genoserotyping system developed in this PhD work and using MOL-PCR & Luminex technology, combined with a DSS for automatic results' interpretation, even though less exhaustive than WGS, have still their utility as they are rapid, accurate and cost-effective. Indeed, at the French NRC, 26% of the *Salmonella* isolates were not genoserotyped by WGS in 2017 because of budget restriction, although this NRC is to be considered to be amongst the more "wealthier" ones. Additionally, the serotyping by slide-agglutination has to be used in case of urgent identification like during outbreaks, because WGS analyses take between 15 to 22 days to be completed (at the sequencing platform of the Pasteur Institute), versus 2 to 7 days with classical methods (Institut Pasteur 2019). However, the target-based molecular methods, such as the MOL-PCR & Luminex assays developed in this PhD research, have the potential to overcome these time and cost limitations. Additionally, the target-based and sequencing methods can be used in complement of each other.

Indeed, the following ideal workflow, optimizing time and cost constraints, is proposed for routine serotyping of *Salmonella* (**Figure 1**). The MOL-PCR BASE can be used for a first screening of every new *Salmonella* isolate to identify. Like this, the serotypes subjected to an official control, as well as the most prevalent serotypes, are rapidly identified in one day and at low cost. If the analysed isolates are only partially identified by this first module, the DSS developed in this PhD research recommends which of the 3 other MOL-PCR assays must be used to complete their identification, when possible, requiring an additional day of analysis. Concerning the remaining isolates which cannot be identified by any of the 4 modules, they can be sent to a sequencing platform for complete identification using WGS.

With the proposed workflow, the most frequent (more than 77% according to the Belgian NRC data) and isolates subjected to an official control are rapidly identified in 1 to 2 days by the MOL-PCR & Luminex method at a low price. WGS, more expensive and time-consuming, is only required for a small part (less than 23%) which are still to be identified for national surveillance. If there is an abnormal rise of a given serotype, among the isolates completely identified by Luminex, and thus a suspicion of outbreak, the isolates belonging to the suspected serotype can retrospectively be analysed by WGS for subsequent subtyping analyses and outbreak investigation. This has the advantage to use the huge amount of data produced by the time-consuming and expensive WGS analysis, only when it is effectively needed (e.g., in case of an abnormal rise of a given serotype), and to not sequence the full genome of thousands of isolates belonging to common and identical serotypes, without the full exploitation of the valuable sequencing data. For the isolates belonging to redundant serotypes (e.g. *S.* Typhimurium and *S.* Enteritidis), only a representative selection of them could be further characterized by WGS when needed.

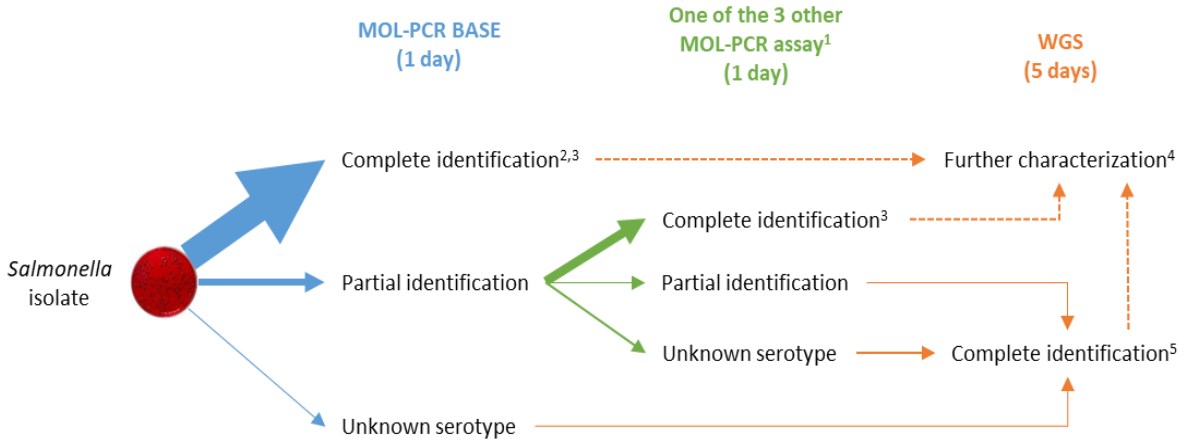

Figure 1: Recommended workflow for routine *Salmonella* genoserotyping.

MOL-PCR: Multiplex Oligonucleotide Ligation-PCR;
WGS: Whole Genome Sequencing.
[1]: MOL-PCR O3-4-21, MOL-PCR O7 or MOL-PCR O9.
[2]: including identification of the serotypes subjected to an official control.
[3]: including identification of highly invasive serotypes.
[4]: for a selection of common and clonal isolates (e.g. *S.* Typhimurium and *S.* Enteritidis) and for other isolates in case of an abnormal rise of a given serotype.
[5]: using *Salmonella In Silico* Typing Resource (SISTR).
The size of the arrows is representative of the number of samples which are proceeded this way. The dotted arrows represent the use of WGS for further characterization, including subtyping, when needed.

Unfortunately, in Belgium, the situation is more complex, with a clear distinction between the food and human sector. Currently, the budget allocated by the competent authorities to *Salmonella* serotype identification is not sufficient for WGS analyses such as elaborated in the 'ideal' workflow above. Therefore, the isolates not identified by one of the 4 modules developed in this PhD work would still need to be serotyped by the classical methods. Moreover, the main objective of the FASFC is the rapid exclusion of the serotypes subjected to an official control along the food chain. The identification of the other serotypes is for FASFC less important, and hence no budget is allocated to this, and certainly not to perform WGS. As anticipated in this project, these serotypes targeted by the legislation, as well as the most invasive and prevalent serotypes coming from poultry and pork, can be easily detected by the new method, which is fast and cost-effective. Therefore, this genoserotyping system has the potential to help the professionals of the food sector to comply with the regulation and improve the surveillance of circulating strains at a national level. Indeed, the cost savings brought along by the use of the MOL-PCR & Luminex method for routine analyses at first line can allow the increase of the sampling points along the food chain and therefore imply an improvement of the *Salmonella* control program (i.e. more controls with the same budget). In addition to be faster than the classical serotyping, the new method is also more accurate and user-friendly and can be directly implemented in the first-line laboratories, avoiding the sending of the samples to the NRC to obtain the serotyping. This will reduce the results'

reporting time. However, the MOL-PCR & Luminex technique is now used only at the Belgium NRC and for samples coming from human sources. Indeed, for the veterinary and food samples, the FASFC has to update the regulation to allow the alternative molecular techniques to be used as replacement methods for the serotyping of *Salmonella*. The data produced in this PhD work showed all the advantages provided by the molecular methods and this helped to convince the competent authorities to start procedures for law adaptations. Once authorized by the FASFC, the method could easily be implemented in the first-line laboratories such as DGZ and Arsia in Belgium. If this is done, all the genoserotyping results obtained this way will be automatically collected by the NRC, for national surveillance, thanks to the database included in the DSS developed with the method. However, the *Salmonella* isolates, not identified by one of the 4 new MOL-PCR assays developed in this work, still have to be sent to the NRC if complete identification and potential further characterisation are needed. In the future, it should be evaluated whether the proposed ideal workflow for *Salmonella* genoserotyping, including the WGS, would not be possible for the Belgian NRC (for humans as well as for food and veterinary samples once the law will be adapted) because of the reduced cost for analysis by MOL-PCR and Luminex (and hence budget savings to be used for WGS), compared to slide-agglutination. For the human isolates, in contrast to food isolates, a full identification remains required for surveillance purposes.

Irrespective of the future of slide-agglutination, there is a consensus in the scientific community stating that the WKL scheme must be updated. Indeed, supported by the main findings of this PhD work, including the complex classification system of *Salmonella*, the genomic heterogeneity inside some serotype populations such as Paratyphi B and the benefits of the molecular techniques compared to the classical methods, the future of the WKL scheme can be put into question. However, the clustering of *Salmonella* isolates into serotypes, according to their antigenic formula, was implemented 80 years ago, before the molecular era, the use of NGS and the knowledge acquired through genomic studies. Consequently, it is difficult to update a classification method such as the serotyping based on the WKL scheme, that is extensively used since so many years in public health laboratories worldwide and mentioned as well in the regulations. The nomenclature of *Salmonella* has certainly to change, taking into account the new clustering tools such as STs and eBGs based on WGS, but at the same time the link with the historical serotype classification should be kept, to not loose decades of epidemiological data. Moreover, although WGS might be proposed by the scientific community as the most accurate *Salmonella* serotyping method, as elaborated above, the sequencing techniques needed for molecular identification and typing, are still not adapted to the time- and cost-effectiveness required for current surveillance programs and rapid outbreak traceability, and only few laboratories in the world can afford this technology on a routine basis. This is especially an issue for pathogen surveillance and outbreak investigation which has to be seen in an international context. If WGS is adopted as the gold standard method for serotype identification in the official control programs, there is a risk of economic

discrimination between countries which can or cannot follow this evolution. Even inside the European Union, such evolution would negatively impact the quality of the surveillance system and limit data exchanges between states using old and new technologies. Consequently, molecular techniques must coexist and be complementary to the phenotypical methods before slowly replacing them when it will be feasible. This represents a challenging paradox between the ideal/scientific and real/in-the-field world of *Salmonella* surveillance. Indeed, NRCs still have to keep their expertise in classical serotyping, which includes to maintain their expensive collection of sera up-to date. Sadly, with the advance of molecular methods in replacement of the slide-agglutination for serotyping, these sera will be increasingly hard to get from the suppliers because less asked for and hence less produced. Furthermore, to some people, the future of *Salmonella* identification might be to analyse the pathotype (e.g. virulence genes) of samples directly on raw matrices, instead of performing the genotyping of isolates, with the aim to improve the clinical relevance of pathogen diagnostics. However, the technologies for these kinds of metagenomics analyses are still at the R&D level. Moreover, the isolation and collection of *Salmonella* strains for national surveillance currently remains one of the main missions of the NRLs and NRCs.

Therefore, it is clear that before a major evolution such as the use of sequencing technologies as reference methods in routine laboratories will take place, the genoserotyping assays developed in this PhD work are the ideal methods to use for *Salmonella* serotyping in a routine setting as they make the perfect link between the historical classification and the new possibilities brought along by the molecular technologies, and this with taking into account the time and budget limitations forced upon the NRCs and first-line laboratories.

ANNEX 1
Development of the Decision Support System (DSS)

Once a Luminex analysis is done, the Single to Noise Ratios (SNRs), and Allele Calls (ACs) when needed, have to be calculated from the generated data, for each molecular marker and sample (see Chapter 4). From this, the user can determine if a combination of molecular markers specific to a serotype was obtained. Doing this result interpretation manually for each run can be complex, laborious and time-consuming for the users. Therefore, a Decision Support System (DSS, accessible as a web-application) was developed during this PhD research. The role of the DSS is to automatically process and convert the raw fluorescence data produced by the MagPix, i.e. the MFI stored in a Comma Separated Value (CSV) file, into identification results (including an extended quality control) through the use of a barcode system: the Gödel Prime Product (GPP) (see Chapter 5). Another role of the DSS is the central storage of all the genoserotyping results produced with it, in a national database accessible by the National Reference Center (NRC) for national surveillance.

In addition to MFIs, the CSV file contains also general information about the run, the configuration of the MagPix and the MOL-PCR assay which was used. Based on this information, including the name of the module (i.e. MOL-PCR BASE, MOL-PCR O3-4-21, MOL-PCR O7 or MOL-PCR O9) and its version, the web-application retrieves automatically in its system the data required for the interpretation of the run (prime numbers, cut-off values, threshold, GPP list, etc…), that were initially configured by the user (to do once for each MOL-PCR assay). The DSS uses also this general information to check if the assay was correctly performed according to the protocol settings and if an automatic interpretation is possible without generating incorrect results. Additionally, from the negative and positive control wells included in the analysis plate, quality controls (QCs) are performed to check that all the probes are functional.

The DSS converts the MFIs into serotype identifications as described in chapter 5 (section 5.2.5) and presented in the general workflow below (**Figure 1**, 8 steps involved**)**. Additionally, during this process, the web-app can monitor if a detection problem occurred during the run, leading for a specific marker and sample, to obtain the value “NaN” (Not a Number) instead of the MFI value, because this one could not be measured. If this situation happens, to not block the complete interpretation of the sample, the DSS considers the molecular markers for which a detection problem occurred, as negative, and it displays a corresponding error message to the user informing what serotype cannot be detected. For example, if a detection problem occurred with the probe STID15, the serotype Virchow cannot be detected. But if the sample concerned by this detection error belongs to the serotype Hadar, the DSS will still be able to identify it as STID15 is not included in the probe combination of Hadar. However, the user will be informed that an issue was observed with STID15 and he/she will need to decide if the analysis needs to be redone.

The DSS was developed in partnership with the IT department of Sciensano and deployed according to the DTAP principle, i.e. following the 4 phases of Development, Testing, Acceptance and Production. This software is hosted by a web-application accessible at the following address: https://salmstid.wiv-isp.be. The DSS was validated with all the possible scenario simulating valid and no valid analyses. Correct and incorrect configurations of the MagPix were extensively tested to check the behavior of the DSS with all the possible cases. All the results were compared with a manual processing of the data using Excel sheets (Chapter 4 and 5).

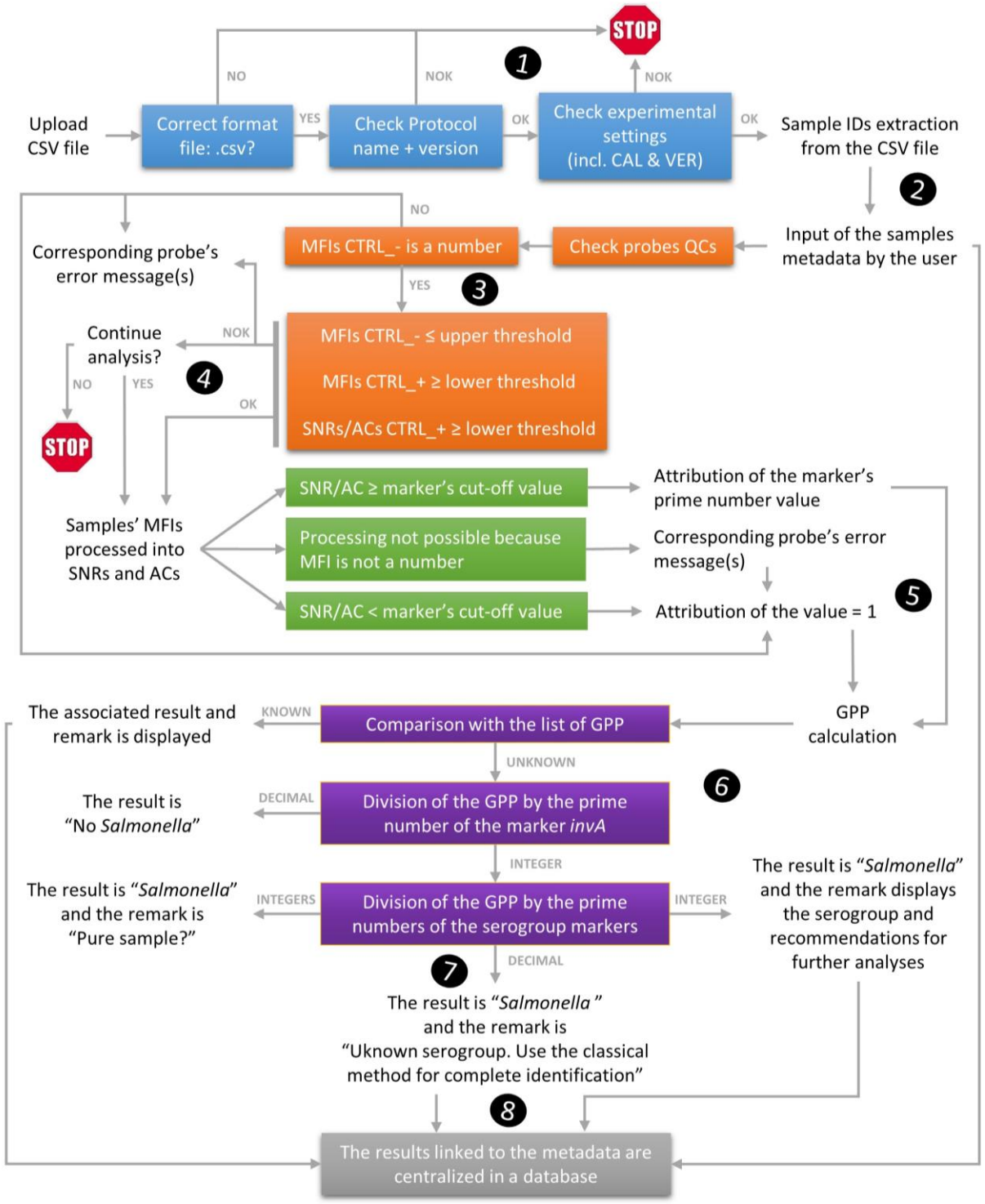

Figure1: General workflow of the DSS

The main steps of the workflow presented in the **Figure 1** are:

❶ Based on the name of the protocol mentioned in the CSV file and its version, the DSS retrieves the necessary information needed for the proper interpretation of the results, like for instance the probes, the prime numbers, the Gödel Prime Product (GPP) list, the cut-off values, the threshold values, etc…Also, based on the data present in the header of the CSV file, the system checks if the MagPix was properly configured, calibrated and verified before the analysis (like recommended by the manufacturer). If not, the analysis stops.

❷ The DSS retrieves the sample IDs from the CSV file. They were previously typed when the assay was configured in the MagPix. At this step, the user can add metadata linked to the nature and history of the samples (date of isolation, origin, clinical symptoms associated, etc…)

❸ Based on negative control (CTRL_-) and positive controls (CTRL_+) the DSS performs a Quality Control (QC) of the probes using raw data, i.e. Median Fluorescence Intensity (MFI), and processed data calculated automatically for the controls, i.e. the Single to Noise Ratio (SNR) and Allele Call (AC). These data are compared to threshold values.

❹ If the QC of a given probe is not ok, an error message linked to the probe is displayed to the user with information about the impact on the assay, e.g. ”Detection Typhimurium not possible”. Based on these messages, the user chooses to continue or stop the analysis.

❺ The MFI of the samples are automatically processed into SNR and AC. If the processed data are above the cut-off values specific to each molecular marker, the prime number of this one is included in the GPP, if not, the value “1” is included instead. If the MFI CTRL_- or any marker’s MFI of a sample is not a number because of detection issues, the value “1” is used in the GPP for the corresponding marker and an error message is displayed to the user like described above.

❻ The calculated GPP is compared to the list of expected GPP configured in the system. If it is known, the linked result and remark are displayed. If not, the presence of the marker *invA* (specific to the *Salmonella* genus) is determined by dividing the GPP by the prime number of the *invA* marker: if an integer number is obtained, the marker is present; if a decimal number is obtained, the marker is absent.

❼ If the sample is a *Salmonella* (*invA* present) but with an unknown GPP, the DSS tries to cluster it among the targeted serogroups (O:3, O:4, O:7, O:8, O:9, O:21) by successively dividing the GPP by each prime number of the serogroup markers. If among the results of these operations, only one integer is obtained, the serogroup linked to the prime number and thus the marker which produced this result, is determined. If more than one integer is obtained, the sample is suspected as “not pure” because the targeted serogroups cannot coexist together in a *Salmonella* isolate. If only decimal numbers are obtained, the serogroup of the sample does not belong to one of those targeted by the method and the remark “Unknown serogroup” is displayed.

❽ The final and partial results are linked to their metadata and accessible in the database. If a new analysis is made with another MOL-PCR assay, the system links the 2 results in the database, as long as the same sample ID is used.

REFERENCES

Achtman, Mark, John Wain, François-Xavier Weill, Satheesh Nair, Zhemin Zhou, Vartual Sangal, Mary G. Krauland, et al. 2012. "Multilocus Sequence Typing as a Replacement for Serotyping in Salmonella Enterica." *PLoS Pathogens* 8 (6). https://doi.org/10.1371/journal.ppat.1002776.

Ågren, Joakim, Anders Sundström, Therese Håfström, and Bo Segerman. 2012. "Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups." *PLoS ONE* 7 (6). https://doi.org/10.1371/journal.pone.0039107.

Alfredsson, G. A., Ruth M. Barker, D. C. Old, and J. P. Duguid. 1972. "Use of Tartaric Acid Isomers and Citric Acid in the Biotyping of Salmonella Typhimurium." *Journal of Hygiene* 70 (4): 651–66. https://doi.org/10.1017/S0022172400022518.

Alikhan, Nabil Fareed, Zhemin Zhou, Martin J. Sergeant, and Mark Achtman. 2018. "A Genomic Overview of the Population Structure of Salmonella." *PLoS Genetics* 14 (4): 1–13. https://doi.org/10.1371/journal.pgen.1007261.

Allard, Marc W. 2016. "The Future of Whole-Genome Sequencing for Public Health and the Clinic." *Journal of Clinical Microbiology* 54 (8): 1946–48. https://doi.org/10.1128/JCM.01082-16.

Alvarez, Juan, Mertxe Sota, Ana Belén Vivanco, Ildefonso Perales, Ramón Cisterna, Aitor Rementeria, and Javier Garaizar. 2004. "Development of a Multiplex PCR Technique for Detection and Epidemiological Typing of Salmonella in Human Clinical Samples." *Journal of Clinical Microbiology* 42 (4): 1734–38. https://doi.org/10.1128/JCM.42.4.1734-1738.2004.

Alves Batista, Diego Felipe, Oliveiro Caetano de Freitas Neto, Adriana Maria de Almeida, Grazieli Maboni, Tatiane Furtado de Carvalho, Thaynara Parente de Carvalho, Paul Andrew Barrow, and Angelo Berchieri. 2018. "Evaluation of Pathogenicity of Salmonella Gallinarum Strains Harbouring Deletions in Genes Whose Orthologues Are Conserved Pseudogenes in S. Pullorum." *PLoS ONE* 13 (7): 1–18. https://doi.org/10.1371/journal.pone.0200585.

Andino, A., and I. Hanning. 2015. "Salmonella Enterica: Survival, Colonization, and Virulence Differences among Serovars." *Scientific World Journal* 2015. https://doi.org/10.1155/2015/520179.

Angelo, Kristina M., Alvina Chu, Madhu Anand, Thai An Nguyen, Lyndsay Bottichio, Matthew Wise, Ian Williams, et al. 2015. "Outbreak of Salmonella Newport Infections Linked to Cucumbers — United States, 2014." *Morbidity and Mortality Weekly Report* 64 (6): 144–47.

Antunes, P., J. Mourão, J. Campos, and L. Peixe. 2016. "Salmonellosis: The Role of Poultry Meat." *Clinical Microbiology and Infection* 22 (2): 110–21. https://doi.org/10.1016/j.cmi.2015.12.004.

Ashton, Philip M., Satheesh Nair, Tansy M. Peters, Janet A. Bale, David G. Powell, Anaïs Painset, Rediat Tewolde, et al. 2016. "Identification of *Salmonella* for Public Health Surveillance Using Whole Genome Sequencing." *PeerJ* 4: e1752. https://doi.org/10.7717/peerj.1752.

Audia, Jonathon P., Cathy C. Webb, and John W. Foster. 2001. "Breaking through the Acid Barrier: An Orchestrated Response to Proton Stress by Enteric Bacteria." *International Journal of Medical Microbiology* 291 (2): 97–106. https://doi.org/10.1078/1438-4221-00106.

Banoo, Shabir, David Bell, Patrick Bossuyt, Alan Herring, David Mabey, Freddie Poole, Peter G Smith, et al. 2010. "Evaluation of Diagnostic Tests for Infectious Diseases: General Principles." *Nature Reviews. Microbiology* 8 (12 Suppl): S17-29.

Barbau-Piednoir, Elodie, Sophie Bertrand, Jacques Mahillon, Nancy H. Roosens, and Nadine Botteldoorn. 2013. "SYBR®Green QPCR Salmonella Detection System Allowing Discrimination at the Genus, Species and Subspecies Levels." *Applied Microbiology and Biotechnology* 97 (22): 9811–24. https://doi.org/10.1007/s00253-013-5234-x.

Barbau-Piednoir, Elodie, Nadine Botteldoorn, Marc Yde, Jacques Mahillon, and Nancy H. Roosens. 2013. "Development and Validation of Qualitative SYBR®Green Real-Time PCR for Detection and Discrimination of Listeria Spp. and Listeria Monocytogenes." *Applied Microbiology and Biotechnology* 97 (9): 4021–37. https://doi.org/10.1007/s00253-012-4477-2.

Barker, Ruth M. 1985. “Utilization of D-Tartaric Acid by Salmonella Paratyphi B and Salmonella Java: Comparison of Anaerobic Plate Test, Lead Acetate Test and Turbidity Test.” *Journal of Hygiene* 95 (1): 107–14. https://doi.org/10.1017/S0022172400062331.

Batista, Diego Felipe Alves, Oliveiro Caetano de Freitas Neto, Priscila Diniz Lopes, Adriana Maria de Almeida, Paul Andrew Barrow, and Angelo Berchieri. 2013. “Polymerase Chain Reaction Assay Based on RatA Gene Allows Differentiation between Salmonella Enterica Subsp. Enterica Serovar Gallinarum Biovars Gallinarum and Pullorum.” *Journal of Veterinary Diagnostic Investigation* 25 (2): 259–62. https://doi.org/10.1177/1040638713479361.

Bell, Rebecca L., Karen G. Jarvis, Andrea R. Ottesen, Melinda A. Mcfarland, and Eric W. Brown. 2016. “Recent and Emerging Innovations in Salmonella Detection: A Food and Environmental Perspective.” *Microbial Biotechnology* 9 (3): 279–92. https://doi.org/10.1111/1751-7915.12359.

Bertrand, Sophie, Katelijne Dierick, Kim Heylen, Thierry De Baere, Brigitte Pochet, Emmanuel Robesyn, Sophie Lokietek, et al. 2016. “Lessons Learned from the Management of a National Outbreak of Salmonella Ohio Linked to Pork Meat Processing and Distribution.” *Journal of Food Protection* 73 (3): 529–34. https://doi.org/10.4315/0362-028x-73.3.529.

Bertrand, Sophie, François Xavier Weill, Axel Cloeckaert, Martine Vrints, Eric Mairiaux, Karine Fraud, Katlijne Dierick, et al. 2006. “Clonal Emergence of Extended-Spectrum β-Lactamase (CTX-M-2)-Producing Salmonella Enterica Serovar Virchow Isolates with Reduced Susceptibilities to Ciprofloxacin among Poultry and Humans in Belgium and France (2000 to 2003).” *Journal of Clinical Microbiology* 44 (8): 2897–2903. https://doi.org/10.1128/JCM.02549-05.

Berwouts, Sarah, Michael A Morris, and Elisabeth Dequeker. 2010. “Approaches to Quality Management and Accreditation in a Genetic Testing Laboratory.” *European Journal of Human Genetics*, 1–19. https://doi.org/10.1038/ejhg.2010.104.

Bopp, Dianna J., Deborah J. Baker, Lisa Thompson, Amy Saylors, Timothy P. Root, Leeanna Armstrong, Kara Mitchell, Nellie B. Dumas, and Kimberlee Arruda Musser. 2016. “Implementation of Salmonella Serotype Determination Using Pulsed-Field Gel Electrophoresis in a State Public Health Laboratory.” *Diagnostic Microbiology and Infectious Disease* 85 (4): 416–18. https://doi.org/10.1016/j.diagmicrobio.2016.04.023.

Braun, Sascha D., Albrecht Ziegler, Ulrich Methner, Peter Slickers, Silke Keiling, Stefan Monecke, and Ralf Ehricht. 2012. “Fast DNA Serotyping and Antimicrobial Resistance Gene Determination of Salmonella Enterica with an Oligonucleotide Microarray-Based Assay.” *PLoS ONE* 7 (10). https://doi.org/10.1371/journal.pone.0046489.

Bulcke, Marc Van Den, Antoon Lievens, Elodie Barbau-Piednoir, Guillaume Mbongolombella, Nancy Roosens, Myriam Sneyers, and Amaya Leunda Casi. 2010. “A Theoretical Introduction to ‘Combinatory SYBR®Green QPCR Screening’, a Matrix-Based Approach for the Detection of Materials Derived from

Genetically Modified Plants." *Analytical and Bioanalytical Chemistry* 396 (6): 2113–23. https://doi.org/10.1007/s00216-009-3286-7.

Bulcke, Marc Van den, APNR Lievens, A Leunda, EG MbongoloMbella, E Barbau-Piednoir, and MJS Sneyers. 2008. Transgenic plant event detection. Patent WO 2008/092866, issued 2008.

Burd, Eileen M. 2010. "Validation of Laboratory-Developed Molecular Assays for Infectious Diseases" 23 (3): 550–76. https://doi.org/10.1128/CMR.00074-09.

Cardona-Castro, Nora, Miryan Sánchez-Jiménez, Lelia Lavalett, Nélida Múñoz, and Jaime Moreno. 2009. "Development and Evaluation of a Multiplex Polymerase Chain Reaction Assay to Identify Salmonella Serogroups and Serotypes." *Diagnostic Microbiology and Infectious Disease* 65 (3): 327–30. https://doi.org/10.1016/j.diagmicrobio.2009.07.003.

Castellanos, L Ricardo, Linda van der Graaf-van Bloois, Pilar Donado-Godoy, Kees Veldman, Francisco Duarte, María T Acuña, Claudia Jarquín, et al. 2020. "Antimicrobial Resistance in Salmonella Enterica Serovar Paratyphi B Variant Java in Poultry from Europe and Latin America." *Emerging Infectious Diseases* 26 (6): 1164–73. https://doi.org/10.3201/eid2606.191121.

Cauteren, Dieter Van. 2016. "Estimation de La Morbidité Des Infections d'origine Alimentaire En France." Université Paris-Sud.

Cauteren, Dieter Van, Henriette De Valk, Cecile Sommen, Lisa A King, Nathalie Jourdan-Da Silva, Francois-Xavier Weill, Simon Le Hello, Francis Megraud, Veronique Vaillant, and Jean C Desenclos. 2015. "Community Incidence of Campylobacteriosis and Nontyphoidal Salmonellosis, France, 2008-2013." *Foodborne Pathogens and Disease* 12 (8): 664–69. https://doi.org/10.1089/fpd.2015.1964.

CDC. 2019. "Serotypes and the Importance of Serotyping Salmonella." 2019. https://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotyping-importance.html.

Ceyssens, Pieter Jan, Cristina Garcia-Graells, Frédéric Fux, Nadine Botteldoorn, Wesley Mattheus, Véronique Wuyts, Sigrid De Keersmaecker, Katelijne Dierick, and Sophie Bertrand. 2016. "Development of a Luminex XTAG® Assay for Cost-Effective Multiplex Detection of β-Lactamases in Gram-Negative Bacteria." *The Journal of Antimicrobial Chemotherapy* 71 (9): 2479–83. https://doi.org/10.1093/jac/dkw201.

Christensen, J. P., J. E. Olsen, H. C. Hansen, and M. Bisgaard. 1992. "Characterization of Salmonella Enterica Serovar Gallinarum Biovars Gallinarum and Pullorum By Plasmid Profiling and Biochemical Analysis." *Avian Pathology* 21 (3): 461–70. https://doi.org/10.1080/03079459208418864.

Collard, J. M., S. Bertrand, K. Dierick, C. Godard, C. Wildemauwe, K. Vermeersch, J. Duculot, et al. 2008. "Drastic Decrease of Salmonella Enteritidis Isolated from Humans in Belgium in 2005, Shift in Phage Types and Influence on Foodborne Outbreaks." *Epidemiology and Infection* 136 (6): 771–81. https://doi.org/10.1017/S095026880700920X.

Connor, Thomas R., Sian V. Owen, Gemma Langridge, Steve Connell, Satheesh Nair, Sandra Reuter, Timothy J. Dallman, et al. 2016. “What’s in a Name? Species-Wide Whole-Genome Sequencing Resolves Invasive and Noninvasive Lineages of Salmonella Enterica Serotype Paratyphi B.” *MBio* 7 (4): 1–9. https://doi.org/10.1128/mBio.00527-16.

Crim, Stacy M, Shua J Chai, Beth E Karp, Michael C Judd, Jared Reynolds, C Krista, Amie Nisler, Andre Mccullough, and L Hannah Gould. 2019. “Salmonella Enterica Serotype Newport Infections in the United States, 2004–2013: Increased Incidence Investigated through Four Surveillance Systems.” *Foodborne Pathogens and Disease* 15 (10): 612–20. https://doi.org/10.1089/fpd.2018.2450.Salmonella.

Demirbilek, Serpil K. 2016. “Salmonellosis in Animals.” In *Salmonella - A Re-Emerging Pathogen*, edited by Maria Teresa Mascellino, 18. IntechOpen. https://doi.org/http://dx.doi.org/10.5772/57353.

Dhama, Kuldeep, Sandip Chakraborty, Mahima, Mohd Yaqoob Wani, Amit Kumar Verma, Rajib Deb, Ruchi Tiwari, and Sanjay Kapoor. 2013. “Novel and Emerging Therapies Safeguarding Health of Humans and Their Companion Animals: A Review.” *Pakistan Journal of Biological Sciences : PJBS* 16 (3): 101–11.

Dieckmann, Ralf, and Burkhard Malorny. 2011. “Rapid Screening of Epidemiologically Important Salmonella Enterica Subsp. Enterica Serovars by Whole-Cell Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry.” *Applied and Environmental Microbiology* 77 (12): 4136–46. https://doi.org/10.1128/AEM.02418-10.

Doublet, Benoît, Karine Praud, Tran Nguyen-Ho-Bao, María Angeles Argudín, Sophie Bertrand, Patrick Butaye, and Axel Cloeckaert. 2013. “Extended-Spectrum β-Lactamase- and AmpC β-Lactamase-Producing d-Tartrate-Positive Salmonella Enterica Serovar Paratyphi B from Broilers and Human Patients in Belgium, 2008–10.” *Journal of Antimicrobial Chemotherapy* 69 (5): 1257–64. https://doi.org/10.1093/jac/dkt504.

ECDC. 2018. “Annual Epidemiological Report for 2016 Typhoid and Paratyphoid Fevers.”

Echeita, M. Aurora, Silvia Herrera, Javier Garaizar, and Miguel A. Usera. 2002. “Multiplex PCR-Based Detection and Identification of the Most Common Salmonella Second-Phase Flagellar Antigens.” *Research in Microbiology* 153 (2): 107–13. https://doi.org/10.1016/S0923-2508(01)01295-5.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

EFSA. 2017. “The European Union Summary Report on Trends and Sources of Zoonoses , Zoonotic Agents and Food-Borne Outbreaks In.” Vol. 15. https://doi.org/10.2903/j.efsa.2017.5077.

EFSA. 2018. “The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-Borne Outbreaks in 2017.” *EFSA Journal*. Vol. 16. https://doi.org/10.2903/j.efsa.2018.5500.

EFSA. 2019a. “Salmonella.” 2019. https://www.efsa.europa.eu/en/topics/topic/salmonella.

EFSA. 2019b. “The European Union One Health 2018 Zoonoses Report.” Vol. 17. https://doi.org/10.2903/j.efsa.2019.5926.

Eng, Shu Kee, Priyia Pusparajah, Nurul Syakima Ab Mutalib, Hooi Leng Ser, Kok Gan Chan, and Learn Han Lee. 2015. “Salmonella: A Review on Pathogenesis, Epidemiology and Antibiotic Resistance.” *Frontiers in Life Science* 8 (3): 284–93. https://doi.org/10.1080/21553769.2015.1051243.

Evangelopoulou, G., S. Kritas, G. Christodoulopoulos, and A. R. Burriel. 2015. “The Commercial Impact of Pig Salmonella Spp. Infections in Border-Free Markets during an Economic Recession.” *Veterinary World*. Vol. 8. https://doi.org/10.14202/vetworld.2015.257-272.

Fabre, Laetitia, Simon Le Hello, Chrystelle Roux, Sylvie Issenhuth-Jeanjean, and François Xavier Weill. 2014. “CRISPR Is an Optimal Target for the Design of Specific PCR Assays for Salmonella Enterica Serotypes Typhi and Paratyphi A.” *PLoS Neglected Tropical Diseases* 8 (1): 14. https://doi.org/10.1371/journal.pntd.0002671.

Fabre, Laëtitia, Jian Zhang, Ghislaine Guigon, Simon Le Hello, Véronique Guibert, Marie Accou-Demartin, Saïana de Romans, et al. 2012. “Crispr Typing and Subtyping for Improved Laboratory Surveillance of Salmonella Infections.” *PLoS ONE* 7 (5). https://doi.org/10.1371/journal.pone.0036995.

Farrell, John J., Laura J. Doyle, Rachel M. Addison, L. Barth Reller, Geraldine S. Hall, and Gary W. Procop. 2005. “Broad-Range (Pan) Salmonella and Salmonella Serotype Typhi-Specific Real-Time PCR Assays: Potential Tools for the Clinical Microbiologist.” *American Journal of Clinical Pathology* 123 (3): 339–45. https://doi.org/10.1309/DP0HY5UT10HQW9YM.

FASFC circular. 2019. “Circulaire Relative à La Lutte Contre Les Salmonelles Zoonotiques Chez Les Volailles de Rente Type Viande (PCCB/S2/589616).”

FASFC circular. 2020a. “Circulaire Relative à La Lutte Contre Les Salmonelles Chez Les Volailles Reproductrices (PCCB/S2/418588).”

FASFC circular. 2020b. “Circulaire Relative à La Lutte Contre Les Salmonelles Zoonotiques Chez Les Poules Pondeuses (PCCB/S2/409035).”

FASFC note. 2019. “Programmes Nationaux de Lutte Contre Les Salmonelles Chez Les Volailles (BP-MN-FDS/LABO/1557457).”

Fitzgerald, Collette, Marcus Collins, Susan Van Duyne, Matthew Mikoleit, Teresa Brown, and Patricia Fields. 2007. “Multiplex, Bead-Based Suspension Array for Molecular Determination of Common Salmonella Serogroups.” *Journal of Clinical Microbiology* 45 (10): 3323–34. https://doi.org/10.1128/JCM.00025-07.

Foley, S. L., T. J. Johnson, S. C. Ricke, R. Nayak, and J. Danzeisen. 2013. “Salmonella Pathogenicity and Host Adaptation in Chicken-Associated Serovars.” *Microbiology and Molecular Biology Reviews* 77 (4): 582–

607. https://doi.org/10.1128/mmbr.00015-13.

Franklin, Kristyn, Erika J. Lingohr, Catherine Yoshida, Muna Anjum, Levente Bodrossy, Clifford G. Clark, Andrew M. Kropinski, and Mohamed A. Karmali. 2011. “Rapid Genoserotyping Tool for Classification of Salmonella Serovars.” *Journal of Clinical Microbiology* 49 (8): 2954–65. https://doi.org/10.1128/JCM.02347-10.

Gambino-Shirley, Kelly J., Adiam Tesfai, Colin A. Schwensohn, Cindy Burnett, Lori Smith, Jennifer M. Wagner, Dana Eikmeier, et al. 2018. “Multistate Outbreak of Salmonella Virchow Infections Linked to a Powdered Meal Replacement Product - United States, 2015-2016.” *Clinical Infectious Diseases* 67 (6): 890–96. https://doi.org/10.1093/cid/ciy195.

Gand, M., W. Mattheus, A. Saltykova, N. Roosens, K. Dierick, K. Marchal, S.C.J. De Keersmaecker, and S. Bertrand. 2019. “Development of a Real-Time PCR Method for the Genoserotyping of Salmonella Paratyphi B Variant Java.” *Applied Microbiology and Biotechnology* 103 (12): 4987–96. https://doi.org/10.1007/s00253-019-09854-4.

Gand, M, W Mattheus, N H.C. Roosens, K Dierick, K Marchal, S.C.J. De Keersmaecker, and S Bertrand. 2020. “A Multiplex Oligonucleotide Ligation-PCR Method for the Genoserotyping of Common Salmonella Using a Liquid Bead Suspension Assay.” *Food Microbiology* 87 (May). https://doi.org/10.1016/j.fm.2019.103394.

Gibbons, Cheryl L., Marie Josée J. Mangen, Dietrich Plass, Arie H. Havelaar, Russell John Brooke, Piotr Kramarz, Karen L. Peterson, et al. 2014. “Measuring Underreporting and Under-Ascertainment in Infectious Disease Datasets: A Comparison of Methods.” *BMC Public Health* 14 (1): 1–17. https://doi.org/10.1186/1471-2458-14-147.

Griffin, Patricia M., and Sarah J. O’Brien. 2013. “The ‘Decline and Fall’ of Nontyphoidal Salmonella in the United Kingdom.” *Clinical Infectious Diseases* 56 (5): 705–10. https://doi.org/10.1093/cid/cis967.

Grimont, Patrick A.D;, and François-Xavier Weill. 2007. *Antigenic Formulae of the Salmonella Serovars*. 9th ed. Vol. 9th editio. Paris: Institut Pasteur.

Guiney, Donald G., and Joshua Fierer. 2011. “The Role of the Spv Genes in Salmonella Pathogenesis.” *Frontiers in Microbiology* 2 (JUNE): 1–10. https://doi.org/10.3389/fmicb.2011.00129.

Guo, Dan, Bin Liu, Fenxia Liu, Boyang Cao, Min Chen, Xiyan Hao, and Lu Feng. 2013. “Development of a DNA Microarray for Molecular Identification of All 46 Salmonella O Serogroups” 79 (11): 3392–99. https://doi.org/10.1128/AEM.00225-13.

Havelaar, A. H., S. Ivarsson, M. Löfdahl, and M. J. Nauta. 2013. “Estimating the True Incidence of Campylobacteriosis and Salmonellosis in the European Union, 2009.” *Epidemiology and Infection* 141 (2): 293–302. https://doi.org/10.1017/S0950268812000568.

Hayward, Matthew R., Vincent A.A. Jansen, and Martin J. Woodward. 2013. "Comparative Genomics of Salmonella Enterica Serovars Derby and Mbandaka, Two Prevalent Serovars Associated with Different Livestock Species in the UK." *BMC Genomics* 14 (1): 1. https://doi.org/10.1186/1471-2164-14-365.

Hello, Simon Le, Amany Bekhit, Sophie A. Granier, Himel Barua, Janine Beutlich, Magdalena Zając, Sebastian Münch, et al. 2013. "The Global Establishment of a Highly-Fluoroquinolone Resistant Salmonella Enterica Serotype Kentucky ST198 Strain." *Frontiers in Microbiology* 4 (DEC): 1–10. https://doi.org/10.3389/fmicb.2013.00395.

Heredia, Norma, and Santos García. 2018. "Animals as Sources of Food-Borne Pathogens: A Review." *Animal Nutrition* 4 (3): 250–55. https://doi.org/10.1016/J.ANINU.2018.04.006.

Herrera-León, Silvia, John R. McQuiston, Miguel A. Usera, Patricia I. Fields, Javier Garaizar, and M. Aurora Echeita. 2004. "Multiplex PCR for Distinguishing the Most Common Phase-1 Flagellar Antigens of Salmonella Spp." *Journal of Clinical Microbiology* 42 (6): 2581–86. https://doi.org/10.1128/JCM.42.6.2581-2586.2004.

Herzberg, M., and J. H. Green. 1964. "Composition and Characteristics of Cell Walls of Smooth Strains of Salmonella Typhimurium and Derived Rough Variants." *Journal of General Microbiology* 35 (3): 421–36. https://doi.org/10.1099/00221287-35-3-421.

Hoszowski, A, A Lalak, M Zając, I Samcik, M Skarżyńska, D Wnuk, and D Wasyl. 2011. "Pokrewieństwo Szorstkich Szczepów Salmonella z Reprezentantami Niektórych Serowarów Stwierdzanych u Zwierząt." *Medycyna Weterynaryjna* 67 (3): 194–97.

Hwa Lee, Su, Byeong Yeal Jung, Nabin Rayamajhi, Hee Soo Lee, Woo Jin Jeon, Kang-Seuk Choi, Chang Kweon, and Han Sang Yoo. 2009. "A Multiplex Real-Time PCR for Differential Detection and Quantification of Salmonella Spp., Salmonella Enterica Serovar Typhimurium and Enteritidis in Meats." *Journal of Veterinary Science* 10 (April): 43–51. https://doi.org/10.4142/jvs.2009.10.1.43.

Ibrahim, George M, and Paul M Morin. 2018. "Salmonella Serotyping Using Whole Genome Sequencing." *Frontiers in Microbiology* 9: 2993. https://doi.org/10.3389/fmicb.2018.02993.

Institut Pasteur. 2018. "Rapport d'activité Annuel 2018 Année d'exercice 2017. Centre National de Référence Des Escherichia Coli, Shigella et Salmonella." https://www.pasteur.fr/fr/file/21346/download.

Institut Pasteur. 2019. "Rapport d'activité Annuel 2019 - Année d'exercice 2018. Centre National de Référence Des Escherichia Coli, Shigella et Salmonella." https://www.pasteur.fr/fr/file/30716/download.

Issenhuth-Jeanjean, Sylvie, Peter Roggentin, Matthew Mikoleit, Martine Guibourdenche, Elizabeth De Pinna, Satheesh Nair, Patricia I. Fields, and François Xavier Weill. 2014. "Supplement 2008-2010 (No. 48) to the White-Kauffmann-Le Minor Scheme." *Research in Microbiology* 165 (7): 526–30. https://doi.org/10.1016/j.resmic.2014.07.004.

Iwamoto, Martha, Jared Reynolds, Beth E. Karp, Heather Tate, Paula J. Fedorka-Cray, Jodie R. Plumblee, Robert M. Hoekstra, Jean M. Whichard, and Barbara E. Mahon. 2017. "Ceftriaxone-Resistant Nontyphoidal Salmonella from Humans, Retail Meats, and Food Animals in the United States, 1996-2013." *Foodborne Pathogens and Disease* 14 (2): 74–83. https://doi.org/10.1089/fpd.2016.2180.

Jain, Sonali, Kakali Mukhopadhyay, and Paul J. Thomassin. 2019. "An Economic Analysis of Salmonella Detection in Fresh Produce, Poultry, and Eggs Using Whole Genome Sequencing Technology in Canada." *Food Research International* 116 (August 2018): 802–9. https://doi.org/10.1016/j.foodres.2018.09.014.

Jean-Gilles Beaubrun, J., L. Ewing, K. Jarvis, K. Dudley, C. Grim, G. Gopinath, M. L. Flamer, et al. 2014. "Comparison of a PCR Serotyping Assay, Check&Trace Assay for Salmonella, and Luminex Salmonella Serotyping Assay for the Characterization of Salmonella Enterica Identified from Fresh and Naturally Contaminated Cilantro." *Food Microbiology* 42: 181–87. https://doi.org/10.1016/j.fm.2014.02.008.

Jessica, M. Healy, and B. Bruce Beau. 2019. "Salmonellosis (Nontyphoidal)." 24/06/2019. 2019. https://wwwnc.cdc.gov/travel/yellowbook/2020/travel-related-infectious-diseases/salmonellosis-nontyphoidal.

Jia, Yibing. 2012. "Real-Time PCR." *Methods in Cell Biology* 112 (January): 55–68. https://doi.org/10.1016/B978-0-12-405914-6.00003-2.

Jourdan-da Silva, Nathalie, Laetitia Fabre, Eve Robinson, Nelly Fournet, Athinna Nisavanh, Mathias Bruyand, Alexandra Mailles, et al. 2018. "Ongoing Nationwide Outbreak of Salmonella Agona Associated with Internationally Distributed Infant Milk Products, France, December 2017." *Eurosurveillance* 23 (2): 1–5. https://doi.org/10.2807/1560-7917.ES.2018.23.2.17-00852.

Kang, Lin, Nan Li, Ping Li, Yang Zhou, Shan Gao, Hongwei Gao, Wenwen Xin, and Jinglin Wang. 2017. "MALDI-TOF Mass Spectrometry Provides High Accuracy in Identification of Salmonella at Species Level but Is Limited to Type or Subtype Salmonella Serovars." *European Journal of Mass Spectrometry* 23 (2): 70–82. https://doi.org/10.1177/1469066717699216.

Karns, Jeffrey S., Bradd J. Haley, and Jo Ann S. Van Kessel. 2015. "Improvements to a PCR-Based Serogrouping Scheme for Salmonella Enterica from Dairy Farm Samples." *Journal of Food Protection* 78 (6): 1182–85. https://doi.org/10.4315/0362-028X.JFP-14-475.

Keefer, Andrea B., Lingzi Xiaoli, Nkuchia M. M'ikanatha, Kuan Yao, Maria Hoffmann, and Edward G. Dudley. 2019. "Retrospective Whole-Genome Sequencing Analysis Distinguished PFGE and Drug-Resistance-Matched Retail Meat and Clinical Salmonella Isolates." *Microbiology* 165 (3): 270–86. https://doi.org/10.1099/mic.0.000768.

Kérouanton, Annaëlle, Muriel Marault, Renaud Lailler, François-Xavier Weill, Carole Feurer, Emmanuelle Espié, and Anne Brisabois. 2007. " Pulsed-Field Gel Electrophoresis Subtyping Database for Foodborne Salmonella Enterica Serotype Discrimination ." *Foodborne Pathogens and Disease* 4 (3): 293–303.

https://doi.org/10.1089/fpd.2007.0090.

Kim, H J, S H Park, T H Lee, B H Nahm, Y H Chung, K H Seo, and H Y Kim. 2006. “Identification of Salmonella Enterica Serovar Typhimurium Using Specific PCR Primers Obtained by Comparative Genomics in Salmonella Serovars.” *Journal of Food Protection* 69 (7): 1653–61. https://doi.org/10.4315/0362-028X-69.7.1653.

Kim, Seonghan, Jonathan G. Frye, Jinxin Hu, Paula J. Fedorka-Cray, Romesh Gautom, and David S. Boyle. 2006. “Multiplex PCR-Based Method for Identification of Common Clinical Serotypes of Salmonella Enterica Subsp. Enterica.” *Journal of Clinical Microbiology* 44 (10): 3608–15. https://doi.org/10.1128/JCM.00701-06.

Koutsoumanis, Kostas, Ana Allende, Avelino Alvarez-Ordóñez, Declan Bolton, Sara Bover-Cid, Marianne Chemaly, Alessandra De Cesare, et al. 2019. “Salmonella Control in Poultry Flocks and Its Public Health Impact.” *EFSA Journal*. Vol. 17. https://doi.org/10.2903/j.efsa.2019.5596.

Kozyreva, Varvara K, John Crandall, Ashley Sabol, Alyssa Poe, Peng Zhang, Jeniffer Concepción-Acevedo, Morgan N Schroeder, et al. 2016. “Laboratory Investigation of Salmonella Enterica Serovar Poona Outbreak in California: Comparison of Pulsed-Field Gel Electrophoresis (PFGE) and Whole Genome Sequencing (WGS) Results.” *PLoS Currents* 8 (November): ecurrents.outbreaks.1bb3e36e74bd5779bc43ac3a8dae52. https://doi.org/10.1371/currents.outbreaks.1bb3e36e74bd5779bc43ac3a8dae52e6.

Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. 2016. “MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.” *Molecular Biology and Evolution* 33 (7): 1870–74. https://doi.org/10.1093/molbev/msw054.

Lalsiamthara, Jonathan, Je Hyoung Kim, and John Hwa Lee. 2018. “Engineering of a Rough Auxotrophic Mutant Salmonella Typhimurium for Effective Delivery.” *Oncotarget* 9 (39): 25441–57. https://doi.org/10.18632/oncotarget.25192.

Lampel, K. A., S. P. Keasler, and D. E. Hanes. 1996. “Specific Detection of Salmonella Enterica Serotype Enteritidis Using the Polymerase Chain Reaction.” *Epidemiology and Infection* 116 (2): 137–45. https://doi.org/10.1017/S0950268800052365.

Leekitcharoenphon, Pimlapas, Gitte Sørensen, Charlotta Löfström, Antonio Battisti, Istvan Szabo, Dariusz Wasyl, Rosemarie Slowey, et al. 2019. “Cross-Border Transmission of Salmonella Choleraesuis Var. Kunzendorf in European Pigs and Wild Boar: Infection, Genetics, and Evolution.” *Frontiers in Microbiology* 10 (feburay): 1–12. https://doi.org/10.3389/fmicb.2019.00179.

Leggett, Richard M., and Matthew D. Clark. 2017. “A World of Opportunities with Nanopore Sequencing.” *Journal of Experimental Botany* 68 (20): 5419–29. https://doi.org/10.1093/jxb/erx289.

Liang, D. W., J. H. Lu, Q. Wu, B. X. Ke, C. H. Jiang, J. Long, Y. P. Fang, et al. 2016. “Comparing the Ability of

Luminex XMAP® Salmonella Serotyping Assay and Traditional Serotyping Method for Serotyping Salmonella Isolated from Southern Chinese Population." *Journal of Applied Microbiology* 120 (6): 1668–76. https://doi.org/10.1111/jam.13106.

Liu, Bin, Lida Zhang, Xinna Zhu, Chunlei Shi, Jing Chen, Weibing Liu, Xiaohua He, and Xianming Shi. 2011. "PCR Identification of Salmonella Serogroups Based on Specific Targets Obtained by Comparative Genomics." *International Journal of Food Microbiology* 144 (3): 511–18. https://doi.org/10.1016/j.ijfoodmicro.2010.11.010.

Luk, J. M.C., U. Kongmuang, P. R. Reeves, and A. A. Lindberg. 1993. "Selective Amplification of Abequose and Paratose Synthase Genes (Rfb) by Polymerase Chain Reaction for Identification of Salmonella Major Serogroups (A, B, C2, and D)." *Journal of Clinical Microbiology* 31 (8): 2118–23.

Majowicz, Shannon E., Jennie Musto, Elaine Scallan, Frederick J. Angulo, Martyn Kirk, Sarah J. O'Brien, Timothy F. Jones, Aamir Fazil, and Robert M. Hoekstra. 2010. "The Global Burden of Nontyphoidal *Salmonella* Gastroenteritis." *Clinical Infectious Diseases* 50 (6): 882–89. https://doi.org/10.1086/650733.

Malorny, Burkhard, Cornelia Bunge, and Reiner Helmuth. 2003. "Discrimination of D-Tartrate-Fermenting and -Nonfermenting Salmonella Enterica Subsp. Enterica Isolates by Genotypic and Phenotypic Methods." *Journal of Clinical Microbiology* 41 (9): 4292–97. https://doi.org/10.1128/JCM.41.9.4292-4297.2003.

Marineli, Filio, Gregory Tsoucalas, Marianna Karamanou, and George Androutsos. 2013. "Mary Mallon (1869-1938) and the History of Typhoid Fever." *Annals of Gastroenterology* 26 (2): 132–34.

Martínez-Avilés, Marta, Macarena Garrido-Estepa, Julio Álvarez, and Ana de la Torre. 2019. "Salmonella Surveillance Systems in Swine and Humans in Spain: A Review." *Veterinary Sciences* 6 (1): 20. https://doi.org/10.3390/vetsci6010020.

Maurischat, Sven, Istvan Szabo, Beatrice Baumann, and Burkhard Malorny. 2015. "Rapid Real-Time PCR Methods to Distinguish Salmonella Enteritidis Wildtype Field Isolates from Vaccine Strains Salmovac SE/Gallivac SE and AviPro SALMONELLA VAC E." *Journal of Microbiological Methods* 112: 92–98. https://doi.org/10.1016/j.mimet.2015.03.015.

Mba-Jonas, Adamma, Wright Culpepper, Thomas Hill, Venessa Cantu, Julie Loera, Julie Borders, Lori Saathoff-Huber, et al. 2018. "A Multistate Outbreak of Human Salmonella Agona Infections Associated with Consumption of Fresh, Whole Papayas Imported from Mexico - United States, 2011." *Clinical Infectious Diseases* 66 (11): 1756–61. https://doi.org/10.1093/cid/cix1094.

Mellou, Kassiani, Theologia Sideroglou, Athina Kallimani, Maria Potamiti-Komi, Danai Pervanidou, Eleni Lillakou, Theano Georgakopoulou, et al. 2013. "Evaluation of Underreporting of Salmonellosis and Shigellosis Hospitalised Cases in Greece, 2011: Results of a Capture-Recapture Study and a Hospital Registry Review." *BMC Public Health* 13 (1): 1. https://doi.org/10.1186/1471-2458-13-875.

Miller, Ruth R., Vincent Montoya, Jennifer L. Gardy, David M. Patrick, and Patrick Tang. 2013. “Metagenomics for Pathogen Detection in Public Health.” *Genome Medicine* 5 (9). https://doi.org/10.1186/gm485.

Minor, Leon Le, and Michel Y Popoff. 1987. “Designation of Salmonella Enterica Sp. Nov., Nom. Rev., as the Type and Only Species of the Genus Salmonella: Request for an Opinion.” *International Journal of Systematic and Evolutionary Microbiology* 37 (4): 465–68. https://doi.org/https://doi.org/10.1099/00207713-37-4-465.

Mølbak, Kåre, Jacob Simonsen, Charlotte S. Jørgensen, Karen A. Krogfelt, Gerhard Falkenhorst, Steen Ethelberg, Johanna Takkinen, and Hanne Dorthe Emborg. 2014. “Seroincidence of Human Infections with Nontyphoid Salmonella Compared with Data from Public Health Surveillance and Food Animals in 13 European Countries.” *Clinical Infectious Diseases* 59 (11): 1599–1606. https://doi.org/10.1093/cid/ciu627.

Nair, Divek V.T., Kumar Venkitanarayanan, and Anup Kollanoor Johny. 2018. “Antibiotic-Resistant Salmonella in the Food Supply and the Potential Role of Antibiotic Alternatives for Control.” *Foods* 7 (10). https://doi.org/10.3390/foods7100167.

O’Regan, Edel, Evonne McCabe, Catherine Burgess, Sheila McGuinness, Thomas Barry, Geraldine Duffy, Paul Whyte, and Séamus Fanning. 2008. “Development of a Real-Time Multiplex PCR Assay for the Detection of Multiple Salmonella Serotypes in Chicken Samples.” *BMC Microbiology* 8: 1–11. https://doi.org/10.1186/1471-2180-8-156.

Oniciuc, Elena A., Eleni Likotrafiti, Adrián Alvarez-Molina, Miguel Prieto, Jesús A. Santos, and Avelino Alvarez-Ordóñez. 2018. “The Present and Future of Whole Genome Sequencing (WGS) and Whole Metagenome Sequencing (WMS) for Surveillance of Antimicrobial Resistant Microorganisms and Antimicrobial Resistance Genes across the Food Chain.” *Genes* 9 (5): 1–28. https://doi.org/10.3390/genes9050268.

Oxford Analytica. 2012. “A Report Produced for the International Federation for Animal Health: The Costs of Animal Disease.” https://doi.org/10.1057/9780230358522_6.

Pedersen, Karl, Gitte Sørensen, Charlotta Löfström, Pimlapas Leekitcharoenphon, Bent Nielsen, Anne Wingstrand, Frank M. Aarestrup, René S. Hendriksen, and Dorte Lau Baggesen. 2015. “Reappearance of Salmonella Serovar Choleraesuis Var. Kunzendorf in Danish Pig Herds.” *Veterinary Microbiology* 176 (3–4): 282–91. https://doi.org/10.1016/j.vetmic.2015.01.004.

Pijnacker, Roan, Timothy J. Dallman, Aloys S.L. Tijsma, Gillian Hawkins, Lesley Larkin, Saara M. Kotila, Giusi Amore, et al. 2019. “An International Outbreak of Salmonella Enterica Serotype Enteritidis Linked to Eggs from Poland: A Microbiological and Epidemiological Study.” *The Lancet Infectious Diseases* 19 (7): 778–86. https://doi.org/10.1016/S1473-3099(19)30047-7.

Pires, Sarah M, Leonardo De Knegt, and Tine Hald. 2011. “Estimation of the Relative Contribution of Different Food and Animal Sources to Human Salmonella Infections in the European Union.” *EFSA*. Vol. 1.

Pornsukarom, Suchawan, Arnoud H M van Vliet, and Siddhartha Thakur. 2018. “Whole Genome Sequencing Analysis of Multiple Salmonella Serovars Provides Insights into Phylogenetic Relatedness, Antimicrobial Resistance, and Virulence Markers across Humans, Food Animals and Agriculture Environmental Sources.” *BMC Genomics* 19 (1): 801. https://doi.org/10.1186/s12864-018-5137-4.

Rajtak, Ursula, Nola Leonard, Declan Bolton, and Séamus Fanning. 2011. “A Real-Time Multiplex SYBR Green I Polymerase Chain Reaction Assay for Rapid Screening of *Salmonella* Serotypes Prevalent in the European Union.” *Foodborne Pathogens and Disease* 8 (7): 769–80. https://doi.org/10.1089/fpd.2010.0768.

Ranjbar, Reza, Seyyed Mojtaba, Ali Mehrabi, Meysam Sarshar, Ali Najafi, and Rahim Soruri. 2017. “Simultaneous Molecular Detection of Salmonella Enterica Sero.” *Iran J Public Health* 46 (1): 103–11. http://ijph.tums.ac.ir.

Rasschaert, G., K. Houf, H. Imberechts, K. Grijspeerdt, L. De Zutter, and M. Heyndrickx. 2005. “Comparison of Five Repetitive-Sequence-Based PCR Typing Methods for Molecular Discrimination of Salmonella Enterica Isolates.” *Journal of Clinical Microbiology* 43 (8): 3615–23. https://doi.org/10.1128/JCM.43.8.3615-3623.2005.

Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the Control of Salmonella and Other Specified Food-Borne Zoonotic Agents.

Robertson, James, Catherine Yoshida, Simone Gurnik, Madison Mcgrogan, Kristin Davis, Gitanjali Arya, Stephanie A Murphy, Anil Nichani, and John H E Nash. 2018. “An Improved DNA Array-Based Classification Method for the Identification of Salmonella Serotypes Shows High Concordance between Traditional and Genotypic Testing,” 1–12.

Robertson, James, Catherine Yoshida, Peter Kruczkiewicz, Celine Nadon, Anil Nichani, Eduardo N. Taboada, and John Howard Eagles Nash. 2018. “Comprehensive Assessment of the Quality of Salmonella Whole Genome Sequence Data Available in Public Sequence Databases Using the Salmonella in Silico Typing Resource (SISTR).” *Microbial Genomics* 4 (2). https://doi.org/10.1099/mgen.0.000151.

Ryan, Michael P., Jean O’Dwyer, and Catherine C. Adley. 2017. “Evaluation of the Complex Nomenclature of the Clinically and Veterinary Significant Pathogen Salmonella.” *BioMed Research International* 2017. https://doi.org/10.1155/2017/3782182.

Rychlik, Ivan, and Paul A. Barrow. 2005. “Salmonella Stress Management and Its Relevance to Behaviour during Intestinal Colonisation and Infection.” *FEMS Microbiology Reviews* 29 (5): 1021–40. https://doi.org/10.1016/j.femsre.2005.03.005.

Shivaprasad, H. L. 2000. “Fowl Typhoid and Pullorum Disease.” *OIE Revue Scientifique et Technique* 19 (2): 405–24. https://doi.org/10.20506/rst.19.2.1222.

Sofos, John N. 2008. “Challenges to Meat Safety in the 21st Century.” *Meat Science* 78 (1–2): 3–13.

https://doi.org/10.1016/j.meatsci.2007.07.027.

Soper, G A. 1939. “The Curious Career of Typhoid Mary.” *Bulletin of the New York Academy of Medicine* 15 (10): 698–712. https://www.ncbi.nlm.nih.gov/pubmed/19312127.

Tamura, Koichiro, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar. 2013. “MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0.” *Molecular Biology and Evolution* 30 (12): 2725–29. https://doi.org/10.1093/molbev/mst197.

Tennant, Sharon M., Deanna Toema, Farah Qamar, Najeeha Iqbal, Mary Adetinuke Boyd, Joanna M. Marshall, William C. Blackwelder, et al. 2015. “Detection of Typhoidal and Paratyphoidal Salmonella in Blood by Real-Time Polymerase Chain Reaction.” *Clinical Infectious Diseases* 61 (Suppl 4): S241–50. https://doi.org/10.1093/cid/civ726.

Thomas, M. Kate, Regan Murray, Logan Flockhart, Katarina Pintar, Frank Pollari, Aamir Fazil, Andrea Nesbitt, and Barbara Marshall. 2013. “Estimates of the Burden of Foodborne Illness in Canada for 30 Specified Pathogens and Unspecified Agents, Circa 2006.” *Foodborne Pathogens and Disease* 10 (7): 639–48. https://doi.org/10.1089/fpd.2012.1389.

Untergasser, Andreas, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C Faircloth, Maido Remm, and Steven G Rozen. 2012. “Primer3 — New Capabilities and Interfaces” 40 (15): 1–12. https://doi.org/10.1093/nar/gks596.

Ventola, Eleonora, Bert Bogaerts, Sigrid C. J. De Keersmaecker, Kevin Vanneste, Nancy H. C. Roosens, Wesley Mattheus, and Pieter-Jan Ceyssens. 2019. “Shifting National Surveillance of Shigella Infections toward Geno-serotyping by the Development of a Tailored Luminex Assay and NGS Workflow.” *MicrobiologyOpen*, no. December 2018: e00807. https://doi.org/10.1002/mbo3.807.

Wattiau, Pierre, Cécile Boland, and Sophie Bertrand. 2011. “Methodologies for Salmonella Enterica Subsp. Enterica Subtyping: Gold Standards and Alternatives.” *Applied and Environmental Microbiology* 77 (22): 7877–85. https://doi.org/10.1128/AEM.05527-11.

Weill, François-Xavier, Laetitia Fabre, Véronique Guibert, Laure Diancourt, and Sylvain Brisse. 2007. Molecular typing and subtyping of salmonella by identification of the variable nucleotide sequences of the CRISPR loci. PCT/B2O08/OO4OO4, issued 2007.

Wessels, E, L G Rusman, M J A W M van Bussel, and E C J Claas. 2014. “Added Value of Multiplex Luminex Gastrointestinal Pathogen Panel (XTAG(R) GPP) Testing in the Diagnosis of Infectious Gastroenteritis.” *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 20 (3): O182-7. https://doi.org/10.1111/1469-0691.12364.

WHO. 2019. “Typhoid.” 2019. https://www.who.int/news-room/fact-sheets/detail/typhoid.

Wise, Mark G., Gregory R. Siragusa, Jodie Plumblee, Mimi Healy, Paula J. Cray, and Bruce S. Seal. 2009. "Predicting Salmonella Enterica Serotypes by Repetitive Sequence-Based PCR." *Journal of Microbiological Methods* 76 (1): 18–24. https://doi.org/10.1016/j.mimet.2008.09.006.

Wuyts, Véronique. 2015. "Development of a Molecular Alternative for Classical Microbiological Subtyping Methods." KU Leuven.

Wuyts, Véronique, Wesley Mattheus, Nancy H.C. Roosens, Kathlee Marchal, Sophie Bertrand, and Sigrid C.J. De Keersmaecker. 2015. "A Multiplex Oligonucleotide Ligation-PCR as a Complementary Tool for Subtyping of Salmonella Typhimurium." *Applied Microbiology and Biotechnology* 99 (19): 8137–49. https://doi.org/10.1007/s00253-015-6831-7.

Wuyts, Véronique, Nancy H C Roosens, Sophie Bertrand, Kathleen Marchal, and Sigrid C J De Keersmaecker. 2015. "Guidelines for Optimisation of a Multiplex Oligonucleotide Ligation-PCR for Characterisation of Microbial Pathogens in a Microsphere Suspension Array." *BioMed Research International* 2015. https://doi.org/10.1155/2015/790170.

Yachison, Chris A., Catherine Yoshida, James Robertson, John H.E. Nash, Peter Kruczkiewicz, Eduardo N. Taboada, Matthew Walker, et al. 2017. "The Validation and Implications of Using Whole Genome Sequencing as a Replacement for Traditional Serotyping for a National Salmonella Reference Laboratory." *Frontiers in Microbiology* 8 (JUN): 1–9. https://doi.org/10.3389/fmicb.2017.01044.

Yoshida, Catherine, Kristyn Franklin, Paulina Konczy, John R Mcquiston, Patricia I Fields, John H Nash, Ed N Taboada, and Kris Rahn. 2007. "Methodologies towards the Development of an Oligonucleotide Microarray for Determination of Salmonella Serotypes" 70: 261–71. https://doi.org/10.1016/j.mimet.2007.04.018.

Yoshida, Catherine, Peter Kruczkiewicz, Chad R. Laing, Erika J. Lingohr, Victor P.J. Gannon, John H.E. Nash, and Eduardo N. Taboada. 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): 1–17. https://doi.org/10.1371/journal.pone.0147101.

Yoshida, Catherine, Erika J. Lingohr, Friederike Trognitz, Nikki MacLaren, Andrea Rosano, Stephanie A. Murphy, Andre Villegas, et al. 2014. "Multi-Laboratory Evaluation of the Rapid Genoserotyping Array (SGSA) for the Identification of Salmonella Serovars." *Diagnostic Microbiology and Infectious Disease* 80 (3): 185–90. https://doi.org/10.1016/j.diagmicrobio.2014.08.006.

Yoshida, Catherine, Gurnik Simone, Aaminah Ahmad, Travis Blimkie, Stephanie A. Murphy, Andrew M. Kropinski, and John H.E. Nash. 2016. "Evaluation of Molecular Methods for Identification Of." *Journal of Clinical Microbiology* 54 (8): 1992–98. https://doi.org/10.1128/JCM.00262-16.Editor.

Zhai, Ligong, Qian Yu, Xiaomei Bie, Zhaoxin Lu, Fengxia Lv, Chong Zhang, Xiaohan Kong, and Haizhen Zhao. 2014. "Development of a PCR Test System for Specific Detection of Salmonella Paratyphi B in Foods." *FEMS Microbiology Letters* 355 (1): 83–89. https://doi.org/10.1111/1574-6968.12443.

Zhang, Shaokang, Yanlong Yin, Marcus B. Jones, Zhenzhen Zhang, Brooke L.Deatherage Kaiser, Blake A. Dinsmore, Collette Fitzgerald, Patricia I. Fields, and Xiangyu Deng. 2015. “Salmonella Serotype Determination Utilizing High-Throughput Genome Sequencing Data.” *Journal of Clinical Microbiology* 53 (5): 1685–92. https://doi.org/10.1128/JCM.00323-15.

Zou, Wen, Wei Jiun Lin, Steven L. Foley, Chun Houh Chen, Rajesh Nayak, and James J. Chen. 2010. “Evaluation of Pulsed-Field Gel Electrophoresis Profiles for Identification of Salmonella Serotypes.” *Journal of Clinical Microbiology* 48 (9): 3122–26. https://doi.org/10.1128/JCM.00645-10.

CURRICULUM VITAE

Mathieu Gand was born on August 19th, 1989, in Metz, France. In 2007, he enrolled at the IUT Louis Pasteur, part of the Université de Strasbourg (Unistra) in France, where he obtained a DUT's degree in Food and Biology Industry Engineering. During his DUT, he worked 2 months at the laboratory of the Lactalis Nestlé Ultra Frais factory in Sarrebourg (France). During the same period, he also worked, first as an intern and then as an employee, 2 months at the Biomedical Analysis Laboratory Kandel in Sarrebourg (France) and 5 months at the Institut des Neurosciences Cellulaires et Intégratives (INCI) - CNRS UPR 3212 in Strasbourg (France). From 2010 to 2013, he studied in France at the Université de Montpellier (UM) and obtained a bachelor's degree in Microbiology, a 1st year master's degree in Microorganisms-Host-Environment Interactions and a 2nd year master's degree in Nutrition, Agro-valorization and Food Safety. During his Master, he did two 6 months internships, one at the Centre International de la Recherche Agronomique pour le Développement (CIRAD) – UMR95 QualiSud in Montpellier (France) and the other at Aérial – Centre de Recherche Technologique (CRT) in Strasbourg (France). Since then, he worked 2 months as a school assistant at the Collège de Hartzviller in Hartzviller (France) and 19 months as a junior scientist at the Department of Sustainable Food Science of the Université de Liège. In November 2015 he started on his PhD project at Sciensano in collaboration with Dr. Sophie Bertrand (FOD Volksgezondheid), Dr. Sigrid De Keersmaecker (Sciensano) and Prof. Kathleen Marchal (UGent). Through this PhD project, he was first author of 2 peer-reviewed publications and co-author of one national publication. One additional manuscript has been submitted for publication, and another is ready to be submitted after acceptance of the latter one. He also participated to 3 international and 3 national congresses.

Courses and training followed:

- How to develop a qPCR (genotyping) assay? One day external training given by Joris Verheyde from Applied Biosystems / ThermoFisher Scientific. 08/05/2018, Sciensano, Brussels.
- Sciensano's internal training: Galaxy. 10/02/2017, Sciensano, Brussels
- Sciensano's internal training: Galaxy. Luminex and MagPix training (including probe design using VisualOMP). Sciensano Internal training 31/12/2015

Attendance to symposium and conferences:

- *Symposium Santé Animal "De l'œuf à la poule"*, 23/10/2018, Brussels. Oral communication: "How to save money and time for *Salmonella* serotype identification in Belgium: a new, fast and accurate genoserotyping system"
- *6th Congress of the European Association of Veterinary Laboratory Diagnosticians (EAVLD)* 15 to 17/10/2018, Brussels. Poster: "A molecular method as replacement for classical serotyping of the most common *Salmonella* from pork and poultry sectors in Belgium".
- *23rd Conference on Food Microbiology (BSFM)*, 4 and 5/10/2018, Brussels. Poster presentation (**3rd award of the poster competition**): "How to save money and time for *Salmonella* serotype identification in Belgium: a new, fast and accurate genoserotyping system".
- *International Symposium Salmonella and Salmonellosis (I3S)*, 24 to 26/09/2018, Saint-Malo. Oral communication: "Combination of multiplex oligonucleotide ligation-PCR methods using a liquid bead suspension array for the genoserotyping of the most common *Salmonella* in Belgium"
- *28th European Congress of Clinical Microbiology & Infectious Diseases (ECCMID)*, 21 to 24/04/2018, Madrid. Oral & Poster:" A multiplex oligonucleotide ligation-PCR method using the Luminex technology for the genoserotyping of the most common *Salmonella* in Belgium"
- *22nd Conference on Food Microbiology (BSFM)*, 21 and 22/09/2017, Brussels. Poster presentation (**1st award of the poster competition**): Development of a real-time PCR test for the genoserotyping of *Salmonella* Paratyphi B and its variant Java.

Presentations of the work of and by the PhD student (in addition to the above mentioned presentations at (inter)national conferences):

- Scientific seminar of the Bacterial Diseases service of Sciensano: presentation of the *Salmonella* genoserotyping Luminex method which will be implemented in the service. 18/10/18, Sciensano, Brussels.
- Technical and practical presentation of the *Salmonella* genoserotyping Luminex method to AFSCA delegation in view to encourage law adaptation for molecular typing in the food and veterinary sector. 18/09/2018, Sciensano, Brussels.
- 3rd SALMSTID steering committee: Presentation of the yearly results (3rd year) and due tasks according to the work packages of the project. 07/09/2018, FOD Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu, Brussels.
- Training given to members of DGZ and Arsia about the Luminex technology for genoserotyping of *Salmonella* in view to implement the method in their laboratories. 19/07/2018, Sciensano, Brussels.

- Scientific seminar of the Transversal activities in Applied Genomics service of Sciensano: presentation of the PhD work. 31/05/2018, Sciensano, Brussels.
- 2nd SALMSTID steering committee: presentation of the yearly results (2nd year) and due tasks according to the work packages of the project. 29/09/2018, FOD Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu, Brussels.
- 1st SALMSTID steering committee: presentation of the yearly results (1st year) and due tasks according to the work packages of the project. 23/09/2016, FOD Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu, Brussels.

Involvement in training of young scientists, students and technicians

- How to select genetic markers, using bioinformatics tools, for the development of a molecular method? Case study: specific detection of *Salmonella enterica* subsp. *enterica* ser. Enteritidis. Applied course given to 3rd year bachelor student at UGent. 21/03/2019
- Training of the laboratory technicians of Bacterial Diseases service, to the Luminex method, in view to implement the genoserotyping system at the Belgian NRC. October 2018

Involvement in organizational tasks in the laboratory/department/faculty/university

- Redaction of standard operating procedures
- Maintenance and management of the MagPix
- Implementation of the Luminex *Salmonella* genoserotyping assay at the Belgian NRC and accreditation following ISO 17025 and ISO15189

List of publications:

- **Gand M,** Mattheus W, Roosens N, Dierick K, Marchal K, Bertrand S*, De Keersmaecker S* (2020) "A genoserotyping system for a fast and objective identification of *Salmonella* serotypes commonly isolated from poultry and pork food sectors" *Food Microbiology* 91 103534 https://doi.org/10.1016/j.fm.2020.103534 (* equal contribution)
- **Gand M,** Mattheus W, Roosens N, Dierick K, Marchal K, De Keersmaecker S*, Bertrand S* (2020) "A multiplex oligonucleotide ligation-PCR method for the genoserotyping of common *Salmonella* using a liquid bead suspension assay" *Food Micro* 87 103394 https://doi.org/10.1016/j.fm.2019.103394 (* equal contribution)
- **Gand M,** Mattheus W, Saltykova A, Roosens N, Dierick K, Marchal K, De Keersmaecker S* et Bertrand S* (2019) "Development of a real-time PCR test for the genoserotyping of *Salmonella* Paratyphi B and its variant Java" *Appl Microbiol Biotechnol* 103(12):4987-4996 (* equal contribution)
- Bertrand S, Mattheus W, Ceyssens PJ, **Gand M**, Vanhoof R, Botteldoorn N, Denayer S, Roosens N et De Keersmaecker S. (2017). "*Salmonella* infections: identification techniques for successful investigations" *Labinfo* 16, 41-47

- Cauchie E., **Gand M.,** Kergourlay G., Taminiau B., Delhalle L., Korsak N. et Daube G. (2015) "The use of 16S rDNA metagenetic monitoring of refrigerated food products for understanding the kinetics of microbial subpopulations at different storage temperatures: the example of white pudding" *International Journal of Food Microbiology* 247:70-78
- Croisé P., Houy S., **Gand M.**, Lanoix J., Calco V., Tóth P., Brunaud L., Lomazzi S., Paramithiotis E., Chelsky D., Ory S et Gasman S. (2015) " Cdc42 and Rac1 activity is reduced in human pheochromocytoma and correlates with FARP1 and ARHGEF1 expression " *Endocrine-Related Cancer* 23(4):281-93
- **Gand M** et al. "Evaluation of a genoserotyping system for the serotype identification of auto-agglutinable *Salmonella* isolates" Ready for submission

ACKNOWLEDGEMENT

Ce travail de thèse a été une grande aventure professionnelle et personnelle ! Mais je n'aurais jamais pu mener ce projet à bien sans l'aide de personnes inestimables que je souhaiterais remercier ici.

Je tiens tout d'abord chaleureusement à remercier mes promotrices de thèse : Sophie Bertrand, Sigrid De Keersmaecker et Kathleen Marchal.

Sophie, tu étais l'investigatrice principale du projet SALMSTID, sur lequel est basé ce travail de thèse, et je te remercie de m'avoir fait confiance pour le mener à bien. Je me souviendrai toujours de mon entretien d'embauche durant lequel tu as su voir mon potentiel et m'as donné ma chance. Pour cela, je ne te remercierai jamais assez. Tu es toujours restée disponible afin de m'apporter ton aide et ton expertise inestimable dans le domaine des Salmonelles, dont je n'aurais pu me passer.

Sigrid, complémentaire à Sophie, tu m'as apporté toutes les connaissances et l'expertise dont j'avais besoin dans le domaine de la biologie moléculaire. Ton aide a été sans pareille dans chacune des étapes de mon doctorat, que ce soit sur le plan scientifique ou administratif. Toujours présente et disponible (soirs et week-end inclus !) pour un suivi méticuleux de mon travail, tu m'as prodigué de précieux conseils lors de la rédaction d'articles scientifiques et de mon manuscrit de thèse, afin que je puisse sans cesse m'améliorer et devenir le scientifique que je suis aujourd'hui. Toujours guidée par un souci de perfection, disons-le franchement, ton investissement dans ce doctorat a été considérable !

Kathleen, je te remercie pour ton aide et tes conseils, ainsi que pour l'expérience enrichissante que tu m'as offerte en me permettant de donner un cours à tes étudiants.

Vous avez, toutes les trois, grandement contribué à la réussite de ce travail !

J'aimerais aussi remercier les membres de mon jury, Prof. Bart Devreese, Dr. Filip Boyen, Prof. Marc Heyndrickx, Dr. Nancy Roosens et Dr. François-Xavier Weill, d'avoir accepté d'évaluer mon travail et d'y dédier une partie de votre temps. Merci pour vos remarques et suggestions pertinentes, ainsi que pour la richesse de nos échanges, qui ont permis d'améliorer mon travail.

Je souhaiterais également remercier toutes les personnes du service des Maladies Bactériennes qui, de près ou de loin, m'ont aidé dans la réussite de ce doctorat. Je suis tout particulièrement reconnaissant à Zahra Boukhouchi et Maïté Boutry pour leur accueil dans le service, leur aide au laboratoire et les bons moments passés ensemble, ponctués de nombreux fous rire ! Je tiens également à remercier les autres techniciennes du CNR pour les nombreux sérotypages par agglutination sur lame qu'elles ont effectués pour moi. Enfin, j'aimerais remercier Tatiana Theeten, Veerle Van Melle, Kévin Charlier, Mehdi Kiass, Jonathan Reniers, Vanessa Mathys et Eugenia Pereira Gonçalves pour leur accueil à Uccle grâce à qui le déménagement du service n'a pas été si terrible que ça !

Je voudrais également remercier tous les membres du service TAG de m'avoir immédiatement adopté comme si j'étais l'un des leurs et de la bonne ambiance dans le service. Je remercie les techniciens de m'avoir aidé à « dompter » le MagPix ainsi que les bio informaticiens pour leur aide dans mes analyses génomiques, et tout particulièrement Assia Saltykova qui, avec ses lignes de commande magiques, m'a souvent aidé à sortir de l'impasse ! Enfin, je voudrais remercier Nancy Roosens de m'avoir toujours considéré comme un membre à part entière de TAG et de s'être toujours démenée pour me garder à l'Institut, même une fois mon projet de thèse terminé.

Une pensée particulière va à Marie-Alice Fraiture, Sarah Nauwelaerts et Assia Saltykova. Que ce soit pour des discussions professionnelles, des after-works endiablés ou, tout simplement, une oreille attentive à qui se confier, j'ai adoré tous les moments passés ensemble, et j'espère qu'il y en aura beaucoup d'autres !

Toujours présents pour moi, dans les meilleurs moments comme dans les pires, j'aimerais sincèrement remercier mes amis et mes proches, et tout spécialement Giovanni Vazzano et Jérôme Bellot pour tout le soutien moral qu'ils m'ont apporté durant ces années de thèse. Merci infiniment à Nicolas Gigant d'avoir enduré mes moments de stress, de fatigue, mes sautes d'humeur (le tout en étant en partie confiné !), m'encourageant sans cesse et me montrant ta fierté.

Enfin, je tenais également à remercier mes parents, Nadine et Didier Gand, qui m'ont donné la chance de faire des études longues et ont cru en ma réussite, ainsi que le reste de ma famille, pour leurs encouragements durant tout ce parcours, avec une pensée toute particulière pour Claire Dubourgnoux et Mélanie Lecacheur.

Merci à vous tous,

Mathieu

FACULTY OF SCIENCES
DEPARTMENT OF PLANT BIOTECHNOLOGY AND BIOINFORMATICS
Krijgslaan 281 - S2
9000 GHENT, BELGIUM
mathieu.gand@ugent.be
https://www.ugent.be/we