KU LEUVEN

ARENBERG DOCTORAL SCHOOL
FACULTY OF BIOSCIENCE ENGINEERING

Development of a molecular alternative for classical microbiological subtyping methods

Salmonella phage typing as a case study

Véronique WUYTS

Dissertation presented in partial
fulfilment of the requirements for
the degree of Doctor
in Bioscience Engineering (KU Leuven)
and of Science: Bioinformatics (UGent)

October 2015

Development of a molecular alternative for classical microbiological subtyping methods

Salmonella phage typing as a case study

Véronique WUYTS

Supervisors:
Prof. K. Marchal, UGent
Prof. J. Vanderleyden, KU Leuven
Dr. S. De Keersmaecker, WIV-ISP
Dr. S. Bertrand, WIV-ISP

Members of the Examination Committee:
Prof. D. Berckmans, chair, KU Leuven
Dr. I. Bergval, KIT, The Netherlands
Prof. J. Fostier, UGent
Prof. R. Jelier, KU Leuven
Dr. N. Roosens, WIV-ISP
Prof. D. Springael, KU Leuven

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Bioscience Engineering (KU Leuven) and of Science: Bioinformatics (UGent)

October 2015

Doctoraatsproefschrift nr. 1294 aan de faculteit Bio-ingenieurswetenschappen van de KU Leuven

Onderzoek gefinancierd met beurs P4044.0103 (SalMolType) van het Wetenschappelijk Instituut Volksgezondheid (WIV-ISP - RP/PJ)

Acknowledgements

In augustus 2011 kreeg ik een e-mail van Kathleen: of ik geen zin had om een doctoraat te maken bij Sigrid. Ik zou wel moeten terugkeren naar het labo en het was in Brussel te doen. De samenvatting van het project sprak over infectieziekten, subtypering, *Salmonella*, genoomsequenties, ... Ok, dit interesseerde me wel, dus ging ik naar Brussel (WIV-ISP, nog nooit van gehoord) om te solliciteren. Daar begon Sophie ineens Frans tegen mij te spreken, o oh. En nu ben ik de laatste woorden aan't schrijven van een 4 jaar lange uitdaging. Dit is dan ook het moment om een speciaal woordje te richten aan de personen die dit werk mee mogelijk gemaakt hebben.

In de eerste plaats wil ik mijn promotoren bedanken.
Sigrid, rots in de branding, steun en toeverlaat, bedankt voor de kansen die je me gegeven hebt in de R&D-ploeg van PBB, om je kennis en ervaring te delen, voor je kritische opmerkingen op mijn schrijfsels en je motiverende woorden. Ik heb veel geleerd en ik apprecieer je vertrouwen.
Kathleen, bedankt om me de mogelijkheid te geven om dit doctoraat te maken. Wetenschappelijke discussies met jou blijven plezant en inspirerend, er komt altijd wel iets nieuws uit.
Sophie, merci pour me soumettre dans le monde miraculeux du (sou)typage de *Salmonella* et pour m'accueillir dans ton labo. Il n'a pas été autant facile, mais je suis contente que nous avons atteint la ligne d'arrivée ensemble.
Jos, je was er van in het begin bij, doch slechts een korte periode als promotor, maar ik wil je bedanken om mij op te vangen in Leuven.

Ik zou graag de voorzitter, Prof. Berckmans, en de leden van mijn examencommissie, Dr. Bergval, Prof. Fostier, Prof. Jelier, Dr. Roosens en Prof. Spingael, bedanken voor hun tijd die ze gestoken hebben in de kritische beoordeling van deze thesis en voor de samenwerkingen tijdens mijn doctoraat. Jullie interessante vragen, opmerkingen en suggesties worden zeer gewaardeerd.

A special word of thanks goes to Nancy for her inexhaustible scientific interest in the SalMolType project, to my lab- and office-mates Elodie, Latifa, Marie-Alice, Xavier, Aurélie, Sam, Sander, Axel, Céline and Linda, and to my other colleagues of PBB, Sylvia, Nina, Els, Dirk, Maud, Loïc, Joost, Kevin and Qiang, for all the good advice, support, listening ears, pep talks, amusing chattering, French courses and the memorable PBB lunches. It is a privilege to be part of this team and you are 'een pakske goed bezig'.

Ook wil ik een dankjulliewel richten aan mijn collega's van de dienst Bacteriële Ziekten, Wesley, Dina, Francis, Marie, Gabriella, Hilde, Christa, Gerald, Joelle, Zahra en Maïté, om mij op sleeptouw te nemen in jullie labo en voor de goede babbels op tijd en stond.

Aan Katelijne en Sarah van de dienst Voedselpathogenen, een bedankt om aan sequentie-analyse te denken bij de interessante *Salmonella*-uitbraken.

Woorden van dank gaan ook naar het CPMG, speciaal naar David voor het *Salmonella* labostammen DNA, naar Toon en Maarten voor hun hulp bij het sequencen en naar Sandra (nu ITG) om mijn congres-maatje te zijn.

Giles and Mahdi at the UGent, thank you for listening to my SalMolType talk and for providing the Brownie corrected sequencing reads (I still have an appetite for chocolate while writing this).

Mijn vriendinnen en vrienden wil ik danken voor de ontspannende 'iets drinken', 'iets eten', moto-ritjes en weekendjes weg. Sorry, dat ik niet altijd tijd maakte, daarin beloof ik verbetering.

Ma en Pa, jullie hebben me altijd mijn eigen weg laten zoeken en me hierin gesteund. In de eerste plaats hebben jullie dit werk mogelijk gemaakt. Sas en Lie, ik ben blij dat ik twee grote zussen heb op wie ik kan rekenen, en bedankt, ook Jan en Kurt, voor de deugddoende vakanties samen. Michiel en Freek, dank je voor jullie interesse en ontspannende tantekes-weekendjes. Voor de volgende thesis is het aan jullie!

Bert, zonder jou was me dit niet gelukt. Bedankt voor alles!

Nik

Dit doctoraatsproefschrift is opgedragen aan Paul Wuyts

(23 mei 1956 – 29 september 2012)

Abstract

Characterisation of pathogens below the subspecies or serovar level, *i.e.* subtyping, is essential in public health for routine surveillance of pathogens, outbreak detection and timely confinement of an outbreak. An ideal subtyping method is rapid, highly discriminative, inexpensive and robust. The classical subtyping methods for the bacterium *Salmonella*, a major foodborne pathogen, are phage typing, multiple-locus variable-number of tandem repeats analysis (MLVA) and pulsed-field gel electrophoresis (PFGE). Each of these techniques has however its intrinsic disadvantages.

In the main part of this PhD, a molecular alternative for the subjective phage typing method was developed. *Salmonella enterica* subsp. *enterica* serovar Typhimurium was taken as a case study, because this is the most frequent *Salmonella* serovar in Belgium. In addition, it was shown by a serial passage experiment that 3 of the 5 the MLVA loci in the *S.* Typhimurium scheme are unstable, which especially complicates outbreaks investigations.

For the time-efficiency of the alternative subtyping assay, several molecular markers had to be combined in a single assay. Therefore, the Luminex technology, which allows multiplexing by microsphere suspension arrays, was implemented in the Scientific Institute of Public Health (WIV-ISP) during this PhD research. The multiplex oligonucleotide ligation-PCR (MOL-PCR) assay technique was selected for the novel method for subtyping of *S.* Typhimurium, since this is a sufficiently fast and relatively inexpensive technique.

The thorough optimisation of the MOL-PCR assay indicated that the DNA isolation, the probe concentration, the amount of microspheres and the concentration of reporter dye are the parameters that have a major influence on the MOL-PCR performance.

The final developed MOL-PCR assay for subtyping of *S.* Typhimurium and *S.* 1,4,[5],12:i:- combines 52 molecular markers, including prophage genes, amplified fragment length polymorphism (AFLP) elements, *Salmonella* genomic island 1 (SGI1), allantoinase gene *allB*, MLVA locus STTR10, antibiotic resistance genes, single nucleotide polymorphisms (SNPs) and phase 2 flagellar

gene *fljB*. The method proved to have a discriminatory power equal to that of phage typing and could deliver results in less than 8 hours, making it suitable for outbreak investigations. Additionally, an R application was created that allows an objective interpretation of the data and summarises the results in a MOL-PCR profile.

During this PhD, whole genome sequencing (WGS) has gained the attention of public health institutes as the ultimate universal subtyping technique for surveillance of pathogens and for outbreak investigations. As such, this technique was explored in the second part of this work.

Four issues associated to pathogen surveillance with WGS were elaborated using data of 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates. The first issue relates to the workflow to be followed. For a SNP-based workflow, we showed that the applied tools, the used reference genome and error correction have an impact on the resulting phylogenetic tree. SNP-based analysis of WGS data proved to have more discriminatory power than the classical subtyping methods, which is especially of interest for isolates with frequently occurring subtypes. The alternative gene-based workflow, also referred to as whole genome multilocus sequence typing (MLST), could not be explored, as currently there is no whole genome MLST scheme available for *Salmonella*. The second issue concerns the maximum number of SNPs between two isolates before they are regarded as distinct subtypes. This number was again highly dependent on the applied tools. Existing web-based tools were used to look into the third and fourth issue, namely the inference of phenotypic characteristics from WGS data and the link between historical subtyping data and WGS data, which showed to be possible, albeit very limited.

As a case study to show the 'universal' applicability of WGS, data of *S.* Enteritidis isolates related to two outbreaks were used, as for each of these outbreaks both food and human isolates were available. The WGS analysis with existing user-friendly tools confirmed the initial outbreak investigation with classical subtyping methods. In addition, the WGS data analysis could discriminate between the source of the two outbreaks, which was not possible with the classical subtyping methods.

To conclude, the developed MOL-PCR assay for *S.* Typhimurium and *S.* 1,4,[5],12:i:- is a multiplex, molecular subtyping method that has the same discriminatory power as phage typing. For surveillance and outbreak investigation, WGS can currently be applied to complement the existing subtyping methods for increased discrimination between isolates. Before WGS can be implemented as a replacement of the existing subtyping methods, further optimisation of the data analysis tools is however required.

Samenvatting

Karakterisering van ziekteverwekkers tot onder het subspecies of serovar niveau, *i.e.* subtypering, is van essentieel belang in de volksgezondheid voor surveillance van ziekteverwekkers, detectie van uitbraken en tijdige inperking van uitbraken. Een ideale subtyperingsmethode is snel, zeer discriminerend, goedkoop en robuust. De klassieke subtyperingsmethoden voor de *Salmonella* bacterie, een belangrijk voedselpathogeen, zijn faagtypering, 'multiple-locus variable-number of tandem repeats analysis' (MLVA) en pulsed-field gelelektroforese (PFGE). Elk van deze technieken heeft echter zijn inherente nadelen.

De ontwikkeling van een moleculair alternatief voor de subjectieve faagtyperingsmethode vormde het voornaamste deel van dit doctoraat. Als casestudy werd *Salmonella enterica* subsp. *enterica* serovar Typhimurium genomen, omdat dit de meest voorkomende serovar in België is. Bovendien werd door een overentingsexperiment aangetoond dat 3 van 5 loci in het MLVA-schema van *S.* Typhimurium instabiel zijn, wat vooral uitbraakonderzoeken met deze subtyperingsmethode bemoeilijkt.

Voor een tijdsefficiënte alternatieve subtyperingsmethode moesten meerdere moleculaire merkers gecombineerd worden in één assay. Daarom werd tijdens dit doctoraatsonderzoek de Luminex technologie, die multiplexing mogelijk maakt met microsferen suspensie arrays, geïmplementeerd in het Wetenschappelijk Instituut Volksgezondheid (WIV-ISP). Als techniek werd de multiplex oligonucleotide ligatie-PCR (MOL-PCR) geselecteerd, omdat deze voldoende snel en relatief goedkoop is.

De optimalisatie van de MOL-PCR assay gaf aan dat de manier van isolatie van het DNA, de hoeveelheid microsferen en de concentratie van de probes en de reporter kleurstof een grote invloed hadden op de performantie van de MOL-PCR assay.

In de uiteindelijke MOL-PCR assay voor subtypering van *S.* Typhimurium en *S.* 1,4,[5],12:i:- worden 52 moleculaire merkers gecombineerd, waaronder profaag genen, 'amplified fragment length polymorphism' (AFLP) elementen, *Salmonella* genomisch eiland 1 (SGI1), het allantoïnase gen *allB*, de MLVA

locus STTR10, antibioticaresistentie genen, 'single nucleotide polymorphisms' (SNPs) en fase 2 flagellair gen *fljB*. De methode bleek hetzelfde discriminerend vermogen te hebben als faagtypering en resultaten werden bekomen binnen de 8 uren, waardoor deze geschikt is voor uitbraakonderzoek. Bovendien werd een R applicatie gecreëerd die een objectieve interpretatie van de data toelaat en die de resultaten samenvat in een MOL-PCR profiel dat uitwisseling van resultaten vergemakkelijkt.
Tijdens dit doctoraat trok 'whole genome sequencing' (WGS) de aandacht van volksgezondheidsinstituten als ultieme universele subtyperingstechniek voor surveillance van ziekteverwekkers en uitbraakonderzoeken. Daarom werd deze techniek verkend in het tweede deel van dit werk.
Vier items in verband met WGS voor surveillance van pathogenen werden uitgewerkt met data van 32 *S.* Typhimurium en *S.* 1,4,[5],12:i:- isolaten. Betreffende de workflow hebben we aangetoond dat de toegepaste tools, de gebruikte referentiegenomen en foutcorrectie een effect hebben op de fylogenetische boom die resulteert uit een op SNPs gebaseerde workflow. SNP-analyse van WGS data bleek beter te discrimineren dan de klassieke subtyperingsmethoden, wat interessant is voor isolaten met veelvoorkomende subtypes. De alternatieve op genen gebaseerde workflow, ook 'whole genome' MLST genoemd, kon niet verkend worden, daar er momenteel geen 'whole genome' MLST schema bestaat voor *Salmonella*. Het tweede item handelt over het maximaal aantal SNPs tussen twee isolaten alvorens deze als verschillende subtypes beschouwd worden. Dit aantal was opnieuw sterk afhankelijk van de gebruikte tools. Bestaande webtoepassingen werden aangewend om meer inzicht te krijgen in het derde en vierde item, namelijk de afleiding van fenotypische kenmerken uit WGS data en het verband tussen historische subtyperingsdata en WGS data. Dit bleek mogelijk, zij het zeer gelimiteerd.
Als casestudy om de 'universele' toepasbaarheid van WGS aan te tonen, werden data van *S.* Enteritidis isolaten van 2 uitbraken gebruikt, omdat voor beide uitbraken zowel voedsel- als humane isolaten beschikbaar waren. De WGS analyse met bestaande gebruiksvriendelijke tools, wat een must is voor een Nationaal Referentie Laboratorium, bevestigde het initiële uitbraakonderzoek met klassieke subtyperingsmethoden. Bovendien kon met WGS analyse een onderscheid gemaakt worden tussen de bron van elk van de uitbraken, wat niet mogelijk was met de klassieke subtyperingsmethoden.
We kunnen besluiten dat de ontwikkelde MOL-PCR assay voor *S.* Typhimurium en *S.* 1,4,[5],12:i:- een multiplex, moleculaire subtyperingsmethode is die hetzelfde discriminerend vermogen heeft dan faagtypering. WGS kan momenteel gebruikt worden als complementaire subtyperingsmethode voor surveillance en uitbraakonderzoek voor een betere discriminatie tussen isolaten. Vooraleer WGS kan toegepast worden als vervanging van de bestaande subtyperingsmethoden, is echter verdere optimalisatie van de data analyse vereist.

Abbreviations

D	Simpson's index of diversity
E	Shannon's index of equitability
H'	Shannon's index of diversity
T	typeability
S. Enteritidis	*Salmonella enterica* subsp. *enterica* serovar Enteritidis
S. Typhi	*Salmonella enterica* subsp. *enterica* serovar Typhi
S. Typhimurium	*Salmonella enterica* subsp. *enterica* serovar Typhimurium
S. 1,4,[5],12:i:-	*Salmonella enterica* subsp. *enterica* serovar 1,4,[5],12:i:-
A	ampicillin
A/S	ampicilin/sulbactam
AFLP	amplified fragment length polymorphism
Amc	amoxicillin plus clavulanic acid
Ami	amikacin
AMR	antimicrobial resistances
ASPE	allele-specific primer extension
Axo	ceftriaxone
Azt	aztreonam
BAM	binary alignment/map
BED	browser extensible data
BHI	brain-heart infusion
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BRIG	BLAST Ring Image Generator
BWA	Burrows-Wheeler alignment

C	chloramphenicol
CCD	charge-coupled device
Cep	cephalothin
cgMLST	core genome MLST
Cip	ciprofloxacin
CLSI	Clinical and Laboratory Standards Institute
CRISPOL	CRISPR polymorphism
CRISPR	clustered regularly interspaced short palindromic repeats
CSI Phylogeny	Call SNPs & Infer Phylogeny
csv	comma-separated values
Ctx	cefotaxime
dCTP	deoxycytidine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTU	Technical University of Denmark (Danmarks Tekniske Universitet)
ECDC	European Centre for Disease Prevention and Control
EDTA	ethylenediaminetetraacetic acid
Eff ΔG°	effective change of Gibb's free energy
EFSA	European Food Safety Authority
ELISA	enzyme-linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
Etp	ertapenem
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FASFC	Federal Agency for the Safety of the Food Chain
Faz	cefazolin
FDA	U.S. Food and Drug Administration
Fep	cefepime
Fox	cefoxitin
Fur	cefuroxime
G	gentamicin
GATK	Genome Analyis Toolkit
gDNA	genomic DNA
GPP	Gödel Prime Product

GQ	genotype quality
GUI	graphical user interface
HEK	hektoen enteric
HPLC	high performance liquid chromatography
IGM	InstaGene Matrix
ISO	International Organization for Standardization
K	kanamycin
kb	kilobase
LB	Luria-Bertani
LED	light-emitting diode
MAC	MacConkey
MB	megabyte
MDR	multidrug resistance
MEGA	Molecular Evolutionary Genetics Analysis
Mero	meropenem
MFI	median fluorescence intensity
MIC	minimal inhibitory concentration
MLPA	multiplex ligation-dependent probe amplification
MLST	multilocus sequence typing
MLVA	multiple-locus variable-number of tandem repeats analysis
MOL-PCR	multiplex oligonucleotide ligation-PCR
mPCR	multiplex PCR
MQ	mapping quality
MUSCLE	multiple sequence comparison by log-expectation
Na	nalidixic acid
NCBI	National Center for Biotechnology Information
ND	not determined
NEG CON	negative control
NGS	next generation sequencing
NRC	National Reference Centre
NRCSS	National Reference Centre for *Salmonella* and *Shigella*
NRL	National Reference Laboratory
NRL-FBO	National Reference Laboratory of Foodborne Outbreaks

NT	not-typable
NTC	no-template-control
OLA	oligo ligation assay
P/T4	piperacillin/tazobactam constant 4
PacBio	Pacific Bioscience
PBB	Platform Biotechnology and Molecular Biology
PC	positive control
PCR	polymerase chain reaction
pdf	portable document format
PFGE	pulsed-field gel electrophoresis
Pod	cefpodoxime
POS CON	positive control
RDNC	reacts-but-does-not-conform
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RP	reverse-phase
S	streptomycin
SAM	sequence alignment/map
SAPE	streptavidin-R-phycoerythrin
SGI1	*Salmonella* genomic island 1
SMRT	single molecule real time
SN	signal-to-noise ratio
SNP	single nucleotide polymorphism
SOP	standard operating procedure
ST	sequence type
Su	sulphonamides
Sxt	trimethoprim plus sulfamethoxazole
T	tetracycline
T_m	melting temperature
Taz	ceftazidime
Tgc	tigecycline
Tim2	ticarcillin/clavulanic acid constant 2
TMAC	tetramethylammonium chloride
Tmp	trimethoprim
Tob	tobramycin
TR	tandem repeat

UPGMA	unweighted pair group method using arithmetic averages
vcf	variant call format
VNTR	variable number tandem repeat
WGS	whole genome sequencing
WHO	World Health Organization
WIV-ISP	Scientific Institute of Public Health (Wetenschappelijk Instituut Volksgezondheid - Institut Scientifique de Santé Publique)
WT	wild-type
XLD	xylose-lysine-deoxylate
xMAP	MAP: multi-analyte profiling, x: the unknown

Contents

List of figures

List of tables

Chapter 1

General introduction, objectives and outline

1.1 Subtyping for public health

Infectious diseases, whether caused by bacteria, viruses, fungi or parasites, pose a major healthcare concern and are therefore surveyed, including in Belgium, by National Reference Laboratories and Centres, which are part of an international network through the European Centre for Disease Prevention and Control (ECDC) and the World Health Organization (WHO). The primary missions of the National Reference Laboratories and Centres are the diagnosis of pathogens, routine surveillance to study the incidence and dissemination of pathogens over time and epidemiological studies for fast detection of outbreaks and tracing their source. A second important task is monitoring of antimicrobial susceptibility to assist treatment of the infection, to determine resistance mechanisms to existing antimicrobials or to investigate new therapeutic molecules. Other responsibilities are conducting scientific research *e.g.* for improving diagnostic tests or prevention and treatment methods, concentration and sharing of expertise *e.g.* by publications or participations in scientific conferences, and maintaining a collection of interesting strains.
For routine surveillance and detection and trace-back of outbreaks, pathogens have to be characterised at a level suitable for fingerprinting the strain, which means for most pathogens that subtyping, *i.e.* characterisation below the subspecies or serovar level, is required. Ideally, such a subtyping method is rapid, inexpensive, sufficiently discriminative, robust, easily implemented and

standardised, universally applicable for a wide range of pathogens, and resulting in objective data which are convenient for interpretation and transfer between different laboratories (Wattiau *et al.* 2011; Sabat *et al.* 2013).

1.1.1 The foodborne pathogen *Salmonella*

Salmonella is a Gram-negative, motile, rod-shaped enterobacterium. The genus *Salmonella* includes two species: *S. enterica* and *S. bongori*. Based on biochemical characteristics, six subspecies can be distinguished in *S. enterica*, *i.e. enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae* and *indica*. Through a technique called serotyping, *Salmonella* is divided in different serovars based on the determination of their somatic O and flagellar H antigens present on the surface of the bacteria. Serotyping is performed by adding specific antisera to a bacterial suspension and examining the presence or absence of an agglutination reaction. The specific antigenic formula of each serovar is reported as O antigens : phase 1 flagellar H antigens : phase 2 flagellar H antigens. *E.g.*, *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) has the antigenic formula 1,4,[5],12:i:1,2, thus 1,4,[5],12 as O antigens, i as phase 1 flagellar H antigen and 1,2 as phase 2 flagellar H antigens. The underlined O antigen indicates that antigen 1 is determined by phage conversion, which means the antigen is only presented after lysogenic conversion of the *Salmonella* culture by the corresponding converting bacteriophage. The square brackets designate that antigen 5 can be present or absent in serovar Typhimurium. Besides O and H antigens, capsular antigen Vi is interestingly for diagnostics, as it gives an indication of the virulence of the isolate. The Vi antigen is generally presented by *Salmonella enterica* subsp. *enterica* serovar Typhi (*S.* Typhi), which causes, possible life-threatening, typhoid fever, but is also observed in serovars Paratyphi C and Dublin (Le Minor and Richard 1993; Morris *et al.* 2003; Grimont and Weill 2007; Gunn *et al.* 2014). According to supplement no. 48 to the White-Kauffmann-Le Minor scheme (Issenhuth-Jeanjean *et al.* 2014), the genus *Salmonella* is subdivided into a total of 2659 serovars, of which 1586 serovars reside in the subspecies *enterica*.

Salmonella is the second most commonly reported cause of zoonosis and the principal causal agent of foodborne outbreaks in Europe. Outbreaks related to consumption of eggs, pork, beef and poultry meat are frequently described (European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2015). Human *Salmonella* infection or salmonellosis leads in most cases to self-limiting gastroenteritis, with symptoms of diarrhoea and vomiting, but more serious bacteremia with presence of *Salmonella* in the blood stream and severe enteric fever have to be treated with

antibiotics (Pui *et al.* 2011). The infective dose of *Salmonella* is estimated to range from less than 10 organisms for *S.* Typhimurium up to 10^6 organisms for *Salmonella enterica* subsp. *enterica* serovar Cubana (Cruickshank and Humphrey 1987). After an acute *Salmonella* infection, a small portion of the patients ($< 5\%$) fails to clear the infection completely from their body and move to an asymptomatic *Salmonella* carrier-state. Chronic human carriers may have *Salmonella* present in their stool for more than 55 days and for *S.* Typhi, this period may be more than one year. Chronic carriage of *Salmonella* after acute salmonellosis is mostly observed with people with a weak immunity system and children, but also otherwise healthy persons have been reported as asymptomatic carriers. In addition, an asymptomatic carrier state has also been reported for people which have not experienced an acute *Salmonella* infection, but which are regularly exposed to animals, such as livestock, poultry and pets (Buchwald and Blaser 1984; Kotova *et al.* 1988; Levine and Robins-Browne 2012; Gunn *et al.* 2014). Estimations of numbers of excreted *Salmonella* from human carriers range from 10^2–10^3 organisms per gram of stool for adults up to 10^6–10^7 organisms per gram of stool for children (Cruickshank and Humphrey 1987).

1.1.2 Isolation of *Salmonella* from human samples and food samples

In Belgium, clinical laboratories isolate *Salmonella* from human samples and then send the pure isolate to the National Reference Centre for *Salmonella* and *Shigella* (NRCSS) on a voluntary basis. Most of the *Salmonella* isolated from human samples is cultured from stool, but *Salmonella* may also be isolated from blood, urine, pus, sputum or other body fluids (Bertrand *et al.* 2014). Microbiological examination of food samples is performed by the National Reference Laboratory. Food samples may be, amongst others, eggs, egg-containing products, meat, milk, vegetables, fruit, spices, food additives or candy (U.S. Food and Drug Administration (FDA) 2015a).

Detection and isolation of *Salmonella* from human stool For detection and isolation of *Salmonella*, at least 5 ml, 1 g or a walnut-sized portion of stool, preferably not contaminated with urine, is required. Swabs of the stool sample are then used to inoculate differential and selective agar media which are based on lactose-fermenting and hydrogen sulphide production, such as MacConkey (MAC) agar, hektoen enteric (HEK) agar, xylose-lysine-deoxylate (XLD) agar and *Salmonella-Shigella* agar, or an enrichment broth, such as selenite-F or

gram-negative broth. The agar media and selenite-F broth are incubated for 24 hours at 35–37°C, gram-negative broth is incubated for 6–8 hours at 35–37°C. From an enrichment broth, a subculture is made on MAC and HEK or XLD agar. After incubation of the agar plates, the morphology of the colonies is examined for typical reactions of *Salmonella*. *Salmonella* colonies are colourless or transparent on MAC and on *Salmonella-Shigella* agar, blue or green on HEK agar and red on XLD agar. The colonies may have black centres on *Salmonella-Shigella*, HEK and XLD agar. One colony with a morphology that suggests a *Salmonella* species is subcultured on Kligler Iron agar for confirmation and further confirmatory biochemical and agglutination tests (York and Rodrigues-Wong 2010).

Detection and isolation of *Salmonella* from food The general procedure to detect and isolate bacterial species from food samples consists of an enrichment step, an isolation step and ends with confirmation of the isolated bacterial species.
For detection and isolation of *Salmonella*, the enrichment step includes a non-selective pre-enrichment and a selective enrichment phase. Pre-enrichment occurs by homogenising 25 g or 25 ml of the food sample in 225 ml buffered peptone water and incubating the suspension for 16–18 hours at 37°C. For subsequent selective enrichment, 0.1 ml of the pre-enriched sample is added to 10 ml of Rappaport-Vassiliadis with soja broth, which is incubated for 24 hours at 41.5°C, and 1 ml of the pre-enriched sample is added to 10 ml of Muller-Kauffmann tetrathionate broth, which is incubated for 24 hours at 37°C. After incubation, a full inoculation loop of each enrichment is inoculated on selective XLD agar and *Salmonella* ID2 agar. After incubation for 24 hours at 37°C, one colony of each plate is subcultured on Kligler Iron agar for confirmation and further confirmatory biochemical, agglutination and Maldi-TOF mass spectometry tests. If the picked colony is negative on Kligler Iron agar, 4 other colonies are subcultured from the selective XLD agar and *Salmonella* ID2 agar plates (Botteldoorn 2015).

1.1.3 Subtyping of *Salmonella*

In 2013, the Belgian NRCSS received 2874 human *Salmonella* isolates, of which 94.1% were cultured from stool. Of these isolates, 54.1% was serotyped as *S.* Typhimurium and its monophasic variant *Salmonella enterica* subsp. *enterica* serovar 1,4,[5],12:i:- (*S.* 1,4,[5],12:i:-), which lacks phase 2 flagellar H antigens, and 20.0% as *Salmonella enterica* subsp. *enterica* serovar Enteritidis (*S.* Enteritidis) (Bertrand *et al.* 2014). These serovars are also most reported for human salmonelloses in Europe, although *S.* Enteritidis represents here

Table 1.1: Overview of main advantages and disadvantages of *Salmonella* subtyping methods used by the Belgian NRCSS.

Phage typing	MLVA	PFGE
+ Inexpensive	+ Rapid	+ Medium discriminative[a]
+ Discriminative	+ Profiles easily compared between laboratories	+ Public database available
− High level of expertise required	− Relatively expensive	− Relatively slow
− Limited reproducibility, since interpretation of lysis patterns is subjective	− Only available for small number of pathogens	− Limited reproducibility
− NT and RDNC strains	− Data analysis with commercial software	− Data analysis with commercial software
	− Too discriminative for *S.* Typhimurium	

MLVA: multiple-locus variable-number of tandem repeats analysis; NT: not-typable; PFGE: pulsed-field gel electrophoresis; RDNC: reacts-but-does-not-conform.
[a]Serovar dependent, *e.g.* not enough discriminative for *S.* Enteritidis.

39.5% of the 73627 human cases for which serotyping data were available and *S.* Typhimurium and *S.* 1,4,[5],12:i:- 28.8% (data from 2013) (European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2015).
Due to the high incidence of these serovars, subtyping techniques which allow discrimination below the serovar level are indispensable for routine surveillance, for detection of outbreaks and for finding and confirming the source of an outbreak. The ECDC supports currently multiple-locus variable-number of tandem repeats analysis (MLVA) and pulsed-field gel electrophoresis (PFGE) for characterisation of *S.* Typhimurium, *S.* 1,4,[5],12:i:- and *S.* Enteritidis, but not all European laboratories apply these techniques as routine subtyping methods. Indeed, the Belgian NRCSS uses phage typing and MLVA for routine characterisation of *S.* Typhimurium, *S.* 1,4,[5],12:i:- and *S.* Enteritidis, and, in case of outbreak situations, PFGE is additionally performed. The main advantages and disadvantages of these subtyping methods are summarised in Table 1.1.

Phage typing Phage typing is a classical, inexpensive and simple phenotypic subtyping method based on the host specificity of bacteriophages (Anderson *et al.* 1977; Threlfall and Frost 1990). A predefined set of bacteriophages is added to a culture of the *Salmonella* isolate to be characterised on a solid medium and after incubation, the lysis reactions are examined (Figure 1.1). Distinguishing between different types of lysis reactions requires extensive expertise (Figure 1.2) and even then the results remain subjective. This was illustrated by a Scandinavian outbreak investigation, in which the *S.* Typhimurium outbreak strain was reported to belong to phage type U288 in Denmark, U302 in Sweden and reacts-but-does-not-conform (RDNC) in Norway (Bruun *et al.* 2009). RDNC and not-typable (NT) isolates are the result of the limited number of bacteriophages that are applied (Wattiau *et al.* 2011). RDNC isolates give lysis reactions with the bacteriophages, but these reactions do not match any of the patterns that define a certain phage type. NT isolates do not react with any of the bacteriophages.

Multiple-locus variable-number of tandem repeats analysis With MLVA the genetic relatedness between isolates is assessed through rapidly evolving genomic loci called tandem repeats (TRs). For *S.* Typhimurium and *S.* Enteritidis, five of such loci are amplified in a multiplex PCR (mPCR) with fluorescently labelled primers. The number of repeats is then determined by estimating the length of the amplicons through capillary electrophoresis, which gives MLVA profiles displayed as a sequence of five numbers (Figure 1.3). Standardised MLVA schemes are available for *S.* Typhimurium and *S.* Enteritidis (Lindstedt *et al.* 2004; Hopkins *et al.* 2011) and these allow fast subtyping, but a major concern for *S.* Typhimurium is the stability of these MLVA loci (Wuyts *et al.* 2013; Dimovski *et al.* 2014), which hampers tracing of outbreaks (Petersen *et al.* 2011; Friesema *et al.* 2012; Garvey *et al.* 2013; Kuhn *et al.* 2013; Paranthaman *et al.* 2013).

Pulse-field gel electrophoresis PFGE is related to the restriction fragment length polymorphism (RFLP) technique and is considered as the gold standard for subtyping of *Salmonella*. A rare cutting restriction enzyme is used to digest the genomic DNA of the isolate, after which the large DNA fragments are separated by gel electrophoresis with a constantly changing electrical field. As it is important to keep the large DNA fragments intact, a culture of the isolate is captured into an agarose plug and extraction and subsequent digestion of the genomic DNA is performed inside this plug, which can be readily loaded into an agarose gel (Peters 2009). Photos of the gels are processed with specialised software, such as BioNumerics (Applied Maths) (Figure 1.4), and by comparison with the international PulseNet database (PulseNet International 2015b), a

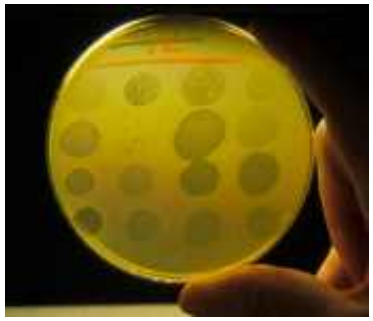

Figure 1.1: Phage typing. A set of bacteriophages is applied to a lawn of bacterial culture, after which the lysis reactions are determined.

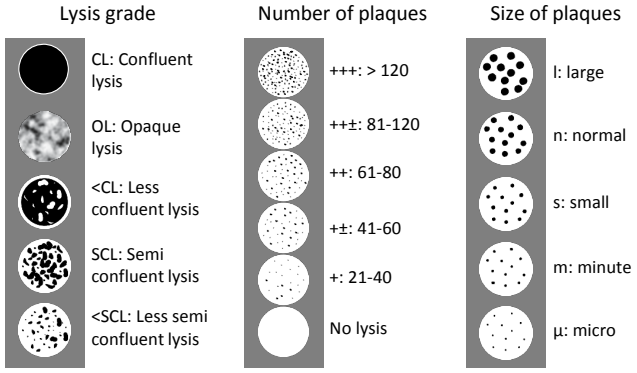

Figure 1.2: Interpretation of lysis patterns in phage typing (adapted from Rabsch (2007)).

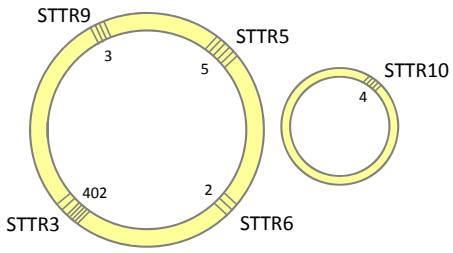

Figure 1.3: Schematic representation of multiple-locus variable-number of tandem repeats analysis for a *S.* Typhimurium isolate with profile 3-5-2-4-402. Locus STTR3 is a combination of two repeats of length 27 bp and 33 bp.

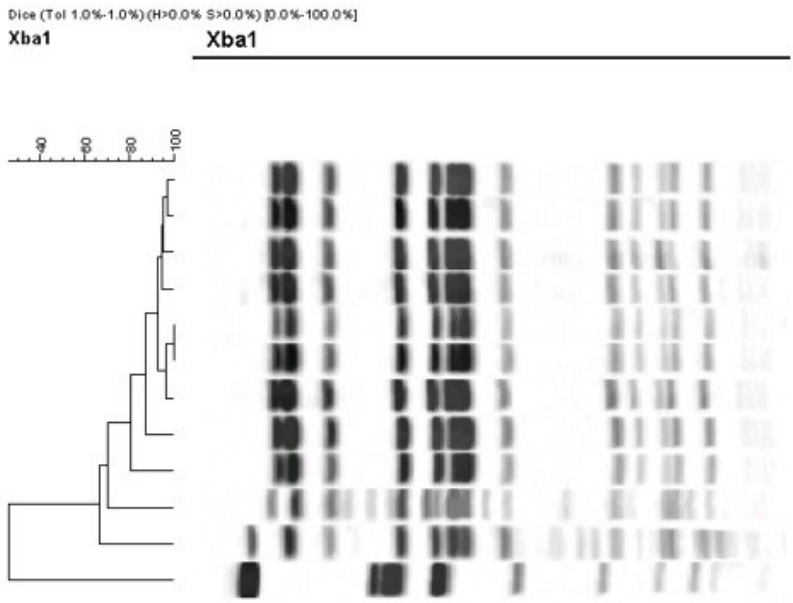

Figure 1.4: Example of processed pulsed-field gel electrophoresis patterns.

PFGE pattern can be assigned. However, a comprehensive standardisation is required for such interlaboratory comparison of PFGE patterns (Hunter *et al.* 2005; Ribot *et al.* 2006). PFGE is not routinely applied in the Belgian NRCSS for subtyping of *Salmonella*, because it is a labour-intensive and time-consuming protocol.

Other *Salmonella* subtyping methods Next to phage typing, MLVA and PFGE, also multilocus sequence typing (MLST) and clustered regularly interspaced short palindromic repeats (CRISPRs) are often reported in literature for subtyping of *Salmonella*.
In MLST, depending on the particular scheme used, a set of mostly seven housekeeping, virulence, fimbrial or flagellar genes or CRISPR loci, are amplified by PCR and Sanger sequenced. By comparison of the sequences of each locus to an allele database, a sequence type (ST) is assigned (Figure 1.5). A sequence type is a unique, arbitrary number that is associated to an allelic profile, *i.e.* a specific combination of sequences. The high cost, the prolonged protocol and the limited discriminatory power of this method preclude implementation of MLST for characterisation of *Salmonella* in routine laboratories (Wattiau *et al.* 2011; Sabat *et al.* 2013).

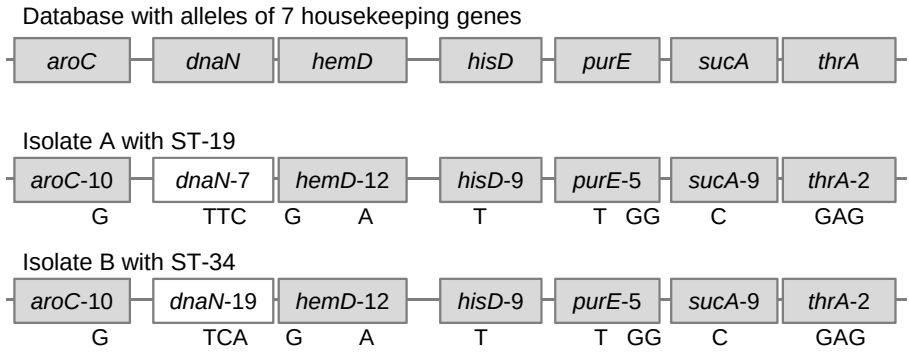

Figure 1.5: Multilocus sequence typing (MLST) for *Salmonella enterica*. Seven housekeeping genes (distributed around the genome) are amplified by PCR and Sanger sequenced. The sequences are compared to an allele database to assign a sequence type (ST) to the isolate.

For subtyping of *S.* Typhimurium, the spacers in two CRISPR loci are determined in a Luminex-based assay called CRISPOL (for CRISPR polymorphism) (Figure 1.6). First, the spacer content of the two loci is amplified by PCR. Subsequently, the spacers are identified by specific probes in a direct hybridisation assay (Fabre *et al.* 2012). The latter assay is elaborated in section 1.2.1. The CRISPOL assay is a fast technique, but has the disadvantages that it demands further refinement for genetically homogeneous populations such as *S.* Typhimurium phage type DT104, and that is has been patented by the French Pasteur Institute (Weill *et al.* 2008). As such, the technique has not yet been implemented as routine subtyping method for *S.* Typhimurium outside France.
In addition to the CRISPOL assay, also Fang *et al.* (2012) have applied the Luminex technology for subtyping of *S.* Typhimurium. In this assay, 30 prophage-related markers are amplified in 2 separate 15-plex PCRs, after which the amplicons are analysed in a direct hybridisation assay.

Quantitative interpretation of subtyping For interpretation of the discriminatory power of subtyping methods, Simpson's index of diversity (D) (Hunter and Gaston 1988) and Shannon's indices of diversity (H') and equitability (E) (Shannon 1948) are frequently reported in literature.
Simpson's index of diversity (D) gives the probability that two randomly sampled isolates from the test population will be assigned different subtypes and is calculated with equation 1.1 where N is the total number of isolates in the test population, S is the total number of different subtypes in the test

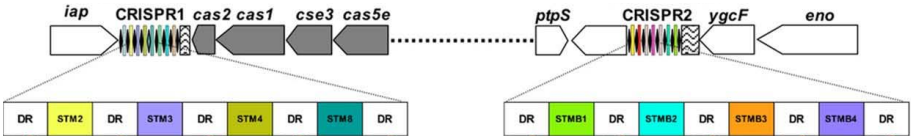

Figure 1.6: Clustered regularly interspaced short palindromic repeats (CRISPR) polymorphism (CRISPOL) typing. The spacers (coloured blocks) of two CRISPR loci are identified by probes, which are specific to a certain spacer, in a Luminex direct hybridisation assay (adapted from Fabre *et al.* (2012)).

population and n_j is the number of isolates from the test population that belong to subtype j.

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{S} n_j(n_j - 1) \tag{1.1}$$

Shannon's index of diversity (H') is an indicator for subtype richness and is calculated with equation 1.2, where S is the total number of different subtypes in the test population and p_i is the proportion of isolates, relative to the test population, that belong to subtype i. Shannon's index of equitability (E), which is calculated with equation 1.3, measures the evenness of the distribution of the different subtypes in the test population.

$$H' = -\sum_{i=1}^{S} p_i \ln p_i \tag{1.2}$$

$$E = \frac{H'}{\ln S} \tag{1.3}$$

In chapters 2 and 4, it is shown and discussed that these indices are not always appropriate for interpretation of the discriminatory power of a subtyping method.

1.1.4 Antimicrobial susceptibility

Besides the above described subtyping methods, antimicrobial susceptibility testing is used for additional characterisation of *Salmonella* isolates and detects resistances against a predefined set of antibiotics. Techniques that are commonly used for *Salmonella* are the disk diffusion test and the broth microdilution test (Humphries and Schuetz 2015).

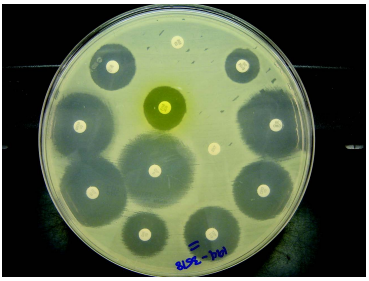

Figure 1.7: Disk diffusion test (Jorgensen and Ferraro 2009). Paper disk impregnated with antibiotics are placed on a lawn of bacterial culture. After incubation, the diameter of the growth inhibition zone around each disk is measured to determine the sensitivity of the bacterial isolate to the antibiotic.

Disk diffusion test For the disk diffusion test, a solid agar medium is first inoculated with a lawn of bacterial culture of the isolate to be characterised. Subsequently, commercial paper disks that are impregnated with a certain antibiotic at a defined concentration are deposited on the inoculated surface of the agar medium. After incubation of the agar plate for 18–24 hours at 35–37°C, the diameters of the growth inhibition zone around each disk (Figure 1.7) are measured (Jorgensen and Ferraro 2009). These diameters allow determination of the sensitivity of the isolate to each of the antibiotics in the paper disks by interpretation according to the criteria of the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and of the Clinical and Laboratory Standards Institute (CLSI) (Clinical and Laboratory Standards Institute (CLSI) 2015; European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2015a).

Broth microdilution test Commercial microtiter plates filled with gradient amounts of different antibiotics are used the broth microdilution test. An example of the layout of such commercial microtiter plate is shown in Figure 1.8. To each well this microtiter plate, a predefined volume of a bacterial culture at a certain turbidity is added, after which the plate is covered and incubated at 37°C for 20–24 hours. The plate is then read out, often with an automated system, to determine the minimal inhibitory concentration (MIC) for these antibiotics by examining if bacterial growth is present in each of the wells. The MIC is defined as the lowest concentration of an antibiotic that prevents bacterial growth (Jorgensen and Ferraro 2009). The observed MICs are then compared to epidemiological cut-off values provided by the European Commission (2013) and the European Committee on Antimicrobial Susceptibility Testing (EUCAST)

	1	2	3	4	5	6	7	8	9	10	11	12
A	Ami	A/S	Faz	Cep	Etp	G	P/T4	Sxt	Taz	Tgc	Axo	T
	64	32/16	32	16	16	16	128/4	4/76	32	8	64	16
B	Ami	A/S	Faz	Cep	Etp	G	P/T4	Sxt	Taz	Tgc	Axo	T
	32	16/8	16	8	8	8	64/4	2/38	16	4	32	8
C	Ami	A/S	Faz	Cep	Etp	G	P/T4	Sxt	Taz	Tgc	Axo	T
	16	8/4	8	4	4	4	32/4	1/19	8	2	16	4
D	Ami	A/S	Faz	Cep	Etp	G	P/T4	Sxt	Taz	Tgc	Axo	T
	8	4/2	4	2	2	2	16/4	0.5/9.5	4	1	8	2
E	A	Azt	Fep	Mero	Fur	Cip	Fox	Pod	Taz	Tim2	Axo	T
	32	32	32	8	32	4	32	16	2	64/2	4	1
F	A	Azt	Fep	Mero	Fur	Cip	Fox	Pod	Taz	Tim2	Axo	T
	16	16	16	4	16	2	16	8	1	32/2	2	0.5
G	A	Azt	Fep	Mero	Fur	Cip	Fox	Pod	Tob	Tim2	Axo	NEG
	8	8	8	2	8	1	8	4	8	16/2	1	CON
H	A	Azt	Fep	Mero	Fur	Cip	Fox	Pod	Tob	POS	POS	POS
	4	4	4	1	4	0.5	4	2	4	CON	CON	CON

Figure 1.8: Plate layout of a Sensititre plate for the broth microdilution antimicrobial susceptibility testing. The numbers denote the concentration of the respective antibiotic if 50 μl of a bacterial culture is added to each well (adapted from Thermo Fisher Scientific (2015)).
A/S: ampicilin/sulbactam; A: ampicilin; Ami: amikacin; Axo: ceftriaxone; Azt: aztreonam; Cep: cephalothin; Cip: ciprofloxacin; Etp: ertapenem; Faz: cefazolin; Fep: cefepime; Fox: cefoxitin; Fur: cefuroxime; G: gentamicin; Mero: meropenem; NEG CON: negative control; P/T4: piperacillin/tazobactam constant 4; Pod: cefpodoxime; POS CON: positive control; Sxt: trimethoprim/sulfamethoxazole; T: tetracycline; Taz: ceftazidime; Tgc: tigecycline; Tim2: ticarcillin/clavulanic acid constant 2; Tob: tobramycin.

(2015b) to determine the sensitivity of the bacterial isolate to the respective antibiotic.

1.2 Luminex technology

The Luminex technology, which emerged in 1999 with the launch of the first Luminex 100 platform, allows multiplexing with liquid bead-suspension arrays. The beads in these arrays are polystyrene microspheres with a diameter of 5.6 to 6.5 μm and which are divided into different spectral sets or regions by colouring them with distinct combinations of red, infra-red and orange-red dyes. These unique spectral addresses enable a Luminex instrument to categorise each microsphere individually in a mixture of microspheres of different regions. As to each of these microsphere regions a particular marker can be associated, a Luminex instrument distinguishes these markers in a multiplex bead-suspension array by analysing the red colour of the microsphere. The presence of the

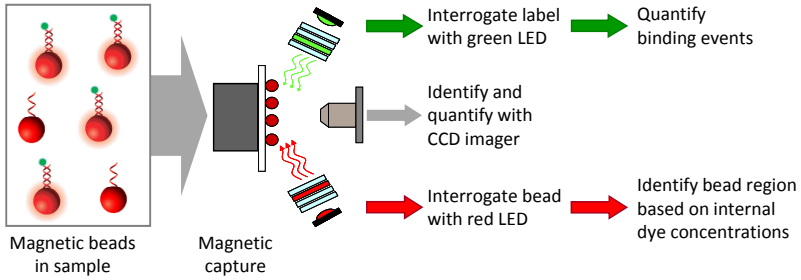

Figure 1.9: Read-out of liquid bead-suspension array with LEDs and CCD imager (adapted from Luminex).

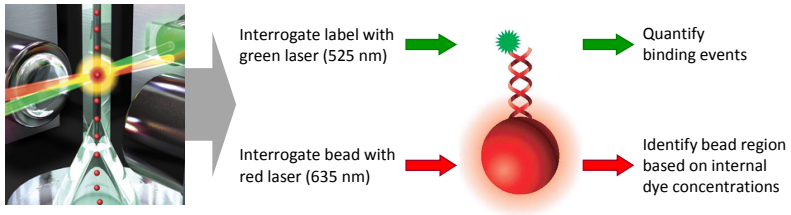

Figure 1.10: Flow cytometry-based read-out of liquid bead-suspension array (adapted from Luminex).

markers in a sample is detected by incorporating a green fluorescent label in the target analytes. The read-out of a multiplex bead-suspension array is thus basically a measurement of the red signal of the microsphere and the green signal of the target. The red signal identifies the microsphere region and thus the marker, presence or absence of the green signal indicates presence or absence of the target in the sample. The analysis of a bead-suspension array is done with LEDs and a CCD camera in the MAGPIX instrument (Figure 1.9) and with lasers in the flow cytometry-based Luminex 100/200 and FLEXMAP 3D instruments (Figure 1.10). The MAGPIX allows multiplexing up to 50, the Luminex 100/200 up to 100 and with the FLEXMAP 3D 500-plex assays can be analysed.

As modified oligonucleotides, proteins or other small molecules can be coupled to the microspheres, the Luminex technology allows both genomic and protein applications, such as sequence detection, single nucleotide polymorphism (SNP) typing, gene expression, microRNA analysis, ELISA and enzyme immunoassays. This introduction only discusses DNA assays, which are implemented through the Luminex xMAP or xTAG technology. An overview of the main characteristics of the different DNA assays are presented in Table 1.2.

Table 1.2: Overview of the main characteristics of Luminex assays commonly used in DNA applications.

Feature	Direct hybridisation assay	ASPE	OLA	MOL-PCR
Usage		Unique sequences, SNPs, multiple polymorphisms		
Technology	xMAP	xTAG	xTAG	xTAG
Microspheres	MagPlex or MicroPlex	MagPlex-TAG	MagPlex-TAG	MagPlex-TAG
Required oligos	Labelled multiplex primer pairs and modified target-specific probes	Multiplex primer pairs and ASPE primers	Multiplex primer pairs and labelled OLA probe pairs	Modified multiplex probe pairs and labelled universal primer pair
Coupling of capture probe to microspheres	Yes	No	No	No
Multiplex step	PCR	PCR and primer extension	PCR and ligation	Ligation
Hybridisation of target oligo to microspheres	Optimise: 45–55°C	Standard: 37°C	Standard: 37°C	Standard: 37°C
Buffer	TMAC	Tm	Tm	Tm
PCR amplicon size	$<$ 300 bp	All sizes	All sizes	All sizes
Patent	No	No	No	Yes
Cost	+	+++	++	+
Total time	$\pm$ 3.5 hours	6-7 hours	6-7 hours	6-7 hours

ASPE: allele-specific primer extension; MOL-PCR: multiplex oligonucleotide ligation-PCR; OLA: oligo ligation assay; SNP: single nucleotide polymorphism; Tm: buffer composed of 0.1 M Tris-HCl pH 8.0, 0.2 M NaCl and 0.08% Triton X-100; TMAC: tetramethylammonium chloride, buffer composed of 3 M TMAC, 0.15% Sarkosyl, 75 mM Tris-HCl and 6 mM EDTA; EDTA: ethylenediaminetetraacetic acid.

1.2.1 Luminex xMAP technology

The Luminex xMAP technology is based on target-specific probes which are coupled to either MagPlex microspheres or MicroPlex microspheres and which are used in a direct hybridisation assay. MagPlex microspheres are magnetic and can be used on all Luminex instruments, including MAGPIX, while the non-magnetic MicroPlex microspheres cannot be analysed on a MAGPIX system. Through a 5' amino modifier, the target-specific probes are covalently coupled to the carboxylated surface of the Magplex or MicroPlex microspheres. In case of SNP typing, the SNP should be located in the middle of the target-specific probe.
In general, a direct hybridisation assay (Figure 1.11) starts with a mPCR in which one of the primers in each primer pair is labelled with biotin or a fluorophore. Subsequently, the amplicons are hybridised to the target-specific probes on the microspheres and, after an incubation with streptavidin-R-phycoerythrin (SAPE) in case of biotinylated primers, read out on a Luminex system. The multiplex limiting factors in a direct hybridisation assay are the mPCR and the target-specific probes. These primers and probes have to be designed so that they do not cross hybridise and so that they have the same melting temperature (T_m), in order to allow annealing in PCR or hybridisation of the different mPCR amplicons to the probes at one temperature. If toxic tetramethylammonium chloride (TMAC) buffer is used in the direct hybridisation assay, only the length of the target-specific probes is important for the T_m. Indeed, TMAC stabilises the hydrogen bond between A-T base pairs and as such raises their T_m to that of G-C base pairs (Melchior and Von Hippel 1973), allowing a single, uniform temperature for hybridisation of the mPCR amplicons to the probes coupled to the microspheres.

1.2.2 Luminex xTAG technology

The basis of the Luminex xTAG technology are MagPlex-TAG microspheres, which are pre-coupled with anti-TAGs. Each microsphere region is associated with a unique anti-TAG. These anti-TAGs are 24 bp oligonucleotides with minimal cross hybridisation (no C, only T, A and G) and they all hybridise with their complementary TAG at 37°C. Hence, this TAG has to be incorporated in the target oligonucleotide through the Luminex assay, of which the most common are discussed below.

Allele-specific primer extension Allele-specific primer extension (ASPE) (Figure 1.12) makes use of an ASPE primer, which consists of a 5' TAG sequence, which is unique for each allele-specific sequence at 3'. For SNP

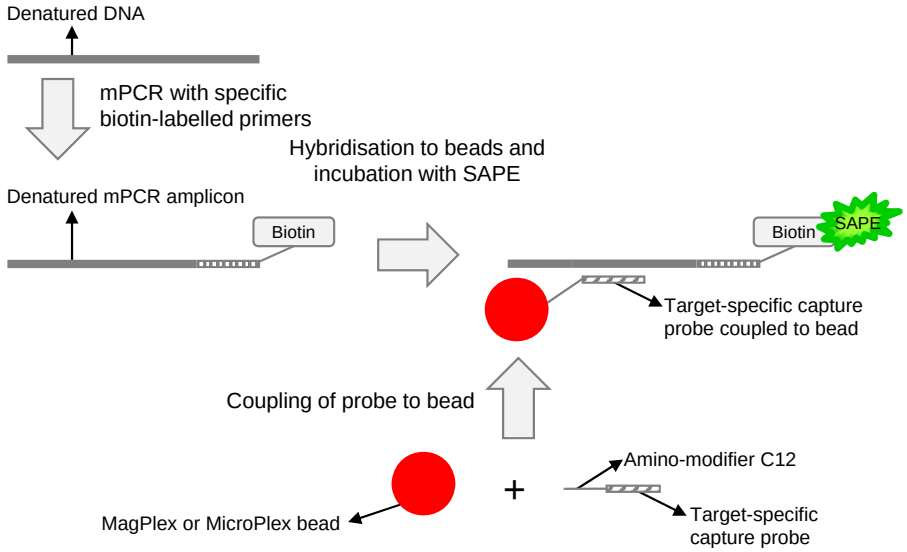

Figure 1.11: Schematic overview of a Luminex direct hybridisation assay.
mPCR: multiplex PCR; SAPE: streptavidin-R-phycoerythrin.

typing, the SNP has to be the 3' base of the ASPE primer. ASPE starts with amplification of the target in the genomic DNA by mPCR. The mPCR product then undergoes a clean-up with ExoSAP-IT reagent to inactivate dNTPs and remove excess primers. Next, ASPE primers anneal to the mPCR amplicons and are elongated by a DNA polymerase with integration of biotin-labelled dCTPs. This elongation only occurs when the 3' base in the ASPE primer is complementary to the target, which makes SNP typing possible. Through the TAG in the ASPE primer, the elongated products are hybridised to the anti-TAG on MagPlex microspheres, which are analysed on a Luminex instrument after incubation with SAPE (Angeloni *et al.* 2014). The multiplexing capacity in this assay format is mainly limited by the mPCR. The ExoSAP-IT treatment of the mPCR products make ASPE a relatively expensive assay.

Oligo ligation assay In an oligo ligation assay (OLA) (Figure 1.13) two probes are used: an upstream OLA-TAG probe which is similar to an ASPE primer and a downstream OLA reporter probe consisting of a target-specific sequence with a 3' biotin. As in ASPE, the target regions in the genomic DNA are first amplified in a mPCR. Subsequently, the OLA-TAG and the OLA reporter probes are annealed to mPCR amplicons and, in case of a strict complementarity of the

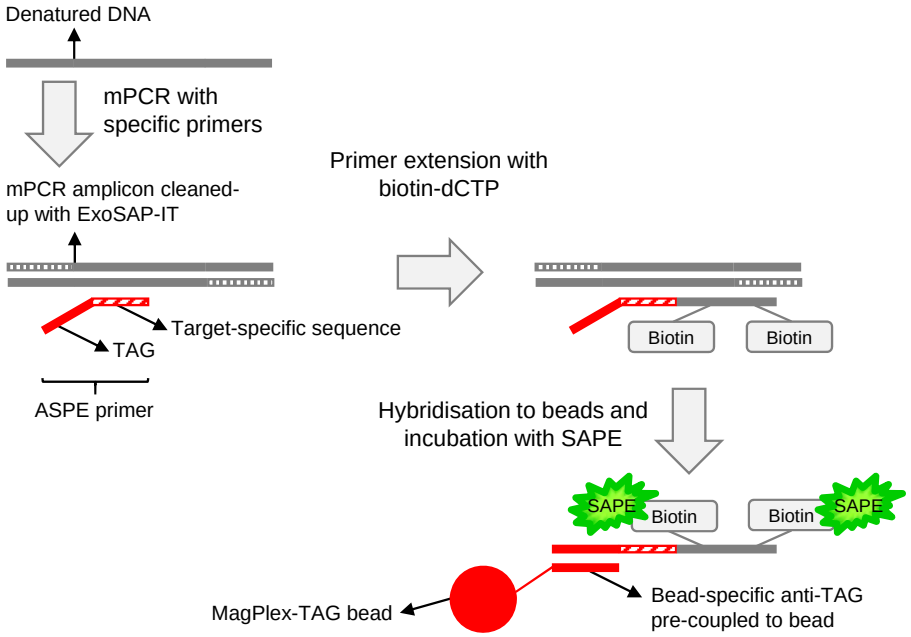

Figure 1.12: Schematic overview of a Luminex allele-specific primer extension (ASPE) assay.
dCTP: deoxycytidine triphosphate; mPCR: multiplex PCR; SAPE: streptavidin-R-phycoerythrin.

base pairs flanking the ligation site, the two probes are ligated by a DNA ligase and the biotin label is linked to the target-specific TAG. Analysis on a Luminex instrument is performed after hybridisation to MagPlex-TAG microspheres and incubation with SAPE (Angeloni *et al.* 2014). Again, the multiplexing capacity in this assay format is mainly limited by the mPCR.

Multiplex oligonucleotide ligation-PCR Multiplex oligonucleotide ligation-PCR (MOL-PCR) utilises two probes for detection of the target: an upstream probe including a 5' universal primer site (*e.g.* T7), an internal anti-TAG, which is unique for each target, and a 3' target-specific sequence; and a downstream probe consisting of a 5' target-specific sequence and a 3' universal primer site (*e.g.* T3). In contrast to ASPE and OLA, the first step in MOL-PCR (Figure 1.14) is a multiplex ligation, in which the upstream and downstream probes are ligated by a DNA ligase, under the severe constraint of complementarity of the base pairs next to the ligation site. Ligated probe pairs form a PCR

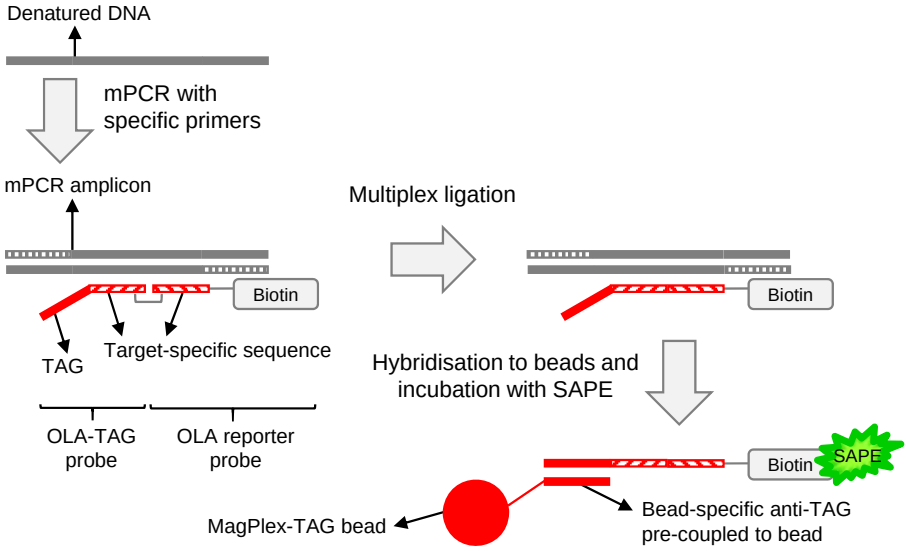

Figure 1.13: Schematic overview of a Luminex oligo ligation assay (OLA).
mPCR: multiplex PCR; SAPE: streptavidin-R-phycoerythrin.

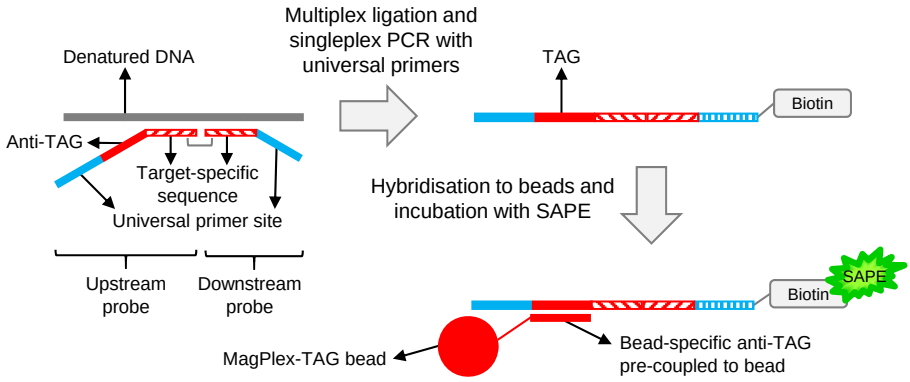

Figure 1.14: Schematic overview of a Luminex multiplex oligonucleotide ligation-PCR (MOL-PCR) assay.
SAPE: streptavidin-R-phycoerythrin.

template and are amplified in the succeeding singleplex PCR with the universal primer pair. One of the primers is biotinylated for later read-out on a Luminex instrument, after hybridisation to MagPlex-TAG microspheres and incubation with SAPE (Deshpande *et al.* 2010; Stucki *et al.* 2012; Thierry *et al.* 2013). As the multiplexing step in this assay is a ligation instead of a mPCR, higher multiplexing is possible with MOL-PCR than in assays with mPCR. Also, as only one primer is labelled with biotin, MOL-PCR is less expensive than ASPE or OLA.

1.3 Whole genome sequencing as subtyping standard for public health

Along with the studied pathogens, also subtyping techniques have evolved over the past decades. Classical phenotypic methods as phage typing are being, or have already been, abandoned and are replaced by molecular assays which have shown their additional value for surveillance and outbreak investigations, because of their improved capabilities to examine relationships between different isolates of the same pathogen. Currently, a next evolution is ongoing, which might take subtyping to the ultimate level of discriminatory power and universal applicability: whole genome sequencing (WGS).

WGS is the determination of the complete DNA sequence of an isolate by high-throughput sequencing, which permits WGS data of a pathogen to be available in a couple of hours up to several days, depending on the sequencing platform. This contrasts to the initial Sanger dideoxy chain termination sequencing method, by which it took several months to complete a typical pathogen genome. However, unlike a Sanger closed genome, WGS data do not constitute the entire DNA sequence of the isolate (often only about 90% of the genome), mostly because the short reads produced by high-throughput sequencing platforms do not allow resolving sequence repeats, which thus result as gaps in the genome (Sabat *et al.* 2013).

Nowadays, WGS is mostly applied for retrospective studies of outbreaks or of pathogen collections to extract informative data that may be useful for future routine implementation of WGS for outbreak detection and surveillance, since as such a baseline of expected diversity for a particular pathogen population is established (Gilmour *et al.* 2013; Ashton *et al.* 2015). This observation illustrates that no longer sequencing itself is an issue, as next generation sequencing (NGS) platforms are available which are suitable for different types of routine laboratories, although the time needed for generation of the sequence reads and the cost of this data generation may not yet be adapted to all routine laboratories. The real issue now is the analysis of the WGS data and how they

can be translated into relevant information for public health.
In this section, an overview of the preparatory work before sequencing is given, the main high-throughput sequencing platforms are discussed, after which the issue of data analysis is elaborated from the view of WGS for subtyping of bacterial pathogens.

1.3.1 Preparation of samples for high-throughput sequencing

WGS starts with a fresh culture of the pathogen from which the genomic DNA is isolated, usually with a commercial kit to have sufficiently pure and concentrated DNA. This first step may already be problematic for some difficult to grow pathogens. Then, a sequencing library is prepared by mechanically or enzymatically fragmenting the isolated DNA, repairing the ends and ligating adaptors. There are also library preparation kits available, such as the Illumina Nextera kit, that do the fragmenting and adaptor ligation in one step for fast sequencing library preparation (Loman *et al.* 2012). The library preparation will determine the length of the DNA fragments that are sequenced. With the adaptors, the fragmented DNA is immobilised on a solid surface, after which the library is amplified by PCR to increase the signal of the following sequencing step (Di Bella *et al.* 2013).
In paired-end sequencing the DNA fragments are sequenced from each end. In contrast to paired-end sequencing, mate pair sequencing requires an additional library preparation step. For a mate pair library, the fragmented DNA, typically at a size of several kb, is repaired with labelled dNTPs, after which the DNA fragments are circularised and fragmented again. The labelled fragments are selected and again end repaired, after which adaptors for paired-end sequencing are ligated. A mate pair library thus includes short DNA fragments composed of two fragments that were initially separated by several kb. This information is valuable in *de novo* assembly, since it can locate reads in the genome (Loman *et al.* 2012; Illumina 2015b; Illumina 2015d).

1.3.2 High-throughput sequencing platforms for WGS

The currently available high-throughput sequencing platforms are based on sequencing by synthesis, but each platform has its specific sequencing chemistry, with its intrinsic read length and error rate. Within the different platforms, there is a distinction between the large high-end instruments which are designed for a massive throughput and which generate more output than the less expensive benchtop instruments, which have shorter run time and which focus more at small genomes in moderate throughput (Loman *et al.* 2012; Di Bella *et al.* 2013).

The requirement of a library preparation, which may introduce a PCR bias through the DNA amplification step, sets apart the second generation sequencing, also known as NGS, techniques and the third generation sequencing techniques, commonly referred to as high-throughput sequencing techniques.

Roche 454 GenomeSequencer The Roche 454 GenomeSequencer uses pyrosequencing, in which the incorporation of a dNTP by a DNA polymerase releases a pyrophosphate and hydrogen ions. In each sequencing cycle, a single type of dNTP is surged over the surface on which the amplified library is attached. By determining the amount of present pyrophosphate through an enzymatic reaction which releases light, the order and number of the incorporated dNTPs is recorded. Roche 454 has the advantages of producing long reads of 700–800 bp in a 1 day runtime, but has a high reagent cost and gives a high error rate, especially in homopolymer regions (Loman *et al.* 2012). Roche 454 sequencing platforms are planned to be discontinued by mid-2016.

Illumina The Illumina technology utilises reversible terminator nucleotides, which are fluorescently labelled depending on the base. All four of these nucleotides are run over the flow cell and when a nucleotide is incorporated, its fluorescent signal is recorded and then quenched. Subsequently, the protection that prevented incorporation of the next nucleotide (the terminator) is inactivated and the next sequencing cycle starts (Loman *et al.* 2012). The Illumina technology has an estimated error rate of 1/1000 sequenced bases, of which most errors are substitutions (Ross *et al.* 2013). The high-end Illumina HiSeq 2000 and HiSeq 2500 have typical short read lengths of 36–100 bp and paired-end runs take 2.5 to 11 days, but they can handle a very high-throughput at a lower cost (Di Bella *et al.* 2013; Illumina 2014; Illumina 2015a). The benchtop MiSeq produces reads of length 36–300 bp in 4 to 56 hours, respectively (Illumina 2015c). At the moment, Illumina is the leading platform for high-throughput WGS, despite the long run times (European Food Safety Authority (EFSA) 2014).

Ion Torrent The Ion Torrent platform uses essentially the same chemistry as the 454, but instead of registering released light, it detects the released hydrogen ions through the change in pH with semiconductor technology. The PGM platform produces reads up to 400 bp and a single-end run takes 8 hours. The disadvantage of the Ion Torrent platform is the high error rate in homopolymer stretches, which is estimated at 1/100 positions (Loman *et al.* 2012; Ross *et al.* 2013).

Pacific Biosciences (PacBio) PacBio introduced single molecule real time (SMRT) sequencing. Here, a highly productive DNA polymerase is attached to the solid surface of a SMRT cell and captures a DNA strand to be sequenced. A mixture of fluorescently labelled dNTPs is constantly flown over the SMRT cell and when a dNTP is incorporated, the fluorescent signal is recorded, after which the fluorophore is released and a next dNTP can be incorporated (Loman *et al.* 2012). The advantages of the PacBio platform are the enormous read lengths up to 14 kb and as no PCR amplification step is required during sample preparation, also epigenetic data, such as DNA methylation, can be obtained. However, PacBio sequencing is not being used currently for routine WGS, because of its high cost and error rate of about 1/10 sequenced bases (Di Bella *et al.* 2013; Ross *et al.* 2013).

1.3.3 WGS data analysis

As sequencing costs are decreasing, the generation of WGS data becomes more accessible for National Reference Centres and Laboratories, although the costs still depend heavily on the number of samples that are simultaneously processed, so that for smaller laboratories, which have to analyse only a small number of pathogenic isolates at a time, WGS may still be too expensive. For National Reference Centres and Laboratories which have the resources to invest in high-throughput sequencing platforms, the extraction of clinically and epidemiologically useful information out of WGS reads is the bottleneck that prevents the adoption of WGS as subtyping standard. Firstly, analysis of millions of short reads requires bioinformaticians and substantial computer power, both of which are not commonly found in the average National Reference Centre or Laboratory. Secondly, as with all subtyping methods, a standard protocol is required to be able to compare results between different laboratories and also to compare results over time. But how should that standard protocol for WGS data analysis look like? Many questions have yet to be answered by the involved community, but answers are requested as WGS will become the standard in Europe (European Food Safety Authority (EFSA) 2014).
The first discussion point is the workflow for interpretation of WGS data. Two complementary methodologies exists: SNP calling from reference-based read mapping and gene-by-gene comparison. For SNP calling a completed reference genome is needed and it does not allow detection of new genetic elements, but results can be straightforwardly visualised in a phylogenetic tree. The gene-based approach relies on *de novo* assembly, of which the quality is very much dependent on the quality of the WGS data, so this should be specified in the standard protocol. A gene-based approach can be worked out as presence or absence of a set of genes or, if mutations are included, as an allele comparison,

also known as whole genome MLST. The set of genes to be compared can be the core genome, defined by Tettelin *et al.* (2008) as "the genes shared by all the strains studied" or the pan-genome, which includes the core genome plus the accessory genome, in which the latter consists of all isolate-specific genes (Tettelin *et al.* 2008). The decision on which set of genes or MLST scheme to use, will depend on the required resolution and may be pathogen-specific. In any case, for whole genome MLST, arrangements have to be made for an international database to define the extended sequence type (ST) (European Food Safety Authority (EFSA) 2014). An architecture for such bacterial pathogen databases, *i.e.* BIGSdb, was already proposed by Jolley and Maiden (2010) and Applied Maths is also developing a whole genome MLST module for several pathogens for the BioNumerics software. A computational advantage of whole genome MLST is that, contrary to a phylogenetic tree that results from SNP calling, not the complete dataset has to be recalculated when a new isolate is added, as the new isolate will receive an extended ST that can be readily compared to that of the isolates already residing in the dataset (Jolley and Maiden 2010).

A second point of consideration is a definition of distinct subtypes, which is a very important issue from a legal viewpoint. *E.g.* during an outbreak investigation, how many SNPs may there be between two isolates before they are regarded as descending from two distinct strains? To answer this question, WGS should be applied to as much as possible related and unrelated isolates of a pathogen, in order to determine differences in the genetic diversity in background strains and in outbreak cases.

A third question is related to the inference of phenotype data, such as antimicrobial susceptibility and virulence, from the WGS genotype data. ResFinder (Zankari *et al.* 2012) and VirulenceFinder (Joensen *et al.* 2014) already allow to identify antimicrobial resistance and virulence genes which are introduced in the underlying databases (VirulenceFinder is available for *Escherichia coli*, *Staphylococcus aureus* and *Enterococcus*), but expression of these genes cannot be inferred from WGS data. Moreover, mechanisms of antimicrobial resistance also includes mutations and all mechanisms have not yet been elucidated.

A last point of consideration is the link with the historical subtyping data. National Reference Centres and Laboratories retain classical phenotypic and molecular data of surveillance and outbreak isolates and compatibility with WGS data is not guaranteed, especially for phenotypic subtyping data as phage type. Classical MLST sequence types based on a few housekeeping genes, can readily be extracted from WGS data (Larsen *et al.* 2012), but Illumina reads are too short to span a MLVA locus, so that counting of tandem repeats for MLVA might be impossible. In addition, a *de novo* assembly results in a number of contigs with gaps between them, which make it impossible to predict a PFGE pattern, unless the genome is closed (Sabat *et al.* 2013). Suggestions are

that we have to move away from the historical subtyping data and certainly from the historical nomenclature, as all results from WGS analysis may not be captured as a single type, but more in the sense of a report (Gilmour *et al.* 2013; European Food Safety Authority (EFSA) 2014).
The eventual standardised protocol for WGS analysis in a National Reference Centre or Laboratory should be implemented in a user-friendly pipeline, so that they can perform the standard analysis without intervention of a bioinformatician. Hence, this user-friendly pipeline should be operated by a graphical user interface (GUI), so that it is manageable in a routine setting, contrary to research environments, where a command line may be a common interface.

1.4 Rationale and outline of the thesis

1.4.1 Rationale of the thesis

The research question of this PhD project, called SalMolType, arises from the Belgian National Reference Centres located at the Scientific Institute for Public Health (WIV-ISP), which were in need for an improved subtyping technique, since the available methods for characterisation of a pathogen did not meet the specifications of an ideal subtyping method as defined in section 1.1. This need applies equally to the average human National Reference Centres and the National Reference Laboratories for food. Especially for the important foodborne pathogen *Salmonella* an alternative subtyping method was needed, since the subjective phage typing technique had to be discontinued, MLVA was not equally performing for each serovar and PFGE was too labour-intensive and time-consuming for routine implementation. As a case study, the most commonly isolated serovar, *i.e. S.* Typhimurium, was taken, because MLVA resulted in so many different profiles (during the PhD work this was attributed to the instability of the MLVA loci) so that analysis was difficult for routine surveillance and outbreak detection.
For *Salmonella*, subtyping means to discriminate between isolates which are classified as the same serovar. The required level of discrimination can vary for different usages of subtyping. *E.g.*, for outbreak investigations, it may be necessary to use a technique that allows a very high resolution, so that outbreak isolates can be distinguished from circulating background isolates, which is most likely not required for routine surveillance purposes. The method will also have to be more rapid for outbreak investigations than for routine surveillance. For both outbreak investigations as for routine surveillance, the subtyping method is, ideally, inexpensive, robust, easy to implement and to standardise, universally

applicable to a wide range of pathogens, and resulting in objective data that are suitable for interpretation and for communication between different laboratories. A convenient nomenclature may be indispensable for the interpretation and communication requirements.
Technologically, this PhD aimed at the development of a multiplex assay based on molecular markers, universally applicable in epidemiology and surveillance. Concomitantly, the Luminex technology was implemented at the Scientific Institute for Public Health (WIV-ISP). From a comparison of different Luminex assays, it was decided that the MOL-PCR was the optimal method, as it allowed a high level of multiplex and cost and time needed (less than 8 hours is requested for outbreak investigations) for the assay were acceptable.
After development of the MOL-PCR assay for subtyping of *S.* Typhimurium, which included a thorough optimisation of the used technology and validation on a large collection of *S.* Typhimurium and its monophasic variant *S.* 1,4,[5],12:i:-, it was clear that the MOL-PCR assay was not yet ideal with respect to discriminatory power, which was observed to be equal to that of phage typing and which might not be sufficient in some cases. Addition of more markers in the MOL-PCR assay would however increase the cost and effort of the assay. Furthermore, the MOL-PCR assay technique is universally applicable to other pathogens, but each pathogen would require the development of a new set of probe pairs, which is not ideal. Therefore, WGS was explored as suggested ideal subtyping method. The use of WGS for outbreak investigation and surveillance by an average National Reference Centre or Laboratory has been evaluated based on selected case studies.

1.4.2 Outline of the thesis

The outline of the thesis with the different chapters is presented in Figure 1.15. The first chapter gives a general introduction on different aspects and techniques used in this PhD and provides the rationale and outline of the thesis. Chapter 2 is a study of 3-year surveillance data on *S.* Typhimurium of the Belgian National Reference Centre for *Salmonella* and *Shigella.* Next to phage type and antimicrobial susceptibility, the surveillance data included MLVA for which the analysis was an issue for the National Reference Centre for *Salmonella* and *Shigella*, because of the large number of distinct MLVA profiles that were obtained. As MLVA is currently a standard subtyping method in Europe (Larsson *et al.* 2013; Lindstedt *et al.* 2013), an analysis scheme adapted to the needs of the National Reference Centre for *Salmonella* and *Shigella* was created. Together, chapters 1 and 2 give the general background information for the thesis.
The optimisation of the MOL-PCR assay is incorporated as chapter 3 of the

thesis, with the actual development and validation of the MOL-PCR assay for subtyping of *S.* Typhimurium as chapter 4.
In chapter 5, WGS data of *S.* Typhimurium and *S.* 1,4,[5],12:i:- are explored in relation to the different questions in section 1.3.3. Chapter 6 is then an application of currently available, user-friendly software and tools for outbreak investigation with WGS. Two *S.* Enteritidis outbreaks were taken as a case study, since both food and human isolates were available and to underline the universal usability of WGS for characterisation of pathogens.
Finally, chapter 7 presents the general conclusions of this PhD and the future perspectives.

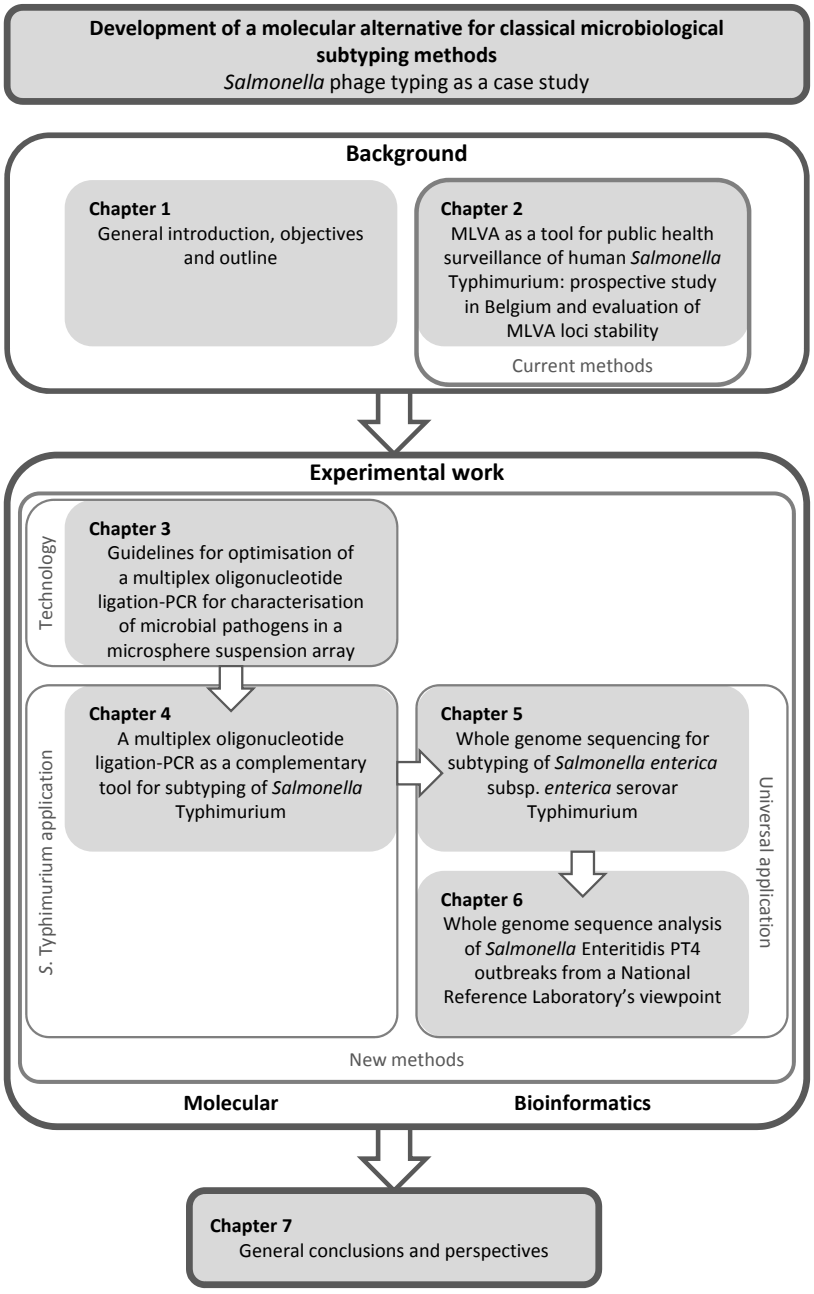

Figure 1.15: Schematic outline of the thesis.

Chapter 2

MLVA as a tool for public health surveillance of human *Salmonella* Typhimurium: prospective study in Belgium and evaluation of MLVA loci stability

This chapter was previously published as:
V. Wuyts, W. Mattheus, G. De Laminne de Bex, C. Wildemauwe, N. H. C. Roosens, K. Marchal, S. C. J. De Keersmaecker, and S. Bertrand (2013). "MLVA as a tool for public health surveillance of human *Salmonella* Typhimurium: prospective study in Belgium and evaluation of MLVA loci stability". *PLoS ONE* 8.12, e84055. DOI: 10.1371/journal.pone.0084055

Authors' contributions

The serial passage experiment was designed by V. Wuyts, W. Mattheus, N. H. C. Roosens, S. C. J. De Keersmaecker and S. Bertrand and was performed by V. Wuyts and G. De Laminne de Bex. Serotype, phage type, MLVA and antimicrobial susceptibility data were provided by W. Mattheus, C. Wildemauwe and S. Bertrand. The data were analysed by V. Wuyts after discussion with W. Mattheus, G. De Laminne de Bex, N. H. C. Roosens, K. Marchal, S. C. J. De Keersmaecker and S. Bertrand. The paper was written by V. Wuyts with critical input of S. C. J. De Keersmaecker. W. Mattheus, N. H. C. Roosens, K. Marchal and S. Bertrand reviewed the paper and also contributed reagents, materials and analysis tools. S. C. J. De Keersmaecker and S. Bertrand contributed equally to this work.

Abstract

Surveillance of *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) is generally considered to benefit from molecular techniques like multiple-locus variable-number of tandem repeats analysis (MLVA), which allow earlier detection and confinement of outbreaks. Here, a surveillance study, including phage typing, antimicrobial susceptibility testing and the in Europe most commonly used 5-loci MLVA on 1,420 *S.* Typhimurium isolates collected between 2010 and 2012 in Belgium, was used to evaluate the added value of MLVA for public health surveillance. Phage types DT193, DT195, DT120, DT104, DT12 and U302 dominate the Belgian *S.* Typhimurium population. A combined resistance to ampicillin, streptomycin, sulphonamides and tetracycline (ASSuT) with or without additional resistances was observed for 42.5% of the isolates. 414 different MLVA profiles were detected, of which 14 frequent profiles included 44.4% of the *S.* Typhimurium population. During a serial passage experiment on selected isolates to investigate the *in vitro* stability of the 5 MLVA loci, variations over time were observed for loci STTR6, STTR10, STTR5 and STTR9. This study demonstrates that MLVA improves public health surveillance of *S.* Typhimurium. However, the 5-loci MLVA should be complemented with other subtyping methods for investigation of possible outbreaks with frequent MLVA profiles. Also, variability in these MLVA loci should be taken into account when investigating extended outbreaks and studying dynamics over longer periods.

2.1 Introduction

Salmonella is the most frequent cause of foodborne outbreaks and human salmonellosis is the second most frequently reported zoonosis in the European Union (European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2013). The most common serovars of *Salmonella* isolated from human outbreaks are *Salmonella enterica* subsp. *enterica* serovar Enteritidis (*S.* Enteritidis) and *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) (European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2013). Typing methods which allow characterization below the serovar level are essential in a surveillance program for this diverse genus. Classical surveillance programs for *Salmonella* rely on phenotyping methods such as phage typing and antimicrobial susceptibility testing. Nowadays, phage typing as a subtyping technique is often complemented with molecular methods like pulsed-field gel electrophoresis (PFGE), which is considered the gold standard for subtyping of *Salmonella* (Wattiau *et al.* 2011). Yet, recent studies suggest that multiple-locus variable-number of tandem repeats analysis (MLVA) improves surveillance, detection of outbreaks and of sources of outbreaks of *Salmonella* and in particular of *S.* Typhimurium (Torpdahl *et al.* 2007; Heck 2009; Sintchenko *et al.* 2012; Lindstedt *et al.* 2013). MLVA targets rapidly evolving genomic elements known as tandem repeats. This allows to use them to study genetic relatedness between isolates. There exist different MLVA schemes for *S.* Typhimurium, all with a different number of variable number of tandem repeat (VNTR) loci used. The first MLVA scheme for *S.* Typhimurium used 8 VNTR loci (Lindstedt *et al.* 2003). Improvements of PCR multiplexing and capillary electrophoresis to this MLVA scheme resulted in a scheme with 5 VNTR loci of which 3 loci were previously included in the 8-loci MLVA scheme and 2 loci were newly added (Lindstedt *et al.* 2004), and which was equally well performing as the more labour intensive 8-loci MLVA scheme. This 5-loci scheme is currently in Europe the most commonly used MLVA scheme for *S.* Typhimurium (Larsson *et al.* 2013; Lindstedt *et al.* 2013). Variations on this 5-loci MLVA scheme have been tested. However, none of these showed to have large added value, compared to the 5-loci MLVA scheme, for routine surveillance and outbreak investigation of *S.* Typhimurium. Indeed, other MLVA schemes with 10 and 6 VNTR loci were developed for simultaneously typing of multiple *Salmonella* serovars (Ramisse *et al.* 2004; Witonski *et al.* 2006), thereby circumventing the disadvantage that most MLVA schemes are dedicated to one specific serovar. However, the small set of studied *S.* Typhimurium isolates only showed variability at the VNTR loci in common with the previously published 8-loci and 5-loci MLVA schemes. Additionally, a recent study which enlarged the 5-loci MLVA scheme with 11 additional VNTR loci for *S.* Typhimurium concluded that the 5-loci MLVA

scheme was suitable to supplement PFGE in routine surveillance and outbreak investigation (Chiou *et al.* 2010). The 5-loci MLVA scheme of Lindstedt *et al.* (2004) was recently validated in a large European inter-laboratory trial (Larsson *et al.* 2013; Lindstedt *et al.* 2013). Outside Europe other MLVA schemes are in use, *e.g.* PulseNet USA (PulseNet USA 2013) developed a 7-loci MLVA protocol for *S.* Typhimurium by adding 2 VNTR loci to the 5-loci MLVA scheme. Whatever MLVA scheme used, the combination of the number of tandem repeats at a predefined number of MLVA loci results in a MLVA profile. Nonetheless, questions are raised related to the stability of MLVA loci (Hopkins *et al.* 2007; Ross *et al.* 2011) and different ways of handling closely related MLVA profiles during outbreak investigations have been proposed (Petersen *et al.* 2011; Friesema *et al.* 2012; Kuhn *et al.* 2013; Garvey *et al.* 2013; Paranthaman *et al.* 2013).

The objective of this study was to evaluate the added value of MLVA typing for surveillance and outbreak detection by comparing MLVA profiles of a large *S.* Typhimurium collection to results obtained through traditional phenotyping methods and by evaluating the *in vitro* stability of the loci of the used MLVA scheme in a serial passage experiment.

The *S.* Typhimurium isolates in this study were collected in Belgium, where *S.* Typhimurium is since 2006 the serovar most frequently isolated from humans with an average of 1,985 isolates (55.9% of all *Salmonella* isolates, average from 2006 up to 2012) reported per year. *S.* Enteritidis follows with an average of 774 isolates (21.8%) reported each year (Bertrand *et al.* 2012).

For routine surveillance of human *S.* Typhimurium infections, the Belgian National Reference Centre for *Salmonella* and *Shigella* (NRCSS) uses phage typing and antimicrobial susceptibility testing. In outbreak situations, these phenotyping techniques are complemented with PFGE. During the 3-year period 2010–2012, the *S.* Typhimurium isolates were also analysed with the 5-loci MLVA scheme commonly used in Europe (Larsson *et al.* 2013; Lindstedt *et al.* 2013). Hence, this large collection of data created the ideal opportunity to investigate the potential and added value of MLVA for surveillance and outbreak detection of an important foodborne pathogen. This study demonstrates that although the discriminatory power of MLVA allows for an improvement of public health surveillance, additional or alternative molecular subtyping methods should be used to detect an outbreak and to uniquely characterize an outbreak isolate. Moreover, as some of the MLVA loci showed to be unstable, the interpretation of these genetic markers for subtyping should be done with caution.

2.2 Materials and methods

2.2.1 Bacterial isolates

In Belgium, peripheral clinical laboratories collect *Salmonella* isolates from human patients and send them voluntarily to the NRCSS for serotyping. In the 3-year period from 1 January 2010 to 31 December 2012, the Belgian NRCSS received a total of 10,055 human *Salmonella* samples. From the 5,698 isolates (56.7%) that were serotyped as Typhimurium, a random subset of 1,439 *S.* Typhimurium isolates were analysed by phage typing, antimicrobial susceptibility testing and MLVA. Exclusion of 24 isolates which gave inconsistent phage types in confirmatory tests, led to a total of 1,415 randomly sampled *S.* Typhimurium isolates in this study ($n_{2010} = 481$, $n_{2011} = 449$, $n_{2012} = 485$). This randomly sampled set included 11 isolates (*i.e.* 0.8%) that were serotyped as the monophasic variant (1,4,[5],12:i:-) of *S.* Typhimurium. Five isolates collected during an outbreak in a day nursery in 2011 were also included in this study, so that the total size of the studied population is 1,420 isolates. All typing data are available as Dataset S1 in the supporting information.

2.2.2 Serotyping and phage typing

Serotyping of *Salmonella* isolates was performed by slide agglutination with commercial antisera by the Kauffmann-White scheme (Grimont and Weill 2007). Phage typing of *S.* Typhimurium was carried out according to the recommendations of the Health Protection Agency (Colindale, United Kingdom) (Threlfall and Frost 1990). A frequent phage type was defined as a phage type that was detected in at least 50 isolates during the 3-year period 2010–2012.

2.2.3 Antimicrobial susceptibility testing

The susceptibility to 13 antibiotics was determined by the disk diffusion (Kirby-Bauer) method following recommendations of the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and using Bio-Rad (Nazareth, Belgium) disks (European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2013a). Inhibition zones were interpreted according to EUCAST criteria (European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2013b) for ampicillin (A), amoxicillin plus clavulanic acid (Amc), cefotaxime (Ctx), chloramphenicol (C), ciprofloxacin (Cip), gentamicin (G), trimethoprim (Tmp) and trimethoprim plus sulfamethoxazole (Sxt), and Clinical and Laboratory Standards Institute's (CLSI) criteria for kanamycin (K),

nalidixic acid (Na), streptomycin (S), sulphonamides (Su) and tetracycline (T). Quality control was performed using the *Escherichia coli* ATCC 25922 strain. Multidrug resistance (MDR) was defined as resistance to 4 or more antimicrobials.

2.2.4 Multiple-locus variable-number of tandem repeats analysis (MLVA)

MLVA was performed as described previously (Lindstedt *et al.* 2004). Liquid cultures were heated at 95°C for 10 minutes and used directly in the PCR reaction after a brief centrifugation at 18,188 g for 10 minutes, or a DNA lysate was prepared by heating a single colony in 300 μl sterile water at 100°C for 10 minutes and collecting the supernatant after centrifugation at 9,300 g for 10 minutes. PCR products were subjected to capillary electrophoresis on a ABI 3130xl Genetic Analyzer (Life Technologies™), after which the size of the PCR products was determined with GeneMapper® software v.1.0 (Life Technologies™). GeneScan™ 600 LIZ® (Life Technologies™) was used as size standard. The calibration strains were the same as described previously (Larsson *et al.* 2009) with the addition of two strains, STm-SSI32 and STm-SSI33 (European Centre for Disease Prevention and Control (ECDC) 2011). Presence of loci which presented a relatively low peak area with respect to other loci in the sample, was confirmed through single target PCR and agarose gel electrophoresis. MLVA profiles are reported as a string of 5 numbers (STTR9-STTR5-STTR6-STTR10-STTR3) representing the number of repeats at the corresponding locus or NA in case a PCR product was not obtained for that locus (Larsson *et al.* 2009). A frequent MLVA profile was defined as a MLVA profile that was detected in at least 20 isolates during the 3-year period 2010–2012. A MLVA profile that was detected in less than 20 isolates during the 3-year period 2010–2012 defined a rare profile.

2.2.5 Stability experiment

The *in vitro* stability of the 5 MLVA loci was evaluated in 20 *S.* Typhimurium isolates with a frequent MLVA profile and a frequent phage type, and in 11 *S.* Typhimurium isolates with a rare MLVA profile but with a frequent phage type. A single colony from a culture grown overnight on LB agar at 37°C was inoculated into 5 ml LB broth and incubated overnight at 37°C without shaking. Next, a series of 50 passages at a rate of two passages per day was performed by inoculating 20 μl of culture into 5 ml fresh LB broth and incubating at 37°C without shaking. Glycerol (25% *v/v*) stocks (−80°C) were made before each

fifth passage. MLVA was performed on heated liquid cultures after every fifth passage, as described above, leading to a total of 310 typing tests (Struelens *et al.* 1996).

2.2.6 Minimum spanning tree and diversity indices

A minimum spanning tree based on MLVA profiles of *S.* Typhimurium isolates was created in BioNumerics 6.5 (Applied Maths) using the categorical coefficient and no priority rules for the algorithm.
The discriminatory power of phage typing, antimicrobial resistance testing and MLVA was evaluated in the 1,415 randomly sampled *S.* Typhimurium isolates with Simpson's index of diversity (D) (Hunter and Gaston 1988) and Shannon's indices of diversity (H') and equitability (E) (Shannon 1948). Shannon's indices were calculated with the Biodiversity Calculator developed by J. Danoff-Burg and C. Xu (Danoff-Burg 2003).

2.3 Results

2.3.1 Phage types

The 1,420 *S.* Typhimurium isolates were first subtyped by phage typing. Forty-one distinct phage types were present, nevertheless 969 (68.2% of the sample population) isolates were attributed to the frequent phage types DT193 (20.0%), DT195 (17.7%), DT120 (16.3%), DT104 (6.3%), DT12 (4.1%) and U302 (3.9%). The isolates with frequent phage types were not equally present over each year of the study period (Figure 2.1). Phage types DT12 and U302 were mainly found in 2010. Phage type DT195 dominated the seasonal peaks in 2010 and 2011, but was replaced by phage type DT120 during the seasonal peak in 2012. Nevertheless, isolates with phage types DT193 and DT104 were detected throughout the 3-year period.
Not-typable (NT) and reacts-but-does-not-conform (RDNC) isolates of *S.* Typhimurium comprised respectively 7.1% and 9.2% of the sample population. Other phage types covered each less than 3% of the sample population.
The monophasic variant isolates were phage typed as DT193 (6 isolates), RDNC (3), DT120 (1) and U302 (1).

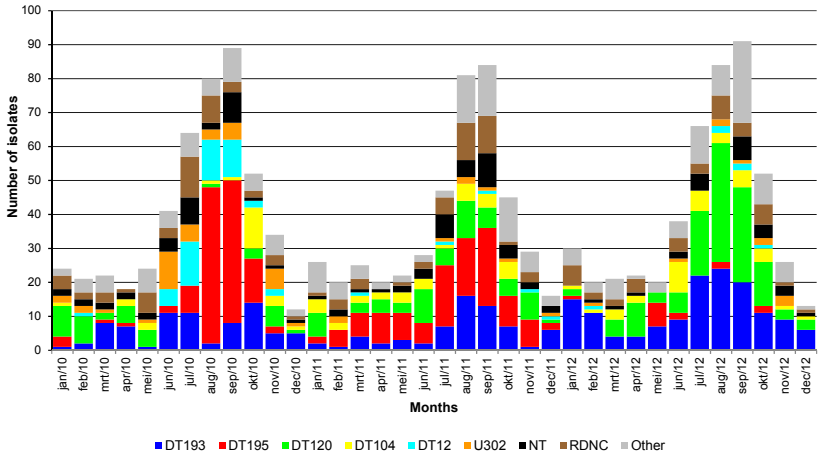

Figure 2.1: Monthly distribution of phage types among the *Salmonella* Typhimurium population over the 3-year period. N = 1,407 as for 13 isolates in this study the isolation date is unknown. NT: not-typable; RDNC: reacts-but-does–not-conform; Other: phage types with less than 3% occurrence in the sample population.

2.3.2 Antimicrobial susceptibility testing

Subsequently, the 1,420 *S.* Typhimurium isolates were further characterised by antimicrobial susceptibility testing. The isolates of *S.* Typhimurium were most frequently resistant to ampicillin (77.3% of the sample population), sulphonamides (63.0%), streptomycin (57.3%) and tetracycline (54.6%) whereas 11.9% of the isolates were susceptible to all antibiotics tested. Frequency of resistance to other antibiotics was between 0.2% for ciprofloxacin and 14.7% for trimethoprim. In our study population, the percentage of resistance remained stable over the 3-year period, except for the resistance to ampicillin and tetracycline, which showed a slightly increasing trend. An increasing trend was also observed for isolates with decreased susceptibility to sulphonamides. However, a declining trend was observed for decreased susceptibility to nalidixic acid (data not shown).

Multidrug resistance (MDR) occurred in 51.1% of the *S.* Typhimurium isolates. The dominant MDR pattern was ampicillin, streptomycin, sulphonamides and tetracycline (ASSuT) with or without additional resistances, which was responsible for 83.3% of the MDR isolates in the 3-year period and which was observed in 6 of the monophasic variant isolates (the 5 other monophasic variant isolates had a ASSu resistance pattern). ASSuT was also the most common

resistance pattern in *S.* Typhimurium belonging to phage types DT193 and DT120. ACSSuT with or without additional resistances was the most frequent pattern for MDR isolates of phage types DT104 and DT12. A combined resistance to 6 or more antibiotics occurred regularly for phage types DT12, DT120 and DT104, which presented a moderate to high percentage of MDR isolates. On the other hand, *S.* Typhimurium with phage types U302 and DT195 presented a low frequency of MDR isolates and were mainly associated with a single resistance to ampicillin (Table 2.1).

2.3.3 MLVA typing

Among the 1,420 *S.* Typhimurium isolates typed with MLVA targeting 5 loci, 414 distinct MLVA profiles were detected. Absence of a PCR amplicon occurred most often at loci STTR10 (75.0% of the sample population), STTR3 (3.3%) and STTR6 (1.8%). The highest number of different alleles was seen at locus STTR6 (27), followed by loci STTR10 (24), STTR5 (22), STTR3 (20) and STTR9 (11). Thirty MLVA profiles were observed throughout the 3-year period, while 131, 93 and 108 MLVA profiles were only identified in 2010, 2011 and 2012, respectively.

263 rare MLVA profiles (63.5% of the MLVA profiles) were detected for only one *S.* Typhimurium isolate (18.5% of the sample population), while 14 frequent MLVA profiles (3.4% of the MLVA profiles) comprised 44.4% of the *S.* Typhimurium isolates.

In order to simplify the data analysis, we have decided to partition the frequent MLVA profiles into 2 groups taking into account the number of 27-bp repeats in locus STTR3: group 1 (32.0% of the sample population) with allele 211 and group 2 (12.3%) with allele 311 (Figure 2.2). The MLVA profiles of each group were single-locus and single-repeat variants of other MLVA profiles in the group. The frequent MLVA profiles were observed throughout the 3-year period (Figure 2.3), except for profile 3-15-13-NA-311, belonging to group 2, which was not detected in 2010. Nine monophasic variant isolates were typed with a MLVA profile belonging to group 1. Rare MLVA profiles 3-11-11-NA-211 and 3-11-3-NA-211 were detected in the other monophasic variant isolates.

The *S.* Typhimurium isolates of both groups differed with respect to the presence of phage types (Figure 2.4). DT193, DT195, U302 and NT isolates were observed in both groups, whilst DT120, DT110, DT138, U311 and RDNC isolates were common in group 1 and scarce in group 2. Interestingly, 30 out of the 48 (62.5%) RDNC isolates in group 1 had MLVA profile 3-12-11-NA-211 (Figure 2.4). Frequent phage types DT104 and DT12 were observed only for a single isolate of group 1 and of group 2, respectively. Within the frequent phage types, the number of distinct MLVA profiles ranged from 41 for U302 up

Table 2.1: Overview of antimicrobial resistance and MLVA characteristics in relation to the phage type (n = 1,420).

Phage type	No. isolates (% of total)	No. susceptible isolates (%)	No. MDR isolates (%)	Most common resistance patterns (% of isolates with resistance pattern)	No. MLVA types	Most common MLVA profiles (% of isolates with MLVA profile)
DT193	284 (20.0)	4 (1.4)	156 (54.9)	ASSuT (35.6), A (28.9), ASSu (7.0)	79	3-12-10-NA-211 (7.7), 3-13-10-NA-211 (7.7)
DT195	251 (17.7)	11 (4.4)	82 (32.7)	A (47.8), ASSuT (14.3), ASSu (6.8)	72	3-12-10-NA-211 (10.0), 3-15-13-NA-311 (8.4)
DT120	231 (16.3)	11 (4.8)	187 (81.0)	ASSuT (48.1), ASSuSxtTTmp (15.6), ASSu (6.9)	76	3-12-9-NA-211 (11.7), 3-12-10-NA-211 (10.4)
DT104	90 (6.3)	2 (2.2)	75 (83.3)	AAmcCSSuT (33.3), ACSSuT (15.6), AAmcCNaSSuT (10.0)	70	3-12-21-14-NA (4.4), 3-15-10-23-410 (4.4)
DT12	58 (4.1)	8 (13.8)	30 (51.7)	ACSSuT (19.0), SSu (13.8), AAmcCSSuT (8.6), AAmcCNaSSuT (8.6)	50	3-13-13-16-311 (5.2), 3-14-19-16-NA (5.2)
U302	55 (3.9)	17 (30.9)	12 (21.8)	A (29.1), ASSu (5.5), T (5.5), ASuSxtTmp (5.5)	41	3-16-10-NA-311 (7.3), 4-9-12-9-211 (5.5)
NT	101 (7.1)	15 (14.9)	57 (56.4)	ASSuT (26.7), A (18.8), ASSuSxtTTmp (5.0)	60	3-13-10-NA-211 (12.9), 3-12-10-NA-211 (5.0), 3-15-11-NA-311 (5.0), 4-9-10-10-111 (5.0)
RDNC	131 (9.2)	37 (28.2)	33 (25.2)	ASSu (25.2), ASSuT (11.5), SSuT (6.9)	64	3-12-11-NA-211 (22.9), 3-17-9-NA-211 (4.6)
Other	219 (15.4)	64 (29.2)	93 (42.5)	ASSuT (25.6), T (9.1), A (6.8), ASSu (6.8)	117	3-12-10-NA-211 (7.8), 3-12-11-NA-211 (7.8)

MDR: multidrug resistant; MLVA: multiple-locus variable-number of tandem repeats analysis; MLVA profile: STTR9-STTR5-STTR6-STTR10-STTR3; NT: not-typable; RDNC: reacts-but-does-not-conform; Other: phage types with less than 3% occurrence in the sample population; A: ampicillin; Amc: amoxicillin plus clavulanic acid; C: chloramphenicol; Na: nalidixic acid; S: streptomycin; Su: sulphonamides; Sxt: trimethoprim plus sulfamethoxazole, T: tetracycline; Tmp: trimethoprim.

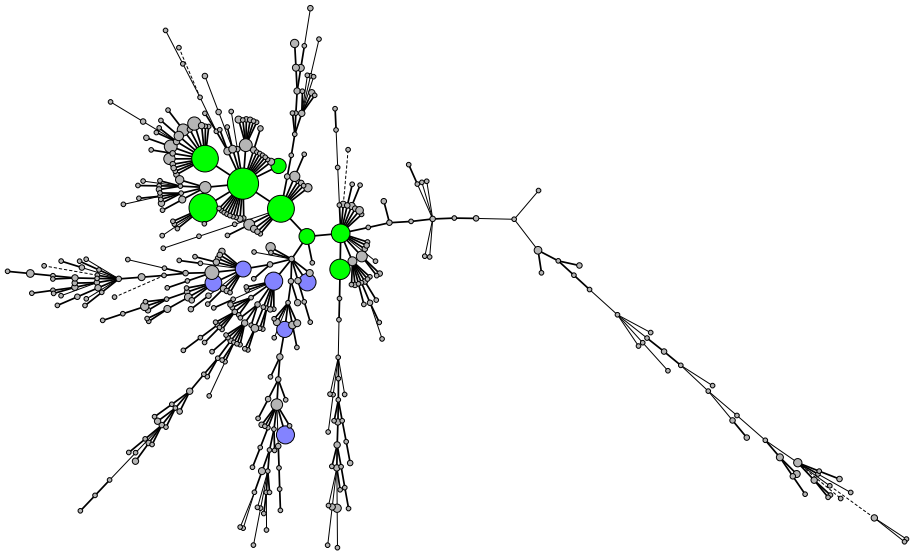

Figure 2.2: Minimum spanning tree calculated of MLVA profiles of *Salmonella* Typhimurium isolates over the 3-year period. N = 1,420. Each node represents a different MLVA profile. Node colour denotes frequent MLVA profiles of group 1 (green) and group 2 (blue), and rare MLVA profiles (grey). Node size is proportional to the number of isolates with that MLVA profile. Branch thickness indicates how many loci are different in the MLVA profiles of the connected nodes. Thick solid lines connect nodes that differ by 1 MLVA locus, thin solid lines connect nodes that differ by 2 MLVA loci and dashed lines connect nodes that differ by 3 MLVA loci. MLVA: multiple-locus variable-number of tandem repeats analysis.

to 79 for DT193 (Table 2.1). Except for phage types DT104 and DT12, which showed many different MLVA profiles, the most common MLVA profile for each phage type belonged to the frequent MLVA groups.
Also with respect to antimicrobial resistance there were dissimilarities between both groups of frequent MLVA profiles. Whereas group 1 comprised 73.6% MDR isolates, group 2 included 6.9% MDR isolates. Correspondingly, the most common antimicrobial resistance patterns were ASSuT (50.4%) in group 1 and a single resistance to ampicillin (86.3%) in group 2. Both groups displayed equivalent numbers of isolates which were susceptible to all antibiotics tested (3.3% and 2.3% for group 1 and 2, respectively).

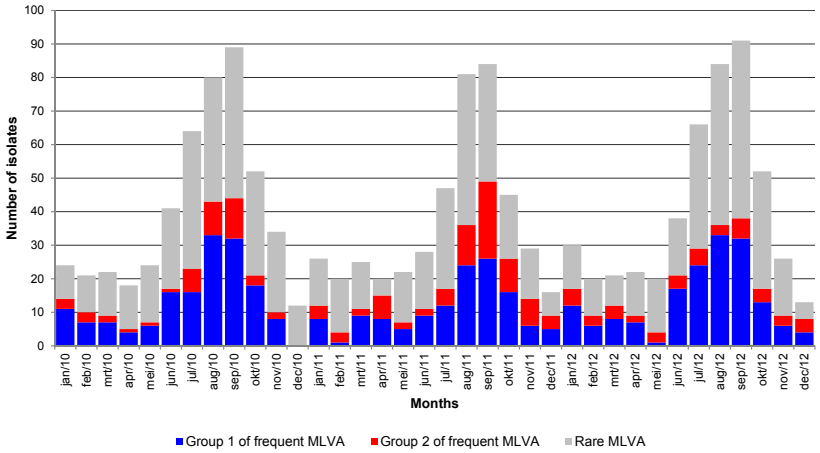

Figure 2.3: Monthly distribution of *Salmonella* Typhimurium isolates over the 3-year period according to MLVA profile. N = 1,407 as for 13 isolates in this study the isolation date is unknown. Group 1 of frequent MLVA profiles: 3-12-8-NA-211, 3-12-9-NA-211, 3-12-10-NA-211, 3-12-11-NA-211, 3-13-8-NA-211, 3-13-9-NA-211, 3-13-10-NA-211 and 3-13-11-NA-211. Group 2 of frequent MLVA profiles: 3-14-11-NA-311, 3-15-10-NA-311, 3-15-11-NA-311, 3-15-12-NA-311, 3-15-13-NA-311 and 3-16-10-NA-311. Rare MLVA: MLVA profiles that have less than 1.4% occurrence in the sample population. MLVA: multiple-locus variable-number of tandem repeats analysis.

2.3.4 Diversity of phage typing, antimicrobial susceptibility testing and MLVA

To compare the discriminatory power of MLVA to that of the phenotyping methods used, Simpson's and Shannon's diversity indices were calculated. The discriminatory power of a subtyping method is defined as the ability to distinguish between unrelated isolates (Hunter and Gaston 1988) and the higher the value of a diversity index, the higher the discriminatory power of the subtyping method. Simpson's D ranges from 0 to 1 and gives the probability that 2 randomly sampled and unrelated isolates will have a different subtype (Hunter and Gaston 1988). Rare subtypes, which apply to only a small number of isolates, will have a small contribution to Simpson's index and as such, the number of subtypes has little influence on Simpson's index (Boxrud *et al.* 2007). Shannon's H' is an indicator for subtype richness (Boxrud *et al.* 2007) and its highest value is $\ln S$, where S is the number of subtypes. Shannon's E is a measure for the evenness of the subtype distribution (Boxrud *et al.* 2007) and

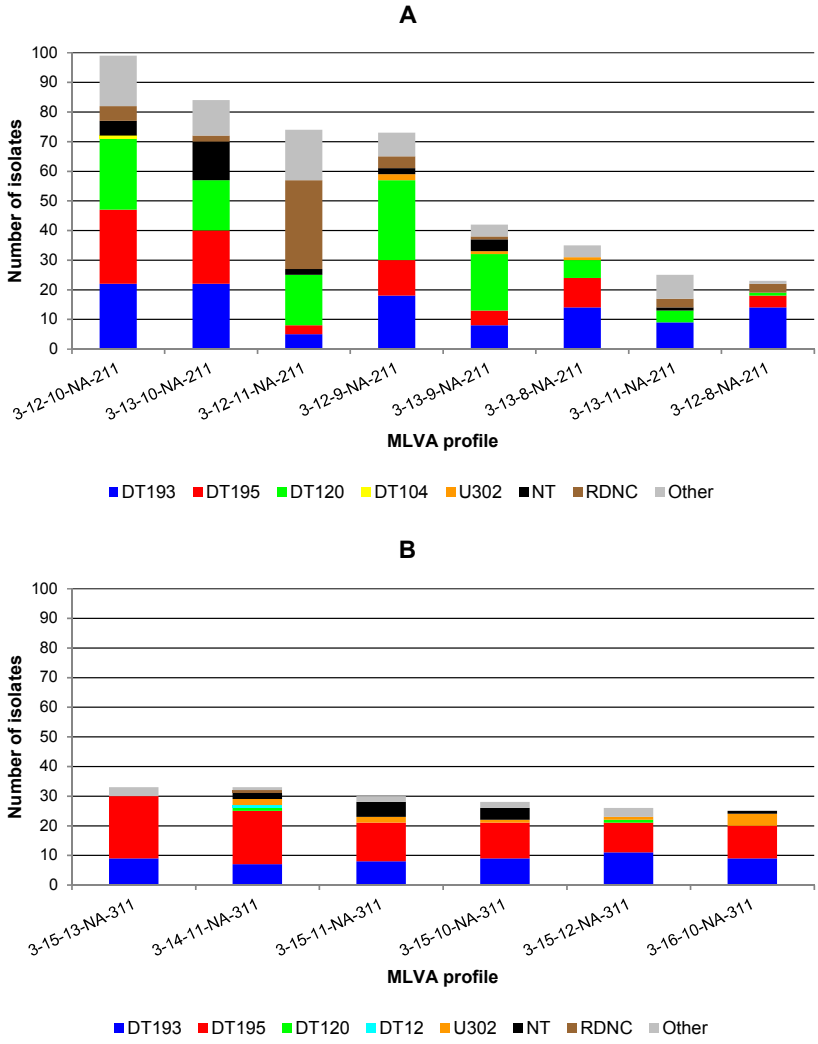

Figure 2.4: Prevalence of *Salmonella* Typhimurium isolates in this study with a frequent MLVA profile. Figure 2.4A: Group 1 of frequent MLVA profiles (n = 455) for which other phage types are DT7, DT20, DT56, DT110, DT116, DT138, DT185, DT194 and U311. Figure 2.4B: Group 2 of frequent MLVA profiles (n = 175) for which other phage types are DT35, DT41, DT99, DT194, DT208 and U311. MLVA: multiple-locus variable-number of tandem repeats analysis; NT: not-typable; RDNC: reacts-but-does–not-conform.

has 1 as maximum value.
Calculated values for Simpson's D, Shannon's H' and E indices were, respectively, 0.88, 2.49 and 0.67 for phage typing, 0.88, 2.84 and 0.62 for antimicrobial susceptibility testing, and 0.98, 4.92 and 0.82 for MLVA. These values suggest that MLVA has a higher discriminatory power than both phage typing and antimicrobial susceptibility testing.

2.3.5 Typing of outbreak isolates

To evaluate the added value of MLVA in case of outbreak detection, five isolates originating from an outbreak in a day nursery in 2011 were included in this study. These isolates were characterised as phage type DT138 with antimicrobial resistance pattern ASSu and MLVA profile 3-13-11-NA-211. Whereas each of these typing results independently are shared with isolates which are not related to the outbreak and hence would not allow to identify specific clusters of outbreak isolates, the combination of their MLVA profile and phage type is unique to the outbreak isolates compared to all other isolates in the 3-year period.
The outbreak isolates were also typed with PFGE (data not shown), which resulted in identical patterns for these isolates. However, this PFGE pattern could not be compared to other isolates in this study, as it was not feasible for the Belgian NRCSS to perform PFGE on all randomly sampled isolates during the 3-year period.

2.3.6 Stability of MLVA loci

To evaluate the *in vitro* stability of the number of tandem repeats in the MLVA loci, 31 *S.* Typhimurium isolates were subjected to a serial passage experiment. The isolates selected for this stability experiment covered all frequent phage types and 17 different MLVA profiles, of which 6 and 3 MLVA profiles belonged to respectively frequent MLVA groups 1 and 2 (Table 2.2).
Among the 20 isolates with a frequent MLVA profile, 3 isolates (15.0%) presented a single-repeat variant at locus STTR6 during the serial passage experiment. Among the *S.* Typhimurium with a rare MLVA profile, variations of the initial MLVA profile were observed in 6 out of 11 isolates (54.5%). These variations were not only noticed in locus STTR6, but also in loci STTR5, STTR9 and STTR10. In one isolate with a rare MLVA profile, variations of the initial MLVA profile were observed in loci STTR5, STTR6 and STTR9, and another isolate with a rare MLVA profile presented 3 different variations at locus STTR10 (Table 2.2). For loci STTR6 and STTR5 only single-repeat variants were seen,

whereas the varying allele for locus STTR9 differed 2 repeats from the original allele and for locus STTR10 there were differences from 1 up to 7 repeats between original and varying alleles.

2.4 Discussion

S. Typhimurium is the most frequently isolated serovar from human patients in Belgium and hence subtyping of this serovar is very important for outbreak detection and tracing outbreak sources. The Belgian NRCSS relies on phage typing and antimicrobial susceptibility testing for routine surveillance of *S.* Typhimurium, complemented with PFGE during outbreak investigations. PFGE, which is widely considered as the gold standard for subtyping of *Salmonella*, is a labour intensive and time consuming technique and therefore implementation of this subtyping method for routine surveillance is not realisable for the Belgian NRCSS. MLVA, which requires less hands-on time and allows faster typing and easy inter-laboratory comparison of results, has been adopted by several European countries for surveillance and detection and investigation of outbreaks (Lindstedt *et al.* 2013). For evaluation of the capability of MLVA typing for surveillance and outbreak detection of human *S.* Typhimurium in Belgium, 1,420 isolates collected over the 3-year period 2010–2012 were characterised by phage typing, antimicrobial susceptibility testing and 5-loci MLVA.

Our study shows that phage types DT193, DT195, DT120, DT104, DT12 and U302 dominated the *S.* Typhimurium population. This predominance of a small number of phage types was also observed in other countries (Weill *et al.* 2006; Torpdahl *et al.* 2007; Prendergast *et al.* 2011; Sintchenko *et al.* 2012) and reduces the capacity of phage typing to discriminate outbreak isolates. Additionally, 16.3% of the isolates are categorized as RDNC or NT, which lowers the proportion of isolates that are subtyped by this technique, and hence its suitability for surveillance and outbreak detection.

Antimicrobial susceptibility testing is another phenotyping method used in public health surveillance. In this study, ASSuT is the leading antimicrobial resistance pattern, which is observed with or without additional resistances for 42.5% of the *S.* Typhimurium isolates. The ASSuT pattern has been reported in France, UK, Spain, Luxembourg, Italy and Germany in association with DT193, DT120 and NT isolates and is often connected to the monophasic variant of *S.* Typhimurium (Hampton *et al.* 1995; Soler *et al.* 2006; Weill *et al.* 2006; Mossong *et al.* 2007; Graziani *et al.* 2008; Hauser *et al.* 2010; Hopkins *et al.* 2012); but also, differently from Belgium, in combination with phage type U302 in Denmark and Italy (Ethelberg *et al.* 2004; Dionisi *et al.* 2009). Contrary to other European countries (Mossong *et al.* 2007; Hauser *et al.* 2010; Hopkins

Table 2.2: Overview of isolates and outcome of the stability experiment (n = 31).

ID number	Phage type	Initial MLVA profile	Frequent MLVA group	Varying number of repeats observed (passage of first occurrence)
11-1129	DT104	3-12-9-NA-211	1	-
12-2475	DT193	3-12-9-NA-211	1	-
12-3110	DT193	3-12-10-NA-211	1	STTR6: 11 (25)
11-0841	DT120	3-12-11-NA-211	1	-
11-2577	DT195	3-12-11-NA-211	1	-
12-3096	DT193	3-13-8-NA-211	1	-
11-1058	DT193	3-13-9-NA-211	1	-
12-0828	DT120	3-13-9-NA-211	1	-
12-1779	DT120	3-13-9-NA-211	1	-
11-0050	DT120	3-13-10-NA-211	1	-
11-2038	DT120	3-13-10-NA-211	1	-
11-2650	DT195	3-13-10-NA-211	1	-
11-2847	DT195	3-13-10-NA-211	1	-
11-3000	U302	3-15-11-NA-311	2	-
12-1651	DT195	3-15-11-NA-311	2	-
11-2321	DT195	3-15-13-NA-311	2	STTR6: 12 (45)
11-3418	DT193	3-15-13-NA-311	2	STTR6: 12 (35)
11-3445	DT193	3-15-13-NA-311	2	-
11-2326	U302	3-16-10-NA-311	2	-
12-1918	U302	3-16-10-NA-311	2	-
11-2630	DT12	3-14-10-NA-311	-	STTR5: 13 (10)
11-0676	DT195	3-14-10-NA-311	-	-
11-3355	DT104	3-14-11-21-311	-	STTR10: 20 (35)
11-1160	DT104	3-14-18-14-311	-	-
10-02975	DT12	3-14-18-14-311	-	STTR10: 9 (15); STTR10: 7 (50); STTR10: 17 (50)
11-0008	DT104	3-15-10-23-311	-	-
11-0210	DT104	3-15-10-23-311	-	STTR6: 11 (30)
11-0444	U302	3-18-16-17-311	-	STTR10: 20 (20)
11-0335	U302	4-14-18-7-211	-	-
11-3448	DT12	5-13-15-8-211	-	-
11-3005	DT12	5-14-11-8-211	-	STTR9: 3 (10); STTR5: 13 (5); STTR6: 10 (10); STTR6: 12 (20)

MLVA: multiple-locus variable-number of tandem repeats analysis; MLVA profile: STTR9-STTR5-STTR6-STTR10-STTR3.

et al. 2010; European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2013), Belgium has a low rate (0.8% of the randomly sampled set) of human monophasic variant isolates of *S.* Typhimurium.

Molecular techniques like MLVA are generally considered to improve surveillance and detection of outbreaks and their sources because of different advantages of this method. The possibility to present the result as a string of numbers is one of the strengths of MLVA and allows easy sharing across country borders and setup of databases of MLVA profiles (Heck 2009; Larsson *et al.* 2009). Yet, laboratories have to agree on the set of calibration strains and on the nomenclature used, so that different laboratories report the same MLVA profiles for the same isolates (Dyet *et al.* 2011; Larsson *et al.* 2013). Another already reported asset of MLVA subtyping is the high discriminatory power of this technology (Heck 2009; Lindstedt *et al.* 2013). MLVA targeting 5 loci divided the investigated *S.* Typhimurium collection into 414 distinct profiles and indeed allowed for discrimination within isolates of the same phage type, which is in concordance with previous studies (Lindstedt *et al.* 2004; Torpdahl *et al.* 2007; Prendergast *et al.* 2011; Ross *et al.* 2011). However, 14 frequent MLVA profiles included 44.4% of the *S.* Typhimurium isolates. These 14 profiles were partitioned into 2 groups based on the number of 27-bp repeats in locus STTR3. The two MLVA groups differ regarding associated phage types and frequency of MDR isolates. Most of the MLVA profiles in the first group have already been described in other European countries (Hopkins *et al.* 2010; Prendergast *et al.* 2011; Gossner *et al.* 2012; Hopkins *et al.* 2012; Kuhn *et al.* 2013) and are present in the MLVA-NET database (Guigon *et al.* 2008). Of the second group, 2 profiles have been reported in literature (Prendergast *et al.* 2011; Fabre *et al.* 2012) and 1 additional profile has been found in the MLVA-NET database (Guigon *et al.* 2008).

As in our study almost half of the *S.* Typhimurium population is characterised by only 14 different MLVA profiles, while the total population is represented by 414 distinct MLVA profiles, questions on the discriminating ability of this subtyping technique could be raised. In order to compare the discriminatory power in an objective manner, diversity indices were calculated for the different subtyping methods used in this study. A higher Simpson's index of diversity (D) was obtained for MLVA as compared to phage typing and antimicrobial susceptibility testing, but the latter techniques showed the same D value. This demonstrates the small influence of the number of subtypes on Simpson's diversity index, as antimicrobial susceptibility testing revealed 99 distinct patterns compared to only 41 detected phage types, while resulting in equal discriminatory power according to Simpson's index. Shannon's diversity (H') and equitability (E) indices, which are indicators of the number of subtypes and of the evenness of the distribution of these subtypes (Boxrud *et al.* 2007),

respectively, offer a more differentiated measure for comparison of discriminatory power, *i.e.* aiming at a high number of evenly distributed subtypes. Shannon's H' denotes that antimicrobial susceptibility testing was more able to discriminate between unrelated isolates than phage typing, but Shannon's E values indicate that phage types are more equally distributed than antimicrobial resistance patterns. Both Simpson's and Shannon's indices suggest that MLVA has a higher discriminatory power than phage typing and antimicrobial resistance testing, as was concluded from earlier studies (Lindstedt *et al.* 2004; Boxrud *et al.* 2007; Torpdahl *et al.* 2007; Prendergast *et al.* 2011), and which would result in an improved surveillance and detection of possible outbreaks. However, as *S.* Typhimurium isolates in this study are dominated by 14 frequent MLVA profiles and 6 frequent phage types, care must be taken with the interpretation of these indices. *S.* Typhimurium isolates with a frequent MLVA profile show from 4 up to 11 different phage types and visa-versa, *S.* Typhimurium isolates with a frequent phage type show from 41 up to 79 different MLVA profiles. This would implicate that all isolates with a same frequent MLVA profile or all isolates with a same frequent phage type do not originate from a single *S.* Typhimurium strain. Consequently, for detection of an outbreak or the source of an outbreak with isolates with a frequent MLVA profile or phage type, when using the in this study applied 5-loci MLVA, a combination of both subtyping techniques might be necessary, which was also previously observed (Ross *et al.* 2011; Kuhn *et al.* 2013). This was also the case for the 5 outbreak isolates presented in this study. On the contrary, MLVA or phage typing might be sufficient to distinguish a cluster of isolates with rare MLVA profiles or phage types. However, for locating the source of a human outbreak in animal, food or environmental isolates, we must take into account that rare MLVA profiles or phage types in human isolates might be common MLVA profiles or phage types in animals, food or the environment (Torpdahl *et al.* 2007), which would complicate the designation of a single source.

In addition to the discriminatory power of a subtyping method, also the stability of the assessed markers should be taken into account. From our serial passage experiment on 31 *S.* Typhimurium isolates, we observed that 71.0% of the MLVA profiles remain stable *in vitro*. However, variations due to passages occurred and most variations were seen among isolates with a rare MLVA profile, which presented variations in more isolates and at more loci. Single-locus variants constituted 8 of the 9 isolates with varying alleles. MLVA loci STTR 6 and STTR5 displayed only single-repeat variants, in contrast to STTR9 and STTR10, where differences up to 7 repeats from the original number of repeats were detected. Locus STTR3 remained stable throughout the experiment, which was also observed in a serial passage experiment by Barua *et al.* (2013) on 4 DT41 *S.* Typhimurium isolates of poultry origin. In contrast to the results of our experiment, in which more sampling points were included, the MLVA profiles

of their DT41 isolates remained stable at locus STTR9 and only single-repeat variants at locus STTR10 and differences up to 4 repeats at locus STTR6 were detected (Barua *et al.* 2013). Nevertheless, these passage experiments point out that we cannot rule out the possibility that isolates with closely related MLVA profiles are not clonal.

In conclusion, based on Simpson's and Shannon's indices, 5-loci MLVA has a higher discriminatory power for the 1,420 *S.* Typhimurium collected during the 3-year period 2010–2012, and can thus improve public health surveillance. However, outbreak detection with MLVA is not straightforward, since for isolates with a frequent MLVA profile, phage typing is still necessary to achieve a unique, combined subtyping result in this study. Also, during investigation of extended outbreaks, variations in MLVA profiles should be taken into account, since *in vitro* stability could not be confirmed for all 5 MLVA loci. Therefore, improvement of the 5-loci MLVA scheme may be desirable. Indeed, the MLVA scheme which is currently implemented in Europe, and which has been used in this study, targets 5 VNTR loci of which 1 locus did not amplify in 75.0% of the 1,420 isolates and 4 loci showed instabilities during a serial passage experiment. Since there are more than 30 VNTR loci described in literature for *S.* Typhimurium (Le Flèche *et al.* 2001; Lindstedt *et al.* 2003; Lindstedt *et al.* 2004; Ramisse *et al.* 2004; Witonski *et al.* 2006; Chiou *et al.* 2010; PulseNet USA 2013), there may be possibilities to improve the assay. Other opportunities may be the further development and implementation of CRISPR genotyping (Fabre *et al.* 2012) for subtyping of *S.* Typhimurium, so that the need of a second subtyping assay for genetically homogeneous populations is eliminated, or the development of a subtyping assay that combines different typing methods in a single molecular assay with stable markers.

Supporting information

Dataset S1 with all typing data of the studied *S.* Typhimurium population (n = 1,420) is available at http://journals.plos.org/plosone/article/asset?unique&id=info:doi/10.1371/journal.pone.0084055.s001 (file format: csv).

Acknowledgements

Capillary electrophoresis for MLVA was performed at the Platform Biotechnology and Molecular Biology at the Scientific Institute of Public Health (WIV-ISP).

Funding

This work was supported by: RP/PJ WIV-ISP (SalMolType), the Federal Public Service of Health, Food Chain Safety and Environment. The National Reference Centre for *Salmonella* and *Shigella* is partially supported by the Belgian Ministry of Social Affairs through a fund within the Health Insurance System. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors have declared that no competing interests exist.

Chapter 3

Guidelines for optimisation of a multiplex oligonucleotide ligation-PCR for characterisation of microbial pathogens in a microsphere suspension array

This chapter was previously published as:
V. Wuyts, N. H. C. Roosens, S. Bertrand, K. Marchal, and S. C. J. De Keersmaecker (2015c). "Guidelines for optimisation of a multiplex oligonucleotide ligation-PCR for characterisation of microbial pathogens in a microsphere suspension array". *BioMed Research International*, article ID 790170. DOI: 10.1155/2015/790170

Authors' contributions

The experimental work was performed by V. Wuyts. The experiments were designed by V. Wuyts, N. H. C. Roosens and S. C. J. De Keersmaecker. Data analysis and writing of the paper was done by V. Wuyts with the critical input of S. C. J. De Keersmaecker. N. H. C. Roosens, S. Bertrand and K. Marchal reviewed the paper and also contributed reagents, materials and analysis tools.

Abstract

With multiplex oligonucleotide ligation-PCR (MOL-PCR) different molecular markers can be simultaneously analysed in a single assay and high levels of multiplexing can be achieved in high-throughput format. As such, MOL-PCR is a convenient solution for microbial detection and identification assays where many markers should be analysed, including for routine further characterisation of an identified microbial pathogenic isolate. For an assay aimed at routine use, optimisation in terms of differentiation between positive and negative results and of cost and effort is indispensable. As MOL-PCR includes a multiplex ligation step, followed by a singleplex PCR and analysis with microspheres on a Luminex device, several parameters are accessible for optimisation. Although MOL-PCR performance may be influenced by the markers used in the assay and the targeted bacterial species, evaluation of the method of DNA isolation, the probe concentration, the amount of microspheres, and the concentration of reporter dye is advisable in the development of any MOL-PCR assay. Therefore, we here describe our observations made during the optimisation of a 20-plex MOL-PCR assay for subtyping of *Salmonella* Typhimurium with the aim to provide a possible workflow as guidance for the development and optimisation of a MOL-PCR assay for the characterisation of other microbial pathogens.

3.1 Introduction

Characterisation of microbial pathogens beyond the species and subspecies level, that is, subtyping, requires more than one marker and these markers are usually combined in a multiplex assay for time-effectiveness of the assay. Evaluation of a multiplex assay can be facilitated by the Luminex technology, which has the capacity to analyse up to 500 markers in a single sample. Luminex assays are bead-based suspension arrays in which, in the case of DNA-based assays, fluorescently labelled oligonucleotides hybridise to probes that are coupled to distinctly coloured microspheres (up to 500 different colours). Fluorescently labelled oligonucleotides can be created by different types of assays, such as multiplex PCR (direct hybridisation assay), oligo ligation assay (OLA), allele-specific primer extension (ASPE) (Angeloni *et al.* 2014), and multiplex oligonucleotide ligation-PCR (MOL-PCR).

MOL-PCR was first described by Deshpande *et al.* (2010) as a powerful tool for detection of microbial pathogens allowing to combine analysis of multiple types of markers like unique sequences, indels, repeats, or single nucleotide polymorphisms (SNPs) in a single multiplex reaction. With MOL-PCR high levels of multiplexing can be achieved, because the multiplexing step is a ligation rather than a PCR and signals are amplified in a singleplex PCR. MOL-PCR is a variant on multiplex ligation-dependent probe amplification (MLPA) (Schouten *et al.* 2002) in which the overnight hybridisation step and subsequent ligation are replaced by cycles of hybridisation and ligation by a thermostable ligase. The read-out of MLPA products occurs through fragment sizing by capillary electrophoresis (Schouten *et al.* 2002), but also applications with analysis on a Luminex device have been reported (Kipp *et al.* 2011; Bergval *et al.* 2012).

Multiplex ligation-based assays as MOL-PCR and MLPA have been reported as efficient assays for the diagnosis of human genetic diseases (Schouten *et al.* 2002; Slater *et al.* 2004; Xu *et al.* 2013; Garin *et al.* 2014; Kasatkar *et al.* 2014; Marcinkowska-Swojak *et al.* 2014), the detection of viruses (Reijans *et al.* 2008; Theelen *et al.* 2010; De Smet *et al.* 2012) and bacteria (Deshpande *et al.* 2010; Chung *et al.* 2012; Berning *et al.* 2014), and characterisation of pathogens, including subtyping (Bergval *et al.* 2008; Beyene *et al.* 2009; Stucki *et al.* 2012; Pham Thanh *et al.* 2013; Thierry *et al.* 2013; Cornelius *et al.* 2014).

Although MOL-PCR is increasingly used for characterisation of microbial pathogens on pure isolates, little is found in literature on the steps taken during optimisation of the assay, leading to the final, published protocol. A paper on *Mycobacterium tuberculosis* refers to a general protocol and gives little detail on the reaction conditions (Bergval *et al.* 2012). Deshpande *et al.* (2010), Stucki *et al.* (2012) and Thierry *et al.* (2013) provide the reaction conditions in detail, but refrain from comprehensive optimisation results, although encountered issues with some aspects in the assay and their solutions are discussed by Thierry

et al. (2013).
As we consider that optimisation experiments might contain valuable information for other scientists starting with the development of a MOL-PCR assay for their microbial pathogen of interest, we describe here the observations we made during the optimisation process of a 20-plex MOL-PCR assay, which is one of three MOL-PCR assays for the subtyping of *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium), as an illustration of which parameters could be optimised for a MOL-PCR assay for other pathogens.
Even though optimisation of MOL-PCR might depend on the markers in the assay and on the bacterial species for which the assay is intended, we point out the main parameters of the MOL-PCR assay that could be evaluated during the design of a new assay for another microbial pathogen.

3.2 Materials and methods

3.2.1 Probe design

The MOL-PCR assay taken as an example in this study contains 18 discriminating markers and 2 markers as internal positive control of the DNA template. One internal positive control marker targets a SNP, while all other markers detect the presence of a unique sequence. Each marker requires an upstream and downstream probe that anneal adjacent to each other on the target DNA. The upstream probes consist of a 5' 20 bp T7 primer sequence, an internal 24 bp anti-TAG sequence, and a 3' 19 bp to 27 bp target-specific sequence. The downstream probes are made up of a 5' 18 bp to 28 bp target-specific sequence and at 3' the 20 bp reverse-complement of a T3 primer. The downstream probes are 5' phosphorylated to enable ligation to the upstream probe by the DNA ligase. All probe pairs were designed with Visual OMP (DNA Software) so that the complex of target-specific DNA, upstream probe, and downstream probe had an effective change of Gibb's free energy (Eff ΔG°) of -31.15 to -25.13 kcal mol^{-1} and a melting temperature of 56 to 64°C at the conditions of the multiplex ligation reaction, except for one discriminating and one internal DNA control marker for which the complex had a melting temperature of 68°C and 52°C, respectively. As the aim of this study is not to provide a new subtyping method for *S.* Typhimurium, but to elaborate on important parameters to be evaluated during the development of a MOL-PCR assay, the sequences of the probes and primers used are less relevant for the key message of this paper. However, sequences of probes (partially based on Fang *et al.* (2012)) and primers are available upon request. All probes and primers were ordered from Eurogentec with a RP-Cartridge-Gold purification

for upstream probes and T7 primer and a reverse phase HPLC purification for downstream probes and biotinylated T3 primer. A HPLC purification was chosen for the downstream probes and T3 primer, since the supplier did not offer the basic RP-Cartridge-Gold purification for these modified oligonucleotides.

3.2.2 Bacterial isolates and DNA isolation

All *S.* Typhimurium isolates in this study were received from the Belgian National Reference Centre for *Salmonella* and *Shigella*. The test panel consisted of 6 *S.* Typhimurium isolates, which were selected so that for each of the 18 discriminating markers at least 1 positive and 1 negative result was obtained.
Different methods for DNA isolation based on heat lysis were tested, that is, heat lysis in water and heat lysis in a commercial product for DNA purification were adapted for a single colony from previous reports (Bergval *et al.* 2012; Luminex Corporation 2012; Stucki *et al.* 2012; Thierry *et al.* 2013) (Table 3.1: DNA isolation). Isolates were grown overnight (14 to 20 hours) at 37℃ on LB agar (Merck). In the first method, a single colony was dissolved in 50 to 300 μl sterile deionised water and incubated at 100℃ for 10 min. After cooling for at least 5 min at 4℃ and centrifugation for 10 min at 11000 g, the supernatant was stored at −20℃ and used for further analysis. In the second method, a single colony was treated as described in the product insert of the xMAP *Salmonella* Serotyping Assay Kit (2012). In short, a single colony was added to 20 μl InstaGene Matrix (Bio-Rad) and in a thermal cycler the following programme was run: 56℃ for 10 min, 99℃ for 5 min, and 4℃ forever. After addition of 100 μl of nuclease free distilled water (Life Technologies), the tube was spun for 5 min in a microcentrifuge and 50 μl of the supernatant was stored at −20℃ for further use.
A positive control template DNA was created by combining a single colony of each of 5 different isolates in one tube, which was treated in the same way as the isolates of the test panel. The combination of these 5 isolates yielded a positive reaction for each of the 20 markers in the assay.

3.2.3 Multiplex oligonucleotide ligation

The multiplex ligation reaction mix combined 1 to 5 nM of each of the 40 probes with 1× *Taq* DNA ligase reaction buffer (New England BioLabs), 2 to 6 units of *Taq* DNA ligase (New England BioLabs), and 2 or 4 μl of template DNA and was brought to a final volume of 10 μl with nuclease free distilled water (Life Technologies). The thermal cycling programme (Swift MaxPro, Esco) included 5 or 10 min of denaturation at 95℃ followed by 30 cycles of 25 s at 94℃ and 30 s at

Table 3.1: Test and reference conditions for the optimisation experiments.

	Preparatory work				Multiplex oligonucleotide ligation									
Condition	DNA isolation[a]				Probe concentration (nM)			DNA volume (μl)		*Taq* DNA ligase (units)			Initial denaturation (min)	
experiment	300	100	50	IGM	1	2	5	2	4	2	4	6	5	10
1	R	ns	**ls**	ls			x	x			x			x
2			x		gs	**R**	gs	x			x			x
3			x			x		**R**	gs	**R**	gs	gs		x
4			x			x		x		x			R	**ls**
5			x			x		x		x				x
6			x			x		x		x				x
7			x			x		x		x				x
8			x			x		x		x				x
9			x			x		x		x				x
10			x			x		x		x				x

R: reference condition; x: condition used in experiment; ns: test condition which is not significant at $\alpha = 0.01$ in two-sided hypothesis test; ls: test condition which is significant at $\alpha = 0.01$ in one-sided hypothesis test with alternative hypothesis true location shift is less than 0; gs: test condition which is significant at $\alpha = 0.01$ in one-sided hypothesis test with alternative hypothesis true location shift is greater than 0; bold type: optimal condition as concluded from the experiment.

[a]300: single colony in 300 μl sterile deionised water; 100: single colony in 100 μl sterile deionised water; 50: single colony in 50 μl sterile deionised water; IGM: InstaGene Matrix.

Table 3.1 *continued*

Condition	PCR						Hybridisation and analysis on MAGPIX									
	Ligation product volume (μl)		*Taq* DNA polymerase (units)		Cycles		PCR product volume (μl)			Microspheres per reaction			SAPE concentration (μg/ml)		Reporter dye	
Experiment	3	5	0.25	0.5	35	40	2.5	5	all	350	750	2500	4	10	SAPE	Alexa 532
1	x		x		x			x			x		x		x	
2	x		x		x			x			x		x		x	
3	x		x		x			x			x		x		x	
4	x		x		x			x			x		x		x	
5	**R**	ns	x		x			x			x		x		x	
6	x		**R**	gs	x			x			x		x		x	
7	x		x		**R**	gs		x			x		x		x	
8	x		x		x		R	**ls**	ls		x		x		x	
9	x		x		x			x		ls	**ls**	R	**ls**	R	x	
10	x		x		x			x			x		x		**R**	gs

R: reference condition; x: condition used in experiment; ls: test condition which is significant at $\alpha = 0.01$ in one-sided hypothesis test with alternative hypothesis true location shift is less than 0; gs: test condition which is significant at $\alpha = 0.01$ in one-sided hypothesis test with alternative hypothesis true location shift is greater than 0; bold type: optimal condition as concluded from the experiment; SAPE: streptavidin-R-phycoerythrin.

58℃ (Table 3.1: Multiplex oligonucleotide ligation). Each experiment included a positive control for the reaction and a no-template-control (NTC) to measure background signal for which the template DNA was replaced by, respectively, positive control template DNA and nuclease free distilled water. The results of neither the positive nor the negative control were taken into account for the statistical analysis and are not included in the figures shown, since these results are not representative of a bacterial isolate to be characterised.

3.2.4 Singleplex polymerase chain reaction

The singleplex PCR reaction (Table 3.1: PCR) was performed in a final volume of 10 μl which included 1× HotStarTaq PCR buffer (Qiagen), 125 nM T7 primer (Eurogentec), 500 nM 5'-biotin-T3 or 5'-Alexa 532-T3 primer (Eurogentec), 200 μM of each dNTP (Thermo Scientific), 0.25 or 0.5 units HotStarTaq DNA polymerase (Qiagen), and 3 or 5 μl of ligase product. The following protocol was run in a thermal cycler (SwiftMaxPro, Esco): 15 min at 95℃, 35 to 40 cycles of 30 s at 94℃, 30 s at 60℃, and 30 s at 72℃, 5 min at 72℃.

3.2.5 Hybridisation to microspheres and analysis on Luminex platform

An adapted version of the manufacturer's no wash protocol (Angeloni *et al.* 2014) was used for the hybridisation reaction (Table 3.1: Hybridisation and analysis on MAGPIX). Twenty regions of MagPlex-TAG microspheres (Luminex) with anti-TAGs specific to each of the probe pairs in the assay were diluted/concentrated to 375 to 2500 microspheres of each region per reaction in Tm hybridisation buffer. In a total volume of 25 μl, 2.5 μl, 5 μl, or all of the PCR product was combined with the microsphere mix to a final concentration of 1× Tm hybridisation buffer (0.1 M Tris-HCl pH 8.0 (Sigma), 0.2 M NaCl (Sigma), and 0.08% Triton X-100 (Sigma) in nuclease free distilled water (Life Technologies)). All of the PCR product is defined as the complete content of the tube after the PCR, which may be less than the initial 10 μl volume attributable to possible evaporation. In a thermal cycler (Swift MaxPro, Esco; or iCycler, Bio-Rad), the samples were denatured for 90 s at 96℃ and hybridisation to anti-TAGs on the microspheres occurred for 30 min at 37℃. In case a 5'-biotin-T3 primer was used in the PCR, 100 μl of a reporter mix including 4 or 10 μg/ml of streptavidin-R-phycoerythrin (SAPE) (Life Technologies) in 1× Tm hybridisation buffer was added to each sample and after incubation for 15 min at 37℃ in a thermal cycler (Swift MaxPro, Esco; or iCycler, Bio-Rad), 100 μl of the sample was analysed on a MAGPIX device (Luminex). In case a 5'-Alexa 532-T3 primer

was used in the PCR, 100 μl of 1× Tm hybridisation buffer was added to each sample before analysis of 100 μl on a MAGPIX device.
The analysis was performed at 37℃. The protocol included a sample wash in the MAGPIX device and the minimum bead count was 50 microspheres of each region.

3.2.6 Statistical analysis

The output of the MAGPIX device includes the median fluorescence intensity (MFI) value for each marker, that is, for each DNA sample or control 20 MFI values are obtained, since the MOL-PCR is a 20-plex assay. These MFI values were read into R software (R Core Team 2014). Signal-to-noise ratios (SN) were calculated by dividing the MFI of the sample by the corresponding MFI of the NTC (3.1). As such, each sample yielded 20 signal-to-noise ratios.

$$\mathrm{SN}_{\mathrm{sample}_{\mathrm{marker}\ a}} = \frac{\mathrm{MFI}_{\mathrm{sample}_{\mathrm{marker}\ a}}}{\mathrm{MFI}_{\mathrm{NTC}_{\mathrm{marker}\ a}}} \tag{3.1}$$

In each experiment, different test conditions were compared to a reference condition. For each condition, the test panel of 6 *S.* Typhimurium isolates generated $6 \times 20 = 120$ signal-to-noise ratios, of which 56 signal-to-noise ratios were expected to be positive and 64 were expected to be negative based on a prior screening with PCR or sequencing. The 56 positive signal-to-noise ratios of each test condition were compared to the 56 positive signal-to-noise ratios of the reference condition using a paired Wilcoxon signed-rank test, for which the two-sided variant tests that the distribution of the difference of signal-to-noise ratios of reference and test condition comes from a distribution symmetric around 0. This nonparametric alternative for the *t*-test was used since, with Kolmogorov-Smirnov tests on each set of 56 positive signal-to-noise ratios, the null hypothesis of normal distribution had to be rejected at significance level 0.01.
Each experiment was performed twice in an independent manner and the results of each test were statistically analysed separately. For each experiment, the results of only one of the 2 independent tests are presented, since each of the 2 independent tests yielded the same statistical results. The results of the positive control for the reaction (using the positive control DNA template) and of the NTC (nuclease free distilled water as template) are not included in the figures. For example, in the first experiment the test conditions are 3 different methods for DNA isolation and those are compared to a previously used method for DNA isolation, which is the reference condition. As the panel of 6 *S.* Typhimurium isolates together with the positive control for the reaction and a NTC were assayed twice for each of the 4 DNA isolation methods, the first

experiment included 8 (6 isolates, 1 positive control, 1 NTC) × 4 (conditions) × 2 (independent repetitions) = 64 MOL-PCR assays, consisting of 1280 signal-to-noise ratios in total (20 × (8 × 4 × 2)). However, only the results of the test panel (6 isolates) in the 4 conditions and of one of 2 independent tests are shown, so that in Figure 3.2 the results of 6 × 4 = 24 MOL-PCR assays (24 × 20 = 480 signal-to-noise ratios) are presented, representative of the 2 independent repetitions. The significance level was set at 0.01 for all hypothesis tests. All statistical analyses were performed in R software (R Core Team 2014).

3.3 Results and discussion

MOL-PCR with separated ligation and PCR (Stucki *et al.* 2012; Thierry *et al.* 2013) includes 3 main steps (see Figure 3.1 for an overview of the assay). The first step after DNA isolation is a multiplex oligonucleotide ligation in which a pair of probes that anneal adjacent to each other on the target DNA sequence is ligated by a thermostable DNA ligase. For each marker in the multiplex assay, a different probe pair is included in the assay. Ligation of probe pairs results in various single-stranded DNA molecules, which function as a template in the subsequent singleplex PCR with a universal primer pair (*e.g.*, T3 and T7). One primer is biotinylated for read-out on a Luminex device. In the third step, the PCR products are hybridised to MagPlex-TAG microspheres, each with a different colour shade of red, through a TAG that is incorporated in the marker-specific probes and that is complementary to the anti-TAG coupled to the microspheres. For each marker a different TAG and a different MagPlex-TAG microsphere is used in the assay. After incubation with streptavidin-R-phycoerythrin (SAPE), a Luminex device will identify the microsphere based on the measurement of the red colour, and thus the marker, and measure the fluorescence signal of the SAPE to detect whether a PCR product has hybridised to the anti-TAG on the microsphere.

The optimisation pursues high signal-to-background ratios for positive results, which can be achieved with low background fluorescence intensity and high fluorescence intensity for positive results. As this MOL-PCR assay is intended for use as a routine subtyping method, cost is also an important factor in the optimisation.

Optimisation of a MOL-PCR assay can be attempted at different parameters of the preparatory work and in each of the 3 steps of the assay (Figure 3.1). Here, we used a MOL-PCR assay in development as a routine subtyping method for *S.* Typhimurium to illustrate which parameters are important to be evaluated and, if needed, to be adapted, to achieve high signal-to-noise ratios. Different conditions were tested to check whether deviations from the standard values (based on literature and Luminex guidelines) had a significant impact on the

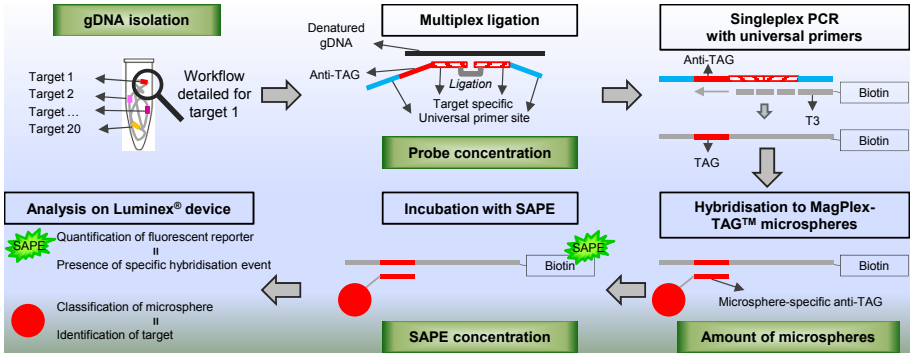

Figure 3.1: Overview of a multiplex oligonucleotide ligation-PCR (MOL-PCR) assay. The workflow is illustrated for 1 target, with the described MOL-PCR being a 20-plex assay. However, a MAGPIX device can discriminate up to 50 different microsphere sets and hence up to 50 different targets. Optimisation of the MOL-PCR assay was attempted for several parameters and the green boxes indicate those parameters which had a major influence on the signal-to-noise ratios, as has been elaborated in the text. gDNA: genomic DNA; SAPE: streptavidin-R-phycoerythrin; see text for more details.

signal-to-noise ratios and would thus be advisable for further evaluation during optimisation of a MOL-PCR assay. If no impact was seen, no further conditions were tested. The optimal conditions for the parameter that was tested in one experiment were further used in the following experiments. Table 3.1 summarises all tested parameters, conditions, and the results of the statistical analysis, which are elaborated below. Also issues that emerged during this evaluation of the assay are discussed below.

3.3.1 Probe premixes

Since the described MOL-PCR assay is intended for routine use, the preparation of a complex ligation mix with addition of each of the 40 probes separately each time the assay is performed, should be avoided. Therefore, it was attempted first to combine different probe pairs in a defined number of different probe premixes. To avoid repetitive freeze-thaw cycles of diluted oligonucleotides, these probe premixes were stored at −20°C in single-use aliquots at a concentration of 5 μM. However, some combinations of probes in one premix resulted in high background signals as measured by the MFI in the NTC (data not shown). These high background signals were eliminated by changing probe pairs from

one probe premix to another or by taking the probe pair out of the probe premixes. Finally, the 40 probes were divided over 3 probe premixes including 8, 5, and 3 probe pairs each and the last 8 probes were separately added to the ligation mix. The latter 8 probes were also stored in a concentration of 5 μM at -20°C in single-use aliquots. It was observed that a combination of these 8 probes in a probe premix that was not frozen before use did not result in high background signals. This clearly illustrates that the storage and combination of probes may have an impact on the background signals and therefore on the signal-to-noise ratios.

3.3.2 DNA isolation

A simple, time- and cost-efficient method for DNA isolation is preferable for routine purposes. Heat lysis of bacterial cultures meets these requirements and was previously described in literature (Bergval *et al.* 2012; Stucki *et al.* 2012; Thierry *et al.* 2013) for preparation of a DNA template for a MOL-PCR assay. Therefore, in the first experiment, different methods for DNA isolation by heat lysis were compared. The reference condition was heat lysis of a single colony in 300 μl sterile deionised water, which was previously used for PCR screening of the markers. Test conditions with more concentrated DNA template were heat lysis of a single colony in 100 μl or 50 μl sterile deionised water or in 20 μl of InstaGene Matrix, which is a commercial kit. By using DNA of a single colony, potential inaccurate characterisation results assignable to a possible mixture of different strains can be avoided, which is not the case when DNA is isolated from, for example, liquid cultures or multiple colonies. Other parameters for this MOL-PCR assay are shown in Table 3.1 as experiment 1.
A boxplot of the signal-to-noise ratios for the different methods of DNA isolation is given in Figure 3.2. The two-sided hypothesis test did not demonstrate a significant difference between the positive signal-to-noise ratios of a single colony in 300 μl and 100 μl sterile deionised water. P values of one-sided hypothesis tests between reference and a single colony in 50 μl sterile deionised water and between reference and a single colony in 20 μl InstaGene Matrix indicated significant improvements by the test conditions. This significant difference can be explained by the higher MFI values that were obtained for the test panel with DNA templates of heat lysis in 50 μl water and in InstaGene Matrix in comparison to the reference condition (Dataset S.1 in the supporting information). More concentrated DNA is thus beneficial for the MOL-PCR assay. The boxplot shown in Figure 3.2 indicates that heat lysis in InstaGene Matrix results in higher signal-to-noise ratios than heat lysis in 50 μl deionised water. However, heat lysis of a single colony in 50 μl sterile deionised water is preferred over heat lysis of a single colony in InstaGene Matrix, since heat lysis

in water is cheaper than heat lysis in InstaGene Matrix, which makes heat lysis in water more convenient for DNA isolation in a routine method for bacterial pathogen characterisation.

3.3.3 Multiplex oligonucleotide ligation

Next, the impact of the ligation mix with the concentration of the probes, the added volume of DNA template and the quantity of *Taq* DNA ligase enzyme, and of the time of initial denaturation of the DNA during the multiplex oligonucleotide ligation was evaluated.
In the literature, the concentration of the probes ranges from 0.25 nM (Bergval *et al.* 2012) up to 10 μM (Thierry *et al.* 2013). To evaluate the effect of the probe concentration on the signal-to-noise ratio, the concentration of each of the 40 probes in the ligation mix was tested at 1 nM and 5 nM (Deshpande *et al.* 2010) with 2 nM (Stucki *et al.* 2012) as reference condition. Other parameters for this MOL-PCR assay are shown in Table 3.1 as experiment 2.
A boxplot of the signal-to-noise ratios for the different probe concentrations is given in Figure 3.3. For both 1 nM and 5 nM probe concentrations, the one-sided hypothesis tests demonstrated significant differences in favour of the reference condition. Compared to the reference condition of 2 nM probe concentration, higher MFI values were seen for the NTC at 1 nM probe concentration (Dataset S.2) and lower MFI values for expected positive results in the test panel at 5 nM probe concentration (Dataset S.3). The total amount of probes in a MOL-PCR reaction seems to be a crucial factor. Therefore, the optimal individual probe concentration might be dependent on the level of multiplex of the MOL-PCR assay, as remarked in literature (Bergval *et al.* 2012; Stucki *et al.* 2012) and based on our own experience. Indeed, the MLPA assay of Bergval *et al.* (2012) included 47 informative probe pairs (high multiplex level) and hybridisation occurred at a concentration of 0.25 nM of each probe, whereas a concentration of 5 nM of each probe gave high signal-to-noise ratios for positive results in our initial MOL-PCR tests with 6 probe pairs (low multiplex level) (data not shown). At an intermediate level of multiplex, the MOL-PCR assay of Stucki *et al.* (2012) which interrogated 8 SNPs with 3 probes per SNP used a protocol with an intermediate concentration of 2 nM of each probe.
As the cost of a routine assay should be kept as low as possible, the number of *Taq* DNA ligase units is an important aspect for optimisation, since enzymes are a key factor in the cost of a reaction. In experiment 3, the combination of the number of ligase enzyme units and the volume of DNA template in the ligation mix was evaluated, since a possible higher yield of ligated probe pairs due to a higher concentration of enzyme might be restricted by the availability of DNA template. As reference condition the combination of 2 μl DNA template with 2

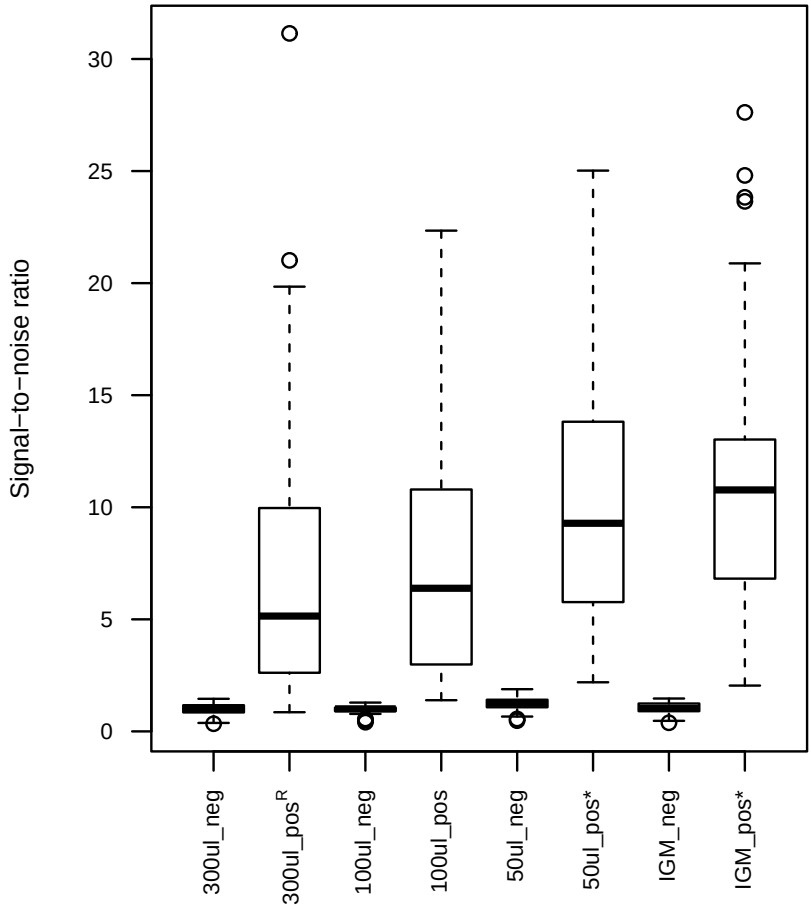

Figure 3.2: Boxplot of signal-to-noise ratios for different methods of DNA isolation. Results of one of 2 independent tests are presented, that is, results of 6 isolates $\times$ 4 conditions = 24 MOL-PCR assays (for each of the 4 conditions: $n_{total} = 120$ with $n_{neg} = 64$ and $n_{pos} = 56$). 300ul: single colony in 300 μl sterile deionised water; 100ul: single colony in 100 μl sterile deionised water; 50ul: single colony in 50 μl sterile deionised water; IGM: single colony in 20 μl InstaGene Matrix; neg: expected negative results; pos: expected positive results; R: reference condition; *: statistically significant difference with the reference condition.

units of ligase enzyme, the least expensive, was taken, whereas test conditions were all 5 other combinations of 2 or 4 μl DNA template (Stucki *et al.* 2012; Thierry *et al.* 2013) with 2, 4 (Stucki *et al.* 2012), or 6 units of ligase enzyme (see Table 3.1 for other parameters of this MOL-PCR assay).
A boxplot of the signal-to-noise ratios for the different conditions is presented in supplementary Figure S.1 in the supporting information. One-sided hypothesis tests indicated significant differences in support of the reference condition for all test conditions. With increasing units of ligase enzyme in combination with 2 μl of DNA template, increasing MFI values for the NTC were seen (Dataset S.4), which caused declining signal-to-noise ratios. At 2 units of ligase enzyme, MFI values for the test panel were generally lower for 4 μl of DNA template than for 2 μl of DNA template (Dataset S.5), which might be explained by inhibition of the ligation reaction because of impurities in the DNA extract. Therefore, pure DNA extracted from liquid cultures with a commercial kit might ameliorate the ligation reaction when more template DNA is used. However, a trade-off should be made between the advantage of using pure DNA and the associated disadvantages of cost, time, and possible strain variations occurred during culturing.
In a fourth experiment, the initial denaturation of the DNA template was considered. An initial denaturation of 10 min at 95℃ was previously used in the PCR for screening of the markers. However, as a longer time at an elevated temperature may have a negative effect on the activity of the ligase enzyme, a shorter initial denaturation time was tested. An initial denaturation of 5 min and 10 min at 95℃ served as reference and test condition, respectively (see Table 3.1 for other parameters of this MOL-PCR assay).
A boxplot of the signal-to-noise ratios for the different initial denaturation times is shown in supplementary Figure S.2. A significant difference in favour of the longer initial denaturation time is demonstrated by a one-sided hypothesis test. This can be explained by slightly higher MFI values for the NTC at 5 min initial denaturation, while the MFI values for the test panel stays at the same level with both denaturation times (Dataset S.6). As such, a potential higher activity of the ligase enzyme due to a shorter time of denaturation at an elevated temperature does not produce high enough MFI values for the test panel to counteract the increased MFI values of the NTC (Dataset S.6).

3.3.4 Singleplex polymerase chain reaction

Subsequently, the singleplex PCR was assessed at the volume of ligation product added, the number of DNA polymerase units used, and the number of PCR cycles performed.
In experiment 5, addition of 3 μl (Stucki *et al.* 2012) or 5 μl (half of the total

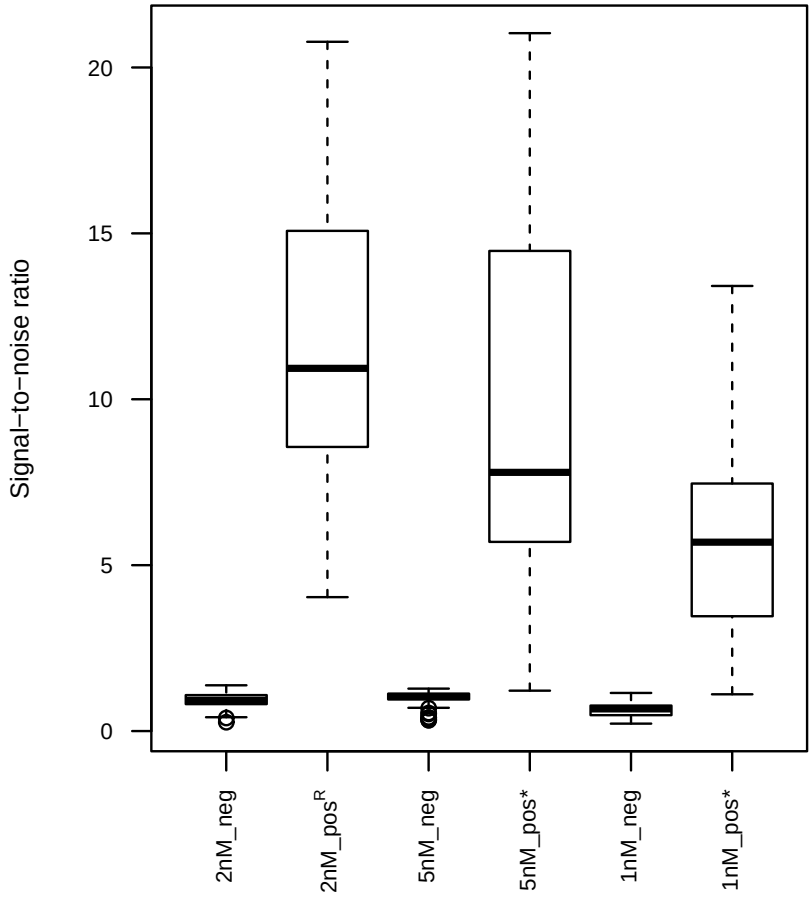

Figure 3.3: Boxplot of signal-to-noise ratios for different probe concentrations in the ligation mix. Results of one of 2 independent tests are presented, that is, results of 6 isolates $\times$ 3 conditions = 18 MOL-PCR assays (for each of the 3 conditions: $n_{total} = 120$ with $n_{neg} = 64$ and $n_{pos} = 56$). neg: expected negative results; pos: expected positive results; R: reference condition; *: statistically significant difference with the reference condition.

volume) of ligation product to the PCR mix was compared (see Table 3.1 for other parameters of this MOL-PCR assay) and supplementary Figure S.3 shows the resulting signal-to-noise ratios. A two-sided hypothesis test did not provide sufficient evidence for a significant difference between both volumes of ligation product. Yet, we selected 3 μl of ligation product as optimal condition, because the ligation reaction volume is 10 μl and after taking 3 μl of the ligation product there would still be enough ligation product left to repeat the PCR if necessary.
As already indicated in the multiplex oligonucleotide ligation reaction, enzymes are an important cost in an assay. However, a balance should be sought between cost and efficacy of the reaction. Therefore, 0.25 units and 0.5 units of HotStarTaq DNA polymerase were evaluated in experiment 6 (see Table 3.1 for other parameters of this MOL-PCR assay), for which the resulting signal-to-noise ratios are given in supplementary Figure S.4. A one-sided hypothesis test pointed out a significant difference in support of 0.25 units of DNA polymerase. Indeed, the MFI values for the NTC with 0.5 units of DNA polymerase were elevated such that, although higher MFI values for the test panel were obtained, the signal-to-noise ratios were considerably lower than those of the PCR with 0.25 units DNA polymerase (Dataset S.7).
These results are in line with the recommendations of Qiagen (Qiagen 2010) to use 2.5 units of HotStarTaq DNA polymerase in a PCR reaction with a total volume of 100 μl, which corresponds to 0.25 units of DNA polymerase in a 10 μl PCR reaction.
A higher number of cycles in the PCR reaction may produce higher MFI values for positive results but also increases the turnaround time of the assay. The number of PCR cycles ranges in the literature from 35 (Bergval *et al.* 2012) up to 45 (Deshpande *et al.* 2010; Thierry *et al.* 2013) and Luminex (Angeloni *et al.* 2014) suggests 35 cycles for PCR. To evaluate if a higher number of PCR cycles had a positive influence on the signal-to-noise ratios, 40 cycles were tested. Supplementary Figure S.5 illustrates the signal-to-noise ratios for 35 and 40 PCR cycles (see Table 3.1 for other parameters of this MOL-PCR assay). A significant difference in favour of 35 PCR cycles was demonstrated by a one-sided hypothesis test. Compared to the PCR with 35 cycles, the PCR with 40 cycles resulted in higher MFI values for the test panel, but they did not lead to increased signal-to-noise ratios caused by raised MFI values for the NTC (Dataset S.8).

3.3.5 Thermal cycler used for the MOL-PCR assay

During the evaluation experiments, it was observed that the thermal cycler had an influence on the MFI values for the NTC (data not shown). When the multiplex oligonucleotide ligation or the PCR or both reactions were performed

on an iCycler (Bio-Rad), higher MFI values for the NTC were obtained than for MOL-PCR assays completely run on a Swift MaxPro (Esco) thermal cycler. An apparent difference between both thermal cyclers is the maximum heating and cooling rate, which is 3.3°C/s and 2.0°C/s for, respectively, heating and cooling in the iCycler, and 4°C/s for both heating and cooling in the Swift MaxPro thermal cycler. A lower ramp rate might thus enhance cross-reactivity of the probes, which is measured through the NTC. This parameter should be checked before starting the MOL-PCR assay development.

3.3.6 Hybridisation to microspheres and analysis on Luminex platform

For optimisation of the hybridisation of the PCR product to the MagPlex-TAG microspheres and read-out on a MAGPIX device, the volume of PCR product added, the amount of microspheres, concentration of SAPE in the reporter mix, and the type of reporter dye were taken into account.

The manufacturer's no-wash protocol for hybridisation to MagPlex-TAG microspheres (Angeloni *et al.* 2014) recommends addition of 2.5 μl of PCR product to the microsphere mix for the hybridisation reaction, while in the literature 5 μl (Thierry *et al.* 2013) to 10 μl (Deshpande *et al.* 2010; Bergval *et al.* 2012; Stucki *et al.* 2012) is reported. A comparison was made with 2.5 μl of PCR product as reference condition and 5 μl of PCR product and all (theoretically 10 μl) of the PCR product as test conditions (see Table 3.1 for other parameters of this MOL-PCR assay). The signal-to-noise ratios of this experiment are shown in supplementary Figure S.6. Both test conditions presented a significant difference against the reference condition as demonstrated by one-sided hypothesis tests.

In view of an assay for routine use, addition of 5 μl of PCR product to the hybridisation reaction is preferred over addition of all PCR product, since standardisation of the assay is more straightforward with well-defined volumes.

The main costs in a MOL-PCR assay are the microspheres and SAPE as a fluorescent reporter. In the literature, 312.5 (Stucki *et al.* 2012) up to 5000 (Deshpande *et al.* 2010) microspheres of each region are used per reaction. In experiment 9, an amount of 2500 microspheres of each region per reaction in combination with a concentration of 10 μg/ml of SAPE in the reporter mix was taken as reference condition, as suggested in the Luminex no-wash protocol (Angeloni *et al.* 2014). The test conditions were all 5 other combinations of 2500, 375, or 750 microspheres of each region per reaction in combination with a concentration of 4 or 10 μg/ml SAPE in the reporter mix (see Table 3.1 for other parameters of this MOL-PCR assay).

Figure 3.4 illustrates the signal-to-noise ratios of this experiment. One-

sided hypothesis tests indicated significant improvements of all test conditions compared to the reference condition. Indeed, MFI values for the test panel were lower for an amount of 2500 microspheres in combination with both 4 and 10 μg/ml SAPE in the reporter mix than for all other combinations (Dataset S.9). In addition, a SAPE concentration of 10 μg/ml in the reporter mix resulted in higher MFI values for the NTC at all 3 amounts of microspheres (Dataset S.10). These findings lead to the conclusion that the reference condition generates the lowest signal-to-noise ratios for expected positive results.
At a SAPE concentration of 4 μg/ml in the reporter mix, an amount of 750 microspheres of each region per reaction is preferred over 375 microspheres of each region per reaction, because microspheres counts below 50, which is the minimum bead count recommended by Luminex, were observed during analysis of reactions with 375 microspheres of each region.
A last experiment compared the use of SAPE and Alexa 532 as fluorescent reporter for the read-out on a MAGPIX device (see Table 3.1 for other parameters of this MOL-PCR assay). SAPE is recommended by Luminex as most suitable fluorescent reporter (100% relative fluorescence intensity) and Alexa 532 (28%) as second and Cy3 (19%) as third most suitable fluorescent reporter (Luminex Corporation 2014). The use of a T3 primer coupled with Alexa 532 (Deshpande *et al.* 2010; Stucki *et al.* 2012) or Cy3 (Bergval *et al.* 2012) has the advantage that an incubation with reporter mix is not needed, in contrast to a T3 primer coupled with biotin which binds to SAPE during the incubation with reporter mix.
The resulting signal-to-noise ratios of experiment 10 are given in supplementary Figure S.7. A one-sided hypothesis test demonstrated a significant difference in support of SAPE as fluorescent reporter. MFI values for both test panel and NTC are considerably lower with Alexa 532 as fluorescent reporter compared to SAPE, with decreased signal-to-noise ratios as a consequence with Alexa 532. As Alexa 532 did not perform better than SAPE and given that the relative fluorescence intensity of Cy3 is even lower than that of Alexa 532, Cy3 was not further tested.

3.4 Conclusion

By systematically optimising different parameters at each step of the assay, we have improved a MOL-PCR assay from an assay with no clear difference between positive and negative results (Figure 3.2) to an assay of which the signal-to-noise ratios present a clear difference between positive and negative results (Figure 3.4). A summary of the effect of each of the optimised parameters on the signal-to-noise ratios and on the cost of the assay is presented in Table 3.2. Evidently, the parameters which have a major effect on signal-to-noise ratios and/or on

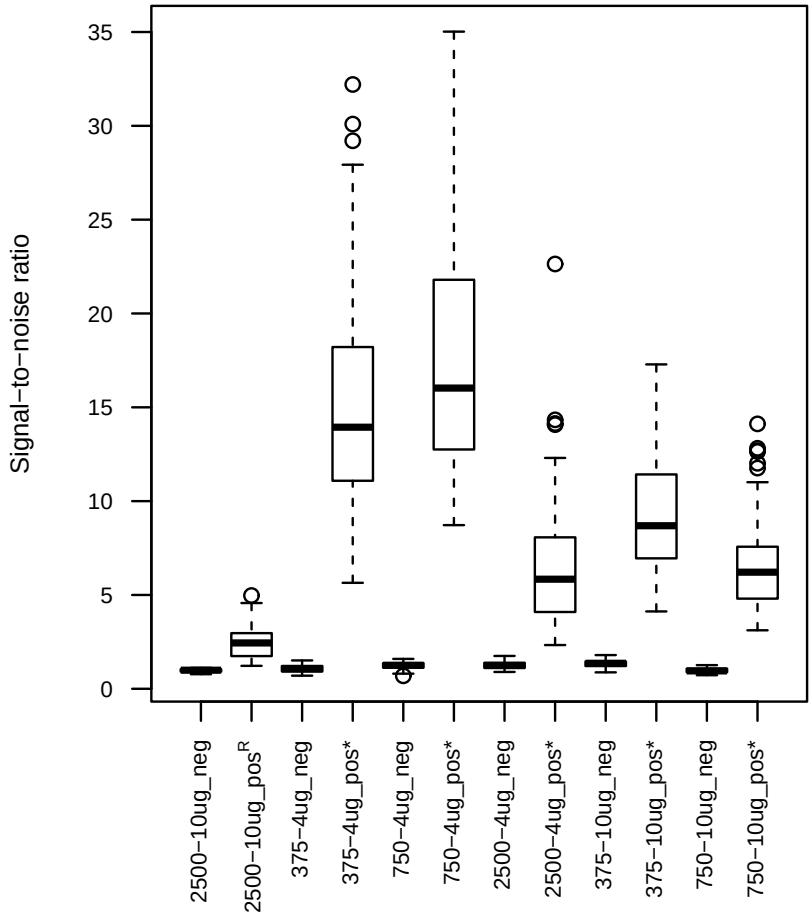

Figure 3.4: Boxplot of signal-to-noise ratios for different combinations of amount of microspheres in the hybridisation reaction and streptavidin-R-phycoerythrin (SAPE) concentration in the reporter mix. Results of one of 2 independent tests are presented, that is, results of 6 isolates $\times$ 6 conditions = 36 MOL-PCR assays (for each of the 6 conditions: $n_{total} = 120$ with $n_{neg} = 64$ and $n_{pos} = 56$). 2500: 2500 microspheres of each region per reaction were combined in the microsphere mix; 375: 375 microspheres of each region per reaction were combined in the microsphere mix; 750: 750 microspheres of each region per reaction were combined in the microsphere mix; 4 ug: the reporter mix contained 4 μg/ml of SAPE; 10 ug: the reporter mix contained 10 μg/ml of SAPE; neg: expected negative results; pos: expected positive results; R: reference condition; *: statistically significant difference with the reference condition.

cost should be prioritised in an optimisation procedure during the development of a similar assay. Major improvements were obtained with changing the method of DNA isolation, lowering the probe concentration in the multiplex ligation mix, and reducing the amount of microspheres of each set per reaction and the concentration of SAPE in the reporter mix (Figure 3.1). We also found that the combination of probe pairs in a frozen premix and the type of thermal cycler used for the multiplex ligation and the singleplex PCR have an influence on the background signals measured through the NTC. These observations are taken into account during the currently ongoing development and validation of a full MOL-PCR assay for the subtyping of *S.* Typhimurium (Wuyts *et al.* 2015b). However, evaluation of the parameters for which a significant impact on the signal-to-noise ratios was seen, is worthwhile for characterisation of any microbial pathogen if the cost and effort of a MOL-PCR assay are important.

Supporting information

Supplementary Figures S.1 up to S.7 and supplementary Datasets S.1 up to S.10 are available at http://downloads.hindawi.com/journals/bmri/2015/790170.f1.pdf (file format: pdf).

Acknowledgements

The authors would like to thank W. Coucke for the fruitful discussions on statistics. Sequencing was performed at the Platform Biotechnology and Molecular Biology at the Scientific Institute of Public Health (WIV-ISP). This work was supported by RP/PJ WIV-ISP (SalMolType).

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Table 3.2: Summary of the effect of optimisation on signal-to-noise ratios and on cost of the multiplex oligonucleotide ligation-PCR assay.

MOL-PCR step	Optimised parameter	Effect of optimisation	
		Signal-to-noise ratios	Cost of the assay
Preparatory work	DNA isolation	++	+
Multiplex oligonucleotide ligation	Probe concentration	++	+
	DNA volume	+	–
	Taq DNA ligase	+	++[a]
	Initial denaturation	+	–
PCR	Ligation product volume	–	–
	Taq DNA polymerase	+	++[b]
	Cycles	+	–
Hybridisation and analysis on MAGPIX	PCR product volume	+	–
	Microspheres per reaction	++	++[c]
	SAPE concentration	++	++[d]
	Reporter dye	+	+

++: major effect; +: minor effect; –: no effect; MOL-PCR: multiplex oligonucleotide ligation-PCR. [a]Factor 2-3 for the cost of the *Taq* DNA ligase used; [b]Factor 2 for the *Taq* DNA polymerase used; [c]Factor 3.3 for the amount of microspheres per reaction used; [d]Factor 2.5 for the amount of SAPE per reaction used.

Chapter 4

A multiplex oligonucleotide ligation-PCR as a complementary tool for subtyping of *Salmonella* Typhimurium

This chapter was previously published as:
V. Wuyts, W. Mattheus, N. H. C. Roosens, K. Marchal, S. Bertrand, and S. C. J. De Keersmaecker (2015b). "A multiplex oligonucleotide ligation-PCR as a complementary tool for subtyping of *Salmonella* Typhimurium". *Applied Microbiology and Biotechnology* 99.19, pp. 8137–8149. DOI: 10.1007/s00253-015-6831-7

Authors' contributions

The experimental work was performed by V. Wuyts, apart from phage typing, MLVA and PFGE for which the data were contributed by W. Mattheus and S. Bertrand. The experiments were designed by V. Wuyts, W. Mattheus, N. H. C. Roosens, K. Marchal, S. Bertrand and S. C. J. De Keersmaecker. Design of primers and probes was done by V. Wuyts. Data analysis and writing of the paper was done by V. Wuyts with the critical input of S. C. J. De Keersmaecker. W. Mattheus, N. H. C. Roosens, K. Marchal and S. Bertrand reviewed the paper and also contributed reagents, materials and analysis tools. S. Bertrand and S. C. J. De Keersmaecker contributed equally to this work.

Abstract

Subtyping below the serovar level is essential for surveillance and outbreak detection and investigation of *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) and its monophasic variant 1,4,[5],12:i:- (*S.* 1,4,[5],12:i:-), frequent causes of foodborne infections. In an attempt to overcome the intrinsic shortcomings of currently used subtyping techniques, a multiplex oligonucleotide ligation-PCR (MOL-PCR) assay was developed which combines different types of molecular markers in a high-throughput microsphere suspension array. The 52 molecular markers include prophage genes, amplified fragment length polymorphism (AFLP) elements, *Salmonella* genomic island 1 (SGI1), allantoinase gene *allB*, MLVA locus STTR10, antibiotic resistance genes, single nucleotide polymorphisms (SNPs) and phase 2 flagellar gene *fljB*. The *in vitro* stability of these markers was confirmed in a serial passage experiment. The validation of the MOL-PCR assay for subtyping of *S.* Typhimurium and *S.* 1,4,[5],12:i:- on 519 isolates shows that the method is rapid, reproducible, flexible, accessible, easy to use and relatively inexpensive. Additionally, a 100% typeability and a discriminatory power equivalent to that of phage typing were observed, and epidemiological concordance was assessed on isolates of 2 different outbreaks. Furthermore, a data analysis method is provided so that the MOL-PCR assay allows for objective, computerised data analysis and data interpretation of which the results can be easily exchanged between different laboratories in an international surveillance network.

4.1 Introduction

Salmonella, one of the major causes of foodborne infections worldwide, is reported to be responsible for about 85,000 human illnesses each year in Europe, with an approximate hospitalisation rate of 36%, and every year, nontyphoidal salmonellosis is accountable for 59 deaths (European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2015).

Salmonella is a complex genus with 2 species, 6 subspecies and 2659 serovars (Grimont and Weill 2007; Issenhuth-Jeanjean *et al.* 2014). In Europe, near 29% of the reported human salmonellosis cases are attributed to *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) and its monophasic variant *S. enterica* subsp. *enterica* serovar 1,4,[5],12:i:- (*S.* 1,4,[5],12:i:-), making them the second and third most commonly reported serovars after *S. enterica* subsp. *enterica* serovar Enteritidis (about 40% of the reported cases) (European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2015). Once serotyped as *S.* Typhimurium or *S.* 1,4,[5],12:i:-, different techniques are applied for further subtyping of the isolate below the serovar level, which is necessary for surveillance, outbreak detection or outbreak investigation. The classical phage typing technique is nowadays complemented with molecular methods for which pulsed-field gel electrophoresis (PFGE) is considered the gold standard. Other molecular methods used for subtyping of *S.* Typhimurium are multiple-locus variable-number of tandem repeats analysis (MLVA), multilocus sequence typing (MLST) and clustered regularly interspaced short palindromic repeats (CRISPR) genotyping. Advantages and disadvantages of each of these techniques have been discussed previously (Boxrud 2010; Wattiau *et al.* 2011; Sabat *et al.* 2013) and, although proven to have additional value for subtyping, each of these techniques has one or more attributes that do not correspond to the ideal subtyping method, which should be inexpensive, rapid, straightforward to execute, highly discriminative, robust, universally applicable for a wide range of bacterial pathogens and generating objective data, which can be easily interpreted and transferred between different laboratories (Wattiau *et al.* 2011; Sabat *et al.* 2013).

In recent years, whole genome sequencing (WGS) has been introduced and promoted as the ultimate subtyping method for each pathogen (Sabat *et al.* 2013), and several *Salmonella* epidemiological investigations reporting its added value have been published (some recent examples: Angelo *et al.* 2015; Ashton *et al.* 2015; Deng *et al.* 2015; Octavia *et al.* 2015). However, although sequencing may eventually have similar costs as other subtyping methods, the turn-around-time from sample to completely analysed data for this technology is not to be neglected (> 24 hours). Moreover, there are still many laboratories which do not have the resources for the substantial investments, both at the level of equipment

as of data analysis, nor have the required high-throughput that is needed to obtain these low sequencing costs (European Food Safety Authority (EFSA) 2014). Therefore, new molecular assays which complement existing subtyping methods and which do not demand high-end equipment nor complex data analysis still have a role to play in these laboratories before WGS will become the gold standard in all European National Reference Laboratories and Centres. Molecular subtyping of *S.* Typhimurium below the serovar level requires multiple markers and for the time-effectiveness of the assay, these markers should be combined in a multiplex assay. A multiplex prophage marker subtyping method was developed by Fang *et al.* (2012). In this assay, 30 prophage-related markers were amplified in 2 separate 15-plex PCRs and the amplicons were analysed with the Luminex xMAP technology in a direct hybridisation assay. In this type of assay, the fluorescent amplicons from a multiplex PCR are hybridised to marker-specific probes, which are covalently coupled to carboxylated microspheres. As different probes are linked to differently coloured microspheres, absence or presence of a marker can be detected in a Luminex device by determining the microsphere colour, and thus the marker, and checking the presence of a hybridised amplicon through its fluorescence. Further refinement of the assay of Fang *et al.* (2012) would necessitate inclusion of more markers, which could be challenging due to the limited multiplexing capacity of a PCR. Multiplex ligation-dependent probe amplification (MLPA) (Schouten *et al.* 2002) allows a higher multiplexing capacity than multiplex PCR, since the multiplex phase is a ligation which is then followed by a singleplex PCR. However, MLPA requires an overnight hybridisation step, which makes it a relatively extended protocol. For the rapid detection of biothreat agents, Deshpande *et al.* (2010) shortened the MLPA protocol by performing the hybridisation and ligation in a thermal cycling step and introduced the Luminex xTAG technology, based on microspheres with anti-TAG sequences pre-coupled to their surface, for the analysis of the PCR products, where the original MLPA protocol relies on fragment sizing by electrophoresis. The resulting multiplex oligonucleotide ligation-PCR (MOL-PCR) allows the detection of a combination of different types of molecular markers, such as single nucleotide polymorphisms (SNPs), unique sequences, insertions and deletions.

MOL-PCR and MLPA have already been described for characterisation and subtyping of pathogens (Beyene *et al.* 2009; Bergval *et al.* 2012; Stucki *et al.* 2012; Pham Thanh *et al.* 2013; Thierry *et al.* 2013; Cornelius *et al.* 2014), but also for the detection of bacteria (Deshpande *et al.* 2010; Chung *et al.* 2012; Berning *et al.* 2014) and viruses (Reijans *et al.* 2008; Theelen *et al.* 2010; De Smet *et al.* 2012) and for diagnosis of human genetic diseases (Schouten *et al.* 2002; Slater *et al.* 2004; Xu *et al.* 2013; Garin *et al.* 2014; Kasatkar *et al.* 2014; Marcinkowska-Swojak *et al.* 2014). Here, we describe a MOL-PCR for subtyping of *S.* Typhimurium that attempts to overcome the major

disadvantages of the currently used subtyping methods, including those of previously described Luminex assays for *Salmonella*. The assay combines markers including prophage genes, amplified fragment length polymorphism (AFLP) elements, *Salmonella* genomic island 1 (SGI1), allantoinase gene *allB*, MLVA locus STTR10, antibiotic resistance, SNPs and phase 2 flagellar gene *fljB*. We elaborate on the development of the assay, report the validation of the subtyping method on a large collection of *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates and provide an analysis method for use in routine subtyping.

4.2 Materials and methods

4.2.1 Bacterial isolates

All *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates were received from the Belgian National Reference Centre for *Salmonella* and *Shigella* and are listed in Dataset S1 in the supporting information. All isolates are available upon request. The validation panel of 519 human *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates (S0001-S0519 in Dataset S1) collected in Belgium in the period 2010-2013 contained 33 different phage types, including 29 not-typable (NT) and 39 reacts-but-does–not-conform (RDNC) isolates, and covered 168 distinct MLVA profiles. Additionally, 13 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates related to 2 different Belgian outbreaks were used in this study. Out-group isolates were isolated around the same time as the corresponding outbreak. Phage typing (Threlfall and Frost 1990) and MLVA (Lindstedt *et al.* 2004; Larsson *et al.* 2009) were performed by the Belgian National Reference Centre for *Salmonella* and *Shigella*.

4.2.2 Pulsed-field gel electrophoresis

PFGE (Hunter *et al.* 2005; Ribot *et al.* 2006) was performed according to the PulseNet Europe protocol (PulseNet International 2015a). Genomic DNA was digested with *Xba*I restriction enzyme and *Xba*I-digested genomic DNA of *S. enterica* subsp. *enterica* serovar Braenderup was used as a size marker. For the PFGE analysis, 53 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates representing one of the 3 most frequently observed MOL-PCR profiles were selected from the validation panel. This selection was made in order to include a high variability of phage types in combination with MLVA profiles. The 53 isolates were run on 4 separate gels. PFGE patterns were analysed with Bionumerics (version 7.1; Applied Maths). A dendrogram was created with following similarity-based

clustering parameters: unweighted pair group method using arithmetic averages (UPGMA) with Dice similarity coefficient and 1.0% optimisation and tolerance settings.

4.2.3 DNA isolation

DNA template was prepared by heat lysis. Hereto, a single colony from an overnight (14 to 20 h) culture at 37°C on LB agar (Merck Millipore) was dissolved in 50 μl sterile deionised water and incubated at 100°C in a heating block for 10 min. After cooling for a minimum of 5 min at 4°C and centrifugation for 10 min at 11,000 $\times$ g, the supernatant was stored at -20°C and used for further analysis.

4.2.4 Selection of molecular markers

The first step in the selection of molecular markers consisted of a literature study (Boyd *et al.* 2001; Hu *et al.* 2002; Lindstedt *et al.* 2004; Mikasová *et al.* 2005; Ross and Heuzenroeder 2005; Hu *et al.* 2006; Drahovská *et al.* 2007; Lan *et al.* 2007; Rychlík *et al.* 2008; Fang *et al.* 2012; Pang *et al.* 2012) to identify molecular markers which could potentially discriminate between *S.* Typhimurium isolates and to find molecular markers which could serve as internal positive control of the *Salmonella* DNA template. In the second step, the molecular markers that could be informative through presence or absence, *i.e.*, SAL-1 up to SAL-55 in Table S1 in the supporting information, were screened by PCR and gel electrophoresis. For this PCR screening, 27 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates of common phage types in Belgium (DT12, DT104, DT120, DT193, DT195 and U302) (Bertrand *et al.* 2014) were selected and complemented with 1 NT isolate, 1 RDNC isolate and 1 isolate of an uncommon phage type in Belgium (DT35) (isolates S0001-S0030 in Dataset S1). DNA was isolated by heat lysis as described above, except that 300 μl sterile deionised water was used instead of 50 μl. The PCR was performed in a final reaction volume of 25 μl including 1$\times$ DreamTaq buffer (Thermo Fisher Scientific), 200 to 500 nM of forward and reverse primer (Table S1), 200 μM of each dNTP (Thermo Fisher Scientific), 0.625 U DreamTaq DNA polymerase (Thermo Fisher Scientific) and 2 μl DNA template. The following protocol was run in a thermal cycler: 10 min at 95°C, 35 cycles of 30 s at 94°C, 30 s at 45 to 60°C (Table S1) and 1 min at 72°C, 10 min at 72°C. PCR products were visualised by agarose gel electrophoresis with ethidium bromide staining.

4.2.5 Probe design

Upstream and downstream probes were designed with Visual OMP (version 7.6.58.0; DNA Software) as previously described (Wuyts *et al.* 2015c). For markers for which a primer pair was reported in literature, it was attempted to take the forward or reverse primer as the target-specific sequence of the upstream probe. For SNP markers, except for an internal positive control marker, a probe with the wild-type allele was also included in the assay.

4.2.6 MOL-PCR assay protocol

The MOL-PCR assay parameters were optimised as previously described (Wuyts *et al.* 2015c).
The selected markers were divided over three MOL-PCRs, *i.e.*, MOL-PCR_1, MOL-PCR_2 and MOL-PCR_SNP, as listed in Table 4.1.
The multiplex oligonucleotide ligation reaction occurred in a 10 μl volume with 1× *Taq* DNA ligase reaction buffer (New England BioLabs), 2 nM of each probe (Tables S2, S3 and S4, Eurogentec), 2 U of *Taq* DNA ligase (New England BioLabs), 2 μl of DNA template and nuclease-free distilled water (Thermo Fisher Scientific). The thermal cycling programme (Swift MaxPro, Esco) included 10 min of denaturation at 95℃ followed by 30 cycles of 25 s at 94℃ and 30 s at 58℃.
The singleplex PCR was performed in a final volume of 10 μl composed of 1× HotStarTaq PCR buffer (Qiagen), 125 nM T7 primer (TAATACGACTCAC-TATAGGG, Eurogentec), 500 nM 5'-biotin-T3 primer (ATTAACCCTCAC-TAAAGGGA, Eurogentec), 200 μM of each dNTP (Thermo Fisher Scientific), 0.25 U HotStarTaq DNA polymerase (Qiagen) and 3 μl of ligase product. The PCR protocol was 15 min at 95℃, 35 cycles of 30 s at 94℃, 30 s at 60℃ and 30 s at 72℃, 5 min at 72℃ (Swift MaxPro, Esco).
The necessary regions (Tables S2, S3 and S4) of MagPlex-TAG microspheres (Luminex) were diluted to 750 microspheres of each region per reaction in 1.25× Tm hybridisation buffer (0.125 M Tris-HCl pH 8.0 (Sigma), 0.25 M NaCl (Sigma), 0.1% Triton X-100 (Sigma) in nuclease-free distilled water (Thermo Fisher Scientific), sterilised by filtration (0.2 μm)). In a total volume of 25 μl, 5 μl of the PCR product was combined with the microsphere mix to a final concentration of 1× Tm hybridisation buffer (0.1 M Tris-HCl pH 8.0 (Sigma), 0.2 M NaCl (Sigma), 0.08% Triton X-100 (Sigma) in nuclease-free distilled water (Thermo Fisher Scientific), sterilised by filtration (0.2 μm)). In a thermal cycler, the samples were denatured for 90 s at 96℃ and hybridisation to anti-TAGs on the microspheres occurred for 30 min at 37℃. Hundred microlitres of a reporter mix including 4 μg/ml of streptavidin-R-phycoerythrin (SAPE) (Thermo Fisher

Scientific) in 1× Tm hybridisation buffer was added to each sample and after incubation for 15 min at 37°C in a thermal cycler, 100 μl of the sample was analysed on a MAGPIX device (Luminex). The analysis was performed at 37°C. The protocol included a sample wash in the MAGPIX device and the minimum bead count was 50 microspheres of each region.
A negative control and a positive control for the reaction were included in each assay, except for the SNP assay in which only a negative control was included, since the wildtype allele acted as a positive control for the reaction. The negative control was a no-template-control (NTC) for which the DNA template was replaced by nuclease-free distilled water (Thermo Fisher Scientific) in the multiplex oligonucleotide ligation reaction. For the positive control DNA template, a single colony of each of five different isolates (*i.e.*, samples S0001, S0002, S0024, S0025 and S0050 in Dataset S1) was mixed in one tube with 50 μl sterile deionised water, which was treated as described in the section DNA isolation. Other isolates, separate or mixed, can be used as positive control for the reaction, as long as the performance of the reaction is verified for all markers in the MOL-PCR assay.

4.2.7 PCR amplicon sequencing

For confirmation of the MOL-PCR results, PCR amplicons were sequenced on an ABI 3130xl Genetic Analyzer (Applied Biosystems). If primers were not available in literature, they were designed with Visual OMP (version 7.6.58.0; DNA Software). The amplicons were obtained by PCR as described above for the PCR screening with primers listed in Table S1 and were cleaned up before sequencing with ExoSAP-IT (Affymetrix) according to the manufacturer's protocol. Sequence alignments were made with ClustalW in MegAlign (version 10.0.1 (3), 419; DNASTAR).

4.2.8 Specificity of the internal positive control markers

The specificity of the internal positive control markers was tested by performing a MOL-PCR reaction with these probes on bacteria that are unrelated and closely related to *Salmonella* and on *Salmonella* isolates of other serovars than Typhimurium, but which are common in Belgium (Table S5) (Bertrand *et al.* 2014). All isolates used in this part of the development of the MOL-PCR assay were available in our laboratory as purified DNA (Barbau-Piednoir *et al.* 2013).

Table 4.1: Type of markers and division over MOL-PCRs in the subtyping assay.

MOL-PCR	Internal PC	Prophage	AFLP	SGI1	Antibiotic resistance	SNP	Other	Plex
MOL-PCR_1	*invA* *rpoB*	8	2	2	4	-	STTR10 *allB*	20
MOL-PCR_2	*rpoB*	9	10	-	1	-	*fljB*	22
MOL-PCR_SNP	-	-	-	-	-	11	-	11[a]
Total	2	17	12	2	5	11	3	50 + 2

[a] MOL-PCR_SNP includes 11 SNP alleles and 11 corresponding wild-type alleles and as such, 22 different regions of MagPlex-TAG microspheres are included in MOL-PCR_SNP.
AFLP: amplified fragment length polymorphism; MOL-PCR: multiplex oligonucleotide ligation-PCR; PC: positive control; SGI1: *Salmonella* genomic island 1; SNP: single nucleotide polymorphism

4.2.9 Stability study

The *in vitro* stability of the selected molecular markers was evaluated in 31 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates (indicated in Dataset S1 in column 'Stability_experiment' with '1' if included) with a common phage type in Belgium as follows: for each isolate, a single colony from a culture grown overnight on LB agar (Merck Millipore) at 37°C was inoculated into 5 ml LB broth (Thermo Fisher Scientific) and incubated overnight at 37°C without shaking. Next, a series of 50 passages at a rate of two passages per day was performed by inoculating 20 μl of culture into 5 ml fresh LB broth and incubating at 37°C without shaking. Glycerol (25% *v/v*) stocks (−80°C) were made before each 5th passage. DNA was isolated after the 50th passage as described above.

4.2.10 Data analysis

The output of the MAGPIX device includes the median fluorescence intensity (MFI) value for each marker in a comma-separated values file. These MFI values were read into R software (version 3.1.2) (R Core Team 2014). Signal-to-noise ratios (SN) were calculated by dividing the MFI of the sample by the corresponding MFI of the NTC (4.1).

$$\mathrm{SN}_{\mathrm{sample}_{\mathrm{marker}\ a}} = \frac{\mathrm{MFI}_{\mathrm{sample}_{\mathrm{marker}\ a}}}{\mathrm{MFI}_{\mathrm{NTC}_{\mathrm{marker}\ a}}} \tag{4.1}$$

For analysis of SNP markers, a SNP allele call was calculated by dividing the signal-to-noise ratio of the SNP marker by the sum of the signal-to-noise ratio of the SNP marker and the signal-to-noise ratio of its corresponding wild-type (WT) marker (4.2). Analogously, a wild-type allele call was calculated by dividing the signal-to-noise ratio of the wild-type marker by the sum of the signal-to-noise ratio of the wild-type marker and the signal-to-noise ratio of its corresponding SNP marker (4.3).

$$\text{Allele call\_SNP}_{\mathrm{sample}\ x_{\mathrm{SNP}\ a}} = \frac{\mathrm{SN}_{\mathrm{sample}\ x_{\mathrm{SNP}\ a}}}{\mathrm{SN}_{\mathrm{sample}\ x_{\mathrm{SNP}\ a}} + \mathrm{SN}_{\mathrm{sample}\ x_{\mathrm{WT}\ a}}} \tag{4.2}$$

$$\text{Allele call\_WT}_{\mathrm{sample}\ x_{\mathrm{WT}\ a}} = \frac{\mathrm{SN}_{\mathrm{sample}\ x_{\mathrm{WT}\ a}}}{\mathrm{SN}_{\mathrm{sample}\ x_{\mathrm{WT}\ a}} + \mathrm{SN}_{\mathrm{sample}\ x_{\mathrm{SNP}\ a}}} \tag{4.3}$$

4.2.11 Data interpretation

During the development of the MOL-PCR assay, a universal cut-off value of 3 on the signal-to-noise ratio was used to determine the positive samples for

markers that discriminate through presence or absence. For future application of the MOL-PCR as a routine subtyping assay, the cut-off values were refined for each of these markers after validation of the method on 519 *S.* Typhimurium and *S.* 1,4,[5],12:i:-. Hereto, the average of the maximum signal-to-noise ratio of the negative samples and the minimum signal-to-noise ratio of the positive samples was calculated (Tables S6 and S7). These average cut-off values were rounded to the nearest integer if the mean was greater than or equal to 3.75, to 3.5 if the mean was greater than or equal to 3.25 and smaller than 3.75 and to 3 if the mean was smaller than 3.25. For the SNP markers, the cut-off was set to 0.6 on the allele call, *i.e.*, if the SNP allele call is greater than 0.6, then the SNP allele is assigned to the sample or if the wild-type allele call is greater than 0.6, then the wild-type allele is assigned to the sample. For the internal positive control markers, the cut-off value for the signal-to-noise ratio was calculated by rounding down the minimum signal-to-noise ratio of the positive samples to the nearest integer.

For the SNP markers, an additional cut-off was calculated on the MFI values to determine if the SNP locus was present, *i.e.*, if the probes could hybridise to the SNP locus and subsequently be ligated and amplified. Hereto, the mean was determined between the maximum MFI of the negative samples for the allele and the minimum of the positive samples for the allele (Table S8). This mean was rounded to the nearest multiple of 100 if the mean was greater than or equal to 375 and to the nearest multiple of 50 if the mean was smaller than 375.

For the interpretation of multiplex data, each marker that discriminates through presence or absence and each SNP marker is assigned a unique prime number. If the marker is present in the sample, *i.e.*, the signal-to-noise ratio or the SNP allele call is higher than the cut-off value, the sample receives the prime number of that marker. Otherwise, if the marker is absent, the sample receives '1', which is the neutral element in the multiplication, for that marker. As such, the Gödel Prime Product (GPP) (Van den Bulcke *et al.* 2008; Van den Bulcke *et al.* 2010) can be calculated as the product of all assigned prime numbers or '1'. Due to the nature of prime numbers, this GPP is a mathematical barcode (Gödel 1931) for the sample so that a large number of results can be assigned to a unique, arbitrary number, thereby simplifying data analysis on bacterial populations. An advantage of the GPP is that through factorisation of the GPP into its dividers, all discriminative markers present in the sample can be identified. Likewise, if the GPP is divided by the prime number of a specific marker, presence or absence of that marker will be indicated by, respectively, an integer or non-integer outcome of the division.

As the subtyping assay combines three MOL-PCRs, a MOL-PCR profile consists of three GPPs, *i.e.*, $\mathrm{GPP}_{\text{MOL-PCR\_1}} - \mathrm{GPP}_{\text{MOL-PCR\_2}} - \mathrm{GPP}_{\text{MOL-PCR\_SNP}}$. Unique prime numbers were assigned to each discriminative marker within each

separate MOL-PCR (Tables S6, S7 and S8). To keep the GPPs as small as possible, the markers that were present in most samples received the lowest prime number; the markers that were present in the least amount of samples received the greatest prime number. As the GPPs may still result in a large number, an in-house code was assigned by ordering the GPPs, separately for each MOL-PCR, from the smallest to largest and numbering the GPPs starting from 1 (Tables S9, S10 and S11). An example of such an in-house code is 16-12-8.

Since a MOL-PCR profile consists of 3 numbers, the profiles were visualised in a 3-dimensional scatterplot in which the number of isolates in each MOL-PCR profile was indicated with a colour code. This type of visualisation may be informative for outbreak detection and can be realised in R software with the package scatterplot3d (version 0.3-35) (Ligges and Mächler 2003). As GPPs may be high numbers, the in-house code was used for each of the 3 axes in the scatterplot. An example of a MOL-PCR profile is 3.91×10^{21}-1.11×10^{13}-4199, which is in-house coded as MOL-PCR profile 33-28-10. The 3-dimensional scatterplot is then generated as follows: the numeral for $GPP_{MOL\text{-}PCR\_1}$, 3.91×10^{21}, is in our example encoded as 33 and is plotted on the x-axis, the numeral for $GPP_{MOL\text{-}PCR\_2}$, 1.11×10^{13}, is in our example encoded as 28 and is plotted on the y-axis, and the numeral for $GPP_{MOL\text{-}PCR\_SNP}$, 4199, is in our example encoded as 10 and is plotted against the z-axis so that finally, MOL-PCR profile 33-28-10 is represented as a point in the 3-dimensional scatterplot. The colour of a point indicates how many of the 519 isolates in the validation panel have that specific MOL-PCR profile. In our example, the MOL-PCR profile 33-28-10 is represented in Figure 4.1 as the blue-green point in the top right corner, which indicates that about 40 isolates of our validation panel had MOL-PCR profile 3.91×10^{21}-1.11×10^{13}-4199, which was in-house coded as 33-28-10 (to be precise, 38 of the 519 isolates in our validation panel had this MOL-PCR profile).

An R-application for data analysis and interpretation, created with the package Shiny (version 0.11.1) (Chang *et al.* 2015) and which takes the MAGPIX output files as input, is available upon request.

4.2.12 Discriminatory power

Discriminatory power was calculated as the average probability that two unrelated strains randomly sampled in the population are assigned a different type using Simpson's index of diversity (Hunter and Gaston 1988).

4.3 Results

4.3.1 Assay design

MOL-PCR consists of three main steps: firstly, a multiplex oligonucleotide ligation of specific probes for detection of the molecular markers, secondly, a PCR for signal amplification and finally, the hybridisation to MagPlex-TAG microspheres and read-out on a Luminex device.
In the multiplex oligonucleotide ligation reaction, a different probe pair is included for each marker in the assay. If both probes of such a probe pair anneal adjacent to each other on the genomic DNA of the bacterial isolate to be tested, they are ligated by a thermostable DNA ligase so that various single-stranded DNA molecules are created that serve as a template in the subsequent singleplex PCR with a universal primer pair, *i.e.*, T7 and T3. The T3 primer is 5' biotinylated for read-out on a Luminex device. The third step starts with hybridisation of the PCR products to MagPlex-TAG microspheres, through a TAG that is integrated in the marker-specific probes and that is complementary to the anti-TAG covalently coupled to the surface of the microsphere. For each marker in the assay, a different TAG and a different MagPlex-TAG microsphere were used. Microspheres with a different anti-TAG have a different red colour code, which allows them to be identified by measurement of the red colour. After incubation with SAPE, a Luminex device will identify the microsphere through its red colour, and thus the marker, and measures the fluorescence signal of the SAPE to detect whether a PCR product has hybridised to the anti-TAG coupled to the microsphere.
A total of 70 potentially discriminative markers, including all 30 markers from the prophage subtyping assay of Fang *et al.* (2012), were selected from literature. The selection consisted of 32 genes of prophages Fels-1, Fels-2, Gifsy-1, Gifsy-2, P22, SopEφ, SLP281, ST64B, ST64T, ST104 and ST104B (Mikasová *et al.* 2005; Ross and Heuzenroeder 2005; Drahovská *et al.* 2007; Rychlík *et al.* 2008; Fang *et al.* 2012), 16 AFLP fragments (Lan *et al.* 2007; Fang *et al.* 2012), the left and right junction of SGI1 (Boyd *et al.* 2000; Rychlík *et al.* 2008), the allantoinase gene *allB* (Rychlík *et al.* 2008), MLVA locus STTR10 (Lindstedt *et al.* 2004), 12 SNPs (Pang *et al.* 2012), 5 antibiotic resistance genes encoded in SGI1 for resistance to ampicillin, chloramphenicol/florfenicol, streptomycin/spectinomycin, sulfonamides and tetracycline (Ng *et al.* 1999; Boyd *et al.* 2002), and the phase 2 flagellar gene *fljB* for identification of *S.* 1,4,[5],12:i:- (European Food Safety Authority (EFSA) Panel on Biological Hazards 2010).

4.3.2 Selection of molecular markers

PCR screening was performed on the prophage genes, AFLP fragments, the left and right junction of SGI1 and the gene *allB*. Those markers which showed variation in at least 2 of the 30 tested isolates, and thus have discriminatory power, were selected for the MOL-PCR assay development. This criterion resulted in the rejection of 15 prophage genes and 7 AFLP fragments, of which, respectively, 6 and 5 markers were included in the prophage subtyping assay of Fang *et al.* (2012). However, although all 30 isolates in the PCR screening were negative for AFLP fragment markers SAL-36 (Fang *et al.* 2012), SAL-40 (Lan *et al.* 2007) and SAL-43 (Fang *et al.* 2012), these markers were not excluded since they showed variation among 8 DT1 *S.* Typhimurium isolates (S0031, S0032, S0036, S0041, S0042, S0043, S0049 and S0050 in Dataset S1), which were screened earlier to determine a positive control for the PCR with these markers.

For each SNP marker, PCR amplicons were sequenced of at least one isolate with the SNP allele and one isolate with the wild-type allele. PCR amplicons were also sequenced for confirmation of MOL-PCR results that did not comply with previous results of the PCR screening. The sequences of all PCR amplicons confirmed the MOL-PCR assay results. For unpredicted negative MOL-PCR results (SAL-20, SAL-38, SAL-47, SAL-58 and SAL-71), mismatches were observed in the alignment of the PCR amplicon sequence and the target-specific sequence of the upstream and downstream probes, which prevented adequate annealing of the probes and explained why no ligation occurred. Polymorphisms in the binding site of primer SAL-10-F may explain the negative results for SAL-10 in the PCR screening for isolates S0007, S0023, S0025 and S0030 while positive results for SAL-10 in the MOL-PCR were seen for these isolates. Indeed, the target-specific sequence of the upstream and downstream probes of marker SAL-10 (SAL-10-U and SAL-10-D) aligned perfectly with the PCR amplicons generated for these isolates with primers SAL-10-R (same position as SAL-10-U) and SAL-10-F-nested, which is located downstream of SAL-10-F. Primer SAL-10-F was used in the PCR screening while SAL-10-F-nested was only used for PCR amplicon sequencing.

During the development of the MOL-PCR assay for subtyping, SNP SAL-57 was also eliminated, since no positive isolate could be identified and high background MFI values, as measured through the NTC, were obtained, even after redesigning of the probes.

As internal positive control markers, invasive gene *invA* (Barbau-Piednoir *et al.* 2013) and a SNP in the β subunit of RNA polymerase encoding gene *rpoB* (Hernández Guijarro *et al.* 2012) were selected. These markers target, respectively, all *Salmonella* species and *S.* Typhimurium and its monophasic variant *S.* 1,4,[5],12:i:-. To verify the specificity of markers *invA* and *rpoB*, the

MOL-PCR assay was performed on isolates of species that are unrelated or closely related to *Salmonella* and on isolates of other serovars of *S. enterica* subsp. *enterica*. The results, summarised in Table S5, confirmed the specificity of the internal positive control makers.
As such, a total of 50 discriminative markers and 2 internal positive control markers were nominated for MOL-PCR assay development and these markers were distributed over 3 MOL-PCRs as indicated in Table 4.1. Twenty-one out of the 50 discriminative markers were also included in the subtyping assay of Fang *et al.* (2012).

4.3.3 Stability study

The *in vitro* stability of the markers in the subtyping assay was examined by comparing the MOL-PCR profiles of 31 *S.* Typhimurium and *S.* 1,4,[5],12:i:- (indicated in Dataset S1 in column Stability_experiment with '1' if included) before and after an experiment of 50 serial passages in LB broth. No changes were observed in the MOL-PCR profiles of all 31 isolates before and after the 50 serial passages.

4.3.4 Validation of the MOL-PCR assay for subtyping

For validation of the MOL-PCR assay for subtyping, the method was performed twice in independent assays on a collection of 519 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with a known phage type and MLVA profile, referred to as the validation panel (S0001-S0519 in Dataset S1) and on 13 isolates related to 2 different outbreaks (S0520-S0532 in Dataset S1). Each isolate was assigned the same MOL-PCR profile in both independent assays.
In the validation panel, 51 different MOL-PCR profiles were observed, which are presented in Figure 4.1 and in Dataset S1. The most common MOL-PCR profiles are profiles 15-1-1 (in-house coded as 2-1-1), 255255-8843835-1155 (16-12-8), 15-3-1 (2-2-1) and 3.91×10^{21}-1.11×10^{13}-4199 (33-28-10) which were observed for, respectively, 157 (30.3% of the isolates in the validation panel), 97 (18.7%), 86 (16.6%) and 38 (7.3%) isolates. All other profiles were detected in less than 15 isolates. The 341 *S.* Typhimurium isolates in the validation panel were grouped into 44 different MOL-PCR profiles and the 178 *S.* 1,4,[5],12:i:- into 7 different MOL-PCR profiles. According to the MOL-PCR results, the *fljB* gene could not be detected in all *S.* 1,4,[5],12:i:- isolates with the probe pair of marker SAL-73.
Markers SAL-50 and SAL-51 for detection of the left and right junction of SGI1 were observed in isolates with phage types DT104, U302, DT12, DT120 and

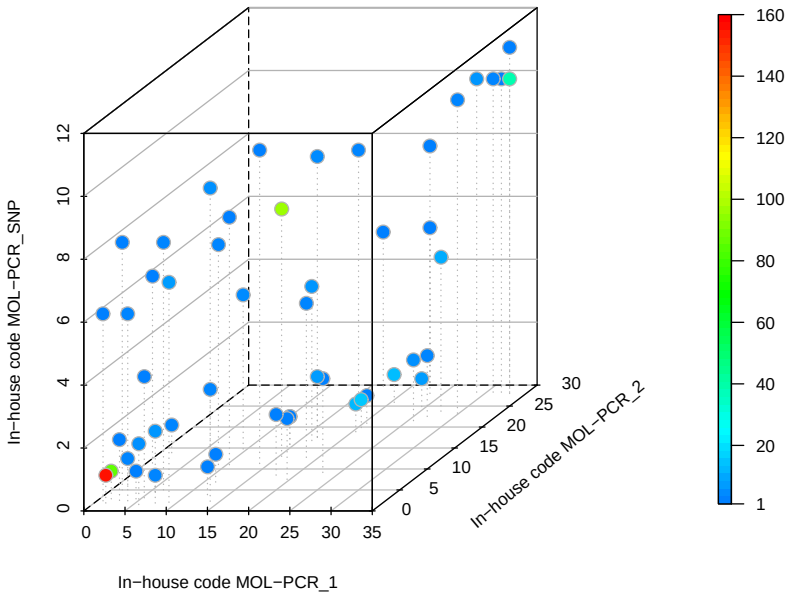

Figure 4.1: Visualisation of the 51 different MOL-PCR profiles (represented by 51 points) observed in the validation panel of 519 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with indication of the number of isolates in each profile, using a colour code. The colour code is automatically adapted to the number of isolates included in the data analysis with the R-application.

DT110, of which, to our knowledge, all but DT110 have already been reported in literature (Boyd *et al.* 2002; Carattoli *et al.* 2002; Lawson *et al.* 2002).
The discriminatory power was calculated as Simpson's index of diversity (D) on the 519 isolates in the validation panel and was 0.84 for the MOL-PCR assay, 0.84 for phage typing and 0.98 for MLVA.
Epidemiological concordance was evaluated by testing isolates related to two different outbreaks with the MOL-PCR assay. The first outbreak included isolates S0520 up to S0524 with S0525 as out-group; the second outbreak consisted of isolates S0526 up to S0530 with isolates S0531 and S0532 as out-group. The developed subtyping assay assigned identical MOL-PCR profiles to the outbreak isolates and separated them from their out-group isolates (Table 4.2).

Table 4.2: Subtyping data of *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates related to 2 different outbreaks.

Sample	Serovar	MOL-PCR profile	MOL-PCR in-house code	Phage type	MLVA profile
S0520[a]	Typhimurium	$1.91 \times 10^7 - 3.26 \times 10^{10} - 1$	$21 - 19 - 1$	DT195	$3 - 12 - 10 - \text{NA} - 311$
S0521[a]	Typhimurium	$1.91 \times 10^7 - 3.26 \times 10^{10} - 1$	$21 - 19 - 1$	DT195	$3 - 12 - 10 - \text{NA} - 311$
S0522[a]	Typhimurium	$1.91 \times 10^7 - 3.26 \times 10^{10} - 1$	$21 - 19 - 1$	DT195	$3 - 12 - 10 - \text{NA} - 311$
S0523[a]	Typhimurium	$1.91 \times 10^7 - 3.26 \times 10^{10} - 1$	$21 - 19 - 1$	DT195	$3 - 12 - 10 - \text{NA} - 311$
S0524[a]	Typhimurium	$1.91 \times 10^7 - 3.26 \times 10^{10} - 1$	$21 - 19 - 1$	DT195	$3 - 12 - 10 - \text{NA} - 311$
S0525[b]	Typhimurium	$15 - 3 - 1$	$2 - 2 - 1$	DT120	$3 - 15 - 5 - \text{NA} - 211$
S0526[c]	1,4,[5],12:i:-	$15 - 1 - 1$	$2 - 1 - 1$	DT138	$3 - 13 - 11 - \text{NA} - 211$
S0527[c]	1,4,[5],12:i:-	$15 - 1 - 1$	$2 - 1 - 1$	DT138	$3 - 13 - 11 - \text{NA} - 211$
S0528[c]	1,4,[5],12:i:-	$15 - 1 - 1$	$2 - 1 - 1$	DT138	$3 - 13 - 11 - \text{NA} - 211$
S0529[c]	1,4,[5],12:i:-	$15 - 1 - 1$	$2 - 1 - 1$	DT138	$3 - 13 - 11 - \text{NA} - 211$
S0530[c]	1,4,[5],12:i:-	$15 - 1 - 1$	$2 - 1 - 1$	DT138	$3 - 13 - 11 - \text{NA} - 211$
S0531[d]	Typhimurium	$15 - 3 - 1$	$2 - 2 - 1$	RDNC	$3 - 14 - 11 - \text{NA} - 211$
S0532[d]	Typhimurium	$1.21 \times 10^{18} - 2.58 \times 10^{11} - 4199$	$30 - 23 - 10$	DT104	$3 - 14 - 18 - 14 - 311$

[a] isolates of the first outbreak; [b] out-group isolate of the first outbreak; [c] isolates of the second outbreak; [d] out-group isolate of the second outbreak;
MLVA: multiple-locus variable-number of tandem repeats analysis; MOL-PCR: multiplex oligonucleotide ligation-PCR; RDNC: reacts-but-does-not-conform

4.3.5 PFGE results

To examine if the isolates of the 3 most observed MOL-PCR profiles, *i.e.*, 15-1-1, 255255-8843835-1155 and 15-3-1, in the validation panel could be further discriminated, PFGE was performed on a total of 53 isolates. Two clusters could be observed (Figure S1 in the supporting information). Cluster A grouped all isolates of MOL-PCR profiles 15-1-1 and 15-3-1, which differ only by marker SAL-73, *i.e.*, *fljB*. Cluster B included all isolates of MOL-PCR profile 255255-8843835-1155. Cluster A was divided into 16 subgroups of which 2 subgroups comprised isolates of both MOL-PCR profiles 15-1-1 and 15-3-1. One subgroup in cluster B comprised 13 of the 17 isolates in this cluster while the other 3 subgroups consisted of only 1 or 2 isolates.
This dissimilarity in variation between cluster A and B for PFGE patterns is in agreement with the difference in number of distinct MLVA profiles in cluster A and B, 18 and 7 MLVA profiles, respectively, but contrasts with phage typing results, as in each cluster, 12 distinct phage types were identified. While PFGE could further divide isolates with the same MOL-PCR profile into separate subgroups, also MOL-PCR could make a distinction between isolates with the same PFGE pattern in 2 subgroups of cluster A. Similarly, isolates with the same MLVA profile (*e.g.*, 3-12-9-NA-211) were assembled into distinct subgroups according to their PFGE patterns and are associated with different MOL-PCR profiles and phage types. Likewise, isolates with the same phage type are spread over clusters A and B with different PFGE patterns, MLVA and MOL-PCR profiles.

4.3.6 Selection of most discriminative markers

It was observed that there were nine groups of two or three markers which were always present or absent together in the isolates of the validation panel: (1) SAL-10/SAL-23/SAL-42, (2) SAL-11/SAL-15, (3) SAL-16/SAL-27, (4) SAL-56/SAL-65, (5) SAL-49/SAL-61/SAL-63, (6) SAL-21/ SAL-33, (7) SAL-50/SAL-51, (8) SAL-37/SAL-38 and (9) SAL-36/SAL-40/SAL-43.
For each marker, Simpson's index of diversity was calculated. The most discriminative markers were SAL-10/SAL-23/SAL-42, SAL-11/SAL-15, SAL-18, SAL-73 and SAL-16/SAL-27, thus 7 prophage gene markers, 1 AFLP marker and 1 SNP marker. The 51 observed MOL-PCR profiles could be reconstructed with a selection of 17 markers: SAL-11, SAL-16, SAL-18, SAL-23, SAL-26, SAL-29, SAL-35, SAL-36, SAL-53, SAL-55, SAL-62, SAL-66, SAL-67, SAL-70, SAL-71, SAL-73 and SAL-74. Hence, these markers encompass 8 prophage genes, 3 antibiotic resistance genes, 2 AFLP fragments, 2 SNPs, MLVA locus STTR10 and *fljB*.

4.4 Discussion

In our attempt to design a complementary subtyping method for *S.* Typhimurium and its monophasic variant *S.* 1,4,[5],12:i:-, we have developed a MOL-PCR method for subtyping which combines different types of molecular markers in a high-throughput multiplex assay. We screen 52 molecular markers in 3 multiplex ligation assays, thereby avoiding the issues associated with multiplex PCR assays. Starting from a single colony, subtyping results are delivered within 8 h, which makes the MOL-PCR assay a convenient subtyping method for outbreak investigations, for which rapidity is of crucial importance. A similar turn-around-time is currently more difficult to be obtained with WGS. Another important aspect for outbreak investigations and for long-term surveillance studies is the stability of the assessed markers, which is not the case for all markers in MLVA (Dimovski *et al.* 2014) but which was demonstrated for our MOL-PCR assay by the results of a serial passage experiment. The data analysis and interpretation are objective and computerised, in contrast to that of phage typing. The data analysis results in a MOL-PCR profile consisting of 3 numerals, which can be easily compared between different laboratories (which is more difficult for PFGE) and straightforwardly stored in an electronic database, so that the developed subtyping method is suitable for use in an international surveillance network. Additionally, the presented visualisation as a 3-dimensional scatterplot is a flexible tool for outbreak detection when used with a limited number of isolates or for surveillance when used for a large collection of isolates, as the colour scheme will adapt itself to the number of isolates included in the data analysis. This objective data analysis is easily done by a non-bioinformatics expert, which might not be the case for WGS data analysis.

Besides rapidity, both Struelens *et al.* (1996) and van Belkum *et al.* (2007) propose flexibility, accessibility, cost and ease of use as convenience criteria for microbiological epidemiologic typing methods. The MOL-PCR assay is flexible in the sense that the technology of the MOL-PCR assay can be applied for (sub)typing of other pathogens. Nevertheless, for other pathogens, a different set of probes will have to be developed, as is also the case for *e.g.*, MLVA and MLST.

The accessibility criterion deals with the availability of reagents and equipment and with the required skills for the method. A ligation reaction requires the same type of reagents, equipment and skills as a regular PCR, which is a generally used laboratory technique. Also for the hybridisation of the MOL-PCR products to MagPlex-TAG microspheres and incubation with SAPE, no special skills or equipment are necessary and required reagents are commonly available. The MOL-PCR assay was developed on a MAGPIX system, which stands at the lower end of the Luminex portfolio regarding cost and skills for use and maintenance

of the system and which is feasible for a routine laboratory. The reagents and consumables cost for subtyping 1 isolate with the designed MOL-PCR method is lower than 10 euros, if the 3 MOL-PCR assays are combined on a 96-well plate so that 29 isolates are subtyped in 1 run.

The MOL-PCR method is designed for processing 96-well plates, which takes about 3.5 h hands-on time and requires no high-level technical skills. Analysis and interpretation of the results is straightforward by using the GPP and an available R-script and is thus not dependent on specialised commercial software. As such, this subtyping method scored well on the ease of use criterion.

In addition to stability of the assessed markers and suitability for computerised analysis and storage of results, both Struelens *et al.* (1996) and van Belkum *et al.* (2007) propose reproducibility, epidemiological concordance, discriminatory power (D) and typeability (T) as performance criteria for microbiological epidemiologic typing methods. As test population for the typing method, both authors recommend a large collection ($n > 100$) and Struelens *et al.* (1996) refine as 'Large size collections of unrelated strains ($n > 100$), not selected on the basis of type characteristics, are recommended for the unbiased and precise comparison of the T and D values of different typing systems'. Our validation panel of 519 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates complied with this recommendation.

A reproducible method is able to assign the same type to an isolate that was tested multiple times and in an independent manner. The reproducibility of the MOL-PCR assay was proven by an independent repeat of the subtyping method, in which all isolates of the validation panel and of the 2 outbreaks received the same MOL-PCR profile as in the initial experiment. In contrast to PFGE, the molecular techniques used in the MOL-PCR assay, *i.e.*, ligation, PCR and hybridisation to microspheres, can be standardised without great effort, so that results may also be reproducible between different laboratories.

In the experiment for the assessment of the epidemiological concordance, identical MOL-PCR profiles were observed for outbreak isolates, which were clearly distinguished from their out-group isolates. These results concurred with results of phage typing and MLVA, although separation of out-group isolate S0531 (Table 4.2) from the respective outbreak isolates would be difficult with only MLVA data since its MLVA profile was different from the MLVA profile of the outbreak isolates in only one repeat at an instable locus (Dimovski *et al.* 2014).

The instability of 3 MLVA loci may also explain the higher value for Simpson's index of diversity (D) for MLVA compared to the MOL-PCR assay, as large numbers of MLVA profiles are produced as a result of these rapidly evolving loci, which make MLVA less suitable for investigation of long-lasting outbreaks and long-term surveillance. According to Simpson's index of diversity, the MOL-PCR method has the same discriminatory power as phage typing. However, whereas

phage typing produces NT and RDNC results, a 100% typability was observed in the MOL-PCR assay since all of the 519 isolates of the validation panel were assigned a MOL-PCR profile and could thus be subtyped by the assay. Compared to the subtyping method described by Fang *et al.* (2012), based on multiplex PCR with detection through a direct hybridisation assay, the developed MOL-PCR assay provides an increased discrimination since the ligation reaction only occurs under strict conditions: the upstream probe has to hybridise exactly adjacent to the downstream probe and a strict complementarity is compulsory for the base pairs flanking the ligation site. Even so, this more stringent discrimination was not reflected in the discriminatory power as calculated by Simpson's index of diversity, which was 0.84 for the MOL-PCR method compared to 0.95 for the multiplex PCR-based method (Fang *et al.* 2012). This might be explained by the dissimilar test panels used for evaluation of both subtyping methods. Our validation panel of 519 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates included 33 distinct phage types, whereas Fang *et al.* (2012) tested a selected panel of 438 *S.* Typhimurium representing 58 phage types and thus calculated Simpson's index of diversity on a smaller collection with a higher variation. Moreover, simple adaptations to the collection of isolates tested may have significant effects on Simpson's index of diversity, *e.g.* if half of the isolates with the 3 most frequently observed MOL-PCR profiles is left out of the validation panel, which would then be reduced to 350 isolates, still above the recommended size of > 100 isolates, Simpson's index of diversity would increase to 0.95 for the MOL-PCR assay and would thus comply with the acceptable discriminatory power for a "more or less 'ideal'" typing system (van Belkum *et al.* 2007). As Simpson's index of diversity is very much dependent on the collection of isolates tested, a more strict definition of such a test panel might be required for evaluation of subtyping methods.

The MOL-PCR assay with its visualisation tool (Figure 4.1) and coupled with insightful epidemiological data, offers a user-friendly and rapid approach for outbreak investigations. If, however, a frequent MOL-PCR profile (*i.e.*, 15-1-1, 255255-8843835-1155, 15-3-1 and 3.91×10^{21}-1.11×10^{13}-4199) would be obtained, the isolate might be further characterised by MLVA, PFGE or WGS. Nonetheless, due to the instability of three MLVA loci in *S.* Typhimurium (Boxrud 2010; Dimovski *et al.* 2014), it may not be clear how to handle isolates that differ in one of the instable loci (Petersen *et al.* 2011; Friesema *et al.* 2012; Garvey *et al.* 2013; Kuhn *et al.* 2013; Paranthaman *et al.* 2013). Therefore, for current routine laboratories, PFGE might be more appropriate. In our case, PFGE could further discriminate isolates with the same MOL-PCR profile, but at the other hand, different MOL-PCR profiles were obtained for isolates within the same PFGE cluster. This is even more an issue with MLVA profiles. This interwoven tangle of subtyping results illustrates the complications that are encountered when comparing different sets of subtyping data, which will

only become more complicated when WGS data will be compared to historical subtyping results. Also, with WGS, agreements will have to be made within the community as to decide whether two isolates are identical or not, as the resolution is at the single nucleotide level and *in vivo*/*in vitro* mutations or sequencing errors might occur.

The discrimination for the most frequently observed MOL-PCR profiles can be increased by including more markers in the MOL-PCR assay. Such additional markers may be identified by WGS comparison of different isolates with the same commonly observed MOL-PCR profile. However, if more markers are included in the assay, this may lead to an increase of the cost and effort of the method. To avoid an expansion of the MOL-PCR assay, redundant markers which were present or absent together in the isolates of the validation panel could be removed. However, the markers that are redundant for our validation panel, could be critical for discrimination when applying the MOL-PCR assay to future collections of *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates. Therefore, it might be more appropriate to evaluate the redundancy of the molecular markers included in the method after routinely subtyping *S.* Typhimurium and its monophasic variant *S.* 1,4,[5],12:i:- with the MOL-PCR assay for a period of *e.g.*, 3 years in multiple reference laboratories.

Ultimately, WGS might become the gold standard for subtyping of any pathogen, but during the time that not all routine laboratories have the resources and data analysis capabilities and agreements on interpretation for WGS of *S.* Typhimurium and its monophasic variant *S.* 1,4,[5],12:i:-, the developed MOL-PCR assay may be considered as an inexpensive complement of currently applied methods in routine subtyping with a readily accessible, computerised data analysis pipeline.

Supporting information

Dataset S1, Tables S1-S11 and Figure S1 are available at http://static-content.springer.com/esm/art%3A10.1007%2Fs00253-015-6831-7/MediaObjects/253_2015_6831_MOESM1_ESM.pdf (file format: pdf).

Acknowledgements

Sequencing and capillary electrophoresis for MLVA was performed at the Platform Biotechnology and Molecular Biology at the Scientific Institute of Public Health (WIV-ISP). We thank E. Barbau-Piednoir for providing the DNA template of the bacterial isolates used for the specificity test of markers *invA*

and *rpoB*. The authors also thank J. Dewit, G. Dupont, M. Thirionet and C. Wildemauwe for the phage typing, MLVA and PFGE analyses.

Compliance with Ethical Standards

Funding This study was funded by grant P4044.0103 (SalMolType) from the Scientific Institute of Public Health (WIV-ISP - RP/PJ). The National Reference Centre for *Salmonella* and *Shigella* is partially supported by the Belgian Ministry of Social Affairs through a fund within the Health Insurance System.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Chapter 5

Whole genome sequencing for subtyping of *Salmonella enterica* subsp. *enterica* serovar Typhimurium

Abstract

In this chapter the current issues concerning the analysis of whole genome sequencing (WGS) data for surveillance in public health are explored with WGS data on 32 *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) and serovar 1,4,[5],12:i:- (*S.* 1,4,[5],12:i:-) isolates.
The first issue is the workflow to be followed. As currently no whole genome multilocus sequence typing (MLST) scheme is available, only a single nucleotide polymorphism (SNP)-based approach was investigated with 3 different workflows. The influence of the reference genome and of error correction is elaborated.
For examination of the second issue, namely the definition of a distinct subtype, the SNP analysis on 3 *S.* Typhimurium strains isolated from the same patient at different time points was more closely considered.
For exploring the third and fourth issue, which are the inference of phenotypic data from WGS analysis and the link between WGS data and historical subtyping data, the tools available on the server of the Center for Genomic Epidemiology were used.
The main conclusion is that a SNP-based analysis has added value for subtyping of *S.* Typhimurium and *S.* 1,4,[5],12:i:-, as isolates with the same multiplex oligonucleotide ligation-PCR (MOL-PCR) profile, the same phage type and the same multiple-locus variable-number of tandem repeats analysis (MLVA) profile could be further discriminated. One has to take into account however that a different result is obtained when applying different SNP-based workflows. This latter observation is also important for defining a distinct subtype. Related to the tools of the Center for Genomic Epidemiology, a good correlation was observed between the detected resistance genes in the WGS data and the phenotypical antimicrobial resistance profile. To a certain extent, a link could be made between historical subtyping data and WGS data, but maybe it would be better to concentrate on a whole new nomenclature instead of holding on to the classical subtypes.

5.1 Aim of this study

Whole genome sequencing (WGS) has been postulated as the universal subtyping technique for pathogens with ultimate resolution. Nonetheless, up to date, no standard WGS subtyping protocol has been issued, since there are still many questions to be answered as evidenced by the European Food Safety Authority (EFSA) Scientific Colloquium Summary Report on the use of WGS of foodborne pathogens for public health protection (European Food Safety Authority (EFSA) 2014). In this chapter we study the WGS data of a selection of 32 *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) and serovar 1,4,[5],12:i:- (*S.* 1,4,[5],12:i:-) isolates for exploring the main discussion points, which were presented in the introductory section 1.3.3.

This chapter focusses on WGS for subtyping of *S.* Typhimurium related to surveillance, while the next chapter discusses a case study of WGS applied to investigate two outbreaks of *S.* Enteritidis.

The first question is the workflow to be followed for interpretation of WGS data: Should this workflow be single nucleotide polymorphism (SNP)-based or gene-based? Both workflows may be complementary, but for implementation of the gene-based approach, also referred to as whole genome multilocus sequence typing (MLST), two requirements have to be met. Firstly, there is a need for a predefined set of genes to compare the *de novo* assembled WGS reads to and secondly, a preferably international database of possible alleles for each gene has to be established, in order to assign a whole genome sequence type (ST) to the isolate that has to be subtyped. In March and April 2015, the first versions of core genome MLST schemes for *Campylobacter jejuni/coli* and *Neisseria meningitidis* were announced on the PubMLST databases website (Jolley 2015). However, since a whole genome MLST scheme is not yet available for *Salmonella*, we have restricted the current study to the SNP-based approach.

For the SNP-based approach different tools are currently being applied for similar applications in Europe. Public Health England uses a workflow based on the Genome Analyis Toolkit (GATK) (Ashton *et al.* 2014; Dallman *et al.* 2014; Ashton *et al.* 2015), while the Call SNPs & Infer Phylogeny (CSI Phylogeny) tool hosted on the server of the Danish Center for Genomic Epidemiology (Technical University of Denmark (DTU) 2015) uses BEDTools and SAMtools in its workflow. Using different tools might result in different phylogenetic trees and hence in different conclusions on the relatedness of isolates. In this study we apply both workflows for subtyping of *S.* Typhimurium and *S.* 1,4,[5],12:i:-, and we also perform SNP calling with the commercial software package CLC Genomics Workbench (CLC Bio). We applied this approach to a selection of *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates that resulted in two frequently occurring multiplex oligonucleotide ligation-PCR (MOL-PCR) profiles (chapter 4) and hence for which a higher discriminatory resolution would be appropriate.

As SNP calling relies on read mapping to a reference genome, we compare the workflows, applied on a subset of the WGS data, when using two different reference genomes, namely *S.* Typhimurium LT2, which is most referred to in literature, and *S.* Typhimurium SL1344, which is described as reference genome in literature for more virulent strains like *S.* Typhimurium ST313 (Okoro *et al.* 2012; Okoro *et al.* 2015).
In addition to different reference genomes, the effect of error correction of FASTQ reads on SNP analysis was evaluated, as it was hypothesised that error correction may result in additional discriminatory SNPs, while avoiding bias of false positive calling of SNPs. Two different error correction algorithms were applied: Blue, which is based on k-mer consensus and context (Greenfield *et al.* 2014), and Brownie, which is under development and which is based on de Bruijn graphs (M. Heydari, G. Miclotte and J. Fostier, personal communication).
The second question related to WGS for subtyping of pathogens for public health is the definition of a distinct subtype, *i.e.* until when can two isolates be considered as the 'same' subtype. The currently reported retrospective WGS studies of outbreaks of pathogens all add value to this issue, however, pathogens may also evolve during the infection within a patient to become more adapted to their host, to escape the immune system or to gain antibiotic resistance (Lieberman *et al.* 2011). Therefore, we concentrate on the SNP analysis of three *S.* Typhimurium strains that were isolated from the same patient over a period of 1.5 months.
The third question is the deduction of phenotypic data like antimicrobial resistance and virulence of WGS data. The ResFinder and PlasmidFinder tools on the server of the Center for Genomic Epidemiology (Technical University of Denmark (DTU) 2015) allow for identifying antibiotic resistance genes and plasmids, respectively. The latter may point out virulence plasmids, which are commonly found in *S.* Typhimurium (Rychlik *et al.* 2006).
The server of the Center for Genomic Epidemiology hosts also tools like MLST server, which can be used related to the fourth question, namely the link between WGS data and historical subtyping data, which is explored in the last part of this study.

5.2 Materials and methods

5.2.1 Bacterial isolates

All 32 human *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates in this study were provided by the Belgian National Reference Centre for *Salmonella* and *Shigella* (NRCSS) and are listed in Table 5.1. The selection of *S.* Typhimurium and

S. 1,4,[5],12:i:- isolates consists firstly of 5 isolates related to an outbreak and 2 out-group isolates that were collected during the same period of this outbreak and that were used in the original outbreak investigation as well. Secondly, 3 *S.* Typhimurium which were isolated at different time points from the same patient were added to the selection. Thirdly, 5 × 3 isolates were selected based on their frequently occurring MOL-PCR profile, namely 15-1-1 and 15-3-1 (Wuyts *et al.* 2015b) (chapter 4), in combination with frequent phage types DT193 and DT120 (Bertrand *et al.* 2014) and frequent multiple-locus variable-number of tandem repeats analysis (MLVA) profiles 3-12-10-NA-211 and 3-13-11-NA-211. Finally, 7 isolates were added which had the same phage type as the outbreak and their out-group isolates and with the same or a frequent MOL-PCR profile. Phage type, MLVA and antimicrobial susceptibility data were provided by the NRCSS and were collected as previously described by Wuyts *et al.* (2013) (chapter 2).

5.2.2 Whole genome sequencing

From each of the 32 selected *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates, a single colony was grown overnight (about 16 hours) in brain-heart infusion (BHI) broth. Genomic DNA was extracted with the Qiagen Genomic-tip 100/G kit. Sequencing was performed at the EMBL GeneCore facility on an Illumina HiSeq 2000 using 100 bp paired-end reads. Forty samples of *Salmonella* genomic DNA were multiplexed on a single lane (*i.e.* 8 *S.* Enteritidis isolates were also included in this run - see chapter 6).

5.2.3 Examining quality of reads

The quality of the raw Illumina FASTQ reads was evaluated using the FastQC 0.11.3 package (Babraham Bioinformatics 2015).

5.2.4 Error correction of FASTQ reads

FASTQ reads of isolates 11-0596, 11-0600, 11-1160, 11-1163 and 11-1164 corrected with the Brownie algorithm were provided by M. Heydari, G. Miclotte and J. Fostier.
The raw FASTQ reads of these isolates were also corrected with Blue 1.1.3 (Greenfield *et al.* 2014) using the default k-mer length of 25.

Table 5.1: Overview and microbiological data of *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates in the study.

ID	Serovar	Isolation date	MOL-PCR profile	Phage type	MLVA profile	Antimicrobial resistances	Additional information
11-0596	1,4,[5],12:i:-	15/01/2011	15-1-1	DT138	3-13-11-NA-211	ASSu	Outbreak
11-1163	1,4,[5],12:i:-	28/03/2011	15-1-1	DT138	3-13-11-NA-211	ASSu	Outbreak
11-1164	1,4,[5],12:i:-	28/03/2011	15-1-1	DT138	3-13-11-NA-211	ASSu	Outbreak
11-1165	1,4,[5],12:i:-	28/03/2011	15-1-1	DT138	3-13-11-NA-211	ASSu	Outbreak
11-1166	1,4,[5],12:i:-	26/03/2011	15-1-1	DT138	3-13-11-NA-211	ASSu	Outbreak
11-0600	Typhimurium	04/02/2011	15-3-1	RDNC	3-14-11-NA-211	ASSuSxtTTmp	out-group
11-1160	Typhimurium	10/04/2011	1.21×10^{18}-2.58×10^{11}-4199	DT104	3-14-18-14-311	AAmcSSu	out-group
S13BD00332	Typhimurium	02/03/2013	1.50×10^{8}-1.69×10^{11}-1155	ND	3-14-14-5-311	A	One patient
S13BD00591	Typhimurium	25/03/2013	1.50×10^{8}-1.69×10^{11}-1155	ND	3-14-14-5-311	A	One patient
S13BD00844	Typhimurium	19/04/2013	1.50×10^{8}-1.69×10^{11}-1155	ND	3-14-14-5-311	A	One patient
12-2003	1,4,[5],12:i:-	29/06/2012	15-1-1	DT120	3-12-10-NA-211	ASSuT	-
12-2203	1,4,[5],12:i:-	11/07/2012	15-1-1	DT120	3-12-10-NA-211	ASSuT	-
12-2460	1,4,[5],12:i:-	22/06/2012	15-1-1	DT120	3-12-10-NA-211	ASSuT	-
12-2455	1,4,[5],12:i:-	26/07/2012	15-1-1	DT193	3-12-10-NA-211	ASSuT	-
12-2599	1,4,[5],12:i:-	07/08/2012	15-1-1	DT193	3-12-10-NA-211	ASSuT	-
12-2730	1,4,[5],12:i:-	14/08/2012	15-1-1	DT193	3-12-10-NA-211	ASSuT	-

Table 5.1 *continued*

ID	Serovar	Isolation date	MOL-PCR profile	Phage type	MLVA profile	Antimicrobial resistances	Additional information
12-1558	1,4,[5],12:i:-	22/05/2012	15-1-1	DT193	3-13-11-NA-211	ASSuT	-
12-2314	1,4,[5],12:i:-	17/07/2012	15-1-1	DT193	3-13-11-NA-211	T	-
12-2379	1,4,[5],12:i:-	29/07/2012	15-1-1	DT193	3-13-11-NA-211	ASSuT	-
12-3792	Typhimurium	08/10/2012	15-3-1	DT120	3-12-10-NA-211	ASSuT	-
12-3907	Typhimurium	14/10/2012	15-3-1	DT120	3-12-10-NA-211	ASSuT	-
12-3990	Typhimurium	23/10/2012	15-3-1	DT120	3-12-10-NA-211	ASSuT	-
12-0084	Typhimurium	13/01/2012	15-3-1	DT193	3-12-10-NA-211	ASSuSxtTTmp	-
12-0161	Typhimurium	21/01/2012	15-3-1	DT193	3-12-10-NA-211	ASSuT	-
12-3663	Typhimurium	30/09/2012	15-3-1	DT193	3-12-10-NA-211	ASSuT	-
12-3558	1,4,[5],12:i:-	25/09/2012	15-1-1	DT138	3-12-11-NA-211	ASSuT	-
12-3582	1,4,[5],12:i:-	11/09/2012	15-1-1	DT138	3-12-11-NA-211	ASSuT	-
12-3583	1,4,[5],12:i:-	11/09/2012	15-1-1	DT138	3-12-11-NA-211	ASSuT	-
12-2984	1,4,[5],12:i:-	27/08/2012	15-1-1	RDNC	3-12-11-NA-211	ASSuT	-
12-2998	1,4,[5],12:i:-	not known	15-1-1	RDNC	3-12-11-NA-211	ASSuT	-
12-3067	1,4,[5],12:i:-	04/09/2012	15-1-1	RDNC	3-12-11-NA-211	ASSuT	-
12-3444	Typhimurium	22/09/2012	1.21×10^{18}-2.58×10^{11}-4199	DT104	3-16-16-13-311	AAmcSu	-

A: ampicillin; Amc: amoxicillin plus clavulanic acid; MLVA: multiple-locus variable-number of tandem repeats analysis; MOL-PCR: multiplex oligonucleotide ligation-PCR; NA: absence of a PCR amplicon in MLVA; ND: not determined; RDNC: reacts-but-does-not-conform; S: streptomycin; Su: sulphonamides; Sxt: trimethoprim plus sulfamethoxazole; T: tetracycline; Tmp: trimethoprim.

5.2.5 SNP analysis

SNP analysis was performed with the commercial software package CLC Genomics Workbench (CLC Bio), with GATK, which is used by Public Health England (Ashton *et al.* 2015), and with the CSI Phylogeny tool, which is available on the server of the Danish Center for Genomic Epidemiology (Technical University of Denmark (DTU) 2015).

CLC Genomics Workbench FASTQ reads were imported into CLC Genomics Workbench 8.0 (CLC Bio) using the Illumina Paired Importer and trimmed with the parameters quality score limit 0.001 (Q30) and maximum 2 ambiguous bases; and reads with a length below 15 nucleotides after quality trimming were discarded. Subsequently, the trimmed reads were mapped to the *S.* Typhimurium reference genome LT2 (NC_003197) or SL1344 (NC_016810) using default parameters. The Fixed Ploidy Variant Detection tool was used for variant calling with the ploidy set to 1, a required variant probability of 95.0%, a minimum frequency of 90.0% and a minimum coverage of 10. Variants were exported as vcf (variant call format) files.

Genome Analyis Toolkit (GATK) Nucleotides with a Phred score below 30 were trimmed from the trailing end of FASTQ reads with Trimmomatic 0.33 (Bolger *et al.* 2014). The processed paired reads were mapped to the *S.* Typhimurium reference genome LT2 (NC_003197) or SL1344 (NC_016810) with bwa mem (BWA 0.7.12) (Li and Durbin 2009). The resulting SAM file was sorted and converted to BAM format. Subsequently, duplicates were marked in the BAM file and the BAM file was indexed. These steps were performed with Picard tools 1.134 (Broad Institute 2015). SNPs were called with GATK 3.4-0 UnifiedGenotyper (DePristo *et al.* 2011) with ploidy set to 1 and default Phred-scaled confidence threshold of 30 at which variants should be called and emitted. The resulting SNPs were filtered to 90% consensus, minimum depth 10, minimum GQ 30 and minimum MQ 30 with SnpSift 4.1 (Cingolani *et al.* 2012).
In an additional experiment with the GATK SNP calling workflow, additional pruning of the SNPs to filter out SNPs within 10 bp of each other was performed with VCFtools 0.1.12b (Danecek *et al.* 2011).

CSI Phylogeny Raw reads were uploaded to the CSI Phylogeny 1.0a server (Kaas *et al.* 2014). As reference genome *S.* Typhimurium LT2 (NC_003197) or SL1344 (NC_016810) was uploaded. The default parameters as described by Kaas *et al.* (2014) were applied: minimum depth of 10 at SNP positions,

minimum relative depth of 10% at SNP positions, minimum 10 bp distance between SNPs, minimum SNP quality of 30, minimum read mapping quality of 25 and a minimum Z-score of 1.96. The CSI Phylogeny server performs read mapping with BWA (Li and Durbin 2009) and SNP calling with mpileup of SAMtools (Li *et al.* 2009). The depth at the mapped positions is calculated with genomeCoverageBed of BEDTools (Quinlan and Hall 2010), the SNP quality with SAMtools and the read mapping quality with BWA. The SNPs are filtered with parameters for minimum depth and minimum relative depth at SNP positions, minimum distance between SNPs (*i.e.* pruning), minimum SNP quality, minimum read mapping quality and a minimum z-score. For each of these parameters, the user of CSI Phylogeny server has to give a value, as described above. FastTree (Price *et al.* 2010) is then used to derive a maximum likelihood phylogeny tree.
The resulting phylogeny, which includes the reference genome, was downloaded as a newick file, which was imported into MEGA 6.06 (Tamura *et al.* 2013) for visualisation of the tree. The called SNPs were downloaded as vcf files.

Alignment and maximum likelihood phylogeny Vcf files exported from CLC Genomics Workbench or resulting from SNP calling with GATK were used to generate a pseudo-genome for each sample by replacing the bases in the reference genome LT2 or SL1344 by the SNPs that are present in the sample. All genome positions with a SNP in a set of samples were combined into a list from the vcf files with the GATK CombineVariants and VariantsToTable tools. The resulting list was used to extract the genome positions from the pseudo-genomes of the samples as a fasta file. The fasta sequences were imported into MEGA 6.06 (Tamura *et al.* 2013) and aligned using MUSCLE (Edgar 2004). A maximum likelihood phylogeny tree was constructed with the Jukes-Cantor substitution model as described by Ashton *et al.* (2015). The labels in the phylogenetic trees are coloured according to the phage type of the isolate, except for isolates S13BD00332, S13BD00591 and S13BD00844, which were isolated from one patient and from which the phage type was not determined (Table 5.1). Additionally, the 5 isolates of the outbreak and the 2 out-group isolates are indicated in boldface.

5.2.6 Tools available at the Center for Genomic Epidemiology

Next to the CSI Phylogeny server (Kaas *et al.* 2014), which was discussed related to the SNP analysis (section 5.2.5), the Center for Genomic Epidemiology (Technical University of Denmark (DTU) 2015) hosts also other tools, such as those that can be used for *in silico* inference of the antibiotic resistome, *i.e.* ResFinder (Zankari *et al.* 2012), or the virulome, *i.e.* VirulenceFinder (Joensen

et al. 2014). VirulenceFinder is however not available for *Salmonella*. As serovar specific virulence plasmids contribute to the virulence of *S.* Typhimurium through the *spv* genes, the presence of plasmids in the *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates was evaluated *in silico* with PlasmidFinder (Carattoli *et al.* 2014).

The MLST server (Larsen *et al.* 2012) of the Center for Genomic Epidemiology can be used to identify the sequence type (ST), based on 7 housekeeping genes, of a pathogen through its WGS data, instead of through the classical technique with PCR and Sanger sequencing (*i.e.* NGS-derived conventional MLST).

ResFinder, PlasmidFinder and MLST server accept raw FASTQ reads compressed with gzip or assembled contigs in multi-fasta format. In case raw FASTQ reads are uploaded, these are assembled with Velvet (Zerbino and Birney 2008) after quality trimming (Larsen *et al.* 2012). As a test, both raw reads and assembled contigs generated with CLC Genomics Workbench 8.0 of the *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates were uploaded to ResFinder and comparison of the results showed that the same or more resistance genes were recovered with the assembled contigs (data not shown). As the assembled contigs in multi-fasta format had a size of about 5 MB per isolate, while paired-end reads took up 2 × about 230 MB, uploading assembled contigs was more time-efficient.

***De novo* assembly** Raw reads were imported and trimmed in CLC Genomics Workbench 8.0 as described in section 5.2.5. The processed reads were *de novo* assembled in mapping mode 'map reads back to contigs' with automatic bubble and word size, with scaffolding and a minimum contig size of 200 nucleotides.

ResFinder Uploaded assembled contigs were searched with BLAST for all resistance genes present in the database with a minimum ID threshold of 98% and a minimum length of 60% of the resistance gene as described by Zankari *et al.* (2012).

PlasmidFinder Uploaded assembled contigs were searched with BLAST for all plasmid replicons present in the Enterobacteriaceae database with the default minimum ID threshold of 95% (Carattoli *et al.* 2014).

MLST server Uploaded assembled contigs were typed by MLST with the *Salmonella enterica* configuration (Larsen *et al.* 2012).

5.3 Results

5.3.1 Examining quality of reads

The raw reads of the 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates were checked with FastQC and did not reveal serious quality issues, except for isolates 11-0600, 11-1165 and 11-1166 where the overrepresented sequences indicated that Illumina paired end primers may be present in the reads. For all isolates it was observed that quality scores lowered towards the end of the reads, which is expected with Illumina data.

5.3.2 SNP analysis with different workflows, different reference genomes and error correction

A selection of *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with known relationships (based on classical subtyping methods) was used to evaluate the influence of different reference genomes, different workflows and error correction on SNP analysis. The selection consisted of the *S.* 1,4,[5],12:i:- outbreak isolates 11-0596, 11-1163 and 11-1164, which are expected to have a close relationship, and the out-group *S.* Typhimurium isolates 11-0600 and 11-1160, which are expected to be more distantly related to the outbreak isolates. The different workflows applied were read mapping and variant calling in CLC Genomics Workbench (referred to as CLC workflow), read mapping with BWA and SNP calling with GATK (referred to as GATK workflow) and the CSI Phylogeny server, which makes use of BWA for read mapping and SAMtools and BEDTools for SNP calling (referred to as CSI workflow). The different reference genomes used for read mapping were *S.* Typhimurium LT2 and *S.* Typhimurium SL1344. Error correction was performed with Blue and Brownie. The latter is an algorithm under development. As an error was encountered with the Brownie corrected reads during the GATK workflow and the CSI workflow (probably due to a bug in the algorithm under development), only results of SNP analysis with Brownie corrected data of the CLC workflow are presented.

Read mapping Uncorrected and corrected quality trimmed reads were mapped to reference genomes *S.* Typhimurium LT2 (NC_003197) or to *S.* Typhimurium SL1344 (NC_016810). Tables 5.2a, 5.2b, 5.3a, 5.3b and 5.4 show the read mapping metrics which were provided by the CLC, GATK and CSI workflow, respectively. No clear differences are observed between uncorrected and corrected reads or between reference genomes LT2 and SL1344. For both the GATK and

Table 5.2a: Read mapping metrics provided by the CLC workflow for reference genome LT2.

Error correction	Uncorrected		Blue		Brownie	
Isolate	Mapped reads (%)	% of reference genome covered	Mapped reads (%)	% of reference genome covered	Mapped reads (%)	% of reference genome covered
11-0596	97.73	97.77	97.72	97.75	97.74	97.78
11-1163	92.95	93.91	92.99	94.29	92.95	93.91
11-1164	96.12	96.16	96.10	96.15	96.12	96.17
11-0600	86.81	87.69	86.86	87.89	86.82	87.69
11-1160	92.65	92.91	92.67	93.03	92.66	92.91

Table 5.2b: Read mapping metrics provided by the CLC workflow for reference genome SL1344.

Error correction	Uncorrected		Blue		Brownie	
Isolate	Mapped reads (%)	% of reference genome covered	Mapped reads (%)	% of reference genome covered	Mapped reads (%)	% of reference genome covered
11-0596	97.90	97.94	97.90	97.94	97.89	97.92
11-1163	93.23	94.21	93.24	94.21	93.28	94.56
11-1164	96.44	96.50	96.45	96.50	96.43	96.50
11-0600	87.09	87.98	87.11	87.98	87.15	88.20
11-1160	92.91	93.18	92.93	93.19	92.95	93.31

Table 5.3a: Read mapping metrics provided by the GATK workflow for reference genome LT2.

Error correction	Uncorrected			Blue		
Isolate	Mean coverage depth	% of reference genome covered	% of reference genome covered with minimum depth 10	Mean coverage depth	% of reference genome covered	% of reference genome covered with minimum depth 10
11-0596	179.39	98.0	91.8	179.70	98.0	91.7
11-1163	44.78	98.0	97.9	45.09	98.0	97.9
11-1164	139.86	98.0	97.9	140.18	98.0	97.9
11-0600	151.90	98.0	97.7	152.40	98.0	97.7
11-1160	156.54	98.1	97.9	157.04	98.1	97.9

Table 5.3b: Read mapping metrics provided by the GATK workflow for reference genome SL1344.

Error correction	Uncorrected			Blue		
Isolate	Mean coverage depth	% of reference genome covered	% of reference genome covered with minimum depth 10	Mean coverage depth	% of reference genome covered	% of reference genome covered with minimum depth 10
11-0596	178.85	98.1	91.8	179.16	98.1	91.8
11-1163	44.73	98.1	98.0	45.03	98.1	98.0
11-1164	139.72	98.1	98.1	140.04	98.1	98.1
11-0600	151.63	98.2	97.8	152.14	98.2	97.8
11-1160	156.33	98.2	98.0	156.82	98.2	98.0

Table 5.4: Read mapping metrics provided by the CSI workflow.

Reference genome	LT2		LT2		SL1344		SL1344	
Error correction	Uncorrected		Blue		Uncorrected		Blue	
Isolate	% of reference genome covered by all isolates	Valid positions (% of reference genome)	% of reference genome covered by all isolates	Valid positions (% of reference genome)	% of reference genome covered by all isolates	Valid positions (% of reference genome)	% of reference genome covered by all isolates	Valid positions (% of reference genome)
11-0596	97.5	92.4	97.4	92.3	97.6	92.4	97.6	92.4
11-1163		97.9		97.9		98.0		98.0
11-1164		97.9		97.9		98.0		98.0
11-0600		97.7		97.7		97.8		97.8
11-1160		97.9		97.9		98.0		98.0

the CSI workflow, it is noted that the percentage of genome covered with minimum depth 10 (GATK) or the valid positions (CSI) for SNP calling is lower in isolate 11-0596 than in the other isolates, although the percentage of the reference genome which is covered is similar to the other isolates.

SNP calling The total number of SNP positions in all 5 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates that were detected with each workflow in combination with each reference genome and with or without error correction are presented in Tables 5.5a and 5.5b. Most SNP positions are observed with the CLC workflow, followed by the GATK workflow and the CSI workflow. More SNP positions are detected when applying error correction.
Matrices of the number of SNPs between the 5 isolates and the respective reference genome are provided for each of the 3 workflows in Tables A.1 up to A.14 in appendix A. For all 3 workflows, more SNPs are observed between the isolates and SL1344 than between the isolates and LT2. For both reference genomes LT2 and SL1344, the CSI workflow emits clearly less SNPs than the CLC and GATK workflow. With the Blue error correction in the GATK workflow, less SNPs were observed between the 3 outbreak isolates for both reference genomes, but also less SNPs were seen between the 3 outbreak isolates and out-group isolate 11-0600, while there were more SNPs between the outbreak isolates and out-group isolate 11-1160. This effect of error correction with Blue was also observed for the CSI workflow with SL1344 as reference genome, but was absent in the CLC workflow.

Table 5.5a: Total number of SNP positions between the 5 selected *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates for different workflows, with or without error correction and LT2 as reference genome.

Error correction	Uncorrected	Blue	Brownie
CLC workflow	1461	1477	1575
GATK workflow	1091	1271	-
CSI workflow	777	782	-

Table 5.5b: Total number of SNP positions between the 5 selected *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates for different workflows, with or without error correction and SL1344 as reference genome.

Error correction	Uncorrected	Blue	Brownie
CLC workflow	2363	2438	2485
GATK workflow	1816	2019	-
CSI workflow	1031	1035	-

For the GATK workflow, a comparison was made between the SNPs emitted with the Blue corrected reads as input and the SNPs emitted with uncorrected reads, for which the results are presented in Tables 5.6a and 5.6b. For isolate 11-1163 no SNPs were observed which were only emitted by the GATK workflow with uncorrected reads, while for the other isolates 2 up to 13 SNPs were only emitted with uncorrected reads. The SNPs that were only emitted with uncorrected reads may be false positive SNPs.

Table 5.6a: Comparison of SNPs emitted with Blue corrected reads versus uncorrected reads in the GATK workflow with reference genome LT2.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	530	585	574	496	609
Only Blue	69	53	94	165	187
Only uncorrected	6	0	4	4	2

Table 5.6b: Comparison of SNPs emitted with Blue corrected reads versus uncorrected reads in the GATK workflow with reference genome SL1344.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	984	1076	1167	908	1109
Only Blue	97	85	150	230	250
Only uncorrected	13	0	3	5	6

A comparison was also made between the SNPs emitted with Blue corrected reads and all unfiltered SNPs that were emitted with uncorrected reads by the

GATK workflow. This comparison is shown in Tables 5.7a and 5.7b. From this comparison it was observed that all, except 2, SNPs emitted with Blue corrected reads were already present in the unfiltered SNPs called with GATK. Thus, in general, error correction with Blue does not allow to find new SNPs, but improves the quality of the SNPs so that more SNPs pass the quality filter.

Table 5.7a: Comparison of SNPs emitted with Blue corrected reads versus unfiltered SNPs emitted with uncorrected reads in the GATK workflow with reference genome LT2.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	597	638	668	661	796
Only Blue	2	0	0	0	0
Only unfiltered uncorrected	238	158	186	283	263

Table 5.7b: Comparison of SNPs emitted with Blue corrected reads versus unfiltered SNPs emitted with uncorrected reads in the GATK workflow with reference genome SL1344.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	997	1076	1170	913	1115
Only Blue	0	0	0	0	0
Only unfiltered uncorrected	518	387	416	850	635

A last comparison was made between the SNPs emitted with the CLC or the CSI workflow on the one hand and all unfiltered SNPs that were emitted with uncorrected reads by the GATK workflow on the other hand. This comparison is presented in Tables 5.8a, 5.8b, 5.9a and 5.9b, respectively. From this comparison it was observed that the CLC workflow emitted additional SNPs which were not present in the unfiltered SNPs called with GATK. Contrary to the CLC workflow, all, except 1, SNPs emitted by the CSI workflow were already present in the unfiltered SNPs called with GATK. This indicates that the main difference between the GATK workflow and the CSI workflow is the filtering of the SNPs.

Table 5.8a: Comparison of SNPs emitted with the CLC workflow versus unfiltered SNPs emitted with uncorrected reads in the GATK workflow with reference genome LT2.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	537	563	669	664	824
Only CLC workflow	12	2	16	8	6
Only unfiltered uncorrected	298	233	185	280	235

Table 5.8b: Comparison of SNPs emitted with the CLC workflow versus unfiltered SNPs emitted with uncorrected reads in the GATK workflow with reference genome SL1344.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	972	1023	1294	1176	1336
Only CLC workflow	123	10	117	104	94
Only unfiltered uncorrected	543	440	292	587	414

Table 5.9a: Comparison of SNPs emitted with the CSI workflow versus unfiltered SNPs emitted with uncorrected reads in the GATK workflow with reference genome LT2.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	394	394	393	392	511
Only CSI	0	0	0	0	0
Only unfiltered uncorrected	441	402	461	552	548

Table 5.9b: Comparison of SNPs emitted with the CSI workflow versus unfiltered SNPs emitted with uncorrected reads in the GATK workflow with reference genome SL1344.

	Isolate				
Observed SNPs	11-0596	11-1163	11-1164	11-0600	11-1160
In common	622	622	621	642	730
Only CSI	0	0	0	0	1
Only unfiltered uncorrected	893	841	965	1121	1020

Phylogenetic trees The phylogenetic trees of the 3 outbreak and 2 out-group isolates are presented in Figures 5.1 up to 5.6 for the CLC, GATK and CSI workflow without error correction in combination with reference genomes *S.* Typhimurium LT2 and *S.* Typhimurium SL1344, and those with error correction are given in Figures A.1–A.8 in appendix A. From these figures we observe that the CLC workflow is not able to correctly cluster the outbreak isolates together, and this clustering does not improve with Blue or Brownie corrected reads. The GATK workflow does correctly cluster the outbreak isolates together with uncorrected reads, but not with Blue corrected reads. The CSI workflow clusters the outbreak isolates correctly together with both uncorrected and Blue corrected reads.

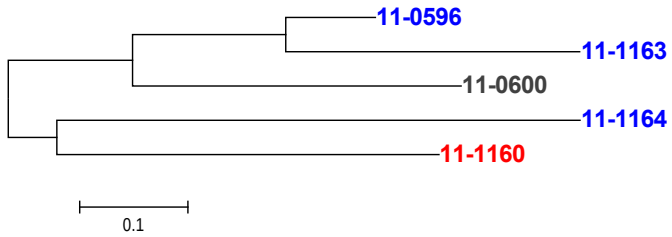

Figure 5.1: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with uncorrected reads in the CLC workflow with reference genome LT2.

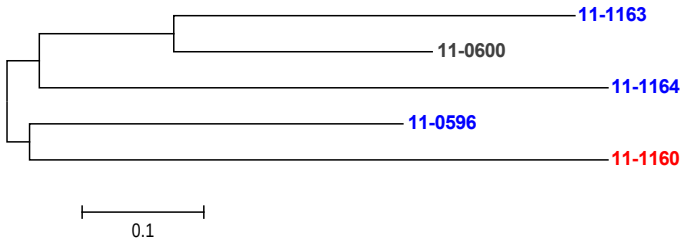

Figure 5.2: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with uncorrected reads in the CLC workflow with reference genome SL1344.

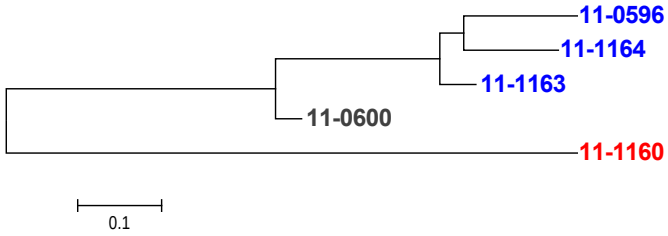

Figure 5.3: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with uncorrected reads in the GATK workflow with reference genome LT2.

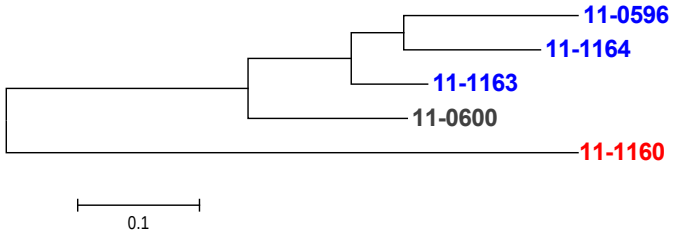

Figure 5.4: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with uncorrected reads in the GATK workflow with reference genome SL1344.

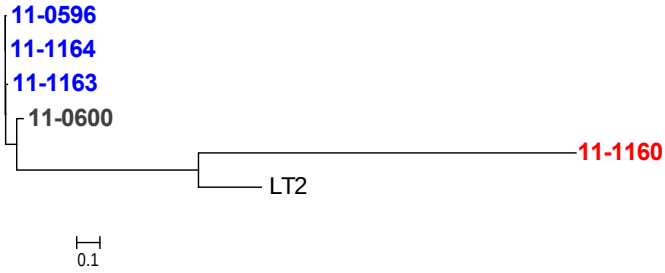

Figure 5.5: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with uncorrected reads in the CSI workflow with reference genome LT2.

Figure 5.6: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with uncorrected reads in the CSI workflow with reference genome SL1344.

5.3.3 SNP analysis for surveillance

The CLC, GATK and CSI workflows were also performed with the WGS data set of all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates to simulate the use of SNP analysis for surveillance. *S.* Typhimurium LT2 was taken as reference genome, since less SNPs were detected between the selection of 5 isolates and LT2 than between the selection of 5 isolates and SL1344. Error correction was not applied, since the comparison of workflows with and without error correction did not show an improvement for the corrected reads.
A total of 2400, 1826 and 1063 of SNP positions in all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates were detected with the CLC, GATK and CSI workflow,

respectively. The final phylogenetic trees are shown in Figures 5.7, 5.8 and 5.9. Here we see that with the CLC workflow neither the 5 outbreak isolates 11-0596, 11-1163, 11-1164, 11-1165 and 11-1166, nor the 3 *S.* Typhimurium isolated from 1 patient, *i.e.* S13BD00332, S13BD00591 and S13BD00844, are clustered together. These latter 3 isolates clustered together with the GATK workflow, but in this extended setting with 32 isolates, the GATK workflow fails to cluster the 5 outbreak isolates. The CSI workflow does cluster correctly both the 5 outbreak isolates and the 3 isolates from 1 patient.

Additionally, the isolates 11-1160 and 12-3444, which have the same MOL-PCR profile (1.21×10^{18}-2.58×10^{11}-4199) and the same phage type (DT104), are clustered with the CSI workflow. Moreover, in the cluster from isolate 12-1558 up to isolate 12-2203, which is a mixture of isolates with phage types DT120 and DT193, the *S.* Typhimurium and *S.* 1,4,[5],12:i:- tend to be grouped together. It was also observed that isolate 12-2460 was separated from isolates 12-2003 and 12-2203, although these 3 isolates have the same MOL-PCR profile (15-1-1), the same phage type (DT120) and the same MLVA profile (3-12-10-NA-211). Hence, SNP calling may provide additional discriminatory power.

To rule out the influence of the possible presence of Illumina paired end primers, as was indicated during the quality check with FastQC, the GATK workflow was repeated without isolates 11-0600, 11-1165 and 11-1166. This did not reduce the total number of SNP positions in the 29 isolates and also did not improve the clustering in the phylogenetic tree (Figure A.9).

In the CSI workflow, a strict filtering of the emitted SNPs is performed and also pruning is applied to remove a SNP that has a position within 10 bp of another SNP (Kaas *et al.* 2014). This pruning step was in an additional experiment also applied to the SNPs emitted with the GATK workflow, so that the total number of SNPs positions in all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates was reduced to 1694 positions. Nonetheless, this pruning did not improve the clustering in the phylogenetic tree (Figure A.10).

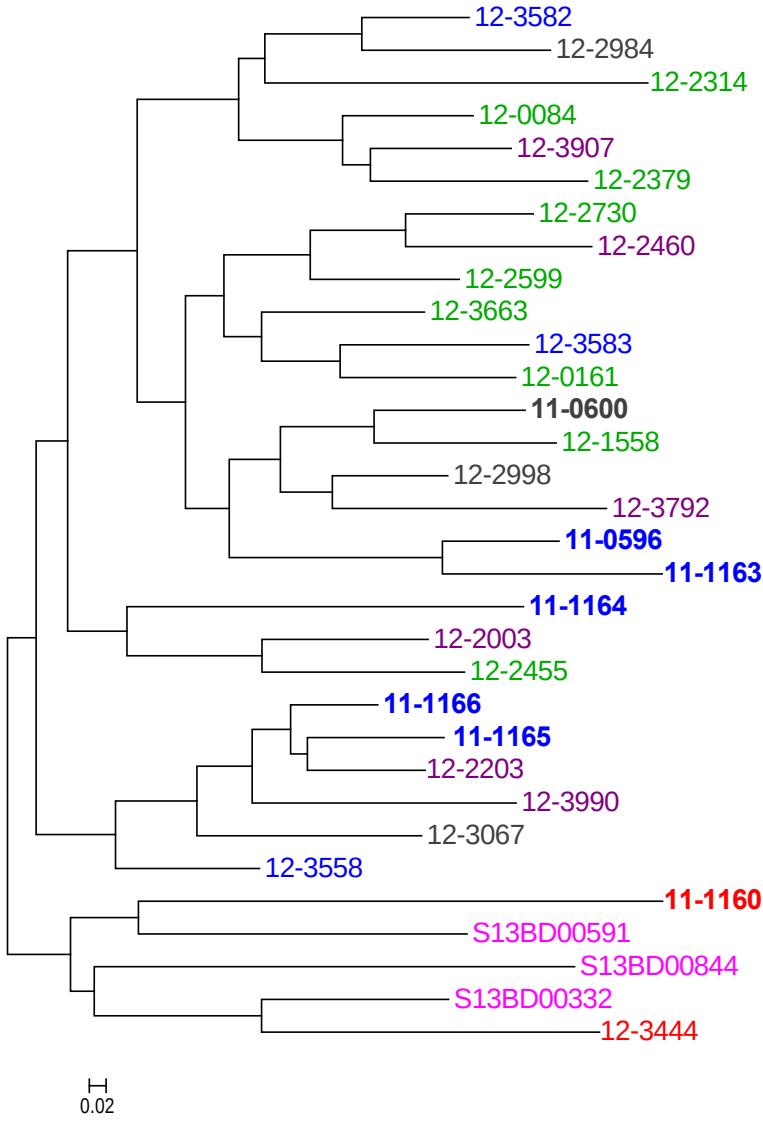

Figure 5.7: Phylogenetic tree of all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with uncorrected reads in the CLC workflow with reference genome LT2. The labels are colour coded according to the phage type of the isolate, and outbreak and out-group isolates are indicated in boldface.

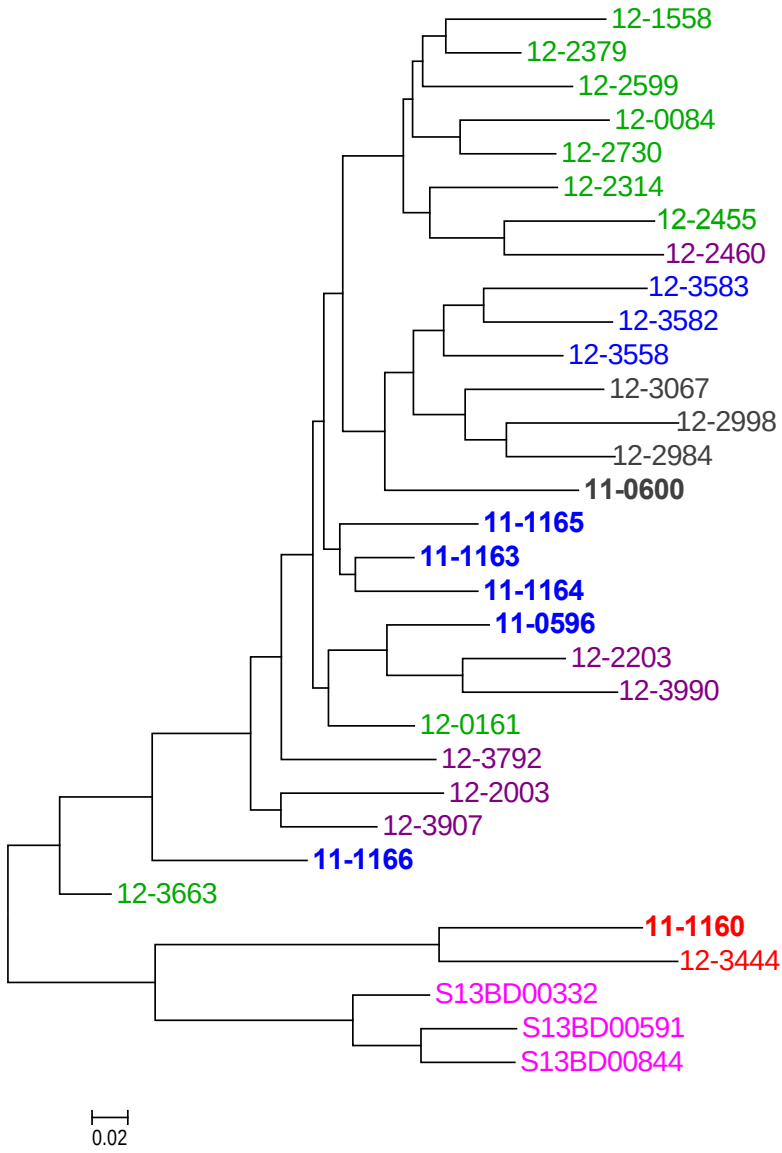

Figure 5.8: Phylogenetic tree of all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with uncorrected reads in the GATK workflow with reference genome LT2. The labels are colour coded according to the phage type of the isolate, and outbreak and out-group isolates are indicated in boldface.

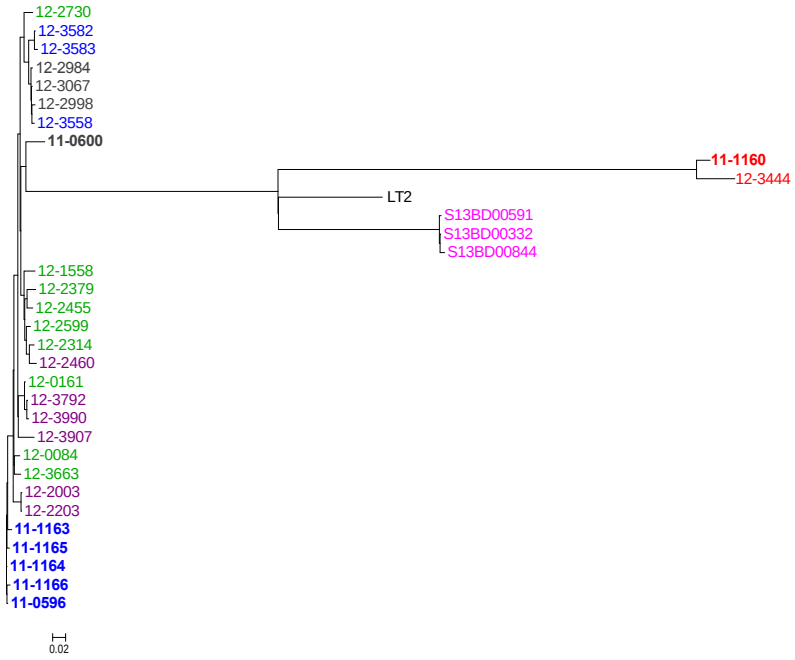

Figure 5.9: Phylogenetic tree of all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with uncorrected reads in the CSI workflow with reference genome LT2. The labels are colour coded according to the phage type of the isolate, and outbreak and out-group isolates are indicated in boldface.

5.3.4 Definition of a distinct subtype with SNP analysis

To explore the evolution of SNPs in *Salmonella* during infection of a human host, which may be an important factor for the definition of a distinct subtype, we took a closer look at the SNP analysis on 3 *S.* Typhimurium strains isolated from the same patient at different time points. Both the GATK workflow and the CSI workflow with LT2 as reference genome were able to cluster together isolates S13BD00332, S13BD00591 and S13BD00844. The GATK workflow resulted in a total of 518 SNP positions in all 3 isolates and put 179 to 209 SNPs difference between the 3 isolates (Table 5.10a). The CSI workflow emitted a total of 284 SNP positions in all 3 isolates and resulted in 6 to 12 SNPs between the 3 isolates (Table 5.10b).

Table 5.10a: Counts of SNPs between 3 *S.* Typhimurium strains isolated from the same patient at different time points for the GATK workflow with LT2 as reference genome.

Isolate	S13BD00332	S13BD00591	S13BD00844	LT2
S13BD00332	0			
S13BD00591	209	0		
S13BD00844	208	179	0	
LT2	335	434	389	0

Table 5.10b: Counts of SNPs between 3 *S.* Typhimurium strains isolated from the same patient at different time points for the CSI workflow with LT2 as reference genome.

Isolate	S13BD00332	S13BD00591	S13BD00844	LT2
S13BD00332	0			
S13BD00591	6	0		
S13BD00844	6	12	0	
LT2	306	306	304	0

5.3.5 Deduction of phenotypic data and link with historical subtyping data

***De novo* assembly** All 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates were *de novo* assembled with CLC Genomics Workbench 8.0, so that the contigs could be uploaded to the server of the Center for Genomic Epidemiology instead of the raw reads, which was more time-efficient. The draft assemblies had an average N50 of 180028 bp (range 25628 to 282662) with an average of 152 contigs (range 65 to 541) (Table A.15). The averages and ranges of the N50 and the contig count were calculated without isolates 11-0600, 11-1165 and 11-1166, as these had extreme high numbers of contigs with respect to all other isolates (2553, 5960 and 30836 respectively).

ResFinder The assembled contigs in multi-fasta format were uploaded to the ResFinder tool on the server of the Center for Genomic Epidemiology. The results are presented in Table 5.11 and correspond to the antibiotic resistance phenotype of the isolates, except for aminoglycoside gene *aac*(6')-Iy, which

was partially present in almost all isolates, but no resistance phenotype to kanamycin was observed.

PlasmidFinder The assembled contigs in multi-fasta format were uploaded to the PlasmidFinder tool on the server of the Center for Genomic Epidemiology. The results are presented in Table 5.12. In all isolates, except for isolate 12-2314, one or more plasmid replicons were identified. The plasmid replicons FIB(S) and FII(S), which were observed in isolates 11-1160 and 12-3444, were identified on the same contig, which indicates that both replicons are located on the same plasmid (Carattoli *et al.* 2014). This plasmid is also a virulence plasmid, as *spv*RABCD genes are located on the plasmids associated with replicons FIB(S) and FII(S) (accession numbers FN432031 and CP000858).

MLST server The assembled contigs in multi-fasta format were uploaded to the MLST tool on the server of the Center for Genomic Epidemiology. All isolates were assigned ST-34 with alleles *aroC*-10/*dnaN*-7/*hemD*-12/*hisD*-9/*purE*-5/*sucA*-9/*thrA*-2, except for isolates 11-1160, 12-3444, S13BD003332, S13BD00591 and S13BD00844, which were assigned ST-19 with alleles *aroC*-10/*dnaN*-19/*hemD*-12/*hisD*-9/*purE*-5/*sucA*-9/*thrA*-2.

Table 5.11: ResFinder results of 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates. The percentage of identity is given for the resistance genes that were detected.

	Aminoglycoside					Beta-lactamase		Sulphonamide		Tetracycline	Trimethoprim		
Resistance gene[a]	*aac*(6')-Iy	*aph*(3')-Ia	*strA*	*strB*	*aadA1*	*bla*$_{\text{TEM-1B}}$	*bla*$_{\text{CARB-2}}$	*sul1*	*sul2*	*tet*(B)	*dfrA1*	*dfrA14*	
Resistance against antibiotic[b]	K	K	S	S	S	A	A Amc	Su	Su	T	Tmp	Tmp	Phenotypic AMR
11-0596	99.54	-	100	100	-	100	-	-	100	-	-	-	ASSu
11-1163	99.54	-	100	100	-	100	-	-	100	-	-	-	ASSu
11-1164	99.54	-	100	100	-	100	-	-	100	-	-	-	ASSu
11-1165	99.54	-	100	100	-	100	-	-	100	-	-	-	ASSu
11-1166	99.54	99.85	100	100	-	100	-	-	100	-	-	-	ASSu
11-0600	99.54	-	100	100	100	100	-	-	100	100	100	-	ASSuSxt TTmp
11-1160	99.54	-	-	-	-	-	100	100	-	-	-	-	AAmcSSu
S13BD00332	99.54	-	-	-	-	100	-	-	-	-	-	-	A
S13BD00591	[c]	-	-	-	-	100	-	-	-	-	-	-	ND
S13BD00844	99.54	-	-	-	-	100	-	-	-	-	-	-	ND
12-2003	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2203	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2460	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2455	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2599	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2730	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-1558	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2314	99.54	-	-	-	-	-	-	-	-	100	-	-	T
12-2379	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3792	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3907	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3990	99.52	-	100	100	-	[d]	-	-	100	100	-	-	ASSuT
12-0084	99.54	-	100	100	-	100	-	-	100	100	-	99.79	ASSuSxt TTmp
12-0161	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3663	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3558	99.36	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3582	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3583	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2984	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-2998	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3067	99.54	-	100	100	-	100	-	-	100	100	-	-	ASSuT
12-3444	99.54	-	-	-	-	-	100	100	-	-	-	-	AAmcSu

[a] Resistance gene retrieved by ResFinder; [b] Only antibiotics which were phenotypically tested are given; [c] Gene *aac*(6')-Iy was partially detected on 2 different contigs when the ResFinder analysis was repeated with the parameter for minimum length of the resistance gene lowered to 40%; [d] Genes *bla*$_{\text{TEM-1A}}$ and *bla*$_{\text{TEM-1B}}$ were partially detected on the same contig but seperated by 25 bp when the ResFinder analysis was repeated with the parameter for minimum length of the resistance gene lowered to 40%. A: ampicillin; Amc: amoxicillin plus clavulanic acid; AMR: antimicrobial resistances; K: kanamycin; ND: not determined; S: streptomycin; Su: sulphonamides; Sxt: trimethoprim plus sulfamethoxazole (combination of trimethoprim and sulphonamides); T: tetracycline; Tmp: trimethoprim.

Table 5.12: PlasmidFinder results of 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates. The percentage of identity is given for the replicons that were observed.

Accession number	FN432031[a]	CP000858[a]	HE654726	AP005147	NC_009781	JN935898
Isolate	FIB(S)	FII(S)	Q1	I1	Col156	X1
11-0596	-	-	100	-	-	-
11-1163	-	-	100	-	-	-
11-1164	-	-	100	-	-	-
11-1165	-	-	100	-	-	-
11-1166	-	-	100	-	-	-
11-0600	-	-	100	-	-	98.94
11-1160	100	100	-	-	-	-
S13BD00332	-	-	-	-	-	98.66
S13BD00591	-	-	-	-	-	98.66
S13BD00844	-	-	-	-	-	98.66
12-2003	-	-	100	-	-	-
12-2203	-	-	100	-	-	-
12-2460	-	-	100	-	-	-
12-2455	-	-	100	-	-	-
12-2599	-	-	100	-	-	-
12-2730	-	-	100	-	-	-
12-1558	-	-	100	-	-	-
12-2314	-	-	-	-	-	-
12-2379	-	-	100	-	-	-
12-3792	-	-	100	-	-	-
12-3907	-	-	100	-	-	
12-3990	-	-	100	-	-	-
12-0084	-	-	100	-	96.71	-
12-0161	-	-	100	-	-	-
12-3663	-	-	100	-	-	-
12-3558	-	-	100	-	-	-
12-3582	-	-	100	-	-	-
12-3583	-	-	100	-	-	-
12-2984	-	-	100	-	-	-
12-2998	-	-	100	-	-	-
12-3067	-	-	100	-	-	-
12-3444	100	100	-	-	-	-

[a] virulence plasmid.

5.4 Discussion

In this study WGS data of 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:-, including isolates related to an outbreak and 3 *S.* Typhimurium isolated from the same patient at different time points, were used to evaluate four discussion points related to subtyping of pathogens by WGS for public health.

Regarding a SNP-based or gene-based workflow, only the SNP-based workflow has been explored, as for *Salmonella* a whole genome MLST scheme with accompanying database has not yet been established. Three different workflows have been applied and it was observed that the CSI Phylogeny produced the best phylogenetic tree, when taking the classical subtyping results into account. This workflow makes use of severe SNP filtering and pruning, so that it was also the workflow which emitted the least number of SNP positions in the set of all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates. This approach will most likely not report every SNP that may exist between different isolates, but for routine surveillance this might not be necessary (Kaas *et al.* 2014). Nonetheless, CSI Phylogeny has the disadvantage of being a black box application.

The study showed also that with SNP calling a higher discrimination can be obtained than with classical subtyping methods, as isolates with the same classical subtype were separated in the phylogenetic tree (*e.g.* Figure 5.9). Therefore, WGS might be an appropriate alternative molecular subtyping method. However, it was also clear that different workflows produce different results. For comparing results between different laboratories, a standardised SNP calling protocol may thus be necessary. More studies may also be needed to develop the most robust protocol. In addition, for easy exchange of subtype information between laboratories, the involved community should agree on a nomenclature to be used. While a nomenclature is not straightforward for a SNP-based approach that results in a phylogeny, a sequence type (ST) can be unambiguously assigned when applying whole genome MLST. So from this point of view, a gene-based workflow may be easier to implement. It will be interesting to repeat all analyses with a standard protocol for SNP calling and certainly to compare the results to those of whole genome MLST, when a scheme and database will be available for *Salmonella* (under development, BioNumerics, Applied Math).

A comparison was made between *S.* Typhimurium LT2 and *S.* Typhimurium SL1344 as a reference genome. All 3 workflows confirmed that, based on SNP counting, LT2 was more related to our set of *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates than SL1344. For SNP calling, this would make LT2 the better candidate to serve as reference genome, because if less SNPs are emitted by a SNP calling workflow, one would also expect less false positive emitted SNPs. This issue should be taken into account if one wants to standardise a SNP calling protocol: Should the reference genome that is closest to the set of isolates under study

be taken or is one reference genome assigned that should be used in all SNP analyses? The reference genome, which inherently introduces a bias in the analysis, could be abandoned by applying a reference-free SNP calling algorithm, *e.g.* with an algorithm based on De Bruijn graphs, or by SNP calling based on *de novo* assembly of the isolates under study (Leggett and MacLean 2014).
The false positive emitted SNPs could also be reduced by applying error correction to the WGS reads. Error correction, however, did not had a clear effect on the read mappings, but error correction with the Blue algorithm in combination with the GATK workflow resulted in less SNPs between the 3 selected outbreak isolates and in more SNPs between the 3 outbreak isolates and 1 out-group isolate. Nonetheless, the final maximum likelihood phylogeny did not cluster the 3 outbreak isolates correctly together, as also less SNPs were observed between the 3 outbreak isolates and the second out-group isolate. Unfortunately, it was not possible within the time frame of this PhD to apply the GATK and CSI workflow on debugged Brownie corrected data, but it would certainly be interesting to run these analyses in the future, especially since the CLC workflow seemed to be the least performing approach.
A comparison of uncorrected and Blue corrected reads showed that for 4 of the 5 isolates there were 2 to 13 SNPs that were only present in the uncorrected reads, and as such may be false positive SNPs. The single isolate that did not have SNPs only present in the uncorrected reads, had a mean coverage depth of about 45, while the mean coverage depth of the other 4 isolates ranged between 140 and 179. A high coverage depth may thus not always be beneficial. Currently, numbers of 30 to 100 are circulating for required coverage depth (European Food Safety Authority (EFSA) 2014) and this should be further examined (*e.g.* by downsampling WGS data) when creating a quality standard for WGS subtyping for public health.
Related to the definition of a distinct subtype based on SNP calling, most reported WGS studies refer to outbreak and out-group isolates for making an evaluation of how many SNPs difference there may be between two related isolates. However, pathogens also evolve during infection of a human host, as was observed by SNP analysis on 3 isolates from the same patient, but sampled over a period of about 1.5 months. Depending on the workflow, there were about 8 or 200 SNPs difference between these 3 isolates. Examination of the accumulation of SNPs in these isolates can be used to investigate whether these isolates were clones of a single strain, if the patient had a mixed infection or if this is just related to sequencing errors. More studies on isolates from the same patient, also sampled at the same time point, are required to estimate the factor of mixed infections and the factor of evolution in the total variation that may be present in an outbreak or in a background population. However, the current isolation method for *Salmonella* from human samples is not adapted to this kind of analysis, as it takes only one colony from a culture.

For inference of the antibiotic resistance phenotype, the available ResFinder tool was used to identify resistance genes in the *de novo* assembled contigs. The detected resistance genes provided a good correlation with the phenotypic resistance profile. Nevertheless, one resistance gene was partially found on 2 different contigs and in another isolate, partial $bla_{\text{TEM-1A}}$ and $bla_{\text{TEM-1B}}$ genes were found on the same contig, but separated by several nucleotides, which may be caused by assembly errors. At the other hand, a kanamycin resistance gene was found in the WGS data, while no such resistance could be phenotypically confirmed. These findings have to be taken into account when phenotype and genotype data do not correspond. Possible virulence plasmids were searched with the PlasmidFinder tool, as the VirulenceFinder tool is not yet available for *Salmonella* (this would be an interesting extension for the future). However, inference of the phenotype from WGS data will remain a difficult issue, as the presence of a gene does not imply that this gene is also expressed. Moreover, especially with respect to antibiotic resistance, not all mechanisms, which include also mutations, are yet known and can thus not be captured in a database. Extension of the current resistance genes databases behind tools like ResFinder with such mutations is to be done in the future.

Linking historical molecular subtyping data is straightforward for classical MLST and tools are already available for extracting the classical ST from WGS data. The MLST tool showed that classical MLST based on 7 housekeeping genes does not provide sufficient discrimination for *S.* Typhimurium and *S.* 1,4,[5],12:i:-, as only 2 different STs were obtained for all 32 isolates. More difficult is MLVA, which counts tandem repeats at a set of predefined loci, as these loci span more than a single Illumina read. These repeat loci form most likely gaps in a *de novo* assembly of short reads, as they are difficult to resolve. The long PacBio reads may bring a solution, but at the moment PacBio sequencing is too expensive (and its error rate is too high) for routine application in public health. PacBio sequencing may also be more suitable for inference of PFGE patterns, but for extracting a PFGE pattern from WGS data, a completely closed genome is required, which is not likely the goal of WGS for pathogen subtyping. For linkage of phage type with WGS data, it has been observed that isolates with the same phage type tend to cluster together in phylogenetic trees (Ashton *et al.* 2015) and this was also observed in our study (Figure 5.9). For validation of observations made related to linking historical subtyping data to WGS results, more WGS data are needed on available collections of pathogens which were characterised with these classical methods. However, one has to evaluate the added value of holding on to historical subtypes of pathogens, sometimes based on subjective methods like phage typing, while it may be necessary to develop a completely new and universal nomenclature for WGS subtyping. This will be indispensable for putting WGS into routine practice, where also the exchange of information on subtypes of pathogens between laboratories is a prerequisite.

Acknowledgements

This work has been made possible by the NGS service of the EMBL GeneCore for which we would like to thank Dr. Vladimir Benes. We would also like to thank M. Heydari, G. Miclotte and Prof. J. Fostier for providing the Brownie corrected sequencing reads.

Chapter 6

Whole genome sequence analysis of *Salmonella* Enteritidis PT4 outbreaks from a National Reference Laboratory's viewpoint

This chapter was previously published as:
V. Wuyts, S. Denayer, N. H. C. Roosens, W. Mattheus, S. Bertrand, K. Marchal, K. Dierick, and S. C. J. De Keersmaecker (2015a). "Whole genome sequence analysis of *Salmonella* Enteritidis PT4 outbreaks from a National Reference Laboratory's viewpoint". *PLOS Currents Outbreaks*. DOI: 10.1371/currents.outbreaks.aa5372d90826e6cb0136ff66bb7a62fc

Authors' contributions

Preparation of samples for whole genome sequencing (WGS) and the WGS analysis was performed by V. Wuyts. S. Denayer and K. Dierick contributed the general epidemiological, phage typing and the antimicrobial susceptibility data. W. Mattheus and S. Bertrand supplied the serotyping and MLVA data. The experiments were designed by V. Wuyts, N. H. C. Roosens, K. Marchal and S. C. J. De Keersmaecker. Writing of the paper was done by V. Wuyts with the critical input of S. C. J. De Keersmaecker. S. Denayer, N. H. C. Roosens, W. Mattheus, S. Bertrand, K. Marchal and K. Dierick reviewed and commented the paper.

Abstract

Introduction In April and May 2014, two suspected egg-related outbreaks of *Salmonella enterica* subsp. *enterica* serovar Enteritidis (*S.* Enteritidis) were investigated by the Belgian National Reference Laboratory of Foodborne Outbreaks. Both the suspected food and human isolates being available, and this for both outbreaks, made these the ideal case study for a retrospective whole genome sequencing (WGS) analysis with the goal to investigate the feasibility of this technology for outbreak investigation by a National Reference Laboratory or National Reference Centre without thorough bioinformatics expertise.

Methods The two outbreaks were originally investigated epidemiologically with a standard questionnaire and with serotyping, phage typing, multiple-locus variable-number of tandem repeats analysis (MLVA) and antimicrobial susceptibility testing as classical microbiological methods. Retrospectively, WGS of six outbreak isolates was done on an Illumina HiSeq. Analysis of the WGS data was performed with currently available, user-friendly software and tools, namely CLC Genomics Workbench, the tools available on the server of the Center for Genomic Epidemiology and BLAST Ring Image Generator (BRIG).

Results To all collected human and food outbreak isolates, classical microbiological investigation assigned phage type PT4 (variant phage type PT4a for one human isolate) and MLVA profile 3-10-5-4-1, both of which are common for human isolates in Belgium. The WGS analysis confirmed the link between food and human isolates for each of the outbreaks and clearly discriminated between the two outbreaks occurring in a same time period, thereby suggesting a non-common source of contamination. Also, an additional plasmid carrying an

antibiotic resistance gene was discovered in the human isolate with the variant phage type PT4a.

Discussion For the two investigated outbreaks occurring at geographically separated locations, the gold standard classical microbiological subtyping methods were not sufficiently discriminative to distinguish between or assign a common origin of contamination for the two outbreaks, while WGS analysis could do so. This case study demonstrated the added value of WGS for outbreak investigations by confirming and/or discriminating food and human isolates between and within outbreaks. It also proved the feasibility of WGS as complementary or even future replacing (sub)typing method for the average routine laboratory.

6.1 Introduction

For 71% of the total number of outbreaks reported within the European Union the causative agent is known. The most frequently reported causative agent of foodborne outbreaks remains *Salmonella*, with *Salmonella enterica* subsp. *enterica* serovar Enteritidis (*S.* Enteritidis) as predominant serovar. Egg and egg-related products are still the most common source (60%) of *S.* Enteritidis outbreaks in Europe (European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) 2015); this despite the implementation of the national control programme on reduction of *Salmonella* in commercial laying hens, with obligatory vaccination for countries with high incidence of *Salmonella* in laying hen flocks since 2008.

During outbreak investigations, in addition to collecting epidemiological information through interrogation of the human cases, bacterial isolates collected from food samples, leftovers and human cases (often via stool samples) are being characterised by the National Reference Laboratories for Food (NRL) and the Human Reference Centres (NRC) in order to find a common source of contamination to be able to control the outbreak as soon as possible. This allows to support a strong relatedness between the isolate from the human case and that from the suspected food, which can have important economic implications. It will also allow to identify other human cases linked to the outbreak, *i.e.* which consumed the same contaminated food. This is especially important for outbreaks with a dispersed geographical distribution of human cases. Several methods, including molecular ones, can be used for characterisation, or subtyping, of a pathogenic isolate. For *S.* Enteritidis, this concerns for example phage typing, multi-locus sequence typing (MLST) and multiple-locus variable-number of tandem repeats analysis (MLVA) (Wattiau *et al.* 2011; Barco *et al.* 2013; Sabat *et al.* 2013). However, in addition to some other disadvantages (Wattiau *et al.* 2011), the resolution of these methods is not always sufficient to discriminate the outbreak isolates from the circulating background strains, especially when it concerns isolates belonging to the most frequently occurring subtypes.

Recently, whole genome sequencing (WGS) has been postulated as the universal, ultimate resolution subtyping technique (Sabat *et al.* 2013; Dunn 2015; Gilchrist *et al.* 2015). However, its data analysis requires appropriate tools, often involving the necessary bioinformatics expertise which is not always present in the average routine laboratory. In general, for bacterial WGS analysis, two workflows are proposed in literature (European Food Safety Authority (EFSA) 2014), *i.e.* allele-based (comparison of allelic variants) or single nucleotide polymorphism (SNP)-based (SNP calling). For an allele-based WGS analysis, which is also called gene-by-gene comparison or whole genome or core genome MLST (cgMLST), depending on the number of genes included, a preferably international database

with a MLST scheme and already known alleles is required so that a sequence type (ST) can be assigned and isolates can be compared (Jolley *et al.* 2012). Such internationally accepted scheme and database on a whole genome scale is not yet available for all pathogens. For SNP-based WGS analysis, numerous software packages are available, but most, if not all, of these have no graphical interface and are run in a command-line environment, which is not feasible for the average routine laboratory. Retrospective WGS analysis of identified outbreaks may contribute to the development of adequate WGS data analysis pipelines for future outbreak detection (Ashton *et al.* 2015).
Here we report on the retrospective WGS analysis of two *S.* Enteritidis outbreaks that were taken as a case study to demonstrate the added value of WGS for outbreak investigation and to evaluate its feasibility for an average NRL or NRC. The two outbreaks were selected since for both the food and human isolates were available. Additionally, it concerned two geographically separated outbreaks, but occurring around the same time, of *S.* Enteritidis PT4 that were linked to non-commercial eggs from privately kept laying hens which were used to prepare desserts for social events. This would allow investigating the possibility of using WGS to distinguish between and within outbreaks or to confirm a common source of contamination. The WGS analysis was performed with user-friendly software and tools to demonstrate the feasibility of characterisation of outbreaks isolates by non-bioinformaticians, thereby facilitating its implementation in a routine NRL.

6.2 Materials and methods

6.2.1 Epidemiological investigation

In Belgium, the Federal Agency for the Safety of the Food Chain (FASFC) is responsible for sampling and investigation of food, while the Health Services of the Belgian Communities collect human samples. Epidemiological information such as age, symptoms, timeline and circumstances of the outbreaks were gathered by local health inspectors and by inspectors of the FASFC using a standard questionnaire. All collected information was transmitted to the Belgian National Reference Laboratory of Foodborne Outbreaks (NRL-FBO). For each of the two outbreaks, a case was defined as an individual who consumed a meal at the respective social event and who suffered from diarrhoea.

6.2.2 Microbiological investigation

The NRL-FBO received food samples, including leftovers, of both outbreaks for detection of *Salmonella*, which was performed according to ISO 6579:2002 (ISO 2012). The Belgian National Reference Centre for *Salmonella* and *Shigella* (NRCSS) received *Salmonella* isolates from human cases of both outbreaks.
Isolates of both outbreaks were serotyped (Grimont and Weill 2007) by the NRCSS and phage typed by Public Health England. MLVA (Hopkins *et al.* 2011) was performed by the Belgian NRCSS.
The antimicrobial susceptibility of the *Salmonella* isolates was tested by determination of the minimal inhibitory concentration (MIC) of 14 antimicrobials in a Sensititre MIC plate EUVSEC with read-out on a Sensititre Vizion system. Following epidemiological cut-off values were applied: ampicillin 8 mg/l, cefotaxime 0.5 mg/l, ceftazidime 2 mg/l, chloramphenicol 16 mg/l, ciprofloxacin 0.064 mg/l, colistin 2 mg/l, gentamicin 2 mg/l, meropenem 0.125 mg/l, nalidixic acid 16 mg/l, tetracycline 8 mg/l, tigecycline 1 mg/l and trimethoprim 2 mg/l (European Commission 2013; European Committee on Antimicrobial Susceptibility Testing (EUCAST) 2015b). Epidemiological cut-off values were not available for azithromycin and sulphamethoxazole.

6.2.3 Whole genome sequencing

Genomic DNA of the outbreak isolates (Table 6.1) was extracted with the Qiagen Genomic-tip 100/G kit. The samples were sequenced at the EMBL GeneCore facility in 40-plex on a single lane of the Illumina HiSeq 2000 using 100 bp paired-end reads. FASTQ reads from all sequences were deposited at the WIV-ISP - *Salmonella* BioProject at NCBI (PRJNA289069).

6.2.4 WGS data analysis

All analyses were performed on a Windows 7 platform.
In CLC Genomics Workbench 8.0 the raw FASTQ reads were first trimmed to quality score limit 0.001 (Q30) with maximum 2 ambiguous nucleotides and reads with length below 15 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 200 nucleotides.
On the server of the Center for Genomic Epidemiology (Technical University of Denmark (DTU) 2015), the resulting contigs were uploaded to MLST 1.7 (Larsen *et al.* 2012) with *Salmonella enterica* as MLST scheme, ResFinder 2.1

(Zankari *et al.* 2012) and PlasmidFinder 1.2 (Carattoli *et al.* 2014). ResFinder was used to find all available antimicrobial resistance genes with minimum 98% identity and minimum 60% of their length. In PlasmidFinder, the database of Enterobacteriaceae was searched with an identity threshold of 95%. Additionally, raw FASTQ reads were uploaded to the CSI Phylogeny 1.0a (Kaas *et al.* 2014) server on which the SNP calling was run with *S.* Enteritidis P125109 (NC_011294) as reference genome, default input parameters as described by Kaas *et al.* (2014) and a minimum Z-score of 1.96. The downloaded Newick file was used for visualisation of the phylogenetic tree in FigTree v1.4.2 (Rambaut 2014). The downloaded vcf (variant call format) files were used for investigation of the position of the SNPs on the chromosome of *S.* Enteritidis P125109.
As PlasmidFinder results indicated that plasmids were present, trimmed reads were mapped to *S.* Enteritidis P125109 (NC_011294) in CLC Genomics Workbench 8.0 with default settings and unmapped reads were *de novo* assembled as described above. The resulting contigs were blasted to plasmids pSLA5 (NC_019002) and pSD107 (JX566770) and visualised as concentric rings with BLAST Ring Image Generator (BRIG) (Alikhan *et al.* 2011).

6.3 Results

The first outbreak occurred in Flanders at a social event with about 220 guests and where food was supplied by a catering service. The onset of the first symptoms was on April 23rd 2014. The outbreak extended to 45 cases with 5 hospitalised individuals and was reported to the Flemish Agency for Care and Health by pharmacists and general practitioners, who treated an unusually high number of people for diarrhoea. Since people who only ate dessert at the event also became ill, freshly prepared chocolate mousse and ice cream were indicated as possible cause of the outbreak. Due to shortage of commercial eggs, also non-commercial eggs from privately kept laying hens were used to prepare the chocolate mousse. Two samples from white and brown chocolate mousse were sent to the NRL-FBO and were tested positive for *Salmonella*. Raw eggs from privately kept laying hens, collected on April 24th and 25th, were also investigated, but these tested negative for *Salmonella*. The NRCSS received 11 *Salmonella* isolates from different human cases linked to this outbreak, which were isolated from stool in clinical laboratories. Two of these isolates were randomly selected for phage typing and WGS.
The second outbreak occurred in Wallonia on a social event which was attended by about 300 people and where a barbecue meal was prepared by volunteers. The onset of the first symptoms was on May 1st 2014. The number of cases was estimated at 40 and some people were hospitalised, but no number is available for the hospitalised cases. The hospital reported the outbreak to the

Walloon-Brussels Health Inspection Service. The NRL-FBO received samples of mascarpone cheese, bacon, sausages, pork ribs and raw eggs (which were stored refrigerated). There were no left-overs of consumed tiramisu, which was prepared with the sampled mascarpone cheese and the sampled non-commercial eggs from privately kept laying hens. Chocolate mousse, prepared with commercial eggs, was also consumed, but no leftovers were available. The raw eggs from privately kept laying hens tested positive for *Salmonella*. One *Salmonella* isolate of a human case, which was isolated from stool in a clinical laboratory, was sent to the NRCSS.

All 15 outbreak isolates were serotyped as *S.* Enteritidis. Six collected food and human isolates related to the outbreak in Flanders and Wallonia (Table 6.1) were selected for sequencing and subtyping with available user-friendly software and tools. All 6 selected isolates (Table 6.1) were phage typed as PT4, except for isolate S14BD01672 which showed a PT4a phage type. MLVA resulted for all isolates in profile 3-10-5-4-1 (SENTR4-SENTR5-SENTR6-SENTR7-SE3).

The *de novo* assemblies consisted of 271 contigs on average (range 35–480) with an average N50 of 174624 (range 28773–405843). The MLST server (Larsen *et al.* 2012) typed all isolates as ST-11 (alleles: *aroC*-5/*dnaN*-2/*hemD*-3/*hisD*-7/*purE*-6/*sucA*-6/*thrA*-11). No resistance genes were found by ResFinder (Zankari *et al.* 2012), with the exception of isolate S14BD01672, for which a perfect match to $bla_{\text{TEM-1B}}$ (JF910132) was detected. These *in silico* results were phenotypically confirmed by the antimicrobial susceptibility tests, as for all outbreak isolates only a single resistance to colistin was observed, except for isolate S14BD01672 which had an additional resistance to ampicillin. As colistin resistance is often linked to mutations (Blair *et al.* 2015), this will not be recognised by ResFinder, which only identifies resistance genes.

Results of PlasmidFinder (Carattoli *et al.* 2014) pointed to the presence of plasmid pSLA5 in all 6 outbreak isolates and additionally to pSD107 in isolate S14BD01672. BRIG analyses are shown in Figures 6.1 and 6.2. Absence of the region *srgB*-SELA5_RS23145-SELA5_p0022 on the pSLA5 plasmid of the outbreak isolates is most likely an artefact, because this region is also present on the chromosome of *S.* Enteritidis, so that reads mapping to this region on the chromosome are absent in the *de novo* assemblies of unmapped reads used for the BRIG analysis. This is also indicated by the coverage of this region in the read mapping to *S.* Enteritidis P125109, which is about 2 to 4 times higher than the average coverage of the read mapping to the chromosome of P125109.

The analysis with CSI Phylogeny (Kaas *et al.* 2014) showed that there was a pairwise distance of 0 to 2 SNPs within the Flemish outbreak and more specifically, no SNPs between the food isolates (S14FP01640 and S14FP01642), 2 SNPs between the human isolates (S14BD01605 and S14BD01672) and 1 SNP between each of the food and each of the human isolates. Within the Walloon outbreak, no SNPs were observed between the food (S14FP01877) and human

Table 6.1: Overview of outbreak isolates selected for sequencing with the microbiological investigation results.

Outbreak	Isolate	Origin	Phage type	MLVA	Antimicrobial resistance
Flanders	S14FP01640	Chocolate mousse	PT4	3-10-5-4-1	Colistin
Flanders	S14FP01642	Chocolate mousse	PT4	3-10-5-4-1	Colistin
Flanders	S14BD01605	Human	PT4	3-10-5-4-1	Colistin
Flanders	S14BD01672	Human	PT4a	3-10-5-4-1	Colistin - ampicillin
Wallonia	S14FP01877	Raw egg (non-commercial)	PT4	3-10-5-4-1	Colistin
Wallonia	S14BD01753	Human	PT4	3-10-5-4-1	Colistin

MLVA: multiple-locus variable-number of tandem repeats analysis.

isolate (S14BD01753). Fifty-one to 53 SNPs were detected between the two outbreaks, which were distributed around the chromosome of *S.* Enteritidis P125109. Forty-five to 47 and 44 SNPs were observed between the reference P125109 and outbreak isolates of, respectively Flanders and Wallonia. A phylogenetic tree is presented in Figure 6.3, clearly linking the human and food isolates within each outbreak and distinguishing between both outbreaks.

6.4 Discussion

S. Enteritidis remains linked to egg-related outbreaks, albeit in the described outbreaks to non-commercial eggs from privately owned laying hens. In response to two geographically separated outbreaks occurring in the same time period in Belgium, the NRL-FBO received food samples from which *Salmonella* was isolated and the NRCSS received *Salmonella* isolates from human cases of these outbreaks. With the traditional epidemiological and microbiological investigations, *i.e.* phage typing and MLVA, the isolates of both outbreaks were classified as PT4 and variant PT4a, and as profile 3-10-5-4-1. Phage type PT4 and MLVA profile 3-10-5-4-1 are frequently observed for human isolates in Belgium. Therefore, the gold standard subtyping methods might not have

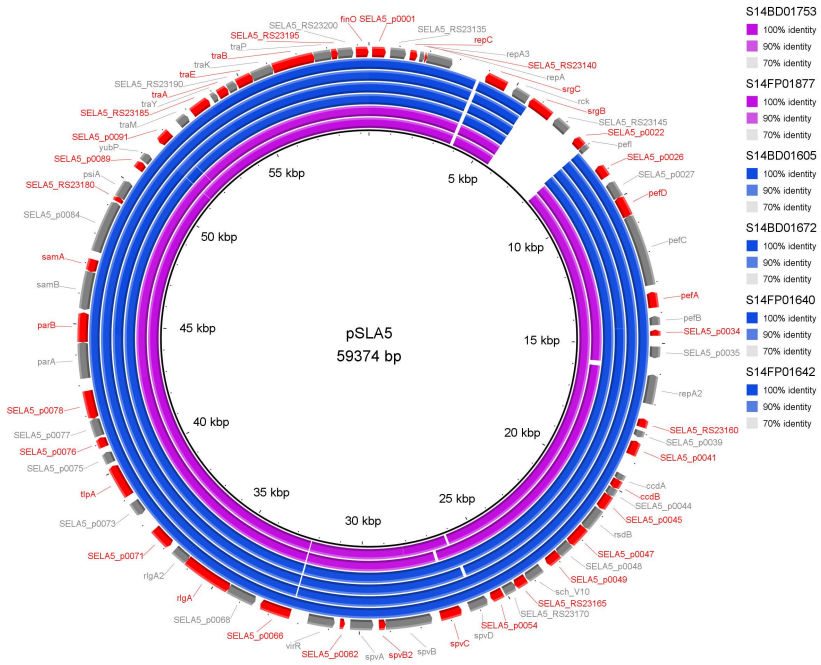

Figure 6.1: *De novo* assemblies of reads of the outbreak isolates that did not map to *S.* Enteritidis P125109 are shown as concentric rings with plasmid pSLA5 as reference on the inner black circle. Absence of colour in a ring indicates absence of the region. Isolates of the outbreak in Flanders are represented by a blue colour, those of the Walloon outbreak by a purple colour.

been sufficiently discriminative to establish or exclude a common source of contamination for both outbreaks, which now could only be distinguished by their separate geographical location. However, there could still be a common origin, *e.g.* a common breeding flock of the laying hens bought by the private persons.

As WGS has been postulated as a universal subtyping method with ultimate resolution, the two outbreaks were retrospectively examined with WGS. Moreover, they were a good case study to investigate the feasibility of using WGS for outbreak investigation by a National Reference Laboratory, as both human and food isolates were available. An additional aspect that was evaluated in this study concerns the WGS data analysis tools, *i.e.* to see whether these are not restricted to expert bioinformaticians which are often not available in an average routine laboratory.

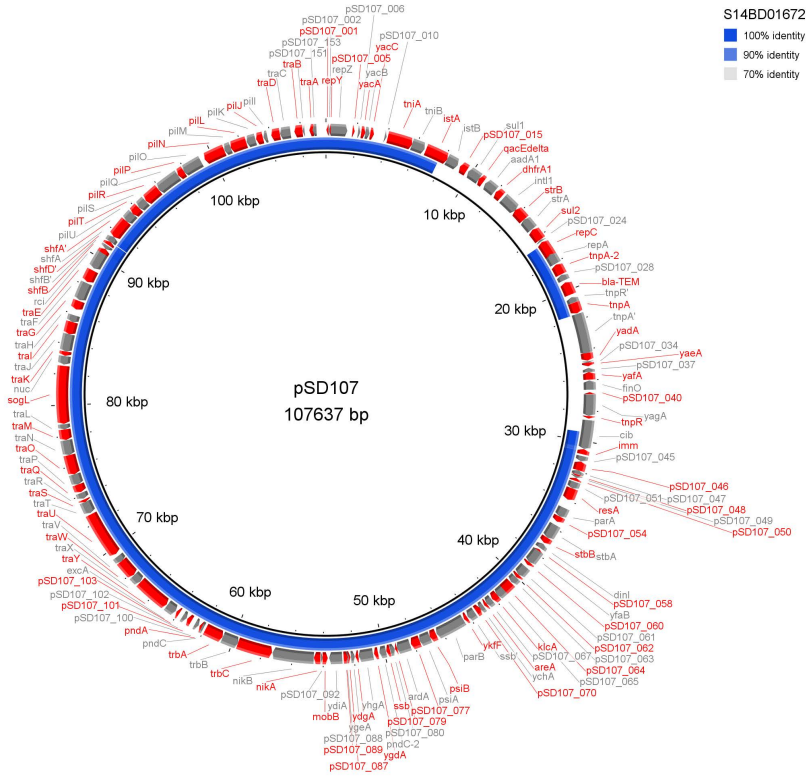

Figure 6.2: The *de novo* assembly of *S.* Enteritidis P125109 unmapped reads of outbreak isolate S14BD01672 is shown as a concentric ring with plasmid pSD107 as reference on the inner black circle. Absence of colour in the ring indicates absence of the region.

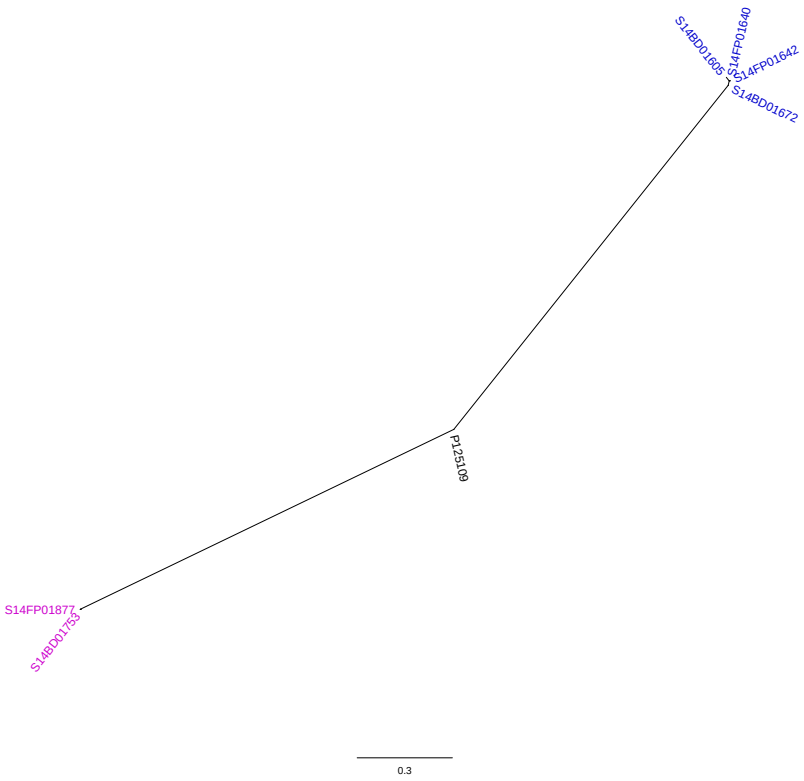

Figure 6.3: Radial phylogenetic tree of the 6 outbreak isolates with *S.* Enteritidis P125109 as reference. Isolates of the outbreak in Flanders are represented by a blue colour, those of the Walloon outbreak by a purple colour.

Therefore, the WGS data were analysed with user-friendly, albeit commercial, software (CLC Genomics Workbench) and the tools publicly available on the server of the Center for Genomic Epidemiology (Technical University of Denmark (DTU) 2015), which not only allows for SNP analysis (Kaas *et al.* 2014), but also, amongst others, to explore the resistome (ResFinder (Zankari *et al.* 2012)), to search for plasmids (PlasmidFinder (Carattoli *et al.* 2014)) and to assign a classical MLST (Larsen *et al.* 2012) sequence type based on 7 housekeeping genes. As mentioned above, an internationally accepted scheme and database on a whole genome scale for the allele-based data analysis workflow is not yet available for *Salmonella*, although commercially driven development efforts are ongoing, so that this type of analysis could not be performed in this study. Once the cgMLST for *S.* Enteritidis will be available, it would be interesting

to reanalyse our dataset and to compare the results to those obtained with the SNP-based analysis. This will allow evaluating the impact of the selected type of data analysis workflow on the efficiency and accuracy of the outbreak investigation.

Similar to the traditional outbreak investigations, SNP analysis of the WGS data confirmed the association of food and human isolates in both outbreaks thereby proving the link between the contaminated eggs and the human cases who consumed these eggs. Moreover, and this in contrast to the gold standard subtyping methods, the SNP analysis was sufficiently discriminative to reveal a clear difference between the two outbreaks, *i.e.* the food isolates of the two outbreaks were not closely related. This clearly illustrates the utility of WGS and SNP analysis for a first indication of the source in the investigation of outbreaks. In our study, we observed about 52 SNPs difference between the two outbreaks, while only 0 to 2 SNPs difference within each outbreak. However, as previously suggested by Ashton *et al.* (2015), it is difficult, if not impossible to set a single diversity threshold within a certain *Salmonella* outbreak, as it would depend on the size of the population that caused the outbreak, and hence it depends on the size of the facilities at the origin of the outbreak. As no sampling was done at the original sources of the outbreaks, namely the laying hens and their environment, or even at the distributor of these laying hens, the genetic diversity of the source population in the outbreaks described in this study could not be investigated. This sampling at the source, and an epidemiological investigation of this source, may be important for future outbreak investigations with WGS as the diversity of the source population may give an indication of the expected genetic diversity within outbreak isolates (Ashton *et al.* 2015). As such, more studies are still needed to contribute to the validation of SNP detection pipelines for this purpose.

The described analysis also shows that examination of mobile elements as plasmids can be useful for fine-tuning the results of a SNP analysis. One human outbreak isolate had a deviating phage type PT4a, which may be explained by presence of an additional plasmid carrying an antibiotic resistance gene (E. de Pinna, personal communication). Indeed, this mobile element also harbours a bla_{TEM} gene, which explains the ampicillin resistance observed phenotypically only in this isolate. As this mobile element was found in a human isolate, a possible hypothesis is that it might have been acquired during the foodborne infection. This could be further studied by analysis of multiple isolates from the same human case, which would also be interesting to examine the possible microevolution of a strain within a host.

The complete WGS analysis in this case study was performed on a Windows platform with currently available user-friendly software and tools which proves that WGS data analysis is not strictly restricted to bioinformaticians. As the use of WGS for characterisation of pathogens will only increase in the

future, studies like these that are creating benchmarking datasets, may lead, in collaboration with hard core bioinformaticians, to the further development of user-friendly pipelines. This would imply that routine laboratories will no longer be solely dependent on bioinformaticians for WGS analyses and that WGS could be applied in real-time for diagnosis and outbreak investigations, if the infrastructure to generate the data in a short period of time is accessible to the routine laboratory.

This case study clearly demonstrated the added value of WGS as a complementary subtyping method during outbreak investigation, for isolates belonging to common circulating subtypes. The fact that the data analysis was done with user-friendly tools illustrates the feasibility of this technology for an average Reference Laboratory or Centre where bioinformatics expertise might be scarce. With the decreasing sequencing costs, WGS might become a replacing subtyping method, also in these environments. In this context, implementation of WGS in the average Reference Laboratories and Centres as routine characterisation method for *Salmonella* and other pathogens for surveillance and outbreak detection and investigation will benefit of sequencing of more outbreak and background isolates to create a database of circulating strains so that the diversity in the background population can be estimated and new outbreak isolates can be better discriminated. As outbreaks are not stopped by country borders, a European and/or international collaboration to set up such real-time WGS database would certainly be invaluable for future *Salmonella* outbreak detection and investigation.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgments

This study was made possible by the NGS service of the EMBL GeneCore for which the authors would like to gratefully thank Dr. Vladimir Benes. The authors wish to thank the Federal Agency for the Safety of the Food Chain (FASFC) and the Flemish and French speaking Communities for the excellent collaboration.

Funding statement

This study was supported by grant P4044.0103 (SalMolType) from the Scientific Institute of Public Health (WIV-ISP - RP/PJ). The NRL-FBO and NRL *Salmonella*-Food are co-funded by the Federal Agency for the Safety of the Food Chain and the Federal Public Service Health, Food Chain Safety and Environment. The National Reference Centre for *Salmonella* and *Shigella* is partially supported by the Belgian Ministry of Social Affairs through a fund within the Health Insurance System.
The funders had no role in study design, data collection, analysis and interpretation, decision to publish or preparation of the manuscript.

Chapter 7

General conclusions and perspectives

Subtyping or characterisation of pathogens below the subspecies or serovar level, is essential for routine surveillance, outbreak detection and outbreak investigation. One of such important pathogens is *Salmonella*, which was taken as a case study in this PhD research. The classical subtyping methods for *Salmonella*, of which phage typing, multiple-locus variable-number of tandem repeats analysis (MLVA) and pulsed-field gel electrophoresis (PFGE) are the most important, all have their intrinsic drawbacks. In the first and main part of this PhD an alternative, molecular method was developed for subtyping of *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium), a major cause of foodborne infections. Therefore, the multiplex oligonucleotide ligation-PCR (MOL-PCR) technique with read-out on a Luminex device was applied. The second part of this PhD focussed on the ultimate universal subtyping of pathogens with whole genome sequencing (WGS), which emerged and gained attention by public health institutes during the course of this PhD.

During the analysis of available *S.* Typhimurium subtyping data of the Belgian National Reference Centre for *Salmonella* and *Shigella* (chapter 2), a solution was provided for interpretation of the numerous distinct MLVA profiles that were generated, so that relevant information for routine surveillance and outbreak detection and investigation could be extracted. This was necessary, as MLVA became a gold standard subtyping method for bacterial pathogens in Europe. By a serial passage experiment, the high number of different MLVA profiles was explained by the instability of 3 of the 5 loci used in the MLVA scheme of *S.* Typhimurium. Later, these findings were confirmed by independent *in vitro*

and *in vivo* experiments of Dimovski *et al.* (2014), who concluded that a cluster algorithm for outbreak investigations should allow inclusion of isolates with variations in one of the 3 unstable loci into the same cluster. This highlighted the need for an alternative molecular subtyping method. A comparison of different possible multiplex assay techniques showed that MOL-PCR was the most suitable assay technique for a multiplex molecular subtyping method intended to be used in public health laboratories.

Although the Luminex technology is already used in different research institutes, this is not the case for public health institutes. This PhD work, including the development of a novel MOL-PCR assay for subtyping of *S.* Typhimurium and its monophasic variant *S.* 1,4,[5],12:i:-, allowed to introduce the Luminex technology in the Scientific Institute of Public Health (WIV-ISP) in Belgium. This technology allows multiplexing with high-throughput microsphere suspension arrays and has a broad range of genomic and proteomic applications. As Luminex assays provide a time- and cost-effective alternative for many singleplex assays, this technology is now also applied in other ongoing research projects within the WIV-ISP. As such, Luminex assays are being developed (or are already being used) for characterisation of shiga toxin-producing *Escherichia coli*, for profiling of indoor airborne fungi, for identification of viruses in human samples, for examining antibiotic resistance genes and mutations in bacteria and for detection of genetically modified organisms in the food and feed chain. Also commercial Luminex assay kits such as those for measuring human cytokine levels have found their way into the WIV-ISP. As such, this PhD gave access to many new technological research applications of Luminex in public health.

When searching the literature on the MOL-PCR assay, not many papers were retrieved that described the impact of changing several parameters in the assay. Therefore, the findings from the optimisation of the MOL-PCR assay for subtyping of *S.* Typhimurium and *S.* 1,4,[5],12:i:- were collected in a manuscript (chapter 3), as these may guide other scientists in the development of their own MOL-PCR assay. This optimisation study indicated that the parameters that have a major influence on the MOL-PCR performance are the DNA isolation, the probe concentration, the amount of microspheres and the concentration of reporter dye.

The complete MOL-PCR assay (chapter 4) consisted of 52 molecular markers, including prophage genes, amplified fragment length polymorphism (AFLP) elements, *Salmonella* genomic island 1 (SGI1), allantoinase gene *allB*, MLVA locus STTR10, antibiotic resistance genes, single nucleotide polymorphisms (SNPs) and phase 2 flagellar gene *fljB*. The MOL-PCR assay showed to have a discriminatory power similar to that of phage typing, thereby reaching the initial goal set. The latter phenotyping method is now no longer being used in the National Reference Centre for *Salmonella* and *Shigella*. With the developed

assay, the time needed to get from a single colony to a MOL-PCR profile is less than 8 hours, which makes the assay suitable for outbreak investigations. Moreover, the data analysis with an R application allows for an objective interpretation of the results. Another significant aspect for a routine subtyping method is the cost. The cost of reagents and consumables for typing 1 isolate is less than 10 euro and the cost of a Luminex MAGPIX instrument is comparable to that of two PCR machines, which is affordable for most public health laboratories. The MOL-PCR assay was also tested on a non-Belgian collection of 89 *S.* Typhimurium isolates, which resulted in 35 distinct MOL-PCR profiles, of which 7 were already observed in Belgian isolates (preliminary data of an ongoing collaboration with S. Le Hello, Pasteur Institute, France). This gave a Simpson's index of diversity of 0.925 for the collection of 89 isolates, which illustrates that Simpson's index of diversity is heavily dependent on the test population. A more objective metric for interpretation of the discriminatory power of a subtyping method may thus be required, especially in view of implementation of WGS as subtyping standard. This study also illustrates that the application of the developed MOL-PCR assay, which is based on assigning a profile to an isolate, is not limited to Belgian *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates. Interlaboratory studies still have to be done. Nonetheless, during this part of the PhD, the foundations have been laid for a validation dossier for demanding accreditation of the method under ISO 17025 or ISO 15189 norms. Introduction of the novel MOL-PCR assay in other laboratories will however be a difficult exercise, since nowadays WGS attracts all the attention. Nonetheless, WGS for subtyping of pathogens is not yet within the reach of all European laboratories. It may also not be necessary to use WGS on all pathogenic isolates that are sent to a National Reference Centre for characterisation. In this case, the novel MOL-PCR assay may provide a relatively inexpensive and rapid, when compared to WGS, alternative subtyping method for *S.* Typhimurium and *S.* 1,4,[5],12:i:-. If an isolate is assigned a same MOL-PCR profile as those of the circulating background population, further analyses may be needed for additional discrimation. Hereto, PFGE was applied during the PhD, but in the future, such further analyses could also be done by WGS, as explored in the second part of this PhD.

Indeed, a selection of *S.* Typhimurium, *S.* 1,4,[5],12:i:- and *S.* Enteritidis isolates from the collection of the National Reference Centre for *Salmonella* and *Shigella* were sequenced at the EMBL GeneCore facility. This process pointed already to a practical issue for WGS as a subtyping technique for public health, since the long turn-around time of about 6 to 8 weeks for outsourcing of NGS is not always acceptable in a routine setting. Commercial NGS facilities also offer quicker turn-around times, often with less service provided, but at a substantial higher cost. A solution would be that public health institutes invest in an in-house sequencing facility, so that turn-around times and costs can be reduced.

Nonetheless, such an investment does not only include the sequencing and computing equipment, it involves also the know-how for producing WGS data of acceptable quality. In addition, a reduction of the sequencing cost can only be obtained when enough isolates are present, so that samples can be pooled in one run. For cost-effectiveness, the throughput should also be high enough to let the sequencing platform run at its maximum capacity, so that general maintenance costs can be spread over more samples. Such a cost-effectiveness may not always be feasible in a public health setting, because in case of a crisis, it may be needed to urgently sequence one or a few isolates, which would be expensive. In this respect, it may be beneficial for a Reference Centre to implement WGS for all pathogenic isolates in routine surveillance. However, at this moment pathogenic subtyping by WGS for routine surveillance is still too expensive for many National Reference Centres. Additionally, bioinformatics knowledge, substantial computing power and considerable data storage are not standard equipment for an average Reference Centre. Furthermore, as no international standard protocol is available, WGS for public health is still in an exploration phase and concentrated in a few big public health centres (*e.g.* UK, Denmark and FDA). Nevertheless, user-friendly, *i.e.* with a graphical user interface, software packages and web-based tools for WGS data analysis already exist, so that also non-bioinformaticians can run a WGS analysis. Many of these available user-friendly software packages and tools are however black box applications, which do not allow the user to intervene in the analysis, with the exception of some parameter values. Such tools can thus not be readily adapted for a special analysis.

The studies on *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates in chapter 5 and on *S.* Enteritidis isolates in chapter 6 show that the available software and tools allow now for improved outbreak investigations by WGS, which then mostly relies on SNP analysis. The phylogenetic tree that is obtained with a SNP workflow, which has to be recalculated each time a new isolate is added, may not be suitable for routine surveillance, as it is difficult to translate a phylogenetic tree into epidemiologically relevant information, such as a named subtype for interlaboratory comparison of results. In this respect, routine surveillance might benefit more from a WGS data analysis which is based on gene-by-gene comparison, such as whole genome multilocus sequence typing (MLST), so that a sequence type can be assigned to an isolate. Based on this sequence type, the isolate could then be added to the corresponding branch of an 'evergreen' tree which should never be recalculated. However, this should be evaluated in the future by re-analysing WGS data that are currently being analysed with a SNP-based approach, so that the results and their epidemiological interpretation can be compared between the two approaches.

The decision of which WGS data analysis workflow is more suitable for outbreak

investigations or for routine surveillance is essentially an evaluation of the required resolution, or discriminatory power, for each research question. As some pathogens are more clonal than others, this required resolution, and its corresponding WGS data analysis workflow, may be species-dependent. SNP calling based on read mapping has always a bias of the reference genome. In addition, which reference genome is most suitable, if there is one available? Moreover, quality filtering and pruning have a significant effect on the resolution of SNP calling and on the produced phylogenetic tree, as was seen in chapter 5. The bias of a reference genome can be avoided with reference-free SNP detection, a field that is under research (Leggett and MacLean 2014), but also with *de novo* assembly, which is required for whole genome MLST. The discriminatory power of whole genome MLST can be adapted by including more or less genes in the MLST scheme, so that *e.g.* only the core genome, the accessory genome or the complete pan-genome is taken into account. *De novo* assembly of WGS reads could also allow to detect other sequence variants, such as rearrangements, in order to improve the resolution of WGS data. This would *e.g.* be interesting for isolates originating from the same patient over time, but may require sequencing with other platforms than Illumina to get longer reads (*e.g.* PacBio).

WGS data may also provide the opportunity to strengthen the discriminatory power of the developed MOL-PCR assay. *E.g. S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates of phage types DT120 and DT193 were often assigned the same MOL-PCR profile and isolates with these phage types tend to cluster together in phylogenetic trees resulting from SNP-based WGS data analysis. For these phage types, it would thus be interesting to search for new discriminating molecular markers and to validate them for use in the MOL-PCR assay.

It is clear that research is still needed on WGS as subtyping technique for public health and for such research, a great time is coming. More and more sequencing data of clinical pathogenic isolates are uploaded to public databases. The NCBI BioProject PRJNA248792 of Public Health England currently holds the WGS reads of more than 8000 *Salmonella* isolates and through the FDA GenomeTrakr Network (U.S. Food and Drug Administration (FDA) 2015b), sequencing data of more than 12000 *Salmonella* isolates and more than 3000 *Listeria* isolates are available on NCBI. However, is this the way to go? WGS data have to be publicly available, but some metadata should be coupled to them. As metadata may be too confidential, databases like NCBI may not be suitable for this purpose. In addition, there is often a delay between deposit and availability of WGS data in NCBI-like databases, whereas in outbreak situations these data have to be available in real-time. This real-time availability is indispensable for finding rapidly a relatedness between clinical, food and environmental pathogenic isolates.

In this PhD we have only explored WGS for subtyping of *Salmonella*. However,

this should be broadened to other pathogens as well, which is a PhD project in itself. Projects like these, which are a collaboration between a public health institute and a university, are an ideal setting to exchange applied and fundamental knowledge. Additionally, it allows to apply the developed methods on clinically relevant isolates, which are available in the valuable collections of public health institutes.

To conclude, the aim of this PhD was to develop a molecular alternative for classical microbiological subtyping methods. More specifically, an alternative for the subjective *Salmonella* phage typing method was set as target. To this end, solutions were provided at several levels. Firstly, an analysis method was presented for MLVA, which is currently the gold standard for subtyping of *Salmonella* in Europe, and insight was gained into the instability of MLVA loci of *S.* Typhimurium. Secondly, a MOL-PCR assay was developed for subtyping of *S.* Typhimurium and *S.* 1,4,[5],12:i:-, which had the same discriminatory power as phage typing. Lastly, WGS, the ultimate universal subtyping method, was explored for routine surveillance and a proof of concept was delivered for outbreak investigations. While different issues still have to be resolved for the implementation of WGS in surveillance, WGS already has an added value for outbreak investigations. Besides confirmation of the source of an outbreak, SNP analysis of WGS data can discriminate between isolates related to different outbreaks and also between food isolates to determine the correct source of the contamination. As such, WGS can currently be used to complement existing subtyping methods, like MOL-PCR, for surveillance and outbreak investigations, *e.g.* if an isolate has to be discriminated from the circulating background population. In the future, WGS may replace all existing subtyping methods, but therefore further finetuning of the data analysis tools is required.

Appendix A

Supporting information for chapter 5

A.1 Supplementary figures

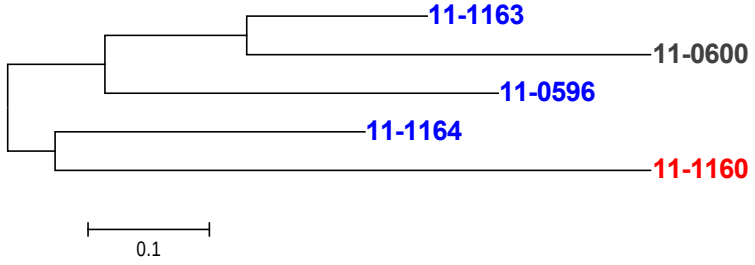

Figure A.1: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Blue corrected reads in the CLC workflow with reference genome LT2.

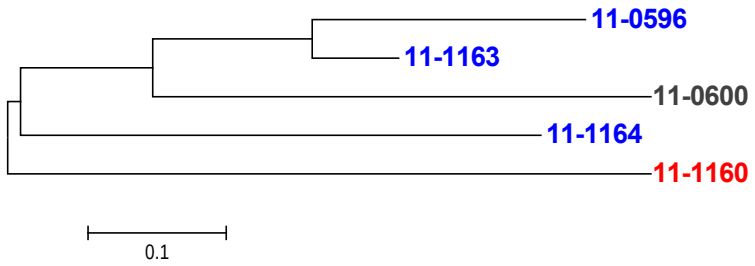

Figure A.2: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Blue corrected reads in the CLC workflow with reference genome SL1344.

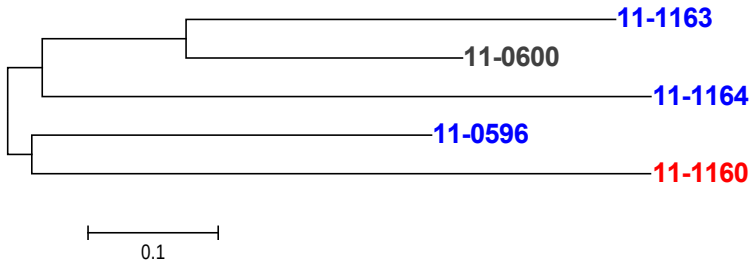

Figure A.3: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Brownie corrected reads in the CLC workflow with reference genome LT2.

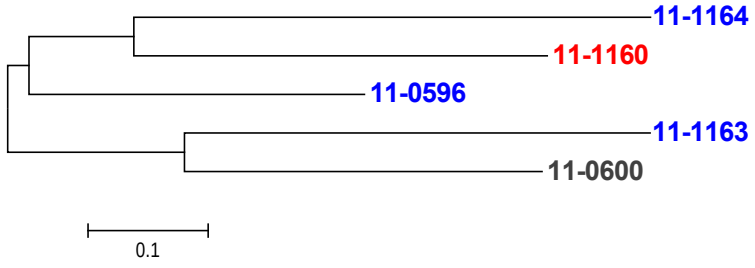

Figure A.4: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Brownie corrected reads in the CLC workflow with reference genome SL1344.

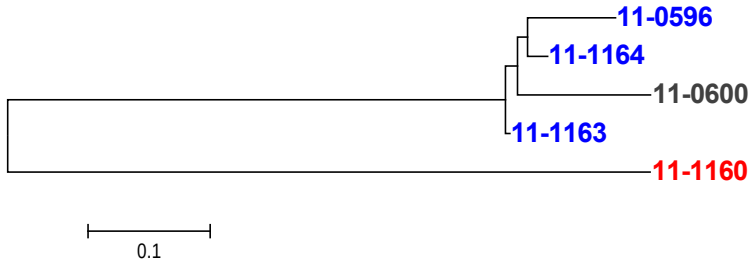

Figure A.5: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Blue corrected reads in the GATK workflow with reference genome LT2.

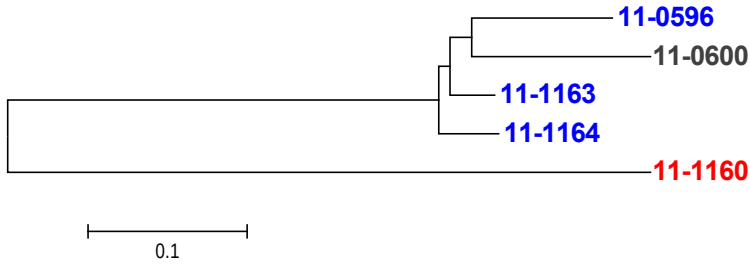

Figure A.6: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Blue corrected reads in the GATK workflow with reference genome SL1344.

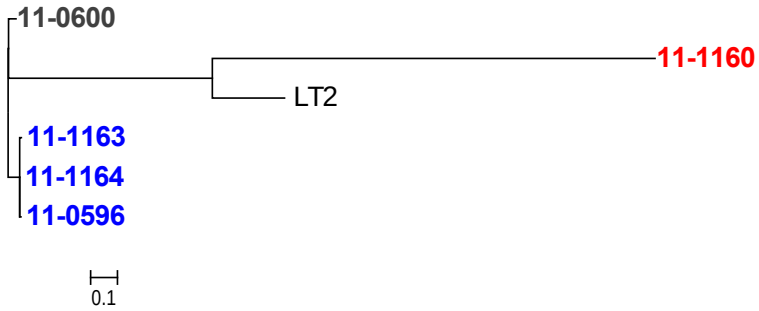

Figure A.7: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Blue corrected reads in the CSI workflow with reference genome LT2.

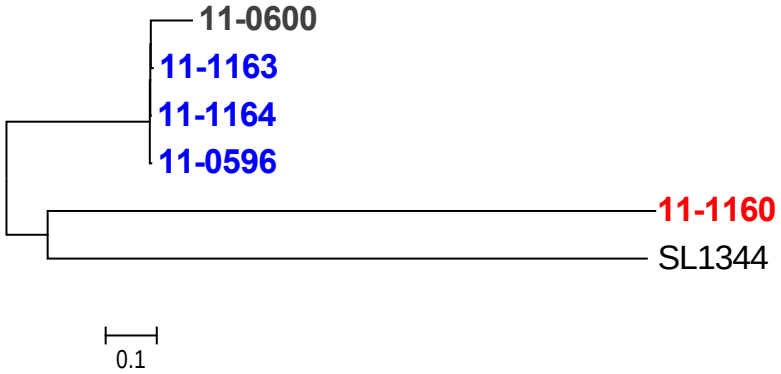

Figure A.8: Phylogenetic tree of outbreak isolates 11-0596, 11-1163 and 11-1164 (blue), and out-group isolates 11-0600 (dark grey) and 11-1160 (red) with Blue corrected reads in the CSI workflow with reference genome SL1344.

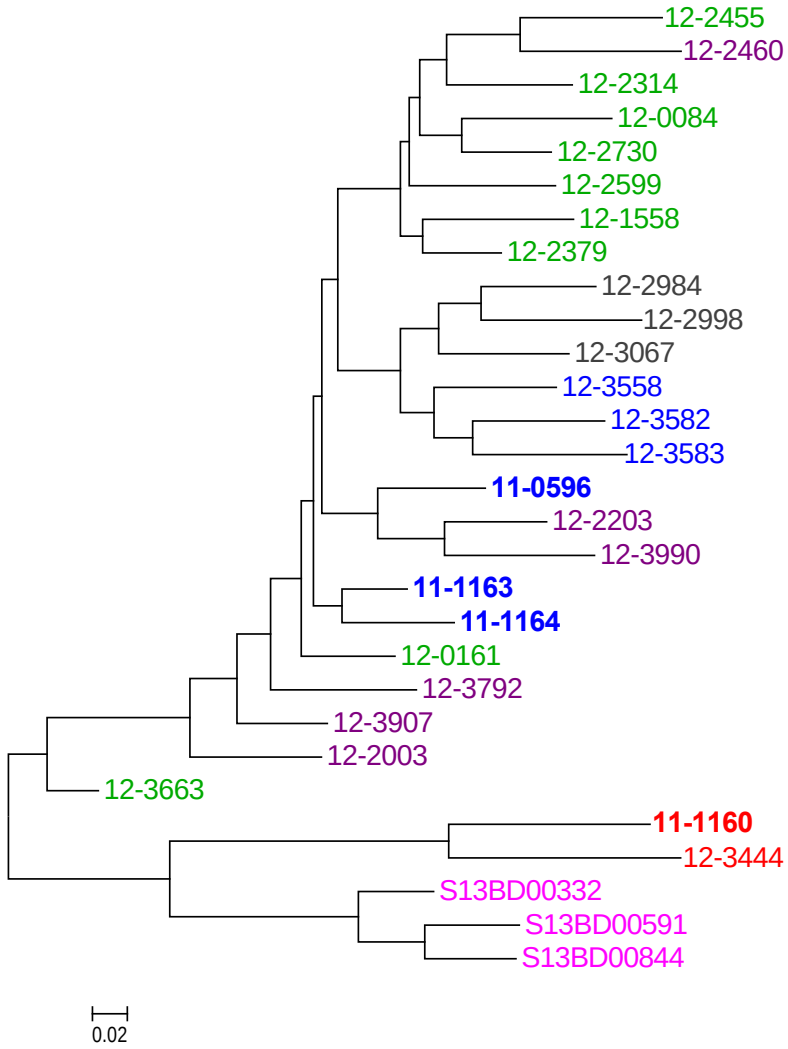

Figure A.9: Phylogenetic tree of 29 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with uncorrected reads in the GATK workflow with reference genome LT2. Isolates 11-0600, 11-1165 and 11-1166 were excluded from the SNP analysis. The labels are colour coded according to the phage type of the isolate, and outbreak and out-group isolates are indicated in boldface.

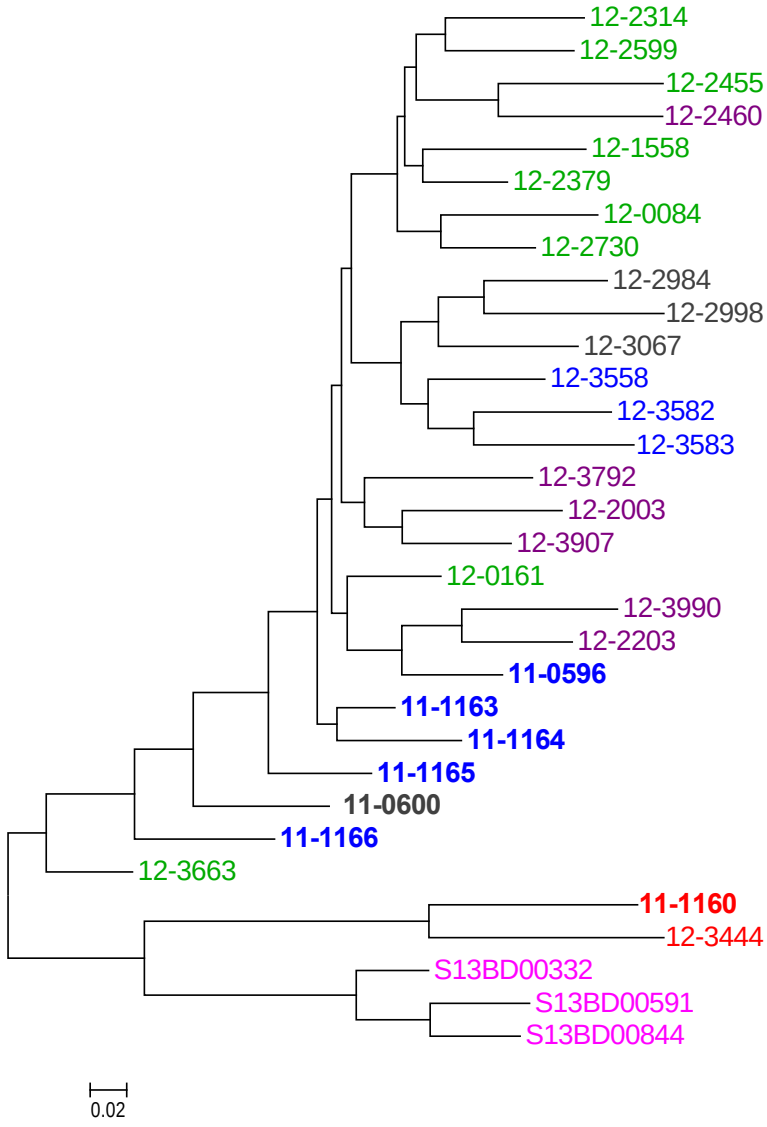

Figure A.10: Phylogenetic tree of all 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with uncorrected reads in the GATK workflow with reference genome LT2 and with additional pruning of SNPs within 10 bp of each other. The labels are colour coded according to the phage type of the isolate, and outbreak and out-group isolates are indicated in boldface.

A.2 Supplementary tables

Table A.1: Counts of SNPs between isolates and between isolates and reference genome LT2 for the CLC workflow with uncorrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	LT2
11-0596	0					
11-1163	164	0				
11-1164	164	124	0			
11-0600	236	190	176	0		
11-1160	947	929	973	957	0	
LT2	549	565	685	671	830	0

Table A.2: Counts of SNPs between isolates and between isolates and reference genome LT2 for the CLC workflow with Blue corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	LT2
11-0596	0					
11-1163	167	0				
11-1164	177	122	0			
11-0600	252	179	169	0		
11-1160	953	932	986	961	0	
LT2	553	582	698	677	830	0

Table A.3: Counts of SNPs between isolates and between isolates and reference genome LT2 for the CLC workflow with Brownie corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	LT2
11-0596	0					
11-1163	208	0				
11-1164	223	125	0			
11-0600	335	231	240	0		
11-1160	996	928	969	987	0	
LT2	595	567	682	702	833	0

Table A.4: Counts of SNPs between isolates and between isolates and reference genome SL1344 for the CLC workflow with uncorrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	SL1344
11-0596	0					
11-1163	522	0				
11-1164	392	384	0			
11-0600	562	473	410	0		
11-1160	1224	1222	1210	1316	0	
SL1344	1095	1033	1411	1280	1430	0

Table A.5: Counts of SNPs between isolates and between isolates and reference genome SL1344 for the CLC workflow with Blue corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	SL1344
11-0596	0					
11-1163	548	0				
11-1164	439	357	0			
11-0600	589	468	410	0		
11-1160	1253	1291	1238	1370	0	
SL1344	1104	1086	1435	1306	1458	0

Table A.6: Counts of SNPs between isolates and between isolates and reference genome SL1344 for the CLC workflow with Brownie corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	SL1344
11-0596	0					
11-1163	582	0				
11-1164	467	381	0			
11-0600	683	519	490	0		
11-1160	1293	1219	1208	1346	0	
SL1344	1134	1036	1409	1311	1430	0

Table A.7: Counts of SNPs between isolates and between isolates and reference genome LT2 for the GATK workflow with uncorrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	LT2
11-0596	0					
11-1163	189	0				
11-1164	232	171	0			
11-0600	318	267	302	0		
11-1160	741	724	751	699	0	
LT2	536	585	578	500	611	0

Table A.8: Counts of SNPs between isolates and between isolates and reference genome LT2 for the GATK workflow with Blue corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	LT2
11-0596	0					
11-1163	117	0				
11-1164	111	54	0			
11-0600	232	165	175	0		
11-1160	889	866	888	919	0	
LT2	599	638	668	661	796	0

Table A.9: Counts of SNPs between isolates and between isolates and reference genome SL1344 for the GATK workflow with uncorrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	SL1344
11-0596	0					
11-1163	405	0				
11-1164	437	362	0			
11-0600	572	467	573	0		
11-1160	1082	1023	1055	1052	0	
SL1344	997	1076	1170	913	1115	0

Table A.10: Counts of SNPs between isolates and between isolates and reference genome SL1344 for the GATK workflow with Blue corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	SL1344
11-0596	0					
11-1163	298	0				
11-1164	270	184	0			
11-0600	417	293	369	0		
11-1160	1190	1138	1156	1239	0	
SL1344	1081	1161	1317	1138	1359	0

Table A.11: Counts of SNPs between isolates and between isolates and reference genome LT2 for the CSI workflow with uncorrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	LT2
11-0596	0					
11-1163	8	0				
11-1164	2	6	0			
11-0600	53	55	51	0		
11-1160	595	593	593	592	0	
LT2	393	389	393	390	506	0

Table A.12: Counts of SNPs between isolates and between isolates and reference genome LT2 for the CSI workflow with Blue corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	LT2
11-0596	0					
11-1163	8	0				
11-1164	3	5	0			
11-0600	52	54	49	0		
11-1160	603	604	601	600	0	
LT2	392	394	391	390	514	0

Table A.13: Counts of SNPs between isolates and between isolates and reference genome SL1344 for the CSI workflow with uncorrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	SL1344
11-0596	0					
11-1163	11	0				
11-1164	4	11	0			
11-0600	80	85	80	0		
11-1160	612	611	612	634	0	
SL1344	621	614	619	635	725	0

Table A.14: Counts of SNPs between isolates and between isolates and reference genome SL1344 for the CSI workflow with Blue corrected reads.

Isolate	11-0596	11-1163	11-1164	11-0600	11-1160	SL1344
11-0596	0					
11-1163	7	0				
11-1164	3	4	0			
11-0600	75	76	72	0		
11-1160	620	620	618	640	0	
SL1344	619	618	616	632	738	0

Table A.15: Metrics of *de novo* assembly of 32 *S.* Typhimurium and *S.* 1,4,[5],12:i:- isolates with CLC Genomics Workbench.

Isolate	Number of contigs	N50	Total number of bases
11-0596	178	73993	4864778
11-1163	85	191828	4866913
11-1164	110	270252	4886320
11-1165	5960	108060	7002418
11-1166	30836	712	20765882
11-0600	2553	149463	5834905
11-1160	259	178471	5030475
S13BD00332	541	50911	4974356
S13BD00591	100	132639	4881450
S13BD00844	153	90686	4875817
12-2003	65	282662	4887137
12-2203	221	64263	4879998
12-2460	197	274692	4980141
12-2455	68	275867	4924384
12-2599	67	275287	4932740
12-2730	80	275053	4898886
12-1558	192	75544	4920258
12-2314	86	193547	4933059
12-2379	90	222811	4929363
12-3792	67	270379	4909689
12-3907	84	275281	4918736
12-3990	536	25628	4863328
12-0084	133	191662	4955497
12-0161	101	174514	4916251
12-3663	149	187516	4938281
12-3558	176	97308	4907832
12-3582	91	222811	4928898
12-3583	121	105445	4924831
12-2984	83	282662	4935670
12-2998	140	102709	4916756
12-3067	74	282569	4928020
12-3444	175	73816	4942552

Bibliography

Alikhan, N.-F., N. K. Petty, N. L. Ben Zakour, and S. A. Beatson (2011). "BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons". *BMC Genomics* 12, 402. DOI: 10.1186/1471-2164-12-402 (p. 135).

Anderson, E., L. R. Ward, M. J. de Saxe, and J. de Sa (1977). "Bacteriophage-typing designations of *Salmonella typhimurium*". *Journal of Hygiene* 78.2, pp. 297–300 (p. 6).

Angelo, K. M., A. Chu, M. Anand, T.-A. Nguyen, L. Bottichio, M. Wise, I. Williams, S. Seelman, R. Bell, M. Fatica, S. Lance, D. Baldwin, K. Shannon, H. Lee, E. Trees, E. Strain, and L. Gieraltowski (2015). "Outbreak of *Salmonella* Newport infections linked to cucumbers - United States, 2014". *MMWR Morbidity and mortality weekly report* 64.6, pp. 144–147 (p. 73).

Angeloni, S., R. Cordes, S. Dunbar, C. Garcia, G. Gibson, C. Martin, and V. Stone (2014). *xMAP Cookbook. A collection of methods and protocols for developing multiplex assays with xMAP Technology.* 2nd edition. URL: http://www.luminexcorp.com (pp. 16, 17, 51, 56, 65, 66).

Ashton, P. M., K. S. Baker, A. Gentle, D. J. Wooldridge, N. R. Thomson, T. J. Dallman, and C. Jenkins (2014). "Draft genome sequences of the type strains of *Shigella flexneri* held at Public Health England: comparison of classical phenotypic and novel molecular assays with whole genome sequence". *Gut Pathogens* 6.1, 7. DOI: 10.1186/1757-4749-6-7 (p. 97).

Ashton, P. M., T. Peters, L. Ameh, R. McAleer, S. Petrie, S. Nair, I. Muscat, E. de Pina, and T. Dallman (2015). "Whole genome sequencing for the retrospective investigation of an outbreak of *Salmonella* Typhimurium DT 8". *PLOS Currents Outbreaks.* DOI: 10.1371/currents.outbreaks.2c05a47d292f376afc5a6fcdd8a7a3b6 (pp. 19, 73, 97, 102, 103, 126, 133, 141).

Babraham Bioinformatics (2015). *FastQC.* URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (visited on 06/26/2015) (p. 99).

Barbau-Piednoir, E., S. Bertrand, J. Mahillon, N. H. Roosens, and N. Botteldoorn (2013). "SYBR®Green qPCR *Salmonella* detection system

allowing discrimination at the genus, species and subspecies levels". *Applied Microbiology and Biotechnology* 97, pp. 9811–9824. DOI: 10.1007/s00253-013-5234-x (pp. 78, 84).

Barco, L., F. Barrucci, J. E. Olsen, and A. Ricci (2013). "*Salmonella* source attribution based on microbial subtyping". *International Journal of Food Microbiology* 163.2–3, pp. 193–203. DOI: 10.1016/j.ijfoodmicro.2013.03.005 (p. 132).

Barua, H., I. L. Lindblom, M. Bisgaard, J. P. Christensen, R. H. Olsen, and H. Christensen (2013). "*In vitro* and *in vivo* investigation on genomic stability of *Salmonella enterica* Typhimurium DT41 obtained from broiler breeders in Denmark". *Veterinary Microbiology* 166.3–4, pp. 607–616. DOI: 10.1016/j.vetmic.2013.06.035 (pp. 46, 47).

Bergval, I. L., R. N. C. P. Vijzelaar, E. R. Dalla Costa, A. R. J. Schuitema, L. Oskam, A. L. Kritski, P. R. Klatser, and R. M. Anthony (2008). "Development of multiplex assay for rapid characterization of *Mycobacterium tuberculosis*". *Journal of Clinical Microbiology* 46.2, pp. 689–699. DOI: 10.1128/JCM.01821-07 (p. 51).

Bergval, I., S. Sengstake, N. Brankova, V. Levterova, E. Abadía, N. Tadumaze, N. Bablishvili, M. Akhalaia, K. Tuin, A. Schuitema, S. Panaiotov, E. Bachiyska, T. Kantardjiev, R. de Zwaan, A. Schürch, D. van Soolingen, A. van 't Hoog, F. Cobelens, R. Aspindzelashvili, C. Sola, P. Klatser, and R. Anthony (2012). "Combined species identification, genotyping, and drug resistance detection of *Mycobacterium tuberculosis* cultures by MLPA on a bead-based array". *PLoS ONE* 7.8, e43240. DOI: 10.1371/journal.pone.0043240 (pp. 51, 53, 60, 61, 65–67, 74).

Berning, L., S. W. Aberle, B. Simon, C. Luger, P. Apfalter, S. Machherndl-Spandl, and H. Kerschner (2014). "Evaluation of a multiplex ligation-dependent probe amplification assay for the detection of respiratory pathogens in oncological patients". *Journal of Clinical Virology* 60.2, pp. 141–146. DOI: 10.1016/j.jcv.2014.02.010 (pp. 51, 74).

Bertrand, S., R. Vanhoof, and W. Mattheus (2012). *Salmonella en Shigella stammen afgezonderd in België in 2011*. [Dutch]. D/2012/2505/30. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP) (p. 32).

Bertrand, S., R. Vanhoof, and W. Mattheus (2014). *Salmonella en Shigella stammen afgezonderd in België in 2013*. [Dutch]. D/2014/2505/13. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP) (pp. 3, 4, 76, 78, 99).

Beyene, D., I. Bergval, E. Hailu, S. Ashenafi, L. Yamuah, A. Aseffa, H. G. Wiker, H. Engers, P. Klatser, and L. Sviland (2009). "Identification and genotyping of the etiological agent of tuberculous lymphadenitis in Ethiopia". *Journal of Infection in Developing Countries* 3.6, pp. 412–419. DOI: 10.3855/jidc.411 (pp. 51, 74).

Blair, J. M. A., M. A. Webber, A. J. Baylay, D. O. Ogbolu, and L. J. V. Piddock (2015). "Molecular mechanisms of antibiotic resistance". *Nature Reviews Microbiology* 13.1, pp. 42–51. DOI: 10.1038/nrmicro3380 (p. 136).

Bolger, A. M., M. Lohse, and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data". *Bioinformatics* 30.15, pp. 2114–2120. DOI: 10.1093/bioinformatics/btu170 (p. 102).

Botteldoorn, N. (2015). *Recherche de Salmonella ISO 6579*. [French]. SOP 11/VM/114/F. Version 5. Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium. 8 pp. (p. 4).

Boxrud, D. (2010). "Advances in subtyping methods of foodborne disease pathogens". *Current Opinion in Biotechnology* 21.2, pp. 137–141. DOI: 10.1016/j.copbio.2010.02.011 (pp. 73, 91).

Boxrud, D., K. Pederson-Gulrud, J. Wotton, C. Medus, E. Lyszkowicz, J. Besser, and J. M. Bartkus (2007). "Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of *Salmonella enterica* serotype Enteritidis". *Journal of Clinical Microbiology* 45.2, pp. 536–543. DOI: 10.1128/JCM.01595-06 (pp. 40, 45, 46).

Boyd, D. A., G. A. Peters, L.-K. Ng, and M. R. Mulvey (2000). "Partial characterization of a genomic island associated with the multidrug resistance region of *Salmonella enterica* Typhymurium DT104". *FEMS Microbiology Letters* 189.2, pp. 285–291 (p. 83).

Boyd, D., A. Cloeckaert, E. Chaslus-Dancla, and M. R. Mulvey (2002). "Characterization of variant *Salmonella* genomic island 1 multidrug resistance regions from serovars Typhimurium DT104 and Agona". *Antimicrobial Agents and Chemotherapy* 46.6, pp. 1714–1722. DOI: 10.1128/AAC.46.6.1714-1722.2002 (pp. 83, 86).

Boyd, D., G. A. Peters, A. Cloeckaert, K. S. Boumedine, E. Chaslus-Dancla, H. Imberechts, and M. R. Mulvey (2001). "Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of *Salmonella enterica* serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona". *Journal of Bacteriology* 183.19, pp. 5725–5732. DOI: 10.1128/JB.183.19.5725-5732.2001 (p. 76).

Broad Institute (2015). *Picard*. URL: http://broadinstitute.github.io/picard/ (visited on 06/26/2015) (p. 102).

Bruun, T., G. Sorensen, L. Forshell, T. Jensen, K. Nygard, G. Kapperud, B. Lindstedt, T. Berglund, A. Wingstrand, R. Petersen, L. Muller, C. Kjelso, S. Ivarsson, M. Hjertqvist, S. Lofdahl, and S. Ethelberg (2009). "An outbreak of *Salmonella* Typhimurium infections in Denmark, Norway and Sweden, 2008". *Eurosurveillance* 14.10, pii=19147 (p. 6).

Buchwald, D. S. and M. J. Blaser (1984). "A review of human salmonellosis: II. Duration of excretion following infection with nontyphi *Salmonella*". *Clinical Infectious Diseases* 6.3, pp. 345–356 (p. 3).

Carattoli, A., E. Filetici, L. Villa, A. M. Dionisi, A. Ricci, and I. Luzzi (2002). "Antibiotic resistance genes and *Salmonella* genomic island 1 in *Salmonella enterica* serovar Typhimurium isolated in Italy". *Antimicrobial Agents and Chemotherapy* 46.9, pp. 2821–2828. DOI: 10.1128/AAC.46.9.2821-2828.2002 (p. 86).

Carattoli, A., E. Zankari, A. García-Fernández, M. V. Larsen, O. Lund, L. Villa, F. M. Aarestrup, and H. Hasman (2014). "*In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing". *Antimicrobial Agents and Chemotherapy* 58.7, pp. 3895–3903. DOI: 10.1128/AAC.02412-14 (pp. 104, 121, 135, 136, 140).

Chang, W., J. Cheng, J. Allaire, Y. Xie, and J. McPherson (2015). *Shiny: web application framework for R*. URL: http://CRAN.R-project.org/package=shiny (visited on 02/23/2015) (p. 82).

Chiou, C.-S., C.-S. Hung, M. Torpdahl, H. Watanabe, S. K. Tung, J. Terajima, S.-Y. Liang, and Y.-W. Wang (2010). "Development and evaluation of multilocus variable number tandem repeat analysis for fine typing and phylogenetic analysis of *Salmonella enterica* serovar Typhimurium". *International Journal of Food Microbiology* 142.1–2, pp. 67–73. DOI: 10.1016/j.ijfoodmicro.2010.06.001 (pp. 32, 47).

Chung, B., G. W. Shin, J. Na, M.-H. Oh, and G. Y. Jung (2012). "Multiplex quantitative foodborne pathogen detection using high resolution CE-SSCP coupled stuffer-free multiplex ligation-dependent probe amplification". *Electrophoresis* 33.9-10, pp. 1477–1481. DOI: 10.1002/elps.201100615 (pp. 51, 74).

Cingolani, P., V. M. Patel, M. Coon, T. Nguyen, S. J. Land, D. M. Ruden, and X. Lu (2012). "Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift". *Frontiers in Genetics* 3, 35. DOI: 10.3389/fgene.2012.00035 (p. 102).

Clinical and Laboratory Standards Institute (CLSI) (2015). *CLSI Standards Center*. URL: http://clsi.org/standards/ (visited on 09/02/2015) (p. 11).

Cornelius, A. J., O. Vandenberg, B. Robson, B. J. Gilpin, S. M. Brandt, P. Scholes, D. Martiny, P. E. Carter, P. van Vught, J. Schouten, and S. L. W. On (2014). "Same-day subtyping of *Campylobacter jejuni* and *C. coli* isolates by use of multiplex ligation-dependent probe amplification-binary typing". *Journal of Clinical Microbiology* 52.9, pp. 3345–3350. DOI: 10.1128/JCM.00815-14 (pp. 51, 74).

Cruickshank, J. G. and T. J. Humphrey (1987). "The carrier food-handler and non-typhoid salmonellosis". *Epidemiology and infection* 98.3, pp. 223–230 (p. 3).

Dallman, T. J., M. A. Chattaway, L. A. Cowley, M. Doumith, R. Tewolde, D. J. Wooldridge, A. Underwood, D. Ready, J. Wain, K. Foster, K. A. Grant, and C. Jenkins (2014). "An investigation of the diversity of strains of enteroaggregative *Escherichia coli* isolated from cases associated with a large multi-pathogen foodborne outbreak in the UK". *PLoS ONE* 9.5, e98103. DOI: 10.1371/journal.pone.0098103 (p. 97).

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group (2011). "The variant call format and VCFtools". *Bioinformatics* 27.15, pp. 2156–2158. DOI: 10.1093/bioinformatics/btr330 (p. 102).

Danoff-Burg, J. (2003). *Measuring Biological Diversity*. Columbia University, New York, USA. URL: http://www.columbia.edu/itc/cerc/danoff-burg/MBD_Links.html (visited on 08/09/2013) (p. 35).

De Smet, L., J. Ravoet, J. R. de Miranda, T. Wenseleers, M. Y. Mueller, R. F. A. Moritz, and D. C. de Graaf (2012). "BeeDoctor, a versatile MLPA-based diagnostic tool for screening bee viruses". *PLoS ONE* 7.10, e47953. DOI: 10.1371/journal.pone.0047953 (pp. 51, 74).

Deng, X., N. Shariat, E. M. Driebe, C. C. Roe, B. Tolar, E. Trees, P. Keim, W. Zhang, E. G. Dudley, P. I. Fields, and D. M. Engelthaler (2015). "Comparative analysis of subtyping methods against a whole-genome-sequencing standard for *Salmonella enterica* serotype Enteritidis". *Journal of Clinical Microbiology* 53.1, pp. 212–218. DOI: 10.1128/JCM.02332-14 (p. 73).

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data". *Nature Genetics* 43.5, pp. 491–498. DOI: 10.1038/ng.806 (p. 102).

Deshpande, A., J. Gans, S. W. Graves, L. Green, L. Taylor, H. B. Kim, Y. A. Kunde, P. M. Leonard, P.-E. Li, J. Mark, J. Song, M. Vuyisich, and P. S. White (2010). "A rapid multiplex assay for nucleic acid-based diagnostics". *Journal of Microbiological Methods* 80.2, pp. 155–163. DOI: 10.1016/j.mimet.2009.12.001 (pp. 19, 51, 61, 65–67, 74).

Di Bella, J. M., Y. Bao, G. B. Gloor, J. P. Burton, and G. Reid (2013). "High throughput sequencing methods and analysis for microbiome research". *Journal of Microbiological Methods* 95.3, pp. 401–414. DOI: 10.1016/j.mimet.2013.08.011 (pp. 20–22).

Dimovski, K., H. Cao, O. Wijburg, R. Strugnell, R. Mantena, M. Whipp, G. Hogg, and K. Holt (2014). "Analysis of *Salmonella enterica* serovar Typhimurium variable-number tandem-repeat data for public health

investigation based on measured mutation rates and whole-genome sequence comparisons". *Journal of Bacteriology* 196.16, pp. 3036–3044. DOI: 10.1128/JB.01820-14 (pp. 6, 89–91, 146).

Dionisi, A. M., C. Graziani, C. Lucarelli, E. Filetici, L. Villa, S. Owczarek, A. Caprioli, and I. Luzzi (2009). "Molecular characterization of multidrug-resistant strains of *Salmonella enterica* serotype Typhimurium and monophasic variant (*S.* 4,[5],12:i:-) isolated from human infections in Italy". *Foodborne Pathogens and Disease* 6.6, pp. 711–717. DOI: 10.1089/fpd.2008.0240 (p. 43).

Drahovská, H., E. Mikasová, T. Szemes, A. Ficek, M. Sásik, V. Majtán, and J. Turna (2007). "Variability in occurrence of multiple prophage genes in *Salmonella* Typhimurium strains isolated in Slovak Republic". *FEMS Microbiology Letters* 270.2, pp. 237–244. DOI: 10.1111/j.1574-6968.2007.00674.x (pp. 76, 83).

Dunn, J. R. (2015). "Whole-genome sequencing: opportunities and challenges for public health, food-borne outbreak investigations, and the global food supply". *Journal of Infectious Diseases*. DOI: 10.1093/infdis/jiv298. In press (p. 132).

Dyet, K. H., E. Turbitt, and P. E. Carter (2011). "Multiple-locus variable-number tandem-repeat analysis for discriminating within *Salmonella enterica* serovar Typhimurium definitive types and investigation of outbreaks". *Epidemiology and infection* 139.7, pp. 1050–1059. DOI: 10.1017/S0950268810002025 (p. 45).

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". *Nucleic Acids Research* 32.5, pp. 1792–1797. DOI: 10.1093/nar/gkh340 (p. 103).

Ethelberg, S., M. Lisby, M. Torpdahl, G. Sørensen, J. Neimann, P. Rasmussen, S. Bang, U. Stamer, H. B. Hansson, K. Nygård, D. L. Baggesen, E. M. Nielsen, K. Mølbak, and M. Helms (2004). "Prolonged restaurant-associated outbreak of multidrug-resistant *Salmonella* Typhimurium among patients from several European countries". *Clinical Microbiology and Infection* 10.10, pp. 904–910. DOI: 10.1111/j.1469-0691.2004.00990.x (p. 43).

European Centre for Disease Prevention and Control (ECDC) (2011). *Laboratory standard operating procedure for MLVA of Salmonella enterica serotype Typhimurium*. Version 1.2. Stockholm, Sweden. DOI: 10.2900/56328 (p. 34).

European Commission (2013). "Decision on the monitoring and reporting of antimicrobial resistance in zoonotic and commensal bacteria (2013/652/EU)". *Official Journal of The European Union* L 303/26 (pp. 11, 134).

European Committee on Antimicrobial Susceptibility Testing (EUCAST) (2013a). *Antimicrobial susceptibility testing: EUCAST disk diffusion method*. Version 3.0 (p. 33).

European Committee on Antimicrobial Susceptibility Testing (EUCAST) (2013b). *Breakpoint tables for interpretation of MICs and zone diameters*. Version 3.1 (p. 33).

European Committee on Antimicrobial Susceptibility Testing (EUCAST) (2015a). *Breakpoint tables for interpretation of MICs and zone diameters*. Version 5.0 (p. 11).

European Committee on Antimicrobial Susceptibility Testing (EUCAST) (2015b). *MIC distributions and ECOFFs*. URL: http://www.eucast.org/ (visited on 05/15/2015) (pp. 11, 134).

European Food Safety Authority (EFSA) (2014). "Use of whole genome sequencing (WGS) of food-borne pathogens for public health protection". *EFSA Scientific Colloquium Summary Report*. EFSA Scientific Colloquium 20. (Parma, Italy, June 16, 2014). ISBN: 978-92-9199-621-6. DOI: 10.2805/66246. URL: http://www.efsa.europa.eu/en/supporting/doc/743e.pdf (pp. 21–24, 74, 97, 125, 132).

European Food Safety Authority (EFSA) Panel on Biological Hazards (2010). "Scientific Opinion on monitoring and assessment of the public health risk of "*Salmonella* Typhimurium-like" strains". *EFSA Journal* 8.10, 1826. DOI: doi:10.2903/j.efsa.2010.1826 (p. 83).

European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) (2013). "The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2011". *EFSA Journal* 11.4, 3129. DOI: 10.2903/j.efsa.2013.3129 (pp. 31, 45).

European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) (2015). "The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013". *EFSA Journal* 13.1, 3991. DOI: 10.2903/j.efsa.2015.3991 (pp. 2, 5, 73, 132).

Fabre, L., J. Zhang, G. Guigon, S. Le Hello, V. Guibert, M. Accou-Demartin, S. de Romans, C. Lim, C. Roux, V. Passet, L. Diancourt, M. Guibourdenche, S. Issenhuth-Jeanjean, M. Achtman, S. Brisse, C. Sola, and F.-X. Weill (2012). "CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections". *PLoS ONE* 7.5, e36995. DOI: 10.1371/journal.pone.0036995 (pp. 9, 10, 45, 47).

Fang, N.-X., B. Huang, L. Hiley, J. Bates, and J. Savill (2012). "A rapid multiplex DNA suspension array method for *Salmonella typhimurium* subtyping using prophage-related markers". *Journal of Microbiological Methods* 88.1, pp. 19–27. DOI: 10.1016/j.mimet.2011.10.002 (pp. 9, 52, 74, 76, 83–85, 91).

Friesema, I. H. M., B. Schimmer, J. A. Ros, H. J. Ober, M. E. O. C. Heck, C. M. Swaan, C. M. de Jager, R. M. Peran i Sala, and W. van Pelt (2012). "A regional *Salmonella enterica* serovar Typhimurium outbreak associated

with raw beef products, The Netherlands, 2010". *Foodborne Pathogens and Disease* 9.2, pp. 102–107. DOI: 10.1089/fpd.2011.0978 (pp. 6, 32, 91).

Garin, I., G. Mantovani, U. Aguirre, A. Barlier, B. Brix, F. M. Elli, K. Freson, V. Grybek, B. Izzi, A. Linglart, G. Perez de Nanclares, C. Silve, S. Thiele, R. Werner, and on behalf of the EuroPHP Consortium (2014). "European guidance for the molecular diagnosis of pseudohypoparathyroidism not caused by point genetic variants at *GNAS*: an EQA study". *European Journal of Human Genetics* 23.4, pp. 438–444. DOI: 10.1038/ejhg.2014.127 (pp. 51, 74).

Garvey, P., P. McKeown, P. Kelly, M. Cormican, W. Anderson, A. Flack, S. Barron, N. De Lappe, J. Buckley, C. Cosgrove, D. Molloy, J. O'Connor, P. O'Sullivan, J. Matthews, M. Ward, A. Breslin, M. B. O'Sullivan, K. Kelleher, A. McNamara, C. Foley-Nolan, H. Pelly, F. Cloak, and Outbreak control team (2013). "Investigation and management of an outbreak of *Salmonella* Typhimurium DT8 associated with duck eggs, Ireland 2009 to 2011". *Eurosurveillance* 18.16, pii=20454 (pp. 6, 32, 91).

Gilchrist, C. A., S. D. Turner, M. F. Riley, J. Petri William A., and E. L. Hewlett (2015). "Whole-genome sequencing in outbreak analysis". *Clinical Microbiology Reviews* 28.3, pp. 541–563. DOI: 10.1128/CMR.00075-13 (p. 132).

Gilmour, M., M. Graham, A. Reimer, and G. Van Domselaar (2013). "Public health genomics and the new molecular epidemiology of bacterial pathogens". *Public Health Genomics* 16.1–2, pp. 25–30. DOI: 10.1159/000342709 (pp. 19, 24).

Gödel, K. (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". [German]. *Monatshefte für Mathematik und Physik* 38, pp. 173–198 (p. 81).

Gossner, C. M., D. van Cauteren, S. Le Hello, F. X. Weill, E. Terrien, S. Tessier, C. Janin, A. Brisabois, V. Dusch, V. Vaillant, and N. Jourdan-da Silva (2012). "Nationwide outbreak of *Salmonella enterica* serotype 4,[5],12:i:- infection associated with consumption of dried pork sausage, France, November to December 2011". *Eurosurveillance* 17.5, pii=20071 (p. 45).

Graziani, C., L. Busani, A. M. Dionisi, C. Lucarelli, S. Owczarek, A. Ricci, M. Mancin, A. Caprioli, and I. Luzzi (2008). "Antimicrobial resistance in *Salmonella enterica* serovar Typhimurium from human and animal sources in Italy". *Veterinary Microbiology* 128.3–4, pp. 414–418. DOI: 10.1016/j.vetmic.2007.10.017 (p. 43).

Greenfield, P., K. Duesing, A. Papanicolaou, and D. C. Bauer (2014). "Blue: correcting sequencing errors using consensus and context". *Bioinformatics* 30.19, pp. 2723–2732. DOI: 10.1093/bioinformatics/btu368 (pp. 98, 99).

Grimont, P. A. D. and F.-X. Weill (2007). *Antigenic formulae of the Salmonella serovars*. 9th edition. Paris, France (pp. 2, 33, 73, 134).

Guigon, G., J. Cheval, R. Cahuzac, and S. Brisse (2008). "MLVA-NET – a standardised web database for bacterial genotyping and surveillance". *Eurosurveillance* 13.19, pii=18863 (p. 45).

Gunn, J. S., J. M. Marshall, S. Baker, S. Dongol, R. C. Charles, and E. T. Ryan (2014). "*Salmonella* chronic carriage: epidemiology, diagnosis, and gallbladder persistence". *Trends in Microbiology* 22.11, pp. 648–655. DOI: 10.1016/j.tim.2014.06.007 (pp. 2, 3).

Hampton, M. D., E. J. Threlfall, J. A. Frost, L. R. Ward, and B. Rowe (1995). "*Salmonella typhimurium* DT 193: differentiation of an epidemic phage type by antibiogram, plasmid profile, plasmid fingerprint and *Salmonella plasmid virulence* (*spv*) gene probe". *Journal of Applied Bacteriology* 78.4, pp. 402–408 (p. 43).

Hauser, E., E. Tietze, R. Helmuth, E. Junker, K. Blank, R. Prager, W. Rabsch, B. Appel, A. Fruth, and B. Malorny (2010). "Pork contaminated with *Salmonella enterica* serovar 4,[5],12:i:-, an emerging health risk for humans". *Applied and Environmental Microbiology* 76.14, pp. 4601–4610. DOI: 10.1128/AEM.02991-09 (p. 43).

Heck, M. (2009). "Multilocus variable number of tandem repeats analysis (MLVA) – a reliable tool for rapid investigation of *Salmonella* Typhimurium outbreaks". *Eurosurveillance* 14.15, pii=19177. DOI: 10.1128/AEM.02991-09 (pp. 31, 45).

Hernández Guijarro, K., S. Feingold, and H. Terzolo (2012). "A single nucleotide polymorphism on *rpoB* gene allows specific identification of *Salmonella enterica* serotype Typhimurium". *Research Journal of Microbiology* 7.7, pp. 344–352. DOI: 10.3923/jm.2012.344.352 (p. 84).

Hopkins, K. L., M. Kirchner, B. Guerra, S. A. Granier, C. Lucarelli, M. C. Porrero, A. Jakubczak, E. J. Threlfall, and D. J. Mevius (2010). "Multiresistant *Salmonella enterica* serovar 4,[5],12:i:- in Europe: a new pandemic strain?" *Eurosurveillance* 15.22, pii=19580 (pp. 43, 45).

Hopkins, K. L., C. Maguire, E. Best, E. Liebana, and E. J. Threlfall (2007). "Stability of multiple-locus variable-number tandem repeats in *Salmonella enterica* serovar Typhimurium". *Journal of Clinical Microbiology* 45.9, pp. 3058–3061. DOI: 10.1128/JCM.00715-07 (p. 32).

Hopkins, K. L., E. de Pinna, and J. Wain (2012). "Prevalence of *Salmonella enterica* serovar 4,[5],12:i:- in England and Wales, 2010". *Eurosurveillance* 17.37, pii=20275 (pp. 43, 45).

Hopkins, K., T. Peters, E. de Pinna, and J. Wain (2011). "Standardisation of multilocus variable-number tandem repeat analysis (MLVA) for subtyping of *Salmonella enterica* serovar Enteritidis". *Eurosurveillance* 16.32, pii=19942 (pp. 6, 134).

Hu, H., R. Lan, and P. R. Reeves (2002). "Fluorescent amplified fragment length polymorphism analysis of *Salmonella enterica* serovar Typhimurium reveals

phage-type-specific markers and potential for microarray typing". *Journal of Clinical Microbiology* 40.9, pp. 3406–3415. DOI: 10.1128/JCM.40.9.3406-3415.2002 (p. 76).

Hu, H., R. Lan, and P. R. Reeves (2006). "Adaptation of multilocus sequencing for studying variation within a major clone: evolutionary relationships of *Salmonella enterica* serovar Typhimurium". *Genetics* 172.2, pp. 743–750. DOI: 10.1534/genetics.105.046466 (p. 76).

Humphries, R. M. and A. N. Schuetz (2015). "Antimicrobial susceptibility testing of bacteria that cause gastroenteritis". *Clinics in Laboratory Medicine* 35.2, pp. 313–331. DOI: 10.1016/j.cll.2015.02.005 (p. 10).

Hunter, P. R. and M. A. Gaston (1988). "Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity". *Journal of Clinical Microbiology* 26.11, pp. 2465–2466 (pp. 9, 35, 40, 82).

Hunter, S. B., P. Vauterin, M. A. Lambert-Fair, M. Van Duyne, K. Kubota, L. Graves, D. Wrigley, T. Barrett, and E. Ribot (2005). "Establishment of a universal size standard strain for use with the PulseNet standardized pulsed-field gel electrophoresis protocols: converting the national databases to the new size standard". *Journal of Clinical Microbiology* 43.3, pp. 1045–1050. DOI: 10.1128/JCM.43.3.1045-1050.2005 (pp. 8, 75).

Illumina (2014). *Documentation & Literature. HiSeq Systems Specification Sheet*. URL: http://support.illumina.com/sequencing/sequencing_instruments/hiseq_2000/documentation.html (visited on 06/22/2015) (p. 21).

Illumina (2015a). *HiSeq 2500 Literature. HiSeq 2500 System Specification Sheet*. URL: http://www.illumina.com/systems/hiseq_2500_1500/literature.html (visited on 06/22/2015) (p. 21).

Illumina (2015b). *Mate pair sequencing*. URL: http://www.illumina.com/technology/next-generation-sequencing/mate-pair-sequencing_assay.html (visited on 06/13/2015) (p. 20).

Illumina (2015c). *MiSeq Literature. MiSeq System Specification Sheet*. URL: http://www.illumina.com/systems/miseq/documentation.html (visited on 06/22/2015) (p. 21).

Illumina (2015d). *Paired-end sequencing*. URL: http://www.illumina.com/technology/next-generation-sequencing/paired-end-sequencing_assay.html (visited on 06/13/2015) (p. 20).

ISO (2012). *ISO 6579:2002 Microbiology of food and animal feeding stuffs - Horizontal method for the detection of Salmonella spp.* ISO. Geneva, Switzerland: International Organization for Standardization (p. 134).

Issenhuth-Jeanjean, S., P. Roggentin, M. Mikoleit, M. Guibourdenche, E. de Pinna, S. Nair, P. I. Fields, and F.-X. Weill (2014). "Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme". *Research in*

Microbiology 165.7, pp. 526–530. DOI: 10.1016/j.resmic.2014.07.004 (pp. 2, 73).

Joensen, K., F. Scheutz, O. Lund, H. Hasman, R. Kaas, E. Nielsen, and F. Aarestrup (2014). "Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*". *Journal of Clinical Microbiology* 52.5, pp. 1501–1510. DOI: 10.1128/JCM.03617-13 (pp. 23, 103).

Jolley, K. (2015). *PubMLST*. Department of Zoology, University of Oxford, UK. URL: http://pubmlst.org (visited on 06/26/2015) (p. 97).

Jolley, K. A., D. M. C. Hill, H. B. Bratcher, O. B. Harrison, I. M. Feavers, J. Parkhill, and M. C. J. Maiden (2012). "Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid web-based analysis methods". *Journal of Clinical Microbiology* 50.9, pp. 3046–3053. DOI: 10.1128/JCM.01312-12 (p. 133).

Jolley, K. A. and M. C. Maiden (2010). "BIGSdb: Scalable analysis of bacterial genome variation at the population level". *BMC Bioinformatics* 11, 595. DOI: 10.1186/1471-2105-11-595 (p. 23).

Jorgensen, J. H. and M. J. Ferraro (2009). "Antimicrobial susceptibility testing: a review of general principles and contemporary practices". *Clinical Infectious Diseases* 49.11, pp. 1749–1755. DOI: 10.1086/647952 (p. 11).

Kaas, R. S., P. Leekitcharoenphon, F. M. Aarestrup, and O. Lund (2014). "Solving the problem of comparing whole bacterial genomes across different sequencing platforms". *PLoS ONE* 9.8, e104984. DOI: 10.1371/journal.pone.0104984 (pp. 102, 103, 116, 124, 135, 136, 140).

Kasatkar, P., S. Shetty, and K. Ghosh (2014). "Prenatal diagnosis in severe von Willebrand disease families from India using combination of phenotypic and genotypic assays". *Prenatal Diagnosis* 34.4, pp. 377–381. DOI: 10.1002/pd.4315 (pp. 51, 74).

Kipp, B. R., S. E. Roellinger, P. A. Lundquist, W. E. Highsmith, and D. B. Dawson (2011). "Development and clinical implementation of a combination deletion PCR and multiplex ligation-dependent probe amplification assay for detecting deletions involving the human α-globin gene cluster". *Journal of Molecular Diagnostics* 13.5, pp. 549–557. DOI: 10.1016/j.jmoldx.2011.04.001 (p. 51).

Kotova, A., S. Kondratskaya, and I. Yasutis (1988). "*Salmonella* carrier state and biological characteristics of the infectious agent." *Journal of Hygiene, Epidemiology, Microbiology, and Immunology* 32.1, pp. 71–78 (p. 3).

Kuhn, K. G., G. Sørensen, M. Torpdahl, M. K. Kjeldsen, T. Jensen, S. Gubbels, G. O. Bjerager, A. Wingstrand, L. J. Porsbo, and S. Ethelberg (2013). "A long-lasting outbreak of *Salmonella* Typhimurium U323 associated with several pork products, Denmark, 2010". *Epidemiology and infection* 141.2, pp. 260–268. DOI: 10.1017/S0950268812000702 (pp. 6, 32, 45, 46, 91).

Lan, R., G. Stevenson, K. Donohoe, L. Ward, and P. R. Reeves (2007). "Molecular markers with potential to replace phage typing for *Salmonella enterica* serovar *typhimurium*". *Journal of Microbiological Methods* 68.1, pp. 145–156. DOI: 10.1016/j.mimet.2006.07.004 (pp. 76, 83, 84).

Larsen, M. V., S. Cosentino, S. Rasmussen, C. Friis, H. Hasman, R. L. Marvig, L. Jelsbak, T. Sicheritz-Pontén, D. W. Ussery, F. M. Aarestrup, and O. Lund (2012). "Multilocus sequence typing of total-genome-sequenced bacteria". *Journal of Clinical Microbiology* 50.4, pp. 1355–1361. DOI: 10.1128/JCM.06094-11 (pp. 23, 104, 134, 136, 140).

Larsson, J. T., M. Torpdahl, MLVA working group, and E. Møller Nielsen (2013). "Proof-of-concept study for successful inter-laboratory comparison of MLVA results". *Eurosurveillance* 18.35, pii=20566 (pp. 25, 31, 32, 45).

Larsson, J. T., M. Torpdahl, R. F. Petersen, G. Sørensen, B. A. Lindstedt, and E. M. Nielsen (2009). "Development of a new nomenclature for *Salmonella* Typhimurium multilocus variable number of tandem repeats analysis (MLVA)". *Eurosurveillance* 14.15, pii=19174 (pp. 34, 45, 75).

Lawson, A. J., M. U. Dassama, L. R. Ward, and J. Threlfall (2002). "Multiply resistant (MR) *Salmonella enterica* serotype Typhimurium DT 12 and DT 120: a case of MR DT 104 in disguise?" *Emerging Infectious Diseases* 8.4, pp. 434–436 (p. 86).

Le Flèche, P., Y. Hauck, L. Onteniente, A. Prieur, F. Denoeud, V. Ramisse, P. Sylvestre, G. Benson, F. Ramisse, and G. Vergnaud (2001). "A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*". *BMC Microbiology* 1, 2 (p. 47).

Le Minor, L. and C. Richard (1993). *Méthodes de laboratoires pour l'identification des entérobactéries*. Paris, France: Institut Pasteur (p. 2).

Leggett, R. M. and D. MacLean (2014). "Reference-free SNP detection: dealing with the data deluge". *BMC Genomics* 15 Suppl 4, S10. DOI: 10.1186/1471-2164-15-S4-S10 (pp. 125, 149).

Levine, M. M. and R. M. Robins-Browne (2012). "Factors that explain excretion of enteric pathogens by persons without diarrhea". *Clinical Infectious Diseases* 55.S4, S303–S311. DOI: 10.1093/cid/cis789 (p. 3).

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows–Wheeler transform". *Bioinformatics* 25.14, pp. 1754–1760. DOI: 10.1093/bioinformatics/btp324 (pp. 102, 103).

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup (2009). "The Sequence Alignment/Map format and SAMtools". *Bioinformatics* 25.16, pp. 2078–2079. DOI: 10.1093/bioinformatics/btp352 (p. 103).

Lieberman, T. D., J.-B. Michel, M. Aingaran, G. Potter-Bynoe, D. Roux, J. Davis Michael R., D. Skurnik, N. Leiby, J. J. LiPuma, J. B. Goldberg, A. J.

McAdam, G. P. Priebe, and R. Kishony (2011). "Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes". *Nature Genetics* 43.12, pp. 1275–1280. DOI: 10.1038/ng.997 (p. 98).

Ligges, U. and M. Mächler (2003). "Scatterplot3d - an R package for visualizing multivariate data". *Journal of Statistical Software* 8.11, pp. 1–20 (p. 82).

Lindstedt, B. A., M. Torpdahl, G. Vergnaud, S. Le Hello, F. X. Weill, E. Tietze, B. Malorny, D. M. Prendergast, E. Ní Ghallchóir, R. F. Lista, L. M. Schouls, R. Söderlund, S. Börjesson, and S. Åkerström (2013). "Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012". *Eurosurveillance* 18.4, pii=20385 (pp. 25, 31, 32, 43, 45).

Lindstedt, B., E. Heir, E. Gjernes, and G. Kapperud (2003). "DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci". *Journal of Clinical Microbiology* 41.4, pp. 1469–1479. DOI: 10.1128/JCM.41.4.1469-1479.2003 (pp. 31, 47).

Lindstedt, B., T. Vardund, L. Aas, and G. Kapperud (2004). "Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis". *Journal of Microbiological Methods* 59.2, pp. 163–172. DOI: 10.1016/j.mimet.2004.06.014 (pp. 6, 31, 32, 34, 45–47, 75, 76, 83).

Loman, N., C. Constantinidou, J. Chan, M. Halachev, M. Sergeant, C. Penn, E. Robinson, and M. Pallen (2012). "High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity". *Nature Reviews Microbiolology* 10.9, pp. 599–606. DOI: 10.1038/nrmicro2850 (pp. 20–22).

Luminex Corporation (2012). *xMAP Salmonella Serotyping Assay Product Insert*. P/N 89-60000-00-091 Rev.A (p. 53).

Luminex Corporation (2014). *Relative reporter intensities*. URL: http://www.luminexcorp.com/prod/groups/public/documents/lmnxcorp/relative-reporter-intensities.pdf (visited on 08/07/2014) (p. 67).

Marcinkowska-Swojak, M., K. Klonowska, M. Figlerowicz, and P. Kozlowski (2014). "An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs". *Gene* 546.2, pp. 257–262. DOI: 10.1016/j.gene.2014.05.072 (pp. 51, 74).

Melchior Jr., W. B. and P. H. Von Hippel (1973). "Alteration of the relative stability of dA-dT and dG-dC base pairs in DNA". *Proceedings of the National Academy of Sciences of the United States of America* 70.2, pp. 298–302 (p. 15).

Mikasová, E., H. Drahovská, T. Szemes, T. Kuchta, R. Karpíšková, M. Sásik, and J. Turna (2005). "Characterization of *Salmonella enterica* serovar Typhimurium strains of veterinary origin by molecular typing methods".

Veterinary Microbiology 109.1–2, pp. 113–120. DOI: 10.1016/j.vetmic.2005.05.006 (pp. 76, 83).

Morris, C., C. K. P. Tam, T. S. Wallis, P. W. Jones, and J. Hackett (2003). "*Salmonella enterica* serovar Dublin strains which are Vi antigen-positive use type IVB pili for bacterial self-association and human intestinal cell entry". *Microbial Pathogenesis* 35.6, pp. 279–284. DOI: 10.1016/j.micpath.2003.08.001 (p. 2).

Mossong, J., P. Marques, C. Ragimbeau, P. Huberty-Krau, S. Losch, G. Meyer, G. Moris, C. Strottner, W. Rabsch, and F. Schneider (2007). "Outbreaks of monophasic *Salmonella enterica* serovar 4,[5],12:i:- in Luxembourg, 2006". *Eurosurveillance* 12.3–6, pp. 156–158 (p. 43).

Ng, L., M. Mulvey, I. Martin, G. Peters, and W. Johnson (1999). "Genetic characterization of antimicrobial resistance in Canadian isolates of *Salmonella* serovar Typhimurium DT104". *Antimicrobial Agents and Chemotherapy* 43.12, pp. 3018–3021 (p. 83).

Octavia, S., Q. Wang, M. M. Tanaka, S. Kaur, V. Sintchenko, and R. Lan (2015). "Delineating community outbreaks of *Salmonella enterica* serovar Typhimurium by use of whole-genome sequencing: insights into genomic variability within an outbreak". *Journal of Clinical Microbiology* 53.4, pp. 1063–1071. DOI: 10.1128/JCM.03235-14 (p. 73).

Okoro, C. K., L. Barquist, T. R. Connor, S. R. Harris, S. Clare, M. P. Stevens, M. J. Arends, C. Hale, L. Kane, D. J. Pickard, J. Hill, K. Harcourt, J. Parkhill, G. Dougan, and R. A. Kingsley (2015). "Signatures of adaptation in human invasive *Salmonella* Typhimurium ST313 populations from Sub-Saharan Africa". *PLoS Neglected Tropical Disease* 9.3, e0003611. DOI: 10.1371/journal.pntd.0003611 (p. 98).

Okoro, C. K., R. A. Kingsley, M. A. Quail, A. M. Kankwatira, N. A. Feasey, J. Parkhill, G. Dougan, and M. A. Gordon (2012). "High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal *Salmonella* Typhimurium disease". *Clinical Infectious Diseases* 54.7, pp. 955–963. DOI: 10.1093/cid/cir1032 (p. 98).

Pang, S., S. Octavia, P. R. Reeves, Q. Wang, G. L. Gilbert, V. Sintchenko, and R. Lan (2012). "Genetic relationships of phage types and single nucleotide polymorphism typing of *Salmonella enterica* serovar Typhimurium". *Journal of Clinical Microbiology* 50.3, pp. 727–734. DOI: 10.1128/JCM.01284-11 (pp. 76, 83).

Paranthaman, K., S. Haroon, S. Latif, N. Vinnyey, V. de Souza, W. Welfare, M. Tahir, E. Cooke, K. Stone, C. Lane, T. Peters, and R. Puleston (2013). "Emergence of a multidrug-resistant (ASSuTTm) strain of *Salmonella enterica* serovar Typhimurium DT120 in England in 2011 and the use of multiple-locus variable-number tandem-repeat analysis in

supporting outbreak investigations". *Foodborne Pathogens and Disease* 10.10, pp. 850–855. DOI: 10.1089/fpd.2013.1513 (pp. 6, 32, 91).

Peters, T. M. (2009). "Pulsed-field gel electrophoresis for molecular epidemiology of food pathogens". *Molecular Epidemiology of Microorganisms*. Ed. by D. A. Caugant. Methods in Molecular Biology 394. Totowa, New Jersey, USA: Humana Press. Chap. 6, pp. 59–70. ISBN: 978-1-60327-998-7. DOI: 10.1007/978-1-60327-999-4_6 (p. 6).

Petersen, R. F., E. Litrup, J. T. Larsson, M. Torpdahl, G. Sørensen, L. Müller, and E. M. Nielsen (2011). "Molecular characterization of *Salmonella* Typhimurium highly successful outbreak strains". *Foodborne Pathogens and Disease* 8.6, pp. 655–661. DOI: 10.1089=fpd.2010.0683 (pp. 6, 32, 91).

Pham Thanh, D., N. Tran Vu Thieu, C. Tran Thuy, M. Lodén, K. Tuin, J. I. Campbell, N. Van Minh Hoang, P. Voong Vinh, J. J. Farrar, K. E. Holt, G. Dougan, and S. Baker (2013). "Identification of *Salmonella enterica* serovar Typhi genotypes by use of rapid multiplex ligation-dependent probe amplification". *Journal of Clinical Microbiology* 51.9, pp. 2950–2958. DOI: 10.1128/JCM.01010-13 (pp. 51, 74).

Prendergast, D. M., D. O'Grady, S. Fanning, M. Cormican, N. Delappe, J. Egan, C. Mannion, J. Fanning, and M. Gutierrez (2011). "Application of multiple locus variable number of tandem repeat analysis (MLVA), phage typing and antimicrobial susceptibility testing to subtype *Salmonella enterica* serovar Typhimurium isolated from pig farms, pork slaughterhouses and meat producing plants in Ireland". *Food Microbiology* 28.5, pp. 1087–1094. DOI: 10.1016/j.fm.2011.02.013 (pp. 43, 45, 46).

Price, M. N., P. S. Dehal, and A. P. Arkin (2010). "FastTree 2 - approximately maximum-likelihood trees for large alignments". *PLoS ONE* 5.3, e9490. DOI: 10.1371/journal.pone.0009490 (p. 103).

Pui, C., W. Wong, L. Chai, R. Tunung, P. Jeyaletchumi, M. Noor Hidayah, A. Ubong, M. Farinazleen, Y. Cheah, and R. Son (2011). "*Salmonella*: A foodborne pathogen". *International Food Research Journal* 18.2, pp. 465–473 (p. 3).

PulseNet International (2015a). *PulseNet Europe*. URL: http://www.pulsenetinternational.org/networks/europe (visited on 05/08/2015) (p. 75).

PulseNet International (2015b). *The international molecular subtyping network for foodborne disease surveillance*. URL: http://www.pulsenetinternational.org/ (visited on 05/30/2015) (p. 6).

PulseNet USA (2013). *PulseNet standard operating procedure for PulseNet MLVA of Salmonella enterica serotype Typhimurium – Applied Biosystems Genetic Analyzer 3130XL platform*. PNL24 (pp. 32, 47).

Qiagen (2010). *HotStarTaq PCR Handbook*. URL: http://www.qiagen.com (p. 65).

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features". *Bioinformatics* 26.6, pp. 841–842. DOI: 10.1093/bioinformatics/btq033 (p. 103).

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org (pp. 57, 58, 80).

Rabsch, W. (2007). "*Salmonella* Typhimurium phage typing for pathogens". *Salmonella: Methods and Protocols*. Ed. by H. Schatten and A. Eisenstark. Methods in Molecular Biology 394. Totowa, New Jersey, USA: Humana Press. Chap. 10, pp. 177–217. ISBN: 978-1-59745-512-1. DOI: 10.1007/978-1-59745-512-1_10 (p. 7).

Rambaut, A. (2014). *FigTree*. URL: http://tree.bio.ed.ac.uk/software/figtree/ (visited on 04/15/2015) (p. 135).

Ramisse, V., P. Houssu, E. Hernandez, F. Denoeud, V. Hilaire, O. Lisanti, F. Ramisse, J.-D. Cavallo, and G. Vergnaud (2004). "Variable number of tandem repeats in *Salmonella enterica* subsp. *enterica* for typing purposes". *Journal of Clinical Microbiology* 42.12, pp. 5722–5730. DOI: 10.1128/JCM.42.12.5722-5730.2004 (pp. 31, 47).

Reijans, M., G. Dingemans, C. H. Klaassen, J. F. Meis, J. Keijdener, B. Mulders, K. Eadie, W. van Leeuwen, A. van Belkum, A. M. Horrevorts, and G. Simons (2008). "RespiFinder: a new multiparameter test to differentially identify fifteen respiratory viruses". *Journal of Clinical Microbiology* 46.4, pp. 1232–1240. DOI: 10.1128/JCM.02294-07 (pp. 51, 74).

Ribot, E., M. Fair, R. Gautom, D. Cameron, S. Hunter, B. Swaminathan, and T. Barrett (2006). "Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet". *Foodborne Pathogens and Disease* 3.1, pp. 59–67. DOI: 10.1089/fpd.2006.3.59 (pp. 8, 75).

Ross, I. L., D. E. Davos, L. Mwanri, J. Raupach, and M. W. Heuzenroeder (2011). "MLVA and phage typing as complementary tools in the epidemiological investigation of *Salmonella enterica* serovar Typhimurium clusters". *Current Microbiology* 62.3, pp. 1034–1038. DOI: 10.1007/s00284-010-9820-1 (pp. 32, 45, 46).

Ross, I. L. and M. W. Heuzenroeder (2005). "Discrimination within phenotypically closely related definitive types of *Salmonella enterica* serovar Typhimurium by the multiple amplification of phage locus typing technique". *Journal of Clinical Microbiology* 43.4, pp. 1604–1611. DOI: 10.1128/JCM.43.4.1604-1611.2005 (pp. 76, 83).

Ross, M. G., C. Russ, M. Costello, A. Hollinger, N. J. Lennon, R. Hegarty, C. Nusbaum, and D. B. Jaffe (2013). "Characterizing and measuring bias in sequence data". *Genome Biology* 14.5, R51. DOI: 10.1186/gb-2013-14-5-r51 (pp. 21, 22).

Rychlik, I., D. Gregorova, and H. Hradecka (2006). "Distribution and function of plasmids in *Salmonella enterica*". *Veterinary Microbiology* 112.1, pp. 1–10. DOI: 10.1016/j.vetmic.2005.10.030 (p. 98).

Rychlík, I., H. Hradecka, and M. Malcova (2008). "*Salmonella enterica* serovar Typhimurium typing by prophage-specific PCR". *Microbiology* 154.5, pp. 1384–1389. DOI: 10.1099/mic.0.2007/015156-0 (pp. 76, 83).

Sabat, A., A. Budimir, D. Nashev, R. Sá-Leão, J. van Dijl, F. Laurent, H. Grundmann, A. Friedrich, and on behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM) (2013). "Overview of molecular typing methods for outbreak detection and epidemiological surveillance". *Eurosurveillance* 18.4, pii=20380 (pp. 2, 8, 19, 23, 73, 132).

Schouten, J. P., C. J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, and G. Pals (2002). "Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification". *Nucleic Acids Research* 30.12, e57 (pp. 51, 74).

Shannon, C. E. (1948). "A mathematical theory of communication". *The Bell System Technical Journal* 27, pp. 379–423 (pp. 9, 35).

Sintchenko, V., Q. Wang, P. Howard, C. W. Y. Ha, K. Kardamanidis, J. Musto, and G. L. Gilbert (2012). "Improving resolution of public health surveillance for human *Salmonella enterica* serovar Typhimurium infection: 3 years of prospective multiple-locus variable-number tandem-repeat analysis (MLVA)". *BMC Infectious Diseases* 12, 78. DOI: 10.1186/1471-2334-12-78 (pp. 31, 43).

Slater, H., D. Bruno, H. Ren, P. La, T. Burgess, L. Hills, S. Nouri, J. Schouten, and K. H. A. Choo (2004). "Improved testing for CMT1A and HNPP using multiplex ligation-dependent probe amplification (MLPA) with rapid DNA preparations: comparison with the interphase FISH method". *Human Mutation* 24.2, pp. 164–171. DOI: 10.1002/humu.20072 (pp. 51, 74).

Soler, P., R. González-Sanz, M. J. Bleda, G. Hernández, A. Echeíta, and M. A. Usera (2006). "Antimicrobial resistance in non-typhoidal *Salmonella* from human sources, Spain, 2001-2003". *Journal of Antimicrobial Chemotherapy* 58.2, pp. 310–314. DOI: 10.1093/jac/dkl223 (p. 43).

Struelens, M. J., European Study Group on Epidemiological Markers (ESGEM), and European Society for Clinical Microbiology and Infectious Diseases (ESCMID) (1996). "Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems". *Clinical Microbiology and Infection* 2.1, pp. 2–11 (pp. 35, 89, 90).

Stucki, D., B. Malla, S. Hostettler, T. Huna, J. Feldmann, D. Yeboah-Manu, S. Borrell, L. Fenner, I. Comas, M. Coscollà, and S. Gagneux (2012). "Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages". *PLoS ONE* 7.7, e41253. DOI: 10.1371/journal.pone.0041253 (pp. 19, 51, 53, 58, 60, 61, 63, 66, 67, 74).

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0". *Molecular Biology and Evolution* 30.12, pp. 2725–2729. DOI: 10.1093/molbev/mst197 (p. 103).

Technical University of Denmark (DTU) (2015). *Center for Genomic Epidemiology*. URL: http://www.genomicepidemiology.org/ (visited on 06/26/2015) (pp. 97, 98, 102, 103, 134, 140).

Tettelin, H., D. Riley, C. Cattuto, and D. Medini (2008). "Comparative genomics: the bacterial pan-genome". *Current Opinion in Microbiology* 12.5, pp. 472–477. DOI: 10.1016/j.mib.2008.09.006 (p. 23).

Theelen, W., E.-J. M. Speel, M. Herfs, M. Reijans, G. Simons, E. V. Meulemans, M. M. Baldewijns, F. C. S. Ramaekers, J. Somja, P. Delvenne, and A. H. N. Hopman (2010). "Increase in viral load, viral integration, and gain of telomerase genes during uterine cervical carcinogenesis can be simultaneously assessed by the HPV 16/18 MLPA-assay". *American Journal of Pathology* 177.4, pp. 2022–2033. DOI: 10.2353/ajpath.2010.090901 (pp. 51, 74).

Thermo Fisher Scientific (2015). *Gram negative plates*. URL: http://www.thermoscientific.com/en/products/gram-negative-plates.html (visited on 09/03/2015) (p. 12).

Thierry, S., R. A. Hamidjaja, G. Girault, C. Löfström, R. Ruuls, and D. Sylviane (2013). "A multiplex bead-based suspension array assay for interrogation of phylogenetically informative single nucleotide polymorphisms for *Bacillus anthracis*". *Journal of Microbiological Methods* 95.3, pp. 357–365. DOI: 10.1016/j.mimet.2013.10.004 (pp. 19, 51, 53, 58, 60, 61, 63, 65, 66, 74).

Threlfall, E. J. and J. A. Frost (1990). "The identification, typing and fingerprinting of *Salmonella*: laboratory aspects and epidemiological applications". *Journal of Applied Bacteriology* 68.1, pp. 5–16 (pp. 6, 33, 75).

Torpdahl, M., G. Sørensen, B. Lindstedt, and E. M. Nielsen (2007). "Tandem repeat analysis for surveillance of human *Salmonella* Typhimurium infections". *Emerging Infectious Diseases* 13.3, pp. 388–395. DOI: 10.3201/eid1303.060460 (pp. 31, 43, 45, 46).

U.S. Food and Drug Administration (FDA) (2015a). *Bacteriological Analytical Manual*. URL: http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm2006949.htm (visited on 09/03/2015) (p. 3).

U.S. Food and Drug Administration (FDA) (2015b). *GenomeTrakr Network*. URL: http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm363134.htm (visited on 06/26/2015) (p. 149).

van Belkum, A., P. Tassios, L. Dijkshoorn, S. Haeggman, B. Cookson, N. Fry, V. Fussing, J. Green, E. Feil, P. Gerner-Smidt, S. Brisse, M. Struelens, and for the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group on Epidemiological Markers (ESGEM) (2007). "Guidelines for the validation and application of typing methods for use

in bacterial epidemiology". *Clinical Microbiology and Infection* 13.Suppl 3, pp. 1–46. DOI: 10.1111/j.1469-0691.2007.01786.x (pp. 89–91).

Van den Bulcke, M. H. G., A. P. N. R. Lievens, A. Leunda, E. G. MbongoloMbella, E. Barbau-Piednoir, and M. J. S. Sneyers (2008). "Transgenic plant event detection". WO 2008/092866 (Brussels, Belgium). Scientific Institute of Public Health (WIV-ISP) (p. 81).

Van den Bulcke, M., A. Lievens, E. Barbau-Piednoir, G. MbongoloMbella, N. Roosens, M. Sneyers, and A. Leunda Casi (2010). "A theoretical introduction to "Combinatory SYBRGreen qPCR Screening", a matrix-based approach for the detection of materials derived from genetically modified plants". *Analytical and Bioanalytical Chemistry* 396.6, pp. 2113–2123. DOI: 10.1007/s00216-009-3286-7 (p. 81).

Wattiau, P., C. Boland, and S. Bertrand (2011). "Methodologies for *Salmonella enterica* subsp. *enterica* subtyping: gold standards and alternatives". *Applied and Environmental Microbiology* 77.22, pp. 7877–7885. DOI: 10.1128/AEM.05527-11 (pp. 2, 6, 8, 31, 73, 132).

Weill, F.-X., L. Fabre-Berland, V. Guibert, L. Diancourt, and S. Brisse (2008). "Molecular typing and subtyping of *Salmonella* by identification of the variable nucleotide sequences of the CRISPR loci". WO 2008/IB04004 (p. 9).

Weill, F.-X., F. Guesnier, V. Guibert, M. Timinouni, M. Demartin, L. Polomack, and P. A. D. Grimont (2006). "Multidrug resistance in *Salmonella enterica* serotype Typhimurium from humans in France (1993 to 2003)". *Journal of Clinical Microbiology* 44.3, pp. 700–708. DOI: 10.1128/JCM.44.3.700-708.2006 (p. 43).

Witonski, D., R. Stefanova, A. Ranganathan, G. E. Schutze, K. D. Eisenach, and M. D. Cave (2006). "Variable-number tandem repeats that are useful in genotyping isolates of *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Newport". *Journal of Clinical Microbiology* 44.11, pp. 3849–3854. DOI: 10.1128/JCM.00469-06 (pp. 31, 47).

Wuyts, V., S. Denayer, N. H. C. Roosens, W. Mattheus, S. Bertrand, K. Marchal, K. Dierick, and S. C. J. De Keersmaecker (2015a). "Whole genome sequence analysis of *Salmonella* Enteritidis PT4 outbreaks from a National Reference Laboratory's viewpoint". *PLOS Currents Outbreaks*. DOI: 10.1371/currents.outbreaks.aa5372d90826e6cb0136ff66bb7a62fc (p. 129).

Wuyts, V., W. Mattheus, G. De Laminne de Bex, C. Wildemauwe, N. H. C. Roosens, K. Marchal, S. C. J. De Keersmaecker, and S. Bertrand (2013). "MLVA as a tool for public health surveillance of human *Salmonella* Typhimurium: prospective study in Belgium and evaluation of MLVA loci stability". *PLoS ONE* 8.12, e84055. DOI: 10.1371/journal.pone.0084055 (pp. 6, 29, 99).

Wuyts, V., W. Mattheus, N. H. C. Roosens, K. Marchal, S. Bertrand, and S. C. J. De Keersmaecker (2015b). "A multiplex oligonucleotide ligation-PCR as a

complementary tool for subtyping of *Salmonella* Typhimurium". *Applied Microbiology and Biotechnology* 99.19, pp. 8137–8149. DOI: 10.1007/s00253-015-6831-7 (pp. 69, 71, 99).

Wuyts, V., N. H. C. Roosens, S. Bertrand, K. Marchal, and S. C. J. De Keersmaecker (2015c). "Guidelines for optimisation of a multiplex oligonucleotide ligation-PCR for characterisation of microbial pathogens in a microsphere suspension array". *BioMed Research International*, article ID 790170. DOI: 10.1155/2015/790170 (pp. 49, 77).

Xu, Z., W. Chen, D. P. Merke, and N. B. McDonnell (2013). "Comprehensive mutation analysis of the *CYP21A2* gene: an efficient multistep approach to the molecular diagnosis of congenital adrenal hyperplasia". *Journal of Molecular Diagnostics* 15.6, pp. 745–753. DOI: 10.1016/j.jmoldx.2013.06.001 (pp. 51, 74).

York, M. K. and P. Rodrigues-Wong (2010). "Fecal culture for aerobic pathogens of gastroenteritis". *Clinical Microbiology Procedures Handbook*. Ed. by L. S. Garcia and H. D. Isenberg. 3th edition. Vol. 1.3. Washington, DC, USA: ASM Press. Chap. 3.8, pp. 3.8.1.1–3.8.1.21. ISBN: 978-1-55581-527-1 (p. 4).

Zankari, E., H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. Aarestrup, and M. Larsen (2012). "Identification of acquired antimicrobial resistance genes". *Journal of Antimicrobial Chemotherapy* 67.11, pp. 2640–2644. DOI: 10.1093/jac/dks261 (pp. 23, 103, 104, 135, 136, 140).

Zerbino, D. R. and E. Birney (2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs". *Genome Research* 18.5, pp. 821–829. DOI: 10.1101/gr.074492.107 (p. 104).

List of publications

Publications in internationally peer-reviewed journals

Wuyts, V., W. Mattheus, G. De Laminne de Bex, C. Wildemauwe, N. H. C. Roosens, K. Marchal, S. C. J. De Keersmaecker* and S. Bertrand* (2013). "MLVA as a tool for public health surveillance of human *Salmonella* Typhimurium: prospective study in Belgium and evaluation of MLVA loci stability". *PLoS ONE* 8.12, e84055. DOI: 10.1371/journal.pone.0084055.

Wuyts, V., N. H. C. Roosens, S. Bertrand, K. Marchal, and S. 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*, article ID 790170. DOI: 10.1155/2015/790170.

Wuyts, V., W. Mattheus, N. H. C. Roosens, K. Marchal, S. Bertrand*, and S. 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, pp. 8137–8149. DOI: 10.1007/s00253-015-6831-7.

Wuyts, V., S. Denayer, N. H. C. Roosens, W. Mattheus, S. Bertrand, K. Marchal, K. Dierick, and S. C. J. De Keersmaecker (2015). "Whole genome sequence analysis of *Salmonella* Enteritidis PT4 outbreaks from a National Reference Laboratory's viewpoint". *PLOS Currents Outbreaks*. DOI: 10.1371/currents.outbreaks.aa5372d90826e6cb0136ff66bb7a62fc

Wuyts, V., N. H. C. Roosens, S. Bertrand, K. Marchal, and S. C. J. De Keersmaecker. "Optimized MOL-PCR for characterization of microbial pathogens". *Current Protocols in Cytometry*, accepted for publication.

*Equally contributed

Publications in other international academic journals

Barbau-Piednoir*, E., S. C. J. De Keersmaecker*, V. Wuyts, C. Gau, W. Pirovano, A. Costessi, P. Philipp*, and N. H. Roosens* (2015). "Genome sequence of EU-unauthorized genetically modified *Bacillus subtilis* strain 2014-3557 overproducing riboflavin, isolated from a vitamin B2 80% feed additive". *Genome Announcements* 3.2, e00214-15. DOI: 10.1128/genomeA.00214-15.

Publications in other journals and reports

Wuyts, V., S. Bertrand, N. Roosens, K. Marchal, and S. De Keersmaecker (2014). "Development of a Luminex®-based molecular alternative for classical microbiological subtyping techniques, *Salmonella* phage typing as a case study". *WIV-ISP scientific report 2012–2013*. D/2014/2505/35. Resp. ed. J. Peeters. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP).

Wuyts, V., S. Denayer, N. H. C. Roosens, W. Mattheus, S. Bertrand, K. Marchal, K. Dierick, and S. C. J. De Keersmaecker. "The usefulness of whole genome sequencing for outbreak investigation of food pathogens, *Salmonella* Enteritidis as a case study". *Labinfo*, submitted.

Communications at international conferences, published as proceedings

Wuyts, V., S. Bertrand, N. Roosens, K. Marchal, and S. De Keersmaecker (2013). "Molecular subtyping of *Salmonella* Typhimurium combining different types of markers in a multiplex liquid bead suspension array". *Proceedings of the International Symposium Salmonella and Salmonellosis I3S*. International Symposium *Salmonella* and Salmonellosis I3S. (Saint-Malo, France, May 27–29, 2013). Ed. by P. Colin and G. Clément. pp. 80–84. Selected for oral presentation.

Abstracts of posters presented at national and international conferences and workshops

Wuyts, V., N. Roosens, S. Bertrand, K. Marchal, and S. De Keersmaecker (2012). "Design of multiplex assays for molecular subtyping of pathogens with the Luminex xMAP technology". *Posttranscriptional regulation and epigenetics in microorganisms*. Contact forum organised by the National Committee for

Microbiology and the Belgian Society for Microbiology. Brussels, Belgium, November 30, 2012. Book of abstracts, pp. 90.

Wuyts, V., S. Bertrand, N. Roosens, K. Marchal, and S. De Keersmaecker (2013). "Development of a molecular subtyping method for *Salmonella* Typhimurium combining different types of markers in a Luminex xTAG assay". Luminex xSAMPLES Benelux. Rotterdam, The Netherlands, June 19–20, 2013.

Wuyts, V., S. Bertrand, W. Mattheus, N. H. C. Roosens, K. Marchal, and S. C. J. De Keersmaecker (2014). "Evaluation of whole genome sequencing for routine subtyping of *Salmonella* Typhimurium for routine surveillance and outbreak investigation". *Applications of Next Generation Sequencing for public health.* Workshop organised by the Scientific Institute of Public Health (WIV-ISP). Brussels, Belgium, October 10, 2014.

Wuyts, V., S. Bertrand, W. Mattheus, N. H. C. Roosens, K. Marchal, and S. C. J. De Keersmaecker (2015). "Evaluation of whole genome sequencing for routine subtyping of *Salmonella* Typhimurium for routine surveillance and outbreak investigation". *Revolutionizing Next-Generation Sequencing: Tools and Technologies.* VIB conference. Leuven, Belgium, January 15–16, 2015.

Wuyts, V., S. Denayer, N. H. C. Roosens, W. Mattheus, S. Bertrand, K. Marchal, K. Dierick, and S. C. J. De Keersmaecker (2015). "User-friendly WGS analysis of *Salmonella* Enteritidis PT4 outbreaks". *Twentieth Conference on Food Microbiology.* Belgian Society for Food Microbiology. Brussels, Belgium, October 8–9, 2015. Selected for oral presentation.

Internal reports

Wuyts, V. (2012). "Technical fiche – ESCO-1: Swift MaxPro thermal cycler". Version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 3 pp.

Wuyts, V. (2012). "Technical fiche – Ultrasoon-1: Ultrasonic cleaner". Version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 3 pp.

Wuyts, V. (2013). "Technical fiche – Thermoshaker-1: Thermoshaker PCMT". Version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 3 pp.

Vandermassen, E., and V. Wuyts (2014). "Standard operating procedure – Management of oligonucleotides". SOP 44/12/E version 6. Verified by S.

Broeders and S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 5 pp.

Wuyts, V. (2014). "Standard operating procedure – Purification of a PCR product". SOP 44/2022/E version 3. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 8 pp.

Wuyts, V. (2014). "Technical fiche – ESCO-2 & ESCO-3: Swift Aeris thermal cycler". Version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 4 pp.

Van Geel, D., and V. Wuyts (2014). "Technical fiche – MAGPIX-1: MAGPIX". Version 6. Verified by S. De Keersmaecker and E. Vandermassen, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 14 pp.

Van Geel, D., and V. Wuyts (2015). "Standard operating procedure – Orders". SOP 44/01/E version 5. Verified by S. Broeders and S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 14 pp.

Lefèvre, L., and V. Wuyts (2015). "Standard operating procedure – Use of ethidium bromide in PBB". SOP 44/2024/E version 2. Verified by M.-A. Fraiture, M. Delvoye and S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 4 pp.

Wuyts, V. (2015). "Standard operating procedure – DNA isolation through heat lysis of a single colony of *Salmonella enterica* subsp. *enterica*". SOP 44/2028/E version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 5 pp.

Wuyts, V. (2015). "Standard operating procedure – MOL-PCR for subtyping of *Salmonella enterica* subsp. *enterica* serovar Typhimurium and serovar 1,4,[5],12:i:-". SOP 44/2029/E version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 20 pp.

Wuyts, V. (2015). "Standard operating procedure – Hybridisation of MOL-PCR products to MagPlex-TAG microspheres and analysis on MAGPIX". SOP 44/2030/E version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 21 pp.

Wuyts, V., and D. Van Geel (2015). "Standard operating procedure – MAGPIX: Use and development of assays". SOP 44/A-19/E version 1. Verified by S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute

of Public Health (WIV-ISP). 38 pp.

Marichal, A., and V. Wuyts (2015). "Technical fiche – Desiccator-1: Desiccator". Version 1. Verified by S. Broeders and S. De Keersmaecker, approved by N. Roosens. Brussels, Belgium: Scientific Institute of Public Health (WIV-ISP). 3 pp.

Curriculum vitae

Véronique Wuyts was born on March 24th, 1980 in Lier, Belgium. In 1998 she enrolled at the KU Leuven where she obtained a bachelor's degree in Chemistry in 2000, followed by a master's degree in Biochemistry in 2002 and a degree in Complementary Studies in Business Economics in 2003. Since then she worked as a stability officer in the pharmaceutical company Schering-Plough Labo NV in Heist-op-den-Berg, Belgium, until she quitted her job in 2008 to take on the bachelor's programme in mathematics. After one year and again used to study university courses, she switched programmes to receive a master's degree in Bioinformatics in 2011. In October 2011 she started on her PhD project SalMolType at the Scientific Institute of Public Health (WIV-ISP) under daily supervision of Dr. Sigrid De Keersmaecker and with support of Prof. Kathleen Marchal and Dr. Sophie Bertrand as additional promotors of this PhD. Later on in the project, Prof. Jos Vanderleyden reinforced the promotor team.

FACULTY OF BIOSCIENCE ENGINEERING
DEPARTMENT OF MICROBIAL AND MOLECULAR SYSTEMS
CENTRE OF MICROBIAL AND PLANT GENETICS
Kasteelpark Arenberg 20 box 2460
B-3001 HEVERLEE, BELGIUM
veronique.wuyts@biw.kuleuven.be
www.biw.kuleuven.be