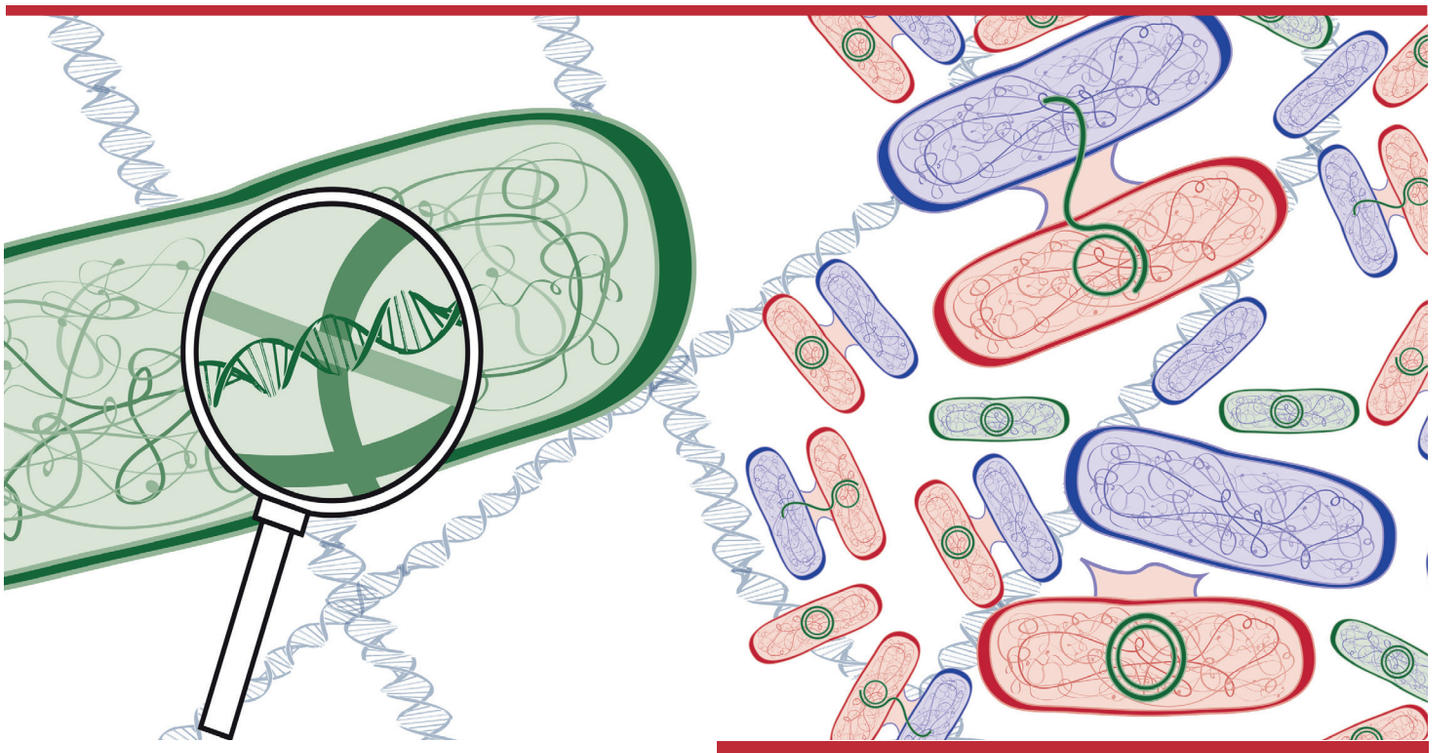
Recipient Determinants Affecting Conjugational Promiscuity in *Enterobacteriaceae*



Louise Roer
PhD Thesis
2016

DTU Food
National Food Institute

Recipient Determinants Affecting Conjugational Promiscuity in *Enterobacteriaceae*

PhD Thesis
by
Louise Roer
December 2015

Submitted: 31.12.2015
Research Group for Genomic Epidemiology
National Food Institute
Technical University of Denmark
Søltofts Plads 221, 2800 Kongens Lyngby, Denmark

PREFACE

The research of this PhD thesis was carried out from July 2012 to December 2015, in the Research Group for Genomic Epidemiology (former Division for Epidemiology and Microbial Genomics) at the National Food Institute, Technical University of Denmark. Part of the research was carried out at the Section of Microbiology, Department of Biology, Faculty of Science, University of Copenhagen. The research was supported by the grant 11-106571 from the Danish Council for Independent Research/Technology and Production Sciences.

SUPERVISORS

Professor, PhD, Frank Møller Aarestrup, Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Denmark

Senior Scientist, PhD, Henrik Hasman, Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Denmark

ASSESSMENT COMMITTEE

Professor John Wain, Norwich Medical School, University of East Anglia, United Kingdom

Associate Professor Lars Bogø Jensen, Research Group for Microbial Food Safety and Quality, National Food Institute, Technical University of Denmark, Denmark

Professor Lars Hestbjerg Hansen, Head of section for Environmental Microbiology and Biotechnology, Aarhus University, Denmark

ACKNOWLEDGMENTS

A wise man once told me to follow my dreams. As a total cliché I've always wanted to make a difference. Along this incredible journey, sometimes tough but mostly fantastic, I found my place. I'm sure that my insatiable curiosity and my eager to learn will guide me forward from this point on.

However, I would never have managed to pull this off without a bunch of people! First of all, I would like to thank my main supervisor Frank Møller Aarestrup for offering me the chance to do a PhD, and for giving me the opportunity to be a part of the research group. Thank you for the guidance through challenging times, your ever ambitious and inspirational ideas, and for all the constructive and educational discussions.

A special thanks to my co-supervisor Henrik Hasman for always taking the time to guide me in the right directions, or simply just answering whatever question that I might have. Thank you for motivating me and keeping me going in some of the though periods during my PhD, it has meant a lot to me. I would like to thank Rene S. Hendriksen for being a great mentor. It has been a pleasure working with you, and I am very grateful for your interest in my project and for you always encouraging me. A great thank to Sünje J. Pamp, for always taking the time, both for constructive discussion of projects as well as for personal conversations during tough times the past months.

Thanks to Søren J. Sørensen and his research group 'Section for Microbiology' at KU, for welcoming me during my external stay. Thanks to my external supervisor Leise Riber for help and guidance and to Karin Vestberg for laboratory guidance during my stay. A special thanks to Stefan M. Milani for spending hours and hours with me at the flow cytometer.

Big thanks to all my colleagues at DTU, to Hanne Mordhorst and Kate Vibefeldt for helping me out in the laboratory, and to Vibeke Hammer for all your administrative help.

Thanks to Katrine Joensen, Gosia (Malgozata Anna Ligowska-Marzeta) and Ea Zankari for being great officemates, always ready for listening, laughing, talking or scientific discussions. Thanks to all the PhD students and Postdocs, old as new, it has been great working with you. Thanks to the three bioinformaticians (or bioinfor-magicians); Rolf S. Kaas, Oksana Lukjancenko, and Shinny (Pimlapas Leekitcharoenphon) for helping me out and contributing to my bioinformatic skills. I can always count on you guys, whenever I need you to do your magic.

Thanks to Berith Knudsen and Marie S. Jensen for all the great fun we had.

Thanks to my family and friends. Lisbeth Andersen the lab has never been the same without you. Thank you for being such a great friend, ready with support whenever I need it! Thanks to my mom and grandma for being supportive, for actually trying to understand what I am doing, and for being proud of me no matter what.

Above all, I would like to thank my amazing husband Kenneth Roer. You have been the greatest support during this process. Thank you for putting up with me, for believing in me and making me feel that anything is possible. Thank you!

This thesis is dedicated to those who never stop believing in you and encourage you to go follow your dreams.

Skovlunde, December 2015
Louise Roer

TABLE OF CONTENTS

RESEARCH ARTICLES INCLUDED IN THE THESIS

I. **Roer L**, Aarestrup FM, Hasman H. The *Eco*KI Type I Restriction-Modification System in *Escherichia coli* Affects but Is Not an Absolute Barrier for Conjugation. J Bacteriol 197: 337–342. Available: http://www.ncbi.nlm.nih.gov/pubmed/25384481.

II. **Roer L**, Hendriksen RS, Leekitcharoenphon P, Lukjancenko O, Kaas RS, Hasman H, Aarestrup FM. Is the evolution of *Salmonella enterica* subspecies *enterica* Linked to Restriction-Modification Systems? *Manuscript for submission to: mSystems*.

III. **Roer L**, Hasman H, Pamp SJ, Kaas RS, Riber L, Sørensen SJ, Aarestrup FM. *Salmonella enterica* recipient genes affecting plasmid conjugation. *Manuscript in preparation.*

ARTICLES NOT INCLUDED IN THE THESIS

1. Hendriksen RS, Leekitcharoenphon P, Mikoleit M, Jensen JD, Kaas RS, **Roer L**, Joshi HB, Pornruangmong S, Pulsrikarn C, Gonzalez-Aviles GD, Reuland EA, Al Naiemi N, Wester AL, Aarestrup FM, Hasman H. 2015. Genomic dissection of travel-associated extended-spectrum-beta-lactamase-producing Salmonella enterica serovar typhi isolates originating from the Philippines: a one-off occurrence or a threat to effective treatment of typhoid fever? J. Clin. Microbiol. 53:677–80.
Available: http://www.ncbi.nlm.nih.gov/pubmed/25428145

SUMMARY

The global emergence and rapid spread of antimicrobial resistance are considered major threats to human health. Horizontal transfer of plasmids is the most important method for transmission of resistance, a method which with the accumulation of sequence data in recent years, was shown to be highly underestimated. Thus, to be able to control and prevent the spread of undesirable genetic traits, it is crucial to understand the factors controlling the conjugational transfer.

With the great attention to whole genome sequencing it is now possible to rapidly predict a variety of different characteristics of pathogens, including antibiotic resistance, plasmid replicons, and the evolutionary relatedness of bacteria. Thus, combinations of the bioinformatic technology and the classical molecular approaches in the laboratory provide new opportunities for research nowadays.

The aim of this PhD study was to improve the knowledge for the implication of restriction-modification systems in horizontal gene transfer and evolution, and to gain new knowledge of recipient determinants included in regulation of plasmid uptake in bacteria by using a combination of bioinformatics and applied microbiology in the laboratory.

In **Roer I** we addressed the dogma of RM systems being a barrier for horizontal gene transfer, which, to our knowledge never has been proven at a genetic level for conjugation. In this study we constructed an isogenic setup, with single gene-knockout mutants of the type I endonuclease *Eco*KI and the cognate methyltransferase M.*Eco*KI. By utilizing the plasmid pOLA52, with 2 recognitions sites for the RM system, it was possible to investigate the uptake and maintenance of an un-methylated plasmid in the wild-type strain (intact *hsdR* gene), and in a *hsdR* deficient derivate. The inactivation of the restriction enzyme caused a 7-fold increase in conjugational transfer for the un-methylated plasmid compared to the wild-type with intact *hsdR* gene. Interestingly, the RM system did not impose an absolute barrier for conjugational transfer of the wild-type. Thus, these findings are leaving a question of the evolutionary impact of RM systems on organisms which preferably exchange DNA by conjugation.

In **Roer II**, we therefore decided to elucidate on the association between RM systems and the evolution of the 221 genomes of the conjugative bacterial species *Salmonella enterica* ssp. *enterica*. In addition to evaluate the evolution, we decided to investigate the connection between RM systems and the content of some of the evolutionary drivers, plasmids and pathogenicity islands. For the purpose of identifying RM systems and Salmonella Pathogenicity Islands, *RM-Finder* and *SPI-Finder* were constructed. The detected RM systems were compared with the evolution of the

species, depicted as a network of core- and pan-genome trees. Though, from the study we were able to conclude that in comparison with natural transformable species, there was no evidence for RM systems shaping the evolution of the conjugational species *Salmonella enterica* ssp. *enterica.* In addition, no pattern between the mobile genetic elements and RM systems was observed.

In **Roer III**, we aimed to identify other recipient determinants controlling the conjugational uptake and maintenance. To identify possible gene candidates we used a combination of modern molecular techniques in the laboratory, and bioinformatic gene prediction. The attempt to identify gene candidates was performed on a collection of 93 *Salmonella enterica* ssp. *enterica* isolates, covering 54 different serovars. An initial conjugation experiment was performed to clarify the recipient potential for the 93 isolates, with subsequent grouping into good and poor recipients. From this rather diverse collection, we were not yet able to detect common gene candidates controlling neither increase nor decrease in conjugation abilities. A parallel study, only focusing on *S.* Enteritidis, proved able to detect 33 gene candidates potentially involved in conjugation restriction abilities which are being investigated and awaiting verification. This approach indicates that on species level no common gene controlling conjugation could be detected and other methods should be considered.

Overall, this PhD study has assessed the old dogma of RM systems being a barrier for horizontal gene transfer. This was previously proven in the transfer by transduction and transformation, however in this thesis we provide evidence that the imposed barriers are indefinite in conjugational transfer. Further, the work in this thesis has provided user-friendly tools for easy *in-silico* detection of RM systems (*RM-Finder*) and the transferable genomic islands conferring pathogenicity in *Salmonella* (*SPI-Finder*). It was furthermore demonstrated that, despite the clear correlation between RM systems and evolution in the natural transformable organism *N. meningitidis*, RM systems where not the main driver of evolution when turning to organisms which preferably exchange their DNA material by conjugation. Additionally, the study proved that the possession of mobile genetic elements like plasmids and SPIs cannot be reflected in the RM systems. The PhD study suggests that recipient genes are involved in the control of conjugation; however, different approaches and serovar and strain specific genes should be investigated.

This PhD thesis has altogether improved the knowledge on the influence of RM systems in conjugation and evolution, and placed a potential for further investigation of recipient determinants affecting conjugation.

RESUMÉ

Den globale udvikling og hurtige spredning af antimikrobiel resistens anses for at være en stor trussel for det menneskelige helbred. Horisontal overførsel af plasmider er en af de mest betydningsfulde mekanismer for overførsel af resistens, en metode som med akkumuleringen af sekvens data de seneste år har vist sig at være særdeles underestimeret. For at kunne kontrollere og forebygge spredningen af uønskede genetiske træk er det afgørende at forstå de faktorer der kontrollerer overførsel ved konjugation. Med den øgede opmærksomhed og udvikling indenfor hel-genom sekventering er det nu muligt hurtigt at forudsige forskellige karaktertræk hos patogene bakterier, hvilket inkluderer antibiotika resistens, plasmid replikon og det evolutionære forhold mellem bakterier. Således kan kombinationen af bioinformatik og klassiske molekylære fremgangsmåder i laboratoriet tilføre nye muligheder til vor tids forskning.

Formålet med dette ph.d.-studie var at øge den allerede eksisterende viden indenfor betydningen af RM systemer i horisontal gen overførsel og evolutionen, og at opnå ny viden indenfor recipient faktorer involveret i reguleringen af plasmid optaget i bakterier, ved at benytte en kombination af bioinformatik og anvendt mikrobiologi i laboratoriet.

I **Roer I** håndterede vi dogmet for RM systemer som værende en barriere for horisontal gen overførsel, hvilket ifølge vores viden aldrig har været vist på et genetisk niveau for konjugation. I dette studie konstruerede vi et iso-genetisk opsæt med enkelt gen knockouts af type I restriktionsenzymet *Eco*KI og dens tilhørende methyltransferase M.*Eco*KI. Ved at bruge plasmidet pOLA52, som besidder 2 genkendelses sekvenser for RM systemet, var det muligt at undersøge optaget og opretholdensen af et ikke methyleret plasmid i vild-type stammen (med intakt *hsdR* gen), og i et derivat med ødelagt *hsdR* gen. Inaktiveringen af restriktionsenzymet forårsagede en 7-folds stigning i konjugations overførsel af det ikke methylerede plasmid, sammenlignet med vild-type stammen med intakt *hsdR* gen. RM systemet viste sig overraskende nok ikke at tilføre en fuldstændig barriere for konjugation til vild-typen. Resultaterne efterlader derved spørgsmål om den evolutionære indvirkning af RM systemer på organismer som fortrinsvis udveksler DNA via konjugation.

I **Roer II** besluttede vi derfor at undersøge sammenhængen mellem RM systemer og evolutionen af 221 genomer af *Salmonella enterica* ssp. *enterica,* som benytter konjugation ved overførsel. Ud over at evaluere evolutionen besluttede vi at undersøge sammenhængen mellem RM systemer og evolutionære drivkræfter som plasmider og patogenetiske øer. Til at identificere RM systemer og

Salmonella Patogenetiske øer (SPI) udviklede vi *RM-Finder* og *SPI-Finder*. De opfangede RM systemer blev sammenlignet med evolutionen af *Salmonella*, gengivet i et netværk af core- og pan-genom træer. Dette studie gjorde det muligt at konkludere, at sammenlignet med naturligt kompetente organismer viste RM systemer sig for ikke at forme evolutionen i konjugationsorganismen *Salmonella enterica* ssp. *enterica.* Yderligere var det heller ikke muligt at observere et mønster mellem mobile genetiske elementer og RM systemer.

I **Roer III** forsøgte vi at identificere andre recipient faktorer involveret i kontrollen af konjugations overførsel. Til at identificere mulige gen-kandidater benyttede vi en kombination af moderne molekylære teknikker i laboratoriet og bioinformatisk forudsigelse af gener. Forsøget på at identificere gen-kandidater blev udført på en kollektion af 93 *Salmonella enterica* ssp. *enterica* isolater af 54 forskellige serovars. For at bestemme recipient potentialet af de 93 isolater, blev et indledende konjugations forsøg udført. Herefter blev isolaterne inddelt i grupper af gode og dårlige recipienter. Fra denne meget diverse kollektion har det endnu ikke været muligt at finde nogen fælles gen-kandidater der kontrollerede hverken op- eller nedregulering i evnen til at konjugere. *S.* Enteritidis blev undersøgt i et parallelt studie, hvor det var muligt at identificere 33 gen-kandidater der potentielt kan være involveret i nedregulering af konjugation, hvilket bliver undersøgt nærmere og afventer verificering. Med denne fremgangsmåde er det ikke er muligt at identificere gener der kontrollerer konjugation på organisme niveau, og andre metoder bør overvejes.

Alt i alt har dette ph.d.-studie behandlet det gamle dogme for RM sytemer som værende en barriere i horisontal gen overførsel. Dette har tidligere været vist i overførslen via transduktion og transformation. I denne afhandling leverer vi et bevis for at de pålagte barrierer ikke er fuldkomne i konjugativ overførsel. Yderligere har dette studie leveret brugervenlige redskaber til in-silico påvisning af RM systemer (*RM-Finder*) og de overførbare genomiske øer som tilfører patogenese i *Salmonella* (*SPI-Finder*). Det blev yderligere demonstreret, på trods af en klar sammenhæng mellem RM systemer og evolution i den naturlig overførbare organisme *N. meningitidis*, så var RM systemer ikke en drivende kraft i evolutionen af organismer der fortrinsvis udveksler DNA via konjugation. Studiet viste yderligere at besiddelsen af mobile genetiske elementer, såsom plasmider og SPIs, ikke reflekteres i RM systemerne. Dette ph.d.-studie antyder at organisme specifikke gener kan være involveret i kontrollen af konjugation, men forskellige metoder bør overvejes og serovar og stamme specifikke gener bør undersøges yderligere. Denne ph.d.-afhandling har overordnet øget den eksisterende viden om RM systemers indvirkning på konjugation og evolution, og givet anledning til yderligere undersøgelser af recipient faktorer der kan have indflydelse på konjugation.

LIST OF ABBREVIATIONS

AdoMet	-	*S*-Adenosyl-Methionine
ALU	-	Arbitrary Light Units
ATP	-	Adenosine Triphosphate
BHR	-	Broad Host Range
BLAST	-	Basic Local Alignment Search Tool
bp	-	Basepair
Cas complex	-	CRISPR-associated complex
CGE	-	Center for Genomic Epidemiology
CRISPR	-	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	-	CRISPR-RNA
DNA	-	Deoxyribonucleic Acid
DR	-	Directed Repeats
FACS	-	Fluorescent Activated Cell Sorting
FSC	-	Forward Scatter
gfp	-	Green Fluorescent Protein encoding gene
GFP	-	Green Fluorescent Protein
HGT	-	Horizontal Gene Transfer
In	-	Integron
Inc	-	Incompatibility
Indel errors	-	Insertion/Deletion Errors
IS	-	Insertion Sequence
MGE	-	Mobile Genetic Elements
mpf	-	Mating-Pair Formation
NHR	-	Narrow host range
ORF	-	Open Reading Frame
oriT	-	Origin of Transfer
oriV	-	Origin of Replication
PAIDB	-	PAthogenicity Island DataBase
PAIs	-	Pathogenicity Islands
RCR	-	Rolling-Circle Replication
REBASE	-	Restriction Enzyme database
RM	-	Restriction-Modification
RNA	-	Ribonucleic Acid
RTX-like exotoxin	-	Repeats in Toxin-like Exotoxin
SDR	-	Strand Displacement-type Replication
SNP	-	Single-Nucleotide Polymorphisms
SPIs	-	*Salmonella* Pathogenicity Islands
SSC	-	Side Scatter
T4CP	-	Type IV Coupling Protein,
T4SS	-	Type IV Secretion System
Tn	-	Transposon
TR	-	Theta-type Replication
T-strand	-	Single-stranded DNA intended for transfer
VFDB	-	Virulence Factors of Pathogenic Bacteria
WGP	-	Whole-Genome Phylogenies
WGS	-	Whole Genome Sequencing

BACKGROUND

The evolutionary pressure from the increased use of antibiotics in human and veterinary medicine has contributed to the emergence and global spread of antimicrobial resistance, which today is considered one of the major human health threats. Invention of new antimicrobial agents targeting bacterial infections might be an expensive and comprehensive process; consequently it is, as far as possible, important to protect the current agents, through improved knowledge of emergence and spread of the genetic traits.

Horizontal transmission of plasmids harbouring antimicrobial resistance genes is one of the most important methods of resistance transmission between bacterial species. However, little is known about the factors controlling the conjugational transfer of plasmids between different species and clones, and without expanding this knowledge it may be impossible to prevent the spread of genetic traits through conjugation.

It has been observed that host-specific and non-host-specific serovars of *Salmonella* isolated from the same reservoir differs in their antibiotic resistance pattern (VAN DUIJKEREN *et al.* 2002; LAUDERDALE *et al.* 2006). Additionally, preliminary experiments for *Escherichia coli* have suggested that pathogenic clonal lineages serve as better plasmid recipients than non-pathogenic clones (unpublished). Thus it is possible that the ability to promote or reject the uptake of foreign DNA, such as plasmids, is defined by the genetic differences between the bacterial linages. Possible gene candidates include restriction-modification (RM) systems which are thought to act as bacterial host barriers against foreign DNA. For RM systems it is practically common knowledge, hence a dogma that they interfere with horizontal gene transfer, however for conjugation this has not yet been confirmed by a systematic experimental approach. Thus, other recipient genes may be included in the regulation of plasmid uptake. If present, identifying these additional genes acting as barriers or enhancers in conjugation might clarify the spread of plasmid transferred resistance.

PURPOSE

The purpose of the PhD project was to elucidate the implication of restriction-modification systems in horizontal gene transfer between bacterial clones, with the focus on conjugational transfer of plasmid borne antimicrobial resistance in related clones of *Escherichia coli.* Furthermore, to establish the relationship between restriction-modification systems and evolution in *Salmonella* in

comparison with the content of mobile genetic elements like plasmids and *Salmonella* pathogenicity islands.

Finally, the purpose was to identify additional genes that might block or enhance uptake in the recipient bacteria.

RESEARCH APPROACH

The specific studies conducted during this PhD project focused on the following objectives:

1. To quantify the importance of restriction-modification systems in conjugational transfer of plasmids.
2. To investigate the influence of restriction-modification systems in long term evolution of *Salmonella enterica* ssp. *enterica.*
3. To identify and assess gene candidates from whole genome sequenced bacterial isolates, which promote or resist plasmid uptake.

Chapter 1

EVOLUTION OF MICROBIAL COMMUNITIES

The environment of microorganisms can vary considerably, thus bacteria need to be dynamic and constantly capable of adapting to new surroundings. Accumulation of these adaptive changes leads to diversity within the different species. To understand, control and fight the spread of pathogens, we need to understand the processes on how bacteria diversify and the mechanisms that control the evolution. The prokaryotic evolution is a result of three main forces: acquisition of genes, loss of genes and changes within genes, Figure 1 (FRASER-LIGGETT 2005; PALLEN and WREN 2007; USSERY *et al.* 2009). Events within these forces happen all the time, leaving evolutionary traces on bacterial genomes.

1.1 Evolution by Single-nucleotide Changes

Single-nucleotide polymorphisms (SNPs) are the simplest and smallest-scale variation within changes of genes (FRASER-LIGGETT 2005). A mutation or base substitution in a genome can be caused by two fundamental processes: incorporation of wrong nucleotides during replication or mutagenesis triggered by physical- or chemical damage or changes. Thus exposure to antimicrobial agents such as rifampicin and ciprofloxacin can cause spontaneous mutations, and in *Helicobacter pylori* frequencies at 1×10^{-8} to 2×10^{-8} per cell per division were detected, conferring rapid adaptation to exposed environments (WANG *et al.* 2001). As spontaneous mutations in general are rare it is unlikely that the same substitution will occur in multiple genomes (unless there is a strong

selective pressure to obtain this given mutation). Hence the presence of the same base SNP in two independent genomes implies that the substitution likely descend from a shared ancestor. However, exceptions do exist e.g. for genes under selective pressure, like resistance genes. Genes under selective pressure are more likely to mutate to make the bacteria capable of adapting to the changed environment, thus for such genes the same base SNP can be found in multiple independent genomes. SNP analysis has been applied to a variety of uniform pathogens with rather conserved genomes (GUTACKER *et al.* 2002; PEARSON *et al.* 2004; MONOT *et al.* 2005; TOUCHMAN *et al.* 2007), thus a problem with the more variable species have been the high proportion of mobile genetic elements (MGE) causing larger insertions, which challenges the one base substitution assumption. However, SNP analysis has proven capable of differentiating between *Escherichia coli* that had diverged for only 200 generations (SHENDURE 2005). Therefore it appears highly suitable for detection of short-term evolution such as antimicrobial adaption, rather than long-term evolution and acquisition of MGEs (ALLARD *et al.* 2012, 2013; GIERALTOWSKI *et al.* 2013; HOFFMANN *et al.* 2013).

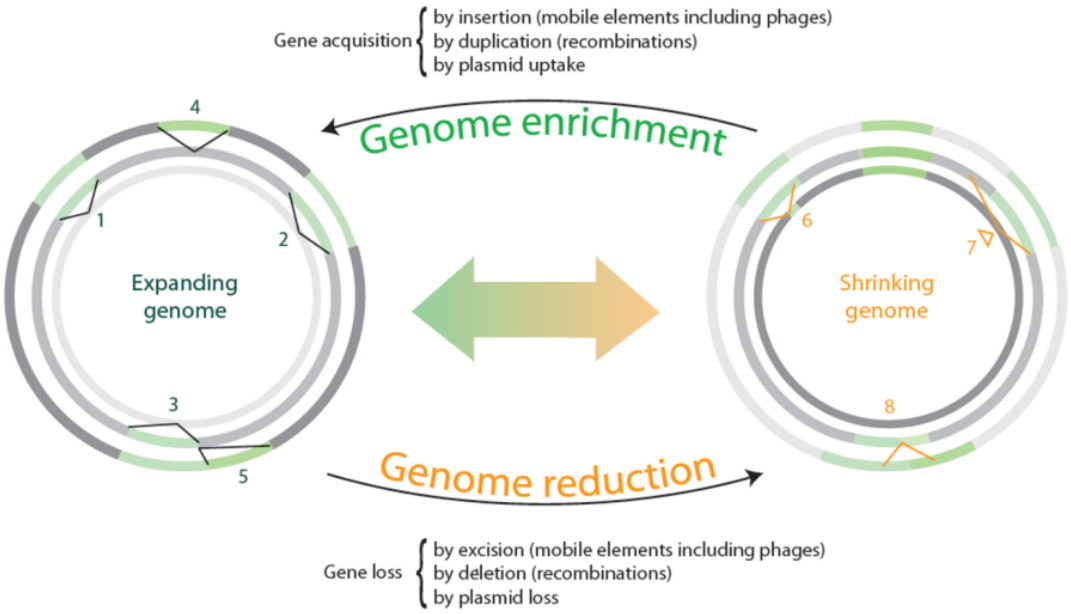


Figure 1 | Genome evolutionary events. Genome size changes due to plasmid exchange, insertion sequence elements, phages and duplication and deletions of genes. The genomes in the figure expand (left) or decrease (right) from the light gray to the dark grey backbone. In the events 1, 2 and 3 from the light gray genome (left) a piece of DNA inserts itself into the chromosome, and from the middle genome additional insertion happens in event 4 and 5, the latter located at the same position as event 3. All these insertions could be caused by an IS element, a phage or another mobile genetic element. During genome reduction (right) the inserted region from event 1 was partly removed in event 6, leaving a residue in the genome. Contrary in event 7, more DNA was deleted than initially introduced in event 2, resulting in a genome reduction in comparison with the original genome (light gray genome left). Further, in event 8 a partial deletion of the two individual insertion from event 3 and 5 was resulting in a novel junction, thus in a second step the remaining insertions was deleted. The inserted DNA from event 4 was permanently fixed in the genome. Figure 1 is adapted from (USSERY *et al.* 2009).

1.2 Evolution by Gene Loss and Deletions

As bacterial genomes remain about the same size, a balance of gene duplication or acquisition and gene loss or deletion must exist (MIRA *et al.* 2001). The most severe form of genome reduction is found in intracellular host-associated bacteria such as endosymbionts. These bacteria are segregated from the surroundings in a stable and undemanding niche, which for *Rickettsia prowazekii* has resulted in a substantial downsizing of the genome, leading to renunciation of many of the inherent genes (ANDERSSON *et al.* 1998; WERNEGREEN 2005). With gene loss and deletion of large DNA segments in the reductive evolution, it was initially thought that genes targeted for loss were only redundant genes not necessary for survival in the new host environment. However an increasing number of comparative genomic studies have suggested that a positive selection for gene loss might also be advantageous for the bacteria. In *Salmonella enterica* (*S. enterica*), different deletions appeared beneficial (KOSKINIEMI *et al.* 2012), and for *Bordetella pertussis* and *Shigella* the loss of a cell surface antigen enhanced the ability to avoid the host immune system (NAKATA *et al.* 1993; PARKHILL *et al.* 2003).

1.3 Evolution by Gene Duplication and Acquisition

Acquisition of additional genes through gene duplication can increase the number of key genes responsible for regulation, secondary metabolisms and transport. As much as 50% of larger genomes can be represented by gene paralogs (KONSTANTINIDIS and TIEDJE 2004; FRASER-LIGGETT 2005). However, acquisition of genes through horizontal gene transfer (HGT) is the most effective source of variation within the genome (PALLEN and WREN 2007), providing a possibility for rapid alterations for the constant changing demands on the bacteria (BOUCHER *et al.* 2003). Different genetic elements can be transferred by HGT, among others the virulence determinants on pathogenicity islands (PAIs) and prophages, which appear to play a major role in the evolution of pathogens. In a comparative analysis of *Staphylococcus aureus* and the closely related *Staphylococcus epidermidis*, it was established that the main cause of variation in virulence and pathogenicity was caused by the acquisition of genomic islands, presumably obtained through HGT (GILL *et al.* 2005). Another prominent source for diversification by HGT is mediated by plasmids, which can rapidly confer antimicrobial resistance or numerous of virulence factors and thereby evolve existing clonal lineages. The opportunistic pathogen *Bacillus cereus* is normally differing from the etiologic agent of anthrax, *Bacillus anthracis*, by two plasmids, toxic-encoding pXO1 and capsule gene coding plasmid pXO2. However, whole genome sequence (WGS) analysis revealed that the *B. cereus* G9241 strain was harbouring a plasmid 99.6% identical with the *B. anthracis*

plasmid pXO1, and a previously unidentified plasmid encoding polysaccharide capsule cluster, resulting in a virulent *B. cereus* strain (HOFFMASTER *et al.* 2004).

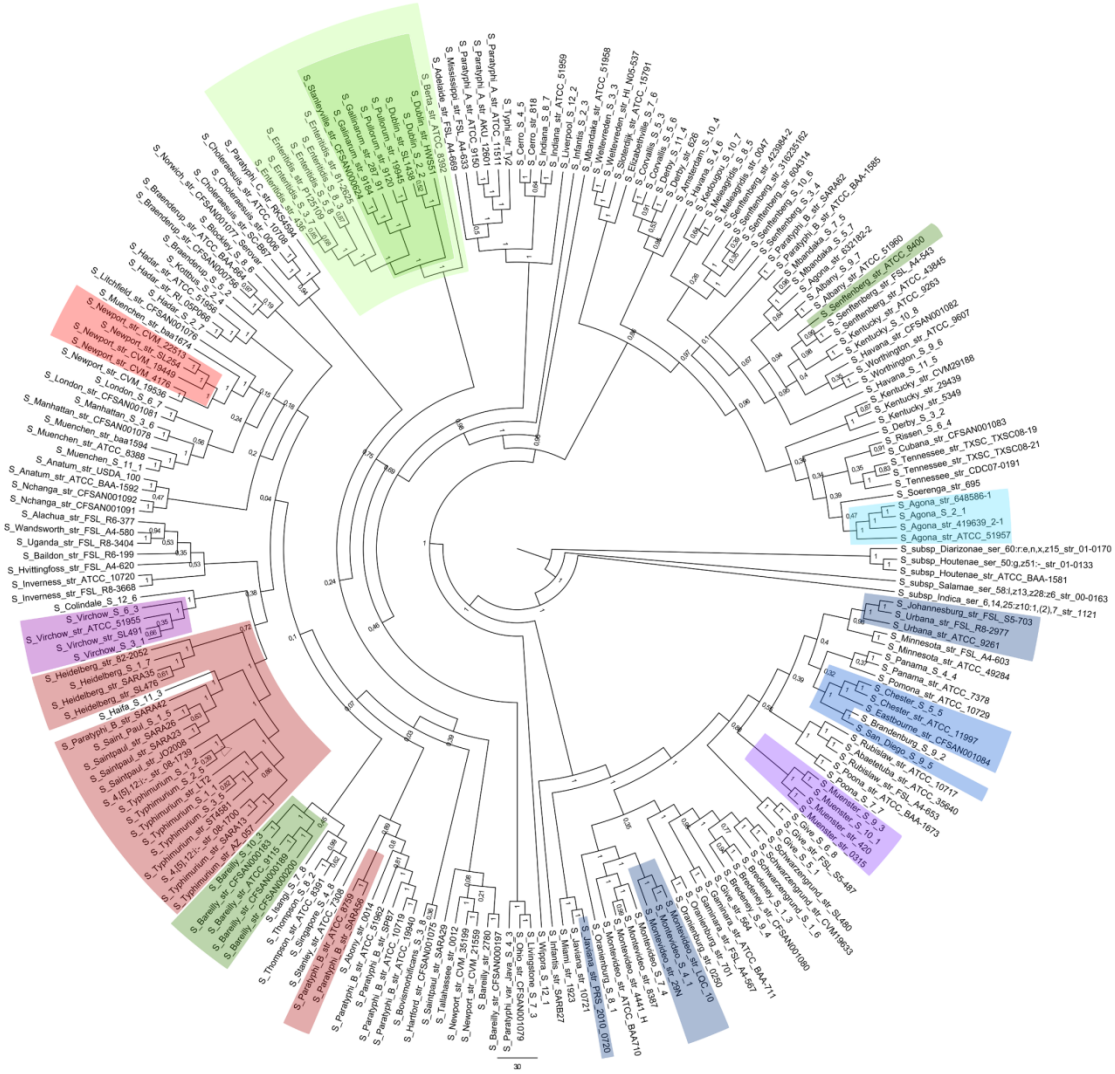


Figure 2 | Concatenated Core-genome tree of 221 *Salmonella enterica* isolates. The tree is based on 1,072 core gene clusters, with colour indications of restriction-modification systems present in the isolates. The bootstrap values are indicated at each branch. Figure 2 is adapted from (**Roer II**).

With the accumulation of prokaryotic genome sequences, the challenge of constructing whole-genome phylogenies (WGP) began. The extent and influence of HGT on the evolution, as indicated in the two examples with *Staphylococcus* and *Bacillus*, together with disagreements between WGP and rRNA phylogenies indicated that a reconstruction of the Tree of Life would be difficult or even impossible (DOOLITTLE 1999). Consequently, rather than illustrating WGP as a single tree, a consensus appeared that the evolution of prokaryotes was more appropriate depicted as a network of trees covering both horizontal and vertical transferred genes (FRASER-LIGGETT 2005). In Roer *et al.* (**Roer II**), the evolution of *S. enterica* was depicted by the network of the core- and pan-genome

trees, presented in Figure 2 and Figure 3. The pan-genome is defined as any gene family in the bacterial species, including both the core-genome and the dispensable genome. Thus, a pan-genome can either be open or closed, dependent on whether addition of new genomes to the analysis reveals new genes (open), or no new genes (closed) to the pan-genome (MEDINI *et al.* 2005; PALLEN and WREN 2007).

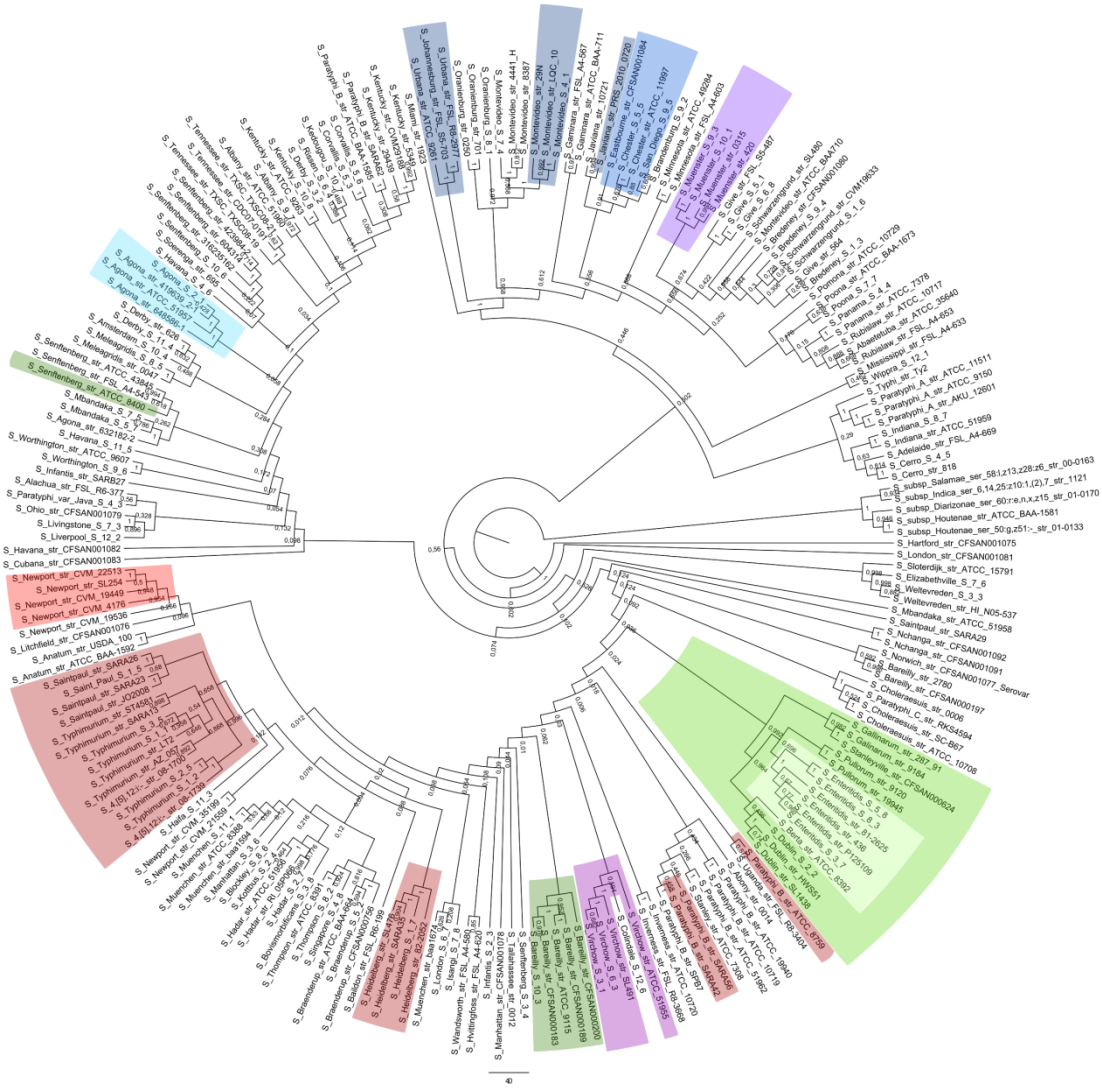


Figure 3 | Pan-genome tree of 221 *Salmonella enterica* isolates. The tree is constructed from a presence or absence matrix of gene clusters in the *Salmonella* genomes. The colours indicate the restriction-modification systems present in each isolate. The bootstrap values are shown at each branch. Figure 3 is adapted from (**Roer II**).

The core-genome is the pool of conserved genes shared by all the strains of the bacterial species, which normally are essential for bacterial growth. The dispensable genome is all the accessory genes present in some but not all strains of the bacterial species, together with the strain-specific genes (MEDINI *et al.* 2005). Further, the accessory genes are contributing to the diversity of the species and consist of genes encoding selective advantage including colonization of new hosts,

adaption to different niches or confer antibiotic resistance. The core-genome tree is constructed from an alignment of the core-genes of the strains, accounting for the vertical evolution. The pan-genome is a picture of the dispensable genes as the core-genes are present in all strains, and not contributing to the pan-genome tree. The pan-genome tree is constructed from a presence or absence matrix of the genes, where the relative Manhattan distance between the genomes is used for hierarchical clustering. The pan- and core-genome tree network is valuable in detecting variations within conserved genes but also HGT.

Chapter 2

PROMISCUITY OF BACTERIA

It is well-established that considerable horizontal gene transfer can occur between bacterial species. However, along with the accumulation of sequence data it became clear that the genetic and biochemical diversity in the prokaryotic world was highly underestimated, thus horizontal gene transfer was far more extensive in some bacterial species than previously anticipated (PERNA *et al.* 2001; FRASER-LIGGETT 2005).

Some bacterial species like *Mycobacterium tuberculosis, Mycobacterium leprae, Yersinia pestis*, and *Bacillus anthracis* were found to be genetically uniform pathogens (PALLEN and WREN 2007). In addition, *B. anthracis* was shown to have a closed pan-genome, where the addition of the fourth genome to the analysis did not add new genes to the pan-genome (TETTELIN *et al.* 2005). Furthermore, the pan-genome of nine *B. anthracis* genomes was shown to consist of 2,893 core genes, 85 accessory genes and seven strain-specific genes (ROULI *et al.* 2014). Uniform species are likely to have closed pan-genomes, thus they are more conserved, they have a low ability to acquire foreign DNA, and are normally found in niches with limited access to the global microbial gene pool (MEDINI *et al.* 2005).

In contrast to the bacterial species with closed pan-genomes, bacterial species that normally are capable of colonizing multiple environments and can exchange genetic material in multiple ways, like *Salmonella, E. coli, Streptococci, Meningococci* and *Helicobacter pylori,* are likely to have an open pan-genome, where each new genome sequence adds new genes to the overall gene pool of

the species. Thus, these bacterial species seem more promiscuous in the sense that they are able to acquire DNA outside their own species. With the completion of the genomic sequence of the pathogenic *E. coli* O157:H7 in 2001, Perna *et al.* compared the genome with the non-pathogenic laboratory *E. coli* K-12 strain and revealed 1,387 new genes encoded in strain-specific clusters of *E. coli* O157:H7. These genes encoded protein candidates of virulence factors, numerous prophages and alternative metabolic abilities. This is a practical example of how two members of the same species can differ almost 30% (PERNA *et al.* 2001), which has also been shown for bacterial species such as *H. pylori* and *Staphylococcus aureus*, both with an average of strain specific genes recorded to ~22% (SALAMA *et al.* 2000; FITZGERALD *et al.* 2001). However, these findings were observed in the years 2000 and 2001, where a limited number of sequenced genomes were available. Therefore, assessing the proportion of strain specific genes in the present time could potentially give a different result, with the ongoing exchange of genetic material.

Various studies have been assessing the pan- and core-genomes of *E. coli*, hence estimating the core-genome to be between 1,500-2,000 gene families (WILLENBROCK *et al.* 2007; CHATTOPADHYAY *et al.* 2009; TOUCHON *et al.* 2009; VIEIRA *et al.* 2011; KAAS *et al.* 2012). Additionally, the pan-genome was found in one early study to consist of 11,432 gene families when assessing 20 genomes (TOUCHON *et al.* 2009), while a study only three years later found 16,676 pan-genes when assessing 186 genomes (KAAS *et al.* 2012).

Another highly diverse species is *Salmonella enterica*, which is divided into six subspecies of which *S. enterica* subspecies *enterica* alone is grouped into more than 1,500 individual serovars. Some serovars are defined as host-specific, hence they are only found in a limited and specific number of hosts, compared to the non-host-specific serovars which can be found in a broad range of hosts. Evaluating the pan- and core-genome of *S. enterica*, one study on 73 genomes showed a core-genome of 2,882 gene families with a cognate pan-genome of 10,581 gene families (LEEKITCHAROENPHON *et al.* 2012), where an another study utilizing the same method, estimated the core-genome on 221 genomes to comprise 2,138 gene families with a cognate pan-genome of 16,375 gene families (**Roer II**). The pan-genomes of both *E. coli* and *S. enterica* appear to be open, and follow the rule that addition of more genomes to the analysis causes larger pan-genomes. However, it has been observed for *Salmonella* that host-specific serovars possess considerably less genes related to antimicrobial resistance than serovars with multiple hosts, even though the isolates were isolated from the same reservoir (VAN DUIJKEREN *et al.* 2002; LAUDERDALE *et al.* 2006). This could indicate, that the difference between host- and non-host-specific *S. enterica* could explain

their ability to acquire new genes, as the need for host-specific serovars to adapt to new surroundings is smaller since they are limited to only a few hosts. Thus, some bacterial species and serovars seem to be capable of diversifying by acquiring new genetic material through plasmids or other MGEs more easily than others. The case with *S. enterica* indicates a segregation of promiscuity is possible within a single but yet diverse species.

2.1 Plasmids

Plasmids play a major role in the exchange of genetic material between bacteria, thereby enhancing the genetic diversity and the ability for bacteria to adapt to changing environments.
The term plasmid was suggested by Joshua Lederberg in 1952, as a common phrase for all extrachromosomal hereditary determinants, both genetically simple or complex (LEDERBERG 1952). However, now plasmids are described as either linear of circular double-stranded DNA elements capable of autonomous replication in their host, and having a genome size ranging from approximately 750 bp to 2.58 Mbp (SHINTANI *et al.* 2015). Plasmids can be classified by several different criteria, including their transmissibility, incompatibility groups, copy number and host range. Some plasmids are thought of as selfish DNA elements, as they are self-transmissible, and can replicate and survive within a host. However, using these criteria, not all plasmids are to be considered as selfish; in a comparison of 1,730 fully sequenced plasmids, only 15% of the plasmids were predicted to be self-transmissible, compared with 24% mobilizable plasmids which need help to be transmitted and 61% non-transmissible plasmids (SMILLIE *et al.* 2010). In 2015, Shintani *et al.* classified 4,602 complete plasmid genomes by their host (data collected August 2014). They found that 137 of the plasmid genomes were from archaea, 47 were from eukaryotes, and the remaining 4,418 plasmids were from bacteria (SHINTANI *et al.* 2015).

The best studied group of self-transmissible plasmids are the IncP-1 plasmids, which initially were found in clinical bacterial isolates (JOBANPUTRA and DATTA 1974; JACOBY *et al.* 1976). They are still of clinical concern due to their ability to carry and facilitate the spread of antibiotic resistance (INGRAM *et al.* 1973; NOVAIS *et al.* 2006). Additionally, IncP plasmids are present in diverse environments like manure (BINH *et al.* 2008), water streams (SMALLA *et al.* 2006), waste-water treatment plants (SCHLÜTER *et al.* 2007), and agricultural soils (TOP *et al.* 1995; SEN *et al.* 2011), which is also the case for the IncP-1 plasmid pKJK5 isolated from a soil environment (BAHL *et al.* 2007). A schematic map of a typical IncP-1 plasmid is depicted in Figure 4. The plasmids consist of two regions; a ‘plasmid backbone’ and the accessory genes. The plasmid backbone includes genes

responsible for replication (*trfA*), transfer and maintenance (*tra* and *ctl*), whereas the accessory regions encode genes that can increase the fitness to the host such as antimicrobial or metal resistance and alternative metabolic pathways (SCHUMANN *et al.* 2008). For the IncP-1 plasmids, the accessory genes are typically found in two different areas, the first located between the *trb* operon that encode mating-pair formation, and the *tra* operon that encode genes responsible for plasmid transfer. Tthe second region is located between the origin of replication (*oriV*) and *trfA* responsible for plasmid replication (SCHUMANN *et al.* 2008; SEN *et al.* 2013). The accessory regions can carry other MGEs such as integrons (Ins), insertion sequences (ISs) and transposons (Tns) that enables the accessory genes to move between the bacterial chromosome and plasmids, contributing to the bacterial evolution (SCHLÜTER *et al.* 2007; SCHUMANN *et al.* 2008). However, plasmids that do not enhance beneficial traits to their hosts also exists, and are named cryptic plasmids (VAN ELSAS *et al.* 1998).

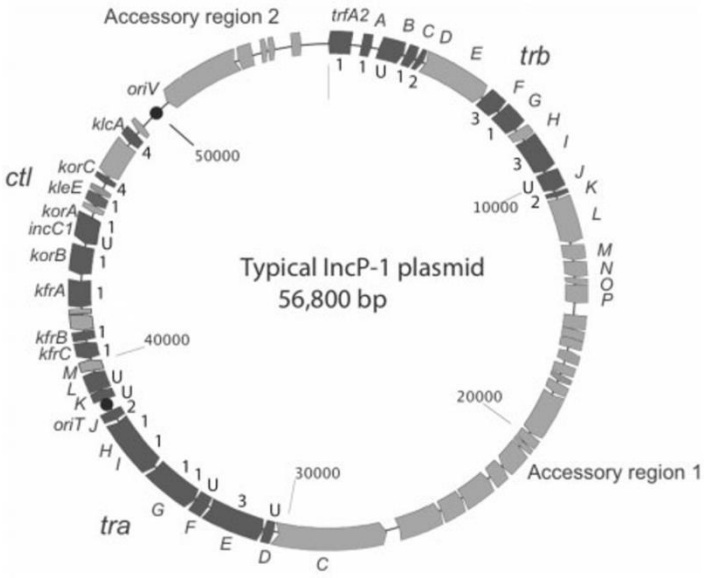


Figure 4 | Genetic map of a typical IncP-1 plasmid. The map displays the different functional elements located on a typical IncP-1 plasmid. Backbone genes are coloured in dark grey, with accessory genes displayed in light grey. The backbone of the plasmid is composed of an origin of replication (*oriV*) and an operon for replication initiation (*trfA*). The *trb* operon is involved in bridge formation and the *tra* operon in DNA processing, both during conjugation. The central control region (*ctl*) comprises regulatory genes involved in maintenance and plasmid stability, whereas the accessory regions can consist of host-beneficial genes, whit potential traits. Figure 4 is adapted from (SEN *et al.* 2013)

2.1.1 Replication

One of the main reasons for the success of plasmids is their ability to self-replicate within their host. In plasmids, the three different well-known mechanisms for replication of circular plasmids includes rolling-circle replication, theta-type replication, and strand displacement-type replication (DEL SOLAR *et al.* 1998).

Rolling-circle replication (RCR) plasmids are found in both Gram-positive and Gram-negative bacteria as well as archaea. RCR is named due to the apparent rolling movement during replication. Plasmids replicating by RCR are generally very small and compact (less than 10 kb) and found in multiple copies (KHAN 2005; SCHUMANN *et al.* 2008). Additionally, RCR plasmids include both conjugative and mobilizable plasmids, often carrying antibiotic or metal resistance genes, however some of the plasmids are cryptic (KHAN 2005). The Rep region of plasmids encodes proteins that are involved in the initiation, elongation and termination of replication (DEL SOLAR *et al.* 1998).

The large plasmids are either replicating through the theta-, or strand displacement-type replication mechanisms (MEIJER *et al.* 1995; GUGLIELMETTI *et al.* 2007). The theta-type replication (TR) is named by the molecule shape of the replication intermediate, which is visible under electron microscopy. The replication initiates with melting of the parental strands and primer RNA synthesis, followed by extension of the primers. In TR the leading strand is synthesized continuously, and the lagging strand discontinuously (DEL SOLAR *et al.* 1998). In contrast, strand displacement-type replication (SDR) synthesizes the two strands bidirectionally. Replication by SDR is initiated by the combination of three proteins, further this mechanism has been associated with mobilizable plasmids of the IncQ family (SAKAI and KOMANO 1996; DEL SOLAR *et al.* 1998).

Besides the size difference between RCR and the two other methods, a substantial difference is the single-stranded DNA intermediate of RCR plasmids (DEL SOLAR *et al.* 1998; KHAN 2005), providing plasmid instability of plasmids replicated by rolling-circle.

Following successful replication of the plasmids, they will segregate to the daughter cells by vertical transfer and establish in the new host cell.

2.1.2 Host Range of Plasmids

The plasmid host range is generally described as the collection of hosts in which a plasmid can replicate. Thus this host range is usually smaller than the transfer range (the range of hosts the plasmid can be transferred to by conjugation), though bigger than the range of host in which the plasmid can be maintained stably without selection (DE GELDER *et al.* 2007). Plasmids can either have a narrow or a broad host range. Narrow host range (NHR) plasmids can be limited in their transfer process, by unsuccessful formation of mating pairs, or in difficulties with plasmid maintenance caused by incorrect expression of the plasmid replication system in the recipient (THOMAS and NIELSEN 2005). Opposite NHR, Broad Host Range (BHR) plasmids are capable of transferring to and replicating within a comprehensive range of hosts. Though, different definitions

of BHR have been utilized, with some researchers defining BHR as plasmids capable of transferring between different genera, whereas others define BHR as those plasmids that can transfer and replicate within at least two subgroups of a bacterial species (SZPIRER *et al.* 1999). As the transfer range of some BHR plasmids are wider than the replication range, the combination of conjugative transfer followed by integration into the recipient chromosome enables the ability to spread genetic material on a broad taxonomic range. Additionally, mobilization of non-conjugative plasmids with a broader host range increases the spread. For the self-transmittable (conjugative) plasmids, the group of IncP-1 are considered those with the broadest host range (THOMAS 1987; ADAMCZYK and JAGURA-BURDZY 2003; SCHLÜTER *et al.* 2007). Klümper *et al.* showed that the IncP-1 plasmid pKJK5 possessed the ability to transfer to a surprisingly diverse group of recipients extracted from soil, thus showing a rather diverse host range (KLÜMPER *et al.* 2014). However transferring pKJK5 to various *S. enterica* recipients showed differences in the ability to take up and maintain the plasmid (**Roer III**).

Small mobilizable (non-self-transmissible) plasmids, like the IncQ plasmid RSF1010, have showed tendencies of broader host range than large plasmids with the ability to conjugate (GUERRY *et al.* 1974). However, the mobilizable plasmids are deeply dependent on the conjugative plasmids to transfer between hosts.

2.1.3 Incompatibility

Incompatibility has not only been used to classify plasmids since the 1970s (SHINTANI *et al.* 2015), but it still has huge impact on the ability of plasmids to be maintained in a new host. Incompatibility (Inc) is defined as the inability for two plasmids to coexist in the same host without a selective pressure. The failure to coexist can be caused by similarity in replication and partition systems. The partition system is covering the active process that controls a consistent distribution of low-copy-number plasmids to daughter cells (BOUET *et al.* 2007). Hence, if a potential new host already possess an analogous replication mechanism, the new host is unable to distinguish the two plasmids and neither will be replicated frequently enough to be stably maintained (NOVICK 1987; BOUET *et al.* 2007). This will in most cases lead to a loss of the newly transferred plasmid, as the initial copy number is lower.

A PCR-based method was previously used to classify the replicon types within *Enterobacteriaceae*. The limitations to identifying novel and divergent replicons resulted in a more accurate classification based on the full-length sequence (CARATTOLI 2009), enabling grouping of

unidentified Inc groups (SHINTANI *et al.* 2015). Online WGS tools like *PlasmidFinder* is now available to classify the replicon type based on sequencing data (CARATTOLI *et al.* 2014).

In a review by Shintani *et al.* from 2015, the current number of Inc groups was reported. For *Enterobacteriaceae* 27 different Inc groups were described, 14 Inc groups for *Pseudomonas* and about 18 Inc groups for *Staphylococcus* (SHINTANI *et al.* 2015), indicating a broad member of MGEs.

2.2 Pathogenicity Islands

Another important group of mobile elements that are presumed to be involved in HGT are the Pathogenicity islands (PAIs), which are decisive in the virulence of bacteria (GILL *et al.* 2005; PALLEN and WREN 2007). The HGT facilitates a rapid acquirement of complex virulence functions from different species. PAIs were first described in pathogenic *E. coli*, where Hacker and collegues investigated the genetics behind virulence. By deletion of two large DNA inserts in the chromosome, they found that deletions affected virulence expression, thus mutants with deletions of both inserts resulted in a entirely avirulent strain (KNAPP *et al.* 1986; HACKER *et al.* 1990). These regions responsible for virulence in *E. coli* 536 led to the term pathogenicity islands (BLUM *et al.* 1994).

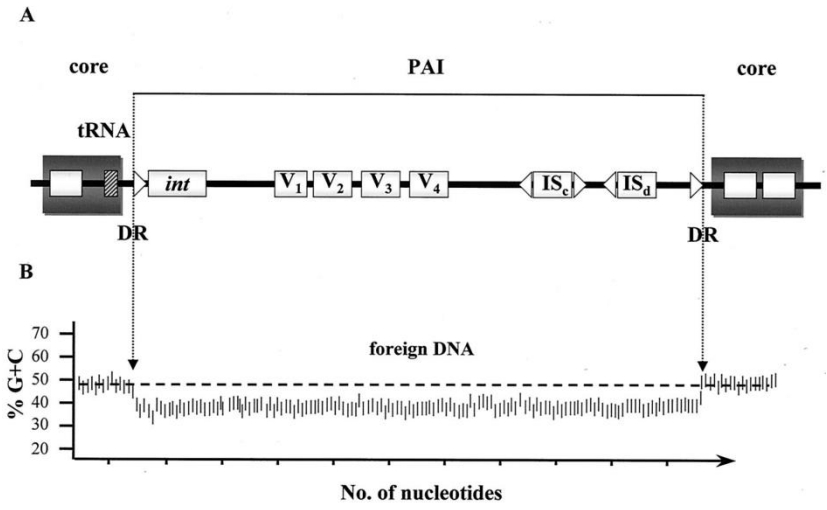


Figure 5 | General structure of Pathogenicity Islands. (A) shows a schematic model of a bacterial genomic island. The PAIs are frequently inserted at sites with tRNA or tRNA-like genes, flanked by directed repeats (DR) which are used in insertion and deletion processes. (B) The content of guanine and cytosine (G+C) differs from the rest of the genome, which is often used to identify new PAIs. Figure 5 is adapted from (SCHMIDT and HENSEL 2004)

PAIs can be diverse in both structure and function, however numerous common genetic features have been observed. For a genomic island to be a PAI it has to carry at least one virulence gene, illustrated by V_1, V_2 etc. in Figure 5. The majority of the islands can be relatively large, with the

majority somewhere between 10 and 200 kb. The PAIs will only be present in pathogens of the same or closely related bacterial species. The guanine and cytosine composistion (base composition) of bacterial DNA can vary from 25 to 75%, however PAI often differ in base composition compared to the rest of the genome in which they are found (HACKER and CARNIEL 2001; HENSEL 2004). As PAI are belived to be acquired by HGT, the reason for this difference in composition is considered to derive from the donor species. However, it has been observed from studying genome evolution that horizontally acquired DNA will tend to shift towards the base composition in the recipient genome, making it difficult to explain why inherent PAIs still differ in their base composition (SCHMIDT and HENSEL 2004).

PAIs are often, like bacteriophages (CAMPBELL 1992), located within an insertion sites such as tRNA, and frequently associated with DNA mobility. Directed repeats (DR) might serve as recognition site for integration, however they can also contribute to the instability of PAIs as DR act as recognition sequences for enzymes that are involved in deletion of MGEs. When PAIs are transferred by transduction (bacteriophages) integrases might mediate the integration into the chromosome of the bacteria, as well as excision when the bacteriophage needs to enter a lytic cycle. Tns are another type of MGEs that are often found in PAIs, thus besides changing the location within the bacterial chromosome, Tns can also jump between chromosome and plasmids. Additionally, the combination of two or more insertion elements (IS) can also result in mobilization of the PAIs, as well as inactivation of genes. Thus, all the MGEs present at PAIs causes unstablility, and deletions of PAIs from chromsomes with distinct frequencies (HACKER *et al.* 1997; HACKER and CARNIEL 2001; SCHMIDT and HENSEL 2004).

The functions encoded by PAIs are dependent on the environment the bacterium lives in. Thus the genetic repertoire encompass several functional groups, with the most common ones being: (i) Iron uptake systems, including siderophores like aerobaction or yersiniabactin used to delivery of iron into microbial cells. (ii) Adhesins counting i.a. type 4 pili, P-pili, S-fimbriae and Sap adhesion. Virulence factors in this group enable bacteria to adhere to the surfaces of the host and promote the infection process in the host. (iii) Pore-forming toxins, exotoxins and enterotoxins such as α-hemolysin, listeriolysin, RTX-like exotoxin and pertussis toxin. The toxins can destroy or affect the function of the eukaryotic host cells. (iv) Protein transport through Type III and IV secretion systems. These organelles deliver effector proteins into the host cells, modulating the functions of the host. (v) Invasion genes, such as the *Salmonella* spp. *inv* genes, that facilitate bacterial access to the eukaryotic host cells (SCHMIDT and HENSEL 2004; GAL-MOR and FINLAY 2006).

For *S. enterica*, most of the virulence factors are encoded by genes carried in the chromosome, many located within PAIs. Genes needed for invasion of *S. enterica* were found to cluster within a defined region in the chromosome by the group of C. A. Lee (MILLS *et al.* 2006). This lead to the PAI named SPI-1. Thus currently 20 different types of PAIs are recorded in PAIDB for *Salmonella* (http://www.paidb.re.kr/ accessed December 4th 2015).

2.2.1 SPI-Finder

As whole genome sequencing more or less has become the new standard, a user-friendly web-tool for an easy *in silico* detection of *Salmonella* pathogenicity islands (SPIs) has been developed (**Roer II**). The *SPI-Finder* tool is publicly available as a web-based service hosted by the Center for Genomic Epidemiology (CGE) (https://cge.cbs.dtu.dk/services/SPIFinder/). *SPI-Finder* is built on a BLAST-based methodology for detection of genes from a customized database containing SPI variants.

There have been various initiatives for constructing databases concerning the virulence potential of bacteria, including the database 'Virulence Factors of Pathogenic Bacteria' (VFDB), a comprehensive database which comprises virulence factors form a broad range of bacterial species of medical importance (CHEN *et al.* 2005). However, a drawback with VFDB is the numerous genes present in the database that encodes hypothetical and unknown protein functions. Furthermore, the database only includes 5 of the known *Salmonella* pathogenicity islands. Another initiative is the PAthogenicity Island DataBase (PAIDB) constructed by Yoon *et al.* An initiative with the purpose of providing a database with comprehensive information on both known and potential pathogenicity island regions in prokaryotic genomes (YOON *et al.* 2007). The updated PAIDB version 2.0 contains a total of 223 types of pathogenicity islands, thus the *Salmonella* records have served as inspiration for the customized database behind the *SPI-Finder* (YOON *et al.* 2015). For some of the islands recorded in PAIDB, multiple copies with 100% identity were represented from different host origins. However, in those cases only one representative was included in the *SPI-Finder* database. Different variants within one island were denoted based on their host origin. The SPI database was constructed in FASTA format, with a descriptive header and nucleotide sequence for each record, thus the database can be updated easily by adding new island variants to the file. Another advantage of using FASTA is the uniform format that makes it easy to share.

The BLAST-based program used in *SPI-Finder* was originally developed by Zankari *et al.* for the web-tool *ResFinder* capable of detecting acquired resistance genes from WGS-data. The program

was developed for detection in both pre-assembled genomes and data of raw reads, with a user-selection of minimum percent identity (%ID) between 80 and 100% (ZANKARI *et al.* 2012).

SPIFinder-1.0 Server - Results

Pathigenic islands									
Gene	Origin	%Identity	HSP/Query length	Contig	Position in contig	Insertion location	Function category	Genome Accession	SPI Accession
SPI-5	Salmonella Choleraesuis str SC-B67	100.00	5689 / 5689	S_Choleraesuis_str_SC-B67_1	1155936..1161624	tRNA-ser	18	NC_006905	gi|62178570:1155936-1161624
SPI-11	Salmonella Choleraesuis str SC-B67	100.00	15686 / 15686	S_Choleraesuis_str_SC-B67_1	1350481..1366166	Gifsy-1_prophage	7	NC_006905	gi|62178570:1350481-1366166
SPI-2	Salmonella Choleraesuis str SC-B67	100.00	41829 / 41829	S_Choleraesuis_str_SC-B67_1	1497670..1539498	tRNA-val	14	NC_006905	gi|62178570:1497670-1539498
SPI-12	Salmonella Choleraesuis str SC-B67	100.00	11075 / 11075	S_Choleraesuis_str_SC-B67_1	2354604..2365678	tRNA-pro	8	NC_006905	gi|62178570:2354604-2365678
SPI-1	Salmonella Choleraesuis str SC-B67	100.00	43488 / 43488	S_Choleraesuis_str_SC-B67_1	2960260..3003747	fhlA-mutS	5	NC_006905	gi|62178570:2960260-3003747
SPI-13	Salmonella Gallinarum SGA-10	99.12	341 / 341	S_Choleraesuis_str_SC-B67_1	3236624..3236964	tRNA-pheV	10		AY956834
SPI-13	Salmonella Gallinarum SGG-1	100.00	404 / 404	S_Choleraesuis_str_SC-B67_1	3238332..3238735	tRNA-pheV	11		AY956833
SPI-13	Salmonella Gallinarum SGD-3	98.82	338 / 338	S_Choleraesuis_str_SC-B67_1	3239043..3239380	tRNA-pheV	9		AY956832
SPI-3	Salmonella Choleraesuis str SC-B67	100.00	12819 / 12819	S_Choleraesuis_str_SC-B67_1	3890879..3903697	tRNA-selC	15	NC_006905	gi|62178570:3890879-3903697
SPI-4	Salmonella Choleraesuis str SC-B67	100.00	26698 / 26698	S_Choleraesuis_str_SC-B67_1	4411902..4438599	ssb-soxSR	17	NC_006905	gi|62178570:4411902-4438599
SPI-14	Salmonella Gallinarum SGA-8	100.00	501 / 501	S_Choleraesuis_str_SC-B67_1	953132..953632	Not_published	12		AY956835
SPI-14	Salmonella Gallinarum SGC-8	99.77	441 / 441	S_Choleraesuis_str_SC-B67_1	960047..960487	Not_published	13		AY956836

Category function table
extended output

Results as text | Hit in genome sequences | Database gene sequences | Results Tab Separated

Selected %ID threshold: *95.00 %*

Selected minimum length: *60 %*

Input Files: *S_Choleraesuis_str_SC-B67.fna*

Figure 6 | *SPI-Finder* output for *Salmonella* Choleraesuis str. SC-B67. For verification of the database, the output for *Salmonella* Choleraesuis str. SC-B67 was correlated with the seven known records in the SPI database for the isolate. Besides the seven records that originated from S. Choleraesuis str. SC-B67, *SPI-Finder* also detected SPI-13 and SPI-14 variants which were first identified in different *S.* Gallinarum isolates.

Subsequently, the original program was updated to only report the best hit within a given position of the data (with allowance of 30% overlap), and to report multiple variants of the same gene in different positions. Further, a second user-selection parameter was added, allowing for selection of a minimum length between 100 and 20%, and the selection interval for %ID was updated to go as low as 30% ID. This updated version of the program was applied to the *SPI-Finder*, allowing only the best hits from the SPI database, which also meet the user-specified parameters, to be reported in the output. The *SPI-Finder* is constructed with an easily understandable output, reporting the

pathogenicity islands detected, the origin of the islands, the %ID of the hit, the length of the hit and database record, the contig followed by the position of the hit in the contig, the insertion location from which the island was originally found, and additionally a function category from which the function of the islands can be interpreted in the category function table located beneath the output table. An output for *S.* Choleraesuis SC-B67 is shown in Figure 6. The hits in the output are colour indicated according to the two parameters; length and identity. Dark green indicates hits of full length and with an identity of 100%, where light green indicates hits with full length but with less than 100% identity. Additionally, results indicated with grey represents results of less than 100% identity that are not of full length (not represented in Figure 6).

In order to examine the performance and stability of *SPI-Finder*, the two WGS isolates with the most records in the SPI database (*S.* Choleraesuis str. SC-B67 with seven records and *S.* Typhi CT18 with 10 records), were tested against the program. For *S.* Choleraesuis SC-B67 (Figure 6), all seven records were found on a 100% ID-level, and in full-length, five additional islands were detected. For S. Typhi CT18 all 10 records were detected with the addition of one island. Thus in both cases, the *SPI-Finder* was capable of detecting all records in the SPI database from the two isolates.

As mentioned, *SPI-Finder* is not the first tool exploring pathogenicity islands; the PAIDB initiative comprises a PAI Finder predicting PAI-like regions. However, a drawback with the PAI Finder is that the user cannot upload WGS data directly to the tool, as it works on open reading frames (ORF). Additionally, the header within the searched file needs to follow a strict pattern: ">ID(integer)|ORF_name|start..end|strand (+ or -)", and the number of query sequences are limited to 1,000. Thus Genebank files of seven different diverse prokaryotic species shows that the number of ORFs are between 1700 and 5566 for a genome, which were also predicted by different gene prediction tools (HYATT *et al.* 2010), indicating that it is impossible to search an entire genome at a time by using PAI Finder.

The attempt with *SPI-Finder* was to design a simple, user-friendly tool to interpret *Salmonella* pathogenicity islands from WGS data. However, different aspects should be considered when using such tool. Notably, searches based on databases are only capable of detecting and reporting records if they are present in the database explored. Consequently, when inquiring a database tool like *SPI-Finder*, the user is to some extend obligated to become acquainted with the content in the database employed. Additionally, it is important to bear in mind that *SPI-Finder* solely determines the

presence of genes in the genome and does not access any information concerning the expression. *SPI-Finder* provides an assessment of pathogenicity islands in *Salmonella*, but no information is provided about the actual level of pathogenicity.

Another consideration when working with WGS data is the quality of the sequence to be explored, thus identifying the ideal level of %ID to be capable of detecting the presence of genes. In the original program constructed for *ResFinder*, Zankari *et al.* reported an optimal %ID at 98% to circumvent noise signals (ZANKARI *et al.* 2012). However the initial study was performed on Illumina sequencing data which, compared to other technologies, contain less indel errors. For sequence data of lower quality it could be an advantage to decrease the %ID to be capable of detecting genes. As a default *SPI-Finder* do not report hits with an ID of less than 95%. In addition, the minimum length should be considered in comparison with data quality. By default, *SPI-Finder* only report hits that cover at least 60% of the gene length. However, poor genome assembly can result in a false-negative detection of a gene, or the sequence of a present gene might be split on multiple contigs, making it difficult to detect the gene with such BLAST-based tool. Thus, quality control and measures of the input data should be provided for the user to be able to have a critical view on the results. Currently, methods that map reads to databases are tested as an alternative to the BLAST-based method, to circumvent the problems with bad quality assemblies and genes split into multiple contigs.

2.3 The Mechanisms of Horizontal Genetic Transfer

Vertical transfer describes the process from which genetic material is inherent from parental cells to offspring which happens naturally during reproduction. Opposite the vertical transfer process, HGT describes the movement of genetic material between and within different populations of prokaryotes in the community (FRANCINO 2012).

The importance of HGT was first recognized together with the observation of heredity of multiple antibiotic resistances in pathogens in the 1940's (WATANABE 1963; DAVIES and DAVIES 2010). Subsequently, substantial genetic exchange between bacteria was revealed by disparities in base composition and biased codon usage, high gene similarity of distantly related organisms, and variation of gene content within the same species (KOONIN *et al.* 2001; HEUER and SMALLA 2007; SCHUMANN *et al.* 2008). HGT is currently accepted as a dominant driving force in rapid adaptation and evolution of bacterial genomes (BERG and KURLAND 2002; DOOLITTLE *et al.* 2003; HEUER and SMALLA 2007; SCHUMANN *et al.* 2008).

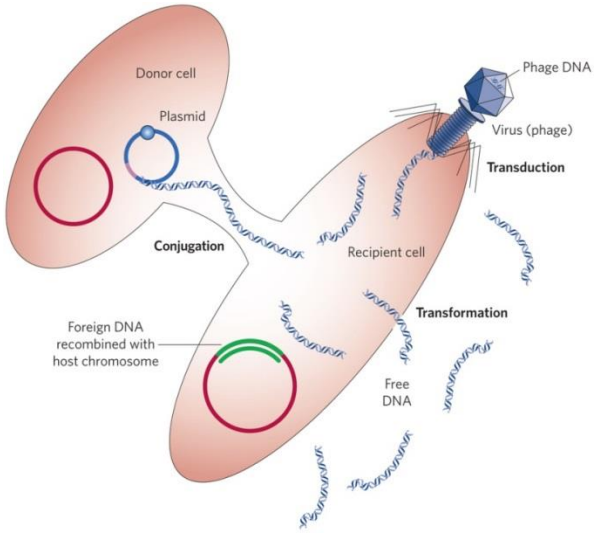


Figure 7 | The three general transfer modes of Horizontal Gene Transfer; Conjugation, Transformation and Transduction. The different mechanisms are explained in the text. Figure 7 is adapted from (STEWART 2013).

Several studies have elucidated the extent of HGT in bacterial communities. In a clinical isolate of the gram-positive bacteria *Enterococcus faecalis* more than 25% of the genome potentially consisted of mobile DNA (PAULSEN *et al.* 2003). As previously mentioned, in the case with the non-pathogenic *E. coli* K-12 strain and the pathogenic O157:H7 strain a difference around 30% was observed (PERNA *et al.* 2001), indicating a big intra-species divergence.

The horizontal transfer of genes is mediated by three general mechanisms; conjugation, transformation and transduction, as illustrated in Figure 7.

The individual contribution to HGT is not known for the three transfer mechanisms. However, the transfer of plasmid mediated by conjugation is believed to be one of the most frequent and disseminated mechanisms for transfer of genetic material among bacterial species (JAIN *et al.* 2002; THOMAS and NIELSEN 2005; SCHUMANN *et al.* 2008). This PhD thesis is focusing on conjugational transfer, the mating process between bacteria. However, transduction and transformation will shortly be described in the sections 2.3.2 and 2.3.3.

2.3.1 Conjugation

The conjugational process is highly specific, where the transfer of DNA is specialized by a multiprotein complex, the conjugation apparatus (GROHMANN *et al.* 2003). The active participation of both mating partners, donor and recipient, was already suggested in 1968 by Curtiss *et al.* (CURTISS *et al.* 1968). A physical contact between the cell surfaces of the mating pair is an

important prerequisite for conjugational transfer, which in gram-negative bacteria is mediated by extracellular filaments termed sex pili (GROHMANN *et al.* 2003).

The type IV secretion systems (T4SS) are responsible for the conjugative transfer of plasmids. One of the best studied conjugative transfer systems is the IncP (*tra*) system of RP4, a broad-host range IncP plasmid, of which the most essential mechanism will be described in this section.

The IncP transfer system is comprised in the two regions; Tra1 and Tra2. For DNA to travel across the cell envelopes of the mating cells, two main protein complexes are involved; the multiprotein-DNA complex relaxosome and a mating-pair formation (mpf) complex. The relaxosome complex is formed from genes encoded by both chromosome and plasmid (FÜRSTE *et al.* 1989; LANKA and WILKINS 1995). The mpf complex is encoded in the plasmid *trb* operon (LESSL *et al.* 1993) as illustrated at the IncP plasmid in Figure 4.

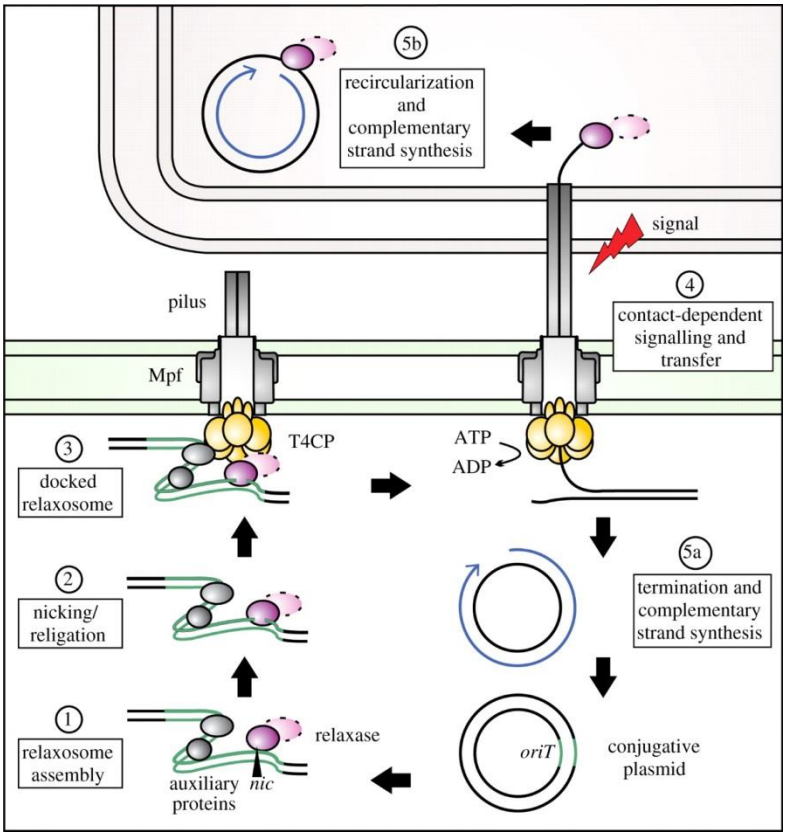


Figure 8 | Overview of general mechanisms for plasmid transfer. Conjugative plasmids are carrying an origin of transfer (*oriT*), genes encoding proteins for pili, mating-pair formation complex (mpf) responsible for the T4 envelope spanning channel, a T4CP (yellow), the relaxsosome conjugation initiation complex with the secretion substrate and DNA nicking relaxase (pink). The different steps are described in the text. Figure 8 is adapted from (ZECHNER *et al.* 2012)

The conjugational transfer is illustrated step-by-step in Figure 8, and described in the following.

(i) The conjugation is initiated by the assembly of multiple proteins forming the relaxosome

complex at the origin of transfer (*oriT*). (ii) The complex is then preparing the single-stranded DNA intended for transfer (T-strand). This is done by nicking activity of the phosphodiester bond at the *nic* site within *oriT*, which in addition, mediates formation of a covalent tyrosinyl-DNA adduct (ZECHNER *et al.* 2012). (iii) The plasmid-encoded TraG-like T4CP (coupling protein) recognizes the nucleoprotein and docks it to the complex (GRAHN *et al.* 2000). (iv) When the initiation signal is registered, as a result of donor-recipient contact, the T-strand with linked relaxase is actively pumped through the mfp transport apparatus. (v) After the termination of transfer, the original plasmid in the donor is stabilized by conjugative replication, where the missing complementary strand is synthesized. (vi) In the recipient, recircularization of the plasmid is done by joining the free 3'-hydroxyl and the 5'-terminus with the covalently bound relaxase. Finally, the transferred plasmid is stabilized by freeing the relaxase and synthesizing the complementary strand. (GROHMANN *et al.* 2003; ZECHNER *et al.* 2012). The transfer genes at IncP plasmids are not constitutively expressed, but is regulated by local auto-regulators and global regulators, which results in a coordinated expression with other plasmid functions (ZATYKA and THOMAS 1998; DANG *et al.* 1999).

The Type IV secretion system (T4SS) involved in conjugational transfer is capable of connecting a wide variety of organisms (GRAHN *et al.* 2000; THOMAS and NIELSEN 2005), and is thereby one of the most significant machineries in bacterial adaptation and evolution (AMINOV 2011). Additionally, conjugative plasmids facilitate the spread of antibiotic resistance between pathogens, and continue to be a crucial player in the emergence of multidrug resistant pathogens in hospitals (BEOVIĆ 2006) and agriculture (ZHU *et al.* 2013).

2.3.1.1 Methods for Detection of Conjugational Plasmid Transfer

Conjugation abilities of bacterial species and transfer range of plasmids have conventionally been evaluated by using single isolates as recipients (LEDERBERG *et al.* 1952). The most common and still widely used method for detection of plasmid transfer is based on selective plating, where plasmid encoded traits act as markers for conjugational events. This requires for the plasmid to confer resistance to antibiotics or heavy metals, or to possess accessory metabolic pathways, which allow only the transconjugants to grow on the selective media. Nevertheless, the method has allowed for identification of several factors affecting conjugational transfer, including the effect of chemical compounds like biocides in transfer of the transposon Tn*916* (SEIER-PETERSEN *et al.* 2014), conjugational inhibition by unsaturated fatty acids (FERNANDEZ-LOPEZ 2005), the

availability of nutrients (SØRENSEN and JENSEN 1998), changes in temperature and pH values (RICHAUME *et al.* 1989; ROCHELLE *et al.* 1989).

The selective detection method does have other drawbacks than the selective requirements of the plasmids, especially when studying a complex microbial community. The method is restricted to the small fraction of bacteria which is culturable and able to grown on specific growth media; a fraction that probably is below 1% of the total bacterial community (AMANN *et al.* 1995). In addition, this method ignores the fact that most bacteria exist in complex matrices and together in communities with hundreds to thousands of other species, when only focusing on a single recipient strain (HONG *et al.* 2006; BROWN KAV *et al.* 2012).

To overcome some of these drawbacks, *in-situ* reporter genes were introduced on plasmids. The detection of plasmid transfer could then be measured as the expression of the reporter genes in the transconjugants, thus avoiding cultivation with selective plating and the enclosed selective advantage. Different reporter gene systems have been utilized for detecting plasmid transfer, including the *lux* system (HOFFMANN *et al.* 1998), the β-galactosidase assay with the *lacZ* gene (JAENECKE *et al.* 1996), and systems with reporter genes expressing fluorescent proteins (CHRISTENSEN *et al.* 1996; DAHLBERG *et al.* 1998).

To avoid expression of the reporter genes in the donor strain, inducible reporter genes were introduced on the conjugative plasmids. This was achived by introducing the reporter genes behind a *lacZ* promoter subjected to inhibition by constitutive *lacI* repression in the donor (either by insertion of *lacI* on the chromosome or present on a non-transferable plasmid) (DAHLBERG *et al.* 1998; FERNANDEZ-LOPEZ 2005). Thus, expression of the reporter genes would only occur after successful transfer to the recipients. The construction of the inducible *lux* system, measuring plasmid transfer as arbitrary light units (ALU), gave rise to a high-throughput conjugational assay, allowing up to 96 samples at the same time by using microplate luminometers for detection (FERNANDEZ-LOPEZ 2005; PÉREZ-MENDOZA and DE LA CRUZ 2009).

The inducible fluorescent system with *gfp* as reporter gene has been widely used to quantify plasmid transfer using flow cytometers, where the transfer was measured as expression of GFP in the recipients (CHRISTENSEN *et al.* 1996; DAHLBERG *et al.* 1998; SØRENSEN *et al.* 2003). The enhanced ability to study plasmid transfer in microbial communities revealed transfer frequencies up to 1,000 times higher than observed with methods dependent on cultivation (MUSOVIC *et al.* 2006). In 2000, the *gfp* reporter system was further improved by implementing a red-fluorescent

marker gene (*DsRed*) at the donor chromosome. This enabled simultaneously detection and quantification of the donors, recipients and transconjugants (TOLKER-NIELSEN *et al.* 2000).

The improved *gfp* reporter system is still available, thus it provides the opportunity to screen a broad collection of strains or even communities for their ability to conjugate and take up a plasmid. With isolation of single transconjugants by fluorescent activated cell sorting (FACS), and whole genome amplification it is possible to determine the range of recipients in a community, without the need of selection and cultivation (MUSOVIC *et al.* 2006; KLÜMPER *et al.* 2014; SHINTANI *et al.* 2014).

The fluorescent properties of the system have been improved by using the advanced *gfpmut3* variant for the green fluorescent reporter and *mCherry* as the red fluorescent donor marker gene, in the construct of i.a. *E. coli* MG1655::*mCherry-lacI*^{q} with the plasmid construct pKJK5::*gfp3mut* (KLÜMPER *et al.* 2014).

This latter *E. coli* pKJK5 system, constructed by Klümper *et al.*, was applied in **Roer III**, as a high-throughput screening method, to identify conjugative abilities in a broad collection of *S. enterica* ssp. *enterica*.

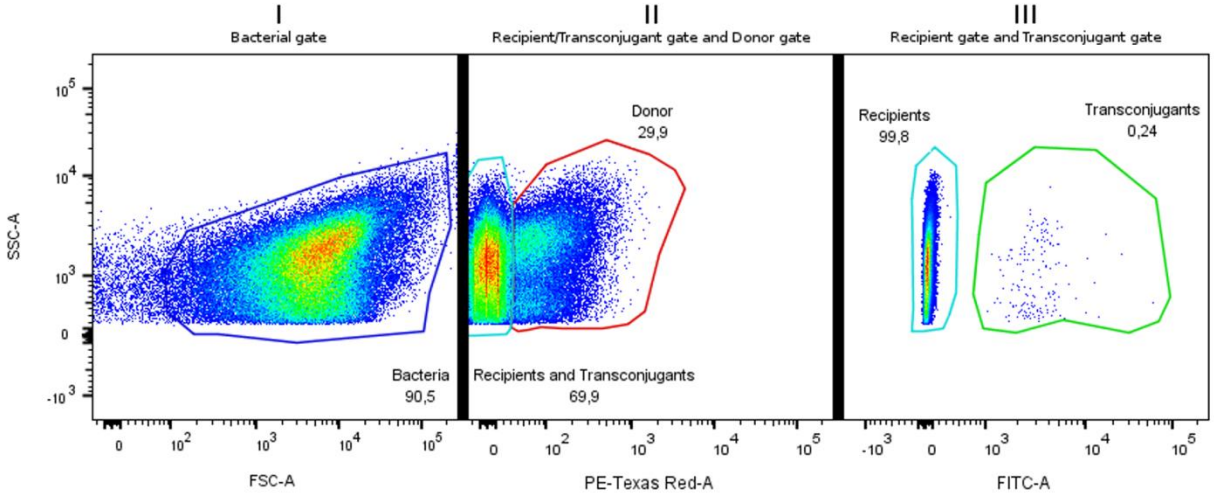


Figure 9 | Quintuple gating for donor, recipient and transconjugant detection by flow cytometry. Flow cytometry detection of donor, recipients, and transconjugants from a conjugation mixture of a *S. enterica* recipient and *E. coli K-12* MG1655 donor carrying pKJK5. The gating procedure consists of five successive gates in the 3 bivariate plots: Plot I with the Bacterial gate, detects all particles of bacterial size based on front scatter (FSC) and side scatter (SSC); Plot II detects all red fluorescence particles (*mCherry*) representing the donors, and all non-red particles representing either recipients or transconjugants by using side scatter and PE-Texas Red; Plot III divides the recipients and transconjugants by gating non-green and green fluorescence particles (*gfp3mut*) based on side scatter and FITC. Figure 9 adapted from (**Roer III)**.

In **Roer III**, the detection of donor, recipients and transconjugants was based on a quintuple gating approach. The five gates are defined in the three bivariate plots in Figure 9. In the first plot, a

combination of side scatter (SSC) and forward scatter (FSC) is used to distinguish bacterial particles from other subpopulations, by using a gate corresponding to particles of bacterial size. In the red fluorescence (PE-Texas Red) vs SSC plot a gate was set to cover all donor cells, with an additional non-red gate covering recipients and transconjugants. In the third plot, green fluorescence (FITC) and SSC was used to distinguish green transconjugants from non-green recipients. This approach allowed us to count 100,000 bacterial events in around 30 seconds, without the need to use selective cultivation.

2.3.2 Transduction

Transduction is another of the three transfer mechanisms comprised by the term HGT. The transfer of non-viral DNA to a new host is mediated by the bacterial viruses, bacteriophages (phage), as illustrated in the upper right corner of Figure 7 (STEWART 2013). The mechanism relies on mistakenly packed phage heads where host DNA is integrated in the phage DNA during the reproduction of the phage particle. When host cells are lysed, defective phage particles are released, ready to adsorb to new host cells, where they can inject the carried DNA from the previous host. When delivered to the recipient host cell, the transferred DNA can be integrated into the chromosome (MAZODIER and DAVIES 1991; HEUER and SMALLA 2007). A review by Brüssow *et al.* describes different bacterial species that have acquired pathogenicity determinants through phages (BRÜSSOW *et al.* 2004). Additionally, PAIs in pathogenic strains is believed to have evolved from lysogenic bacteriophages (HACKER *et al.* 2003; DOBRINDT *et al.* 2004).

It was originally assumed that phages was limited to a small range of hosts (BERGH *et al.* 1989; MAZODIER and DAVIES 1991), however phages like P1 and Mu have shown to interconnect with a broad range of bacteria (JENSEN *et al.* 1998; WOMMACK and COLWELL 2000; CHEN and NOVICK 2009). Transfer by transduction has the advantage that DNA is protected from the environment, and the phage particles can persist relatively long under environmental conditions (ZEPH *et al.* 1988; WOMMACK and COLWELL 2000).

The number of bacteriophages in seawater is found to be 10 times higher than the number of bacteria (BRÜSSOW and HENDRIX 2002; DANOVARO *et al.* 2008), and with transduction events of 20 million billions per second in the oceans (CHIBANI-CHENNOUFI *et al.* 2004) marine viruses are believed to be a major contributor in the global ecosystem (SUTTLE 2007; DANOVARO *et al.* 2008; ZHAO *et al.* 2013). However, as the transduction mechanism relies on mistakes in phage packing,

most of the transferred DNA will not be functional genes. Thus, the evolutionary impact of transduction will probably be rather low.

2.3.3 Transformation

The third HGT mechanism is transformation (Figure 7). Transformation covers the uptake of free exogenous DNA by competent cells (LORENZ and WACKERNAGEL 1994; DUBNAU 1999). Natural competence is a physiological phase, genetically programmed, where efficient uptake of DNA is permitted. This can either be a persistent phase, as for *N. meningitidis* (HAMILTON and DILLARD 2006), or a phase induced by environmental factors (HANAHAN 1983; NIELSEN and VAN ELSAS 2001). Thus, the prerequisites for transformation are the bacterial change to a competent phase, the availability of exogenous free DNA, followed by the uptake and stable integration of the DNA (HEUER and SMALLA 2007).

Transformation is a commonly used method in gene technologies, as artificially constructed changes of environments can trigger the desired bacterial cells to enter a competent phase, ready to act as recipients for uptake of free DNA. However, transformation frequencies among isolates from the same species can differ significantly, indicating that the ability to transform is not shared systematically between isolates from the same bacterial species (SIKORSKI *et al.* 2002; MAAMAR and DUBNAU 2005). Additionally, the efficiency of integrating the newly introduced DNA to the bacterial genome might vary between bacterial species (SIKORSKI *et al.* 2002).

Transformation has been confirmed in various bacterial species and environments (BRÄUTIGAM *et al.* 1997; NIELSEN and VAN ELSAS 2001; AVERHOFF and FRIEDRICH 2003; SØRENSEN *et al.* 2005; HAMILTON and DILLARD 2006), however very limited knowledge exists on the importance of transformation, the subsequent ability of the bacteria to adapt to different environments (HEUER and SMALLA 2007), as well as the contribution to bacterial evolution (JOHNSBORG *et al.* 2007).

Chapter 3

TRANSFER DEFENCE MECHANISMS

The evolutionary rate in bacteria is affected by both the ability to exchange DNA as well as the frequencies of transfer, thus any genetic determinant restricting or enhancing this, may influence evolution. A number of well-described mechanisms affecting conjugative transfer include plasmid incompatibility (described in section 2.1.3), entry exclusion, clustered regularly interspaced short palindromic repeats (CRISPR), and restriction-modification systems (RM systems). This PhD thesis is focusing on the barriers of RM systems, and their influence on evolution. However, entry exclusion and CRISPR will shortly be described in the sections 3.2.1 and 3.2.2.

3.1 Restriction-Modification Systems

Restriction-modification systems are transfer defence mechanisms endorsed by the bacterial hosts. Generally all bacteria mark their DNA with an identity signature, thus when DNA is transferred from one bacterium to another, DNA lacking the signature of the recipient strains will be perceived as 'foreign' DNA rather than 'self'. DNA perceived as 'foreign' is generally enzymatically cleaved in the recipient (KING and MURRAY 1994; WILLIAMS 2003; ROBERTS *et al.* 2003).

The discriminatory barrier was first demonstrated in the 1950's with an *E. coli* K-12 strain infected by bacteriophage λ, transmitted through either *E. coli* strain C or B (BERTANI and WEIGLE 1953). Around a decade later, it was proven that the establishment of bacteriophages in *E. coli* K-12 strains

could be 'restricted' by an enzyme that attacked 'foreign' DNA (MESELSON and YUAN 1968; LINN and ARBER 1968).

Traditionally, the restriction enzyme is found together with a cognate methyltransferase, modifying the DNA by adding a methyl group in a given position. The activities from the two genes constitute the restriction-modification systems. However, there are systems only comprised by a modification-dependent restriction enzyme, which only attack DNA, when the DNA is methylated in a specific pattern (ROBERTS *et al.* 2003). Both variants share the ability to attack DNA recognized as 'foreign' and thereby prevent DNA establishment in the recipient. Thus, while the methyltransferase from a classical RM system is required to protect the cells own DNA from degradation by the cognate restriction enzyme, the modification can act as substrate for restriction activity from a modification-dependent system in another strain (MURRAY 2002). The classical RM systems have been subdivided into three groups (type I-III), based on their complexity, the requirement of cofactors, their recognition sequence, and DNA cleavage site, as illustrated in Figure 10. The modification-dependent systems account for a fourth group (type IV). However, some RM systems do not fit readily into any of the four types (KING and MURRAY 1994).

The key characteristics for the type I RM systems are the hetero-oligomeric complexes which catalyses both restriction and modification, when the recognition sequence is recognized by the specificity subunit (S). They are usually comprised of two R subunits, two M subunits and one S subunit.

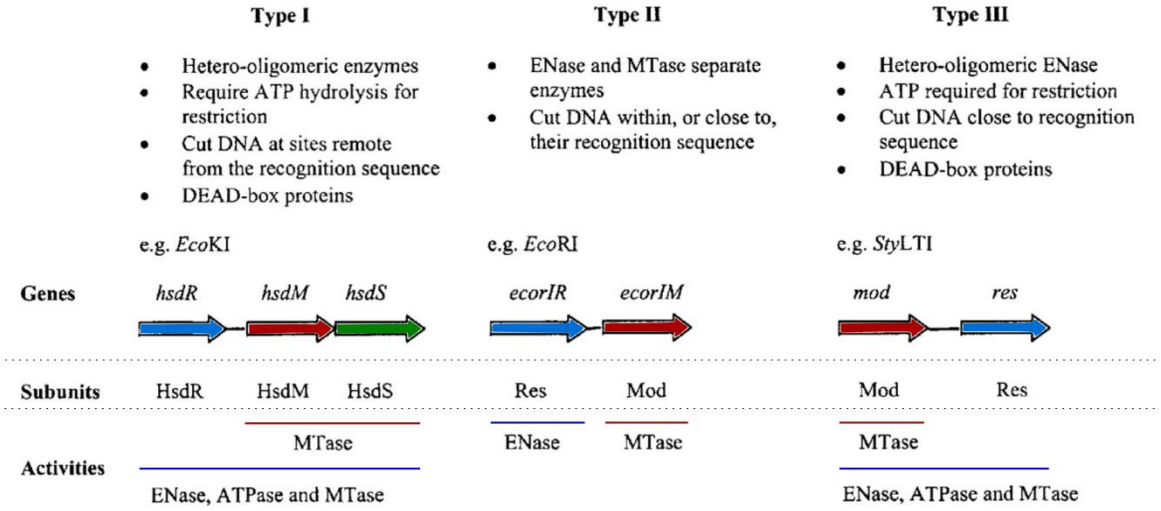


Figure 10 | Characteristics and gene organization for the three classical RM systems; Type I-III. RM systems are subdivided according to their complexity, cofactor requirement, recognition sequence and DNA cleavage site. Type I, type II and type III are the classical RM systems. Type IV Modification-dependent systems are not classical systems, as they do not possess a cognate methyltransferase, and only cleave DNA with specific modifications. ENase, endonuclease activity; MTase, methyltransferase activity. Figure 10 is modified from (MURRAY 2002).

The three type I subunits are encoded by the closely related genes; *hsdR*, *hsdM* and *hsdS*. The genes are transcribed from two promotors; *hsdM* and *hsdS* from the same promotor, and *hsdR* from a separate promotor (LOENEN *et al.* 1987). For methyltransferase activity, only the two subunits M and S are necessary, whereas the third subunit R is required for restriction activity. Thus, the restriction activity will not be active without the M and S subunits. Two functional complexes exists; one with only methyltransferase activity (M_2S_1), and the RM system which comprises all three subunits ($R_2M_2S_1$) (MURRAY 2000). In the modification reaction, *S*-adenosyl-methionine (AdoMet) act as the methyl source, whereas AdoMet, ATP, and Mg^{2+} all are required for restriction activity (BICKLE and KRÜGER 1993). In type I RM systems, the recognition sequences are asymmetric and are constructed by two components, the first of 3 or 4 bp and the second of 4 or 5 bp. The two components are separated by a non-specific spacer region of 6-8 bp. When the recognition sequence is hemimethylated the DNA will act as substrate for modification, whereas unmodified recognition sequences will trigger DNA cleavage at sites remote from the recognition sequence (BICKLE and KRÜGER 1993; MURRAY 2002).

Compared to the type I RM systems, the type II systems are less complicated, which is probably why this is also the type of restriction enzymes used in gene technology. The systems are comprised of two separate genes; *Res* encoding the restriction activity and *Mod* encoding the methyltransferase activity. As for the type I systems, the modification requires AdoMet as methyl donor, while the restriction activity for type II systems only is dependent on Mg^{2+} (BICKLE and KRÜGER 1993; ROBERTS *et al.* 2003). The restriction enzyme and the methyltransferase recognize the same recognition sequence, which generally is a palindromic sequence of 4-8 bp. The methyltransferase modify a specific base within the recognition sequence on each strand. The structure of the recognition sequence facilitates cleavage of both DNA strands at the same time, generally within or close to the recognition sequence (ROBERTS *et al.* 2003).

The type III systems are not as simple as type II systems, but are sharing some traits with the type I systems. A hetero-oligomeric complex of the *Res* and *Mod* genes catalyse both restriction and modification activities in the type III systems. As for the type I and II systems, the AdoMet is required for modification, while both Mg^{2+} and ATP are required for restriction activity in type III systems. The recognition sequences for type III systems are asymmetrical, and the cleavage site is located close to the recognition sequence (BICKLE and KRÜGER 1993; ROBERTS *et al.* 2003).

The non-classical RM systems, type IV, restrict modified DNA. However their recognition sequences have usually not been defined (ROBERTS *et al.* 2003).

For decades, the restriction-modification systems have been known to act as barriers, not only on bacteriophages, but for horizontal gene transfer as well (KING and MURRAY 1994; TOCK and DRYDEN 2005; VEIGA and PINHO 2009; VASU *et al.* 2012). Generally, the belief is that DNA lacking the signature of the recipient will act as substrate for restriction, irrespective of whether the 'foreign' DNA would enter the recipient as single- or double-stranded (MURRAY 2002). Plasmid DNA, which enters the recipient as single-stranded DNA through conjugation, will become sensitive to cleavage after stabilization from the synthesis of the complementary strand (THOMAS and NIELSEN 2005).

However, this has, to our knowledge, never been demonstrated at a genotypic level. In **Roer I**, the impact of the type I RM system *Eco*KI, in conjugational transfer, was elucidated by utilizing a combination of isogenic RM variants of the *E. coli* K-12 strain MG1655 and plasmids pOLA52 (SØRENSEN and HANSEN 2003) and pHHA45 (DOLEJSKA *et al.* 2012), respectively harboring either two or zero recognition sequences for the *Eco*KI system.

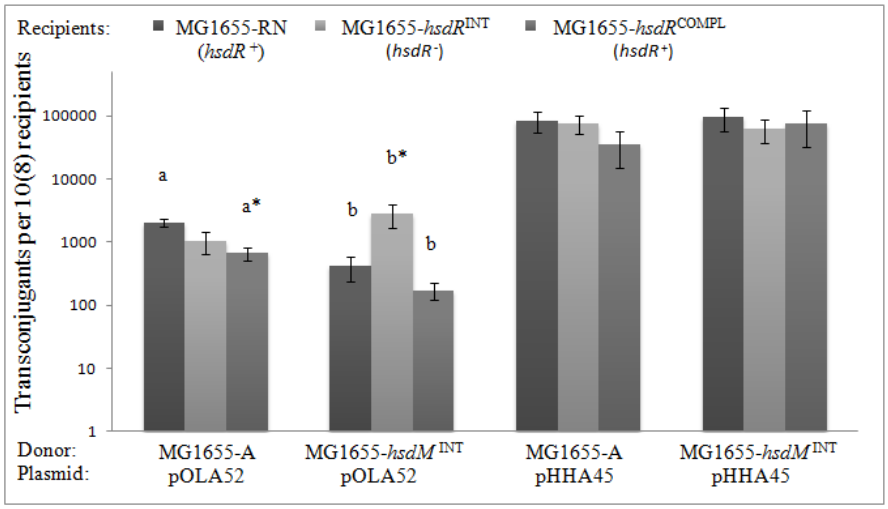


Figure 11 | Conjugational transfer with isogenic donor and recipient strains of *E. coli* K-12.
Each transfer experiment was performed at least in triplicate. The data is visualized as the means of transconjugants per 10^8 recipients, with standard errors of the mean (SEM). For un-methylated plasmid pOLA52, with two recognition sequences for the *Eco*KI system, statistically significant differences were observed between the restriction deficient recipient MG1655-*hsdR*^{INT} and the remaining two recipients (b*, *P*0.029). No statistically significant differences were observed for the un-methylated pHHA45 plasmid without recognition sequences for *Eco*KI.

In **Roer I** we show that a *hsdR* deficient mutant caused a 7-fold increase in uptake of an unmethylated plasmid with two recognition sites. Though, most importantly we demonstrated that

the barrier imposed by the *Eco*KI type I restriction system, had an effect, but that the barrier was not absolute in relation to conjugational plasmid transfer.

3.1.1 Restriction-ModificationFinder

In **Roer II**, the content of RM systems was elucidated in a comprehensive collection of *S. enterica* WGS genomes. For that purpose, another user-friendly web-tool, *Restriction-ModificationFinder* (*RM-Finder*) was developed for an easy *in silico* detection of RM systems. The *RM-Finder* was built on the same BLAST-based methodology as *SPI-Finder*, with a customized database containing restriction enzymes, methyltransferases and specificity subunits for the classical systems, as well as the modification-dependent type IV systems. *RM-Finder* is likewise available as a web-based service hosted by CGE (https://cge.cbs.dtu.dk//services/Restriction-ModificationFinder/).

For databases on RM systems, the Restriction Enzyme database (REBASE) is the only authoritative source, a dedicated work by R. Roberts and D. Macelis (ROBERTS *et al.* 2015). REBASE is a comprehensive database, currently with more than 5,000 complete descriptions of RM systems. The database is containing information about restriction enzymes and related proteins like DNA methyltransferases, specificity subunits, homing endonucleases, helicase domain proteins etc. and contains both published and unpublished references. The data in REBASE is based on genome sequence data subjected to analysis, where predicted components from RM systems are labelled with the suffix 'P' to indicate that they are putative and included in the database (ROBERTS *et al.* 2015). When the systems become biochemically characterized, the suffix is replaced with a Roman number. Though, with the Single molecule Real Time (SMRT) sequencing, the characterization of especially the methyltransferases are rapidly increasing (FLUSBERG *et al.* 2010; KORLACH *et al.* 2010; KORLACH and TURNER 2012). Impressively, REBASE is continuously updated and their web interface refreshed on a nightly basis. Their ftp access to the data is refreshed once per month, providing data to external sources, like *RM-Finder*. However, the continuous expansion of REBASE indicates that far from all RM systems are known. This can be a problem when searching for RM systems, as it will be impossible to know if all systems are detected in the searched genomes.

Additionally, REBASE is providing various tools for investigating the determinants in RM systems, including a BLAST tool to search sequences against the REBASE database. This tool is capable of using either DNA or protein sequences as input file. However, few major drawbacks from this tool exist; the lack of ability to BLAST either raw data or assembled genomes, and absence of user-

selective parameters. The limitations from REBASE lead to the construction of *RM-Finder* (**Roer II**). REBASE is the source of data behind the database in *RM-Finder*, and includes the relevant restriction enzymes, methyltransferases, and specificity subunits for type I-IV RM systems. The RM database was constructed in the uniform and easily sharable FASTA format, containing a descriptive header, and the nucleotide sequence for each record. The database is automatically updated according to the monthly REBASE update.

The *RM-Finder* is constructed with the updated user-selection parameters, allowing for selection of minimum length and minimum %ID. The default settings for *RM-Finder* are set to minimum ID of 95%, and a minimum length of 60%. Furthermore, it is possible to search only one type of RM systems, and to search all records in the database, including putative genes.

Restriction-ModificationFinder-1.1 Server - Results

Type I Restriction enzymes

Gene	%Identity	HSP/Query length	Contig	Position in contig	Type	Function	Recognition Seq	Accession number
S.EcoKI	100.00	1395 / 1395	NC_000913	4578091..4579485	Type I	specificity subunit	AACNNNNNNGTGC	U00096.
M.EcoKI	100.00	1590 / 1590	NC_000913	4579482..4581071	Type I	methyltransferase	AACNNNNNNGTGC	U00096.
EcoKI	100.00	3567 / 3567	NC_000913	4581272..4584838	Type I	restriction enzyme	AACNNNNNNGTGC	U00096.

Type II Restriction enzymes

Gene	%Identity	HSP/Query length	Contig	Position in contig	Type	Function	Recognition Seq	Accession number
M.Eco3609Dcm	99.22	1419 / 1419	NC_000913	2028923..2030341	Type II	methyltransferase	CCWGG	JASV01000004.
M.EcoKII	100.00	885 / 885	NC_000913	3409675..3410559	Type II	methyltransferase	ATGCAT	U00096.

Type III Restriction enzymes

No restriction enzyme genes found.

Type IV Restriction enzymes

Gene	%Identity	HSP/Query length	Contig	Position in contig	Type	Function	Recognition Seq	Accession number
EcoKMcrA	100.00	834 / 834	NC_000913	1209569..1210402	Type IV	methyl-directed restriction enzyme	YCGR	U00096.
EcoKMcrBC(EcoKMcrC)	100.00	1047 / 1047	NC_000913	4574935..4575981	Type IV	methyl-directed restriction enzyme		U00096.
EcoKMcrBC(EcoKMcrB)	100.00	1380 / 1380	NC_000913	4575981..4577360	Type IV	methyl-directed restriction enzyme		U00096.
EcoKMrr	100.00	915 / 915	NC_000913	4584972..4585886	Type IV	methyl-directed restriction enzyme		U00096.

extended output

Results as text

Results tab separated | Hit in genome sequences | Restriction enzyme gene sequences

Selected %ID threshold: *95.00 %*

Selected minimum length: *60 %*

Input Files: *MG1655.fasta*

Figure 12 | *RM-Finder* output for *Escherichia coli* K-12 strain MG1655. For demonstration of *RM-Finder*, the complete genome of *E. coli* K-12 strain MG1655 was evaluated.

Figure 12 illustrates the output constructed by *RM-Finder*, reporting hits of each of the four systems in an individual section. The output reports the RMS genes detected, the %ID of the hit, the length of the hit compared to the database record, the contig in which the hit was found followed by the position in the contig. Additionally, the type of system, the function of the gene, the recognition sequence (if known), and the accession number are reported in the output. As for *SPI-Finder*, the hits are indicated by colour according to the two user-selection parameters; %ID and minimum length.

To evaluate the performance of *RM-Finder*, the *E. coli* K-12 strain MG1655 was evaluated, and compared with the putative predictions from REBASE (http://rebase.neb.com/cgi-bin/onumget?17068). It can be argued that the MG1655 strain is not the most complicated control strain. However, if *RM-Finder* is not capable of detecting all the known RM systems in a well-characterized strain, then the results predicted for uncharacterized strains will be unreliable.

Table 1 | Comparison of REBASE and *RM-Finder*

Predicted by REBASE	**Recognition Sequence**	**Predicted by *RM-Finder***
M.Eco1655DamP	GATC	Found by including putative
M.Eco1655DcmP	CCWGG	M.Eco3609Dcm
Eco1655McrBP	unknown	EcoKMcrB
Eco1655McrCP	unknown	EcoKMcrC
Eco1655MrrP	unknown	EcoMrr
Eco1655ORF300P	AACNNNNNNGTGC	EcoKI
M.Eco1655ORF300P	AACNNNNNNGTGC	M.EcoKI
S.Eco1655ORF300P	AACNNNNNNGTGC	S.EcoKI
M.Eco1655ORF6040P	ATGCAT	M.EcoKII
Eco1655ORF16910P	YCGR	EcoKMcrA

By only searching the biochemically characterized systems in *RM-Finder*, 90% of the genes were identified in concordance with REBASE. By inspecting the predictions from REBASE M.Eco1655DamP was predicted by the putative M.UbaC1152DamP. When searching the RM-Finder, including putative variants, the M.Eco1655DamP enzyme is predicted from *M.Eco321DcmP* with 100% ID and 100% length.

When using the *RM-Finder tool*, the same precaution should be taken as mentioned for *SPI-Finder*. The WGS data quality should be considered, and even though the data included in *RM-Finder* is comprehensive, database searches can only find what is included in the database.

3.2 Other Defence Mechanisms

In addition to RM systems, other defence mechanisms exist. Some can be encoded or caused by the transferred plasmid to prevent further uptake of MGEs, as for incapability and plasmid entry exclusion, others can, as for RM systems and CRISPR, be encoded on the bacterial chromosome to protect the bacteria. In the following sections, CRISPR and entry exclusion will briefly be introduced.

3.2.1 Host CRISPR Systems

CRISPR is referred to as the bacterial adaptive immune systems, as it has the benefit of continuously adjusting and adapting its reach and defence mechanisms against invasion (GOREN *et al.* 2012). According to the CRISPR database approximately 45% (1,176 out of 2,612) of the bacterial genomes possess CRISPRs (http://crispr.u-psud.fr/crispr/ Accessed December 22nd, 2015).

The CRISPR systems are combined by two different mechanisms; Immunization of 'foreign' incoming DNA like plasmids or bacteriophages, and secondly subsequent immunity to the same DNA. In the immunization process, 'foreign' plasmid or virus DNA is recognized by a Cas complex, and cleaved into novel spacer unit. The novel spacer unit is inserted at the leader end (5'-end) of the CRISPR locus, which induces acquired immunity against subsequent invasion (VAN DER OOST *et al.* 2009; HORVATH and BARRANGOU 2010). As the new spacer units generally are inserted at the leader end of the CRISPR locus, a chronological record of the encountered DNA exists. However, swapping and loos of spacers occur (LILLESTØL *et al.* 2006).

For the immunity process, the repeat-spacer array constructed from the immunization is transcribed into pre-crRNA, which is further processed into mature crRNAs. Subsequently, the crRNAs are guiding a Cas complex to the invading DNA, where inactivation is conducted (HORVATH and BARRANGOU 2010).

CRISPRs are defence mechanisms only recently discovered, however they are now applied to molecular biology research, enabling targeted genome editing (CONG *et al.* 2013; DOUDNA and CHARPENTIER 2014).

3.2.2 Entry Exclusion

Entry exclusion facilitated by plasmids can turn the bacteria into poor recipients to avoid addition of extra genetic elements (GARCILLÁN-BARCIA and DE LA CRUZ 2008). This ability can provide an

evolutionary advantage for the plasmid, as intra-cell competition with other conjugative plasmids or MGEs can be minimized or avoided (THOMAS and NIELSEN 2005).

The exclusion mechanism was shown to operate at both the inner- and outer recipient membrane, and be caused by the two plasmid-encoded genes, *traT* and *traS* (ACHTMAN *et al.* 1977). The protein of *traT* operates at the outer membrane, where modifications of the recipient cell membrane prevent the adhesion of transfer pili encoded by additional plasmids. This results in reduced abilities for the recipient to form mating aggregates with new donor cells (THOMAS and NIELSEN 2005).

As described in section 2.3.1, the initiation of conjugational transfer is dependent on a signal. This signal is proposed to be induced by TraG. After translocation of TraG from the donor to the recipient cell, contact with the inner membrane of the recipient initiates conjugation (AUDETTE *et al.* 2007). The plasmid-encoded *traS* can interfere with this process, thus blocking additional DNA uptake in the recipient (GARCILLÁN-BARCIA and DE LA CRUZ 2008).

The barrier of entry exclusion was shown not to be absolute, and uptake of new plasmids in a host was not completely excluded. However, a decrease of more than 500-fold caused by entry exclusion was observed for conjugation (PÉREZ-MENDOZA and DE LA CRUZ 2009)

3.3 Host Defence and Evolution

Defence mechanisms like RM systems and CRISPR are predicted to have an important influence on the uptake and establishment of DNA. It is therefore reasonable to think that their function can be reflected in the evolution of bacteria.

In was suggested by Budroni *et al.* that the evolution of bacteria was associated with the genomic content of RM systems (BUDRONI *et al.* 2011). The *Neisseria* genus serves as a paradigm for natural transformation, where exchange of genetic material is generally due to the persistent competence in the organism. Budroni *et al.* investigated 20 *Neisseria meningitidis* genomes, covering five serogroups of the recorded 13 for the species (ROSENSTEIN *et al.* 2001; CENTERS FOR DISEASE CONTROL AND PREVENTION 2012). They found that, for the phylogenetic network constructed based on the core genes; genomes from the same clonal complex were forming distinct phylogenetic clades. Further, they provided evidence that the phylogeny, thus evolution, in *N. meningitidis* could be associated with the RM systems identified in the genomes (BUDRONI *et al.* 2011).

In addition to the study by Budroni *et al*., the connection between RM systems and evolution has been elucidated for the organism *S. enterica* ssp. *enterica* (**Roer II**), which is known to generally

exchange genetic material through conjugation (FERGUSON *et al.* 2002). The association was investigated for 221 *S. enterica* genomes, covering 97 *S. enterica* ssp. *enterica* serovars, where the evolution was depicted as a phylogenetic network based on a core-genome three (Figure 2) and a pan-genome three (Figure 3). Assessing the content of RM systems in correlation to the two trees, we did not observe significant linkage, though we observed sub-linage correlation and serovar specific patterns in cases with e.g. *S.* Enteritidis and *S.* Typhymurium (**Roer II**). Additionally, we observed that plasmid replicons, SPIs, and AMR showed a better correlation to serovars than to RM systems. Thus, the study presented in **Roer II** suggests a limited influence of RM systems on the evolution of the conjugative organism *Salmonella enterica* ssp. *enterica* (**Roer II**), which correlates very well with the findings of RM systems not producing an absolute barrier for conjugation (**Roer I**).

In a recent study by Fricke *et al.* the evolutionary influence of CRISPR was investigated in *S. enterica* (FRICKE *et al.* 2011). From the 28 *S. enterica* genomes included in the study, the phylogeny was only partially reflected in the composition of the CRISPR arrays. However matches between CRISPR spacers and the content of plasmids and prophages could indicate a potential for CRISPR to be involved in sub-linage evolution (FRICKE *et al.* 2011).

Even though, barriers like RM systems seem to have an influence on the evolution of transformable organisms (BUDRONI *et al.* 2011), the indefinite barrier of RM systems in conjugation (**Roer I),** the lack of association between evolution and RM systems (**Roer II),** and the only partial association between CRISPR and evolution (FRICKE *et al.* 2011), implies that for conjugal organisms, other factors might be involved in shaping the evolution and importantly the spread of genetic traits like AMR.

Chapter 4

OTHER RECIPIENT DETERMINANTS AFFECTING CONJUGATION

Numerous studies have elucidated on environmental factors or exogenous components affecting conjugations (RICHAUME *et al.* 1989; ROCHELLE *et al.* 1989; FERNANDEZ-LOPEZ 2005; SEIER-PETERSEN *et al.* 2014). However, only few studies have investigated the contribution and ability of bacteria to act as recipients.

In 1974 Skurray *et al.* showed that some *E. coli* K-12 mutants were defective in conjugation, due to the lack of a major outer membrane protein (SKURRAY *et al.* 1974). The missing protein in the recipient precluded mating pair formation. Thus, the defective recipient inspired to the term "Con$^-$ mutant" (conjugation negative mutant). Further, evidence was provided, that different recipient cell surface components like LPS and the outer membrane protein OmpA were involved in Con$^-$ recipients for conjugation (HAVEKES *et al.* 1977; MANOIL and ROSENBUSCH 1982). However, mutants generated at that time, especially cell envelope mutants, were generated by random mutagens like ethylmethanesulfonate (EMS), which made it difficult to identify the exact site of mutation (HAVEKES *et al.* 1977; MANOIL and ROSENBUSCH 1982).

In a recent study, Pérez-Mendoza and de la Cruz tried to identify recipient genes in *E. coli* affecting the ability in plasmid conjugation, by using a more systematic approach (PÉREZ-MENDOZA and DE LA CRUZ 2009). By using the inducible *lux* system for conjugational detection (described in 2.3.1.1), they were capable of performing high-throughput screening in the attempt to identify recipient factors involved in conjugation. The *E. coli* mutant collection, the Keio collection (BABA

et al. 2006), enabled systematic screening of more than 99% of the non-essential genes from the *E. coli* K-12 genome. However, none of the 3,908 recipients tested, showed a dramatic drop in conjugation frequency compared to the wild-type *E. coli* K-12 strain BW25113. As a second attempt to identify recipient genes involved in conjugation, Pérez-Mendoza and de la Cruz constructed a transposon insertion mutant library of the DH5α *E. coli* strain. From this, they confirmed the previous findings of LPS being involved in conjugational transfer (PÉREZ-MENDOZA and DE LA CRUZ 2009). Therefore, they concluded that recipient bacterial cells cannot evade from being used as recipients in conjugation. However, their study was only based on two *E. coli* strains, both with wild-types showing reasonable conjugation abilities. Thus, these two isolates could potentially lack the restricting or enhancing abilities for conjugation.

As described in Chapter 3 with transfer defence mechanisms, recipient systems like RM and CRISPR do exist. However, RM systems was both shown not to be an absolute barrier in conjugation between isogenic *E. coli* strains (**Roer I**), and not to shape the evolution of *S. enterica* ssp. *enterica* (**Roer II)**. Thus, in **Roer III**, we aimed to identify other recipient determinants either enhancing or restricting plasmid uptake in *S. enterica* ssp. *enterica*. To overcome the drawback of not knowing if the isolates investigated, potentially could possess determinants either restricting or enhancing conjugation, the study was initiated with identifying the recipient potential of all isolates included. 93 isolates, covering 54 serovars were included in the study, to provide a broad genetic pool for investigation. From the initial conjugation, the isolates were subdivided into two groups of good (10 isolates) and poor (83 isolates) recipients. However, the comparative bioinformatic approaches applied could not yet identify gene candidates for neither enhancing nor restricting plasmid uptake. Thus other approaches should be considered in the identification of genes in *Salmonella* subspecies *enterica* that might control conjugation (**Roer III**). A parallel study on *S.* Enteritidis identified 33 gene candidates potentially restricting conjugation, which are being investigated and awaiting verification (**Roer III**).

Though, little is still known about recipient determinants affecting conjugational transfer. So far, membrane proteins have demonstrated to be involved in conjugation (SKURRAY *et al.* 1974; HAVEKES *et al.* 1977; MANOIL and ROSENBUSCH 1982). The findings in **Roer III**, with transfer frequencies from the same donor construct, varying from 0.0 to 2.5×10^{-1} transconjugants per potential recipients indicates that recipient determinants must exist (**Roer III**). This contradicts the conclusion by Pérez-Mendoza and de la Cruz, which states that in conjugation, bacteria cannot avoid being used as recipients (PÉREZ-MENDOZA and DE LA CRUZ 2009).

Chapter 5

CONCLUSION AND FUTURE PERSPECTIVES

For decades, it has been a dogma that RM systems act as a barrier for horizontal gene transfer. This has previously been demonstrated for transfer of DNA by transduction and transformation. The studies within this PhD thesis have illustrated that, the barriers imposed by the RM system *Eco*KI, are not absolute in conjugational transfer of an IncP-1 plasmid (**Roer I**). However, an inactivation of the restriction enzyme could cause a 7-fold increase in plasmid uptake. As the barriers of RM systems was demonstrated to be indefinite, it was interesting to investigate if the observed reduction in transfer after all could be reflected in the evolution of an organisms using conjugation as the preferred method in DNA exchange (**Roer II**). By developing and using the easily interpreted and user-friendly tools, *RM-Finder* and *SPI-Finder*, the PhD study demonstrated that the evolution of the conjugative organisms *S. enterica* ssp. *enterica*, was not associated with RM systems, and that, plasmid replicons and SPIs could not be reflected in the content of RM systems either. This led to the thought of other recipient determinants being involved in the control of plasmid uptake and in the evolution of bacteria (**Roer III**). By combining results obtained in the laboratory with bioinformatic approaches, we were able to predict 33 gene candidates restricting plasmid uptake in *S.* Enteritidis, which are being investigated and awaiting verification. However, we were not yet able to detect common genes on species level that potentially could control conjugation.

There is still an enormous lack of knowledge regarding the recipient determinants that control the uptake and establishment of plasmids. Thus, the global emergence and rapid spread of undesirable genetic traits are of significant importance, with antimicrobial resistance being an obvious example of such traits. The usage of antimicrobials in treatment of both human and animals has created a reservoir of resistant organisms and a pool of resistance genes available for further transmission.

One of the most striking and current examples is the rent observations of the plasmid-mediated *mcr-1* gene, which is conferring resistance towards colistin. A resistance which previously was mediated by chromosomal mutations with subsequent vertical heredity to daughter cells (HASMAN *et al.* 2015; LIU *et al.* 2015). However, the plasmid-mediated gene now enables rapid horizontal spread between bacteria. In Denmark, the gene was detected in an *E. coli* isolate only susceptible to a narrow range of antimicrobial classes (HASMAN *et al.* 2015). Thus, acquisition of additional resistances will leave a limited number of (if any) suitable options for treatment of bacterial infections. This emphasizes the importance of understanding the mechanisms and genes involved in controlling the uptake and maintenance of plasmids. Through improved knowledge we can hopefully predict and control the conjugational spread of generic threats in the future.

When elucidating on conjugational transfer, a big effort has been made on explaining the conjugative apparatus encoded on plasmids, and environmental settings affecting the transfer. However, little is known about the ability for bacteria to act as recipients. Different membrane proteins and barriers in the bacteria have been proposed to be involved in regulation of horizontal gene transfer. With new knowledge of RM systems not providing an absolute barrier, it would be interesting to elucidate on the mechanisms which allow some plasmids to evade the digestion of the restriction enzymes. Since most restriction enzymes works on double stranded DNA, one theory could be that the methyltransferase methylates some of the plasmids as the bacterial 'self', simultaneously with the synthesis of the stabilizing complementary strand right after transfer to the recipient, thus protecting the DNA from the restriction enzyme. This theory should be investigated to further improve the knowledge of the effect of RM systems in conjugation.

The development of benchtop sequencing technologies in the recent years has made whole-genome sequencing applicable in modern research, which also enabled a comprehensive analysis of a diverse dataset comprising 221 genomes in **Roer II**. With the continuous decrease in both purchasing and sequencing cost and the increased knowledge within this field, sequencing will most likely be standard in most research laboratories in the future. However, to enable practical usage of

the new technology, the knowledge obtained in the laboratory needs to be converted into bioinformatic tools like the *RM-Finder* and *SPI-Finder*. Currently, tools to predict plasmid replicons, antimicrobial resistance genes, virulence factors, serotypes etc. are developed (ZANKARI *et al.* 2012; CARATTOLI *et al.* 2014; JOENSEN *et al.* 2014, 2015), and over the next years we will probably see a great increase in the development of bioinformatic tools and usage. However, everything cannot be predicted by *in-silico* analysis alone, to improve our knowledge of mechanisms and bacterial behaviour we still need to do research in the laboratory. But the technology definitely has the potential in diagnostics and routine surveillance settings, where bacteria are screen for known characteristics.

Combining molecular technology and advanced bioinformatics is thus a new strategy in contemporary research, an approach pursued in **Roer III**. However, for now we were not capable of linking genes to neither enhanced or restricted conjugational abilities on species level, but a comprehensive investigation and characterization of the 93 isolates included in the study could potentially reveal a connection between isolates within the two groups of recipients. Another interesting analysis could be a SNP comparison of the genomes, thus unique SNPs within genes could explain a functional inactivation of proteins involved in conjugation. If the combination with bioinformatics cannot predict or answer why differences are observed in conjugational abilities, alternative approaches would be to turn the focus back to functional analysis in the laboratory. One such approach could be to screen the bacterial genes by performing random gene-knockouts of the genes in the recipient. However, the genes responsible for the recipient potential could as well be an essential gene for the bacteria, in which case a knock-out of the gene would be fatal for the bacteria. The knock-out construct should be accompanied by the construction of a transposon cloning library of the recipient. Different approaches combining the technologies, as well as sticking to molecular microbiology can be followed in the attempt to identify the recipient determinants that are involved in the control of conjugational uptake of plasmids.

In summary, this PhD study is improving the knowledge of RM systems and their impact on conjugation and evolution. Moreover, this PhD implies that recipient genes could be involved in controlling conjugation, and are leaves great opportunities for revealing of genes involved in plasmid uptake.

REFERENCES

ACHTMAN M., KENNEDY N., SKURRAY R., 1977 Cell--cell interactions in conjugating Escherichia coli: role of traT protein in surface exclusion. Proc. Natl. Acad. Sci. U. S. A. **74**: 5104–8.

ADAMCZYK M., JAGURA-BURDZY G., 2003 Spread and survival of promiscuous IncP-1 plasmids. Acta Biochim. Pol. **50**: 425–453.

ALLARD M. W., LUO Y., STRAIN E., LI C., KEYS C. E., SON I., STONES R., MUSSER S. M., BROWN E. W., 2012 High resolution clustering of Salmonella enterica serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **13**: 32.

ALLARD M. W., LUO Y., STRAIN E., PETTENGILL J., TIMME R., WANG C., LI C., KEYS C. E., ZHENG J., STONES R., WILSON M. R., MUSSER S. M., BROWN E. W., 2013 On the Evolutionary History, Population Genetics and Diversity among Isolates of Salmonella Enteritidis PFGE Pattern JEGX01.0004. PLoS One **8**.

AMANN R. I., LUDWIG W., SCHLEIFER K. H., 1995 Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59**: 143–169.

AMINOV R. I., 2011 Horizontal Gene Exchange in Environmental Microbiota. Front. Microbiol. **2**: 1–19.

ANDERSSON S. G. E., ZOMORODIPOUR A., ANDERSSON J. O., SICHERITZ-PONTÉN T., ALSMARK U. C., PODOWSKI R. M., NÄSLUND A. K., ERIKSSON A.-S., WINKLER H. H., KURLAND C. G., 1998 The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature **396**: 133–140.

AUDETTE G. F., MANCHAK J., BEATTY P., KLIMKE W. A., FROST L. S., 2007 Entry exclusion in F-like plasmids requires intact TraG in the donor that recognizes its cognate TraS in the recipient. Microbiology **153**: 442–51.

AVERHOFF B., FRIEDRICH A., 2003 Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. Arch. Microbiol. **180**: 385–393.

BABA T., ARA T., HASEGAWA M., TAKAI Y., OKUMURA Y., BABA M., DATSENKO K. a, TOMITA M., WANNER B. L., MORI H., 2006 Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. **2**: 2006.0008.

BAHL M. I., HANSEN L. H., GOESMANN A., SØRENSEN S. J., 2007 The multiple antibiotic resistance IncP-1 plasmid pKJK5 isolated from a soil environment is phylogenetically divergent from members of the previously established α, β and δ sub-groups. Plasmid **58**: 31–43.

BEOVIĆ B., 2006 The issue of antimicrobial resistance in human medicine. Int. J. Food Microbiol. **112**: 280–7.

BERG O. G., KURLAND C. G., 2002 Evolution of microbial genomes: sequence acquisition and loss. Mol. Biol. Evol. **19**: 2265–76.

BERGH O., BØRSHEIM K. Y., BRATBAK G., HELDAL M., 1989 High abundance of viruses found in aquatic environments. Nature **340**: 467–8.

BERTANI G., WEIGLE J., 1953 Host controlled variation in bacterial viruses. J. Bacteriol. **65**.

BICKLE T. a, KRÜGER D. H., 1993 Biology of DNA restriction. Microbiol. Rev. **57**: 434–50.

BINH C. T. T., HEUER H., KAUPENJOHANN M., SMALLA K., 2008 Piggery manure used for soil fertilization is a reservoir for transferable antibiotic resistance plasmids. FEMS Microbiol. Ecol. **66**: 25–37.

BLUM G., OTT M., LISCHEWSKI A., RITTER A., IMRICH H., TSCHÄPE H., HACKER J., 1994 Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an Escherichia coli wild-type pathogen. Infect. Immun. **62**: 606–14.

BOUCHER Y., DOUADY C. J., PAPKE R. T., WALSH D. a., BOUDREAU M. E. R., NESBØ C. L., CASE R. J., DOOLITTLE W. F., 2003 Lateral Gene Transfer and the Origins of Prokaryotic Groups. Annu. Rev. Genet. **37**: 283–328.

BOUET J.-Y., NORDSTRÖM K., LANE D., 2007 Plasmid partition and incompatibility - the focus shifts. Mol. Microbiol. **65**: 1405–1414.

BROWN KAV A., SASSON G., JAMI E., DORON-FAIGENBOIM A., BENHAR I., MIZRAHI I., 2012 Insights into the bovine rumen plasmidome. Proc. Natl. Acad. Sci. U. S. A. **109**: 5452–7.

BRÜSSOW H., HENDRIX R. W., 2002 Phage genomics: small is beautiful. Cell **108**: 13–6.

BRÜSSOW H., CANCHAYA C., HARDT W., 2004 Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. **68**: 560–602, table of contents.

BRÄUTIGAM M., HERTEL C., HAMMES W. P., 1997 Evidence for natural transformation of Bacillus subtilis in foodstuffs. FEMS Microbiol. Lett. **155**: 93–98.

BUDRONI S., SIENA E., DUNNING HOTOPP J. C., SEIB K. L., SERRUTO D., NOFRONI C., COMANDUCCI M., RILEY D. R., DAUGHERTY S. C., ANGIUOLI S. V, COVACCI A., PIZZA M., RAPPUOLI R., MOXON E. R., TETTELIN H., MEDINI D., 2011 Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination. Proc. Natl. Acad. Sci. U. S. A. **108**: 4494–9.

CAMPBELL a. M., 1992 Chromosomal insertion sites for phages and plasmids. J. Bacteriol. **174**: 7495–9.

CARATTOLI A., 2009 Resistance Plasmid Families in Enterobacteriaceae. Antimicrob. Agents Chemother. **53**: 2227–2238.

CARATTOLI A., ZANKARI E., GARCIÁ-FERNÁNDEZ A., LARSEN M. V., LUND O., VILLA L., AARESTRUP F. M., HASMAN H., 2014 In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. **58**: 3895–3903.

CENTERS FOR DISEASE CONTROL AND PREVENTION, 2012 *Manual for the Surveillance of Vaccine-Preventable Diseases* (SW Roush and LM Baldy, Eds.).

CHATTOPADHYAY S., WEISSMAN S. J., MININ V. N., RUSSO T. a, DYKHUIZEN D. E., SOKURENKO E. V, 2009 High frequency of hotspot mutations in core genes of Escherichia coli due to short-term positive selection. Proc. Natl. Acad. Sci. U. S. A. **106**: 12412–12417.

CHEN L., YANG J., YU J., YAO Z., SUN L., SHEN Y., JIN Q., 2005 VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. **33**: D325–D328.

CHEN J., NOVICK R. P., 2009 Phage-mediated intergeneric transfer of toxin genes. Science **323**: 139–41.

CHIBANI-CHENNOUFI S., BRUTTIN A., DILLMANN M.-L., BRUSSOW H., 2004 Phage-Host Interaction: an Ecological Perspective. J. Bacteriol. **186**: 3677–3686.

CHRISTENSEN B. B., STERNBERG C., MOLIN S., 1996 Bacterial plasmid conjugation on semi-solid surfaces monitored with the green fluorescent protein (GFP) from Aequorea victoria as a marker. Gene **173**: 59–65.

CONG L., RAN F. A., COX D., LIN S., BARRETTO R., HABIB N., HSU P. D., WU X., JIANG W., MARRAFFINI L. A., ZHANG F., 2013 Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-.). **339**: 819–823.

CURTISS R. I., CHARAMELLA L. J., STALLIONS D. R., MAYS J. A., 1968 Parental Functions During Conjugation in Escherishia coli K-12. Bacteriol. Rev. **32**: 320–348.

DAHLBERG C., BERGSTROM M., ANDREASEN M., CHRISTENSEN B. B., MOLIN S., HERMANSSON M., 1998 Interspecies bacterial conjugation by plasmids from marine environments visualized by gfp expression. Mol. Biol. Evol. **15 (4)**: 385–390.

DANG T. A., ZHOU X. R., GRAF B., CHRISTIE P. J., 1999 Dimerization of the Agrobacterium tumefaciens VirB4 ATPase and the effect of ATP-binding cassette mutations on the assembly and function of the T-DNA transporter. Mol. Microbiol. **32**: 1239–53.

DANOVARO R., DELL'ANNO A., CORINALDESI C., MAGAGNINI M., NOBLE R., TAMBURINI C., WEINBAUER M., 2008 Major viral impact on the functioning of benthic deep-sea ecosystems. Nature **454**: 1084–7.

DAVIES J., DAVIES D., 2010 Origins and evolution of antibiotic resistance. Microbiol. Mol. Biol. Rev. **74**: 417–33.

DOBRINDT U., HOCHHUT B., HENTSCHEL U., HACKER J., 2004 Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. **2**: 414–424.

DOLEJSKA M., VILLA L., HASMAN H., HANSEN L., CARATTOLI A., 2012 Characterization of IncN plasmids carrying bla CTX-M-1 and qnr genes in Escherichia coli and Salmonella from animals, the environment and humans. J. Antimicrob. Chemother. **68**: 333–339.

DOOLITTLE W. F., 1999 Phylogenetic Classification and the Universal Tree. Science (80-.). **284**: 2124–2128.

DOOLITTLE W. F., BOUCHER Y., NESBØ C. L., DOUADY C. J., ANDERSSON J. O., ROGER a J., 2003 How big is the iceberg of which organellar genes in nuclear genomes are but the tip? Philos. Trans. R. Soc. Lond. B. Biol. Sci. **358**: 39–57; discussion 57–8.

DOUDNA J. A., CHARPENTIER E., 2014 The new frontier of genome engineering with CRISPR-Cas9. Science (80-.). **346**: 1258096–1258096.

DUBNAU D., 1999 DNA Uptake in Bacteria. Annu. Rev. Microbiol. **53**: 217–44.

DUIJKEREN E. VAN, WANNET W. J. B., HECK M. E. O. C., PELT W. VAN, SLOET VAN OLDRUITENBORGH-OOSTERBAAN M. M., SMIT J. a H., HOUWERS D. J., 2002 Sero types, phage types and antibiotic susceptibilities of Salmonella strains isolated from horses in The Netherlands from 1993 to 2000. Vet. Microbiol. **86**: 203–12.

ELSAS J. D. VAN, MCSPADDEN GARDENER B. B., WOLTERS A. C., SMIT E., DIRK J. a N., GARDENER B. B. M. C. S., WOLTERS A. C., ELSAS J. D. VAN, MCSPADDEN GARDENER B. B., SMIT E., ELSAS J. D. Van, 1998 Isolation, characterization, and transfer of cryptic gene-mobilizing plasmids in the wheat rhizosphere. Appl. Environ. Microbiol. **64**: 880–889.

FERGUSON G. C., HEINEMANN J. A., KENNEDY M. A., 2002 Gene transfer between Salmonella enterica serovar Typhimurium inside epithelial cells. J. Bacteriol. **184**: 2235–42.

FERNANDEZ-LOPEZ R., 2005 Unsaturated fatty acids are inhibitors of bacterial conjugation. Microbiology **151**: 3517–3526.

FITZGERALD J. R., STURDEVANT D. E., MACKIE S. M., GILL S. R., MUSSER J. M., 2001 Evolutionary genomics of Staphylococcus aureus: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. U. S. A. **98**: 8821–8826.

FLUSBERG B. A., WEBSTER D., LEE J., TRAVERS K., OLIVARES E., CLARK A., KORLACH J., TURNER S. W., BIOSCIENCES P., DRIVE A., PARK M., 2010 Direct detection of DNA methylation during single-molecule, real- time sequencing. Nat. Methods **7**: 461–465.

FRANCINO M. P., 2012 The Ecology of Bacterial Genes and the Survival of the New. Int. J. Evol. Biol. **2012**: 1–14.

FRASER-LIGGETT C. M., 2005 Insights on biology and evolution from microbial genome sequencing. Genome Res. **15**: 1603–1610.

FRICKE W. F., MAMMEL M. K., MCDERMOTT P. F., TARTERA C., WHITE D. G., LECLERC J. E., RAVEL J., CEBULA T. a., 2011 Comparative genomics of 28 Salmonella enterica isolates: Evidence for crispr-mediated adaptive sublineage evolution. J. Bacteriol. **193**: 3556–3568.

FÜRSTE J. P., PANSEGRAU W., ZIEGELIN G., KRÖGER M., LANKA E., 1989 Conjugative transfer of promiscuous IncP plasmids: interaction of plasmid-encoded products with the transfer origin. Proc. Natl. Acad. Sci. U. S. A. **86**: 1771–5.

GAL-MOR O., FINLAY B. B., 2006 Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell. Microbiol. **8**: 1707–1719.

GARCILLÁN-BARCIA M. P., LA CRUZ F. DE, 2008 Why is entry exclusion an essential feature of conjugative plasmids? Plasmid **60**: 1–18.

GELDER L. DE, PONCIANO J. M., JOYCE P., TOP E. M., 2007 Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. Microbiology **153**: 452–463.

GIERALTOWSKI L., JULIAN E., PRINGLE J., MACDONALD K., QUILLIAM D., MARSDEN-HAUG N., SAATHOFF-HUBER L., STEIN D. VON, KISSLER B., PARISH M., ELDER D., HOWARD-KING V., BESSER J., SODHA S., LOHARIKAR a, DALTON S., WILLIAMS I., BARTON BEHRAVESH C., 2013 Nationwide outbreak of Salmonella Montevideo infections associated with contaminated imported black and red pepper: warehouse membership cards provide critical clues to identify the source. Epidemiol. Infect. **141**: 1244–52.

GILL S. R., FOUTS D. E., ARCHER G. L., MONGODIN E. F., DEBOY R. T., RAVEL J., PAULSEN I. T., KOLONAY J. F., BRINKAC L., BEANAN M., DODSON R. J., DAUGHERTY S. C., MADUPU R., ANGIUOLI S. V, DURKIN a S., HAFT D. H., VAMATHEVAN J., KHOURI H., UTTERBACK T., LEE C., DIMITROV G., JIANG L., QIN H., WEIDMAN J., TRAN K., KANG K., HANCE I. R., NELSON K. E., FRASER C. M., 2005 Insights on evolution of virulence and resistance from the complete

genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain. J. Bacteriol. **187**: 2426–38.

GOREN M., YOSEF I., EDGAR R., QIMRON U., 2012 The bacterial CRISPR/Cas system as analog of the mammalian adaptive immune system. RNA Biol. **9**: 549–54.

GRAHN A. M., HAASE J., BAMFORD D. H., LANKA E., 2000 Components of the RP4 conjugative transfer apparatus form an envelope structure bridging inner and outer membranes of donor cells: Implications for related macromolecule transport systems. J. Bacteriol. **182**: 1564–1574.

GROHMANN E., MUTH G., ESPINOSA M., 2003 Conjugative Plasmid Transfer in Gram-Positive Bacteria. Microbiol. Mol. Biol. Rev. **67**: 277–301.

GUERRY P., EMBDEN J. VAN, FALKOW S., 1974 Molecular nature of two nonconjugative plasmids carrying drug resistance genes. J. Bacteriol. **117**: 619–30.

GUGLIELMETTI S., MORA D., PARINI C., 2007 Small rolling circle plasmids in Bacillus subtilis and related species: Organization, distribution, and their possible role in host physiology. Plasmid **57**: 245–264.

GUTACKER M. M., SMOOT J. C., MIGLIACCIO C. a L., RICKLEFS S. M., HUA S., COUSINS D. V., GRAVISS E. a., SHASHKINA E., KREISWIRTH B. N., MUSSER J. M., 2002 Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: Resolution of genetic relationships among closely related microbial strains. Genetics **162**: 1533–1543.

HACKER J., BENDER L., OTT M., WINGENDER J., LUND B., MARRE R., GOEBEL W., 1990 Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal Escherichia coli isolates. Microb. Pathog. **8**: 213–25.

HACKER J., BLUM-OEHLER G., MÜHLDORFER I., TSCHÄPE H., 1997 Pathogenicity islands of virulent bacteria: Structure, function and impact on microbial evolution. Mol. Microbiol. **23**: 1089–1097.

HACKER J., CARNIEL E., 2001 Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep. **2**: 376–81.

HACKER J., HENTSCHEL U., DOBRINDT U., 2003 Prokaryotic chromosomes and disease. Science **301**: 790–3.

HAMILTON H. L., DILLARD J. P., 2006 Natural transformation of Neisseria gonorrhoeae: From DNA donation to homologous recombination. Mol. Microbiol. **59**: 376–385.

HANAHAN D., 1983 Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. **166**: 557–80.

HASMAN H., HAMMERUM A., HANSEN F., HENDRIKSEN R., OLESEN B., AGERSØ Y., ZANKARI E., LEEKITCHAROENPHON P., STEGGER M., KAAS R., CAVACO L., HANSEN D., AARESTRUP F., SKOV R., 2015 Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. Eurosurveillance **20**: Article 1.

HAVEKES L., TOMMASSEN J., HOEKSTRA W., LUGTENBERG B., 1977 Isolation and characterization of Escherichia coli K-12 F- mutants defective in conjugation with an I-type donor. J. Bacteriol. **129**: 1–8.

HENSEL M., 2004 Evolution of pathogenicity islands of Salmonella enterica. Int. J. Med. Microbiol. **294**: 95–102.

HEUER H., SMALLA K., 2007 Horizontal gene transfer between bacteria. Environ. Biosafety Res. **6**: 3–13.

HOFFMANN A., THIMM T., DRÖGE M., MOORE E. R., MUNCH J. C., TEBBE C. C., 1998 Intergeneric transfer of conjugative and mobilizable plasmids harbored by Escherichia coli in the gut of the soil microarthropod Folsomia candida (Collembola). Appl. Environ. Microbiol. **64**: 2652–9.

HOFFMANN M., LUO Y., LAFON P. C., TIMME R., ALLARD M. W., MCDERMOTT P. F., BROWN E. W., ZHAO S., 2013 Genome Sequences of Salmonella enterica Serovar Heidelberg Isolates Isolated in the United States from a Multistate Outbreak of Human Salmonella Infections. Genome Announc. **1**: 1–2.

HOFFMASTER A. R., RAVEL J., RASKO D. a, CHAPMAN G. D., CHUTE M. D., MARSTON C. K., DE B. K., SACCHI C. T., FITZGERALD C., MAYER L. W., MAIDEN M. C. J., PRIEST F. G., BARKER M., JIANG L., CER R. Z., RILSTONE J., PETERSON S. N., WEYANT R. S., GALLOWAY D. R., READ T. D., POPOVIC T., FRASER C. M., 2004 Identification of anthrax toxin genes in a Bacillus cereus associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. U. S. A. **101**: 8449–54.

HONG S.-H., BUNGE J., JEON S.-O., EPSTEIN S. S., 2006 Predicting microbial species richness. Proc. Natl. Acad. Sci. U. S. A. **103**: 117–122.

HORVATH P., BARRANGOU R., 2010 CRISPR/Cas, the immune system of bacteria and archaea. Science **327**: 167–170.

HYATT D., CHEN G.-L., LOCASCIO P. F., LAND M. L., LARIMER F. W., HAUSER L. J., 2010 Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**: 119.

INGRAM L. C., RICHMOND M. H., SYKES R. B., 1973 Molecular characterization of the R factors implicated in the carbenicillin resistance of a sequence of Pseudomonas aeruginosa strains isolated from burns. Antimicrob. Agents Chemother. **3**: 279–88.

JACOBY G. a., JACOB a. E., HEDGES R. W., 1976 Recombination between plasmids of incompatibility groups P-1 and P-2. J. Bacteriol. **127**: 1278–1285.

JAENECKE S., LORENZO V. DE, TIMMIS K. N., DIAZ E., 1996 A stringently controlled expression system for analysing lateral gene transfer between bacteria. Mol. Microbiol. **21**: 293–300.

JAIN R., RIVERA M. C., MOORE J. E., LAKE J. a, 2002 Horizontal gene transfer in microbial genome evolution. Theor. Popul. Biol. **61**: 489–495.

JENSEN E. C., SCHRADER H. S., RIELAND B., THOMPSON T. L., LEE K. W., NICKERSON K. W., KOKJOHN T. a., 1998 Prevalence of broad-host-range lytic bacteriophages of Sphaerotilus natans, Escherichia coli, and Pseudomonas aeruginosa. Appl. Environ. Microbiol. **64**: 575–80.

JOBANPUTRA R. S., DATTA N., 1974 Trimethoprim R Factors in Enterobacteria from Clinical Specimens. J. Med. Microbiol. **7**: 169–177.

JOENSEN K. G., SCHEUTZ F., LUND O., HASMAN H., KAAS R. S., NIELSEN E. M., AARESTRUP F. M., 2014 Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J. Clin. Microbiol. **52**: 1501–10.

JOENSEN K. G., TETZSCHNER A. M. M., IGUCHI A., AARESTRUP F. M., SCHEUTZ F., 2015 Rapid and easy *in silico* serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. J. Clin. Microbiol. **53**: JCM.00008–15.

JOHNSBORG O., ELDHOLM V., HÅVARSTEIN L. S., 2007 Natural genetic transformation: prevalence, mechanisms and function. Res. Microbiol. **158**: 767–778.

KHAN S. a., 2005 Plasmid rolling-circle replication: Highlights of two decades of research. Plasmid **53**: 126–136.

KING G., MURRAY N. E., 1994 Restriction enzymes in cells, not eppendorfs. Trends Microbiol. **2**: 465–469.

KLÜMPER U., RIBER L., DECHESNE A., SANNAZZARRO A., HANSEN L. H., SØRENSEN S. J., SMETS B. F., 2014 Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. ISME J.: 934–945.

KNAPP S., HACKER J., JARCHAU T., GOEBEL W., 1986 Large, unstable inserts in the chromosome affect virulence properties of uropathogenic Escherichia coli O6 strain 536. J. Bacteriol. **168**: 22–30.

KONSTANTINIDIS K. T., TIEDJE J. M., 2004 Trends between gene content and genome size in prokaryotic species with larger genomes. Proc. Natl. Acad. Sci. U. S. A. **101**: 3160–5.

KOONIN E. V, MAKAROVA K. S., ARAVIND L., 2001 Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. **55**: 709–42.

KORLACH J., BJORNSON K. P., CHAUDHURI B. P., CICERO R. L., FLUSBERG B. A., GRAY J. J., HOLDEN D., SAXENA R., WEGENER J., TURNER S. W., 2010 Real-Time DNA Sequencing from Single Polymerase Molecules. In: *Single Molecule Tools: Fluorescence Based Approaches, Part A*, Elsevier Inc., pp. 431–455.

KORLACH J., TURNER S. W., 2012 Going beyond five bases in DNA sequencing. Curr. Opin. Struct. Biol. **22**: 251–261.

KOSKINIEMI S., SUN S., BERG O. G., ANDERSSON D. I., 2012 Selection-driven gene loss in bacteria. PLoS Genet. **8**: 1–7.

KAAS R. S., FRIIS C., USSERY D. W., AARESTRUP F. M., 2012 Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genomics **13**: 577.

LANKA E., WILKINS B. M., 1995 DNA Processing Reactions in Bacterial Conjugation. Annu. Rev. Biochem. **64**: 141–169.

LAUDERDALE T., AARESTRUP F. M., CHEN P., LAI J., WANG H.-Y., SHIAU Y.-R., HUANG I.-W., HUNG C.-L., 2006 Multidrug resistance among different serotypes of clinical Salmonella isolates in Taiwan. Diagn. Microbiol. Infect. Dis. **55**: 149–55.

LEDERBERG J., 1952 Cell genetics and hereditary symbiosis. Physiol. Rev. **32**: 403–30.

LEDERBERG J., CAVALLI L. L., LEDERBERG E. M., 1952 Sex Compatibility in Escherichia Coli. Genetics **37**: 720–30.

LEEKITCHAROENPHON P., LUKJANCENKO O., FRIIS C., AARESTRUP F. M., USSERY D. W., 2012 Genomic variation in Salmonella enterica core genes for epidemiological typing. BMC

Genomics **13**: 88.

LESSL M., BALZER D., WEYRAUCH K., LANKA E., 1993 The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. J. Bacteriol. **175**: 6415–25.

LILLESTØL R. K., REDDER P., GARRETT R. A., BRÜGGER K., 2006 A putative viral defence mechanism in archaeal cells. Archaea **2**: 59–72.

LINN S., ARBER W., 1968 Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proc. Natl. Acad. Sci. U. S. A. **59**: 1300–6.

LIU Y.-Y., WANG Y., WALSH T. R., YI L.-X., ZHANG R., SPENCER J., DOI Y., TIAN G., DONG B., HUANG X., YU L.-F., GU D., REN H., CHEN X., LV L., HE D., ZHOU H., LIANG Z., LIU J.-H., SHEN J., 2015 Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet. Infect. Dis. **3099**: 1–8.

LOENEN W. a, DANIEL a S., BRAYMER H. D., MURRAY N. E., 1987 Organization and sequence of the hsd genes of Escherichia coli K-12. J. Mol. Biol. **198**: 159–70.

LORENZ M. G., WACKERNAGEL W., 1994 Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. **58**: 563–602.

MANOIL C., ROSENBUSCH J. P., 1982 Conjugation-deficient mutants of Escherichia coli distinguish classes of functions of the outer membrane OmpA protein. MGG Mol. Gen. Genet. **187**: 148–156.

MAZODIER P., DAVIES J., 1991 Gene transfer between distantly related bacteria. Annu. Rev. Genet. **25**: 147–171.

MEDINI D., DONATI C., TETTELIN H., MASIGNANI V., RAPPUOLI R., 2005 The microbial pan-genome. Curr. Opin. Genet. Dev. **15**: 589–594.

MEIJER W. J., BOER a J. DE, TONGEREN S. VAN, VENEMA G., BRON S., 1995 Characterization of the replication region of the Bacillus subtilis plasmid pLS20: a novel type of replicon. Nucleic Acids Res. **23**: 3214–3223.

MESELSON M., YUAN R., 1968 DNA restriction enzyme from E. coli. Nature.

MILLS D. M., BAJAJ V., LEE C. A., 2006 A 40 kb chromosomal fragment encoding Salmonella typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome. Mol. Microbiol. **15**: 749–759.

MIRA A., OCHMAN H., MORAN N. a., 2001 Deletional bias and the evolution of bacterial genomes. Trends Genet. **17**: 589–96.

MONOT M., HONORÉ N., GARNIER T., ARAOZ R., COPPÉE J.-Y., LACROIX C., SOW S., SPENCER J. S., TRUMAN R. W., WILLIAMS D. L., GELBER R., VIRMOND M., FLAGEUL B., CHO S.-N., JI B., PANIZ-MONDOLFI A., CONVIT J., YOUNG S., FINE P. E., RASOLOFO V., BRENNAN P. J., COLE S. T., 2005 On the origin of leprosy. Science **308**: 1040–1042.

MURRAY N., 2000 Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol. Mol. Biol. Rev. **64**.

MURRAY N. E., 2002 Immigration control of DNA in bacteria: Self versus non-self. Microbiology

148: 3–20.

MUSOVIC S., OREGAARD G., KROER N., SORENSEN S. J., 2006 Cultivation-Independent Examination of Horizontal Transfer and Host Range of an IncP-1 Plasmid among Gram-Positive and Gram-Negative Bacteria Indigenous to the Barley Rhizosphere. Appl. Environ. Microbiol. **72**: 6687–6692.

MAAMAR H., DUBNAU D., 2005 Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop. Mol. Microbiol. **56**: 615–24.

NAKATA N., TOBE T., FUKUDA I., SUZUKI T., KOMATSU K., YOSHIKAWA M., SASAKAWA C., 1993 The absence of a surface protease, OmpT, determines the intercellular spreading ability of Shigella: the relationship between the ompT and kcpA loci. Mol. Microbiol. **9**: 459–468.

NIELSEN K. M., ELSAS J. D. VAN, 2001 Stimulatory effects of compounds present in the rhizosphere on natural transformation of Acinetobacter sp. BD413 in soil. Soil Biol. Biochem. **33**: 345–357.

NOVAIS a., CANTON R., VALVERDE A., MACHADO E., GALAN J.-C., PEIXE L., CARATTOLI A., BAQUERO F., COQUE T. M., 2006 Dissemination and Persistence of blaCTX-M-9 Are Linked to Class 1 Integrons Containing CR1 Associated with Defective Transposon Derivatives from Tn402 Located in Early Antibiotic Resistance Plasmids of IncHI2, IncP1- , and IncFI Groups. Antimicrob. Agents Chemother. **50**: 2741–2750.

NOVICK R. P., 1987 Plasmid incompatibility. Microbiol. Rev. **51**: 381–395.

OOST J. VAN DER, JORE M. M., WESTRA E. R., LUNDGREN M., BROUNS S. J. J., 2009 CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci. **34**: 401–407.

PALLEN M. J., WREN B. W., 2007 Bacterial pathogenomics. Nature **449**: 835–842.

PARKHILL J., SEBAIHIA M., PRESTON A., MURPHY L. D., THOMSON N., HARRIS D. E., HOLDEN M. T. G., CHURCHER C. M., BENTLEY S. D., MUNGALL K. L., CERDEÑO-TÁRRAGA A. M., TEMPLE L., JAMES K., HARRIS B., QUAIL M. a, ACHTMAN M., ATKIN R., BAKER S., BASHAM D., BASON N., CHEREVACH I., CHILLINGWORTH T., COLLINS M., CRONIN A., DAVIS P., DOGGETT J., FELTWELL T., GOBLE A., HAMLIN N., HAUSER H., HOLROYD S., JAGELS K., LEATHER S., MOULE S., NORBERCZAK H., O'NEIL S., ORMOND D., PRICE C., RABBINOWITSCH E., RUTTER S., SANDERS M., SAUNDERS D., SEEGER K., SHARP S., SIMMONDS M., SKELTON J., SQUARES R., SQUARES S., STEVENS K., UNWIN L., WHITEHEAD S., BARRELL B. G., MASKELL D. J., 2003 Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat. Genet. **35**: 32–40.

PAULSEN I. T., BANERJEI L., MYERS G. S. a, NELSON K. E., SESHADRI R., READ T. D., FOUTS D. E., EISEN J. a, GILL S. R., HEIDELBERG J. F., TETTELIN H., DODSON R. J., UMAYAM L., BRINKAC L., BEANAN M., DAUGHERTY S., DEBOY R. T., DURKIN S., KOLONAY J., MADUPU R., NELSON W., VAMATHEVAN J., TRAN B., UPTON J., HANSEN T., SHETTY J., KHOURI H., UTTERBACK T., RADUNE D., KETCHUM K. a, DOUGHERTY B. a, FRASER C. M., 2003 Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. Science **299**: 2071–2074.

PEARSON T., BUSCH J. D., RAVEL J., READ T. D., RHOTON S. D., U'REN J. M., SIMONSON T. S., KACHUR S. M., LEADEM R. R., CARDON M. L., ERT M. N. VAN, HUYNH L. Y., FRASER C. M., KEIM P., 2004 Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. Proc. Natl. Acad. Sci. U. S. A. **101**: 13536–13541.

PÉREZ-MENDOZA D., LA CRUZ F. DE, 2009 Escherichia coli genes affecting recipient ability in plasmid conjugation: Are there any? BMC Genomics **10**: 71.

PERNA N. T., PLUNKETT G., BURLAND V., MAU B., GLASNER J. D., ROSE D. J., MAYHEW G. F., EVANS P. S., GREGOR J., KIRKPATRICK H. a, PÓSFAI G., HACKETT J., KLINK S., BOUTIN a, SHAO Y., MILLER L., GROTBECK E. J., DAVIS N. W., LIM a, DIMALANTA E. T., POTAMOUSIS K. D., APODACA J., ANANTHARAMAN T. S., LIN J., YEN G., SCHWARTZ D. C., WELCH R. a, BLATTNER F. R., 2001 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature **409**: 529–533.

RICHAUME A., ANGLE J. S., SADOWSKY M. J., 1989 Influence of soil variables on in situ plasmid transfer from Escherichia coli to Rhizobium fredii. Appl. Environ. Microbiol. **55**: 1730–4.

ROBERTS R. J., BELFORT M., BESTOR T., BHAGWAT A. S., BICKLE T. A., BITINATE J., BLUMENTHAL R. M., DEGTYAREV S. K., DRYDEN D. T. F., DYBVIG K., FIRMAN K., GROMOVA E. S., GUMPORT R. I., HALFORD S. E., HATTMAN S., HEITMAN J., HORNBY D. P., JANULAITIS A., JETSCH A., JOSEPHSEN J., KISS A., KLAENHAMMER T. R., KOBAYASHI I., KONG H., KRÜGER D. H., LACKS S., MARINUS M. G., MIYAHARA M., MORGAN R. D., MURRAY N. E., NAGARAJA V., PIEKAROWICZ A., PINGOUD A., RALEIGH E., RAO D. N., REICH N., REPIN V. E., SELKER E. U., SHAW P.-C., STEIN D. C., STODDARD B. L., SZYBALSKI W., TRAUTNER T. A., ETTEN J. L. VAN, VITOR J. M. B., WILSON G. G., XU S., 2003 A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. **31**: 1805–1812.

ROBERTS R. J., VINCZE T., POSFAI J., MACELIS D., 2015 REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. **43**: D298–D299.

ROCHELLE P. a, FRY J. C., DAY M. J., 1989 Factors affecting conjugal transfer of plasmids encoding mercury resistance from pure cultures and mixed natural suspensions of epilithic bacteria. J. Gen. Microbiol. **135**: 409–424.

ROSENSTEIN N. E., PERKINS B. A., STEPHENS D. S., POPOVIC T., HUGHES J. M., 2001 Meningococcal disease. N. Engl. J. Med. **344**: 1378–1388.

ROULI L., MBENGUE M., ROBERT C., NDIAYE M., SCOLA B. LA, RAOULT D., 2014 Genomic analysis of three African strains of Bacillus anthracis demonstrates that they are part of the clonal expansion of an exclusively pathogenic bacterium. New microbes new Infect. **2**: 161–9.

SAKAI H., KOMANO T., 1996 DNA Replication of IncQ Broad-host-range Plasmids in Gram-negative Bacteria. Biosci. Biotechnol. Biochem. **60**: 377–382.

SALAMA N., GUILLEMIN K., MCDANIEL T. K., SHERLOCK G., TOMPKINS L., FALKOW S., 2000 A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. Proc. Natl. Acad. Sci. U. S. A. **97**: 14668–14673.

SCALLAN E., HOEKSTRA R. M., ANGULO F. J., TAUXE R. V., WIDDOWSON M.-A., ROY S. L., JONES J. L., GRIFFIN P. M., 2011 Foodborne Illness Acquired in the United States—Major Pathogens. Emerg. Infect. Dis. **17**: 7–15.

SCHLÜTER A., SZCZEPANOWSKI R., PÜHLER A., TOP E. M., 2007 Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. FEMS Microbiol. Rev. **31**: 449–477.

SCHMIDT H., HENSEL M., 2004 Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol.

Rev. **17**: 14–56.

SCHUMANN W., LIPPS G., BAIKER A., HAASE R., LIPPS H. J., GRANT J. R., STOTHARD P., SOTA M., TOP E. M., RUSSI S., BOER R., COLL M., KANDAVELOU K., CHANDRASEGARAN S., LOWE D. B., SHEARER M. H., JUMPER C. A., KENNEDY R. C., ZHOU E.-M., 2008 *Plasmids: Current Research and Future Trends* (G Lipps, Ed.). Caister Academic Press.

SEIER-PETERSEN M. A., JASNI A., AARESTRUP F. M., VIGRE H., MULLANY P., ROBERTS A. P., AGERSO Y., 2014 Effect of subinhibitory concentrations of four commonly used biocides on the conjugative transfer of Tn916 in Bacillus subtilis. J. Antimicrob. Chemother. **69**: 343–348.

SEN D., AUWERA G. a. VAN DER, ROGERS L. M., THOMAS C. M., BROWN C. J., TOP E. M., 2011 Broad-Host-Range Plasmids from Agricultural Soils Have IncP-1 Backbones with Diverse Accessory Genes. Appl. Environ. Microbiol. **77**: 7975–7983.

SEN D., BROWN C. J., TOP E. M., SULLIVAN J., 2013 Inferring the Evolutionary History of IncP-1 Plasmids Despite Incongruence among Backbone Gene Trees. Mol. Biol. Evol. **30**: 154–166.

SHENDURE J., 2005 Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. Science (80-.). **309**: 1728–1732.

SHINTANI M., MATSUI K., INOUE J. -i., HOSOYAMA A., OHJI S., YAMAZOE A., NOJIRI H., KIMBARA K., OHKUMA M., 2014 Single-Cell Analyses Revealed Transfer Ranges of IncP-1, IncP-7, and IncP-9 Plasmids in a Soil Bacterial Community. Appl. Environ. Microbiol. **80**: 138–145.

SHINTANI M., SANCHEZ Z. K., KIMBARA K., 2015 Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front. Microbiol. **6**: 1–16.

SIKORSKI J., TESCHNER N., WACKERNAGEL W., 2002 Highly different levels of natural transformation are associated with genomic subgroups within a local population of Pseudomonas stutzeri from soil. Appl. Environ. Microbiol. **68**: 865–873.

SKURRAY R. A., HANCOCK R. E. W., REEVES P., 1974 Con − Mutants : Class of Mutants in Escherichia coli K-12 Lacking a Major Cell Wall Protein and Defective in Conjugation and Adsorption of a Bacteriophage. J. Bacteriol. **119**: 726–735.

SMALLA K., HAINES a. S., JONES K., KROGERRECKLENFORT E., HEUER H., SCHLOTER M., THOMAS C. M., 2006 Increased Abundance of IncP-1 Plasmids and Mercury Resistance Genes in Mercury-Polluted River Sediments: First Discovery of IncP-1 Plasmids with a Complex mer Transposon as the Sole Accessory Element. Appl. Environ. Microbiol. **72**: 7253–7259.

SMILLIE C., GARCILLAN-BARCIA M. P., FRANCIA M. V., ROCHA E. P. C., LA CRUZ F. DE, 2010 Mobility of Plasmids. Microbiol. Mol. Biol. Rev. **74**: 434–452.

SOLAR G. DEL, GIRALDO R., RUIZ-ECHEVARRÍA M. J., ESPINOSA M., DÍAZ-OREJAS R., 1998 Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. **62**: 434–464.

STEWART F. J., 2013 Where the genes flow. Nat. Geosci. **6**: 688–690.

SUTTLE C. A., 2007 Marine viruses — major players in the global ecosystem. Nat. Rev. Microbiol. **5**: 801–812.

SZPIRER C., TOP E., COUTURIER M., MERGEAY M., 1999 Retrotransfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance. Microbiology **145 (Pt 1**: 3321–9.

SØRENSEN S. J., JENSEN L. E., 1998 Transfer of plasmid RP4 in the spermosphere and rhizosphere of barley seedling. Antonie Van Leeuwenhoek **73**: 69–77.

SØRENSEN A., HANSEN L., 2003 Conjugative Plasmid Conferring Resistance to Olaquindox. Antimicrob. Agents Chemother. **47**: 798–799.

SØRENSEN S. J., SØRENSEN A. H., HANSEN L. H., OREGAARD G., VEAL D., 2003 Direct Detection and Quantification of Horizontal Gene Transfer by Using Flow Cytometry and gfp as a Reporter Gene. Curr. Microbiol. **47**: 129–133.

SØRENSEN S. J., BAILEY M., HANSEN L. H., KROER N., WUERTZ S., 2005 Studying plasmid horizontal transfer in situ: a critical review. Nat. Rev. Microbiol. **3**: 700–710.

TETTELIN H., MASIGNANI V., CIESLEWICZ M. J., DONATI C., MEDINI D., WARD N. L., ANGIUOLI S. V, CRABTREE J., JONES A. L., DURKIN a S., DEBOY R. T., DAVIDSEN T. M., MORA M., SCARSELLI M., MARGARIT Y ROS I., PETERSON J. D., HAUSER C. R., SUNDARAM J. P., NELSON W. C., MADUPU R., BRINKAC L. M., DODSON R. J., ROSOVITZ M. J., SULLIVAN S. a, DAUGHERTY S. C., HAFT D. H., SELENGUT J., GWINN M. L., ZHOU L., ZAFAR N., KHOURI H., RADUNE D., DIMITROV G., WATKINS K., O'CONNOR K. J. B., SMITH S., UTTERBACK T. R., WHITE O., RUBENS C. E., GRANDI G., MADOFF L. C., KASPER D. L., TELFORD J. L., WESSELS M. R., RAPPUOLI R., FRASER C. M., 2005 Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. U. S. A. **102**: 13950–13955.

THOMAS C., 1987 Incompatibility Group P Plasmids: Genetics, Evolution, And Use In Genetic Manipulation. Annu. Rev. Microbiol. **41**: 77–101.

THOMAS C. M., NIELSEN K. M., 2005 Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. **3**: 711–721.

TOCK M. R., DRYDEN D. T., 2005 The biology of restriction and anti-restriction. Curr. Opin. Microbiol. **8**: 466–472.

TOLKER-NIELSEN T., BRINCH U. C., RAGAS P. C., ANDERSEN J. B., JACOBSEN C. S., MOLIN S., 2000 Development and dynamics of Pseudomonas sp. biofilms. J. Bacteriol. **182**: 6482–6489.

TOP E. M., HOLBEN W. E., FORNEY L. J., 1995 Characterization of diverse 2,4-dichlorophenoxyacetic acid-degradative plasmids isolated from soil by complementation. Appl. Environ. Microbiol. **61**: 1691–8.

TOUCHMAN J. W., WAGNER D. M., HAO J., MASTRIAN S. D., SHAH M. K., VOGLER A. J., ALLENDER C. J., CLARK E. a., BENITEZ D. S., YOUNGKIN D. J., GIRARD J. M., AUERBACH R. K., BECKSTROM-STERNBERG S. M., KEIM P., 2007 A North American Yersinia pestisdraft genome sequence: SNPs and phylogenetic analysis. PLoS One **2**: 1–5.

TOUCHON M., HOEDE C., TENAILLON O., BARBE V., BAERISWYL S., BIDET P., BINGEN E., BONACORSI S., BOUCHIER C., BOUVET O., CALTEAU A., CHIAPELLO H., CLERMONT O., CRUVEILLER S., DANCHIN A., DIARD M., DOSSAT C., KAROUI M. El, FRAPY E., GARRY L., GHIGO J. M., GILLES A. M., JOHNSON J., BOUGUÉNEC C. LE, LESCAT M., MANGENOT S., MARTINEZ-JÉHANNE V., MATIC I., NASSIF X., OZTAS S., PETIT M. A., PICHON C., ROUY Z., RUF C. Saint, SCHNEIDER D., TOURRET J., VACHERIE B., VALLENET D., MÉDIGUE C., ROCHA E. P. C., DENAMUR E., 2009 Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths (J Casadesús, Ed.). PLoS Genet. **5**: e1000344.

USSERY D. W., WASSENAAR T. M., BORINI S., 2009 *Computing for Comparative Microbial Genomics: Bioinformatics for Microbiologists*.

VASU K., NAGAMALLESWARI E., NAGARAJA V., 2012 Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. Proc. Natl. Acad. Sci. U. S. A. **109**: E1287–E1293.

VEIGA H., PINHO M. G., 2009 Inactivation of the SauI Type I Restriction-Modification System Is Not Sufficient To Generate Staphylococcus aureus Strains Capable of Efficiently Accepting Foreign DNA. Appl. Environ. Microbiol. **75**: 3034–3038.

VIEIRA G., SABARLY V., BOURGUIGNON P.-Y., DUROT M., FEVRE F. LE, MORNICO D., VALLENET D., BOUVET O., DENAMUR E., SCHACHTER V., MEDIGUE C., 2011 Core and Panmetabolism in Escherichia coli. J. Bacteriol. **193**: 1461–1472.

WANG G., WILSON T. J. M., JIANG Q., TAYLOR D. E., 2001 Spontaneous Mutations That Confer Antibiotic Resistance in Helicobacter pylori. Antimicrob. Agents Chemother. **45**: 727–733.

WATANABE T., 1963 Infective heredity of multiple drug resistance in bacteria. Bacteriol. Rev. **27**: 87–115.

WERNEGREEN J. J., 2005 For better or worse: genomic consequences of intracellular mutualism and parasitism. Curr. Opin. Genet. Dev. **15**: 572–583.

WILLENBROCK H., HALLIN P. F., WASSENAAR T. M., USSERY D. W., 2007 Characterization of probiotic Escherichia coli isolates with a novel pan-genome microarray. Genome Biol. **8**: R267.

WILLIAMS R. J., 2003 Restriction Endonucleases: Classification, Properties, and Applications. Mol. Biotechnol. **23**: 225–243.

WOMMACK K. E., COLWELL R. R., 2000 Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. **64**: 69–114.

YOON S. H., PARK Y.-K., LEE S., CHOI D., OH T. K., HUR C.-G., KIM J. F., 2007 Towards pathogenomics: a web-based resource for pathogenicity islands. Nucleic Acids Res. **35**: D395–400.

YOON S. H., PARK Y.-K., KIM J. F., 2015 PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. Nucleic Acids Res. **43**: D624–D630.

ZANKARI E., HASMAN H., COSENTINO S., VESTERGAARD M., RASMUSSEN S., LUND O., AARESTRUP F. M., LARSEN M. V., 2012 Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. **67**: 2640–2644.

ZATYKA M., THOMAS C. M., 1998 Control of genes for conjugative transfer of plasmids and other mobile elements. FEMS Microbiol. Rev. **21**: 291–319.

ZECHNER E. L., LANG S., SCHILDBACH J. F., 2012 Assembly and mechanisms of bacterial type IV secretion machines. Philos. Trans. R. Soc. B Biol. Sci. **367**: 1073–1087.

ZEPH L. R., ONAGA M. a., STOTZKY G., 1988 Transduction of Escherichia coli by bacteriophage P1 in soil. Appl. Environ. Microbiol. **54**: 1731–7.

ZHAO Y., TEMPERTON B., THRASH J. C., SCHWALBACH M. S., VERGIN K. L., LANDRY Z. C., ELLISMAN M., DEERINCK T., SULLIVAN M. B., GIOVANNONI S. J., 2013 Abundant SAR11

viruses in the ocean. Nature **494**: 357–360.

ZHU Y.-G., JOHNSON T. A., SU J.-Q., QIAO M., GUO G.-X., STEDTFELD R. D., HASHSHAM S. A., TIEDJE J. M., 2013 Diverse and abundant antibiotic resistance genes in Chinese swine farms. Proc. Natl. Acad. Sci. **110**: 3435–3440.

ARTICLES

Roer I

The *Eco*KI Type I Restriction-Modification System in *Escherichia coli* Affects but Is Not an Absolute Barrier for Conjugation

Louise Roer, Frank M. Aarestrup, Henrik Hasman

Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark

Published

Journal of Bacteriology. 2015 Jan; 197(2): 337–342

JB
Journals.ASM.org

The EcoKI Type I Restriction-Modification System in *Escherichia coli* Affects but Is Not an Absolute Barrier for Conjugation

Louise Roer, Frank M. Aarestrup, Henrik Hasman

Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark

The rapid evolution of bacteria is crucial to their survival and is caused by exchange, transfer, and uptake of DNA, among other things. Conjugation is one of the main mechanisms by which bacteria share their DNA, and it is thought to be controlled by varied bacterial immune systems. Contradictory results about restriction-modification systems based on phenotypic studies have been presented as reasons for a barrier to conjugation with and other means of uptake of exogenous DNA. In this study, we show that inactivation of the R.EcoKI restriction enzyme in strain *Escherichia coli* K-12 strain MG1655 increases the conjugational transfer of plasmid pOLA52, which carriers two EcoKI recognition sites. Interestingly, the results were not absolute, and uptake of unmethylated pOLA52 was still observed in the wild-type strain (with an intact *hsdR* gene) but at a reduction of 85% compared to the uptake of the mutant recipient with a disrupted *hsdR* gene. This leads to the conclusion that EcoKI restriction-modification affects the uptake of DNA by conjugation but is not a major barrier to plasmid transfer.

The exchange of chromosomal and/or extrachromosomal DNA, such as plasmids, viruses, and transposons, is crucial for the evolution of bacteria and their ability to adapt to new environments. Exchange of genetic material occurs among both related and unrelated species of bacteria and is driven by the three horizontal gene transfer (HGT) mechanisms: conjugation, transformation, and transduction (1–3). Restriction-modification (RM) systems are described as major barriers to HGT (4–6) and comprise restriction endonucleases with a cognate methyltransferase. These recognize and cleave DNA not modified by the methyltransferase, thereby making the bacterium able to distinguish between its own (methylated) DNA and incoming nonmethylated DNA.

Based on their protein-complex subunit composition and functionality, RM systems can be divided into four types; this study focuses on type I. Type I systems require products of the three genes *hsdR* (restriction), *hsdM* (methylation), and *hsdS* (sequence specificity) and cleave randomly at a remote distance from the recognition sequence. Restriction occurs only when a protein complex of all three gene products (R_2M_2S) is formed, whereas methylation of the DNA requires formation of a complex of only the HsdM and HsdS proteins (M_2S) (7).

Some studies have indicated that transfer by conjugation is unaffected by RM systems but that unmodified phage or free DNA in transformation is readily degraded (8–11). This has led to the view that the conjugational transfer of plasmids through a single-stranded DNA intermediate is immune to restriction by RM systems, as the great majority of these recognize only nonmethylated double-stranded DNA (12–17). Other studies have, however, contradicted this. In 1964, Arber and Morse (18) proposed that host specificity (RM systems) might play a role in the acceptance or rejection of DNA transferred by conjugation in *Escherichia coli*. In many studies from the 1960s, the transfer was measured with recombinants of Hfr strains (18–20), but Arber and Morse made a phenotypic study showing that the conjugational transfer of episomes (with the ability to express genes without integration into the bacterial chromosome) was affected in the same manner as in phages (18). Other experiments have shown reduced conjugational transfer between different bacterial species with diverse restriction-modification systems, indicating that they may be the cause of this reduction in transfer (21, 22), but none of these observations have been confirmed with isogenic strains by modern molecular techniques. Recent studies indicated that SauI, a type I RM system for *Staphylococcus aureus*, may be a barrier to transfer into and between *S. aureus* isolates (23), but Veiga and Phino showed that inactivation of the SauI system was not sufficient for producing strains that efficiently take up foreign DNA (6), again questioning the importance of RM systems as barriers to conjugational transfer.

In the current study, we aimed to clarify the impact of a restriction-modification system in the conjugational gene transfer of single-stranded plasmidic DNA (24, 25) at the genotypic level. We focused on the impact of the type I RM system EcoKI, with the recognition sequence $AACN_6GTGC$, in the transfer of conjugative plasmids between RM variants of the *E. coli* K-12 strain MG1655. In addition to possessing the type I RM system, which we examine in the current study, MG1655 possess three different methylation-requiring type IV systems, EcoKMcrA, EcoKMcrBC, and EcoKMrr. All three systems are sequence specific and will not interfere with plasmid transfer between isogenic strains used in the current study (26–29). We provide evidence that the RM system EcoKI has a significant and negative effect on conjugation but also that this is not a major barrier to conjugation.

Received 20 October 2014 **Accepted** 1 November 2014

Accepted manuscript posted online 10 November 2014

Citation Roer L, Aarestrup FM, Hasman H. 2015. The EcoKI type I restriction-modification system in *Escherichia coli* affects but is not an absolute barrier for conjugation. J Bacteriol 197:337–342. doi:10.1128/JB.02418-14.

Editor: R. L. Gourse

Address correspondence to Louise Roer, lroe@food.dtu.dk.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.02418-14.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.02418-14

TABLE 1 Bacterial strains, plasmids, and oligonucleotides used in this study

Bacterial strain, plasmid, or oligonucleotide	Description (relevant genotype and/or phenotype) or sequence (5′ to 3′)[a]	Source or reference
Escherichia coli strains		
MG1655 (K-12 strain)	F$^-$ λ$^-$ *ilvG rfb-50 rph-1*	CGSC
MG1655-RN	Spontaneous Rif^r and Nal^r derivate of MG1655	This study
MG1655-A	Spontaneous $NaN_3{}^r$ derivate of MG1655	This study
MG1655-*hsdR*^{INT}	TargeTron insertion at nucleotide 1740|1741 of *hsdR*, Rif^r Nal^r	This study
MG1655-*hsdM*^{INT}	TargeTron insertion at nucleotide 720|721 of *hsdM*, Kan^r	This study
MG1655-*hsdR*^{COMPL}	MG1655-*hsdR*^{INT} with p*hsdR* for complementation, Rif^r Nal^r Tet^r	This study
MG1655-*hsdR*^{CONTROL}	MG1655-*hsdR*^{INT} with expression vector pMSC83, Rif^r Nal^r Tet^r	This study
DH10B	F$^-$ *endA1 recA1 galE15 galK16 nupG rpsL* Δ*lacX74* ϕ80*lacZ*ΔM15 *araD139* Δ(*ara leu*)*7697 mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ$^-$	Invitrogen
Plasmids		
pACD4K-C-loxP	Linearized TargeTron plasmid with a T7 promoter; Cam^r Kan-Δ*td*	Sigma
pACD4K-C-loxP (*hsdR*)	pACD4K-C-loxP retargeted for *hsdR* of MG1655 (LR1/LR2/LR3)	This study
pACD4K-C-loxP (*hsdM*)	pACD4K-C-loxP retargeted for *hsdM* of MG1655 (LR4/LR5/LR6)	This study
pAR1219	Expresses T7 RNA polymerase under the control of the IPTG-inducible *lac* UV5 promoter; Amp^r	Sigma
706-Cre	Expression plasmid for Cre recombinase driven by the thermosensitive promoter *c*I578; Tet^r	Gene Bridges GmbH
pOLA52	Plasmid of 45.7 kb with two restriction sites for *hsdR* (EcoKI); Amp^r	41
pHHA45	Plasmid of 51.6 kb without restriction sites for *hsdR* (EcoKI); Amp^r	31
pMSC83	Cloning vector used for complementation; Tet^r	This study
p*hsdR*	R.EcoKI from MG1655 cloned into pMSC83	This study
Oligonucleotides[b]		
LR1 (*hsdR* IBS)	AAAAAAGCTTATAATTATCCTTACATCGCGGCTATGTGCGCCCAGATAGGGTG	Sigma
LR2 (*hsdR* EBS1d)	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGGCTATATTAACTTACCTTTCTTTGT	Sigma
LR3 (*hsdR* EBS2)	TGAACGCAAGTTTCTAATTTCGGTTCGATGTCGATAGAGGAAAGTGTCT	Sigma
LR4 (*hsdR*-V-R)	TCCAGCTGGCTGCGGAACTGC	TAGC
LR5 (*hsdM* IBS)	AAAAAAGCTTATAATTATCCTTAGATTGCGCCGCCGTGCGCCCAGATAGGGTG	Sigma
LR6 (*hsdM* EBS1d)	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGCCGCCGTTAACTTACCTTTCTTTGT	Sigma
LR7 (*hsdM* EBS2)	TGAACGCAAGTTTCTAATTTCGATTCAATCTCGATAGAGGAAAGTGTCT	Sigma
LR8 (*hsdM*-V-F)	CCAATGATCTGGACGACCTT	TAGC
LR9 (*hsdR*-C-F)	GGTCATTGCCCGGAAAGGTA	TAGC
LR10 (*hsdR*-C-R)	GGCAGCCTGAAGGATGAAGT	TAGC

[a] For bacterial strains and plasmids, the relevant genotype, phenotype, and other characteristics are shown. Abbreviations: INT, interruption; COMPL, complementation.
[b] The genes targeted by the primers used in the construction of knockout strains or complementary plasmids are shown in parentheses at the end of the entry.

MATERIALS AND METHODS

Media and reagents. *E. coli* cells were cultured in brain heart infusion (BHI) broth at 37°C. For growth on agar, Luria-Bertani (LB) or BHI agar plates were used. The following antibiotics and concentrations were used: ampicillin (Amp), 50 μg/ml for cloning or 100 μg/ml in HGT assays; chloramphenicol (Cam), 25 μg/ml; kanamycin (Kan), 25 μg/ml; tetracycline (Tet), 5 μg/ml or 10 μg/ml; rifampin (Rif), 25 μg/ml; and nalidixic acid (Nal), 25 μg/ml.

Strains and plasmids. Bacterial strains, plasmids, and oligonucleotides used in this study are listed in Table 1. 32 The conjugation experiments were performed from donors with (*hsdM*$^+$) and without (Δ*hsdM*) the ability to methylate the DNA and with functional, disrupted, and complemented restriction abilities. Two plasmids with a published DNA sequence, pOLA52 and pHHA45 (GenBank accession numbers EU370913.1 and JX065630.1, respectively), were used to assess the influence of the RM systems on conjugational transfer. pOLA52, belonging to the IncXI incompatibility group, has two recognition sites for the EcoKI system, while pHHA45, belonging to the IncN incompatibility group, does not contain any sites. Both IncXI and IncN plasmids are known to transfer in the single-stranded form (24, 25, 30, 31).

Construction of TargeTron insertion mutants of *E. coli* MG1655. Two TargeTron mutants with interruptions, MG1655-*hsdR*^{INT} and MG1655-*hsdM*^{INT}, were created by following the guidelines from Sigma-Aldrich (32) for insertion mutations in *E. coli* strains, with the plasmid pAR1219 as the source of T7 RNA polymerase. Plasmid pACD4K-C-loxP was used as the donor for the group II intron, retargeted by PCR with primers designed for position 1740|1741 in *hsdR* (primers LR1, LR2, and LR3) and position 720|721 in *hsdM* (primers LR5, LR6, and LR7). Gene disruptions were induced by the addition of 20 μl of a 10 mM stock solution of isopropyl-β-D-thiogalactopyranoside (IPTG) to 2 ml of culture.

The plasmids (retargeted pACD4K-C-loxP and pAR1219) were cured by overnight growth in broth, followed by plating on BHI, and patched on BHI with 50 μg/ml Amp and BHI plus 25 μg/ml Cam to identify plasmid-free isolates. MG1655-*hsdR*^{INT} was made electrocompetent as described by D. O'Callaghan and A. Charbit (33), but the glycerol washing step was performed with the full-strength original volume (100 ml). MG1655-*hsdR*^{INT} was further transformed with the plasmid 706-Cre to remove the kanamycin resistance marker (Gene Bridges GmbH). Single colonies of both MG1655-*hsdM*^{INT} and MG1655-*hsdR*^{INT} were tested by PCR for the TargeTron insert and removal of the kanamycin gene (MG1655-*hsdR*^{INT} only) by using *Taq* polymerase (Fermentas) and internal and external primers LR3 and LR4 (*hsdR*) or LR7 and LR8 (*hsdM*). For further verification, the PCR products were purified using a GFX purification kit (GE Healthcare) and sequenced by Macrogen Korea.

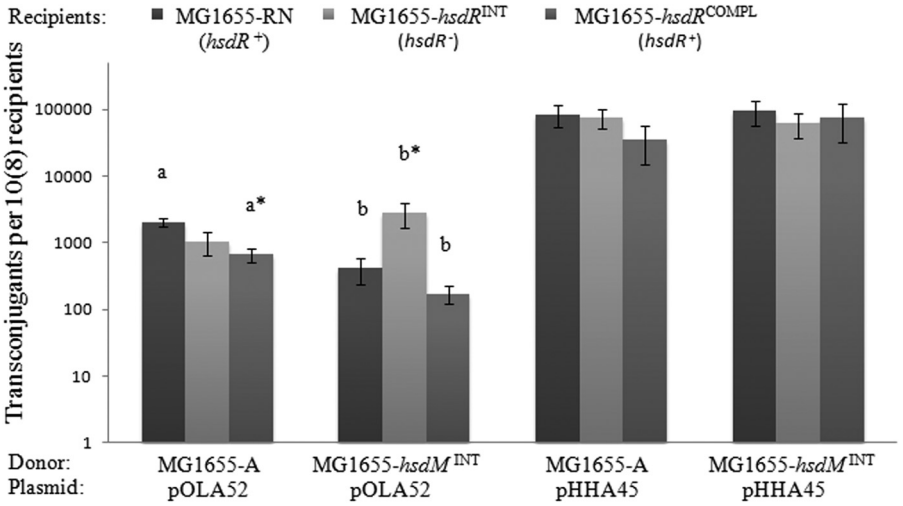


FIG 1 Single-stranded horizontal gene transfer with isogenic donor and recipient strains of *E. coli* K-12 MG1655. All experiments were performed at least in triplicate, and data are means of transconjugants per 10^8 recipients, with standard errors of the mean (SEM). The levels of transfer of the plasmids for the three recipients were compared individually for each donor (strain and plasmid). Statistically significant differences were observed between the recipients MG1655-*RN* and MG1655-*hsdR*^{COMPL} with the donor MG1655-*A* (a*, $P = 0.021$) and between MG1655-*hsdR*^{INT} and the remaining two recipients with the donor MG1655-*hsdM*^{INT} (b*, $P = 0.029$) with the plasmid pOLA52, which harbors two recognition sites for EcoKI. No statistically significant differences were observed with the two donors with the plasmid without recognition sites for EcoKI, pHHA45.

Complementation of *hsdR* with p*hsdR*. For complementation in the *hsdR* mutant, we amplified a 3,633-bp fragment encompassing the *hsdR* gene with the primers LR9 and LR10 by PCR and cloned the resulting fragment into the PvuII site of pMSC83. For a detailed description, see the supplemental material.

Horizontal gene transfer of single-stranded DNA by conjugation. The strains MG1655-*A* and MG1655-*hsdM*^{INT} were transformed with the plasmid pOLA52, containing two recognition sites for EcoKI, or pHHA45, lacking EcoKI recognition sites, to act as donors in the conjugation experiments. The strains MG1655-*RN*, with a functional *hsdR* gene on the chromosome, MG1655-*hsdR*^{INT}, with an interrupted restriction gene, and MG1655-*hsdR*^{COMPL}, with a complemented restriction gene, were used as recipients.

Two individual conjugation experiments were performed, one with transfer of pOLA52 and one with pHHA45. In both cases, overnight cultures of donor and recipients were reinoculated into fresh preheated BHI medium and grown to an optical density at 600 nm (OD_{600}) of 0.5. Then, 1-ml samples of each donor and recipient were mixed in 24-well microtiter plates and incubated at 37°C for 5 h. Conjugation mixtures were diluted and plated on selective plates for CFU counting.

RESULTS

Horizontal gene transfer of single-stranded DNA by conjugation. The results of the conjugational transfers are presented in Fig. 1.

All three recipients, MG1655-*RN*, MG1655-*hsdR*^{INT}, and MG1655-*hsdR*^{COMPL}, accepted the methylated plasmid pOLA52 from donor MG1655-*A* at high ratios (transconjugants per 10^8 recipients) of 2.05×10^3, 1.04×10^3, and 0.67×10^3, respectively. The level of conjugational transfer into the MG1655-*hsdR*^{COMPL} strain complemented with the *hsdR* gene in *trans* was significantly lower ($P = 0.021$) than in the MG1655-*RN* strain, with a wild-type functional *hsdR* gene on the chromosome, possibly due to the higher copy number, stronger promoter, or both.

For the transfer of the unmethylated plasmid pOLA52, with two recognition sites, from the donor MG1655-*hsdM*^{INT} to the three recipients, MG1655-*RN*, MG1655-*hsdR*^{INT}, and MG1655-*hsdR*^{COMPL}, the ratios were 0.42×10^3, 2.79×10^3, and 0.17×10^3, respectively. The recipient MG1655-*hsdR*^{INT} showed a statistically significantly ($P = 0.029$) higher level of transfer, which was more than 6.5 times higher than those of the two recipients with functional *hsdR* genes.

In the transfer of pHHA45, no significant difference was observed between the ratios of 8.45×10^4, 7.73×10^4, and 3.58×10^4 from the donor MG1655-*A* or between the ratios of 9.47×10^4, 6.25×10^4, and 7.66×10^4 from the donor MG1655-*hsdM*^{INT} to the recipients MG1655-*RN*, MG1655-*hsdR*^{INT}, and MG1655-*hsdR*^{COMPL}, respectively (see Table S1 in the supplemental material).

Complementation of *hsdR* restores restriction activity. The restriction gene *hsdR* was cloned into the expression vector pMSC83 under the control of the arabinose promoter pBAD, which is known to be leaky in rich media (34). The conjugation experiment was therefore performed without addition of arabinose to avoid overexpression of the *hsdR* gene, which could potentially be harmful to the cell. As a control for sufficient *hsdR* expression and to verify that the decrease in transfer observed in Fig. 1 was caused by expression of the *hsdR* gene alone and not the vector pMSC83, conjugation with the complemented strain and a control strain with the pMSC83 vector was performed. For the control experiment, the methylation-deficient donor MG1655-*hsdM*^{INT} was used with each plasmid (pHHA45 and pOLA52). The results of the conjugative control experiment are presented in Fig. 2.

In experiments with the unmethylated plasmids pHHA45 and pOLA52, the conjugational transfer of pOLA52 to MG1655-*hsdR*^{COMPL} was significantly decreased ($P = 0.00031$) compared with that to MG1655-*hsdR*^{CONTROL}, with transfer ratios of 0.21×10^3 and 5.89×10^3, respectively. With the transfer of pHHA45, no significant difference was observed between the transfer ratios of 5.70×10^4 and 3.37×10^4 for MG1655-*hsdR*^{COMPL} and MG1655-*hsdR*^{CONTROL}, respectively.

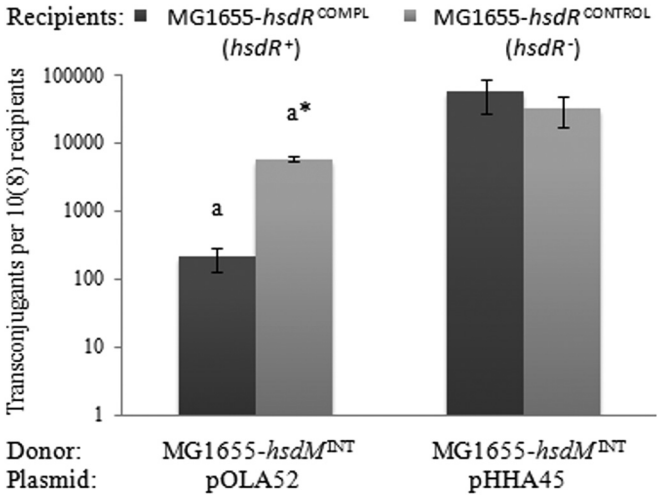


FIG 2 Conjugative effect of *hsdR* from the unmethylated donor strain MG1655-*hsdM*^{INT} inserted into the complemented strain MG1655-*hsdR*^{COMPL} compared to that in a control strain harboring the empty vector pMSC83. For conjugation with the unmethylated plasmid pHHA45, which lacks recognition sites for MG1655-*hsdR*, no significant difference was observed between the complementation and control strains. For conjugation with the plasmid pOLA52, which has two recognition sites, a significant difference was observed between the two recipients (a*, $P = 0.00031$).

DISCUSSION

Previous studies of the influence of RM systems in conjugation have generated conflicting conclusions in relation to the effect of RM systems on plasmid transfer (8–15). This has led to some controversy on how restriction-modification systems act on the uptake of single-stranded DNA, such as plasmids transferred by conjugation. An obvious driver of the conflicting conclusions from these studies is the fact that many of these studies were carried out in an era before the emergence of molecular techniques in microbiology and before the genetic determinants responsible for the RM phenotypes were identified.

The current study aimed to utilize isogenic strains and defined knockout genetic constructs to study how RM systems influence plasmidic transfer and showed that the type I restriction-modification system EcoKI in *E. coli* K-12 MG1655 affects conjugational transfer if the transferred DNA includes nonmethylated recognition sites.

Transfer of the methylated plasmid pOLA52 from the wild-type host to the three different recipients, with different restriction abilities, was not expected to have any significant influence on its uptake, as the plasmid was modified as "self" and in all three cases should have been immune from degradation. Transfer of the nonmethylated plasmid into the complemented strain was significantly lower than into the wild-type strain; more surprisingly, the same was true for the methylated plasmid. One explanation for this decrease might be incompatibility between pOLA52 and the vector carrying the complementation gene *hsdR*, but this was not expected, as the two replicons belong to IncXI and ColE10, respectively. This was also verified by the control experiment (Fig. 2), where the results clearly showed that the decrease in transfer was caused by the restriction gene alone and not the vector pMSC83. Therefore, a more likely explanation may be that overexpression of EcoKI, resulting in the formation of R_2M_2S complexes rather than M_2S complexes, leads to degradation of hemi-methylated DNA, as in the type II systems described by Nelson et al. (35).

The unmethylated plasmid pOLA52 was taken up significantly less in the wild-type strain than in strain MG1655-hsdR^{INT}, which has an interrupted restriction gene. The plasmid without recognition sites for EcoKI can transfer efficiently between donor and recipients independently of methylation and restriction abilities. This was shown by the transfer of the plasmid pHHA45 from both the methylation-deficient donor MG1655-*hsdM*^{INT} and the MG1655-*A* donor with the functional *hsdM* gene to three different recipients with functional, disrupted, and complemented restriction genes.

Even though the transfer efficiencies of the two plasmids, pOLA52 and pHHA45, differ by nearly 2 logs in the *hsdR*-disrupted recipient, we find pHHA45 an appropriate control plasmid, as both plasmids are narrow-host-range plasmids and about the same size. More plasmids could be investigated to confirm our observation that the conjugational transfer or uptake in *E. coli* K-12 strain MG1655 is in fact dependent not only on the donor's methylation and the recipient's restriction abilities but also on the presence of recognition sequences on the plasmid. The results of this genotypic study are in agreement with the study performed by Arber and Morse (18), where the RM-deficient strains were selected based on phenotypes by testing their ability to restrict different phages, but they were not further characterized.

From the present study, we have shown that a type I RM system can act as a barrier to the conjugational transfer of plasmids, but in many of the previous studies, the restriction is described only phenotypically (10, 18) or as a transfer between unrelated species (8, 11, 12) without our knowing the type of RM system involved. There is evidence that a type-III-like RM system may act as barrier to transformation in *S. aureus* strains (36) but not to conjugational transfer. Further, the present study focuses on only a single system, in a single isolate, with only two different plasmids. This leads to the questions of whether all type I systems influence conjugational transfer and whether the three remaining systems have the same ability to protect hosts from invading foreign DNA. Murray and colleagues suggest that the protection from foreign DNA might be altered by alleviation of chromosomal restriction genes, which might lead to uptake (37–39).

Pérez-Mendoza and de la Cruz (40) investigated two different knockout libraries, the Keio collection of single knockouts and a random insertion library, to determine how recipient cells contribute to bacterial conjugation (40). Their only finding was that the lipopolysaccharide (LPS) showed strong conjugation inhibition when conjugation was performed in liquid, but with filter mating, the reduction was restored. The Keio collection is based on a restriction-deficient K-12 strain variant (BW25113), which explains why they did not find EcoKI as a contributing factor in conjugation, as we did in this study.

Even though Pérez-Mendoza and de la Cruz did not find any genes responsible for the conjugational uptake in the recipient, this does not preclude the possibility that such genes exists. The two parental *E. coli* strains used in their study may be missing regulatory genes responsible for conjugational transfer. To identify possible barriers to conjugational uptake, as well as uptake by the other HGT mechanisms, good and poor recipients must be identified by phenotype and compared at the genetic level.

In summary, the EcoKI RM system found in *E. coli* K-12 strain MG1655 affects the conjugational transfer of plasmid pOLA52,

harboring two recognition sites, but the results imply that this effect is not absolute and that uptake is still possible, though at a lower level. The results showed 4.88-times-higher uptake of methylated pOLA52 than of the unmethylated plasmid in wild-type MG1655. This leads to the conclusion that plasmids with the same methylation pattern as the recipient can have a competitive advantage when entering a new host.

ACKNOWLEDGMENTS

This work was supported by the grant 11-106571 from the Danish Council for Independent Research/Technology and Production Sciences.

We thank Lina Cavaco and Mariane Schmidt for construction and sharing of the cloning vectors pBAD-Tet 2.1 and pMSC83.

REFERENCES

1. **Furuya E, Lowy F.** 2006. Antimicrobial-resistant bacteria in the community setting. Nat Rev Microbiol **4:**36–45. http://dx.doi.org/10.1038/nrmicro1325.
2. **Heuer H, Smalla K.** 2007. Horizontal gene transfer between bacteria. Environ Biosafety Res **6:**3–13. http://dx.doi.org/10.1051/ebr:2007034.
3. **Mazodier P, Davies J.** 1991. Gene transfer between distantly related bacteria. Annu Rev Genet **25:**147–171. http://dx.doi.org/10.1146/annurev.ge.25.120191.001051.
4. **Tock MR, Dryden DTF.** 2005. The biology of restriction and anti-restriction. Curr Opin Microbiol **8:**466–472. http://dx.doi.org/10.1016/j.mib.2005.06.003.
5. **Vasu K, Nagamalleswari E, Nagaraja V.** 2012. Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. Proc Natl Acad Sci U S A **109:**E1287–E1293. http://dx.doi.org/10.1073/pnas.1119226109.
6. **Veiga H, Pinho MG.** 2009. Inactivation of the SauI type I restriction-modification system is not sufficient to generate Staphylococcus aureus strains capable of efficiently accepting foreign DNA. Appl Environ Microbiol **75:**3034–3038. http://dx.doi.org/10.1128/AEM.01862-08.
7. **Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev SK, Dryden DTF, Dybvig K, Firman K, Gromova ES, Gumport RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jetsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Krüger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw P-C, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JMB, Wilson GG, Xu S.** 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res **31:**1805–1812. http://dx.doi.org/10.1093/nar/gkg274.
8. **Matsushima P, Broughton MC, Turner JR, Baltz RH.** 1994. Conjugal transfer of cosmid DNA from Escherichia coli to Saccharopolyspora spinosa: effects of chromosomal insertions on macrolide A83543 production. Gene **146:**39–45. http://dx.doi.org/10.1016/0378-1119(94)90831-1.
9. **Read TD, Thomas AT, Wilkins BM.** 1992. Evasion of type I and type II DNA restriction systems by Incl1 plasmid Collb-P9 during transfer by bacterial conjugation. Mol Microbiol **6:**1933–1941. http://dx.doi.org/10.1111/j.1365-2958.1992.tb01366.x.
10. **Stein DC, Gregoire S, Piekarowicz A.** 1988. Restriction of plasmid DNA during transformation but not conjugation in Neisseria gonorrhoeae. Infect Immun **56:**112–116.
11. **Eden PA, Blakemore RP.** 1991. Electroporation and conjugal plasmid transfer to members of the genus Aquaspirillum. Arch Microbiol **155:**449–452. http://dx.doi.org/10.1007/BF00244960.
12. **Schäfer A, Kalinowski J, Simon R, Seep-Feldhaus AH, Pühler A.** 1990. High-frequency conjugal plasmid transfer from gram-negative Escherichia coli to various gram-positive coryneform bacteria. J Bacteriol **172:**1663–1666.
13. **Elhai J, Vepritskiy A, Muro-Pastor AM, Flores E, Wolk CP.** 1997. Reduction of conjugal transfer efficiency by three restriction activities of Anabaena sp. strain PCC 7120. J Bacteriol **179:**1997–2005.
14. **Berndt C.** 2003. DNA restriction is a barrier to natural transformation in Pseudomonas stutzeri JM300. Microbiology **149:**895–901. http://dx.doi.org/10.1099/mic.0.26033-0.
15. **Dubnau D.** 1999. DNA uptake in bacteria. Annu Rev Microbiol **53:**217–244. http://dx.doi.org/10.1146/annurev.micro.53.1.217.
16. **Bischofberger N, Ng P.** 1987. Cleavage of single stranded oligonucleotides by EcoRI restriction endonuclease. Nucleic Acids **15:**709–716. http://dx.doi.org/10.1093/nar/15.2.709.
17. **Yoo O, Agarwal K.** 1980. Cleavage of single strand oligonucleotides and bacteriophage phi X174 DNA by MspI endonuclease. J Biol Chem **255:**10559–10562.
18. **Arber W, Morse ML.** 1964. Host specificity of DNA produced by Escherichia coli. VI. Effects on bacterial conjugation. Genetics **51:**137–148.
19. **Arber W.** 1965. Host-controlled modification of bacteriophage. Annu Rev Microbiol **19:**365–378. http://dx.doi.org/10.1146/annurev.mi.19.100165.002053.
20. **Wood WB.** 1966. Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA. J Mol Biol **16:**118–133. http://dx.doi.org/10.1016/S0022-2836(66)80267-X.
21. **Moser DP, Zarka D, Kallas T.** 1993. Characterization of a restriction barrier and electrotransformation of the cyanobacterium Nostoc PCC 7121. Arch Microbiol **160:**229–237.
22. **Purdy D, O'Keeffe TAT, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP.** 2002. Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier. Mol Microbiol **46:**439–452. http://dx.doi.org/10.1046/j.1365-2958.2002.03134.x.
23. **Waldron DE, Lindsay JA.** 2006. SauI: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into Staphylococcus aureus and between S. aureus isolates of different lineages. J Bacteriol **188:**5578–5585. http://dx.doi.org/10.1128/JB.00418-06.
24. **Willetts N, Wilkins B.** 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol Rev **48:**24–41.
25. **Bresler S, Lanzov V, Lukjaniec-Blinkova A.** 1968. On the mechanism of conjugation in Escherichia coli K 12. Mol Gen Genet **102:**269–284. http://dx.doi.org/10.1007/BF00433718.
26. **Raleigh EA.** 1992. Organization and function of the mcrBC genes of Escherichia coli K-12. Mol Microbiol **6:**1079–1086. http://dx.doi.org/10.1111/j.1365-2958.1992.tb01546.x.
27. **Heitman J, Model P.** 1987. Site-specific methylases induce the SOS DNA repair response in Escherichia coli. J Bacteriol **169:**3243–3250.
28. **Kelleher JE, Raleigh EA.** 1991. A novel activity in Escherichia coli K-12 that directs restriction of DNA modified at CG dinucleotides. J Bacteriol **173:**5220–5223.
29. **Waite-Rees PA, Keating CJ, Moran LS, Slatko BE, Hornstra LJ, Benner JS.** 1991. Characterization and expression of the Escherichia coli Mrr restriction system. J Bacteriol **173:**5207–5219.
30. **Norman A, Hansen LH, She Q, Sørensen SJ.** 2008. Nucleotide sequence of pOLA52: a conjugative IncX1 plasmid from Escherichia coli which enables biofilm formation and multidrug efflux. Plasmid **60:**59–74. http://dx.doi.org/10.1016/j.plasmid.2008.03.003.
31. **Dolejska M, Villa L, Hasman H, Hansen L, Carattoli A.** 2013. Characterization of IncN plasmids carrying bla CTX-M-1 and qnr genes in Escherichia coli and Salmonella from animals, the environment and humans. J Antimicrob Chemother **68:**333–339. http://dx.doi.org/10.1093/jac/dks387.
32. **Sigma-Aldrich.** 2011. TargeTron gene knockout system—user guide. Sigma-Aldrich, St. Louis, MO.
33. **O'Callaghan D, Charbit A.** 1990. High efficiency transformation of Salmonella typhimurium and Salmonella typhi by electroporation. Mol Gen Genet **223:**156–158. http://dx.doi.org/10.1007/BF00315809.
34. **Ham TS, Lee SK, Keasling JD, Arkin AP.** 2006. A tightly regulated inducible expression system utilizing the fim inversion recombination switch. Biotechnol Bioeng **94:**1–4. http://dx.doi.org/10.1002/bit.20916.
35. **Nelson PS, Papas TS, Schweinfest CW.** 1993. Restriction endonuclease cleavage of 5-methyl-deoxycytosine hemimethylated DNA at high enzyme-to-substrate ratios. Nucleic Acids Res **21:**681–686. http://dx.doi.org/10.1093/nar/21.3.681.
36. **Corvaglia AR, François P, Hernandez D, Perron K, Linder P, Schrenzel J.** 2010. A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical Staphylococcus aureus strains. Proc Natl Acad Sci U S A **107:**11954–11958. http://dx.doi.org/10.1073/pnas.1000489107.
37. **O'Neill M, Chen A, Murray NE.** 1997. The restriction-modification genes of Escherichia coli K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities. Proc Natl Acad Sci U S A **94:**14596–14601. http://dx.doi.org/10.1073/pnas.94.26.14596.
38. **O'Neill M, Powell LM, Murray NE.** 2001. Target recognition by EcoKI: the recognition domain is robust and restriction-deficiency commonly

results from the proteolytic control of enzyme activity. J Mol Biol **307:** 951–963. http://dx.doi.org/10.1006/jmbi.2001.4543.
39. **Blakely GW, Murray NE.** 2006. Control of the endonuclease activity of type I restriction-modification systems is required to maintain chromosome integrity following homologous recombination. Mol Microbiol **60:** 883–893. http://dx.doi.org/10.1111/j.1365-2958.2006.05144.x.
40. **Pérez-Mendoza D, de la Cruz F.** 2009. Escherichia coli genes affecting recipient ability in plasmid conjugation: are there any? BMC Genomics **10:**71. http://dx.doi.org/10.1186/1471-2164-10-71.
41. **Sørensen A, Hansen L.** 2003. Conjugative plasmid conferring resistance to olaquindox. Antimicrob Agents Chemother **47:**798–799. http://dx.doi.org/10.1128/AAC.47.2.798-799.2003.

Roer II

Is the Evolution of *Salmonella enterica* subspecies *enterica* Linked to Restriction-Modification Systems?

Louise Roer*, Rene S. Hendriksen*, Pimlapas Leekitcharoenphon*, Oksana Lukjancenko*, Rolf Sommer Kaas*, Henrik Hasman*†, Frank M. Aarestrup*

*Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark

† Section for microbiology and diagnostics, Statens Serum Institut, Copenhagen, Denmark

For Submission to mSystems

Supplemental material can be accessed at:
https://www.dropbox.com/sh/wz4v9rzem87eccn/AAAPJqt4m_TWX-Bhv7M_1PI_a?dl=0

Is the evolution of *Salmonella enterica* subspecies *enterica* Linked to Restriction-Modification Systems?

Running title: RMs partly linked to *S. enterica* evolution

Louise Roer*, Rene S. Hendriksen*, Pimlapas Leekitcharoenphon*, Oksana Lukjancenko*, Rolf Sommer Kaas*, Henrik Hasman*†, Frank M. Aarestrup*

*Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark

† Section for microbiology and diagnostics, Statens Serum Institut, Copenhagen, Denmark

Keywords: Restriction-Modification systems, evolution, *Salmonella* phylogenetic analysis, next generation sequencing, whole genome sequencing,

Corresponding author: Louise Roer
National Food Institute, Technical University of Denmark
Søltoft Plads, Building 221, DK-2800 Kgs. Lyngby, Denmark
E-mail: lroe@food.dtu.dk

ABSTRACT

Salmonella enterica ssp. *enterica* is a highly diverse foodborne pathogen that is sub-divided into more than 1,500 serovars. The diversity is believed result from mutational evolution as well as intra- and interspecies recombination that potentially could be influenced by Restriction-Modification systems (RM systems). The aim of this study was to investigate if RM systems were linked to the evolution of *Salmonella enterica* ssp. *enterica*.

The study included 221 *Salmonella enterica* genomes of which 68 were *de novo* sequenced, in addition to 153 public available genomes from ENA. The dataset covered 97 different serovars of *Salmonella enterica* ssp. *enterica,* and additional five genomes from four other *Salmonella* subspecies as an outgroup for constructing the phylogenetic trees. The phylogenetic trees were constructed based on similarity in core-genes as well as the presence or absence of pan-genes. The topology of the trees was compared to the presence of RM systems, antimicrobial resistance (AMR) genes, *Salmonella* Pathogenicity Islands (SPIs), and plasmid replicons.

We did not observe any significant correlation between the evolution and the RM systems in *S. enterica* ssp. *enterica*. However, sub-lineage correlations and serovar specific patterns was observed in few cases. Additionally, the results implied that plasmids replicons, SPIs, and AMR all were better correlated to serovars than to RM systems.

This study suggests a limited influence of RM systems on evolution of *Salmonella enterica* ssp. *enterica,* very likely due to the conjugational mode of horizontal gene transfer in *Salmonella*. Thus, other factors must be involved in shaping the evolution of bacteria.

IMPORTANCE

The evolution of bacterial pathogens, their plasticity and ability to constantly change and adapt to new surroundings are crucial for understanding the epidemiology and public health. Along with the

application of genomics, it became clear that horizontal gene transfer played a key role in evolution. Thus, to understand, prevent, fight, and control the spread of undesirable pathogens, we need to elucidate the processes that drive the horizontal gene transfer, the evolution and diversification. Restriction-modification systems are thought to cause rearrangements within the chromosome, as well as act as a barrier to horizontal gene transfer. However, here we show that the very persuasive correlation between restriction-modification systems and evolution in other bacterial species is not present in *Salmonella enterica* ssp. *enterica.* In summary, this work illustrates that other mechanisms must be involved in controlling the evolution.

INTRODUCTION

The *Salmonella* genus is highly diverse and comprised by the two species; *Salmonella enterica* and *Salmonella bongori*. The species *S. enterica* contains six subspecies of which the highly diverse *S. enterica* ssp. *enterica* can be sub-divided into more than 1,500 serovars. The subspecies is a common cause of diseases in humans and domestic animals (1, 2) and one of the leading causes of foodborne illness worldwide (3).

Recombination between genomes is thought to be a major driver in evolution (4), and to contribute to the diversity within the *Salmonella* genus (5, 6). In *Neisseria meningitidis* it was suggested that the phylogenetic relationship is associated with the content of Restriction-Modification systems (RM systems) (7). Further, rearrangements of genomes caused by RM systems are described as factors that could influence the evolution of pathogens (4). In addition to their role in rearrangements, RM systems are also considered to be a barrier for horizontal gene transfer between bacteria, thus serving as an immune defence system for uptake of foreign(4) DNA (8–10). The precise contribution have to our knowledge never been quantified, but we have recently shown that for conjugational transfer between isogenic *Escherichia coli isolates,* RM systems are not absolutely barrier (11).

RM systems are comprised of a restriction enzyme (RE) and a cognate methyltransferase. The restriction enzyme recognizes and digests foreign incoming DNA, whereas the methyltransferase perform methylation of the bacterium's own DNA to protect itself from degradation from the cognate restriction enzyme (12, 13). This enables the bacteria to distinguish between own (methylated) DNA and incoming non-methylated DNA.

The RM systems are divided into four types (I-IV), based on their protein-complexes, the sub-unit composition, and the functionality of the system (14). The Type I systems are complexes of three gene products; *hsd*R (R: restriction), *hsd*M (M: methylation), and *hsd*S (S: sequence specificity).

This type cleaves non-methylated DNA randomly at a remote distance from the recognition sequence determined by the specificity subunit. The protein-complex of all three gene products; R_2M_2S (two sub-units of R and M and one sub-unit of S) must be established prior to restriction, whereas a complex of only the *hsd*M and *hsd*S proteins (M_2S) is needed for methylation of the DNA (15). Type II systems are only made up of methyltransferases and restriction enzymes, where the function and composition of the M and R varies dependent on the sub-type of the system. Type II systems modulate (cleave and/or methylate) unmethylated DNA at specific recognition sites, making them suitable as molecular biological tools to cut DNA for cloning or other analysis where only a piece of DNA is needed (12, 15). Type III systems, consisting of the gene products Res and Mod, hemi-methylate the DNA and cleave DNA at specific sites 25-27 bp downstream of the recognition sequence (16), whereas Type IV, compared to Type I-III, does not encode a methyltransferase and only cleaves methylated DNA (12, 15).

In this study, we elucidated the potential association between RM systems and the phylogeny of *S. enterica* ssp. *enterica* serovars. We tested the hypothesis that RM systems might be linked to the evolution of *S. enterica* ssp. *enterica* and hereby be responsible for the diversification of the species. Most bacterial DNA exchange depends on mobile genetic elements (MGE), and in *Salmonella*; *Salmonella* pathogenicity islands (SPI) are believed to be acquired by horizontal gene transfer and have an effect on the structure of the genome (17, 18). Thus, we also elucidated the content of plasmids, antimicrobial resistance genes (AMR) and SPIs in correlation to the RM systems and the phylogeny of the species, to examine their effect on the evolution.

RESULTS

Genomes

A total of 221 *Salmonella* genomes were included in the analysis represented by 153 genomes previously described by Timme *et al.* (1) and retrieved from The European Nucleotide Archive. This collection was merged with 68 genomes sequenced as part of the 100K Foodborne Pathogen Genome Project (http://100kgenome.vetmed.ucdavis.edu/, NCBI BioProject PRJNA186441). The final collection consisted of 216 *S. enterica* ssp. *enterica* genomes and five genomes of other subspecies. The 221 *Salmonella* genomes were summarized in the supplemental material (Table S1).

Characterization of Restriction-Modification systems

To characterize the RM systems of the 221 genomes, a WGS analysis was performed using the newly developed tool Restriction-ModificationFinder 1.0. Partial systems were completed by individual BLAST analysis of up- and downstream sequences against the REBASE database (19). In total, we identified 113 putative RM systems i.e.; 58 Type I RM systems (TI) with 43 of unknown recognition sequence, 23 Type II RM systems (TII), two Type III RM system (TIII), and 30 Type IV RM systems (TIV), respectively. In addition, numerous methyltransferases outside RM systems were identified, including Type I, Type, II and Type III methyltransferases (Table S2). The additional methyltransferases (without associated restriction enzyme) identified were not include in the analysis, as the barrier for horizontal gene transfer (HGT) is thought to be caused by the cleavage from RE.

Two of the genomes only contained one RM system, where the other genomes contained between two and seven systems. One of the Type III RM systems were shared by 191 genomes, the type I system TI-1 was shared by 203 of the genomes and 37 RM systems were specific to a single

genome. The remaining systems were shared by two to 38 genomes. The distribution of the RM systems was illustrated in the presence/absence matrix in Fig. 1, and in details in Fig. S1. The analysis revealed very diverse content of RM systems, and assessing the highest level of discrimination in the cladogram, 120 distinct clusters were formed with 77 clusters containing a single genome.

The Salmonella Pan- and Core-Genomes

The pan- and core-genomes were estimated based on the 221 *S. enterica* genomes. The progression of the pan- and core-genomes is shown in Fig. S2, as increasing numbers of the *S. enterica* genomes were added to the analysis. Analysing the pan-genome, consisting of any gene families found, the gene families increased gradually with the one by one addition of the different *S. enterica* ssp. *enterica* serovars compared to a distinct increase in the number of gene families with the inclusion of *S. enterica* ssp. *diarizonae*. In contrast, the number of conserved gene families across the *S. enterica* genomes in the core-genome analysis seemed relative consistent but with similarly drops when introducing other *S. enterica* subspecies to the analysis. The final analysis compressing all 221 *S. enterica* genomes contained 16,375 gene families in the pan-genome (File S1) and 2,138 gene families in the core-genome (File S2), respectively. Analysing the total number of gene families in the pan-genome, each *S. enterica* genome contribute in average with 65 new gene families, increasing the diversity within the *S. enterica* species.

The link between evolution and Restriction-Modification systems

To study the genomic evolution of *S. enterica* ssp. *enterica*, differences within the core-genes were examined for all 221 genomes and illustrated by the phylogenetic core-genome tree in Fig. 2. This evolution of *S. enterica* ssp. *enterica* is not only formed by the differences in the genes shared by

the genomes, but also the loss of genes leading to differences in gene content. Fig. 3 displays the pan-genome tree based on the presence and absence of genes across all the genomes included in the study.

For both the core- and pan-genome trees, the high bootstrap values of 1 near the root of the trees reflected a general trustable structure for the phylogenetic trees of the *Salmonella* genus, whereas the low bootstrap values found in some of the branches likely indicated that some of the genomes in those branches were highly similar in their core- or pan-gene content, thus it was difficult to determine the definite arrangement within the branch. However, both the core- and pan-genome trees separated well the serovars amongst one another with relatively high bootstrap values at the tips of the trees.

To examine the hypothesis of RM systems being linked to the evolution of *S. enterica* ssp. *enterica*, parallel analyses were performed to identify groups of genomes with highly similar RM pattern that were also forming discrete phylogenetic clusters on the core- and pan-genome trees. Convergence was indicated by colours in the matrix of RM systems (Fig. 1), and the core- and pan-genome trees (Fig. 2 and Fig. 3). A few small clusters with almost identical RM content, partly clustering together in the core- and pan-genome trees were observed. In most cases the genomes within a cluster belonged to the same serovar, as for *S.* Bareilly or *S.* Agona (dark green and light blue). However, three larger clades, with 21, 8 and 7 genomes, were identified in the RM matrix. Of the 21 genomes in the first clade, containing serovars of Typhimurium, Saintpaul, Paratyphi B and Heidelberg, 90% of the genomes were located together in one distinct cluster in the core-genome tree and the remaining 10% were also located together. For the pan-genome tree, the 21 genomes were identified at 4 different locations with the distribution of 67%, 19%, 9% and 5% of the genomes, respectively. The two larger clades of 8 and 7 genomes were located adjacent to each other in the RM matrix, the first containing serovars of Stanleyville, Gallinarum, Pullorum and Dublin, and the

second containing serovars of Enteritidis and Berta. Comparing to the core- and pan-genome trees, in both cases the two clades were located together with all 100% of the genomes.

Plasmid Replicons, Antimicrobial resistance and Pathogenicity Islands in *Salmonella enterica*

All 221 *Salmonella* genomes were analysed for the content of plasmid replicons by using the Center for Genomic Epidemiology (CGE) web-tool PlasmidFinder 1.2 (20), of which 118 of the genomes did not contain any replicons present in the PlasmidFinder database. In the remaining 103 genomes, 40 different plasmid replicons were identified with each genome containing up to seven different replicons.

Assessing the replicons in comparison with the RM systems observed in the *S. enterica* ssp. *enterica* serovars, no convergence was observed (Fig. 4) and RM systems and plasmid replicons were never observed on the same contigs. Evaluating the plasmid replicons in correlation to serovars, no correlation was observed between serovars and quantity of replicons; the 11 genomes with the highest replicon content represented 10 different serovars. However, for multiple isolates with the same serovar common replicons are observed, such as with the *inc*FII and *inc*X1 replicons present in all the *S*. Dublin genomes. Additionally, *inc*FII was observed in 50% of the *S*. Enteritidis genomes, and together with the *inc*FIB replicon in 50% of the *S*. Typhimurium and in both 4,[5],12:i:- genomes. Interestingly, even though the numbers of replicons are not equal in identical serovars, the results imply a better association between replicons and serovars compared to the association between replicons and RM systems..

AMR genes were found in 220 of the genomes, varying from one to 19 different genes per genome, with *aac(6')-Iy* present as the only resistance gene in 140 of the genomes. The correlation of AMR and RM systems was examined (Fig. S3), however no congruency was observed most likely due to

196 the in-built bias of the strain collection. Despite this potential bias, the assessment of correlation
197 between AMR and plasmid replicons revealed different resistance genes located on the same
198 contigs as plasmid replicons. Few genomes contained more than four plasmid replicons and
199 between 10 and 16 resistance genes, correlating high resistance to the number of replicons present
200 in the genomes.
201
202 The presences of SPIs were assessed in all 221 *Salmonella* genomes utilizing the newly developed
203 web-tool SPI-Finder 1.0, and visualized in Fig. 5. SPIs were found in all 221 genomes, and the
204 number of SPI's in each genome varied from one to 14 islands/genes of islands.
205 The comparison of the SPIs and the content of RM systems revealed no clear association (Fig. 5).
206 However, associating the SPIs to individual *Salmonella* serovars such as *S. enterica ssp. enterica*
207 serovars Typhimurium, Paratyphi A, Choleraesuis, Heidelberg, and Saintpaul, some association was
208 apparent within a serovar.

DISCUSSION

For decades, RM systems have been recognized for their ability to act as "immune systems" for bacteria, determining whether or not foreign DNA was established in the cell. For *N. meningitidis* the impact of RM systems in evolution was recently elucidated, and the purpose of the current study was to clarify if a similar association could be identified in *S. enterica* ssp. *enterica*, by investigating a large subset of different whole genome sequenced *S. enterica* ssp. *enterica* serovars and an outgroup of five genomes from other subspecies. However, it was not possible to draw any significant association between RM systems and the overall evolution of *Salmonella enterica* ssp. *enterica,* even though we did observe genomes from the same RM clades in discrete phylogenetic clusters of both the core- and pan-genome trees.

The *Neisseria* genus, including *N. meningitidis*, is known to serve as a paradigm for natural transformation, where genetic exchange happens frequently due to their persistent competence independent of the phase of their life cycle (21). As RM systems have been recognized as barriers for transformation in multiple species (22–24), the link between a natural transformable species and the RM systems, as shown for *N. meningitidis*, seems reasonable. However transformation in *Salmonella* is unlikely as they are not natural transformable, and the transfer of genetic material happens mainly through conjugation (25). In addition, a recent study performed in *E. coli* (11) indicated that the barriers imposed by the RM systems in conjugational plasmid transfer were not absolute, which could explain the lack of significant association in our study when trying to associate RM systems to the plasmid replicons, AMR, SPI's, and their influence on the evolution.

The core- and pan-genome trees were constructed with a method previously described by Leekitcharoenphon *et al.* (26), where 73 genomes were evaluated. They found a core-genome of 2,882 genes and a cognate pan-genome of 10,581 genes. In our study, the core-genome was found to comprise 2,138 genes and the cognate pan-genome to contain 16,375 genes of the 221 genomes

assessed. This indicates an open pan-genome for *S. enterica* ssp. *enterica*, where addition of genomes to the analysis increases the total gene pool in the pan-genome, compared to the very uniform species *Bacillus anthracis*, having a closed genome, where the addition of genomes to the analysis will not increase the gene pool (27). This also supports the knowledge of *S. enterica* ssp. *enterica* being a highly diverse bacterial species.

Assessing the RM systems, we identified 113 RM systems and numerous individual methyltransferases, with each genome harbouring two to seven RM systems. Vasu and Nagaraja (4) recently described how changes in specificity or acquisition of new RM systems could alter the strains genetically from the original clonal population, as the methyltransferase modifies the genome of its new host, and the RE prevents genetic exchange between closely related strains. Thus, mutations accumulate in the "new strain", leading to genetic diversity. We found that 191 of the 221 genomes analysed shared a Type-III RM recognition sequence. Taking this result into consideration it is plausible that the Type-III system was the first RM system introduced in *S. enterica* ssp. *enterica* with subsequently a larger diversity following the later acquisition of new RM systems as described by Vasu and Nagaraja. However the same RM systems could also have been introduced in different branches at different time points, resulting in identical RM systems clade across the trees.

Assessing the pan- and core-genome trees, there are indications of some clustering of genomes with similar RM systems i.e. the cluster of *S.* Enteritidis and *S.* Dublin as well as the red RM systems clade consisting of *S.* Typhimurium, and *S.* Heidelberg located together in both trees. However, the influence is not significant, indicating that the evolution could be driven by several factors. For instance, a previous study compared 28 *Salmonella enterica* isolates and provided evidence that clustered regularly interspaced short palindromic repeats (CRISPR) mediated sub-lineage evolution (28). Other drivers in evolution are host and environmental adaptations, which besides gene

acquisition can be caused by gene loss and deletions, gene duplication and changes within genes by e.g. mutations (29, 30).

In the study on *N. meningitidis* they investigated the association of RM systems, homologous recombination and the phylogenetic network (7). The main study was performed on 20 genomes, covering five serogroups out of the recorded 13 serogroups for *N. meningitidis* (31, 32). Budroni *et al*. found that genomes from the same clonal complex (CC) were located together in phylogenetic clades based on their core genes. In addition, the clades could be associated with the RM systems identified (7). In our study, we investigated 217 genomes of *S. enterica* spp. *enterica* containing 97 different serovars with an outgroup of five genomes from four other subspecies. Considering the highly diverse dataset investigated in this study, compared to the one of *N. meningitidis*, we observed small trends of sub-lineage association of RM systems and evolution. This could indicate that even though the dataset investigated in this study was comprehensive, more genomes of each serovar should be included to cover the complete picture of the influence of RM systems in evolution of *S. enterica* ssp. *enterica*. Thus, with the current speed in WGS, this might be realistic in the nearest future. Thus, even in the ideal data scenario the lack of association is very likely due to the incomplete barrier of RM systems in conjugation.

As for all database dependent approaches, the methods are only capable of detecting and reporting records present in the database explored. Our analysis for detecting the RM systems was limited to the current knowledge presented in the REBASE (33), where the recognition sequences of various number of Type I specificity subunits were not yet determined. Thus it is likely that some strains have acquired RM systems with identical recognition pattern, however this is presently unknown. Thus, this is why the all database dependent analyses should be interpreted with care.

The plasmid replicons which potentially could have an effect on the bacterial diversity due to horizontal gene transfer were also identified, but no clear correlation between the RM systems and

the content of harboured plasmid replicons was observed. However, this approach might be complicated by the fact that plasmids are transferable and affected by factors such as fitness costs, selective pressure (34–37), and again the fact that the RM barrier is not absolute (11). Thus, the analysis performed on the plasmid replicons illustrates the current status of the time of isolation in contrast to what plasmids potentially could be acquired. AMR can be encoded by genes located on transferable plasmids, why this potentially could reflect the promiscuity of the genomes reflected in current time which could explain the lack of association between AMR and RM systems. This might also be explained by the possible biased dataset in respect of selective criteria e.g. susceptibility to antimicrobial agents, plasmid content or virulence (SPI) for the isolates. Though, the content of the RM systems is not believed to be affected by the possible biases.

The mechanism behind the acquisition of SPIs is horizontal gene transfer (17, 18). Nevertheless, the maintenance of SPIs within the genomes are considered stable (38), and is therefore a good measure of the barriers of RM systems, compared to plasmids replicons and AMR which can easily be lost if they do not confer any beneficial traits to the host. Despite this speculation, the influence of RM systems on the distribution of the SPI's was not supported by our analysis - on the contrary there were indications of some SPI's being serovar specific, which corresponds to previous findings (38).

In conclusion, recombination and rearrangement events caused by RM systems are, in several cases, described as driven factors for evolution, contributing to the diversity within a species (4, 39–44). However, high recombination between two distantly related lineages of *S. enterica* is exceptional (6, 45), thus explaining the difficulties of linking the RM systems to the evolution of *S. enterica* ssp. *enterica*. Thus, recombination occurs within and between closely related serovars (6).

In this study, we showed that RM systems could not be linked to the evolution of *S. enterica* ssp. *enterica,* very likely due to the incomplete barriers of RM systems in conjugation. However, we provided evidence of closely related serovars with identical RM systems; i.e. *S.* Dublin and *S.*

Enteritidis, suggesting that to elaborate further on the hypothesis of RM systems being involved in the evolution of *Salmonella enterica* ssp. *enterica*, either a collection of closely related serovars or a more comprehensive dataset with multiple representatives from each serovar could be assessed to expand on the hypothesis if the evolution of sub-groups of *S. enterica* ssp. *enterica* RM systems could have stronger links between their genomic evolution and the presence of RM systems compared to the lack of association for the entire subspecies *enterica*.

MATERIAL AND METHODS

Salmonella Strains

From an in-house strain collection at the Technical University of Denmark, National Food Institute (DTU FOOD), a sub-collection of 68 *S. enterica* ssp. *enterica* isolates with a global origin and a focus on multidrug resistance was submitted to the 100K Food Pathogen Genome Project (http://100kgenome.vetmed.ucdavis.edu/, NCBI BioProject PRJNA186441) for WGS. Subsequently, the genomes from that project were merged with a genomic collection consisting of 105 *Salmonella* strains; with a majority originating from the American Type Culture Collection – often pan-susceptible, sequenced by the Center for Food Safety and Applied Nutrition (FDA-CFSAN) and US Department of Argiculture (USDA) (1) and with 48 public available *Salmonella* genomes retrieved from the European Nucleotide Archive and included this study. The final dataset of 216 *S.* ssp. *enterica* genomes was constructed with focus on diversity including a total of 97 different *Salmonella* serovars. Additionally, five genomes of four other subspecies were included in the dataset forming an outgroup. This dataset might have an in-build bias in respect of the selective criteria e.g. susceptibility to antimicrobial agents. Full genomic information is shown in Table S1 in the supplemental material.

Whole genome sequencing

Genomic DNA was extracted from the 68 *Salmonella* isolates using an Invitrogen Easy-DNA kit (Invitrogen, Carlsbad, CA, USA), and DNA concentrations were determined using a Qubit double-stranded DNA (dsDNA) BR assay kit (Invitrogen). The genomic DNA was prepared for Illumina pair-end sequencing using Illumina (Illumina, Inc., San Diego, CA) NexteraXT Guide 150319425031942 and following protocol revision C (http://support.illumina.com/downloads/nextera_xt_sample _preparation_guide_15031942.html).

The DNA was dispatched for Illumina HiSeq whole genome sequencing at the School of Veterinary Medicine, UC Davis, USA in relation to the 100K Foodborne Pathogen Genome Project. The raw reads of the 68 sequenced genomes, received from UC Davis, were assembled using the Assembler 1.0 pipeline from the CGE available on http://cge.cbs.dtu.dk/services/all.php, which is based on the Velvet algorithms for *de novo* short-read assembly. A complete list of genomic sequence data is available in Table S1 in the supplemental material.

Construction of core-pan-genome plot and pan- and core-genome trees

Open reading frames (ORFs) were predicted on the contigs using Prodigal software (46) and same gene predictor was subsequently used to eliminate biases in annotation quality and to standardize the genes found in all genomes (47).

The predicted genes were translated into amino acid sequences and aligned all-against-all using BLASTP (48). Genes were in this study considered to belong to the same gene family if the alignment length was at least 50% of the longest sequence and more than 50% in similarity ("the 50/50 rule") (27).

Pan- and Core-Genome Plot

The core-pan-genome plot was constructed by comparing the gene families from all genomes. The pan-genome was constructed from the union of the genes from the genomes under consideration, while the core-genome was built from the intersection of genes families shared by every genome under analysis (26, 27).

Pan-genome Tree

The pan-genome tree was reconstructed from a matrix consisting of gene families (rows) and genomes (columns). In the matrix, the absence and presence of genes across the genomes were represented by 0's and 1's, respectively. The genomes were clustered using hierarchical clustering

of the relative Manhattan distance between genomes and the bootstrap values were calculated to represent the confidence of branches (26, 49).

Core-genome Tree

The core genes were aligned in a blast-like manner using BLAT v. 35 (50) to the predicted genes of each genome. The genes found in all genomes were then aligned using MUSCLE v. 3.8.31 (51) and concatenated to a single alignment. 500 resamples of the alignment were created with Seqboot version 3.67 (part of the PHYLIP package (52)).

A gene was considered "identified" according to the "50/50" rule. DNADist (52) was employed to calculate the genomic distances from the initial alignment as well as each of the 500 resamples. FastMe (53) was used to calculate trees from the distance matrices. The tree from the original alignment was compared to the 500 trees created from the resamples using CompareToBootstrap (54). The final tree was visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Construction of RM-Finder and SPI-Finder

To be able to analyse the genomes for their content of RM systems and SPI's, two public available online tools were developed. Both tools were build on a BLAST-based methodology for detection of genes from customized databases, originally developed by Zankari *et al.* for in silico detection of acquired resistance genes (55). The tools were developed to process both pre-assembled genomes and data of raw reads from different sequencing platforms, with user-selections parameters for minimum percent identity (%ID) and minimum length. The default settings were chosen as minimum ID at 95% to avoid noise and fragments of the genes, and a minimum length of 60% to be able to detect genes in the start or end of contigs from bad assemblies.

The database behind RM-Finder originate from the authoritative source REBASE (19, 33), and includes Type I-IV restriction genes, methyltransferases and specificity units. The database is categorised into two groups, one only including genes with confirmed function, and one where

putative genes are included. The RM-Finder database is monthly updated.

The *Salmonella enterica* records from the PAthogenicity Island DataBase (PAIDB) served as inspiration for the customized database behind SPI-Finder (56, 57).

Extensive explanations of the output can be assessed as separate tabs at the online tools; https://cge.cbs.dtu.dk/services/Restriction-ModificationFinder/ and https://cge.cbs.dtu.dk/services/SPIFinder/.

Identification of RMS-genes, Plasmid Replicons, SPI's and Antimicrobial resistance

To analyse the content of RM systems in the 221 genomes, all ORFs were submitted to the Restriction-ModificationFinder version 1.1 available on the CGE website (https://cge.cbs.dtu.dk/services/Restriction-ModificationFinder/), to identify restriction- (R), methyltransferase- (M) and specificity (S) genes. Subsequently, the R, M and S genes identified were individually inspected to form RM systems, and putative systems were assigned when all genes required were present and adjacent on the contig, even if truncated or frame-shifted. For systems with unknown specificity, systems were assigned according to the specificity subunit present. However, incomplete systems were investigated for truncated genes. Additionally, contigs with incomplete systems were inspected by BLAST against REBASE, and putative genes for completion were revealed. Based on the predicted recognition sequences, the systems were merged and named according to the type of system. A RM systems presence/absence matrix was constructed in R-2.14 (http://cran.r-project.org/bin/windows/base/old/2.14.0/) with hierarchical clustering, and euclidean distance (Fig. 1 and Fig. S1).

The 221 draft genomes were analyzed for the content of plasmid replicons, pathogenicity islands and antimicrobial resistance genes, by using the CGE web-tools; PlasmidFinder 1.1 (20), SPIFinder 1.0 (https://cge.cbs.dtu.dk/services/SPIFinder/) and ResFinder 2.0 (55) with %ID = 80.00 and

minimum length at 60.00%. The content of plasmid replicons and pathogenicity islands was interpreted by Circos plots (http://circos.ca/). For the antimicrobial resistance genes, a presence/absence matrix was constructed as described for the RM systems (Fig. S2).

ACKNOWLEDGEMENT

This work was supported by the grant 11-106571 from the Danish Council for Independent Research | Technology and Production Sciences.

REFERENCES

1. **Timme RE**, **Pettengill JB**, **Allard MW**, **Strain E**, **Barrangou R**, **Wehnes C**, **Van Kessel JS**, **Karns JS**, **Musser SM**, **Brown EW**. 2013. Phylogenetic diversity of the enteric pathogen Salmonella enterica subsp. enterica inferred from genome-wide reference-free SNP characters. Genome Biol. Evol. **5**:2109–23.

2. **Porwollik S**, **Boyd EF**, **Choy C**, **Cheng P**, **Florea L**, **Proctor E**, **McClelland M**. 2004. Characterization of Salmonella enterica subspecies I genovars by use of microarrays. J. Bacteriol. **186**:5883–98.

3. **Scallan E**, **Hoekstra RM**, **Angulo FJ**, **Tauxe R V.**, **Widdowson M-A**, **Roy SL**, **Jones JL**, **Griffin PM**. 2011. Foodborne Illness Acquired in the United States—Major Pathogens. Emerg. Infect. Dis. **17**:7–15.

4. **Vasu K**, **Nagaraja V**. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol. Mol. Biol. Rev. **77**:53–72.

5. **Octavia S**, **Lan R**. 2006. Frequent recombination and low level of clonality within Salmonella enterica subspecies I. Microbiology **152**:1099–1108.

6. **Didelot X**, **Bowden R**, **Street T**, **Golubchik T**, **Spencer C**, **McVean G**, **Sangal V**, **Anjum MF**, **Achtman M**, **Falush D**, **Donnelly P**. 2011. Recombination and Population Structure in Salmonella enterica. PLoS Genet. **7**.

7. **Budroni S**, **Siena E**, **Dunning Hotopp JC**, **Seib KL**, **Serruto D**, **Nofroni C**, **Comanducci M**, **Riley DR**, **Daugherty SC**, **Angiuoli S V**, **Covacci A**, **Pizza M**, **Rappuoli R**, **Moxon ER**, **Tettelin H**, **Medini D**. 2011. Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination. Proc. Natl. Acad. Sci. U. S. A. **108**:4494–9.

8. **Tock MR**, **Dryden DT**. 2005. The biology of restriction and anti-restriction. Curr. Opin. Microbiol. **8**:466–472.

9. **Vasu K**, **Nagamalleswari E**, **Nagaraja V**. 2012. Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. Proc. Natl. Acad. Sci. U. S. A. **109**:E1287–E1293.

10. **Veiga H**, **Pinho MG**. 2009. Inactivation of the SauI Type I Restriction-Modification System Is Not Sufficient To Generate Staphylococcus aureus Strains Capable of Efficiently Accepting Foreign DNA. Appl. Environ. Microbiol. **75**:3034–3038.

11. **Roer L**, **Aarestrup FM**, **Hasman H**. 2015. The EcoKI Type I Restriction-Modification System in Escherichia coli Affects but Is Not an Absolute Barrier for Conjugation. J. Bacteriol. **197**:337–42.

12. **Williams RJ**. 2003. Restriction Endonucleases: Classification, Properties, and Applications. Mol. Biotechnol. **23**:225–243.

13. **Murray N**. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol. Mol. Biol. Rev. **64**.

14. **Bickle T a**, **Krüger DH**. 1993. Biology of DNA restriction. Microbiol. Rev. **57**:434–50.

15. **Roberts RJ**, **Belfort M**, **Bestor T**, **Bhagwat AS**, **Bickle TA**, **Bitinate J**, **Blumenthal RM**, **Degtyarev SK**, **Dryden DTF**, **Dybvig K**, **Firman K**, **Gromova ES**, **Gumport RI**, **Halford SE**, **Hattman S**, **Heitman J**, **Hornby DP**, **Janulaitis A**, **Jetsch A**, **Josephsen J**, **Kiss A**, **Klaenhammer TR**, **Kobayashi I**, **Kong H**, **Krüger DH**, **Lacks S**, **Marinus MG**, **Miyahara M**, **Morgan RD**, **Murray NE**, **Nagaraja V**, **Piekarowicz A**, **Pingoud A**, **Raleigh E**, **Rao DN**, **Reich N**, **Repin VE**, **Selker EU**, **Shaw P-C**, **Stein DC**, **Stoddard BL**, **Szybalski W**, **Trautner TA**, **Van Etten JL**, **Vitor JMB**, **Wilson GG**, **Xu S**. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. **31**:1805–1812.

16. **Rao DN**, **Dryden DTF**, **Bheemanaik S**. 2014. Type III restriction-modification enzymes: A historical perspective. Nucleic Acids Res. **42**:45–55.

17. **Juhas M**, **van der Meer JR**, **Gaillard M**, **Harding RM**, **Hood DW**, **Crook DW**. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. **33**:376–93.

18. **Garcia-Vallve S**, **Romeu A**, **Palau J**. 2000. Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. Genome Res. **10**:1719–1725.

19. **Roberts RJ**, **Vincze T**, **Posfai J**, **Macelis D**. Rebase.

20. **Carattoli A**, **Zankari E**, **Garciá-Fernández A**, **Larsen MV**, **Lund O**, **Villa L**, **Aarestrup FM**, **Hasman H**. 2014. In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. **58**:3895–3903.

21. **Hamilton HL**, **Dillard JP**. 2006. Natural transformation of Neisseria gonorrhoeae: From DNA donation to homologous recombination. Mol. Microbiol. **59**:376–385.

22. **Moser DP**, **Zarka D**, **Kallas T**. 1993. Characterization of a restriction barrier and electrotransformation of the cyanobacterium Nostoc PCC 7121. Arch. Microbiol. **160**:229–237.

23. **Ando T**, **Xu Q**, **Torres M**, **Kusugami K**, **Israel D a**, **Blaser MJ**. 2000. Restriction-modification system differences in Helicobacter pylori are a barrier to interstrain plasmid transfer. Mol. Microbiol. **37**:1052–65.

24. **Monk I**, **Shah I**, **Xu M**, **Tan M**, **Foster T**. 2012. Transforming the untransformable: application of direct transformation to manipulate genetically Staphylococcus aureus and Staphylococcus epidermidis. MBio.

25. **Ferguson GC**, **Heinemann JA**, **Kennedy MA**. 2002. Gene transfer between Salmonella enterica serovar Typhimurium inside epithelial cells. J. Bacteriol. **184**:2235–42.

26. **Leekitcharoenphon P**, **Lukjancenko O**, **Friis C**, **Aarestrup FM**, **Ussery DW**. 2012. Genomic variation in Salmonella enterica core genes for epidemiological typing. BMC Genomics **13**:88.

27. **Tettelin H**, **Masignani V**, **Cieslewicz MJ**, **Donati C**, **Medini D**, **Ward NL**, **Angiuoli S V**, **Crabtree J**, **Jones AL**, **Durkin a S**, **Deboy RT**, **Davidsen TM**, **Mora M**, **Scarselli M**, **Margarit y Ros I**, **Peterson JD**, **Hauser CR**, **Sundaram JP**, **Nelson WC**, **Madupu R**, **Brinkac LM**, **Dodson RJ**, **Rosovitz MJ**, **Sullivan S a**, **Daugherty SC**, **Haft DH**, **Selengut J**, **Gwinn ML**, **Zhou L**, **Zafar N**, **Khouri H**, **Radune D**, **Dimitrov G**, **Watkins K**, **O'Connor KJB**, **Smith S**, **Utterback TR**, **White O**, **Rubens CE**, **Grandi G**, **Madoff LC**,

Kasper DL, **Telford JL**, **Wessels MR**, **Rappuoli R**, **Fraser CM**. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. U. S. A. **102**:13950–13955.

28. **Fricke WF**, **Mammel MK**, **McDermott PF**, **Tartera C**, **White DG**, **LeClerc JE**, **Ravel J**, **Cebula T a.** 2011. Comparative genomics of 28 Salmonella enterica isolates: Evidence for crispr-mediated adaptive sublineage evolution. J. Bacteriol. **193**:3556–3568.

29. **Fraser-Liggett CM**. 2005. Insights on biology and evolution from microbial genome sequencing. Genome Res.

30. **Pallen MJ**, **Wren BW**. 2007. Bacterial pathogenomics. Nature **449**:835–842.

31. **Centers for Disease Control and Prevention**. 2012. Manual for the Surveillance of Vaccine-Preventable DiseasesVPD Surveillance Manual.

32. **Rosenstein NE**, **Perkins BA**, **Stephens DS**, **Popovic T**, **Hughes JM**. 2001. Meningococcal disease. N. Engl. J. Med. **344**:1378–1388.

33. **Roberts RJ**, **Vincze T**, **Posfai J**, **Macelis D**. 2010. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. **38**:D234–6.

34. **Dahlberg C**, **Chao L**. 2003. Amelioration of the Cost of Conjugative Plasmid Carriage in Eschericha coli K12. Genetics **165**:1641–1649.

35. **Bergstrom CT**, **Lipsitch M**, **Levin BR**. 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics **155**:1505–1519.

36. **Platt TG**, **Bever JD**, **Fuqua C**. 2012. A cooperative virulence plasmid imposes a high fitness cost under conditions that induce pathogenesis. Proc. R. Soc. B Biol. Sci.

37. **Müller K**, **Aabo S**, **Birk T**, **Mordhorst H**, **Bjarnadóttir B**, **Agersø Y**. 2012. Survival and Growth of Epidemically Successful and Nonsuccessful <I>Salmonella</I> <I>enterica</I> Clones after Freezing and Dehydration. J. Food Prot. **75**:456–464.

38. **Hensel M**. 2004. Evolution of pathogenicity islands of Salmonella enterica. Int. J. Med. Microbiol. **294**:95–102.

39. **Dressler D**, **Potter H**. 1982. Molecular mechanisms in genetic recombination. Annu. Rev. Biochem. **51**:727–761.

40. **Milkman R**, **Raleigh E a.**, **McKane M**, **Cryderman D**, **Bilodeau P**, **McWeeny K**. 1999. Molecular evolution of the Escherichia coli chromosome. V. Recombination patterns among strains of diverse origin. Genetics **153**:539–554.

41. **King G**, **Murray NE**. 1994. Restriction enzymes in cells, not eppendorfs. Trends Microbiol. **2**:465–469.

42. **Chang S**, **Cohen SN**. 1977. In vivo site-specific genetic recombination promoted by the EcoRI restriction endonuclease. Proc. Natl. Acad. Sci. U. S. A. **74**:4811–4815.

43. **Arber W**. 2000. Genetic variation: Molecular mechanisms and impact on microbial evolution. FEMS Microbiol. Rev. **24**:1–7.

44. **Furuta Y**, **Abe K**, **Kobayashi I**. 2010. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. Nucleic Acids Res. **38**:2428–2443.

45. **Didelot X**, **Achtman M**, **Parkhill J**, **Thomson NR**, **Falush D**. 2007. A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: Convergence or divergence by homologous recombination? Genome Res. **17**:61–68.

46. **Hyatt D**, **Chen G-L**, **Locascio PF**, **Land ML**, **Larimer FW**, **Hauser LJ**. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119.

47. **Jacobsen A**, **Hendriksen RS**, **Aaresturp FM**, **Ussery DW**, **Friis C**. 2011. The Salmonella enterica Pan-genome. Microb. Ecol. **62**:487–504.

48. **Altschul S**, **Gish W**, **Miller W**, **Myers E**, **Lipman D**. 1990. Basic local alignment search tool. J Mol Biol **215**:403–410.

49. **Snipen L**, **Ussery DW**. 2010. Standard operating procedure for computing pangenome trees. Stand. Genomic Sci. **2**:135–141.

50. **Kent WJ**. 2002. BLAT — The BLAST -Like Alignment Tool. Genome Res. 656–664.

51. **Edgar RC**. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

52. **Felsenstein J**. 1989. PHYLIP - Phylogeny Inference Package. 3.6. Cladistics **5**:164–166.

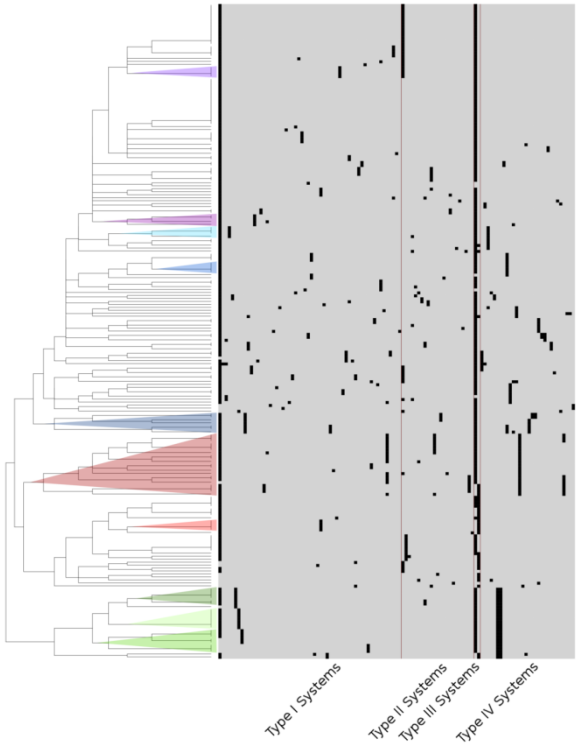
53. **Desper R**, **Gascuel O**. 2002. Fast and accurate phylogeny minimum-evolution principle. J Comput Biol **9**:687–705.

54. **Price MN**, **Dehal PS**, **Arkin AP**. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One **5**:e9490.

55. **Zankari E**, **Hasman H**, **Cosentino S**, **Vestergaard M**, **Rasmussen S**, **Lund O**, **Aarestrup FM**, **Larsen MV**. 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. **67**:2640–2644.

56. **Yoon SH**, **Park Y-K**, **Lee S**, **Choi D**, **Oh TK**, **Hur C-G**, **Kim JF**. 2007. Towards pathogenomics: a web-based resource for pathogenicity islands. Nucleic Acids Res. **35**:D395–400.

57. **Yoon SH**, **Park Y-K**, **Kim JF**. 2015. PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. Nucleic Acids Res. **43**:D624–D630.



564

565 **Fig. 1. Presence or absence matrix of the 113 restriction-modification systems.**

566 In the matrix, each row represents one genome analysed, and each column represents one of 113 RM systems. The

567 cladogram is a hierarchical clustering of the genomes based on the Euclidean algorithm. Convergence of genomes with

568 highly similar RM system content and discrete phylogenetic clades on the core- and pan-genome trees was indicated by

569 the colours in the cladogram.

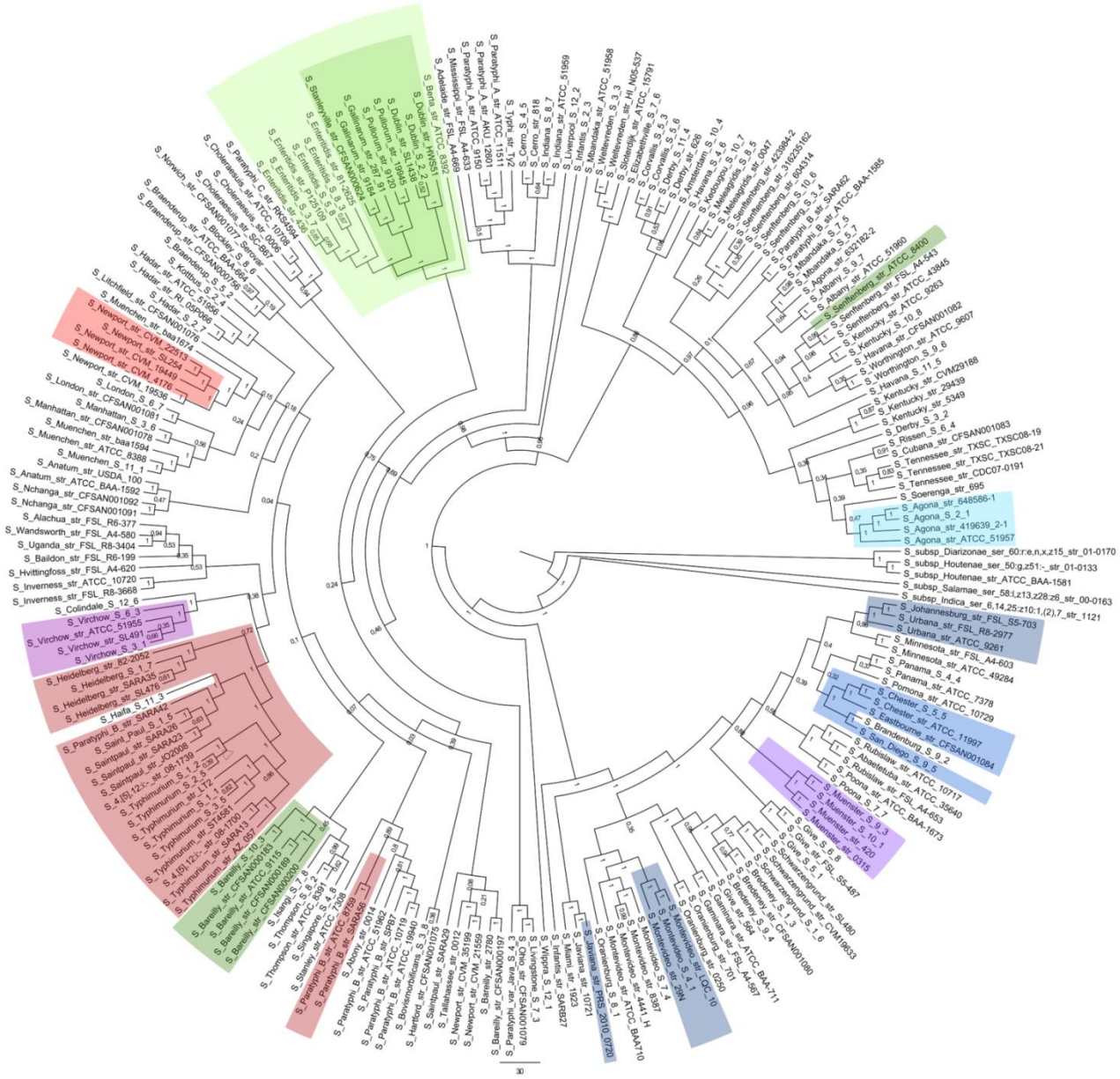


Fig. 2. Concatenated core-genome tree of *Salmonella enterica* serovars constructed on 1,072 core gene clusters. Phylogenetic tree constructed on core genes. Discrete phylogenetic clades with highly identical RM system content were indicated by the colours defined in Fig. 1.

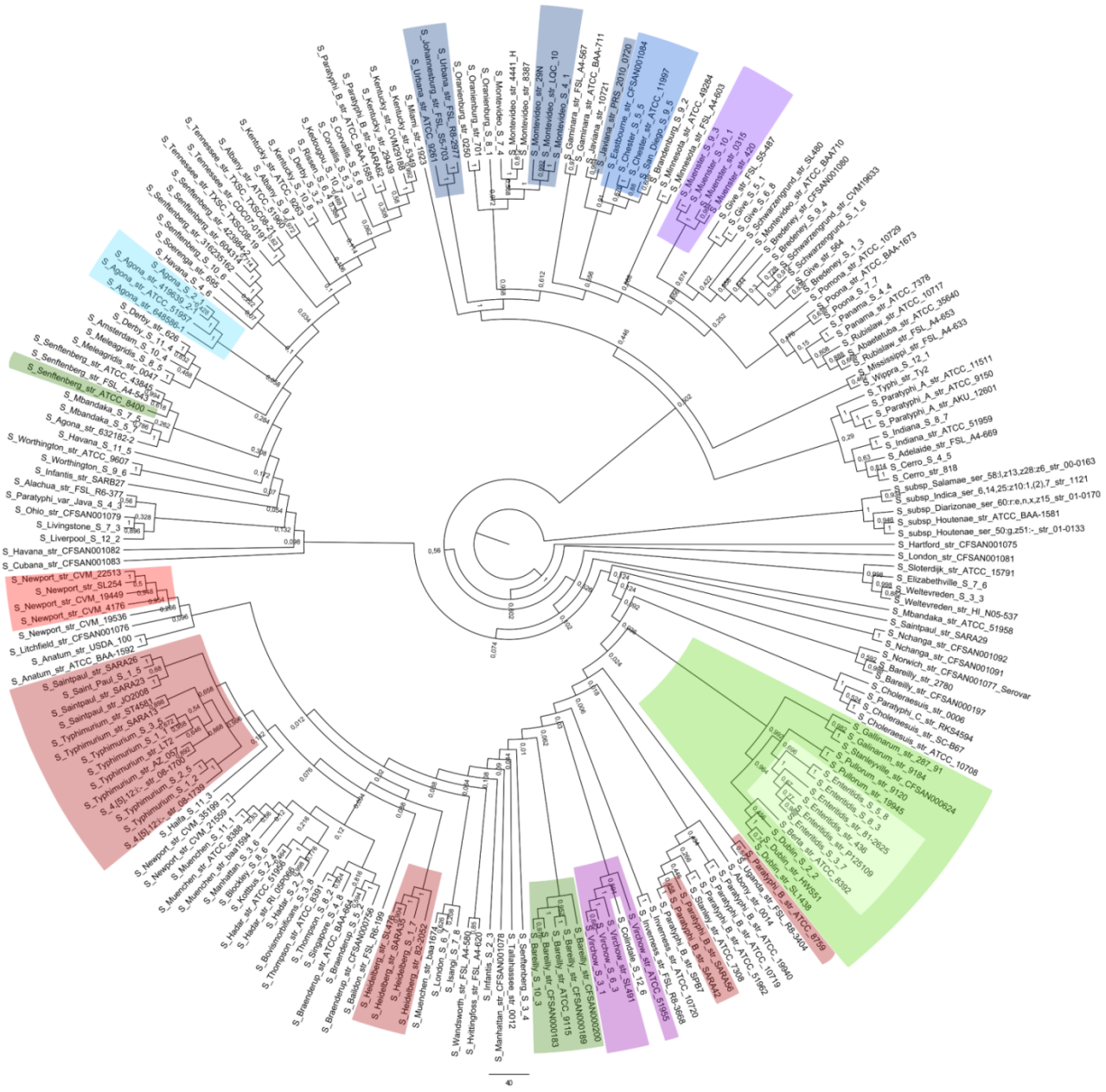


Fig. 3. Pan-genome tree of *Salmonella enterica* serovars with colour indicated RM systems. Phylogenetic tree constructed from presence/absence matrix of genes across genomes. The colours represent the different groups of RM systems defined in Fig. 1**.**

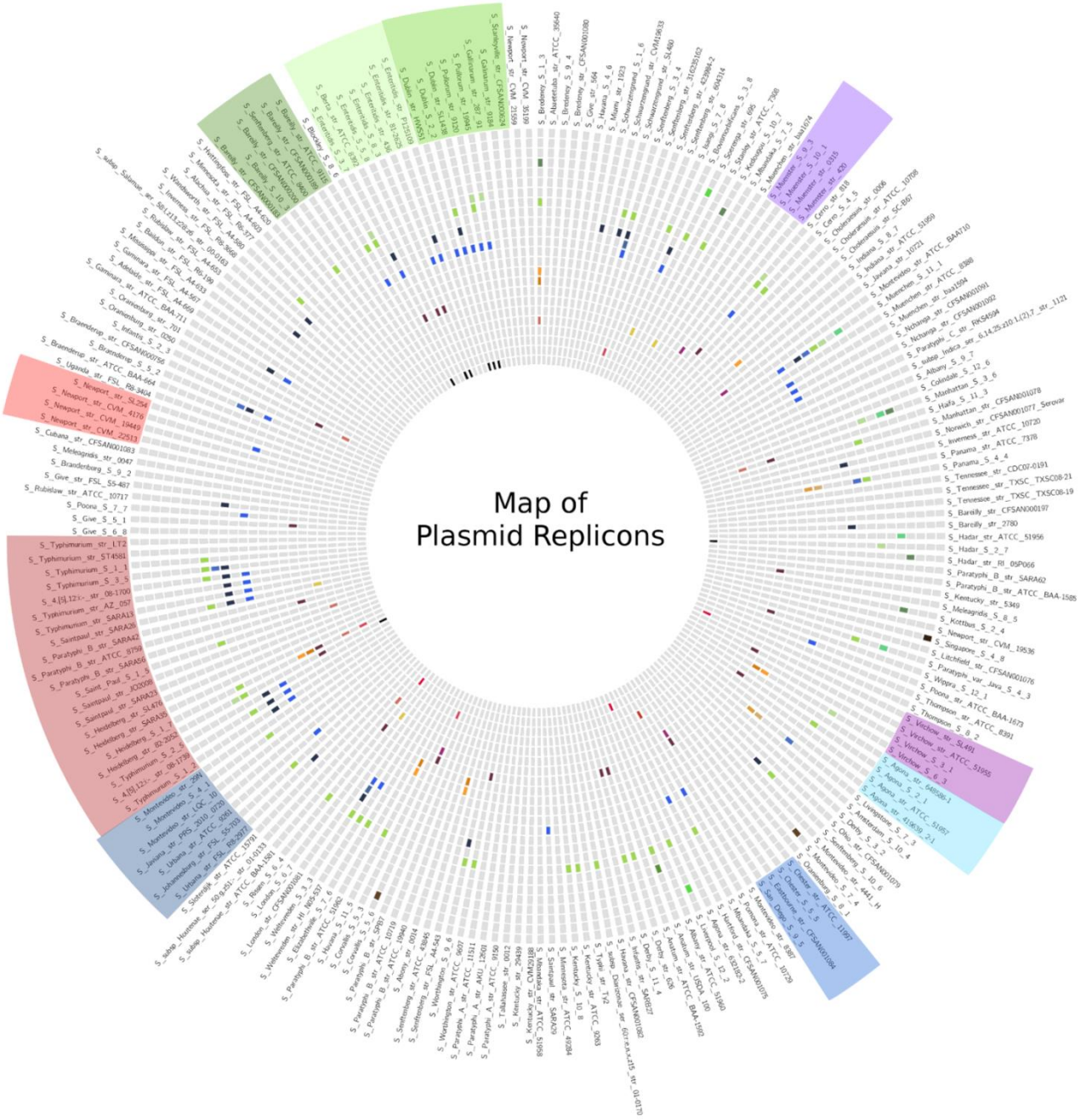


Fig. 4. Map of plasmid replicons in *Salmonella enterica.* The genomes in the map are ordered according to their RM systems, and coloured as in Fig. 1. The bright gray colour indicates absence of replicon in the given genomes, where presence is indicated by a colour specific for the replicon. From the outside to the middle the order of the replicons are: *inc*A/C, *inc*A/C2, ColMGD2, Col156, Col8282, ColE10, ColpVC, ColRNAI, *inc*FIA, *inc*FIB, *inc*FIC, *inc*FII, *inc*HI1A, *inc*HI1B, *inc*HI2, *inc*HI2A, *inc*I1, *inc*I2, *inc*N, *inc*P, p0111, *inc*Q1, *inc*R, *inc*X1, *inc*X4.

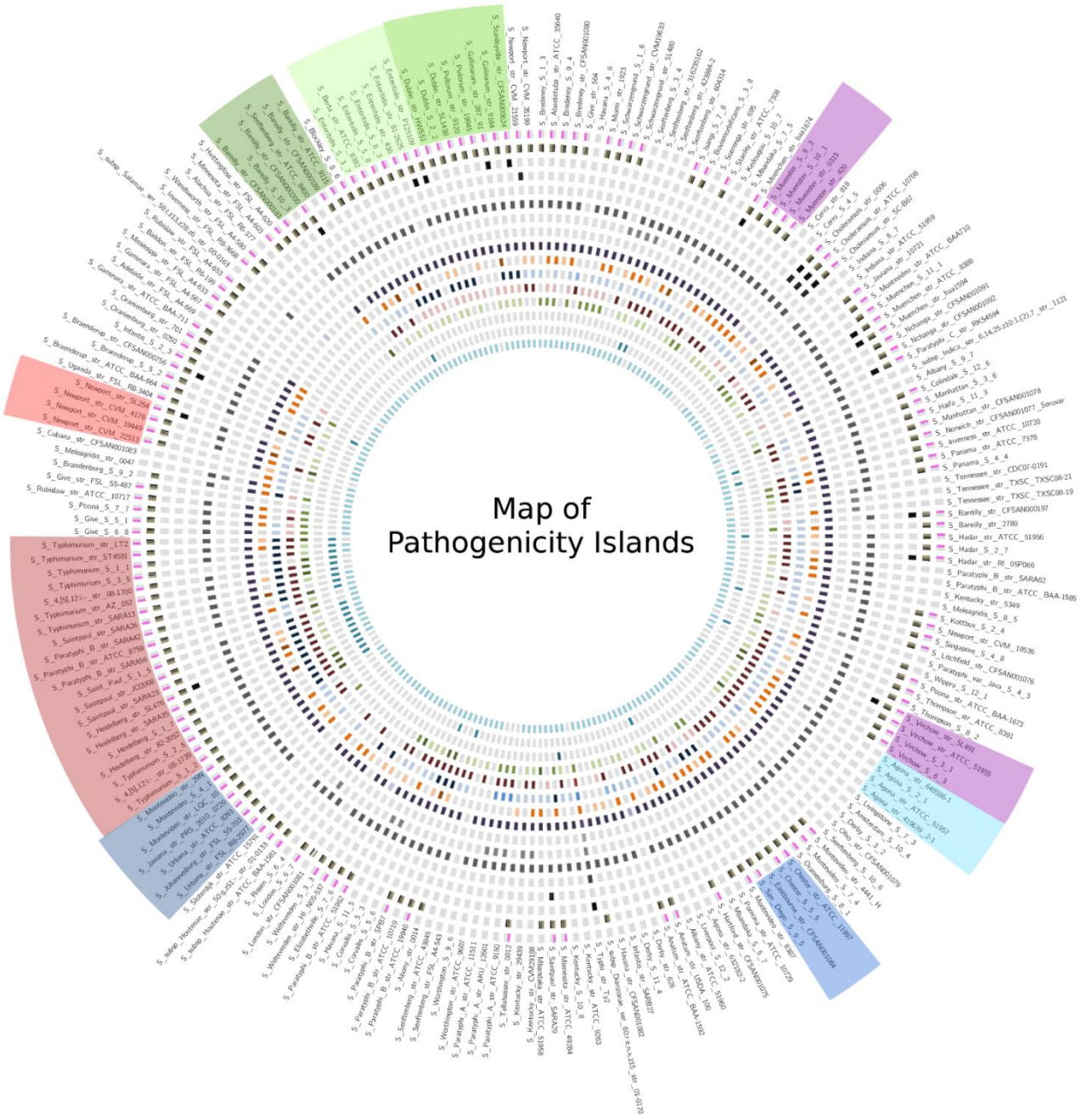


Fig. 5. Map of *Salmonella* Pathogenicity Islands found in *Salmonella enterica* genomes. The genomes are ordered by their RM system profile, indicated by the colours from Fig. 1. Each SPI is defined by a circle with different variants indicated by colour in the given circle.

From the inner circle and outwards, the order of the SPI's is: C63PI, CS54-island, SGI1, SPI-1, SPI-2, SPI-3, SPI-4, SPI-5, SPI-7, SPI-8, SPI-9, SPI-10, SPI-11, SPI-12, genes of SPI-13, genes of SPI-14.

Supporting Information

Table S1: Full genomic information

Table S2: Overview of genes in the RM systems

Fig. S1: Distribution of RM systems in *Salmonella enterica* isolates

Fig. S2: Pan- and Core-genome Plot of 221 *Salmonella enterica* gnomes

Fig. S3: Distribution of Antimicrobial resistance genes in *Salmonella enterica* isolates

File S1: The 16,375 *Salmonella* Pan genes in FASTA format

File S2: The 2,138 Salmonella Core genes in FASTA format

Roer III

Identification of *Salmonella enterica* recipient genes affecting plasmid conjugation

Louise Roer[1], Henrik Hasman[1,3], Sünje J. Pamp[1], Rolf S. Kaas[1], Leise Riber[2], Søren J. Sørensen[2], Frank M. Aarestrup[1]

[1]Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Section for Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[3] Section for microbiology and diagnostics, Statens Serum Institut, Copenhagen, Denmark

Manuscript in preparation

Supplemental material can be accessed at:
https://www.dropbox.com/sh/src29dglfq2h0pu/AADsmz81L9M5VzfKpVNORf0Ha?dl=0

Title: Identification of *Salmonella enterica* recipient genes affecting plasmid conjugation

Running title: Recipient genes affecting conjugation

Louise Roer*[1], Henrik Hasman[1], Sünje J. Pamp[1], Rolf S. Kaas[1], Leise Riber[2], Søren J. Sørensen[2], Frank M. Aarestrup[1]

[1]Division for Epidemiology and Microbial Genomics, National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Section for Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark

*: Corresponding author: Louise Roer

National Food Institute, Technical University of Denmark

Søltoft Plads, Building 221, DK-2800 Kgs. Lyngby, Denmark

E-mail: lroe@food.dtu.dk

Abstract

The spread of genetic threats through horizontal gene transfer is of very high importance. By understanding the mechanisms and genes associated with conjugational ability in recipients we might be able to predict, delay or even control the spread of antimicrobial resistance. The aim of this study was to identify species-specific gene family candidates from whole genome sequenced *Salmonella enterica*, which are enhancing or restricting plasmid uptake and maintenance in recipient cells.

The study included 93 *Salmonella enterica* isolates, covering 54 different serovars. To evaluate the ability of the 93 isolates to take up a plasmid, an initial conjugation experiment was performed with the *mCherry* marked *Escherichia coli* K-12 strain MG1655 donor, and the *gfp* reporter constructed IncP-1β plasmid pKJK5. The conjugation frequencies, ranging from 0.0 to 2.5×10^{-1} transconjugants per potential recipients, facilitated a subdivision of the genomes by using the kmeans algorithm. The predicted gene families of the resulting groups of ten good and 83 poor recipients were compared by different approaches to identify gene family that might be involved in either enhancing or restricting conjugation. In addition, a parallel analysis was performed, on 11 *S.* Enteritidis isolates, to elucidate if serotype specific genes could be involved in conjugation.

For the entire collection of 93 *Salmonella enterica* isolates, we were not able to detect unique gene family candidates associated with the either high or low transfer frequencies. However, when assessing the 11 *S.* Enteritidis genomes separately, 33 gene family candidates were identified that were associated with restricting conjugational abilities.

This study suggests that no common genes exist for *Salmonella enterica*, which neither enhancing nor restricting conjugation. However, serovar specific gene family candidates should be

investigated to clarify their impact on conjugation. Additionally, the results obtained in this study indicate that bacteria can avoid being used as recipients.

Introduction

Horizontal gene transfer has played a major role in both long- and short term bacterial evolution. Mobile genetic elements such as plasmids and transposons are widespread in all bacterial species [1]. With the global emergence and rapid spread of multiple antibiotic resistance in recent years, it has become clear that the importance of horizontal gene transfer in evolution is of very high significance, perhaps even higher than previously expected [2–4], especially with the global transmission of resistance plasmids into several bacterial species [5]. We also observe seemingly a difference in the ability of specific clonal lineages and bacterial species to acquire resistance plasmid, which cannot just be explained by differences in reservoirs [6, 7]. If we can identify genetic traits associated with reduced or increased ability to acquire resistance genes we might better predict where resistance will emerge.

In conjugation, there is a need for active participation of both mating partners [8]. In a recent study, Pérez-Mendoza and de la Cruz tried to identify recipient genes in *Escherichia coli* affecting the ability in plasmid conjugation [9], but the results indicated that none of the non-essential genes from the *E. coli* recipient played an important role in conjugation. Different host defense mechanisms have been suggested to impose barriers for conjugation, including the RM systems. However a study performed on *E. coli* revealed that these barriers were not absolute [10], and in addition could not be reflected in the evolution of *Salmonella enterica* ssp. *enterica* (unpublished data from Roer *et al*., 2015). It has been observed that host-specific *Salmonella* serovars are carrying considerably less resistance genes compared to the serovars with multiple hosts, although they have been isolated from the same reservoir [6, 7]. This phenomenon has so far not been explained, thus it is likely that a varying ability to receive DNA could be caused by genetic difference among the clones.

In this study we aimed to identify genetic traits associated with uptake of IncP-1 plasmids. IncP-1 plasmids were originally found in clinical bacterial isolates [11, 12], and are still of clinical concern because of their ability to carry and facilitate the spread of antibiotic resistance [13, 14] and the IncP transfer (*tra*) system of the broad-host-range plasmid RP4 is one of the best-studied conjugative systems [15]. We investigated several *Salmonella* recipients of various serovars, and divided them into groups of good and poor recipients based on their transconjugant frequencies using an *E. coli* donor. Different bioinformatical approaches were taken in the attempt to identify genes responsible for promoting or resisting conjugational transfer.

Material and Methods

Bacterial donor strain and conjugative plasmid

The mCherry tagged donor *E. coli* MG1655::*lacI*^{q}*-pLpp-mCherry-Km*^{R} with the conjugative *gfpmut3*-tagged pKJK5 plasmid was kindly provided by Uli Klümper *et. al* [16]. Essentially, the plasmid was genetic marked with a gfpmut3 gene, encoding green fluorescent protein (GFP), conditionally expressible by a $lacI_q$ repressible promoter located upstream of gfpmut3 gene [17]. Further, the plasmid donor strain *E. coli* MG1655 was chromosomally marked with a gene cassette encoding constitutive expression of both red fluorescence (mCherry gene) and $lacI_q$. As a result, the constitutive $lacI_q$ expression ensures a gfp repression on the plasmid in the donor strain, but upon plasmid transfer to the recipients, gfp expression is possible resulting in green fluorescent transconjugants detectable by flow cytometry. Donor strain and conjugative plasmid are listed in Table 1.

Salmonella enterica Recipients

From an in-house strain collection at the Technical University of Denmark, National Food Institute (DTU FOOD), a collection of 35 *S. enterica* spp. *enterica* isolates with a global origin and a focus on serovar diversity was merged with a sub-set of 58 genomes previously sequenced as part of the so-called 100K project (NCBI BioProject PRJNA186441). The final recipient collection consisted of 93 *Salmonella* spp. *enterica* genomes covering 54 different serovars. A complete list with genomic information is avalible in S1 Table in the supplemental material.

Whole genome sequencing

From the in-house collection, genomic DNA was extracted from the 35 *Salmonella* isolates using Invitrogen Easy-DNA kit (Invitrogen, Carlsbad, CA, USA). DNA concentrations were determined using Qubit double-stranded DNA (dsDNA) BR assay kit (Invitrogen). The genomic DNA was

prepared for Illumina pair-end sequencing using Illumina (Illumina, Inc., San Diego, CA) NexteraXT Guide 150319425031942 and following protocol revision C (http://support.illumina.com/downloads/nextera_xt_sample_preparation_guide_15031942.html), and submitted for whole genome sequencing at the in-house Illumina MiSeq sequencer.

The raw reads of both in-house 35 genomes and the 58 genomes from 100K project were assembled using the Assembler v1.0, which is based on the Velvet algorithms for *de novo* short-read assembly, available from the Center for Genomic Epidemiology (CGE), http://cge.cbs.dtu.dk/services/all.php.

Conjugation with pKJK5::*gfp*

A single colony of the donor strain was grown in LB with TMP at 37^{o}C overnight. The individual recipients were inoculated in 24 well plates (Nunc) and grown overnight at 37^{o}C. The growth was stopped by incubating the overnight cultures on ice for 3 hours. The optical density (OD) was measured at 600 nm and donor and recipients were diluted to a final OD at 0.5. The donor was mixed with the individual recipients in a ratio of 1:1 in 24 well plates, and incubated 18 hours for conjugation at 37^{o}C. The conjugations were interrupted by properly mixing the liquid, and in the initial screening plates were stored on ice for 30 minutes. Donor, recipients and transconjugants were detected and counted by flow cytometry. All samples were diluted in 0.9% NaCl to reach approximately 2,000 counting events per second before running on flow cytometer.

Flow cytometric detection

Detection of transconjugants, donor, and recipient cells was carried out by using a FACSAria IIIu (Becton Dickinson Biosciences, San Jose, CA, USA). The following settings and voltages were used for detection: forward scatter (FSC) = 505V, side scatter (SSC) = 308V, the bandpass filter 530/30 nm and 508V were used for detection of green fluorescence (GFP) whereas the bandpass filter 610/20 nm and 500V were used for red fluorescence detection. A 70 mm nozzle was used with

a sheath fluid pressure of 70 psi. For operating the system the BD FACSDiva software v.6.1.3 was used, whereas FlowJo v.10 was used for analyzing the data. Five gates were defined in three bivariate plots to distinguish donors, recipients and transconjugants. On the SSC-A vs. FSC-A plot, one gate was defined for particles of bacterial size. On the PE-Texas Red-A vs. SSC-A plot two gates were defined, a gate covering all red fluorescent particles defining the donor bacterial cells, and a non-red gate for the recipient- and transconjugant cells. From the non-red gate, on a FITC-A vs. SSC-A plot, a gate was set to cover all green fluorescent particles, while an additional non-green gate was set to account for the recipient cells (Figure 1). Each sample was set to analyze a total of 100,000 events with the size of bacteria.

Defining good and poor recipients

From the conjugation experiments, transfer frequencies were defined as number of transconjugants per number of potential recipients, and the mean transfer frequencies (TFs) were calculated for each isolate. The algorithm kmeans in R-2.14 (http://cran.r-project.org/bin/windows/base/old/2.14.0/) was used to divide the isolates into two groups (good and poor recipients, respectively) based on their mean TF. The algorithm partitions each point into the group where the sum of squares from the point to the center of the cluster gives the lowest value. To illustrate that it was reasonable to group by the TF mean, a clusplot was drawn based on data from donor, recipient, transconjugant, TF and transconjugant/donor for each isolate.

Identification of recipient determinants involved in conjugation

For identification of recipient determinant of the 93 genomes that might be involved in conjugation, two different approaches were taken. In the first attempt all 93 genomes were analysed together. The Prodigal software [18] was used to predict open reading frames (ORFs) on the scaffolds to eliminate biases in annotation quality and to standardize the genes found [19]. Further, the predicted genes were translated into amino acid sequences and aligned all-against-all using BLASTP [20].

Genes were considered as part of a gene family if the alignment length was at least 80% of the longest sequence and had more than 80% in similarity (80/80 rule) [21]. A common core-pan-genome plot was constructed by comparing the gene families from all the genomes. The pan-genomes were constructed by the union of all the genes, while the core-genomes were assembled from the intersection of gene families shared by at least 90% of genomes [21, 22]. By using the perl script for comparative genomic analysis from the CMG-biotool [23], unique genes was determined by comparing the intersection of genes in one group with the ORFs of the other group.

In the second attempt the genomes were, prior to the subjected analysis, divided into the two groups of good and poor recipients. ORFs, gene families, pan- and core-genomes were predicted for each group, as described above. To identify unique genes, the core-genome of one group was compared with pan-genome of the other group by using BLASTP and the 80/80 rule. This method will ignore orthologues. The resulting gene family candidates were further inspected by using BLASTP against the predicted ORF from each genome from the opposite group, to ensure that the predicted gene families would be unique.

In addition to the analyses of all 93 *S. enterica* genomes, the 11 *S.* Enteritidis genomes spanning both groups (good and poor) were analysed separately, to search for serovar specific gene family candidates. For this analysis, the first approach was used, where ORFs, gene families, and core- and pan-genomes were constructed based on all 11 genomes.

Results

High throughput conjugational screening and grouping of isolates

The high throughput conjugational approach was conducted to screen a high number of varying *S. enterica* isolates for their conjugational ability. This was done to identify recipients with poor versus enhanced ability to take up a plasmid. The plasmid pKJK5, encoding the green fluorescence gene *gfp,* was introduced to the *S. enterica* recipients via the red fluorescent tagged donor *E. coli* K-12 MG1655 which represses *gfp*. Flow cytometry detection revealed transfer frequencies (TF) of the 93 *S. enterica* recipients varying from 0.0 to 2.5×10^{-1} transconjugants per potential recipients, illustrated in the boxplot in Figure 2A.

From the TF mean for each isolate, the kmean algorithm was used to divide the isolates into two groups, resulting in a group of 10 good recipients, and a group of 83 poor recipients. The clustering based on the TF mean for each isolate was represented by a principal component analysis (PCA), illustrated in Figure 3. Isolates from the group of good recipients were illustrated by the red triangles, and isolates from the poor group of recipients by blue circles. The two groups were located apart without overlap in the 2D representation in Figure 3.

Comparative genomics and identification of gene family candidates involved in conjugation

By investigating the gene families found in the 83 bad recipients, absent in the 10 good recipients, and vice versa, we attempted to identify gene families involved in enhancing or restricting plasmid uptake. The pan- and core-genome of the 93 recipients were predicted to 22,161 gene families and 1,933 gene families, respectively. However, in the attempt to identify group dependent gene families, no families were identified by this approach.

In the second attempt, separate core- and pan-genomes were constructed for the two groups. The core-genome (conserved genes) from one group of recipients was compared with the pan-genome

(entire gene-pool including core-genome) from the other groups of recipients, to identify unique gene families only shared within one of the group. The pan- and core-genomes of the good recipient group were predicted to 8,733 and 3,249 gene families by prodigal with following all-against-alignment. For the group of poor recipients, the pan- and core-genomes were predicted to 20,458 and 2,052 gene families.

By using the second approach to compare the gene families between the two groups of recipients, we identified 6 gene families from the good recipients, potentially encoding genes facilitating plasmid transfer. Additionally, 4 gene families were identified in the group of poor recipients, as gene candidates for resisting plasmid transfer. However surprisingly, when manually searching for the identified ORFs in the opposite group by using BLASTP, none of the gene family candidates were found to be unique for the groups.

Additionally, to narrow down the diversity within the dataset and to look for serovar-specific gene candidates, the analysis was performed on a single serovar as well. For the 11 *S.* Enteritidis genomes included in the study, four genomes were predicted to belong to the good recipient group and seven genomes was predicted to belong to the poor recipient group. The TF for the 11 genomes were illustrated in Figure 2B, with the TF values for the 18 S. Typhimurim genomes in Figure 2C for comparison. The common pan- and core-genomes were predicted to 6,149 and 3,736 gene families, respectively. The interception of gene families from one group, totally absent from the other group, revealed that the bad group was sharing 34 gene families not identified in the group of good recipients. Contrary, no unique gene families were identified from in the good group. To verify the findings, the 34 predicted proteins from the poor recipient group were manually searched against the four genomes from the good recipients. This revealed that one of the genes was found in one of the four genomes with an identity of 99.6%, and the remaining 33 was found with less than 66% identity at amino acid level. Thus, 33 gene family candidates, possible restricting conjugation,

were identified in *S.* Enteritidis. The protein sequences of the 33 unique genes can be found in supplemental material (S9 File), together with the sequences of all the predicted pan- and core-genomes (S1 File - S8 File).

Discussion

In recent time, it has become clear that the genetic spread through horizontal gene transfer is highly important in evolution of bacteria [2–4]. However, it is equally important in global emergence and rapid spread of undesirable genetic traits, like the spread of antibiotic resistance. A striking example is the recent finding of the plasmid-mediated colistin resistance gene, *mcr-1*, a resistance which until now only was mediated by chromosomal mutations with following vertical transmission to daughter cells [24, 25]. Hasman *et al.* detected the *mcr*-1 gene in a highly resistant *E. coli* isolate from a human bloodstream infection. The isolate was only susceptible to a limited number of antimicrobial classes, including the carbapenems. Additional acquisition of resistance will leave a limited number of suitable treatment options [25], which underlines why understanding how genes are controlling the uptake and establishment of the plasmids in the recipients is of great importance.

The objective of this study was to identify genetic traits of recipient strains, associated with uptake of IncP-1 plasmids, a mediator for spread of antimicrobial resistance. Based on transconjugant frequencies from a conjugational setup, the 93 *S. enterica* recipients clustered into two groups based on high and low transconjugant frequencies. By using different approaches for comparative genomics of the two groups, we initially identified 6 gene family candidates for enhancing conjugation, in addition to 4 gene family candidates restricting conjugation. However, further analysis with BLASTP against all predicted ORFs from the opposite group, revealed that none of the candidates were in fact unique to any of the two defined groups. The main dataset of 93 recipients was covering 54 different serovars, which is arguably a very diverse dataset and could have contributed to the lack of success to identify unique gene candidates. Thus, to decrease the genetic diversity, a single serovar was assessed separately. Assessing only the *S.* Enteritidis serovar, it was possible to identify 33 unique gene families in the genomes belonging to the low transfer frequency group, indicating the potential for serovar specific genes involved in the restriction of

conjugational transfer. Adding additional *S.* Enteritidis serovars to the analysis could increase the serovar specific diversity and perhaps decrease the number of significant genes found in the analysis.

The attempt to identify genes affecting recipient abilities in plasmid conjugation has previously been done by Pérez-Mendoza and de la Cruz [9]. Their study was performed with *E. coli* as recipient, and their strategy was based on high-throughput screening for systematic evaluation of individual *E. coli* genes in bacterial conjugation, where 20,000 random Tn-insertion mutants as well as the Keio collection of 3,908 individual deletion mutants, were screened. The attempt to identify recipient genes was focusing on two different K-12 strains, without knowing if genes affecting conjugation would be present at all [9]. They concluded that conjugation occur with little regard to the components of the recipient cells, which cannot avoid being used as recipient [9].

In the present study, we tried to deal with the drawback of not knowing the level of conjugational ability of the recipients investigated, by performing an initial conjugation experiment of 93 *S. enterica* ssp. *enterica* isolates, with following grouping of good and poor recipients. Unfortunately, no gene families could be associated with either the good recipient- or poor recipient groups when assessing the entire *enterica* subtype of *Salmonella*. However, other analysis identifying unique SNPs could perhaps reveal an association not recorded with the utilized methods. Further, a comprehensive characterization including plasmid replicons, antimicrobial resistance, MLST profiles, etc. could as well guide the analysis in a direction or give an impression of where to search for differences that could cause the different recipient abilities.

As mentioned above, an explanation could be the high complexity within the dataset, as the initial study was performed on a variety of *S. enterica* ssp. *enterica* isolates. To reduce the diversity, the 11 genomes of *S.* Enteritidis were assessed in an individual analysis, with the grouping defined

from the analysis of the 93 genomes. However, when assessing the conjugational frequencies of *S.* Enteritidis, it became clear that it can be difficult to define an arbitrary threshold separating good recipients from poor recipients. It can be questioned if we need the threshold to identify genes involved in conjugation, or if a strain by strain comparison, taking the TF into account, would be more appropriate. Comparing the conjugational frequencies of *S.* Enteritidis and *S.* Typhimurium, the frequencies for the two serovars are not in the same range, however the same pattern of increase in frequencies can be observed within each of the two serovars (Figure 2). Thus, a comparison of the gene content of the two serovars could be a subject for identification of gene-candidates. From the kmean algorithm none of the *S.* Typhimurium but four out of 11 *S.* Enteritidis isolates were determined as good recipients. Assessing the continuously increase between of the isolates illustrated in Figure 2, this classification into good and poor recipients do not seem reasonable, as no distinct grouping is observed within neither of the two serovars. Additionally, the similar pattern for the two serovars, but different range of transfer, indicates that the ratio of conjugational transfer is very serovar specific. An explanation for this difference in range could be the individual sub-linage content of RM systems for the two serovars (unpublished data from Roer *et al*., 2015). However, this does not explain the difference observed within each serovar. Even though there is no clear grouping of good and poor recipients within a serovar, differences in recipient abilities can be observed which even appear significant between the two isolates with the highest and lowest frequencies. Comparing these results to the conclusion by Pérez-Mendoza and de la Cruz [9], it is not obvious that the bacteria cannot avoid being used as recipients, especially as some of the transfer frequencies in this study are recorded to zero and some are as high as 2.5×10^{-1} transconjugants per potential recipients, from the same donor and conjugation apparatus. The very high TF observed for some of the isolates is very unusual in conjugation, indicating that these isolates indeed have good recipient abilities. Thus, some recipient components must be involved in

the recipient conjugational ability. As RM systems did not impose absolute barriers, other factors must be involved in the control of conjugation.

The continuously increase in *S.* Enteritidis and *S.* Typhimurium could indicate that multiple genes are involved in determining the ability to take up plasmids, causing a synergy effect increasing or decreasing of the transfer frequency. However, it could also be different genes accounting for the differences in recipient abilities for each isolate, which probably could be answered by protein annotation and sequence alignment of the predicted proteins. Changing the arbitrary threshold for *S.* Enteritids could potentially reveal more or less gene family candidates. An interesting approach with this continuously increase in frequency could be to do multiple comparisons, where each "jump" is investigated in correspondence to the remaining genomes. Thus, for each increase/decrease the responding gene could be predicted.

Interestingly, only gene families predicted to be involved in reducing plasmid uptake were observed from this analysis. However, considering the fitness costs of harbouring plasmids, it is reasonable to think that bacteria in general develop methods for restricting plasmid uptake as a protection or defence mechanism, and not develop enhancing abilities.

Playing the devil's advocate it could be argued that it is possible that the 33 genes only are predicted as unique due to the size difference between the two groups, as the poor recipient group contain seven genomes compared to the four in the good group. Thus, maybe the 33 gene family candidates are found due to natural diversity between the strains, and do not possess any impact on conjugational transfer. Therefore, the 33 gene family candidates predicted in this study should be further investigated, both by function prediction, prevalence of genes, and in molecular techniques in the laboratory.

The current study was constructed by using a combination of molecular technologies and bioinformatics to predict gene family candidates involved in regulation of recipient abilities in

306 conjugation. Though, some drawbacks from using bioinformatics in prediction of bacterial
307 behaviour do exist. The prediction of gene families is based on the "80/80" rule. However a single
308 nucleotide difference can cause alteration of the protein which in some cases can lead to functional
309 inactivation without truncating the protein. An alternative approach could be to screen the bacterial
310 genes for examining enhanced or restricted abilities to uptake plasmids. One example would be to
311 perform random gene-knockout of a good recipient, with following identification of genes either
312 decreasing or increasing the conjugational abilities, as performed in by Pérez-Mendoza and de la
313 Cruz in *E. coli* [9]. However, in this case the conjugational ability would be known in advance.
314 Another approach could be a transposon-library of the bacterial genome, which can be introduce to
315 a recipient with the opposite conjugation ability, than the wildtype.

316 In summary, the findings from this study indicate that differences in conjugation ability can be
317 observed, both at species level and serovar level for *Salmonella,* with differences varying from 0.0
318 to 2.5×10^{-1} transconjugants per potential recipients. No common genes could yet be linked to
319 enhanced or restricted conjugational abilities, when comparing a broad community of *Salmonella*
320 *enterica* ssp. *enterica* isolates, but additional bioinformatic approaches are ongoing to explore this
321 dataset further. With regards to *S.* Enteritidis, 33 gene family candidates could be predicted, which
322 are being investigated and awaiting verification. Thus, the collection of genomes used in this study,
323 together with the results from the initial conjugation experiment leaves great opportunities for
324 revealing of genes involved in plasmid uptake.

Acknowledgement

We thank S.M. Milani for assistance with flow cytometry detection in the conjugational assay, and Pimlapas Leekitcharoenphon and Oksana Lukjancenko for bioinformatic guidance and support.

This work was founded by the grant 11-106571 from the Danish Council for Independent Research/Technology and Production Sciences.

References

1. Shintani M, Sanchez ZK, Kimbara K: **Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy**. *Front Microbiol* 2015, **6**(March):1–16.

2. Norman A, Hansen LH, Sørensen SJ: **Conjugative plasmids: vessels of the communal gene pool.** *Philos Trans R Soc Lond B Biol Sci* 2009, **364**:2275–2289.

3. Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick H a, Pósfai G, Hackett J, Klink S, Boutin a, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim a, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch R a, Blattner FR: **Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.** *Nature* 2001, **409**:529–533.

4. Fraser-Liggett CM: **Insights on biology and evolution from microbial genome sequencing**. *Genome Research* 2005:1603–1610.

5. Tijet N, Richardson D, MacMullin G, Patel SN, Melano RG: **Characterization of Multiple NDM-1-Producing Enterobacteriaceae Isolates from the Same Patient**. *Antimicrob Agents Chemother* 2015, **59**:3648–3651.

6. Lauderdale T, Aarestrup FM, Chen P, Lai J, Wang H-Y, Shiau Y-R, Huang I-W, Hung C-L: **Multidrug resistance among different serotypes of clinical Salmonella isolates in Taiwan.** *Diagn Microbiol Infect Dis* 2006, **55**:149–55.

7. van Duijkeren E, Wannet WJB, Heck MEOC, van Pelt W, Sloet van Oldruitenborgh-Oosterbaan MM, Smit J a H, Houwers DJ: **Sero types, phage types and antibiotic susceptibilities of Salmonella strains isolated from horses in The Netherlands from 1993 to 2000.** *Vet Microbiol* 2002, **86**:203–12.

8. Curtiss RI, Charamella LJ, Stallions DR, Mays JA: **Parental Functions During Conjugation in Escherishia coli K-12**. *Bacteriol Rev* 1968, **32**:320–348.

9. Pérez-Mendoza D, de la Cruz F: **Escherichia coli genes affecting recipient ability in plasmid conjugation: Are there any?** *BMC Genomics* 2009, **10**:71.

10. Roer L, Aarestrup FM, Hasman H: **The EcoKI Type I Restriction-Modification System in Escherichia coli Affects but Is Not an Absolute Barrier for Conjugation.** *J Bacteriol* 2015, **197**:337–42.

11. Jobanputra RS, Datta N: **Trimethoprim R Factors in Enterobacteria from Clinical Specimens**. *J Med Microbiol* 1974, **7**:169–177.

12. Jacoby G a., Jacob a. E, Hedges RW: **Recombination between plasmids of incompatibility groups P-1 and P-2**. *J Bacteriol* 1976, **127**:1278–1285.

13. Ingram LC, Richmond MH, Sykes RB: **Molecular characterization of the R factors implicated in the carbenicillin resistance of a sequence of Pseudomonas aeruginosa strains isolated from burns.** *Antimicrob Agents Chemother* 1973, **3**:279–88.

14. Novais a., Canton R, Valverde A, Machado E, Galan J-C, Peixe L, Carattoli A, Baquero F, Coque TM: **Dissemination and Persistence of blaCTX-M-9 Are Linked to Class 1 Integrons Containing CR1 Associated with Defective Transposon Derivatives from Tn402 Located in**

Early Antibiotic Resistance Plasmids of IncHI2, IncP1- , and IncFI Groups. *Antimicrob Agents Chemother* 2006, **50**:2741–2750.

15. Grohmann E, Muth G, Espinosa M: **Conjugative Plasmid Transfer in Gram-Positive Bacteria**. *Microbiol Mol Biol Rev* 2003, **67**:277–301.

16. Klümper U, Riber L, Dechesne A, Sannazzarro A, Hansen LH, Sørensen SJ, Smets BF: **Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community**. *ISME J* 2014:934–945.

17. Bahl MI, Oregaard G, Sørensen SJ, Hansen LH: **Construction and use of flow cytometry optimized plasmid-sensor strains.** *Methods Mol Biol* 2009, **532**:257–268.

18. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification.** *BMC Bioinformatics* 2010, **11**:119.

19. Jacobsen A, Hendriksen RS, Aaresturp FM, Ussery DW, Friis C: **The Salmonella enterica Pan-genome**. *Microb Ecol* 2011, **62**:487–504.

20. Altschul S, Gish W, Miller W, Myers E, Lipman D: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.

21. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli S V, Crabtree J, Jones AL, Durkin a S, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan S a, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, et al.: **Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome".** *Proc Natl Acad Sci U S A* 2005, **102**:13950–13955.

22. Leekitcharoenphon P, Lukjancenko O, Friis C, Aarestrup FM, Ussery DW: **Genomic variation in Salmonella enterica core genes for epidemiological typing**. *BMC Genomics* 2012, **13**:88.

23. Vesth T, Lagesen K, Acar Ö, Ussery D: **CMG-biotools, a free workbench for basic comparative microbial genomics.** *PLoS One* 2013, **8**:e60120.

24. Liu Y-Y, Wang Y, Walsh TR, Yi L-X, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu L-F, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu J-H, Shen J: **Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study.** *Lancet Infect Dis* 2015, **3099**:1–8.

25. Hasman H, Hammerum A, Hansen F, Hendriksen R, Olesen B, Agersø Y, Zankari E, Leekitcharoenphon P, Stegger M, Kaas R, Cavaco L, Hansen D, Aarestrup F, Skov R: **Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015**. *Eurosurveillance* 2015, **20**:Article 1.

405 **Table 1 | Bacterial donor strain and plasmid used in this study.**

Bacterial strain and plasmid	Description (relevant genotype or phenotype)	Source or reference
Bacterial strains		
Escherichia coli MG1655 mCherry	*F⁻ λ⁻ ilvG- rfb-50 rph-1, lacIq-pLpp-mCherry-Kan*^{R}	[16]
Plasmids		
pKJK5::P*lac*::*gfp*	IncP-1β, Tmp^R	[16]

406 [1] For bacterial strains and plasmids the relevant genotype, phenotype and other characteristics are shown.
407

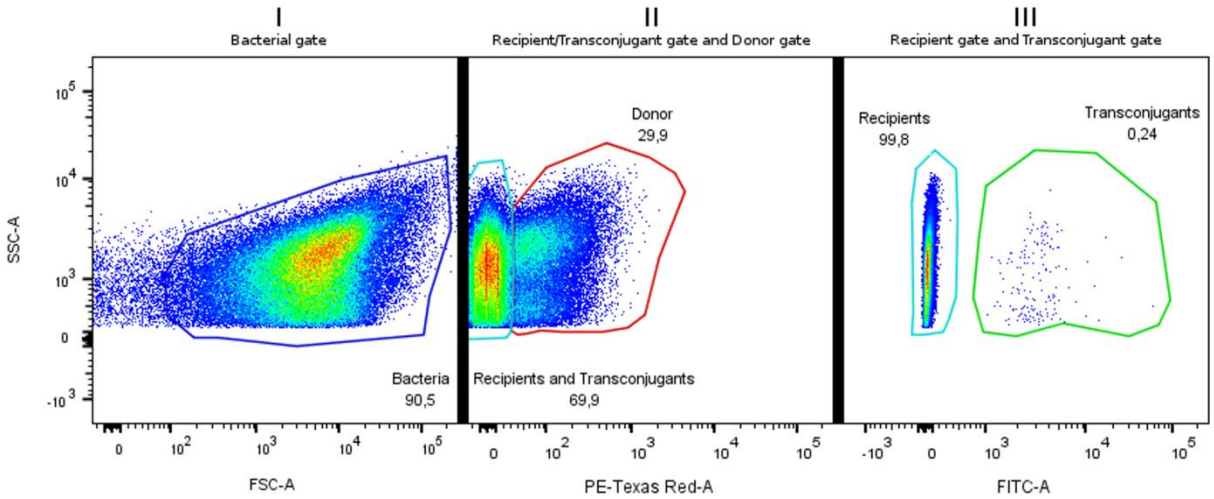


Figure 1 | Conjugational screening detected by flow cytometry. Flow cytometry detection of Donor, Recipients, and Transconjugants from a conjugation mixture of *S. enterica* and *E. coli K-12* MG1655 carrying pKJK5. The gating procedure consists of five successive gates in the 3 different plots: Plot I with the Bacterial gate, detects all particles of bacterial size based on front and side scatter; Plot II detects all red fluorescence particles representing the donors, and all non-red particles representing either recipients or transconjugants by using side scatter and PE-Texas Red; Plot III divides the Recipients and Transconjugants by gating non-green and green fluorescence particles based on side scatter and FITC.

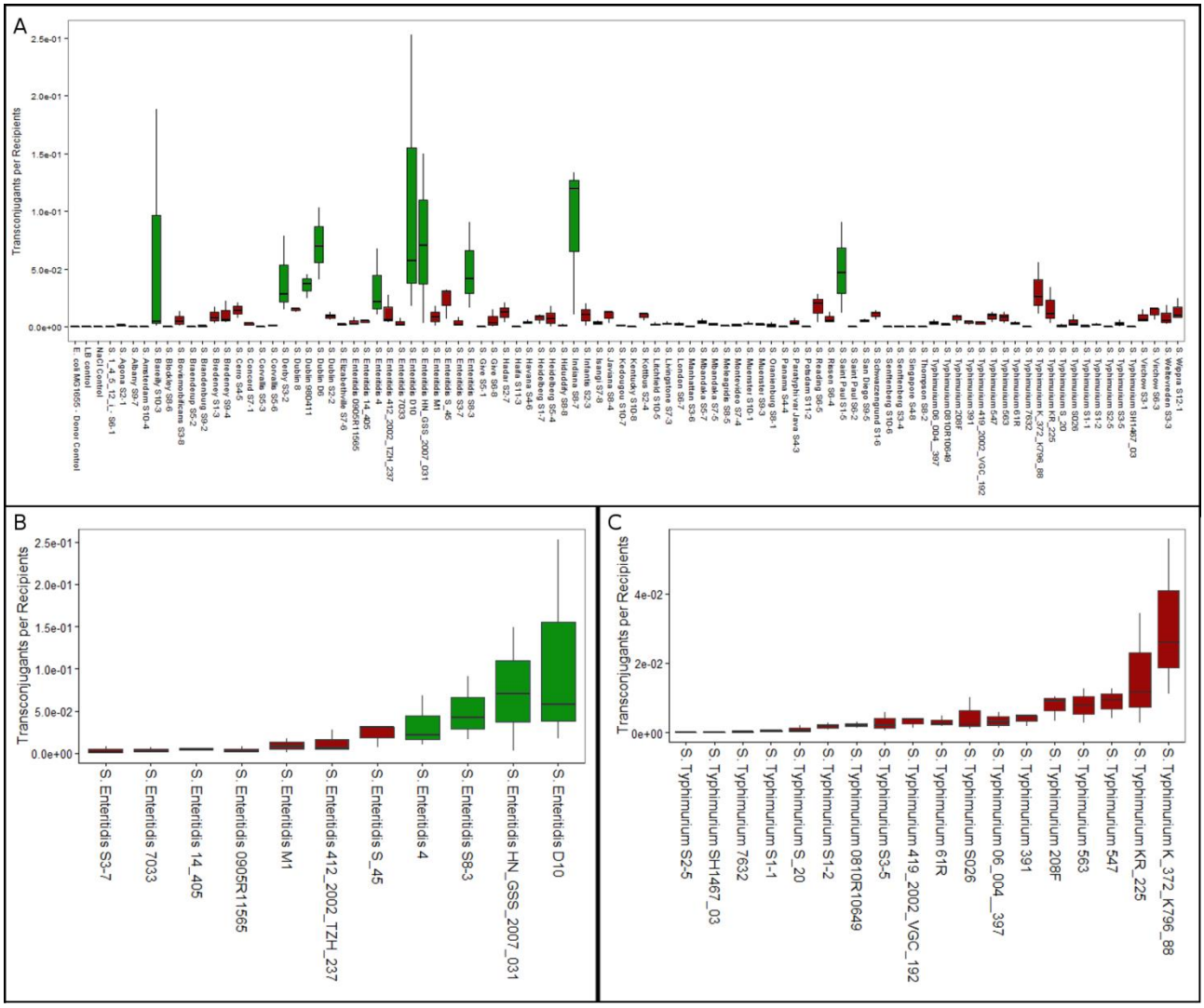


Figure 2 | Conjugation frequencies for *S. enterica* ssp. *enterica*. A, Transconjugant frequencies of the initial screening of recipients with good or poor uptake. The frequencies are represented as a box for each isolate, illustrating the results of the three repetitions. The frequencies are represented as a box for each isolate, illustrating the results of the three repetitions. **B**, Conjugation frequencies for the 11 *S.* Enteritidis isolates included in the study. **C**, Conjugation frequencies for the 18 *S.* Typhimurium isolates included in the study. The isolates in B and C are ordered according to their mean transfer frequency.

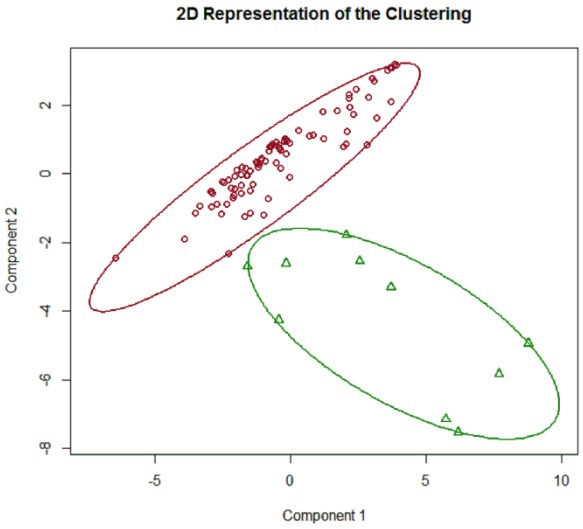


Figure 3 | 2D Representation of Clustering. The PCA analysis is based on the donor, recipients and transconjugants in each replicate, as well as their mean of the three repetitions of the conjugations experiment. Additionally, the representation is based on the TF mean and the donor/transconjugant ratio. The bad recipients are marked with blue circles and the good recipients are marked with red triangles. The two components together explain 75.22% of the point viability of the data.

Supporting information

S1 Table: Genomic information of recipient strains

S1 File: The 22,161 *Salmonella* Pan genes from all 93 genomes, in FASTA format

S2 File: The 1,933 *Salmonella* Core genes from all 93 genomes, in FASTA format

S3 File: The 8,733 *Salmonella* Pan genes from group with high frequency, in FASTA format

S4 File: The 3,249 Salmonella Core genes from group with high frequency, in FASTA format

S5 File: The 20,458 *Salmonella* Pan genes from group with low frequency, in FASTA format

S6 File: The 2,052 *Salmonella* Core genes from group with low frequency, in FASTA format

S7 File: The 6,149 *Salmonella* Enteritidis Pan genes from 11 genomes, in FASTA format

S8 File: The 3,736 *Salmonella* Enteritidis Core genes from 11 genomes, in FASTA format

S9 File: The 34 Unique *Salmonella* Enteritidis genes from group with low frequency, in FASTA format

National Food Institute
Technical University of Denmark
Mørkhøj Bygade 19
DK - 2860 Søborg

Tel. 35 88 70 00
Fax 35 88 70 01

www.food.dtu.dk

ISBN: 978-87-93109-77-3