Aus dem Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie

Lehrstuhl: Bakteriologie

der Ludwig-Maximilians-Universität München

komm. Vorstand: Prof. Dr. Rainer Haas

Genome-based characterization of *Yersinia enterocolitica*: patho-evolution and adaptation of a versatile bacterium

Dissertation

zum Erwerb des Doktorgrades der Naturwissenschaften

an der Medizinischen Fakultät

der Ludwig-Maximilians-Universität München

vorgelegt von

Debora Garzetti

aus Treviglio/Italien

2014

Gedruckt mit Genehmigung der Medizinischen Fakultät

der Ludwig-Maximilians-Universität München

Betreuer: Prof. Dr. Dr. Jürgen Heesemann

Zweigutachter: Priv. Doz. Dr. Ralf Heermann

Dekan: Prof. Dr. med. Dr. h.c. Maximilian Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 21.01.2015

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema “Genome-based characterization of *Yersinia enterocolitica*: patho-evolution and adaptation of a versatile bacterium” selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.
Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, den 19.08.2014

_____

Debora Garzetti

"*The important thing is not to stop questioning.*

Curiosity has its own reason for existing.

One cannot help but be in awe when he contemplates the mysteries

of eternity, of life, of the marvelous structure of reality.

It is enough if one tries merely to comprehend a little of this mystery every day.

Never lose a healthy curiosity."

Albert Einstein

To my little sister

Table of Contents

List of Figures

List of Tables

Abbreviations

ABC	ATP-binding cassette
Aat	Aggregative adherence transporter
Ail	Attachment invasion locus
CIN	Cefsulodin-irgasan-novobiocin
DIG	Digoxigenin
GI	Gastrointestinal
GSP	General secretion pathway
HPI	High pathogenicity island
IVIAT	*In vivo* induced antigen technology
LPS	Lipopolysaccharide
LRR	Leucine-rich repeat
Myf	Mucoid *Yersinia* factor
NGS	Next generation sequencing
O-ag	O-antigen
PFGE	Pulsed-field gel electrophoresis
pYV	*Yersinia* virulence plasmid
PZ	Plasticity zone
T2SS	Type two secretion system
T3SS	Type three secretion system
Tc	Toxin complex
YAPI	*Yersinia* adhesion pathogenicity island
YE	*Yersinia enterocolitica*
YEE	*Yersinia enterocolitica* subsp. *enterocolitica*
YEP	*Yersinia enterocolitica* subsp. *palearctica*
YGI	*Yersinia* genomic island
Yop	*Yersinia* outer membrane protein
Ysa-T3SS	*Yersinia* secretion apparatus-T3SS
Ysc-T3SS	Yop secretion-T3SS
Yts	*Yersinia* type II secretion
Yst	*Yersinia* heat-stable toxin

Summary

Yersinia enterocolitica is a foodborne pathogen comprising a versatile group of organisms, which cause yersiniosis with diverse clinical manifestations, are found all over the world, and survive in a variety of habitats, with high prevalence in animal and food sources. Molecular pathogenesis of yersiniosis and the strategies used by enteropathogenic *Yersinia* to overcome host defenses are well understood, especially for the mouse highly-virulent bioserotype 1B/O:8 strains. Nevertheless, *Y. enterocolitica* isolates from patients mainly belong to the mouse weakly-virulent serotypes O:3, O:9 and O:5,27. In particular, serotype O:3 strains are responsible for 80-90% of yersiniosis cases worldwide, and are frequently isolated from pigs, considered as their asymptomatic reservoir.

Genomic investigations have been mainly performed using the reference genome of bioserotype 1B/O:8 strain 8081 as representative of the highly-virulent group, typically detected in America. Beside strain 8081, another bioserotype 1B/O:8 strain, namely WA-314, is studied in respect of pathogenicity, and strains closely related to WA-314 have recently emerged in Europe. In this context, it is worthwhile to unravel the genome sequence of strain WA-314 and compare it with that of strain 8081, together with their pathogenic potential. Not only did strain WA-314 overcome strain 8081 in mouse colonization, but also their genomes showed high variability, with evidences of recombination and horizontal gene transfer events, reflecting a population adapted to a broad range of niches. Discovered genetic differences in adhesion proteins and in genes encoded by the pYV virulence plasmid, together with the presence of a colicin cluster and differing prophages, may account for the higher virulence and fitness properties of strain WA-314 in comparison with strain 8081.

Since a detailed characterization of the genetic variability within the whole *Y. enterocolitica* species is missing, the uncovered highly-virulent specific genetic features have been analyzed together with available sequences from other serotypes. Whole genome comparison stressed the importance of anaerobic respiration in the physiology of weakly-virulent and non-virulent strains, due to specific nitrite and DMSO reductases. Notably, autotransporters and fimbriae constitute the most relevant virulence determinants for weakly-virulent and non-virulent strains, confirming that pathogenesis-associated abilities of *Y. enterocolitica* are strongly dependent on adhesion to and interaction with host cells. Genome comparison made also possible the establishment of an efficient PCR-based method for *Y. enterocolitica* patho-serotyping, by identifying new O-antigen gene clusters and designing primers targeting serotype-specific genes. This test has proved to be especially useful for typing strains which lost expression of the O-antigen and could not be tested by conventional methods.

By means of phylogenetic and comparative analyses of 10 genomes, representing the broad diversity of *Y. enterocolitica*, the high genome variability of the entire species has been discovered. The open pan-genome of the whole still evolving species reflects the wide range of habitats where these versatile bacteria can proliferate, especially those belonging to the highly- and non-virulent groups, in contrast to the lower diversity within the lineage of the weakly-virulent strains, which occupy more specialized niches. In particular, analyzing additional 20 genomes of the pig-adapted serotype O:3 revealed an overall clonal-like microevolution, driven by genetic drift and gene loss, and proposed the definition of "genetically monomorphic" for *Y. enterocolitica* serotype O:3 strains. No apparent differences between pig and human isolates have been identified, while a SNP-based phylogeny showed more recent mutational events in the human strains. Together with the genetic isolation, these results are consistent with the assumed role of pig asymptomatic carriers as the main reservoir of human infections by serotype O:3 strains. *Y. enterocolitica* serotype O:3, therefore, can be regarded as a genetically monomorphic group of bacteria, corresponding to a lineage within a species of greater diversity, probably because of adaptation to its major host, the swine.

Finally, since the patho-adaptation mechanisms of *Y. enterocolitica* serotype O:3 to humans and swine are not known, an immuno-based approach has been applied to identify genes specifically induced during *in vivo* growth. Confirming the findings from the genome comparison analyses, genes belonging to the anaerobic pathways of cobalamin synthesis, nitrogen and sulfur metabolisms revealed immunogenic reactions with human sera, as well as flagella and fimbrial genes.

In conclusion, this study reinforces the role of physiology and adhesion properties during host colonization and adaptation of *Y. enterocolitica* to specific hosts or environmental niches. New insights into the evolution of *Y. enterocolitica* as whole species and of the serotype O:3 sub-population support the heterogeneity of this species, with the broad-host-range highly- and non-virulent strains on one side and the host-restricted weakly-virulent ones on the other. The basis for a more detailed epidemiological knowledge of *Y. enterocolitica* has been established and future developments are expected.

Zusammenfassung

Yersinia enterocolitica ist ein Krankheitserreger, der durch den Konsum kontaminierter Lebensmittel übertragen wird und aus einer heterogenen Gruppe verschiedener Organismen besteht. *Y. enterocolitica* verursacht das Krankheitsbild einer Yersiniose, die durch verschiedenste klinische Manifestationen charakterisiert ist. *Y. enterocolitica* ist weltweit verbreitet und kann in einer Vielzahl unterschiedlicher Habitate überleben, vor allem aber in Tieren und tierischen Lebensmitteln. Der molekulare Pathogenitätsmechanismus der Yersiniose und die Strategien, mit denen enteropathogene *Yersinien* die Immunabwehr des Wirtsorganismus entkommen, sind gut erforscht. Dies gilt vor allem für den im Mausmodell verwendeten hoch-virulenten Bio-/Serotyp 1B/O:8 Stamm. Klinische *Y. enterocolitica* Isolate gehören hingegen meist zu den Serotypen O:3, O:9 und O:5,27, die im Mausmodell nur schwach-virulent sind. Weltweit werden 80-90% der Yersiniosen durch Serotyp O:3 Stämme verursacht. Diese werden häufig aus Schweinen isoliert, weshalb Schweine als asymptomatisches Hauptreservoir für *Yersinien* gelten.

Genomische Untersuchungen wurden vor allem mit dem Bio-/Serotyp 1B/O:8 Referenzstamm 8081 durchgeführt. Dieser Stamm gehört zur Gruppe der im Mausmodell hoch-virulenten *Yersinia* Spezies und wird vor allem in Amerika isoliert. Neben Stamm 8081 wird WA-314, ein weiterer Bio-/Serotyp 1B/O:8 Stamm, in Bezug auf sein Pathogenitätsverhalten untersucht. Stämme die nah verwandt mit WA-314 sind, breiten sich mittlerweile in Europa aus. Von daher ist es von besonderem Interesse, die Genomsequenz von Stamm WA-314 zu entschlüsseln und sowohl das Genom als auch sein Pathogenitätsverhalten mit Stamm 8081 zu vergleichen. In dieser Arbeit konnte gezeigt werden, dass Stamm WA-314 im Mausmodell eine bessere Kolonisierung als Stamm 8081 aufweist und dass eine starke Genomvariabilität zwischen diesen beiden Stämmen vorliegt. Spuren von Rekombinationsereignissen und horizontalem Gentransfer spiegeln somit eine Population wider, die an eine Vielzahl verschiedener ökologischer Nischen angepasst ist. Genetische Unterschiede von Adhäsionsproteinen und Genen, welche auf dem pYV Virulenzplasmid kodiert sind, sowie die Existenz eines Colicin Genclusters und verschiedener Prophagen, könnten für die starke Virulenz und die Fitness von Stamm WA-314 gegenüber Stamm 8081 im Mausmodel verantwortlich sein.

Bislang fehlt eine detaillierte Charakterisierung der genetischen Variabilität innerhalb der gesamten *Y. enterocolitica* Spezies. Deshalb, wurden die entdeckten Genbereiche, die mit dem hoch-virulenten Phänotyp in Zusammenhang gebracht werden, mit vorhandenen Genomsequenzen anderer Serotypen verglichen. Gesamtgenomvergleiche zeigten die Bedeutung der anaeroben Atmung für die Physiologie der schwach-virulenten und nicht-

virulenten Stämme. Dies ist auf das Vorhandensein von spezifischen Nitrit und DMSO Reduktasen zurückzuführen. Von besonderer Bedeutung für die Virulenz von schwach-virulenten und nicht-virulenten Stämmen zeigten sich Gene kodierend für Autotransporter und Fimbrien. Damit konnte bestätigt werden, dass das Pathogenitätsverhalten von *Y. enterocolitica* stark von Adhäsion und der Interaktion mit der Wirtszelle abhängt. Weiterhin ermöglichten die Genomvergleiche, die Etablierung einer effizienten PCR-basierte Methode, zur Identifizierung verschiedener *Yersinia* Serotypen. Neue O-Antigen Gencluster wurden identifiziert, die mit entsprechend designten Primern amplifiziert werden können. Es zeigte sich, dass dieser Test besonders geeignet ist, um Stämme zu identifizieren, die das O-Antigen nicht mehr exprimieren können. Diese Stämme konnten bislang nicht mit konventionellen Methoden erfasst werden.

Durch die phylogenetische und vergleichende Analyse von 10 verschiedenen *Y. enterocolitica* Genomsequenzen konnte die hohe Genomvariabilität innerhalb der gesamten Spezies gezeigt werden. Das Pangenom dieser gesamten sich weiterhin entwickelnden Spezies spiegelt die verschiedenen Habitate wider, wo sich diese Bakterien vermehren können. Diese Variabilität ist besonders bei jenen Stämmen ausgeprägt, die zur Gruppe der hoch-virulenten und nicht-virulenten Gruppen gehören. Im Gegensatz dazu ist das Genom der schwach-virulenten Stämme weniger variabel, welches mit der Spezialisierung auf bestimme ökologische Nischen zu erklären ist. Die Analyse von 20 weiteren Genomen von Schwein-adaptierten Serotyp O:3 Stämmen ergab eine Art übergeordnete, klonähnliche Mikroevolution, die durch Gendrift und Genverlust entstanden ist. Daher, führt dieses zu der Definition „genetisch monomorphen“ *Y. enterocolitica* Serotyp O:3 Stämmen. Es konnten keine genetischen Unterschiede zwischen *Yersinia* Isolaten aus Schwein und Mensch identifiziert werden. SNP-basierte phylogenetische Untersuchungen hingegen zeigten, dass kürzlich ereignete Mutationen vor allem in den Isolaten vom Menschen zu finden sind. Zusammen mit der genetischen Isolierung unterstützen diese Ergebnisse die Hypothese, dass asymptomatische Schweine das Hauptreservoir für Infektionen von Menschen mit Serotyp O:3 Stämmen darstellen. Aufgrund dessen können *Y. enterocolitica* Serotyp O:3 Stämme als eine genetisch monomorphe Gruppe innerhalb einer Spezies mit größerer genetischer Variabilität angesehen werden. Dies ist vermutlich auf die Adaption an ihren Hauptwirt, das Schwein, zurückzuführen.

Der Pathogenitätsmechanismus, wie *Y. enterocolitica* Serotyp O:3 sich Menschen oder Schweinen anpasst ist bislang nicht bekannt. Demzufolge wurde in dieser Arbeit ein immunologischer Ansatz verwendet, um Gene zu identifizieren, die während des Wachstums *in vivo* induziert werden. Gene die mit dem anaeroben Weg der Cobalamin Synthese, des Nitrit oder Sulfat Metabolismus, sowie Flagellen und Fimbrien assozierte Gene zeigten

immunologische Reaktionen gegenüber Humanserum. Dieser Zusammenhang ergab sich bereits aus den Ergebnissen der Genomvergleiche.

Zusammenfassend bestärkt diese Arbeit die Rolle der Physiologie und der adhäsiven Eigenschaften von *Y. enterocolitica* während der Kolonisierung des Wirts und der Anpassung an spezifische Wirtseigenschaften bzw. an ökologische Nischen. Neue Einblicke in die Evolution von *Y. enterocolitica* als gesamte Spezies sowie der Serotyp O:3 Subpopulation untermauern die Hypothese der Heterogenität dieser Spezies: auf der einen Seite die hoch-virulenten und nicht-virulenten Stämme mit einem breiten Wirtsspektrum und auf der anderen Seite die schwach-virulenten Stämme mit einer eingeschränkten Wirtsspezifität. Mit dieser Arbeit wurde die Basis für detailliertere epidemiologische Erkenntnisse von *Y. enterocolitica* geschaffen und wird zukünftige Entwicklungen vorantreiben.

1. Introduction

1.1. *Yersinia enterocolitica* taxonomy, classification and ecology

Yersinia enterocolitica (YE) is a member of the genus *Yersinia* in the family Enterobacteriaceae, class gamma-proteobacteria. Together with *Y. pseudotuberculosis* and *Y. pestis*, YE is one of the three *Yersinia* species documented as virulent for humans, whereas the remaining 14 species (*Y. aldovae, Y. aleksiciae, Y. bercovieri, Y. entomophaga, Y. frederiksenii, Y. intermedia, Y. kristensenii, Y. massiliensis, Y. mollaretii, Y. nurmii, Y. pekkanenii, Y. rohdei, Y. ruckeri,* and *Y. similis*) are mostly environmental and rarely associated with human disease (Merhej et al., 2011; Murros-Kontiainen et al., 2011a; Murros-Kontiainen et al., 2011b; Sprague and Neubauer, 2005; Sprague et al., 2008; Sulakvelidze, 2000).

Y. enterocolitica is a gram-negative facultative anaerobe rod-shaped bacterium, which is able to survive and actively proliferate at low temperatures (1-4 °C). YE is flagellated and motile at 25 °C, while it is non-flagellated, hence non-motile, at 37 °C (Bottone, 1999). YE is a highly heterogeneous group of organisms that can be classified into 6 biotypes according to biochemical reactions as regards to metabolism (Bottone, 1997) (Table 1), while antigenic variations in the O-antigen molecule of the lipopolysaccharide (LPS) allows differentiation of more than 70 serotypes (Wauters et al., 1987) (section 1.1.1). A common nomenclature based on the geographical origin of the first isolation in Europe/Japan or North America allows a further classification of the isolates into “Old World” and “New World” strains, respectively. Moreover, grouping of YE into 2 subspecies, *palearctica* (YEP) and *enterocolitica* (YEE), have been justified by differences in the 16S rRNA gene sequences (Neubauer et al., 2000a). Based on the bacterial virulence in the mouse model, YE can be also classified into 3 groups: mouse- or highly-virulent, weakly-virulent and non-virulent (Table 2).

Y. enterocolitica is a ubiquitous microorganism, which has been isolated from animals, water, soil, food, environment and humans, and especially swine serve as main reservoir for human pathogenic strains (Bottone, 1999). Interestingly, there is a close association between the bioserotype of the isolates and their ecological niches. Even though the facultative-pathogenic and opportunistic biotype 1A strains are the ones most frequently recovered from the environment, also pathogenic strains of biotype 1B have been isolated from water. Bioserotype 4/O:3 strains colonize pigs with high frequency, while strains of bioserotype 2/O:9 are isolated from goats and cows. Strains of the very rare bioserotype 5/O:2,3 are isolated from sheep, hares

and goats, and bioserotype 3/O:1,2a,3 strains from chinchillas (Rahman et al., 2011). These data suggest that animals are the origin of YE human infections.

Table 1 Biotyping scheme of *Y. enterocolitica*. Adapted from (Bottone, 1997).

	Biotype reaction					
Test	**1A**	**1B**	**2**	**3**	**4**	**5**
Esculin hydrolysis (24 h)	+/-	-	-	-	-	-
Pyrazinamidase activity	+	-	-	-	-	-
Salicin (acid production in 24 h)	+	-	-	-	-	-
Lipase activity	+	+	-	-	-	-
Indole production	+	+	v	-	-	-
Xylose (acid production)	+	+	+	+	-	v
Trehalose (acid production)	+	+	+	+	+	-
Sorbitol (acid production)	+	+	+	+	+	-
Nitrate reduction	+	+	+	+	+	-
Ornithine decarboxylase	+	+	+	+	+	+(+)

Symbols: +, positive result (detected activity); -, negative result (no activity); (+), delayed positive; v, variable, depends on the examined bacterial strain.

Table 2 Classification of *Yersinia enterocolitica* isolates (Bottone, 1997; Sabina et al., 2011).

Group	**Subspecies**	**Geographical origin**	**Biotype**	**Serotype**
Highly-virulent (mouse-virulent)[a]	*enterocolitica* (YEE)	North America ("New World" strains)	1B	O:4,32; O:8; O:13a,13b; O:16; O:18; O:20; O:21; O:25; O:41,42; NT[b]
Weakly-virulent	*palearctica* (YEP)	Europe and Japan ("Old world" strains)	2	O:9; O:5,27; O:27
			3	O:1,2,3; O:5,27; O:3
			4	O:3
			5	O:2,3
Non-virulent			1A[c]	O:4; O:5; O:6,30; O:6,31; O:7,8; O:7,13; O:10; O:14; O:16; O:21; O:22; O:25; O:37; O:41,42; O:46; O:47; O:57; NT[b]

[a] In this thesis, the high-virulence pathotype refers to mouse-virulence and will be used in this sense.

[b] NT: non-typeable.

[c] Growing evidences support the idea that *Y. enterocolitica* 1A isolates are opportunistic pathogens in patients with underlying medical disorders and immune deficiencies.

1.1.1. *Y. enterocolitica* O-antigens and serotypes

Serotype-based classification of *Y. enterocolitica* was first applied in 1967 (Winblad, 1973). The serotyping scheme is mainly defined by the variability of the O-antigen (O-ag), allowing determination of more than 70 serotypes in YE and related species (Wauters et al., 1991), with biotype 1A strains being the more heterogeneous and belonging to at least 17 serotypes

(Bhagat and Virdi, 2011). Serological classification is performed by agglutination of strains with serotype-specific antisera, produced by immunized rabbits, after injection of heat-killed bacterial antigenic saline suspension (Aleksic et al., 1986). The O-ag epitopes are determined by the type and arrangement of sugar residues in the repeated O-units that compose the LPS. O-ags can be classified as homopolymeric, when the repeating unit is a single sugar, or heteropolymeric, if it is composed of two or more different sugar residues (Figure 1). Chemical approaches have elucidated the structures of a number of YE O-ags (Gorshkova et al., 1985, 1986; Ovodov et al., 1992; Radziejewska-Lebrecht et al., 1994; Skurnik, 2006). Interestingly, in YE serotypes O:3 and O:9 the O-ag and the outer-core hexasaccharide are both attached to the inner core, forming a short branch in the LPS molecule, a peculiarity rarely seen in other enterobacteria (Skurnik and Bengoechea, 2003). The genetic cluster and genomic location of YE O-ags are partly solely known for serotypes O:3, O:9 and O:8. The O-ag gene cluster of serotype O:8 is located between the *hemH* (ferrochelatase) and *gsk* (inosine-guanosine kinase) genes, similarly to *Y. pestis* and *Y. pseudotuberculosis*. This locus is occupied by the outer core gene cluster in YE serotypes O:3 and O:9, while the genomic location of their O-ag clusters was unknown (Skurnik and Bengoechea, 2003) before this thesis.

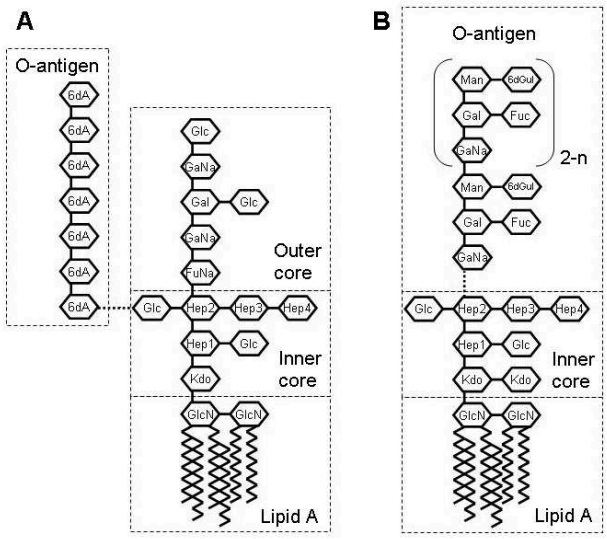


Figure 1 Schematic structure of the lipopolysaccharide of *Y. enterocolitica*.
LPS of serotype O:3 (A) and serotype O:8 (B) are shown as example of homopolymeric and heteropolymeric O-antigens, respectively. Modified from (Skurnik, 2004).

Serotyping of YE is a first approach for diagnostics, prognosis of disease and epidemiological studies (section 1.3.1). In fact, only 11 serotypes have been associated with human yersiniosis, namely O:8, O:3, O:9, O:5,27, O:13a,13b, O:4, O:18, O:20, O:21, O:1,2,3, O:2,3 (Bottone,

1997). In particular, pathogenic strains mainly belong to a restricted group of serotypes (O:8, O:3, O:9 and O:5,27) and their mouse virulence varies, as serotype O:8 strains are more virulent than O:3 or O:9 strains. Notably, *in vitro* binding and invasion of different host cells have been recently shown to possess serotype-specific characteristics (Schaake et al., 2013). Transfusion-associated YE bacteremia have been related to few serotypes, O:3, O:9, O:5,27, O:1,2,3, and O:20, whereas secondary sequelae of acute YE infection have been correlated to serotypes O:8, O:3 and O:9 (Bottone, 1997).

1.2. Pathogenesis of *Yersinia enterocolitica*

1.2.1. Pathogenesis model and clinical manifestations

Y. enterocolitica is a foodborne pathogen, which causes human infections after ingestion of food or contaminated water (Black et al., 1978; Keet, 1974). Once ingested, the bacteria colonize the intestinal tract, especially the terminal ileum and proximal colon, where they cross the mucus layer covering the mucosal epithelial cells via binding to M cells of the Peyer's patches. Attachment to M cells, mediated by interaction between the bacterial outer membrane protein Inv (invasin) and β1 integrins, leads to internalization and translocation of YE by the M cells across the mucosal barrier. Bacteria are then expelled from the basolateral side of the M cells into the dome region of the follicle-associated epithelium (FAE) and are phagocyted by dendritic cells, macrophages and lymphocytes (Autenrieth and Firsching, 1996; Schulte et al., 2000). Through phagocyte migration, YE gain access to the subjacent tissues, disseminating to the mesenteric lymph nodes, liver, and spleen, where it proliferates in an extracellular form within micro-abscesses (Dube, 2009) (Figure 2). Invasion and survival of YE within different host tissues are multifactorial processes which require a large set of virulence factors encoded by the chromosome and by the *Yersinia* virulence plasmid (pYV) (Cornelis et al., 1998; Revell and Miller, 2001) (sections 1.2.2 and 1.2.3 and Table 3).

Illness usually begins within 24-72 hours after infection with YE, causing mainly gastrointestinal disease, as well as extra-intestinal disorders, under defined host conditions (Revell and Miller, 2001). The manifestation of different intestinal syndromes is associated with serotype of the invading strain, age of host, and host status. In children, acute enteritis, fever and diarrhea are the most frequent occurrences, while in young adults acute terminal ileitis and mesenteric lymphadenitis mimicking appendicitis are more common clinical symptoms (Black et al., 1978; Chandler and Parisi, 1994). The disease generally evolves as a self-limiting gastroenteritis, which may rarely results in fatal septicemia and abscesses formation in deep organs (Savin and

Carniel, 2008). Reactive arthritis is a post-infection phenomenon highly associated with YE; glomerulonephritis, endocarditis, erythema nodosum, uveitis and thyroid disorders are suggested to be secondary immunologically-induced diseases rarely occurring (Simonet, 1999).

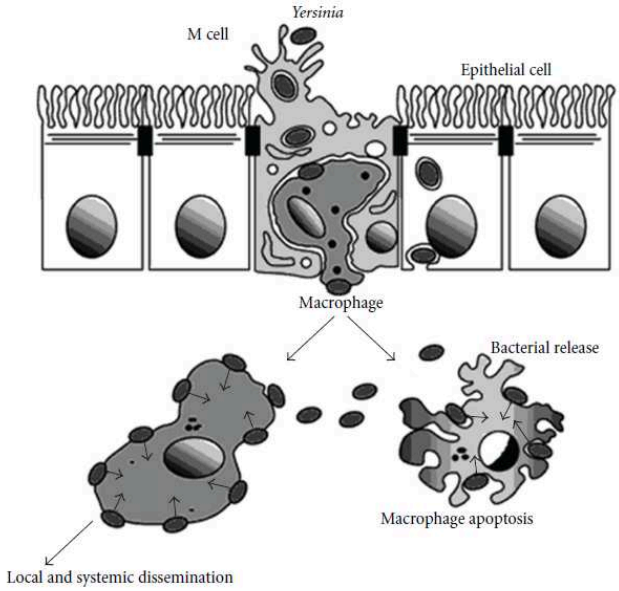


Figure 2 Pathogenesis model of *Y. enterocolitica* infections.
After adhesion to M cells, bacteria are internalized and transported towards the basolateral side of the M cell, where they are expelled into the sub-mucosa to gain access to and multiply in subjacent tissues. Adapted from (Sabina et al., 2011).

1.2.2. Plasmid-encoded virulence factors

The presence of the pYV virulence plasmid, common to all the three pathogenic *Yersinia* species, is a critical element for the manifestation of a pathogenic phenotype in *Y. enterocolitica*. In fact, virulent strains of both subsp. *enterocolitica* and *palearctica* (biotypes 1B and 2-5) carry this plasmid, while the non-virulent biotype 1A strains do not have it. The 70-kb pYV plasmid governs the synthesis of a set of effector proteins called Yops (*Yersinia* outer membrane proteins), which are secreted by a type 3 secretion system (T3SS), namely the *Yersinia* secretion (Ysc-T3SS) locus, also encoded by the pYV plasmid. The adhesion protein YadA is another major virulence determinant synthetized by a pYV-encoded gene (see below).

The Ysc-T3SS secretion machinery inhibits the host immune defenses by injecting cytotoxic Yops into the host cell cytosol, after contact between *Yersinia* and eukaryotic target cells (Cornelis, 2010). The Ysc-T3SS, whose basal structure resembles the bacterial flagellum, is composed of scaffold proteins, export apparatus proteins, ATPase complex, cytoplasmic

components, needle and translocator complex (Dewoody et al., 2013b). The complete injectisome comprised a basal body, a needle and a translocator complex, which span the cytoplasmic membrane, the peptidoglycan layer, and the outer membrane of the bacterium, extend to the extracellular milieu, and inserts into the host cell cytoplasmic membrane (Cornelis, 2002). The efficient translocation of effector Yops into the cytosol of target cells requires the three translocator proteins, YopB, YopD and LcrV, which form a pore in the host cell membrane (Håkansson et al., 1993; Pettersson et al., 1999). The Yop translocation is tightly regulated in order to optimize the infectious process by *Yersinia*. *In vitro*, a temperature shift from 26 °C to 37 °C and low calcium levels activate the Ysc-T3SS, whereas *in vivo* a physical contact with host cells is necessary (Heesemann et al., 1986a; Rosqvist et al., 1994).

Once translocated into the host cell, Yop effector proteins enable survival of the invading *Yersinia* and their proliferation in lymphoid tissues (Table 4). The seven Yop effectors identified so far impair the dynamics of the cytoskeleton, therefore blocking phagocytosis by macrophages and polymorphonuclear leukocytes (YopH, YopT, YopE and YopO/YpkA); induce apoptosis in macrophages (YopP/YopJ); and interfere with the production of pro-inflammatory cytokines, chemokines and adhesion molecules (YopH) (Cornelis and Wolf-Watz, 1997). YopQ/YopK mainly regulates the translocation rate of Yop effectors into host cells and probably prevents the inflammasome activation (Brodsky et al., 2010; Dewoody et al., 2013a; Holmström et al., 1997; Thorslund et al., 2013). YopM, whose function still remains unclear, is the only effector detected into the nucleus of target cells (Skrzypek et al., 1998) and has been shown to be necessary for virulence in a mouse model (Leung et al., 1990).

The non-fimbrial homotrimeric adhesin YadA plays a significant role in adherence and invasion of target cells by YE (Heesemann and Grüter, 1987; Pepe et al., 1995). YadA interacts with extracellular matrix proteins, such as cellular fibronectin, collagen and laminin (El Tahir and Skurnik, 2001), and protects the bacteria against the innate immune response of the host (Heise and Dersch, 2006). Transcription of *yadA* is controlled by two regulators, the plasmid-borne VirF and the chromosome-encoded YmoA, and *yadA* expression is upregulated at 37 °C. Osmolarity, pH, Fe^{3+} ion concentration and contact between bacterium and target cell also affect *yadA* transcription (Portnoy et al., 1984).

1.2.3. Chromosome-encoded virulence factors

Besides the important role played by the pYV plasmid, numerous chromosomal factors contribute to the virulence of *Y. enterocolitica*. Importantly, as summarized in Table 3, differences in the distribution of virulence determinants exist among the five YE biotypes.

Virulence factors found in all isolates include the invasin InvA, the LPS molecule, the Myf antigen, the Yst toxin, the urease and the flagella clusters. The Ail protein is present in pathogenic biotypes (1B and 2-5), whereas the high pathogenicity island HPI, the Ysa-T3SS and the Yts1-T2SS are harbored only by highly-virulent strains of biotype 1B. Lastly, the weakly-virulent group of biotypes 2-5 specifically carries the Ysp/Ysa2-T3SS, Flag-2, Rtx and Tc clusters (Table 3).

The invasin InvA is a 92 kDa outer membrane protein required for the entry of bacteria into the intestinal epithelial cells, by attachment to β1 integrins and stimulation of remodeling of actin filaments in the cytoskeleton of M cells (Miller et al., 1988; Pepe and Miller, 1993). Additionally, InvA is able to induce phagocytosis of *Yersiniae* by macrophages (Deuretzbacher et al., 2009). Temperature, pH and growth phase have a great influence on *invA* expression, which is significantly reduced at 37 °C (Pepe and Miller, 1993; Uliczka et al., 2011). InvA production has been shown to be different among various YE serotypes. Strains of serotypes O:8 and O:9 efficiently express *invA* at environmental temperatures, while InvA is constitutively produced at both 22 °C and 37 °C in serotype O:3 strains (Uliczka et al., 2011).

The lipopolysaccharide (LPS) is a major component of the membrane in gram-negative bacteria. Composed of lipid A, core and O-antigen, its completeness is necessary for full virulence of YE and the outer core of serotype O:3 has demonstrated virulence activity (al-Hendy et al., 1992; Skurnik et al., 1999; Zhang et al., 1997). Mutations in the O-ag gene cluster of YE serotype O:8 were frequent in signature-tagged-mutagenesis screens, indicating a critical role for the O-ag during infection (Gort and Miller, 2000; Young and Miller, 1997).

The Myf antigen, encoded by several *myf* genes, constitutes a fibrillar structure, with MyfA being the main subunit. The other proteins of the cluster include MyfB, a putative chaperone, MyfC, the membrane usher protein, and two transcriptional regulators, MyfE and MyfF (Iriarte et al., 1993). MyfA shows high similarity to the pH 6 antigen Psa, whose role in thermo-inducible binding and haemagglutination in *Y. pseudotuberculosis* is uncertain (Yang et al., 1996). In YE, however, the myf operon is not able to mediate haemagglutination and, therefore, its role in pathogenesis has to be clarified.

YE is able to synthetize three highly identical heat-stable enterotoxins, encoded by the genes *ystA*, *ystB* and *ystC*. YstA, which is specifically produced by virulent strains, has a controversial role in YE pathogenesis. It induces fluid accumulation by increasing the concentration of cyclic GMP levels within intestinal epithelial cells (Robins-Browne et al., 1979), and stimulates both the nuclear and cytosolic levels of calcium and inositol triphosphate in rat intestinal epithelial cells, acting as causative agent of secretory diarrhea (Saha et al., 2009). On the other hand, a YE

serotype O:3 *ystA*-mutant maintained its virulence in a piglet model, and no toxin production was detected in diarrheal stool samples of experimental animals with YE-induced diarrhea (Robins-Browne et al., 1985). The enterotoxin YstB is detected in isolates of the non-virulent biotype 1A (Ramamurthy et al., 1997), whereas YstC, whose gene is present in both virulent and non-virulent isolates (Ramamurthy et al., 1997), demonstrated the highest toxicity of the three enterotoxins (Yoshino et al., 1995).

Table 3 Relevant *Y. enterocolitica* virulence determinants and comparison among the three virulent groups. Modified from (Rakin et al., 2015).

Genomic region / Gene cluster	**Function**	**Highly-virulent (biotype 1B)**	**Weakly-virulent (biotypes 2-5)**	**Non-virulent (biotype 1A)**
pYV plasmid (Yop, Ysc, YadA)	Invasion of the host	+	+	-
InvA	Invasion	+	+	+
Myf cluster	Fibrillious adhesins	+	+	+
Urease cluster	Protection from acidic environment	+	+	+
Flag-1	Motility	+	+	+
T2SS *yts2*	GSP	+	+	+
Ail	Attachment and invasion	+	+	-
YstA	Enterotoxin	+	+	-
YstB	Enterotoxin	-	-	+
HPI-PZ (Yersiniabactin)	Dissemination in the host	+	-	-
T3SS *ysa*-PZ	Colonization of GI tract	+	-	-
T2SS *yts1*-PZ	GSP	+	-	-
T3SS *ysp/ysa2*	GSP	-	+	+
Rtx cluster	Toxin	-	+	+/-[a]
Flag-2-PZ	Motility	-	+	-
Tc proteins	Insecticidal toxin	-	+	-

[a]+/-: detected in one 1A genome with low protein homology to the cluster of biotype 2-5 strains.

Abbreviations: Ail, attachment invasion locus; GI, gastrointestinal; GSP, general secretion pathway; HPI, high pathogenicity island; Myf, mucoid *Yersinia* factor; pYV, *Yersinia* virulence plasmid; PZ, Plasticity zone; RTX, Repeats in Toxin; T2SS, type II secretion system; T3SS, type III secretion system; Tc, toxin complex; Yop, *Yersinia* outer membrane protein; *ysa*, *Yersinia* secretion apparatus; Ysc, Yop secretion; *yts*, *Yersinia* type II secretion; Yst, *Yersinia* heat-stable toxin.

Table 4 *Yersinia* Yop effectors and their functions during human infection. Modified from (Rakin et al., 2015).

Effector	Biological activity	Target	Effect/feature
YopE	GAP	Rac1, RhoG	Actin depolymerization, inhibition of phagocytosis, of reactive oxygen species (ROS) production, and of the inflammatory response
YopT	Cysteine protease	Rac, RhoA, Cdc42	Actin depolymerization, inhibition of phagocytosis
YopO/YpkA	Ser/Thr kinase, GDI	Gαq, RhoA, Rac	Actin depolymerization, inhibition of phagocytosis
YopH	PTPase	$P130^{Cas}$, FAK, Paxillin, Fyb, SKAP-HOM, p85, Lck	Disruption of peripheral focal complexes, impairing of T- and B-cell activation, inhibition of phagocytosis, of reactive oxygen species (ROS) production, and of the inflammatory response
YopP/YopJ	Cysteine protease Acetyltransferase ?	MAKKs, IKKβ, TRAF2, TRAF6, IKKα, IKKβ, and IkBα	Prevention of NF-kB and MAPK activation, inhibition of inflammatory response, apoptosis in macrophages and dendritic cells
YopM	LRR protein	PRK2, RSK1	Disruption of inflammasome formation and caspase-1 activation, depletion of NK cells, induction of IL-10 production
YopQ/YopK	Unknown	Inflammasome	Inhibition of inflammasome activation by regulating the rate of Yops translocation and preventing detection of T3SS by innate immune system

Abbreviations: GAP, GTPase activating protein; GDI, guanine nucleotide dissociation inhibitor; PTPase, protein tyrosine phosphatase; FAK, focal adhesion kinase; Fyb, Fyn-binding protein; MAKK, mitogen-activated kinase kinase; IKKβ, inhibitor-kappa B kinase β; NF-kB, nuclear factor kB; MAPK, mitogen-activated protein kinase; LRR, leucine-rich repeat; PRK2, protein kinase C-like 2; RSK1, ribosomal S6 protein kinase 1; NK, natural killer.

All *Yersinia* species, except *Y. pestis*, show a positive phenotype for urease, a trimeric cytosolic enzyme able to hydrolyze urea to ammonia and carbon dioxide (de Koning-Ward and Robins-Browne, 1996). The gene cluster is homologous between the three pathogenic *Yersinia* species and it comprises structural (*ureA*, *ureB*, *ureC*) and accessory (*ureE*, *ureF*, *ureG*, *ureD*) genes (de Koning-Ward et al., 1994). It is suggested that these genes are arranged in an operon-like manner, but whether they form a single unit (*ureABCEFGD*) or two polycistronic units (*ureABC* and *ureEFGD*) is not understood. Importantly, YE biotype 1A strains harbor a smaller *ureB* gene and larger intergenic regions (Bhagat and Virdi, 2009). *Yersinia* urease is a novel type of bacterial urease, since it is optimally active at acidic pH, it is maximally expressed at 28 °C rather than 37 °C, and it has a high affinity for urea (de Koning-Ward and Robins-Browne, 1997).

Yersinia may primarily use urease to live as saprophytes, degrading urea from soil and water. Urease may also contribute to the pathogenesis of YE by protecting the bacteria during their passage through the stomach, as supported by the fact that serotype O:3 and O:8 strains require urease to be fully virulent in infection models (Gripenberg-Lerche et al., 2000).

Flagella are organelles that provide means for motility and for the export of proteins to the environment. The ability to produce flagella distinguishes the gastrointestinal pathogens YE and Y*. pseudotuberculosis* from *Y. pestis*, which has permanently lost this capability by mutations in the gene cluster (Young, 2004). Flagellum biosynthesis involves circa 50 genes, organized in the flagellar system 1 (Flag-1), which are expressed in a hierarchical cascade in three major flagellar gene classes. Class I genes consist of the *flhDC* operon and encode the transcriptional activators FlhD and FlhC, which are required for induction of all other flagellar genes (Young et al., 1999). Class II genes encode basal body and hook structural proteins as well as two regulatory components, FlgM and FliA (Iriarte et al., 1995; Kapatral et al., 1996). Class III genes encode proteins necessary for formation of the flagellum and chemosensory system. The production of flagella is regulated by environmental stimuli, such as temperature and salt concentration (Kapatral and Minnich, 1995; Kapatral et al., 1996; Young et al., 1999). Flagella may contribute to disease pathogenesis by initiating host cell invasion, helping the bacteria to migrate and attach to epithelial cells (Young et al., 2000). Experimental evidences support a reverse co-regulation of Inv and flagella expression, so that *Yersinia* are not motile after attachment to the host gut tissues (Badger and Miller, 1998).

The attachment invasion locus (Ail) protein is a 17 kDa adhesin highly correlated with virulence, as it is present only in pathogenic strains. Its expression occurs exclusively at 37 °C and is affected by pH and oxygen level (Pederson and Pierson, 1995). Ail is a polypeptide folded into eight amphipathic β-sheets in the bacterial outer membrane (Miller et al., 1990). Ail is involved in adhesion to and invasion of tissue culture cells, as well as in resistance to serum, together with the pYV-encoded YadA (Miller et al., 2001).

Presence of the High-Pathogenicity Island (HPI), encoding an iron acquisition system, is a main prerequisite for the highly pathogenic phenotype in *Yersinia*, as loss or inactivation of this system results in significant virulence attenuation in a mouse model (Heesemann et al., 1993; Perry and Fetherston, 2011). The HPI is restricted to some subgroups of the three pathogenic *Yersinia* species: serotypes O:1 and partly O:3 of *Y. pseudotuberculosis*, biotype 1B of YE, and all three biotypes of *Y. pestis* (de Almeida et al., 1993; Rakin and Garzetti, 2013; Rakin and Heesemann, 1995). The YE HPI genomic region is 43.1 kb long and can be divided into two distinct parts: the highly conserved right-hand part, named the “yersiniabactin” locus, and the

variable left-hand part (Rakin et al., 1999). The genetic locus of the high-affinity, iron-chelating system yersiniabactin is composed of 11 genes, organized in four operons, which can be divided into three functional parts, namely biosynthesis, transport into the bacterial cell, and regulation (Pelludat et al., 1998). Yersiniabactin biosynthesis is strictly regulated by the transcriptional regulator YbtA and the FurA repressor (Perry et al., 1999). The left-hand part of YE biotype 1B HPI contains a cluster of four IS elements and seven other ORFs that have no significant similarity to any known gene (Carniel et al., 1996; Rakin et al., 1999). The HPI confers on the bacteria the ability to capture iron ions bound to eukaryotic molecules (such as hemoglobin, transferrin, ferritin and myoglobin) via the yersiniabactin, and subsequently to disseminate in mammalian hosts and to cause systemic infections. Therefore, the ability of YE biotype 1B strains to proliferate and be virulent in mice strongly correlate with the capability to capture iron from host tissues through yersiniabactin (Heesemann, 1987).

Besides the pYV-encoded Ysc T3SS, to colonize the GI tract YE biotype 1B can utilize a chromosomal T3SS, named Ysa-T3SS (*Yersinia* secretion apparatus) (Haller et al., 2000). The Ysa-T3SS genetic cluster is located within the so called “plasticity zone” (PZ) region in serotype O:8 strains (Thomson et al., 2006) and resembles the T3SS-1 encoded within the *Salmonella enterica* SPI-1 and the *Shigella* Mxi-Spa T3SS (Foultier et al., 2002). Eight Ysp (*Yersinia* secreted protein) effectors have been so far identified: YspA, YspE, YspF, YspI, YspK, YspL, YspM, and YspP (Foultier et al., 2003); however, functional activity has been clarified for only few of them (Matsumoto and Young, 2006, 2009). The Ysa-T3SS has been shown to play an important role during the earliest stages of infection in the GI tract, and to facilitate the overcoming of the host immune barriers (Venecia and Young, 2005).

Two T2SS, Yts1 and Yts2, have been identified in YE (von Tils et al., 2012). Yts2 is present in all YE isolates, while Yts1 can be found only in highly-virulent YEE strains, as part of the PZ (Iwobi et al., 2003; Thomson et al., 2006). Yts1, mainly active at low temperatures and in presence of high concentrations of Mg^{2+}, is involved in dissemination and colonization of deeper tissues, such as liver and spleen (von Tils et al., 2012). Additionally, a role of the Yts1 system in survival of the bacteria in an environmental habitat has been suggested by *in vitro* functional characterization of the Ysp secreted proteins (Shutinoski et al., 2010). Concerning the Yst2 T2SS, its biological function is still unknown.

Y. enterocolitica subsp. *palearctica* lack the HPI and the two Ysa and Yts1 secretion systems specifically present in the highly-virulent YEE strains. Nevertheless, strains of biotypes 1A and 2-5 carry additional factors, potentially associated with their adaptation to the host. The Ysp/Ysa2-T3SS, homologous to the SPI-2 T3SS in *Salmonella* and collinear to respective regions in

Y. intermedia and *Y. frederiksenii*, lack identified effector proteins (Batzilla et al., 2011a). Its biological activity and, therefore, its possible role in YEP pathogenesis require a better characterization.

An Rtx cluster, detected in biotypes 2-5 and in some biotype 1A genomes, encodes an RtxC activator, an RtxH peptide chain release factor, and a large RtxA cytotoxin (Batzilla et al., 2011a; Fuchs et al., 2011; Wang et al., 2011). This toxin may contribute to YE pathogenesis, by forming pores in the host cell membranes and/or inducing Ca^{2+} release, as it has been demonstrated for Rtx toxins produced by other Gram-negative bacteria (Linhartová et al., 2010; Liu et al., 2009; Lo et al., 2011).

For another toxin complex, the insecticidal toxin cluster (Tc), biological activity has been demonstrated in a weakly-virulent YE serotype O:9 strain, which showed higher toxicity toward the larvae *M. sexta*, compared to toxin-depleted strains (Fuchs et al., 2008). However, a function of the insecticidal Tc toxin in *Yersinia* virulence to mammals cannot be excluded.

A second flagellar cluster, Flag-2, absent from biotypes 1A and 1B strains, may contribute to motility of weakly-virulent strains. Functionality of the genes from a YE serotype O:9 strain has been proved by swarming assay, and their expression was high at 20 °C and dependent on the Flag-1 master regulator FlhC (Bresolin et al., 2008). The possible role of this cluster in virulence and host cell invasion is still unproven.

1.3. Isolation and diagnosis of *Yersinia enterocolitica*

1.3.1. Culture-based methods and serotyping

Cultivation and isolation of *Y. enterocolitica* from clinical, food and environmental samples is the first step for identification of pathogenic strains and diagnosis of yersiniosis, whose causative agents include not only YE but also *Y. pseudotuberculosis* (EFSA and ECDC, 2013). Isolation methods are affected by the source and stage of infection. Direct plating of samples on conventional enteric media is usually successful for pathogenic strains from patients with acute gastroenteritis or organ abscesses, where YE is the dominant bacterium. On the other hand, in asymptomatic carriers, food or environmental samples, this method is rarely successful and enrichment in liquid media before isolation on solid media is necessary (Fredriksson-Ahomaa and Korkeala, 2003). Cold enrichment (Aulisio et al., 1980) takes advantage of the psychrotrophic nature of YE. It is performed incubating the specimens at 4 °C for 1-3 weeks in phosphate buffer or in phosphate-buffered saline (PBS) with sorbitol and bile salts (PSB), with or without treatment with potassium hydroxide, and plating the inoculum onto selective agar plates,

such as CIN (Cefsulodin-irgasan-novobiocin) or SSDC (*Salmonella-Shigella* deoxycholate calcium chloride) agar (Schiemann, 1979; Wauters, 1973). Alternatively, selective enrichment in media culture with antimicrobial agents can be used. Several media have been reported: MRB (modified Rappaport broth) and ITC (irgasan, ticarcilin, and chlorate) media successfully recovered YE strains of bioserotype 4/O:3 (de Boer, 1992), while the BOS (bile-oxalate-sorbose) medium is particularly useful for bioserotype 1B/O:8 (Schiemann, 1982). Enrichment in PSB broth has shown to give better results for non-pathogenic strains, whereas for pathogenic strains enrichment in ITC broth is more efficient (Wauters et al., 1988). Identification of YE from colonies on CIN agar plates is then carried out by biochemical tests such as fermentation of sucrose, rhamnose, and melibiose. Commercial systems like API 20E and ID 32E provide reliable alternatives to conventional tube tests (Manafi and Holzhammer, 1994; Neubauer et al., 1998).

Pathogenicity of YE isolates can be assessed by the pyrazinamidase and autoagglutination tests (Kandolo and Wauters, 1985; Laird and Cavanaugh, 1980). The pyrazinamidase test allows differentiation between potentially pathogenic and nonpathogenic strains, showing negative and positive reactions, respectively. Autoagglutination at 37 °C is a phenotypic characteristic associated with the virulence plasmid, and in particular with the YadA protein, together with calcium dependence and uptake of Congo red and crystal violet, which are all indirect markers used to identify pathogenic YE isolates. Notably, the pYV plasmid can be lost after a few subcultures, especially at 37 °C, of YE strains, thus showing negative reactions in tests fro pYV-associated phenotypes (Fukushima et al., 2011). Therefore, direct methods are required to assess the potential pathogenicity of a strain, even if pYV-negative.

Since a correlation exists between the serotype of YE isolates and their ability to cause infection, serotyping is an important diagnostic tool for this species (see section 1.1.1). Classical serotyping of YE is performed by slide agglutination with O-antigen specific typing antisera. Commercial monovalent (serotypes O:3, O:8, O:9, O:5) or polyvalent (serotypes O:1 and O:2) antisera from immunized rabbits are nowadays extensively used. If the tested YE strain possesses the O-ag recognized by the antisera antibody, a clear agglutination reaction of the bacterial mass will appear. Despite the feasibility and simplicity of this method, it has several drawbacks: isolates which do not express the O-ag (rough LPS) in experimental conditions cannot be typed with the traditional agglutination test; cross-reactions may occur, as between YE serotype O:9 and *Brucella* sp. (Godfroid et al., 2002); it is not possible to infer virulence by serotyping itself, since, for example, some strains of serotype O:8, frequently O:8,7, belong to the non-virulent group 1A (Bottone, 1997); subjective interpretation of the results is required. Even though a PCR-based O-genotyping method has been developed for identification and

serotyping of *Y. pseudotuberculosis* and *Y. pestis* (Bogdanovich et al., 2003), such a test is not yet available for *Y. enterocolitica*.

1.3.2. Molecular methods for *Y. enterocolitica* detection

With the conventional culture-dependent methods, identification of *Y. enterocolitica* can be supposedly done in 36-48 hours, but confirmation may take more than 14-21 days. Moreover, biochemical reactions do not guarantee good identification at species level. In the past decades, researchers have attempted to develop rapid and reliable molecular methods for detecting YE directly in natural samples (Fredriksson-Ahomaa et al., 2006). Nucleic acid-based methods permit fast detection of DNA with high specificity and sensitivity. Several PCR, real-time PCR and colony hybridization assays have been designed to verify the pathogenicity of YE isolates, targeting specific region of the virulence plasmid and/or the chromosome (Fredriksson-Ahomaa and Korkeala, 2003). The most used pYV-encoded genetic targets are *yadA*, encoding the essential virulence protein YadA, and *virF*, a regulatory gene involved in the coordinated thermal induction of virulence determinants (Brubaker, 1991). Due to the possible plasmid loss during laboratory subculturing, chromosomal gene targets, including *ail*, *inv*, *yst* and *Yersinia*-specific region of the 16S rRNA gene, have been applied to molecular diagnosis of YE (Fredriksson-Ahomaa and Korkeala, 2003). Multiplex PCR assays able to detect various genotypic features have been also designed, combining pYV- and chromosome-encoded determinants (Harnett et al., 1996; Kaneko et al., 1995). Single PCR assays targeting the *rfbC* (dTDP-4-deoxyrhamnose-3,5-epimerase) and *per* (perosamine synthetase) genes of the O-antigen clusters have allowed differentiation of YE serotypes O:3 and O:9, respectively, from other pathogenic serotypes (Jacobsen et al., 2005; Thisted Lambertz and Danielsson-Tham, 2005). Importantly, due to the complex composition of natural samples such as blood, feces, meat and soil, PCR inhibitory compounds may be present and sensitivity of the test reduced (Rossen et al., 1992). Therefore, sample preparation steps prior PCR are recommended, including enrichment, dilution, filtration, centrifugation, and adsorption (Fredriksson-Ahomaa and Korkeala, 2003).

1.4. Epidemiology of *Yersinia enterocolitica*

In Europe, yersiniosis is the third most frequently reported zoonotic gastrointestinal disease in humans, after campylobacteriosis and salmonellosis (ECDC, 2013). The number of yersiniosis cases has dropped with a statistically significant five-year trend in 2007-2011 (Figure 3). In particular, six countries reported this reduction (Denmark, Germany, Lithuania, Slovenia, Spain

and Sweden), while an increasing trend was noted in Hungary, Romania and Slovakia (EFSA and ECDC, 2013).

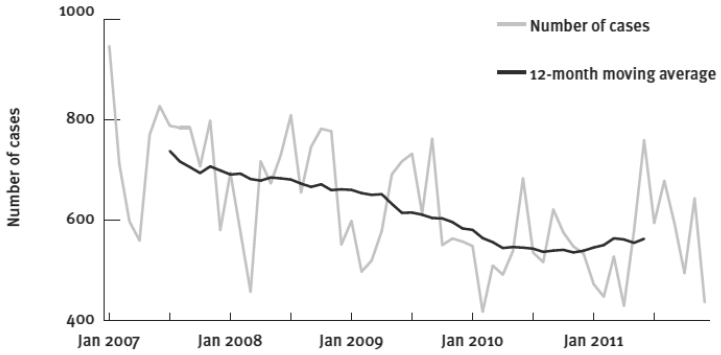


Figure 3 Trend of confirmed cases of yersiniosis reported in Europe. Modified from (ECDC, 2013).

In 2011, a total of new 7,041 confirmed cases (2.20 cases per 100,000) of human yersiniosis were reported in the EU, increasing by 2% on 2010 (ECDC, 2013). In the same year in the USA the number of newly confirmed *Yersinia* infections was 163 (0.34 per 100,000 population) (CDC, 2012), indicating a clearly higher incidence of YE foodborne infections in European countries compared to USA. Importantly, yersiniosis cases in Germany accounted for almost half of all reported European yersiniosis events (47.5%, 3,345 confirmed cases), while Lithuania and Finland showed the greatest notification rates (11.40 and 10.31 cases per 100,000 population, respectively) (ECDC, 2013), supporting the high incidence of yersiniosis in temperate and cold countries. High numbers of cases of gastrointestinal illness due to yersiniosis have been also observed in many developing countries like Bangladesh, Iraq, Iran, and Nigeria, indicating probable food safety problems (Rahman et al., 2011). Overall, *Y. enterocolitica* has been isolated in various countries on all continents (Carniel, 2006), and account for 98.4% of confirmed yersiniosis cases (*Y. pseudotuberculosis* represents circa 1%) (EFSA and ECDC, 2013). Biotype 4 is the most prevalent biotype isolated from humans, followed by biotype 2 and biotype 3, with serotype O:3 predominating in Europe, Japan and North America, serotype O:9 in England and serotype O:5,27 in Russia (Griffin, 2008). Also biotype 1A strains, even though lacking most of the classical virulence markers, have been isolated from patients with diarrhea, supporting the pathogenicity of some biotype 1A strains (Burnens et al., 1996; Greenwood and Hooper, 1987; Huovinen et al., 2010; Ratnam et al., 1982).

Human infections most frequently occur as sporadic cases or, rarely, as small outbreaks. Direct person-to-person contact has been suspected as source of infection in a nosocomial outbreak due to YE bioserotype 1A/O:5 in hospitalized patients (Ratnam et al., 1982), and indirect person-

to-person transmission by transfusion of blood products from donors suffering from latent yersiniosis has often occurred (Bottone, 1999). However, transmission of YE occurs most commonly by the fecal-oral route. Although transmission of pathogenic YE from pigs to humans has not been proven, the main sources of human infection are thought to be through the handling of swine and the consumption of raw pork meat. Pigs do not develop diseases, but are known to carry the bacteria in their oropharynx, nasopharynx and their intestine (Nesbakken et al., 2003; Thibodeau et al., 1999). Consumption of contaminated and inadequately heated food has been suggested as the most common source of intestinal yersiniosis, even though pathogenic isolates have rarely been recovered from food samples (Prpic et al., 1985), probably due to limited sensitivity of culture methods (Fredriksson-Ahomaa and Korkeala, 2003). Nevertheless, YE has been detected in milk and milk products, egg products, beef, pork, lamb and poultry meats, vegetables, and other food products (Rahman et al., 2011). Drinking water has been reported as a source of YE biotype 1B infection (Ostroff et al., 1994). Overall, the majority of YE isolates recovered from environmental samples, such as from slaughterhouses, soil, and water, belong to the non-virulent biotype 1A (Berzero et al., 1991; Cork et al., 1995).

Y. enterocolitica is able to endure repeated freezing and thawing cycles and to survive in frozen foods (Bhaduri, 2006), making refrigerated aliments potential vehicles for infection. Additionally, YE grows, even though with low efficiency, at refrigeration temperatures thanks to its psychotropic nature. Notably, poor sanitation, inadequate sterilization techniques and improper storage may contribute to food contamination. Therefore, general food hygiene guidelines, like accurate washing of raw fruits and vegetables, adequate food cooking, and consumption of pasteurized milk, might significantly reduce the risk of YE contamination (Kapperud, 1991; Lampel et al., 2009).

1.4.1. Molecular subtyping of *Yersinia enterocolitica*

Subtyping is an important tool in epidemiology for recognizing outbreaks, detecting cross-transmission of pathogens and determining the source of an infection. *Y. enterocolitica* strain differentiation has typically relied on phenotypic characteristics, such as biotype, serotype, phage type and antibiogram type. These methods have provided valuable information on the epidemiology and taxonomy of this microorganism; however, they have some limitations. The biotyping scheme well correlates with the pathogenicity of YE, but atypical strains are frequent. Preparing monospecific antisera for serotyping is tedious and technically laborious; moreover, commercial antisera are available for only some serotypes, and a complete set of antisera is accessible in few laboratories in the world. Phage typing, being also a technically demanding

procedure, is performed in specialized laboratories and is not commonly used. Antibiotic susceptibility tests have demonstrated limited efficacy in assessing heterogeneity among YE isolates (Virdi and Sachdeva, 2005). Therefore, also due to technological advances in molecular techniques, genotyping methods have recently become more reliable typing tools. Molecular subtyping of YE has been performed by different approaches, allowing characterization of strains belonging to the same bioserotype and, consequently, tracking the spread of clones and the source of an infection (Table 5).

Multi-locus enzyme electrophoresis (MLEE) relies on the relative electrophoretic mobility of a large number of intracellular enzymes. Differences in mobility are directly related to amino acid substitutions, caused by mutations at the corresponding gene locus, which affect the net charge of the protein. The unique profile produced by each strain is called electromorph type (ET). Application of MLEE for YE subtyping has shown clustering of strains into two groups, represented by serotypes O:1,2, O:3, O:5,27 and O:9 on one side, and O:8, O:13 and O:21 on the other (Caugant et al., 1989). Other studies have indicated that YE strains belong to one cluster, and subgroups within this cluster most often equal the bioserotypes of the isolates (Dolina and Peduzzi, 1993; Goullet and Picard, 1988). However, within one bioserotype the discriminatory power was not enough satisfactory, as strains of different serotypes could have the same ET.

Restriction endonuclease analysis of chromosome (REAC) requires cleavage of genomic DNA by restriction enzymes and separation of the resulting fragments by gel electrophoresis. REAC has been used to study polymorphisms among YE strains belonging to different bioserotypes, resulting in 22 bioserotype-specific distinct fragment patterns amongst 72 *Yersinia* strains (Kapperud et al., 1990). Some DNA profiles also varied within the same bioserotype. This method showed the best discriminatory power within bioserotype 1B/O:8 strains, whereas bioserotypes 3/O:3, 4/O:3, 2/O:5,27 and 2/O:9 were more homogeneous. Moreover, most of the strains of bioserotype 4/O:3 isolated in Norway could not be distinguished, indicating an overall genetic homogeneity. The interpretation of complex profiles, consisting of hundreds of unresolved and overlapping bands, is a major limitation of this technique. Restriction endonuclease analysis of the plasmid (REAP) is based on the comparison of banding patterns which result from extraction and endonuclease-digestion of the pYV virulence plasmid present in pathogenic YE strains. Mostly, REAP studies have demonstrated serotype-specific profiles, with O:8 strains displaying the highest degree of polymorphism, and O:3 and O:9 strains exhibiting low diversity (Iteman et al., 1996; Kapperud et al., 1990; Kwaga and Iversen, 1993; Nesbakken et al., 1987). Another study reported close correlation between REAC and REAP patterns of

serotype O:5,27 strains (Fukushima et al., 1993). The REAP method is rapid and easy to perform, but it is not applicable to non-virulent strains or to virulent strains cured of their plasmid. Overall, it appears that the use of REAC and REAP for molecular epidemiological tracing of YE might not bring valuable information.

Ribotyping refers to the analysis of the restriction fragment length polymorphisms generated after hybridization of digested bacterial genomic DNA with a ribosomal probe. Strain-specific fingerprints can be obtained based on the number and position of the rRNA operons. Several studies have applied this technique to YE typing, but the results are not comparable since different enzymes have been used (Fukushima et al., 1998; Lobato et al., 1998; Mendoza et al., 1996). In these analyses ribotyping has partly allowed the distinction of subtypes within a certain bioserotype, with good correlation between phenotypic and ribotypic traits. However, ribotypic polymorphisms appeared to be higher for biotype 1A or 1B strains detected in the environment, than for those adapted to animal hosts, such as bioserotypes 4/O:3 and 2/O:9. This technique could therefore be useful to trace the spread of YE over the world but it has limited interest for outbreak investigations. Moreover, the use of ribotyping on routine basis is hindered by technical unfeasibility, long experimental time and costs.

Pulsed-field gel electrophoresis (PFGE) is considered the gold-standard for different bacteria molecular subtyping. It permits analysis of the whole bacterial genome, using various rare-cutting restriction enzymes and separation of the resulting large-size DNA fragments by special electrophoretic apparatus. This technique has shown the highest discriminatory power for YE epidemiological tracing (Iteman et al., 1996). In different studies, strains demonstrated low genomic variability within each bioserotype, and therefore major serotype-specific pulsotypes could be delineated (Asplund et al., 1998; Najdenski et al., 1994, 1995; Saken et al., 1994). Pulsotypes seemed also to be more closely related to the biotype than to the serotype (Buchrieser et al., 1994; Najdenski et al., 1994). PFGE produces highly reproducible restriction profiles that usually show distinct and well-resolved fragments. However, optimal enzyme and gel electrophoresis conditions have not yet been established; the enzymes produce a large number of closely spaced restriction fragments, which make analysis of patterns difficult.

PCR is a versatile technique that has been used in several variations for molecular genotyping of bacteria. The randomly amplified polymorphic DNA (RAPD) assay makes use of the random amplification of genomic DNA at low annealing temperatures with arbitrarily selected primers, which do not target any specific bacterial sequence. Studies that used RAPD for typing YE strains showed a relatively weak correlation between their RAPD clusters and bioserotypes, and only in some cases the technique was able to efficiently discriminate among isolates belonging

to the same bioserotype (Leal et al., 1999; Odinot et al., 1995; Rasmussen et al., 1994). RAPD-defined subtypes of pathogenic YE strains correlated well with the geographical origin (Blixt et al., 2003). RAPD is an easy and quick technique, but has low reproducibility and standardization of the technique is difficult. Repetitive element PCR (rep-PCR) is based on primers targeting interspersed repetitive DNA elements spread through the whole bacterial genome to generate specific genomic fingerprints. Two families of repetitive elements, the repetitive extragenic palindromic (REP) and enterobacterial repetitive intergenic consensus (ERIC) sequences, have been used to differentiate YE strains (Sachdeva and Virdi, 2004; Wojciech et al., 2004). YE isolates with different serotypes produced identical rep-profiles, indicating limited genetic heterogeneity. Overall, although both REP and ERIC genotyping gave comparable results, REP-PCR proved to be more discriminatory than ERIC-PCR. PCR ribotyping, or 16S–23S intergenic spacer region typing, takes advantage of variations in the number of copies and sizes of 16S–23S spacer region of the *rrn* operon. DNA is amplified by PCR with defined primers, the PCR products are digested with a restriction enzyme, and the resulting fragments electrophoretically resolved. YE strains with similar ribotypes have been differentiated into specific sub-types with this method, even though same PCR-ribopatterns resulted from strains belonging to different serotypes (Lobato et al., 1998). The discriminatory power of PCR-ribotyping was shown to be similar to ERIC-PCR but lower than REP-PCR when strains of different bioserotypes were analyzed (Wojciech et al., 2004).

Variable number tandem repeats (VNTRs) are short nucleotide sequences repeated in tandem in variable numbers, causing polymorphisms in these loci. VNTR-based typing is considered a valuable epidemiological tool to characterize bacterial species, especially if genetically homogeneous and difficult to type with other techniques (van Belkum, 1999). A VNTR locus identified and characterized in several strains of YE was found to be polymorphic in strains of bioserotype 4/O:3 (de Benito et al., 2004). Interestingly, analysis of this locus in eight strains belonging to the same pulsotype showed seven different alleles, improving the discriminatory power of PFGE. Multi-Locus VNTR (MLVA) approaches have proved to be a promising tool for YE outbreak investigations, due to their high discriminatory power (Gierczyński et al., 2007; Sihvonen et al., 2011; Virtanen et al., 2013). One study, specific for strains of YEP, could distinguish 76 genotypes among 91 isolates of different bioserotypes and geographical origins, and 41 genotypes among 51 bioserotype 4/O:3 isolates, demonstrating high resolution power. A minor correlation between genotypes and serotypes was observed. A more recent MLVA method has detected 312 different MLVA patterns when genotyping 379 YE isolates from human patients, pigs at slaughterhouses, and pork samples from meat stores. The method was

applicable for serotypes O:3, O:5,27, and O:9 and appeared to be a highly discriminating tool for distinguishing sporadic and outbreak-related strains.

Amplified fragment length polymorphism (AFLP) is an approach based on selective PCR amplification of restriction fragments from digested genomic DNA. Its steps require restriction of DNA, ligation of oligonucleotide adapters, selective amplification of sets of restriction fragments, and gel analysis of the amplified products. AFLP is normally performed with a combination of a 6-base-specific and a 4-base-specific restriction enzyme and fluorescent-labelled primers. Results are interpreted using gene scan analysis software, rendering AFLP an effective means of DNA fingerprinting for bacterial DNA. YE studied by an AFLP method (Fearnley et al., 2005), which included 70 strains belonging to different bioserotypes and sources, showed clustering of strains into two groups, biotypes 2–5 and biotypes 1A and 1B, respectively. AFLP profiles further differentiated strains within these two clusters and within the resulting serotype-related subclusters, indicating a high discriminatory power of this method for YE. Another recent AFLP genotyping study demonstrated high clonality of strains isolated from American pigs, regardless of geographical origin or production system (Tadesse et al., 2013). In some cases, genetic diversity was observed in isolates that originated from the same farm with the same phenotype.

Rapid development of cost-effective NGS (next generation sequencing) techniques has recently made sequencing a good alternative method for genotyping bacterial isolates. A PCR strategy based on 16S rRNA gene sequencing, for example, has been developed for unambiguous identification of isolates identified as YE by biochemical tests (Neubauer et al., 2000b). Moreover, based on the sequencing of specific housekeeping genes, multilocus sequence typing (MLST) methods have been developed for *Yersinia* typing (Duan et al., 2014; Kotetishvili et al., 2005; Laukkanen-Ninios et al., 2011). YE strains showed some polymorphisms between different bioserotypes and within the same bioserotype; however, the pathogenic YE strains formed a relatively conserved group, with serotypes O:3, O:8, and O:9 separated into three distinct blocks.

To conclude, PFGE typing is highly effective in molecular epidemiological studies of *Y. enterocolitica* and is more efficient than other molecular methods in discriminating among isolates of the same bioserotype. The highly-discriminatory AFLP and MLST techniques could provide a means of differentiating YE strains detected as identical with other genotyping tools. The recent progresses in sequencing technologies will allow rapid strain comparison on a single nucleotide level (e.g. SNPs), providing higher sensitivity and discriminatory power for strains within the same serotype.

Table 5 Comparison of methods for molecular subtyping of *Y. enterocolitica* isolates.
Modified from (Fredriksson-Ahomaa et al., 2006; Virdi and Sachdeva, 2005).

Method	Typeability	Reproducibility	Discriminatory power	Use	Interpretation
MLEE	Excellent	Good	Good	Moderate	Moderate
REAP	Variable	Good	Poor	Easy	Easy
REAC	Excellent	Moderate	Moderate	Easy	Difficult
Ribotyping	Excellent	Excellent	Variable	Moderate	Easy
PFGE	Excellent	Excellent	Good	Moderate	Easy
PCR	Excellent	Moderate	Variable	Easy	Moderate
VNTR	Variable	Good	Moderate	Moderate	Moderate
AFLP	Excellent	Good	Good	Moderate	Moderate
DNA sequencing	Excellent	Excellent	Good	Difficult	Moderate

1.4.2. *Yersinia enterocolitica* bioserotype 4/O:3

In Europe, Japan and Canada, *Y. enterocolitica* strains most frequently isolated from human patients belong to bioserotype 4/O:3 (80 to 90%); even though in USA bioserotype 1B/O:8 has traditionally been the most predominant, recently 4/O:3 strains have also emerged (Fredriksson-Ahomaa, 2007). YE serotype O:3 has been responsible for several outbreaks in different countries, but the source of infections remained undetected (Jones, 2003; Lee et al., 1990; Maruyama, 1987; Olsovsky et al., 1975). In general, outbreaks of YE O:3 gastroenteritis are more uncommon than sporadic infections (Fredriksson-Ahomaa, 2007). YE bioserotype 4/O:3 strains pathogenic for humans have been isolated from mostly asymptomatic pigs, which are recognized as the main reservoir of this bioserotype (Kapperud, 1991). The bacteria mainly colonize tonsils, pharynx and intestinal tract of the pigs without inducing disease (Nesbakken et al., 2003). The prevalence of this bioserotype in pig slaughterhouses has been reported to be higher than 50% in Finland (Korte et al., 2004) and in southern Germany (Fredriksson-Ahomaa et al., 2001). Moreover, serotype O:3 is predominant among isolates recovered from slaughter pigs in the USA (Bhaduri et al., 2005; Tadesse et al., 2013) and in Brazil (Paixao et al., 2013a).

Consumption of contaminated pork products has been often suspected to be the main source of YE bioserotype 4/O:3 infections, but the connection between human and animal isolates is not well understood. Contamination of meat with YE probably occurs during slaughtering and processing of carcasses (Laukkanen et al., 2009). Bioserotype 4/O:3 strains have been detected from raw pork products, such as pig tongues and minced pork, at retail level (Bucher et al., 2008; Fredriksson-Ahomaa et al., 1999; Messelhausser et al., 2011). Moreover, YE bioserotype 4/O:3 is able to grow in modified atmosphere-packed pig cheek meat at 6 °C (Fredriksson-Ahomaa et al., 2012).

Several DNA fingerprinting approaches have been used to study genotypic diversity among bioserotype 4/O:3 strains, indicating a general limited genetic diversity and a high similarity between human and pig isolates (Fredriksson-Ahomaa et al., 2006). In particular, the pulsotypes obtained from PFGE analyses on YE bioserotype 4/O:3 strains have shown high homogeneity (Asplund et al., 1998; Iteman et al., 1996; Najdenski et al., 1994; Saken et al., 1994). PFGE characterization of bioserotype 4/O:3 strains recovered from tonsils of Bavarian slaughter pigs revealed the presence of one genotype common to most of the farms (Fredriksson-Ahomaa et al., 2010). However, PFGE genotypes differed from German and Finnish strains, indicating a different geographical distribution (Fredriksson-Ahomaa et al., 2003). AFLP, ERIC-PCR, and PFGE have been recently applied to characterize YE bioserotype 4/O:3 isolates from Brazilian slaughterhouses, showing more than 80% identity among the 61 tested isolates (Paixao et al., 2013a). AFLP presented higher discriminatory power than ERIC-PCR, and PFGE was the best technique for genetic differentiation of bioserotype 4/O:3 strains, even though all the isolates presented identity higher than 91%. In all, the isolates tended to be grouped according to the slaughterhouse of origin. In Chinese abattoirs, YE pathogenic isolates from pigs mainly belonged to bioserotype 3/O:3 and could be differentiated into 49 PFGE patterns (Liang et al., 2012). Human and pig isolates from the same geographic area gave indistinguishable patterns, suggesting that swine are the source of YE human infections. The same conclusion has been recently drawn by a study from Italy, where 22 YE bioserotype 4/O:3 pig isolates showed a moderate degree of genetic diversity by PFGE and clustered together with 4 human isolates (Bonardi et al., 2014). More comprehensive studies on the genetics of human and porcine YE bioserotype 4/O:3 strains are therefore necessary to confirm these preliminary evidences that swine are the main reservoir of human pathogenic bioserotype 4/O:3 strains.

1.5. *Yersinia enterocolitica* genomes

During the last decade, several *Y. enterocolitica* genomes have been sequenced and published. To date, three complete genome sequences from human isolates are available: YE strain 8081, belonging to bioserotype 1B/O:8 (Thomson et al., 2006); YE strain Y11, of bioserotype 4/O:3 (Batzilla et al., 2011c); and YE strain 105.5R(r), of bioserotype 3/O:9 (Wang et al., 2011). Moreover, numerous whole genome shotgun sequences have been obtained from clinical, environmental and animal strains of different bioserotypes, including weakly-virulent serotypes O:5,27, O:3 and O:9 and non-virulent serotypes O:5 and O:36 (Batzilla et al., 2011a; Batzilla et al., 2011b; Fuchs et al., 2011; Klinzing et al., 2012; Reuter et al., 2012; Savin et al., 2013). Comparative genome analyses have mainly been performed in order to identify intra-species regions of difference among YE strains, using the highly-virulent strain 8081 as reference.

The single circular chromosome of YE has a size of about 4.5-4.8 Mb, with a G+C content of 47-48% and 7 rRNA operons. Besides harboring dissimilar virulence determinants (as described in section 1.2.3 and in Table 3), other metabolism-associated genetic clusters are differentially present in the sequenced genomes. An ATP binding cassette transporter system, encoded by the Aat cluster, has been identified in the weakly-virulent YE serotypes O:3 and O:9 strains, while it is absent from the highly-virulent strain 8081 (Batzilla et al., 2011a; Wang et al., 2011). Even though the functionality of this cluster has not been characterized in YE, a possible role in pathogenesis cannot be excluded, since in enteroaggregative *E. coli* this system transports dispersin protein out of the bacterial cells, therefore participating in virulence (Nishi et al., 2003). The *agaVWEF* operon, which encodes the enzymes necessary for utilization of the intestinal mucin component N-acetyl-galactosamine, is absent from YEE (Batzilla et al., 2011a). Since the mucus of the pig's small intestine is mainly composed of N-acetyl-galactosamine (Mantle et al., 1980), the *aga* cluster may serve as fitness factor in YEP, especially for the pig-adapted bioserotype 4/O:3 strains. Multiple serotype- or strain-specific bacteriocins and toxin-antitoxin systems have been also detected in YE genomes, conferring the bacteria further advantages for efficient colonization of their ecological niches.

The most relevant differences in YE genomes are mobile elements, specifically genomic islands and prophages. The *Yersinia* genomic island YGI-1 is restricted to the virulent biotypes 1B and 2-5; YGI-2, highly conserved in various members of *Enterobacteriaceae*, has been only detected in biotypes 1A and 1B; YGI-3, a putative integrated plasmid, is unique to strain 8081; lastly, the integrated plasmid YGI-4 is variably present in biotype 1B only (Thomson et al., 2006). On the other side, YE bioserotype 4/O:3 strains harbor the novel GIYep-01 island, which has been proved to be actively mobile through its functional integrase in strain Y11 (Batzilla et al., 2011a). Most of the YE novel genomic regions are composed of prophages and prophage-like elements. None of the four prophages identified in strain 8081 are conserved among YEP genomes. In particular, ΦYE250 is strain-specific, while ΦYE98, ΦYE185 and ΦYE200 are partially detected in various 1B isolates (Thomson et al., 2006). Prophage regions have been also specifically acquired by YE strain Y11. PhiYep-1 has been found to be highly similar to prophages CUS-1 of *E. coli* and Ypf of *Y. pestis*, but it is not complete in other bioserotype 4/O:3 strains (Batzilla et al., 2011a). Two P2-like prophages, PhiYep-2 and PhiYep-3, are similar to a prophage region in *Y. pseudotuberculosis*. Interestingly, it has been demonstrated that PhiYep-3 is able to precisely excise from its attachment site in strain Y11; however, it is not conserved among bioserotype 4/O:3 isolates, suggesting a recent acquisition and possible use as epidemiological marker (Batzilla et al., 2011a). Similarly, five specific prophages have been identified in YE strain

105.5R(r) and suspected to be partly responsible for the structural rearrangements and low synteny with strain 8081 (Wang et al., 2011).

The pYV virulence plasmid, carried by virulent YE strains of biotypes 1B and 2-5, is 67.7-72.5 kb in size and has a G+C content of 43-44%. The sequences of pYV_{8081} (bioserotype 1B/O:8) and pYV_{Y11} (4/O:3) demonstrated higher divergence than the corresponding chromosomes, while pYV_{Y11} and pYV_{W22703} (2/O:9) showed closely-related sequences (Batzilla et al., 2011a). An arsenic resistance *arsABRH* gene cluster has been identified in pYV_{Y11} and pYV_{W22703}, with important differences in the sequences of *arsB* and *arsR* genes, but not in pYV_{8081}; moreover, mutations have been discovered in the *yadA* gene of pYV_{Y11} and pYV_{W22703}, causing structural changes in YadA of YE strain Y11 (Batzilla et al., 2011a). Comparative analyses among three pYV sequences of YE bioserotype 1B/O:8 strains (8081, WA-314 and A127/90) have shown general similarity with evident differences in the amino acid sequences of YadA and some T3SS proteins, such as SycH, YopM, LcrV and YscP (Foultier and Cornelis, 2003; Oberhettinger et al., 2011; Snellings et al., 2001).

Overall, the sequenced genomes of *Y. enterocolitica* reinforce the current differentiation into two subspecies and demonstrate both similarities and specificities, reflecting the evolutionary history of the different groups.

2. Aims of the study

The foodborne pathogen *Y. enterocolitica* represents a noticeably versatile group of organisms which cause intestinal and extraintestinal diseases, differ in prevalence around the world, and survive in a wide range of natural and host environments. Genomics and molecular pathogenesis of yersiniosis have been thoroughly studied and the strategies used by *Yersinia* to overcome host defenses are well understood, especially for highly-virulent strains of bioserotype 1B/O:8. However, the majority of YE isolates from patients belong to the weakly-virulent serotypes O:3, O:9 and O:5,27, which are frequently detected also in animals and food, and lead to diseases with divergent severity among the hosts. In particular, serotype O:3 strains are responsible for 80-90% of yersiniosis cases worldwide, and are repeatedly isolated from pigs, considered as the asymptomatic reservoir of these strains. Evidences of a direct connection between animal and human strains are not satisfactory, even if consumption of contaminated pork has been suspected to be the main source of clinical cases caused by YE serotype O:3. Molecular epidemiology methods have proven to be inadequate in terms of discriminatory power for studying the transmission of this group of organisms, preventing efficient control strategies and identification of sources of infection. Moreover, colonization and virulence-associated factors of serotype O:3 have not been exhaustively characterized, mainly because of its low virulence and the inadequacy of the mouse model.

Whole-genome sequencing has recently demonstrated high impact on all life sciences, providing thorough insights into the complete genetic capacity of an organism, its physiology, virulence behavior, adaptation to a preferred niche, and evolution. Moreover, genomics is prerequisite for the application of further omics technologies. This study, in fact, aims at investigating the pathogenesis and evolutionary aspects of YE with a genome-based approach. To appreciate the virulence mechanisms of the heterogeneous YE species it is necessary to concentrate on more than a single serotype or strain. The findings obtained by comparing highly-virulent strains should be used to identify the specific disease-associated abilities and genomic features of the weakly-virulent and non-virulent groups, thanks to the availability of multiple sequences. This should then allow analysis of the evolution of the whole YE species, and identification of genetic markers for a molecular diagnostic test. Focusing on the intra-species microevolution of serotype O:3 should provide a better understanding about the population structure, and thus epidemiology, of these homogeneous strains. The mechanisms putatively regulating the *in vivo* infection in humans of serotype O:3 strains should also be studied, bringing new insight into the low-virulence and adaptation properties of this group.

3. Materials and Methods

3.1. Instruments and consumables

Standard laboratory devices were used in this study (Table 6). If special instruments, software or methods were necessary, they will be explained in the further sections when appropriate.

Table 6 List of instruments used.

Instrument	Model	Supplier
Agarose gel documentation	GelDoc EQ	Bio-Rad, Munich
Centrifuge	3-30KS	Sigma GmbH, Osterode
Centrifuge	1K15	Sigma GmbH, Osterode
Centrifuge	1-15	Sigma GmbH, Osterode
DNA workstation	L020-GC	Kisker Biotech GmbH, Steinfurt
Electrophoresis chambers DNA	MGU-402T-FL	C.B.S. Scientific, Del Mar, CA, US
Electrophoresis chambers SDS-PAGE	Mini PROTEAN-3	Bio-Rad, Munich
Electrophoresis power supply	PowerPac 300	Bio-Rad, Munich
Electroporation device	MicroPulser™	Bio-Rad, Munich
Film development device	Fujifilm FPM-100A	Fuji, Düsseldorf
Hybridization oven	PersonalHyb	Stratagene, La Jolla, CA, US
Incubator	IPP400	Memmert, Schwabach
Incubator	INP-500	Memmert, Schwabach
Incubator	Type B20	Heraeus, Hanau
Incubator shaker	Excella E24	New Brunswick GmbH, Nürtingen
Laminar flow hood	Herasafe HS12	Heraeus, Hanau
Mixer	Thermomixer compact	Eppendorf, Hamburg
Mixer	Thermomixer 5436	Eppendorf, Hamburg
PCR cycler	Veriti 96-w	Life Technologies GmbH, Darmstadt
PCR cycler	2720	Life Technologies GmbH, Darmstadt
pH meter	MP220	Mettler Toledo, Giessen
Pipettes	Labmate	Abimed, Langenfeld
Pipettes	2.5, 10, 100 µl	Eppendorf, Hamburg
Semi-dry blot device	Trans-Blot SD	Bio-Rad, Munich
Spectrophotometer	Ultrospec 3100 pro	Amersham Biosciences, Freiburg
Tissue homogenizer	MM2000	Retsch GmbH, Haan
Ultrasonic processor	UP 50H	Hilscher Ultrasonic GmbH, Teltow

For consumables, standard products from Becton Dickinson (Heidelberg), Biozym Scientific GmbH (Hessisch Oldendorf), Brand GmbH (Wertheim), Braun (Melsungen), Corning (Wiesbaden), Eppendorf (Hamburg), Kisker Biotech GmbH (Steinfurt), Merck Millipore (Cork, IE), Nunc (Roskilde, DK), and PEQLAB Biotech GmbH (Erlangen) were used.

3.2. Chemicals, kits and reagents

Routine tasks were performed with commercially available kits, chemicals and reagents purchased from Amersham Biosciences (Freiburg), AppliChem GmbH (Darmstadt), Bio-Rad (Munich), Life Technologies GmbH (Darmstadt), Macherey-Nagel (Düren), Merck (Darmstadt), Oxoid (Wesel), Roche Diagnostics (Mannheim), Roth (Karlsruhe), Sigma-Aldrich (Taufkirchen) and VWR (Darmstadt). Restriction enzymes, T4 DNA ligase and alkaline phosphatase were purchased from ThermoScientific (St.Leon-Rot) and used according to the protocols provided by the manufacturer. Relevant materials are listed in Table 7.

Table 7 Commercial chemicals, kits and reagents used in this study.

Chemical, reagent and kit	Supplier
Anti Human-HRP polyclonal antibody	Roth, Karlsruhe
DIG Nucleic Acid Detection kit	Roche Diagnostics, Mannheim
ECL Western blotting detection reagents	GE Healthcare, Freiburg
dNTP Mix, 10 mM each	ThermoScientific, St.Leon-Rot
GeneRuler 1 kb DNA Ladder	ThermoScientific, St.Leon-Rot
Nitrocellulose membrane	Roth, Karlsruhe
NucleoBond PC 100 kit	Macherey-Nagel, Düren
NucleoSpin Plasmid kit	Macherey-Nagel, Düren
NucleoSpin Tissue kit	Macherey-Nagel, Düren
Nylon membranes	Roche Diagnostics, Mannheim
PCR DIG Probe Synthesis kit	Roche Diagnostics, Mannheim
peqGOLD Gel Extraction kit	PEQLAB Biotech GmbH, Erlangen
Phusion High-Fidelity DNA Polymerase	ThermoScientific, St.Leon-Rot
Protease Inhibitor Cocktail	Sigma-Aldrich, Taufkirchen
recomLine Yersinia IgG 2.0	Mikrogen Diagnostik, Neuried
Spectra™ Multicolor Broad Range Protein Ladder	ThermoScientific, St.Leon-Rot
Taq DNA Polymerase Master Mix 2.0	VWR, Darmstadt

3.3. Microbiological methods

3.3.1. Bacterial strains and plasmids

In Table 8 and Table 9 bacterial strains and plasmids used in this study are listed. The additional *Y. enterocolitica* strains tested for validation of the patho-serotyping PCR are listed in Table 10.

Table 8 Bacterial strains used in this study.
(FBI-Zoo: Food-Borne Zoonotic Infections of Humans, BMBF project; FLI: Friedrich-Löffler Institute; MvPI: Max-von-Pettenkofer Institute)

***Y. enterocolitica* strains**		
Bacterial strain ID	**Features**	**Reference**
8081	Bioserotype 1B/O:8, American human isolate	(Portnoy et al., 1981)
WA-314 (ATCC 51871)	Bioserotype 1B/O:8, clinical isolate	(Heesemann and Laufs, 1983)
Y11 (DSMZ 13030)	Bioserotype 4/O:3, human isolate	(Neubauer et al., 2000a)
W22703	Bioserotype 2/O:9, mutant from strain W227	(Cornelis and Colson, 1975)
Y5.27P	Bioserotype 3/O:5,27	MvPI collection
NF-O	Bioserotype 1A/O:5, clinical isolate	(Ratnam et al., 1982)
IP2222	Bioserotype 1A/O:36, Japanese isolate from water	(Grant et al., 1999)
Y8265/556	Bioserotype 4/O:3, French human isolate	MvPI collection
Y5307/647	Bioserotype 4/O:3, Finnish human isolate	MvPI collection
P1 (BR6319)	Bioserotype O:3, German isolate from pig	Prof. Neubauer, FLI Jena
P4 (BR0055)	Bioserotype O:3, German isolate from pig	Prof. Neubauer, FLI Jena
149 (CN-Y641)	Bioserotype 4/O:3, Belgian isolate from puppy	MvPI collection
150 (CN-Y486)	Bioserotype 4/O:3, Belgian isolate form calf	MvPI collection
D1 (M1-S3405)	Bioserotype 4/O:3, German human isolate	MvPI collection
D2 (M2-S3547)	Bioserotype 4/O:3, German human isolate, obtained 45 days after D1 from same patient	MvPI collection
D3 (CN-35)	Bioserotype 4/O:3, German human isolate	MvPI collection
D4 (CN-142)	Bioserotype 4/O:3, Belgian isolate from pig	MvPI collection
D5 (CN-146)	Bioserotype 4/O:3, Danish isolate from pig	MvPI collection
D6 (CN-243/Y745)	Bioserotype 3/O:3, Japanese isolate	MvPI collection
D7 (P8/FBI-00033)	Bioserotype 4/O:3, German isolate from pig	FBI-Zoo collection
D8 (P14/FBI-03871)	Bioserotype 4/O:3, German isolate from pig	FBI-Zoo collection
D9 (P15/04247)	Bioserotype 4/O:3, German isolate from pig	FBI-Zoo collection

Other strains		
Species	**Features**	**Reference**
E. coli T7 express (BL21 derivative, no λ-prophage)	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10*--TetS*)2 [dcm] R(zgb-210::Tn10*--TetS*) endA1* D*(mcrCmrr)114::IS10*	New England Biolabs, Frankfurt
Y. frederiksenii CN-113	Strain H56-36/81, serotype O:60	MvPI collection
Y. intermedia CN-114	Strain H9-36/83, serotype O:17	MvPI collection
Y. kristensenii CN-115	Strain H17-36/83, serotype O:12.25	MvPI collection
Y. rohdei CN-116	Strain H274-36/85, serotype O:76	MvPI collection
Y. bercovieri CN-117	Strain H632-36/85, serotype O:16	MvPI collection
Y. mollaretii CN-118	Strain H279-36/86, serotype O:59	MvPI collection
Y. pseudotuberculosis	Strain IP32953, serotype O:1	(Chain et al., 2004)
Y. pestis CN-691	Strain KIM, biotype Medievalis	(Fetherston and Perry, 1994)
E. coli	Strain UTI89, serotype O18:K1:H7	Prof. Stecher, MvPI
E. coli	Strain CFT073, serotype O6:H1	Prof. Stecher, MvPI
S. enterica subsp.I	Strain C5, serovar Typhimurium	Prof. Stecher, MvPI
S. enterica subsp.I	Strain ATCC 14028, serovar Typhimurium	Prof. Stecher, MvPI
C. jejuni	Strain L11, serotype Lior 11	MvPI collection
C. jejuni	Strain G23	MvPI collection
L. pneumophila	Strain 2020	Prof. Hilbi, MvPI

Table 9 List of plasmids.

Plasmid	**Features**	**Reference**
pET-30a	Expression vector, T7 promoter, N-ter and C-ter His·Tag, Kan^R	Novagen, Darmstadt
pET-30b	Expression vector, T7 promoter, N-ter and C-ter His·Tag, Kan^R	Novagen, Darmstadt
pET-30c	Expression vector, T7 promoter, N-ter and C-ter His·Tag, Kan^R	Novagen, Darmstadt

Table 10 *Y. enterocolitica* strains used for validation of the patho-serotyping PCR-based method.
When necessary, re-serotyping of these strains was performed by Dr. Bader (MvPI, Munich).
(CVUA: Chemisches und Veterinäruntersuchungsamt Stuttgart; FBI-Zoo: Food-Borne Zoonotic Infections of Humans, BMBF project; FLI: Friedrich-Löffler-Institute; MvPI: Max-von-Pettenkofer Institute; RKI: Robert Koch Institute, Wernigerode)

Strain ID	Features	Source
CN-32	Serotype 0:3, strain 108P	MvPI collection
CN-36	Bioserotype 1A/O:5, human isolate	MvPI collection
CN-38	Bioserotype 1A/O:4,33, isolate from milk	MvPI collection
CN-39	Bioserotype 1A/O:7,8, isolate from milk	MvPI collection
CN-40	Bioserotype 1A/O:6.30, isolate from milk	MvPI collection
CN-41	Bioserotype 1A/O:10,K1, isolate from milk	MvPI collection
CN-42	Bioserotype 2/O:5,27	MvPI collection
CN-44	Bioserotype 4/O:9, human isolate	MvPI collection
CN-45	Bioserotype 4/O:5,27, human isolate	MvPI collection
CN-47	Bioserotype 3/O:5,27, human isolate	MvPI collection
CN-49	Bioserotype 2/O:5,27, human isolate	MvPI collection
CN-52	Bioserotype 1A/O:48	MvPI collection
CN-53	Bioserotype 1A/O:41,43	MvPI collection
CN-54	Bioserotype 1A/O:5	MvPI collection
CN-55	Bioserotype 1B/O:8, strain CDC 544	MvPI collection
CN-56	Bioserotype 1B/O:8, strain CDC 540/84	MvPI collection
CN-57	Bioserotype 1B/O:8, strain E 661 Montana	MvPI collection
CN-58	Bioserotype 1B/O:13, strain USA 329/82	MvPI collection
CN-59	Bioserotype 1B/O:20, strain CDC 3039/85	MvPI collection
CN-60	Bioserotype 1B/O:21, strain CDC 209/84	MvPI collection
CN-101	Bioserotype 1A/O:36	MvPI collection
CN-170	Serotype O:3, strain 59/03	MvPI collection
CN-233	Bioserotype 1B/O:8, strain IP636	MvPI collection
CN-235	Bioserotype 2/O:9, strain IP22393	MvPI collection
CN-237	Bioserotype 2-3/O:5,27, strain IP199	MvPI collection
CN-238	Bioserotype 2-3/O:5,27, strain IP885	MvPI collection
CN-239	Bioserotype 2-3/O:5,27, strain IP1607	MvPI collection
CN-240	Bioserotype 2-3/O:5,27, strain IP22460	MvPI collection
CN-242	Bioserotype 4/O:3, strain IP24231	MvPI collection
CN-244	Bioserotype 4/O:3, strain IP24231	MvPI collection
CN-245	Bioserotype 4/O:3, strain IP21981	MvPI collection

CN-249	Bioserotype 4/O:3, strain IP23357	MvPI collection
CN-252	Bioserotype 4/O:3, strain IP7032	MvPI collection
CN-254	Bioserotype 4/O:3, strain IP3692	MvPI collection
CN-256	Bioserotype 4/O:3, strain IP4115	MvPI collection
CN-262	Bioserotype 4/O:3, strain IP22274	MvPI collection
CN-266	Bioserotype 4/O:3, strain IP23227	MvPI collection
CN-272	Bioserotype 1A/O:5, strain IP124	MvPI collection
CN-584	Bioserotype 1A/O:5, strain NF-O	MvPI collection
260	Not typeable, strain FBI-4465	Klinikum der Universität München
267	Serotype O:3, strain FBI-4848	Klinikum der Universität München
282	Not typeable, strain FBI-5620	Klinikum der Universität München
292	Serotype O:27, strain FBI-6584	Klinikum der Universität München
343	Serotype O:9, strain FBI-5634	Klinikum der Universität München
377	Serotype O:3, strain FBI-6821	Klinikum der Universität München
H-82	Strain Ty 4726/85	MvPI collection
H-89	Strain Ty 448/87	MvPI collection
H-93	Strain Ty 822/88	MvPI collection
H-395	Serotype O:5,27	MvPI collection
H-522	Serotype O:1,2a,3, strain H 523/88	MvPI collection
H-523	Serotype O:2a,2b,3, strain H 518/88	MvPI collection
H-534	Bioserotype 2/O:5,27, strain D113	MvPI collection
H-556	Bioserotype 4/O:3, strain 8265	MvPI collection
H-568	Bioserotype 2/O:5,27, strain YE873	MvPI collection
H-647	Bioserotype 4/O:3, strain 5307	MvPI collection
H-678	Bioserotype 2/O:5,27, strain SZ 3801/03	MvPI collection
H-679	Bioserotype 2/O:5,27, strain SZ 81/04	MvPI collection
H-744	Strain S11136	MvPI collection
H-745	Strain S11149	MvPI collection
H-746	Serotype O:3, strain S11925	MvPI collection
H-747	Serotype O:3, strain S12723	MvPI collection
H-748	Serotype O:3, strain S12891	MvPI collection
H-749	Serotype O:3, strain S12982	MvPI collection
H-750	Serotype O:3, strain S13559	MvPI collection
P2	Bioserotype 4/O:3, strain 0350, from pig	Dr. Rau, CVUA, Stuttgart
P3	Bioserotype 4/O:3, strain 5819, from pig	Dr. Rau, CVUA, Stuttgart
P5	Serotype O:3, strain BR0056, from pig	Prof. Neubauer, FLI, Jena
P6	Serotype O:3, strain BR0057, from pig	Prof. Neubauer, FLI, Jena

P7	Serotype O:3, strain BR0058, from pig	Prof. Neubauer, FLI, Jena
P9	Bioserotype 4/O:3, strain FBI-03524, from pig	FBI-Zoo collection
P10	Bioserotype 4/O:3, strain FBI-03540, from pig	FBI-Zoo collection
P11	Bioserotype 4/O:3, strain FBI-03585, from pig	FBI-Zoo collection
P12	Bioserotype 4/O:3, strain FBI-03855, from pig	FBI-Zoo collection
P13	Bioserotype 4/O:3, strain FBI-03870, from pig	FBI-Zoo collection
P16	Bioserotype 4/O:3, strain FBI-05198, from pig	FBI-Zoo collection
P17	Bioserotype 4/O:3, strain FBI-05680, from pig	FBI-Zoo collection
06-03972	Bioserotype 3/rough	Dr. Tietze, RKI
06-04957	Bioserotype 3/rough	Dr. Tietze, RKI
07-05684	Bioserotype 3/rough	Dr. Tietze, RKI
08-00920	Bioserotype 1A/rough	Dr. Tietze, RKI
08-02627	Bioserotype 1A/O:5	Dr. Tietze, RKI
09-05064	Bioserotype 4/O:3	Dr. Tietze, RKI
09-05074	Bioserotype 4/O:3	Dr. Tietze, RKI
09-05075	Bioserotype 4/O:3	Dr. Tietze, RKI
09-05076	Bioserotype 4/O:3	Dr. Tietze, RKI
09-05259	Bioserotype 3/rough	Dr. Tietze, RKI
09-08364	Bioserotype 3/rough	Dr. Tietze, RKI
10-00826	Bioserotype 1B/O:8	Dr. Tietze, RKI
10-00827	Bioserotype 1A/O:8	Dr. Tietze, RKI
11-01204	Bioserotype 3/O:5,27	Dr. Tietze, RKI
11-06758	Bioserotype 3/O:5,27	Dr. Tietze, RKI
11-07058	Bioserotype 1A/rough	Dr. Tietze, RKI
11-08263	Bioserotype 3/O:5,27	Dr. Tietze, RKI
11-08388	Bioserotype 3/O:5,27	Dr. Tietze, RKI
11-08389	Bioserotype 3/O:5,27	Dr. Tietze, RKI
12-00565	Bioserotype 3/O:9	Dr. Tietze, RKI
12-00586	Bioserotype 3/O:9	Dr. Tietze, RKI
12-00587	Bioserotype 3/O:9	Dr. Tietze, RKI
12-00821	Bioserotype 1A/rough	Dr. Tietze, RKI
12-01019	Bioserotype 3/O:9	Dr. Tietze, RKI
12-01336	Bioserotype 1A/O:8	Dr. Tietze, RKI
12-01511	Bioserotype 1A/O:5	Dr. Tietze, RKI
12-01512	Bioserotype 1A/O:8	Dr. Tietze, RKI
12-01527	Bioserotype 3/O:9	Dr. Tietze, RKI
12-01557	Bioserotype 1A/O:8 (re-serotyped as rough)	Dr. Tietze, RKI

12-01616	Bioserotype 1A/O:8	Dr. Tietze, RKI
12-01779	Bioserotype 1A/O:5	Dr. Tietze, RKI
12-01805	Bioserotype 3/O:9	Dr. Tietze, RKI
12-01910	Bioserotype 3/O:9	Dr. Tietze, RKI
12-02272	Bioserotype 1A/O:5	Dr. Tietze, RKI
12-02439	Bioserotype 1A/O:5	Dr. Tietze, RKI
12-02966	Bioserotype 1A/O:8	Dr. Tietze, RKI
12-03062	Bioserotype 1A/O:5	Dr. Tietze, RKI
12-03266	Bioserotype 1A/rough	Dr. Tietze, RKI
12-03276	Bioserotype 4/O:3	Dr. Tietze, RKI
12-03288	Bioserotype 1A/rough	Dr. Tietze, RKI
12-03324	Bioserotype 1A/rough	Dr. Tietze, RKI
12-03401	Bioserotype 1A/rough	Dr. Tietze, RKI
12-03475	Bioserotype 4/O:3	Dr. Tietze, RKI
12-03489	Bioserotype 4/O:3	Dr. Tietze, RKI
12-03561	Bioserotype 4/O:3	Dr. Tietze, RKI
12-03562	Bioserotype 4/O:3	Dr. Tietze, RKI
12-03803	Bioserotype 3/rough	Dr. Tietze, RKI
12-04470	Bioserotype 1A/rough	Dr. Tietze, RKI
12-05220	Bioserotype 1B/O:8	Dr. Tietze, RKI
13-00330	Bioserotype 4/O:5,27	Dr. Tietze, RKI

3.3.2. Culture media

Liquid media were prepared by standard formulation (Table 11) in batches and sterilized by autoclaving at 121 °C, 1 bar, 20 min. Similarly, solid media were made with the addition of 16 g/l agar to the respective media, unless otherwise stated.

Table 11 Liquid media for microbiological cultures.

Media	**Composition**
Lysogeny Broth (LB)	1% tryptone, 0.5% yeast extract, 0.5% NaCl, pH 7.4
Brain Heart Infusion medium (BHI)	37 g/l BHI powder
Müller Hinton medium (MH)	21 g/l MH powder
SOC medium	2% tryptone, 0.5% yeast extract, 0.05% NaCl, 250 mM KCl, 20 mM Glucose
Cryopreservation medium	LB medium with 20% glycerol

3.3.3. Bacterial growth and conservation

Bacteria were routinely cultured under aerobic conditions in LB medium at 27 °C or 37 °C with 180-200 rpm shaking. Cultivation of *Y. enterocolitica* at 37 °C in BHI medium was carried out to induce Yop secretion. MH liquid or solid medium was used for antimicrobial susceptibility test and growth curve determination, at either 27 °C or 37 °C, as explained below. The optical density at 600 nm (OD_{600}) of 900 µl bacterial suspension was determined by spectrophotometry. Bacterial pellets from overnight (o/n) culture were stored at -80 °C in cryopreservation medium (Table 11) and subsequently recovered streaking out frozen bacteria onto Columbia Blood agar, *Yersinia* selective agar (CIN) or LB agar plates.

3.3.4. Determining bacterial growth curves

Bacteria were recovered from frozen stocks, inoculated in MH or LB broth and allowed to grow o/n at 27 °C. A fresh sub-culture was then prepared with OD_{600}=0.02 and bacteria were grown at 27 °C for 6 hours. Every 30 min, OD_{600} was measured, and serial dilutions in sterile 0.9% NaCl solution were prepared from 10^{-4} to 10^{-8} and plated on MH or LB agar plates. After o/n growth (16-20 h), the colonies on the plates were manually counted and only plates with a colony number included between 50 and 400 were taken into account for calculating the colony forming unit (CFU)/ml: $\frac{\text{n° of colonies per plate} \times 10}{10^{-\text{n° of dilution}}}$.

3.3.5. Antibiotics and antimicrobial susceptibility test

When appropriate, selection of transformants was obtained by adding kanamycin to liquid media or solid agar at the concentration of 30 µg/ml.

To test antimicrobial susceptibility of *Y. enterocolitica*, the broth macro-dilution (a) and the disk diffusion (b) methods were used, with experiments performed in duplicate. In both techniques, o/n cultures from single colonies were freshly inoculated in MH medium at 1:50 dilution and grown at 27 °C until the logarithmic phase (OD_{600}=0.5-0.6) had been attained. (a) Serial two-fold antibiotic dilutions (from 256 µg/ml to 0.25 µg/ml) were prepared in sterile MH broth in 12-well plates. According to previously acquired data (section 3.3.4), bacterial suspension were inoculated at a final inoculum of 10^4 CFU/ml. Plates were incubated at 27 °C for 16-20 h with shaking and results determined as absorbance at OD_{600}. (b) Inocula of 10^4 CFU were prepared and spread onto MHA plates, and two antibiotic disks (Oxoid, Wesel) were then placed onto each plate. After incubation o/n at 27 °C and 37 °C, results were recorded as diameter of inhibition zone.

3.3.6. Electrocompetent cells and transformation

Preparation of electrocompetent cells. In order to make bacterial cells able to receive exogenous DNA, bacteria were grown o/n under standard culture conditions and inoculated at a 1:50 dilution in fresh LB media the following day. The culture was then allowed to grow until bacteria reached the logarithmic phase (OD_{600}=0.5-0.6) and placed immediately on ice for 15 min. Cells were then harvested at 5,000 x g for 15 min at 4 °C, washed twice in cold sterile water and once in cold 10% glycerol solution, and finally resuspended in the residual supernatant. Aliquots of 55 µl were used for electroporation or stored at -80 °C until further use.

Transformation by electroporation. For transformation of plasmid DNA into bacteria, electrocompetent cells from frozen stocks were thawed on ice, gently mixed with 1 µl of ligation mixture (10-30 ng) and transferred into 2 mm cuvettes. Electroporation was performed with a MicroPulser, with time=4 ms, potential=2.5 kV, resistance=200 Ω and capacitance=25 µF. Immediately after exposure to electrical field, bacteria were re-suspended in 1 ml SOC medium at room temperature (RT), incubated for 1 h at 37 °C, and subsequently plated on the appropriate selective LB agar plates for o/n growth.

3.3.7. Isolation of released proteins from *Y. enterocolitica*

To isolate secreted Yop proteins, *Y. enterocolitica* cells were grown o/n in BHI broth at 27 °C. The bacteria were then inoculated 1:20 in fresh BHI broth and incubated 2 h at 37 °C. Secretion of Yops was stimulated by adding 10 mM $MgCl_2$, 5 mM EGTA, and 0.2 % glucose, allowing further growth until OD_{600}=0.7-0.9. The culture was centrifuged at 7,500 rpm, 4 °C for 15 min. The supernatant was collected and mixed with TCA (trichloroacetic acid, final concentration 10% v/v) for protein precipitation. After 1 h on ice, proteins were pelleted at 14,000 x g, 4 °C, for 20 min, and washed three times with ice-cold acetone on ice for 20 min, with a centrifugation step at 12,000 x g, 4 °C for 10 min. A final wash in distilled water was performed before pelleting the proteins (12,000 x g, 4 °C for 10 min) and dissolving them in 30 µl SDS-PAGE sample buffer (section 3.5.1).

3.4. Molecular biology methods

3.4.1. Isolation of nucleic acids

Genomic DNA from bacterial suspensions was extracted using the commercial NucleoSpin Tissue kit, while plasmid DNA was purified by NucleoSpin Plasmid or NucleoBond PC 100 kits, following manufacturer's instructions. When pure DNA was not necessary, DNA was directly

isolated by the boiling methods using two different procedures. As standard method, a single colony, or bacterial pellet from 1 ml o/n culture (5 min centrifugation at 6,000 x g), were re-suspended in 60 µl distilled water and heated at 95-100 °C for 10 min. After centrifugation at 11,000 x g, 4 °C, 5 min, 50 µl of supernatant were transfer to a new tube and store at 4 °C till use. Alternatively, higher quality DNA could be obtained by inoculation of a single colony in 50 µl of 25 mM NaOH, heating at 100 °C for 10 min, neutralization of the suspension with 50 µl of 80 mM Tris/HCl buffer (pH 7.5), and centrifugation at 20,000 x g, 4 °C for 5 min. DNA concentration was measured spectrophotometrically at 260 nm absorbance (A_{260}) and the A_{260}/A_{280} ratio determined. An A_{260}/A_{280} ratio between 1.8 and 2.0 was considered indication of pure DNA.

3.4.2. Agarose gel electrophoresis

DNA was separated by electrophoresis in 0.7-2% agarose gels in TAE buffer (40 mM Tris-Acetate, 1mM EDTA, pH 7.6). Samples were mixed with loading buffer, when necessary, loaded onto the gel and electrophoretically separated at varying voltages and times, ranging between 70-110 V for 60-90 min, depending on the expected DNA size. Gels were stained in an ethidium bromide bath for approximately 10 min and the separated DNA was visualized by UV radiation using the GelDoc imaging system.

3.4.3. Oligonucleotides

All the primers used for PCR amplification in this study were obtained from Eurofins MWG Operon (Ebersberg) as salt-free lyophilized oligonucleotides (Table 12).

Table 12 List of oligonucleotides.

Primer	Sequence (5'-3')	Purpose
T7 promoter	TAATACGACTCACTATAGGG	Sequencing of pET-30 insert
T7 terminator	GCTAGTTATTGCTCAGCGG	Sequencing of pET-30 insert
pYVO3_6900_FW	TAGGTATGATAGGAGTTACATATCA	Specific region for pYV_O:3
pYVO3_7130_RV	CTTATTATTAATATCAATCGACGAT	Specific region for pYV_O:3
yopM_FW	TTCAGAAGGGCATTCAATATGTTTA	Amplification of *yopM*
yopM_RV	ACTCAAATACATCATCTTCAAGTTTG	Amplification of *yopM*
yscP_FW	TCTACAGCATCACAGGAACGCCTA	Amplification of *yscP*
yscP_RV	ATGGCTGAGCCGTTGACGTAGC	Amplification of *yscP*
16S_Yen_FW	AATACCGCATAACGTCTTCGGA	Identification multiplex
16S_Yen_RV	CTTCTTCTGCGAGTAACGTCAAT	Identification multiplex
inv_FW	TGGCATCAATCTCGTGATTTCG	Identification multiplex
inv_RV	GTTGCCCCTGAATATCTAAAGTGAC	Identification multiplex

ail_FW	TGTTAATGTGTACGCTGCGAGT	Identification multiplex
ail_RV	GTTTGGAGTATTCATATGAAGCGTC	Identification multiplex
ystB_FW	AACTTTTTGGACACCGCACAG	Identification multiplex
ystB_RV	GTCTGAGTATCGCACGCT	Identification multiplex
wbcA_O8_FW	TGATGAACGAGGCGAGTTTGTT	O:8 genotyping
wbcA_O8_RV	TACTCCGTCTGTTATGCGGATTTAG	O:8 genotyping
wbbU_O3_FW	ACCTCGTATTTTTGAAGATGATCGC	O:3 genotyping
wbbU_O3_RV	GTACTCAATAACTTGCTGTTCGGA	O:3 genotyping
per_O9_FW	TCCTTCTCCAAATATATAGGTGCCA	O:9 genotyping
per_O9_RV	ATGCGGCATTAGATGAGATGGA	O:9 genotyping
wzt_FW	GTTAGTTCCTGCATCTGATCGCC	O:5/O:5,27 genotyping
wzt_RV	ATCCAGCATCCATGGCTCC	O:5/O:5,27 genotyping
RM_O5,27_FW	TTCCAGCACACGTCGAACAAGTTC	O:5,27 genotyping
RM_O5,27_RV	AGGAAGATATCCAGTGCCGCT	O:5,27 genotyping
16S_FW_D1	CGATATCTCTAGAAGAGTTTGATCCTGGCTCAG	Universal 16S rRNA primer
16S_FW_D2	CGATATCTCTAGAAGAGTTTGATCATGGCTCAG	Universal 16S rRNA primer
16S_RV_P1	GATATCGGATCCACGGTTACCTTGTTACGACTT	Universal 16S rRNA primer

3.4.4. Polymerase chain reaction

Polymerase chain reaction (PCR) allows amplification of specific nucleic acids by polymerase enzymes, nucleoside triphosphates (dNTPs) and specific primers binding the DNA target. The three steps involved in one cycle of a PCR reaction are denaturation, annealing and extension.

In this study, most of the amplification experiments were accomplished with the VWR Red Taq DNA polymerase master mix, which is a ready-to-use reaction mix containing the polymerase, the dNTPs and a dye with a stabilizer to enable direct loading of PCR products onto agarose gels. The master mix contains 150 mM Tris-HCl pH 8.5, 40 mM $(NH_4)_2SO_4$, 4.0 mM $MgCl_2$, 0.2% Tween 20®; 0.4 mM dNTPs; 0.05 units/µl Amplicon Taq polymerase. The final reaction composition and the protocol used as standard are described in Table 13. When accurate amplification was required, the Phusion Hot Start High-Fidelity DNA polymerase (ThermoScientific) was used. This polymerase possesses a 3'->5' exonuclease proofreading activity, which is inactivated at RT by a proteic inhibitor. Reaction mix and amplification protocol are given in Table 14.

Table 13 PCR reaction and cycling conditions used with VWR Red Taq Master Mix.

Reagent	**Volume**	**Final concentration**
VWR Master Mix	25 µl	0.2 mM dNTPs, 2.0 mM $MgCl_2$, 1.25 U Taq
Forward primer	2 µl	0.4 µM
Reverse primer	2 µl	0.4 µM
Template DNA	1-5 µl	10-500 ng
Denatured water	up to 50 µl	
Step	**Temperature and Time**	**Number of Cycles**
Initial denaturation	95 °C, 5 min	1
Denaturation	95 °C, 30-40 sec	
Annealing	55-60 °C, 30-60 sec	30-40
Extension	72 °C, 1 min/kb template	
Final extension	72 °C, 8 min	1
Hold	8 °C, indefinite	1

Table 14 PCR reaction and cycling conditions used with Phusion High Fidelity DNA polymerase.

Reagent	**Volume**	**Final concentration**
5X Phusion HF buffer	10 µl	1X
2 mM dNTPs	5 µl	200 µM
Forward primer	2.5 µl	0.5 µM
Reverse primer	2.5 µl	0.5 µM
Template DNA	1-5 µl	10-500 ng
Phusion DNA polymerase	0.5 µl	0.02 U/µl
Denatured water	up to 50 µl	
Step	**Temperature and Time**	**Number of Cycles**
Initial denaturation	98 °C, 30 sec	1
Denaturation	98 °C, 10 sec	
Annealing	65-75 °C, 30 sec	25-35
Extension	72 °C, 30 sec/kb template	
Final extension	72 °C, 8 min	1
Hold	8 °C, indefinite	1

3.4.5. Patho-serotyping PCR

Primer design. Genome sequences of different *Y. enterocolitica* serotypes were used as references for specific primer design. For both multiplexes 1 and 2 (see below), primer pairs were manually designed in order to satisfy the following criteria: (a) primer length is 18–24 bp; (b) primers have 40 to 60% G+C content and similar annealing temperature (between 54.5 °C and 59.7 °C); (c) primer sequences in one reaction do not form self- or hetero-dimers; (d) primer sequences uniquely align to the target region; (e) the produced amplification products have unambiguously different sizes. Different annealing temperatures were tested with the designed primer pairs in single PCR reactions. Products showing the expected lengths and precise bands on agarose gels, without non-specific amplification, were obtained with an annealing temperature of 58 °C.

PCR assays. Two multiplex PCRs were developed for patho-serotyping of *Y. enterocolitica*. Multiplex 1, designed for species identification, was performed with primer pair inv_FW/RV (amplicon size: 1,009 bp), ail_FW/RV (431 bp), 16S_Yen_FW/RV (330 bp) and ystB_FW/RV (208 bp), while multiplex 2, aimed at molecular serotyping, with primers per_O9_FW/RV (837 bp), wbbU_O3_FW/RV (463 bp), wbcA_O8_FW/RV (269 bp) and wzt_FW/RV (662 bp). The 50 µl PCR mixtures were prepared as described in Table 13 with 100 ng of DNA template, and amplification was achieved by initial denaturation at 95 °C for 5 min, 30 cycles of denaturation at 95 °C for 40 sec, primer annealing at 58 °C for 40 sec, extension at 72 °C for 60 sec, and final extension at 72 °C for 8 min. The single PCR for discrimination between serotypes O:5 and O:5,27 was done as described above with primers RM_O:5,27_FW/RV, resulting in amplification of a 627 bp product only in serotype O:5,27 strains. The detection limit of the test was evaluated on five YE reference strains, conducting both multiplexes 1 and 2 on diluted DNA samples (1 ng/µl, 100 pg/µl, 10 pg/µl).

3.4.6. Synthesis of digoxigenin-labeled probes

Two digoxigenin (DIG)-dUTP-labeled probes were designed by selecting strain-specific regions in the 8081 and WA-314 genomes. The probe specific for detecting strain 8081 is 162 bp long and targets part of a putative hemolysin gene encoded by the YAPI island, whereas the 234-bp probe specific for strain WA-314 targets a region inside the strain-specific colicin cluster (section 4.1.1). Both probes were generated using the PCR DIG Probe Synthesis Kit (Roche), which produces PCR products with DIG-11-dUTP incorporated as a labeling molecule. Primers and amplification protocol are given in Table 15.

Table 15 Primer sequences and PCR conditions for synthesis of DIG-labeled probes.

Primer	Sequence (5'-3')	Probe	PCR conditions
Hem_8081_FW Hem_8081_RV	CAATATGACTACCGACCCGGTTAC GGATACATCTGCTGGGCGATATAC	Hem_8081 162 bp	- Denaturation 95 °C, 2 min - 30 cycles: denaturation 95 °C, 30 sec annealing 60 °C, 30 sec elongation 72 °C, 40 sec - Final elongation 72 °C, 7 min
Col_WA_FW Col_WA_RV	CGATCGTAGTAGTAAGGCAACTCC GACGGTATCATGCCCATAACTG	Col_WA 234 bp	

3.4.7. Protein expression library generation

Bacterial inserts. To create an expression library from *Y. enterocolitica* strain Y11, genomic DNA was purified from a saturated o/n liquid culture with the NucleoSpin Tissue kit, partially digested with the enzyme FastDigest Sau3AI (4 min at 37 °C) and separated by electrophoresis on a 0.7% agarose gel. DNA fragments of 0.5-3.0 kb were collected from the agarose gel and purified using a gel extraction kit.

Expression vectors. The pET-30a, pET-30b and pET-30c expression vectors were extracted from *E. coli* JM109 (stored at –80 °C at the Max von Pettenkofer-Institute strain collection) with a Midi kit for plasmid purification, digested with the FastDigest BamHI enzyme and loaded on a 0.7% agarose gel. The plasmid linear forms were purified by means of a gel extraction kit, dephosphorylated by alkaline phosphatase, and pooled together for further applications.

Ligation and transformation. The bacterial genomic DNA fragments were ligated into pET-30 expression vectors. Electrocompetent *E. coli* T7 express cells were used for electroporation of the ligation product and the transformants were selected on LB-kanamycin agar plates. After o/n incubation at 37 °C, colonies were scraped into LB medium containing 20% glycerol and frozen at -80 °C in 0.5 ml aliquots. The library was considered representative (containing at least one copy of all sequences) when the number of clones (N) satisfied the following formula: $N = \frac{\ln(1-P)}{\ln(1-f)}$, where P is the probability for a gene to be represented (0.99) and f is the fraction of the genome in one insert ($f = \frac{\text{insert length}}{\text{bp genome}}$).

3.5. Biochemical and analytical methods

3.5.1. SDS-PAGE

For separation of proteins based on their molecular weight, sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed in acrylamide gel under denaturing conditions (Laemmli, 1970). Gels were composed of two layers having different buffer contents, namely a 5% stacking gel and a 12% resolving gel. Protein samples were dissolved in the sample buffer and loaded into the gel. Electrophoresis was carried out in running buffer at 80 V for 20 min and 180 V for approximately 30-50 min, depending on the protein size. For protein visualization, gels were stained for 1 h in Coomassie solution and subsequently destained in water o/n. Proteins were transferred to nitrocellulose membranes by semi-dry electro-transfer at 15 V for 15 min, after equilibrating both SDS-PAGE gels and membranes in transfer buffer. All solutions are given in Table 16.

3.5.2. Western Blot

Nitrocellulose membranes were incubated o/n (or for 4 h) in blocking buffer (Table 16) at 4 °C and washed 2-3 times in PBST for 10 min, RT (Laemmli, 1970). Proteins were detected by immuno-staining with primary antibody diluted in PBST + 0.5% blocking buffer(1:5,000 for 1 h, RT), three washing steps for 10 min, incubation with horseradish peroxidase (HRP)-labelled anti-human secondary antibody (1:10,000, 40-45 min, RT), followed by three final washing steps. Chemo-luminescence reaction by the HRP was detected by ECL Western blotting detection reagents and visualized on x-ray films. When re-probing was necessary, the membranes were incubated 1 h at RT in 5-10 ml stripping buffer, blocked, washed in PBST and then subject to first and second antibodies as described above.

Table 16 Solutions and buffers used for SDS-PAGE and Western Blot experiments.

Solution	Composition
5X SDS-PAGE sample buffer	225 mM Tris/HCL (pH 6.8), 50% glycerol, 5% SDS, 0.05% bromophenol blue, 250 mM DTT
SDS-PAGE running buffer	25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3
Coomassie solution	0.15% Coomassie brilliant blue, 50% MeOH, 10% acetic acid
Transfer buffer (semi-dry)	48 mM Tris, 39 mM glycine, 20% methanol
Wash buffer (PBST)	PBS, 0.2% Tween20
Blocking buffer	5% milk powder, PBS, 0.2% Tween20
Stripping buffer	2% SDS, 100 mM NaOH, 50 mM DTT

3.5.3. Quantification of proteins

Proteins were quantified by the Bradford method, which is based on the binding of the Coomassie Brilliant Blue dye to protein samples and the comparison to a standard curve generated by the reaction of known amounts of a standard protein (Bradford, 1976). BSA (bovine serum albumin) was used for the calibration curve, with dilutions (0.2, 0.4, …, 1.8, 2.0 mg/ml) prepared in the same buffer where the protein samples were dissolved (e.g. lysis buffer). For each dilution, 20 µl were mixed with 1 ml of 1:5 diluted Protein Assay dye Reagent (Bio-Rad, Munich), OD_{595} was measured and the standard curve plotted. Dilutions of the protein samples were then prepared in duplicate. The protein concentration was finally determined by comparing the OD_{595} with the standard curve, taking into account the dilution factor.

3.5.4. Colony hybridization

To specifically detect colonies belonging to either *Y. enterocolitica* strain WA-314 or 8081, a DNA-DNA colony hybridization protocol based on DIG-labeled probes was established (Grunstein and Hogness, 1975). All solutions needed are listed in Table 17.

Preparation of membranes. Nylon membrane disks were placed for 5 min onto the surface of LB agar plates from o/n bacterial growth, transferred onto new plates and incubated at 27 °C for 4 h. Membranes were then placed colony-side up on filter papers soaked with 10% w/v sodium dodecyl sulphate (SDS) for 10 min. This procedure was repeatedly done with a denaturation solution for 15 min, a neutralization solution for 15 min, and 2X SSC for 10 min. After air-dry, membranes were baked at 80 °C for 60 min to cross-link the transferred DNA, and stored at 4 °C until use.

Hybridization and detection. Hybridization was carried out under high stringency conditions, calculating the temperature of hybridization as $T_{hyb} = 49.82 + 0.41\ GC\%_{probe} - \frac{600}{bp_{probe}}$, and the optimal temperature as $T_{opt} = T_{hyb} - 20/25$ °C. Since T_{hyb} resulted to be 70.16 °C for the probe Hem_8081 and 69.12 °C for the probe Col_WA, the T_{opt} for the whole experiment was fixed at 50 °C. The prepared nylon membranes were placed in glass bottles or 50 ml Falcon tubes and pre-hybridized at 50 °C in a hybridization oven for 1 h with 10 ml of hybridization buffer. Labeled DIG-probes (section 3.4.6), after denaturation at 97 °C for 5 min, were placed on ice, mixed with 5 ml of pre-warmed hybridization buffer at a concentration of 2 µl probe/ml, and transferred to sterile tubes with the appropriate membrane for the hybridization step at 50 °C for 3 h. Nylon membranes were then subject to stringent wash steps in two washing solutions with constant agitation: washing solution 1 for 2 x 5 min, RT; washing solution 2 for 2 x 15 min at 67 °C. The

DIG Nucleic Acid Detection Kit (Roche) was used to allow colorimetric detection (with NBT/BCIP) of the hybridized probes, according to manufacturer's instructions. To record the results, membranes were photographed, scanned and wet stored in plastic bags at 4 °C for any further stripping and re-hybridization. Hybridization buffers containing the labeled probes were stored at -20 °C and, when needed, re-used after denaturation at 65 °C.

Table 17 Solutions used in the colony hybridization experiment.

Solution	Composition	Use
Filter saturation solution	10% w/v SDS	Saturating the filter paper before transferring the colonies
Denaturation solution	0.5 M NaOH, 1.5 M NaCl, pH 11.5	Soaking the filter paper for denaturing the bacterial DNA
Neutralization solution	1.5 M NaCl, 1.0 M Tris-HCl, pH 7.4	Soaking the filter paper to neutralize the nylon membrane
20X SSC	3 M NaCl, 0.3 M Na3-citrate, pH 7.0	Soaking the filter paper to rinse the nylon membrane
Hybridization buffer	50% v/v formamide, 5X SSC, 1% v/v blocking reagent diluted in Buffer 1, 0.1% v/v N-lauroylsarcosine, 0.02% v/v SDS	Hybridization
N-lauroylsarcosine stock solution	10% w/v N-lauroylsarcosine	Stock solution
SDS stock solution	10% SDS	Stock solution
Washing solution 1	2X SSC, 0.1% SDS w/v	Washing the filters before blocking
Washing solution 2	0.5X SSC, 0.1% SDS w/v	Washing the filters before blocking
Maleic acid buffer	0.1 M maleic acid, 0.15 M NaCl, pH 7.5	Dilution of Blocking solution
Blocking solution 10X	10% Blocking reagent in maleic acid buffer	Stock solution
Blocking solution 1X	1:10 dilution of blocking solution 10X in maleic acid buffer (1% final concentration)	Blocking of unspecific binding sites
Antibody solution	Dilute anti-DIG-AP antibody 1:5000 (150 mU/ml) in Blocking solution	Binding to the DIG-labeled probe
Washing buffer	Buffer 1, 0.3% Tween 20, pH 7.5	Washing of membrane
Detection buffer	0.1 M Tris-HCl, 0.1 M NaCl, 0.05 M MgCl2, pH 9.5	Alkaline phosphatase buffer; for dilution of color substrate
Color substrate solution	80 µl NBT/BCIP stock solution in 4 ml detection buffer	Color reaction
TE-buffer	10 mM Tris-HCl, 1mM EDTA, pH 8.0	Stopping the color reaction
Stripping buffer	0.2 M NaOH, 0.1% SDS	Stripping membranes

Stripping reaction. To decolorize membranes after hybridization and colorimetric detection for further re-probing, a large beaker containing 150 ml of N,N-Dimethylformamide was heated to 60 °C. Membranes were placed into the hot solution and incubated for 30-60 min until the blue/violet color had been removed. They were then rinsed with ample distilled water and washed twice at 37 °C in stripping buffer for 20 min. After rinsing with 2X SSC for 5 min, membranes were stored wet in 10 ml 2X SSC at 4 °C until hybridization.

3.5.5. Serum selection and adsorption

Selection. Sera for the IVIAT (*in vivo* induced antigen technology) experiment were obtained from anonymous clinical human patients with confirmed *Y. enterocolitica* serotype O:3 infection in Poland and Germany (provided by Dr. Waldemar Rastawicki, Warsaw, and by Dr. Konrad Trülzsch, Munich, respectively). Immune responsiveness of each individual serum was tested with a line immunoassay (Mikrogen, Neuried). Serum samples lacking a significant immune response were excluded from further analyses, while the remaining positive samples were pooled together to a final volume of 3 ml (Table 18).

Adsorption. Sera were subject to adsorption with both *Y. enterocolitica* Y11 and *E. coli* T7 express whole cells and cell lysates, in order to remove antibodies which recognize antigens expressed *in vitro*. After growth at 27 °C in LB broth, bacterial cells from 600 ml fresh liquid culture (OD_{600}=0.5-0.6) were collected by centrifugation at 10,000 x g, 4 °C, 20 min, and washed in cold PBS. Pellets were weighted to add Protease Inhibitor Cocktail accordingly, and used directly for adsorption or lysed. For whole cell adsorption, the pellets were dissolved in PBS (1 ml for 100 ml culture). Bacterial suspension were centrifuged at 15,000 x g, 5 min, 4 °C, pellets were re-suspended in an equal volume of sera and incubated at 4 °C for 1 h, with continuous shaking. After centrifugation (15,000 x g, 15 min, 4 °C), unadsorbed antibodies were discarded from the supernatant, and adsorption was repeated three additional times. For adsorption with cell lysates, bacterial pellets from 100 ml culture were dissolved in 1 ml lysis buffer (50 mM Tris-HCl pH 8, 100 mM NaCl, 10% glycerol, 0.1% NP-40, 1mM PMSF, 0.1 mg/ml lysozyme), sonicated on ice three times for 20 sec, and centrifuged at 15,000 x g, 4 °C for 15 min. Supernatants (soluble proteins) were transferred to a new tube together with the pellets (insoluble proteins), which had been re-suspended in 1 ml lysis buffer. One aliquot was heated at 100 °C for 5 min to have heat-denatured cell lysates, and protein concentration of all cell lysates was determined (section 3.5.3). According to standard dot-blot protocols, nitrocellulose strips were saturated with 25 µg cell lysates, and incubated for 1 h at 4 °C with the sera previously adsorbed against whole cells, twice for cell lysates and once for heat-denatured cell

lysates. Aliquots of the final adsorbed sera were stored at -20 °C. Efficiency of the adsorption process was assessed by SDS-PAGE and Western Blot with unadsorbed, partially adsorbed and completely adsorbed sera.

Table 18 Sera from human patients and relative information (n.a.: not available).

ID	Date of collection	Serological and clinical info	Age/Gender of patient	Source	Pooled for IVIAT
S1	n.a.	Agglutination with O:3 anti-serum	n.a.	K. Trülzsch	Yes
S2	n.a.	Agglutination with O:3 anti-serum	n.a.	K. Trülzsch	Yes
S3	n.a.	Agglutination with O:3 anti-serum	n.a.	K. Trülzsch	No
S4	20.07.2011	Agglutination with O:3 anti-serum, 2 weeks after infection	n.a.	K. Trülzsch	No
S5	2003	Agglutination with O:3 anti-serum, 1 week after infection	n.a.	K. Trülzsch	Yes
S6	Feb 2012	Agglutination with O:3 anti-serum	21 years/n.a.	K. Trülzsch	No
S7	Feb 2012	Agglutination with O:3 anti-serum, 10 days after S6	21 years/n.a. (same as S6)	K. Trülzsch	Yes
S10	Jun 2009	High level of IgA, IgG, IgM to O:3 LPS and Yops	57 years/female	W. Rastawicki	No
S11	Oct 2009	High level of IgA, IgG, IgM to O:3 LPS and Yops	33 years/female	W. Rastawicki	No
S12	04.02.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	n.a./female	W. Rastawicki	Yes
S13	Feb 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	26 years/female	W. Rastawicki	No
S14	02.03.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	30 years/male	W. Rastawicki	Yes
S15	26.03.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	37 years/female	W. Rastawicki	Yes
S16	May 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	40 years/female	W. Rastawicki	No
S17	May 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	41 years/female	W. Rastawicki	No
S18	12.05.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	36 years/female	W. Rastawicki	Yes
S19	18.05.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	20 years/male	W. Rastawicki	Yes
S20	May 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	1 year/female	W. Rastawicki	No
S21	24.05.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	41 years/male	W. Rastawicki	Yes
S22	Jun 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	29 years/male	W. Rastawicki	No

S23	22.06.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	23 years/female	W. Rastawicki	Yes
S24	Jul 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	29 years/female	W. Rastawicki	No
S25	Aug 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	49 years/female	W. Rastawicki	No
S26	Sep 2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	41 years/male	W. Rastawicki	No
S27	25.10.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	51 years/male	W. Rastawicki	Yes
S28	04.11.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	51 years/female	W. Rastawicki	Yes
S29	16.11.2010	High level of IgA, IgG, IgM to O:3 LPS and Yops	12 years/male	W. Rastawicki	Yes

3.5.6. Screening of expression library

Expression library was screened by immunoblotting as follows. Aliquots of the *E. coli* T7 express library glycerol stock were thawed, diluted, and spread onto LB-kanamycin agar to obtain 300-500 colonies per plate. Master plates were incubated at 30 °C o/n before transfer onto nitrocellulose filter disks. Filter disks were placed colony side up on new LB agar containing 1 mM IPTG and 30 µg/ml kanamycin, and incubated at 37 °C for 4 h. Following induction, nitrocellulose filters were removed and washed in water-soaked filter papers. Colony immunoblotting was performed as described (section 3.5.2) using the pooled adsorbed sera as primary antibody. Clones identified during primary screening were picked from the master plate and subject to secondary and tertiary screening, similarly to the first screen. Clones that maintained reactivity were considered positive identifications. Antigens were identified by isolating the pET-30 plasmids from the positive clones, sequencing the inserts with T7 promoter and T7 terminator primers and performing a Blast search against *Y. enterocolitica* strain Y11.

3.6. Mouse experiments

3.6.1. Preparation of microorganisms

The mouse-virulent *Y. enterocolitica* strains 8081 and WA-314 used for animal experiments were obtained from Prof. Heesemann (Max-von-Pettenkofer Institute). Both strains had been mouse-passaged to select bacteria still maintaining virulence after extended manipulation in laboratory conditions, by challenging mice by intra-peritoneal (i.p.) injection and re-isolating the bacteria 1 day post-infection.

3.6.2. Infection of animals

In order to evaluate the *in vivo* virulence behavior of *Y. enterocolitica* strains 8081 and WA-314, single-strain and co-infection experiments in a mouse model were carried out. Mouse infection was performed with seven-week old C57/BL6 female mice, which were identified with ear tags. Groups of 8 mice were used for both single-strain- and co-infection experiments. For this purpose, bacteria were o/n grown at 27 °C as described in section 3.3.3, diluted 1:50 in LB medium and allowed to further grow at 37 °C for 80 min. Cultures were centrifuged at 4,500 x g for 10 min at 4 °C and the harvested bacteria were washed in sterile Dulbecco's PBS (DPBS). Using previously acquired data (section 3.3.4), an infection dose of 1.7 x 10^4 bacteria was prepared for single-strain infection, and 1.2 x 10^4 bacteria for the co-infection experiment. The number of CFU was confirmed by plating dilutions of the infection samples for single-strain mixes, counting colonies after 24 h growth at 27 °C, and by colony hybridization for the co-infection mix (section 3.5.4). Mice were subject to i.p. injection of 600 µl of the appropriate DPBS bacterial suspension and weighted every day.

3.6.3. Recovery of bacteria from mouse organs

Mice were sacrificed by inhalation of CO_2 five days post infection. Livers and spleens were immediately removed under sterile conditions and stored in 1 ml DPBS on ice. Organs were homogenized by a tissue blender for 4 min. Plating serial dilutions of the homogenized livers (10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}) and spleens (10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}) on LB agar plates, followed by incubation at 27 °C for 20-24 h, allowed determination of bacterial counts.

3.7. Bioinformatics

3.7.1. Bioinformatics tools

All the tools used in this study for bioinformatics analysis and their respective references/sources are listed in Table 19.

Table 19 List of used bioinformatics tools.

Software/ Platform	Description	Reference/Link
BioEdit	Biological sequence alignment editor	(Hall, 1999)
Blast	Basic Local Alignment Search Tool	(Altschul et al., 1990)
BRIG	Interactive generation of comparative genomic images	(Alikhan et al., 2011)
BWA	Mapping of short read using Burrows-Wheeler transform	(Li and Durbin, 2009)
CDD	Conserved Domain Database of the NCBI	(Marchler-Bauer et al., 2013)
CLC DNA Workbench	Comprehensive workbench for DNA, RNA, and protein analyses	CLC bio, Aarhus, Denmark
CLC Genomic Workbench	High-throughput sequencing data analysis	CLC bio, Aarhus, Denmark
ClustalW	Multiple sequence alignment	(Larkin et al., 2007)
Edgar	Comparative analysis of annotated genomes	(Blom et al., 2009)
FastQC	Quality control of reads from all sequencing platforms	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FreeBayes	Bayesian genetic variant detector	http://arxiv.org/abs/1207.3907
Galaxy	Web-based platform for data intensive biomedical research	(Goecks et al., 2010)
GATK	Post-processing and SNP calling	(DePristo et al., 2011)
Gene Ontology	Ontology of defined terms representing gene product properties	(Ashburner et al., 2000)
KAAS	Server for automatic genome annotation and pathway reconstruction	(Moriya et al., 2007)
Maple	Metabolic And Physiological potential Evaluator	(Takami et al., 2012)
Mauve	Generation of multiple genome alignments for comparative genomics	(Darling et al., 2010)
MLST	Web-based method for MLST of bacterial species based on WGS data	(Larsen et al., 2012)
MUSCLE	Multiple sequence alignment	(Edgar, 2004)
NCBI	Biomedical and genomic databases	www.ncbi.nlm.nih.gov
Pfam	Large collection of protein families	(Finn et al., 2006)

PGAAP	Prokaryotic Genome Automatic Annotation Pipeline	http://www.ncbi.nlm.nih.gov/genome/annotation_prok/
PHAST	Rapid and accurate identification, annotation and graphically displaying of prophages	(Zhou et al., 2011)
Phylip	Programs for inferring phylogenies	(Felsenstein, 1989)
Picard	Manipulation of SAM files	http://picard.sourceforge.net/
Prophage Finder	Prediction of prophage loci in prokaryotic genomes	(Bose and Barber, 2006)
PsortB	Bacterial localization prediction tool	(Yu et al., 2010)
PubMLST	Public databases for molecular typing and microbial genome diversity	(Jolley and Maiden, 2010)
RAST	Automated service for high-quality annotation of bacterial genomes	(Aziz et al., 2008)
SAMtools	Manipulation of SAM files, SNP calling	(Li et al., 2009)
SEED	Comparative genomic environment	(Overbeek et al., 2005)
SignalP	Predicts of signal peptide cleavage sites	(Petersen et al., 2011)
tRNAscan-SE	Web-server for tRNA gene search in genomic sequences	(Schattner et al., 2005)

3.7.2. General sequence analysis

Whole genome, gene and protein sequences were acquired from the online NCBI databases. *Y. enterocolitica* strains and accession numbers of their sequenced genomes are given in Table 20. Basic sequence manipulation and visualization were performed using BioEdit Sequence Alignment Editor, version 7.0.5.3, and CLC DNA Workbench, version 6.0. Homology searches in public databases were done using Blast, while identification of conserved protein domains was obtained by CDD and Pfam database search. MLST profiles of *Y. enterocolitica* strains were determined *in silico* using the MLST (MultiLocus Sequence Typing) service (version 1.7); the *Yersinia* MLST allele sequences, in particular, are based on the scheme hosted at the PubMLST database.

3.7.3. Sequencing of bacterial DNA and read quality control

The genomic sequence of *Y. enterocolitica* strain WA-314 was acquired by BGI-Hongkong Co. (Hong Kong), who performed high-throughput Illumina paired-end sequencing, quality control of raw data and assembly of the short reads with SOAPdenovo. The methods described in sections 3.7.4 and 3.7.5 were therefore not applied to this sample. All the YE serotype O:3 genomes generated in this study were obtained by re-sequencing with the Illumina MiSeq technology, with a 150-bp paired-end library (IMGM Laboratory, Martinsried) or a 250-bp paired-end library

(collaboration with Prof. Kalinowski, University of Bielefeld). Reads were subject to quality control using the program CLC Genomics Workbench versions 6 and 7 and the FastQC tool embedded in the Galaxy platform (Table 19). In particular, sequencing accuracy was assessed by measuring the base calling accuracy using the Phred quality score (Q score), which indicates the probability (P) that a given base is incorrect in a logarithmic scale ($Q = -10 \log_{10} P$). For example, a Phred score of 30 means that the chances that a base is called incorrectly are 1 in 1,000 with an accuracy of 99.9%, while a Phred score of 40 corresponds to a base call accuracy of 99.99%.

Table 20 *Y. enterocolitica* genome sequences used in this study.

Strain	Bioserotype	Genome Accession N°	Reference
8081	1B/O:8	NC_008800 and NC_008791 (pYV)	(Thomson et al., 2006)
WA-314	1B/O:8	AKKR00000000	This study (Garzetti et al., 2012)
a127/90	1B/O:8	NC_004564 (pYV)	(Foultier and Cornelis, 2003)
Y11	4/O:3	NC_017564 and NC_017565 (pYV)	(Batzilla et al., 2011c)
Y8265	4/O:3	CACU01000000	(Batzilla et al., 2011a)
Y5307	4/O:3	CACV01000000	(Batzilla et al., 2011a)
W22703	2/O:9	FR718488-FR718797	(Fuchs et al., 2011)
105.5R(r)	3/O:9	CP002246.1	(Wang et al., 2011)
Y5.27P	3/O:5,27	CACW00000000	(Batzilla et al., 2011a)
NF-O	1A/O:5	CACY00000000	(Batzilla et al., 2011b)
IP2222	1A/O:36	CACZ00000000	(Batzilla et al., 2011b)
IP10393	4/O:3	CAOV01000000	(Savin et al., 2013)
PhRBD_Ye1	n.a./O:3	AGQO00000000	(Klinzing et al., 2012)
YE12/03	4/O:3	HF933425	(Reuter et al., 2012)
O3-gb	4/O:3	n.a.	Provided by Dr. A. McNally, Nottingham

3.7.4. Read mapping and alignment post-processing

Reads from each isolate were mapped against the chromosome of *Y. enterocolitica* strain Y11 serotype O:3 reference genome using the BWA (Burrows-Wheeler Aligner) Galaxy wrapper with default parameters. Mapping quality reports were generated by CLC Genomics Workbench, providing coverage statistics and read distributions. Alignments were filtered using SAMtools (Galaxy wrapper for sam_bitwise_flag_filter), in order to remove unmapped and unpaired reads,

and Picard (Galaxy wrapper for MarkDuplicates), to reduce the data set by including only one copy of the duplicate sequences, which may arise from PCR amplification artifacts during library construction. The GATK tools embedded in galaxy were then used to perform a local re-alignment, to transform regions with mismatches around indels into clean reads. This step is recommended, since these misalignments are easily mistaken as SNPs. A base quality score recalibration is also performed with GATK tools, in order to make error rates more accurate and informative than the ones provided by the sequencers.

In case a *de novo* assembly of genomes was required, paired reads were assembled by the CLC Genomics Workbench, with the following settings: Minimum contig length = 200; Mismatch cost = 2; Insertion cost = 3; Deletion cost = 3; Length fraction = 0.5; Similarity fraction = 0.8.

3.7.5. Variant calling and filtering

Variants were identified with different approaches, depending on the type of genome sequence. For assembled genomes downloaded from public databases, variants were called using the web server snpTree version 1.1, with 10 bp as minimum distance between SNPs and 20 bp as minimum distance to end of sequence. For the genomes obtained in this study and mapped against a reference, variants were detected combining the results from two tools (FreeBayes and SAMtools). The FreeBayes algorithm was run in Galaxy with the following parameters: Minimum mapping quality = 30; Minimum base quality = 20; Minimum coverage = 10; Minimum variant probability = 0.8. Variants were identified by SAMtools with Minimum mapping quality = 30 and Minimum base quality = 20. GATK tools allowed merging of the variants called by both methods, to obtain a more robust variant detection procedure. Variants called in assembled genomes were then combined with the ones obtained from mapped sequences, and filtered to exclude the ones within phage regions and repetitive elements. After manual inspection, low-confidence alleles with a quality score < 130 or heterozygous base call were removed. Functional consequences (e.g. non-synonymous mutations and amino acid changes) of the detected variants were computed by CLC Genomics Workbench.

3.7.6. Genome annotation

For annotating the assembled bacterial genomes, the RAST (Rapid Annotations using Subsystems Technology) online server was employed. This automatic annotation service is based on a library of manually curated subsystems and on protein families, mainly derived from the subsystems, called FIGfams. When genomes were to be submitted to the GenBank database, the PGAAP pipeline of the NCBI was used, combining *ab initio* gene prediction

algorithms with homology based methods. Identification of tRNA genes was done by tRNAscan-SE and prophage regions were determined by PHAST and Prophage Finder.

3.7.7. Whole genome comparison

Comparison of multiple whole genome sequences was conducted by means of various programs. To align regions conserved in subsets of the genomes, the Mauve genome alignment package (version 2.3.1) was used. In particular, alignments were done by the progressiveMauve algorithm, which is able to align a large number of genomes with high accuracy, also taking care of regions differentially conserved among the analyzed organisms. For the generation of comparative genomic circular images, the BRIG (BLAST Ring Image Generator) software was chosen. Genome comparison based on the identification of orthologous genes, using bidirectional best hits (BBHs) as orthology criterion, was achieved by Edgar (Efficient Database framework for comparative Genome Analyses using BLAST score Ratios), allowing, among other features, classification of genes as core or pan genes. In particular, core genome development plots were calculated in Edgar according to the approach described in (Tettelin et al., 2005), while pan-genome statistics, based on a Heaps' law approach (Tettelin et al., 2008), were kindly provided by Dr. Jochen Blom (Justus-Liebig-University Giessen). Comparison analyses were also performed within the SEED environment, connected to the RAST server.

3.7.8. Functional annotation and comparative functionome analysis

Genes were functionally classified using KAAS (KEGG Automatic Annotation Server) by BLAST best-hit comparisons against KEGG Genes, a database created from public resources. The KAAS results contain KO (KEGG Orthology) assignments, which consist of manually defined ortholog groups for all proteins and functional RNAs, allowing automatic generation of KEGG Pathways and KEGG Modules (Moriya et al., 2007). Comparison among comprehensive functional categories (functionomes) could be obtained by MAPLE (Metabolic And Physiological potentiaL Evaluator), which calculates the completion ratio of all KEGG functional modules in each organisms based on KEGG Module (Takami et al., 2012). KEGG Module is a collection of pathway modules, which are manually defined functional units (such as consecutive reaction steps or operons); structural complex modules, which comprise multiple molecules (such as the subunits of transporters); functional essential sets; and signature modules. The query sequences are assigned to KO identifiers, whose combinations define each module, allowing functional annotation.

3.7.9. Phylogenetic analyses

Construction of the phylogenetic tree for analysis of the *Y. enterocolitica* species was performed using the protein sequences of all the core genes (calculated by the Edgar software, section 3.7.8) of a selection of genomes. Orthologous CDSs found in all genomes were concatenated and aligned using the multiple alignment tool MUSCLE. Uninformative characters not matching the alignment were removed by Gblocks. A distance matrix was then calculated from this alignment and finally a phylogenetic tree was constructed under the Neighbor-Joining method. The two latter methods were used in the Phylip implementations. A majority rule-consensus tree of 100 bootstrap replicates was also computed to evaluate node support.

A SNP-based phylogenetic tree was reconstructed using the filtered variants vcf file from section 3.7.5 as input and the sequence from YE strain Y11 as reference, removing insertions and deletions. For each strain, a nucleotide sequence containing the concatenated base calls at each SNP location was obtained. After performing a multiple alignment by MUSCLE, a DNA-based phylogeny was calculated using the Neighbor-Joining method with Phylip.

3.8. Statistics

Statistical analyses of results from the mouse experiments were performed with Prism software (GraphPAd) using the two-tailed unpaired Student's *t* test with Welch's correction and one way ANOVA. A P value < 0.05 was considered statistically significant.

Statistics on the called variants were also done with Prism software using a non-linear regression, with 95% confidence intervals and weighted sum of squares (1/Y).

4. Results

Y. enterocolitica is an important foodborne pathogen strongly associated with consumption of food products of porcine origin. Serotypes O:8, O:3, O:9 and O:5,27 are the most prevalent in humans, but genomics and pathogenesis studies have been mainly carried out on the mouse-virulent serotype O:8, especially strain 8081. In order to understand the pathogenicity and evolutionary mechanisms of a bacterial species, however, there is a need to analyze more than a single genome (section 4.1) and to characterize genomic features of multiple serotypes (section 4.2). Moreover, molecular epidemiology has shown low resolution for the microevolution of the most frequently isolated strains of serotype O:3, whose virulence and colonization factors are unknown. The precise intra-species relationships of this serotype, therefore, need further analysis (section 4.3), as well as the mechanisms underlying the pathogenicity in humans (section 4.4).

4.1. Highly-virulent *Y. enterocolitica* subsp. *enterocolitica*

Two *Y. enterocolitica* subsp. *enterocolitica* bioserotype 1B/O:8 strains, namely 8081 and WA-314, are commonly used in *Yersinia* research laboratories for their high-virulence properties in mouse infection models. The genome of strain 8081, the first YE to be sequenced (Thomson et al., 2006), was compared to the genome sequence of strain WA-314, obtained in this study, to disclose the virulence and fitness determinants shared by the high-virulence group, together with *in vitro* and *in vivo* experimental analyses.

4.1.1. Whole genome comparison of *Y. enterocolitica* bioserotype 1B/O:8

General features. The sequencing and assembly of the high-quality *Y. enterocolitica* strain WA-314 whole-genome, obtained in cooperation with BGI-Hongkong Co., produced 129 contigs, which could be further assembled in 4 large scaffolds and 8 contigs, with an additional scaffold constituting the pYV virulence plasmid. After head-to-tail concatenation of contigs with no definite order, the chromosome size resulted to have ~4,459,631 bp, with a G+C content of 47.24%, while the pYV is ~65,033 bp long, having 43.6% of G+C content (Table 21). These sizes are underestimated due to the inter-scaffold gaps, which represent repetitive regions not resolved by the used sequencing approach. RAST automatic annotation identified 4,086 coding sequences (CDSs) in the genome, and 87 CDSs in the pYV plasmid. Overall, 87.1% of the proteins have length < 500 aa, the bigger protein is 3,054 aa long and the smallest is formed by 38 aa (Figure 4). The chromosome also harbors 65 tRNA genes of 70-91 bp, and 7 prophage-like regions, which are 10-42 kb large. The contig sequences have been deposited in public

databases (DDBJ/EMBL/GenBank) as Whole Genome Shotgun project under Accession Numbers AKKR01000001-AKKR01000129.

The main features of the chromosomal and plasmidic sequences of YE strains 8081 and WA-314 are summarized in Table 21. The smaller size of strain WA-314 may be explained by the incompleteness of the genome and the presence of gaps between contigs, while the higher number of CDSs is probably due to the different annotation procedures adopted. Comparison with SEED considering only bidirectional best hit proteins with a minimum similarity score of 40% showed that 3,702 orthologous CDSs are shared between strain WA-314 and 8081 chromosomes, including the main known virulence factors. Genes specifically found in either of the two genomes mainly belong to hypothetical proteins or proteins encoded by mobile elements, such as prophages, but include also protein-encoding genes with specific assigned functions (Table 22).

Table 21 Properties of the chromosomes and pYV plasmids of *Y. enterocolitica* strains WA-314 and 8081.

Property	**Chromosome strain WA-314**	**Chromosome strain 8081**	**pYV_{WA-314}**	**pYV_{8081}**
Size	~ 4,459,631 bp	4,615,899 bp	~ 65,033 bp	67,721 bp
G+C content	47.24%	47.27%	43.6%	43%
Number of CDS	4,086	4,037	87	72
Average gene size	909 bp	968 bp	550 bp	675 bp
tRNA	~ 65	81	0	0
Prophage regions	7	4	0	0

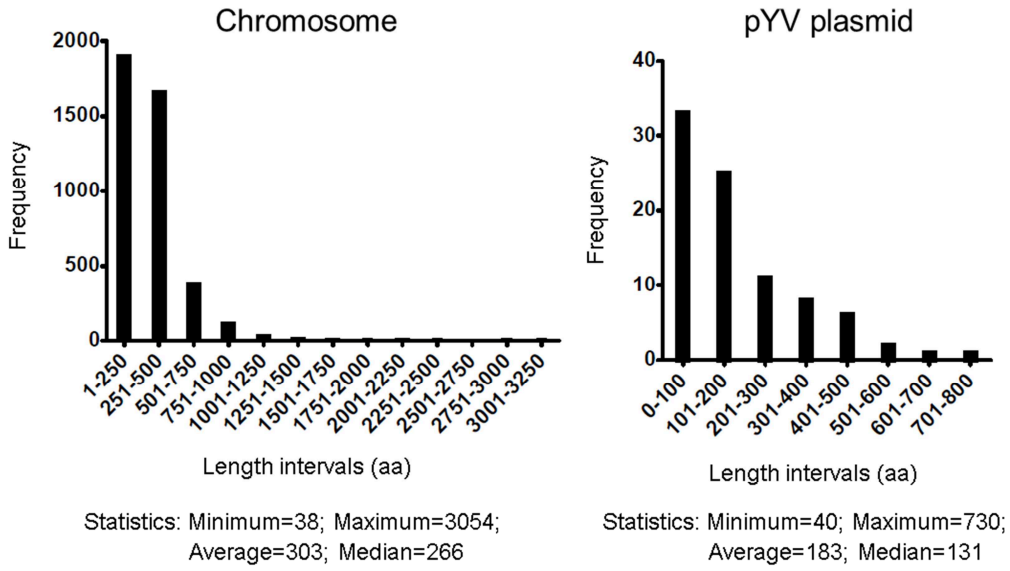


Figure 4 Histograms and statistics of *Y. enterocolitica* strain WA-314 protein length.

Mobile elements. Apart from the conserved YGI-1 and YGI-2, genomic islands and prophage-like regions represent the largest source of genetic variation between genomes of YE strains 8081 and WA-314 (Figure 5). Strain WA-314 is devoid of the YAPI (*Yersinia* Adhesion Pathogenicity Island), a 66-kb virulence-associated region located within the Plasticity Zone (PZ) of strain 8081 and integrated between an intact and a partial tRNA-Phe copy, which remains complete in strain WA-314. This island carries a type IV secretion system, a putative hemolysin, a toxin-antitoxin system (CcdAB family) and an arsenic resistance operon. The YGI-4, which is variably present in YE 1B, is absent from strain WA-314, while in strain 8081 it is substituted by a putative integrated plasmid. In contrast to the 4 prophages of strain 8081, 7 prophage-like regions were identified in strain WA-314 (Figure 5). Analysis with PHAST classified prophages YWA-1 and YWA-5 as intact, YWA-7 and YWA-2 as questionable, and YWA-3, YWA-4 and YWA-6 as incomplete. Prophages detected in strain 8081, ΦYE185 and ΦYE200, and ΦYE98, are partly similar to YWA-1 and YWA-2, respectively. A strain-independent acquisition of these prophages is likely, as these regions are not located in the same chromosomal context. Curiously, prophage YWA-1 is 70% identical to the bacteriophage L-413C of *Y. pestis* (Accession Number NC_004745), while YWA-2 shares CDSs with phage HP2 of *Haemophilus influenzae* (Accession Number NC_003315). The P4-like prophage YWA-4, located in a chromosomal region corresponding to the putative integrated plasmid YGI-3 in strain 8081 (see also Figure 20), harbors a PilV-like protein (locus tag: YWA314_12491) containing a Shufflon N-terminal region, similarly to adhesins in *Y. enterocolitica* subsp. *palearctica* and in *Y. intermedia*. As part of a probable operon encoding ancient type IV pili, this protein may be involved in adhesion to and invasion of the human intestinal cells.

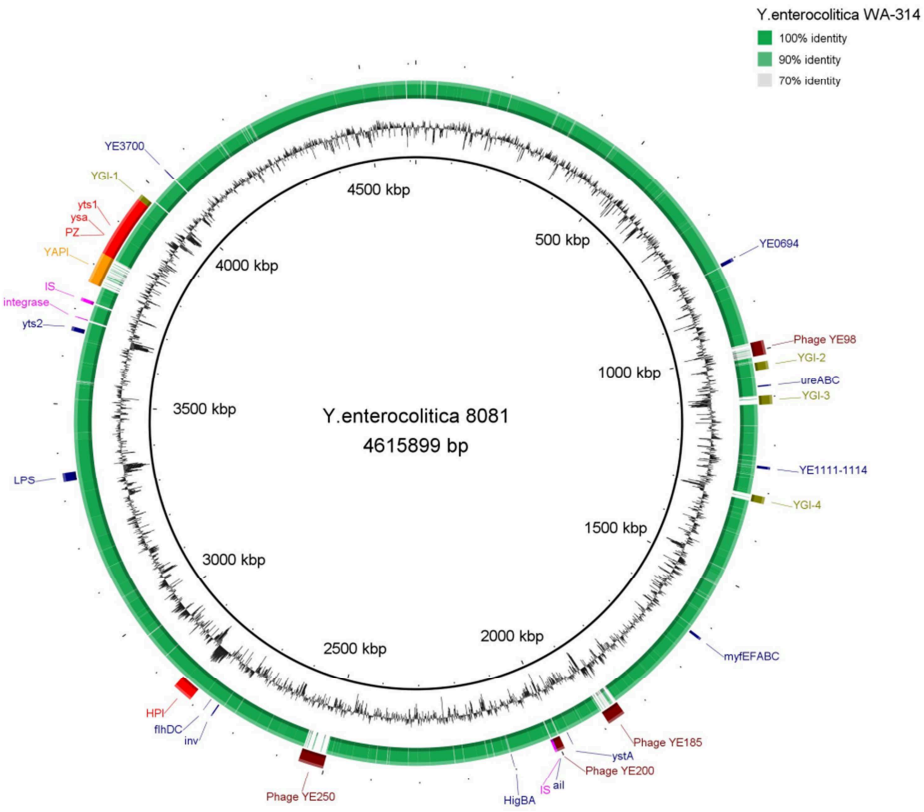
Table 22 Significant regions and genes of difference in highly-virulent *Y. enterocolitica*.

Region or gene cluster	Strain 8081	Strain WA-314	Description
YAPI (PZ)	Present	Absent	Pathogenicity island
YGI-3	Present (plasmid)	Present (prophage)	Putative hot-spot region for HGT
YGI-4	Present	Absent	Integrated plasmid
YE0694 YWA314_00878	83% protein sequence identity		Putative adhesion
YE3700 YWA314_14949	32% protein sequence identity		Autotransporter
YE1111-1114 YWA314_11901-11886	85% DNA sequence identity		Fimbrial operon
YWA314_17584-17599	Absent	Present	Restriction modification system
YWA314_20244-20259	Absent	Present	Colicin operon
YWA314_07469	Absent	Present	XAT acetyltransferase

Strain WA-314 specific genes. A restriction-modification (RM) system, a putative colicin cluster and a xenobiotic-acyltransferase (XAT)-encoding gene are specific acquisitions of YE strain WA-314. RM systems are defense mechanisms developed by bacteria to defend themselves against external DNA, which is cleaved by an endonuclease, after recognizing non-methylated cytosines. The host DNA is protected by methylated cytosines, and C to T mutations are repaired by very-short-patch repair proteins. The RM cluster of strain WA-314 is composed of 4 CDSs (locus tags: YWA314_17584-17599), which encode an EcoRII-like type II restriction endonuclease, a Vsr endonuclease, a cytosine methylase and a DNA-binding protein. This cluster is 95% identical to a RM system in YE serotype O:5,27 (section 4.2.8) and 86% identical to RM clusters carried by genomes of *Y. bercovieri* and *Y. frederiksenii*, indicating a probable parallel acquisition of this cluster from the environment.

Colicins are a type of bacteriocins produced by some bacteria to compete with closely related strains, with the producing bacteria being protected against the colicin activity by antagonist immunity proteins. They can act as pore-forming toxins or as endonucleases and may need a lysis protein to be released (Cascales et al., 2007). The colicin cluster detected in strain WA-314 consists of four CDSs (locus tags: YWA314_20244-20259), encoding three putative immunity proteins and one putative colicin. In particular, the bacteriocin sequence contains a cloacin-like tRNAse C-ter domain and an S-pyocin-like N-ter domain, suggesting an enzymatic endonuclease activity. The operon is located between a phage anti-termination protein- and a pilus chaperone-encoding gene, similarly to colicin clusters of YEP isolates, which share the same chromosomal background (section 4.2.2 and Figure 16). Moreover, according to Blast analysis, putative colicin-encoding genes are carried by other *Yersinia* species, indicating an ancient evolutionary origin of bacteriocin-like clusters in *Yersinia* and subsequent genetic loss in strain 8081.

The 221-aa XAT protein of strain WA-314 (locus tag: YWA314_07469), also annotated as chloramphenicol acetyltransferase and similar to streptogramin-A/virginiamycin/antibiotic acetyltransferases from various proteobacteria, is curiously present in YEP strains, but not in other bioserotype 1B/O:8 strains, as revealed by Blast search. In the same genomic location, strain 8081 harbors a 119-aa hypothetical protein (locus tag: YE2013), which seems to be strain-specific. Regarding the XAT protein of strain WA-314, the detected LbH (Left-handed parallel beta-Helix) domain, containing complete active sites and hexapeptide repeats, is typically present in the XAT class of hexapeptide acyltransferases, which inactivate antibiotics via enzymatic CoA-dependent acetylation. This gene might be responsible for the different resistance to streptogramins observed (see below).


Y.enterocolitica WA-314
100% identity
90% identity
70% identity
Y.enterocolitica 8081
4615899 bp
500 kbp
1000 kbp
1500 kbp
2000 kbp
2500 kbp
3000 kbp
3500 kbp
4000 kbp
4500 kbp
YE3700
YGI-1
yts1
ysa
PZ
YAPI
IS
integrase
yts2
LPS
HPI
flhDC
inv
Phage YE250
HigBA
IS
ail
Phage YE200
ystA
Phage YE185
myfEFABC
YE1111-1114
YGI-4
Phage YE98
YGI-2
ureABC
YGI-3
YE0694

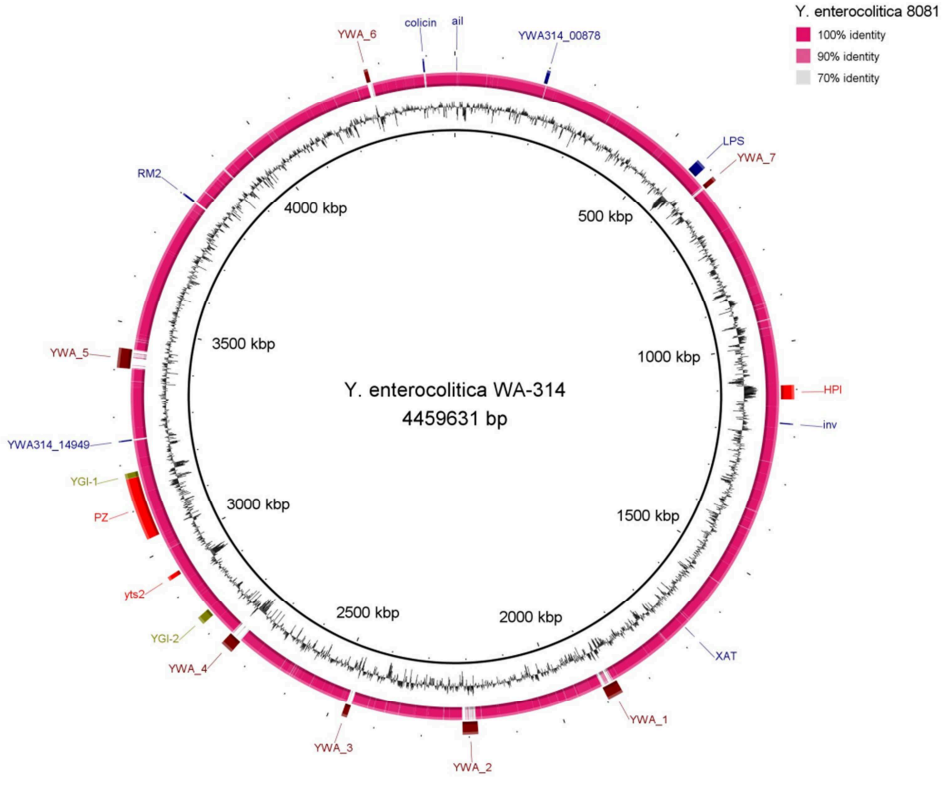

Y. enterocolitica 8081
100% identity
90% identity
70% identity
Y. enterocolitica WA-314
4459631 bp
500 kbp
1000 kbp
1500 kbp
2000 kbp
2500 kbp
3000 kbp
3500 kbp
4000 kbp
YWA_6
colicin
ail
YWA314_00878
LPS
YWA_7
RM2
YWA_5
YWA314_14949
YGI-1
PZ
yts2
YGI-2
YWA_4
YWA_3
YWA_2
YWA_1
XAT
HPI
inv

Figure 5 Comparative circular representation of the genomes of *Y. enterocolitica* strains 8081 vs WA-314 (top) and WA-314 vs 8081 (bottom). Significant regions and genes are shown in the outer ring. Modified from (Garzetti et al., 2012).

Identification of new putative virulence determinants. Genomic comparison of the two highly-virulent YE strains 8081 and WA-314 revealed presence of genes with low-sequence similarity, which might be associated with virulence. A 635-aa outer membrane protein (locus tag: YE3700) in strain 8081 is only 32% identical to its orthologous protein in strain WA-314, a 902-aa putative autotransporter (locus tag: YWA314_14949). The two genes share the same genomic region and have both low G+C content (43.1% in strain 8081 and 39.2% in strain WA-314). While YWA314_14949 has homologous genes in *Y. pestis*, *Y. similis* and *Y. pseudotuberculosis*, YE3700 can be found only in YE. Autotransporters in Gram-negative bacteria are highly diverse and perform a variety of functions that promote virulence by catalyzing proteolysis, serving as adhesins or cytotoxins, and mediating actin-promoted motility. These proteins are commonly organized in a signal peptide, followed by an N-terminal passenger domain, which is transported through the outer membrane by a pore formed by the C-terminal translocator domain (Benz and Schmidt, 2011). Both proteins of strains 8081 and WA-314 contain at the C-terminus an autotransporter beta-domain, traditionally involved in type V secretion of the protein itself. Additionally, a pertactin-like passenger domain of T5SS autotransporters could be recognized at the central region of YWA314_14949, similarly to the homologous AidA-I protein in *E. coli*, whereas no significant hits were found for the N-ter of YE3700. SignalP (version 4.1) could not predict any signal peptide in either of the two proteins, probably due to the low sequence identity among autotransporter sequences (Dautin and Bernstein, 2007).

A long adhesin shares 83% sequence identity between YE strains 8081 (locus tag: YE0694, 2,484 aa) and WA-314 (locus tag: YWA314_00878, 2,490 aa), but is missing in YEP, as previously described (Thomson et al., 2006), in *Y. pestis* and *Y. pseudotuberculosis*. Differences mainly lie in the C-terminal part of the proteins, where no known domains were detected. Both proteins contain, from the N-terminal to the C-terminal, a lysine motif (a small globular domain involved in peptidoglycan-binding), a domain of unknown function DUF3442 (intimin/invasin), and multiple bacterial Ig-like domains (typically found in bacterial surface proteins associated with cell-adhesion and pathogenicity). Interestingly, these adhesins are classified into the KEGG pathway “Bacterial invasion of epithelial cells”, together with the Ysa-T3SS proteins YspD and YspB.

A type-1 fimbrial operon has been identified as chromosomal region of difference among *Yersinia* species, since it is specifically carried by YE bioserotype 1B/O:8 (Thomson et al., 2006). In strains 8081 (locus tags: YE1111-1114) and WA-314 (locus tags: YWA314_11901-11886) the DNA sequence of the whole operon is 84.7% identical, with the most differences located in the first and last genes of the operon, encoding two fimbrial subunits. In particular, homologous proteins YE1111 (343 aa) and YWA314_11901 (274 aa) show an amino acid sequence similarity score of 36.7%, and proteins YE1114 (175 aa) and YWA314_11886 (177 aa) are 42.3% identical. While a signal peptide was recognized only in one WA-314 protein, YWA314_11886, both fimbrial proteins of strain 8081 contain a signal peptide at the N-terminal, suggesting a role in motility and secretion of these three proteins and unclear function of protein YWA314_11901.

Plasmid-encoded features. The DNA sequence of the pYV_{WA-314} was found to be identical to that previously published (Oberhettinger et al., 2011), as expected. Whole-sequence comparison between pYV_{8081} and pYV_{WA-314} (Figure 6) shows significant differences in repeat regions and in three Ysc-T3SS protein-encoding genes (*yopM*, *yscP* and *lcrV*), minor mutations in *yadA* and the presence of the *ylpA* gene in pYV_{WA-314}.

The LRR (leucine-rich repeat) effector protein YopM has shown interesting heterogeneity among YE bioserotype 1B/O:8 isolates (Boland et al., 1998; Heesemann et al., 1986b). Consistently, the $yopM_{8081}$ gene (locus tag: YEP0010) is 1,104 bp long and encodes a 367-aa protein with 13 LRRs, while $yopM_{WA-314}$ (locus tag: YWA314_20912) is 1,518 bp long and the encoded protein is composed of 505 residues. Being 100% identical to that of $YopM_{a127/90}$, $YopM_{WA-314}$ possesses 3 additional copies of LRR12-LRR13-LRR14 compared to $YopM_{8081}$ (Figure 7).

The $lcrV_{WA-314}$ (locus tag: YWA314_20937, 1,002 bp) and $lcrV_{8081}$ (locus tag: YEP0015, 975 bp) genes demonstrated an overall similarity of 95.9%. As previously observed (Roggenkamp et al., 1997), the V-antigen $LcrV_{WA-314}$ protein contains a 9-aa insertion (LHEVGVIAG) between residues 229 and 237, similarly to $LcrV_{a127/90}$ (Foultier and Cornelis, 2003). This insertion is missing in $LcrV_{8081}$, in weakly-virulent YE, in *Y. pestis* and *Y. pseudotuberculosis*, but is carried by LcrV proteins of serotype O:13 and O:20 strains (also biotype 1B) (Roggenkamp et al., 1997).

The YscP protein sequence of YE isolates contains a variable number of three repeated motifs of 14 aa, 25 aa and 46 aa (Wagner et al., 2009). Conforming to these data, the $yscP_{WA-314}$ is a 1,617-bp gene, encoding a 538-aa protein (locus tag: YWA314_20992), while $yscP_{8081}$ is 1,359 bp long and encodes a protein of 452 residues (locus tag: YEP0026). The longer $YcsP_{WA-314}$ protein possesses 3 additional repeats, absent from YE strain 8081 and from strains belonging to the weakly-virulent group, which in turn have 2 extra specific repeats (Figure 8).

The 422-aa virulence adhesin YadA of YE strains 8081 and WA-314 showed 96.9% of protein sequence similarity. In particular, between residues 53 and 67 there are six differing amino acids: AAEaLggtnalakSI in strains 8081 and EVRaPggtnalakGT in strain WA-314. This region is located in the head domain of the protein and is surface-exposed (Roggenkamp et al., 2003).

A notable difference found by comparison of the pYV sequences is the presence of the *ylpA* gene in pYV_{WA-314} (locus tag: YWA314_20862), which is missing from pYV_{8081} but present in $pYV_{a127/90}$, in pYV plasmids of YEP, of *Y. pestis* and *Y. pseudotuberculosis*. The encoded YlpA protein is an outer membrane lipoprotein homologous to the enterobacterial TraT protein, which inhibits phagocytosis in *E. coli* (Aguero et al., 1984).

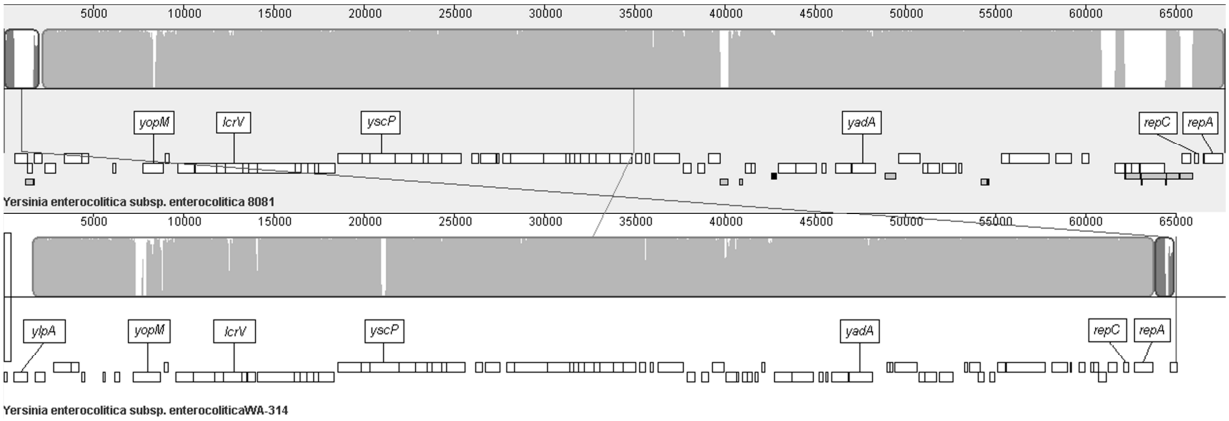


Figure 6 Mauve comparison between pYV_{8081} and pYV_{WA-314}.
Each sequence is arranged in a horizontal track filled with a colored similarity plot, the height of which is proportional to the level of sequence identity in that region (white shades therefore indicate regions with low or no sequence homology). Below the tracks, annotated CDSs are shown as white boxes on two different levels according to the coding strand, while repeated regions are designed as colored boxes.

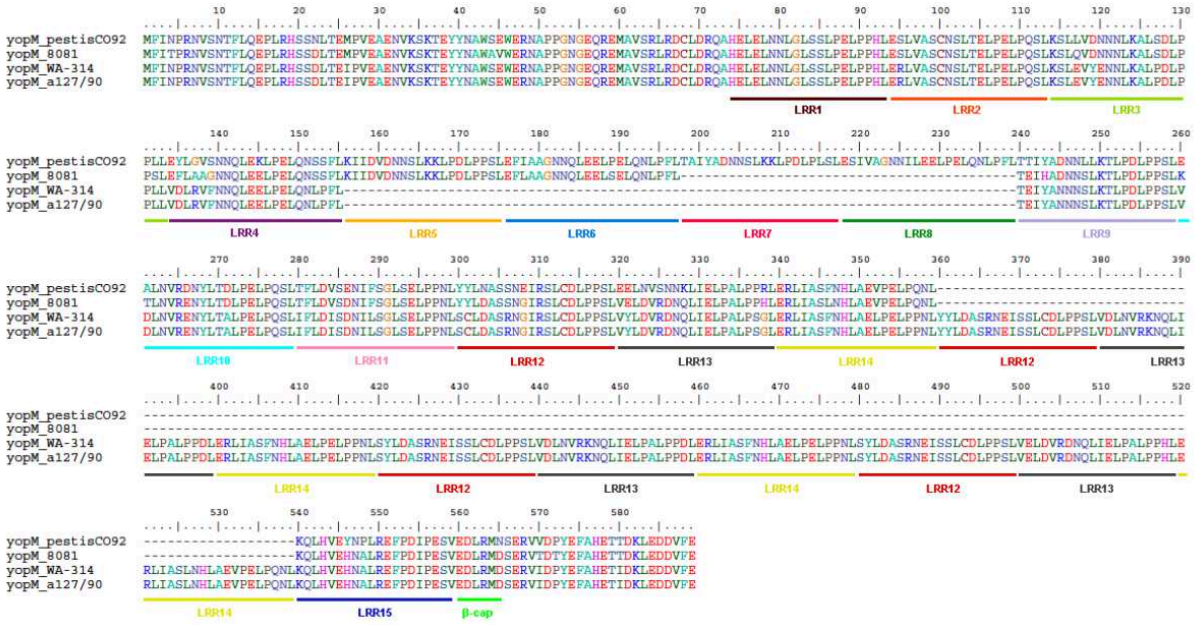


Figure 7 Alignment of YopM proteins and LRR structure.

Sequences from *Y. enterocolitica* bioserotype 1B/O:8 strains are aligned against *Y. pestis* CO92 as reference strain. LRR numbers as in (Vieux and Barrick, 2011). Modified from (Garzetti et al., 2012).

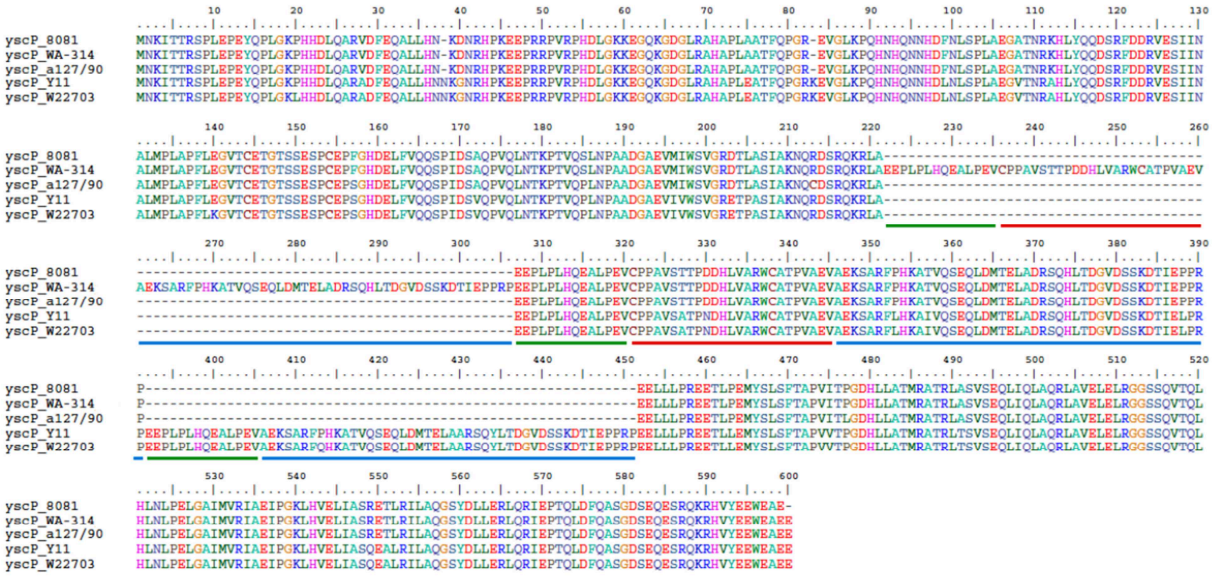


Figure 8 Alignment of YscP proteins.

Sequences from *Y. enterocolitica* highly-virulent strains (8081, WA-314 and a127/90) and weakly-virulent strains (Y11 and W22703) are displayed. Below the alignment, green, red and blue lines highlight the repeats of 14 aa, 25 aa, and 46 aa, respectively (Garzetti et al., 2012).

4.1.2. Functional characterization of strains 8081 and WA-314

Bacterial growth curve. The correlation between CFU/ml and OD_{600} during growth at 27 °C was determined for both *Y. enterocolitica* strains 8081 and WA-314, as described in section 3.3.4, revealing small reproducible differences (Figure 9). In LB and MH liquid media, both strains exhibited similar growth properties, reaching an OD_{600} of 0.5 (8081) and 0.56 (WA-314) after 6 h. However, when considering the CFU/ml ratio, strain WA-314 showed a slightly faster growth behavior, achieving a CFU/ml of $4.76x10^8$, compared to a ratio of $2.1x10^8$ for strain 8081. To note, no clear differences in metabolic and regulatory factors, which would explain the fastest growth rate of strain WA-314, were identified by genome analysis (section 4.1.1).

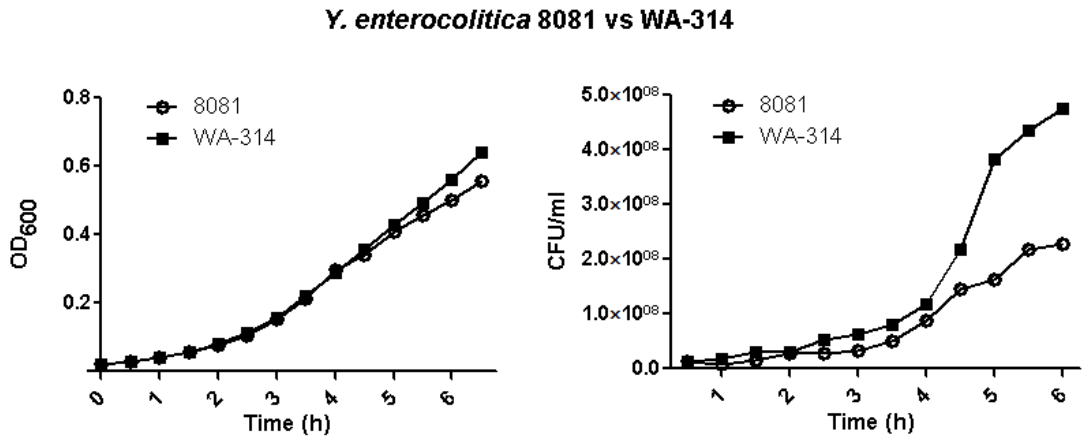


Figure 9 Growth curves of *Y. enterocolitica* strains 8081 and WA-314.

Resistance to chloramphenicol and streptogramins. Streptogramins are cyclic peptide antibiotics inhibiting protein synthesis. Being poorly hydrophilic molecules, they cannot enter the permeability barrier of Gram-negative bacteria, which are therefore intrinsically resistant. Chloramphenicol, on the other side, is a broad-spectrum antibiotic, also used in the treatment of yersiniosis. Due to the presence of a specific chloramphenicol/streptogramin acetyltransferase XAT gene in strain WA-314 (section 4.1.1), antimicrobial susceptibility tests to these classes of antibiotics were performed by broth dilution and disk diffusion methods, comparing the behavior of YE strains WA-314 and 8081. Chloramphenicol and virginiamycin, a streptogramin antibiotic, were added to the liquid media at serial two-fold dilutions. The two strains grew slightly differently in the presence of chloramphenicol, showing an MIC (minimal inhibitory concentration) of 4 µg/ml for strain 8081 and of 8 µg/ml for strain WA-314 (Figure 10). Virginiamycin, on the other hand, seemed to have low antimicrobial effect on strain 8081 and no effect on strain WA-314, whose growth was drastically reduced only at the antibiotic

concentration of 256 µg/ml. The disk diffusion method was also utilized to check the antibacterial activity of chloramphenicol and quinupristin/dalfopristin, two streptogramin antibiotics which show bactericidal activity only when used in combination with each other. Commercially available chloramphenicol (30 µg) and quinupristin/dalfopristin (15 µg) disks were applied to the surface of dried agar plates, previously inoculated with YE strains 8081 and WA-314. After o/n growth at 27 and 37 °C, the diameter of inhibition zones showed that both strains were clearly susceptible to chloramphenicol. The resistance to streptogramins is evident for strain WA-314 at both tested temperatures, while strain 8081 showed a lower resistance level (Table 23). The observed differences may be related to the XAT gene present in strain WA-314 but absent from strain 8081 (see above). Interestingly, this gene is 99% similar to a streptogramin A acetyltransferase (*sat*) gene from YE, which has been characterized, proving its functionality and acetylating activity (Seoane and Garcia Lobo, 2000). Curiously, virginiamycin has been used in farm animal feed as a growth promoter, but is now banned in Europe because of the selection for streptogramin resistance, which may be transferred to humans (Simjee et al., 2006). The close contact of YE with pigs might have selected for a streptogramin acetyltransferase gene with high affinity for this class of antibiotics.

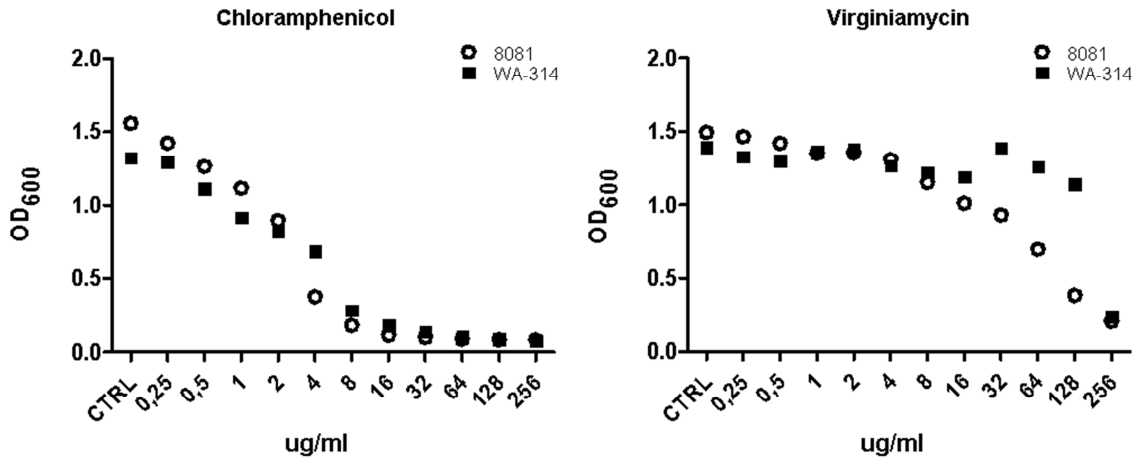


Figure 10 Antimicrobial susceptibility test by broth dilutions.

Growth of *Y. enterocolitica* strains 8081 and WA-314 with chloramphenicol and virginiamycin was tested. Data are represented as mean of OD_{600} absorbance values between two independent experiments after o/n growth at 27 °C in MH broth. CTRL: negative control (no antibiotics added).

Table 23 Disk diffusion test results.

Y. enterocolitica strains 8081 and WA-314 were tested for susceptibility to chloramphenicol and quinupristin/dalfopristin. Results are expressed as diameters of inhibition zone (cm).

Antibiotic	Temperature	Strain 8081		Strain WA-314	
Chloramphenicol	27 °C	2.4	2.3	2	2
	37 °C	2.5	2.6	2.1	2.2
Quinupristin/Dalfopristin	27 °C	0.9	1	0	0
	37 °C	1.1	1	0	0

4.1.3. Different virulence properties of highly-virulent *Y. enterocolitica*

Experiment optimization. The *in vivo* virulence properties of *Y. enterocolitica* strains 8081 and WA-314 were tested in a mouse model and compared by means of a co-infection experiment. Discrimination between YE strains 8081 and WA-314 was successfully obtained by a colony hybridization method using DIG-labeled probes (sections 3.4.6 and 3.5.4). Genomic comparison results allowed designing of strain-specific probes, Hem_8081 and Col_WA, targeting the putative hemolysin gene (locus tag: YE3454) in the YAPI region of strain 8081 and the colicin gene (YWA314_20259) of strain WA-314, respectively. Preliminary experiments to test the performance of the developed protocol were carried out analyzing plates with spotted known bacterial suspensions of either strain 8081 or WA-314. Once the colony hybridization protocol proved to be optimally working, cultures of the two strains were mixed, plated on the surface of agar plates and subject to the hybridization method. As demonstrated in Figure 11 A and B, probe Hem_8081 recognized colonies which were not targeted by Col_WA, and vice versa. Superimposing the membranes and examining the image in comparison with the original plate (Figure 11 C and D) revealed that all the bacterial colonies were recognized by either of the probes, and that no colonies were detected by both probes. This demonstrated the specificity of the developed method and its capability to differentiate between the two YE strains.

Infection in a mouse model. Three groups of 8 mice were infected with YE strain 8081, strain WA-314, or a 1:1 mix of both microorganisms, via i.p. injection of bacterial suspensions (section 3.6.2). Mice were weighted everyday, before they were sacrificed five days post infection (Figure 12). Except for mice n° 9 and 21, all the animals lost from 13.5% (mouse n° 11, co-infected) to 32% (mouse n° 5, infected with strain 8081) of their body weight, with no significant differences among the three groups.

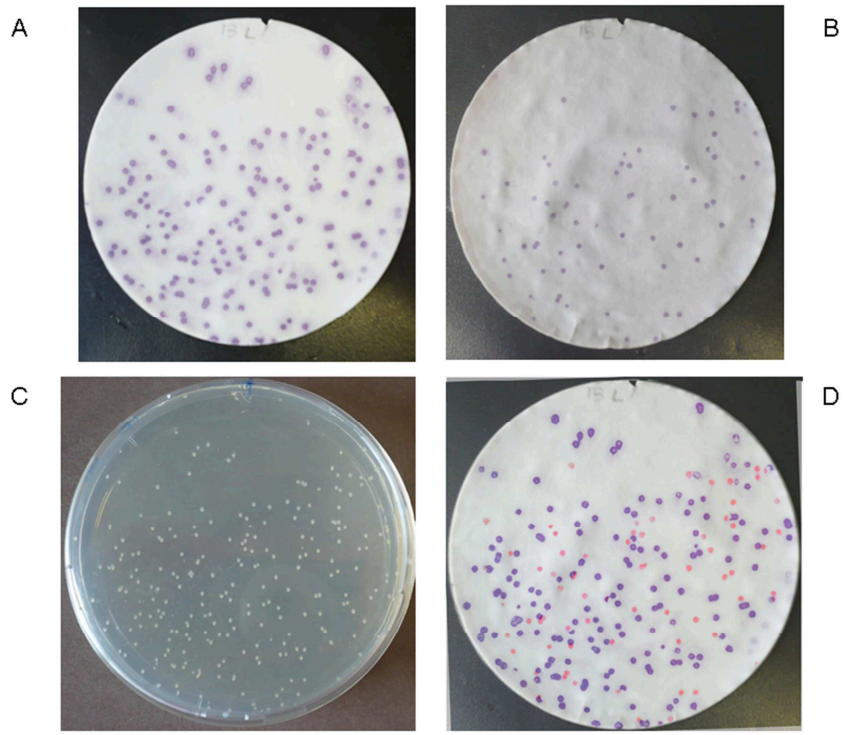


Figure 11 Example of colony hybridization experiment.
One nylon membrane from an LB-plate containing both *Y. enterocolitica* strains 8081 and WA-314 (C) was probed with Hem_8081 (A), stripped and re-probed with Col_WA (B). Superimposing images from A, in violet, and B, in pink, confirmed specificity of the probes (D). Modified from (Garzetti et al., 2012).

To prove that the two strains were equally present in the co-infection mixture, the colony hybridization protocol was conducted on agar plates incubated with dilutions of the mix. After colorimetric detection, colonies were manually counted and the results confirmed the 1:1 proportion between YE strain 8081 and strain WA-314 (out of 217 colonies, 100 were recognized by probe Col-WA and 117 by Hem_8081).

Counting of bacteria recovered from spleens and livers of the three groups of mice allowed comparison of bacterial loads and, consequently, of the virulence efficiency of the tested YE strains (Figure 13). Overall, bacterial counts were higher in spleen samples than in livers, as expected from previous experiments with YE done in our Institute. No bacteria were extracted from both organs of mice n° 9 and 21, indicating an unsuccessful i.p. injection, as confirmed by the lack of weight body loss in these two mice (Figure 12). Moreover, the bacterial count in the spleen of mouse n° 17 was zero, while few bacteria ($3x10^6$) were still present in the liver. Statistically significant differences were found between bacterial loads from spleens of mice independently infected with strain 8081 and strain WA-314 (P=0.011), and from livers of mice

infected with strain 8081 and co-infected mice (P=0.006). Bacterial colonization of mice infected with strain WA-314 alone and with combined strains was not significantly different in either organs (P_{spleen}=0.86 and P_{liver}=0.27). Livers from mice infected with strain 8081 contained fewer bacteria than WA-314 single infection, but the difference was not statistically significant (P=0.16).

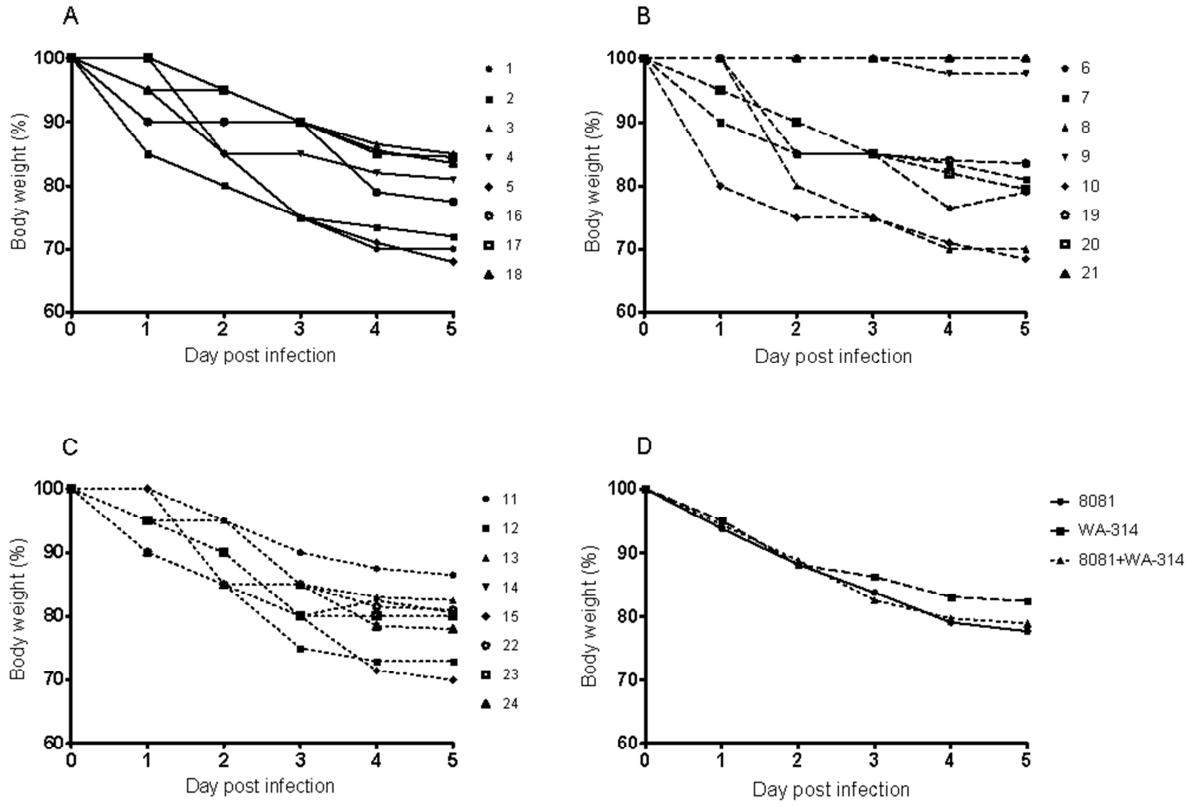


Figure 12 Mouse weight loss.
A total of 24 mice were divided into 3 groups and infected with *Y. enterocolitica* strain 8081, mice number 1, 2, 3, 4, 5, 16, 17, 18 (A); strain WA-314, mice number 6, 7, 8, 9, 10, 19, 20, 21 (B); and a mixture of both strains, mice number 11, 12, 13, 14, 15, 22, 23, 24 (C). The weight loss average of the mice belonging to each group is shown in (D).

To investigate whether the colonization of YE strains 8081 and WA-314 was equally efficient during co-infection, the optimized colony hybridization experiment (sections 3.5.4 and 4.1.2) was applied to bacteria recovered from spleens and livers of co-infected mice (Figure 14). Overall, YE strain WA-314 outperformed strain 8081. Strain 8081 exceeded strain WA-314 only in mouse n° 13, whose spleen contained the lowest bacterial count of the whole study ($5.81x10^6$), suggesting an incorrect inoculation of the bacterial mix in this mouse or a pre-induction of

inflammatory response because of mesenterial irritation. Without taking this mouse into consideration, on average strain WA-314 accounted for the 80.5% of the total bacterial loads in the organs of co-infected mice, indicating that strain WA-314 overcame strain 8081 in colonization of host tissues due to increased mouse virulence.

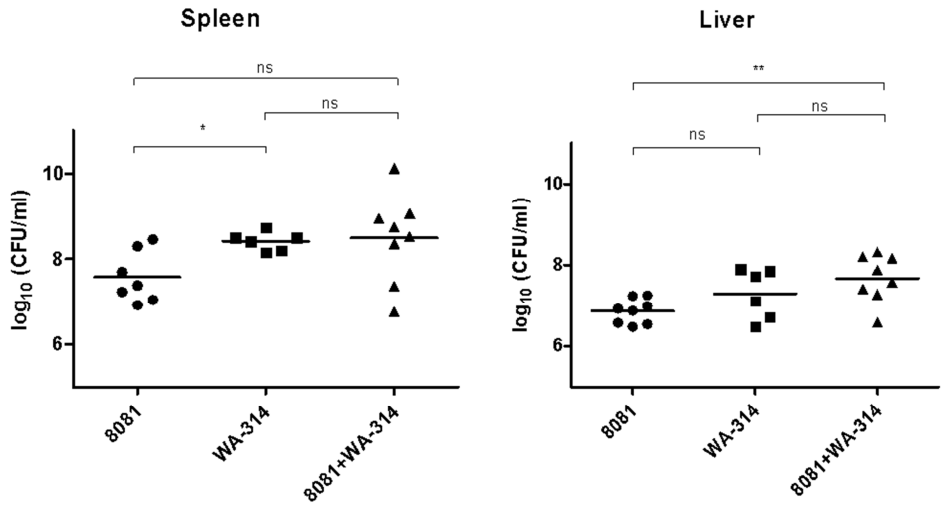


Figure 13 Mouse infection experiments.

Graphs show log-transformed bacterial counts (CFU/ml) recovered from spleens and livers of mice infected with *Y. enterocolitica* strain 8081, strain WA-314 and co-infected with both strains. The horizontal bars represent the mean number of bacteria for each group. Overall ANOVA analyses gave *P*=0.06 for spleen data, and *P*=0.016 for liver data. Individual groups were compared using two-tailed unpaired *t* test with Welch's correction: ns, not significant; *, *P*<0.05; **, *P*<0.01. Modified from (Garzetti et al., 2012).

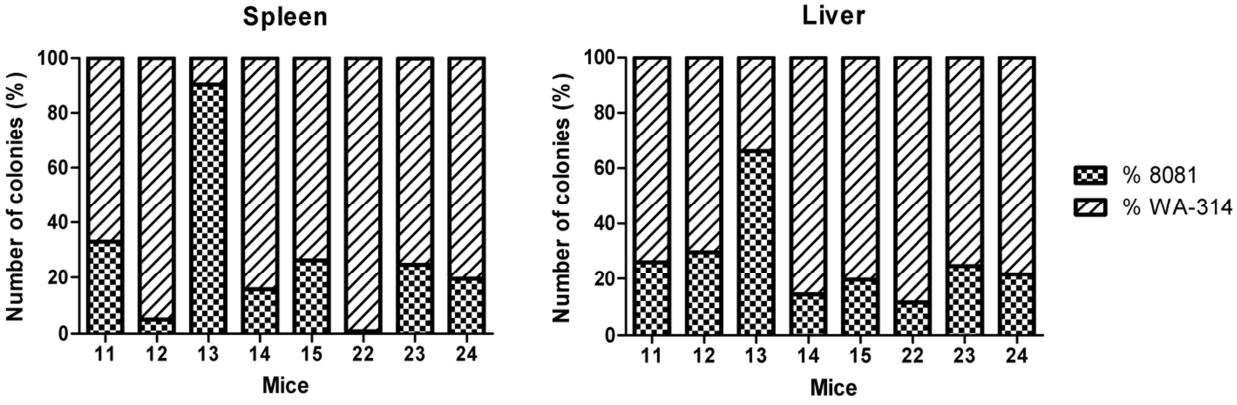


Figure 14 Differentiation of *Y. enterocolitica* strains 8081 and WA-314 in co-infected mice.

Data were obtained applying the colony hybridization method on bacteria recovered form spleens and livers of the 8 co-infected mice. Strain WA-314 overcame strain 8081 in 7 out of 8 mice. (Garzetti et al., 2012).

4.2. Genome diversity of the *Y. enterocolitica* species

Whole genome sequences of different strains belonging to *Y. enterocolitica* subsp. *palearctica* are available in public databases, and have been analyzed in previous studies to identify novel virulence genes and fitness factors of the weakly-virulent group, emphasizing the dissimilarities with the highly-virulent strain 8081 (section 1.5). However, one genome is not sufficient to gather the completeness of the variable population structure of *Y. enterocolitica* subsp. *enterocolitica*, as shown in this study (section 4.1). In addition, undisclosed differences within YEP genomes may explain the adaptation of these bacteria to a variety of ecological niches and their wide dissemination in humans, despite the low virulence (section 1.4). In this section, genome sequences of weakly-virulent and non-virulent strains are analyzed between each other and in relation to the mouse-virulent YEE genomic traits (section 4.1.1), allowing also a thorough investigation of the species YE in its entirety.

The 10 genomes used in this study and their main characteristics are summarized in Table 21 (YEE) and Table 24 (YEP): 2 highly-virulent strains of bioserotype 1B/O:8 (8081 and WA-314); 6 weakly-virulent strains of serotype O:9 (W22703 and 105.5R(r)), bioserotype 4/O:3 (Y11, Y8265, Y5307) and serotype O:5,27 (Y5.27P); 2 non-virulent strains of bioserotype 1A/O:5 (NF-O) and 1A/O:36 (IP2222). These strains cover the main serotypes isolated from diseased humans and include also an environmental isolate and a non-virulent strain isolated from a patient, representing the diversity of the species YE and the isolates with clinical and non-clinical significance.

4.2.1. *Y. enterocolitica* subsp. *palearctica* whole-genome comparison

To gain a general understanding of the main genomic differences in *Y. enterocolitica* subsp. *palearctica*, BRIG comparisons were computed using the two complete genomes of strains Y11 and 105.5R(r) as references (Figure 15). As expected, prophages represent the main source of genomic variation, even if some of them (e.g. PhiYep-2 and PhiYep-3) are partly conserved among different serotypes. The LPS/O-ag cluster is clearly serotype-specific (see section 4.2.7), while the Aat, Flag-2 and Tc clusters are exclusively carried by weakly-virulent strains of serotypes O:3, O:9 and O:5,27, as previously determined (section 1.5). Interestingly, apart from prophage regions, hypothetical proteins and O-ag-encoding genes, the three YE serotype O:3 strains carry other serotype-specific regions, such as the genomic island named GIYep-01 (Batzilla et al., 2011a), a colicin type 7 Cja (locus tag: Y11_38121), a putative N-acetylglucosamine-binding protein (Y11_42051) and a HipBA system (Y11_28521-28531)

located within prophage Phi-b. On the other hand, regions specifically acquired by serotype O:9 strains only include prophages, hypothetical proteins and the O-ag gene clusters.

Table 24 Main features of the *Y. enterocolitica* subsp. *palearctica* genomes analyzed in this study. Data correspond to information published in the GenBank database or calculated by the RAST server.

Feature	***Y. enterocolitica* strain**			
	Y11	**105.5R(r)**	**Y5.27P**	**NF-O**
Bioserotype	4/O:3	3/O:9	3/O:5,27	1A/O:5
Origin	human, Germany	human, China	human, Germany	human, Canada
GenBank Acc. No.	FR729477	CP002246, CP002247	CACW01000001-CACW01000020	CACY01000001-CACY01000097
No. of contigs	1	1	20	97
Genome Size (Mbp)	4.55	4.55	≥ 4.66	≥ 4.66
G+C content (%)	47.0	47.0	46.9	47.1
Predicted CDSs	4,349	4,016	≥ 4,351	≥ 4,160
tRNA	70	71	≥ 98	≥ 76
	IP2222	**W22703**	**Y8265**	**Y5307**
Bioserotype	1A/O:36	2/O:9	4/O:3	4/O:3
Origin	water, Japan	human, Belgium	human, France	human, Finland
GenBank Acc. No.	CACZ01000001-CACZ01000074	FR718488-FR718797	CACU01000001-CACU01000014	CACV01000001-CACV01000018
No. of contigs	74	305	14	18
Genome Size (Mbp)	≥ 4.75	≥ 4.75	≥ 4.52	≥ 4.55
G+C content (%)	47.1	46.9	47	46.9
Predicted CDSs	≥ 4,234	≥ 5,006	4,296	4,393
tRNA	≥ 82	≥ 68	≥ 61	≥ 94

Concentrating the investigation into serotype- and subspecies-specific features allowed discovering of highly variable regions in YEP and YEE genomes located within or near mobile elements. An interesting region, from now on called VR1 (variable region 1) for more clearness, is located in the same chromosomal location between transposase-encoding genes (locus tags: Y11_22211-22491), and carries functional units differentially present in the analyzed strains (Table 25). In particular, a four-CDS cluster (Y11_22231-22261), which encodes anaerobic DMSO reductases in YE serotypes O:3, O:9 and O:5,27, is replaced by nine genes encoding carbohydrate ABC transporters and transketolases in biotype 1B (YE0818-0826). Strains of biotype 1A carry both gene clusters and an additional 1A-specific cluster involved in fucose metabolism and carbohydrate transport. A cluster (Y11_22281-22341) with homologous functions to the LIV (leucine, isoleucine, and valine) branched-chain amino acid transport system (encoded by the *livKHMGF* operon) is then only present in YEP strains. It is flanked by a sorbose uptake system (Y11_22371-22421), which is universal for YE, by an enterotoxin-

encoding gene, unique for 1A strains, and by an ABC transporter cluster, specific for 1A and 1B strains (YE0801-0806). Curiously, another ABC transport system, the phosphate transporter PstBAC (Y11_43361-43381), is inserted in YEP genomes in a region corresponding to transposase-genes of the IS1400 in YEE (YE1122-1123). To note, all YE carry a LIV transport system (YE0230-0234 and Y11_34321-34361); therefore, strains of the YEP possess two LIV system-encoding gene clusters, denoting nutritional versatility and ability to catabolize a diverse array of organic compounds.

Flanked by a tRNA-Ser-encoding gene (locus tag: Y11_t25250), genomes of YEP strains harbor nine CDSs that are absent from strain 8081 and WA-314. These genes (Y11_04791-04871) belong to the *nrfABCDEFG* operon, which encodes enzymes involved in the anaerobic reduction of nitrite to ammonium in the so-called dissimilatory nitrate reduction process. The nitrite reductase gene remnant YE1598 of strain 8081 is a pseudogene of *nrfG*. Overall, YEP have two systems for nitrite reduction, NrfAH and NirBD, while YEE can use only the NirBD pathway, as supported by KEGG and SEED metabolic analyses.

A known species-specific region of hyper-variability in YE is the 200-kb locus named plasticity zone (PZ) in strain 8081 (Table 3 and Table 25). The second large flagellar operon Flag-2 is found within the PZ of genomes of weakly-virulent strains but it is absent from biotypes 1A and 1B. The Flag-2 cluster is complete in serotypes O:9 (locus tags: YE105_C3260-C3302) and O:5,27, and is divided into two parts by a transposase-gene insertion. The second part of the Flag-2 (YE105_C3289-C3302) is missing in serotype O:3 strains, which are therefore devoid of the components for flagellum transport and assembly. Serotype O:3 strains also lack the adjacent genes (YE105_C3304-C3311), encoding the *fep-fes* gene cluster required for uptake and utilization of the siderophore enterochelin, which are present in all other analyzed serotypes, as previously discovered (Schubert et al., 1999). YE biotypes 1B and 1A carry also the *fepA* gene (YE3624), containing a frameshift mutation in highly-virulent strains, and five genes encoding exported and outer membrane proteins (YE3626-3630). In the region corresponding to these genes, strains of biotypes 2-5 uniquely harbor the Aat (aggregative adherence transporter) system (YE105_C3315-C3324), followed by a Gifsy-2 prophage protein (YE105_C3326).

In YE serotype O:3 genomes, two transposase-encoding genes are inserted where the other strains possess a genetic cluster (locus tags: YE3357-3361) involved in sucrose transfer and metabolism. Interestingly, analyzing the KEGG “Starch and sucrose metabolism” pathway, it appears that strain Y11 has only one maltodextrin glucosidase (Y11_20981) for regulating intracellular levels of maltose, in contrast to strains 8081 and 105.5R(r), which can count on two enzymes (YE3359 and YE3177-MalZ).

In addition to the previously-discovered and above-mentioned gene clusters, YEP-specific acquisitions include: sucrose permease and hydrolase (locus tags: Y11_05791-05801); multidrug resistance proteins (Y11_09551-09561), flanked by three biotype 1A-specific nitric oxide reductases; chloramphenicol-acetyltransferase (locus tag: Y11_12731); 10-CDS cluster encoding amino acid transporters (Y11_21641-21741, with Y11_21711 being only present in serotype O:3 and biotype 1A strains); metyltransferases and transporters (Y11_23741-23771); Na^+/bile acid symporter, hemolysin activator and hemolysin protein (Y11_27781-22801); regulator of cell morphogenesis and NO signaling (Y11_36341). The presence of gene clusters shared between the weakly-virulent and non-virulent groups, however, is not correlated with the YE phylogenetic relationships (section 4.2.5), and may suggest an independent parallel acquisition of these features.

Table 25 Functional units encoded by the variable regions VR1 and PZ in *Y. enterocolitica*.

Gene clusters	Highly-virulent biotype 1B	Weakly-virulent biotypes 2-5	Non-virulent biotype 1A
VR1			
Y11_22231-22261 DMSO reductases	-	+	+
YE0818-0826 Carbohydrate ABC transporters	+	-	+
Fucose metabolism and carbohydrate transport system	-	-	+
Y11_22281-22341 LIV transport system	-	+	+
Y11_22371-22421 sorbose Uptake system	+	+	+
Outer membrane autotransporter	-	-	+
Enterotoxin gene	-	-	+
YE0801-0806 ABC transporters	+	-	+
PZ			
YE105_C3260-C3302 Flag-2	-	+ (partly absent in O:3)	-
YE105_C3304-C3311 *fep-fes* cluster	+	+ (absent in O:3)	+
YE3624 *fepA* gene	+	-	+
YE3626-3630 Outer membrane proteins	+	-	+
YE105_C3315-C3324 Aat transporters	-	+	-

4.2.2. Putative virulence determinants in *Y. enterocolitica* subsp. *palearctica*

Comparison of multiple genomes belonging to each subspecies permitted identification of yet undiscovered putative disease-associated factors in *Y. enterocolitica* subsp. *palearctica*.

An RTX-toxin cluster (locus tags: Y11_18761-18781) has been proposed as YEP specific virulence factor (section 1.2.3, page 12); this cluster is absent from strain IP2222 (1A/O:36), while genes in strain NF-O (1A/O:5) show low sequence similarity. Notably, all YE genomes contain a variety of other genes annotated as “putative-RTX family protein”. One cluster (locus tags: YE1998-2000) is conserved among all strains and encodes three transporters probably implicated in RTX toxin/hemolysin secretion. A large RTX-family protein-encoding region (locus tag: YE1322) is present in all pathogenic strains and in the environmental strain IP2222, while it is absent from the genome of strain NF-O. Another protein (locus tag: YE2966, Y11_18861), encoding a putative hemolysin, has a high sequence similarity score (99%) among YEE and among YEP genomes, but between the two subspecies the sequence identity of the protein is lower (59-67%).

A five-CDS beta-fimbriae cluster (locus tags: Y11_14931-14971) has been specifically detected in YEP genomes and is located near a transposase for IS1667, in a region where YEE strains harbor a gene fragment encoding a putative fimbriae chaperone (locus tag: YE2602). The genes of this cluster encode a PapD-like chaperone protein, a large PapC-like usher protein and three beta-fimbriae major subunits. The biogenesis of fimbriae requires the two-component assembly and transport system consisting of the chaperone/usher pathway, suggesting that the three beta-fimbriae subunits of the YE cluster might be correctly assembled and used by the bacteria to attach to host surfaces. Similarly, another 4-genes fimbrial cluster (locus tags: Y11_26051-26081) is specific for the weakly-virulent strains of serotypes O:3, O:5,27 and O:9, while it is absent from biotypes 1A and 1B. It encodes two fimbrial subunits, a fimbrial usher porin and a chaperone protein. As the other cluster, these usher and chaperone proteins share sequence similarity with the PapC and PapD super-families, respectively. Notably, this cluster is partly carried by the genomes of *Y. intermedia*, *Y. mollaretii*, *Y. massiliensis* and *Y. frederiksenii*, which are typically considered as non-pathogenic but whose clinical significance remains controversial.

A putative Ig-like invasin (locus tag: Y11_38661) is specific for YEP. The 1,024-aa protein contains a DUF3442 domain of uncharacterized bacterial proteins and three bacterial Ig-like domains, indicating a probable location on the cell surface and a role in pathogenicity. In fact, Ig-like domains in enterobacteria are known to play important functions during host cell adhesion and invasion, and to be structural components of fimbrial systems and of the intimin/invasin family of outer membrane adhesins (Bodelon et al., 2013).

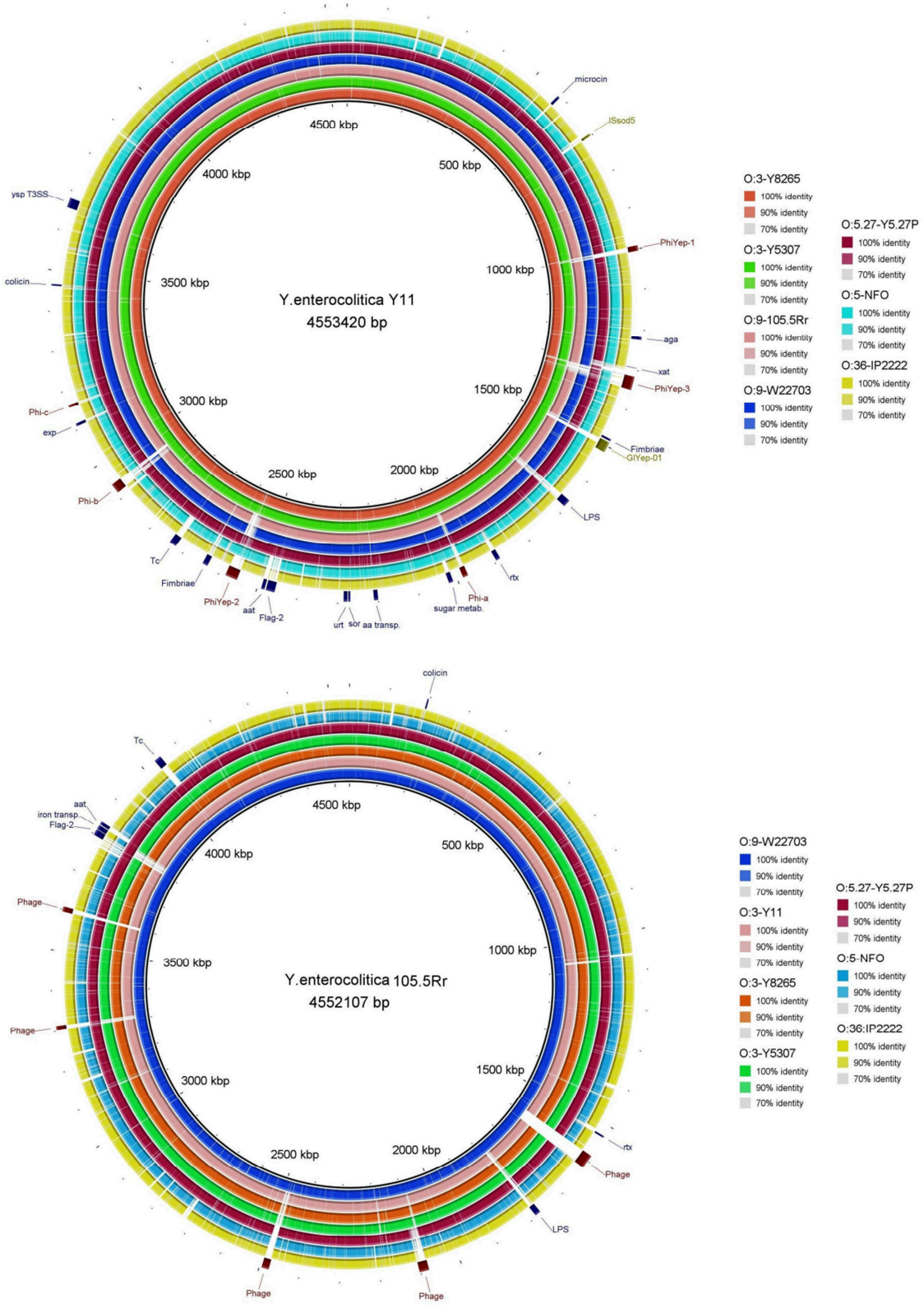


Figure 15 Genomic comparison of *Y. enterocolitica* subsp. *palearctica* based on the sequence of strains Y11, bioserotype 4/O:3 (top), and 105.5R(r), bioserotype 3/O:9 (bottom).

An intriguing protein family which might play a role in virulence of YEP includes autotransporters. In fact, these secreted and cell-surface-exposed proteins are produced by a large variety of pathogenic Gram-negative bacteria and have virulence-related functions, including adhesion, biofilm formation and apoptosis (Grijpstra et al., 2013). One gene (locus tag: Y11_42801, YE105_C3026), whose sequence is weakly conserved among the analyzed strains, is absent from YEE genomes. This protein contains a C-ter autotransporter beta-domain which enables the N-ter pertactin-like passenger domain to be exported across the membrane. Curiously, the YEP genes have homologous in *Y. pestis*, *Y. pseudotuberculosis* and *Y. similis*. A protein with similar structure has been previously discussed in biotype 1B strains (locus tag: YE3700 and YWA314_14949) as putative virulence factor (section 4.1.1). Orthologous to this gene are found in all YE genomes. The short genes (1,274 bp) from weakly-virulent strains (locus tag: Y11_25751, YE105_C3400) show high sequence similarity among each other (99.9%), but are 47% identical to YE3700 and 27% to YWA314_14949, being mainly similar to the C-ter regions and having homologous genes only in YE genomes. In contrast, the long biotype 1A genes (2,825 bp) are found in *Y. pestis*, *Y. pseudotuberculosis* and *Y. similis*, and exhibit 37.7% of sequence similarity with YE3700 and 72.6% with YWA314_14949. Interestingly, weakly-virulent YE carry a second adjacent gene encoding a long putative pertactin-family protein (locus tag: Y11_25741), containing both passenger and beta-barrel autotransporter domains. A third gene encoding an outer membrane autotransporter is specific for the two analyzed non-virulent strains NF-O and IP2222. It is located between genes YE0798 and YE0799 in the highly variable region VR1 discussed above. The gene shares 76% identity with genes in *Y. pestis*, *Y. pseudotuberculosis* and *Y. similis*, while the protein sequence is 80% identical to autotransporters in *Y. rohdei* and *Y. frederiksenii*. Pertactin-like passenger and autotransporter domains of the adhesin AidA are recognized. Finally, it is worth mention that all analyzed YE genomes carry the highly conserved YapE autotransporter protein (locus tag: YE4059), which has a demonstrated important role in efficiently colonizing mammalian host tissues by *Y. pestis* (Lawrenz et al., 2009).

A colicin-like locus absent from YEE strain 8081 has been detected in strains Y11 (4/O:3) and 105.5R(r) (3/O:9) (Batzilla et al., 2011a; Wang et al., 2011). The cluster in serotypes O:3 and O:9 strains is located between the *ilvC* gene (locus tag: Y11_33501) and a pilus assembly chaperone PapD (locus tag: Y11_33551), flanked by a phage antitermination protein-encoding gene; it is composed of four/five CDSs, depending on the bacterial strain (Figure 16). The homologous colicin locus in biotype 1A strains is inserted downstream the phage antitermination protein, near the PapD-encoding CDS. Surprisingly, a similar cluster has been detected in the YEE strain WA-314 (section 4.1.1), and is 95% and 74% identical to the cluster carried by the

non-virulent and weakly-virulent strains, respectively. Interestingly, the Y11_33511 colicin protein of strain Y11 contains not only a cloacin-like tRNAse C-ter domain and an S-pyocin-like domain, as in protein YWA314_20244 of strain WA-314, but also a conserved domain belonging to the apolipophorin-III family, which plays a critical role in the transport of lipids during insect flight. Overall, the colicin cluster in YE resembles a locus in *Y. similis* and *Y. pestis*, explaining the presence of an insect-related protein domain and supporting the hypothesis of a common origin of the bacteriocin cluster in *Yersinia* species. To note, all the analyzed YE strains carry the Tol and TonB transit machineries required for translocation of colicin-like proteins, and a colicin V production membrane protein CvpA, indicating a putative successful functionality of the detected bacteriocin gene clusters.

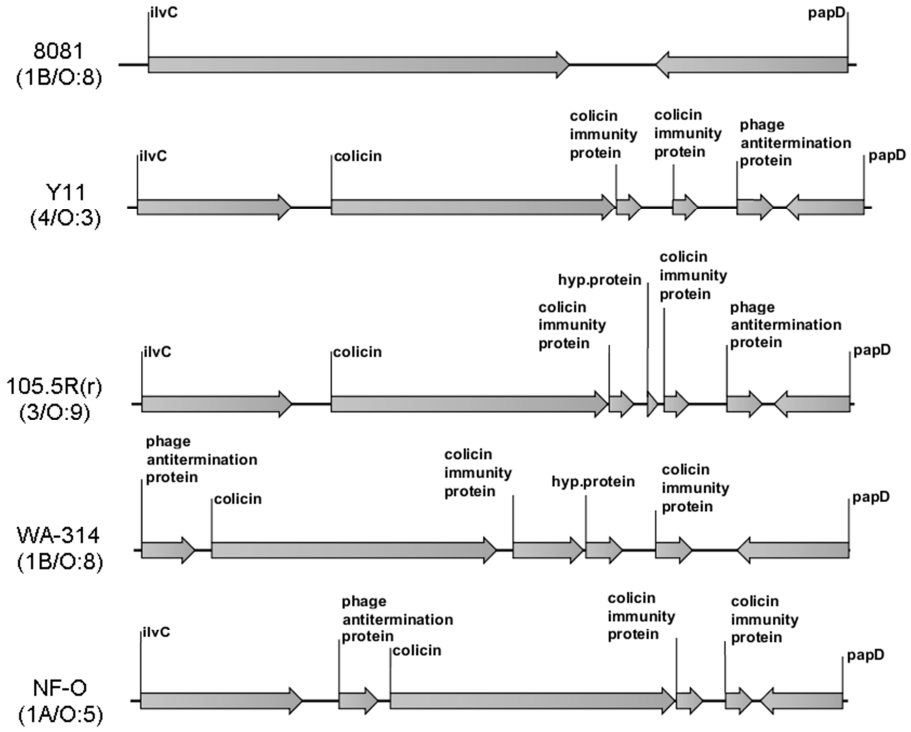


Figure 16 Colicin clusters in *Y. enterocolitica*.

The colicin operon structure is similar in all serotypes, with small differences in the insertion locus and in protein length.

4.2.3. Core and pan-genomes of *Y. enterocolitica*

The core and pan-genomes of the whole species *Y. enterocolitica* were obtained by iterative pairwise comparison of the 10 selected genomes, using the orthology-based software Edgar (section 3.7.7). The core genes are defined as the genes common to all the analyzed genomes, while the pan-genome comprises the core genes, the dispensable genes (present in two or more genomes) and the unique genes specific to single strains. The core genome is relevant for a bacterial species as it is linked to the fundamental biological processes and derived traits from a common ancestor, while the dispensable and unique genomes more often govern the genetic features related to variation in virulence and adaptation (Tettelin et al., 2008).

Considering that some YE strains do not carry the pYV virulence plasmid, this analysis was accomplished using only the chromosomal sequences. The core genome of YE, identified using strain 8081 as reference, consists of 3,090 orthologous genes, accounting on average for 73.7% of the whole coding genome. Overall, there is a descending trend of the core genome size with increasing sequence number, as predicted by the Edgar software using a non-linear least squares model fit (Figure 17). The core genome quickly reaches a plateau at value 3,061 with 20 genomes, indicating that even a small genome size of 10 is sufficient to satisfactorily describe the core genome of the species YE. Intriguingly, calculating the core genome without one strain of bioserotypes 4/O:3 or 1A/O:5 does not considerably increase the number of genes shared by all isolates (Table 26), suggesting a stable genetic backbone in these strains. In contrast, removing the bioserotype 2/O:9 strain W22703 causes the biggest increment of the core gene number, in accordance with the outlying position of this strain in the YE phylogeny (section 4.2.5). Since the core genome is based on the orthologous genes and is, therefore, dependent on the annotations, the anomalous high number of CDSs in strain W22703 (Table 24) may account for the important increment in the core genome observed when this strain is missing.

Annotating the core genes by means of the KEGG Automatic Annotation Server (KAAS) server (section 3.7.8), 2,346 genes out of 3,090 (75.9%) can be functionally assigned to a KEGG orthology. The majority of these genes are predicted to be involved in various housekeeping functions, especially those related to carbohydrate and amino acid metabolism, and associated transport activities (ABC transporters and two-component systems).

The pan-genome calculated on the 10 YE genomes and with strain 8081 as reference consists of 8,102 genes, of which 5,012 are dispensable/unique genes. Thus, the wide pool of non-core genes mainly contributes to the large pan-genome of YE. As expected, reducing the number of genomes, the number of pan genes also decreases (Table 26). In contrast to the core genome, the pan-genome does not display an asymptotic trend and increases with the incorporation of

additional genomes (Figure 18). Moreover, the plot of singletons/new genes is well fit by a decaying function and the extrapolated curve reaches an asymptotic value of 122.6. This means that every newly sequenced genome could bring on average 122 new genes. These data reveals that YE possesses an open pan-genome, indicating strong genetic plasticity, which is typical of bacterial species able to colonize multiple environments and having multiple ways of exchanging genetic material.

Table 26 Core and pan-genome sizes of *Y. enterocolitica*.
Core and pan genes were calculated using strain 8081 as reference on 10 or 9 genomes, removing one sequence at a time. When strain 8081 itself was excluded, strain Y11 was utilized as reference.

***Y. enterocolitica* strain excluded**	**Number of core genes**	**Number of pan genes**
none	3,090	8,102
8081 (1B/O:8)	3,110	7,844
Y11 (4/O:3)	3,095	7,937
105.5R(r) (3/O:9)	3,121	8,058
Y5.27P (3/O:5,27)	3,102	7,902
W22703 (2/O:9)	3,168	7,095
Y8265 (4/O:3)	3,099	8,032
Y5307 (4/O:3)	3,096	7,949
WA-314 (1B/O:8)	3,112	7,893
IP2222 (1A/O:36)	3,105	7,676
NF-O (1A/O:5)	3,097	7,781

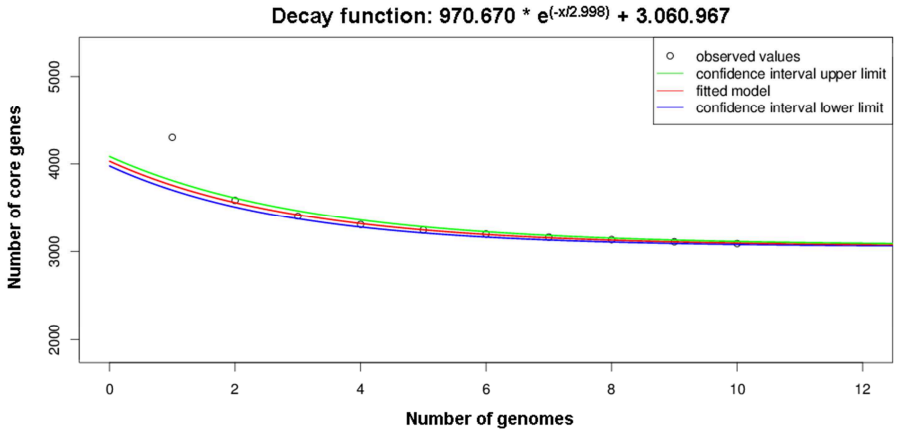


Figure 17 Trend of the core genome size of *Y. enterocolitica*.
The number of genes is plotted as mean function of the number of genomic sequences (x) added.

4.2.4. Strain-specific genes and clusters

An indicative parameter about the genetic variability of a species is the set of singletons unique per genome (genes without any hit against any other genome). As shown in Figure 19, *Y. enterocolitica* bioserotypes 4/O:3 and 3/O:9 strains do not greatly participate in the species variability, while the majority of new discovered genes come from strains of biotypes 1A and 1B. The exception of strain W22703 may be explained by the high number of CDSs predicted in this genome (Table 24). Interestingly, the number of singletons corresponds to the phylogenetic relationships between the analyzed strains (section 4.2.5), in that closely-related strains, such as the ones belonging to serotype O:3, have a low number of singletons and, therefore, lower genetic diversity.

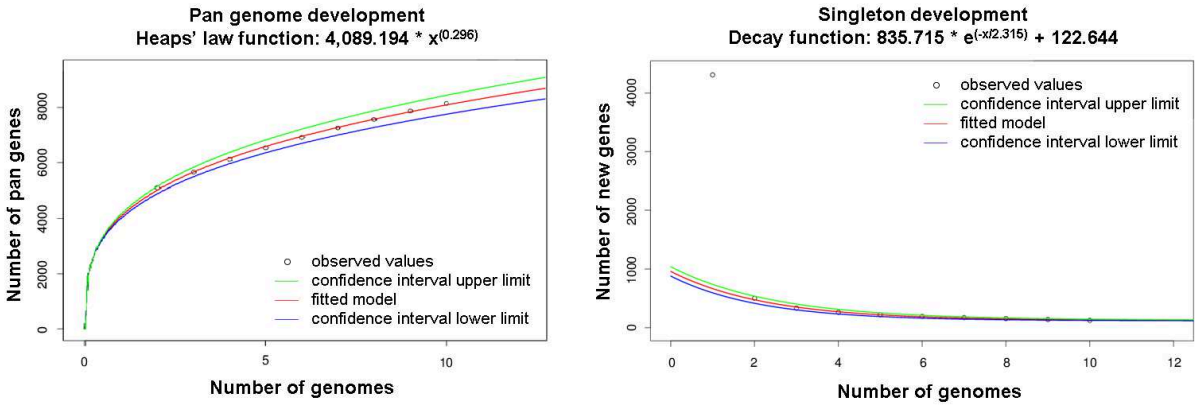


Figure 18 *Y. enterocolitica* pan-genome and singleton development plots.
Graphs show the accumulated number (mean) of pan and new genes against the number of genomes (x).

The unique genes detected in the analyzed 10 genomes mainly encode hypothetical proteins or are carried by mobile regions. In particular, 100% of the singletons detected in the three YEP strains Y8265, Y5307 and 105.5R(r) encode hypothetical proteins. Only one gene exclusive for strain Y11, found within the PhiYep3 phage, is assigned to a specific feature (locus tag: Y11_13111, NAD-dependent DNA ligase), while the other 24 genes are annotated as hypothetical proteins. In strain W22703, 86.5% of unique genes encode hypothetical/unknown proteins. The remaining 13.5% include transposase-encoding genes and, mostly, genes truncated or annotated differently than in the other strains.

As expected by the previous comparison between YE bioserotype 1B/O:8 (section 4.1.1), the genes exclusively found in strains 8081 and WA-314 belong to strain-specific prophages and genomic islands (e.g. YAPI in strain 8081). The non-virulent human isolate NF-O possesses 87.3% of unique genes carried by phages or insertion sequences or encoding hypothetical

proteins. Interestingly, 7 adjacent genes specific for strain NF-O, inserted between the YE2671 and YE2672 genes in strain 8081, encode ABC transporter proteins and an alkylhydroperoxidase-like protein, AhdP family. Performing a Blast search across public DNA sequences, it appears that this gene cluster is not present in genomes of other *Yersinia* strains, but resembles regions in a plasmid from the γ-proteobacterium *Pantoea* sp., in the α-proteobacterium *Mesorhizobium australicum*, and from the β-proteobacterium *Verminephrobacter eiseniae*. This is evidence of a strain-independent acquisition of these genes by lateral gene transfer, probably during co-existence with other microorganisms in the same environment/host. Similarly, the singletons of the environmental YE strain IP2222 encode mainly hypothetical proteins or are found in mobile elements (90%). A 12-CDS region, inserted near a prophage and in the same chromosomal location as the YGI-3 in strain 8081, carries proteins homologous to the *srfABC* virulence cluster in *Salmonella* sp. and to ABC-type antimicrobial peptide transporters. Similar clusters have been detected by SEED in other γ-proteobacteria like *Azotobacter vinelandii*, *Pseudomonas* sp. and *Pasteurella* sp., suggesting that this genomic region is an important hot-spot for acquisition of genetic material in YE (section 4.1.1 and Figure 20), with possible gene flows among proteobacteria.

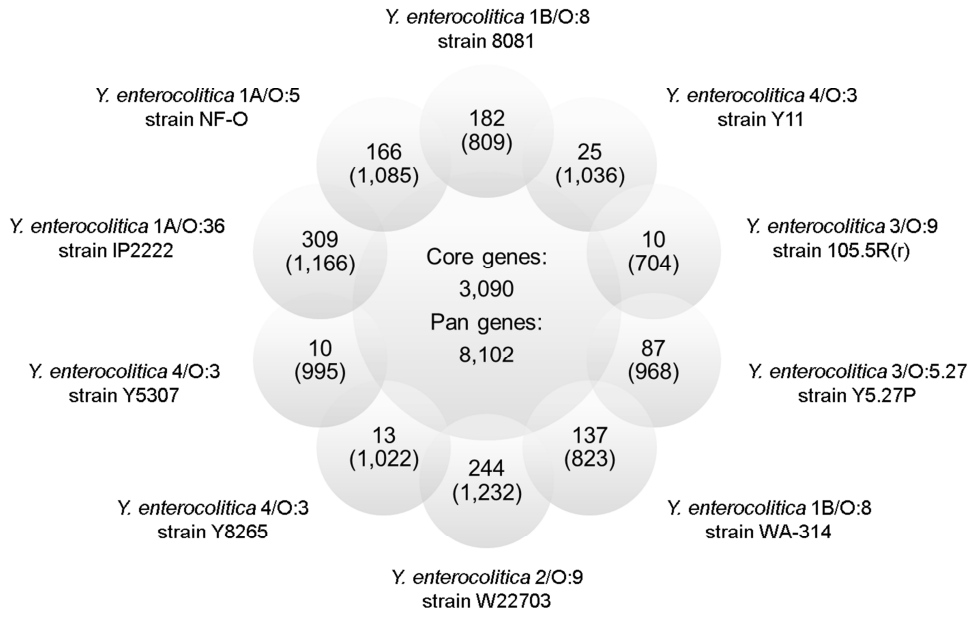


Figure 19 Unique genes in *Y. enterocolitica* genomes.

The chart contains the number of singletons per each genome against all the other genomes; numbers in brackets refer to the singleton genes against the whole core genome.

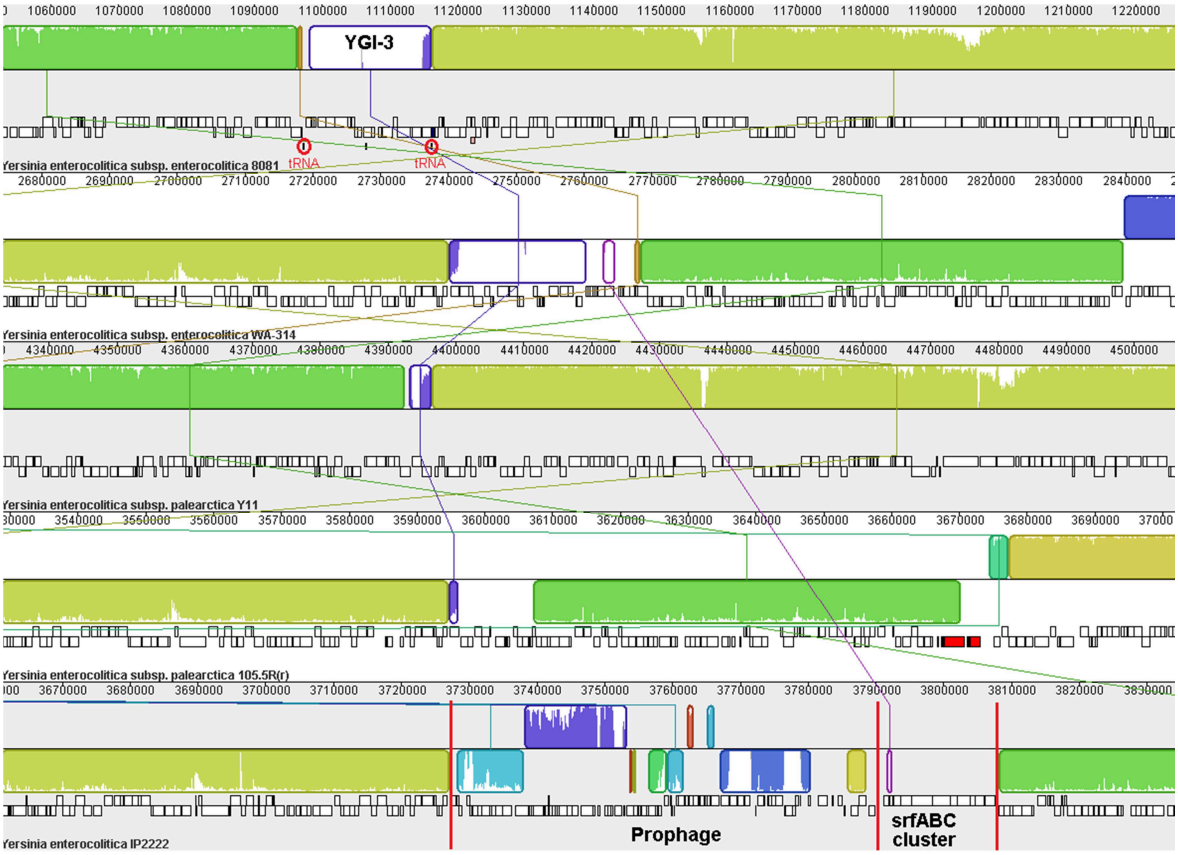


Figure 20 Mauve comparison of the YGI-3 genomic region across five *Y. enterocolitica* isolates.
The YGI-3 is a putative plasmid inserted in strain 8081 between two tRNA genes (between the green and yellow similarity blocks), while in strain IP2222 a large prophage and the *srfABC* cluster replace the island.

4.2.5. Core genome-based phylogeny of *Y. enterocolitica*

To address the phylogeny of the whole species *Y. enterocolitica*, a phylogenetic tree was computed based on the core proteins of the analyzed 10 sequences (section 4.2.1) to maximize the sequence support. In fact, despite the differences in gene prediction algorithms, genes that are well conserved among a species are easier to annotate; therefore, the conserved core genome likely provides sufficient quality to reveal accurate relations among the analyzed strains. The concatenated multiple alignment of the core protein sequences consisted of 1,014,040 aa residues, which were used to construct a neighbor-joining tree (Figure 21). Examining the phylogenetic relationships of the species revealed three main distinct clusters, which reflect the classification in non-virulent, weakly-virulent and mouse-virulent strains, and five different lineages (biotypes 1A and 1B, and serotypes O:9, O:5,27 and O:3). Therefore, for the weakly-virulent group, the phylogeny is congruent with the serotype, suggesting that serotyping might be an accurate criterion for determining grouping and typing of YE. Notably, the serotype O:3

strains are tightly clustered together with short terminal branches compared with the other lineages, indicating low genetic diversity and likely efficient restriction barrier.

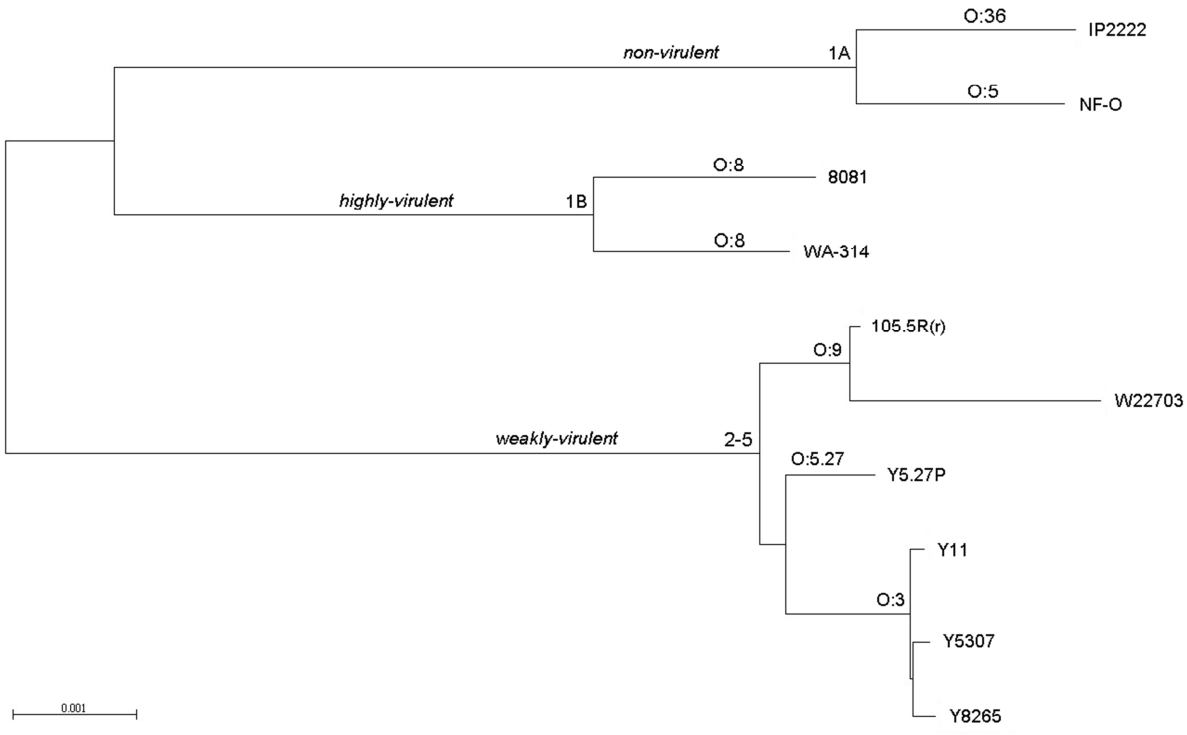


Figure 21 Phylogeny of *Y. enterocolitica*.

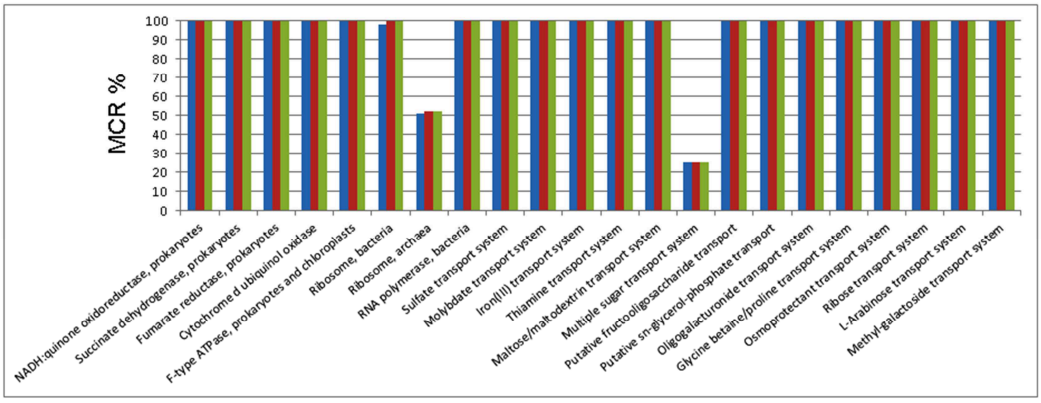
Neighbor-joining tree based on the core genome alignment of the 10 analyzed sequences. Strains are clustered according to their virulence-based classification and to their biotypes/serotypes. Branch lengths are measured in units of the expected fraction of changed amino acids.

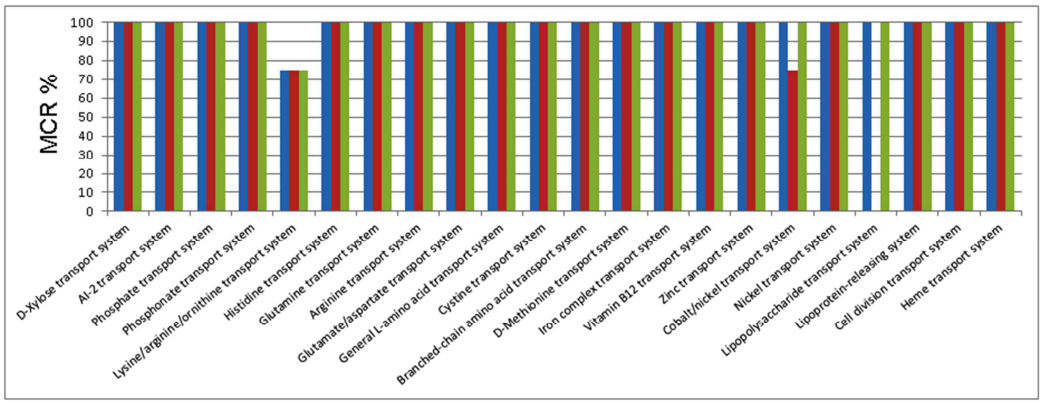
4.2.6. Characterization of the *Y. enterocolitica* functionome

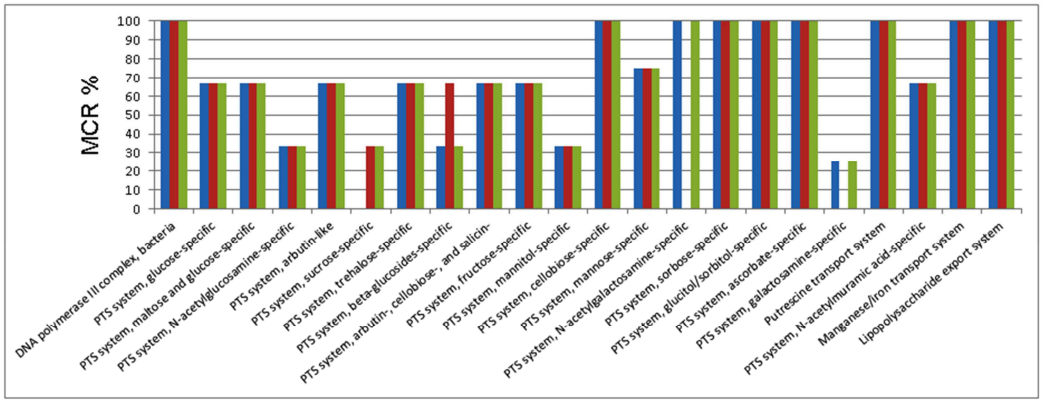
To evaluate the metabolic and physiological potential of *Y. enterocolitica*, the functionomes of strain Y11, 8081 and 105.5R(r) complete genomes, already available as KEGG annotated genomes, were compared by the Metabolic And Physiological potentiaL Evaluator (MAPLE) system. This program is useful to map genes of individual genomes with KEGG-defined functional modules and to calculate module completion ratios (MCRs), corresponding to the percentage of a module completed by the input genes (section 3.7.8). The gene products of the three analyzed strains were assigned to KO identifiers constituting 125 pathway and 87 structural complex modules. Interestingly, no significant differences were identified in terms of pathway modules, while differences in the MCR of some structural complexes appeared (Figure 22). The highly-virulent strain 8081 contains all the three genes which complete the “Erythritol transport system” module, while strains Y11 and 105.5R(r) have a MCR of 0.0% and, therefore,

no genes involved in this biological system. Corroborating the genome-comparison results (section 4.2.1), this module belongs to the monosaccharide ABC transporters composed of genes YE0818, YE0821 and YE0820, which are specifically present in biotypes 1B and 1A strains. Moreover, the "Type II general secretion system" module is complete in strain 8081, while in the two weakly-virulent strains the MCR is 81.8%. This 11-component module corresponds to the *yts1* and *yts2* loci (Table 3), whose genes may be assigned to the same KO identifier, when having the same function; for example, genes YE3349 (in *yts2*) and YE3565 (in *yts1*) are both membrane-associated secretins and belong to KO2453-GspD (general secretion protein D). Strain 8081 harbors both T2SSs, confirming the completeness of the module, while strains Y11 and 105.5R(r) possess only the *yts2* system, and therefore lack the *yts1*-specific GspM (YE3573) and GspH (YE3569) proteins, resulting in a lower MCR for this module. On the other hand, 2 modules are not represented in strain 8081 but are 100.0% complete in strains Y11 and 105.5R(r). Firstly, the "Lipopolysaccharide transport system" complex includes two ABC-2-transporter encoding-genes located within the serotype O:3 and O:9 O-ag clusters, which show high similarity (see section 4.2.7). Secondly, genomic comparison analyses have identified the *agaVWEF* operon (utilization of N-acetyl-galactosamine) as specific YEP acquisition (section 1.5). Accordingly, the MCR for the complex module "PTS system, N-acetylgalactosamine-specific" is 100.0% for the YEP strains Y11 and 105.5R(r) and 0.0% for the YEE strain 8081, which is devoid of this cluster.

To note, the 25% MCR for the module "PTS system, galactosamine-specific" in strains Y11 and 105.5R(r) is due to the presence of one of the four KO identifiers composing this module, KO2744, corresponding to the AgaF transporter, which is also assigned to the N-acetylgalactosamine structural complex. The other three proteins of this module, which is involved in the conversion of galactosamine to galactosamine-6-phosphate, are not encoded by strains Y11 and 105.5R(r), and therefore this module is not functional in any YE strain. The "Cobalt/Nickel transport system" module consists of 4 KO components, corresponding to the *cbiOQNM* genes belonging to the *cob*/*pdu* operon for cobalamin (vitamin B12) biosynthesis in YE (Thomson et al., 2006). In strain 8081 the *cbiN* gene is not annotated, leading to the observed MCR ratio of 75.0%; however, in the other bioserotype 1B/O:8 strain, namely WA-314, the *cbiN* gene is present (locus tag: YWA314_04428), suggesting that this module is complete also in highly-virulent strains. Curiously, the pathway module "Cobalamin biosynthesis" shows a MCR ratio of 87.5% in all three strains, since 2 out of 9 components related to the CobR and CobC proteins are missing. Nevertheless, production of vitamin B12 has been demonstrated in YE strain 8081, when grown in anaerobic conditions (Prentice et al., 2003).


MCR %
100
90
80
70
60
50
40
30
20
10
0
NADH:quinone oxidoreductase, prokaryotes
Succinate dehydrogenase, prokaryotes
Fumarate reductase, prokaryotes
Cytochrome d ubiquinol oxidase
F-type ATPase, prokaryotes and chloroplasts
Ribosome, bacteria
Ribosome, archaea
RNA polymerase, bacteria
Sulfate transport system
Molybdate transport system
Iron(III) transport system
Thiamine transport system
Maltose/maltodextrin transport system
Multiple sugar transport system
Putative fructooligosaccharide transport
Putative sn-glycerol-phosphate transport
Oligogalacturonide transport system
Glycine betaine/proline transport system
Osmoprotectant transport system
Ribose transport system
L-Arabinose transport system
Methyl-galactoside transport system
Y11
8081
105.5R(r)


MCR %
100
90
80
70
60
50
40
30
20
10
0
D-Xylose transport system
AI-2 transport system
Phosphate transport system
Phosphonate transport system
Lysine/arginine/ornithine transport system
Histidine transport system
Glutamine transport system
Arginine transport system
Glutamate/aspartate transport system
General L-amino acid transport system
Cystine transport system
Branched-chain amino acid transport system
D-Methionine transport system
Iron complex transport system
Vitamin B12 transport system
Zinc transport system
Cobalt/nickel transport system
Nickel transport system
Lipopolysaccharide transport system
Lipoprotein-releasing system
Cell division transport system
Heme transport system
Y11
8081
105.5R(r)


MCR %
100
90
80
70
60
50
40
30
20
10
0
DNA polymerase III complex, bacteria
PTS system, glucose-specific
PTS system, maltose and glucose-specific
PTS system, N-acetylglucosamine-specific
PTS system, arbutin-like
PTS system, sucrose-specific
PTS system, trehalose-specific
PTS system, beta-glucosides-specific
PTS system, arbutin-, cellobiose-, and salicin-
PTS system, fructose-specific
PTS system, mannitol-specific
PTS system, cellobiose-specific
PTS system, mannose-specific
PTS system, N-acetylgalactosamine-specific
PTS system, sorbose-specific
PTS system, glucitol/sorbitol-specific
PTS system, ascorbate-specific
PTS system, galactosamine-specific
Putrescine transport system
PTS system, N-acetylmuramic acid-specific
Manganese/iron transport system
Lipopolysaccharide export system
Y11
8081
105.5R(r)

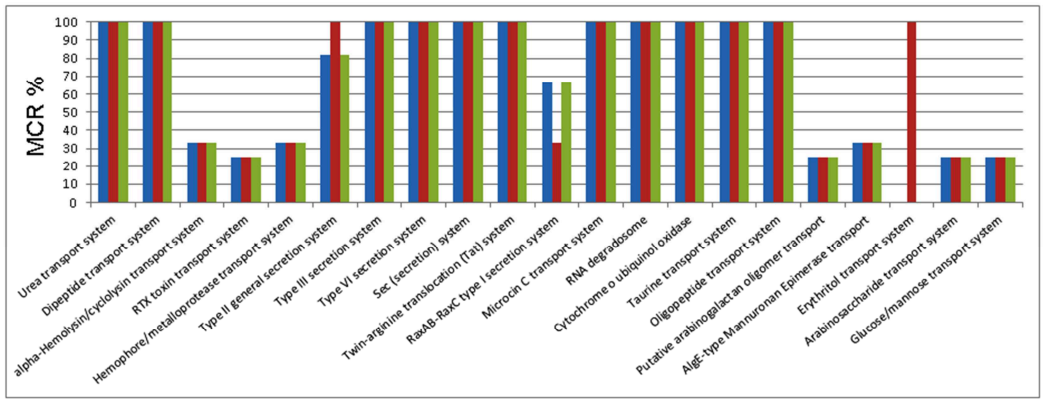

MCR %
100
90
80
70
60
50
40
30
20
10
0
Urea transport system
Dipeptide transport system
alpha-Hemolysin/cyclolysin transport system
RTX toxin transport system
Hemophore/metalloprotease transport system
Type II general secretion system
Type III secretion system
Type VI secretion system
Sec (secretion) system
Twin-arginine translocation (Tat) system
RaxAB-RaxC type I secretion system
Microcin C transport system
RNA degradosome
Cytochrome o ubiquinol oxidase
Taurine transport system
Oligopeptide transport system
Putative arabinogalactan oligomer transport
Alg8-type Mannuronan Epimerase transport
Erythritol transport system
Arabinosaccharide transport system
Glucose/mannose transport system
Y11
8081
105.5R(r)

Figure 22 *Y. enterocolitica* functionome analysis.
Module completion ratios (MCR %) of structural complexes in *Y. enterocolitica* strains Y11, 8081 and 105.5R(r) are compared, showing differences in the metabolic potential of these strains. An MCR of 100% means that all the genetic components of the module were detected in the corresponding genome.

4.2.7. Serotype-specific O-antigen clusters

Serotype is an important characteristic for *Y. enterocolitica* subtyping which is based on the O-antigen entity of the LPS (section 1.1.1 and Figure 1). In this study, availability of whole-genome sequences from YE belonging to the main pathogenic serotypes (O:8, O:3, O:9 and O:5,27) and to two non-virulent ones (O:5 and O:36) allowed identification of so far unknown O-ag cluster genomic locations and genetic structures. In contrast to YE serotype O:8 and O:36, whose O-ag loci are clustered between the *hemH* and *gsk* genes, the O-ag cluster of the other analyzed serotypes is located elsewhere in the genome, while the *hemH-gsk* locus is occupied by the outer core (OC) genes (Figure 23). Interestingly, the OC clusters of serotypes O:3, O:9, O:5,27 and O:5 show a similarity score of 99%.

The O-ag gene region of YE serotype O:9 (locus tags: YE105_C1500-1511) is situated between the *gnd* and *galF* genes, as in *E. coli* and *S. enterica*. The serotype O:3 O-ag gene cluster (locus tags: Y11_16641-16781) lies also in this region, in particular between the *galU* gene and a transposase-encoding gene (Figure 24). The *wbbS-wbbX* locus, which is involved in the synthesis of nucleotide sugar precursors and processing the O-ag, is separated from the glycosyltransferase-encoding genes by a second transposase-gene. The corresponding region in serotype O:8 also harbors several transposase-genes (locus tags: YE2773-YE2777), suggesting probable lateral gene transfer events in this recombinational hot-spot region. This hypothesis is reinforced by the atypical G+C content of the O-ag cluster sequences in serotypes O:3 and O:9 (38% and 35%, respectively, compared to the average YE genomic G+C content of 47%). In contrast to serotypes O:8, O:3 and O:9, whose O-ag gene organization has been already described, nothing has been studied about the O-ag genetic loci of YE serotypes O:5,27 and O:5, even if their structures have been determined (Gorshkova et al., 1986). Comparison of genome sequences discovered different locations for these two clusters, despite their similarity score of 97.3%. The O-ag genes of serotype O:5,27 are situated near a phage-remnant (Figure 24), in a region where also YE serotypes O:8 and O:3 carry transposase- and phage-related genes (locus tags: YE0505-0512 and Y11_37371-37451). Importantly, no gene cluster specific for the factor 27 was identified, as well as its chemical structure was not determined during chemical elucidation of the lipopolysaccharides (Gorshkova et al., 1986). The O-ag gene cluster

of serotype O:5 is inserted between the *gor* and *yapE* genes (locus tags: YE4057-4058). The specific cluster includes also a transposase-encoding gene and part of the *gor* gene (Figure 24); three additional genes, encoding a permease, a DNA-binding protein and a quinone oxidoreductase, are also present in the other non-virulent strain IP2222 but absent from all the other virulent strains. Finally, the so far unknown O-ag cluster of YE serotype O:36 was identified by genomic analysis of strain IP2222. These genes occupy the *hemH-gsk* locus, as in serotype O:8, and encode enzymes participating in the synthesis of sucrose, paratose and tyvelose; additional 4 genes, which are located upstream between the *adk* and the *hemK* genes, encode enzymes involved in mannose metabolism. Curiously, 40% of this cluster is 70% identical to O-ag gene cluster in *Y. pseudotuberculosis* and *Y. pestis* strains.

4.2.8. Patho-serotyping of *Y. enterocolitica*

The genome-wide study of the O-ag genetic clusters of the most common *Y. enterocolitica* serotypes facilitated the identification of serotype-specific genes for developing a serotyping PCR multiplex. Combined with another multiplex for species identification and discrimination between virulent and non-virulent strains, the established PCR-based method could be used for patho-serotyping of YE.

Selection of targets for the identification multiplex 1. To detect YE isolates with an unambiguous PCR method, together with the genetically stable 16S rRNA gene the *inv* gene was chosen, since it has been detected by molecular methods in YE strains with 100% frequency, regardless of isolation source, serotype or pathogenicity degree. For differentiation between virulent (biotypes 1B and 2–5) and non-virulent (biotype 1A) strains, the *ail* and *ystB* genes were added. The chromosomally-encoded *ail* gene has been shown to be a stable marker for virulent strains, while the *ystB* gene is specifically carried by strains of biotype 1A and, therefore, is a distinguishing marker for non-virulent strains (section 1.2.3 and Table 3). Notably, *ystB* is 76.5% identical to the *ystA* gene present in genomes of virulent strains.

Selection of targets for the serotyping multiplex 2. Serotype-specific regions were selected based on the results obtained from the genome comparison of different YE (section 4.2.7). The *rfbC*/*wbbU* gene (locus tag: Y11-16731), previously used for detection of serotype O:3 by a PCR method (Weynants et al., 1996), was confirmed as appropriate serotype O:3-specific target. In fact, no significant hits with genes in publicly available *Yersinia* and no-*Yersinia* genomes were found by BLAST analysis. Similarly, the *per* gene (locus tag: YE105_C1505) was selected as distinctive target for serotype O:9, as described in (Jacobsen et al., 2005). To note, the DNA sequences of *per* genes of YE serotype O:9 and *Brucella* sp. show a similarity score of 64%. In

order to avoid cross-reaction, the designed per_O9_FW/RV primers were subject to BLAST search against *Brucella*, revealing no significant matches. To unambiguously detect strains belonging to serotype O:8, the *rfbC*/*wbcA* gene (locus tag: YE3085) was selected, because of its similar role but low sequence similarity (51.7%) with the orthologous *rfbC*/*wbbU* gene in YE serotype O:3. As preannounced by the results on the genome analyses, target regions carried within the YE serotype O:5,27 O-ag gene cluster but absent from that of serotype O:5 could not be identified. A gene (*rfbE*/*wzt*) was therefore selected as specific amplification target for both serotypes O:5,27 and O:5. The two gene sequences showed 97.2% sequence identity, and no polymorphisms could be used to design primers specific for either serotype according to the chosen criteria (see below). A suitable target for unique detection of serotype O:5,27 was found thanks to the whole-genome comparison, which revealed a restriction modification (RM) system (inserted between locus tags YE4166 and YE4167) absent from serotype O:5, as well as from serotype O:3 and O:9 genomes and from strain 8081. Curiously, it demonstrated 95.2% identity to the RM system detected in strain WA-314 as strain-specific region (section 4.1.1). Hence, this locus could not be applied to the serotyping multiplex, since amplification would occur in strains of serotype O:5,27 and in some strains of serotype O:8. Primers targeting the RM in serotype O:5,27 genomes, however, could be used in a single PCR reaction to differentiate between YE serotypes O:5,27 and O:5 strains giving a positive signal for the *wzt* gene.

Specificity of the PCR assays. Genomic DNAs extracted from six YE reference strains (WA-314, Y11, W22703, Y5.27P, NF-O and IP2222, Table 8) were subject to both multiplexes 1 and 2, using 100 ng of template for all reactions. As shown in Figure 25, the sizes of the amplification products sufficiently differed from each other and allowed a clear discrimination of the bands on an electrophoresis gel, without non-specific amplicons. According to the designed method, multiplex 1 produced different patterns depending on the virulence of the tested strains, in that all YE strains showed a product for the 16S rRNA and the *inv* targets; also, the *ail* amplicon was positive in the virulent strains (WA-314, Y11, W22703 and Y5.27P) but negative in the non-virulent ones (NF-O and IP2222), with opposite results from the *ystB* amplification. In general, testing virulent strains with the developed assay would give pattern inv(+), 16S rRNA(+), ail(+), ystB(-), while pattern inv(+), 16S rRNA(+), ail(-), ystB(+) would result from non-virulent strains. To check that the identification multiplex is species-specific, non-YE and non-*Yersinia* strains were tested (Table 8 and Figure 26). *Y. intermedia* showed a weak band corresponding to the 16S rRNA amplicon, while feeble non-specific bands were present in the reactions from *C. jejuni*, *S. enterica* and *L. pneumophila*; a miss-identification is anyway avoided by the *inv* gene product, which is not amplified in any of the tested species. Lastly, sensitivity tests showed that amplification of all products was obtained with at least 1 ng/µl of DNA

template, while the 16S rRNA amplicon could be detected also at a dilution of 0.01 ng/µl (Figure 27 A).

The serotyping multiplex PCR has been successfully designed for identification of the main YE serotypes (O:3, O:9, O:8, O:5,27) based on the amplification of a single PCR product in each reaction (Figure 25). *Y. enterocolitica* strain IP2222 of serotype O:36, which is not recognized by the assay, gave no amplification. Similarly, negative results would be produced by isolates belonging to serotypes distinct from the targeted ones. Amplification of the wzt_O:5,27 target occurred not only in serotype O:5,27 but, as anticipated, also in serotype O:5. Discrimination between these two strains was achieved by combining the results of the multiplexes 1 and 2. In fact, the non-virulent strain of serotype O:5 had pattern inv(+), 16S rRNA(+), ail(-), ystB(+), wzt(+), while pattern inv(+),16S rRNA(+), ail(+), ystB(-), wzt(+) resulted from the virulent serotype O:5,27 strain. An evident distinction derived by the application of the single PCR targeting the RM system in YE serotype O:5,27, which resulted in amplification of the 627 bp product only in strain Y5.27P. This reaction was also tested on the other reference strains, resulting in a positive amplification only in strain WA-314, as expected.

Validation of the patho-serotyping assay. To test the validity of the developed PCR assays, 133 YE isolates (Table 10), belonging to known or unknown serotypes and originating from different hosts and countries, were analyzed. Phenotypic information about the strains were obtained from the respective supplier institutions upon receipt of the microbiological materials or, in case of the 50 strains from Dr. Tietze, RKI, after completion of the genotypic experiments, in order to perform a blind validation study. PCR mixtures were always prepared using 100 ng of the gDNA templates, as done with the reference strains, so that results were equivalent. Data resulting from the patho-serotyping PCR-based method were therefore compared to the phenotypic data (Table 27).

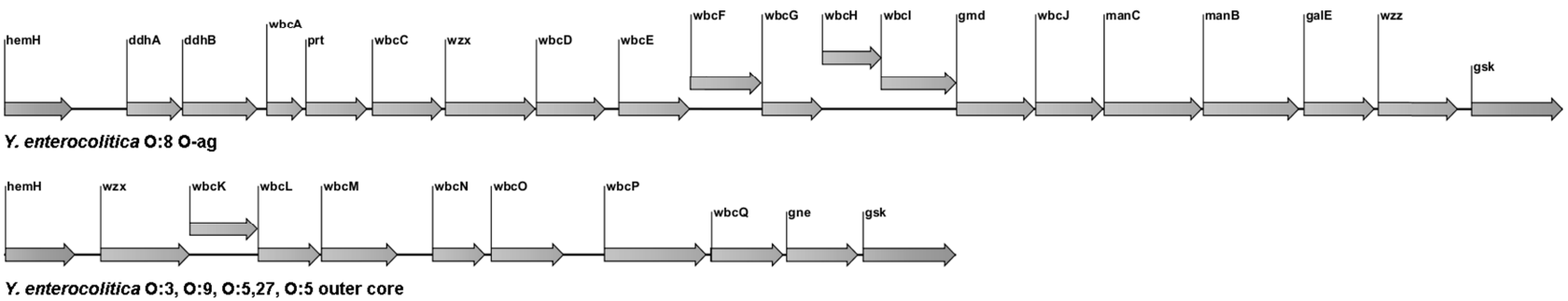


Figure 23 *hemH-gsk* locus in *Y. enterocolitica*.

This locus is occupied by the O-ag cluster in serotype O:8 strains and by the OC cluster in serotypes O:3, O:9, O:5,27 and O:5. Drawn to scale.

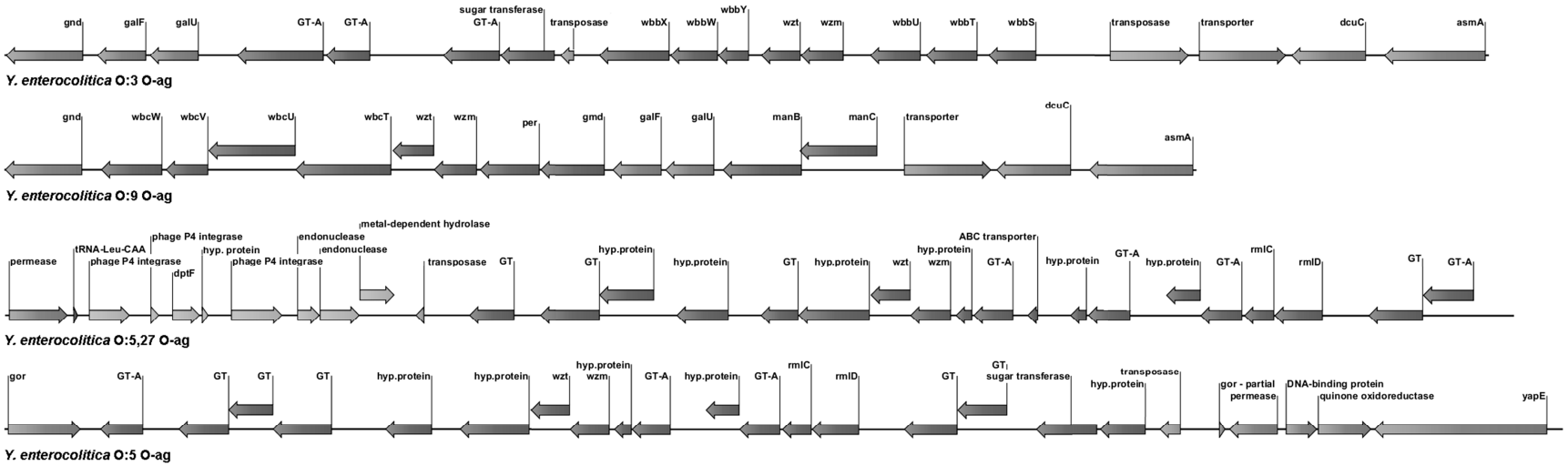


Figure 24 O-ag genetic regions of *Y. enterocolitica* subsp. *palearctica*.

The organization of the O-ag clusters and flanking regions of serotype O:3, O:9, O:5,27 and O:5 are displayed. Drawn to scale. (Garzetti et al., 2014).

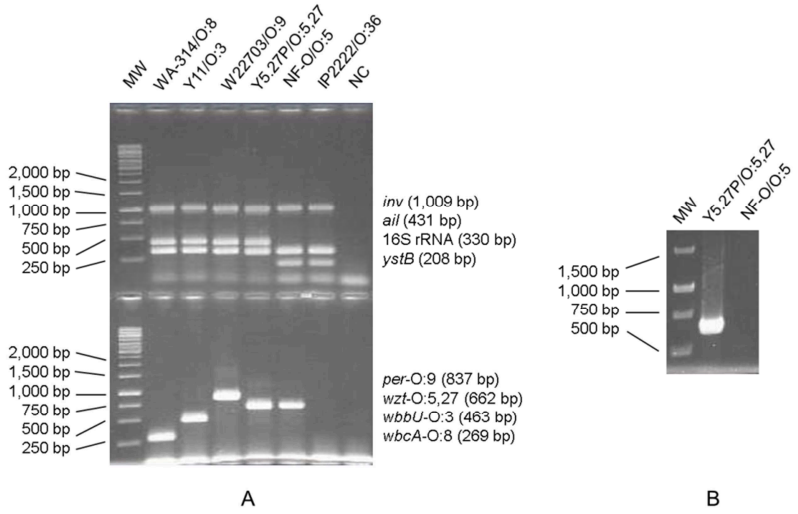


Figure 25 Identification and serotyping of *Y. enterocolitica* reference strains.
Results from the identification and serotyping multiplexes are displayed (A). A negative control (NC, no genomic DNA) was added to confirm absence of non-specific amplification. The amplification of the RM_O:5,27 system in strain Y5.27P was compared to that of strain NF-O (B). MW: molecular weight. Modified from (Garzetti et al., 2014).

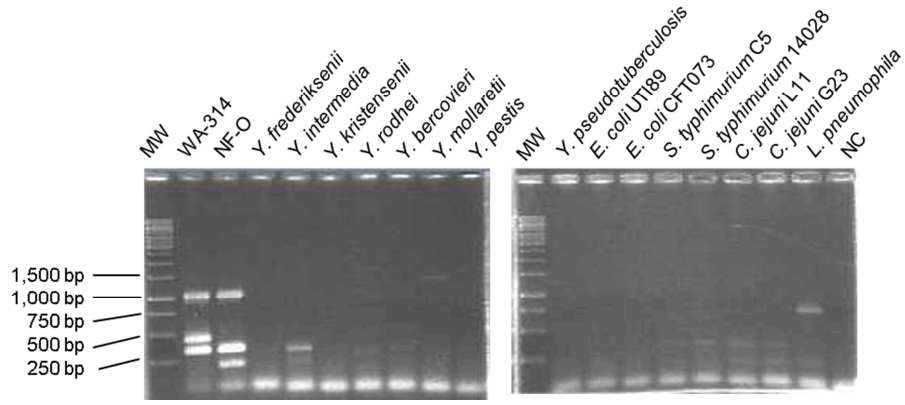


Figure 26 Specificity test of the identification multiplex.
Non-*Y. enterocolitica* and non-*Yersinia* strains were subject to multiplex 1 amplification starting from 100 ng of purified gDNA. *Y. enterocolitica* strains WA-314 and NF-O were used as positive controls.

The identification multiplex produced concordant results in 130 out of 133 YE strains. From strains P10 and P11, two pig-isolates classified as bioserotype 4/O:3 by conventional methods, no amplification was detected, while the 16S rRNA fragment was weakly amplified from strain 12-03803, typed as rough biotype 3. The absence of the *inv* amplicon suggested that these 3 strains do not belong to the species YE; therefore, the 16S rRNA gene was sequenced with the

universal primers 16S_FW_D1,D2 and 16S_RV_P1 for bacterial identification. Blast analysis of the sequences from strains P10 and P11 gave significant hits with *Y. frederiksenii* 16S rRNA, while strain 12-03803 sequence was identical to the one from *Y. kristensenii*. Accordingly, *Y. frederiksenii* isolates belonging to serotype O:3 have been reported (Aleksic et al., 1986). Since the discordant strains were found to be non-YE, the identification multiplex could be considered 100% specific.

Results from molecular and conventional serotyping methods were concordant in 126 out of 133 examined strains. Since a completely negative amplification was again obtained from strains P10 and P11, they were re-serotyped by phenotypic methods, which confirmed their positive agglutination to serotype O:3 antiserum. Together with the results from the 16S rRNA sequencing, this suggests that a dissimilar sequence of the wbbU gene is present in the O-ag gene cluster of these YE-like O:3 strains. No amplification was also observed in strain 12-01557, of bioserotype 1A/O:8. Having pattern inv(+), 16S rRNA(+), ail(–), ystB(+), this strain was genotyped as YE biotype 1A, probably belonging to serotype O:7,8, which is rarely found because of the difficult access to the corresponding antiserum. Accordingly, re-serotyping could not confirm the previously determined serotype O:8 and classified this strain as rough. Strain 292, conventionally designated as serotype O:27, gave negative results from both multiplex 2 and the O:5,27-specific PCR. Seeing that the identification multiplex for this strain resulted in pattern inv(+), 16S rRNA(+), ail(–), ystB(+), it was classified as non-virulent (biotype 1A), consistently with re-serotyping, that produced no reaction with known anti-sera (O:3, O:9, O:8, O:5, O:27). Bioserotype 1A/O:7,8 strain 39 gave a 1A/O:8-like pattern, namely inv(+), 16S rRNA(+), ail(-), ystB(+), wbcA-O:8 (+). Moreover, amplification of the wbbU-O:3 product resulted from strains of serotypes O:1,2a,3 and O:2a,2b,3 (522 and 523), leading to an inaccurate genotyping of these rare strains as virulent YE serotype O:3. YE strains of serotypes O:13 (58), O:20 (59) and O:21 (60), which were found positive for the *ail* target but negative for any product in the serotyping multiplex, were genotyped as virulent YE.

The third O:5,27-specific PCR was tested on 29 YE strains belonging to either serotypes O:5 or O:5,27 (Table 25). Considering that strain 292 turned out to be not-typeable, the selected RM cluster was detected with 100% prevalence in all serotype O:5.27 strains, while no reaction was produced from serotype O:5 strains, confirming its specificity.

A total number of 19 rough (14) and unknown (5) YE strains were tested during the validation process. Except for strain 12-03803, which turned out to belong to the species *Y. kristensenii*, the patho-serotyping assay confirmed the identification as YE of the other 18 strains. In particular, 8 strains (08-00920, 11-07058, 12-00821, 12-03266, 12-03288, 12-03324, 12-03401,

12 04470), producing negative results from the serotyping multiplex, were genotyped as biotype 1A; 5 rough strains (06-03972, 06-04957, 07-05684, 09-05259, 09-08364) were serotyped by our molecular assay as YE serotype O:9; 3 (82, 89, 93) and 2 (260, 282) unknown strains were typed as YE serotype O:3 and biotype 1A, respectively.

Colony PCR experiment. To check the feasibility of a faster patho-serotyping assay, colony PCR protocols were applied, eliminating the gDNA extraction step (section 3.4.1). Satisfactory results were obtained with both standard and NaOH-based methods for the identification multiplex, while non-specific products were occasionally amplified in the serotyping multiplex. The expected bands were anyway brighter than the non-specific ones and allowed identification of the serotype (Figure 27 B and C).

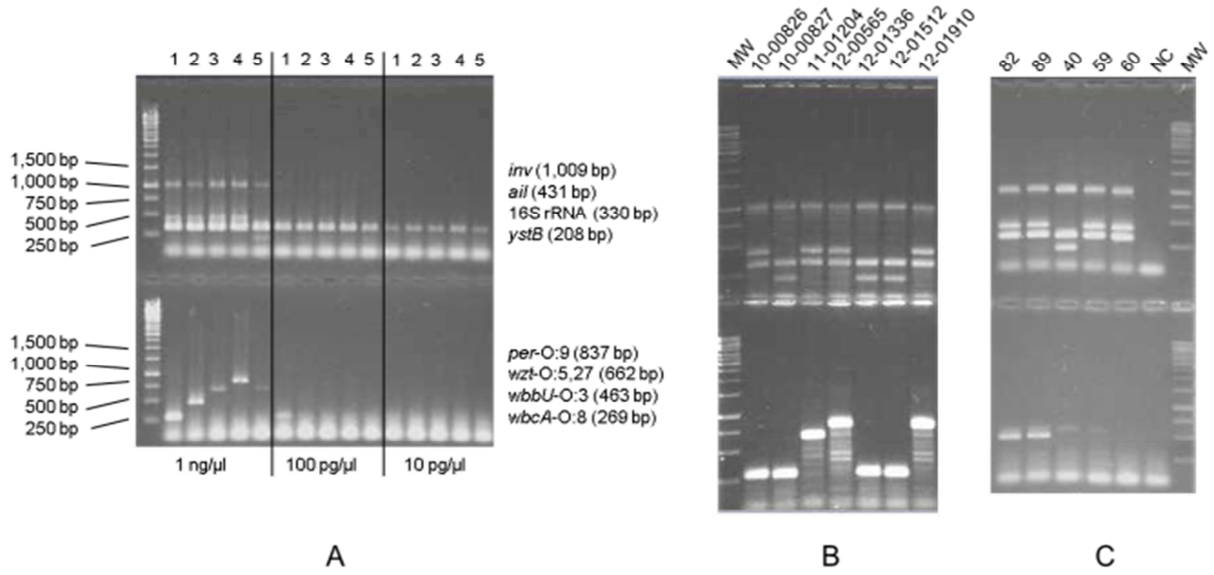


Figure 27 Sensitivity of the patho-serotyping assay and colony PCR experiments.
Genomic DNA samples extracted from five *Y. enterocolitica* strains (1: WA-314; 2: Y11, 3: Y5.27P; 4: W22703; 5: NF-O) were diluted to 1, 0.1 and 0.01 ng/µl and used as template for both identification and serotyping multiplexes. The detection limit of the test resulted to be 1 ng/µl (A). Random samples were processed according to colony PCR standard (B) or NaOH-based (C) methods.

Table 27 Validation of the *Y. enterocolitica* patho-serotyping method.

Results from the 133 tested strains are summarized as positive (+) or negative (-) amplification for each product and classified according to their phenotype. Discordant results between phenotypic and genotypic methods are marked by *. Modified from (Garzetti et al., 2014).

Strain count	Multiplex PCR assays									Identification and serotyping	
	inv	*ail*	16S rRNA	*ystB*	*wbcA* O:8	*wbbU* O:3	*wzt* O:5,27	*per* O:9	RM O:5,27	Phenotypic methods	Genotypic methods
31	+	+	+	-	-	+	-	-		*Y. enterocolitica* 4/O:3	*Y. enterocolitica* O:3
1	+	+	+	-	-	+	-	-		*Y. enterocolitica* 3/O:3	*Y. enterocolitica* O:3
16	+	+	+	-	-	+	-	-		*Y. enterocolitica* O:3	*Y. enterocolitica* O:3
3	+	+	+	-	-	+	-	-		*Y. enterocolitica* unknown	*Y. enterocolitica* O:3
1*	+	+	+	-	-	+	-	-		*Y. enterocolitica* O:1,2a,3	*Y. enterocolitica* O:3
1*	+	+	+	-	-	+	-	-		*Y. enterocolitica* O:2a,2b,3	*Y. enterocolitica* O:3
8	+	+	+	-	-	-	-	+		*Y. enterocolitica* 3/O:9	*Y. enterocolitica* O:9
1	+	+	+	-	-	-	-	+		*Y. enterocolitica* 2/O:9	*Y. enterocolitica* O:9
1	+	+	+	-	-	-	-	+		*Y. enterocolitica* 4/O:9	*Y. enterocolitica* O:9
1	+	+	+	-	-	-	-	+		*Y. enterocolitica* O:9	*Y. enterocolitica* O:9
5	+	+	+	-	-	-	-	+		*Y. enterocolitica* 3/rough	*Y. enterocolitica* O:9
3	+	+	+	-	+	-	-	-		*Y. enterocolitica* 1B/O:8	*Y. enterocolitica* 1B/O:8
3	+	+	+	-	+	-	-	-		*Y. enterocolitica* O:8	*Y. enterocolitica* 1B/O:8
5	+	-	+	+	+	-	-	-		*Y. enterocolitica* 1A/O:8	*Y. enterocolitica* 1A/O:8
1*	+	-	+	+	+	-	-	-		*Y. enterocolitica* 1A/O:7,8	*Y. enterocolitica* 1A/O:8
6	+	+	+	-	-	-	+	-	+	*Y. enterocolitica* 3/O:5,27	*Y. enterocolitica* O:5,27
10	+	+	+	-	-	-	+	-	+	*Y. enterocolitica* 2-3/O:5,27	*Y. enterocolitica* O:5,27
2	+	+	+	-	-	-	+	-	+	*Y. enterocolitica* 4/O:5,27	*Y. enterocolitica* O:5,27
10	+	-	+	+	-	-	+	-	-	*Y. enterocolitica* 1A/O:5	*Y. enterocolitica* 1A/O:5
1*	+	-	+	+	-	-	-	-	-	*Y. enterocolitica* O:27	*Y. enterocolitica* 1A
8	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/rough	*Y. enterocolitica* 1A
2	+	-	+	+	-	-	-	-		*Y. enterocolitica* unknown	*Y. enterocolitica* 1A
1	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/O:8	*Y. enterocolitica* 1A
1	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/O:36	*Y. enterocolitica* 1A
1	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/O:6,30	*Y. enterocolitica* 1A
1	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/O:4,33	*Y. enterocolitica* 1A
1	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/O:10	*Y. enterocolitica* 1A
1	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/O:48	*Y. enterocolitica* 1A
1	+	-	+	+	-	-	-	-		*Y. enterocolitica* 1A/O:41,43	*Y. enterocolitica* 1A
2*	-	-	-	-	-	-	-	-		*Y. enterocolitica* 4/O:3	*Y. frederiksenii*
1*	-	-	+	-	-	-	-	-		*Y. enterocolitica* 3/rough	*Y. kristensenii*
1	+	+	+	-	-	-	-	-		*Y. enterocolitica* 1B/O:13	*Y. enterocolitica* virulent
1	+	+	+	-	-	-	-	-		*Y. enterocolitica* 1B/O:20	*Y. enterocolitica* virulent
1	+	+	+	-	-	-	-	-		*Y. enterocolitica* 1B/O:21	*Y. enterocolitica* virulent

4.3. Microevolution of *Y. enterocolitica* serotype O:3

The most predominant serotype among *Y. enterocolitica* isolates recovered from humans and animals, especially wild and slaughter pigs in Europe, is serotype O:3 (section 1.4.2). A thoroughly comparison of YE serotype O:3 genome sequences might bring new insights into the epidemiology and evolution of these closely related strains, but such analysis is currently missing. Therefore, to create a representative serotype O:3 population framework, a total number of 20 whole-genome sequences were analyzed in this study, including 7 genomes retrieved from the GenBank public database (Y11, Y8265, Y5307, YE12/03, O3-gb, IP10393, PhRBD_Ye1) and 13 newly-sequenced samples (P1, P4, 149, 150, D1, D2, D3, D4, D5, D6, D7, D8, D9), which were isolated between 1982 and 2011 from all over the world and from different sources (Table 8, Table 20 and Table 29).

4.3.1. Re-sequencing of *Y. enterocolitica* serotype O:3 genomes

Whole-genome sequencing of the new 13 *Y. enterocolitica* strains produced 2.02 to 5.41 million reads, with an average Phred score of 32.4 to 34.9 and an average coverage ranging between 62.5X and 200.3X (Table 28). The differences in the average read length between the first 4 strains and the remaining 9 is due to the Illumina technology adopted from the two sequencing centers (section 3.7.3). Mapping was performed against the chromosome of YE strain Y11, used as reference in all the following analyses. The unmapped reads were aligned to the pYV sequence of strain Y11, confirming presence or absence of the pYV plasmid in the samples, previously detected by PCR with primers pYVO3_6900_FW and pYVO3_7130_RV.

4.3.2. *Y. enterocolitica* serotype O:3 genome features

Genome features of the 20 *Y. enterocolitica* strains included in the study, representing the ecological and geographical diversity of serotype O:3, are summarized in Table 29. The genomes have an average sequence length of 4,523,761 bp, with the highly fragmented sequence of strain PhRBD_Ye1 representing the smallest chromosome and the complete genome of strain Y11 the largest one. Interestingly, despite the genome size close to the average, strain YE12/03 possesses the lowest number of CDSs (3,893), and the highest G+C content (48.22%). Strains PhRBD_Ye1 and O3-gb also have a low number of CDSs (3,926 and 4,098, respectively), in contrast to the others, which have approximately 4,200-4,400 genes. The newly-sequenced genomes have the same number of tRNAs as the reference strain Y11, as consequence of the mapping assembly process. Among them, the Japanese strain D6 shows the smallest chromosome, suggesting high diversity or genome loss in relation to the German

reference strain. Performing a *de novo* assembly of strain D6 reads (section 3.7.4) resulted in 220 contigs with a total genome length of 4,513,546 bp. Subtracting the 72,389 bp of the pYV plasmid, the genome size is similar to the one obtained by read mapping, excluding the high diversity supposition. Together with the small size of the Philippines strain PhRBD_Ye1, this may indicate a geographical-dependent genomic decay/gain of serotype O:3 isolates.

Table 28 Sequencing statistics of the 13 *Y. enterocolitica* genomes obtained in this study. Data refers to the mapping against the chromosome of strain Y11.

Strain	pYV plasmid	Total n° of reads	Average read length	Average Phred score	Mapped reads	Mapped reads %	Average coverage
P1	yes	2,024,398	147.4	32.4	1,946,897	96.2	62.5
P4	no	5,410,706	142.5	33.3	5,293,958	97.8	164.02
149	yes	4,030,740	146.4	33.2	3,860,211	95.8	123.05
150	no	3,326,812	147	32.4	3,239,522	97.4	103.45
D1	yes	2,701,604	250	34.8	2,587,891	95.8	140.08
D2	yes	3,045,848	250	34.7	2,906,181	95.4	156.92
D3	no	3,904,574	250	34.8	3,733,035	95.6	200.3
D4	no	3,224,206	250	34.6	3,108,133	96.4	167.85
D5	no	3,464,498	250	34.9	3,364,572	97.1	181.9
D6	yes	2,617,798	250	34.7	2,466,742	94.2	133.2
D7	yes	2,656,406	250	34.7	2,550,860	96.0	138.17
D8	yes	3,692,876	250	34.5	3,542,267	95.9	191.14
D9	yes	2,608,096	250	34.8	2,511,661	96.3	136.16

4.3.3. *Y. enterocolitica* serotype O:3 core and pan-genomes

Determination of the core and pan-genomes of serotype O:3 strains was done as described in section 4.2.1 for the whole species, using *Y. enterocolitica* strain Y11 as reference genome and excluding the pYV plasmid sequences. The core genome of the 20 serotype O:3 strains is formed by 3,313 CDSs, while calculating it separately on the 10 human and 10 animal isolates produced 3,411 and 3,652 genes, respectively. According to the KAAS metabolic annotation, 2,401 out of 3,313 core genes (72.5%) are associated with a KO term, mainly belonging to metabolism- and transport-related KEGG pathways. Analyzing the core development plot and the corresponding decay function (Figure 28 A) results in continue reduction of the core genes with an increasing number of genomes. An asymptote is reached at 3,209 genes with 110 genomes, in contrast to the core plot of the whole YE species, which achieves a plateau with only 20 genomes (section 4.2.1). The pan-genome of YE serotype O:3, composed of 5,613 CDSs, is remarkably small if compared to the whole species pan-genome of 8,102 genes, calculated on 10 instead of 20 genomes (section 4.2.1). The pan-genome curve increases with the sequential addition of genomes (Figure 28 B), but with a small and decreasing average rate

(96 added genes per genome at the first 5 iterations, 54 at 10, 30 genes at 20, and 13 genes at 50). This suggests that, after the inclusion of ten genomes, incorporating additional samples leads to a minor increase in the pan-genome size. This is supported by the expected number of singletons for each newly sequenced strain, which is predicted to be highly variable when adding 1 to 4 genomes, while it is asymptotic after 17 iterations at a value of 25.85 (Figure 28 C). In fact, singletons mostly encode mobile element-related and hypothetical proteins. YE serotype O:3, therefore, displays a rather close pan-genome, confirming the low genetic diversity of this group.

Table 29 Genome features of the *Y. enterocolitica* serotype O:3 used in this analysis.

Strain	Source	Country	Year	N° contigs	Total bp	G+C %	CDSs	tRNAs
Y11	human	Germany	n.a.	1	4,553,420	47.01	4,349	70
Y8265	human	France	1987	14	4,521,393	47.0	4,296	61
Y5307	human	Finland	1995	18	4,553,105	46.9	4,393	94
IP10393	human	France	1982	12	4,463,212	47.0	4,181	64
YE12/03	human	UK	2003	4	4,527,945	48.22	3,893	72
O3-gb	human	UK	2003	2	4,526,243	46.99	4,098	70
PhRBD_Ye1	swine	Philippines	n.a.	112	4,377,740	46.75	3,926	59
P1	swine	Germany	2003	44	4,536,730	47.0	4,333	70
P4	swine	Germany	2011	57	4,522,265	47.0	4,290	70
149	puppy	Belgium	2003	27	4,545,697	47.0	4,335	70
150	calf	Belgium	n.a.	28	4,544,272	47.0	4,333	70
D1	human	Germany	1998	31	4,547,778	47.02	4,342	70
D2	human	Germany	1998	30	4,547,263	47.02	4,338	70
D3	human	Germany	1994	44	4,521,714	47.02	4,307	70
D4	swine	Belgium	1992	20	4,552,170	47.01	4,338	70
D5	swine	Denmark	1992	50	4,508,397	47.02	4,290	70
D6	human	Japan	n.a.	51	4,471,866	46.97	4,227	70
D7	swine	Germany	2008	22	4,549,711	47.01	4,330	70
D8	swine	Germany	2008	15	4,552,293	47.02	4,344	70
D9	swine	Germany	2009	22	4,552,015	47.01	4,334	70

4.3.4. Whole-genome comparison

Genome sequences of the 20 *Y. enterocolitica* serotype O:3 strains were analyzed by Mauve and BRIG comparisons (section 3.7.7). Confirming the results from the previous analyses (section 4.2), serotype O:3 strains appear genetically closely related, with the main detected genomic differences being situated within or near prophage regions (Figure 29). The Japanese strain D6 suffered the most significant prophage loss, especially in the PhiYep3 and PhiYep2 regions. In particular, PhiYep3 is only partly present, while PhiYep2 is completely not inserted at the corresponding tRNA-Met (locus tag: Y11_t04610), in contrast to the other strains.

Interestingly, strain D5 seems to have lost a 12.2 kb region carrying 17 CDSs (locus tags: Y11_09381-09541), near a transposase-gene (Y11_09361) and the PhiYep1 prophage. The encoded proteins have functions in purine metabolism, transcription regulation, enzymatic hydrolysis and histidine transport. Similarly, strain 150 is devoid of a 8-gene locus (Y11_13551-13621) inserted downstream the prophage PhiYep3, between the tRNAs Y11_t16450-t16440-t16430 and the transposase Y11_13631. The missing encoded proteins are mainly hypothetical, but include also a response regulator UvrY, an alpha-amylase and a biotin synthesis protein.

Notably, the serotype O:3-specific GIYep-01 genomic island, which is inserted in the same chromosomal background of the HPI in serotype O:8 strains, is conserved in all the 20 analyzed strains, despite its ability to excise from the chromosome (Batzilla et al., 2011a). This may suggest that a strong selection pressure has occurred during the evolution of serotype O:3 strains to maintain the GIYep-01 island and spread it through the whole population. In fact, genes encoded by this island include a metallo-beta-lactamase (locus tag: Y11_15111), a two-component response system (Y11_15131-15141), and a putative fimbrial cluster (Y11_15151-15191), which may confer advantages in patho-adaptation.

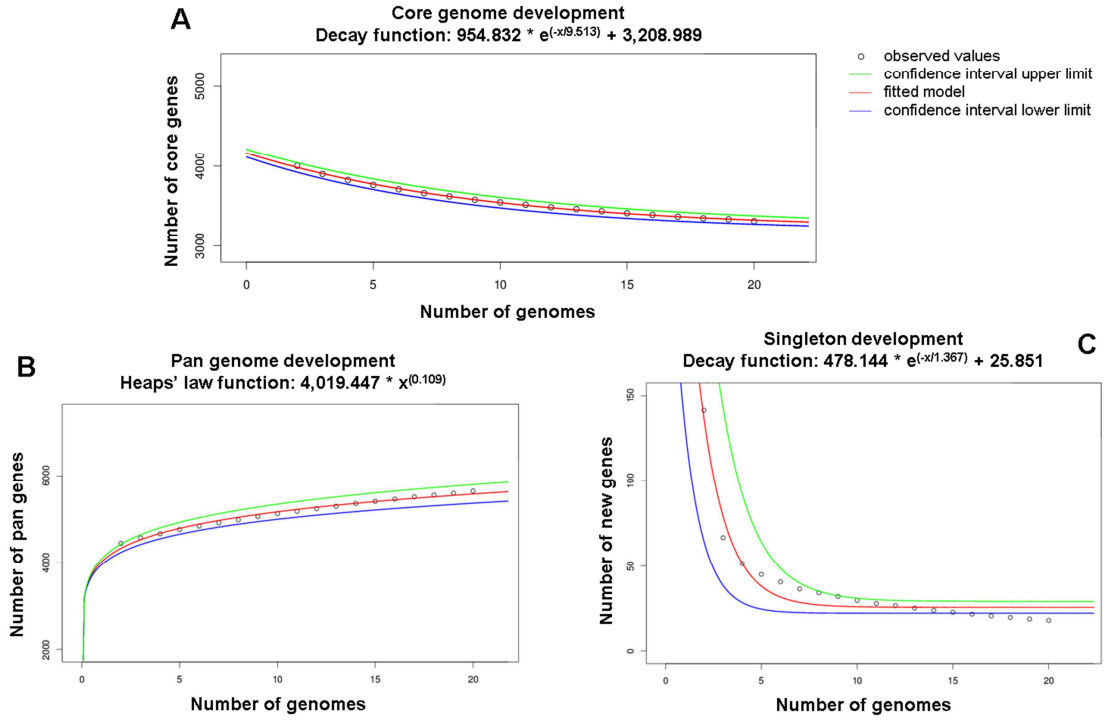


Figure 28 Core, pan-genome and singleton development plots of *Y. enterocolitica* serotype O:3. Mean values for each genome count are shown.

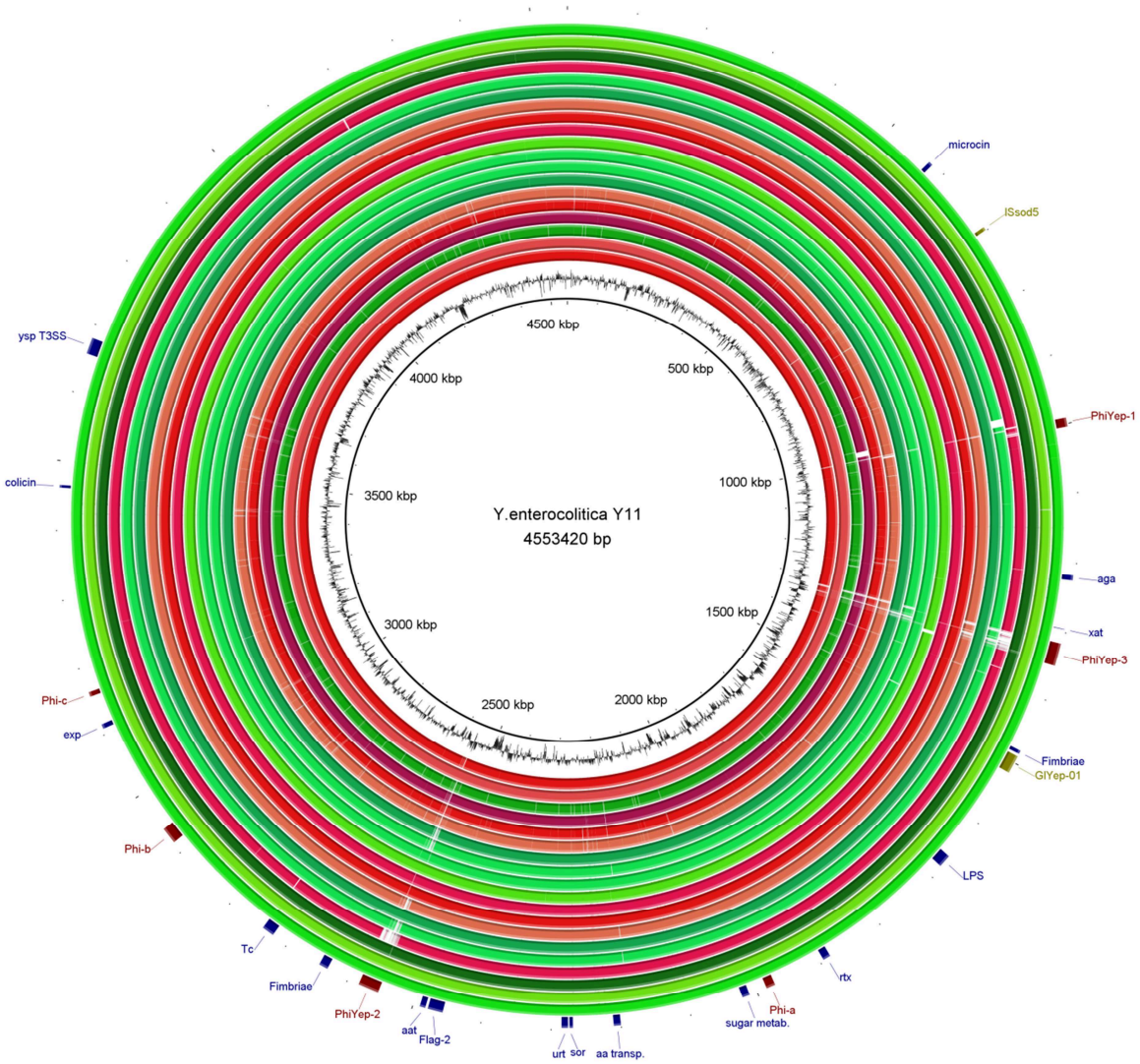


Figure 29 Whole genome comparison of *Y. enterocolitica* serotype O:3.
From the inside out, circles represent G+C content of the reference strain Y11 and comparison of strains YE12/03, O3-gb, PhRBD_Ye1, IP10393, Y8265, Y5307, P1, P4, 149, 150, D1 to D9. The outermost ring shows the main features of serotype O:3 genomes. Strains isolated from humans are drawn in shades of red, while animal-isolates in shades of green.

The low genomic variability in YE serotype O:3 is coupled with a conserved functionome, as calculated by MAPLE (section 3.7.8). Functional comparison among the 13 newly-sequenced strains and the reference strain Y11 showed no differences in the MCR of the 121 pathway, 92 complex and 3 signature modules. A small variation in 1 out of 28 functional sets was detected in strain 150, which has a MCR of 50.0% instead of 100.0% for the module “BarA-UvrY two-component regulatory system”, since the UvrY protein is missing (see above).

Two YE strains, which were isolated from the same human patient with 40 days of difference, namely D1 and D2 (Table 8), were included in the sequencing project to study whether prolonged contact with the host could influence mutations in the stable serotype O:3 genome. Confirming the strong stability of serotype O:3 genomes, no significant variations were found by whole-genome comparison between strains D1 and D2, except for small errors in the annotations, due to the automatic process adopted (section 3.7.6).

4.3.5. *In silico* MLST typing

In order to relate the WGS data with conventional typing scheme, the MLST profile was defined for all 20 *Y. enterocolitica* strains analyzed, using a public online service (section 3.7.2). The *Yersinia* sp. MLST scheme is based on the genetic variability of seven house-keeping genes (*aarF*, *dfp*, *galR*, *glnS*, *hemA*, *rfaE* and *speA*) (Jolley and Maiden, 2010); unique combinations of alleles at each locus specify the 158 *Yersinia* sequence types (ST). Strains P1 and 149 belong to ST23, while all the other eighteen strains are of ST18, with 1 SNP in the *galR* locus for strain Y8265 and 2 SNPs in the loci *rfaE* and *speA* for strain D6. Interestingly, searching the PubMLST database for YE isolates revealed that the majority of serotype O:3 strains belong to ST18, and that ST23 is found only in German isolates. The MLST method, therefore, does not have enough discriminatory power to distinguish among serotype O:3 strains.

4.3.6. Whole genome phylogeny of *Y. enterocolitica* serotype O:3

The core genome calculated on the 20 *Y. enterocolitica* serotype O:3 strains was used to infer the phylogenetic relationship of this group, as previously described (section 3.7.9 and section 4.2.5). The Japanese strain D6 was chosen as outgroup to root the tree, due to its higher divergence to the reference strain Y11, compared to the other strains. The resulting phylogeny (Figure 30) revealed three main lineages (A, B and C) with a total of 14 clusters, corresponding to the terminal nodes. Except for strains D1 and D2, which belong to the same discrete cluster as they were isolated from the same patient, there is no evident clustering according to the isolation source, while the tree seems to be partly congruent with the geographical origin. In fact, the two North-European isolates (D5 and Y5307) are clustered in lineage B together with the two French (Y8265 and IP10393) ones, whereas the three Belgian strains (149, 150 and D4) are all in lineage A. On the contrary, isolates from Germany and Great Britain are found in both lineages. A third group is formed by the Japanese strain D6, which is clearly distant from the other analyzed samples, Notably, the two strains belonging to ST23 (P1 and 149) form a discrete cluster at terminal branches.

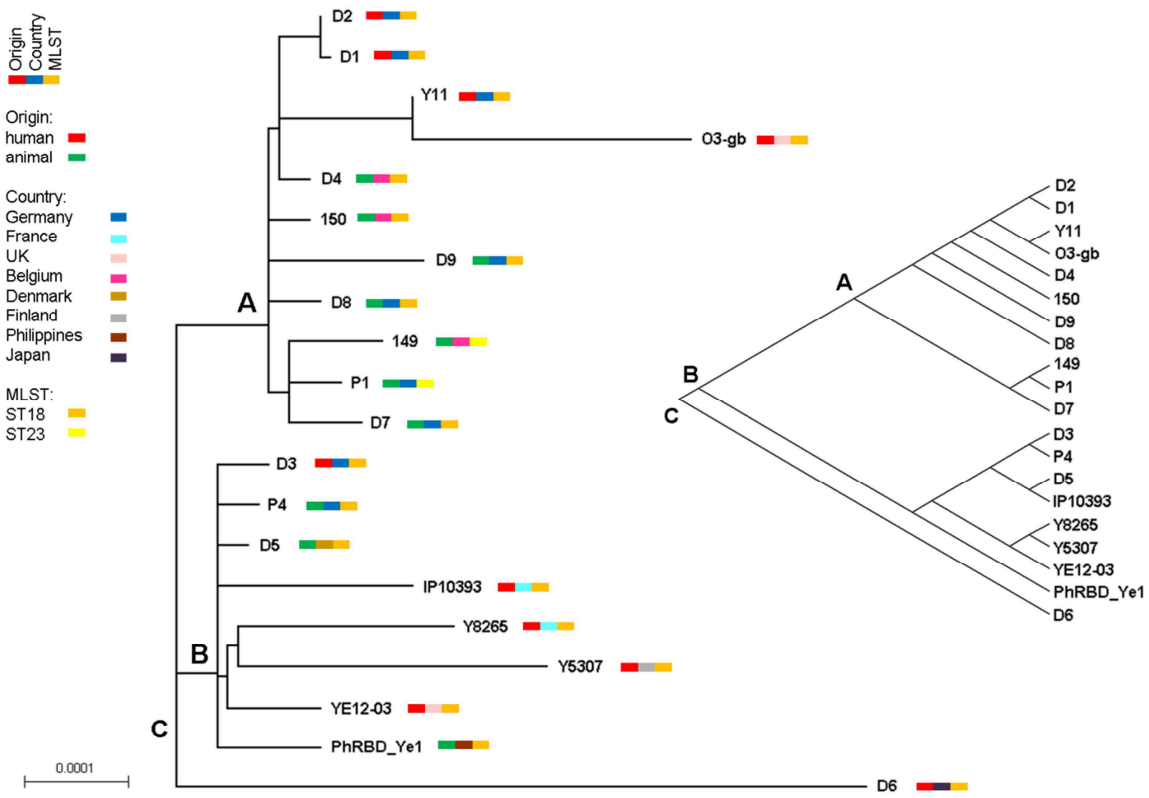


Figure 30 Whole-genome phylogeny of *Y. enterocolitica* serotype O:3.
Neighbor-joining phylogram based on core genes, rooted with the Japanese strain D6 as outgroup. Information about host origin, country and MLST is given in colored blocks for each strain. Inset shows the same tree as slanted cladogram, to clearer demonstrate the relationships between tree leaves and nodes.

4.3.7. Variant-based evolution of *Y. enterocolitica* serotype O:3

Variants in *Y. enterocolitica* serotype O:3 were detected from a total of 19 strains using the complete genome sequence of strain Y11 as reference (section 3.7.5). Initially, 3,256 variants were called; after filtering, e.g. excluding variants within repetitive components of the genome (phages, insertion sequences and rRNAs) and with low quality, the final total number of high-quality variants in the non-repetitive genome resulted to be 2,758 (Table 30 and Table 31). Considering the whole genome (4.55 Mb) and the core genome (3.23 Mb) sizes of strain Y11, the variant ratio is 1 per 1.65 kb (61.3 per 100 kb) and 1 per 1.2 kb (85.4 per 100 kb), respectively. Importantly, these values well correlate with the level of genomic sequence diversity of bacterial pathogens considered as “genetically monomorphic”, which have between 226 and 7 SNPs per 100 kb (Achtman, 2008, 2012).

Table 30 General information of called variants in *Y. enterocolitica* serotype O:3.

	Number	Frequency
Total Variants	2,758	100
SNPs	2,607	94.53
Deletions	41	1.49
Insertions	110	3.99
Intergenic	621	22.51
Synonymous	759	27.52
Nonsynonymous	1,262	45.76
Nonsense	53	1.92
Frameshift	62	2.25
In tRNAs	1	0.04
In CDSs	2,136	77.45
In core CDSs	1,795	65.08

Adaptive selection. Approximately 77.5% of variants occur within protein-coding sequences, and almost 50% cause modifications (nonsynonymous, nonsense, frameshift) within the affected protein sequences. Overall, 1,457 genes (32.8%) contain at least one variant, including 1 tRNA-Leu gene and 1,456 CDSs, while the remaining 67.2% of genes are not affected. The distribution of variants per gene is shown in Figure 31 and follows a Poisson distribution in the range of 0-6 variants per gene. Curiously, there is one gene which deviates from the model, as it contains 22 SNPs, namely Y11_15631, encoding a hypothetical protein. Y11_15631 is a large gene (2,744 bp), which may explain the high variability, and only 3 of the 22 SNPs are nonsynonymous, occurring singularly in strain Y5307. To note, analysis with CDD identified phage-baseplate assembly domains within the amino acid sequence, even though it is not located within a prophage region nor adjacent genes encode phage-related proteins. Therefore, the high number of SNPs cannot be considered a signal of diversifying selection in this gene. Homoplasic variants which occurred multiple times independently in different lineages may also be evidence of selection. To identify potential evidences of adaptive selection in localized genetic regions, clusters of nonsynonymous variants located within five amino acids of the encoded proteins were investigated. Nine genes satisfied these criteria, containing a total number of 30 nonsynonymous variants (1 insertion and 29 SNPs) (Table 32). In particular, two SNPs occur in the same codon of the Y11_26741 protein in two different strains, D8 (A335G) and D6 (T334C), causing the amino acid changes Tyr112Cys and Tyr112His, respectively. Codon Glu39 of Y11_15081 contains two SNPs, which cause change Glu39Val in strain D6 and a nonsense mutation in strain PhRBD_Ye1. Moreover, gene Y11_38851 is subject to two nearby

mutations, Ser41Arg and Asp44Asn in the same isolate D6. Except for gene Y11_15081, encoding a hypothetical protein, the other eight CDSs belong to the core genome of serotype O:3 isolates and encode proteins involved in metabolic and transport functions.

Table 31 Included and excluded regions from variant detection analysis.
Data refers to the reference genome of *Y. enterocolitica* strain Y11.

Genes	**Excluded**	**Included**	**Total**	**Included (%)**
rRNA	22	0	22	0.0
tRNA	4	66	70	94.3
CDS	391	3,958	4,349	91.0
IS elements	107	0	107	0.0
Phages	284	0	284	0.0
Total genes	417	4,024	4,441	90.6
Genomic bp	**Excluded bp**	**Included bp**	**Total bp**	**Included bp (%)**
rRNA	32,370	0	32,370	0.0
tRNA	315	5,121	5,435	94.2
CDS	268,637	3,569,249	3,837,886	93.0
IS elements	86,457	0	86,457	0.0
Phages	228,841	0	228,841	0.0
Intergenic	46,662	631,067	677,729	93.1
Total genome	347,983	4,205,437	4,553,420	92.4

Functional gene loss. A total number of 132 variants causing gene-inactivating mutations, like nonsense, frameshifts and truncations, were identified. They consist of 69 SNPs, 41 insertions and 22 deletions, and are found in 1 genome (78 variants), 2-4 (15), 8-13 (33) and more than 14 sequences (6). The 127 affected genes, including 69 core genes, can then be considered as pseudogenes, and constitute approximately 2.9% of YE serotype O:3 CDSs, which is a high percentage, if compared to *S. typhimurium* (0.9%) and *E. coli* strain K12 (0.7%) (McClelland et al., 2001). Importantly, high pseudogene frequencies have been associated with bacterial organisms with a restricted range of hosts, including *Yersinia* sp. (Parkhill et al., 2003; Thomson et al., 2006).

4.3.8. SNP-based phylogeny

Using SNPs called in the nonrepetitive component of the genome against the reference strain Y11, a phylogenetic tree was reconstructed to study the microevolution of *Y. enterocolitica* serotype O:3 (section 3.7.9). Insertions and deletions were excluded from the analysis, since the concordance of indels called from different pipelines is low (O'Rawe et al., 2013), and the analyzed genomes were not all obtained from the same sequencing technology. Overall, SNPs traced a similar phylogeny (Figure 32) to the one defined by the core-genome-based tree (Figure 30), recognizing the three main lineages A, B and C and 15 clusters, with the two ST23 strains

forming one cluster and the other ST18 strains the remaining ones. SNP-based phylogeny, therefore, shows a discriminatory power to the strain level higher than the MLST scheme, being useful for typing of YE serotype O:3. The Japanese strain D6 looks even more distantly related than the other samples; in fact, 1,116 SNPs out of 2,607 (42.8%) are singularly carried by this strain. A sub-grouping according to new or downloaded sequences appears, probably due to the different methods applied for variant calling on already assembled and re-sequenced genomes (section 3.7.6). Interestingly, the English strain O3-gb contains only 2 SNPs and is, therefore, strictly clustered with the German reference strain Y11 furthest from the root, indicating recent genetic diversification. In contrast, isolate 150 forms a discrete cluster on a branch closest to the root, compared to the previous whole-genome phylogeny. The two strains D1 and D2, which have been isolated from the same patient, are clustered together, as expected. Considering the 10 isolates from animals and the 10 isolates from humans, no detected SNPs are able to distinguish between the two groups, as well as no SNPs are shared by all 8 isolates from pigs. Moreover, the human isolates are mostly located on distal branches, indicating recent mutations, while the animal samples, especially from swine, mainly lie on branches closest to the root with older mutations, in both core-genome- and SNP-based trees. If pig carriers constitute the reservoir for YE serotype O:3, this may account for the different SNP-based phylogenetic ages observed in strains of human and animal origin.

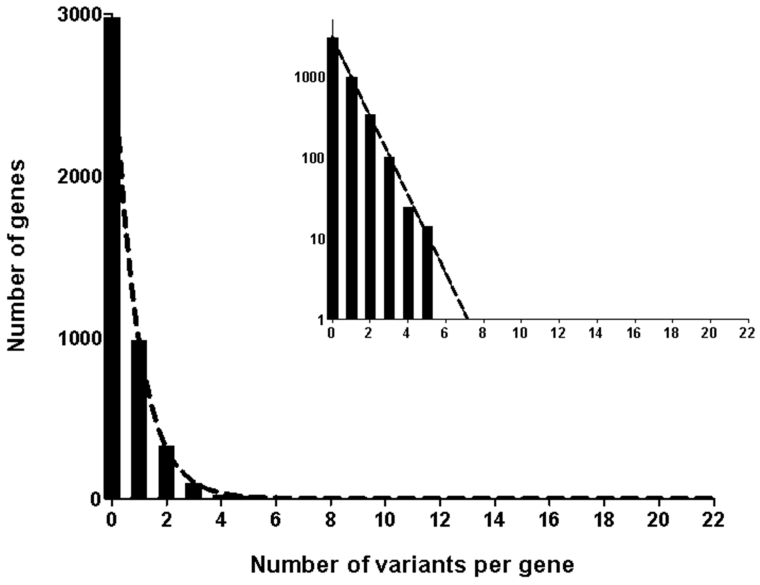


Figure 31 Distribution of serotype O:3 variants per gene.
Inset graph shows gene number on a log scale. Lines indicate the Poisson distribution fitted to the data.

Table 32 Genes with potential signal of adaptive selection.

(ns: nonsynonymous variants)

Gene Locus	Number of ns	Gene product	Core gene
Y11_05871	4	O-acetylhomoserine sulfhydrylase; O-succinylhomoserine sulfhydrylase	Yes
Y11_09291	3	N-acetylmuramoyl-L-alanine amidase	Yes
Y11_15081	3	Hypothetical protein	No
Y11_15171	4	Putative outer membrane usher protein	Yes
Y11_17061	3	Hydrogenase-4 component F	Yes
Y11_20111	3	Potassium efflux system KefA protein	Yes
Y11_26741	4	L-seryl-tRNA(Sec) selenium transferase-related protein	Yes
Y11_38851	3	3-isopropylmalate dehydrogenase	Yes
Y11_43381	3	Phosphate transport system permease protein PstC	Yes

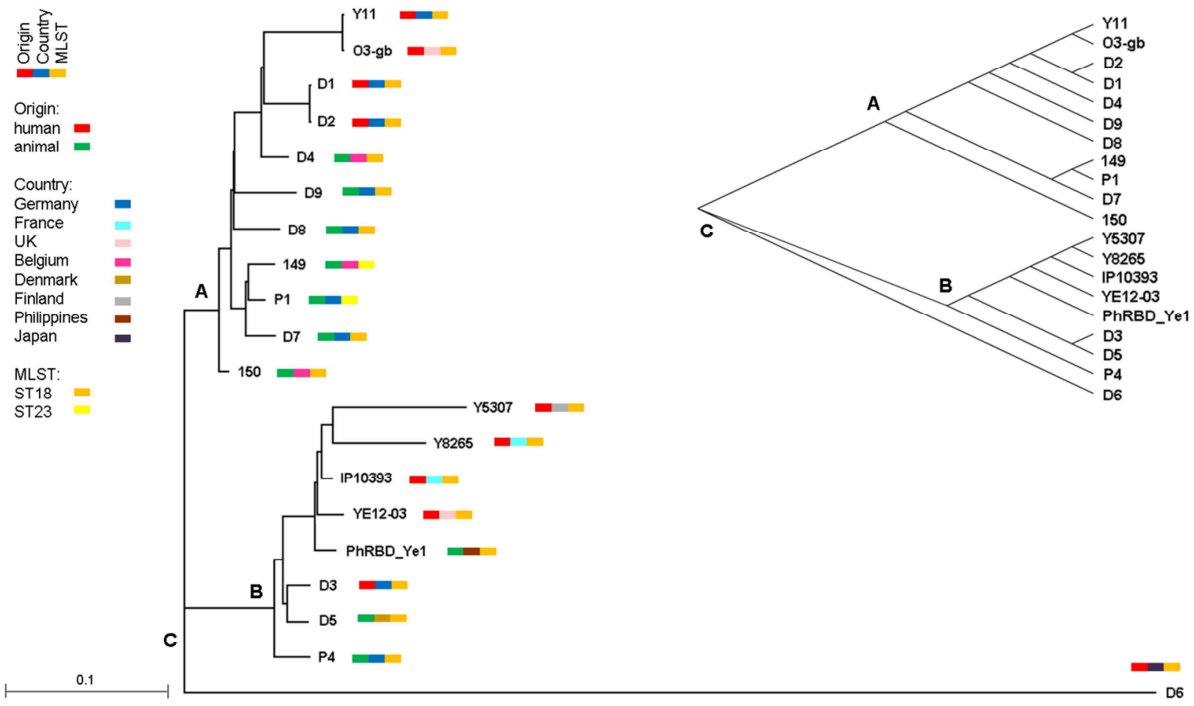


Figure 32 Phylogeny of *Y. enterocolitica* serotype O:3 based on SNPs.

The neighbor-joining tree was constructed from concatenated SNPs for each strain, producing 2,607-bp long sequences. Branch lengths are measured in units of the expected fraction of changed nucleotides, scale as indicated. A slanted cladogram is shown as inset to better reveal the associations among samples. See Figure 30 for comparison.

4.4. Screening of genes expressed during *Y. enterocolitica* O:3 infection

In this study the IVIAT (*in vivo* induced antigen technology) immuno-screening method was applied to strain Y11 to identify bacterial antigens which are expressed during *Y. enterocolitica* serotype O:3 infection and cause a humoral immune response in human patients. Briefly (Figure 33), IVIAT utilizes sera from patients with infection caused by the pathogen of interest. Sera are pooled from several individuals and adsorbed against bacteria grown under standard laboratory culture conditions, in order to remove antibodies that bind bacterial antigens expressed *in vitro*, while retaining antibodies that recognize bacterial antigens specifically produced *in vivo*. Adsorbed serum is then used to probe an inducible protein expression library of the pathogen.

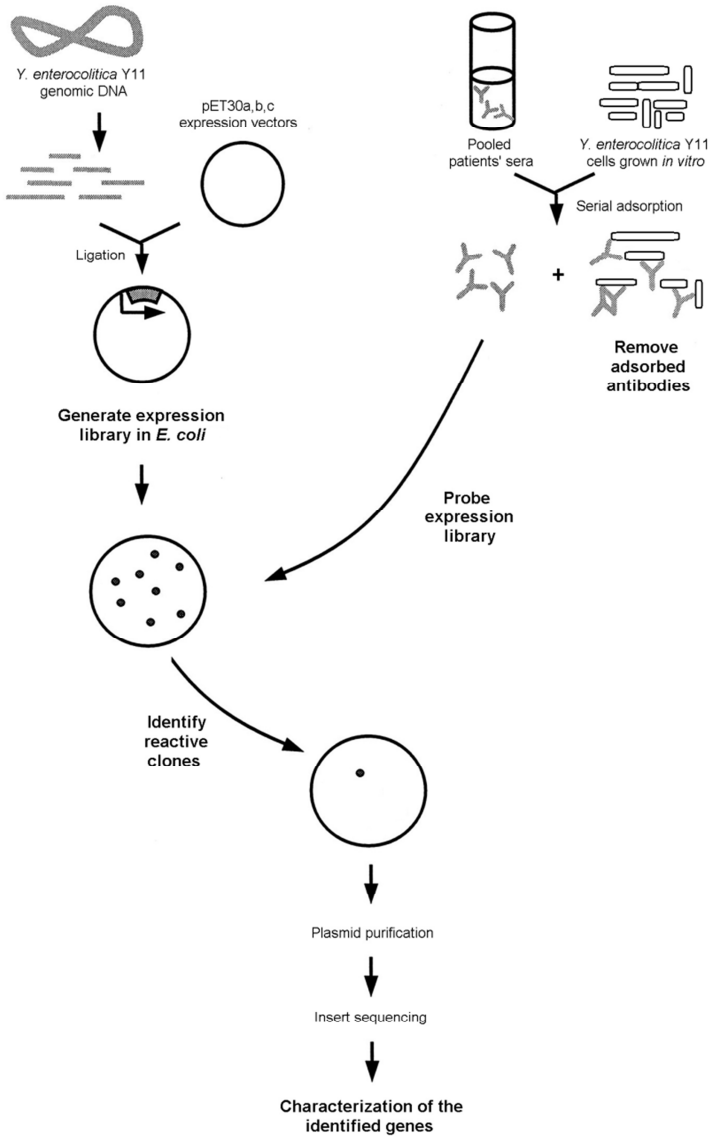


Figure 33 Schematic representation of IVIAT experiment. Modified from (Rollins et al., 2005).

4.4.1. Selection and adsorption of human sera

Sera from 27 human patients with positive *Y. enterocolitica* serotype O:3 infection were collected and tested for positive immune response to antigens YopM, V-Ag, YopD, MyfA, YopE (section 3.5.5). Only 14 sera gave clear and unambiguous results, with significant immune response, and were retained for the following adsorption step (Table 18). Pooled sera were adsorbed against whole cells, native and denatured cell lysates of YE strain Y11 and of the host expression strain *E. coli* T7 express, in order to expose the sera to both surface-exposed and non-exposed proteins synthesized during *in vitro* growth. Multiple rounds of adsorption were conducted and reactivity of the treated sera was compared against the unabsorbed sera by western blot to ascertain the efficiency of the adsorption. As shown in Figure 34, the adsorption steps were successful and antibodies specific to *in vitro* antigens had been completely eliminated. In fact, loss of immuno-reactivity of whole cell and lysate antigens was evident, while immuno-reactivity of known *in vivo* expressed antigens, like the Yop effector proteins (extracted as in section 3.3.7), was nearly preserved after complete adsorption.

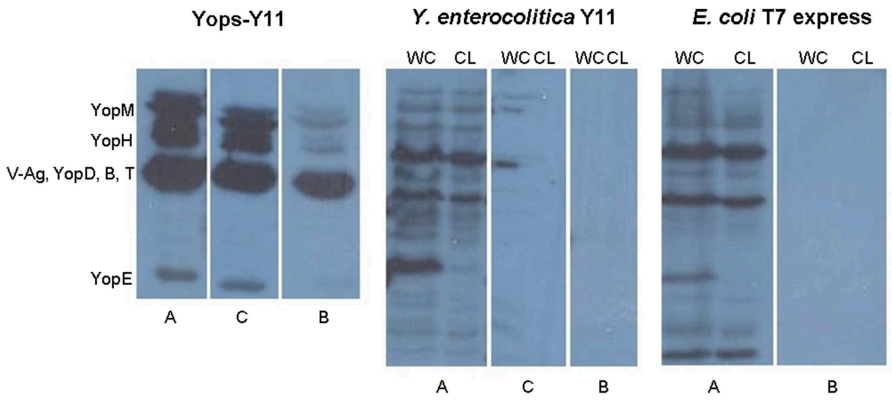


Figure 34 Evaluation of the serum adsorption process efficiency.
Immune reaction against Yops, whole cells (WC) and cell lysates (CL) of unadsorbed (A), adsorbed (B) and partially adsorbed (C) sera is shown.

4.4.2. Screening of a *Y. enterocolitica* strain Y11 expression library

An IPTG-inducible expression library in *E. coli* T7 express was created from ~ 70,000 clones transformed with genomic fragments of *Y. enterocolitica* strain Y11 (sections 3.4.7 and 3.3.6). Considering the minimum insert size of 0.5 kb, a representative library should be obtained from 41,860 clones, while inserts of 3 kb require 6,975 clones. The expression library could therefore be considered as representative and containing at least one copy of all possible sequences. Screening of a significant number of clones (~ 50,000) identified 219 reactive clones, but only 34

of them preserved immune-reactivity after second and third screening and were therefore processed for plasmid isolation and sequencing of the inserts.

4.4.3. *In vivo* induced antigens identified by IVIAT

Insert sequences expressed by the 34 reactive clones were subject to Blast search to determine the encoded genes. Sub-cellular localization was then determined by the PsortB version 3.0 software, together with analysis of the Gene Ontology cellular component term (Table 19). A total number of 45 *Y. enterocolitica* serotype O:3 genes or gene fragments were represented by the immuno-reactive clones (Table 33). These *in vivo* induced (IVI) genes are predicted to be especially associated with different metabolic and biosynthetic functions, but also in motility, host adhesion and invasion, regulation, transport and cell wall synthesis/remodeling.

Table 33 *Y. enterocolitica* serotype O:3 gene products identified by IVIAT.

Functional category	ORF	Gene product, function	Predicted cellular localization
Host adaptation and invasion	Y11_p0711	type III secretion inner membrane protein YscU, T3SS	Cytoplasmic membrane
	Y11_41911	Acid stress chaperone, acid resistance	Periplasmic
Metabolism	Y11_05151	N-methyl-L-tryptophan oxidase, oxidoreductase activity	Cytoplasmic
	Y11_06801	hydroxypyruvate isomerase, hydroxypyruvate isomerase activity	Cytoplasmic
	Y11_20631	cytochrome O ubiquinol oxidase subunit III, aerobic electron transport chain	Cytoplasmic membrane
	Y11_21501	acyl-CoA dehydrogenase short-chain specific, beta-oxidation	Cytoplasmic membrane
	Y11_22271	class I fumarate hydratase, TCA cycle	Cytoplasmic
	Y11_24891	ADP-ribose pyrophosphatase, purine metabolism	Cytoplasmic
	Y11_27461	glutathione S-transferase, glutathione metabolism	Cytoplasmic
	Y11_29751	ATPase subunit beta, ATP synthesis	Cytoplasmic membrane
	Y11_32421	nitrite reductase [NAD(P)H] small subunit NirD, nitrate assimilation	Cytoplasmic
	Y11_32431	nitrite reductase [NAD(P)H] large subunit NirB, nitrate assimilation	Cytoplasmic
Regulation	Y11_00711	putative two-component system sensor kinase, two-component regulatory system	Cytoplasmic membrane
	Y11_20391	nitrogen regulatory protein P-II, transcription regulation	Cytoplasmic membrane
	Y11_29371	nitrogen regulation protein GlnG, transcription regulation	Cytoplasmic
	Y11_15401	transcriptional regulatory protein UhpA, transcription regulation	Cytoplasmic

	Y11_15411	sensor histidine protein kinase UhpB G6P specific, regulatory system	Cytoplasmic membrane
	Y11_38441	transcriptional activator NhaR, transcription regulation	Cytoplasmic
Transport	Y11_00941	oxalate/formate antiporter, transmembrane transport	Cytoplasmic membrane
	Y11_22281	BCAA/urea binding protein, ABC transporter	Periplasmic/ Cytoplasmic membrane
	Y11_32661	taurine transport system permease TauC, transporter activity	Cytoplasmic membrane
	Y11_40611	lysophospholipid transporter LplT, lipid transport	Cytoplasmic membrane
	Y11_41901	hoxn/HupN/NixA family nickel/cobalt transporter, metal ion transport	Cytoplasmic membrane
Motility and cell adhesion	Y11_14281	flagellar motor switch protein FliN, motor activity	Unknown
	Y11_14291	flagellar biosynthesis protein FliO, flagellum organization	Cytoplasmic membrane
	Y11_14301	flagellar biosynthesis protein FliP, protein secretion	Cytoplasmic membrane
	Y11_14311	flagellar biosynthesis protein FliQ, flagellum assembly	Cytoplasmic membrane
	Y11_22651	fimbriae usher protein MrfC, fimbrium biogenesis	Outer Membrane
	Y11_22661	MrfB fimbrial protein, pilus	Extracellular
Molecular biosynthesis and degradation	Y11_03341	endonuclease IV Nfo, DNA binding/repair	Cytoplasmic
	Y11_08121	Phospho-2-dehydro-3-deoxyheptonate aldolase, amino acid biosynthesis	Cytoplasmic
	Y11_16351	uroporphyrinogen-III methyltransferase CobA, porphyrin-containing compound biosynthesis	Cytoplasmic
	Y11_16361	L-threonine 3-O-phosphate decarboxylase CobD, porphyrin-containing compound biosynthesis	Cytoplasmic
	Y11_23401	radical SAM family enzyme, porphyrin-containing compound biosynthesis	Cytoplasmic
	Y11_26251	glutamate synthase [NADPH] large chain, glutamate biosynthesis	Cytoplasmic
	Y11_29191	ATP-dependent DNA helicase RecG, DNA recombination and repair	Cytoplasmic
	Y11_29621	aspartate--ammonia ligase AsnA, amino acid biosynthesis	Cytoplasmic
	Y11_38531	carbamoyl-phosphate synthase large chain CarB, amino acid biosynthesis	Periplasmic/ Cytoplasmic membrane
Cell wall synthesis	Y11_39031	UDP-N-acetylmuramoylalanine--D-glutamate ligase MurD, cell wall biosynthesis	Periplasmic/ Cytoplasmic membrane
	Y11_42431	N-acetylmuramic acid 6-phosphate etherase, cell wall biogenesis and recycling	Cytoplasmic

	Y11_42851	penicillin-insensitive transglycosylase & transpeptidase PBP-1C, peptidoglycan biosynthesis	Cytoplasmic membrane
Putative	Y11_24881	putative cytoplasmic protein	Unknown
Function	Y11_27471	uncharacterized protein, DNA binding	Unknown
	Y11_27481	HipA protein	Unknown
	Y11_33161	uncharacterized protein	Unknown

Virulence-related IVI genes. The only virulence plasmid-encoded IVI antigen observed by IVIAT is the *yscU* gene. YscU is an integral inner-membrane protein, which belongs to the export apparatus of the T3SS injectisome and, therefore, is involved in host invasion. It contains a transmembrane domain and a soluble cytosolic domain (Allaoui et al., 1994); interestingly, the gene insert sequenced from the reactive clone corresponds to the C-terminal cytoplasmic domain of this protein.

A periplasmic acid stress chaperone (locus tag: Y11_41911) was identified and assigned to a host adaptation function, since it represents a strategy developed by food-borne enteric pathogens to survive the gastric acidic environment, when travelling through the stomach to the intestine (Hong et al., 2012). This protein contains domains belonging to the acid-resistance protein HdeB and to the HdeA/HdeB family, and has been annotated as HdeB. In *E. coli* and *Shigella* sp. the *hdeAB* genes constitute an operon and are located in the acid fitness island (Carter et al., 2012). The *hdeA* gene could not be found in any *Yersinia* species, nor was the acid fitness island. The IVI *hdeB* gene, on the other hand, is 48% identical to *hdeB* in *E. coli* and is conserved among YE strains. It is located downstream the urease cluster (Y11_41821-41901), which is also involved in acid tolerance, neutralizing H^+ by converting urea to ammonia.

Two IVI antigens (Y11_22651 and Y11_22661), belonging to a large fimbrial cluster composed of 8 CDSs (Y11_22601-22691), are conserved among YE strains and resemble the Mrf/Stf/Ste fimbrial systems in other *Enterobacteriaceae*. The IVI usher protein MrfC contains the PapC N-ter and C-ter domains and has homologous proteins in *Serratia* sp. and YE-like genomes, but not in *Y. pestis* or in *Y. pseudotuberculosis*. Remarkably, a StbC fimbrial usher protein has been identified by IVIAT performed with the poultry-adapted *S. enterica* serovar *pullorum* (Li et al., 2013). This protein has a similarity score of 42.5% with the IVI MrfC protein identified in YE strain Y11; however, both proteins contain the same conserved PapC N-ter usher and PapC C-ter domains.

IVIAT identified another virulence-related gene, Y11_00711, which encodes a membrane-bound sensor kinase. This protein putatively belongs to a two-component transduction system, together

with the adjacent response regulator Y11_00701. Two-component systems are used by bacteria to respond to environmental stimuli, by transferring a phosphate from the sensor histidine kinase to an aspartate residue within the conserved regulatory domain of the response regulator. Phosphorylation then activates a variable effector domain, triggering the cellular response and, thus, regulating processes such as chemotaxis, osmoregulation and metabolism, as well as expression of toxins and other pathogenic proteins (West and Stock, 2001). Interestingly, two IVI gene products (Y11_15401 and Y11_15411) correspond to the C-ter and N-ter part of the UhpA and UhpB proteins, respectively. These proteins form another two-component regulatory system, which is stimulated by the presence of glucose-6-phosphate and enables the bacteria to adsorb organophosphate compounds from the environment and use them as energy source.

Flagella are considered virulence factors for various pathogens and, indeed, four immunogenic products involved in flagellar assembly were detected by IVIAT. FliN is a component of the motor switch and consists of a SpoA (surface presentation of antigens) domain, while FliO, FliP and FliQ are integral membrane proteins essential for the flagellar export machinery. The genes are located within the Flag-1 cluster, common to all YE.

Macromolecular biosynthesis and metabolism. Proteins associated with biosynthesis and metabolism are essential for bacterial survival and growth *in vivo* and, consequently, for bacterial infection. IVIAT detected several enzymes involved in the synthesis of cell molecules, including amino acids, glutathione and cofactors. In particular, three IVI proteins (CobA, CobD and Y11_23401) are components of the porphyrin/cobalamin metabolic pathway, with genes *cobA* and *cobD* belonging to the cob/pdu operon uniquely acquired by YE (Thomson et al., 2006). Cobalamin, an essential cofactor in animals and plants, is produced by YE only in anaerobic conditions and is used to ferment propanediol as important source of energy (Prentice et al., 2003).

Other three IVI proteins involved in energy metabolism were classified as part of the nitrogen metabolism pathway of YE strain Y11. The NirBD nitrite reductases allow the bacteria to reduce nitrite to ammonium, while the glutamate synthase (NADPH) Y11_26251 assimilates ammonia by converting L-glutamine into L-glutamate. In addition, two nitrogen regulatory proteins were identified by IVIAT. The GlnG protein is a two-component response regulator of the NtrC family. This system responds to nitrogen-limited conditions and induces the transcription of the GlnA glutamine synthetase, facilitating the assimilation of nitrogen-containing compounds. Overall, YE may be assisted by these nitrogen metabolism-related proteins to persist in anaerobic conditions. Two other IVI proteins participate in energy metabolism and, specifically, in the aerobic oxidative phosphorylation. The cytochrome O ubiquinol oxidase (Y11_20631) is located

in the prokaryotic electron transport chain and reduces oxygen into water. The final enzyme in this pathway is the ATP synthase (the IVI gene Y11_29751), a multi-subunit enzyme which drives the synthesis of ATP from ADP and phosphate.

Among the three IVI proteins involved in cell wall synthesis, the PBP-1C penicillin-binding protein has a bifunctional action of transglycosylase and transpeptidase, containing the respective domains at the N- and C-termini. This membrane-associated macromolecule is required not only for peptidoglycan and cell wall biosynthesis but also for cell division. Interestingly, the YE PBP-1C protein has a similarity score of 64% with the orthologous IVI protein in *S. pullorum* (Li et al., 2013).

DNA repair and recombination have essential roles in maintaining prokaryotic cell viability and survival in response to significant environmental stress. Accordingly, IVIAT identified two enzymes implicated in DNA repair. The Nfo protein, an apurinic/apyrimidinic (AP) endonuclease IV, is involved in the DNA base excision repair pathway, while the RecG helicase belongs to the homologous recombination system.

Transport proteins. Among the IVI transporters, a protein with predicted short-chain amide and urea transport activity was detected. This protein is a periplasmic binding protein-dependent ATP-Binding Cassette (ABC) transporter. Interestingly, the corresponding gene locus Y11_22281 is located within the second LIV gene cluster specifically carried by YEP genomes and absent from the highly-virulent strains (section 4.2.1). KEGG analyses classified this protein, together with the products of the loci Y11_22341 and Y11_34321, as LivK high-affinity system for branched-chain amino acid (BCAA) transport. Importantly, the change in intracellular BCAA concentration is a trophic signal playing a role in bacterial regulatory processes related to chromosome structural organization and proteolysis (D'Ari et al., 1993; Guedon et al., 2001). The TauC ABC transporter belongs to the sulfur metabolism pathway and allows bacteria to assimilate extracellular taurine, which can be then converted into sulfite by the taurine dioxygenase TauD, providing sulfur for growth. Taurine is an organic acid which constitutes the human bile and can be found in the large intestine and in different animal tissues; it is also used as food additive for cats and dogs.

IVI antigens with putative functions. Four proteins with unknown function were identified by IVIAT. One of them is predicted to be a HipA-like toxin, with serine/threonine kinase activity, containing both HipA N- and C-ter conserved domains. The HipA activity is usually overcome by the antitoxin HipB, a transcription repressor which also regulates *hipBA* expression. In fact, the adjacent gene, Y11_27471, is the hypothetical HipB HTH-type transcriptional regulator, belonging to the xenobiotic response element family. HipA has demonstrated multidrug

tolerance, characteristic of dormant bacteria or persisters, suggesting that the HipBA system acts as a bacterial persistence factor (Schumacher et al., 2009).

4.4.4. Genomic analysis of IVI genes

Making use of the genome analysis of YE serotype O:3 (section 4.3), IVI genes could be assigned to core genes and genes with variants. Among the 45 identified IVI genes, 40 belong to the core genome of serotype O:3; as expected, the pYV-encoded IVI locus Y11_p0711 is not included, as the core genes were calculated using only the chromosomal sequences (section 4.3.3). The other 4 non-core IVI genes are Y11_00941, Y11_15411, Y11_22661 and Y11_26251. Curiously, 20 IVI genes contain a total of 31 variants, consisting of 19 nonsynonymous, 2 frameshift (Table 34) and 10 synonymous mutations. Considering the distribution across the serotype O:3 phylogeny (Figure 32), 25 out of 31 variants (80.6%) occur in single samples, 2 are specific for the cluster of strains 149, P1 and D7 (in genes Y11_20631, synonymous, and Y11_24881, nonsynonymous) and 1 for strain PhRBD_Ye1 and YE12-03 (the nonsynonymous SNP in the RecG protein). Interestingly, a synonymous mutation in Y11_42431 in present only in lineage B, while the two mutations in Y11_15411 and Y11_40611 are shared by lineages B and C.

Table 34 IVI genes containing nonsynonymous SNPs and Indels (fs: frameshift).

IVI gene	Core gene	Variant mutation	Product
Y11_00941	No	Glu37fs	oxalate/formate antiporter
Y11_05151	Yes	Gly144Ser Pro120Ser	N-methyl-L-tryptophan oxidase
Y11_15401	Yes	Asp152Glu	transcriptional regulatory protein UhpA
Y11_15411	No	Thr409Ala	sensor histidine protein kinase UhpB
Y11_16351	Yes	Gly10Ser	uroporphyrinogen-III methyltransferase
Y11_21501	Yes	Gln104Leu	acyl-CoA dehydrogenase, short-chain specific
Y11_22651	Yes	Ala380Val Thr204Ser	fimbriae usher protein StfC
Y11_24881	Yes	Val78Leu	putative cytoplasmic protein
Y11_26251	No	Val444Gly	glutamate synthase [NADPH] large chain
Y11_29191	Yes	Glu254Lys	ATP-dependent DNA helicase RecG
Y11_29621	Yes	Ala206Val	aspartate--ammonia ligase
Y11_32431	Yes	Val429Ala Ala7Thr	nitrite reductase [NAD(P)H] large subunit
Y11_38531	Yes	Pro17Leu	carbamoyl-phosphate synthase large chain
Y11_39031	Yes	Ala27Val	UDP-N-acetylmuramoylalanine--D-glutamate ligase
Y11_40611	Yes	Leu255Pro Ala374Glu	lysophospholipid transporter LplT
Y11_41901	Yes	Asp167Gly	hoxn/HupN/NixA family nickel/cobalt transporter
Y11_42851	Yes	Leu29fs	penicillin-insensitive PBP-1C

5. Discussion

Comparative genomics investigation into the heterogeneity of different *Y. enterocolitica* serotypes showed a general great diversity as whole species, corresponding to the wide range of niches where these bacteria can proliferate. In contrast to the mouse-virulent and non-virulent biotypes 1B and 1A, the weakly-virulent serotypes demonstrated lower genetic heterogeneity and, especially, serotype O:3 strains proved to be genetically monomorphic, in accordance with their host specialization. Despite the known restricted metabolic capacity of the genetic nondiverse weakly-virulent group, metabolic gene clusters which might provide anaerobic growth advantages in specific ecological niches have been discovered. Moreover, genome-wide investigations allowed development of a molecular serotyping scheme, which could be used as basis for an efficient epidemiological and typing tool. Finally, genome analyses, together with results from the IVIAT technique, also revealed the importance of fimbriae, flagella and autotransporters in virulence/fitness and host adaptation properties of YE.

5.1. High heterogeneity of *Y. enterocolitica* subsp. *enterocolitica*

Highly-virulent *Y. enterocolitica* strains of bioserotype 1B/O:8 have been largely and interchangeably used to elucidate the virulence/fitness mechanisms of this gastrointestinal pathogen. According to the results obtained in this study, YE strains 8081 and WA-314 demonstrated not only high genomic diversity, but also different behaviors during *in vitro* and *in vivo* growth.

5.1.1. Genomic diversity of *Y. enterocolitica* subsp. *enterocolitica*

The newly sequenced genome of *Y. enterocolitica* strain WA-314 was analyzed and compared to the reference genome of strain 8081 (Thomson et al., 2006). Overall, the two highly-virulent strains have similar genome properties (Table 21), with strain WA-314 having smaller chromosome and plasmid sizes because of the incompleteness of the sequences. In fact, shotgun sequencing approaches are based on libraries created by the fragmented target genome, producing sequence reads which have to be ordered and assembled. When the sequence contains repeats, correct assemblies are not easily obtained, since algorithms are not able to place the read in the different alternative positions through the genome. The resulting assemblies are therefore formed by scaffolds, with gaps usually representing the un-assembled repetitive regions (e.g. transposases, insertion sequences and tRNAs). Since these gaps have not been closed in the genome of YE strain WA-314, the resulting sequence sizes and the number of tRNAs are underestimated and smaller than the related strain 8081.

Approximately 91% of genes in strain WA-314 have orthologs in the genome of strain 8081, with the remaining 9% mostly encoding hypothetical and mobile element-related proteins. In fact, important differences have been detected in prophages and genomic islands. Strain WA-314 is devoid of the islands YGI-3, YGI-4 and YAPI, which is involved in *Y. pseudotuberculosis* virulence (Collyn et al., 2004). Additionally, strain WA-314 harbors seven prophages, which show low sequence similarity with prophage regions in strain 8081. In particular, a PilV-like adhesin, carried by the prophage YWA-4, may increase intestinal colonization of strain WA-314, as it has been shown for other *Enterobacteriaceae* (Bieber et al., 1998; Zhang et al., 2000). Curiously, the YWA-4 prophage occupies the same genomic region as the YGI-3 in strain 8081, between two tRNA genes. As demonstrated also by genome comparison of the whole YE species (section 4.2.4 and Figure 20), this region is highly variable and may be considered a "hot-spot" for the integration of genetic material acquired by horizontal gene transfer (HGT) in YE biotypes 1A and 1B. Genomic islands and prophages greatly contribute to inter- and intra-species genetic variability and evolution of a variety of pathogenic bacteria (Dobrindt and Hacker, 2001), stressing the main role of horizontal gene transfer and mobile elements in shaping the genomic versatility of highly-virulent YE. Besides mobile elements, YE highly-virulent strains 8081 and WA-314 differ in the presence and sequence similarity of certain gene clusters and protein-encoding genes. Strain WA-314 specifically possesses a xenobiotic-acetyltransferase-encoding gene, a restriction modification system and a putative colicin cluster. Whether these proteins are efficiently expressed and functional *in vitro* and/or *in vivo* needs to be clarify. From an evolutionary perspective, being isolated from human, animal and environmental sources (Fredriksson-Ahomaa et al., 2006), YE serotype O:8 strains undergo numerous opportunities for exchanging genetic material. These flexible lifestyles make these bacteria be in contact with many microbial communities, and may therefore explain the genome plasticity as a result of continue adaptation for survival in different niches.

5.1.2. New virulence factors for the highly-virulent phenotype

Comparing and analyzing CDSs with low-sequence similarity between strains 8081 and WA-314 allowed identification of new factors putatively involved in *Y. enterocolitica* pathogenesis. Autotransporters represent the largest protein family in pathogenic Gram-negative bacteria and perform heterogeneous functions, such as mediation of biofilm formation, bacterial aggregation and adhesion to epithelial cells. This functional diversity is due to the variable N-ter passenger domain, which is responsible for the pathogenesis of bacteria (Benz and Schmidt, 2011). Two orthologous autotransporters, showing only 32% sequence identity between strains 8081 and WA-314 (respective locus tags: YE3700 and YWA314_14949), contain dissimilar passenger

domains, which may confer strain-specific adhesion properties. While this gene is conserved among all YE, even if with low sequence similarity (section 4.2.2), another adhesin is specific for YEE (locus tags: YE0694 and YWA314_00878), and may contribute to the highly-virulent phenotype of strains 8081 and WA-314. In addition, a biotype 1B-specific fimbrial operon with differences between strains 8081 and WA-314 (locus tags: YE1111-1114 and YWA314_11901-11886) has been identified, and may also be considered a specific virulence factor for highly-virulent strains. In fact, fimbriae (or pili) have demonstrated adhesion properties in different Gram-negative bacteria, leading to colonization of host tissues in the urinary, genital and gastrointestinal tracts (Proft and Baker, 2009). Genome comparison of two YE biotype 1B strains, therefore, allowed identification of yet unknown genes which may be involved in pathogenesis of this highly-virulent group of bacteria.

5.1.3. Differences in the virulence plasmid sequences

Considering the established virulence determinants of *Y. enterocolitica* (Table 3 and section 1.2), significant differences between strains 8081 and WA-314 were found in certain pYV-encoded proteins. YopM is a T3SS-dependent effector protein indispensable for full virulence of pathogenic *Yersinia*, even though its molecular functions still remain largely obscure. In *Y. pseudotuberculosis* YopM modulates virulence by interaction with two intracellular serine/threonine kinases, PRK2 and RSK1, involving the LRR6-LRR15 and the LRR12-C-ter regions, respectively (McCoy et al., 2010; McDonald et al., 2003; McPhee et al., 2010). Notably, $YopM_{8081}$ is 367-aa long with 13 LRRs, while $YopM_{WA-314}$ has 505 residues and 24 LRRs (Figure 7). However, it is not clear whether the different number of LRRs in YopM proteins of strain 8081 and WA-314 has different consequences on the virulence of YE in the mouse model.

LcrV is one of the three translocator proteins which form the pore complex connecting the needle tip of the T3SS injectisome to the target host cell (Dewoody et al., 2013b). The 9-aa insertion identified in $LcrV_{WA-314}$ has been previously observed in strain WA-314 and in other biotype 1B strains but is absent from strain 8081 and from weakly-virulent YE (Foultier and Cornelis, 2003; Roggenkamp et al., 1997). The role of this polymorphic region seems to be related to a humoral immunosuppression function, in that antibodies directed against *Y. pestis* LcrV are unable to block the T3SS-based virulence of YE strain WA-314 and strain W22703 (serotype O:9) (Miller et al., 2012; Motin et al., 1994). Therefore, antibodies produced against $LcrV_{8081}$ might not protect from YE strain WA-314 infections. However, the observed dissimilarities are independent from the amino acid polymorphisms in the LcrV sequence, but may be due to differences in injectisome assembly between *Y. pestis* and YE (Ligtenberg et al.,

2013), suggesting that the different LcrV sequences of strains 8081 and WA-314 might not be involved in different virulence properties.

YadA is an important adhesin involved in serum resistance and in mediation of adherence to epithelial cells and to extracellular matrix (ECM) proteins (Balligand et al., 1985; Heesemann and Grüter, 1987). The fundamental role of YadA for YE virulence has been demonstrated in the mouse infection model (Roggenkamp et al., 1995). The domains involved in these virulence-associated functions are located in the N-terminal head region of the YadA protein (El Tahir and Skurnik, 2001); in particular, the amino acids 29 to 81 are important for neutrophil interaction and virulence in mice (Roggenkamp et al., 1996). Interestingly, $YadA_{8081}$ and $YadA_{WA-314}$ sequences possess six different amino acids at positions 53 to 67 (section 4.1.1, page 61). Whether the detected mutations alter neutrophil binding and/or virulence of YE strains 8081 and WA-314 needs further investigations.

YscP regulates the *Yersinia* injectisome needle length, which is directly correlated to the size of the YscP protein (Wagner et al., 2009). A minimal needle length is required for a fully functional injectisome, also in relation to the length of the YadA adhesin, providing optimal contact with the host cell (Mota et al., 2005). Yop translocation efficiency might be therefore reduced by shorter YscP proteins or longer YadA proteins. Sequence comparison of pYV_{8081} and pYV_{WA-314} discovered a longer YscP protein in YE strain WA-314 compared to strain 8081 (Figure 8), due to three additional amino acid repeats. Since the length of YadA is the same, the longer YscP protein in strain WA-314 might improve the Yop translocation and, therefore, provide better virulence properties than strain 8081.

Apart from differences in the YopM, LcrV, YscP and YadA sequences, YE WA-314 carries the *ylpA* gene, which is completely missing in pYV_{8081}. This gene is conserved among the three pathogenic *Yersinia* species and encodes a lipoprotein related to the TraT protein. TraT is a cell-surface-exposed lipoprotein carried by plasmids of Gram-negative bacteria. Probable roles in pathogenesis are supported by the observation that TraT confers serum-resistance in *E. coli* (Sukupolvi and O'Connor, 1990). However, no differences were found between YE strain W22703 and the corresponding *ylpA* mutant in terms of resistance to human serum (Balligand et al., 1985), nor in bacterial counts from organs of mice infected with these strains (China et al., 1990). An implication of YlpA in YE strain WA-314 virulence is therefore unlikely.

5.1.4. Different colonization properties of highly-virulent *Y. enterocolitica*

Intra-species variation among *Y. enterocolitica* biotype 1B strains is evident from a genomic perspective, but is also confirmed by phenotypical diversification of strains 8081 and WA-314

during *in vivo* growth. In fact, these two highly-virulent strains demonstrated different virulence behaviors in a mouse model, especially in competition assays (section 4.1.3). A statistically significant number of mice were used for single strain and co-infection experiments via the intraperitoneal route, which was preferred to oral infection to avoid different dissemination efficiency from the intestine into spleen and liver, and to enable an accurate and controlled injection of the bacterial suspension. Nevertheless, two mice were not successfully injected with strain WA-314, while another mouse showed no bacteria in the spleen and low counts in the liver, probably because of a faster splenic clearance of bacteria. Overall, bacterial growth was more efficient for strain WA-314 than strain 8081, even though the loads were statistically significantly higher in the spleens and not in the livers. In co-infected mice, strain WA-314 outcompeted strain 8081, demonstrating improved colonization efficiency and adaptation to the host environment. YE mouse-virulent strains, therefore, seem to differ in invasion and proliferation in host niches, affecting their pathogenicity level.

The genetic differences detected in strain 8081 and WA-314 may partly explain the higher virulence/fitness of strain WA-314. Apart from strain-specific prophages, which may carry hypothetical proteins with virulence-associated functions, adhesion proteins and known pYV-encoded virulence factors showed significant differences between the two analyzed strains, as discussed in the previous sections. Moreover, a colicin operon has been identified in strain WA-314. Although its activity has not been proven, it can be postulated that this colicin confers additional advantage to strain WA-314 against strain 8081, by inhibiting growth of competitor organisms. Colicins from *Yersinia* species are known and activity against closely-related bacteria have been demonstrated (Bosak et al., 2012; Bosak et al., 2013; Ferber and Brubaker, 1979).

The small differences observed in *in vitro* growth phenotypes (section 4.1.2) could also account for the *in vivo* behavior of YE strains 8081 and WA-314, independently of their virulence properties. However, the bacterial growth experiment was done in LB medium, at 27 °C and for some hours, in contrast to the *in vivo* growth rate, which was measured at the host temperature (37 °C) during a period of days. Most *Yersinia* virulence factors are not expressed at 27 °C but only at 37 °C, and, in general, the conditions of the *in vitro* growth experiment are totally different than the environmental conditions of mouse body-tissues and -fluids. Accordingly, differential expression profiles of metabolic pathways and virulence factors have been shown in *Yersinia* cultured in human plasma and in LB medium (Rosso et al., 2008). In conclusion, there are many factors that can determine the *in vivo* survival and growth of *Yersinia*, including nutrient composition and availability in the host niches, ability of the bacteria to gain access to the necessary nutrients, virulence factors and host defense mechanisms.

5.2. *Y. enterocolitica* subsp. *palearctica* genomics

By comparison against the highly-virulent strain 8081, in recent works the genomic features involved in the low virulence of some *Y. enterocolitica* subsp. *palearctica* strains have been partly elucidated (section 1.5). In this study, comparison among genomic sequences of different YEP strains was accomplished taking into consideration the general mouse-virulent features discovered from the analysis of strains 8081 and WA-314 (section 4.1). In addition to significant differences in metabolic clusters, genes with putative disease-associated roles were identified. Importantly, the used genome sequences, which derived from both pathogenic and non-pathogenic isolates, well exemplify the broad diversity of this species in terms of serotype, geographical origin and clinical relevance.

5.2.1. Respiratory flexibility of *Y. enterocolitica* subsp. *palearctica*

Utilization of N-acetyl-galactosamine (Aga) as carbon source has been demonstrated in *Y. enterocolitica* serotype O:3 and is related to the *agaVWEF* gene cluster (Batzilla et al., 2011a), which is present in YEP but absent from the highly-virulent YEE. Since Aga is a major component of the pig intestinal mucin, the *aga* operon has been denoted as important fitness and host adaptation factor, especially for the pig-associated serotype O:3 strains. Similarly, YEP strains carry additional metabolic genetic clusters which confer respiratory flexibility, stressing the importance of anaerobic respiration in YE patho-adaptation and proliferation in different ecological niches.

YE is a facultative anaerobic bacterium and, thus, can metabolize energy from oxygen, if present, or from other electron acceptors, by means of reductases. Nitrate is an important component of the nitrogen cycle and a source of energy during bacterial anaerobic growth. Nitrate and nitrite are commonly found in water, soil, vegetables and meat, even though nitrite is easily oxidized to nitrate, which is therefore the predominant compound in the environment. In many bacteria the nitrite generated from nitrate respiration can be further reduced to ammonium by nitrite reductases during the dissimilatory nitrate reduction (Richardson, 2000). Two independent pathways allow this process: the periplasmic respiratory nitrite reductase encoded by the *nrf* operon and the cytoplasmic NADH-dependent nitrite reductase encoded by the *nirBD* genes. The non-respiratory Nir pathway, which does not generate a proton motive force, provides a mechanism for detoxifying nitrite in the cytoplasm, while the electrogenic reduction of nitrite by the Nrf pathway offers a secondary source of energy during anaerobic respiration (Page et al., 1990). As discovered by genome comparison (section 4.2.1), YEP is equipped with both Nrf and Nir systems, while YEE only possesses the Nir pathway. This implies that the

proliferation of weakly-virulent and non-virulent strains in anaerobic ecological niches, such as food, the intestinal tract and tonsils, may be supported by nitrite as source of energy. For example, YE serotype O:3 can grow on raw pork packaged in modified atmosphere (Fredriksson-Ahomaa et al., 2012); as nitrites are used for processing and curing meat, the expression of the Nrf reductase may explain the ability of serotype O:3 strains to survive in these oxygen-deficient conditions. More importantly, Nrf has also demonstrated an activity in nitric oxide reduction during *in vitro* and *in vivo* growth of *E. coli* (Poock et al., 2002). This suggests an additional pathogenic role of this cluster in detoxification of the nitric oxide, which is produced for instance by macrophages as microbicidal agent (Murray and Wynn, 2011), encountered during YE invasion of the host.

Likewise, YEP strains might be able to grow in anaerobic conditions by using dimethylsulphoxide (DMSO) as alternate electron acceptor, since a DMSO reductase system has been identified in this group of organisms (Table 25). DMSO is found in marine environments, in water surface layers, in fruits and vegetables (Griebler and Slezak, 2001), which are all specialized ecological niches for YE. Interestingly, impairment of the function of the DMSO reductase-encoding gene *dmsA* in *Actinobacillus pleuropneumoniae*, which is able to persist on respiratory epithelia, in tonsils, and in the anaerobic environment of encapsulated lung cells, causing porcine pleuropneumonia, resulted in reduced virulence (Baltes et al., 2003). The *dmsA* gene, moreover, is regulated by the HlyX protein, which is involved in persistence of this bacterium in pig lungs and tonsils (Baltes et al., 2005). A role of the DMSO reductases in YE serotype O:3 pig-adaptation and tonsil colonization can logically be proposed.

5.2.2. Virulence-associated factors in weakly-virulent and non-virulent strains

Bacterial genome sequencing and comparative genomics are useful tools for discovering candidate virulence genes and for better understanding pathogenesis mechanisms. Virulence factors are not only involved in adhesion/invasion of the host and inhibition of immune responses, but also in successful proliferation and competitive advantage against other bacteria. Previous genome analyses of weakly-virulent strains have revealed several toxin-related gene clusters, including RTX-like toxins and colicins (Batzilla et al., 2011a; Wang et al., 2011). Despite the question whether these clusters are actively expressed in the host and able to exert pathogenic functions, toxins seem to have a relevant role for the whole species *Y. enterocolitica*, since also mouse-virulent biotype 1B and non-virulent biotype 1A strains carry genes belonging to RTX toxin and bacteriocin families. Apart from putative “classical” functions in virulence, toxins

more probably contribute to YE niche adaptation and defense against bacteria and fungi sharing the same environments.

YE virulence is strongly dependent on the pYV-encoded adhesin YadA for adherence to host surfaces (section 1.2.2 and 5.1.3). Additionally, genomes of YEP revealed that other chromosomal-encoded proteins might play a role in efficient anchoring of these bacteria to host tissues or surfaces. Fimbrial adhesins and autotransporters appeared as newly discovered candidate factors for the high-virulence phenotype. Curiously, also weakly-virulent and non-virulent strains carry specific fimbriae- and autotransporter-encoding genes (section 4.2.2). Surface fimbriae/pili are important for the colonization of many pathogenic bacteria and can also function as receptors for bacteriophages (Proft and Baker, 2009). Pili assembled by the chaperone-usher pathway, as the ones identified in YE, are commonly found in the family of *Enterobacteriaceae*, and require specific receptors exposed on the host cells, such as glycolipids, collagen proteins and integrins, which are recognized attachment sites for the YE YadA and Inv proteins (section 1.2). Therefore, the discovered fimbrial clusters may be additional adhesive structures for host recognition and adherence to receptor molecules on host cells and tissues and, in particular, for colonization of specific niches, such as pig tonsils and intestinal epithelial cells.

Autotransporters are cell-surface-exposed or released proteins produced by a large variety of Gram-negative bacteria. Due to their virulence-related functions, they have been intensively studied, with their synthesis and secretion mechanisms being well characterized (Benz and Schmidt, 2011; Grijpstra et al., 2013). Their virulence-associated roles comprise invasion, adhesion, actin-nucleating functions, cytotoxicity and immune evasion. In YE, one of the major virulence factors, YadA, is a trimeric autotransporter, whose crystal structure has been elucidated, allowing identification of putative collagen-binding regions (Nummelin et al., 2004). The autotransporters specifically detected in different YE serotypes belong to the class of classical autotransporters, which consist of an N-terminal signal sequence for transport across the inner membrane, a passenger domain, and a C-terminal translocation domain that likely forms a pore in the outer membrane for the transport of the passenger part to the cell surface. Examples of classical autotransporters in pathogenic bacteria include the adhesin AidA of enteropathogenic *E. coli*, pertactin and BrkA of *Bordetella pertussis* and the cytotoxin VacA of *Helicobacter pylori* (Benz and Schmidt, 2011). Characterization of the YE newly detected autotransporters may bring new insights into their biological functions and putative implications in YE infection and/or adaptation to specific host niches.

Another important class of bacterial transmembrane proteins includes ABC transporters, which have a wide variety of substrates and a demonstrated role in virulence of several bacterial pathogens, for example being involved in export of toxins or import of nutrients (Lewis et al., 2012). ABC transporters constitute a major part of the core genes (section 4.2.3) and the variable regions (section 4.2.1) of YE. In particular, YEP strains carry two ABC transport systems involved in acquisition of the branched-chain amino acids (BCAA) leucine, isoleucine and valine (LIV), while YEE only possesses one. It has been shown that the LivJHMGF ABC transporter is necessary for *Streptococcus pneumoniae* pathogenesis, even though the decreased virulence mechanisms related to the loss of BCAA are not clear (Basavanna et al., 2009). BCAA limitation seems also to be involved in reduced virulence and survival of *A. pleuropneumoniae* in swine, as mutants unable to synthesize BCAA were attenuated in a porcine infection model (Subashchandrabose et al., 2009). The role of BCAA in YE pathogenesis, therefore, requires more clarification.

Overall, disease-associated determinants in weakly-virulent and non-virulent YE serotypes seem to be strongly related to adherence and host interaction abilities, since autotransporters and fimbriae constitute the main findings. Moreover, being extracellular surface-exposed structures, pili and autotransporters may be used as effective recombinant vaccines and for YE diagnosis.

5.3. A new assay for *Y. enterocolitica* patho-serotyping

Serotype is one of the main characteristics of *Y. enterocolitica*, as demonstrated also by phylogenetic analysis (section 4.3.6), and may be useful for epidemiological and diagnostic purposes, especially because strains enteropathogenic for humans are limited to a few serotypes. The chemical structure of the O-ag, the LPS part which determines the serotypes, has been characterized for some YE serotypes, but the O-ag genetic clusters and their genomic locations are only partly known for serotypes O:3, O:8 and O:9 (section 1.1.1). The availability of YE genome sequences from strains of the major pathogenic serotypes enabled comparison among O-ag gene clusters and identification of serotype-specific genes, with subsequent development and validation of a PCR-based method for simultaneous identification and patho-serotyping of YE.

5.3.1. Newly discovered O-ag genetic clusters

Y. enterocolitica heterogeneous genetic nature is also reflected in the genomic position of the genes involved in the O-ag biosynthesis. In contrast to *Y. pseudotuberculosis* and *Y. pestis*, where the O-ag clusters are located in the same chromosomal location in the *hemH-gsk* locus

(Bogdanovich et al., 2003), only serotype O:8 and O:36 O-ag clusters are found in this region. The O-ag genes of YE serotypes O:3 and O:9 share another location, whereas the genetic clusters of serotypes O:5,27 and O:5, despite their high DNA sequence similarity, are carried in different positions (Figure 24). Interestingly, no gene cluster encoding a second epitope related to the postulated O-ag factor 27 could be identified in the genome of YE strain Y5.27P, suggesting that factor 27 may not be of LPS origin. In fact, serotype-specific antisera are obtained against heat-stable antigens, which are mainly O-antigens (section 1.1.1). However, antisera may also recognize other specific surface-exposed molecules, which may account for factor 27 of serotype O:5,27. The uncommon chromosomal positions and the huge genetic diversity of YE O-antigens may be explained by recent inter-species lateral gene transfer, as demonstrated for other Gram-negative bacteria (Samuel and Reeves, 2003). Supporting this hypothesis, various transposase-encoding genes have been identified in the analyzed O-ag clusters. Further sequencing aimed at yet understudied serotypes may reveal new heat-stable antigenic gene regions and increase the scientific knowledge on YE serotype classification.

5.3.2. Efficient and specific multiplex PCR assays

A PCR-based test for simultaneous identification and patho-serotyping of *Y. enterocolitica* was established and validated in this study (section 4.2.8). The assay is able to identify YE strains at a species-level, distinguishing between the four clinically prevailing serotypes O:3, O:9, O:8 and O:5,27. It can also differentiate between virulent and non-virulent strains, detecting the genetic targets *ail* and *ystB*, respectively. However, recent epidemiological studies have shown that some YE biotype 1A isolates from pigs and pork products carry the *ail* gene (Bonardi et al., 2013; Paixao et al., 2013b). Conversely, the frequency of YE biotype 1A clinical samples harboring the *ail* gene is not significant (Stephan et al., 2013). The identification multiplex would therefore correctly genotype as “non-virulent” biotype 1A human isolates, while a misclassification may occur when testing isolates from animals. Likewise, the *ystB* gene has not been detected in 27.6% non-virulent YE isolates from pigs (Bonardi et al., 2013), while in fecal samples from humans the *ystB* frequency is 95% (Stephan et al., 2013). The *ail*-positive and *ystB*-negative biotype 1A strains identified with this molecular scheme should, therefore, be further studied for presence of other virulence-associated factors.

The corroborated identification and patho-serotyping method can be applied as diagnostic test, and avoid the low objectivity of the conventional agglutination test and the possible misinterpretation of the results. Enabling higher efficiency, sensitivity and specificity, this assay may be especially useful for typing YE isolates which have lost expression of the O-ag. Indeed,

validation of the assay was carried out also on 19 strains with unknown serotype, and demonstrated higher typing capacity than the conventional methods.

The developed typing scheme is neither complete nor perfect, but can be used as starting point for further improvements. Availability of new genome sequences from YE strains belonging to less-common serotypes will allow a broader molecular serotyping, adding new targets to the present scheme.

5.4. Genomic evolution of the *Y. enterocolitica* species

The genome sequences of a number of *Y. enterocolitica* strains have been analyzed, providing not only an expanded view of the adaptation and virulence capacities (sections 5.1 and 5.2), but also presenting novel insights into the intra-species diversity and genome evolution of this bacterial species. Phylogenesis, pan genomics, core genomics, and singleton analyses revealed a high variability among YE genomes as entire species, with lower diversity in the biotypes 2-5 groups. The results reflect the niche restriction of weakly-virulent strains in contrast to the wider ecology of the highly- and non-virulent groups.

5.4.1. Open pan-genome of the species *Y. enterocolitica*

Genome sequencing and comparison of ten strains allowed the description of the core and pan-genomes of *Y. enterocolitica*, providing a glimpse of the genetic diversity within this species. The pan-genome of a particular species represents its whole gene repertoire and corresponds to the sum of the core and dispensable/unique genomes (Medini et al., 2005; Tettelin et al., 2008). The core genome includes genes that are present in all strains of a given species, thus including essential functions; on the other side, the dispensable and unique genes are absent in at least one strain or are present in single strains, and may confer selective advantages, such as adaptation to specific niches, stress resistance, and pathogenicity. Importantly, the calculations of the core and pan-genomes may be slightly biased by the use of draft genomes, since CDS prediction and, therefore, comparison are influenced by the gaps between contigs and by the quality of the assembled sequence.

The analysis of the YE pan-genome using the EDGAR comparative genomics tool (Blom et al., 2009) identified 8,102 CDSs, of which 3,090 are core genes, while 5,012 are dispensable and unique genes (section 4.2.3), thus showing a strong contribution of the non-core genes. YE demonstrated a large pan-genome, in comparison with other bacteria, such as the plant pathogen *Erwinia amylovora*, with 5,751 pan genes calculated on 12 genomes (Mann et al., 2013), the swine pathogen *Streptococcus suis*, with 3,585 pan genes in 13 strains (Zhang et al.,

2011), the human pathogen *Listeria monocytogenes*, having 4,052 pan genes in 26 genomes (Deng et al., 2010), and the human and zoonotic enteric pathogen *S. enterica*, with 10,015 CDSs in 35 genomes (Jacobsen et al., 2011). In contrast, considering that the pan-genome of YE consists of 9,804 genes when calculated on 20 genomes using the Heap' law function of Figure 18, the close relative *E. coli* demonstrated a bigger pan-genome (circa 13,000 genes in 17 sequences) (Rasko et al., 2008).

Nearly 74% of CDSs per YE genome are classified as core, meaning that a relatively high proportion of the genomes are conserved among the whole species. This is consistent with the analysis of the core genome development plot (Figure 17), which represents a decay function predicting the development of the number of core genes with an increasing number of genomes (Tettelin et al., 2005). Based on the 10 YE chromosomal genomes analyzed, the predicted number of core genes converges to 3,060 for an infinite number of sequenced genomes. Similarly, the slope for the singleton number approaches an asymptote, predicting the number of additional singletons (122) putatively retrieved by future genome sequencing (Figure 18). Together with the continuously increasing number of pan genes (Figure 18), such genomic structure indicates that the pan-genome of YE is open, meaning that the species is still evolving (Medini et al., 2005). Bacterial species with an open pan-genome are able to colonize multiple environmental niches, and gene gain and loss play a prominent role in the process of host adaptation, as well as in lifestyle changes (Medini et al., 2005). In fact, YE has been isolated from a variety of hosts, such as water, soil, vegetables and mammals, experiencing great opportunities for contacts with potential gene donors. This seems to be particularly true for the non-virulent biotype 1A and the mouse-virulent biotype 1B strains, which contribute to the species variability with a high amount of strain-specific features (> 130 genes), while the weakly-virulent serotype O:3 strains add less than 50 singletons all together, suggesting a clonal-like behavior (Figure 19). Genes unique to each strain comprise a high proportion of the non-core genome (about 24%, corresponding to 1,183 CDSs), and mainly encode for hypothetical and mobile element-related proteins, supporting the hypothesis of genomic exchange and evolution by gene flow and horizontal gene transfer, a process which tends to involve a large number of hypothetical proteins (Hsiao et al., 2005). Finally, observing that uncharacterized genes are over-represented among the unique, strain-specific genes suggests the possibility that non-essential genes are responsible for driving the evolutionary diversification among YE strains.

5.4.2. *Y. enterocolitica* phylogenetic relationships

The species *Y. enterocolitica* is traditionally classified into 6 biotypes, more than 70 serotypes, and two subspecies (section 1.1). Observing the phylogeny obtained by using the whole core genome of the species revealed a main clustering according to the virulence degree of the strains, as the non-virulent biotype 1A and the weakly-virulent biotypes 2-5 (*Y. enterocolitica* subsp. *palearctica*), and the highly-virulent biotype 1B (*Y. enterocolitica* subsp. *enterocolitica*) form three distinct lineages (Figure 21). In contrast, no apparent grouping appears based on the subspecies classification, with biotype 1A strains being the progenitors and biotypes 2-5 the most recently diverged. Among the weakly-virulent group, clustering occurs according to the serotype of the strains, and the divergence is relatively low within each cluster. Serotype O:3 strains show a particularly low genetic diversity, with short terminal branches. The long branch of strain W22703 is congruent with the high number of CDSs annotated in this genome, but may not really mirror the phylogenetic distance among serotype O:9 strains. In fact, a newly published phylogeny determined on 118 strains, which is strongly consistent with the one constructed in this study with 10 genomes, has shown a more prominent tight clustering among the weakly-virulent strains of serotypes O:5,27, O:3 and O:9 (Reuter et al., 2014). The authors have suggested that this may be due to ecological specialization events, which shaped the early evolution of these niche-restricted groups. This hypothesis is supported by the low number of unique genes in weakly-virulent strains, as obtained by the genome comparison and singleton analyses performed in this study (section 4.2.4). To note, the higher genetic variation of biotypes 1A and 1B correlates with a broader metabolic capacity of these strains, in contrast to the metabolic restriction and low genetic divergence of biotypes 2-5 (Reuter et al., 2014). Having higher metabolic capability may supply biotype 1A and 1B strains with a competitive growth advantage in a variety of ecological niches, by using carbon sources that are not readily fermented. Overall, although the species YE possesses an open pan-genome and high genetic variability, it also includes discrete and genetically isolated clades, especially significant for strains of serotype O:3.

5.5. *Y. enterocolitica* serotype O:3 microevolution

Y. enterocolitica serotype O:3 is a typical zoonotic pathogen, in that it is frequently associated with food contamination, it is transmissible to humans and its usual host is an animal, the pig (section 1.4.2). As a consequence of their high genomic homogeneity, molecular methods have low sensitivity and discriminatory power in typing these trains, thus preventing a complete understanding of their epidemiology and evolution and, therefore, of the association among

clinical cases, of the reservoirs of infection and of the transmission vehicle. Further studies of YE serotype O:3 thus required a whole-genome approach. Comparative genomic analysis of 20 isolates showed little evidence of recombination and adaptive selection, whereas microevolution of this group seems to be driven by high pseudogene frequency and, consequently, by loss of gene function. No clear differences between pig and human isolates were identified, while more recent mutational events appeared in the human strains. Together with the genetic isolation, this is consistent with the assumed role of pig asymptomatic carriers as the main reservoir of human infections by YE serotype O:3. Overall, YE serotype O:3 can be regarded as a genetically monomorphic group of bacteria, corresponding to a lineage within a species of greater diversity.

5.5.1. Limited variation of serotype O:3 genomes

According to extrapolations based on the pan-genomes, core genomes and singletons of the whole species *Y. enterocolitica*, serotype O:3 showed a more clonal-like behavior than the other lineages (section 4.2). To confirm this, it was necessary to capture as much variation as possible in the serotype O:3 population, based not only on three but on twenty genomes, including thirteen re-sequenced new strains. Since a previous phylogenetic tree of these subspecies was not available, the choice of isolates could not be driven by unbiased sampling from a reliable phylogeny (Pearson et al., 2009), but only on clinical phenotype, geographical origin and source and year of isolation. Spanning a time-window of almost 30 years and deriving from both European and Asian humans and animals, the selected strains provide a satisfactory starting pool of genomes. Moreover, due to the known low diversity of serotype O:3 strains, the read-mapping approach was considered sufficiently valid for phylogeny reconstruction and comparative genomics, and preferred to the slower, more complex and memory intensive *de novo* assembly methods.

Investigations into the core and pan-genomes of YE serotype O:3 revealed interesting clues about the genomic conservation of this subgroup. Indeed, in contrast to the whole YE species, which clearly possesses an open pan-genome, the small serotype O:3 pan-genome seems to be essentially close (section 4.3.3). The number of core genes reaches an asymptote in the development plot after adding 110 genomes, while the pan-genome slowly continues to grow with minor increase when adding additional genomes, though it does not arrive at plateau. If an open pan-genome is defined by having an increasing pan-genome and a decreasing core genome, the calculations obtained with YE serotype O:3 signify a weakly close pan-genome. This indicates that introducing new sequences in the serotype O:3 pan-genome does not significantly affect the general population homogeneity, confirming the relatively close nature of

these genomes. Whole genome comparison also supported this finding, as large-scale recombinational events are absent. In fact, all sequenced serotype O:3 isolates showed conservation in the GIYep-01 genomic island and in the prophage-like elements, with small differences (Figure 29). Additionally, in the recombinational hot-spot region corresponding to the genomic island YGI-3 in YE strain 8081, the analyzed serotype O:3 strains do not carry any alternative prophage or plasmid (Figure 20). Overall, it appears that horizontal gene transfer and gene flow do not play a relevant role in the evolutionary changes in YE serotype O:3 genomes.

The incompleteness of the genomes, which are mainly composed of contigs or scaffolds, may have slightly affected the calculations in this study, especially in terms of genome size and gene prediction. For example, it cannot be excluded that the smallest genome size of the Philippines strain PhRBD_Ye1 is due to the highest degree of sequence fragmentation among the analyzed genomes (section 4.3.2 and Table 29), and not to a real genome decay in the Asian strains.

5.5.2. Evolution of the genetically monomorphic *Y. enterocolitica* serotype O:3

An understanding of the genetic structure of microbial populations provides a framework within which epidemiology of pathogens can be monitored and traced. To elucidate the structure of the serotype O:3 population and its microevolution, a variant-based comparative analysis was applied, capturing genomic differences of these very closely related strains.

The performed analyses support the hypothesis of genetic isolation and drift in *Y. enterocolitica* serotype O:3, and provide new insights into its microevolution. Bacterial evolution may occur according to four processes: natural selection, mutation, gene flow/recombination, and genetic drift. Selection for greater fitness during adaptation seems to be the predominant mechanism for evolutionary dynamic under laboratory conditions (Barrick et al., 2009), and has also been observed in humans and natural bacterial populations, especially because of antibiotic treatments (Mwangi et al., 2007; Roumagnac et al., 2006). On the other side, signals of adaptive selection have not emerged in genetically monomorphic bacteria, which are characterized by low level of sequence homoplasy (mutation of a character found in other independent lineages) and rare recombination events (Achtman, 2012). In YE serotype O:3 there is a lack of evidence for adaptive selection, and the population appears to be genetically monomorphic (section 4.3.7). No genes contain multiple homoplasic variants, indicating low level of recombination among serotype O:3 isolates. Only one gene deviates from the Poisson distribution of the number of variants per gene, even though only 3 SNPs are nonsynonymous. Moreover, pseudogenes have a high frequency in YE serotype O:3, as confirmed by a bigger scale study on YE (Reuter et al., 2014), being signal of host-restricted organisms (section 4.3.7). Together with the results

obtained by genome comparison, which proved the absence of genomic recombinational events (section 5.4.1), these findings suggest that evolution in these pig-adapted bacteria is dominated by genetic drift (random sampling of gene variants) and loss of functions instead of natural selection by recombination. Therefore, considering also the low genetic diversity, confirmed by a low mutation rate, YE serotype O:3 can be reasonably considered a genetically monomorphic or clonal microbial population, part of a more diverse species. The reduced genetic variation of serotype O:3 might be the result of a population bottleneck, due to transmission events and ecological specialization. Similarly, both *Y. pestis*, which typically colonizes fleas in order to be transmitted to mammals, and *S. enterica* serovar Typhi, a human-restricted pathogen, have been also designated as genetically monomorphic organisms (Achtman, 2008).

The weak signal of selection can be also noticed by the lack of antigenic variation, defined as the diversification of mainly surface-exposed antigens for escaping the immune response mechanisms (van der Woude and Baumler, 2004). In fact, only few genes, which are metabolism and transport-related, have independent nonsynonymous variants in nearby amino acids (Table 32). Although metabolic and transport functions seem to be important for specific serotype O:3 antigenic proteins (section 4.4.3) and, therefore, the detected genes may represent cases of antigenic variation, the level of variability is low. The lack of evidence for immune evasion suggests that YE serotype O:3 is not under selective pressure from the human immune system, being humans not its usual habitat.

The main source of YE serotype O:3 infections has been often suspected to be contaminated pork products, but the connection between human and animal isolates is not well established. Several DNA-based typing molecular methods have indicated an overall limited genetic diversity of these strains and a high similarity between human and pig isolates (Fredriksson-Ahomaa et al., 2006), supporting the hypothesis that asymptomatic pigs are the main reservoir for serotype O:3 strains. The SNP-based phylogeny obtained in this study seems to confirm this assumption. In fact, no SNPs allow a distinction between the two groups of pig/animal and human isolates. In addition, different phylogenetic ages are evident, with animal strains lying mainly on internal branches, because of old mutations, and the human ones on terminal branches, denoting recent mutations (Figure 32). The patterns of genetic drift, low level of recombination/gene flow and lack of antigenic variations may thus be explained by the fact that pigs are a restricted and persistent reservoir for YE serotype O:3 strains. Mutations occurring during a human infection might have no fitness advantage in the animal carrier condition and may not be maintained in the population.

Importantly, the representative strain collection used in this analysis has been chosen on the basis of multiple criteria, intending to maximize the amount of variation present in the YE serotype O:3 population. The depicted discernible diversity is thus limited by the selected strains, and some subgroupings may have been excluded. Nevertheless, it is unlikely that the variant calls and the identified phylogenetic structure greatly differ from the reality. Indeed, variant detection has been performed on the nonrepetitive component of the genome, thus excluding phages, insertion sequences and rRNAs, since assembly and mapping of short reads are unreliable in repetitive regions, and mobile elements may be under different selective pressure compared to the backbone genome. Only high quality variants have been used for the analysis, while indels have been omitted from the SNP-based phylogeny calculation, to reduce possible errors. SNPs, indeed, are the most valuable markers for studying phylogenetic relationships of isolates in homogenous pathogens, since they have relatively low mutation rates and are evolutionary stable (Achtman et al., 2004; Pearson et al., 2004). The discovered SNPs can be validated with genotyping approaches using a larger collection of strains, in order to develop a molecular SNP-based typing assay, capable of discrimination to the strain level. This may lead to an improved tracking of the spread of YE serotype O:3 between swine and human hosts.

5.6. Gene expressed during *Y. enterocolitica* serotype O:3 infection

Y. enterocolitica is able to respond to environmental stimuli by inducing or repressing genes, for example to optimize proliferation within the host during infection. Genes specifically induced during *in vivo* growth within the host may have a role in pathogenesis, being virulence or fitness factors. Techniques such as IVET (*in vivo* expression technology) and STM (signature-tagged mutagenesis) have been applied to the highly-virulent strain 8081 in order to identify novel virulence genes (Darwin and Miller, 1999; Gort and Miller, 2000; Young and Miller, 1997). In contrast, little is known about virulence factors responsible for the apparent adaptation to the swine of YE serotype O:3 and its ability to cause disease in humans. IVET and STM methods require an animal model, usually mice, which may not satisfactorily represent events during growth in humans or swine. To study serotype O:3 pathogenesis, the IVIAT (*in vivo* induced antigen technology) technique seemed to be a suitable alternative approach (Figure 33), and provided a portrait of possible mechanisms by which these bacteria can defeat host defenses and establish itself within the host.

Applying the immuno-based IVIAT screening to the completely sequenced YE strain Y11 identified 45 immunogenic antigens by probing sera from humans with serotype O:3 infections.

The detected IVI genes are mainly involved in metabolic and biosynthetic functions, similarly to the results from the IVET and STM studies with strain 8081, but also in “classical” virulence activities, such as motility and adhesion (Table 33). Overall, the IVIAT results largely confirm the conclusions drawn by the genome comparative analyses (section 5.2), emphasizing even more the importance of bacterial physiology during host infection.

As suggested by genomic findings, growth in anaerobic conditions seems to play a major role in *in vivo* proliferation of YE serotype O:3, since genes belonging to cobalamin synthesis, nitrogen and sulfur metabolism pathways resulted as IVI antigens. The proposed function of BCAA in YE virulence (section 5.2.2) is supported by the LivK BCAA transporter, which was discovered by IVIAT. Being absent from the highly-virulent group, this protein might play a role in the host-restricted phenotype of serotype O:3 strains. Future characterization of these metabolic-related IVI genes would be of great interest for developing efficient antibacterial strategies against *Yersinia*, as well as for better understanding how these bacteria compete for nutrients, using alternative carbon sources.

Among the IVI antigens with known virulence functions in YE, the plasmid-encoded YscU undergoes an autoproteolysis which generates a C-terminal peptide responsible for correct regulation of Yops synthesis and secretion (Bjornfot et al., 2009). The antigenic region identified by IVIAT within the *yscU* gene corresponds to the C-ter domain, confirming the essential role of the pYV-encoded T3SS in *Yersinia* virulence and supporting the validity of the IVIAT screening. Notably, none of the other Ysc or Yop proteins has been detected, probably because of their expression in both *in vitro* and *in vivo* conditions, with consequent exclusion by the IVIAT approach. A large functional category of putative IVI virulence factors include flagella- and fimbria-associated proteins. These results are consistent with findings from various experiments, in which flagella appeared to be required for migration and initiation of host cell contact, for epithelial cell invasion and for virulence in a mouse model (Young, 2004). Even though the expression of flagella in YE is known to be decreased at 37 °C, temperature regulation seems to be primarily centered on repression of class III genes (Horne and Pruss, 2006). The detected IVI genes *fliN*, *fliO*, *fliP* and *fliQ* belong to the class I and remain expressed at the host temperature (Minnich and Rohde, 2007), confirming their role as immuno-antigens. The identification of Mrf fimbrial IVI antigens is also of particular interest, as genomic analyses have proposed a yet underestimated function in pathogenesis of these adhesion proteins (section 5.2.2). The role of MrfC in YE serotype O:3 infection is supported by the fact that both *stbC* and the adjacent *stbD* fimbrial genes have been involved in *S. typhimurium* virulence (Chuang et al., 2008), and that four fimbrial proteins were identified by IVIAT in *S. paratyphi* (Alam et al., 2013).

In general, different IVI antigens discovered in YE serotype O:3 have similar function as genes previously detected in other IVIAT studies, reinforcing the validity of the method. For example, penicillin-binding proteins have been identified in the chicken-adapted *S. pullorum* (Li et al., 2013) and in *Bacillus anthracis* (Rollins et al., 2008). Sensor kinases resulted to be IVI antigens in the EHEC *E. coli* O157:H7 (John et al., 2005), in the swine pathogen *Streptococcus suis* (Gu et al., 2009), in *C. jejuni* probed with chicken sera (Hu et al., 2014), and in *S. pullorum*. DNA recombination proteins, helicases and endonucleases have been detected in the zoonotic agents *S. pullorum* and *C. jejuni* and in *E. coli* O157:H7.

In conclusion, IVIAT has been applied to identify proteins specifically expressed *in vivo* during both human and animal infection with a variety of pathogens. However, this technique has some limitations, such as (i) exclusion of potentially important antigens expressed both *in vitro* and *in vivo* and of non-proteic antigens; (ii) cross-reactivity with homologous antigens from other bacteria; (iii) variable expression due to the genomic library in an *E. coli* host. Determining the real roles of the detected genes in pathogenesis requires additional studies; however, the IVI genes, especially if located on the bacterial surface, may be used for development of diagnostic tests and vaccines.

6. Conclusions

The species *Y. enterocolitica* contains a very heterogeneous collection of organisms. One end of the spectrum is formed by broad-host-range pathogens such as the highly-virulent biotype 1B strains, which are not only frequently associated with yersiniosis in humans, but are also isolated with water and rodents. At the other end of the spectrum there are pathogens of serotype O:3, whose ecological habitat is mainly restricted to humans and pigs.

Whole genome comparisons provided clear evidences of the wide genetic variability within the highly-virulent group and the whole YE species, in contrast to the genetic isolation of serotype O:3 strains. Genomic insights into the microevolution of the YE serotype O:3 population showed their clonal-like nature and suggested the definition of “genetically monomorphic” for these strains. Recombinational events and adaptive selection play little role in serotype O:3 evolution, which is rather driven by genetic drift, while horizontal gene transfer seems to be the main evolutionary process for the entire YE species. These findings reflect the different degree of niche specialization of this heterogeneous group of microorganisms, and clearly demonstrate the host restriction of the pig-adapted serotype O:3.

Both genome analyses and the IVIAT immuno-screening highlighted the putative essential role of anaerobic respiration abilities of the facultative anaerobic YE and, especially, of serotype O:3 strains, during host colonization and/or pathogenesis. Despite the need of experimental proofs, it can be postulated that the capacity of growing anaerobically might bring an advantage to *Yersinia* for surviving and competing with microbial communities encountered in host niches (gut, tonsils) and in the environment (water, soil). Moreover, considering that anaerobic pathways are less efficient than the aerobic ones, they might be used by YE serotype O:3 more for survival and colonization than for invasion. This is also supported by the fact that the main virulence factors in the YE weakly-virulent group are fimbriae and transporters, having essentially functions in adhesion and contact with host cells.

A thorough knowledge on the epidemiology of serotype O:3 isolates might unquestionably confirm the hypothesis that pigs are asymptomatic reservoir, with human infections being a mere consequence of contaminated-food consumption. Selection of informative SNPs evenly distributed through the chromosome and localized in core genes will allow the design of a rapid and cost-effective diagnostic scheme for strain genotyping.

7. References

Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62, 53-70.

Achtman, M., 2012. Insights from genomic comparisons of genetically monomorphic bacterial pathogens. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 367, 860-867.

Achtman, M., Morelli, G., Zhu, P., Wirth, T., Diehl, I., Kusecek, B., Vogler, A.J., Wagner, D.M., Allender, C.J., Easterday, W.R., Chenal-Francisque, V., Worsham, P., Thomson, N.R., Parkhill, J., Lindler, L.E., Carniel, E., Keim, P., 2004. Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc Natl Acad Sci U S A 101, 17837-17842.

Aguero, M.E., Aron, L., DeLuca, A.G., Timmis, K.N., Cabello, F.C., 1984. A plasmid-encoded outer membrane protein, TraT, enhances resistance of *Escherichia coli* to phagocytosis. Infect Immun 46, 740-746.

al-Hendy, A., Toivanen, P., Skurnik, M., 1992. Lipopolysaccharide O side chain of *Yersinia enterocolitica* O:3 is an essential virulence factor in an orally infected murine model. Infect Immun 60, 870-875.

Alam, M.M., Tsai, L.L., Rollins, S.M., Sheikh, A., Khanam, F., Bufano, M.K., Yu, Y., Wu-Freeman, Y., Kalsy, A., Sultana, T., Sayeed, M.A., Jahan, N., LaRocque, R.C., Harris, J.B., Leung, D.T., Brooks, W.A., Calderwood, S.B., Charles, R.C., Qadri, F., Ryan, E.T., 2013. Identification of in vivo-induced bacterial proteins during human infection with *Salmonella enterica* serotype Paratyphi A. Clinical and vaccine immunology : CVI 20, 712-719.

Aleksic, S., Bockemuhl, J., Lange, F., 1986. Studies on the serology of flagellar antigens of *Yersinia enterocolitica* and related *Yersinia* species. Zentralblatt fur Bakteriologie, Mikrobiologie, und Hygiene. Series A, Medical microbiology, infectious diseases, virology, parasitology 261, 299-310.

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A., 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12, 402.

Allaoui, A., Woestyn, S., Sluiters, C., Cornelis, G.R., 1994. YscU, a *Yersinia enterocolitica* inner membrane protein involved in Yop secretion. J Bacteriol 176, 4534-4542.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403-410.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics 25, 25-29.

Asplund, K., Johansson, T., Siitonen, A., 1998. Evaluation of pulsed-field gel electrophoresis of genomic restriction fragments in the discrimination of *Yersinia enterocolitica* O:3. Epidemiol Infect 121, 579-586.

Aulisio, C.C., Mehlman, I.J., Sanders, A.C., 1980. Alkali method for rapid recovery of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* from foods. Appl Environ Microbiol 39, 135-140.

Autenrieth, I.B., Firsching, R., 1996. Penetration of M cells and destruction of Peyer's patches by *Yersinia enterocolitica*: an ultrastructural and histological study. J Med Microbiol 44, 285-294.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9, 75.

Badger, J.L., Miller, V.L., 1998. Expression of invasin and motility are coordinately regulated in *Yersinia enterocolitica*. J Bacteriol 180, 793-800.

Balligand, G., Laroche, Y., Cornelis, G., 1985. Genetic analysis of virulence plasmid from a serogroup 9 *Yersinia enterocolitica* strain: role of outer membrane protein P1 in resistance to human serum and autoagglutination. Infect Immun 48, 782-786.

Baltes, N., Hennig-Pauka, I., Jacobsen, I., Gruber, A.D., Gerlach, G.F., 2003. Identification of dimethyl sulfoxide reductase in *Actinobacillus pleuropneumoniae* and its role in infection. Infect Immun 71, 6784-6792.

Baltes, N., N'Diaye, M., Jacobsen, I.D., Maas, A., Buettner, F.F., Gerlach, G.F., 2005. Deletion of the anaerobic regulator HlyX causes reduced colonization and persistence of *Actinobacillus pleuropneumoniae* in the porcine respiratory tract. Infect Immun 73, 4614-4619.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., Kim, J.F., 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461, 1243-1247.

Basavanna, S., Khandavilli, S., Yuste, J., Cohen, J.M., Hosie, A.H., Webb, A.J., Thomas, G.H., Brown, J.S., 2009. Screening of *Streptococcus pneumoniae* ABC transporter mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. Infect Immun 77, 3412-3423.

Batzilla, J., Antonenka, U., Höper, D., Heesemann, J., Rakin, A., 2011a. *Yersinia enterocolitica palearctica* serobiotype O:3/4 - a successful group of emerging zoonotic pathogens. BMC Genomics 6, 348.

Batzilla, J., Heesemann, J., Rakin, A., 2011b. The pathogenic potential of *Yersinia enterocolitica* 1A. Int J Med Microbiol 301, 556-561.

Batzilla, J., Hoper, D., Antonenka, U., Heesemann, J., Rakin, A., 2011c. Complete genome sequence of *Yersinia enterocolitica* subsp. *palearctica* serogroup O:3. J Bacteriol 193, 2067.

Benz, I., Schmidt, M.A., 2011. Structures and functions of autotransporter proteins in microbial pathogens. Int J Med Microbiol 301, 461-468.

Berzero, R., Volterra, L., Pacifico, L., Chiesa, C., 1991. Isolation of yersiniae from sewage. Contrib Microbiol Immunol 12, 40-43.

Bhaduri, S., 2006. Enrichment, isolation, and virulence of freeze-stressed plasmid-bearing virulent strains of *Yersinia enterocolitica* on pork. J Food Prot 69, 1983-1985.

Bhaduri, S., Wesley, I.V., Bush, E.J., 2005. Prevalence of pathogenic *Yersinia enterocolitica* strains in pigs in the United States. Appl Environ Microbiol 71, 7117-7121.

Bhagat, N., Virdi, J.S., 2009. Molecular and biochemical characterization of urease and survival of *Yersinia enterocolitica* biovar 1A in acidic pH in vitro. BMC Microbiol 9, 262.

Bhagat, N., Virdi, J.S., 2011. The enigma of *Yersinia enterocolitica* biovar 1A. Crit Rev Microbiol 37, 25-39.

Bieber, D., Ramer, S.W., Wu, C.Y., Toru, M.W.J.T., Fernandez, R., Schoolnik, G.K., 1998. Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. Science 280, 2114–2118.

Bjornfot, A.C., Lavander, M., Forsberg, A., Wolf-Watz, H., 2009. Autoproteolysis of YscU of *Yersinia pseudotuberculosis* is important for regulation of expression and secretion of Yop proteins. J Bacteriol 191, 4259-4267.

Black, R.E., Jackson, R.J., Tsai, T., Medvesky, M., Shayegani, M., Feeley, J.C., MacLeod, K.I., Wakelee, A.M., 1978. Epidemic *Yersinia enterocolitica* infection due to contaminated chocolate milk. The New England journal of medicine 298, 76-79.

Blixt, Y., Knutsson, R., Borch, E., Radstrom, P., 2003. Interlaboratory random amplified polymorphic DNA typing of *Yersinia enterocolitica* and *Y. enterocolitica*-like bacteria. Int J Food Microbiol 83, 15-26.

Blom, J., Albaum, S.P., Doppmeier, D., Puhler, A., Vorholter, F.J., Zakrzewski, M., Goesmann, A., 2009. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC bioinformatics 10, 154.

Bodelon, G., Palomino, C., Fernandez, L.A., 2013. Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. FEMS Microbiol Rev 37, 204-250.

Bogdanovich, T., Carniel, E., Fukushima, H., Skurnik, M., 2003. Use of O-antigen gene cluster-specific PCRs for the identification and O-genotyping of *Yersinia pseudotuberculosis* and *Yersinia pestis*. J Clin Microbiol 41, 5103-5112.

Boland, A., Havaux, S., Cornelis, G.R., 1998. Heterogeneity of the *Yersinia* YopM protein. Microb Pathog 25, 343-348.

Bonardi, S., Alpigiani, I., Pongolini, S., Morganti, M., Tagliabue, S., Bacci, C., Brindani, F., 2014. Detection, enumeration and characterization of *Yersinia enterocolitica* 4/O:3 in pig tonsils at slaughter in Northern Italy. Int J Food Microbiol 177C, 9-15.

Bonardi, S., Bassi, L., Brindani, F., D'Incau, M., Barco, L., Carra, E., Pongolini, S., 2013. Prevalence, characterization and antimicrobial susceptibility of *Salmonella enterica* and *Yersinia enterocolitica* in pigs at slaughter in Italy. Int J Food Microbiol 163, 248-257.

Bosak, J., Laiblova, P., Smarda, J., Dedicova, D., Smajs, D., 2012. Novel colicin Fy of *Yersinia frederiksenii* inhibits pathogenic *Yersinia* strains via YiuR-mediated reception, TonB import, and cell membrane pore formation. J Bacteriol 194, 1950-1959.

Bosak, J., Micenkova, L., Vrba, M., Sevcikova, A., Dedicova, D., Garzetti, D., Smajs, D., 2013. Unique activity spectrum of colicin FY: all 110 characterized *Yersinia enterocolitica* isolates were colicin FY susceptible. PLoS One 8, e81829.

Bose, M., Barber, R.D., 2006. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. In silico biology 6, 223-227.

Bottone, E.J., 1997. *Yersinia enterocolitica*: The Charisma Continues. Clin Microbiol Rev 10, 257-276.

Bottone, E.J., 1999. *Yersinia enterocolitica*: overview and epidemiologic correlates. Microb Infect 1, 323-333.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry 72, 248-254.

Bresolin, G., Trcek, J., Scherer, S., Fuchs, T.M., 2008. Presence of a functional flagellar cluster Flag-2 and low-temperature expression of flagellar genes in *Yersinia enterocolitica* W22703. Microbiology 154, 196-206.

Brodsky, I.E., Palm, N.W., Sadanand, S., Ryndak, M.B., Sutterwala, F.S., Flavell, R.A., Bliska, J.B., Medzhitov, R., 2010. A *Yersinia* effector protein promotes virulence by preventing inflammasome recognition of the type III secretion system. Cell Host Microbe 7, 376-387.

Brubaker, R.R., 1991. Factors promoting acute and chronic diseases caused by yersiniae. Clin Microbiol Rev 4, 309-324.

Bucher, M., Meyer, C., Grötzbach, B., Wacheck, S., Stolle, A., Fredriksson-Ahomaa, M., 2008. Epidemiological data on pathogenic *Yersinia enterocolitica* in Southern Germany during 2000-2006. Foodborne Pathog Dis 5, 273-280.

Buchrieser, C., Weagant, S.D., Kaspar, C.W., 1994. Molecular characterization of *Yersinia enterocolitica* by pulsed-field gel electrophoresis and hybridization of DNA fragments to *ail* and pYV probes. Appl Environ Microbiol 60, 4371-4379.

Burnens, A.P., Frey, A., Nicolet, J., 1996. Association between clinical presentation, biogroups and virulence attributes of *Yersinia enterocolitica* strains in human diarrhoeal disease. Epidemiol Infect 116, 27-34.

Carniel, E., 2006. Yersinia enterocolitica - Epidemiology, in: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., Stackebrandt, E. (Eds.), The Prokaryotes, Third ed. Vol. 6, 3.3.13: *Y. enterocolitica* and *Y. pseudotuberculosis*. Springer Science+Business Media, USA, pp. 271-276.

Carniel, E., Guilvout, I., Prentice, M., 1996. Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. J Bacteriol 178, 6743-6751.

Carter, M.Q., Louie, J.W., Fagerquist, C.K., Sultan, O., Miller, W.G., Mandrell, R.E., 2012. Evolutionary silence of the acid chaperone protein HdeB in enterohemorrhagic *Escherichia coli* O157:H7. Appl Environ Microbiol 78, 1004-1014.

Cascales, E., Buchanan, S.K., Duché, D., Kleanthous, C., Lloubès, R., Postle, K., Riley, M., Slatin, S., Cavard, D., 2007. Colicin Biology. Microbiol Mol Biol Rev 71, 158-229.

Caugant, D.A., Aleksic, S., Mollaret, H.H., Selander, R.K., Kapperud, G., 1989. Clonal diversity and relationships among strains of *Yersinia enterocolitica*. J Clin Microbiol 27, 2678-2683.

CDC, 2012. Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet Surveillance Report for 2011 (Final Report). Atlanta, Georgia: U.S. Department of Health and Human Services, CDC.

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., Brubaker, R.R., Fowler, J., Hinnebusch, J., Marceau, M., Medigue, C., Simonet, M., Chenal-Francisque, V., Souza, B., Dacheux, D., Elliott, J.M., Derbise, A., Hauser, L.J., Garcia, E., 2004. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 101, 13826-13831.

Chandler, N.D., Parisi, M.T., 1994. Radiological case of the month. *Yersinia enterocolitica* masquerading as appendicitis. Archives of pediatrics & adolescent medicine 148, 527-528.

China, B., Michiels, T., Cornelis, G.R., 1990. The pYV plasmid of *Yersinia* encodes a lipoprotein, YlpA, related to TraT. Mol Microbiol 4, 1585-1593.

Chuang, Y.C., Wang, K.C., Chen, Y.T., Yang, C.H., Men, S.C., Fan, C.C., Chang, L.H., Yeh, K.S., 2008. Identification of the genetic determinants of *Salmonella enterica* serotype Typhimurium that may regulate the expression of the type 1 fimbriae in response to solid agar and static broth culture conditions. BMC Microbiol 8, 126.

Collyn, F., Billault, A., Mullet, C., Simonet, M., Marceau, M., 2004. YAPI, a new *Yersinia pseudotuberculosis* pathogenicity island. Infect Immun 72, 4784-4790.

Cork, S.C., Marshall, R.B., Madie, P., Fenwick, S.G., 1995. The role of wild birds and the environment in the epidemiology of *Yersiniae* in New Zealand. New Zealand veterinary journal 43, 169-174.

Cornelis, G., Colson, C., 1975. Restriction of DNA in *Yersinia enterocolitica* detected by recipient ability for a derepressed R factor from *Escherichia coli*. J Gen Microbiol 87, 285-291.

Cornelis, G.R., 2002. The *Yersinia* Ysc-Yop 'type III' weaponry. Nat Rev Mol Cell Biol 3, 742-752.

Cornelis, G.R., 2010. The type III secretion injectisome, a complex nanomachine for intracellular 'toxin' delivery. Biol Chem 391, 745-751.

Cornelis, G.R., Boland, A., Boyd, A.P., Geuijen, C., Iriarte, M., Neyt, C., Sory, M.P., Stainier, I., 1998. The Virulence Plasmid of *Yersinia*, an Antihost Genome. Microbiol Mol Biol Rev 62, 1315-1352.

Cornelis, G.R., Wolf-Watz, H., 1997. The *Yersinia* Yop virulon: a bacterial system for subverting eukaryotic cells. Mol Microbiol 23, 861-867.

D'Ari, R., Lin, R.T., Newman, E.B., 1993. The leucine-responsive regulatory protein: more than a regulator? Trends Biochem Sci 18, 260-263.

Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS One 5, e11147.

Darwin, A.J., Miller, V.L., 1999. Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis. Mol Microbiol 32, 51-62.

Dautin, N., Bernstein, H.D., 2007. Protein secretion in gram-negative bacteria via the autotransporter pathway. Annu. Rev. Microbiol. 61, 89-112.

de Almeida, A.M., Guiyoule, A., Guilvout, I., Iteman, I., Baranton, G., Carniel, E., 1993. Chromosomal *irp2* gene in *Yersinia*: distribution, expression, deletion and impact on virulence. Microb Pathog 14, 9-21.

de Benito, I., Cano, M.E., Aguero, J., Garcia Lobo, J.M., 2004. A polymorphic tandem repeat potentially useful for typing in the chromosome of *Yersinia enterocolitica*. Microbiology 150, 199-204.

de Boer, E., 1992. Isolation of *Yersinia enterocolitica* from foods. Int J Food Microbiol 17, 75-84.

de Koning-Ward, T.F., Robins-Browne, R.M., 1996. Analysis of the urease gene complex of members of the genus *Yersinia*. Gene 182, 225-228.

de Koning-Ward, T.F., Robins-Browne, R.M., 1997. A novel mechanism of urease regulation in *Yersinia enterocolitica*. FEMS Microbiol Lett 147, 221-226.

de Koning-Ward, T.F., Ward, A.C., Robins-Browne, R.M., 1994. Characterisation of the urease-encoding gene complex of *Yersinia enterocolitica*. Gene 145, 25-32.

Deng, X., Phillippy, A.M., Li, Z., Salzberg, S.L., Zhang, W., 2010. Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification. BMC Genomics 11, 500.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics 43, 491-498.

Deuretzbacher, A., Czymmeck, N., Reimer, R., Trülzsch, K., Gaus, K., Hohenberg, H., Heesemann, J., Aepfelbacher, M., Ruckdeschel, K., 2009. Beta1 integrin-dependent engulfment of *Yersinia enterocolitica* by macrophages is coupled to the activation of autophagy and suppressed by type III protein secretion. J Immunol 183, 5847-5860.

Dewoody, R., Merritt, P.M., Marketon, M.M., 2013a. YopK controls both rate and fidelity of Yop translocation. Mol Microbiol 87, 301-317.

Dewoody, R.S., Merritt, P.M., Marketon, M.M., 2013b. Regulation of the *Yersinia* type III secretion system: traffic control. Front Cell Infect Microbiol 3.

Dobrindt, U., Hacker, J., 2001. Whole genome plasticity in pathogenic bacteria. Curr Opin Microbiol 4, 550-557.

Dolina, M., Peduzzi, R., 1993. Population genetics of human, animal, and environmental *Yersinia* strains. Appl Environ Microbiol 59, 442-450.

Duan, R., Liang, J., Shi, G., Cui, Z., Hai, R., Wang, P., Xiao, Y., Li, K., Qiu, H., Gu, W., Du, X., Jing, H., Wang, X., 2014. Homology analysis of pathogenic Yersinia species *Yersinia enterocolitica*, *Yersinia pseudotuberculosis*, and *Yersinia pestis* based on multilocus sequence typing. J Clin Microbiol 52, 20-29.

Dube, P., 2009. Interaction of *Yersinia* with the gut: mechanisms of pathogenesis and immune evasion. Curr Top Microbiol Immunol 337, 61-91.

ECDC, 2013. Annual Epidemiological Report 2013. Reporting on 2011 surveillance data and 2012 epidemic intelligence data. ECDC.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

EFSA, ECDC, 2013. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2011. EFSA Journal 11, 250.

El Tahir, Y., Skurnik, M., 2001. YadA, the multifaceted *Yersinia* adhesin. Int J Med Microbiol 291, 209-218.

Fearnley, C., On, S.L., Kokotovic, B., Manning, G., Cheasty, T., Newell, D.G., 2005. Application of fluorescent amplified fragment length polymorphism for comparison of human and animal isolates of *Yersinia enterocolitica*. Appl Environ Microbiol 71, 4960-4965.

Felsenstein, J., 1989. Phylip: Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.

Ferber, D.M., Brubaker, R.R., 1979. The effects of pesticin on *Yersinia enterocolitica* and *Y. pseudotuberculosis*. Contrib Microbiol Immunol 5, 366-371.

Fetherston, J.D., Perry, R.D., 1994. The pigmentation locus of *Yersinia pestis* KIM6+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. Mol Microbiol 13, 697-708.

Finn, R.D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L., Bateman, A., 2006. Pfam: clans, web tools and services. Nucleic Acids Res 34, D247-251.

Foultier, B., Cornelis, G.R., 2003. DNA sequence and analysis of the pYVa127/90 virulence plasmid of *Yersinia enterocolitica* strain A127/90. Research in microbiology 154, 553-557.

Foultier, B., Troisfontaines, P., Müller, S., Opperdoes, F.R., Cornelis, G.R., 2002. Characterization of the *ysa* pathogenicity locus in the chromosome of *Yersinia enterocolitica* and phylogeny analysis of type III secretion systems. J Mol Evol 55, 37-55.

Foultier, B., Troisfontaines, P., Vertommen, D., Marenne, M.N., Rider, M., Parsot, C., Cornelis, G.R., 2003. Identification of substrates and chaperone from the *Yersinia enterocolitica* 1B Ysa type III secretion system. Infect Immun 71, 242-253.

Fredriksson-Ahomaa, M., 2007. *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*, in: Simjee, S. (Ed.), Infectious disease: Foodborne diseases. Humana Press Inc., Totowa, NJ, USA.

Fredriksson-Ahomaa, M., Bucher, M., Hank, C., Stolle, A., Korkeala, H., 2001. High prevalence of *Yersinia enterocolitica* 4:O3 on pig offal in southern Germany: a slaughtering technique problem. Syst Appl Microbiol 24, 457-463.

Fredriksson-Ahomaa, M., Hielm, S., Korkeala, H., 1999. High prevalence of *yadA*-positive *Yersinia enterocolitica* in pig tongues and minced meat at the retail level in Finland. J Food Prot 62, 123-127.

Fredriksson-Ahomaa, M., Korkeala, H., 2003. Low Occurrence of Pathogenic *Yersinia enterocolitica* in Clinical, Food, and Environmental Samples: a Methodological Problem. Clin Microbiol Rev 16, 220-229.

Fredriksson-Ahomaa, M., Meyer, C., Bonke, R., Stüber, E., Wacheck, S., 2010. Characterization of *Yersinia enterocolitica* 4/O:3 isolates from tonsils of Bavarian slaughter pigs. Lett Appl Microbiol 50, 412-418.

Fredriksson-Ahomaa, M., Murros-Kontiainen, A., Sade, E., Puolanne, E., Bjorkroth, J., 2012. High number of *Yersinia enterocolitica* 4/O:3 in cold-stored modified atmosphere-packed pig cheek meat. Int J Food Microbiol 155, 69-72.

Fredriksson-Ahomaa, M., Niskanen, T., Bucher, M., Korte, T., Stolle, A., Korkeala, H., 2003. Different *Yersinia enterocolitica* 4:O3 genotypes found in pig tonsils in Southern Germany and Finland. Syst Appl Microbiol 26, 132-137.

Fredriksson-Ahomaa, M., Stolle, A., Korkeala, H., 2006. Molecular epidemiology of *Yersinia enterocolitica* infections. FEMS Immunol Med Microbiol 47, 315-329.

Fuchs, T., Brandt, K., Starke, M., Rattei, T., 2011. Shotgun sequencing of *Yersinia enterocolitica* strain W22703 (biotype 2, serotype O:9): genomic evidence for oscillation between invertebrates and mammals. BMC Genomics 12, 1-14.

Fuchs, T.M., Bresolin, G., Marcinowski, L., Schachtner, J., Scherer, S., 2008. Insecticidal genes of *Yersinia* spp.: taxonomical distribution, contribution to toxicity towards Manduca sexta and Galleria mellonella, and evolution. BMC Microbiol 8.

Fukushima, H., Gomyoda, M., Aleksic, S., 1998. Genetic variation of *Yersinia enterocolitica* serotype O:9 strains detected in samples from western and eastern countries. Zentralbl Bakteriol 288, 167-174.

Fukushima, H., Gomyoda, M., Aleksic, S., Tsubokura, M., 1993. Differentiation of *Yersinia enterocolitica* serotype O:5,27 strains by phenotypic and molecular techniques. J Clin Microbiol 31, 1672-1674.

Fukushima, H., Shimizu, S., Inatsu, Y., 2011. *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* Detection in Foods. J Pathog 2011, 735308.

Garzetti, D., Bouabe, H., Heesemann, J., Rakin, A., 2012. Tracing genomic variations in two highly virulent *Yersinia enterocolitica* strains with unequal ability to compete for host colonization. BMC Genomics 13, 1-15.

Garzetti, D., Susen, R., Fruth, A., Tietze, E., Heesemann, J., Rakin, A., 2014. A molecular scheme for *Yersinia enterocolitica* patho-serotyping derived from genome-wide analysis. Int J Med Microbiol 304, 275-283.

Gierczyński, R., Golubov, A., Neubauer, H., Pham, J.N., Rakin, A., 2007. Development of multiple-locus variable-number tandem-repeat analysis for *Yersinia enterocolitica* subsp. palearctica and its application to bioserogroup 4/O3 subtyping. J Clin Microbiol 45, 2508-2515.

Godfroid, J., Saegerman, C., Wellemans, V., Walravens, K., Letesson, J.J., Tibor, A., Mc Millan, A., Spencer, S., Sanna, M., Bakker, D., Pouillot, R., Garin-Bastuji, B., 2002. How to substantiate eradication of bovine brucellosis when aspecific serological reactions occur in the course of brucellosis testing. Vet Microbiol 90, 461-477.

Goecks, J., Nekrutenko, A., Taylor, J., Galaxy, T., 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11, R86.

Gorshkova, R.P., Kalmykova, E.N., Isakov, V.V., Ovodov, Y.S., 1985. Structural studies on O-specific polysaccharides of lipopolysaccharides from *Yersinia enterocolitica* serovars O:1,2a,3, O:2a,2b,3 and O:3. European journal of biochemistry / FEBS 150, 527-531.

Gorshkova, R.P., Kalmykova, E.N., Isakov, V.V., Ovodov, Y.S., 1986. Structural studies on O-specific polysaccharides of lipopolysaccharides from *Yersinia enterocolitica* serovars O:5 and O:5,27. European journal of biochemistry / FEBS 156, 391-397.

Gort, A.S., Miller, V.L., 2000. Identification and characterization of *Yersinia enterocolitica* genes induced during systemic infection. Infect Immun 68, 6633-6642.

Goullet, P., Picard, B., 1988. Characterization of *Yersinia enterocolitica*, *Y. intermedia*, *Y. aldovae*, *Y. frederiksenii*, *Y. kristensenii* and *Y. pseudotuberculosis* by electrophoretic polymorphism of acid phosphatase, esterases, and glutamate and malate dehydrogenases. J Gen Microbiol 134, 317-325.

Grant, T., Bennett-Wood, V., Robins-Browne, R.M., 1999. Characterization of the interaction between *Yersinia enterocolitica* biotype 1A and phagocytes and epithelial cells in vitro. Infect Immun 67, 4367-4375.

Greenwood, M., Hooper, W.L., 1987. Human carriage of *Yersinia* spp. J Med Microbiol 23, 345-348.

Griebler, C., Slezak, D., 2001. Microbial activity in aquatic environments measured by dimethyl sulfoxide reduction and intercomparison with commonly used methods. Appl Environ Microbiol 67, 100-109.

Griffin, P., 2008. Yersiniosis, in: Heymann, D.L. (Ed.), Control of Communicable Diseases Manual, 19th ed. American Public Health Association.

Grijpstra, J., Arenas, J., Rutten, L., Tommassen, J., 2013. Autotransporter secretion: varying on a theme. Research in microbiology 164, 562-582.

Gripenberg-Lerche, C., Zhang, L., Ahtonen, P., Toivanen, P., Skurnik, M., 2000. Construction of urease-negative mutants of *Yersinia enterocolitica* serotypes O:3 and o:8: role of urease in virulence and arthritogenicity. Infect Immun 68, 942-947.

Grunstein, M., Hogness, D.S., 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci U S A 72, 3961-3965.

Gu, H., Zhu, H., Lu, C., 2009. Use of in vivo-induced antigen technology (IVIAT) for the identification of *Streptococcus suis* serotype 2 in vivo-induced bacterial protein antigens. BMC Microbiol 9, 201.

Guedon, E., Serror, P., Ehrlich, S.D., Renault, P., Delorme, C., 2001. Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in *Lactococcus lactis*. Mol Microbiol 40, 1227-1239.

Håkansson, S., Bergman, T., Vanooteghem, J.C., Cornelis, G., Wolf-Watz, H., 1993. YopB and YopD constitute a novel class of *Yersinia* Yop proteins. Infect Immun 61, 71-80.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95-98.

Haller, J.C., Carlson, S., Pederson, K.J., Pierson, D.E., 2000. A chromosomally encoded type III secretion pathway in *Yersinia enterocolitica* is important in virulence. Mol Microbiol 36, 1436-1446.

Harnett, N., Lin, Y.P., Krishnan, C., 1996. Detection of pathogenic *Yersinia enterocolitica* using the multiplex polymerase chain reaction. Epidemiol Infect 117, 59-67.

Heesemann, J., 1987. Chromosomal-encoded siderophores are required for mouse virulence of enteropathogenic *Yersinia* species. FEMS Microbiol Lett 48, 229-233.

Heesemann, J., Gross, U., Schmidt, N., Laufs, R., 1986a. Immunochemical analysis of plasmid-encoded proteins released by enteropathogenic *Yersinia* sp. grown in calcium-deficient media. Infect Immun 54, 561-567.

Heesemann, J., Grüter, L., 1987. Genetic evidence that the outer membraneprotein Yop1 of *Yersinia enterocolitica* mediates adherence and phagocytosis resistance to human epithelial cells. FEMS Microbiol Lett 40, 37-41.

Heesemann, J., Hantke, K., Vocke, T., Saken, E., Rakin, A., Stojiljkovic, I., Berner, R., 1993. Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65,000 Da and pesticin sensitivity. Mol Microbiol 8, 397-408.

Heesemann, J., Kalthoff, H., Koch, F., 1986b. Monoclonal antibodies directed against plasmid-encoded released proteins of enteropathogenic *Yersinia*. FEMS Microbiol Lett 36, 15-19.

Heesemann, J., Laufs, R., 1983. Construction of a mobilizable *Yersinia enterocolitica* virulence plasmid. J Bacteriol 155, 761-767.

Heise, T., Dersch, P., 2006. Identification of a domain in *Yersinia* virulence factor YadA that is crucial for extracellular matrix-specific cell adhesion and uptake. Proc Natl Acad Sci U S A 103, 3375-3380.

Holmström, A., Petterson, J., Rosqvist, R., Håkansson, S., Tafazoli, F., Fällman, M., Magnusson, K.E., Wolf-Watz, H., Forsberg, A., 1997. YopK of *Yersinia pseudotuberculosis* controls translocation of Yop effectors across the eukaryotic cell membrane. Mol Microbiol 24, 73-91.

Hong, W., Wu, Y.E., Fu, X., Chang, Z., 2012. Chaperone-dependent mechanisms for acid resistance in enteric bacteria. Trends Microbiol 20, 328-335.

Horne, S.M., Pruss, B.M., 2006. Global gene regulation in *Yersinia enterocolitica*: effect of FliA on the expression levels of flagellar and plasmid-encoded virulence genes. Archives of microbiology 185, 115-126.

Hsiao, W.W., Ung, K., Aeschliman, D., Bryan, J., Finlay, B.B., Brinkman, F.S., 2005. Evidence of a large novel gene pool associated with prokaryotic genomic islands. PLoS Genet 1, e62.

Hu, Y., Huang, J., Jiao, X.A., 2014. Screening of genes expressed in vivo during interaction between chicken and *Campylobacter jejuni*. Journal of microbiology and biotechnology 24, 217-224.

Huovinen, E., Sihvonen, L.M., Virtanen, M.J., Haukka, K., Siitonen, A., Kuusi, M., 2010. Symptoms and sources of *Yersinia enterocolitica*-infection: a case-control study. BMC Infect Dis 10.

Iriarte, M., Stainier, I., Mikulskis, A.V., Cornelis, G.R., 1995. The *fliA* gene encoding sigma 28 in *Yersinia enterocolitica*. J Bacteriol 177, 2299-2304.

Iriarte, M., Vanooteghem, J.C., Delor, I., Díaz, R., Knutton, S., Cornelis, G.R., 1993. The Myf fibrillae of *Yersinia enterocolitica*. Mol Microbiol 9, 507-520.

Iteman, I., Guiyoule, A., Carniel, E., 1996. Comparison of three molecular methods for typing and subtyping pathogenic *Yersinia enterocolitica* strains. J Med Microbiol 45, 48-56.

Iwobi, A., Heesemann, J., Garcia, E., Igwe, E., Noelting, C., Rakin, A., 2003. Novel Virulence-Associated Type II Secretion System Unique to High-Pathogenicity *Yersinia enterocolitica*. Infect Immun 71, 1872-1879.

Jacobsen, A., Hendriksen, R.S., Aaresturp, F.M., Ussery, D.W., Friis, C., 2011. The *Salmonella enterica* pan-genome. Microbial ecology 62, 487-504.

Jacobsen, N.R., Bogdanovich, T., Skurnik, M., Lubeck, P.S., Ahrens, P., Hoorfar, J., 2005. A real-time PCR assay for the specific identification of serotype O:9 of *Yersinia enterocolitica*. J Microbiol Methods 63, 151-156.

John, M., Kudva, I.T., Griffin, R.W., Dodson, A.W., McManus, B., Krastins, B., Sarracino, D., Progulske-Fox, A., Hillman, J.D., Handfield, M., Tarr, P.I., Calderwood, S.B., 2005. Use of in vivo-induced antigen technology for identification of *Escherichia coli* O157:H7 proteins expressed during human infection. Infect Immun 73, 2665-2679.

Jolley, K.A., Maiden, M.C., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC bioinformatics 11, 595.

Jones, T.F., 2003. From pig to pacifier: chitterling-associated yersiniosis outbreak among black infants. Emerg Infect Dis 9, 1007-1009.

Kandolo, K., Wauters, G., 1985. Pyrazinamidase activity in *Yersinia enterocolitica* and related organisms. J Clin Microbiol 21, 980-982.

Kaneko, S., Ishizaki, N., Kokubo, Y., 1995. Detection of pathogenic *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* from pork using the polymerase chain reaction. Contrib Microbiol Immunol 13, 153-155.

Kapatral, V., Minnich, S.A., 1995. Co-ordinate, temperature-sensitive regulation of the three *Yersinia enterocolitica* flagellin genes. Mol Microbiol 17, 49-56.

Kapatral, V., Olson, J.W., Pepe, J.C., Miller, V.L., Minnich, S.A., 1996. Temperature-dependent regulation of *Yersinia enterocolitica* Class III flagellar genes. Mol Microbiol 19, 1061-1071.

Kapperud, G., 1991. *Yersinia enterocolitica* in food hygiene. Int J Food Microbiol 12, 53-65.

Kapperud, G., Nesbakken, T., Aleksic, S., Mollaret, H.H., 1990. Comparison of restriction endonuclease analysis and phenotypic typing methods for differentiation of *Yersinia enterocolitica* isolates. J Clin Microbiol 28, 1125-1131.

Keet, E.E., 1974. *Yersinia enterocolitica* septicemia. Source of infection and incubation period identified. New York state journal of medicine 74, 2226-2230.

Klinzing, D.C., Matias, R.R., Skowronski, E., Alvarez, M., Liles, V., Dimamay, M.P., Natividad, F.F., 2012. Shotgun genome sequence of a *Yersinia enterocolitica* isolate from the Philippines. J Bacteriol 194, 542-543.

Korte, T., Fredriksson-Ahomaa, M., Niskanen, T., Korkeala, H., 2004. Low prevalence of *yadA*-positive *Yersinia enterocolitica* in sows. Foodborne Pathog Dis 1, 45-52.

Kotetishvili, M., Kreger, A., Wauters, G., Morris, J.G., Jr., Sulakvelidze, A., Stine, O.C., 2005. Multilocus sequence typing for studying genetic relationships among *Yersinia* species. J Clin Microbiol 43, 2674-2684.

Kwaga, J., Iversen, J.O., 1993. Plasmids and outer membrane proteins of *Yersinia enterocolitica* and related species of swine origin. Vet Microbiol 36, 205-214.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.

Laird, W.J., Cavanaugh, D.C., 1980. Correlation of autoagglutination and virulence of yersiniae. J Clin Microbiol 11, 430-432.

Lampel, K.A., Al-Khaldi, S., Cahill, S.M., eds, 2009. Bad Bug Book - Foodborne Pathogenic Microorganisms and Natural Toxins, Second ed. FDA.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50, 1355-1361.

Laukkanen-Ninios, R., Didelot, X., Jolley, K.A., Morelli, G., Sangal, V., Kristo, P., Brehony, C., Imori, P.F., Fukushima, H., Siitonen, A., Tseneva, G., Voskressenskaya, E., Falcao, J.P., Korkeala, H., Maiden, M.C., Mazzoni, C., Carniel, E., Skurnik, M., Achtman, M., 2011. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. Environmental microbiology 13, 3114-3127.

Laukkanen, R., Martínez, P.O., Siekkinen, K.M., Ranta, J., Maijala, R., Korkeala, H., 2009. Contamination of carcasses with human pathogenic *Yersinia enterocolitica* 4/O:3 originates from pigs infected on farms. Foodborne Pathog Dis 6, 681-688.

Lawrenz, M.B., Lenz, J.D., Miller, V.L., 2009. A novel autotransporter adhesin is required for efficient colonization during bubonic plague. Infect Immun 77, 317-326.

Leal, T.C.A., Leal, N.C., De Almeida, A.M.P., 1999. RAPD-PCR typing of *Yersinia enterocolitica* O:3 serotype strains isolated from pigs and humans. Gen Mol Biol 22, 315-319.

Lee, L.A., Gerber, A.R., Lonsway, D.R., Smith, J.D., Carter, G.P., Puhr, N.D., Parrish, C.M., Sikes, R.K., Finton, R.J., Tauxe, R.V., 1990. *Yersinia enterocolitica* O:3 infections in infants and children, associated with the household preparation of chitterlings. The New England journal of medicine 322, 984-987.

Leung, K.Y., Reisner, B.S., Straley, S.C., 1990. YopM inhibits platelet aggregation and is necessary for virulence of Yersinia pestis in mice. Infect Immun 58, 3262-3271.

Lewis, V.G., Ween, M.P., McDevitt, C.A., 2012. The role of ATP-binding cassette transporters in bacterial pathogenicity. Protoplasma 249, 919-942.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Li, Q., Hu, Y., Chen, J., Liu, Z., Han, J., Sun, L., Jiao, X., 2013. Identification of *Salmonella enterica* serovar Pullorum antigenic determinants expressed in vivo. Infect Immun 81, 3119-3127.

Liang, J., Wang, X., Xiao, Y., Cui, Z., Xia, S., Hao, Q., Yang, J., Luo, L., Wang, S., Li, K., Yang, H., Gu, W., Xu, J., Kan, B., Jing, H., 2012. Prevalence of *Yersinia enterocolitica* in pigs slaughtered in Chinese abattoirs. Appl Environ Microbiol 78, 2949-2956.

Ligtenberg, K.G., Miller, N.C., Mitchell, A., Plano, G.V., Schneewind, O., 2013. LcrV mutants that abolish *Yersinia* type III injectisome function. J Bacteriol 195, 777-787.

Linhartová, I., Bumba, L., Mašín, J., Basler, M., Osička, R., Kamanová, J., Procházková, K., Adkins, I., Hejnová-Holubová, J., Sadílková, L., Morová, J., Sebo, P., 2010. RTX proteins: a highly diverse family secreted by a common mechanism. FEMS Microbiol Rev 34, 1076-1112.

Liu, J., Chen, X., Tan, C., Guo, Y., Chen, Y., Fu, S., Bei, W., Chen, H., 2009. In vivo induced RTX toxin ApxIVA is essential for the full virulence of *Actinobacillus pleuropneumoniae*. Vet Microbiol 137, 282-289.

Lo, H.R., Lin, J.H., Chen, Y.H., Chen, C.L., Shao, C.P., Lai, Y.C., Hor, L.I., 2011. RTX toxin enhances the survival of *Vibrio vulnificus* during infection by protecting the organism from phagocytosis. J Infect Dis 203, 1866-1874.

Lobato, M.J., Landeras, E., Gonzalez-Hevia, M.A., Mendoza, M.C., 1998. Genetic heterogeneity of clinical strains of *Yersinia enterocolitica* traced by ribotyping and relationships between ribotypes, serotypes, and biotypes. J Clin Microbiol 36, 3297-3302.

Manafi, M., Holzhammer, E., 1994. Comparison of the Vitek, API 20E and Gene-Track systems for identification of *Yersinia enterocolitica*. Lett. Appl. Microbiol. 18, 90-92.

Mann, R.A., Smits, T.H., Buhlmann, A., Blom, J., Goesmann, A., Frey, J.E., Plummer, K.M., Beer, S.V., Luck, J., Duffy, B., Rodoni, B., 2013. Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. PLoS One 8, e55644.

Mantle, M., Pearson, J., Allen, A., 1980. Pig gastric and small-intestinal mucus glycoproteins: proposed role in polymeric structure for protein joined by disulphide bridges. Biochem Soc Trans 8, 715-716.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Bryant, S.H., 2013. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res 41, D348-352.

Maruyama, T., 1987. *Yersinia enterocolitica* infection in humans and isolation of the microorganism from pigs in Japan. Contrib Microbiol Immunol 9, 48-55.

Matsumoto, H., Young, G.M., 2006. Proteomic and functional analysis of the suite of Ysp proteins exported by the Ysa type III secretion system of *Yersinia enterocolitica* Biovar 1B. Mol Microbiol 59, 689-706.

Matsumoto, H., Young, G.M., 2009. Translocated effectors of *Yersinia*. Curr Opin Microbiol 12, 94-100.

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., Wilson, R.K., 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature 413, 852-856.

McCoy, M.W., Marré, M.L., Lesser, C.F., Mecsas, J., 2010. The C-Terminal Tail of *Yersinia pseudotuberculosis* YopM Is Critical for Interacting with RSK1 and for Virulence. Infect. Immun. 78, 2584–2598.

McDonald, C., Vacratsis, P.O., Bliska, J.B., Dixon, J.E., 2003. The *Yersinia* Virulence Factor YopM Forms a Novel Protein Complex with Two Cellular Kinases. J. Biol. Chem. 278, 18514–18523.

McPhee, J.B., Mena, P., Bliska, J.B., 2010. Delineation of Regions of the *Yersinia* YopM Protein Required for Interaction with the RSK1 and PRK2 Host Kinases and Their Requirement for Interleukin-10 Production and Virulence. Infect Immun 78, 3529–3539.

Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pan-genome. Current opinion in genetics & development 15, 589-594.

Mendoza, M.C., Alzugaray, R., Landeras, E., Gonzalez-Hevia, M.A., 1996. Discriminatory power and application of ribotyping of *Yersinia enterocolitica* O:3 in an epidemiological study. Eur J Clin Microbiol Infect Dis 15, 220-226.

Merhej, V., Adékambi, T., Pagnier, I., Raoult, D., Drancourt, M., 2011. *Yersinia massiliensis* sp. nov., isolated from fresh water. Int J Syst Evol Microbiol 58, 779-784.

Messelhausser, U., Kampf, P., Colditz, J., Bauer, H., Schreiner, H., Holler, C., Busch, U., 2011. Qualitative and quantitative detection of human pathogenic *Yersinia enterocolitica* in different food matrices at retail level in Bavaria. Foodborne Pathog Dis 8, 39-44.

Miller, N.C., Quenee, L.E., Elli, D., Ciletti, N.A., Schneewind, O., 2012. Polymorphisms in the *lcrV* gene of *Yersinia enterocolitica* and their effect on plague protective immunity. Infect Immun 80, 1572-1582.

Miller, V.L., Beer, K.B., Heusipp, G., Young, B.M., Wachtel, M.R., 2001. Identification of regions of Ail required for the invasion and serum resistance phenotypes. Mol Microbiol 41, 1053-1062.

Miller, V.L., Bliska, J.B., Falkow, S., 1990. Nucleotide sequence of the *Yersinia enterocolitica ail* gene and characterization of the Ail protein product. J Bacteriol 172, 1062-1069.

Miller, V.L., Finlay, B.B., Falkow, S., 1988. Factors essential for the penetration of mammalian cells by *Yersinia*. Curr Top Microbiol Immunol 138, 15-39.

Minnich, S.A., Rohde, H.N., 2007. A rationale for repression and/or loss of motility by pathogenic *Yersinia* in the mammalian host. Adv Exp Med Biol 603, 298-310.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35, W182-185.

Mota, L.J., Journet, L., Sorg, I., Agrain, C., Cornelis, G.R., 2005. Bacterial Injectisomes: Needle Length Does Matter. Science 307, 1278.

Motin, V.L., Nakajima, R., Smirnov, G.B., Brubaker, R.R., 1994. Passive immunity to yersiniae mediated by anti-recombinant V antigen and protein A-V antigen fusion peptide. Infect Immun 62, 4192-4201.

Murray, P.J., Wynn, T.A., 2011. Protective and pathogenic functions of macrophage subsets. Nature reviews. Immunology 11, 723-737.

Murros-Kontiainen, A., Fredriksson-Ahomaa, M., Korkeala, H., Johansson, P., Rahkila, R., Björkroth, J., 2011a. *Yersinia nurmii* sp. nov. Int J Syst Evol Microbiol 61, 2368-2372.

Murros-Kontiainen, A., Johansson, P., Niskanen, T., Fredriksson-Ahomaa, M., Korkeala, H., Björkroth, J., 2011b. *Yersinia pekkanenii* sp. nov. Int J Syst Evol Microbiol 61, 2363-2367.

Mwangi, M.M., Wu, S.W., Zhou, Y., Sieradzki, K., de Lencastre, H., Richardson, P., Bruce, D., Rubin, E., Myers, E., Siggia, E.D., Tomasz, A., 2007. Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. Proc Natl Acad Sci U S A 104, 9451-9456.

Najdenski, H., Iteman, I., Carniel, E., 1994. Efficient subtyping of pathogenic *Yersinia enterocolitica* strains by pulsed-field gel electrophoresis. J Clin Microbiol 32, 2913-2920.

Najdenski, H., Iteman, I., Carniel, E., 1995. The genome of *Yersinia enterocolitica* is the most stable of the three pathogenic species. Contrib Microbiol Immunol 13, 281-284.

Nesbakken, T., Eckner, K., Høidal, H.K., Røtterud, O.J., 2003. Occurrence of *Y. enterocolitica* in slaughter pigs and consequences for meat inspection, slaughtering and dressing procedures. Adv Exp Med Biol 529, 303-308.

Nesbakken, T., Kapperud, G., Sorum, H., Dommarsnes, K., 1987. Structural variability of 40-50 Mdal virulence plasmids from *Yersinia enterocolitica*. Geographical and ecological distribution of plasmid variants. Acta Pathol Microbiol Immunol Scand B 95, 167-173.

Neubauer, H., Aleksic, S., Hensel, A., Finke, E., Meyer, H., 2000a. *Yersinia enterocolitica* 16S rRNA gene types belong to the same genospecies but form three homology groups. International Journal of Medical Microbiology 290, 61-64.

Neubauer, H., Hensel, A., Aleksic, S., Meyer, H., 2000b. Identification of *Yersinia enterocolitica* within the genus *Yersinia*. Syst Appl Microbiol 23, 58-62.

Neubauer, H., Sauer, T., Becker, H., Aleksic, S., Meyer, H., 1998. Comparison of systems for identification and differentiation of species within the genus *Yersinia*. J Clin Microbiol 36, 3366-3368.

Nishi, J., Sheikh, J., Mizuguchi, K., Luisi, B., Burland, V., Boutin, A., Rose, D.J., Blattner, F.R., Nataro, J.P., 2003. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. J Biol Chem 278, 45680-45689.

Nummelin, H., Merckel, M.C., Leo, J.C., Lankinen, H., Skurnik, M., Goldman, A., 2004. The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. EMBO J 23, 701-711.

O'Rawe, J., Jiang, T., Sun, G., Wu, Y., Wang, W., Hu, J., Bodily, P., Tian, L., Hakonarson, H., Johnson, W.E., Wei, Z., Wang, K., Lyon, G.J., 2013. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome medicine 5, 28.

Oberhettinger, P., Schütz, M., Raddatz, G., Keller, H., Autenrieth, I.B., Linke, D., 2011. The sequence of the pYV virulence plasmid from *Yersinia enterocolitica* strain WA-314 biogroup 1B serotype O:8. Plasmid 65, 20-24.

Odinot, P.T., Meis, J.F., Van den Hurk, P.J., Hoogkamp-Korstanje, J.A., Melchers, W.J., 1995. PCR-based characterization of *Yersinia enterocolitica*: comparison with biotyping and serotyping. Epidemiol Infect 115, 269-277.

Olsovsky, Z., Olsakova, V., Chobot, S., Sviridov, V., 1975. Mass occurrence of *Yersinia enterocolitica* in two establishments of collective care of children. Journal of hygiene, epidemiology, microbiology, and immunology 19, 22-29.

Ostroff, S.M., Kapperud, G., Hutwagner, L.C., Nesbakken, T., Bean, N.H., Lassen, J., Tauxe, R.V., 1994. Sources of sporadic *Yersinia enterocolitica* infections in Norway: a prospective case-control study. Epidemiol Infect 112, 133-141.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33, 5691-5702.

Ovodov, Y.S., Gorshkova, R.P., Tomshich, S.V., Komandrova, N.A., Zubkov, V.A., Kalmykova, E.N., Isakov, V.V., 1992. Chemical and immunochemical studies on lipopolysaccharides of some *Yersinia* species: A review of some recent investigations. J. Carbohydr. Chem. 11, 21-35.

Page, L., Griffiths, L., Cole, J.A., 1990. Different physiological roles of two independent pathways for nitrite reduction to ammonia by enteric bacteria. Archives of microbiology 154, 349-354.

Paixao, R., Moreno, L.Z., Sena de Gobbi, D.D., Raimundo, D.C., Ferreira, T.S., Spindola, M.G., Hofer, E., Falavina Dos Reis, C.M., Matte, M.H., Micke Moreno, A., 2013a. Genotypic Characterization of *Yersinia enterocolitica* Biotype 4/O:3 Isolates from Pigs and Slaughterhouses Using SE-AFLP, ERIC-PCR, and PFGE. J Pathog 2013, 521510.

Paixao, R., Moreno, L.Z., Sena de Gobbi, D.D., Raimundo, D.C., Hofer, E., Matte, M.H., Ferreira, T.S., de Moura Gomes, V.T., Costa, B.L., Moreno, A.M., 2013b. Characterization of *Yersinia enterocolitica* biotype 1A strains isolated from swine slaughterhouses and markets. TheScientificWorldJournal 2013, 769097.

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., Cerdeno-Tarraga, 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*. Nature genetics 35, 32-40.

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., Van Ert, M.N., 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.

Pearson, T., Okinaka, R.T., Foster, J.T., Keim, P., 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 9, 1010-1019.

Pederson, K.J., Pierson, D.E., 1995. Ail expression in *Yersinia enterocolitica* is affected by oxygen tension. Infect Immun 63, 4199-4201.

Pelludat, C., Rakin, A., Jacobi, C.A., Schubert, S., Heesemann, J., 1998. The yersiniabactin biosynthetic gene cluster of *Yersinia enterocolitica*: organization and siderophore-dependent regulation. J Bacteriol 180, 538-546.

Pepe, J.C., Miller, V.L., 1993. *Yersinia enterocolitica* invasin: a primary role in the initiation of infection. Proc Natl Acad Sci U S A 90, 6473-6477.

Pepe, J.C., Wachtel, M.R., Wagar, E., Miller, V.L., 1995. Pathogenesis of defined invasion mutants of *Yersinia enterocolitica* in a BALB/c mouse model of infection. Infect Immun 63, 4837-4848.

Perry, R.D., Balbo, P.B., Jones, H.A., Fetherston, J.D., DeMoll, E., 1999. Yersiniabactin from *Yersinia pestis*: biochemical characterization of the siderophore and its role in iron transport and regulation. Microbiology 145, 1181-1190.

Perry, R.D., Fetherston, J.D., 2011. Yersiniabactin iron uptake: mechanisms and role in *Yersinia pestis* pathogenesis. Microbes Infect 13, 808-817.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8, 785-786.

Pettersson, J., Holmström, A., Hill, J., Leary, S., Frithz-Lindsten, E., von Euler-Matell, A., Carlsson, E., Titball, R., Forsberg, A., Wolf-Watz, H., 1999. The V-antigen of *Yersinia* is surface exposed before target cell contact and involved in virulence protein translocation. Mol Microbiol 32, 961-976.

Poock, S.R., Leach, E.R., Moir, J.W., Cole, J.A., Richardson, D.J., 2002. Respiratory detoxification of nitric oxide by the cytochrome c nitrite reductase of *Escherichia coli*. J Biol Chem 277, 23664-23669.

Portnoy, D.A., Moseley, S.L., S, F., 1981. Characterization of Plasmids and Plasmid-Associated Determinants of *Yersinia enterocolitica* Pathogenesis. Infect Immun 31, 775-782.

Portnoy, D.A., Wolf-Watz, H., Bolin, I., Beeder, A.B., Falkow, S., 1984. Characterization of common virulence plasmids in *Yersinia* species and their role in the expression of outer membrane proteins. Infect Immun 43, 108-114.

Prentice, M.B., Cuccui, J., Thomson, N., Parkhill, J., Deery, E., Warren, M.J., 2003. Cobalamin synthesis in *Yersinia enterocolitica* 8081. Functional aspects of a putative metabolic island. Adv Exp Med Biol 529, 43-46.

Proft, T., Baker, E.N., 2009. Pili in Gram-negative and Gram-positive bacteria – structure, assembly and their role in disease. Cell. Mol. Life Sci. 66, 613-635.

Prpic, J.K., Robins-Browne, R.M., Davey, R.B., 1985. In vitro assessment of virulence in *Yersinia enterocolitica* and related species. J Clin Microbiol 22, 105-110.

Radziejewska-Lebrecht, J., Shashkov, A.S., Stroobant, V., Wartenberg, K., Warth, C., Mayer, H., 1994. The inner core region of *Yersinia enterocolitica* Ye75R (0:3) lipopolysaccharide. European journal of biochemistry / FEBS 221, 343-351.

Rahman, A., Bonny, T.S., Stonsaovapak, S., Ananchaipattana, C., 2011. *Yersinia enterocolitica*: Epidemiological Studies and Outbreaks. J Pathog 2011, 239391.

Rakin, A., Garzetti, D., 2013. Different siderophores contribute to the high-pathogenicity phenotype in *Yersinia*. Problems of Particularly Dangerous Infections 3.

Rakin, A., Garzetti, D., Bouabe, H., Sprague, L.D., 2015. *Yersinia enterocolitica*, in: Academic Press. Ed. Sussman (Ed.), Molecular Medical Microbiology – 2nd Edition. In press.

Rakin, A., Heesemann, J., 1995. Yersiniabactin/pesticin receptor: a component of an iron uptake system of highly pathogenic *Yersinia*. Contrib Microbiol Immunol 13, 244-247.

Rakin, A., Noelting, C., Schubert, S., Heesemann, J., 1999. Common and specific characteristics of the high-pathogenicity island of *Yersinia enterocolitica*. Infect Immun 67, 5265-5274.

Ramamurthy, T., Yoshino, K., Huang, X., Balakrish Nair, G., Carniel, E., Maruyama, T., Fukushima, H., Takeda, T., 1997. The novel heat-stable enterotoxin subtype gene (ystB) of *Yersinia enterocolitica*: nucleotide sequence and distribution of the yst genes. Microb Pathog 23, 189-200.

Rasko, D.A., Rosovitz, M.J., Myers, G.S., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., Henderson, I.R., Sperandio, V., Ravel, J., 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol 190, 6881-6893.

Rasmussen, H.N., Olsen, J.E., Rasmussen, O.F., 1994. RAPD analysis of *Yersinia enterocolitica*. Lett Appl Microbiol 19, 359-362.

Ratnam, S., Mercer, E., Picco, B., Parsons, S., Butler, R., 1982. A nosocomial outbreak of diarrheal disease due to *Yersinia enterocolitica* serotype 0:5, biotype 1. J Infect Dis 145, 242-247.

Reuter, S., Connor, T.R., Barquist, L., Walker, D., Feltwell, T., Harris, S.R., Fookes, M., Hall, M.E., Petty, N.K., Fuchs, T.M., Corander, J., Dufour, M., Ringwood, T., Savin, C., Bouchier, C., Martin, L., Miettinen, M., Shubin, M., Riehm, J.M., Laukkanen-Ninios, R., Sihvonen, L.M., Siitonen, A., Skurnik, M., Falcao, J.P., Fukushima, H., Scholz, H.C., Prentice, M.B., Wren, B.W., Parkhill, J., Carniel, E., Achtman, M., McNally, A., Thomson, N.R., 2014. Parallel independent evolution of pathogenicity within the genus *Yersinia*. Proc Natl Acad Sci U S A 111, 6768-6773.

Reuter, S., Thomson, N.R., McNally, A., 2012. Evolutionary dynamics of the *Yersinia enterocolitica* complex. Adv Exp Med Biol 954, 15-22.

Revell, P.A., Miller, V.L., 2001. *Yersinia* virulence: more than a plasmid. FEMS Microbiol Lett 205, 159-164.

Richardson, D.J., 2000. Bacterial respiration: a flexible process for a changing environment. Microbiology 146 (Pt 3), 551-571.

Robins-Browne, R.M., Still, C.S., Miliotis, M.D., Koornhof, H.J., 1979. Mechanism of action of *Yersinia enterocolitica* enterotoxin. Infect Immun 25, 680-684.

Robins-Browne, R.M., Tzipori, S., Gonis, G., Hayes, J., Withers, M., Prpic, J.K., 1985. The pathogenesis of *Yersinia enterocolitica* infection in gnotobiotic piglets. J Med Microbiol 19, 297-308.

Roggenkamp, A., Ackermann, N., Jacobi, C.A., Truelzsch, K., Hoffmann, H., Heesemann, J., 2003. Molecular analysis of transport and oligomerization of the *Yersinia enterocolitica* adhesin YadA. J Bacteriol 185, 3735-3744.

Roggenkamp, A., Geiger, A.M., Leitritz, L., Kessler, A., Heesemann, J., 1997. Passive immunity to infection with *Yersinia* spp. mediated by anti-recombinant V antigen is dependent on polymorphism of V antigen. Infect Immun 65, 446-451.

Roggenkamp, A., Neuberger, H.R., Flugel, A., Schmoll, T., Heesemann, J., 1995. Substitution of two histidine residues in YadA protein of *Yersinia enterocolitica* abrogates collagen binding, cell adherence and mouse virulence. Mol Microbiol 16, 1207-1219.

Roggenkamp, A., Ruckdeschel, K., Leitritz, L., Schmitt, R., Heesemann, J., 1996. Deletion of amino acids 29 to 81 in adhesion protein YadA of *Yersinia enterocolitica* serotype O:8 results in selective abrogation of adherence to neutrophils. Infect Immun 64, 2506-2514.

Rollins, S.M., Peppercorn, A., Hang, L., Hillman, J.D., Calderwood, S.B., Handfield, M., Ryan, E.T., 2005. In vivo induced antigen technology (IVIAT). Cell Microbiol 7, 1-9.

Rollins, S.M., Peppercorn, A., Young, J.S., Drysdale, M., Baresch, A., Bikowski, M.V., Ashford, D.A., Quinn, C.P., Handfield, M., Hillman, J.D., Lyons, C.R., Koehler, T.M., Calderwood, S.B., Ryan, E.T., 2008. Application of in vivo induced antigen technology (IVIAT) to *Bacillus anthracis*. PLoS One 3, e1824.

Rosqvist, R., Magnusson, K.E., Wolf-Watz, H., 1994. Target cell contact triggers expression and polarized transfer of *Yersinia* YopE cytotoxin into mammalian cells. EMBO J 13, 946-972.

Rossen, L., Norskov, P., Holmstrom, K., Rasmussen, O.F., 1992. Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. Int J Food Microbiol 17, 37-45.

Rosso, M.L., Chauvaux, S., Dessein, R., Laurans, C., Frangeul, L., Lacroix, C., Schiavo, A., Dillies, M.A., Foulon, J., Coppee, J.Y., Medigue, C., Carniel, E., Simonet, M., Marceau, M., 2008. Growth of *Yersinia pseudotuberculosis* in human plasma: impacts on virulence and metabolic gene expression. BMC Microbiol 8, 211.

Roumagnac, P., Weill, F.X., Dolecek, C., Baker, S., Brisse, S., Chinh, N.T., Le, T.A., Acosta, C.J., Farrar, J., Dougan, G., Achtman, M., 2006. Evolutionary history of *Salmonella typhi*. Science 314, 1301-1304.

Sabina, J., Rahman, A., Chandra Ray, R., Montet, D., 2011. *Yersinia enterocolitica*: Mode of Transmission, Molecular Insights of Virulence, and Pathogenesis of Infection. J Pathog.

Sachdeva, P., Virdi, J.S., 2004. Repetitive elements sequence (REP/ERIC)-PCR based genotyping of clinical and environmental strains of *Yersinia enterocolitica* biotype 1A reveal existence of limited number of clonal groups. FEMS Microbiol Lett 240, 193-201.

Saha, S., Chowdhury, P., Mazumdar, A., Pal, A., Das, P., Chakrabarti, M.K., 2009. Role of *Yersinia enterocolitica* heat-stable enterotoxin (Y-STa) on differential regulation of nuclear and cytosolic calcium signaling in rat intestinal epithelial cells. Cell biology and toxicology 25, 297-308.

Saken, E., Roggenkamp, A., Aleksic, S., Heesemann, J., 1994. Characterisation of pathogenic *Yersinia enterocolitica* serogroups by pulsed-field gel electrophoresis of genomic NotI restriction fragments. J Med Microbiol 41, 329-338.

Samuel, G., Reeves, P., 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydrate research 338, 2503-2519.

Savin, C., Carniel, E., 2008. Les diarrhées d'origine bactérienne: le cas de *Yersinia enterocolitica*. Revue Francophone des Laboratoires 38, 49-58.

Savin, C., Frangeul, L., Ma, L., Bouchier, C., Moszer, I., Carniel, E., 2013. Draft Genome Sequence of a Clinical Strain of *Yersinia enterocolitica* (IP10393) of Bioserotype 4/O:3 from France. Genome announcements 1.

Schaake, J., Kronshage, M., Uliczka, F., Rohde, M., Knuuti, T., Strauch, E., Fruth, A., Wos-Oxley, M., Dersch, P., 2013. Human and animal isolates of *Yersinia enterocolitica* show significant serotype-specific colonization and host-specific immune defense properties. Infect Immun 81, 4013-4025.

Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33, W686-689.

Schiemann, D.A., 1979. Synthesis of a selective agar medium for *Yersinia enterocolitica*. Can J Microbiol 25, 1298-1304.

Schiemann, D.A., 1982. Development of a two-step enrichment procedure for recovery of *Yersinia enterocolitica* from food. Appl Environ Microbiol 43, 14-27.

Schubert, S., Fischer, D., Heesemann, J., 1999. Ferric enterochelin transport in *Yersinia enterocolitica*: molecular and evolutionary aspects. J Bacteriol 181, 6387-6395.

Schulte, R., Kerneis, S., Klinke, S., Bartels, H., Preger, S., Kraehenbuhl, J.P., Pringault, E., Autenrieth, I.B., 2000. Translocation of *Yersinia entrocolitica* across reconstituted intestinal

epithelial monolayers is triggered by *Yersinia* invasin binding to beta1 integrins apically expressed on M-like cells. Cell Microbiol 2, 173-185.

Schumacher, M.A., Piro, K.M., Xu, W., Hansen, S., Lewis, K., Brennan, R.G., 2009. Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. Science 323, 396-401.

Seoane, A., Garcia Lobo, J.M., 2000. Identification of a streptogramin A acetyltransferase gene in the chromosome of *Yersinia enterocolitica*. Antimicrob Agents Chemother 44, 905-909.

Shutinoski, B., Schmidt, M.A., Heusipp, G., 2010. Transcriptional regulation of the Yts1 type II secretion system of *Yersinia enterocolitica* and identification of secretion substrates. Mol Microbiol 75, 676-691.

Sihvonen, L.M., Toivonen, S., Haukka, K., Kuusi, M., Skurnik, M., Siitonen, A., 2011. Multilocus variable-number tandem-repeat analysis, pulsed-field gel electrophoresis, and antimicrobial susceptibility patterns in discrimination of sporadic and outbreak-related strains of *Yersinia enterocolitica*. BMC Microbiol 11, 42.

Simjee, S., Zhang, Y., McDermott, P.F., Donabedian, S.M., Zervos, M.J., Meng, J., 2006. Heterogeneity of vat(E)-carrying plasmids in *Enterococcus faecium* recovered from human and animal sources. International journal of antimicrobial agents 28, 200-205.

Simonet, M.L., 1999. Enterobacteria in reactive arthritis: *Yersinia*, *Shigella*, and *Salmonella*. Rev Rhum Engl Ed 66, 14S-18S.

Skrzypek, E., Cowan, C., Straley, S.C., 1998. Targeting of the *Yersinia pestis* YopM protein into HeLa cells and intracellular trafficking to the nucleus. Mol Microbiol 30, 1051-1065.

Skurnik, M., 2004. Lipopolysaccharides of *Yersinia*, in: Carniel, E., Hinnebusch, B.J. (Eds.), *Yersinia* molecular and cellular biology. 11. Horizon Bioscience, UK, pp. 215-241.

Skurnik, M., 2006. Lipopolysaccharide, in: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., Stackebrandt, E. (Eds.), The Prokaryotes, Third ed. Vol. 6, 3.3.13: *Y. enterocolitica* and *Y. pseudotuberculosis*. Springer Science+Business Media, USA, pp. 333-346.

Skurnik, M., Bengoechea, J.A., 2003. The biosynthesis and biological role of lipopolysaccharide O-antigens of pathogenic *Yersiniae*. Carbohydrate research 338, 2521-2529.

Skurnik, M., Venho, R., Bengoechea, J.A., Moriyón, I., 1999. The lipopolysaccharide outer core of *Yersinia enterocolitica* serotype O:3 is required for virulence and plays a role in outer membrane integrity. Mol Microbiol 31, 1443-1462.

Snellings, N.J., Popek, M., Lindler, L.E., 2001. Complete DNA sequence of *Yersinia enterocolitica* serotype 0:8 low-calcium-response plasmid reveals a new virulence plasmid-associated replicon. Infect Immun 69, 4627-4638.

Sprague, L.D., Neubauer, H., 2005. *Yersinia aleksiciae* sp. nov. Int J Syst Evol Microbiol 55, 831-835.

Sprague, L.D., Scholz, H.C., Amann, S., Busse, H.J., Neubauer, H., 2008. *Yersinia similis* sp. nov. Int J Syst Evol Microbiol 58, 952-958.

Stephan, R., Joutsen, S., Hofer, E., Säde, E., Björkroth, J., Ziegler, D., Fredriksson-Ahomaa, M., 2013. Characteristics of *Yersinia enterocolitica* biotype 1A strains isolated from patients and asymptomatic carriers. Eur J Clin Microbiol Infect Dis 32, 869-875.

Subashchandrabose, S., LeVeque, R.M., Wagner, T.K., Kirkwood, R.N., Kiupel, M., Mulks, M.H., 2009. Branched-chain amino acids are required for the survival and virulence of *Actinobacillus pleuropneumoniae* in swine. Infect Immun 77, 4925-4933.

Sukupolvi, S., O'Connor, C.D., 1990. TraT lipoprotein, a plasmid-specified mediator of interactions between gram-negative bacteria and their environment. Microbiological reviews 54, 331-341.

Sulakvelidze, A., 2000. *Yersiniae* other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: the ignored species. Microbes Infect. 2, 497-513.

Tadesse, D.A., Bahnson, P.B., Funk, J.A., Morrow, W.E., Abley, M.J., Ponte, V.A., Thakur, S., Wittum, T., DeGraves, F.J., Rajala-Schultz, P.J., Gebreyes, W.A., 2013. *Yersinia enterocolitica* of porcine origin: carriage of virulence genes and genotypic diversity. Foodborne Pathog Dis 10, 80-86.

Takami, H., Taniguchi, T., Moriya, Y., Kuwahara, T., Kanehisa, M., Goto, S., 2012. Evaluation method for the potential functionome harbored in the genome and metagenome. BMC Genomics 13, 699.

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., 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.

Tettelin, H., Riley, D., Cattuto, C., Medini, D., 2008. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11, 472-477.

Thibodeau, V., Frost, E.H., Chénier, S., Quessy, S., 1999. Presence of *Yersinia enterocolitica* in tissues of orally-inoculated pigs and the tonsils and feces of pigs at slaughter. Can J Vet Res 63, 96-100.

Thisted Lambertz, S., Danielsson-Tham, M.L., 2005. Identification and characterization of pathogenic *Yersinia enterocolitica* isolates by PCR and pulsed-field gel electrophoresis. Appl Environ Microbiol 71, 3674-3681.

Thomson, N.R., Howard, S., Wren, B.W., Holden, M.T.G., Crossman, L., Challis, G.L., Churcher, C., Mungall, K., Brooks, K., Chillingworth, T., Feltwell, T., Abdellah, Z., Hauser, H., Jagels, K., Maddison, M., Moule, S., Sanders, M., Whitehead, S., Quail, M.A., Dougan, G., Parkhill, J., Prentice, M.B., 2006. The Complete Genome Sequence and Comparative Genome Analysis of the High Pathogenicity *Yersinia enterocolitica* Strain 8081. PLoS Genet 2, e206.

Thorslund, S.E., Ermert, D., Fahlgren, A., Erttmann, S.F., Nilsson, K., Hosseinzadeh, A., Urban, C.F., Fällman, M., 2013. Role of YopK in *Yersinia pseudotuberculosis* resistance against polymorphonuclear leukocyte defense. Infect Immun 81, 11-22.

Uliczka, F., Pisano, F., Schaake, J., Stolz, T., Rohde, M., Fruth, A., Strauch, E., Skurnik, M., Batzilla, J., Rakin, A., Heesemann, J., Dersch, P., 2011. Unique Cell Adhesion and Invasion Properties of *Yersinia enterocolitica* O:3, the Most Frequent Cause of Human Yersiniosis. Plos Pathogens 7.

van Belkum, A., 1999. The role of short sequence repeats in epidemiologic typing. Curr Opin Microbiol 2, 306-311.

van der Woude, M.W., Baumler, A.J., 2004. Phase and antigenic variation in bacteria. Clin Microbiol Rev 17, 581-611, table of contents.

Venecia, K., Young, G.M., 2005. Environmental regulation and virulence attributes of the Ysa type III secretion system of *Yersinia enterocolitica* biovar 1B. Infect Immun 73, 5961-5977.

Vieux, E.F., Barrick, D., 2011. Deletion of internal structured repeats increases the stability of a leucine-rich repeat protein, YopM. Biophys. Chem. 159, 152-161.

Virdi, J.S., Sachdeva, P., 2005. Molecular heterogeneity in *Yersinia enterocolitica* and '*Y. enterocolitica*-like' species--Implications for epidemiology, typing and taxonomy. FEMS Immunol Med Microbiol 45, 1-10.

Virtanen, S., Laukkanen-Ninios, R., Ortiz Martinez, P., Siitonen, A., Fredriksson-Ahomaa, M., Korkeala, H., 2013. Multiple-locus variable-number tandem-repeat analysis in genotyping *Yersinia enterocolitica* strains from human and porcine origins. J Clin Microbiol 51, 2154-2159.

von Tils, D., Blädel, I., Schmidt, M.A., Heusipp, G., 2012. Type II secretion in *Yersinia*-a secretion system for pathogenicity and environmental fitness. Front Cell Infect Microbiol 2.

Wagner, S., Sorg, I., Degiacomi, M., Journet, L., Dal Peraro, M., Cornelis, G.R., 2009. The helical content of the YscP molecular ruler determines the length of the *Yersinia* injectisome. Molec Microbiol 71, 692-701.

Wang, X., Li, Y., Jing, H., Ren, Y., Zhou, Z., Wang, S., Kan, B., Xu, J., Wang, L., 2011. Complete Genome Sequence of a *Yersinia enterocolitica* "Old World" (3/O:9) Strain and Comparison with the "New World" (1B/O:8) Strain. J. Clin. Microbiol. 49, 1251-1259.

Wauters, G., 1973. Improved methods for the isolation and the recognition of *Yersinia enterocolitica*. Contrib. Microbiol. Immunol 2, 68-70.

Wauters, G., Aleksic, S., Charlier, J., Schulze, G., 1991. Somatic and flagellar antigens of *Yersinia enterocolitica* and related species. Contrib Microbiol Immunol 12, 239-243.

Wauters, G., Goossens, V., Janssens, M., Vandepitte, J., 1988. New enrichment method for isolation of pathogenic *Yersinia enterocolitica* serogroup O:3 from pork. Appl Environ Microbiol 54, 851-854.

Wauters, G., Kandolo, K., Janssens, K., 1987. Revised biogrouping scheme of *Yersinia enterocolitica*. Contrib Microbiol Immunol 9, 14-21.

West, A.H., Stock, A.M., 2001. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci 26, 369-376.

Weynants, V., Jadot, V., Denoel, P.A., Tibor, A., Letesson, J.J., 1996. Detection of *Yersinia enterocolitica* serogroup O:3 by a PCR method. J Clin Microbiol 34, 1224-1227.

Winblad, S., 1973. Studies on the O serotypes of *Yersinia enterocolitica*. Contrib Microbiol Immunol 2, 27-37.

Wojciech, L., Staroniewicz, Z., Jakubczak, A., Ugorski, M., 2004. Typing of *Yersinia Enterocolitica* Isolates by ITS profiling, REP- and ERIC-PCR. Journal of veterinary medicine. B, Infectious diseases and veterinary public health 51, 238-244.

Yang, Y., Merriam, J.J., Mueller, J.P., Isberg, R.R., 1996. The *psa* locus is responsible for thermoinducible binding of *Yersinia pseudotuberculosis* to cultured cells. Infect Immun 64, 2483-2489.

Yoshino, K., Takao, T., Huang, X., Murata, H., Nakao, H., Takeda, T., Shimonishi, Y., 1995. Characterization of a highly toxic, large molecular size heat-stable enterotoxin produced by a clinical isolate of *Yersinia enterocolitica*. FEBS letters 362, 319-322.

Young, G.M., 2004. Flagella: Organelles for Motility and Protein Secretion, in: Carniel, E., Hinnebusch, B.J. (Eds.), *Yersinia* molecular and cellular biology. 12. Horizon Bioscience, UK, pp. 243-256.

Young, G.M., Badger, J.L., Miller, V.L., 2000. Motility is required to initiate host cell invasion by *Yersinia enterocolitica*. Infect Immun 68, 4323-4326.

Young, G.M., Miller, V.L., 1997. Identification of novel chromosomal loci affecting *Yersinia enterocolitica* pathogenesis. Mol Microbiol 25, 319-328.

Young, G.M., Smith, M.J., Minnich, S.A., Miller, V.L., 1999. The *Yersinia enterocolitica* motility master regulatory operon, *flhDC*, is required for flagellin production, swimming motility, and swarming motility. J Bacteriol 181, 2823-2833.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., Brinkman, F.S., 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608-1615.

Zhang, A., Yang, M., Hu, P., Wu, J., Chen, B., Hua, Y., Yu, J., Chen, H., Xiao, J., Jin, M., 2011. Comparative genomic analysis of *Streptococcus suis* reveals significant genomic diversity among different serotypes. BMC Genomics 12, 523.

Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P., Skurnik, M., 1997. Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O:8. Mol Microbiol 23, 63-76.

Zhang, X.L., Tsui, I.S.M., Yip, C.M.C., Fung, A.W.Y., Wong, D.K.H., Dai, X., Yang, Y., Hackett, J., Morris, C., 2000. *Salmonella enterica* serovar typhi uses type IVB pili to enter human intestinal epithelial cells. Infect Immun 68, 3067–3073.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39, W347-352.

8. Acknowledgments

The PhD period has been really pleasant for me and many people have made this possible.

First, I would like to thank Dr. A. Rakin and Prof. J. Heesemann for their constant support and encouraging trust in me. Their advices and ideas have inspired my work and helped me become a finer researcher.

Many thanks to all the collaborators who spent time in collecting samples and performing analyses or experiments for my project, in particular Dr. H. Bouabe, Dr. K. Trülzsch, Dr. J. Blom, Dr. E. Tietze and Dr. W. Rastawicki.

I would like to acknowledge the “Yersinia Club” people for their ideas and guidance, especially Dr. O. Rossier and Prof. B. Stecher.

I am sincerely grateful to all the people who contributed in maintaining a nice and helpful atmosphere in the lab, Dr. J. Niefnecker, E. Storz, Dr. E. Rogulin and, in particular, Dr. E.M. Schriefer for her patience in answering my questions and her assistance with the German language. A special thanks to Lukas Schneider, not only for being my funny American group-mate, who liked moka coffee, talked with a funny Italian accent and easily became a kicker champion, but also for the useful scientific discussions and help in the lab.

Thanks to Luba and Sandrine for being friendly colleagues, listening to my outbursts and dealing with me during difficult times. I cannot forget the coworkers who somehow helped me during these years and made me feel welcomed, Dr. W. Fischer, Dr. C. Harrison, Dr. L.F. Jimenez Soto, the whole AG Stecher, A. Kessler, the 4th floor and many others.

I would also like to thank the secretaries, especially Frau U. Gotterbauer, and the EDV, in particular Herr K. Szczepaniak, for their patient assistance.

I am truly grateful to all my Italian friends who have always being supporting my decisions, even if they mean living abroad and being absent when they need me, particularly to Fede and Flavia. A genuine and loving thank to Andrea, who has been always encouraging and guiding me in this challenging period of my life.

Finally, I wouldn’t be the person I am today without the support of my parents, who deserve thanks for so many things that this page wouldn’t be enough to write them all.

9. List of Publications

The work described in this thesis has been partly published in the following publications in peer reviewed journals:

- Garzetti, D., Bouabe, H., Heesemann, J., Rakin, A., 2012. Tracing genomic variations in two highly virulent Yersinia enterocolitica strains with unequal ability to compete for host colonization. BMC Genomics 13, 1-15.

- Garzetti, D., Heesemann, J., Rakin, A., 2013. Genome Sequences of Four Yersinia enterocolitica Bioserotype 4/O:3 Isolates from Mammals. Genome Announc 11, 1(4).

- Garzetti, D., Susen, R., Fruth, A., Tietze, E., Heesemann, J., Rakin, A., 2014. A molecular scheme for Yersinia enterocolitica patho-serotyping derived from genome-wide analysis. Int J Med Microbiol 304, 275-283.

Other publications during the time of dissertation:

- Rakin, A., Batzilla, J., Garzetti, D., Heesemann, J., 2012. Gains and Losses in Yersinia enterocolitica subsp. palearctica Genomes. Adv Exp Med Biol 954, 23-9.

- Rakin, A., Garzetti, D., 2013. Different siderophores contribute to the high-pathogenicity phenotype in Yersinia. Problems of Particularly Dangerous Infections, 3.

- Bosák, J., Micenková, L., Vrba, M., Ševčíková, A., Dědičová, D., Garzetti, D., Šmajs, D., 2013. Unique activity spectrum of colicin FY: All 110 characterized Yersinia enterocolitica isolates were colicin FY susceptible. PLoS One, 8(12).

- Rakin, A., Garzetti, D., Bouabe, H., Sprague, L.D., 2015. Chapter 82, Yersinia enterocolitica. Molecular Medical Microbiology, 2nd Edition, (In press).