Complete genome sequence of *Yersinia enterocolitica* subspecies *palearctica* serotype O:3: Identification of novel virulence-associated genes and evolutionary aspects

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von

Julia Batzilla

aus Münster

Würzburg 2011

Eingereicht am: .....

Mitglieder der Promotionskommission:

Vorsitzender: .....

Gutachter : .....

Gutachter: .....

Tag des Promotionskolloquiums: .....

Doktorurkunde ausgehändigt am: .....

Erklärungen

Ich versichere, dass ich diese Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe. Weiterhin versichere ich, dass ich die Dissertation bisher nicht in gleicher oder ähnlicher Form in einem anderen Prüfungsverfahren vorgelegt habe und ich bisher keine anderen akademischen Grade als B.Sc. und M.Sc. in Biomedizin erworben oder zu erwerben versucht habe.

Regensburg,

Julia Batzilla

1. Zusammenfassung

Yersinia enterocolitica subsp. *palearctica* Serobiotyp O:3/4 ist verantwortlich für 80-90 % aller Yersiniosen beim Menschen in Deutschland und Europa. *Y. enterocolitica* Infektionen zeigen vielfältige Krankheitsbilder wie Gastroenteritis, Lymphadenitis und verschiedene Spätkomplikationen wie reaktive Arthritis. Das wichtigste Tierreservoir stellt das Hausschwein dar. Rohes Schweinefleisch in Metzgereien in Deutschland und anderen Regionen in Nord-Ost Europa ist häufig mit Yersinien kontaminiert (Bayern: 25 %).

Da sich Serobiotyp O:3/4-Stämme geografisch und phylogenetisch deutlich von dem bisher sequenzierten Serobiotyp O:8/1B Stamm 8081 unterscheiden, wurde eine komplette Genomsequenzierung des europäischen Serobiotyp O:3/4 DSMZ Referenzstammes Y11 (aus Patientenstuhl isoliert) durchgeführt. Um einen genaueren Einblick in die *Y. enterocolitica* subsp. *palearctica* Gruppe zu erhalten, wurden zusätzlich zwei weitere Serobiotyp O:3/4 Isolate (Stamm Y8265, Patientenisolat, und Stamm Y5307, mit reaktiver Arthritis assoziiertes Patientenisolat), sowie ein eng verwandtes *Y. enterocolitica* subsp. *palearctica* Serobiotyp O:5,27/3 Isolat, Stamm Y527P, und zwei Biotyp 1A Isolate (ein Isolat nosokomialer Herkunft (Serogruppe O:5) und ein Umwelt-Isolat (O:36)) unvollständig sequenziert. Die nicht mausvirulenten Stämme wurden mit dem mausvirulenten *Y. enterocolitica* subsp. *enterocolitica* Serobiotyp O:8/1B Stamm 8081 verglichen, um genetische Besonderheiten von Stamm Y11 und der *Y. enterocolitica* subsp. *palearctica* Gruppe zu identifizieren. Besonderer Fokus lag hierbei auf dem pathogenen Potential von Stamm Y11, um neue potentielle Virulenz Faktoren und Fitnessfaktoren zu identifizieren, darunter vor allem solche, die eine Rolle bei der Wirtsspezifität von Serobiotyp O:3/4 spielen könnten.

Y. enterocolitica subsp. *palearctica* Serobiotyp O:3/4 Stämmen fehlen einige der Charakteristika der mausvirulenten Gruppe *Y. enterocolitica* subsp. *enterocolitica*, beispielsweise die Yersiniabactin kodierende, High-Pathogenicity Island (HPI), das Yts1 Typ 2 Sekretionssystem und das Ysa Typ 3 Sekretionssystem. Die Serobiotyp O:3/4-Stämme haben ein anderes Repertoir von Virulenz Faktoren erworben, darunter Gene bzw. genomische Inseln für das Ysp Typ 3 Sekretionssystem, Rtx-ähnliches putatives Toxin, Insektizid-Toxine und ein funktionelles PTS System für die Aufnahme von N-acetyl-galactosamin, dem *aga*-Operon. Nach dem Transfer des *aga*-Operons in *Y. enterocolitica* subsp. *enterocolitica* O:8/1B konnte Wachstum auf N-acetyl-galactosamin festgestellt werden. Neben diesen Genen können möglicherweise auch zwei Prophagen (PhiYep-2 und PhiYep-3) und eine *asn* tRNA assoziierte genomische Insel (GIYep-01) zur Pathoadaptation von *Y. enterocolitica* subsp. *palearctica* Serobiotyp O:3/4 beitragen. Der

PhiYep-3 Prophage und die GIYep-01 Insel weisen Rekombinationsaktivität auf, und PhiYep-3 wurde nicht in allen untersuchten Serobiotyp O:3/4 Stämmen gefunden.

Y. enterocolitica subsp. *palearctica* Serobiotyp O:5,27/3 Stamm Y527P ist genetisch eng verwandt zu allen Serobiotyp O:3/4 Isolaten, wohingegen die Biotyp 1A Isolate ein mehr Mosaik-artiges Genom aufweisen und potentielle Virulenzgene sowohl mit Serobiotyp O:8/1B als auch O:3/4 gemeinsam haben, was einen gemeinsamen Vorfahren impliziert. Neben dem pYV Virulenz-Plasmid fehlen den Biotyp 1A Isolaten klassische Virulenzmarker wie das Ail Adhesin, das YstA Enterotoxin und das Virulenz-assoziierte Protein C (VapC). Interessanterweise gibt es keine beträchtlichen Unterschiede zwischen den bekannten Virulenzfaktoren des nosokomialen Isolats und dem Umweltisolat der Biotyp 1A-Gruppe, abgesehen von einem verkürzten Rtx Toxin-ähnlichem Genkluster und Überresten eines P2-ähnlichen Phagen im Krankenhausisolat der Serogruppe O:5.

2. Summary

Yersinia enterocolitica subsp. *palearctica* serobiotype O:3/4 comprises about 80-90 % of all human patient isolates in Germany and Europe and is responsible for sporadic cases worldwide. Even though this serobiotype is low pathogenic, *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 is involved in gastroenteritis, lymphadenitis and various extraintestinal sequelae as reactive arthritis. The main animal reservoir of this serobiotype are pigs, causing a high rate of O:3/4 contaminations of raw pork in butcher shops in Germany (e.g. Bavaria 25 %) and countries in north-east Europe.

As *Y. enterocolitica* O:3/4 is geographically and phylogenetically distinct from the so far sequenced mouse-virulent O:8/1B strain, complete genome sequencing has been performed for the European serobiotype O:3/4 DSMZ reference strain Y11, which has been isolated from a patient stool. To gain greater insight into the *Y. enterocolitica* subspecies *palearctica* group, also draft genome sequences of two other human O:3/4 isolates (strains Y8265, patient isolate, and Y5307, patient isolate associated with reactive arthritis), a closely related *Y. enterocolitica palearctica* serobiotype O:5,27/3 (strain Y527P), and two biotype 1A strains (a nosocomial strain of serogroup O:5 and an environmental serogroup O:36 isolate) have been performed. Those strains were compared to the high-pathogenic *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B strain 8081 to address the peculiarities of the strain Y11 and the *Y. enterocolitica* subspecies *palearctica* group. The main focus was to unravel the pathogenic potential of strain Y11 and thus to identify novel putative virulence genes and fitness factors, especially those that may constitute host specificity of serobiotype O:3/4.

Y. enterocolitica subspecies *palearctica* serobiotype O:3/4 strains lack most of the mouse-virulence-associated determinants of *Y. enterocolitica* subsp. *enterocolitica* serotype O:8, for example the HPI, Yts1 type 2 and Ysa type three secretion systems. In comparison, serobiotype O:3/4 strains obviously acquired a different set of genes and genomic islands for virulence and fitness such as the Ysp type three secretion system, an RtxA-like putative toxin, insecticidal toxins and a functional PTS system for N-acetyl-galactosamine uptake, named *aga*-operon. The *aga*-operon is able to support the growth of the *Y. enterocolitica* subsp. *enterocolitica* O:8/1B on N-acetyl-galactosamine after transformation with the *aga* operon. Besides these genes, also two prophages, PhiYep-2 and PhiYep-3, and a *asn* tRNA-associated GIYep-01 genomic island might influence the *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 pathoadaptation. The PhiYep-3 prophage and the GIYep-01 island show recombination activity and PhiYep-3 was not found in all O:3/4 strains of a small strain collection tested.

Y. enterocolitica subsp. *palearctica* serobiotype O:5,27/3 strain Y527P was found to be closely related to all serobiotype O:3/4 strains, whereas the biotype 1A isolates have more mosaic-segmented genomes and share putative virulence genes both with serobiotypes O:8/1B and O:3/4, which implies their common descent. Besides the pYV virulence plasmid, biotype 1A strains lack classical virulence markers as the Ail adhesin, the YstA enterotoxin, and the virulence-associated protein C. Interestingly, there are no notable differences between the known virulence factors present in nosocomial and environmental strains, except the presence of a truncated Rtx toxin-like gene cluster and remnants of a P2-like prophage in the hospital serogroup O:5 isolate.

3. Introduction

3.1. The genus *Yersinia*

3.1.1. General characteristics of the human pathogenic Yersinia species

The bacteriologist Alexandre Emile Jean Yersin (1863 - 1943) first discovered the etiological agent of plague to be caused by a bacterium, named *Yersinia pestis* in 1964, and demonstrated that this bacterium was involved in both the rodent and human disease, thus unravelling the putative mechanisms of transmission. Species of the *Yersinia* genus are Gram-negative, facultative anaerobe, non-sporulating and coccobacillus shaped bacteria, featuring a wide growth temperature range of 4-42 °C around the optimum temperature of 28°C. With its ability to grow at low temperatures as 4°C, *Yersinia* species are characterised psychrophilic.

The genus *Yersinia* belongs to the family of *Enterobacteriaceae* and today consists of 18 species (Garrity 2004), including the recently discovered *Y. aleksiciae* (Sprague and Neubauer 2005), *Y. entomophaga* (Hurst *et al.* 2011), *Y. massiliensis* (Merhej *et al.* 2008), *Y. nurmii* (Murros-Kontiainen *et al.* 2010a), *Y. pekkanenii* (Murros-Kontiainen *et al.* 2010b) and *Y. similis* (Sprague *et al.* 2008). *Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica* are the only species considered to be human pathogenic (reviewed in Brubaker 1991). *Y. pseudotuberculosis* and *Y. pestis* are closely related species, and *Y. pestis* is reported to have evolved from *Y. pseudotuberculosis* 1,500-20,000 years ago, shortly before the first known pandemics of human plague (Achtman *et al.* 1999). In contrast, *Y. enterocolitica* presents a more variable genomic structure with less than 65 % DNA homology when compared to *Y. pestis* and *Y. pseudotuberculosis* (Achtman *et al.* 1999;reviewed in Bottone 1999).

Yersinia pestis, usually transmitted by flea bites (*Xenopsylla cheopis*), disseminated to regional lymph nodes and then spread systemically via the bloodstream, caused three pandemics (reviewed in Perry and Fetherston 1997): Justinian's plague (541–542), which started in Africa and spread from Egypt around countries of the Mediterranean Sea (Brossolet and Mollaret 1994), the Black Death (starting from 1347, causing declining epidemic waves up to 1750), which spread from the Caspian Sea to Europe and caused the death of about one third of the European population (Biraben 1975) and a third plague pandemic that started from China and spread globally via shipping from Hong Kong in 1894, the year A. Yersin discovered the bacteria to be the etiological cause of the plague (Yersin 1894b,a;Achtman *et al.* 1999). Three biovars (Antiqua, Medievalis, and Orientalis) have been yet distinguished within *Y. pestis*, based on minor phenotypic differences (Achtman *et al.* 1999). Since *Y. pestis* does not produce O-antigens, classification based on serotyping is precluded for this species (Skurnik *et al.* 2000). Recent findings

ascertained that the Black Death plague was imported to Europe on at least two occasions. These two clades of the different import routes are ancestral to the modern isolates of *Y. pestis* biovars Orientalis and Medievalis (Haensch *et al.* 2010).

Yersinia enterocolitica and *Yersinia pseudotuberculosis* are important food-borne pathogens that both cause gastroenteritis and lymphadenitis and are generally transmitted by the oral-faecal route (mostly through contaminated food and water sources). Immuno-compromised patients develop even septicaemia with multiple liver abscesses (Gayraud *et al.* 1993). Both species enter the intestinal epithelium via the M cells and then disseminate to the lymphatic system (Peyer's patches and subsequently mesenteric lymph nodes) where bacterial replication occurs extracellular (Marra and Isberg 1997).

Y. pseudotuberculosis can be classified into 21 different serogroups based on surface antigens (reviewed in Wren 2003). Serogroups O I-III are most prevalent in Europe, whereas O II, O IV and O V(a,b) occur mostly in Japan (Aleksic 1995;Tsubokura *et al.* 1989).

Y. enterocolitica strains are isolated from both aquatic environments and animal reservoirs, with domestic pigs and wild boars providing the most human pathogenic strains (reviewed in Bottone 1997). *Y. enterocolitica* strains can be classified based on biochemical metabolism reactions into six different biotypes (BT): 1A, 1B and biotypes 2-5 (Wauters *et al.* 1987). The majority of isolates from environmental sources are of the non-pathogenic BT 1A. Members of the BT 1A are generally considered to be apathogenic in humans and animals, lacking some of the established virulence determinants (reviewed in Revell and Miller 2001), whereas the human virulent BT 1B is highly pathogenic and lethal in mice. The other biotypes 2-5 are reported to be low pathogenic, because they are non-lethal for humans (reviewed in Carniel 2002). Based on serological tests (addressing the O-antigens and H-antigens), about 60 different serogroups (SG) can be distinguished, of which eleven have been described to correlate with virulence (Wauters *et al.* 1987;reviewed in Bottone 1997). The main recognized pathogenic *Y. enterocolitica* serobiotypes (SG/BT) within human isolates reported in Great Britain were O:9/2 and O:3/4 (McNally *et al.* 2004), likely to be referred to a high prevalence of both lines in slaughter pigs. Serobiotype O:3/4 is predominant in Europe, but nowadays emerging also globally (Bonardi *et al.* 2003;Martinez *et al.* 2011;Rosner *et al.* 2010). Serobiotype O:8/1B is the most abundant isolated in the United States. Together with the less frequently isolated serobiotype O:5,27/3, serobiotypes O:3/4 and O:9/2 are designated “Old World” strains, due to their prevalence in Europe and Japan. In contrast, serobiotype O:8/1B and other serogroups of the BT 1B that are mostly disseminated in the north of America, are thus designated “New World” strains (reviewed in Bottone 1997;reviewed in Schubert *et al.* 2004).

In the last decades, also serogroups ascribed to the biotype 1A class have been shown to cause gastroenteritis and diarrhoea, especially in humans with underlying disorders (Burnens *et al.* 1996;Scheftel 2002). Also, at least two gastrointestinal outbreaks associated with 1A strains are reported (Greenwood and Hooper 1990;Ratnam *et al.* 1982). There are several controversial reports on the pathogenicity potential of clinical and non-clinical BT 1A strains (Grant *et al.* 1998;Tennant *et al.* 2003b,2005;Thoerner *et al.* 2003). Some BT 1A strains of clinical origin were reported to penetrate epithelial cells, to survive within macrophages, and to colonise the intestinal tract of mice (Tennant *et al.* 2003a;Grant *et al.* 1998). However, there is also evidence that infections with BT 1A yersiniae are more common in patients that are generally predisposed to infections (Ratnam *et al.* 1982;reviewed in Bhagat and Virdi 2011). So most probably the pathogenicity potential of the BT 1A group might be mainly an attribute of the host that fails to resist the attack of a relatively innocuous agent.

3.1.2. Yersinia enterocolitica palearctica

The differentiation between biotype 1B ("New World" strains) and the "Old World" strains introduced in the last section has been further pursued in the assembly of two subspecies, *Y. enterocolitica* subsp. *enterocolitica* for biotype 1B predominant in the United States and *Y. enterocolitica* subsp. *palearctica* for the "Old World" strains (BT 1A, 2-5), based on different 16S rRNA types (Trebesius *et al.* 1998). The new *Yersinia* subspecies *palearctica* comprehends the bacterium's major residence in northern "arctic" environments (Neubauer *et al.* 2000).

3.1.3. Clinical manifestations and epidemiology of Y. enterocolitica infections

After the uptake of contaminated food or water, or also after the administration of contaminated blood transfusions (Haverly *et al.* 1996;Hoelen *et al.* 2007;Leclercq *et al.* 2005), the primarily gastrointestinal tract pathogens enter the M cells, replicate extracellular in the Payer's patches and spread to the lymphatic system, causing a variety of symptoms, compendiously expressed as yersiniosis (reviewed in Cornelis *et al.* 1998). The pathogens can cause acute enteritis (prevalent in young children), enterocolitis, mesenteric lymphadenitis, and terminal ileitis. Common symptoms are fever and abdominal pain, attended by (often bloody) diarrhoea. Those symptoms typically develop within 4 to 7 days after the exposure and can last up to 4 weeks or even longer in young children, whereas in adults, symptoms abate normally within two weeks (L. A. Lee *et al.* 1990). In adults, a right-sided abdominal pain together with fever can dominate, often leading to a misdiagnosis of appendicitis (Chandler and Parisi 1994). Septicaemia can develop from extra intestinal spreading of the pathogens, preferably in immuno-compromised patients, including abscess formation in liver or spleen, meningitis and others (Bottone 1997;Rabson *et al.* 1975;Sonnenwirth 1970;Gayraud *et al.* 1993;Schuchmann *et al.* 1997). Also, post-infection

secondary immunologically mediated sequelae are often reported, including the most common, reactive arthritis and erythema nodosum (Bottone 1997).

The reactive arthritis due to a previous *Y. enterocolitica* infection normally occurs after a latent period of 1-3 weeks post infection. It is an extra intestinal, immunologically mediated sterile inflammation of joints (especially knee and ankle, lower extremities), accompanied with painful swelling and heat development in the respective joints (Wollenhaupt 2007). It can involve the inflammation of the eyes (conjunctivitis or uveitis anterior) as well as urethritis (men) or cervicitis (women). The causative pathogens cannot be isolated from joint fluids, but a persistence of the bacteria within other tissues as lymph nodes is discussed. This may lead to the stimulation of the production of specific antibodies directed against the *Yersinia* as well as the presence of specific antigens (especially lipopolysaccharides and heat-shock proteins), which was reported to be detected within peripheral blood mononuclear cells and polymorphonuclear phagocytes of the joint fluids. After four years post infection, bacterial antigens were still detected in the peripheral blood cells of most patients suffering from reactive arthritis (Granfors *et al.* 1998;Marsal *et al.* 1981). *Yersinia* post infection arthritis was also studied in Lewis rats. It could be demonstrated that complement activation leads to joint inflammation (K. Gaede *et al.* 1992;K. I. Gaede *et al.* 1995). Interestingly, a connection between Human Leukocyte Antigen (HLA)-B27 has been described for the *Yersinia* based reactive arthritis (Granfors 1998;Schiellerup *et al.* 2008;Wollenhaupt 2007). HLA-B27 is a class I surface antigen encoded by the major histocompatibility complex (MHC) and provides antigenic peptides to T-cells. It is associated also with other spondyloarthropathy diseases, as the ankylosing spondylitis (AS). In general, the therapy of reactive arthritis is based on a prolonged administration with the same antibiotics used to eliminate *Yersinia* infections.

Erythema nodosum is a painful inflammation of subcuticular fat cells (panniculitis) characterized by tender red nodules or lumps as an immunologic response to a variety of different causes, including bacterial infections. It is likely to appear 2 weeks after the onset of gastrointestinal symptoms (Yotsu *et al.* 2010). Biopsy of the nodules often approves the diagnosis.

Interestingly, the established serobiotypes of *Y. enterocolitica* differ in their clinical manifestation, with serobiotype O:8/1B causing more frequently septicaemia with liver abscesses, extensive ulceration of the gastrointestinal tract and fatal casualties (Gutman *et al.* 1973), whereas serobiotypes O:3/4 and O:9/2 cause less destructive symptoms in the gastrointestinal tract (L. A. Lee *et al.* 1990). In contrast, reactive arthritis and erythema nodosum have been reported predominantly for infections caused by *Y. enterocolitica* serobiotype O:3/4 (Bottone 1997), which can be referred to a higher incidence of the HLA type B27. Due to the higher incidence of cases associated with *Y. enterocolitica* serobiotype O:3/4 in Europe, clinical cases caused by

contaminated blood transfusions have been reported especially for this serobiotype (Leclercq *et al.* 2005;Roussos *et al.* 2001). Even though, considering that *Y. enterocolitica* serobiotype O:9/2 is isolated from stool samples in only 6 % of enteric patients (see Table 3.1), but in about 20 % of septic cases (Guinet *et al.* 2011), one can conclude a higher virulence potential for O:9/2.

In Germany, infections caused by *Yersinia enterocolitica* have to be reported (referring to the German § 7 IfSG). The number of reported cases of yersiniosis in Germany reported by the RKI (Robert Koch Institut) was 3,731 in 2009, 4,354 cases in 2008 and 4,987 in 2007. The incidence per 100,000 inhabitants was 4.5, 5.3 and 6.1, respectively, demonstrating a reduction in the recorded cases for Germany during the last years. Interestingly, the number of reported cases for serobiotype O:8/1B increased slightly in Germany. Cases of yersiniosis predominate within the group of young children, with a peak beneath the one-year old. In addition, incidences are slightly raised for men and in specific parts of Germany. Table 3.1 shows the incidences of yersiniosis and reported serogroups (as available) in Germany in the last three years (data summarised from: Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2007-2009, Robert Koch Institut).

Table 3.1 Incidences of reported cases of yersiniosis in Germany, 2007-2009

	2007	**2008**	**2009**
Number of reported cases in Germany	4,987	4,354	3,731
General Incidence per 100,000 inhabitants	6.1	5.3	4.5
Incidence of serogroup O:3 in %	90	88	89
Incidence of serogroup O:9 in %	6	7	6
Incidence of serogroup O:5,27 in %	0.8	0.8	0.7
Incidence of serogroup O:8 in %	0.2	0.5	0.6

3.2. Pathogenicity factors of *Yersinia enterocolitica*

Human virulent strains of *Y. enterocolitica* share the presence of a 64-75 kb large yersinial virulence plasmid (pYV), encoding specific proteins (Ysc) of a type three secretion system (T3SS) as well as *Yersinia* outer protein effector proteins (Yops) which are substrates of type three secretion systems (T3SS) and other virulence determinants (as YadA and LcrV) which are indispensable for the pathogenicity of *Y. enterocolitica*. The pYV plasmid is reported to be absent from BT 1A strains, demonstrating their mostly apathogenic phenotype. During the early stages of infection, several plasmid or chromosomally encoded genes respond to the temperature shift to 37°C, leading to the modification of surface antigens (Portnoy and Martinez 1985). For example, Yops are expressed only at 37°C (Bottone 1997;Cornelis 1998;Iriarte and Cornelis 1996). Virulence gene expression of *Yersinia* is also controlled by the concentration of calcium in the environment. Thus, the T3SS mediated secretion of Yop effector proteins occurs, apart from the

temperature shift to 37°C, preferentially in the absence of calcium (Heesemann *et al.* 1987). In addition, infection is mediated by the interplay with several chromosome-encoded virulence genes, including among others the invasin gene *invA*, the mucoid *Yersinia* factor *myfA*, the *ail* attachment and invasion locus gene and the Yst enterotoxins (see the following sections). Notably, *Y. enterocolitica* O:8/1B strains also harbour a High-Pathogenicity Island (HPI), encoding the yersiniabactin system for uptake and utilisation of iron (Perry *et al.* 1999;Schubert *et al.* 2004;Heesemann *et al.* 1993;Carniel *et al.* 1996;Schubert *et al.* 1999). This genomic island is essential and causative for the mouse lethal phenotype of *Yersinia enterocolitica* BT 1B. This High-Pathogenicity Island is absent from *Y. enterocolitica palearctica.* In the following sections, some of the actual most important factors of *Y. enterocolitica* virulence are discussed in particular.

3.2.1. Chromosomally encoded virulence factors

3.2.1.1. The yst enterotoxins

Y. enterocolitica strains produce a heat-stable enterotoxin (Yst) with similar features compared to the methanol-soluble heat-stable enterotoxin of enterotoxigenic *Escherichia coli* (ETEC) (Delor *et al.* 1990;K. Okamoto *et al.* 1983). Six different Yst proteins have been detected among the *Yersinia* species, each either of 16 to 20 amino acid residues or with 30 amino acid residues after the cleavage of an N-terminal signal sequence (Takao *et al.* 1984;Takao *et al.* 1985). Even though *yst* genes have been described both in pathogenic and apathogenic members of the species, the production of the Yst toxins has been reported only for pathogenic species (Delor *et al.* 1990). Nevertheless, the toxins are absent from *Y. pestis* and *Y. pseudotuberculosis*. The toxins can be isolated from cultures grown below 30°C (Feeley *et al.* 1979). The heat-stable Yst toxins are ligands for the guanylyl cyclase receptor and considered as superagonists, since they activate the guanylyl cyclase receptor maximally. Thus increasing cGMP levels in the intestine, the toxins are suggested to play a role in the secretion of electrolytes during watery diarrhoea (Rao *et al.* 1979). It matches the finding that Yst toxins exhibit secretory activity in the suckling mice model (Boyce *et al.* 1979;K. Okamoto *et al.* 1982), causing watery diarrhoea and death in the suckling mice. An *yst* mutant strain was found defective in the induction of diarrhoea and death in comparison to wild-type *Y. enterocolitica* (Delor and Cornelis 1992). Thus, the Yst enterotoxins play an important role in the mediation of virulence of *Y. enterocolitica*.

3.2.1.2. Invasin InvA

The Invasin encoded by *inv* is an outer membrane protein which enables attachment and entry into the M cells of the intestinal mucosa (Isberg *et al.* 1987;Marra and Isberg 1997). The N-terminal part of the protein forms a β-barrel that is anchored in the outer membrane, exposing the C-

terminal part on the bacterial surface. The C-terminal domains of InvA have folding topologies that are similar to immunoglobulin superfamily domains. InvA is responsible for promoting both cell adhesion and uptake by binding tightly to the family of α/β1- integrins (α3β1, α4β1, α5β1, α6β1 and αvβ1), mediating the specific interaction of *Yersinia* with non-phagocytic M cells. The tight binding to integrins mediates the subsequent bacterial uptake via the "zipper"-like mechanism, enclosing the host cell membrane around the bacteria (Cossart and Sansonetti 2004;Isberg and Leong 1990;Clark *et al.* 1998). Different invasin domain interactions have been characterised and described for *Y. pseudotuberculosis* to build the adhesion complex and to permit the multimeric attachment of single invasin molecules (Hamburger *et al.* 1999;Isberg and Barnes 2000). The globular domain D2 responsible for the multimeric composition of the invasion molecule in *Y. pseudotuberculosis* is absent in the 92 kDa invasin protein in *Y. enterocolitica* (V. B. Young *et al.* 1990), which can elucidate the observed lower invasiveness of *Y. enterocolitica* compared to *Y. pseudotuberculosis* (Dersch and Isberg 2000). The invasin is regulated by different mechanisms, illustrating its particular role in pathogenesis. The expression of *inv* is controlled by temperature (reduced expression at 37°C, optimum for expression at 23°C) and pH (optimum observed at neutral pH) (Pepe *et al.* 1994). Transcription is activated upon RovA binding, a MarR-type transcriptional regulator, and superposed by binding the repressive histone-like H-NS protein, which in turn is down-regulating the transcription at 37°C, blocking the activation via RovA (Ellison and Miller 2006;Heroven *et al.* 2004;Nagel *et al.* 2001). It has been shown that RovA transcription is negatively influenced by the LysR-type regulatory protein RovM, and the disruption of the *rovM* gene leads to increased RovA levels and invasin synthesis and thus enhances the internalization of *Y. pseudotuberculosis* into host cells (Heroven and Dersch 2006;Revell and Miller 2000). In addition to H-NS, the YmoA protein is reported to play a role in the negative regulation of *inv* (Ellison *et al.* 2003).

The invasin encoded by *inv* is only produced by *Y. enterocolitica* and *Y. pseudotuberculosis* (Isberg *et al.* 1987;Miller *et al.* 1988). In *Y. pestis*, the *inv* gene is interrupted by an IS200-like element. It has been discussed to be an adaptation to a different lifestyle (parasitism), with the loss of expression of genes that are no more essentially required, including also invasion and motility genes (Simonet *et al.* 1996). Likewise favouring a particular role in the adaptation to a different niche or lifestyle, the invasin is produced constitutively in serobiotype O:3/4 strains and thus is enhanced compared to other *Y. enterocolitica* serobiotypes, in which the expression of *invA* is significantly reduced at 37°C. The atypical increase of invasin levels has been ascribed to the insertion of an IS1667 element into the *invA* promoter region and a P98S substitution within the *invA* activator protein RovA, rendering the regulator less susceptible to proteolysis. Variations in both features were shown to influence colonization of the bacteria in mice (Uliczka *et al.* 2011).

3.2.1.3. Mucoid Yersinia factor MyfA

Virulent *Y. enterocolitica* serobiotypes produce a 21 kDa Myf (mucoid *Yersinia* factor) fibrillious adhesin. MyfA involves a chaperone/usher secretion pathway which is encoded by the *myfEFABC* gene cluster. Gene *myfB* encodes a chaperone and *myfC* a membrane usher protein, that are both utilised in MyfA pilus assembly and transport, whereas *myfEF* codes for two transcriptional regulators, essential for *myfA* transcription. The whole Myf machinery is about 44 % identical to the pH6 (Psa) antigen of *Y. pestis* and *Y. pseudotuberculosis,* an antigen which is synthesized by *Y. pestis* at 37°C host body temperature and acidic pH values (Ben-Efraim *et al.* 1961). The Psa antigen functions as an adhesin toward erythrocytes and mammalian epithelial cells (Galvan *et al.* 2007;Isberg 1989b,a;Liu *et al.* 2006). In contrast, Myf does not mediate haemagglutination in *Y. enterocolitica* as described for the *Y. pseudotuberculosis* pH6 antigen (Iriarte and Cornelis 1995). Thus, the actual role of the MyfA antigen in the pathogenicity of *Y. enterocolitica* remains unclear.

3.2.1.4. Attachment and invasion locus gene ail

The *ail* gene (attachment-invasion locus) is present only in virulent strains of *Y. enterocolitica* and plays a role in the invasion of specific mammalian cells such as CHO (Chinese hamster ovary cells) (Miller and Falkow 1988;Miller *et al.* 1989). The 17 kDa Ail outer membrane protein is, like several other yersinial virulence factors, mainly expressed at 37°C (Pierson and Falkow 1993). It is predicted to comprise eight membrane spanning β-strands and four extracellular loops located close to the cell membrane (Beer and Miller 1992;Miller *et al.* 1990). Ail is, together with YadA, reported to mediate resistance to complement killing, and even the recombinant expression of Ail in *E. coli* cells mediated complement resistance. This resistance was reported to be evoked by the interaction of the C4b binding protein of both the lectin and classical pathway of the complement cascade to the Ail protein (Biedzka-Sarek *et al.* 2008a;Biedzka-Sarek *et al.* 2008b;Kirjavainen *et al.* 2008;Pierson and Falkow 1993). Ail allowed the invasin-independent invasion of *E. coli* into different cell lines, with invasion rates dependent on the specific cell types as HEp-2 (epidermoid carcinoma cells of male human), HEC-1B (adenocarcinoma cells of human uterus) and CHO cells (Miller and Falkow 1988), but no underlying molecular mechanism has yet been established for Ail. It has been proposed that Ail could interfere with the formation of complement attack complexes (Biedzka-Sarek *et al.* 2005;Bliska and Falkow 1992).

3.2.2. Virulence plasmid (pYV) encoded pathogenicity factors

3.2.2.1. Adhesin YadA

The *Yersinia enterocolitica* Adhesin A (YadA) mediates cell adherence (Heesemann and Grüter 1987). It is the prototype of trimeric autotransporter adhesins, consisting of a head, stalk and a β-barrel membrane anchor. It was denoted oligomeric coiled-coli adhesin (Hoiczyk *et al.* 2000;Roggenkamp *et al.* 2003). The YadA homotrimers of about 200-240 kDa are at their C-termini anchored with the β-barrel structures in the outer membrane, exposing "lollipop"-like globular head N-termini on the bacterial surface, creating a hydrophobic area. Thus, autoagglutination occurred by the presence of YadA in stationary phase cultures and was referred to the hydrophobic properties. This autoagglutination can play a role in the formation of micro-abscesses in different tissues (Balligand *et al.* 1985;reviewed in El Tahir and Skurnik 2001;Lachica and Zink 1984;Skurnik *et al.* 1984). The YadA fibrillae are 1.5-2 nm in diameter and exhibit an overall length of ~23 nm (Hoiczyk *et al.* 2000). Recently, the trimer stability of YadA has been shown to be indispensable for pathogenicity of *Y. enterocolitica*. The amino acid residue G389 was defined as the critical residue of YadA required for the formation of the beta-barrel translocator pore as well as for serum resistance and trimer stability *in vitro* (Schutz *et al.* 2010). YadA seems to be not essential for the pathogenicity of *Y. pseudotuberculosis*, and in *Y. pestis* the *yadA* gene has been reported to harbour a frameshift mutation silencing the gene (Rosqvist *et al.* 1988;Skurnik and Wolf-Watz 1989). The *Yersinia* adhesin YadA was previously named P1 (Bolin *et al.* 1982) or autoagglutination protein (Skurnik *et al.* 1984) and has been associated to a multitude of biological functions during the last decades of research. The gene *yadA* is transcribed at 37°C host body temperature but not at temperatures below 30°C, dependent on the YadA expression activator VirF (LcrF in *Y. pseudotuberculosis*) (Lambert de Rouvroit *et al.* 1992;Skurnik and Toivanen 1992). Up-regulation of the transcription and expression can be detected at the surface of the bacteria within minutes after the temperature shift to 37°C (Bolin *et al.* 1982). The periplasmic Bam (Omp85) complex, a highly conserved multi-protein machine essential for the assembly of β-barrel outer membrane proteins, has been shown to be essential for the correct assembly of YadA (Lehr *et al.* 2010). In contrast to invasin and other yersinial virulence genes, YadA expression is not Ca^{2+}-dependent (Bolin *et al.* 1982;Skurnik 1985;Skurnik *et al.* 1984). Similar to the Ail protein, YadA mediates complement resistance and thus plays a role in serum resistance during infection (Biedzka-Sarek *et al.* 2008b;Kirjavainen *et al.* 2008;Heesemann and Laufs 1983).

Within the various biological functions of the YadA, its essential role in colonisation by binding of collagen structures has been reported for *Y. enterocolitica*, whereas in *Y. pseudotuberculosis* and *Y. pestis*, YadA has negative effects on virulence (Pepe *et al.* 1995;Tamm *et al.* 1993).

Collagen binding is mediated by a NSVAIGXXS motif of the YadA head domain which binds to the α_1-chain of collagens (Emody *et al.* 1989;Tahir *et al.* 2000). In addition, binding to fibronectin and laminin has been described (reviewed in El Tahir and Skurnik 2001), and adhesion to epithelial cells and neutrophils has been described (Heesemann *et al.* 1987;Roggenkamp *et al.* 1996;Ruckdeschel *et al.* 1996). Recently it was shown that only YadA-expressing *Y. enterocolitica* adhered to neutrophils extracellular traps and were thus killed. The presence of collagen within these neutrophils extracellular traps was responsible for this binding and subsequent killing (Casutt-Meyer *et al.* 2010). Interestingly, the presence of YadA was crucial for the development of reactive arthritis in a rat model (Gripenberg-Lerche *et al.* 1995;Gripenberg-Lerche *et al.* 1994;K. I. Gaede and Heesemann 1995).

3.2.2.2. Yops and the type three secretion system

Type three secretion systems (T3SS) are complex protein export machineries that are used by many Gram-negative bacteria, either to export flagella (Fla-T3SS) or to secrete effector proteins. Effector proteins can be secreted into the extracellular milieu or into the cytosol of target host cells (reviewed in Plano *et al.* 2001). The type three secretion system of *Yersinia* is plasmid encoded and belongs to the Yop virulon. The *Yersinia* Yop virulon consists of the type three secretion system apparatus, named Ysc injectisome (Ysc-T3SS), and a set of proteins secreted by this machinery, named Yops. The basal part of the Ysc injectisome is similar to the basal body of flagella, spanning the bacterial membrane, while the distal part is similar to needle like structures, formed by the polymerization of a 200-300 6 kDa YscF proteins. The needles were isolated and a length of 60–80 nm and a width of 6–7 nm were measured. The needle tip has a diameter of about 20 Å (Hoiczyk and Blobel 2001;reviewed in Cornelis 2002a;Cornelis 2002b). The T3SS in *Yersinia* performs tight injection of the effectors to the host cells. Even though, Yops are secreted to the culture medium by Ca^{2+} chelation *in vitro*.

In *Yersinia*, the type three secretion system effectors affect the regulation of the innate immune response. Macrophages and neutrophils are important elements during early innate immune defence and phagocytise invading bacteria. The recognition of pathogen-associated molecular patterns (PAMPS), such as LPS (lipopolysaccharide), via the family of pattern-recognition receptors (PRRs) by those cells activates several MAPKs and the NF-κB signalling pathways. This leads to an inflammatory response producing proinflammatory cytokines, as TNF-α, IL-12 and IL-18. The Yops mainly antagonize the uptake of *Yersinia* by phagocytic immune cells and are also required for *Yersinia* to counteract the production of proinflammatory cytokines during the infection (reviewed in Brubaker 2003;Cornelis 2002a). There are five Yops detected in all *Yersinia* species: YopH, YopM, YopE, YpkA/YopO and YopJ/YopP. YopT is found only in

Y. enterocolitica. Three other Yops, YopB, YopD and LcrV, resemble not typical host effector proteins but play a role in the formation of a pore like structure in the target cell and are thus required for the translocation of others Yops. LcrV, in addition to its role as a translocator protein, plays also a role in Yop synthesis regulation together with LcrG. Also YopN seems to take part in the complex regulation of Yop secretion (Mueller *et al.* 2005;Nilles *et al.* 1998;reviewed in Hamad and Nilles 2007).

YopH is a potent tyrosine phosphatase, able to dephosphorylate proteins that were post-translational phosphorylated at tyrosine residues. It has a putative mammalian origin and is an essential role in virulence (Bolin and Wolf-Watz 1988;Guan and Dixon 1990). Since crystallographic analysis showed similarity to mammalian phosphatases, YopH is likely to attack host cell substrates (Phan *et al.* 2003;Sun *et al.* 2003). It is likely that YopH is kept silent in the bacterium by the association of YopH with its chaperone SycH. After secretion, the active 50 kDa protein causes a decrease in tyrosine phosphorylated host proteins, among the focal adhesion kinase (FAK), paxillin, Src-like kinases (for example Lck) and others especially for $p130^{Cas}$, a focal adhesion protein required for phagocytosis (reviewed in Cornelis 2002a;Viboud and Bliska 2005). The YopH mediated inactivation leads to a loss of focal adhesions and thus to disruptions in actin cytoskeleton structures (Andersson *et al.* 1996). The terminal region of the protein was shown to direct the YopH to the substrates in the host's focal adhesions (Black *et al.* 1998). In addition to these anti-phagocytotic effects, YopH was shown to play a role in the suppression of the oxidative burst in polymorphonuclear leucocytes (PMN cells) (Ruckdeschel *et al.* 1996), to counteract the activation of B- and T- lymphocytes and to participate in the induction of apoptosis in T-lymphocytes (Alonso *et al.* 2004;Bruckner *et al.* 2005).

YopE is a non-canonical GTPase activating protein (GAP) that is both an essential virulence factor and cytotoxic to host cells, as it disrupts actin microfilaments in infected cells (Rosqvist *et al.* 1994;Rosqvist *et al.* 1995). It harbours an N-terminal localisation sequence to direct the protein to host membranes (Krall *et al.* 2004). The protein inhibits small G-proteins of the Rho-family, including RhoA, Cdc42, Rac1 and especially RhoG, by inducing their GTPase activity and leading to the accumulation of the inactive GDP bound state, thus being essential in virulence and anti-phagocytotic activity (Aili *et al.* 2002;Aili *et al.* 2006;Andor *et al.* 2001;Von Pawel-Rammingen *et al.* 2000;Roppenser *et al.* 2009). Crystallographic analysis revealed that YopE is only functionally redundant with mammalian GAPs, denying a mammalian origin of the protein (Evdokimov *et al.* 2002). Beside the GAP mimicry, YopE is reported to be involved in the resealing of pores previously made by the T3SS, thus preventing the delivery of cellular contents to the extracellular milieu, which would activate the host immune system (Viboud *et al.* 2006). In

addition, a YopE mutant results in a hyper-translocation phenotype, indicating a role for the protein to regulate effector protein secretion rates (Aili *et al.* 2008).

YopM is a member of the leucine-rich repeat protein superfamily, involved in scaffolding. Presently, a catalytic activity of YopM has not been identified, but the protein contains 3–20 repeats of a 19-residue leucine-rich repeat (LRR), known to mediate protein-protein interactions (reviewed in Cornelis 2002a). About 4 YopM monomers assemble to a hollow cylinder with an inner diameter of 35 Å (Evdokimov *et al.* 2001). YopM is translocated into the nucleus via an endocytic, vesicle associated pathway, and has the relevant nuclear transport signal at the C-terminal end (Benabdillah *et al.* 2004;Skrzypek *et al.* 2003). In addition, the N-terminus mediates autotranslocation across the cytoplasmic membrane of mammalian cells (Ruter *et al.* 2010). However, if the protein plays a role in altered transcriptional regulation is not clarified, and oppositional studies exist (Hoffmann *et al.* 2004;Sauvonnet *et al.* 2002). Instead, protein targets have been identified to be activated by YopM, namely a ribosomal S6 protein kinase 1 (Rsk1) and a protein kinase c-like 2 protein (Prk2). The YopM effector is proposed to bind Rsk1 and to act as a scaffold binding Prk2, leading to the activation of different kinases that take part downstream of the YopJ blocked MAPK pathway (mitogen-activated protein kinase pathway) (McDonald *et al.* 2003). Even though the distinct function of YopM needs further scientific attention, at least a contribution to virulence has been reported, since a YopM deficient strain showed decreased virulence in the mouse model (Leung *et al.* 1990;Mulder *et al.* 1989).

YopT is a 35 kDa papain-like cysteine protease which occurs exclusively in *Y. enterocolitica* and *Y. pseudotuberculosis* and has been shown to cleave a cysteine residue of post-translational prenylated Rho, Rac, and Cdc42 near the C-terminus, releasing them from the membrane (Iriarte and Cornelis 1998;Shao and Dixon 2003;Shao *et al.* 2002;Shao *et al.* 2003;reviewed in Cornelis 2002a). YopT thus leads to the mislocalisation of those G-proteins, followed by the disruption of Rho-dependent actin microfilaments during infection (Shao *et al.* 2003;Zumbihl *et al.* 1999). This in turn inhibits the cellular cytoskeleton rearrangements essential for phagocytosis, especially shown for macrophages and neutrophils (Grosdent *et al.* 2002). In addition, interplay with YopE, leading to reduced pore formation, has been reported (Viboud *et al.* 2006).

YopO (named YpkA in *Y. pseudotuberculosis* and *Y. pestis*) is an 80 kDa multifunctional protein with an N-terminal serine/threonine kinase and a Rho-GTPase binding domain and a C-terminal actin binding domain (Barz *et al.* 2000;Galyov *et al.* 1993;Juris *et al.* 2000;reviewed in Trosky *et al.* 2008). As shown for YopE, YopT and YopH, it affects the actin cytoskeleton and contributes to the ability of *Yersinia* to resist phagocytosis (Grosdent *et al.* 2002;Trasak *et al.* 2007;Wiley *et al.* 2006;reviewed in Navarro *et al.* 2005). Interestingly, YpkA is reported to be an inactive kinase until actin binds to the last amino acid residues of the protein. The protein is localised at the host

cell membranes (Dukuzumuremyi *et al.* 2000;Hakansson *et al.* 1996;Juris *et al.* 2000;reviewed in Navarro *et al.* 2005;Viboud and Bliska 2005). It has been shown that the C-terminal GTPase binding domain (amino acid pos. 434-732) of YpkA interacting with Rac1 at its switch I and switch II regions is structural similar to a host guanidine nucleotide dissociation inhibitor (GDI). YpkA thus may inhibit nucleotide exchange in Rac1 and RhoA, and Prehna and colleagues have shown that mutations disrupting the YpkA-GTPase interface dispose this activity *in vitro* (Prehna *et al.* 2006). In addition, the interaction with switch II imitates the inactive conformation of the protein bound to GDP, resulting both in the inhibition of Rac1 functions and of the formation of actin fibres. Recently, it was discovered that the YpkA kinase domain inhibits the Gαq signalling pathways, following the stimulation of G protein coupled receptors (GPCRs), by phosphorylating a serine residue in the GTP binding pocket. Members of the Gαq family of G proteins mediate the activation of the phospholipase-C-β and further can act on RhoA-mediated pathways (Navarro *et al.* 2007). Studies performed with YopO showed a specific block of the Rac-dependent Fcγ receptor internalization pathway and a prevention of Rac activation. In addition, the plasma membrane localization and the guanine-nucleotide dissociation inhibitor (GDI)-like domain of YopO are, like in YpkA, associated with anti-phagocytosis (Groves *et al.* 2010).

YopP of *Y. enterocolitica* (YopJ in *Y. pseudotuberculosis* and *Y. pestis*) is a serine/threonine acetyltransferase but also exhibits cysteine protease and deubiquitinylation activities (Mukherjee *et al.* 2006;Orth *et al.* 2000;Zhou *et al.* 2005). However, the exact mode of action for YopP/YopJ in target cells remains unsolved. YopJ is, as other YopJ-like proteins, supposed to attenuate signalling pathways in the host by its acetyltransferase activity (Mittal *et al.* 2006;Mukherjee *et al.* 2006;reviewed in Trosky *et al.* 2008). YopP has been also shown to downregulate the onset of the host's inflammatory response by blocking the NF-κB and MAPK pathways, which may further induces the activation of the apoptotic cascade in macrophages and dendritic cells (Erfurth *et al.* 2004;reviewed in Viboud and Bliska 2005;Cornelis 2002a).

3.2.3. Chromosomally located genomic islands of Yersinia enterocolitica subsp. enterocolitica

3.2.3.1. The plasticity zone of Y. enterocolitica subsp. *enterocolitica serobiotype O:8/1B strain 8081*

The plasticity zone (PZ) is a large chromosomal region in the genome of *Yersinia* species that has undergone a series of independent insertions at this site. It is located adjacent to the *phe* tRNA gene and accounts for about 16 % of the coding sequences (CDSs) unique to *Y. enterocolitica* (about 199 kb, chromosomal position 3,761,922-3,960,673 in *Y. enterocolitica* subsp. *enterocolitica* strain 8081, encoding 186 CDSs as described by Thomson (Thomson *et al.* 2006)). Several functional units are found within this region in strain 8081, including a region similar to

the *Y. pseudotuberculosis* adhesion pathogenicity island (named YAPIytb in *Y. pseudotuberculosis* and YAPIye in *Y. enterocolitica*), the Ysa-T3SS (discussed below) and a general secretion pathway named Yst1, as well as several metal-uptake operons and resistance-gene loci (Thomson *et al.* 2006). In *Y. pseudotuberculosis*, the YAPIytb island encodes a type IV pilus operon essential for virulence (Collyn *et al.* 2004). The homologous island in *Yersinia enterocolitica* strain 8081 is smaller than the YAPIytb counterpart, and encodes a type IV pilus operon, as well as a variable region that encodes among others a putative haemolysin (YE3454) and a second arsenic-resistance operon (YE3472–YE3475). Additional genes of the PZ have been described as niche specific gene clusters, including a hydrogenase 2 biosynthetic operon, an orthologue of ProP, a chitinase that could be secreted by Yst1, a ferric enterochelin operon including *fepBDGC, fes* and *fepA*. Microarray studies have confirmed the massive genomic variations in the PZ, even among the BT 1B isolates. Large parts of the PZ have been reported as absent from *Y. enterocolitica palearctica* strains, such as the YAPIye, the Ysa-T3SS and the Yst1 general secretion pathway (Thomson *et al.* 2006).

3.2.3.2. The Ysa type three secretion system

Highly pathogenic *Y. enterocolitica* strains of BT 1B carry a second type three secretion system, located on the Ysa pathogenicity island, which is part of the chromosomal plasticity zone (PZ). The Ysa-T3SS is more related to the Spa-T3SS of *Shigella* species (Foultier *et al.* 2003;Haller *et al.* 2000;G. M. Young 2007), and absent from *Y. enterocolitica palearctica* strain Y11. Unexpectedly, among the new identified Ysa effector proteins (namely YspA, YspE, YspF, YspI, YspK, YspL, YspM and YspP) the Ysc-T3SS proteins YopE, YopN and YopP were found (B. M. Young and Young 2002). YopN was previously described to be involved in the regulation of the Ysc-T3SS (reviewed in Hamad and Nilles 2007), but it may also have an unrecognized effector function (Matsumoto and Young 2009). Matsumoto and Young discuss the possibility that the Ysc-T3SS plays the major role in systemic stages of the infection, whereas the Ysa-T3SS is selective for gastrointestinal infection. This hypothesis needs more evaluation of the virulence effectors delivered by the Ysa-T3SS . The Ysa effector protein genes are located throughout the genome. One *ysp* gene is located within the PZ, three are found on the pYV plasmid (as part of the Ysc-T3SS), and the seven remaining are neither PZ nor pYV associated. This observation leads to the hypothesis of severe lateral gene transfer and a selective pressure to maintain the Ysa-T3SS (Matsumoto and Young 2006).

3.2.3.3. The High-Pathogenicity Island

The High-Pathogenicity Island (HPI) is a genomic island reported to be essential for pathogenicity and the mouse-virulent phenotype in *Yersinia*. Pathogenicity islands represent a subset of different

vehicles for horizontal gene transfer. They are usually already immobilized, and so the HPI lacks the properties necessary for its replication and self-transmission, but a phage P4-like integrase, an excisionase and recombination sites encoded on the HPI illustrate former mobility aspects of the island. The HPI encodes for the biosynthesis and transport of a functional siderophore, designated yersiniabactin (Ybt), with genes for biosynthesis and uptake, which enables elevated iron scavenging and thus contributes to the pathogen's fitness and virulence (Carniel *et al.* 1996;Rakin *et al.* 1999;reviewed in Perry and Fetherston 1997). Ybt was shown to reduce reactive oxygen species (ROS), and among different tested iron chelators, yersiniabactin was the most effective in reducing the ROS production in innate immune cells (Paauw *et al.* 2009).Two evolutionary groups of HPIs were established, designated Yps HPI for *Yersinia pestis* and *Yersinia pseudotuberculosis* and Yen HPI for *Yersinia enterocolitica enterocolitica* BT 1B (Rakin *et al.* 1995). The Yen HPI is restricted to *Y. enterocolitica enterocolitica* BT 1B serogroups O:8, O:13, O:20 and O:21, whereas the Yps HPI is widely disseminated among members of the family of *Enterobacteriaceae* (Bach *et al.* 2000;Rakin *et al.* 1999;Schubert *et al.* 2000a;Schubert *et al.* 2000b).

3.2.4. Additional aspects of the Yersinia genus relevant for this study

3.2.4.1. Lipopolysaccharides of Yersinia enterocolitica (LPS)

The lipopolysaccharide (LPS) constitutes the major bacterial surface antigen of Gram-negative bacteria. It is composed of the lipid A core and the O-polysaccharide (O-antigen). For serogroups O:8 and O:3, the composition of the LPS has been resolved. The structure of the *Yersinia* lipid A seems to be structural identical to those of other members of the family of *Enterobacteriaceae* (Muller-Loennies *et al.* 1999). Biological effects associated with *Yersinia* LPS mostly involve the lipid A core. Nevertheless, the O-antigen of pathogenic Gram-negative bacteria has been reported to play an important role in the resistance to the innate immune response, resulting in the effective colonization of host tissues and to complement-mediated.

The O-antigen of *Y. enterocolitica* serogroup O:8 consists of branched pentasaccharide O-units of N-acetyl-D-galactosamine (GalNAc), L -fucose (Fuc), D -galactose (Gal), D -mannose (Man) and 6-deoxy-D -gulose (6d-Gul). The O-units are linked with (1→4) glycosidic bonds between the GalNAc and Man residues (Tomshich *et al.* 1976). The O-antigen gene cluster of *Y. enterocolitica* serobiotype O:8/1B comprises 18 genes, including genes for the biosynthesis of the O-antigen, the transfer of N-acetyl-D-galactosamine to a lipid carrier molecule, undecaprenyl phosphate (Und-P), by specific glycosyltransferases and genes encoding enzymes required for heteropolymeric O-antigens building. One protein, Wzz, is required to control the O-antigen length of 7–10 repeats (Bengoechea *et al.* 2002;Zhang *et al.* 1996).

The *Y. enterocolitica* serogroup O:3 O-antigen is a homopolymer of (1→2)-linked 6-deoxy-L -altrose that is attached to the inner core region of the LPS (Hoffman *et al.* 1980). In addition to the O-antigen, also the outer core (OC) hexasaccharide is attached to the inner core, generating a branch in the LPS molecule. This specific structure, which is absent from serogroup O:8, has allowed the generation of mutants missing either the O-antigen, the OC or both (Sirisena and Skurnik 2003;Skurnik *et al.* 1999;Skurnik *et al.* 1995).

Both the *Y. enterocolitica* serogroup O:3 and O:8 O-antigens are indispensable for virulence, and thus mutants without the O-antigen remained attenuated in mice. Neither of the mutants colonized the Peyer's patches as efficiently as the wild type strains and no microcolonies were found for spleen, liver and mesenteric lymph nodes. An OC mutant of *Y. enterocolitica* serogroup O:3 did colonize the Peyer's patches as efficiently as the wild type but was eliminated within 5 days and was less efficient in the colonisation of deeper tissues. Thus, it seems that the O-antigen in serogroup O:3 is essential within the first hours of infection, whereas the OC is required for the prolonged survival in Peyer's patches and for the invasion of deeper organs as liver and spleen (Skurnik and Bengoechea 2003). However, it is generally accepted that O-antigens are involved in the resistance to complement-mediated killing as well as to antimicrobial peptides, and some virulence factors anchored in the outer membrane require the presence of the O-antigens. Even though, the explicit role of O-antigens in virulence is still unclear (Skurnik *et al.* 1999).

3.2.4.2. Rtx toxins

The large family of Rtx toxins of Gram-negative bacteria comprises many different classes of effector proteins that all share glycine- and aspartate-rich sequence motifs at the carboxy-terminal end of the protein, normally nonapeptides. The consensus domain structure of this motif is GGXGXDX[L/I/V/W/Y/F]X or [L/I/F]XGGXG[N/D]DX. These motifs are important for the binding of calcium and constitute its name: Rtx - repeat in toxin. The number of repeats varies and can reach up to 40 nonapeptides (Satchell 2007;Welch 2001;reviewed in Linhartova *et al.* 2010). In addition, also the toxin secretion via the type one secretion system is a common feature of this group of proteins, but secretion via atypical type one secretion systems (T1SS) can occur (Boardman and Satchell 2004;B. C. Lee *et al.* 2008). The transport via these systems is ATP dependent, SecYEG independent, directs the proteins across both the inner and the outer membrane in one step and recognises target proteins at their C-terminal secretion signal sequences. Those transport genes are mostly located clustered with an activator (*rtxC*) and the pretoxin (*rtxA*) that build the mature Rtx toxin. The protein is supposed to be secreted unfolded or in a floppy conformation, followed by conformational changes promoted by extracellular calcium that binds to the nonapeptide motifs (Felmlee and Welch 1988;Ludwig *et al.* 1988;Rhodes *et al.*

2001;Rose *et al.* 1995). Even similar in genetic organisation, the mode of action differs among the Rtx family members, and the biological functions have been used to further subdivide the group of Rtx toxins (reviewed in Linhartova *et al.* 2010). There are cytotoxic Rtx proteins as well as proteases, lipases, bacteriocins, nodulation proteins and proteins involved in the S-layer of bacterial membranes (Linhartova *et al.* 2010).

The group of cytotoxic proteins can be even more subdivided into pore forming toxins and a multifunctional, autoprocessing group of Rtx toxins (MARTX) (Satchell 2007;Li *et al.* 2008). The pore forming Rtx toxins act by integrating pores in the host cell membrane. They are primarily leukotoxins that typically show a restricted host and cell type range and are 100-120 kDa in size. Even though, haemolysins with a broad spectrum are reported. The action of these leukotoxins is dose dependent. They can bind β_2- integrins, but also the binding to glycoprotein structures without cell specific receptors using electrostatic forces is discussed (Bakas *et al.* 1996;Lally *et al.* 1997;Ostolaza *et al.* 1997). The pore forming Rtx toxins are mostly acylated at specific lysine residues (fatty acylation), assumed to be essential for activation, but also non-acylated toxins form pores with similar properties (Ludwig *et al.* 1996;Masin *et al.* 2005). Still the formation of the pores is not completely understood. Obviously, cell damage occurs via the uncontrolled in- and efflux of electrolytes through the pores, and massive calcium influx has been observed (Welch 2001). In general, MARTX family members are large proteins consisting of 3,000 to 5,000 amino acid (aa) residues in size. They exhibit prolonged repeat variations of the common nonapeptide motif (X(V/I)XXGXXNX(V/I)XXGDGXDX) as well as two new classes of repeats: a 20-residue consensus sequence GXXG(N/D)(L/I)(T/S)FXGAG(A/G) XNX(L/I)X(RH) and a 19-residue consensus T(K/H)VGDGX(S/T)VAVMXGXAN(I/V)X (Lin *et al.* 1999;Satchell 2007). In addition, the genes encoding these Rtx proteins are organised differently. They comprise two operons, *rtxHCA* and *rtxBDE*, with the former operon containing a protein of unknown function (RtxH) and the latter operon encoding the atypical T1SS, which is mostly not linked to the first cluster (Boardman and Satchell 2004;reviewed in Linhartova *et al.* 2010). Modifications and the mode of action are still unclear, but models have been raised that assume the proteins to integrate into the host cell membrane, translocating specific effector domains into the cytosol by an autoprocessing domain (CPD, cysteine protease domain) that cleaves the effectors to release them into the target cell (Satchell 2007). The autoprocessing is induced by binding inositol hexakisphosphate ($InsP_6$), which is present at the inner surface of the cell membrane and reaches up to 40–60 μm in several cell types (Prochazkova and Satchell 2008;Reineke *et al.* 2007). In general, the effects observed for different members of the MARTX family include for example cell rounding and the disintegration of the actin cytoskeleton and thus evoke an acute

inflammatory response in the host (Cordero *et al.* 2006;Kudryashov *et al.* 2008a;Kudryashov *et al.* 2008b).

3.2.4.3. Insecticidal toxin complex

The insecticidal toxin complex has been firstly described for *Photorhabdus luminescens*, and subsequently for *Y. pestis, Y. pseudotuberculosis* and other members of the *Yersinia* genus (Bowen *et al.* 1998;N. R. Waterfield *et al.* 2001;reviewed in N. Waterfield *et al.* 2007). Four high molecular weight toxin complexes have been described yet, termed *tca*, *tcb*, *tcc* and *tcd*. These Tc toxins may be active against different tissues or cell types, for example Tcb against haemocytes and Tcd as well as Tca against cells of the insect gut (reviewed in ffrench-Constant *et al.* 2003;Heermann and Fuchs 2008). The genetic context as well as the number of toxin genes (*tc*) varies among the different species. In the already sequenced isolate of *Y. enterocolitica enterocolitica*, strain 8081, a Tc cluster is absent (Thomson *et al.* 2006), whereas serobiotype O:9/2 harbours *tc* genes of the *tcd* operon type (Fuchs *et al.* 2008;reviewed in N. Waterfield *et al.* 2007). The *tc* genes of this serobiotype are located on a genomic island (*tc*-PAI^{Ye}), and are low-temperature induced, whereas their transcription is repressed at 37°C (Bresolin *et al.* 2006). In contrast, *tcaABC* expression in *Y. pseudotuberculosis* strain IP 32953 was observed at 15°C and at 37°C (Pinheiro and Ellar 2007). Recently, the different *tc* genes in yersiniae have been compared (Fuchs *et al.* 2008). According to the classification of Waterfield (reviewed in N. Waterfield *et al.* 2007), the *tc* genes of the *Yersinia* species all belong to the class of *tcd* operons. The length of the different *tc* clusters ranges from ~17 kb to ~26 kb, due to the presence of one or more *tccC* homologues. Interestingly, the backbone localisation of the clusters is the same in all genomes. The presence of the *tc*-PAI^{Ye} correlates with a higher toxicity of yersiniae towards larvae of the tobacco hornworm, and strains lacking the *tc* genes are thus less insecticidal in assays performed by the authors (Fuchs *et al.* 2008). In addition to the Tc specific PAI, one or two *tccC* genes located outside the *tc*-PAI^{Ye} were identified in different *Y. pestis* and *Y. pseudotuberculosis* strains. These *tccC* genes might have been acquired independently by a further horizontal gene transfer (HGT) event.

3.2.4.4. Mobile genetic elements and genome plasticity

Since complete genome sequencing projects of prokaryotes increased, the expanded set of available genomes enables a closer investigation of the genome architecture of bacterial species. Genome plasticity is one key word which explains the evolutionary rapid adaptation to different environments and lifestyles of bacteria. Intraspecies and interspecies exchange of genetic information (including transformation, transduction and conjugation), which normally is advantageous for fitness and/or virulence, depends on the species' barriers to interact with foreign

genetic material and other species, and the observed frequent exchange of genetic material of several enterobacterial species reflects their mosaic genome structure (Dobrindt *et al.* 2010a;Dobrindt *et al.* 2010b;Preston *et al.* 1998). This interplay hampers the classical taxonomy investigations, and the concepts of a core gene pool and a flexible gene pool have been established (reviewed in Dobrindt *et al.* 2004). The size of the core gene pool is not a constant value, but depends on the number and diversity of the organisms being compared and comprises all genes shared by all organisms investigated. The remaining genes make up the flexible gene pool, thus containing genes that are shared only between subgroups of the organisms studied as well as unique genes of one organism. The flexible gene pool is shaped by genomic rearrangements, insertions, (frequent) point mutations as well as deletion events (genome reduction), the latter being difficult to be traced but assumed especially for obligate intracellular pathogens or symbionts (reviewed in Moran 2002). Many genetic elements forming the modules that are exchanged between species and foreign organisms by HGT have been identified by analysing variations in the GC-ratio throughout the genomes and were named on the basis of their function, including GEIs (genomic islands, mostly used for islands lacking virulence genes), PAIs (pathogenicity islands, comprising mobile elements of different types encoding virulence factors), and MEIs (metabolic islands). Other genetic elements are bacteriophages, IS elements (insertion sequence elements) and ICEs (integrative and conjugative elements) (Dobrindt *et al.* 2010a;Dobrindt *et al.* 2010b;Hacker *et al.* 1990;Knapp *et al.* 1986). These elements replicate episomally or integrate into the chromosome, which is typically afforded by specific mobility enzymes as recombinases and integrases, and mostly depends on site-specific recombination at special anchor sites in the chromosome, for example phage attachment sites (named *attB* for bacterial attachment site and *attP* for the phage attachment site) or direct repeats. Genomic islands that have integrated into the bacterial chromosome can undergo rearrangements, insertions and deletions and may evolve from virulence lacking to pathogenicity islands or lose their ability to excise, leading to immobilisation and lateral transmission. One example, a large PAI in the *Yersinia* genus is the HPI (see section 3.2.3.3) which carries genes for the biosynthesis and uptake of the siderophore yersiniabactin, present only in the highly pathogenic, mouse lethal *Yersinia* species (Rakin *et al.* 1999). PAIs are mostly located adjacent to specific tRNA genes that are highly conserved between various bacterial species (Williams 2002), and the HPI of *Yersinia* species is found inserted to one of the *asn* tRNA genes. It has been shown for *Y. pseudotuberculosis* that any of the three *asn* tRNA sites in the *Y. pseudotuberculosis* genome are suitable for HPI integration (Buchrieser *et al.* 1998). Since the HPI-integrase is a tyrosine recombinase closely related to P4-like phage integrases it was concluded that the HPI originated from a former bacteriophage, but no flanking phage genes adjacent to the integrase gene have been observed (reviewed in Carniel 1999). However, the frequent acquisition and rearrangement

of foreign genetic material which leads to the major genome plasticity of the *Yersinia* genus and contributes to the different niche and host adaptation of the species, awaits further analysis by extensive complete genome comparisons.

3.3. Genome sequencing technologies

In 1977, Frederick Sanger published a new method for fast and efficient DNA sequencing, the chain-termination or dideoxy method (Sanger *et al.* 1977). This method uses the inclusion of dideoxynucleotide triphosphates (ddNTPs) among the standard dNTPs in the DNA synthesis process to inhibit DNA strand extension at random positions of the template. The ddNTPs lack the 3' hydroxyl (OH) group that is essential for the formation of phosphodiester bonds between the nucleotides during the new DNA strand synthesis, aborting further elongation and resulting in variable length of the synthesised DNA strands. These fragments are subsequently separated on sensitive gel matrices, allowing the discrimination of single nucleotide length variations of the DNA fragments and hence to read out the genetic code of the template DNA. This process has to be performed in parallel reactions with each all four deoxynucleotides (dATP, dGTP, dCTP, and dTTP) but only one type of the respective dideoxynucleotides at a time. The comparison of the different gel separation pattern thus allows the reconstruction of the template DNA. The reaction depends on a specific recognition sequence in the template DNA strand to bind a suitable oligonucleotide to start the reaction. Thus, in case of genome sequencing approaches, the genomic DNA has to be fragmented and fused to appropriate recognition sequences, which is enabled by vector based cloning techniques (for example BACs, bacterial artificial chromosomes, or other plasmids that can carry larger insert constructs), where the genomic DNA is integrated into a vector backbone to generate sequencing templates. After the sequencing process, these bases have to be removed manually or automatically (vector clipping). This Sanger method, over years the gold standard of (industrial) DNA sequencing, was improved consistently, using dye-coupled nucleotides and the respective detection equipment for automation. Still, the system suffers from disadvantages, mainly the poor data quality in the first 15–40 bases (requires base calling and trimming analysis) and the descending quality after ~900 bp read length. In addition, homopolymeric stretches can disturb the DNA strand elongation and thus result in wrong nucleotide numbers. The normally vector-bound DNA templates lead to high contamination of the output data with vector sequences which have to be excluded from the assembly of the read data in case of complete genome sequencing.

Eleven years after the introduction of the Sanger method, in 1988, the pyrosequencing or “454” method has been developed, using a single-nucleotide addition (SNA) method for faster DNA sequencing (Hyman 1988;Ronaghi *et al.* 1996;reviewed in Ronaghi 2001). Independent of the

elongation inhibition using ddNTPs, the four individual dNTPs are supplied and removed successive in a fixed order to the DNA synthesis reaction (the template DNA strand is immobilised), and each incorporation event of a single nucleotide to the new DNA strand generates a chemiluminescent signal which can be recorded by a camera. This signal is based on the detection of a released pyrophosphate (PPi) during the nucleotide incorporation in the new DNA strand synthesis, followed by a cascade of enzymatic reactions: the free PPi is converted to ATP by the ATP sulfurylase enzyme, and the energy is used to oxidate luciferin by the luciferase enzyme and generates light signals measured as peaks. This series of detected peaks is named pyrogram and corresponds to the order of nucleotides that are added and allows the reconstruction of the template DNA. In standard pyrosequencing methods, one pmol of DNA yields 6 x 10^{11} ATP molecules which leads to the emission of more than 6 x 10^{9} photons at a wavelength of 560 nanometers (reviewed in Metzker 2005;Ronaghi 2001). The light emissions are proportional to the number of incorporated nucleotides, thus generating larger peaks in case of homopolymorphic stretches. Still, the incorporation of dATP generated false signals. It was shown they were attributed to the fact that dATP itself is a substrate for luciferase. An inert form of dATP was developed that could be incorporated efficiently by all DNA polymerases tested but was no longer catalysed by the luciferase enzyme (Ronaghi *et al.* 1996). The improvement of this method in the last years was mainly concentrating on the average read length, which designates the length of the new synthesised DNA strand and thus the length of information read-out. Today, the read length is converging to the Sanger read length. One disadvantage of the pyrosequencing system is the difficulty in determining the number of incorporated nucleotides in homopolymeric stretches larger than 5-6 nucleotides, and specific software programs are needed for accurate *de novo* sequencing of these stretches.

Today, many different sequencing platforms (454, Illumina (Solexa), SOLiD, MegaBACE) are available to provide suitable applications for different sequencing approaches, including single nucleotide polymorphism sequencing, genotyping as well as rapid high-throughput *de novo* sequencing. In the field of complete genome sequencing, the shotgun approaches (using genomic DNA fragments broken randomly prior to vector dependent or independent sequencing) have been expanded by the invention of paired-end sequencing, which supplies additional data about read orientation and distances between two sequenced templates. The technique is based on the sequencing of the ends of a fragment of a defined length (as the case for the defined insert length in vector based libraries). This technique facilitates the assembly of the raw sequencing data and bypasses some of the difficulties in complete genome sequences, including the presence of massive repetitive elements. However, still the incredible large sets of raw data from whole

genome sequencing approaches need further improvements in the downstream processes, including data handling and assembly as well as the complete genome reconstruction.

3.3.1. Vector based sequencing methods of this study

3.3.1.1. MegaBACE sequencing

The MegaBACE™ DNA analysis system is a fluorescence-based system utilizing capillary electrophoresis related to the Sanger sequencing method. It produces long read lengths with average read lengths of ≥ 500 bases with calculated accuracy scores of ≥ 20. For complete genome shotgun sequencing approaches, vectors are needed to provide the template DNA for sequencing. In this study, the pCR® 4 Blunt TOPO® vector system was used (Invitrogen, Karlsruhe).

3.3.1.2. Illumina sequencing

Illumina's sequencing by synthesis (also known as Solexa or SBS) technology is based on reversible dye-terminators. It enables the detection of single bases as they are attached to the growing DNA strands. As for 454 sequencing, the DNA molecules are attached to oligonucleotides on a sample slide to be clonally amplified for enhanced signals (bridge amplification). A fluorescent-dye labelled terminator is detected for the incorporation of each dNTP and subsequently removed to allow further elongation. All four reversible terminator-bound dNTPs are present in each sequencing cycle, which is differing to the 454 method, and so DNA strands can be extended by only one dNTP at a time, minimising the incorporation bias. The read length is about 75x2 bp, which means the system can sequence 75 bases on either end of a 100-200 bp template. Thus, this technique is mainly used for paired-end sequencing.

3.3.2. Vector independent sequencing methods of this study

3.3.2.1. 454 GS-20 and FLX Titanium sequencing

The 454 system relies on a water-in-oil emulsion reaction mix with fixed, adapter-ligated DNA templates (that have been generated by nebulisation of genomic DNA, shearing the DNA into 300-800 bp) that are bound to small streptavidin-coated beads. The adaptors provide the recognition sequences for both the amplification and sequencing of the fragments. Clonally amplification of the template DNA allows an optimised signal in the sequencing process, thus enhancing the accuracy. (Zheng *et al.* 2010). The GS-20 system was introduced in 2005 and allowed the sequencing of up to 100 base pairs per read. In 2008, the 454 GS FLX Titanium sequencing technology was introduced, producing average reading length of more than 350 bp.

3.4. Aims of the thesis

Y. enterocolitica subsp. *palearctica* serobiotype O:3/4 comprises about 80-90 % of all human patient isolates in Germany and Europe, with rising global relevance (Rosner *et al.* 2010;Bonardi *et al.* 2003;Martinez *et al.* 2011). As other serobiotypes, it is involved in gastroenteritis, lymphadenitis and various extra intestinal sequelae as reactive arthritis (Granfors 1998;reviewed in Bottone 1997). The main animal reservoir of this serobiotype are asymptomatic and ill pigs (Fredriksson-Ahomaa *et al.* 2006), causing a high rate of O:3/4 contaminations in butcher shops in Germany and countries in north-east Europe (Bucher *et al.* 2008;Fredriksson-Ahomaa *et al.* 2004;Fredriksson-Ahomaa *et al.* 2003;Laukkanen *et al.* 2009).

As *Y. enterocolitica* O:3/4 is not only geographically but also phylogenetically distinct from the so far sequenced O:8/1B strain (Thomson *et al.* 2006), this thesis followed the aim 1) to determine the complete genome sequence of the European serobiotype O:3/4 DSMZ reference strain Y11 (isolated from a patient stool) as well as draft genome sequences of two other human O:3/4 isolates (strains Y8265, patient isolate, and Y5307, patient isolate associated with reactive arthritis) and a closely related *Y. enterocolitica palearctica* serobiotype O:5,27/3 (strain Y527P), 2) to perform comparative genome analyses, with the main focus on differences to the already sequenced strain 8081 (serobiotype O:8/1B), 3) to identify novel putative virulence genes and fitness factors, especially those that may constitute host specificity of serobiotype O:3/4, and 4) to investigate selected genes by mutagenesis, metabolic analyses and functional genomics.

4. Materials and Methods

4.1. Equipment

Table 4.1 contains specific laboratory equipment. If not mentioned in this table, standard laboratory devices have been used. For plastic consumables, standard products from Eppendorf (Hamburg), Nunc (Wiesbaden), Becton Dickinson (Heidelberg), Brand (Wertheim) and Renner (for TPP culture flasks, Dannstadt) have been used.

Table 4.1 Equipment

Machine	Model	Supplier
Centrifuge	Megafuge 3.0R	Heraeus, Hanau
Centrifuge	1K15	Sigma, Osterode
Electrophoresis chambers	Mini-PROTEAN 3	Bio-Rad, Munich
Electroporation device	GenePulser II	Bio-Rad, Munich
Film development device	Fujifilm FPM-100A	Fuji, Düsseldorf
Gel documentation	GelDoc EQ	Bio-Rad, Munich
Incubator for 27°C	INP-500	Memmert, Schwabach
Incubator for 37°C	Type B20	Heraeus, Hanau
PCR cycler	GeneAmp 2400	PE Applied Biosystems, Weiterstadt
PCR cycler	Unocycler	VWR, Darmstadt
pH meter	pH211	Hanna Instruments, Kehl
Photometer	Ultraspec 3100 pro	Amersham Biosciences, Freiburg
Semi-Dry Blot device	Trans-Blot SD	Bio-Rad, Munich
Shaking incubators	Certomat BS-1	B. Braun Biotech, Melsungen
Sterile flow hood	Herasafe HS12	Heraeus, Hanau
Transilluminator	UVT-20M/W	Herolab, Wiesloch
Pipettes	Labmate	Abimed, Langenfeld
Power Supply Electrophoresis	PowerPac 300	Bio-Rad, Munich

4.2. Chemicals and kits

If not mentioned elsewise, chemicals, enzymes and antibiotics have been purchased from Amersham Biosciences (Freiburg), Invitrogen (Karlsruhe), MBI Fermentas (St. Leon-Rot), Merck Biosciences (Darmstadt), Roche Diagnostics (Mannheim), Roth (Karlsruhe), Sigma-Aldrich (Taufkirchen) and VWR (Darmstadt). Restriction enzymes have been delivered solely by MBI Fermentas (St. Leon-Rot) and New England Biolabs (Frankfurt a. M.). Broth powder has been supplied by Becton Dickinson (Heidelberg) and cell culture media (including additives) by Invitrogen (Karlsruhe). Selective agar plates for *Yersinia* (CIN) and blood agar plates have been

obtained from Oxoid (Wesel). Nitrocellulose membranes and Whatman-paper have been purchased from Schleicher & Schüll (Dassel). Sugars for growth experiments have been obtained from Sigma-Aldrich (Taufkirchen). Specific kits used are listed in Table 4.2.

Table 4.2 Chemicals and kits used in this study

Chemicals and kits	Supplier
1 kb DNA-ladder GeneRuler™ and GeneRuler™ plus	Fermentas, St. Leon-Rot
GoTaq DNA polymerase	Promega, Mannheim
LA-Taq	TaKaRa, Saint-Germain-en-Laye, France
Maxima™ First Strand cDNA synthesis kit	Fermentas, St. Leon-Rot
Methyltransferases	New England Biolabs, Frankfurt a. M
MonsterScript™ 1st strand cDNA synthesis kit	Epicentre, Biozym, Hess. Oldendorf
NucleoBond® AXG	Macherey - Nagel, Düren
NucleoBond® PC 100	Macherey - Nagel, Düren
NucleoSpin® Extract II	Macherey - Nagel, Düren
NucleoSpin® Plasmid QuickPure	Macherey - Nagel, Düren
NucleoSpin® Tissue	Macherey - Nagel, Düren
PerfectPro Ni-NTA MagBeads Manual	5Prime, Hamburg
pGEM®-T Easy kit	Promega, Mannheim
Phusion und PhusionFlash Mastermix	Finnzymes, New England Biolabs, Frankfurt a. M.
Red Taq DNA polymerase Mastermix	VWR, Darmstadt
RNAProtect® Bacteria Reagent	Qiagen, Hilden
RNeasy Mini kit	Qiagen, Hilden
RTS™ 100 E. coli HY kit	5Prime, Hamburg
RTS™ 100 E. coli LinTempGenSet	5Prime, Hamburg
Spectra™ Multicolour Broad Range Protein Ladder	Fermentas, St. Leon-Rot
StrataClone Ultra Blunt PCR Cloning Kit	Stratagene, Agilent, Waldbronn
T4 DNA Ligase	New England Biolabs, Frankfurt a. M.
Test sera for agglutination of O:3, O:8	Sifin, Berlin
TypeOne Restriction Inhibitor	Epicentre, Biozym, Hess. Oldendorf

4.3. Bacterial strains, mammalian cell lines and plasmids

Bacterial strains, plasmids and mammalian cells are listed in Table 4.3 - Table 4.6. Strains used for genome comparison studies are listed with the respective accession numbers, if available. Nomenclatures of the strain characteristics for *E. coli* refer to Bachmann (Bachmann 1990).

Table 4.3 Bacterial strains: *Yersinia*

Bacterial strains: *Yersinia*	**Characteristics**	**Reference**	**Accession numbers**
Y11 (13030 DMSZ)	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, carrying pYVO:3	(Neubauer *et al.* 2000)	FR729477, FR745874 (plasmid)
Y11, Nal100	Nalidixic acid 100 resistant	This study	
Y11, Nal10	Nalidixic acid 10 resistant	This study	
Y-108C	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, pYV-cured	(Heesemann *et al.* 1984)	
Y11 Δ*rtxA*	Y11 derivate, deletion within *rtxA*	This study	
WA-314	*Yersinia enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B, carrying pYVO:8, U.S.A.	(Heesemann *et al.* 1984)	
8081	*Yersinia enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B, carrying pYV, U.S.A.	(Thomson *et al.* 2006)	NC_008800, NC_008791 (plasmid)
MRS40	Yersinia enterocolitica serobiotype O:9/2	(Sory *et al.* 1995)	
MRS40 Δ*rtxA*	Deletion within *rtxA*	This study	
IP2222	Environmental isolate of *Yersinia enterocolitica* subsp. *palearctica*, serobiotype O:36/1A	Wauters	CACZ01000001-CACZ01000074
NF-O	Clinical isolate of *Yersinia enterocolitica* subsp. *palearctica* serobiotype O:5/1A, Newfoundland	(Ratnam *et al.* 1982)	CACY01000001-CACY01000097
Y8265	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, human isolate, France	(Saken *et al.* 1994)	CACU01000001-CACU01000014
Y5307	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, reactive arthritis patient	SaBw strain collection, Dr. Neubauer	CACV01000001-CACV01000018
Y5,27	*Yersinia enterocolitica* serobiotype O:5,27/3	MvP-strain collection	CACW01000001-CACW01000020
Y751/IP23222	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Great Britain	SESAHS strain collection, Dr. Pham	

Bacterial strains: ***Yersinia***	**Characteristics**	**Reference**	**Accession numbers**
Y748/IP21981	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, France	SESAHS strain collection, Dr. Pham	
Y746/IP24232	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Canada	SESAHS strain collection, Dr. Pham	
PL-15 353/96	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Poland, Warszawa	PZH collection, Dr. Gierczyński	
Y765/IP22274	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Australia	SESAHS strain collection, Dr. Pham	
Y756/IP7036	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Australia	SESAHS strain collection, Dr. Pham	
Y745/IP24231	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/3, Japan	SESAHS strain collection, Dr. Pham	
Y752/IP23357	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Brazil	SESAHS strain collection, Dr. Pham	
PL-6 128/99	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Poland, Szczecin	PZH collection, Dr. Gierczyński	
PL-20 254/97	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, Poland, Lublin	PZH collection, Dr. Gierczyński	
Y641	*Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4, dog puppet	SaBw strain collection, Dr. Neubauer	
Y485	Yersinia enterocolitica subsp. palearctica serobiotype O:3/4, pig	SaBw strain collection, Dr. Neubauer	
STM Nölting 113	Y. frederiksenii 0:60 H56-36/81	SaBw strain collection, Dr. Neubauer	
STM Nölting 114	*Y. intermedia* 0:17 H9-36/83	SaBw strain collection, Dr. Neubauer	
STM Nölting 115	*Y. kristensenii* 0:12,25 H17-36/83	SaBw strain collection, Dr. Neubauer	

Bacterial strains: ***Yersinia***	**Characteristics**	**Reference**	**Accession numbers**
STM Nölting 116	*Y. rhodei* 0:76 H274-36/78	SaBw strain collection, Dr. Neubauer	
STM Nölting 117	*Y. bercovieri* 0:16 H632-36/85	SaBw strain collection, Dr. Neubauer	
STM Nölting 118	*Y. mollaretii* 0:59 H279-36/86	SaBw strain collection, Dr. Neubauer	
STM Nölting 119	*Y. ruckeri* 529-36/85	SaBw strain collection, Dr. Neubauer	

Table 4.4 Bacterial strains: *E. coli*

Bacterial strains: ***E. coli***	**Characteristics**	**Reference**
DH5α	F-, Φ80 *lac*ZΔM15 Δ*(lac*ZYA-*arg*F)U169, *deo*R, *end*A1, *hsd*R17,(rk-,mk+), *sup*E44, *thi*-1,λ-, *rec*A1, *gyr*A96, *rel*A1	(Hanahan 1983)
JM109	*rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17, *sup*E44, *rel*A1 Δ(*lac*-*pro*AB) / F- [*tra*D36 *pro*AB* *lac*I^q *lac*ZΔM15]	(Yanisch-Perron *et al.* 1985)
K12 HB101	F^- *mcr*B *mrr hsd*S20(r_B^- m_B^-) *rec*A13 *leu*B6 *ara*-14 *pro*A2 *lac*Y1 *gal*K2 *xyl*-5 *mtl*-1 *rps*L20(Sm^R) *gln*V44 λ^-	(Boyer and Roulland-Dussoix 1969)
BL21 (DE3) pLys	F- *omp*T, *hsd*S_B (r_B - m_B -) *gal, dcm,* pLysS (Cm^R)	(Studier and Moffatt 1986)
S17-1 λpir	pir+ tra+	(Simon *et al.* 1983)
EPI300™-$T1^R$	F^- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ^- rpsL (Str^R) nupG trfA tonA	Epicentre

Table 4.5 Plasmids

Plasmids	**Characteristics**	**Reference**
pACYC184	Cloning vector, Cm^R, Tet^R	(Chang and Cohen 1978)
pACYC184 bac/immu	Derivate of pACYC184. JB373/JB374 bacteriocin with immunity genes, Cm^R, Tet^R	This study
pAJD434	Mutagenesis plasmid carrying red recombinase system genes, Tp^R	(Maxson and Darwin 2004)
pCB 3-6	Cointegrate of pYVO:3 and pRK290B, Tet^R	(Heesemann and Laufs 1983)
pCR® 4 Blunt TOPO®	Cloning vector for MegaBACE sequencing	Invitrogen, Karlsruhe
pET21b	Expression vector with C-terminal His_6-tag under T7-promotor control, Amp^R	Novagen, Darmstadt
pET24b	Expression vector with C-terminal His_6-tag under T7-promotor control, Km^R	Novagen, Darmstadt

Plasmids	Characteristics	Reference
pGEM®-aga	Derivate of pGEM-T easy. JB506/JB507 *aga*-operon	This study
pGEM®-T-easy	Cloning vector, lacZ, T7-Promotor, SP6-Promotor, poly-T-overlap, Amp^R	Promega, Mannheim
pKD3	Plasmid with Cm^R, used as template for PCR	(Datsenko and Wanner 2000)
pKD46	Mutagenesis plasmid carrying red recombinase system genes, Amp^R	(Datsenko and Wanner 2000)
pKNG101	Suicide plasmid, *sacB*, Sm^R	(Kaniga *et al.* 1991)
pKNG101-1	Derivate of pKNG101 with *rtxA* knockout cassette	A. Golubov, unpublished
pMOS	*lacZ*, f1 ori, blunt end ligation, Amp^R	Amersham Biosciences, Freiburg
pPirK	Helper plasmid with *pir* genes, Km^R	(Posfai *et al.* 1997)
pRK2013	Replicon: ColE1. Vector for *E. coli*. Carries Tn 903. Contains RK2 transfer genes. Helper plasmid for mobilisation of non-self-transmissible plasmids, Km^R	DMSZ no. 5599
pRK290B	BamHI-linker cloned into the EcoRI site of pRK290, Tet^R	(Heesemann and Laufs 1983)
pRK290B3-6	pRK290B carrying the sixth BamHI fragment of pYVO:3, Tet^R	Heesemann, unpublished
pSC-B	Cloning vector of StrataClone™ Ultra Blunt PCR Cloning Kit	Stratagene, Agilent, Waldbronn
pSC-B *rtxA*	Derivate of pSC-B. JB329/JB80 *rtxA*	This study
pWS	*Yersinia* expression vector, pMS470Δ8 derivate, *oriV*, $lacI^q$-Repressor, P_{tac}, *rrnB*-Terminator, Amp^R	(Locher *et al.* 2005)

Table 4.6 Mammalian cell lines

Mammalian cell lines	Description	Reference
IPEC-J2	Porcine intestinal epithelia cells	(Lu *et al.* 2002)
IPI-2I	Porcine intestinal epithelia cells	(Schierack *et al.* 2006)
HEK	Human Embryonic Kidney 293 cells	(Graham *et al.* 1977)
HeLa	Human cervical cancer cells	(Scherer *et al.* 1953)

4.4. Primers

Table 4.7 Primers used in this study

Primer name	Sequence	Characterisation
JB422	cgcaacaattgaatatcgatctgt	Gap closure: next to last gap
JB423	tgaagtcgtaacaaggtaaccgtagg	Gap closure: next to last gap
JB444	ggttatcttgccgaataagcaat	P2 phage gap2
JB446	gtcggtatctacgcgccgcat	P2 phage gap2
JB451	gagcgaatgatggttgatgcccttgccaagaa	P2 phage genomic
JB452	gtcggtatctatgcgcctcac	P2 phage genomic
JB606	ggcgtgttgtggatgtaat	PhiYep-3 phage excision
JB608	atgtcagtatatttggcgat	PhiYep-3 phage excision
JB470	agaatcggaaactttgaatggttt	GI Yep-1 excision
JB471	ttgagccgttaagagacatttgg	GI Yep-1 excision
JB472	cacatcaggcacttctccagg	GI Yep-1 excision
JB473	ttaacagaaatagcgcccat	GI Yep-1 excision
JB563	tgaccagctcgatattgagct	PhiYep-3 phage excision
JB569	gtaggttggcttcaagcgaac	PhiYep-3 phage excision
JB585	gacaggctgatgtactttca	PhiYep-3 phage excision
JB586	tggttgctgaacattaccat	PhiYep-3 phage excision
JB392	cgagttgcttgatgaaatta	*lcrV* SNPs
JB393	tgacactttagagcaactct	*lcrV* SNPs
JB394	gcgtccgctttaattaatat	*yadA* SNPs
JB395	tgtgcttttgccatattcaa	*yadA* SNPs
JB396	aatcggcaatacaacatgataa	*rtxA* SNPs 1
JB397	ttatcatgttgtattgccgatt	*rtxA* SNPs 2
JB398	gagatggtatttctatcggattac	*rtxA* SNPs 2
JB400	caacggcttgattgatagcgtt	*rtxA* SNPs 1
JB401	aga gtt gct cta aag tgt ca	*lcrV* new with JB392
JB402	atc cta tcc ctc ttc tat gg	*yadA* new with JB395
JB407	gggctagatgctaacagcag	*rtxA* SNPs 3
JB408	taacgcctgtttcacccaat	*rtxA* SNPs 3r
JB506	cagcgtcgtacttgatgatttgc	*aga*-operon
JB507	atcatctgttgggcgacacg	*aga*-operon
JB379	ctacggcccaattcacaact	Bacteriocin cDNA test
JB386	atcttcaatgcgcttttgct	Bacteriocin cDNA test
JB351	gcgcg*aagctt*atgggagagggacatagtaa	Bacteriocin for pWS with HindIII site
JB352	gcgcg*aagctt*caccatcaccatcaccatatgggagaggg acatagtaa	JB351 with His6-tag and HindIII site
JB353	gcgcgc*gtcgac*ttatttaacctctttatgaa	Bacteriocin for pWS, SalI site

Primer name	Sequence	Characterisation
JB373	atatat*ccatgg*tgtcggtattgcaaagctga	Bacteriocin gene with immunity proteins in pACYC184, NcoI site
JB374	atatat*ccatgg*tcattctttaaatccgggtt	Bacteriocin gene with immunity proteins in pACYC184, NcoI site
JB371	atatat*catatg*gggagagggacatagtaa	Bacteriocin in pET21, NdeI site
JB372	atatat*ctcgag*tttaacctctttatgaatt	Bacteriocin without stop codon in pET21, XhoI site
JB411	gcgcgc*ggatcc*caccaccaccaccaccactga	pet24b for bacteriocin, BamHI site
JB412	gcgcgc*ctcgag*cgacccatttgctgtccacc	pet24b for bacteriocin, XhoI site
JB413	gcgcgc*ctcgag*atgggagagggacatagtaa	Bacteriocin, XhoI site
JB414	gcgcgc*ggatcc*tttaacctctttatgaattt	Bacteriocin without stop codon, BamHI site
JB516	ctttaagaaggagatatacc*atgggagagggacata*	RTS bacteriocin
JB517	tgatgatgagaaccccccccc*tttaacctctttatgaattt*	RTS bacteriocin
JB80	acggtattccggttgtgaatgc	*rtxA* knockout screening
JB291	actttcaggtgagcgtgatattaga	*rtxA* knockout screening
JB329	agctggcgataagcagaaatcctgatg	*rtxA* for pSC-B
JB80	acggtattccggttgtgaatgc	*rtxA* for pSC-B
JB43for	ccaaccacagaatccgagac	*rtxA* mRNA experiments
JB44rev	atttgcttgtccaccactcc	*rtxA* mRNA experiments

4.5. Broth and antibiotics

4.5.1. Culture media

For liquid media, autoclaving was performed for 20 minutes at 121°C and 1 bar. For solid media, 15 g agar was added to one litre of liquid medium prior to heat autoclaving. Standard media used for cultivation are listed in Table 4.8.

Table 4.8 Media preparation

Medium	Component	Amount
Luria Bertani Medium (LB)	Tryptone	10 g
	Yeast powder extract	5 g
	NaCl	5 g
	H_2O ad 1 L	adjust pH to 7.4
Brain heart infusion medium (BHI)		37 g BHI powder per 1 L
Minimal medium salts	Na_2HPO_4	6 g
	KH_2PO_4	3 g
	NaCl	0.5 g
	NH_4Cl	1 g
	H_2O ad 1 L	adjust pH to 7.4
M9 minimal medium	Minimal salt medium	100 mL
	1 M $MgSO_4$	1 mL
	0.1 M $CaCl_2$	1 mL
	H_2O ad 1 L	
	Addition of the appropriate carbon source.	
	$MgSO_4$ and $CaCl_2$ sterilised by filtration.	
SOC medium	Tryptone	20 g
	Yeast powder extract	5 g
	NaCl	0.5 g
	KCl	250 mM
	Glucose	20 mM
	H_2O ad 1 L	
Medium for cryopreservation of bacteria		LB with 20 % Glycerol

4.5.2. Antibiotics

Antibiotics dissolved either in H_2O, NaOH or Ethanol were sterilised by filtration (0.22 µm filters) and stored diluted at -20°C. The appropriate concentrations are listed in Table 4.9.

Table 4.9 Antibiotics

Antibiotics	Solvent	Final concentration (µg/mL)
Ampicillin (Amp)	H_2O	100 (E. coli), 400 (Yersinia)
Carbenicillin	70 % Ethanol	25 (E. coli), 200 (Yersinia)
Chloramphenicol (Cm)	70 % Ethanol	30
Kanamycin (Km)	H_2O	50
Nalidixic acid (Nal)	0.5 N NaOH	100
Streptomycin (Sm)	H_2O	100
Tetracycline (Tet)	70 % Ethanol	20

4.6. Culture conditions

4.6.1. Growth conditions

4.6.1.1. Mammalian cells

Mammalian cell lines have been cultured in sterile plastic culture flasks at 37°C with 5 % CO_2. IPEC-J2 cells were cultivated in DMEM-F12 (1:1) with 5 % FCS, IPI-2I cells in DMEM-F12 (1:1) with 10 % FCS. Frozen samples of the cells were thawed on ice and resuspended in 5 mL medium in a falcon tube. Cells were spinned down at 1.000 rpm for 2 minutes (Heraeus Megafuge 3.0R), resuspended in 2 mL fresh medium and transferred to a prepared flask with 10 mL medium. Cells were splitted, if the bottom of the flask was over 70 % populated and/or too many rounded cells were visible in the medium. Cells were washed with PBS (see below, Table 4.10), rounded by trypsinisation, harvested with medium to block the trypsinisation process and spinned down at 1.000 rpm before resuspended and splitted into new flasks.

Table 4.10 PBS solution

1 x PBS (phosphate buffered saline)	
NaCl	8 g
KCl	0.2 g
Na_2HPO_4 x 2 H_2O	1.44 g
KH_2PO_4	0.24 g
H_2O ad 1 L	pH 7.4

4.6.1.2. Bacteria

Bacteria were grown either over-night on solid agar plates or in liquid medium (with shaking at 200 rpm), both with appropriate antibiotics if required. *Yersinia* cultures and plates were incubated at 27°C, those of *E. coli* at 37°C. Only in case of conjugation or Yop secretion, *Yersinia* cultures and plates were shifted to 37°C.

4.6.2. Cryopreservation

4.6.2.1. Cells

Mammalian cells were harvested, washed with PBS and frozen in their appropriate medium containing 10 % DMSO. Cells were stored in a chamber surrounded by liquid isopropanol, cooling the cells 1°C per minute when placed at -80°C. After cooling, cells were frozen for short time storage at -80°C. For long time storage, freezing in liquid nitrogen was performed.

4.6.2.2. Bacteria

Bacteria were grown over-night on agar plates or in liquid medium and harvested either by cotton-sticks or centrifugation at about 4,000 g, respectively. Bacterial pellets were resuspended in cryopreservation medium (LB containing 20 % glycerol) and stored at -20°C for short time storage or at -80°C for long term storage of the strains.

4.6.3. Optical density and cell measurement

4.6.3.1. Mammalian cells

Mammalian cells were stained with a Trypan-blue solution (0.4 %) to distinguish between live and dead cells (5 µL of cells mixed with 45 µL of the staining solution). This stain can penetrate only dead cells, and thus dead cells occur in deep blue. The number of cells per mL was estimated using a Neubauer chamber, counting all four of the larger squares.

4.6.3.2. Bacteria

The concentration of bacteria liquid cultures (cfu, colony forming units) was estimated by determining the optical density at 600 nm (OD_{600}). Bacterial suspensions have been measured against a suitable medium blank in the Ultraspec photometer. For a rough estimation of *Y. enterocolitica* cultures, the following formula can be used: $OD_{600} = 0.36 \approx 5 \times 10^7$ cfu per 15 µL of the culture (oral communication, Hicham Bouabe).

4.7. Competent cells

The ability of cells to incorporate foreign, naked DNA molecules is termed competence. Bacterial cells have been made electro-competent prior to electroporation. A suitable amount of medium was freshly inoculated with bacteria from an over-night culture (amount depending on the optical density, normally cells were diluted 1:50 or 1:100). The cultures were incubated with vigorous shaking either at 27°C (*Yersinia*) or 37°C (*E. coli*) until they reached an OD_{600} of about 0.6. The following steps were carried out under sterile conditions. Cultures were placed on ice, transferred to falcon tubes and centrifuged at 4,500 g for 15 minutes at 4°C. The supernatant was removed and the bacterial pellet was resuspended in sterile, ice-cold H_2O. Centrifugation was repeated under the same conditions as described. This washing step was performed twice. A final washing of the bacterial pellet was performed using sterile, ice-cold H_2O with 10 % glycerol. The supernatant was decanted and subsequently, cells were splitted into 50 µL aliquots that were directly frozen at -80°C.

4.8. Transformation by electroporation

4.8.1. Transformation by electroporation

Electroporation was achieved using Gene Pulser II. Short electrical impulses generate pores in the cell membrane that facilitate the incorporation of foreign DNA into the cell (Dower *et al.* 1988). Parameters were set to 25 µF capacitance at 2.5 kV and 200 ohms. Electro-competent cells were thawed on ice, and purified DNA or plasmid DNA was added and incubated on ice for 30 minutes. Cells were transferred to a suitable electroporation cuvette and transformed. After electroporation, transformed cells were mixed with prewarmed 1 mL SOC medium and incubated at either 27°C or 37°C with shaking for about 60 min. Bacterial cells were then plated as 25-300 µl aliquots (dependent on competence and the kind of DNA) on LB-agar plates containing the required antibiotics for selection of positive transformation.

4.8.2. X-gal/IPTG LB-agar plates for blue-white screening of recombinants

In case of vector-based gene cloning that allowed blue-white screening of recombinants for the detection of successful ligations (for example, pGEM-T Easy cloning kit), agar plates containing X-gal and IPTG have been prepared to induce LacZ activity. For one agar plate, 35 µl of a 50 mg/mL X-gal solution and 20 µl of a 100 mM IPTG solution were added to 30 mL of LB-agar with appropriate antibiotics and poured into plastic Petri dishes. Bacterial cells were then plated as 25-300 µl aliquots on the LB agar X-gal/IPTG plates. After incubation over-night at 37°C, the agar plates were shifted to 4°C, which enabled a better selection of white colonies.

4.9. Conjugation

The horizontal transfer of genetic material between donor and recipient bacterial cells through cell-to-cell contact is termed conjugation. For conjugation between selected donor and recipient strains, 1 mL of a well grown overnight culture of the recipient strain was mixed with each 1 mL of an early logarithmic phase culture of one or more donor strains. Cells were washed twice in 0.9 % NaCl solution to remove traces of antibiotics. The strain mixture was spilled on blood agar plates, incubated at 37°C for either 2 hours or over-night (normally both conditions were carried out in parallel), washed down from the plate using 0.9 % NaCl, diluted serially and plated on selective agar plates.

4.10. Yop secretion

For the isolation of Yops (*Yersinia* outer proteins), cells were grown over-night in BHI medium shaking at 27°C. Fresh BHI medium was then inoculated 1:40 and incubated for 2 hours shaking at 37°C. The secretion of Yops was stimulated by the addition of 5 mM EGTA, 0.2 % glucose and 10 mM $MgCl_2$ to the culture and proceeding growth for up to 3 hours at 37°C. The culture was centrifuged at 4°C for 15 minutes with 4,000 rpm (centrifuge Sigma 1K15) and the supernatant was collected. The pellet was dissolved in sample buffer for SDS gels (Table 4.11). For the precipitation of the proteins, 1,800 µL of the supernatant was mixed with 10 % trichloroacetic acid (TCA) over-night at 4°C. Proteins were pelleted at 4°C for 30 minutes at 10,000 rpm (centrifuge Sigma 1K15), incubated with acetone on ice for 20 minutes and then washed twice with ice-cold acetone with centrifugation steps at 10,000 rpm for 10 minutes. Protein pellets were dried and dissolved in sample buffer for SDS gels.

Table 4.11 SDS sample buffer

1 x SDS sample buffer	
1M $MgCl_2$	5 mL
10 % SDS	20 mL
Glycerol	5 mL
ß-Mercaptoethanol	2.5 mL
Bromophenol blue	25 mg
1 M Tris, pH 8,8	4.9 mL
H_2O ad 50 mL	

4.11. Amino-sugar metabolism experiments

Experiments with different amino-sugars to determine the functionality of specific phosphotransferase systems (PTS) have been carried out using sterile falcon tubes to prevent contamination. Over-night cultures of *Yersinia* grown in BHI were grown freshly diluted in

normal LB until the cultures reached an OD_{600} of 0.6. Cells were harvested by centrifugation at 4,500 g and washed in 0.9 % NaCl solution. Washed cells were again harvested by centrifugation (4,500 g) and the obtained pellet was resuspended in glucose free M9 minimal medium with 200 µg/mL tryptophan as an additional nutrient. Cells were splitted equally into filter sterilised M9 medium with the respective amino-sugars obtained from Sigma-Aldrich (0.2 %) and 200 µg/mL tryptophan. Cultures were measured at OD_{600} and adjusted to an OD_{600} of about 0.3. The growth curve was monitored by OD_{600} determination and finally plated on LB and blood agar plates to check the cultures for contaminations. For the *aga*-complemented serobiotype O:8/1B strain, carbenicillin was added.

4.12. Phage excision experiments

Phage excision was checked by either direct or nested PCR with primers that were set in the phage chromosome and directed against the attachment sites, thus giving only positive PCR results for the circularised, excised phage. Likewise, primers set in the *Yersinia* chromosome, directed against the insertion site, gave only positive results in the absence of the phage.

4.13. Overlay assays

Bacteriocin production and activity tests were performed with soft agar (0.8 % LB agar). Plastic Petri dishes were poured with normal 1.5 % LB agar (10 mL), air dried under the laminar flow bench. About 50-100 µL of bacterial cultures grown over-night in normal LB medium were used to inoculate 3-5 mL of soft agar, which was poured onto the thin LB agar plates after vigorous mixing. After 15 minutes, the dried soft-agar plates containing the tester strains were streaked with the bacteriocin harbouring strain Y11. The agar plates were incubated over-night (at 27°C or 37°C, and at 4°C for several days). Agar plates were checked for bacteriocin sensitive strains by visual inspection. Tester strains in soft agar that are sensitive to bacteriocin proteins of strain Y11 should be unable to grow in the direct neighbourhood of the Y11 lawn and thus present a halo around the colonies of strain Y11.

4.14. Agglutination

Agglutination was carried out to determine or check the serogroup status of different *Yersinia* or to check the presence of *Yersinia* in a culture. A needle tip small amount of bacteria harvested from an agar plate was dissolved in about 2-5 µL of the respective antiserum. Agglutination was positive, if the cloudy suspension turned into a crumbled solution. To avoid false positive results, cells were also suspended in 0.9 % NaCl to check the culture for autoagglutination.

4.15. Cell assays

Mammalian cells were washed, harvested and quantified as described under 4.6. The cells were seeded in 6-well Petri dishes appropriate for cell culture. To obtain about 5 x 10^6 cells per well, 2.5 x 10^6 cells per well were seeded one day before the experiment. The *Yersinia* culture was pregrown after the dilution of an over-night culture to an OD_{600} of 0.6 and used in appropriate amounts to get a similar multiplicity of infection (MOI) for all experiment samples. The bacteria were pipetted into the 6-well plates with seeded cells. Bacteria were then centrifuged onto the cells at 1.000 rpm for 2 minutes (Heraeus Megafuge 3.0R) to enable direct contact. After centrifugation, cells were incubated at 37°C for two hours (at 5 % CO^2) and harvested for mRNA isolation.

4.16. Molecular biology and protein methods

4.16.1. DNA isolation

Genomic DNA was isolated using either the NucleoSpin tissue kit or the NucleoBond AXG column kit of Macherey-Nagel, corresponding to the manufacturer's protocols.

4.16.2. RNA isolation, quantification and Reverse Transcription

Whole cell mRNA was purified using the RNeasy Mini kit from Qiagen (protocols 4 and 7 of the manufacturer's protocol handbook) and the RNAprotect reagent (according to the manufacturer's protocol for bacterial cells). In addition to the standard protocol, DNAse digestion was performed directly on the purification columns as described in the appendix of the manufacturer's protocol, but prolonged to at least 30 minutes of incubation. The mRNA concentration was determined as described for DNA (see 4.16.4). The purified mRNA was checked for DNA contamination by PCR, using the same primer pairs as for the detection experiments. DNA free mRNA was transcribed into cDNA using the MonsterScript™ 1st strand cDNA synthesis kit from Epicentre. The obtained cDNA was then used for PCR under standard conditions to address the presence of different mRNA transcripts under specific conditions.

4.16.3. Plasmid DNA isolation

4.16.3.1. Plasmid purification in small and large scale

Plasmid DNA was isolated using either the NucleoSpin Plasmid quick pure kit or the NucleoBond PC100 column kit of Macherey-Nagel, corresponding to the manufacturer's protocols.

4.16.3.2. Kado plasmid isolation

This protocol, modified from the protocol of Kado and Liu (Kado and Liu 1981) for the preparation of large plasmids has been used to check *Yersinia* cultures for the presence of the large pYV plasmid. A larger needle tip amount of bacteria was rubbed into 100 µL of the Kado solution (3 % SDS, 50 mM Tris at pH 12.6, 80 mg/mL NaOH) and heated 30 minutes to 1 hour at 60°C. This step was followed by phenol extraction, in which twice the volume of phenol-chloroform was added, mixed and spinned down at maximal speed. The upper phase was collected and mixed with loading buffer, then pipetted on 0.8 % agarose gels.

4.16.4. DNA quantification

Nucleic acids have their maximum light absorption at 260 nm wavelength. The isolated DNA was diluted with distilled water (1:50) and the absorbance at 260 nm (A_{260}) against H_2O was measured spectro-photometrically in a quartz cuvette (light path length 1 cm). According to the Beer Lambert Law (A = e c l; A: absorbance, e: extinction coefficient, c, concentration, l: path length) the concentration can be calculated from the absorbance value ($c = A\ (e\ l)^{-1}$). At the wavelength of 260 nm, the extinction coefficient for double-stranded DNA is 0.020 $(\mu g/ml)^{-1}\ cm^{-1}$, and for single-stranded RNA 0.027 $(\mu g/ml)^{-1}\ cm^{-1}$. For the determination of DNA purity, the A_{260}/A_{280} ratio was determined. A ratio of $A_{260}/A_{280} < 1.8$ indicates possible contamination of the DNA with protein or remnants of aromatic substances such as phenol, whilst a ratio of $A_{260}/A_{280} > 2.0$ indicates possible contamination with RNA.

4.16.5. Horizontal DNA agarose gel electrophoresis

Agarose gels were prepared by mixing an appropriate amount of agarose (to a final concentration of 0.7-2 % depending on the molecular weight of the sample DNA) with 1 x TAE buffer, boiling the solution until the agarose is melted and pouring the liquid into agarose gel chambers. The DNA sample was then mixed with loading buffer, loaded onto the gel and electrophoretically separated at 80-100 V with 1 x TAE solution as the running buffer for 30-90 minutes, depending on the DNA size expected (Table 4.12).

To estimate fragment sizes separated in the gel, a DNA marker with different fragment sizes was applied (see Figure 4.1). Following the electrophoretic run, gels were stained in ethidium bromide solution (1 µg ethidium bromide/mL H_2O) and the DNA fragments were visualized under ultraviolet radiation of the GelDoc EQ device.

Table 4.12 Gel electrophoresis buffers

Buffer	Component
1x TAE buffer	40 mM Tris/HCl, pH 8.2
	20 mM Acetic acid
	2 mM EDTA, pH 7.6
10 x Loading buffer for agarose gels	0.25 % (w/v) Bromophenol blue
	10 % (v/v) Glycerol

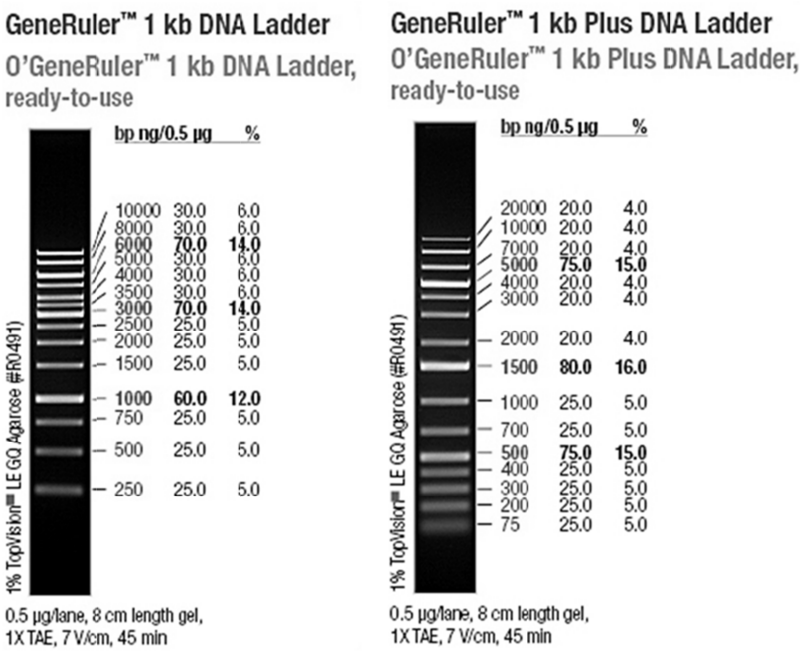


Figure 4.1 GeneRuler 1kb DNA ladder and DNA ladder plus. The column next to the picture of the DNA fragments used as length markers show the respective fragment size in base pairs (bp). The second column refers to the amount of the specific DNA fragments applied onto the gel lane (in ng per 0.5 µg marker per lane), which enables a rough estimation of the amount of an applied DNA fragment solution, for example a PCR product. The last column refers to the percentage of the respective DNA fragment amount in the applied DNA ladder solution.

4.16.6. PCR and nested PCR

The polymerase chain reaction (PCR) is used to amplify specific DNA targets *in vitro* using DNA polymerase enzymes, nucleoside triphosphate molecules (dNTPs) and target specific primers. Three steps are involved in one cycle of a PCR reaction. Denaturation achieves the dissociation of the double stranded DNA molecules into single strands; annealing allows the single strand primers to bind to the complementary sites on the specific DNA target and elongation enables the extension of the amplified DNA strands. Amplified products were visualised on agarose gels.

Nested polymerase chain reaction is a modification of the standard polymerase chain reaction that involves two sets of primers and PCR runs. Nested PCR can be used to reduce the possibility of contamination due to an amplification of products from unexpected primer binding sites or to enhance traces of template DNA prior to the second amplification. Nested PCRs have been performed using products of the first PCR reaction sample for the reaction of the second PCR.

4.16.6.1. GoTaq and VWR mastermix

The GoTaq enzyme is a modified form of the standard *Taq* polymerase which lacks 5'→3' exonuclease activity. *Taq* DNA polymerases generate poly-A overlaps that can be used for direct cloning into vectors with poly-T overlaps. The standard amount of the components of a 50 µL reaction is listed below.

Table 4.13 Reaction components GoTaq

Component	Volume	Final Concentration
GoTaq reaction buffer	10 µL	1.5 mM $MgCl_2$
dNTPs	1 µL	0.2 mM each dNTP
Upstream primer	1 µL	0.5 µM
Downstream primer	1 µL	0.5 µM
GoTaq polymerase	0.25 µL	1.25 units
Template DNA	1 µL	0.1-100 ng
	H_2O ad 50 µL	

The VWR Red Taq DNA polymerase master mix is a ready-to-use mix containing the enzyme and dNTPs, as well as a dye with a stabilizer to enable direct load on agarose gels.

Table 4.14 Reaction components VWR master mix

Component	Volume	Final concentration
VWR Red master mix	25 µL	1.25 units, 1.5 mM $MgCl_2$, 0.2 mM dNTPs
Upstream primer	1 µL	0.5 µM
Downstream primer	1 µL	0.5 µM
Template DNA	1 µL	0.1-100 ng
	H_2O ad 50 µL	

For both GoTaq and VWR Red master mix the following protocol was used:

Table 4.15 Cycling conditions

Step	Temperature	Time	Number of cycles
Initial denaturation	95°C	2 minutes	1
Denaturation	95°C	30 seconds	
Annealing	55-60°C	30 seconds	30 - 35
Extension	72°C	1 minute/kb DNA	
Final extension	72°C	5 minutes	1
Cooling	4°C	indefinite	1

4.16.6.2. Phusion and Phusion Flash

The Phusion enzymes are based on novel *Pyrococcus*-like enzymes and offer 5'→3' exonuclease activity. It is especially suitable for long PCR products that need proofreading activity. Phusion Flash mastermix is based on the modified Phusion enzyme and allows very short cycle protocols (15 seconds per kb of DNA). Both standard reaction components and the recommended reaction protocol are listed below.

Protocols for Phusion DNA polymerase:

Table 4.16 Reaction components Phusion

Component	Volume	Final Concentration
Phusion buffer HF	10 µL	1.5 mM $MgCl_2$
dNTPs	1 µL	0.2 mM each dNTP
Upstream primer	1 µL	0.5 µM
Downstream primer	1 µL	0.5 µM
Phusion polymerase	0.5 µL	1 unit
Template DNA	1 µL	0.1-100 ng
	H_2O ad 50 µL	

Table 4.17 Cycling conditions

Step	Temperature	Time	Number of cycles
Initial denaturation	98°C	30 seconds	1
Denaturation	98°C	10 seconds	
Annealing	55-60°C	30 seconds	30 - 35
Extension	72°C	15 - 30 seconds/kb DNA	
Final extension	72°C	5 minutes	1
Cooling	4°C	indefinite	1

Protocols for Phusion Flash DNA polymerase master mix:

Table 4.18 Reaction components Phusion Flash

Component	Volume	Final concentration
Phusion Flash master mix	25 µL	not declared
Upstream primer	1 µL	0.5 µM
Downstream primer	1 µL	0.5 µM
Template DNA	1 µL	0.1-100 ng
	H_2O ad 50 µL	

Table 4.19 Cycling conditions Phusion Flash

Step	Temperature	Time	Number of cycles
Initial denaturation	98°C	10 seconds	1
Denaturation	98°C	1 second	
Annealing	55-60°C	5 seconds	30
Extension	72°C	15 seconds/kb DNA	
Final extension	72°C	1 minute	1
Cooling	4°C	indefinite	1

4.16.6.3. SNP PCRs

PCRs that were made to detect single nucleotide polymorphisms (SNPs) have been exclusively performed using DNA polymerase enzymes with 5'→3' exonuclease activity (proofreading). PCR products have been purified and sequenced in both directions. Sequences have been clipped at both ends to remove low quality bases at the edge.

4.16.7. PCR purification

PCR reaction products that were used for manipulation or sequencing have been purified directly or after electrophoresis from agarose gel slices using the PCR extract kit from Macherey-Nagel, following the manufacturer's protocols.

4.16.8. Small scale DNA sequencing

Sequencing reactions have been carried out by AGOWA (LGC Genomics, Berlin), GATC (Konstanz) and Eurofins MWG Operon (Ebersberg). PCR products have been purified and the quantity has been measured spectro-photometrically in a quartz cuvette. Recommended amounts of DNA were either premixed with primer dilutions or send separately. PCR products were sequenced in both directions to minimise sequencing errors.

4.16.9. Large scale DNA sequencing

Large scale sequencing (complete genome sequencing) has been performed using either MegaBACE, 454 or Illumina technologies. MegaBACE has been performed at Integrated Genomics (Jena), 454 based methods were performed at 454 (Branford CT, USA) and Seq-IT (Kaiserslautern) and Illumina was performed with BGI-Hongkong Co. (Hong Kong). The sequencing run outputs are announced in the result sections (5.1 and 5.2).

4.16.10. Sequencing analysis

Sequences have been clipped at both ends to remove low quality bases at the edge and chromatogram data or quality scores have been checked for low quality base callings. In case of ambiguous results, sequencing reactions were repeated. Clipped accurate sequence results have been further analysed using alignments or BLAST searches.

4.16.11. DNA digestion

Digestion of DNA templates (either plasmid DNA or PCR products) has been performed in accordance with the manufacturer's protocols. In general, one unit of enzyme can normally digest up to 1 µg of DNA in one hour. If available, FastDigest enzymes have been used that reduce incubation times. For double digestions, a suitable buffer for both enzymes was selected as recommended by the company, or, in case of FastDigest enzymes, a standard buffer suitable for all enzymes has been used.

4.16.12. DNA modifications

4.16.12.1. Uladzimirase: poly-A overlap generation

In case of PCR products generated by DNA polymerases that generate blunt end DNA fragments, a simple protocol for the generation of poly-A overlaps has been applied: One volume of the purified blunt end PCR product was mixed with one volume of VWR Red polymerase mastermix (generating poly-A nucleotide overlaps) and incubated at 72°C for 10 minutes. The reaction was purified again and used directly for ligation.

4.16.13. Ligation

For all ligations, the T4 DNA Ligase from New England Biolabs was used. In general, a vector/insert ratio of 1/5 was utilized. Ligations were carried out at room temperature for 20 minutes, followed by heat inactivation at 65°C for 10 minutes.

4.16.14. Fosmid library

For the preparation of a fosmid library for *Y. enterocolitica* strain Y11, a pCC2FOS vector system with copy control was used, according to the manufacturer's protocols. This system combines clone stability by single-copy cloning with the advantage of high yield of DNA by selective induction of the clones to high-copy number. Fosmid clones can thus be induced from single-copy to up to 200 copies per cell to improve DNA yields for sequencing and other downstream applications. The pCC2FOS vector contains both a single-copy origin and the high-copy *oriV* origin of replication. Initiation of replication from *oriV* requires the *trfA* gene product that is supplied by the EPI300™-T1^R *E. coli* strain. This *E. coli* strain provides a mutant *trfA* gene whose gene product is required for initiation of replication from *oriV*. The gene is under tight control of an inducible promoter. Features of the pCC2FOS vector supplied by the manufacturer are listed below.

Features of pCC2FOS:

Chloramphenicol resistance marker.

E. coli F factor-based partitioning and single-copy origin of replication.

oriV high-copy origin of replication.

Bacteriophage lambda cos site for lambda packaging or lambda-terminase cleavage

Bacteriophage P1 loxP site for Cre-recombinase cleavage.

Bacteriophage T7 RNA polymerase promoter flanking the cloning site.

A picture illustrating the fosmid bank construction taken from the manual of the manufacturer's protocol is shown in Figure 4.2 below.

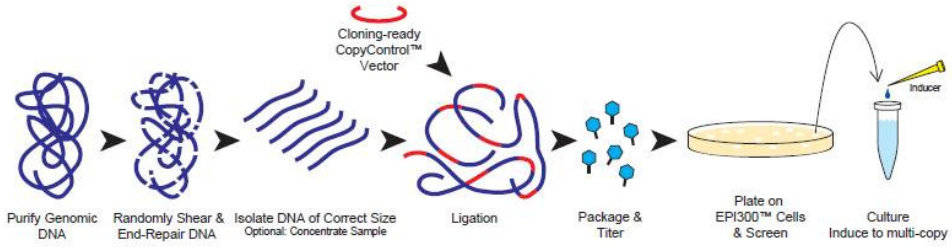


Figure 4.2 Overview of the fosmid library construction steps. Picture copied from the manual of the pCC2FOS fosmid kit from Epicentre.

4.16.14.1. Library construction

To prepare the DNA for the fosmid library, 12 mL of an over-night culture of *Y. enterocolitica* strain Y11 were used for three columns of the NucleoBond AXG kit. To obtain only DNA of best quality (high molecular weight, unfragmented), only the visible precipitate of the isopropanol-precipitation step has been isolated and dissolved in 1 mL of H_2O. DNA has been adjusted to

about 500 ng/µL and 1-10 µL were analysed on a large size 1 % agarose gel at 30 V over-night. As a marker, 1 µL of the fosmid control DNA (40-42 kb) was applied. DNA was then sheared pipetting up and down until a homogenous fragment length of about 42 kb was reached. The DNA was subsequently used for end-repair to blunt 5′-phosphorylated ends, as recommended from the manufacturer. The DNA was again analysed on an agarose gel as described above and used without size range selection for ligation into the CopyControl pCC2FOS vector, following the standard protocol. The ligation reaction was packaged in phage particles (MaxPlax Lambda Packaging Extracts) according to the instructions of the manufacturers and plated in an appropriate dilution with EPI300-T1^{R} plating cells on LB agar with 12.5 µg/mL chloramphenicol. 960 colonies were picked individually and grown in a sterile, deep-well 96-well plate (550 µL LB medium with antibiotics per well) with gentle shaking to avoid cross contamination. By calculation based on the manufacturer's protocol, a colony number of 500 clones should represent the complete genome of *Y. enterocolitica* strain Y11. The fosmid library was directly frozen at -80°C. All samples have been streaked on fresh agar plates with appropriate amounts of antibiotics from the frozen stocks before screening.

4.16.14.2. Library screening and induction

Cells have been screened for the insert of interest in batches of 10 samples to reduce expenses. For each batch, 50 µL of water was inoculated with needle-tip small amounts of 10 bacterial samples, boiled for 5 minutes and spinned down at maximum speed. About 2 µL of the supernatant have been used as the template DNA for the screening PCR. Batches positive in PCR have been screened individually for each strain, and positive strains were induced to high copy plasmid numbers with the autoinduction solution for exactly 16 hours before isolating the respective plasmids.

4.16.15. SDS PAGE and Western Blot

4.16.15.1. SDS PAGE

SDS polyacrylamide gel electrophoresis (SDS PAGE) allows the separation of proteins through migration through a sodium-dodecyl-sulphate (SDS) polyacrylamide gel on the basis of their molecular weights. SDS is an anionic detergent that denatures proteins. The SDS also effects the disruption of hydrogen bonds and blocks hydrophobic interactions. In general, also the tertiary and secondary structures are unfolded after boiling, except heat stable proteins. The SDS gel is based on two layers with different buffer contents, generating best resolution during electrophoresis: a non-restrictive large-pore gel named stacking gel (low acrylamide content, pH 6.8) is layered on top of a high percentage resolving gel (high acrylamide concentration, pH 8.8).

The proteins migrate first through the stacking gel as a compact front (due to the field voltage reduction at the border to the resolving gel) and then enter the separating gel with different speed based on molecular weights (Laemmli 1970). With a protein marker applied in neighboured lanes of the gel, the molecular weight of the proteins of interest can be estimated. The following is the pipetting scheme applied for the preparation of two 12 % acryl amide SDS-gels (0.75 mm, 8 x 7 cm) and the components of the buffers used:

Table 4.20 SDS gel components, resolving gel

Resolving Gel	
Solution component	Volume
H_2O	3.3 mL
1.5 M Tris pH 8.8	2.5 mL
10 % SDS	0.1 mL
30 % acryl amide mix	4.0 mL
10 % ammonium persulfate	0.1 mL
TEMED	4 µL

Table 4.21 SDS gel components, stacking gel

Stacking gel	
Solution component	Volume
H_2O	1.4 mL
1.0 M Tris pH 6.8	0.25 mL
10 % SDS	20 µL
30 % acryl amide mix	0.33 mL
10 % ammonium persulfate	20 µL
TEMED	2 µL

Table 4.22 Buffers and solutions for SDS gel electrophoresis

10x SDS Electrophoresis running buffer	
Tris	30.2 g
Glycine	142.6 g
H_2O ad 1 L	

Table 4.23 Buffers and solutions for SDS gel electrophoresis

4x SDS-loading buffer (pH 6.8)	
Tris	0.4 g
SDS	1.2 g
Glycerol	7.5 mL
ß-Mercaptoethanol	2.5 mL
Bromophenol blue (2 % solution)	0.5 mL
H_2O ad 50 mL	

Table 4.24 Buffers and solutions for SDS gel electrophoresis

Coomassie dye solution	
Coomassie Brilliant blue (Serva)	0.15 %
Methanol	500 mL
Acetic acid	100 mL
H_2O ad 1 L	

Table 4.25 Buffers and solutions for SDS gel electrophoresis

Destaining solution	
Methanol	500 mL
Acetic acid	75 mL
H_2O ad 1 L	

The gels were run in Mini-PROTEAN 3 chambers (Bio-Rad, Munich) in 1 x SDS electrophoresis buffer at 80 V for 20 minutes followed by 180 V for 45 minutes, until the front end of the staining solution left the gel.

4.16.15.2. Western Blot

The separated proteins on the SDS gel were blotted onto nitrocellulose membranes in the Mini Trans-Blot chambers of Bio-Rad (Munich) at 200 mA for 1 h in 1 x blotting buffer. Specific protein bands were then visualised by antibody interactions coupled to chemiluminescence: The membrane was blocked for unspecific reactions using 5 % milk powder in washing solution (0.5 % Tween in 1xPBS) either shaking for 1 hour at room temperature or over-night at 4°C. The first, protein specific antibody was diluted in 5 % milk powder in washing solution and applied on the membrane with shaking for 1 hour at room temperature, followed by serial washing steps, 3 times for 10 minutes with washing solution. The secondary antibody (coupled with peroxidase) directed against the first, protein specific one was applied for shaking for 45 minutes at room temperature. After performing washing steps as described previously, the luminescence reaction

of the peroxidase was activated using the ECL Western Blotting System (GE Healthcare, Munich) following the manufacturer's protocol and directly developed on x-ray films (Fuji-film).

Table 4.26 Western Blot buffer

10 x Blotting-Puffer	
Tris	30.3 g
Glycine	144.1 g
H_2O ad 1 L	

4.16.16. In vitro translation

The *Y. enterocolitica* strain Y11 specific bacteriocin gene was labelled with a C-terminal His_6-tag and translated *in vitro* to circumvent toxic side effects of the protein in bacterial over expression systems. The template for translation was generated using the LinTempGenSet of 5Prime (Hamburg), following the manufacturer's protocol. A first PCR was generated with bacteriocin specific primers (JB516/JB517) that were coupled to specific sequences for the generation of the desired His_6-tag. The second PCR (using primers delivered by the manufacturer) generated a linear fragment of the bacteriocin gene, the regulatory elements and the His_6-tag. The purified template was directly used for the *in vitro* translation or frozen at -20°C. About 100 ng template per 50 µL reaction was used and 1 µg of the GFP control vector. The reaction solution was pipetted following the manufacturer's instructions and the translation reaction was incubated for 6 hours at exactly 30°C. The successful translation was controlled using Western Blot detection of the His-tagged protein.

4.16.17. His tag purification

His-tagged proteins were purified using Ni-NTA magnetic beads of 5Prime (Hamburg), following the manufacturer's protocol.

4.16.18. Protein quantification

Protein quantification was performed using the Bradford based Bio-Rad Protein Assay Kit attending the manufacturer's protocols. BSA (Bio-Rad, Munich) was used for the calibration curve. Measurements were performed at 595 nm of the spectrophotometer.

4.17. Bioinformatics

4.17.1. BLAST

The Basic Local Alignment Search Tool (BLAST) identifies regions of local similarity between two or more sequences. It compares nucleotide (BLASTn) or protein sequences (BLASTp) to

sequence databases and calculates a statistical significance of matches, denoted in the output by a score and E-value. This E-value, or expectation value, describes how many hits one can "expect" statistically by chance when searching a database of a specific size. It decreases exponentially as the score of the match increases. An E value of 1 denotes that in a database of the current size 1 match with a similar score can be expected as an output by chance. The lower the E-value is (the closer this value is to zero), the more significant the match can be ranked. It is important to mention that the E-value takes into account the length of a query sequence to reduce false positive hits due to a higher statistical occurrence of hits to the database of shorter sequences by chance.

4.17.1.1. BLAST at NCBI

4.17.1.1.1. BLASTn

BLASTn was used for standard nucleotide sequences that were compared against the nucleotide collection database. The parameters were optimised for megablast (searching for highly similar sequences). Mismatch penalties were kept as initialised. Parameters for short input sequences were automatically adjusted. In case of no specific hits, the parameters were set to find somewhat similar sequences (BLASTn parameters).

4.17.1.1.2. BLASTp

BLASTp was used for standard protein-protein BLAST searches against the nucleotide collection database. Mismatch penalties and matrices were kept as initialised. If needed, searches were adjusted to either PSI-BLAST (this parameter setting allows to build a position-specific scoring matrix using the results of the first BLASTp run) or to PHI-BLAST (this setting performs a standard search but limits the output alignments to those that match a pattern in the query).

4.17.1.1.3. Other BLAST applications

Blast does not only allow searches of nucleotides or proteins against the respective databases, but also a combination of these. Thus, one can search the protein database using a translated nucleotide query sequence (BLASTx) or the translated nucleotide database using a protein sequence as query input (tBLASTn). In necessary, also a translated nucleotide query can be used to fish out matches of the translated nucleotide database (tBLASTx). In addition, BLAST offers several trailed tools that allow the alignment of two sequences (bl2seq), screens for vector sequences, SNPs or other sequence characteristics. These tools are found as specialised BLAST applications.

4.17.1.2. BLAST standalone version

BLAST offers also the possibility to be run against a defined database of interest as a standalone version. For that purpose, the BLAST tools have to be downloaded (use the binaries for the respective platform - in this case for the Linux distribution Fedora) and unpacked. The program that converts FASTA formatted files (for example, contig sequence data) to the binary files used by BLAST tools is *formatdb*. To create a new database, the command "formatdb" was used, following specific arguments as described below:

formatdb -i Y11.fasta -p F -o T

Arguments are: -i, specifies the input file(s) to be formatted (Y11.fasta, for example), -p, specifies the input type (T/F): T: true, input is protein or F, false, input is nucleotide. Argument -o parses deflines and indexes sequence IDs. FASTA sequence files with NCBI styled deflines need -o T arguments, otherwise one uses -o F (no indexes will be created).

For standard BLAST searches against the new formatted database, the command "blastall" was used, followed by the arguments described below:

blastall -p BLASTn -d Y11 -i QUERY -o out.QUERY

The argument -p defines the program used (for example, BLASTn or BLASTp), -d defines the database made by the "formatdb" command (for example, Y11 contig sequences), -i defines the file of the query sequence (in fasta format) and -o defines the file name of the BLAST output data. Further optional arguments are -e, the expectation value, and other arguments for search filters and output styles.

Using the standalone BLAST on Linux, it was possible to search for specific genes and sequences in the large database of contig sequence data from the early stage sequencing results with maximised speed and significance.

4.17.2. DNAstar Lasergene

Lasergene is a DNA and protein sequence analysis software suite, including functions ranging from sequence assembly and SNP detection, to sequence editing and primer design. In this study, Version 7.2.0 was used.

4.17.2.1. EditSeq

This program was invented for importing and editing unusual file types, but has been used for all kinds of editing and sequence conversions (translation, reverse complement sequences etc.). The program allowed the copy and paste application from non-Lasergene formats to save sequences in the recommended .seq format, enabled position specific changes and searches within the sequence

(through commands as “go to position”, “find sequence” and the respective sequence coordinates). The program allowed statistical analyses (GC content and others) as well as the editing of features. More details are found in the Lasergene EditSeq manual.

4.17.2.2. SeqMan

This program was used for contig assembly and analysis, including SNP discovery and coverage evaluation, as well as for single gene comparisons (alignment), sequencing analysis and gap closure projects. Data obtained from Roche 454 technology or Sanger data as well as unspecified sequence contigs could be loaded into the program for assembly. Parameters could be adjusted prior to the assembly process, including the assembly parameters (for example, minimal match percentage, minimal sequence length, match size and penalties). Sequences could be trimmed even after the assembly process to eliminate mismatches due to sequencing ends. A consensus could be exported with adjustable parameters to generate new contigs from the respective assemblies. This tool has been used to generate contigs from the last sequencing assembly before gap closure was performed.

4.17.3. Artemis

Artemis is a free, six-frame translation DNA sequence viewer and annotation tool, distributed under the terms of the GNU General Public License, which allows the visualisation and editing of DNA sequence features and the results of analyses as BLAST searches within the context of the respective sequence. Artemis is written in Java, reads EMBL or GENBANK format sequences and feature tables, and can work on sequences of any size (Carver *et al.* 2008).

This software has been used for sequence comparisons as described under 4.17.4, and for the annotation of locus tags for the publication of the *Y. enterocolitica* strain Y11 sequence at NCBI. Locus tags are unique prefixes for every strain in the database, placed ahead the gene numbers of the respective organism. Also, this tool was used to shift the start of the sequence to *dnaA*, the unofficial typical start of a chromosomal sequence of a bacterium in a database.

4.17.4. ACT and WebACT

The Artemis Comparison Tool ACT has been used for the pairwise comparison of two or more genome sequences, generating handsome figures illustrating rearrangements and inversions of shared sequences between the different genomes. To run an ACT comparison, a comparison file is needed between each of the entry files. These files can be generated by MSPcrunch or BLAST, and free online tools are available. If annotated sequences are loaded, the tool allows the selection and specific comparison of the annotated features. If only contig data are used as an entry, the tool

enables to view or save output files containing unaligned sequences of a particular minimal size for further analysis. In addition, as for the Artemis program, editing of the sequence annotation is also possible, as well as the prediction of open reading frames.

4.17.5. Mauve

Mauve (Darling *et al.* 2010) is a software suite that visualises the alignment of orthologous and xenologous regions among two or more genome sequences. It is a tool to compare the local and large scale changes between a set of related organisms, as for example the integrity of a region of phage genes or a region of high plasticity through the tree of descendants. Mauve allows different file formats as input files, and is able to visualise annotations within the comparison window. The presence of similar gene clusters is illustrated as coloured blocks that are spaced by white stripes in case of sequence differences or absence of smaller parts within a block.

4.17.6. Multalign

This free online software tool (http://multalin.toulouse.inra.fr/multalin/) allows the alignment of either DNA or protein data in fasta format with different parameters and output styles. If not mentioned else, the standard parameters have been applied (Corpet 1988).

4.17.7. Pedant

The pedant genome database (Frishman *et al.* 2003) gives access to a pre-calculated analysis of predicted protein sequences and thus provides analysis results of the respective genomic sequences by a variety of bioinformatics tools, including BLAST similarity searches (against the nr database), motif searches (against Pfam, BLOCKS and PROSITE) and predictions of cellular functions (based on high stringency BLAST searches). This multiple sequence analysis tool has been used for early sequencing stage genome versions. The database has been used for a preliminary but accurate genome annotation and thus allowed similarity searches as well as a detailed analysis of specific sequence regions.

4.17.8. RAST

RAST (Rapid Annotation using Subsystem Technology, (Aziz *et al.* 2008)) is a free online service for automatically annotation of bacterial and archaeal genomes. It provides automated rapid gene calling and functional annotation, both for complete genomes as well as draft genomes in fasta format. The RAST server normally makes the annotated genome available within 24 hours after submission to the server. During the upload, several parameters can be adjusted, including the fixing of frame shifts, backfill of gaps and error fixing. For the genome of *Y. enterocolitica* strain Y11, only the automatically fixing of errors has been applied. The

annotated genome can be modified with personal annotation changes or browsed using the SEED viewer technology described below.

4.17.9. SEED

The SEED viewer (Overbeek *et al.* 2005) can be used to compare the RAST annotated genomes to so far published genomes (including alignments of protein and nucleotide data, tabular output of whole genome protein comparisons) and to visualise annotated genes (as a table or as graphical output including the gene orientation, providing an interactive window for browsing). In case of protein comparisons, identity values are given in %. The SEED viewer comparison tools do not offer parameters for adjusted searches, but tabular output can be downloaded for further analyses.

4.17.10. DNAplotter

The DNAplotter generates individual images of circular and linear DNA maps (using fasta, EMBL, GenBank and other formats as input) to display selected regions and features of interest in different styles. Circles representing the calculated GC skew as well as RNA features can be displayed. This tool has been used to illustrate the location of specific genetic elements as prophages (Carver *et al.* 2009).

4.17.11. 454 data analysis

The 454 Newbler assembly software (Roche, Mannheim) was used not only for assembling sequence data generated by the two different 454 runs of pyrosequencing, but also for the quality control of the complete sets of raw and assembled data. All analyses with this software have been generated by Dr. Dirk Höper (FLI institute, Riems). Generated output files containing differences with statistical significance (HCDiff, HCStructVars) delivered by Dr. Höper have been checked manually and repeated after the correction of sequencing errors that have been checked by PCR and subsequent sequencing whenever needed. In addition, this software has been used to calculate statistical data concerning genome coverage and quality scores.

5. Results

Parts of the results have been published before (Batzilla *et al.* 2011a;Batzilla *et al.* 2011b;Batzilla *et al.* 2011c) and were taken over closely.

5.1. Sequencing of *Yersinia enterocolitica* subsp. *palearctica* strain Y11

Sequencing of the complete genome of the European serobiotype O:3/4 DSMZ reference strain Y11 (DSMZ number 13030, isolated from a patient stool) was performed in three individual genome sequencing runs that have been compiled in a final assembly.

5.1.1. MegaBACE sequencing run

MegaBACE Sequencing performed 2003 by Integrated Genomics (Jena) based on a shotgun clone library of *Y. enterocolitica* strain Y11 whole genomic DNA. The genomic DNA was sheared and cloned into the pCR® 4 Blunt TOPO® vector (Invitrogen). The company generated about 73,848 reads with 64,366,908 bases, assembled into 1,334 contigs. Due to a coassembly of these data with the fragmented genome of *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B, strain 8081, contigs were reduced to less than 400 contigs, but without any reduction of gaps. These data have been firstly used for further assemblies with the 454 GS-20 data (see below), but after the generation of the high-quality 454 FLX titanium dataset only for the chromosome replicon they were rejected due to the lower sequencing quality of this run. Only plasmid associated sequencing data used for the finishing of the pYV plasmid sequence have been included.

5.1.2. 454 GS-20 sequencing run

The 454 Genome Sequencer (GS) 20 pyrosequencing run (performed 2007 at 454, Branford CT, USA) yielded 927,998 reads with a total of 94,454,598 bp. The pyrosequencing method does not require the use of cloning vectors, thus reducing the amount of (vector-) contaminations and the disadvantages of cloning biases. Even so, the GS-20 system produces reads of about 100 bp (average read length in this sequencing run: 102 bp), which can cause problems in the assembly of repetitive sequence data, as IS elements. In addition, homopolymeric stretches can likewise result in misleading assemblies: The accuracy of pyrosequencing technology devices in base calling on homopolymers diminishes markedly after a few bases in a homopolymeric stretch.

5.1.3. 454 GS FLX Titanium sequencing run

In 2008, the 454 GS FLX Titanium sequencing technology was introduced, producing better average reading length of more than 350 bp. Performed 2009 at Seq-IT (Kaiserslautern), the 454

GS FLX Titanium run added 240,813 reads with 105,539,453 bp (average read length: 438 bp) of a derivate of strain Y11 cured of the pYV plasmid.

5.1.4. Assembly of the raw data

The two 454 derived pyrosequencing datasets were assembled with the DNAstar Lasergene software (SeqMan Pro, Version 7.2.0) to a draft genome of 105 contigs with a total length of 4,464,482 bp. based on 448 large contigs derived from both 454 pyrosequencing runs. Settings for this assembly were finally as follows:

Match Size	12
Minimum Match Percentage	80
Minimum Sequence Length	100
Maximum Added Gaps per kb in Contig	70
Maximum Added Gaps per kb in Sequence	70
Maximum Register Shift Difference	70
Lastgroup Considered	2
Gap Penalty	0.00
Gap Length Penalty	0.70

The new generated contigs have been used for manual sequence gap closure by PCR and sequencing of the respective amplicons.

5.1.5. Gap closure

The gap closure has been performed by manual PCR, primer walking and sequencing as well as in parts by an industrial gap closure project at AGOWA (LGC Genomics, Berlin). Gap closure results and sequencing biases are discussed in the sections below.

5.1.5.1. Gap closure biases

5.1.5.1.1. IS elements

The genome of *Yersinia enterocolitica* strain Y11 harbours many repetitive sequences, especially IS elements. The most often found IS element was ISYen1 (about 1,380 bp), found finally in 53 copies throughout the genome. Beside the large group of ISYen1 elements, seven copies of the *palearctica* group-specific IS element Yen2 were found. ISYen2 is related to IS elements of the IS21 family and is present in two isoforms in *Y. enterocolitica* strain Y11 (designated ISYen2A/B). The occurrence of these IS elements restrained the efficient closure of gaps, since PCR often resulted in false positive products due to mispriming of the primers.

5.1.5.1.2. P2-like phages

A second bias in gap closure process appeared in form of two nearly identical copies of a P2-like prophage, each of more than 29 kb in size. Due to the strong sequence similarity of more than 90 %, previous assembly performances resulted in the misassembly of all P2-like prophage sequences in the same region. To solve this problem, a fosmid library to obtain single clones harbouring only one of the prophages was initialised.

5.1.5.2. PCR based gap closure

Gap closure PCRs have been made based on genome comparisons with *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B, strain 8081, using manually generated predicted contig order and orientation maps. This PCR approach did succeed only in some of the predicted gaps, and failures were mostly due to unexpected gap length or mispriming of the primers. The cooperation with B. Fartmann at LGC Genomics, Berlin, reduced the gaps to only two unresolved issues. One of the gaps could be subsequently identified by PCR, generating a 4.7 kb product using Phusion Flash polymerase (at 60°C annealing temperature and one minute elongation time). The PCR product was modified with sticky poly-A overlaps and subsequently cloned in the pGEM-T Easy vector backbone (Promega). The gap spanning product was sequenced by primer walking.

5.1.5.3. Fosmid clone based gap closure

The last remaining gap in the genome of *Y. palearctica* strain Y11 was flanked by sequence regions with high similarity to the P2 prophage in the middle of the single genomic contig. Specific primers located in regions of difference in these duplicated regions proved the presence of two different P2 prophages, as illustrated in the alignment of Figure 5.1. Primer pairs JB444 and JB446 amplified the prophage region near the genome contig end (Seq.gap), whereas JB451 and JB452 were specifically designed for the region within the completely sequenced prophage in the middle of the genome (Seq.genome). Sequencing affirmed the differences within both prophage regions.

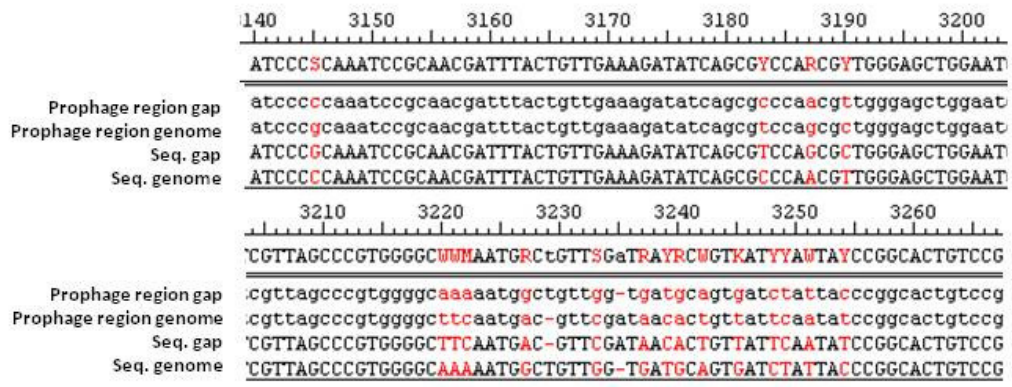


Figure 5.1 Two different P2-like prophages in the *Y. enterocolitica* strain Y11 genome. Regions of difference were proven by PCR product sequencing. Seq.gap, sequenced product of the remaining gap. Seq.genome, sequenced product of the P2-like prophage found in the middle of the single contig of the draft genome of strain Y11. Prophage region gap, the P2-like prophage found in the last remaining gap of the chromosome replicon. Prophage region genome, the P2-like prophage found in the middle of the single contigs of the draft genome. Differences are coloured in red.

Since most approaches utilising specific primers to amplify parts of the gap spanning prophage failed and a single long range PCR was impossible to cover the complete gap (of estimated 25-30 kb), a fosmid library was generated and screened to isolate a single clone covering the complete gap spanning the phage. Fosmid vectors can incorporate up to 40 kb of foreign DNA, which suited the need for at least 30 kb to cover the complete P2-like prophage. The fosmid bank comprises 960 clones frozen in individual wells. Batches of 10 clones were used for each PCR with the primer pair JB444 and JB446 for the gap phage fragment described above. A fosmid clone named G1 was positive in this gap-specific PCR and sequencing proved the fosmid insert to be a part of the gap spanning phage. Fosmid vector specific primers were used to sequence both edges of the insert and proved the complete coverage of the gap, resulting in sequences specific for both contig ends of the single contig draft genome. Subsequently, the fosmid clone G1 was sequenced by primer walking from both sides and the obtained sequence was transferred to the genome sequence to complete the chromosomal replicon.

5.1.6. Complete genome quality control

Finally, data were assembled into a complete genome sequence of 4,553,420 bp for the genome of *Y. enterocolitica* strain Y11 and an additional 72,460 bp contig representing the pYVeY11 virulence plasmid. For quality control, all raw data were mapped along the two contigs using the 454 reference mapper software (version 2.3, Roche, performed by Dr. Höper, FLI Institute,

Riems). Significant differences within the assembled raw data and the genome sequence were documented in the 454HCDiff output file. Likewise, structural differences between the finished genome and the raw data were reported in the 454HCStructVars output file. These reports have been analysed for every significant difference of the read data (divergence above 50 %). An exemplary illustration of these reports is given in Figure 5.2. In this example, the pYV plasmid sequence is used as the reference template for raw data mapping. In the first line, the name of the reference is given, followed by the position of difference and the nucleotide difference outcome itself and the percentage of reads of difference. Below, the original sequence of the reference is given. This sequence is superseded with four different reads in the next paragraph. Given under the heading “other reads”, the reference supporting reads are listed (reads that have the same start and end point are grouped as indicated in the output file with the number of reads in brackets). In this example, only 14 % of all reads mapped to this region are different from the reference sequence. For this low number of varying reads, no quality control PCR is needed. Normally, regions with differences above 50 % of all raw data mapped occurred clustered and have been checked for quality control by PCR in a single reaction. Primers have been designed at least 100 bp up- and downstream of the region of interest to exclude false base callings at sequencing edges. PCR products have been revised by sequencing in both directions and the sequence was edited whenever necessary.

```
>pYV_Y11_review05/2010  864   864   A     T     29    14%

Reads with Difference:
pYV_Y11_review0             837+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 879
                                                              **
DXOHH53O2FH6CT                1+             A----GCCAA-TTAAA-TT-GGCCAATTAA-------------CTTAAA 29
DXOHH53O1BOO9J               40-                   CCAA-TTAAA-TT-GGCCAATTAA-------------CTTAAA 14
DZ1PP4MO2FG7VL               39+                   CCAA-TTAAA-TT-GGCCAATTAA-------------CTTAAA 65
DZ1PP4MO1A3WBG               93-                   CCAA-TTAAA-TT-GGCCAATTAA-------------CTTAAA 67
                                                              **

Other Reads:
                                                              **
149-2-38-E04.r.             199- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 157
149-2-1-C12.r.1             385+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 427
149-2-38-E04.r.             174- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 132
149-2-198-C09.f             122+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 164
149-2-198-C09.f             116+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 158
149-2-71-E06.f.             144- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 102
149-2-71-E06.f.             205- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 163
DYROZ6WO2FWGFK               38+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 80
149-2-166-E09.r              29+ GCA-CGGAACC-AAAA--C-A-GT-A--CT-AG-CCA-T-AA-------------CTTAAA 63
DX5RM8GO2JGRS3              115- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 60
FL8F2BVO4YJ6ZV              195+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAA-------------TT    234
DZ1PP4MO1CKXF3               80- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAA-------------TT    41
DYAXGIWO1CDY6O               43+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 98
FL8F2BVO4YI5AH               92- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 37
DYROZ6WO2F77RT               99- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 44
FL8F2BVO4ZF55R              112- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 57
FL8F2BVO4X253J  (2)         142- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 87
149-2-68-C08.f. (2)         117+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 172
149-2-319-E09.f              29+ GCA-CGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 83
FL8F2BVO4Y203A  (2)         225- GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 170
FL8F2BVO4ZFJR2               85+ GCAACGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 140
149-2-319-E09.f              12+ GCA-CGGAACCCAAAAA-CCA-GTTAA-CT-AG-CCAATTAAATTGGCCAATTAACTTAAA 66
149-2-166-E09.r               1+       GGAACCCAAAAA-CCA-GTTA--CT-AG-CCAATTAA-------------CTTAAA 37
DZ1PP4MO1A2R7E  (2)           1+                        A-GTTAA-CT-AG-CCAATTAA-------------CTTAAA 24
149-2-166-E09.r               1+                           TTA--CT-AG-CCAATTAA-------------CTTAAA 21
                                                              **
```

Figure 5.2 454HCDiff output file screenshot. >pYV_Y11_review05/2010 is the name of the reference sequence, followed by the position and feature of the divergent nucleotide as well as the percentage of divergent reads (position 864, A/T discrepancy, 14 %). Reads with difference represent the divergent A/T substitution and are followed by the list of other reads, comprising the reference sequence confirming reads.

Using the quality checked sequences, the median sequence depth for the genome contig was reported as 37 (1st quartile 30, 3rd quartile 48), and the proportion of Q40+ bases (Q40+ base calls are a sign of high quality) was 99.58 %. For the plasmid sequence, a median depth of 12 (1st quartile 8, 3rd quartile 17) was achieved with the proportion of Q40+ bases being 98.01 %. These data have been generated in cooperation with Dr. Dirk Höper (FLI Institute, Riems).

5.1.7. Annotation

The annotation of the finished, quality checked genome of *Y. enterocolitica* strain Y11 was done using the rapid annotation using subsystem technology online tool, RAST (Aziz *et al.* 2008). This tool calls and annotates the genes of a complete or draft prokaryotic genome in a short run time. RAST returns also an analysis of the annotated genes and subsystems, as supported by comparative approaches and other forms of evidence. The annotation of the strain Y11 genome was downloaded from the RAST server as an EMBL file. A file with an EMBL file extension is

an IEMBL Nucleotide Format file that can contain several sequences. Each sequence entry generally begins with an identifier line (ID), followed by further information featured in annotation lines. Each line begins with a line code of two characters, which indicates the kind of information contained in the line. The currently used line types and line codes have been taken from the ENA guideline and are listed below:

ID - identification	(begins each entry; 1 per entry)
AC - accession number	(>=1 per entry)
PR - project identifier	(0 or 1 per entry)
DT - date	(2 per entry)
DE - description	(>=1 per entry)
KW - keyword	(>=1 per entry)
OS - organism species	(>=1 per entry)
OC - organism classification	(>=1 per entry)
OG - organelle	(0 or 1 per entry)
RN - reference number	(>=1 per entry)
RC - reference comment	(>=0 per entry)
RP - reference positions	(>=1 per entry)
RX - reference cross-reference	(>=0 per entry)
RG - reference group	(>=0 per entry)
RA - reference author(s)	(>=0 per entry)
RT - reference title	(>=1 per entry)
RL - reference location	(>=1 per entry)
DR - database cross-reference	(>=0 per entry)
CC - comments or notes	(>=0 per entry)
AH - assembly header	(0 or 1 per entry)
AS - assembly information	(0 or >=1 per entry)
FH - feature table header	(2 per entry)
FT - feature table data	(>=2 per entry)
XX - spacer line	(many per entry)
SQ - sequence header	(1 per entry)
CO - contig/construct line	(0 or >=1 per entry)
bb - (blanks) sequence data	(>=1 per entry)
// - termination line	(ends each entry; 1 per entry)

However, the annotated genome of strain Y11 had to be adjusted to recommended settings of the ENA institute. The European Nucleotide Archive, ENA, collects and presents information referring to nucleotide sequencing data, for example from genome sequencing projects. ENA collects this information, covering input information (about the sample, the experimental setup, devices and setups used), output data (for example sequence traces, raw data reads and quality

scores of a sequence) and subjectively processed information (sequence assembly, mapping of the reads and annotation). Therefore, different requirements were addressed:

Data concerning the sequencing project were added (organism and source information, sequencing technologies), the taxonomy was adjusted (RAST used a different taxonomy designation for strain Y11), translation exceptions were included in the annotation, the start point for the complete sequence was shifted to the *dnaA* gene as recommended by the ENA curators and Y11 specific locus tags were added. Locus tags are prefixes that are unique for one strain in the databases and occur as a label prior to the ascending gene numbers. For the genome of *Y. enterocolitica* strainY11, the term "Y11_" was chosen as the locus tag for the chromosome sequence and "Y11_p" for the plasmid sequence (for the sake of completeness: the locus tag of strain 8081 was previously designated "YE"). These tags were added using the Artemis software. Likewise, Artemis was also used to shift the start of the sequence to the *dnaA* gene. For one exception in translation found in strain Y11, the following line was added in the annotation of the formate dehydrogenase chain D:

FT /transl_except=(pos:3134720..3134722,aa:Sec)

Screenshots illustrating some of the changes of the EMBL format are shown in Figure 5.3 and Figure 5.4.

```
ID   unknown; SV 1; linear; unassigned DNA; STD; UNC; 4553420 BP.
XX
AC   unknown;
XX
DE   Contig Y11_anfang_wie_8081_23_11_10 from Yersinia enterocolitica Y11
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..4553420
FT                   /mol_type="genomic DNA"
FT                   /db_xref="taxon: 150053"
FT                   /genome_md5=""
FT                   /project="pesticin_150053"
FT                   /genome_id="150053.22"
FT                   /organism="Yersinia enterocolitica Y11"
FT   CDS             complement(270..710)
FT                   /db_xref="GO:0010181"
FT                   /db_xref="GO:0016491"
FT                   /translation="MADITLISGSTLGSAEYVAEHLAEKLEEAGFTTETLHGPELDELT
FT                   LDGMWLIITSTHGAGDLPDNLQPLLEQIEQQKPDLSQVRFGAVGLGSSEYDTFCGAVRK
FT                   LDQQLIIQGAMRVGDILEIDVIKHEIPEDPAEIWVKNWINLL"
FT                   /product="Flavoprotein MioC"
FT   CDS             complement(802..1263)
FT                   /translation="MSEIYQIDNLDRGILNALMENARTPYAELAKNFGVSPGTIHVRVE
FT                   KMRQAGIITAACVHVNPKQLGYDVCCFIGIILKSAKDYPSALKKLESLEEVVEAYYTG
```

Figure 5.3 Screenshot of the EMBL annotation file downloaded from the RAST server, before manual adjustments were done. The ID character begins the first entry, AC codes for the accession number (at this time missing), DE names the description lines, FH the feature header lines and the FT character codes indicate feature table contents.

```
ID   FR729477; SV 1; circular; genomic DNA; STD; PRO; 4553420 BP.
XX
ST * public
XX
AC   FR729477;
XX
PR   Project:60483;
XX
DT   30-NOV-2010 (Rel. 107, Created)
DT   30-NOV-2010 (Rel. 107, Last updated, Version 0)
XX
DE   Yersinia enterocolitica subsp. palearctica Y11
XX
KW   complete genome.
XX
OS   Yersinia enterocolitica subsp. palearctica Y11
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
XX
RN   [1]
RP   1-4553420
RA   Batzilla J.;
RT   ;
RL   Submitted (24-NOV-2010) to the EMBL/GenBank/DDBJ databases.
RL   Batzilla J., Max von Pettenkofer-Institute, LMU, Pettenkoferstrasse 9a,
RL   Munich, D-80336, GERMANY.
XX
RN   [2]
RA   Batzilla J., Hoeper D., Heesemann J., Rakin A.;
RT   "Complete genome sequence of Y. enterocolitica subsp. palearctica serogroup
RT   O:3";
RL   Unpublished.
XX
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..4553420
FT                   /organism="Yersinia enterocolitica subsp. palearctica Y11"
FT                   /sub_species="palearctica"
FT                   /strain="Y11"
FT                   /serotype="O:3"
FT                   /mol_type="genomic DNA"
FT                   /db_xref="taxon:930944"
FT   CDS             1..720
FT                   /transl_table=11
FT                   /product="chromosomal replication initiator protein dnaA"
FT                   /translation="MLVEVLLNTPAQLSLPLYLPDDETFASFYPGENPSLLAAIQSAVR
FT                   QSHGSYIYFWSREGGGRSHLLHAACAELSQKGEAVGYVPLDKRAYFVPEVLDGMEQLAL
FT                   VCIDNIECIAGDEQWEMAMFNLYNRIVETGRTRLLITGDRPPRQLNLGLPDLASRLDWG
```

Figure 5.4 Screenshot of the EMBL annotation after manual adjustments. Several information have been added, including the status (ST character code, non-official designation), the project name (PR), dates (DT), organism species (OS), organism classification (OC) and other information as keywords or publication status. For a complete character description, see explanations in the text.

After the adjustment of these changes, curators of the ENA confirmed the successful submission of the annotated genome and provided the accession numbers, each a unique designation to find the genome and plasmid sequence in the databases at, for example, NCBI or EMBL. The *Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4 strain Y11 genome and plasmid sequences can now be found under FR729477 and FR745874 (plasmid) accession numbers.

5.1.8. Publication of the complete genome sequence of Y. enterocolitica *strain Y11*

The complete genome sequence of *Y. enterocolitica* strain Y11, serobiotype O:3/4, was published as a genome announcement at the Journal of Bacteriology (Batzilla *et al.* 2011c).

5.2. Draft genome sequencing of different *Yersinia enterocolitica* subsp. *palearctica* serobiotypes

During the complete genome sequencing of *Y. enterocolitica* strain Y11, serobiotype O:3/4, other genomes have been selected for draft genome sequencing. Genomic DNA has been isolated as described under 4.16.1, and sequencing was performed at BGI-Hongkong Co. (Hong Kong, China) using the Illumina sequencing technology.

5.2.1. Draft sequencing results of different Yersinia enterocolitica subsp. palearctica strains

Five *Y. enterocolitica* subsp. *palearctica* isolates, namely strains Y8265 and Y5307 of serobiotype O:3/4, strain Y5,27 of serobiotype O:5,27/3 and biotype 1A strains NF-O (serobiotype O:5/1A) and IP2222 (serobiotype O:36/1A) were sequenced as high coverage draft genomes during this study. The two additional O:3/4 strains have been used to unravel strain specific differences in this serobiotype, and the additional *palearctica* subspecies representatives of different serobiotypes have been used to enhance the information value about this subspecies and to discriminate between strain-, serobiotype- and subspecies-specific genetic elements. In addition, the draft genome sequencing of BT 1A strains, a nosocomial isolate of serobiotype O:5/1A (which caused an outbreak of diarrheal disease in Newfoundland, Canada) and an environmental isolate of serobiotype O:36/1A was performed to analyse the ambivalent pathogenic profile of BT 1A strains, both reported to be non-pathogenic but involved in registered outbreaks.

High-throughput Illumina sequencing of the five *Y. enterocolitica* subsp. *palearctica* isolates was used to conduct paired-end sequencing based on a 500 bp library with extended data of 500-1,000 Mb, and an additional 2 kb (NF-O, IP2222) or 6 kb (Y8265, Y5307, Y5,27) library with data of 250-900 Mb. Genome assembly resulted in 14 large scaffolds with 215 contigs for strain Y8265, 18 scaffolds with 256 contigs for strain Y5307, 20 scaffolds with 408 contigs for strain Y5,27, 10 large scaffolds with 74 contigs for strain IP2222 and 14 scaffolds with 97 contigs for strain NF-O. Genome coverage based on reads mapping was about 97 % (Y8265, Y5307, Y5,27) to 99 % (NF-O, IP2222) for all genomes. The draft genome of strain Y8265 comprised 4,521,393 bp with 4296 CDS (83 % coding density), an average CDS size of 868 bp and an overall GC content of 45.41 %. The draft genome of strain Y5307 consists of 4,553,105 bp with 4393 CDS (83 % coding density), an average CDS size of 861 bp and an overall GC content of 45.6 %. For strain Y5,27, we maintained a draft genome size of 4,658,995 bp with 82 % coding regions and 4351 CDS. The average CDS size was 875 bp and the overall GC content 45.97 %. In case of the environmental strain IP2222, genome sequencing revealed 86.16 % coding sequences, a GC content of 47.14 % and a draft genome size of 4,796,259 bp. For strain NF-O, we maintained

85.55 % coding regions and an overall GC content of 47.08 %. The draft sequence genome size was 4,695,527 bp. All genomes were annotated with RAST (Aziz *et al.* 2008).

5.3. Accession numbers of sequenced *Yersinia enterocolitica* subsp. *palearctica* strains

Table 5.1 Accession numbers of strain sequenced in this study. The sequence of strain Y11 is a published as a complete genome, whereas the sequences of the other strains are draft genomes published as whole genome shotgun projects.

Strain name and serobiotype	Accession numbers in EMBL database
Y11, O:3/4	FR729477 and FR745874 (plasmid)
Y8265, O:3/4	CACU01000001- CACU01000014
Y5307, O:3/4	CACV01000001- CACV01000018
Y5,27P, O:5,27/3	CACW01000001- CACW01000020
NF-O, O:5/1A	CACY01000001–CACY01000097
IP2222, O:5/1A	CACZ01000001–CACZ01000074

5.4. General genomic features of *Yersinia enterocolitica* subsp. *palearctica* strain Y11, serobiotype O:3/4

5.4.1. The virulence plasmid pYV

The virulence plasmid pYV is present in all human pathogenic *Yersinia*. It encodes the Ysc-T3SS, with the effector proteins, and the adhesion protein YadA (Y11_p0221 in strain Y11). The pYV plasmids of serobiotype O:3/4 and O:8/1B are more divergent than the corresponding chromosome sequences. In contrast, the pYV plasmid sequence of serobiotype O:9/2 (accession no. AF102990.1) is highly homologous to the plasmid of O:3/4. Thus, the replication origin was predicted based on the pYVe227 plasmid sequence of serobiotype O:9/2. The virulence associated pYVeY11 plasmid covers 72,460 bp with an overall GC content of 43.99 %. For the plasmid sequence of strain Y11, the number of CDS and the coding density are higher than for the pYV plasmid of *Y. enterocolitica* strain 8081 (serobiotype O:8/1B). From 112 CDS in the Y11 plasmid, 18 CDS were found annotated as hypothetical proteins, constituting 16 % of all annotated CDS. One copy of the *palearctica* group specific IS element Yen2A/B was found within the pYVeY11 plasmid sequence, together with a large number of transposase-like genes that may have caused rearrangements of the plasmid sequence. Beside the Yop machinery genes and *yadA*, an *arsABRH* gene cluster for arsenic resistance facilities was found.

5.4.2. The chromosomal replicon

The chromosome replicon of *Y. enterocolitica* subsp. *palearctica* strain Y11 consists of 4,553,420 bp with a GC content of 47.01 %. The start of the sequence was set to the *dnaA* gene as

agreed by the scientific community. Putative rRNA-coding sequences were predicted and annotated by the RAST annotation tool, and the reported features of "Large Subunit Ribosomal RNA", "Small Subunit Ribosomal RNA" and "5S RNA" can be traced back to seven copies of rRNA clusters present in strain Y11, which is in common with the number of reported clusters in serobiotype O:8/1B strain 8081. In contrast, the number of tRNAs in strain Y11 is remarkably lower (70 tRNAs) than in strain 8081 (81 tRNAs), but still they cover cognates for all 20 amino acids. The complete genome size of strain Y11 is smaller compared to the chromosome replicon sequence of strain 8081, but even though the overall number and length of CDS, and therefore also the coding density, is higher in strain Y11. From 4,355 CDS in the strain Y11 genome, 623 CDS annotated as hypothetical proteins were found, constituting 14.3 % of all annotated CDS. About 64 IS elements were detected, of which 53 are derivates of the ISYen1 IS element and seven of ISYen2A/B, a specific IS element found so far only in genomes of the *palearctica* subspecies.

Features of both the chromosome replicon and the pYV virulence plasmid of *Y. enterocolitica* subsp. *palearctica* strain Y11 and the already sequenced representative of the *enterocolitica* subspecies, strain 8081 (Thomson *et al.* 2006), are listed in Table 5.2. In Figure 5.5 and Figure 5.6, DNA plots of the *Y. enterocolitica* strain Y11 genome and plasmid are shown.

Table 5.2 General features of the genome and plasmid sequences of *Y. enterocolitica* strain Y11 (O:3/4) and strain 8081 (O:8/1B)

Feature	Genome Y11	Genome 8081	Plasmid pYVeY11	Plasmid pYVe8081
Genome size	4,553,420	4,615,899	72,460	67,721
GC content in %	47.01	47.27	43.99	43
Number of CDS	4,355	3,978	112*	72
Average CDS size	884	968	506	672
% Coding density	85	83	78	71
rRNA clusters	7	7	none	none
tRNAs	70	81	none	none

*) The higher number of coding sequences (CDS) for the O:3/4 pYV plasmid compared to O:8/1B is evoked by more transposase-like CDS (including also remnant or fragmented transposases) and 18 hypothetical CDS with no strong similarity on protein level to serobiotype O:8/1B. In addition, an *arsABRH* gene cluster is present in the pYV plasmid of serobiotypes O:3/4 and O:9/2, but not in O:8/1B

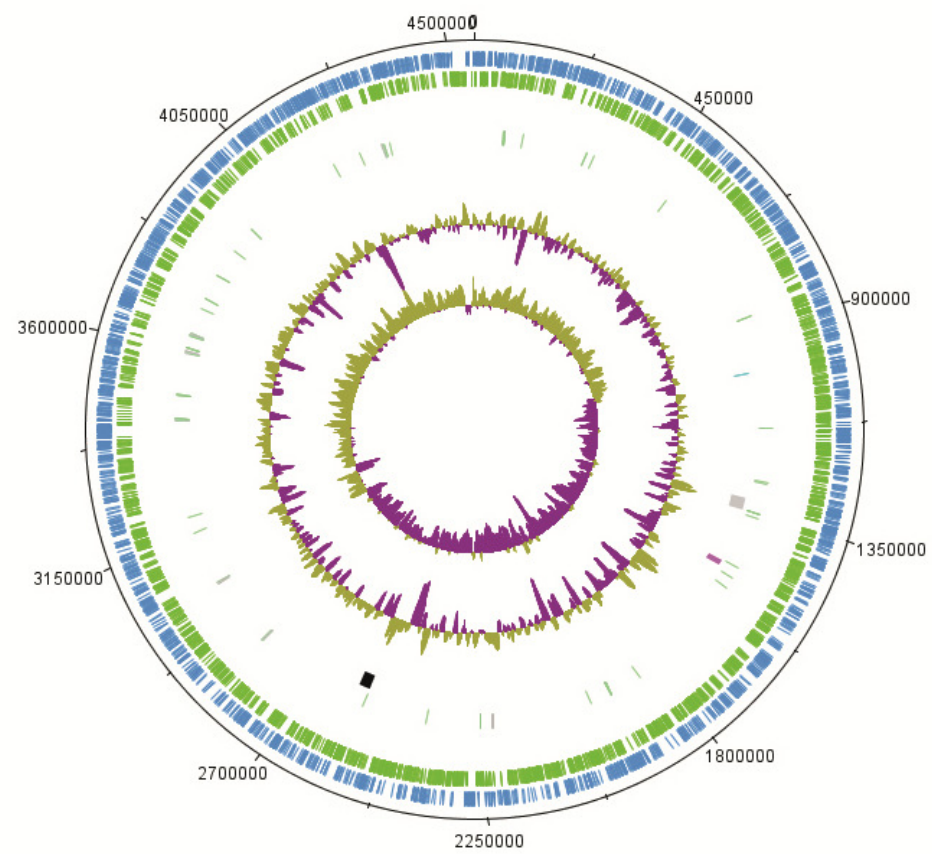


Figure 5.5 DNA plot of the strain Y11 genome sequence. The outer scale shows the size in bps. From the outside in, circles 1 and 2 show the position of CDS transcribed in a clockwise and anticlockwise direction, respectively. Circle one (blue) shows CDS transcribed in a clockwise direction, and circle two (green) illustrates CDS transcribed in an anticlockwise direction. Circle 3 represents tRNA and rRNA features in light green and dark grey, respectively, and circle 4 shows the location of particular genomic island and prophages specific for serobiotype O:3/4 (described in detail in section 5.5.6): PhiYep-1 (turquoise), PhiYep-2 (black), PhiYep-3 (grey) and GIYep-01 (pink). Circle 5 represents the GC plot (in %), with a window size of 10,000 and a step size of 200, and circle 6 represents the calculated GC skew [(GC)/(G+C)], with the same window and step size. The DNA plot was made using the DNAplotter software.

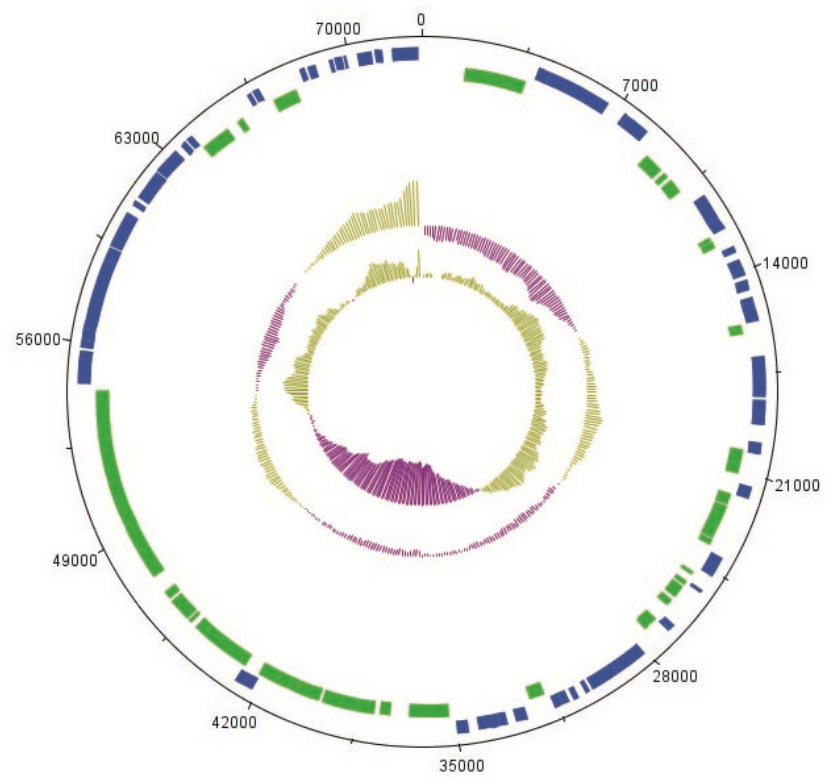


Figure 5.6 DNA plot of the strain Y11 plasmid sequence. The outer scale shows the size in bps. From the outside in, circles 1 and 2 show the position of CDS transcribed in a clockwise and anticlockwise direction, respectively. Circle one (blue) shows CDS transcribed in a clockwise direction, and circle two (green) illustrates CDS transcribed in an anticlockwise direction. Circle 3 represents the GC plot (in %), with a window size of 10,000 and a step size of 200, and circle 4 represents the calculated GC skew [(GC)/(G+C)], with the same window and step size. The DNA plot was made using the DNAplotter.

5.5. Genome comparisons of *Yersinia enterocolitica* subsp. *palearctica* strain Y11, serobiotype O:3/4

5.5.1. Differences between Y. enterocolitica subsp. palearctica strain Y11 and the two draft sequences of serobiotype O:3/4 strains Y8265 and Y5307

During the strain Y11 sequencing project, a draft genome sequencing of two further O:3/4 isolates has been initiated. Strain Y8265 is a human isolate from France (Saken et al., 1994) and strain Y5307 was isolated from a reactive arthritis patient. Among the three serobiotype O:3/4 isolates, only 57 proteins in strain Y11 do not have orthologs in both strains, Y8265 and Y5307, and of these only 9 proteins were larger than 100 amino acids (aa). Proteins that were not annotated as hypothetical or transposase/phage derived longer than 100 aa are: a carbonic anhydrase of 165 aa, Y11_02911, and a NAD-dependent DNA ligase of 295 aa, Y11_13111, part of the PhiYep-3 prophage (described in detail under 5.5.6). There have been 96 Y11-specific predicted proteins found when comparing all annotated Y11 proteins to strain Y8265, but 60 of them are hypothetical and smaller than 100 aa, 15 of them are transposases and only 8 genes (not belonging to transposons/prophages and not hypothetical ones) encode proteins longer than 100 aa. Likewise, when compared to strain Y5307, there are 87 Y11-specific proteins, and again a large number of 72 proteins are annotated as hypothetical proteins (including 61 proteins smaller than 100 aa), and only 3 proteins (that do not belong to transposons/prophages and are not annotated as hypothetical) are found to be larger than 100 aa (namely, again the carbonic anhydrase of 165 amino acids, Y11_02911, again the NAD-dependent DNA ligase of 295 aa, Y11_13111, which is part of the PhiYep-3 prophage, and a DNA primase TraC of 253 aa, Y11_15061). When strain Y5307 was compared with strain Y11, much more hypothetical proteins smaller than 100 aa that do not have orthologs in strain Y11 were found (140 proteins) compared to 61 found as Y11-specific. This can be due to artificially shortened protein products at the ends of the contigs that can occur twice, due to expected overlaps between the contigs. Figure 5.7 illustrates the differences between strain Y11 and the draft serobiotype O:3/4 genomes. Two circles represent the annotated proteins of the draft genomes of strains Y8265 and Y5307, respectively, and the level of homology of every predicted protein compared to strain Y11 is shown by a colour code. Interestingly, two regions on both circles directly attract the attention - located in these light coloured regions are the annotated proteins of the two P2-like prophages PhiYep-2 and PhiYep-3. Both prophages share a large number of orthologous proteins with the two serobiotype O:3/4 genomes (strains Y8265 and Y5307), but only one copy of the P2-like phage is present and the level of homology is noticeable reduced compared to the rest of the genomes. The P2-like phage present in the two draft genomes was assumed to be PhiYep-2, based on the comparison of

neighbouring genes and excision PCRs for the two serobiotype O:3/4 genomes of strain Y8265 and strain Y5307.

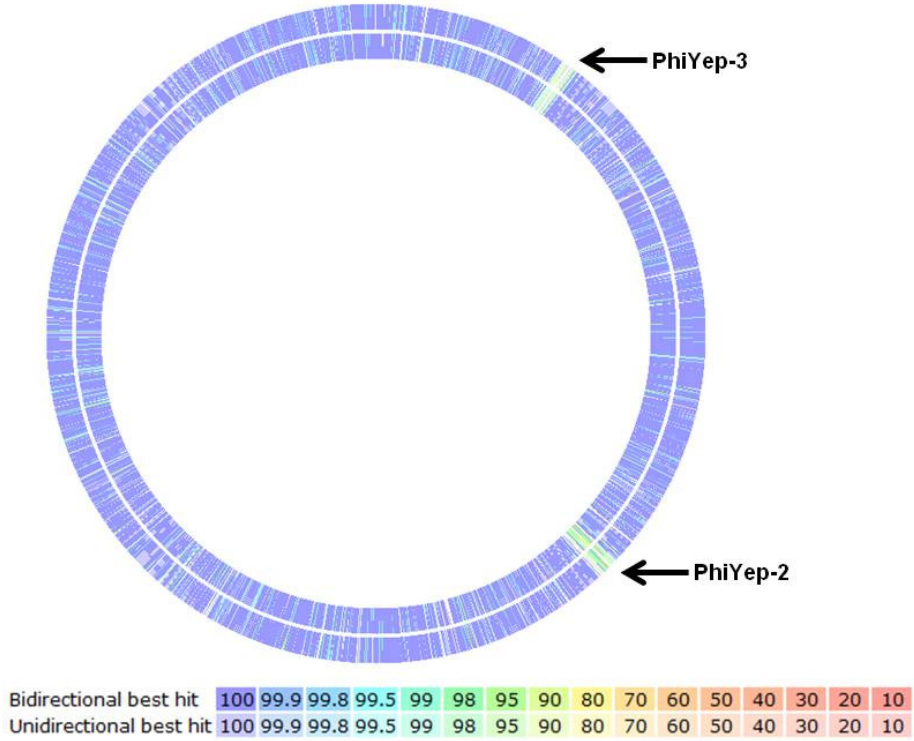


Figure 5.7 RAST comparison map of strain Y11 with the two serobiotype O:3/4 draft genomes. Two circles represent the annotated proteins of the draft genomes of strains Y8265 and Y5307, respectively, and the level of homology of every predicted protein compared to strain Y11 is shown by a colour code as indicated by the box below the circles.

5.5.2. Virulence pattern differences between Y. enterocolitica subsp. palearctica strain Y11, Y. enterocolitica subsp. enterocolitica strain 8081 and biotype 1A strains NF-O and IP2222

The draft genome sequences of two biotype 1A strains, a nosocomial serobiotype O:5/1A isolate (which caused an outbreak of diarrheal disease in Newfoundland, Canada) and an environmental serobiotype O:36/1A isolate, have a mosaic genome organization and share genes both with serobiotypes O:8/1B and O:3/4 which suggests their common descent. Even though 16S rRNA sequencing of both BT 1A strains confirmed their affiliation to *Y. enterocolitica* subsp. *palearctica*, the fact that these strains harbour serobiotypes O:8/1B and O:3/4 specific genes hampers an exact differentiation between both subspecies, namely *Y. enterocolitica* subsp. *palearctica* and *Y. enterocolitica* subsp. *enterocolitica*, and the definition of a unique core genome for each of these subspecies. Putative virulence and fitness genes shared with serobiotype O:8/1B are, among others, invasin-like genes, an arsenical gene cluster as well as metabolic genes, as for

example the glucoside-operon *arbBFG* and amino acid transport and modification genes. Putative virulence and fitness genes shared only with the low virulent serobiotype O:3/4 are a specific chromosomal type three secretion system (T3SS), a beta-fimbriae gene cluster, the N-acetyl-D-galactosamine PTS *aga*-operon and haemolysin-, invasin- and pertactin-like genes. As expected, in addition to the pYV virulence plasmid, other established markers of both virulent serobiotypes O:8/1B and O:3/4 are absent, for example the Ail adhesin, the siderophore yersiniabactin (encoded as a part of the High-Pathogenicity Island in biotype 1B), the virulence associated proteins VapC and VagC and the enterotoxin YstA. Interestingly, no insecticidal toxin cluster genes as described for serobiotype O:3/4 have been found in both BT 1A isolates, even though these genes are described for other low pathogenic serobiotypes and have been reported to be present also in 1A isolates (Tennant *et al.* 2005).

In contrast to previous studies, the strains of the 1A biotype were found to harbour a gene encoding the MyfA antigen, although with low similarity to each other and to the highly conserved MyfA of both virulent serobiotypes O:8/1B and O:3/4. This suggests potential different roles for the Myf fibrillae in different *Yersinia*. It has been previously reported that Myf does not mediate haemagglutination in *Y. enterocolitica* in contrast to the *Y. pseudotuberculosis* pH6 antigen (Iriarte and Cornelis 1995). The genetic differences revealed are not uniform along the *myf* cluster, and while the MyfA sequences have evidently diverged, the MyfF and MyfC proteins remain highly conserved (97 % and 94 % positive residues, respectively). Thus the MyfB chaperone seems to co-evolve with the MyfA adhesin, while the MyfC usher and the MyfF transcriptional regulator appear to be less constraint and dependent on MyfA sequence alterations. Indeed, the ushers act as multifunctional assembly and secretion platforms in the bacterial outer membrane (Henderson *et al.* 2011). On the other hand, one of the two conserved regions (PCRM1, periplasm chaperone recognition motif, see Figure 5.8) that might be involved in MyfA interactions with its chaperone has undergone evident changes in the O:36/1A non-clinical strain with the largest MyfA fibrillae. The promoter sequences preceding the *myfA* sequence are rather conserved in all strains (see Figure 5.9). Besides the indels occurred, the distance between the -10 region and the ATG start codon is well preserved in all isolates.

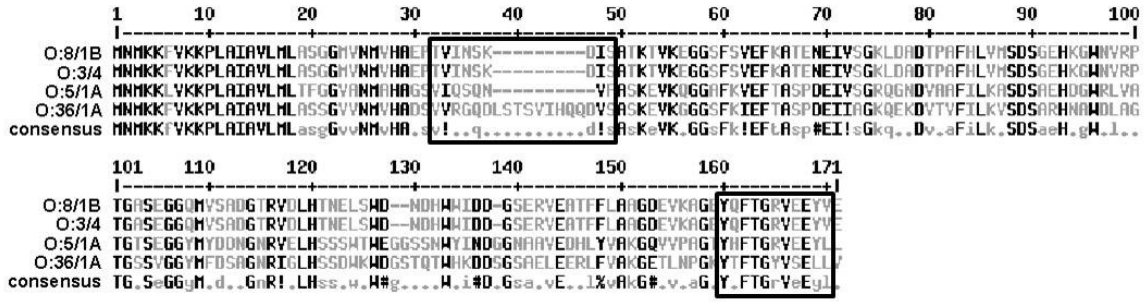


Figure 5.8 Alignment of the amino acids of MyfA or MyfA-like antigens of serobiotypes O:8/1B (strain 8081), O:3/4 (strain Y11), O:5/1A (strain NF-O) and O:36/1A (strain IP2222). The residues conserved in all sequences are coloured in black. Black boxes indicate the PCRMs, the periplasmic chaperone recognition motifs (PCRM1, first box, PCRM2, second box) (Torres-Escobar *et al.* 2010). The alignment has been done using the Multalign online tool using the given parameters.

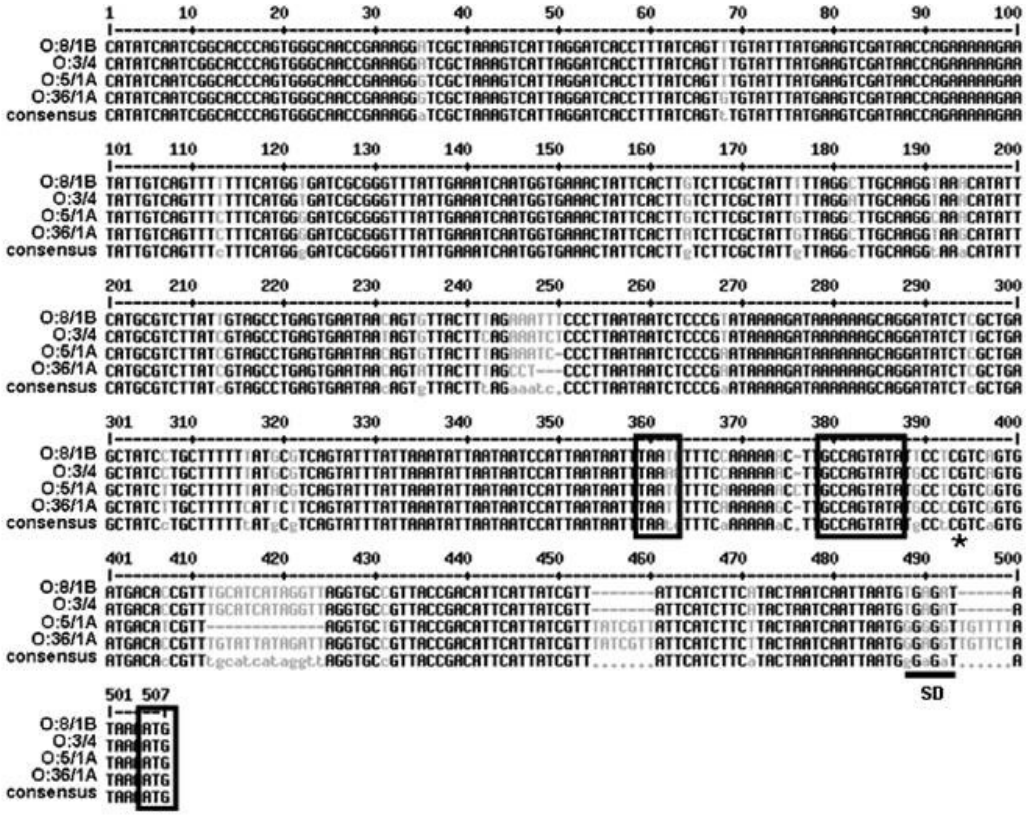


Figure 5.9 Alignment of the promoter nucleotide sequence of *myfA* and *myfA*-like genes of serobiotypes O:3/4 (strain Y11), O:5/1A (strain NF-O) and O:36/1A (strain IP2222). Exact matches in all sequences are coloured in black. Black boxes indicate the predicted -10 and -35 regions as well as the start codon. The transcription start point is marked by an asterisk (Iriarte and Cornelis 1995). SD, Shine Dalgarno sequence. The alignment has been done using the Multalign online tool using the given parameters.

Many virulence-associated and putative genes that might be involved in pathogenicity and fitness are present and shared by both the clinical and non-clinical BT 1A isolates. Namely these were the genes for the enterotoxin *ystB*, the invasin *invA*, the antigen *myfA* (even though with less similarity), the chromosomal gene modulating the expression of virulence functions *ymoA* (Cornelis *et al.* 1991), the in vivo-expressed protease *hreP* (Heusipp *et al.* 2001), the enterochelin utilisation gene cluster *fepBDGC/fepA/fes*, the arsenic cluster *arsCBR*, several putative haemolysin-, invasin- and pertactin-like gene clusters, the chromosomally encoded serobiotype O:3/4-like Ysp-T3SS, the beta-fimbriae cluster, the N-acetyl-D-galactosamine PTS genes, and the bacteriocin genes. The InvA proteins of both the clinical and non-clinical BT 1A isolates are highly homologous, with only single amino acid exchanges, but less homologous to InvA of strains Y11 and 8081 (see Table 5.3).

Table 5.3 Similarities in % between the InvA proteins of strains NF-O (O:5/1A), IP2222 (O:36/1A), 8081 (O:8/1B) and Y11 (O:3/4)

	InvA NF-O	**InvA IP2222**	**InvA 8081**	**InvA Y11**
InvA NF-O	-	96 %	85 %	84 %
InvA IP2222	96 %	-	85 %	84 %

The main difference in the assorted virulence pattern between the two 1A strains was the presence of putative fragments of the O:3/4-like *rtxA* gene cluster (Y11_18761-18781) only in O:5/1A, an *rtxA*-like putative leukotoxin gene (Y11_01911-01921) only in O:36/1A and remnants of one of the O:3/4-specific prophages, PhiYep-2/PhiYep-3 (Y11_25141-25551/ Y11_13081-13511), only in the NF-O isolate of clinical origin. The *rtxA* gene is split into four fragments in O:5/1A and suspected to be non-functional. Even though the *rtxC* and *rtxH* genes are 80 % identical in O:5 and O:3. The conserved regions of the N- and C-terminal regions of RtxA contain Rtx-family-specific domains and repeats (with a 130 bp extension of this repeat region in strain Y11 in comparison to O:5/1A). The more dissimilar region in the center of the protein alignment represents a putative effector domain region, with a peptidase C58 motif, a thiol protease cysteine motif and a putative signal profile found in Y11. This might impose a profound effect on the functionality of the protein. However, the impact of these putative virulence or fitness associated gene clusters is not resolved, and thus the two representatives of the BT 1A seem to be armed with some known potential virulence factors that are also found in the human pathogenic yersiniae of biotypes 1B and 2-5.

5.5.3. Y. enterocolitica subsp. enterocolitica strain 8081 genes absent from Y. enterocolitica subsp. palearctica strain Y11

Some of the virulence determinants that are associated with the high-pathogenicity phenotype of *Y. enterocolitica* subsp. *enterocolitica* are missing in the sequenced strains of serobiotype O:3/4. Most prominent are the High-pathogenicity Island (HPI), involved in yersiniabactin-mediated iron uptake, and two chromosomally encoded secretion systems (the Ysa-T3SS and the Yst1 T2SS), located within the Plasticity Zone (PZ). Within the large PZ, an arsenical resistance operon is located in *Y. enterocolitica* O:8/1B, as well as many putative exported proteins, a putative haemolysin and a ferric enterochelin operon. Those genes are all absent from *Y. enterocolitica* O:3/4. Besides HPI and PZ, also several genomic islands (namely, YGI-2 and YGI-3) and prophages (e.g. ΦYE250, ΦYE185 (encoding the Vaps, virulence-associated proteins)) are absent from the genome sequence of strain Y11. Most of the genome differences of strain 8081 to other serobiotypes mentioned above have been described in detail previously (Thomson *et al.* 2006). However, some so far unnoted putative virulence-associated genes have been identified:

The haemolysin cluster YE2407-2408 of strain 8081 is absent from *Y. enterocolitica* O:3/4 but found in *Y. pestis* and *Y. pseudotuberculosis*. In addition, the putative haemolysin gene YE2966 (annotated by RAST as haemolysin A) is completely different from the respective putative haemolysin found in *Y. enterocolitica* O:3/4 at the same location and genetically background (Y11_18861). The adjacent genes surrounding this putative haemolysin are quite homologous - the protein identity of the adjacent coding sequences varies between 94 % and 98 %, whereas the identity of the putative haemolysin is only 66 % (shown in Figure 5.10), with the main differences in length and amino acid composition at the N terminus of the protein. A different role or impact in pathogenesis could be assumed for both genes.

```
 CLUSTAL W (1.8.3) multiple sequence alignment

---------------MNMKVKIRKTSIASTLIKNGEFKSIYIKNAPPKKI
MIITDISKKEIILSKKTSKNKSKTIPITGILITNGKMKSIFLDYISPDKV
                * *   * *   ** **  ***      * *

FFEYSDAILSIKFLDKNNNVYAIRKYTVVKDGQFTINSIFLREKNGIRHE
FFQYLDSKLKVIIVADNNHFFS-KTYHVIDNGNYIVKSVILRDKNGVKTE
** * *  *       **       * *   *     *  ** ***   *

LFLTLTETYKRENFYDDPSKDLLLTREDLPHHSAHHLITNTTPSKSALLT
LFLTLKETKKKEGFYDGPSKDILLIREDLKYNSAHHLITNTRPVERFSLN
***** ** * * *** **** ** ****   ********* *     *

DNDDFIGHKYIRNKVIGGDGNNIIQGGDKSDNLAGGAGNDIIFGHASFDS
GDDTFSGHKYIRNEVTGGEGNNTIITGGDKNDNLAGGNGNDLIMGYDSFDS
  * * ******* * ** *** * **** ****** *** * *  ****

LFGQDGNDILIGGEGNDGLTGGAGNDILDGGDGNDQLHGDADGLASFASE
LFGEGGNDVLVGGNGNDGLTGGAGDDFLDGGDGDDDLEGDAGGL--YASE
***  *** * ** ********** * ****** * * *** **   ***

SSWRGNDIIFGGKGNDRIEGEKNNDYLAGGTGNDSYIFSAYDGINMIVEY
AQWRGNDILFGGKGNDRIRGGKNNDYLAGGTGDDLYVFSAYDGINMIVEY
  ****** ********* * *********** * * **************

SNEENTISINDHFFHQLKFDRYGNHLMISSTEVHPNNLVIVIKDQYAEDG
ADEENTISIDDYFFHQLKFTRYGNHLIISSTEDHPNNLVIVIKDQYTPDG
  ******* * ******* ****** ***** *************  **

YKIKHLQTKSYLRYGSENDKKIHEAIINNLVMKYTKDPDSLVNELNNIKD
YKVKNLQTKSYLRHGSDDDKNNHTVMINNLVRKHTDDAEKIVAKLSNIKN
** * ******** **  **  *   ***** * * *    *  * ***

FGDYYRIKLSEFYCEKTPNIEEITGKNLVDVLIEKNNALREMYEDSAFYF
FGDYYPTDLSDIFCEKIPDIEKTTGENLIDMLIEQNNTLREMYAKRAFYF
*****   **   *** * **  ** ** * *** ** *****   ****

GDNLPDITYITQALSSFAPLEASQSDIKYLKSPIPMDNLANSLVTGKH
GDNLPDITYIIQALSSFAPKEASQYDIKYLTSPIPMDSFAGGLVTVKH
********** ******** **** ***** ******  *  *** **
```

Figure 5.10 Protein alignment (ClustalW) between YE2966 and Y11_18861. Positive matches are indicated by asterisks. The identity level between both proteins is 66 %.

Another haemolysin gene cluster is found in all *Yersinia* genomes analysed (YE4084-4086). Precursor and activator genes of the cluster are closely related in all strains, but the haemolysin structural gene may be no longer active in serobiotype O:3/4 (Y11_31011), *Y. pestis* and *Y. pseudotuberculosis*. The gene is interrupted by a stop codon in strain Y11 and completely absent from *Y. pestis* and *Y. pseudotuberculosis*, putting their functionality and contribution to virulence into question.

In strain 8081, the gene YE1873 is a putative adhesion protein of 639 aa, named the YadA C-terminal domain protein. This protein is unique for *Y. enterocolitica* O:8/1B (the gene was found identical in strains 8081 and WA-314) and is absent from all other yersiniae, including serobiotype O:3/4. It is located near a transposase gene, a possible explanation for the route of acquisition.

5.5.4. Y. enterocolitica subsp. palearctica serobiotype O:3/4 specific genes absent from Y. enterocolitica subsp. enterocolitica serobiotype O:8/1B

5.5.4.1. Putative toxins

One of the declared intentions of the genome comparisons made in this study was the detection of yet unknown genes with a possible impact on pathogenicity or on the adaptation process to the most prominent host of the serobiotype O:3/4, the pig. In the following section, putative novel toxins and their characteristics will be described. Specially denoted features will be revisited in chapter 5.6.

5.5.4.1.1. RtxA

A large gene cluster which codes for a large RtxA-like protein was detected in *Y. enterocolitica* serobiotype O:3/4 (Y11_18761, Y11_18771 and Y11_18781), whereas *Y. enterocolitica* serobiotype O:8/1B does not possess the gene cluster (see Figure 5.11).

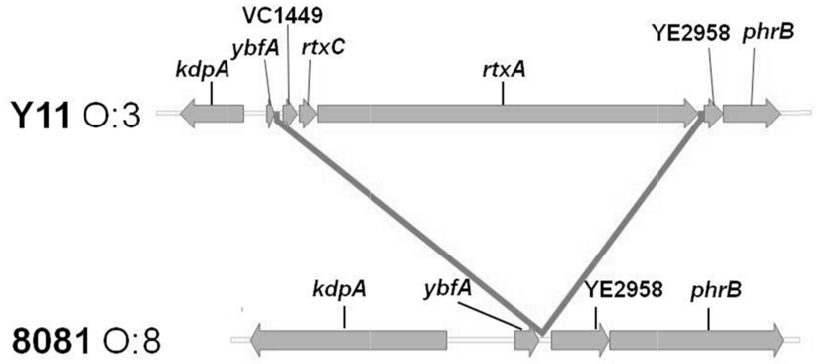


Figure 5.11 The upper panel shows the *rtxA* gene cluster of strain Y11 comprising VC1449 (*rtxH*), *rtxC* and *rtxA* with their adjacent genes *kdpA* (putative potassium-transporting ATPase A chain encoding gene), *ybfA* (a putative membrane or exported protein) and *YE2958* (encoding a hypothetical protein) and the lower panel shows the genetic background shared with strain 8081 which lacks the *rtx* gene cluster.

The typical genetic organisation of *rtx* toxin gene clusters shows an operon with an activator (*rtxC*, a putative acyltransferase), pretoxin (*rtxA*) and transport genes (atypical T1SS). A homolog to a conserved hypothetical gene, (annotated as VC1449 or *rtxH*, similar to the gene in the toxin cluster of *V. cholerae*) is also found in serobiotype O:3/4 strain Y11. Former studies have shown that *rtxA* is found as one mRNA transcript with *ybfA*, *rtxH* and *rtxC* (Golubov 2005). The transport genes, unlikely, are not located near this gene cluster in serobiotype O:3/4 (Y11_10141-10171). In addition, the genes encoding the putative RtxA secretion system found in strain Y11 are interrupted by a frameshift, elevating the number of genes from three to four (compared to strain 8081 and other related species), raising the question of RtxA export (see Figure 5.12). The

DNA cluster for RtxA was confirmed and sequenced also in strain MRS40 of serobiotype O:9/2 and O:5,27/3 (in which *rtxA* is interrupted into three smaller genes). Both genes are highly homologous, with minor amino acid differences, but the regions that could form the functional domains in the middle of the gene are almost identical.

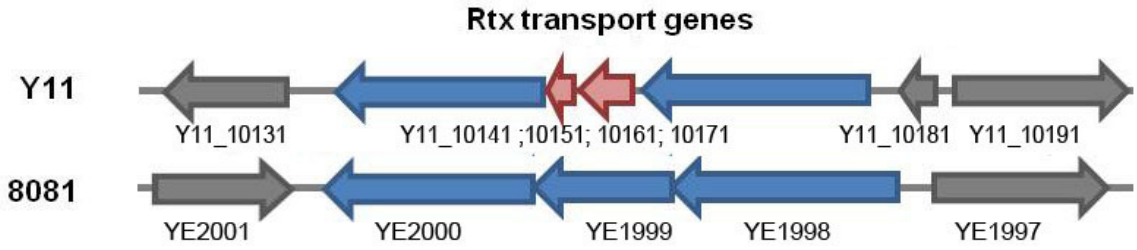


Figure 5.12 Comparison of the *rtx* transport gene cluster of strain Y11 (O:3/4) and strain 8081 (O:8/1B). Arrows in blue depict the homologous transport genes of both strains, indicating its orientation, whereas red arrows show the interrupted middle transport gene found in the cluster of strain Y11. Predicted functions and gene length: Y11_10131, putative ATPase gene (1179 bp); Y11_10141, putative *rtx* transport gene (1908 bp); Y11_10151, putative *rtx* transport gene (186 bp); Y11_10161, putative *rtx* transport gene (474 bp); Y11_10171, putative *rtx* transport gene (1944 bp); Y11_10181, hypothetical gene (162 bp); Y11_10191, putative formate-tetrahydrofolate ligase gene (1758 bp), YE1997, probable formate--tetrahydrofolate ligase gene (1805 bp); YE1998, putative RTX-family toxin transporter gene (2126 bp); YE1999, hemolysin transport protein (1394 bp); YE2000, putative toxin transport protein (2132 bp); YE2001, L-arabinose isomerase (1502 bp).

The large RtxA protein (Y11_18761, length: 3246 aa, about 350 kDa) could play a special role in the pathogenesis of serobiotype O:3/4, since well-studied members of the wide family of Rtx proteins have been shown to be either cytotoxic, haemolytic, or able to induce cell rounding and thus being major contributors to the pathogenicity of different bacterial species (Kudryashov *et al.* 2008b;Lally *et al.* 1999;Welch 2001;reviewed in Linhartova *et al.* 2010). The activity of this protein could be, as among related family members, cell type specific (Vanden Bergh *et al.* 2008). The *rtxA* gene encodes an inactive form of the toxin, which is modified post-translationally to a cytotoxic form by a helper gene product (assumed to be encoded by *rtxC*) before secretion. Most Rtx toxins are activated by fatty acid acylation. After procession, the Rtx toxins of the MARTX family (Multifunctional Autoprocessing Repeats-in-Toxin) are predicted to integrate into the membrane. The large protein of serobiotype O:3/4 has been shown to be a member of this family (Satchell 2007). The *rtxA* repeat motifs in the first and last third of the protein have been assumed to form an anchor in the cytoplasmic membrane of the host cell, enabling the exposure of specific effector domains into the cell. These putative effector domains in the middle third of the protein of strain Y11 show no homology to already known family members. However, motif searches

detected three independent regions of cysteine peptidase motifs (see Figure 5.13), which could act as endopeptidases at specific cleavage sites to release the effector domains into the cytosol of the host cell. Rtx toxins require Ca^{2+} binding for their cytotoxicity. Regions of haemolysin type calcium binding motifs were found at the C-terminal end of the protein. Glycine rich nonapeptide repeats are supposed to be responsible for the binding of Ca^{2+}. The consensus domain structure of the family members of this motif is GGXGXDX[L/I/V/W/Y/F]X or [L/I/F]XGGXG[N/D]DX (where X is any amino acid). One cell attachment (RGD) sequence motif appears at the amino acid position 2234-2236 of RtxA. The RtxA of *V. cholerae* contains two similar motifs, whereas it is absent in HlyA of *E. coli*. This short peptide sequence, found also in adhesive proteins, facilitates binding to the integrin family of cell surface receptors. However, despite the vast amount of data obtained from other Rtx proteins, the putative mode of action of the RtxA-like protein is unclear in strain Y11. During this study, the biological role of RtxA in strain Y11 has been addressed and a mutant lacking the *rtxA* gene has been generated (see chapter 5.6.6).

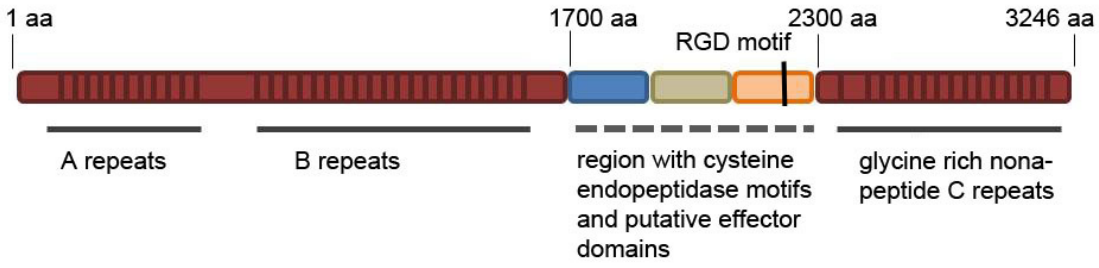


Figure 5.13 Schematical overview of selected putative amino acid sequence motifs of the RtxA protein of strain Y11 (about 350 kDa) as predicted by the MOTIF search tool (http://motif.genome.jp/). Repeats that are characteristic for Rtx toxins are indicated by dashed regions and horizontal designations (A, B and C repeats established by (Satchell 2007)). The putative effector domains are located in a region of three cysteine endopeptidase motifs.

5.5.4.1.2. Bacteriocin

Bacteriocins are toxins produced by bacteria to inhibit the growth of similar or closely related bacterial species and strains. The family of bacteriocins can vary vastly in the mode of action and specificity. In general, bacteriocin producers protect themselves from bacteriocin attack by bacteriocin-immunity proteins secreted into the periplasmic space, which capture the active site of the toxins (reviewed in Michel-Briand and Baysse 2002). The genome of *Y. enterocolitica* O:3/4 harbours a bacteriocin cluster composed of three genes, one bacteriocin and two putative immunity proteins, located next to a phage antitermination gene (Y11_33511-33531). The bacteriocin gene cluster and adjacent genes are missing in the genome of strain 8081. The

bacteriocin protein of 905 amino acids detected in serobiotype O:3/4 (Y11_33511) shows sequence homology to pyocin-like proteins. It exhibits also similarity to DNAses at the C-terminus, therefore likely being a potential endonuclease enzyme. Interestingly, differing numbers of the genes coding for the immunity proteins of different size in the other *Yersinia* species have been found. In addition, a similar cluster was detected in biotype 1A strains in the same genomic environment, but the bacteriocin genes are interrupted by frameshifts into three smaller proteins in strain IP2222 (O:36/1A), raising the question of functionality (see Figure 5.14). The adjacent phage antitermination gene in both biotype 1A strains is found flanking the bacteriocin gene cluster on the other side compared to strain Y11, possibly indicating a hot spot for integration and transfer events. Nevertheless, all clusters found in different yersiniae are linked to *papD*, encoding a pilus chaperone, and thus are situated in a similar chromosomal location. The duplication of immunity protein like genes could indicate a particular toxin with extreme toxicity for its host. Since *Y. enterocolitica* O:3/4 has adapted to a very narrow and specific niche, the pig nostrils, this bacteriocin cluster would provide a serious advantage in colonisation by excluding other bacteria to colonise the same niche. The toxicity and host range of this bacteriocin has been addressed in chapter 5.6.5.

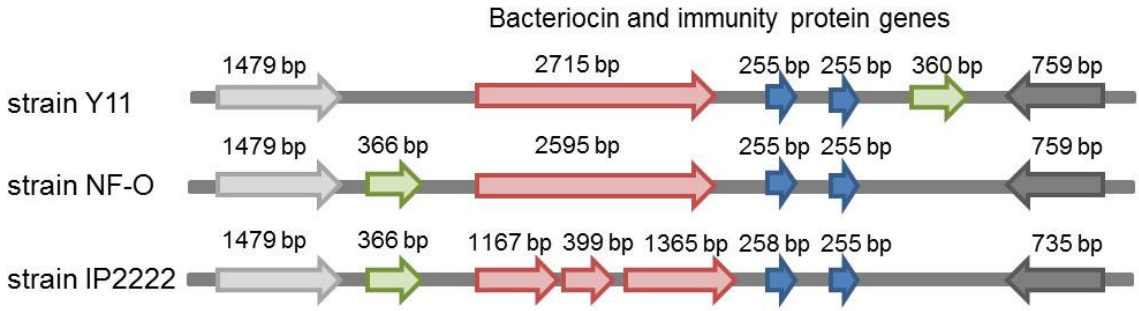


Figure 5.14 The bacteriocin gene cluster in strain Y11 compared to those of strains NF-O (serobiotype O:5/1A) and IP2222 (serobiotype O:36/1A). The gene orientation is indicated by the respective arrow direction and gene lengths are given in base pairs (bp). Bacteriocin genes are indicated by red, immunity genes by blue arrows. Green arrows mark the hypothetical phage antitermination gene Q. This gene is flanking the bacteriocin gene cluster on the right side in strain Y11 and on the left side in the two biotype 1A strains. Light grey arrows mark the ketol-acid reductoisomerase gene, dark grey arrows the pilus assembly protein gene *papD*.

5.5.4.1.3. Other putative toxins

Two pertactin like proteins were found in strain Y11 (Y11_25741, Y11_42801). Both are absent from *Y. enterocolitica* O:8/1B, only one gene directly adjacent to one of the pertactin-like protein

encoding genes (Y11_25751) shows minor similarity (less than 65 % identity) to the beta-barrel outer membrane protein in *Y. enterocolitica* O:8/1B (YE3700). For Y11_25751, *Y. pestis* and *Y. pseudotuberculosis* share the same backbone, but have just a single homologous protein matching both with Y11_25741 and Y11_25751. The protein of the annotated gene Y11_42801 shows similarities to *Y. bercovieri* and *Y. intermedia*, and the backbone chromosomal region is the same, but the protein of both other species is larger in size and shows no convincing homology (72 % and 60 % identity on protein level, respectively). Pertactin like proteins are known to play a role in adhesion and virulence, but the actual role of these proteins in O:3/4 has to be investigated.

The classical cytolethal distending toxin (CDT) of *E. coli* arrests mammalian cells in either G2 or early M phase (Peres *et al.* 1997). Normally, three subunits (ABC) are needed for its activity. The potential CDT protein found in serobiotype O:3/4 strain Y11 (Y11_06011) is about 300 amino acids in size and is present also in the draft genome of *Y. enterocolitica* O:5,27/3, but absent from other members of the *Yersinia* genus. However, only a single gene with homology to the subunit B of the classical CDT was found in strain Y11. Since a cluster of at least three subunits is needed for functionality, the putative CDT protein in Y11 is unlikely to be biologically active.

A member of cell-wall associated hydrolases of about 422 amino acids is located in a more than 2 kb large genomic region absent from strain 8081 (Y11_03361, a putative invasion-associated protein). In *Y. mollaretii*, there is a similar organized cluster with a smaller homologous protein of 357 amino acids. The two proteins both share homologies to cell-wall associated hydrolases, but have only limited homology to each other. The genes are located between YeiH and YeiE, a potential membrane protein and an adjacent transcriptional regulator, but the function of these proteins is also unknown. Therefore, a connection between the Yei proteins and the putative cell-wall associated hydrolase is just speculative and its role in pathogenesis remains unclear.

A 3,075 bp putative invasin precursor gene found in serobiotype O:3/4 (Y11_38661) is absent in *Y. enterocolitica* serobiotype O:8/1B, but present in *Y. enterocolitica* serobiotype O:5,27/3, *Y. pseudotuberculosis* and *Y aldovae*, and (as a more dissimilar one) in *Y. pestis*. Motif search revealed homology to Ig-like domains and similar domain structures all over the 1,025 amino acid protein that are typical for invasin or intimin proteins, including domains that refer to membrane integration characteristics. The gene in O:3/4 could therefore be a bacterial surface protein and take part in the invasion or adhesion process during host infection.

A *hlyD* like gene has been found in *Y. enterocolitica* O:3/4 (Y11_09551), together with two adjacent genes that are all absent from O:8/1B. These genes could form a regulated transport cluster, involved in the transport of HlyA. Proteins of the HlyA family are typically secreted into the medium by a mechanism that does not require a cleaved N-terminal signal sequence for

processing (Felmlee *et al.* 1985b;Felmlee *et al.* 1985a). They require two or more proteins for their secretion across the membrane. These secretion proteins include members of the ABC transporter family that are found as adjacent genes to Y11_09551 in strain Y11.

5.5.4.2. HasA diversity

The hemophore like *has* gene cluster is present in all species of the genus *Yersinia*. It has been described as homolog to the hemophore-dependent heme acquisition system (Has system) of *Serratia marcescens* (Rossi *et al.* 2001). This gene locus consists of the *hasR* receptor gene, the *hasA* hemophore gene and genes encoding an ABC transporter as well as a *tonB* homologue. It has been shown that the *has* promoter is Fur regulated and has increased activity at 37°C. Nevertheless, no contribution of the system to heme uptake was observed for *Y. pestis*. Interestingly, the copy number of *hasA* genes differs among the different species of the *Yersinia* genus (five - in *Y. mollaretii*, four - in *Y. enterocolitica* strain 8081, three - in *Y. intermedia*, and two - in *Y. bercovieri* and *Y. frederiksenii*), and though the homology of the transport and receptor genes is high, the HasA proteins show weak homology (also on DNA level) between *Y. enterocolitica* O:3/4 and O:8/1B. In *Y. enterocolitica* subsp. *palearctica* O:3/4, different HasA protein length and numbers were observed. In strains Y8265 and Y5307 three HasA homologs were found, whereas strain Y11 carries four HasA proteins (Y11_28011 - Y11_28041). The differences are due to a stretch of 147 amino acids (441 base pairs) missing in strain Y8265. This prolonged sequence in strain Y11 harbours a stop codon for the first HasA protein and a start codon where the second protein translation begins, whereas in strain Y8265 only a single protein is encoded.

5.5.4.3. Insecticidal toxin cluster

An insecticidal toxin cluster (Tc, Y11_26921-Y11_27061) is found in *Y. enterocolitica* strain Y11 downstream of *tldD*, Y11_27071. This cluster was initially described in serobiotype O:9/2 strain W22703 and has been identified afore by PCR and Southern hybridisation also in *Y. enterocolitica* strains of serobiotype O:3/4 (Bresolin *et al.* 2006). The insecticidal toxin cluster comprises in general four proteins, Tca, Tcb, Tcc and Tcd and is represented in *Y. enterocolitica* serobiotype O:9/2 strain W22703 by TcaA, TcaB, TcaC and TccC proteins. It could be shown that the toxins are active at temperatures below 30°C; and best expression levels were found at 10°C. Lysates of the serobiotype O:9/2 bacteria could kill *M. sexta* within 5 days (Fuchs *et al.* 2008). Interestingly, TcaB is splitted into two proteins in serobiotype O:9/2, TcaB1 and TcaB2, and *tcaC* is interrupted by a stop codon at pos. 314. Still the system could be proven to be functional in *Y. enterocolitica* serobiotype O:9/2 strains. In addition, in *Y. pestis* strain CO92 the *tcaB* is inactivated by a frame shift mutation, but the remaining Tc homologs remain functional (Gendlina

et al. 2007). Therefore, this cluster seems likely to undergo frequent genetic rearrangements and mutations without disturbing the overall functionality of this cluster. The insecticidal toxin cluster found in *Y. enterocolitica* strains of serobiotype O:3/4 also differs in length and gene number (see Figure 5.15). One of the two putative *tcaA* regulatory genes is splitted into two smaller ones, and TcaA (Y11_26951) is a truncated version of the protein of *Y. enterocolitica* serobiotype O:9/2 strain W22703. Interestingly, in serobiotype O:3/4 strains Y8265 and Y5307 TcaA is not truncated but prolonged compared to strain W22703, and the proteins are less homologous in the C-terminus. TcaB (Y11_26971) is in contrast to strains of O:9/2 one single protein, and the *tcaC'(1)* gene (Y11_26981) undergoes a frame shift at position 3498, not at position 314 as in denoted genes of *Y. enterocolitica* of serobiotype O:9/2. Therefore, the remaining part of TcaC'(2) (Y11_26991) is truncated compared to *Y. enterocolitica* serobiotype O:9/2. In *Y. pseudotuberculosis* and *Y. pestis*, a single gene codes for the respective TcaC proteins. TccC of strain Y11 (Y11_27041) is similar to that of serobiotype O:9/2 strains, whereas a second homolog of this protein is present in *Y. pestis*. Automated annotation also identified further putative ORFs, which could be part of this Tc cluster in serobiotype O:3/4 strain Y11.

Y. enterocolitica strain Y5,25 of serobiotype O:5,27/3 also harbours an insecticidal toxin complex which is closely related to the ones described for serobiotypes O:3/4 strains and O:9/2 strain W22703. The cluster in O:5,27/3 harbours a single *tcaC* gene (see Figure 5.15). In addition, *tcaR2* and *tcaB* genes are complete and not fragmented as in serobiotypes O:3/4 and O:9/2. So it seems likely that this toxin cluster represents an uninterrupted prototype version with the full size of all proteins.

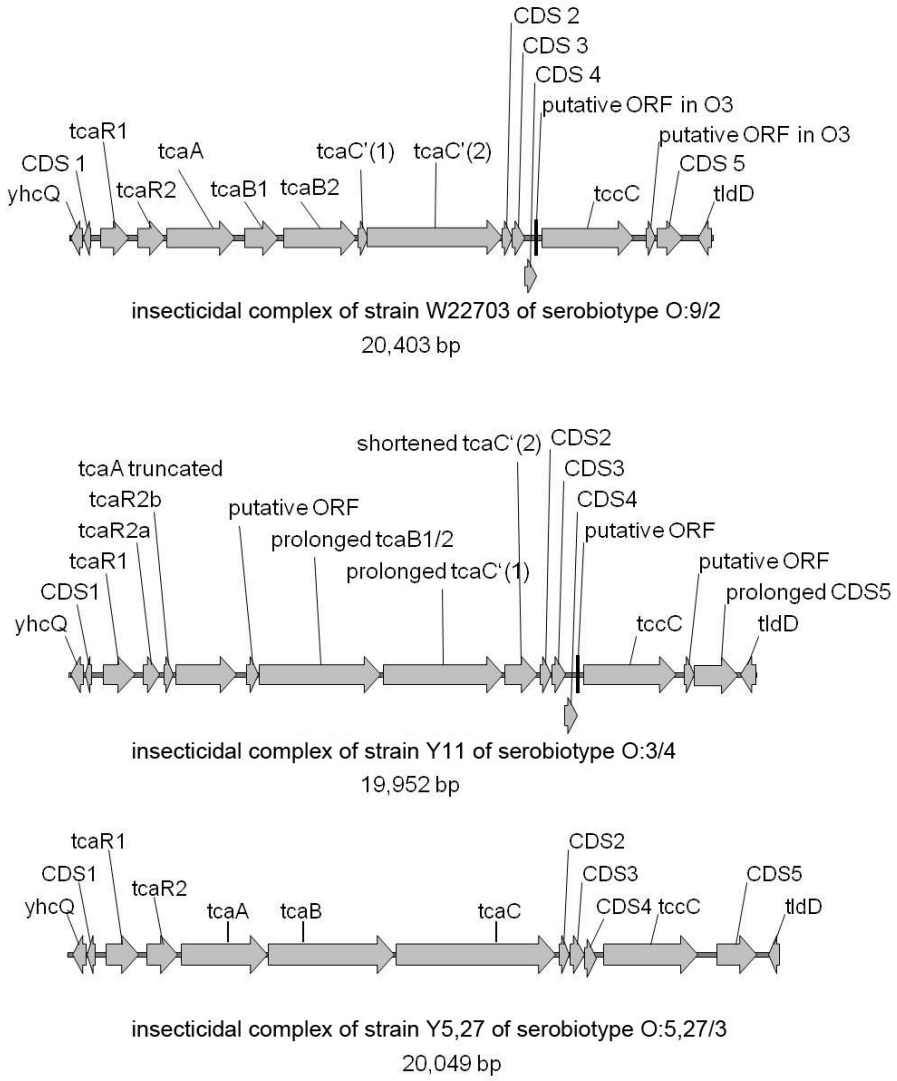


Figure 5.15 Comparison of the insecticidal complex gene clusters of strain W22703 of serobiotype O:9/2, strain Y11 of serobiotype O:3/4 and strain Y5,27 of serobiotype O:5,27/3. The genes of these Tc clusters differ in size and number. Interestingly, the serobiotype O:5,27/3 toxin cluster may represent an uninterrupted prototype version with the full size of all proteins, since this cluster is the only one with single, uninterrupted *tc* genes.

Unassociated with this insecticidal gene cluster, another putative protein encoding ORF (Y11_05031) is present in strains of *Y. enterocolitica* serobiotypes O:3/4 and O:5,27/3, but not in those of serobiotype O:8/1B. This protein is slightly similar to a homologous gene product of the same length in *Y. pseudotuberculosis*. It has high homology to the insecticidal TccC2 and TccC3 proteins of *Y. enterocolitica* strains of serobiotype O:9/2. PCR analysis revealed the presence of *tccC*2 and *tccC3* homologs in all biotype 2–5 strains tested, and absence in most representatives of biotypes 1A and 1B (Fuchs *et al.* 2008). A frame shift seems to have splitted one *tccC* homolog into two ORFs. The *tccC* genes located outside of the insecticidal gene cluster might be acquired

by a further horizontal gene transfer event following the separation of *Y. pseudotuberculosis* and *Y. enterocolitica*, since all genomes of the *Y. pseudotuberculosis* and *Y. pestis* sublines analysed by Fuchs and colleagues share two *tccC* insertion sites.

5.5.4.4. Toxin-antitoxin systems

Several toxin-antitoxin (TA) systems are found in *Y. enterocolitica* O:3/4 that are absent or different from those that can be found in strain 8081 of serobiotype O:8/1B. Classical toxin-antitoxin systems are plasmid- or chromosome-encoded protein complexes composed of a stable toxin and a short-living inhibitor (anti-toxin) of the toxin. Current hypotheses propose that TA systems are involved in the stress management of bacteria. They act either by promoting the altruistic death of a larger fraction of the population (programmed cell death hypothesis) or by inducing a dormant stage (stasis) (reviewed in Van Melderen 2010). In addition, chromosomally encoded TA systems could serve as a protection against the invasion of mobile genetic elements. However, Van Melderen alludes that chromosomally encoded TA systems might lose their addictive properties, when selective pressure disappears, which could be a first step of degeneration (Van Melderen 2010).

A toxin-antitoxin system annotated as YgiT- (antitoxin, Y11_40161) and YgiU- (toxin, Y11_40151) like proteins was found in strains Y11 and Y5,27 of serobiotypes O:3/4 and O:5,27/3, respectively, as well as in *Y. pseudotuberculosis* and *Y. pestis*. In *Y. pestis* strain CO92, the two genes are not linked to the two phage genes that were detected downstream of the TA system in strain Y11. Nevertheless, it could be a prophage-associated gene cluster. Kasari and colleagues (Kasari *et al.* 2010) reported that the protein YgiU (also known as MqsR) with multiple TA-like features inhibits growth and induces rapid shutdown of protein synthesis *in vivo*. YgiT is the antitoxin, which protects cells from the effects of MqsR. They constitute a single operon which is transcriptionally repressed by YgiT. It has been shown that transcription occurs in response to HipA activation (Kasari *et al.* 2010). The observation that HipA activation leads to induction of the *mqsR-ygiT* system is not the only example of sequential or simultaneous activation of different TA systems. In strain Y11, there are three different HipA annotated CDS found (Y11_27061, Y11_27481, Y11_28521). However, a HipB homolog is missing in this strain.

Directly next to a haemolysin gene cluster (Y11_31011) another TA cluster in the genome of strain Y11 was found, annotated as YfjZ (antitoxin, Y11_30951) and YpjF (toxin, Y11_30941). It is absent from strain 8081 (serobiotype O:8/1B) and *Y. pestis*, but present in strain Y5,27 (serobiotype O:5,27/3) and *Y. pseudotuberculosis* strains IP 32953 and YPIII. However, the TA system in strain Y11 and strain Y5,27 is truncated compared to the strains of

Y. pseudotuberculosis. Interestingly, the TA system of strain Y11 is located in a region subjected to frequent chromosomal rearrangements which is evident by comparing strain Y11 with the genome sequences of strain 8081. This TA system has been described as a putative part of a defective prophage with unknown function (Garcia-Contreras *et al.* 2008).

5.5.4.5. Flagellar genes, beta-fimbrial genes and other fimbriae

5.5.4.5.1. Flag-2 flagellar cluster

A large flagellar cluster (Y11_24071-24371) present in strain Y11 is similar to the *flag-2* gene cluster of *Y. enterocolitica* strain W22703 of serobiotype O:9/2 (Bresolin *et al.* 2008). The *flag-2* cluster of serobiotype O:9/2 comprises 12 ORFs (11.2 kb) and is thus much smaller than the large flagellar gene cluster found in strain Y11 of serobiotype O:3/4 (22 kb) (see Figure 5.16 and Figure 5.17). The strain Y11 specific cluster is present also in *Y. enterocolitica* strain Y5,27, in *Y. pestis* strain CO92 and *Y. pseudotuberculosis* strain IP 32953 (where it is found homologous to a less extent, but with similar genetic organisations). Figure 5.18 shows the genetic map of several *flag-2* genes of different species: In the first line, several annotated genes of the strain Y11 *flag-2* gene cluster are shown. The last two organisms in this window are *Y. pestis* strains that share the same genetically organisation of this flagellar cluster. Only the *flaA* genes (light green, right edge of the window) are present in three copies in the *Y. pestis* strain CO92. Only adjacent genes of this flagellar cluster are homologous to *Y. enterocolitica* O:8/1B genes (see Figure 5.19). In the Mauve comparison viewer, coloured boxes indicate strong homologies between the genetic regions, whereas white shades within a coloured box indicate low homology regions. Regions that are not boxed are detected as unique or weakly homologous. The velvet and pink box in the first line belongs to strain Y11, representing the *flag-2* gene cluster. The second line boxes belong to *Y. pestis* strain CO92, and the last line shows genomic regions of strain 8081. Thin lines between the shaded boxes connect homologous regions, which is the case for strain CO92, but not strain 8081. Even though, the homology in this genetic region is rather weak, and not even the complete *flag-2* cluster is present (indicated by shaded boxes: homologous regions; not shaded or not marked by boxes: regions that are not homologous, see Figure 5.18) in the *Y. pestis* strain CO92. For strain 8081, only slight homologies are reported (pink shaded box), that are referred to the adjacent genes of the *flag-2* in strain Y11.

The functionality and role of this additional flagellar gene cluster in the pathogenesis of serobiotype O:3/4 are questionable, since different experimental observations indicate a weak motility for this strain *in vitro*. However, experiments will have to be done to unravel the impact of this cluster on O:3/4 motility and pathogenesis *in vitro* and *in vivo*.

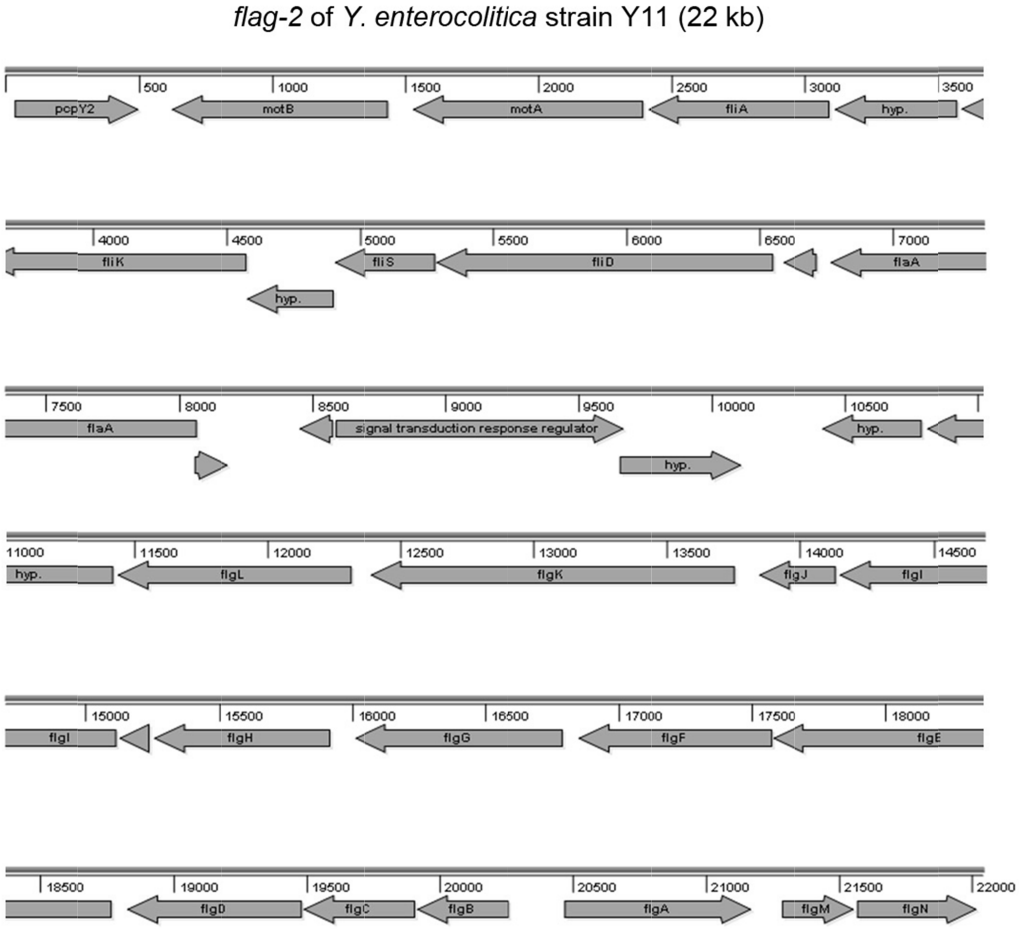


Figure 5.16 Map of the second flagellar system (22 kb) in strain Y11 (O:3/4) absent from serobiotype O:8/1B strain 8081.

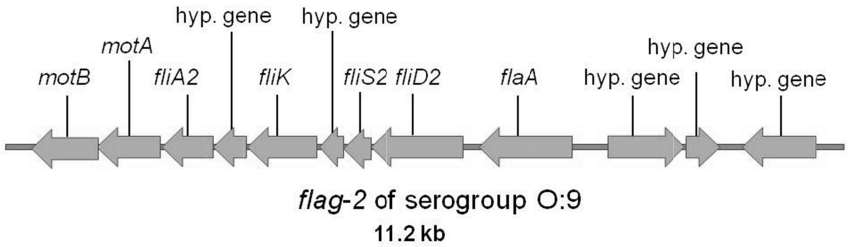


Figure 5.17 Map of the second flagellar system in serobiotype O:9/2. The *flag-2* cluster of O:9/2 comprises 12 ORFs (11.2 kb) and is thus much smaller than the large flagellar gene cluster found in serobiotype O:3/4 (about 22 kb, see Figure 5.16)

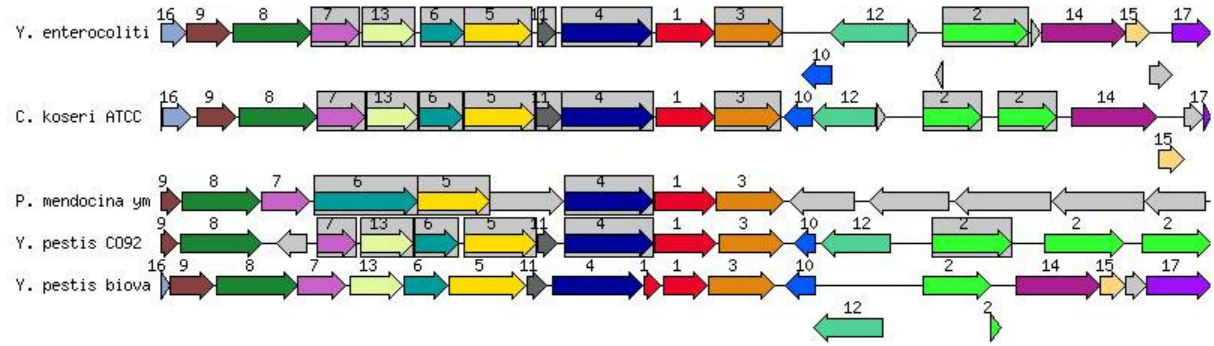


Figure 5.18 Screenshot of the SEED comparison window for parts of the *flag-2* gene cluster of different species: In the first line, several annotated genes of the strain Y11 *flag-2* gene cluster are shown. In the second line, similar organised genes of *Citrobacter* (*C.*) *koseri* strain ATCC BAA-895 are shown, in the third line genes of *Pseudomonas* (*P.*) *mendocina* strain ymp. The last two organisms in this screenshot are *Yersinia pestis* strains CO92 and biovar Medievalis strain 91001 that share a similar genetically organisation of this flagellar cluster. Only the *flaA* genes (light green, right edge of the window) are present in two or three copies in the *Y. pestis* strains.

Figure 5.19 Screenshot of the Mauve comparison window depicting the *flag-2* gene cluster of strain *Y. enterocolitica* strain Y11, *Y. pestis* strain CO92 and *Y. enterocolitica* strain 8081. Coloured boxes indicate strong homologies between the genetic regions, whereas white shades within a coloured box indicate low homology regions. Regions that are not boxed are detected as unique or weakly homologous. The velvet and pink box in the first line represents the *flag-2* gene cluster of strain Y11. The second line boxes belong to *Y. pestis* strain CO92, and the last line shows genomic regions of strain 8081. Thin lines between the shaded boxes connect homologous regions, which is the case for strain CO92, but not for strain 8081.

5.5.4.5.2. Beta-fimbrial genes and other fimbriae

Fimbrial genes are nonflagellar bacterial filaments, which have been classified earlier based on their morphology and functional characteristics. Today they are classified based on their assembly mechanism, for example conjugative fertility fimbriae (F pili), type IV fimbriae (as toxin-coregulated pili), curli (assembled by the extracellular nucleation/precipitation pathway), and fimbriae that are assembled by the chaperone/usher-dependent pathway. The last group has been divided into six major phylogenetic clades, α-, β-, γ-, κ-, π-, and σ-fimbriae, which are based on the fimbrial usher protein phylogenetic relationships. Operons of the class of β-fimbriae do not contain genes that resemble typical tip adhesins, thus assumed to form nonfimbrial or fibrillar structures. They were so far identified by whole-genome sequencing approaches and remain uncharacterized (reviewed in Nuccio and Baumler 2007). In the strain Y11 genome, a cluster of three putative beta-fimbriae major subunits with both a chaperone and usher protein (Y11_14931-14971) was found. These genes and their protein sequences are found in other *Yersinia*, including also (with lower homology) the other human pathogenic species *Y. pseudotuberculosis* and *Y. pestis*, but the number of genes, orientation and structure of the clusters are different in all strains compared. The cluster is completely absent from strain 8081 of serobiotype O:8/1B. However, in all strains compared, the genes encoding a chaperone and an usher are preceded by at least one open reading frame that encodes a protein with a domain of unknown function, probably a fimbrial subunit.

Another cluster of four fimbriae-associated genes is found in strain Y11, encoding a fimbrial gene, a chaperone, an usher protein and *fimA* (Y11_26051-26081). This cluster is identical to those found in *Y. mollaretii*, *Y. intermedia* and others, but the genes are absent from *Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica* strain 8081 (O:8/1B). The cluster is conserved in protein length and cluster organisation in those strains, but the function is still unknown.

5.5.4.6. ABC transporter genes aatPABCD

Y. enterocolitica strain Y11 harbours a putative ABC transporter complex, comprising an interrupted cluster of four genes. In *Citrobacter* and *E. coli* five genes are reported, designated *aatPABCD*. The cluster includes a homolog of an inner-membrane permease (AatP), an ATP-binding cassette (AatC) and an outer membrane protein (AatA). The *aat* cluster in enteroaggregative *E. coli* (EAEC) encodes a specialised ABC transporter, which plays a role in virulence of these pathogens by transporting dispersin out of the bacterial cell (Nishi *et al.* 2003). In strain Y11, one gene similar to both *aatB* and *aatA* was found (Y11_24431, about 90% identity on protein level compared to *Citrobacter rodentium* strain ICC168 proteins AatA and AatB), indicating a possible bifunctional enzyme. The first gene of the cluster, *aatD* (Y11_24511), is

separated from the other ones by several small hypothetical genes, *gadW* and a transposase gene, and the last gene of the cluster, similar to *aatC* (Y11_24421), is also flanked by several transposase genes. The functionality of the interspaced cluster is yet uncharacterised, but a potential role in pathogenesis as for the EAEC can be considered.

5.5.4.7. Chromosomally encoded type three secretion system (T3SS)

The virulence-associated chromosomal Ysa-T3SS is absent in *Y. enterocolitica* strain Y11. Instead, strain Y11 harbours an alternative, more than 24 kb spanning T3SS (designated Ysp-T3SS, Y11_35171-35491, see Figure 5.20), which is found also present in strain Y5,27 of serobiotype O:5,27/3 and with more dissimilar proteins and a modified cluster organization in *Y. pseudotuberculosis* IP 32953 and *Y. pestis* strain KIM. Protein based sequence comparison reveals that at least two third of the proteins are homologous to the clusters in *Y. pseudotuberculosis* IP 32953 and *Y. pestis* strain KIM. Best homologs are found within *Y. frederiksenii* strain ATCC 33641, where the complete cluster is present, except for few differences: The putative gene *yscT* (*ssaT*) is interrupted in strain Y11 into two smaller ORFs, whereas strain ATCC 33641 harbours the complete gene. In addition, *ssaL* (858 bp) is missing in strain Y11 but present in strain ATCC 33641. Three putative ORFs annotated for strain Y11 are missing in strain ATCC 33641. Interestingly, strain Y5,27 of serobiotype O:5,27/3 shares similarities with both strains – the *ssaL* gene is present in this strain and *ssaT* is a single full length gene, likewise in *Y. frederiksenii* strain ATCC 33641. But all three putative genes not detected in strain ATCC 33641 are present in this serobiotype.

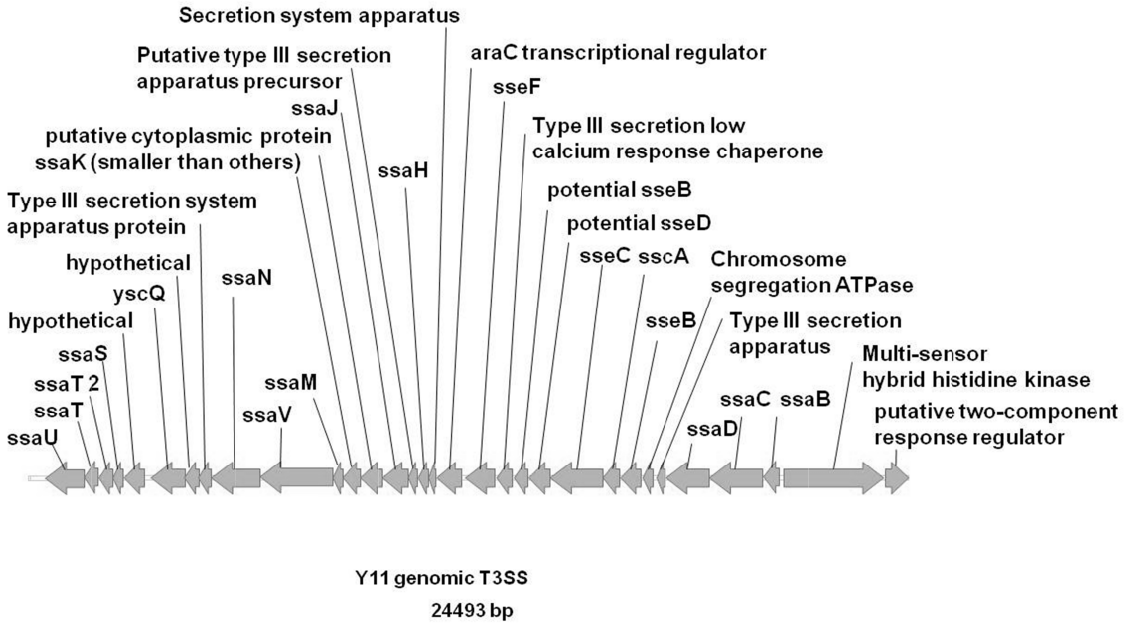


Figure 5.20 The Ysp-T3SS of strain Y11 (O:3/4). The Ysp-T3SS is homologous to the *Salmonella* SPI2-T3SS (Waterman and Holden 2003). Even though a large part of the SPI2 system is represented in the Ysp-T3SS, still some functional parts are missing in strain Y11, including the SsaG and SsaH proteins of the needle and SsaR from the inner membrane ring. In addition, no genes for effector proteins are directly linked to this secretion system gene cluster.

In strain 8081 (O:8/1B), the strain Y11-like Ysp-T3SS system is absent. Instead, the Ysp-T3SS in strain Y11 is located in the same genetic background when compared to strain 8081. The Ysp-T3SS shows some similarity to the *Salmonella* SPI2-T3SS (Waterman and Holden 2003). When the predicted proteins of the Ysp-T3SS were compared to proteins of this SPI2 system from *Salmonella*, a large part of the SPI2 system is represented in Ysp-T3SS, but still some functional parts are missing in strain Y11. Even though the translocon of the apparatus (SseBCD) is present, the SsaG and SsaH proteins of the needle are missing, as well as SsaR from the inner membrane ring. In addition, there are no effector proteins directly linked to this secretion system gene cluster. Thus, the characterisation of this SPI2 like T3SS awaits the identification of putative effectors and their downstream function for pathogenicity or for cell-cell interactions.

5.5.4.8. LPS biosynthesis

A specific glycosyltransferase-cluster is present in strain (O:3/4). A much larger glycosyltransferase-cluster is located in *Y. enterocolitica* strain 8081 (O:8/1B) in the same genomic background. These clusters encode major enzymes for the O-antigen in LPS biosynthesis of these strains (Skurnik 2003;Skurnik and Bengoechea 2003). The cluster of *Y. enterocolitica* O:8/1B is nearly twice in size compared to strain Y11, due to the different structure of the O-

antigen. However, in strain Y11, two separate clusters are present; one representing genes for the outer core gene cluster (the above one, OC, Y11_19901-20011) and another for the O-antigen encoding genes (Y11_16711-16781).

5.5.4.9. Antibiotic resistance genes

Two unrelated potential antibiotic resistance genes (chloramphenicol acetyltransferase genes, Y11_09991 with 666 bp respective Y11_12731 with 645 bp) were found in the genome of *Y. enterocolitica* O:3/4, of which only the first one (chloramphenicol transferase 1, Y11_09991) has homologous genes in each *Y. enterocolitica* strain Y5,27 (O:5,27/3), *Y. intermedia* and *Y. kristensenii*. In *Y. enterocolitica* serobiotype O:5,27/3 the gene has a stronger homology to a streptogramin A acetyltransferase, whereas in *Y. intermedia* a virginiamycin acetyltransferase is annotated. It has been reported by Seoane and Lobo (Seoane and Garcia Lobo 2000) that this gene would be present in all *Y. enterocolitica* genomes. However, a homolog of this gene was not found in *Y. enterocolitica* strain 8081 (O:8/1B). The proteins of *Y. enterocolitica* strain Y11 (O:3/4) belong to the family of xenobiotic acyltransferases (XAT): The XAT class of hexapeptide acyltransferases consists of a large number of microbial enzymes that are able to catalyze the CoA-dependent acetylation of a variety of hydroxyl-bearing acceptors such as chloramphenicol and streptogramin. The protein length in both proteins is comparable to various proteins in NCBI, including best matches to *E. coli* (for Y11_12731) and *Serratia* proteins. BLASTp elicits 65 % identity for the *E. coli* hexapeptide repeat-containing transferase (84 % positives) compared with chloramphenicol transferase 1. Members of this class of enzymes contain repeated copies of a six-residue hexapeptide repeat sequence motif (X-[STAV]-X-[LIV]-[GAED]-X) and adopt a left-handed parallel beta helix (LbH) structure. Four motifs were found in common in both the *E. coli* strain EC4024 hexapeptide repeat-containing transferase and *Y. enterocolitica* strain Y11 chloramphenicol transferase 2 (Y11_12731) and one motif was found to occur at the same location but different between the two proteins. For *E. coli* strain EC4024, one additional hexapeptide motif was detected, which was absent from O:3/4 (see Figure 5.21). For the strain Y11 chloramphenicol transferase 1 (Y11_09991), also 5 motifs were indicated. The active acyltransferase is reported to be a trimer with CoA and substrate binding sites at the interface of two separate LbH subunits. It has to be elucidated, if these proteins are functional in strain Y11 and against which particular xenobiotic they are active.

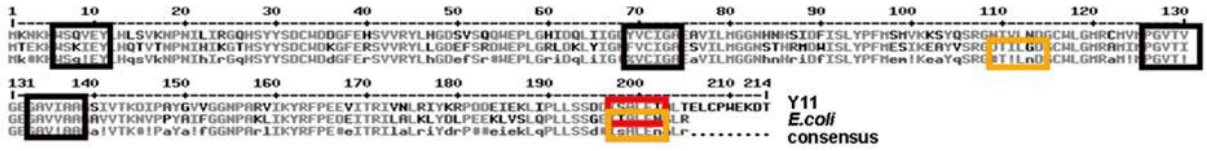


Figure 5.21 Alignment between the hexapeptide repeat-containing transferase of *E. coli* strain EC4024 and chloramphenicol transferases of *Y. enterocolitica* strain Y11 (O:3/4) (Y11_12731). Black boxes mark the hexapeptide repeat sequence motifs (X-[STAV]-X-[LIV]-[GAED]-X) shared by both bacterial species, orange boxes indicate motifs only present in *E. coli* and the red box surrounds the motif only found in *Yersinia enterocolitica* strain Y11. This sequence motif is conserved in the class of XAT acyltransferases in normally six copies, but one motif is missing in strain Y11.

Methylenomycin can be produced as an antibiotic compound (together with resistance genes) by *Streptomyces*. Two adjacent proteins with similarity to methylenomycin A resistance proteins (Mmr proteins) were found in *Y. enterocolitica* strain Y11 (Y11_09761 and Y11_09771). However, they are rather short (38 and 83 amino acids, respectively, likewise found in the draft genomes of strain Y5307 and Y8265) and show no similarity to any compared genome. Nevertheless, in *Y. enterocolitica* strain 8081 a large protein of 492 amino acids with the same predicted function exists at the same position. There are phage related genes located downstream of these genes, suggesting that the loss of a potential larger and functional protein could be due to insertions and recombination events. In the downstream region in *Y. enterocolitica* strain 8081, plasmid maintenance/toxin genes are located and a putative LuxR-family transcriptional regulatory protein next to a tetracycline-resistance determinant (*tetV*).

The genes for the two different β-lactamases of strain 8081(*ampC* and *blaA*) (Bent and Young 2010) can be both found in strain Y11 (with protein similarities about 98 % and 94 %, respectively). Interestingly, the upstream regions of the *blaA* gene in strain Y11 and strain 8081 are less homologous. Since strain Y11 is less susceptible to β-lactam antibiotics, the promoter region of strain Y11 may constitute a different regulation of *blaA* gene expression. In addition, the *blaA* gene has been shown to occur in multiple copies (Seoane *et al.* 2003), which might also take place in the different susceptibility to β-lactam antibiotics of both strains (see also 5.5.6.3).

5.5.4.10. Carbohydrate uptake

N-acetyl-D-galactosamine and N-acetyl-D-glucosamine are amino-sugar components of the intestinal mucin in pigs and humans, as well as of the peptidoglycan of several bacteria. N-acetyl-D-galactosamine is, together with fucose, a major terminal sugar of the porcine gastric mucin. The amount of N-acetyl-D-galactosamine is nearly twice that of any of the other sugars present in the pig's small intestinal mucin and higher than the amount in the human or rat mucin (Mantle and

Allen 1981). In the human mucin, N-acetyl-D-glucosamine is the major amino-sugar component (Forstner *et al.* 1979). The composition and modification of intestinal mucin is a critical defence mechanism in the prevention against pathogenic bacteria in the intestine. These amino-sugars are therefore an attractive source of carbon with potential connections to intestinal pathogenesis for *Y. enterocolitica* O:3/4. Enteric bacteria have been reported to differ in their ability to grow on N-acetyl-D-galactosamine (GalNAc or Aga) and on D-galactosamine (GalN or Gam). N-acetyl-D-galactosamine utilized in *Yersinia* as a carbon source can be taken up by a specific phosphoenolpyruvate-dependent phosphotransferase systems (PTS system). The enzymes that build this PTS transport complex are AgaVWEF, with their respective genes clustered in an operon in the genome of *Y. enterocolitica* strain Y11 (O:3/4) (named *aga*-operon, see Figure 5.22). The Y11 specific N-acetyl-D-galactosamine PTS system is completely absent in *Y. enterocolitica* strain 8081 (O:8/1B).

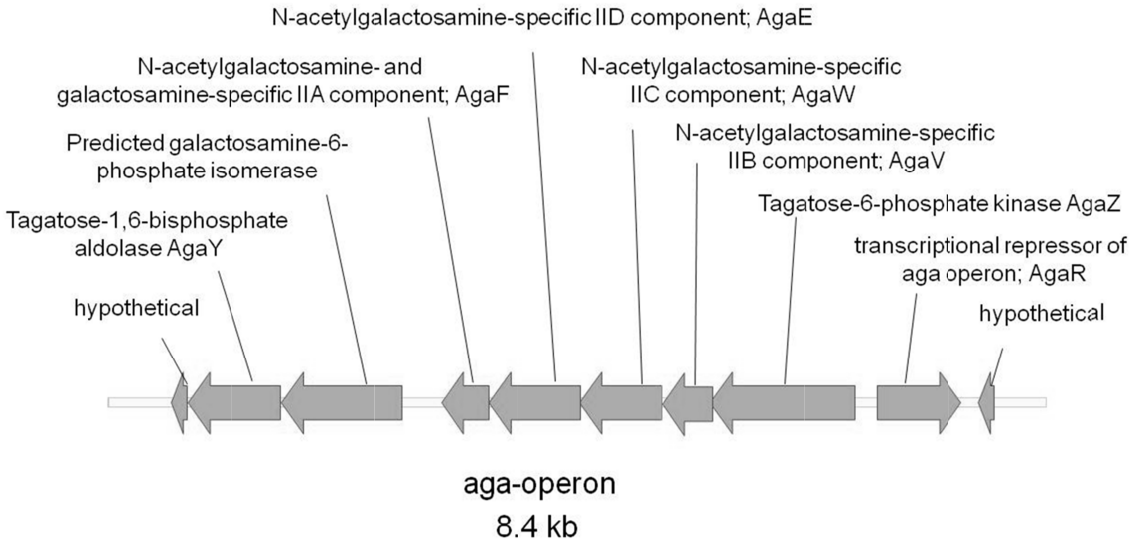


Figure 5.22 Genetic map of the N-acetyl-D-galactosamine phosphoenolpyruvate-dependent phosphotransferase system (PTS) (also named *aga*-operon) found in *Y. enterocolitica* serobiotype strain Y11 of O:3/4. The enzymes that build the PTS transport complex are AgaVWEF.

5.5.4.11. Other metabolic differences

Y. enterocolitica strain Y11 (O:3/4) harbours a complete cluster for the cytochrome C maturation required for NrfA activity (a formate-dependent nitrite reductase). This *nrf*-like cluster is weakly similar to the *ccm* cluster (the system I of cytochrome C biogenesis, known as operon *ccmABCDEFGH* in *E. coli*), which is also present in strain Y11. The *nrf*-like cluster is absent from strain 8081. The *ccm* genes are essential for the maturation of all C-type cytochromes (cytochrome C) in *E. coli*, whereas the *nrf* gene clusters (operon *nrfABCDEFG* (Hussain *et al.*

1994), all genes present in strain Y11, Y11_04801-04871, and strain Y5,27 (O:5,27/3)) are required specifically for the maturation of one specific type of cytochrome C in NrfA that contains an unusual heme-binding motif. The ability to reduce nitrite could therefore differ in the related subspecies and should be compared between the different strains.

Y. enterocolitica strain Y11 and strain 8081 share a urea transporter system. However, a second urea transporter system was detected in *Y. enterocolitica* strain Y11 (Y11_22281-22341), which is independent and different from the first one shared with strain 8081. This second system is also present as an additional cluster in *Y. intermedia* and *Y. frederiksenii*. There is no obvious explanation for the presence of two different urea clusters in strain Y11. However, since the two urea systems are unrelated in their protein composition, they must have been acquired independently.

A *dms* gene cluster (DMSO reductase, Y11_22241-22261) is present in *Y. frederiksenii* strain ATCC 33641, *Y. pestis* strain CO92, *Y. pseudotuberculosis* strain IP32953 and *Y. enterocolitica* strain Y5,27 of serobiotyoe O:5,27/3, but not *Y. enterocolitica* strain 8081 of serobiotype O:8/1B. Many enteric and soil bacteria are able to respire anaerobically by using a variety of amine-N-oxides and methyl-sulfoxides as electron acceptors (Bearson *et al.* 2002). This ability depends on the regulated synthesis of a membrane bound DMSO and/or TMAO (trimethylamine N-oxide) reductase enzyme encoded by the *dmsABC* gene cluster. Next to this cluster, a *torD* homolog is found, which has homologs in *Y. frederiksenii* strain ATCC 33641, *Y. pestis* strain CO92, *Y. pseudotuberculosis* strain IP32953 and *Y. enterocolitica* strain Y5,27, but is absent from *Y. enterocolitica* strain 8081. The reduction of trimethylamine N-oxide in *E. coli* strain MC4100 has been reported to involve the TorCAD complex. The terminal molybdoreductase TorA is located within the periplasm, whereas TorC, the C-type cytochrome, is anchored in the membrane. TorD is a cytoplasmic chaperone which interacts with the unfolded state of TorA (Pommier *et al.* 1998). Interestingly, TorA and TorC are not present in strain Y11 and *Y. frederiksenii* strain ATCC 33641, *Y. pestis* strain CO92, *Y. pseudotuberculosis* strain IP32953 and *Y. enterocolitica* strain Y5,27. Thus, the role of the single *torD*-like gene remains unclear.

Many species of the human intestinal microbiota show bile salt hydrolase (BSH) activity mediated by the choloylglycine hydrolase enzyme (CGH). Whether this enzyme contributes to pathogenicity or fitness of bacteria in the gastrointestinal tract is not known. Studies have shown that the CGH enzyme confers to the ability to resist the antimicrobial action of bile salts (Delpino *et al.* 2007). So the CGH may contribute to the ability of bacteria to infect the host through the oral route. There is one CGH enzyme present in *Y. enterocolitica* strain Y11 (Y11_23571), similar to proteins of *Y. intermedia*, *Y. kristensenii* and *Y. rhodei*. The gene for CGH is absent from strain 8081 and the other human pathogenic species *Y. pestis* and *Y. pseudotuberculosis*, likely reflecting

different niches and/or host infection routes. The activity of the enzyme in strain Y11, however, has to be elucidated.

5.5.5. The pYV virulence associated plasmid

The virulence plasmid pYV is present in all human pathogenic *Yersinia*. It encodes the Ysc - T3SS, with an adjacent set of effector proteins, and the adhesion protein YadA (Y11_p0221 in Y11). The pYV plasmids of strain Y11 and strain 8081 are more divergent than the corresponding genome sequences. In contrast, the pYVe227 plasmid of *Y. enterocolitica* strain W22703 of serobiotype O:9/2 (accession no. AF102990.1) is highly homologous to the plasmid of strain Y11 (see Figure 5.23). This may be due to an independent acquisition or evolution of the pYV plasmid within the different subspecies. For all proteins annotated for the strain W22703 pYVe227, homologous proteins were found in the pYV plasmid of strain Y11. About 200 single nucleotide differences were identified within the annotated genes from which 82 were deletions or insertions. These changes affected the sequence of especially *yadA* (Y11_p0221), which harbours an additional cluster of 15 amino acids in strain Y11 and several minor changes, altering the protein size and sequence. Also the sequence of *arsB* (Y11_p0281) and *arsR* (Y11_p0291) responsible for the arsenic resistance are significantly different between the pYV plasmids of strain Y11 and strain W22703. The gene *arsB* is altered in its sequence, whereas *arsR* is additionally shortened in strain Y11 compared to strain W22703. A substantial difference between the genes found in the pYV plasmids of strain Y11 and strain W22703 is represented by *traX* (Y11_p1031), a gene which encodes an inner membrane protein that acetylates the N-terminal residue of propilin (the precursor of F-pilin subunits). However, the functionality of *traX* in strain Y11 is questionable.

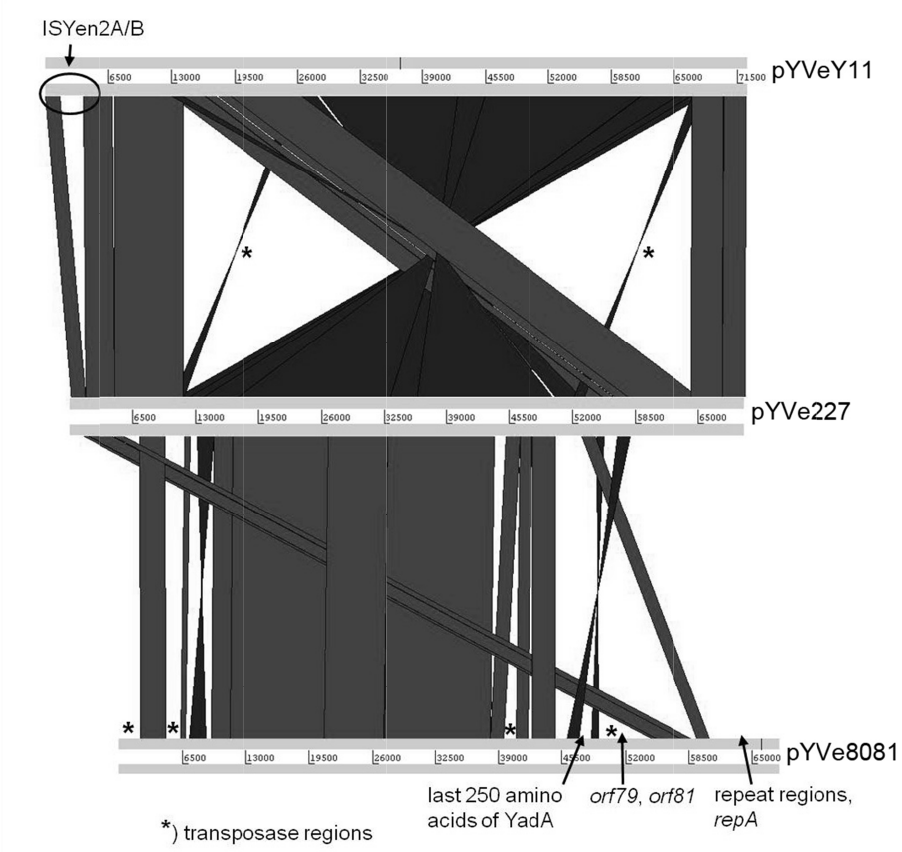


Figure 5.23 Comparison of the pYV plasmids of strain Y11 (pYVeY11, serobiotype O:3/4), strain W22703 (pYVe227, serobiotype O:9/2) and strain 8081 (pYVe8081, serobiotype O:8/1B) using the Artemis Comparison Tool (ACT). Dark grey shaded areas are inverted but homologous, whereas light grey shaded areas are direct homologs. Regions that are not connected between the three pYV plasmid sequences by shaded areas are not homologous.

5.5.6. Mobile elements shaping the Y. enterocolitica subsp. palearctica genome

5.5.6.1. Integrases

Mobile genetic elements (genomic islands, plasmids, prophages, transposons, including insertion sequences, IS) are known to be involved in horizontal gene transfer (HGT). They utilize site-specific integrases for recombination with the core genome and use small tRNA genes as the attachment sites for integration. Thus, the presence of full-length integrases indicates potential recent gene acquisitions while truncated copies are suggestive for more ancient genetic events. In the chromosome of strain Y11 of serobiotype O:3/4, 13 automatically annotated copies of integrase genes were detected (Y11_06541, Y11_12191, Y11_13091, Y11_14181, Y11_15011, Y11_19481, Y11_19641, Y11_23521, Y11_25551, Y11_28701, Y11_29471, Y11_30661 and Y11_42091), whereas the genome of strain 8081 of serobiotype O:8/1B harbours at least 21. However, most of the strain Y11 integrases seem to be truncated and so represent remnants of ancient acquisitions. Two of them were smaller than 50 amino acids in length, denying any

functionality. Presence of intact integrases in turn speaks in favour of their potential activity and impact on genome stability and rearrangements.

5.5.6.2. Genomic islands

Many tRNA sequences that are associated with the P4-like integrases are known to be associated with genomic islands or prophages (5 of the 13 annotated integrase genes are located adjacent to tRNA genes). This was also the case for the *asn* tRNA which has acquired a novel genomic island of unknown function with several ORFans (a humorous term for "orphan" genes). In *Y. enterocolitica* strain Y11 (O:3/4) and strain 8081 (O:8/1B), three different *asn* tRNA loci are found within a narrow region comprising less than 100 genes. Remarkably, one of these *asn* tRNA copies is occupied by the High-pathogenicity Island (HPI) encoding the yersiniabactin iron acquisition system in strain 8081 (Carniel *et al.* 1996). The GIYep-01 island denoted mobile element (Y11_15011-Y11_15121) found within a *asn* tRNA copy in strain Y11, has a GC content of about 40%, lower than that of the core genome sequence (47%). Even though, there is no prediction of a CpG island by the Geneious software tool (see Figure 5.24). The HPI, on the other hand, has a higher GC content than the backbone chromosome (about 58% in the HPI compared to 47% in the backbone genome) and, in contrast to GIYep-01, an inactivated integrase (Bach *et al.* 2000). The novel GIYep-01 genomic island is about 14.9 kb in length and contains up to 11 potential genes besides the integrase (see Figure 5.24). Translated ORFs show homologies to a metallo-β-lactamase domain containing protein (351 amino acids), a SbcC protein (involved in DNA repair), a protease like protein, an antirestriction protein and to transition helper proteins, with the latter ones as typical members of mobile elements. Nested PCR (see section 5.6.1) demonstrated the precise excision of the GIYep-1 with complete restoration of the *attB* attachment site. Moreover, when the integrase and its attachment sites were analysed in two other strains of serobiotype O:3/4, Y8265 and Y5307, both harboured a full-length integrase and the intact attachment sites. This speaks in favour of a recent acquisition of this genetically active mobile element by *Y. enterocolitica* subsp. *palearctica* O:3/4 strains.

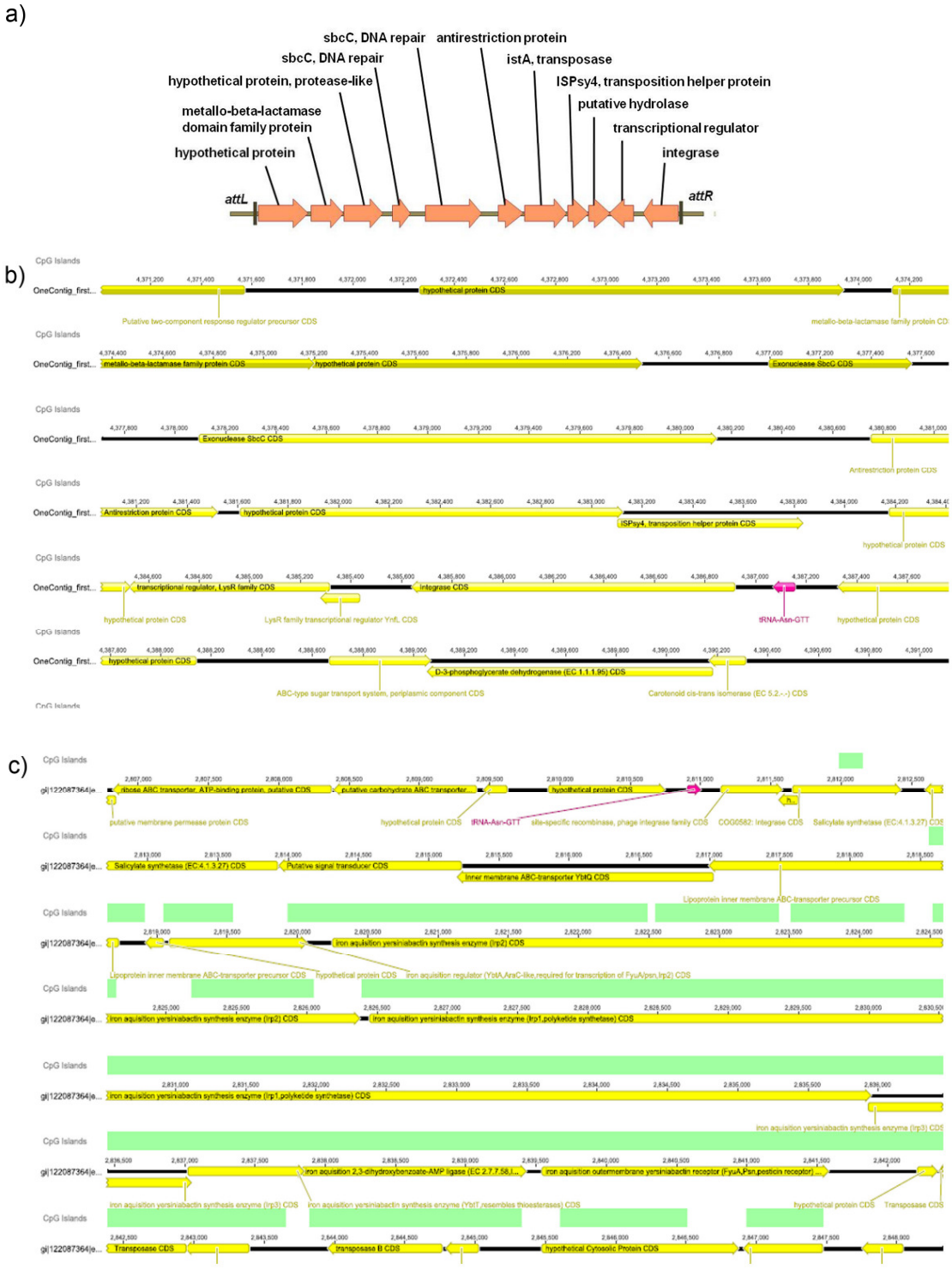


Figure 5.24 Genomic islands integrated into *asn* tRNA in strain Y11 and strain 8081. a) Map of GIYep-01 genes found in *Y. enterocolitica* subsp. *palearctica* strain Y11 (O:3/4). b) Screenshot of the map of GIYep-01 genes (yellow) of strain Y11 and the *asn* tRNA gene integration site (pink) generated by the Geneious software tool (http://www.geneious.com/). CpG islands have been predicted by an additional plugin tool (Geneious, based on the model and parameterisation described by Durbin and colleagues (Durbin 1998)). CpG islands are not predicted for the GIYep-01 of strain Y11. c) Screenshot of the map of parts of the HPI1 genes (yellow) of strain 8081 and the *asn* tRNA gene integration site (pink) generated by the Geneious software tool. Predicted CpG islands are coloured in light green above the respective regions.

5.5.6.3. Prophages in Y. enterocolitica: the dominant genetic acquisitions

A filamentous PhiYep-1 prophage (Y11_09601-09661) that is highly homologous to the Ypf prophage of *Y. pestis* (13 ORFs, YPO2271-YPO2281 (Gonzalez *et al.* 2002)) was found in the genome of strain Y11 (see Figure 5.25). In *Y. pestis* strain CO92 the Ypf genome contains all functional modules needed for its assembly and encodes viable phage particles. Ypf has been found integrated into the genome in tandem repeats that is typical for filamentous phages. It was suggested that Ypf plays an important role in *Y. pestis* virulence (Derbise *et al.* 2007). However, the deletion of the Ypf prophage genome does not affect *Y. pestis* in colonizing and blocking the flea proventriculus, but results in a slight alteration of its pathogenicity in mice (Derbise *et al.* 2007). The Ypf prophage is also similar to the CUS-1 phage described in *E. coli*. The gene encoding the YPO2277 homolog (Y11_09621) is also homologous but increased in size compared to the *puvA* gene in *E. coli*. Gonzalez *et al*. reported a decreased *in vivo* fitness of an *E. coli* K1 *puvA* mutant (Gonzalez *et al.* 2002), suggesting a potential role of this prophage in pathogenicity. In *Y. enterocolitica* strain Y11, the 3,900 bp (comprising the genes YPO2275-9 in *Y. pestis*, see Figure 5.25) PhiYep-1 fragment is 98 % identical to the *Y. pestis* Ypf and 99 % to the CUS-1 prophage of *E. coli*. The regions up- and downstream of these genes in the strain Y11 genome are found similar but less related to each Ypf and CUS-1 (90 % and 98 % identity in about 5 kb, respectively). This prophage is absent from *Y. enterocolitica* strain 8081; even though the infection with this prophage has been demonstrated for all three pathogenic *Yersinia* species (Chouikha *et al.* 2010).

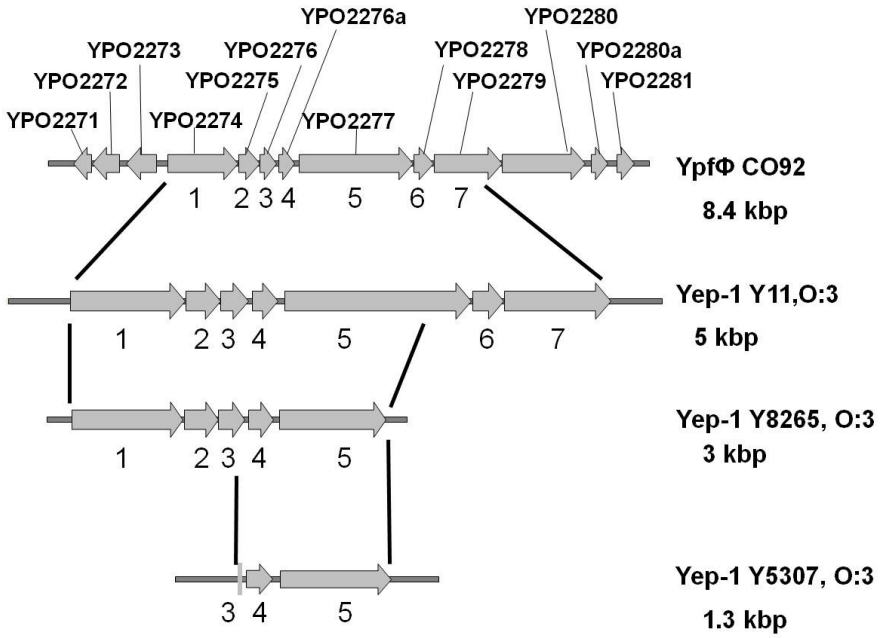


Figure 5.25 The PhiYep-1 prophage of strain Y11 compared to the Ypf prophage of *Y. pestis* strain CO92 and remnants of PhiYep-1 in other serobiotype O:3/4 strains. The prophage sequence has already suffered successive deletions in these isolates.

The PhiYep-1 prophage seems to be more widely disseminated in yersiniae than expected. However, the comparison of PhiYep-1 prophage sequences in the two other serobiotype O:3/4 strains, Y8265 and Y5307, showed that the prophage sequence has already suffered successive deletions in these isolates that might reflect a high recombination activity in this region (see Figure 5.25).

Moreover, the PhiYep-1 prophage constitutes a part of a 28 kb tandem repeat amplified in *Y. enterocolitica* after the elevation of ampicillin levels as described by Seoane and colleagues (Seoane *et al.* 2003). The 28 kb region was shown to harbour prolonged tandem repeats after growth in medium with elevated ampicillin levels. This repeat region harbours the *blaA* gene (encoding a penicillinase) and at least portions of the PhiYep-1 prophage which indicates a link between PhiYep-1 multiplication and elevated ampicillin resistance (see Figure 5.26). It could be proven by PCR that the multiplication of the 28 kb fragment occurs in strain Y11, even in the absence of the ampicillin as a selective pressure. Moreover, when raising the ampicillin concentration from 100 µg/mL to 1000 µg/mL both for strain Y11 and Y5307 that lacks major parts of the PhiYep-1 sequence, the strains were able to grow in 1000 µg/mL ampicillin containing LB media to similar overnight densities. Thus, the tandem multiplication of the

PhiYep-1 prophage together with the nearby *blaA* region seems not to be the main mechanism of rapid ampicillin resistance acquisition in strains of serobiotype O:3/4.

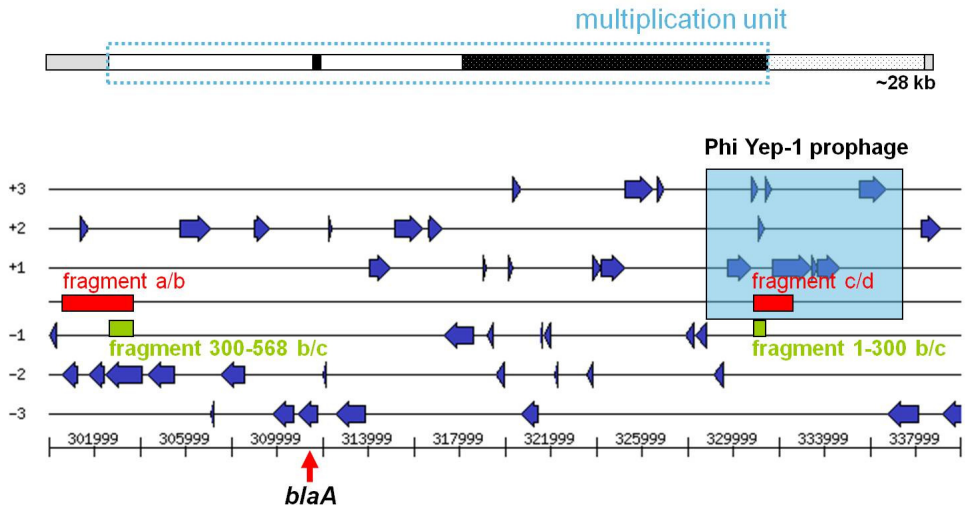


Figure 5.26 The multiplication unit of a part of the PhiYep-1 prophage in strain Y11. The blue dotted line indicates the multiplication unit. Blue box, prophage. PhiYep-1. Red boxes, fragments a/b and c/d that have been generated by PCR to cover the borders of the multiplication unit with the chromosome backbone. Green boxes, parts of fragments b/c found as a single fragment at each border of the serial multiplication unit. Red arrow, localisation of *blaA*.

Two highly homologous P2-like prophages (see Figure 5.27 and Figure 5.28), PhiYep-2 (Y11_25141- Y11_25551) and PhiYep-3 (Y11_13081- Y11_13511), are integrated in two different tRNA genes in strain Y11; PhiYep-2 in *met* tRNA and PhiYep-3 in *leu* tRNA. The phages are highly homologous to the P2-like prophage found in *Y. pseudotuberculosis* IP 32953. In contrast to PhiYep-2, PhiYep-3 contains a full-length integrase and was proven for its ability to leave its attachment site (see section 5.6.1). However, the PhiYep-3 prophage was absent in the two other *Y. enterocolitica* strains, Y8265 and Y5307. Screening of 15 strains of serobiotype O:3/4 for the occupied *leu* tRNA gene (using primers designed for the amplification of the *attL* site, see section 5.6.1) demonstrated the PhiYep-3 presence in nine of the *Y. enterocolitica* O:3/4 strains tested. Thus presence of the PhiYep-3 prophage may serve as an additional epidemiological marker to differentiate serobiotype O:3/4 strains by prophage acquisition events.

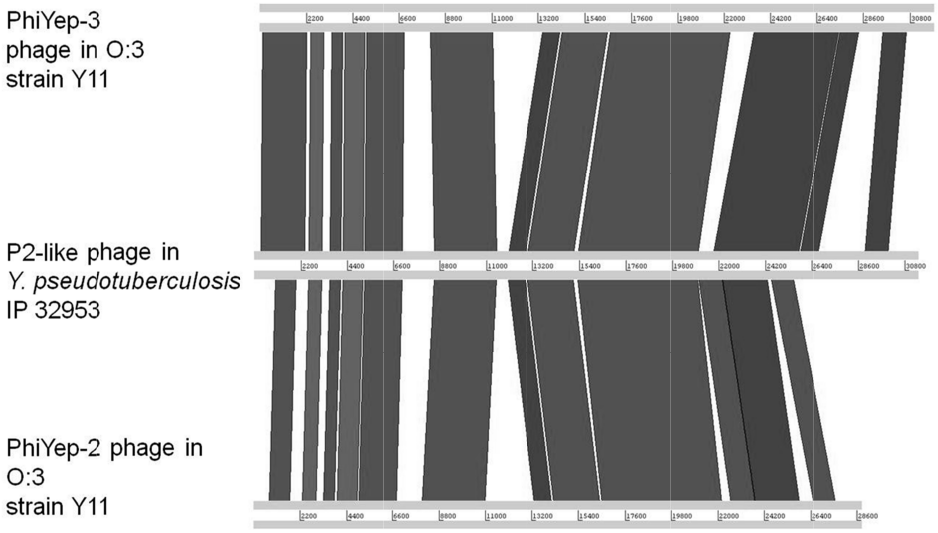


Figure 5.27 PhiYep-2 and PhiYep-3 compared to the P2-like phage of *Y. pseudotuberculosis*. The comparison was generated using ACT. Regions that are not connected to the other phages are unique for the respective strain.

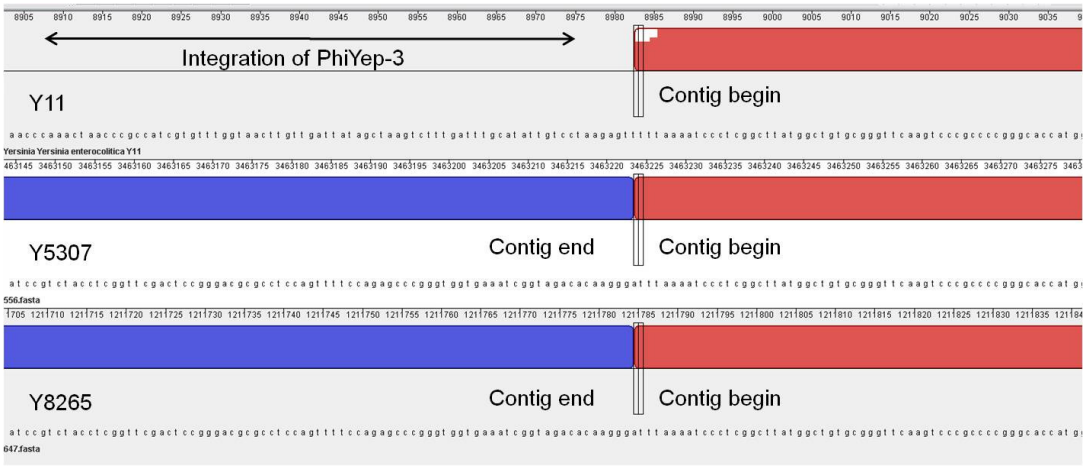


Figure 5.28 Mauve comparison tool illustration screenshot. Homologous regions shown as velvet and red boxes. The screenshot illustrates the regions flanking the PhiYep-3 prophage in strain Y11 (velvet and red, less homologous sequences indicated by white colour) as adjacent regions in the draft genomes of strains Y5307 and Y8265 lacking the prophage and the PhiYep-3 prophage in strain Y11 (lacking homologs and thus showing no coloured homology boxes).

5.5.6.4. IS-elements as footprints of bacterial communication history.

Total IS element copy numbers in the *Yersinia* genus vary between 12 to 1,147 (Chen *et al.* 2010). The number of IS elements discovered in the strain Y11 genome using the ISfinder database search tool (http://www-is.biotoul.fr/is.html, (Siguier *et al.* 2006)) was slightly higher than that in strain 8081 (about 64 full length IS elements). Interestingly, the variety of IS element families was much higher in strain 8081 and includes IS4 and IS200 family members that are absent from the

genome of strains of serobiotype O:3/4. In *Y. enterocolitica* subsp. *palearctica* strain Y11 ISYen1 is the most frequent IS element (more than 50 copies, see Table 5.4). Since the ISYen1 elements are the most abundant, they are potentially affecting genes in their neighbourhood. These overrepresented elements could induce various crossing-overs of the complete genome. One of the ISYen1 elements located in the promoter region of *inv* affects its regulation (Uliczka *et al.* 2011). Furthermore, many other genes are found in strain Y11 in the vicinity of ISYen1 or other IS elements, which can interfere with their activity. Examples are *cbpA*, *yidE*, a formate efflux transporter and dehydrogenase, *argO*, *ycaD*, putative virulence factors (e.g. a toxin subunit S1 precursor), probable transcription regulators and others. Interestingly, one IS element was detected to be located within a conserved genomic region of members of the family of *Enterobacteria* that contains the *yidE* gene (Y11_30211), coding for a mediator of hyper adherence. Perhaps this widespread genomic region is altered in *Y. enterocolitica* O:3/4 by the insertion of ISYen1.

Table 5.4 Insertion sequence elements found in *Y. enterocolitica* O:3/4 strain Y11.

IS element	**Number of elements found in the chromosome of *Y. enterocolitica* strain Y11**
ISYen2A/B	7
IS1400	1
ISYen4	1
ISYal1	2
ISYen1	53

The ISYen2A/B elements are subspecies specific and have not been found in other yersiniae. Beside the large group of ISYen1 elements, seven copies of the *Y. enterocolitica* subsp. *palearctica* group-specific IS-element Yen2 (Golubov *et al.* 2005) were counted throughout the genome. ISYen2 is related to IS elements of the IS21 family and is present in two isoforms in strain Y11 (designated ISYen2A/B). It has been shown that the number of elements varies in yersiniae, and seven genomic copies is the highest number yet reported. One further copy of ISYen2 is located in the virulence plasmid pYVeY11 (Golubov *et al.* 2005). At least one ISYen2A/B copy was found also in *Y. enterocolitica* strain Y5,27 of the serobiotype O:5,27/3 genome.

5.6. Experimental investigations of strain Y11 specific features

5.6.1. Phage excisions

Prophage elements with full length integrases are likely to be still active, that is to say the phage is still able to excise from and reintegrate in the chromosomal backbone of the host bacterium. To analyse the excision capability of PhiYep-3 and GIYep-01, conventional and nested PCRs have been performed, amplifying only the excised, circular form of the phage or the chromosomal region lacking the phage. The latter is mostly applicable if the phage element is large enough to prevent double bands with less discriminative power due to a similar size of the amplified product. The primers designed for restoration of the 123 bp *attB* attachment site demonstrated high frequency precise excision of the P2-like prophage PhiYep-3. The primer pairs JB585 and JB586 as well as JB585 and JB569 generated clearly visible fragment bands even under standard conditions (55°C annealing temperature, 1 minute elongation time, using the VWR Red polymerase). Generated fragments have been sequenced (see Figure 5.29) and compared with the sequenced genome harbouring the prophage.

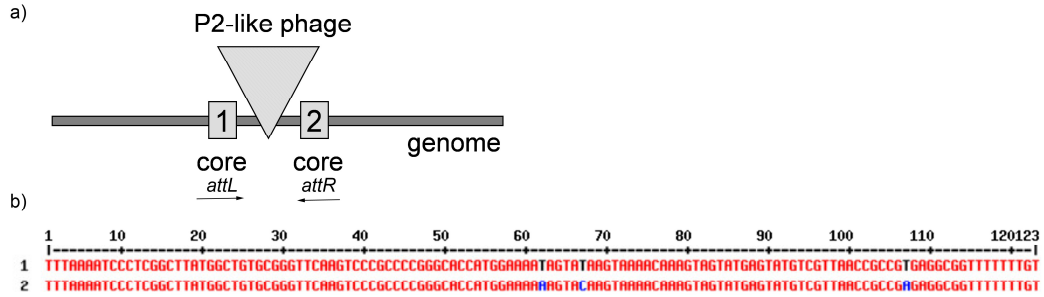


Figure 5.29 Illustration of the *attB* sites flanking the P2-like phage in strain Y11. a) A simplified illustration of the *attB* core sites and the site of integration of PhiYep-3. b) The alignment of both cores 1 and 2 (123 bp), reporting their minor differences in composition.

PCRs performed with primer pair JB606 and JB608 (protocol adjusted to 56°C annealing temperature and 30 seconds elongation time, using the VWR Red polymerase) amplified the *attP* sites of the excised, circular phage (see Figure 5.30). To address the question if the presence of the PhiYep-3 is an occasional event in strain Y11 or if the prophage is present and active also in other strains of this serobiotype, a set of 12 strains of serobiotype O:3/4 from different countries and hosts (also used for SNP studies) was screened for both the presence of the phage and its absence. Beside the isolates used for draft genome sequencing (strains Y5307 and Y8265), the African, Japanese, Australian and Canadian strains of O:3/4 were negative for the occurrence of the PhiYep-3 prophage (all isolates are listed in Figure 5.31). All other strains were positive for the presence and excision of the phage (see Figure 5.31).

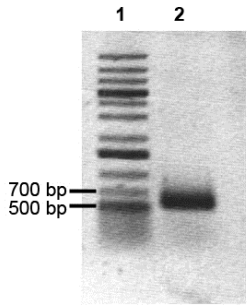


Figure 5.30 P2-like phage excision PCR on a 1 % agarose gel for strain Y11. The generated fragment is based on the excised, circular phage genome (the amplified fragment covers the *attP* region). Lane 1, DNA ladder, lane 2, PCR product of JB606 and JB608 primers.

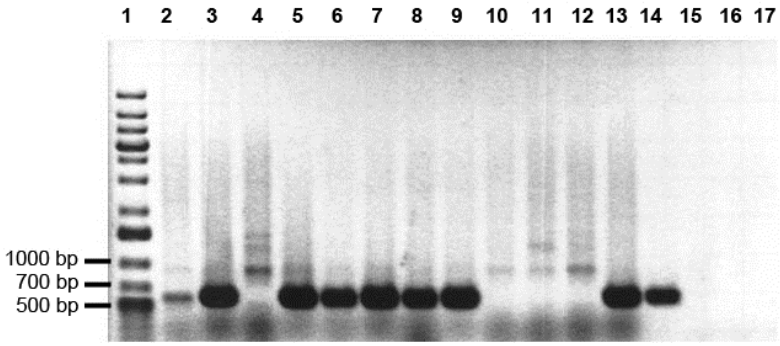


Figure 5.31 P2-like phage excision PCR on a 1 % agarose gel for different *Y. enterocolitica* serobiotype O:3/4 strains. Lane 1, DNA ladder, lane 2-17, PCR samples of different strains generated with primers JB606 and JB608. Lane 2, strain Y641, dog puppet, lane 3, strain Y752, Brazil, lane 4, strain Y756, South Africa, lane 5, strain PL-6, Poland, lane 6, strain PL-15, Poland, lane 7, strain PL-20, Poland, lane 8, strain Y751, Great Britain, lane 9, strain Y748, France, lane 10, strain Y745, Japan, lane 11, strain Y765, Australia, lane 12, strain Y764, Canada, lane 13, strain Y485, pig, lane 14, strain Y11, lane 15, strain Y8265, lane 16, strain Y5307, lane 17, negative control (water).

For GIYep-01, nested PCRs were performed using the primer pairs JB470/JB472 and JB471/JB473. The first primer pair (JB470/JB472) was used for the performance of 35 cycles at 55°C annealing temperature and 1 minute elongation time, and the generated reaction sample was directly used for the second PCR (with primers JB471/JB473), applying 30 cycles with 55°C annealing temperature and 20 seconds elongation time, respectively. For both experiments, the normal VWR Red polymerase was used. The small fragments (<500 bp) generated in the second PCR were subcloned into pGEM-T Easy (Promega), screened with IPTG and X-gal and a batch of clones was revised for the presence of the insert and proven by sequencing (see Figure 5.32).

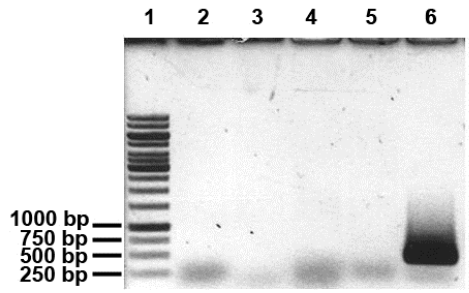


Figure 5.32 GIYep-01 excision PCR on a 1.5 % agarose gel for strain Y11. Lane 1, DNA ladder, lane 2-5, negative PCR samples of other phage excision PCRs, lane 6, positive PCR sample for strain Y11.

5.6.2. Microarray studies

Previously, A. Golubov applied the method of suppression subtractive hybridization (SSH) to identify possible genetic markers for the low pathogenic *Y. enterocolitica* subsp. *palearctica* group. To uncover such differences he subtracted genomic DNA of the mouse non-virulent strain Y-108C, serobiotype O:3/4, with DNA of the mouse virulent strain WA-C, serobiotype O:8/1B, by using WA-C DNA as a driver to subtract common genetic sequences. In total, 83 Y-108C - specific sequences were identified and used to design and to evaluate a serobiotype specific DNA-microarray. The array covers also a number of established virulence-associated genes of the *Yersinia* genus. As the complete genome sequence of serobiotype O:3/4 strain Y11 is now available, the SSH-data have now been optimised using the genome sequence to erase duplicates (or too close adjacent microarray samples) and enhancing the discriminating power. Duplicated sequences present in the same regions of genome differences were filtered by BLASTn search against the Y11 genome. Thus the number of strain Y11- specific probes obtained by SSH has been considerably reduced resulting in a small scale DNA microarray. Whenever possible, Y11-specific sequences were switched to gene-covering regions and customised in fragment size. The annotation and genome comparison between serobiotypes O:3/4, O:8/1B and the related serobiotype O:5,27/3 also provided new information about unique gene clusters and sequences, including also several putative virulence determinants. These data, together with the established and optimized DNA microarray dataset, could be used both for genotyping of *Yersinia enterocolitica* and for expression studies to uncover novel virulence determinants. The sequences of the microarray are attached as supplementary material under section 1.1.

5.6.3. SNPs

A single-nucleotide polymorphism (SNP) is a variation in DNA sequence; a single nucleotide in the genomic equipment differs between members of a species. These polymorphisms find utilisation in finetyping of bacterial species, such as MLST and DNA Fingerprinting (RFLP Analysis). Since housekeeping genes are known to exhibit slower rates of single nucleotide

mutations (and therefore are especially used in subtyping), virulence genes and genes coding for proteins exposed to the bacterial surface (or secreted) are supposed to be more susceptible for polymorphisms. They may serve as an adaptation tool for fitness, survival and pathogenicity, with the possibility to be a selective advantage in environment or hosts.

To analyse the potential of SNPs in *Y. enterocolitica* subsp. *palearctica* O:3/4, a collection of 12 strains from different countries, different cities within one country (three Polish isolates) and two different animal sources (pig, dog puppet) was selected for a first screening (see Table 5.2). Polymorphisms found between the two related genomes of O:3/4 and O:5,27/3 were considered to be good candidates for polymorphisms also between members of the O:3/4 serobiotype. Interestingly, in a set of 20 outer membrane proteins, OMPs, only 4 genes showed more than one SNP between the two organisms. In contrast, *lcrV*, *yadA* and *rtxA* exhibit a large number of single nucleotide differences. In case of the >9 kb comprising *rtxA* gene, over 70 differences were present, whereas in the 1,413 bp comprising *yadA* gene 10 polymorphisms were detected. For *lcrV* (975 bp), 5 differences were found.

Three regions of *rtxA*, encoding a putative toxin, and two regions of the pYV plasmid-encoded virulence determinants *lcrV* and *yadA* were chosen, based on promising differences between the *Y. enterocolitica* representatives Y11 and Y5,27 of serobiotypes O:3/4 and O:5,27/3, respectively. The respective regions have been amplified using primers JB392-408 (see table under 4.4). Within the analysed regions of *rtxA* (3 different loci covering more than 2,400 bp of the 9 kb *rtxA* gene), only one SNP was found in the isolate from GB (strain Y751). In one strain (Japan, Y745) an additional sequence of 9 bp was found, exactly as in strain Y5,27 of serobiotype O:5,27/3. Beside these differences, none of the SNPs between strains of serobiotype O:3/4 and O:5,27/3 could be confirmed as SNP within the O:3/4 serobiotype. Within the region of about 750 bp covering more than one half of the 1,413 bp long *yadA*, 4 SNPs in strain Y745 (Japan) were detected, of which one SNP was identical to the only SNP of strain Y765 (Australia) and strain Y746 (Canada). In strain Y745, also a region of 12 bp was missing (starting at position 75), reducing the protein length. For *lcrV*, the primer pair used covers the complete gene (975 bp). In all strains analysed, not a single SNP was found.

Table 5.5 SNP study results for *rtxA*, *yadA* and *lcrV*

Country	Strain no.	SNPs
GB	Y751	1 SNP in *rtxA*
France	Y748	-
Canada	Y746	1 SNPs in *yadA*
Poland	PL-15, Warszawa	-
Australia	Y765	1 SNP in *yadA*

Country	Strain no.	SNPs
South Africa	Y756	-
Japan	Y745	9 additional bp in *rtxA*, 4 SNPs in *yadA*, loss of 12 bp in *yadA*
Brazil	Y752	-
Poland	PL-6, Szczecin	-
Poland	PL-20, Lubin	-
Dog	Y641	-
Pig	Y485	-
Germany	Y11 (reference)	-

5.6.4. Carbon source transport - the aga-operon

The group translocation of amino-sugars as N-acetyl-D-galactosamine and N-acetyl-D-glucosamine is enabled via specific phosphotransferase-systems (PTS). Those systems generally transfer phosphate residues of phosphoenolpyruvate to the respective substrates, mostly hexoses, which than can be transported across the membrane into the cytoplasm of the bacterium. For this process, different enzymes are necessary: phosphoenolpyruvate (PEP) represents the phosphoryl group donor, whereas enzyme I (EI) and the histidine phosphocarrier protein (HPr) are transferring the phosphate to the Aga complex (belonging to the enzyme class II, EII), which converts the substrate specifically to a phosphorylated one and enables the translocation across the membrane (see Figure 5.33) (Siebold *et al.* 2001;Tchieu *et al.* 2001).

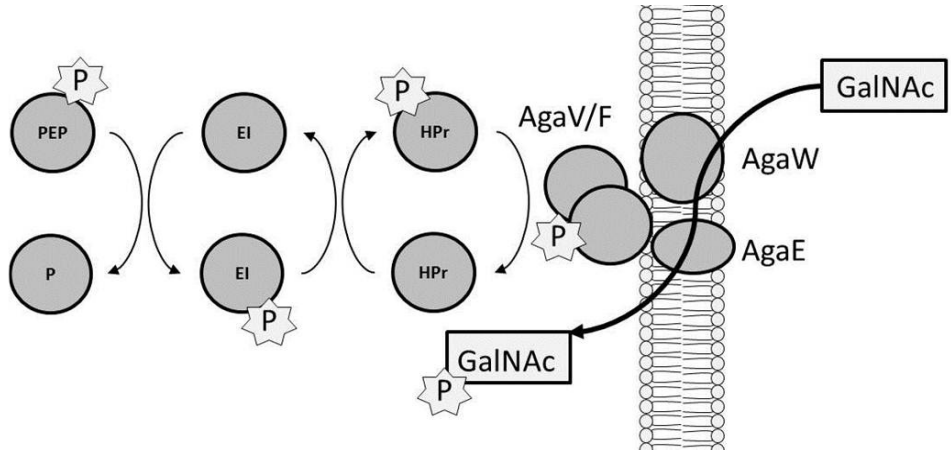


Figure 5.33 Schematic diagram of the AgaVWEF-dependent amino-sugar transport across a membrane. Phosphoenolpyruvate (PEP) transfers a phosphoryl group to an enzyme class I protein (EI), resulting in pyruvate and phosphorylated EI. EI transfers this phosphate to a histidine phosphocarrier protein (HPr), which in turn transfers the phosphate to the Aga complex, which consists of a cytoplasmic AgaV/F dimer and the membrane spanning AgaW and AgaE. The phosphoryl group finally is transferred to the substrate, in this case GalNAc.

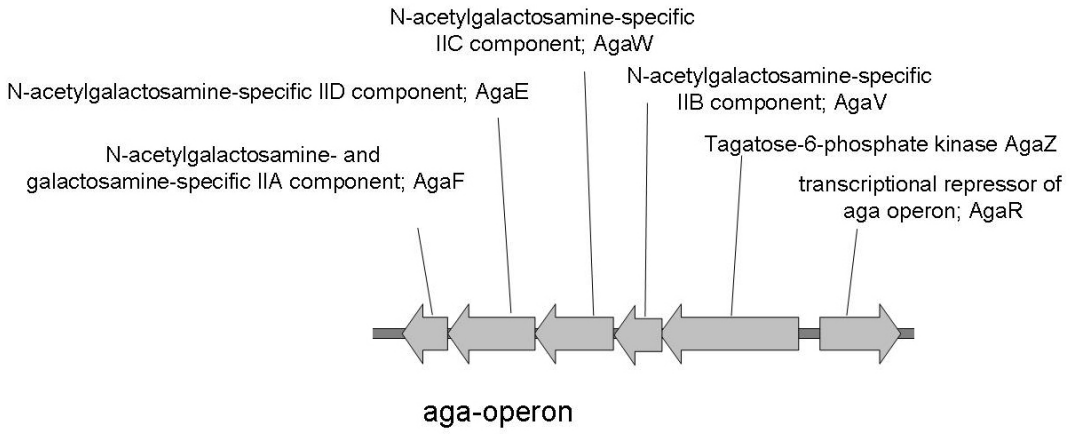


Figure 5.34 The *aga*-operon genes, comprising *agaZVWEF* and the repressor gene *agaR*. AgaZ is involved in the carbohydrate metabolism; AgaV, AgaW, AgaE and AgaF belong to the class II enzymes involved in N-acetyl-D-galactosamine (GalNAc) transport.

The genes that encode the EII proteins are organised in an operon (see Figure 5.34), which is present only in strain Y11 of serobiotype O:3/4 and absent from strain 8081 and WA-314 of serobiotype O:8/1B. To address the functionality of the carbohydrate transport, growth experiments and genetic complementation experiments with different amino-sugars have been performed.

The utilisation of the amino-sugars N-acetyl-D-galactosamine (GalNAc) N-acetyl-D-glucosamine (GlcNAc), galactosamine (Gam) and D-glucosamine (Nag), all obtained from Sigma-Aldrich, was analysed in the two different *Yersinia* serobiotypes O:3/4 and O:8/1B in minimal M9 medium without carbohydrate sources other than 0.2 % amino-sugars (w/v). As expected, both serobiotypes were able to grow in 0.2 % N-acetyl-D-glucosamine and glucosamine. Since the Gam-phosphoenolpyruvate-dependent phosphotransferase system (PTS) encoding genes (*agaBCD*) are absent from both serobiotypes O:3/4 and O:8/1B, it was not surprising that both strains were found unable to grow in galactosamine as the only source for carbon. *Y. enterocolitica* O:3/4 was able to use N-acetyl-D-galactosamine (GalNAc) as the only carbohydrate source, even though there was no homolog for the downstream processing *agaA* gene detected (see Table 5.6). To prove the sufficiency of *agaVWEF* to support growth in the presence of GalNAc the *Y. enterocolitica* O:3/4 *aga*-operon was cloned together with its promoter sequence (Y11_11961-Y11_12031) into the vector pGEM-T Easy (Promega). *Y. enterocolitica* strain WA-314 transformed with *agaVWEF* was then able to grow in the presence of N-acetyl-D-galactosamine (see 5.6.4.3). Thus, the *aga*-operon is able to support utilisation of N-acetyl-D-galactosamine *in vitro* and supports *Y. enterocolitica* with the ability to efficiently utilize this amino-sugar of the intestinal mucin.

Table 5.6 Sugar specific PTS genes in *Y. enterocolitica*

Sugar/amino-sugar	PTS, sugar specific gene	Presence in Y. enterocolitica
Glucose	ptsG/crr	present
Glucosamine	gamP	present
Galactosamine	agaBCD(F)	absent
N-acetyl-D-glucosamine	nagE	present
N-acetyl-D-galactosamine	agaVWE(F)	present in O:3/4, absent in O:8/1B

5.6.4.1. Carbohydrate growth experiments

To address the uptake and utilisation of different carbohydrates and amino-carbohydrates of the glucose (GlcNAc, Nag, glucose) and mannose (GalNAc, Gam) families, *Y. enterocolitica* strain Y11 was grown in minimal salts M9 medium, supplemented with only a single carbohydrate source. The first experiment was performed with a starter concentration of OD_{600} 0.02 over-night. In this experiment, only the positive controls (1mM glucose as the carbohydrate source) and N-acetyl-D-glucosamine samples were grown to at least logarithmic densities (OD_{600} between 0.4 and 0.8). The optical density of strain Y11 in the N-acetyl-D-galactosamine medium was only slightly increased to 0.034, whereas that of strain WA-314 remained constant, indicating a lack of proliferation (see Figure 5.35). These experiments were repeated, using bacteria directly from BHI over-night cultures instead of the fresh grown, logarithmic phase bacteria. Even though the growth rates were increased slightly after over-night incubation in the respective carbohydrate-M9 media, the GalNAc culture of strain Y11 reached not more than an OD_{600} of 0.15. In a second approach, bacteria were grown on M9 agar plates containing 0.2 % of the respective carbohydrates. Before plating, the bacteria harvested from CIN agar plates were washed twice with 0.9 % NaCl solution. After one week at 27°C as well as 37°C, only the positive control (0,2 % glucose-M9 agar plates) was able to grow to visible colonies. Only when yersiniae were directly streaked from CIN to M9 agar plates, growth of strain Y11 on GalNAc plates was detected (at 27°C, no growth at 37°C). Thus, a prolonged growth time in liquid medium together with an additionally higher inoculation rate was applied (OD_{600} about 0.25). To support the growth in the amino acid free medium, 200 µg/mL tryptophan was added (Trp) to stabilise growth rates.

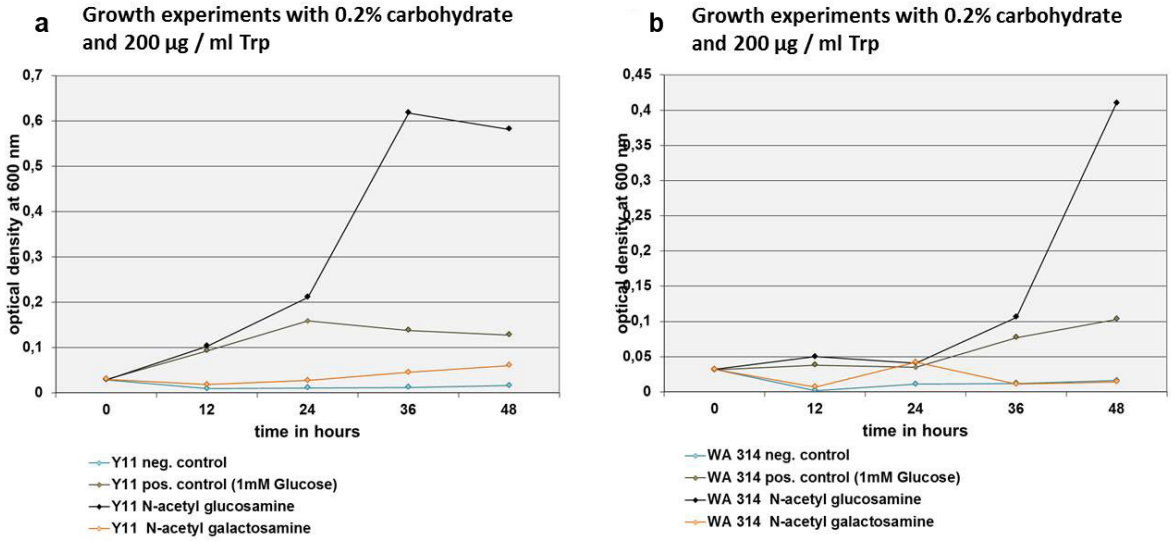


Figure 5.35 Carbohydrate-dependent growth of *Yersinia* in M9 medium supplemented with different carbohydrates. Strains used for the experiments are Y11 (a) and WA-314 (b). The experiment was performed with an inoculation resulting in OD_{600} = 0.02. The legends below the respective diagrams show the different growth conditions. In this experiment, only the positive controls (1 mM glucose, brown curves) and N-acetyl-D-glucosamine samples (black curves) were grown to at least logarithmic densities (OD_{600} between 0.4 and 0.8). The optical density of strain Y11 in the N-acetyl-D-galactosamine medium (orange curve, a) was only slightly increased to 0.034, whereas the strain WA-314 GalNAc culture (orange curve, b) did not show any growth.

5.6.4.2. Cloning of the aga-operon

An 8.4 kb region encoding the complete *aga*-operon of *Y. enterocolitica*, comprising *agaZVWEF* and the additional regions of *agaR* and *aga(X/S?)Y*, was amplified with the Phusion polymerase using the primer pair JB506/JB507 (see Figure 5.36, primer JB506 located before the flanking hypothetical gene, JB507 within the flanking hypothetical gene) at 60°C annealing temperature and 3.3 minutes for the elongation time. Uladzimirase reaction was performed to generate overlaps needed for subcloning of the gene cluster into pGEM-T Easy (Promega). The *aga*-operon plasmid was transformed into strain WA-314 using carbenicillin as the selective antibiotic.

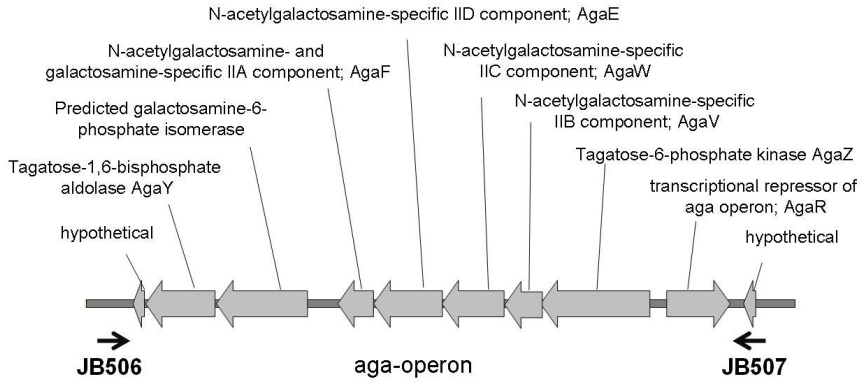


Figure 5.36 The *aga*-operon illustrated with the location of primers used for subcloning into the pGEM-T Easy vector.

5.6.4.3. Carbohydrate growth experiments with strain WA-314 aga+

Growth experiments using the WA-314 serobiotype O:8/1B strain transformed with the *aga*-operon (see Figure 5.36) were carried out referring to the experimental set up described above, but in the presence of carbenicillin to ensure the maintenance of the plasmid. Since the cloning vector pGEM-T easy was utilised, the *aga*-operon was only under the regulation by intrinsic promoters. The *aga*-operon enabled growth of WA-314 with GalNAc as the only carbon source better than compared to strain Y11. This is due to the likely presence of more than one copy of the plasmid. Figure 5.37 shows the growth curve of all strains during the long time period of 48 hours. Figure 5.38 shows the same experiment in a more comprehensible bar diagram.

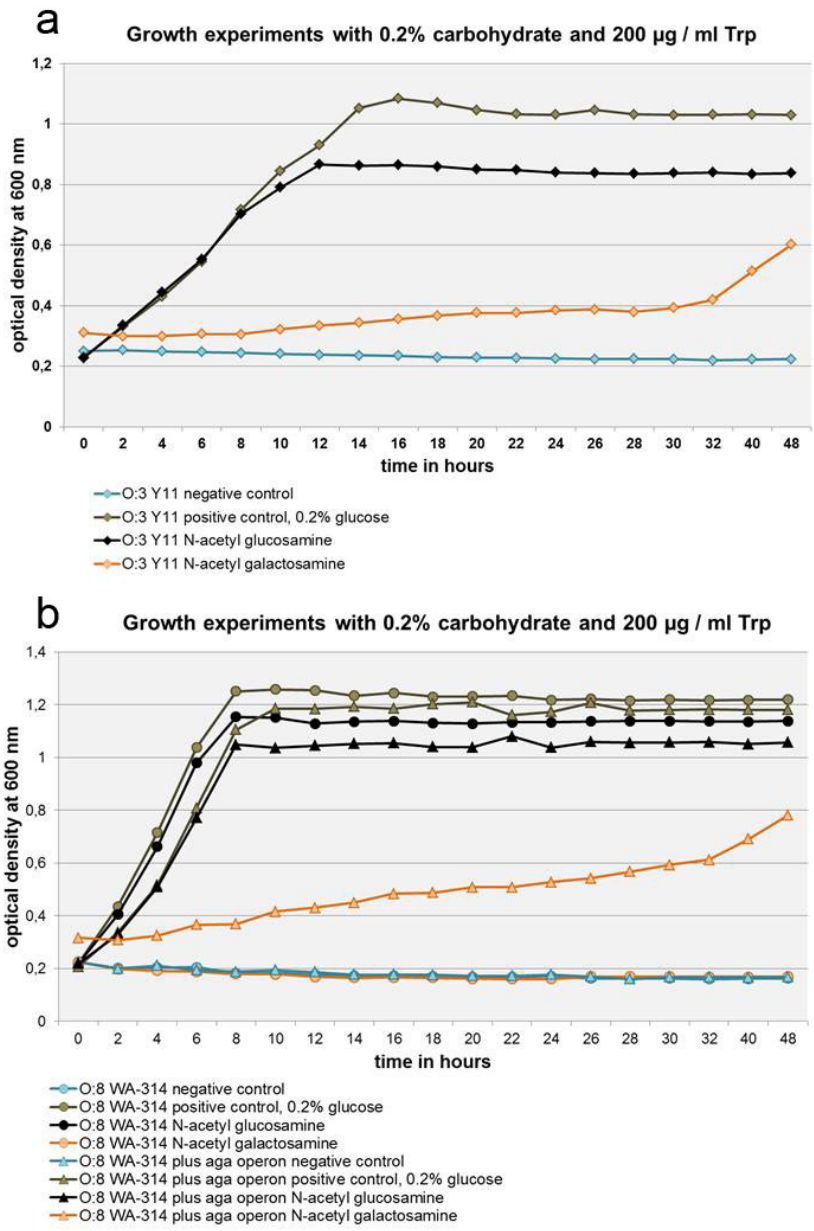


Figure 5.37 Carbohydrate-dependent growth of *Yersinia* in M9 medium supplemented with different carbohydrates and appropriate antibiotics if required. Strains used for the experiments are Y11 (a) and WA-314 (b). The legends below the respective diagrams show the different growth conditions. The experiment was performed with an inoculation resulting in about OD_{600} = 0.25. Negative control was M9 without any carbon source but supplemented with tryptophan; the positive control was 0.2 % glucose. The strain WA-314 plus *aga*-operon refers to the WA-314 serobiotype O:8/1B strain transformed with the complete *aga*-operon.

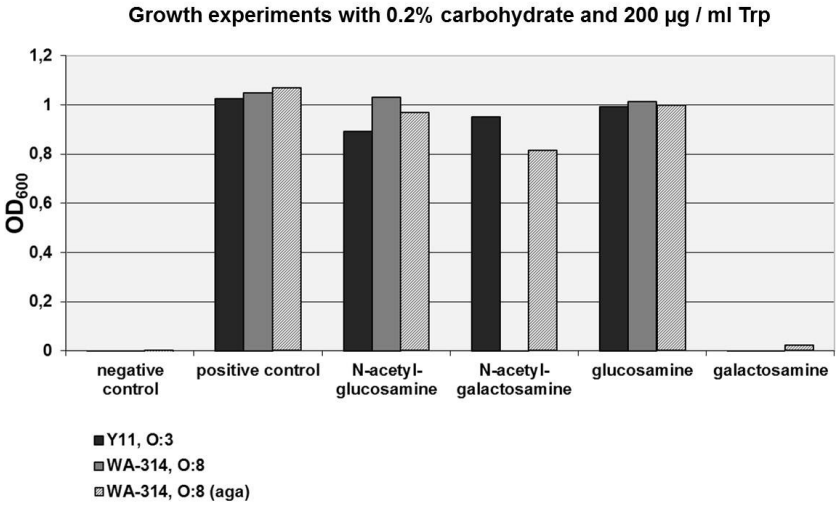


Figure 5.38 Carbohydrate-dependent growth of *Yersinia* in M9 medium supplemented with different carbohydrates after 48 hours. The inoculation OD_{600} has been subtracted from the OD_{600} values measured after 48 hours. Strains used for the experiments are Y11 (O:3/4) and WA-314 (O:8/1B) as well as a derivative of WA-314, supplemented with the complete *aga*-operon (WA-314 (aga)). The legend below the diagram shows the different strains. Negative control was M9 without any carbon source but supplemented with tryptophan; the positive control was 0.2 % glucose. The diagram shows the measurement of the optical density at 600 nm (OD_{600}) directly after inoculation.

5.6.5. Bacteriocin

The unique bacteriocin gene cluster found in strain Y11 (Y11_33511-33531) was analysed for the putative functions using motif and domain comparison tools. The protein is similar to S-type pyocins and shares several homologies to endonuclease and hydrolase motifs (see Figure 5.39). Bacteriocins have been described to be toxic mostly against related species. To address the function and host range of this Y11-specific protein, overlay assays have been performed and the protein was purified.

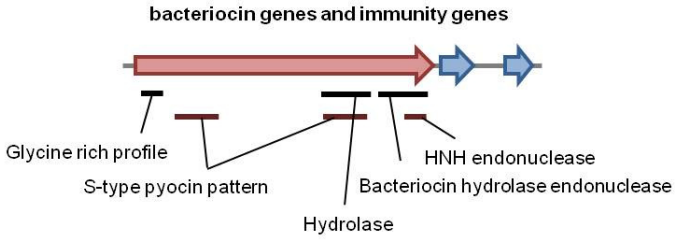


Figure 5.39 Motifs of the strain Y11 specific bacteriocin as predicted by the MOTIF search tool (http://motif.genome.jp/). The bacteriocin gene in strain Y11 is indicated by a red, the immunity genes by blue arrows.

5.6.5.1. Overlay experiments

In order to ascertain the specific species sensitive to the strain Y11 bacteriocin, several *Yersinia* species (STM Nölting strains 113-119, as well as strain WA-314, *E. coli* strain DH5α and others) were analysed as tester strain in the soft agar overlay assays. In addition, nasopharyngeal swabs of a feeding pig were streaked on blood agar plates and single colonies of the flora were used as tester strains. All tester strains were diluted in soft agar as described in 4.13 and exposed to *Y. enterocolitica* strain Y11. Different incubation temperatures have been used (4°C, room temperature, 27°C and 37°C), and agar plates were inspected for growth inhibition of the tester strains after over-night incubation as well as after several days. However, none of the tester strains used was inhibited in growth in theses soft agar overlay assays.

5.6.5.2. mRNA experiments

Co-culture experiments to induce the bacteriocin production were made with *E. coli* DH5α and *S. aureus* (random nose isolate). Strains were diluted in fresh LB medium from over-night cultures and grown to a mid-logarithmic phase before used for co-culture. The strains were cultured 1:2 (one part, inducer strain, two parts, strain Y11) for one hour at 27°C (*E. coli*) or 37°C (*S. aureus*). The complete mRNA was isolated and the RNA was tested for DNA contamination. The mRNA was transcribed to cDNA and analysed for the presence of *gapA* (housekeeping gene, primers of the institute) and bacteriocin (primer pair JB379 and JB386). The housekeeping gene was present in both coculture experiments, but still no cDNA for the Y11 bacteriocin could be detected (see Figure 5.40).

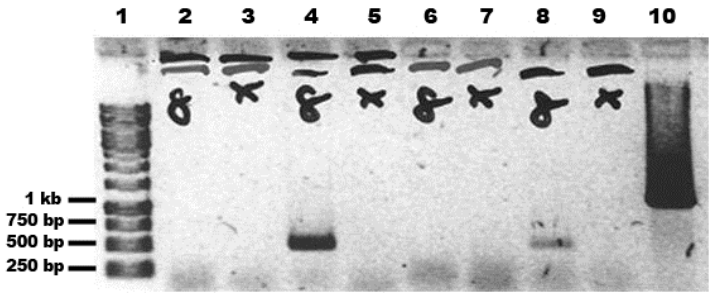
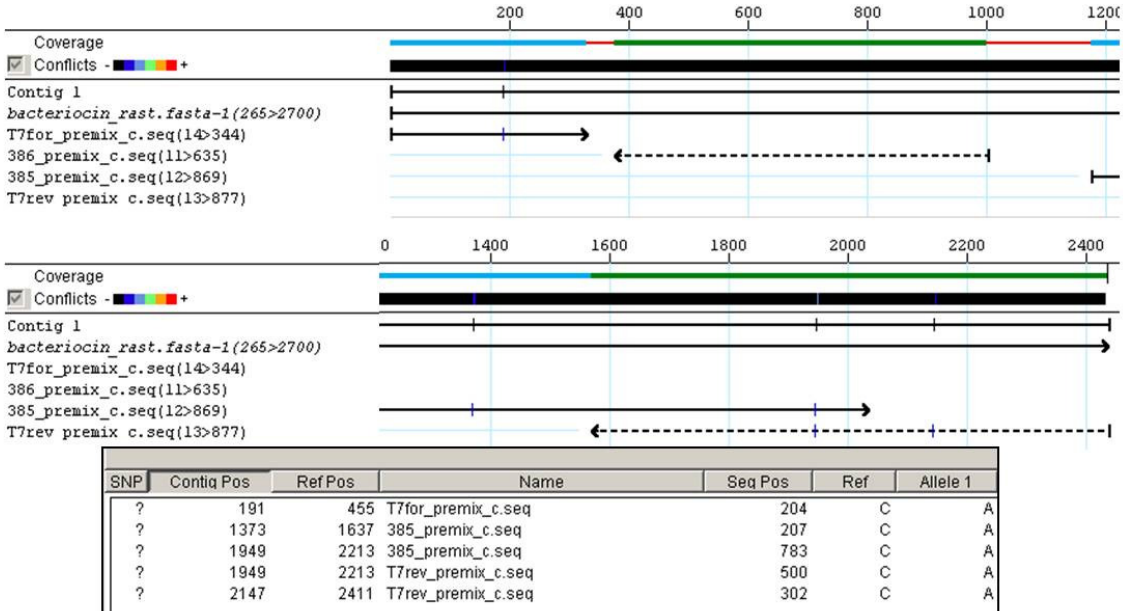


Figure 5.40 PCR results of mRNA experiments for the bacteriocin gene in strain Y11. Lane 1, DNA ladder, lanes 2-5, *S. aureus* experimental data, lanes 6-9, *E. coli* experimental data, lane 10, positive control performed with strain Y11 genomic DNA and primers JB379 and JB386. Lanes 2 and 3 (*S. aureus*) were performed with untranscribed mRNA (first lane, *gapA*, second lane, bacteriocin), lanes 4 and 5 with cDNA, showing only positive results for the *gapA* control gene. This is also the case for the *E. coli* experiments, where only the *gapA* positive control for cDNA is detectable in the PCR.

5.6.5.3. Cloning expenses for the bacteriocin gene

To enable a stable expression or even over expression of the bacteriocin gene, the gene was subcloned into different vectors: pWS (bacteriocin with His_6-tag, primer pair JB352/JB353, bacteriocin without His_6-tag, JB351 and JB353, both performed with Phusion polymerase at 57°C annealing temperature and for one minute elongation time), pGEM-T Easy (same primers as for pWS), pACYC184 (Bacteriocin and immunity proteins, JB373 and JB374, 35 cycles with Phusion polymerase, conditions set to 55°C and 2 minutes) and pET vectors pET21b (Bacteriocin, JB371 and JB372, 35 cycles with Phusion polymerase, adjusted conditions to 55°C, 1.5 minutes) and pET24b (Primer pair JB411/JB412 used to amplify the pET24b vector with new XhoI and BamHI restriction sites, primer pair JB413/JB414 for the bacteriocin gene without the stop codon, both amplified with Phusion flash polymerase, adjusted conditions 55°C and 1.5 minutes). The cloning of bacteriocin alone was not possible for pWS and pGEM-T Easy vectors. Thus, a two-plasmid strategy was followed: The bacteriocin and immunity genes were cloned into pACYC184, transformed into JM109 *E. coli* cells and used as the donor strain for transformed ligations with pET21b, containing only the full length bacteriocin gene. This strategy was supposed to overcome the toxic side effects of the bacteriocin gene, since the immunity proteins with their natural promoter region were provided to the *E. coli* host strain. In parallel, pET24b was used to follow the strategy of different antibiotic resistance cassettes (the ampicillin resistance of pET21b is not recommended for use in *Yersinia enterocolitica* because of the presence of two beta-lactamases). Unfortunately, all these cloning strategies failed, resulting in premature stop codon containing vectors or leading to altered sequences with changes in different proline residues in the gene (up to 4 changed prolins, see Figure 5.41). Proline residues are reported to alter the structure of a protein, thus likely to have an effect on the functionality of the protein. So even using tight cloning vectors, there must have been a force to select for clones harbouring plasmids with altered bacteriocin sequences. Finally, the bacteriocin was used for *in vitro* expression.



SNP	Contig Pos	Ref Pos	Name	Seq Pos	Ref	Allele 1
?	191	455	T7for_premix_c.seq	204	C	A
?	1373	1637	385_premix_c.seq	207	C	A
?	1949	2213	385_premix_c.seq	783	C	A
?	1949	2213	T7rev_premix_c.seq	500	C	A
?	2147	2411	T7rev_premix_c.seq	302	C	A

Figure 5.41 Proline residues altered in bacteriocin clones. Screenshot made from Lasergene, SeqMan. The screenshot illustrates the location of the nucleotide differences of the altered amino acids (upper panels, vertical lines on the sequencing results shown as arrows) and gives the exact position and kind of the exchanged base (lower box, Seq Pos, position of the nucleotide, Ref, nucleotide in the original bacteriocin sequence, Allele 1, changed base).). Proline residues are reported to alter the structure of a protein, thus likely to have an effect on the functionality of the protein.

5.6.5.4. In vitro translation of the bacteriocin

The strain Y11 bacteriocin gene was amplified with customised primers to perform *in vitro* translation using the 5Prime RTS system (JB516 and JB517, ordered following the manufacturer's protocol). Two PCRs were performed to generate the RTS template, the first one generated specific overlaps adjacent to the bacteriocin gene, the second PCR (using 150 ng of the first PCR product) used these overlaps to amplify the gene with the appropriate regulatory elements and the His_6-tag with primers supplied by 5Prime. The product of this PCR was then directly used for *in vitro* translation with the RTS system, 6 hours at 30°C with gentle shaking. The Protein was purified with magnetic beads and checked for presence and size with Western Blot detection, using a Gfp protein as the positive control. Detection was performed with anti-His_6-antibodies. Figure 5.42 shows the coomassie staining of the gel and the Western Blot developed film.

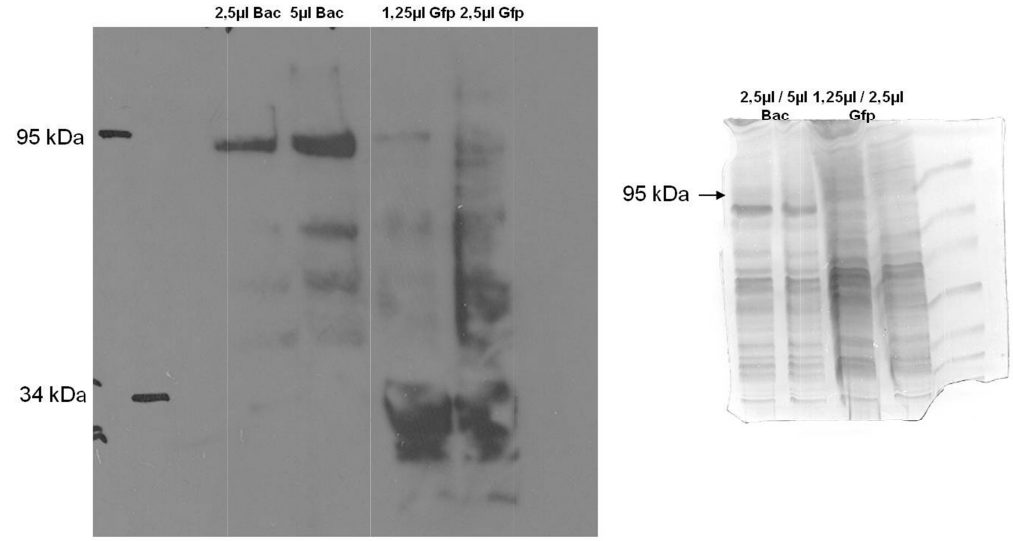


Figure 5.42 Bacteriocin (Bac) western blot detection (left) and coomassie staining (right). Different amounts of the purified protein have been applied to the SDS gel (2.5 and 5 µL of the bacteriocin protein, 1.25 and 2.5 µL of the Gfp control product). Bac, bacteriocin protein, Gfp, Gfp control protein. The molecular weight marker has been manually marked at ~ 95 kDa (indicating the bacteriocin protein band) and ~ 34 kDa (indicating the Gfp control protein band).

5.6.5.5. In vitro DNA digestion experiments and agar assays

The purified protein obtained from *in vitro* translation was used for host specificity and functionality tests. PCR products of random origin (about 150-200 ng) were incubated with the bacteriocin (5-8 µL of the purified protein) in FastDigest restriction enzyme buffer (Fermentas) for 1.5 hours at 37°C. Interestingly, gel electrophoresis showed a slight shift of the PCR product band when incubated with the protein, but this could be due to complex formation. Purification of the reaction mix (following the standard PCR purification protocol of the manufacturer) showed no significant decrease in PCR product template, disqualifying the enzyme to exhibit DNA endonuclease activity under these conditions. Likewise, performing overlay assays as described with again different strains and species, using 5 µL of the purified bacteriocin dropped onto the soft agar, no growth inhibition could be shown. Also the DNAse-test agar (with methyl green as indicator for DNA degradation, obtained from Becton Dickinson (Heidelberg)) did not show any degradation of DNA. Either this enzyme is not active after the *in vitro* translation, or the conditions used did not enable functionality.

5.6.6. RtxA

The RtxA protein (Y11_18761, 9738 bp, about 350 kDa) found to be specific for serobiotype O:3/4 has been described as a member of the MARTX family and shares several characteristics with these group of proteins. But still the function and role of RtxA in *Y. enterocolitica* O:3/4 is

unknown. To address these questions, the transcription of this gene was analysed and an *rtxA* knockout mutant was generated.

5.6.6.1. RtxA mutagenesis

The knockout mutant of *rtxA* was generated using suicide plasmid conjugation. Suicide plasmids, as pKNG101, are unable to reproduce themselves in the host. They need the additional presence of the *pir* genes to replicate, for example delivered with the pPirK plasmid. Thus, once the *pir* genes are missing, the only way to obtain the advantage of antibiotic resistance cassettes of the plasmid, the plasmid has to integrate into the host's genome. This is facilitated by homologous regions provided on the plasmid, flanking the resistance cassette, directing the integration of the plasmid to one specific region of the host genome. In addition, a single sided recombination (single crossover) can be forced to both sides by adding sucrose to the medium. The *sacB* gene located on the suicide vector provides a positive selection marker for the excision of the vector, as it confers sensitivity to sucrose. Figure 5.43 shows the map of pKNG101-1, including the *sacB* gene location and the homology regions 1 and 2, comprising nucleotides 70-791 and 8902-9794 (overlapping the end of the gene) of *rtxA*, respectively (see also Figure 5.44).

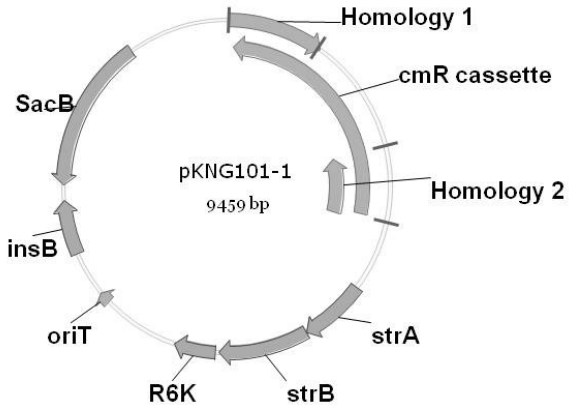


Figure 5.43 Map of pKNG101-1. Homology regions 1 and 2 refer to the *rtxA* gene (nucleotides 70-791 and 8902-9794 (overlapping the end of the gene) of *rtxA*).

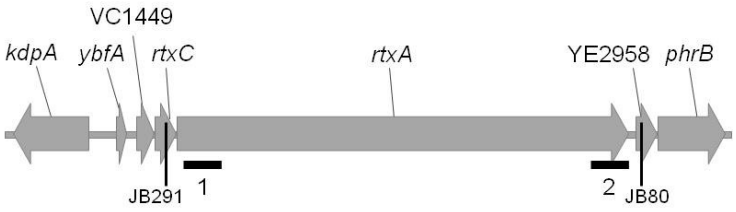


Figure 5.44 Genetic map of the *rtx* cluster depicting the localisation of *rtxA* mutagenesis homology regions 1 and 2 and the localisation of sequencing primers JB291 and JB80.

Conjugation was performed with a nalidixic acid resistant derivate of strain Y11 (Y11 (Nal^R)) to select against *E. coli* after mating on blood agar plates. The pKNG101-1 plasmid was delivered by S17-1 λpir. Conjugation was performed for 2 hours on blood agar, followed by the serial dilution and plating of the mating mixture. Clones were selected on nalidixic acid and chloramphenicol agar plates, and the excision of the pKNG101-1 backbone after the successful recombination of the plasmid with the strain Y11 genome in one region of homology was enabled selecting against the *sacB* gene on sucrose containing agar plates (5 % sucrose). Clones grown on sucrose were tested for the deletion of *rtxA* by PCR and sequencing with primers JB291 and JB80, confirming the presence of the chloramphenicol resistance cassette (Cm^R, see Figure 5.45).

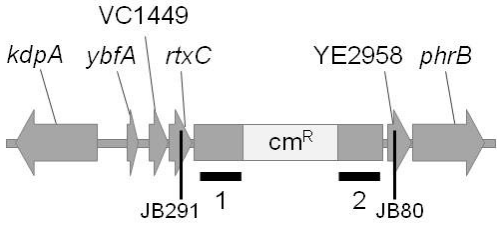


Figure 5.45 Genetic map of the mutated *rtx* cluster in strain Y11 (Nal^R). The region flanked by JB291 and JB80 has been amplified and sequenced to prove correct mutagenesis.

Later experiments with the generated *rtxA* mutant strain, named Y11 Δ*rtxA*, showed the absence of the pYV plasmid in this strain. This could be due to the temperature shift (37°C) in combination with the physical stress of co-culture during conjugation. The pYV plasmid was restored by conjugation of the strain Y11 mutant with pCB 3-6 containing the pYVO:3 co-integrate fused with the mobilisable vector pRK290 and the pRK2013 helper plasmid.

Complementation of the Δ*rtxA* mutant was performed using the pSC-B vector and primers JB329 and JB80, generating an 11.926 kb fragment with the complete *rtxA* gene (see Figure 5.46). The PCR was done with the Phusion polymerase, executed with 35 cycles and the following adjusted conditions within the 35 cycles: denaturation for 7 seconds at 98°C, annealing at 56°C for 15 seconds and elongation time 4 minutes at 72°C. This fragment was ligated into pSC-B (following the manufacturer's protocol) and clones were inspected at both borders of the ligation.

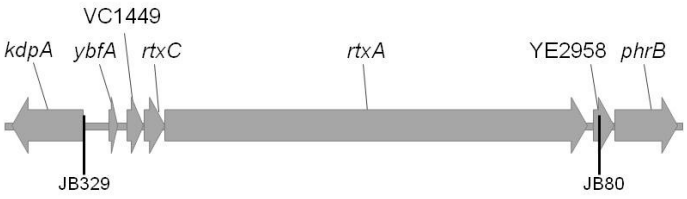


Figure 5.46 Genetic map of the complementation region for the *rtxA* mutant strain. The region between JB329 and JB80 has been amplified and subcloned into the vector plasmid pSC-B.

5.6.6.2. Transcriptional analysis of rtx gene expression

To investigate the presence of the putative toxin RtxA, mRNA experiments were performed using the complete mRNA of the bacterial cells, followed by DNA contamination testing by PCR. After transcription into cDNA, PCRs for both a housekeeping gene and the *rtxA* gene were performed.

The hypothesis was that the exposure of different mammalian cells with strain Y11 bacteria could trigger or induce the expression of *rtxA* by receptor or ligand binding. To address this hypothesis, different cells lines were incubated with strain Y11 with a multiplicity of infection (MOI) of about of 10 for 2 hours at 37°C and 5 % CO_2 in 6-well dishes (2 mL cell suspensions have been seeded one day before the experiments with about 2.5 x 10^6 cells per well). The mRNA of the infected cell culture was purified and analysed as described under 4.16.2. Interestingly, different cell lines used (including among standard HEK and HeLa cell lines the porcine intestinal epithelia cell lines IPI-2I cells and IPEC-J2) resulted in positive mRNA transcripts for *rtxA*. Therefore, experiments were reproduced only with cell culture medium and standard LB medium. Still mRNA of *rtxA* could be traced by cDNA based PCR (with primers JB43/ JB44). Finally, strain Y11 was grown over-night (inoculated early, so growth for about 16 hours was enabled) at 27°C in M9 minimal medium. Even under these conditions, cDNA based PCR gave results for *rtxA* (see Figure 5.47), indicating a weak regulation and the constitutive transcription of the *rtxA* gene.

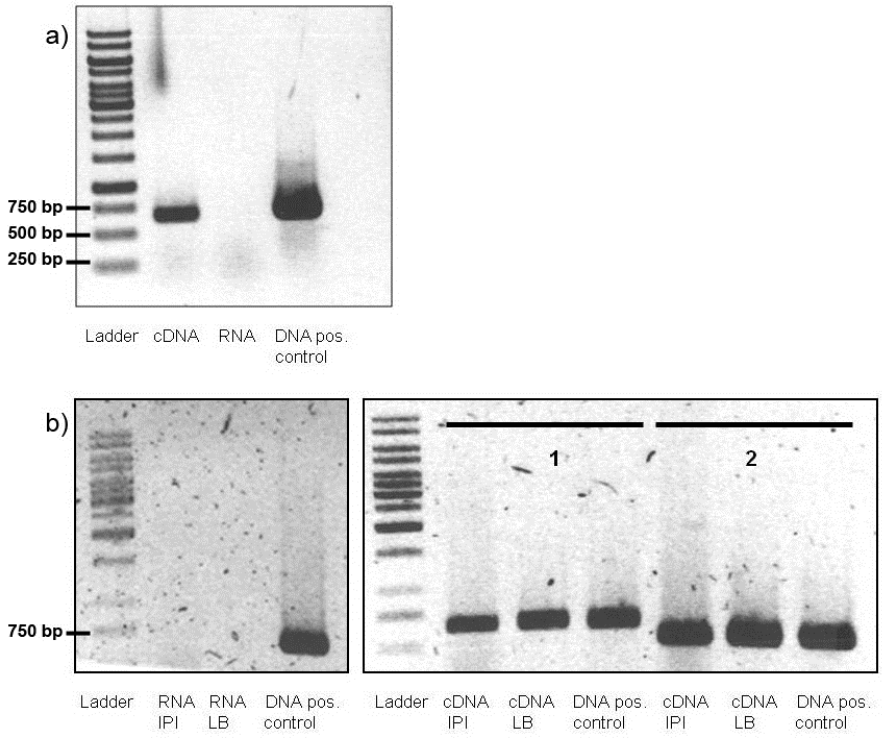


Figure 5.47 PCR results of mRNA experiments for *rtxA* with primers JB43/ JB44. a) Experiments carried out in M9 medium. First lane, marker, second lane, the transcribed mRNA PCR (cDNA), third lane, DNA contamination control (RNA), last lane, positive control. DNA marker: GeneRuler 1kb ladder. DNA positive control performed with strain Y11 wt DNA. B) mRNA experiments performed with mammalian cells (abbreviated IPI, IPI-2I cells) or LB medium, respectively. The left picture shows the DNA contamination control for both mRNA preparations. DNA marker: GeneRuler 1kb ladder. DNA positive control performed with strain Y11 wt DNA. The figure on the right side shows the successful detection of *rtxA* cDNA in all samples (1) and the housekeeping gene control (*gapA*, 2). DNA marker: GeneRuler 1kb ladder. DNA positive control performed with strain Y11 wt DNA.

6. Discussion

Y. enterocolitica subsp. *palearctica* serobiotype O:3/4 represents the majority of human stool isolates in Germany as well as northern Europe. Moreover, this pathogenic serobiotype made a step forward to global relevance (Bonardi *et al.* 2003;Martinez *et al.* 2011;Rosner *et al.* 2010), which can be referred to a high contamination of butcher shops and slaughter pigs (Bucher *et al.* 2008;Fredriksson-Ahomaa *et al.* 2006;Laukkanen *et al.* 2009). Since *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 is geographically and phylogenetically distinct from the so far sequenced O:8/1B isolate (Thomson *et al.* 2006), the purpose of this study was to determine the complete genome sequence of the European serobiotype O:3/4 DSMZ reference strain Y11 (isolated from a patient stool) and to address the host-specificity and susceptibility for *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 infection. In addition, the draft sequencing of two other human serobiotype O:3/4 isolates, strains Y8265 (patient isolate) and Y5307 (patient isolate associated with reactive arthritis) as well as a closely related *Y. enterocolitica* subsp. *palearctica* serobiotype O:5,27/3 isolate, strain Y5,27, was performed. Together with the draft genome sequencing of two biotype 1A strains, a nosocomial serogroup O:5 isolate (patient isolate derived from an outbreak of diarrheal disease in Newfoundland, Canada) and an environmental serogroup O:36 isolate (both performed to analyse the ambivalent pathogenic profile of clinical and non-clinical BT 1A strains), these draft genomes were used to add more information value to the *Y. enterocolitica* subsp. *palearctica* characteristics. The completed genome sequence of strain Y11 (O:3/4) was mainly compared with the complete strain 8081 (serobiotype O:8/1B) genome available at NCBI (Accession numbers NC_008800, NC_008791 (plasmid)). Initial experimental approaches build the foundation for phenotypic analyses that now have to supply the theoretical genome comparison findings with relevant applied investigations.

6.1. Applied sequencing strategies for *Yersinia enterocolitica* genomes

6.1.1. Revision of the complete genome sequencing of Y. enterocolitica subsp. palearctica strain Y11 (O:3/4)

The complete genome sequencing project of *Y. enterocolitica* subsp. *palearctica* DSMZ strain Y11 (serobiotype O:3/4) has been initiated in 2003, when whole genome sequencing methods were developing but still high in consumption of money and less reliable in sequencing quality. The first sequencing run performed (MegaBACE) was based on a vector clone library containing fragments of the genomic DNA of strain Y11. Even though the read length for this method are outstanding compared to the pyrosequencing average read length, the quality of base calling especially for extended reads is low, and the clipping (rejection) of contaminating vector

backbone sequences was time consuming and inaccurate. In addition, the tools generated for the assembly of large data sets as generated by complete genome sequencing were in the early stages (for example, the Staden package) - thus demanding manual data curation and elimination of program bugs. The performance of the additional sequencing run in 2007 (454 pyrosequencing, GS-20) generated even more data that were of a novel file type and thus not ready for the implementation into the Staden software tools. Even though, the combination of the two datasets could be used for first whole genome comparisons with the genome of the so far sequenced O:8/1B isolate, and several key genes that could be of interest for host an niche adaptation of *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 were identified, including the *rtxA* gene. In addition, the datasets were screened for pYV plasmid sequences and gaps have been closed by PCR based sequencing briefly after the sequencing project was supported by the BMBF grant FBI-Zoo. Since the sequencing techniques improved rapidly and bacterial genome sequencing biases were rated more critical, another sequencing run was performed in 2009 (454 GS FLX Titanium), generating the most bases of all three runs (105,539,453 bp with average read length of 438 bp). For this sequencing run, strain Y11 was cured of the pYV plasmid to obtain only chromosomal reads, since at this time the plasmid was assembled into a single contig. But after the last sequencing run the updated 454 mapper software unravelled the first assembly of pYV sequences of the old datasets to be of poor quality, and doubtful positions had to be resequenced in 2010, based on PCR product sequencing. For the chromosomal replicon, sequences were assembled using the Lasergene software (DNAstar), and based on the comparison of the generated contigs with the reference genome of strain 8081, prediction maps of the contig localisation and orientation were generated and gap length were estimated. However, primers designed to amplify the respective gaps failed immoderate. This issue most likely occurred by inappropriate amplification conditions, regarding primer design or the polymerase enzyme and program used. Many of the gaps were prolonged by the insertion of IS elements, which hampered the amplification under the selected cycle conditions. A calculation of costs was convincing to release the gap closure process to industrial approaches. The output was of high quality and reduced the genome fragments to two contigs, leaving only two gaps. With an additional analysis of the reference genome of strain 8081, one of the gaps could be closed manually by PCR. Sequence analysis of the remaining contigs ends uncovered nearly perfect matches to other regions of the contigs, indicating a repetitive region. PCR experiments proved the duplicated regions (by the verification of minor differences between the duplicates) and sequence assemblies predicted the presence of a second P2-like prophage in the remaining gap. Since the prophage genome was far too large to be covered by PCR, a fosmid library was initiated that can carry up to 40 kb of DNA, feasible to cover the complete prophage with a single clone of the library. About 500 clones are suitable to cover the complete genome sequence of 4.5 Mb in size by chance, and so a library of

960 clones was more than enough to detect a suitable clone carrying the complete prophage of the remaining gap. Plasmid based sequencing thus closed the last remaining gap. In retrospection, bacterial whole genome sequencing has developed rapidly the last years, reducing not only the cost of sequencing projects (to less than 1/10 of the early projects), but also the manual efforts and method restrictions, since the selection of methods and software tools is now optimised for the different approaches, based on the growing experience (discussed in more detail below).

6.1.2. De novo sequencing of bacterial genomes - strategies and efforts

The sequencing methods available today are more reliable in base calling quality and generate large amounts of raw data in short time, reducing the necessary sequencing runs and thus time and costs. Thus, the number of completed bacterial genomes increased significantly the last years (see Figure 6.1). Depending on the amount of repetitive sequences and homopolymeric stretches of the bacterial genome, the invention of paired end sequencing or the Illumina technique simplifies the efficient de novo sequencing. The reversible terminator-bound dNTPs of the Illumina method can extend the new generated DNA strand by only one nucleotide at a time, minimising the incorporation bias in homopolymeric stretches. The technique is mainly used for paired-end sequencing, which provides eminent additional information about the orientation, distance and context of each output read, thus reducing the effort in gap closure and enhancing the success in PCR performance. Likewise, the software tools for data curation and assembly improved and increased. Together with the improved hardware possibilities, the tools can handle now even largest data sets more rapidly and accurately, generating also helpful quality revisions and automated primer design. Thus, the gap closure phase can be markedly reduced by the synchronisation of data assembly, gap prediction and quality control.

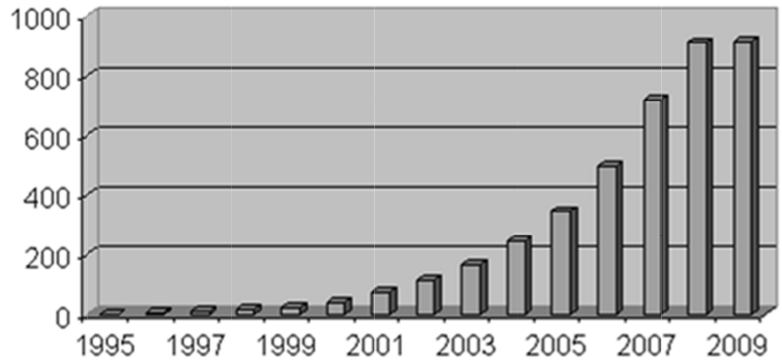


Figure 6.1 Number of completely sequenced genomes deposited at GenBank over the years. This statistical illustration has been obtained from the Genomes On Line Database (GOLD), available on http://www.genomesonline.org. More statistics about GOLD and completely sequenced genomes have been published recently (Liolios *et al.* 2010).

6.2. Complete genome comparisons

6.2.1. Main findings - novel and expected genomic features of Y. enterocolitica subsp. palearctica strain Y11, serobiotype O:3/4

Y. enterocolitica subsp. *palearctica* strain Y11 lacks several of the defined high-pathogenicity associated features of the mouse virulent *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B strain 8081, including the High-pathogenicity Island (HPI) and the type two and type three secretion systems that are part of the so-called Plasticity Zone (Howard *et al.* 2006). Also within the large PZ of strain 8081, the arsenical resistance operon is located, a putative haemolysin and the ferric enterochelin operon, all absent from *Y. enterocolitica* O:3/4. In addition, many genomic islands and prophages described for strain 8081 are missing. As described previously (Skurnik 2003), also the genetic regions coding for the enzymes involved in the biosynthesis of the different O-antigens differ markedly between the two strains. Interestingly, also yet undescribed putative virulence factors of *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B were found to be absent from strain Y11, including the putative haemolysin gene YE2966, the haemolysin gene cluster YE4084-YE4086 and YE1873, a putative adhesion protein. These so far disregarded putative virulence genes could be promising to describe additional phenotypic differences between both strains. In contrast, *Y. enterocolitica* subsp. *palearctica* strain Y11 carries putative virulence associated determinants that are absent from *Y. enterocolitica* subsp. *enterocolitica* strain 8081, as the RtxA toxin, a dual functional insecticidal toxin gene cluster (Tc) and a flag-2 flagellar system, several putative invasins and haemolysins, a beta-fimbrial gene cluster and a chromosomal type three secretion system (Ysp-T3SS) different from the Ysa system of strain 8081. The effectors of this novel T3SS have to be identified. Another interesting aspect of the genomic features of strain Y11 is the presence of several toxin-antitoxin systems that are all absent from strain 8081. These systems have been described to play a role in the bacterial stress management and may therefore reflect niche adaptation processes. Likewise, a set of metabolic clusters has been found which is absent from *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B strain 8081, promising to mirror the lifestyle of the serobiotype O:3/4 (see section 6.5). The pYV virulence plasmid was found more related to the virulence plasmid of the serobiotype O:9/2. The presence of an arsenical resistance gene cluster was apart from that a major difference in the genetic content between the virulence plasmids pYVeY11 and pYVe8081.

It was not surprising that the genome of *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 strain Y11 did not harbour a specific gene or gene cluster which can be adopted to explain the particular niche adaptation, but in fact there is a complex pattern of altered or new virulence factors found which can play a role in pathoadaptation and niche development. In addition, the

regulation of particular virulence genes has also an important effect on pathogenicity and the pathogens life style. It has been shown recently that the insertion of an IS element in the promoter region of *invA* completely changed the gene regulation, resulting in severe phenotypic changes (Uliczka *et al.* 2011). Thus, even small changes in the genetic context can have a strong effect on virulence gene expression and evoke important virulence property alterations, which may in turn be of advantage in the adaptation to the host environment. Taken together, the deciphered *Y. enterocolitica* subsp. *palearctica* O:3/4 genome demonstrates the absence of the high-pathogenicity-associated features of the mouse virulent *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B but contains an alternative pattern of virulence and fitness-associated determinants that might be involved in the on-going process of pathoadaptation to its main host, the pig, and successful dissemination of this emerging zoonotic enteropathogen.

6.2.2. Limitations of genome comparisons

Even though sequencing methods and efficiency of the sequencing runs has been improved over the years, comparative genomics experiences different limitations. The most important is the massive amount of information that needs extensive data mining, self-evidently increasing extremely with the number of genomes compared. This issue has been immediately addressed, and numerous software tools have been invented to visualise complete genome comparisons (as, for example, ACT and Mauve). However, these programs can't provide the scientists with important but inconspicuous findings as premature stop codons or single nucleotide differences in the promoter sequences of specific genes, which would not even be found with a protein comparison tool. Therefore, a manual inspection, preferentially hypothesis-driven, of critical genes and clusters is indispensable in the comparison of complete genomes. Another limitation is the gene prediction and annotation of the sequenced genomes. Automated annotation tools such as RAST provide a reliable tool for quick and basic annotations, but since the program depends on so far annotated proteins the rate of misleadingly annotated genes is high. For example, genes that have been identified to encode major pathogenicity factors may be annotated with a more general designation or marked as a "putative" protein, although a related gene has been connected with an established and adequate designation. To circumvent the problem of unequal annotations made by different automated programs or due to manual changes and slightly differing gene predictions (based on alterations in the bioinformatics' parameters), complete genome comparisons between serobiotype O:3/4 and O:8/1B have been performed using an automated RAST annotation of strain 8081 (O:8/1B), although the strain has been published annotated before. In addition, as the programs are guided by known genetic functions, a large number of genes is still annotated as "hypothetical", likely obliterating a number of interesting unknown virulence or fitness genes. From 4,355 CDS in the serobiotype O:3/4 strain Y11 genome, 623 CDS annotated as hypothetical

proteins were found, constituting 14.3 % of all annotated CDS. Today, the manual inspection of automated annotation forms a bottleneck for high-throughput genome projects. For the genome of strain Y11, the manual curation of the automated RAST annotation is still premature and awaits further interaction. But beside the method associated limitations in comparative genomics, it should be not forget that most of the detailed information about genomic differences will be invalid without pertinent, hypothesis-driven wet laboratory based research.

6.3. Two promising toxins of *Yersinia enterocolitica* O:3: bacteriocin and *rtxA*

6.3.1. Experimental setups and perspectives for the bacteriocin

Since bacteriocins are potent toxins produced by many bacteria to inhibit the growth of closely related bacterial species (reviewed in Cascales *et al.* 2007), the presence of a specific bacteriocin gene cluster in *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 (Y11_33511-33531) has been assumed to contribute to the niche adaptation and defence. Bacteriocins vary in their mode of action and specificity. The strain Y11 bacteriocin protein (Y11_33511, 905 amino acids) shows sequence homology to pyocin-like proteins and DNAses, thus likely being a potential endonuclease enzyme (reviewed in Michel-Briand and Baysse 2002). It was supposable to address the role of the bacteriocin by screening closely related species for sensitivity by overlay assays. However, none of the species tested was inhibited in growth by the presence of strain Y11. It cannot be excluded that the single representatives of one species mask the sensitivity of other serobiotypes of the species, as described previously (Strauch *et al.* 2001). The expansion of the strain collection used for the screening, including defined species of the pig's epithelium flora, could reveal the susceptible serobiotypes. In addition, a forced induction of bacteriocin expression by mitomycin C or H_2O_2 as well as lowering the susceptibility of tester strains by using iron depleted medium could be promising (reviewed in Michel-Briand and Baysse 2002). Even though the target organisms were not resolved, the bacteriocin of strain Y11 indicates a severe toxic effect on bacterial cells. All plasmid based cloning and over expression experiments were abolished, even using tight vector backbones, which indicate that even small amounts of the putative endonuclease killed the host cells immediately or led to extensive insertion of point mutations in proline coding residues. Finally, the translation of the bacteriocin mRNA *in vitro* was performed successfully and the protein could be detected on Western blots, but the activity of the protein could not be proven with the parameters used in the experimental setup (different buffers and temperatures, use of DNA and bacterial cells as targets). This can be either depending on a misfolded protein produced in the cell-free assay (nothing is known about post-translational modifications, chaperones or other enzymes required for the correct folding) or on the parameters and buffers used. Since motif predictions indicate endonuclease activity, structural analyses of the

in vitro translated protein could evaluate the correct folding of the protein. Beside the cloning expenses, mRNA experiments were performed to investigate the regulation of the bacteriocin gene, but the conditions needed for transcriptional activity of the genes could not be resolved. Further coculture experiments with a wider set of different defined parameters, including mitomycin C induction, could successfully reveal the regulation mechanisms. In addition, marker gene fusion experiments using the promoter sequence of the bacteriocin may circumvent the methodical fragility of the RNA based methods.

6.3.2. Experimental setups and perspectives for rtxA

To address the role of the RtxA protein in the pathogenicity of *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 strain Y11, a knockout mutant of *rtxA* was generated. We conclude that serobiotype O:3/4 is more restricted in the uptake of foreign DNA compared to serobiotype O:8/1B, as established approaches using the RED recombinase system failed (Datsenko and Wanner 2000). It has been assumed that the limitation in the introduction of foreign DNA could be due to the presence of specific restriction modification systems, but until now no promising candidate genes have been identified. In addition, the performance of DNA transformation in the presence of type one restriction inhibitor reagents (Epicentre) failed in the same manner as the application of DNA fragments that have been passaged in strain Y11 as a plasmid before they were re-digested and transformed into electro-competent strain Y11 cells. Thus, conjugation remained as the method with the most recombinants. This method required the use of a nalidixic acid resistant derivate of strain Y11 for selective purposes. This resistance feature led to an elevated general resistance to other antibiotics, hampering the transformation of the mutant with other plasmids as the complementation with *rtxA*. In addition, conjugation performed with *E. coli* at 37°C somehow evoked a loss of the pYV virulence plasmid, which was reintroduced using conjugation with cointegrates. To eliminate these disadvantages, the use of CIN agar plates instead of nalidixic acid as a selective pressure was promising, even though recombinant clones grew less efficient on these triple antibiotic agar plates. Cooperation partners using the CIN agar plates did not experience the loss of the pYV plasmid either. The *rtxA* mutant derivate of strain Y11, together with a wild type derivate containing the pYV cointegrate and the complementation strain is now ready for experimental approaches in cell culture and animal models. However, since the putative effector domains of the *rtxA* show no homology to any known genes, enzymes or proteins (reviewed in Linhartova *et al.* 2010;Satchell 2007), the role of this gene in pathogenicity is just speculative. Secondary, the mRNA experiments demonstrated an unrestricted transcription of the gene, which was not expected for toxins or even metabolic genes of this size. These aspects suggest a deliberated experimental setup, as the use of a mouse model prior to the more expensive and elaborated pig model. The MyD88 knockout mice are deficient in the MyD88

adapter protein that is involved in both the Toll-like receptor and IL-1 receptor signalling pathways in the innate immune response (Takeuchi and Akira 2002). Different experiments revealed this mouse lineage to be more susceptible to infections with *Yersinia*. In addition, the use of mice pre-treated with desferrioxamine could overcome the necessity of pig model experiments (Autenrieth *et al.* 1994;Robins-Browne and Prpic 1983). Thus, the serobiotype O:3/4 strain Y11, which is normally not able to kill mice could be tested in this model to obtain first results addressing the question of altered pathogenicity in the *rtxA* mutant. To obtain reliable results especially for the outcome of the infection, general infection studies with the Y11 wild type strain are highly recommended. Another approach would be the isolation of the putative effector domains and their purification for biochemical and structural analyses. Most Rtx toxins are analysed using the isolated effectors; the *Vibrio cholerae* Rtx toxin for example causes depolymerisation of the actin cytoskeleton through covalent actin crosslinking only after the autoprocessing of the effector domains of the protein, thus requiring the use of purified proteins for activity experiments (Fullner and Mekalanos 2000;Sheahan *et al.* 2007;Sheahan and Satchell 2007). The proposed model for the processing of the RtxA toxin supposes the binding of RtxA to specific receptors, followed by endocytosis. This requires the use of suitable cell types exposing the specific receptor structures on their surface. As long as the receptors are not identified, the selection of the appropriate cell line remains a game of chance. LDH (lactate dehydrogenase) assays used for infection experiments with the plasmid-less derivate of strain Y11 (wild type) and the *rtxA* mutant did not show any cell toxicity (data not shown). The use of isolated effector molecules injected into mammalian cells could circumvent this issue of cell specificity and give first results regarding cell toxicity. Another promising aspect is the use of inositol hexakisphosphate ($InsP_6$), since the autoprocessing of the toxin is induced by $InsP_6$ binding (see the postulated model of MARTX toxin processing in Figure 6.2, (Reineke *et al.* 2007;Satchell 2007)). An *in vitro* cleavage assay could be performed according to the method of Prochazkova and Satchell (Prochazkova and Satchell 2008) to prove the autoprocessing of the protein and to determine the size of the effector molecules on SDS gels.

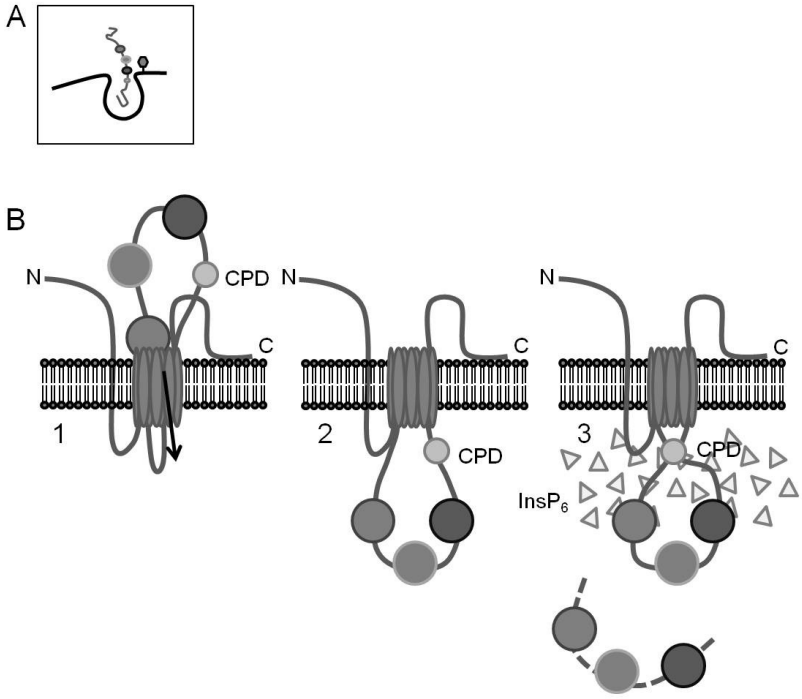


Figure 6.2 Postulated model of MARTX toxin processing (modified from Satchell (Satchell 2007) and Reineke (Reineke *et al.* 2007)). A. The secreted Rtx toxin binds to a particular receptor presented on the host cell surface, leading to the internalisation of the protein. B. The repeat containing regions of the protein that flank the effector domains are assumed to integrate into the host cell membrane to facilitate the translocation of the effector domains and the CPD motif (cysteine protease domain) into the cytosol of the host cell (1/2). After the translocation of these domains, inositol hexakisphosphate ($InsP_6$) located at the inner cell membrane builds a binding pocket that allows the conformational change of the protein, thus activating the autocleavage by the CPD motif and the release of the effector molecules into the cytosol.

Another question which was not finally addressed in this study is the secretion of the RtxA toxin by the atypical type one secretion system (T1SS) cluster. The genes for the secretion in strain Y11 are interrupted by a premature stop codon, resulting in the presence of four smaller genes. However, it can't be excluded that the system is still functional or replaced by other secretion mechanisms. The secretion signal for RtxA is supposed to be located at the C-terminus of the protein. A fusion of the putative signal sequence for export by the T1SS appropriate reporter genes could unravel the question of functionality of the RtxA secretion in strain Y11. Since the genes for RtxA transport are present also in *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B strain 8081, this strain could be used as a promising control of the fusion approach.

6.4. Genome plasticity and phage mobility in *Yersinia enterocolitica* O:3/4

The presence of genomic islands, prophages and IS elements has a deep impact on the genome plasticity of *Y. enterocolitica* and reflects the pathogens' interactions with other species and their host or niche environment. As reported previously, several genomic islands (namely, HPI, PZ,

YGI-2 and YGI-3) and prophages of serobiotype O:8/1B strain 8081are absent from the genome sequence of serobiotype O:3/4 strain Y11. In contrast, strain Y11 harbours its own, unique collection of incorporated mobile elements.

6.4.1. Genomic Islands and prophages

The novel genomic island GIYep-01, found in the genome of strain Y11, occupies the same *asn* tRNA as the HPI in *Y. enterocolitica* subsp. *enterocolitica* serobiotype O:8/1B strain 8081 (Thomson *et al.* 2006). This 14.9 kb island encodes a putative metallo-β-lactamase and a protease. These factors might be involved both in fitness and pathogenicity of yersiniae. For its integration, GIYep-01 utilizes a P4-like integrase, likewise to most genomic islands, including the HPI. However, the integrase of the integrated HPI in serobiotype O:8/1B is inactive (Rakin *et al.* 1999), whereas the GIYep-01 integrase displays recombination activity being able to precisely excise the island with the restoration of the initial *attB* attachment site. Whether the integrase of one mobility element can affect recombination of another and thus arrest its mobility still has to be clarified. In any way, the acquisition of the HPI by the *Y. enterocolitica* subspecies that was followed by the inactivation of its integrase and multiple rearrangements in its AT-rich part seems to be a more ancient event compared to acquisition of the GIYep-01 island, which is still mobile, by the *palearctica* subspecies. In contrast to the HPI, the GIYep-01 has an average GC content similar to that of the *Yersinia* backbone genome sequence, indicating that it might be a result of a recent acquisition from a microorganism with a similar GC content profile or a melioration of the acquired sequence. The possible role of this novel genomic island in virulence and fitness awaits its elucidation. Future experiments could address the transcriptional regulation of the putative metallo-β-lactamase by mimicking different conditions that could be suitable for the activation of transcription (mRNA detection) and over-expression experiments to determine the catalytic function of the enzyme. In addition, the potential of this mobile genomic island to insert at another *asn* tRNA of serobiotype O:8/1B (or the potential to occupy the resident *asn* tRNA of the HPI) could be investigated by conjugation experiments with the circularised, excised island.

The filamentous prophage PhiYep-1 was found in all three sequenced *Yersinia enterocolitica* subsp. *palearctica* isolates of serobiotype O:3/4. This prophage has a high similarity to the *Y. pestis* Ypf prophage and the CUS-1 prophage disseminated in pathogenic *E. coli*. Both Ypf and CUS-1 prophages are suspected to play a role in the pathogenicity of their bacterial hosts (Gonzalez *et al.* 2002). However, the PhiYep-1 prophage has suffered different substantial deletions in serobiotype O:3/4 strains Y8265 and Y5307, leaving only a few truncated ORFs in the latter one, barely alluding to the prophage origin. Thus, the possible impact of this prophage on the pathogenicity of *Y. enterocolitica* subsp. *palearctica* serobiotype O:3/4 is questionable. The

comparison of more genomes of this serobiotype would be appealing to unravel the mechanisms and sequences of degradation of this prophage. This PhiYep-1 prophage might be also associated with the rapid elevated resistance to ampicillin in *Y. enterocolitica*, since the presence of elevated copy numbers of the *blaA* gene (Seoane *et al.* 2003) was found to be coupled with likewise elevated copy numbers of parts of the prophage (this study, see 5.5.6.3). The generation of serobiotype O:3/4 strains that lack even the smallest remnants of the prophage could shed light into the impact of the phage on the elevation of ampicillin resistance, and the mechanism underlying the duplication of large genomic sequences associated with resistance could be at least curtailed.

Two copies of highly similar P2-like prophages are present in the genome of strain Y11 (see Figure 6.3). PhiYep-2 inserted into *met* tRNA while PhiYep-3 integrated into *leu* tRNA. It could be demonstrated that PhiYep-3 is, in contrast to PhiYep-2, able to excise precisely with high frequency and to leave the intact *attB* attachment site in the bacterial host. In contrast, the integrase of the PhiYep-2 is truncated by a stop codon and thus is inactive. However, the PhiYep-3 prophage is absent in both other sequenced serobiotype O:3/4 isolates, strains Y8265 and Y5307. Thus, the mobile PhiYep-3 prophage is likely to represent a more recent acquisition of strain Y11 and serves as an additional epidemiological marker for a certain group within the *Y. enterocolitica* subsp. *palearctica* strains. The coexistence of two closely related P2-like prophages and their potential of immunity to superinfection raise an additional question to be answered.

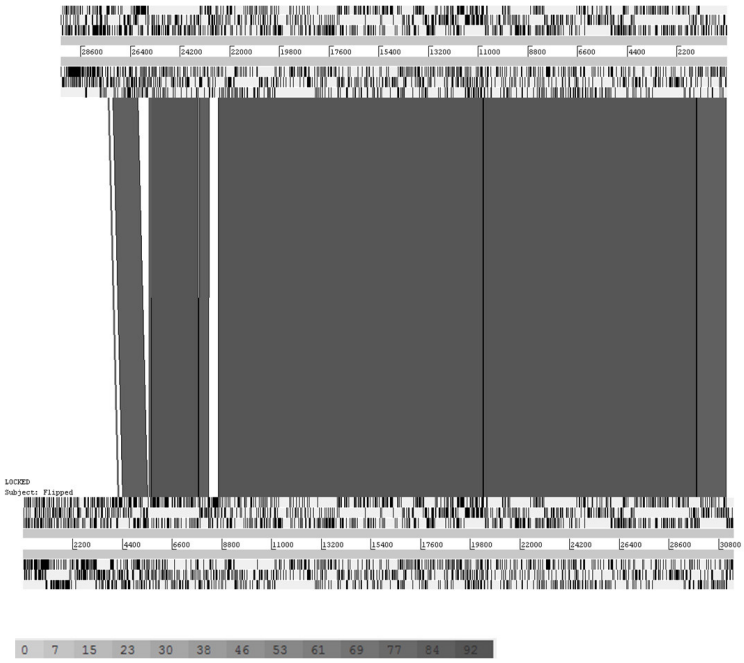


Figure 6.3 PhiYep-2 (upper panel) and PhiYep-3 (lower panel, sequence flipped) comparison. The illustration was generated using ACT. The region in the beginning of the flipped PhiYep-3 sequence contains the integrase, which is truncated and inactive in PhiYep-2. The scale below the comparison window shows the colour code with numbers indicating the identity per cent.

6.4.2. IS elements

Beside the mobile genomic elements, also IS elements take part in shaping the genome of strain Y11. Multiple IS elements found in strain Y11 can tell us a different story of its interactions with neighbours. The spectrum of IS elements differs in the serobiotype O:3/4 and O:8/1B strains, with ISYen2A/B being a *palearctica* subspecies-specific insertion sequence. In contrast, a wide variety of IS families IS3, IS4 and IS200 (also IS family IS110 (ISYen1), which is predominant in serobiotype O:3/4) constitutes the high-pathogenicity specific group of the *enterocolitica* subspecies (see 5.5.6.4). These differences can be used not only for subspecies identification but also for tracing the history of interbacterial contacts throughout the history. The presence of similar sets of IS elements and their evolution in different bacterial strains may reflect their potential close contact over the time by similar niche occupation. In addition, the excessive presence of a single IS family member (ISYen1, found 53 times in strain Y11) leaves open questions, especially regarding the impact on gene interruption or promoter sequence alterations, as reported for *invA* (Uliczka *et al.* 2011). It is likely that the presence of multiple IS element copy numbers may force evolutionary adaptation to novel environmental conditions or niches.

6.4.3. Mobile genetic element hot spots

Genome plasticity based on the insertion of mobile elements is well known to be accumulated at specific insertion sites as tRNA genes. But it is remarkable that several gene clusters that do not show any association with phages or mobile elements tend to occupy the same positions (hot spots) in the *Y. enterocolitica* backbone genome as independent clusters with potentially closely related functions in other representatives of the species. For example, the two different T3SS Ysp and Ysa are located in the same genomic background in strains Y11 and 8081. Similar hot spots seem to exist for the O-antigen clusters, the AidA adhesin, a haemolysin and others (this study and Thomson *et al.* 2006). However, the remarkable divergence in these clusters might be the result of both vertical and horizontal evolution changes.

6.5. Metabolic differences as a niche adaptation evidence in *Yersinia enterocolitica* subsp. *palearctica* serobiotype O:3/4

The genome of *Y. enterocolitica* subsp. *palearctica* strain Y11 harbours several subspecies specific metabolic features that are absent from the serobiotype O8/1B strain 8081. Strain Y11 carries *nrf* genes for nitrite reduction (Y11_04801-04871), a second urea transportation gene cluster (Y11_22281-22341), a *dms* DMSO reductase gene cluster (Y11_22241-22261), a choloylglycine hydrolase enzyme (CGH) gene (Y11_23571) and the N-acetyl-D-galactosamine PTS gene cluster (Y11_11961-Y11_12031). The ability to reduce nitrite should be compared between strain Y11 and strain 8081 to investigate the functionality of this gene cluster. Since the verification of nitrite reduction is sensitive, a special anaerobe experimental setup has to be applied for nitrite reduction experiments. In addition, the function of the putative CGH enzyme could be analysed for its impact on resistance to antimicrobial bile salts. It can be speculated that this enzyme could take part in the ability of the bacteria to infect its host through the oral route. The function of the second urea transport gene cluster can be addressed only using mutants with deletions in either one of the gene clusters or in both urea transport clusters. For the N-acetyl-D-galactosamine PTS gene cluster it could be demonstrated that, in contrast to serobiotype O:3/4 strain Y11, serobiotype O:8/1B strain 8081 is not able to utilize N-acetyl-D-galactosamine (GalNAc) amino-sugar as a carbon source, most probably due to a deletion of the genes responsible for this trait. The amount of GalNAc, that is one of the main components of the small intestinal mucin, is twice higher than that of the other sugars in the pig's intestinal mucin and higher than in the human mucin (Mantle *et al.* 1980). Moreover, mucin is an important defence factor against bacteria in the gut. The complete *aga*-operon was subcloned and transferred to strain 8081 and this was sufficient to support growth on GalNAc as the only carbon source. Since the availability of mucin is elevated in pigs as in the main host of serobiotype O:3/4, the ability of

strains of this serobiotype to use this aminosugar might reflect one of the possible mechanisms for its specific host adaptation.

6.6. *Yersinia enterocolitica* subsp. *palearctica* biotype 1A: the opportunistic pathogen

The draft genome comparison of two *Y. enterocolitica* subsp. *palearctica* biotype (BT) 1A strains, a hospital and environmental one, confirmed the absence of the established classical *Y. enterocolitica* virulence markers of the pathogenic serobiotypes (reviewed in Revell and Miller 2001) in both isolates. This might explain the restricted pathogenic potential that has been associated with BT 1A isolates, neglecting their pathogenic potential for a long time in history. Besides the pYV virulence plasmid, BT 1A strains lack the genes for the Ail adhesin, the YstA enterotoxin (instead, 1A group strains acquired a close related YstB enterotoxin), a serine protease, a leukotoxin, protein VapC and VagC, as well as a putative HipA regulator. However, both 1A isolates still possess at least closely related genes of other well-established yersiniae virulence-associated genes like *invA*, *myfA*, *ymoA*, and *hreP* that could play a considerable role in virulence. Infections evoked by BT 1A strains often result in symptoms that are indistinguishable from that produced by the pathogenic *Yersinia* biotypes (Burnens *et al.* 1996;Morris *et al.* 1991). Thus, it is important to understand the underlying genetic mechanisms that coordinate few known virulence-associated genes to create an infection indistinguishable from other biotypes. Interestingly, only few differences between the environmental and clinical isolates have been found, indicating it is most likely that the interplay of the BT 1A virulence gene pattern with an altered environment or host defence system could have the main influence on the establishment of an infection.

Pili adhesins are involved in a variety of surface interactions including the colonization of both biotic and abiotic surfaces, biofilm formation and the initial contacts with host cells. This wide scope of interactions implies sufficient variability of the MyfA fibrillae structure in bacterial species exploiting different ecological niches. The virulence associated gene *myfA* has been reported to be present only in the pathogenic serobiotypes of the *Yersinia* and thus served as a reliable virulence marker (Gierczynski *et al.* 2002). The present genome sequencing of the two BT 1A isolates revealed instead the presence of *myfA* homologues in both the clinical and environmental strain. The MyfA amino acid sequence is highly conserved in the two pathogenic yersinial serobiotypes O:3/4 and O:8/1B, but occurs rather diverged in both 1A isolates. The comparison of the genes involved in regulation and assembly of *myfA* indicated the co-evolution of the assembly machinery and transcription regulators together with the fimbriae secretion system, although to a different extent. This extreme sequence variability may explain multiple

failures in the detection of the *myf* sequences in non-virulent *Yersinia* serobiotypes (Gierczynski *et al.* 2002).

Biotype 1A stains that are classified by 16S rRNA sequencing as *Y. enterocolitica* subsp. *palearctica* have mosaic genomes and share genes both with highly pathogenic BT 1B group and the low pathogenic BT 3 and BT 4 groups. However, the BT 1A group is phylogenetically closer related to the BT 2-5 groups than to BT 1B and BT 1A is not pathogenic in the mouse model (Heesemann and Laufs 1983). Only strains of 1B and 1A biotypes carry ribose transport and the *arbBFG* glucoside utilization operons. This could hint at a potential wider scope of environmental interactions between biotype 1B and 1A strains. On the other hand, BT 1A possesses a lot of genes shared with the low-pathogenic biotypes but absent in BT 1B, including the genes coding for the low pathogen-specific Ysp-T3SS, the second urea-transport system, specific beta fimbriae genes, the Aga proteins for N-acetyl-D-galactosamine utilisation, the HlyD secretion protein, a putative cytolethal distending toxin and others. The presence of the insecticidal toxin gene cluster, the RtxA-like toxin, and three putative pertactin-like genes turns out to be the hallmark of the low-pathogenic *Y. enterocolitica* subspecies *palearctica* representatives and supports different life styles and ecological niches of these closely related *Y. enterocolitica* representatives.

Not any evident structural differences were found between the genomes of the nosocomial and environmental isolates that might be responsible for the pathogenicity of the hospital O:5/1A isolate. The possible impact and activity of the defined minor differences between the two isolates, like the presence of the *rtxC* and *rtxH* genes in O:5, have to be verified by further experiments as mRNA detection and mutagenesis. However, it seems unlikely that these few genes alone might be responsible for the virulence phenotype of the clinical 1A isolate. Thus, the environmental, non-clinical 1A isolate possesses the same pathogenic potential as the one derived from a clinical outbreak. Taken together, the *Y. enterocolitica* subsp. *palearctica* BT 1A group represents opportunistic pathogens whose opportunity to establish an infection seems to rely mainly on the state of the host defence system. The absence of the main established virulence factors restricts BT 1A strains in their ability to overcome the host defences and to efficiently multiply within the host. However, BT 1A strains still contain virulence-associated factors that might cause an infection with symptoms similar to that caused by the pathogenic BT s, when the first line of the host's defence is weakened. Thus the scope of virulence markers found in BT 1A strains is sufficient to cause a clinical outcome similar to that of the virulent *Yersinia* and thus it is important to reconsider the pathogenic potential of this group of emerging pathogens.

6.7. The *palearctica* subspecies

As expected, the genomes of the *Y. enterocolitica* subsp. *palearctica* strains do not carry the already defined high-pathogenicity associated features of the O:8/1B strain (e.g. the HPI and type two and type three secretion systems that are part of the Plasticity Zone). The absence of these determinants might in part be able to explain the inability of *Y. enterocolitica* subsp. *palearctica* strains to kill mice and to establish generalized infections in contrast to the highly virulent *Y. enterocolitica* subsp. *enterocolitica* strains (Pelludat *et al.* 2002Pelludat *et al.* 2002). The *palearctica* subspecies seems to be less armed with established virulence factors when compared with the *enterocolitica* subspecies isolates, but this decreased pathogenicity may enable this subspecies to balance its interactions with the host to support further dissemination. However, the *Y. enterocolitica* subsp. *palearctica* group is not devoid of all potent pathogenicity factors. Besides the highly similar pYV virulence plasmid shared with most *Yersinia* species, *Y. enterocolitica* subsp. *palearctica* strains carry many of the well described virulence genes, including e.g. *myfA* and *invA*. Also a set of different group- or serobiotype-specific putative virulence associated determinants like the RtxA toxin, a dual functional insecticidal toxin gene cluster, several putative invasins and haemolysins, a beta-fimbrial gene cluster and an alternative to the *Y. enterocolitica* subsp. *enterocolitica* Ysa chromosomal T3SS make up the virulence pattern of this group. Some of these subspecies specific genes and a large number of hypothetical genes that are potentially virulence-associated need further investigation in wet laboratory experiments. Nevertheless, the heterogeneity of the *Y. enterocolitica* subsp. *palearctica* group, especially underlined by the presence of putative pathogenicity-associated genes shared only between *Y. enterocolitica* subsp. *palearctica* BT 1A strains and the *Y. enterocolitica* subsp. *enterocolitica* strains make it difficult to finally define a complete group-specific virulence profile. The increasing number of sequenced isolates of the *Yersinia* species may further refine the set of genes designated to be subspecies-specific and the experimental investigation of putative virulence genes will shed light into the interplay of particular genes to establish an infection, providing new models for subspecies specific virulence pattern.

7. Supplementary material

7.1. Abbreviations

(v/v)	Volume per volume
(w/v)	Weight per volume
°C	Degree in Celsius
µ	micro
Å	Ångström
aa	Amino acid
ACT	Artemis Comparison Tool
Amp^R	Ampicilline resistance
AS	Ankylosing spondylitis
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BHI	Brain heart infusion
BLAST	Basic Local Alignment Search Tool
BMBF	Bundesministerium für Bildung und Forschung
bp	Base pairs
BSA	Bovine serum albumin
BT	Biotype
cDNA	Complementary DNA
CDS	Coding sequence
cfu	Colony forming units
cGMP	Cyclic guanosine monophosphate
CIN	*Yersinia* Selective Agar (with antibiotics Cefsulodin-Irgasan-Novobiocin)
Cm^R	Chloramphenicol resistance
DMEM-F12	Dulbecco's Modified Eagle Medium with nutrient F12
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Desoxyribonukleosidtriphosphate
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EGTA	Ethylene glycol tetraacetic acid
EMBL	European Molecular Biology Laboratory
et al.	Latin: et alii (and others)
ETEC	Enterotoxigenic *Escherichia coli*
F	Capacitance
FCS	Fetal Calf Serum
g	Gram/ gravitational

GalN or Gam	D-galactosamine
GalNAc or Aga	N-acetyl-D-galactosamine
GC	Two of four bases that occur in the DNA (adenine (A), cytosine (C), guanine (G) and thymine (T).
Gfp	Green fluorescent protein
GlcNAc	N-acetyl-D-glucosamine
h	Hour
HGT	Horizontal Gene Transfer
HLA	Human leukocyte antigen
IS	Insertion sequence (element)
kb	Kilo-base pairs
kDa	Kilo-Dalton
Km^R	Kanamycin resistance
L	Liter
LB	Lysogeny broth/ Luria-Bertani broth
M	Molar mass
m	Meter
M9	Minimal medium with M9 salts
MARTX	Multifunctional Autoprocessing Repeats-in-Toxin
MOI	Multiplicity of infection
n	Nano
Nal^R	Nalidixic acid resistance
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PTS	Phosphotransferase system
PZ	Plasticity zone
RAST	Rapid Annotation using Subsystem Technology
RNA	Ribonucleic acid
rpm	Rounds per minute
SDS	Sodium dodecyl sulfate
SG	Serogroup
Sm^R	Streptomycin resistance
SNPs	Single nucleotide polymorphisms
subsp.	Subspecies
TAE	Tris base, acetic acid and EDTA containing buffer
TCA	Trichlor acetic acid
TEMED	Tetramethylethylenediamine
Tet^R	Tetracycline resistance
tRNA	Transfer RNA
Trp	Tryptophan

V	Volt
wt	Wild type
X-gal/IPTG	5-bromo-4-chloro-indolyl-galactopyranoside/ Isopropyl β-D-1-thiogalactopyranoside

7.2. Sequences selected for the microarray for *Y. enterocolitica* finetyping

>putative_SeC_C_containing_protein_Y11
ATGTTAAAAAAGATTATAAAAGAGACAGGATCAAATTCATCAGAAGAAAAACTTGCTAGTTATGCCGATGATGCATTTTTGG
GTTTGTGGGCTTATCCCAATGTTTATAGTGACGAAGGTCTTAGTAAAAATAAAACAGGTGTGGAAGTCTGTGATTTATTAGTT
GTTTTTAATAATAATGTGATTATTTTTTCAGACAAGAATTGTACTTTCAATCATGATATAGATGTTGGAGTAGCATGGTCGAGA
TGGTTTAAAAAATCGGTAATTGCTTCATCAAAACAATTATATGGTGCTGAAAAATTCATTCGTGAATATCCTAATAGATTATTT
TTAGATAAAAAATGTACTGTTAATTTTCCAGTCGAAATAAAAAATGACATGCGTTTTTTTCTTATTGCCGTTACAAATAACTCA
GCAAAAAGTGCCAGCAAATATTATGATAATGTGGCTGTTGGAAGCACGGGATCTTTGATGAGTTGTTTTCATTTTAATGCTGA
>gamma-glutamyltranspeptidase_Y11
ATGATCAACAGCAACATGGCACCGCAAGGAATGGCGGTTACCCCACATCATTTAGCCAGTGAAAGTGCCCTCGCTGTATTAC
GTGAAGGCGGCAACGCAATTGAAGCAATGGTCGCAGCGGCGGCCACCATAGCGGTGGTTTATCCACATATGAATGGACTGGG
TGGTGATGGTTTCTGGCTGATTGTCCCGCCCAATGGCGATCCGATTGCCATCGATGCCAGTGGAGCCGCAGGTGAGCTCGCGA
CGCTTGAGCGCTACAGCGCATATCAACAGATCCCGCACCGTGGGCCGCAAGCCGCATTGACGGTTGCCGGAACTGTCGGCGG
CTGGGAAGAAGCGCTGAATGTTGCTCGCGAACTCGGTGAAAATACCCTGCCACTGCGCCGCTTACTGACAGATGCGATTACCT
ACGCTACACGCGGCATTCCGGTCACCCCTTCCCAAGAAGCAGCCACTCGTGCTAAATCCAATGAGCTGGCTGAATTCACTGAT
TTTAG
>hypothetical_protein_Y11
CAATTATTAGTATACAAAATCAATAATGAATATTTTTCTCTGGATAAAAAATGTGTTGAACGCGTATACCTGCGTGAAAATAG
GCCTATATTTTATTTAAATGATAATGTAATTATAGAGCTAGATAACAAAAAAGAGTGTGTGCAACAATTTTCTAAGTTTATCA
CTTCGAATATTGGTGAGCAAATGGAAATACGATTTAAGGGAGTGGAAATAGTTCCTGTTACCTGGATAAGATCTGCAATGCTT
TTTGACAATGGGATGCTTTCTCAAGCAGTGACTAAAAGGGAGATGGCATTAGAAATAGTCAAGGGATATGCTCAATAAAAAC
CCCGTGGTGTTTATATAAATGAATTTCATTATTTATAAACATAAATAAGCCACGGGATTTTATTTTACTTAATAGTTAACTCTA
ATAACCTGACCGACAGAATGACTTTTAAATCGTATTCTGTCTGTCCCGGAACCACTTGCTTTAATCTGATAATTACCCGGCCCC
A
>DNA_cytosine_methyltransferase_Y11
GATTTATCCGTTGCTGGAGCTAGAGCTGGGCTAGCGGGAGTGGAGTCTGGAACTCTTAATGCAGTACTCAAGATTCTCGAGTA
TAAGAAATCTGTCGGGGATTCCCCGAAGGTCGTGATGCTTGAAAATGTGAGTGGATTTCTTACTTCACATAACGGAAATGATG
TTGCCGAAACTGTAAGGGCTTTATCAAACCTCAATTATATCGTTGATATTATAGAACTAGATGCAGCTTTGTTCACCCCTCAAA
GTAGACCACGGGTTTTTGTGATTGCAGTTCTTAGAGATATTGCTTCATCAGTAATGCATATCAAGCAGACAGATAATATTTTC
GATGATTGGTGGCAAAAATTTGATGAACAACCTAGTTTACGTTCGAATAAACTGAAGAGAATAATAGCTTCAAACGAAGGGC
TTAACTGGGGACTGATTGACGTCCCTCTTCCAACCCAAAGAACAAGTGAACTAGCTGATATTATTGATATTGATATCCAAGAC
GA
>hypothetical_exonuclease_SbCC_Y11
AAACTGAAAGAGTTAGTTGCAGTTACTACGGAAACAGATGCAGCACTTAGCCTTATAAAAGCTCAACAAGAGATTACTCAAA
CTGAACTAAATAAAATCACATCCTTAAATATAGACTGTGATTCTCTAATAAAGAGCAGTTCTGATTTAGATGGTTTTGATCCA
GAACGTATCAAAGATTTAAATTTACTTGCTAGAGAGTTATCCACCTTACAGTCCCATGAACAAACTATTCGCCAAACTCAAAA
TACTCTTAACCAACAAATGAGTTTGCATGAGAAATTAATATCAGCAGGATTAGAATTCTTAAGTATTTCACCTACAAATATTT
GTCCACTATGCCGTACTAATCATGAATCGACATTATCCCTTAAAGCAATCATAGAAAATAATGACCTAGTGTCAGGATTGGCT
AAGGAAAACTATCGATTAATAGAACAATCTTTGCTTCAACAAAAAAATATTAATTCAAAAATGCTATTGATTGTGAGTGAAG
CTCT
>wbbX_Y11
TATTTTTCTATGGTCGCCCTTCAACAGAACGGCGCTGTTTCGATTTGGGCATGGCCTCTTTGCTGAAGATTTCGCAACATTATC
CAGACGTTGAAATTATTATTGCTGGTCTTGATTTGAAAGCGAAGCCGCCGTTTCCTGCAACGTTACTCGGCAATATGTCATTG
AAAGATACTGGCGATTTGTATCGTACATGTGATATCGGCATGGCTTTCTCTGGAACGAATCTGTCATATCTACCGGTTGAACT
CATGGCCTCCGGCGTTCCCGTTATTTCAAATAACGGCCCTCAGGTTGAGTGGCATTGTAAACATGGGGAAAATGCATATTTGG
CTGACCCAACTCCTGCATCCATGCTTGAGGCGTTTACAGTACTTTATGAGAGCAAAGAACTTCGTCAAAAACTTGTTGATGGA
GGGTTAGCTGTAATGGAACCGCTAACTTGGGAATCACAAATGACAAGTATTTATGAATATGTAAGTGATAATATACATAAAT
AA
>lysR_family_Transcriptional_regulator_Y11
ATGCGTCCGGCTCTTGATTTTAATACCCTGTAAGTGTTTATAGCCGTGGTTGTAAGTGAAAGTTTTGTTGGGGCATCGTAAGTC
CTTGAAATGCCGACATCAAATGTGAGTCGTTGTATTTCTCAGTTAGAAGAAAAACTGAATCTCCAGCTCATTGAGCGCAGCAC
TCGACACATGAAACTCACCCAGGCCGGGCACCTACTTTATACCCGAGCGAAGCCATTATTGGAGGCGCTTGAGCAAACTGAG
ACAGAATTGACGTTGCGGCAAATGCAACTCAAGGGCCCATTACGAATCTGTATTCCCAATGAAATTGGCCCAGCATTGCTGG
GTTCGATTGTTGCCGATTTCGCCTGCCAGTACCCGGATTTGGAGGTAAGCTGTGTCACAAATTTGTCAGGTTTTGAATCCCTGC
GAGATGATCTGGATTTAGCTATCATTATTACCCGTGGTCAGATGGATGACAGTGACTACATAGCGCGCCATCTGCTGACTATC
CC
>hypothetical_metallo_beta_lactamase_Y11
ATGAGTATAAGCATACGTATCCTCAAGGCGAATCATGGTGACTGCATTCTAGTTACTCATCATGCAAACGACACAGTTTTTCA
TCTATTGATTGATGGTGGAAACGGTGCAACCTTTAAGTATGGACCTCGTGGGCGTTTTAAGGGGGCACTATGCTCGATGCTGG
ATGAGTTAAAGGCAAAGGGACAGCATATAGACTTAGCAATTTTGACTCATATTGATGATGATCATATTGGTGGATTGTTAAGT
GCATTCAAAGATCCTGGTTATCTAAATGATATGGTTAAATCAATTTGGTTTAACTCTTCTCGGCTTATTACCAAACATTTTAAG
GAAAATGAAATTCCTGAAAATGATATCTACCTTAGTGATGATTCCCCTCTAACTAGTGTCAGACAAGGGAAAAAATTTGAGTC
ATTGCTTGATGAAATAGGATGTAAGAGAGGTCCCTTAATTAGTGCTGGAAAAGTAATGAATGAGGGGCCATTCAAACTTACG
AT
>hipA_domain_containing_protein_Y11
ATGAAGTTAGATGTTCAAATTAGCGGCCAAAATGTTGCCAAGCTATTTCGCGAAAGGGATGAGTACCATTTGCAATATTTACC
GGGGATATCCGAAAATAATTTTATTAGTCTAACAATGCCTGTTCGCGACTTAGCATGGCGTTGGCCGAGAGATCTTCATCCTT
TCTTTCGACAAAACTTACCGGAAGGTTATTTACTGGGTATTATTCGAGAGGAATTCGGTCCTTTATTAGACGGTACAGATCTTT
CATTATTGGCAGTCATTGGCGGAACGGGAATAGGGCGCGTTACCGTGAGCCCTGAAGGTGTGGCATCAGGTTTTGAATTAGA
TCCTTTGCAAATTGAAGATCTTCTTAAAGGCAATAATACTACGGCTCATTTTGCTTCCTTGGTTCGGCGATATGCGCGAGCAGC
AGTGTCTGGTGCCGTACCCAAGTTTTTAGCTCCAGAACGCACAAACACTGTCTCATTAGGCAAGCCTACTCTTCGAACGAGCA
A

>predicted_orf_with_similarity_to_CI_repressor_protein_Y11
GAGAATTAGCGGCTTATCTAGGTGCAGGAACGGGAACAATCAGCACTTGGATCAAACGTGATTACTTTCCGGGCAAAGAGAT
TGTGCTCTGTGCTTTAGAAACTGGTGTTTCATTACATTGGCTTGCAACTGGTGAGGGGGAACCTCAGGAACCGGTTAAACCTG
CCGCACATGAAACCGCAAAATCTATTGCTCATAAAAGCCTTGAAGATGGCTTGCTGATAGATGTATCCCCTATATTGCTAGAT
TCAGAGTTACTGCCCGTTCAGATTATTGAACCTGAATTAATTTCGTTCCCTAATGAAAAACGTTATTTTCTGGTCGAACGTCAA
TTTAGAAATGTCGCTGACGGCCTTTGGTTGATTGAAAAAGCAGGTGTGACCTCGATTAGCAATATTGCTCGTTTACCCGGTGA
TGTATGGCGTATCAATGATGTTAATTGGCCGGTTAGTGAGGTTAATATACTGGCGAAAGTTGTCGGCGAAATTACGGGCTATT
GA
>antirestriction_protein_ArdC_Y11
ATGACCATTTCCCTGCATACCCAGACTAGCGCCCCTAACGCCGCAGCGGCGACCAGCGTATCCCCGCTGGAGCAGTCAGGCTC
CTCAAAAACGAAATTTTCTAAAGCCAAAACTGATATTTATCAGACTGTGACCAATAACATCATTGCAGCGCTTAAATCCGGTG
TAAAACCTTGGTTTTGCCCGTGGAAACGAGTGCCGGGTATGTCTGGGTTGCCTTCCAACTTCGCAACCGGTATCGCGTATAGC
GGTATGAATATCATGCTGTTGTGGTGCAGTGCGTCAGAACAGGGCTTCGGTGATTCACGTTGGATGACCTACAAACAGGCGC
AGGCAGTAGGTGGGCAGGTTCGAAAAGGTGAACACGGCACGACAGCCATTTTCTACACAACCTTAGAGAAAGAAAACGAAG
ACGGTGACATCGACCAGATTCCGATGCTGAAAACCTTCAATGTGTTCAATGTTGAACAAATTGATGGCCTTCCGCTGACAACT
GAAAC
>putative_cytidylyltransferase_Y11
ATGAAAGTATATGCATTTGTTCCAGCCAAGGGTTCCAGCGAAAGAAATCAAAATAAGAACATGCGCTTTCTGGAGGGTGAAA
GACTCTACATCAAAGCACTTAAAACACTGCTGCAATGTACAGAAATCGATCGAGTGTTCCTTGATTCGGAATCCGAGGAAAT
GTACAGTCATGCCGATTATCTGCCTATTACATTTTTACAGCGTGACTCTGCCTTGGCTACCAACAAAACCGATGGGCATCAGA
TGTTCCTCAATGAAGTCAACTCATATCCTGACGCGGACATCTATGTACAACTGCTGTGCACCTCGCCATTTATTAAACCACAG
ACCATTGATGATGCAATTCGTAAGCTGTAAAGCGATCCGAGCTACGATTCAGCCATATTGATGAAAAAAGATAAGTACTATTT
CTGGAAAGATGGTATTCCCGAATATGGTCCCGACCACATACCAAACAGTAAAGACTTACCTGATACCATCTCAGAAAGTATG
GGCTT
>putative_membrane_protein_Y11
TACGTTTAGCCAGTTTCTTTTGAAGTATTATGACCATTATTGGAAATCCAGATGCGGAAATTGCACATACCGCCAGCAACATG
TTCACATTTTGTGGTGCCATACCATGATCAGGTAATAGAAAAGCAGACAGGAAAAATGCAACGGTTGTAATCAGATACATTA
TTGATGTCTGTAATGAAGAAATTCCAGCTCGTTGACTATCATCAGGAAACTGAATCGAAACTGCCGAAGATGAAACCAGGCG
ACTGTAAGATGCACCGAGAAATAAAGTCATAAATAACGCCGCATACTGGTAGCCCAGCATTGGTATTAGCAAACTCAGTATA
AAAACTAGGGTTGAACCTGTAGCCAGAATCAAAGCAGAAACACGGGAGGTTAATATGCCTGCCATTTTTGTAGACAGGTAAC
CTGTAACCCCCCCGACGAAGAACAGCCAGGGCAACAGGTCTTGCGAAGCGCCCAACTGCTGTGTCACTAATGGCACCAGAAT
AGGAATG
>putative_phage_related_membrane_protein_Y11
GGGGAAGGGAGCTTTTGCCAGTTTGGTGATTTAATTGTTATTGATGAAGCATGGCGAATCTTCGGTAGCGATAAGGATATGAC
GGCTGAGAAGAAATCATTTATTGCTGAACATCGTCATTTTACGCACCCTGAAACGGGTATTAGCTGTGATTTGGTTATTGTAA
ATCAGTCACTTTCTAATATTGCTCGCTTTCTGAAAGACAAAATAGAAACAACTTACCGGATGCGCAAGCTGAAAGCGTTGGGC
CTGAATAATCATTACTGCATTGACGTATATTCAGGCCACAAAATCTATAAAAGCAACCTCGTCACCAGTTATCGCAATAAATA
TAACCCTGATATTTTTGAACTTTATAAAAGCTATGAAGGAAATAACGGTAATGAAAAGCAGACAGATAAACGCCAGAGCATC
TGGAATTCTGGCAAAGTCAGGTTCTTTCTTGTGCTTTTTCCATTGATGTTTATCGGGTCAGGCTGGCTGATTTACTCATTTTTCA
G
>hypothetical_protein_Y11
TACTAATCATGAATCGACATTATCCCTTAAAGCAATCATAGAAAATAATGACCTAGTGTCAGGATTGGCTAAGGAAAACTATC
GATTAATAGAACAATCTTTGCTTCAACAAAAAAGTATTAATTCAAAAATGCTATTGATTGTGAGTGAAGCTCTTGACTATCGA
AGCAATAAAATATCCACCTTACAAATAAAACTAAAAGAATTGGCAGAGGAAATATCAACCGAGGTGTATAAGAATAAAACT
GCAGCTGCACAGATAAAAAATGTAACAGCGGTTCTTGATAAACTTCAAAACGAAACTTTTAATGTTGAAAAGAGTGAGTTTA
TTACAAGTATAAATAATGAAATTAATTTGTTCTCGACAGAGATTCATAAACATGAATCTATCATTAATGAATGTAGATTATTA
CTTTCCAAGAATGAAACTACAATTATTGATATAAATTCCGAATTGAATTCAGTTACCAACAAAATAAATTCATTAACTACCTC
GGAGT
>hypothetical_secretion_protein_HlyD_family_protein_O527_Y11
ACGTCCTGATGGTGCTGGATGAGGTCGATACTGACATTGCCATTGAACGTCTTAAGGCGGTTATTCGTTCTCAGGAGGCATTG
ATCCGTCAGGAGGAATCGGCTATCCGTATCGCCACCAATGAAACCACATTAACCGAGGTTTCATCCTGGCGTAAAATTGAAG
AAATTCAGGCTACATTGAGCGCTGCCCGTGCCAGTGAAAATCTGGCGCGCACGGATTTCAATCGCATTGCTAAGTTGCATAAT
ACCGGTAGTGTTTCGCAATCGGTGCTGGATAATGCCAAAAGCAGTTTGACCACAACCCATGCTGCAGTGATACAAGCAGAGC
GTCAATTGGCTTCAGCGATGATTGGCACTACACCAGCACAAGTGAAAAATCTGTCAGAAAAAGAGAGTGCGCAGGGCATGAC
GCTACAGTCAATTGCTAACTCACGCGAATCTATCAAAAACCGCCAGAACATTCTGGATCAGTTACTGGCTCAATTAGATCAAT
CTCAT
>YscQ_O527_Y11
ATGGTGGGGACACTATCGGCTGAATTACGCCAGCTCAGCGCCATCATTGGGCGTGGGCGTCATAGTGATGGCATTTACGCCAC
ACTGGCGCAGGCCGAGGGGGAGGGTATTTACCTGCCACTTCATTATGACGGCAAGGTATGCGGTTTCTGGTTATCCCAATGCT
GTTGGAATAACTGGCTGGCATCTTCGCTGGCAACAGATGAACCCCATTTGCTGGCGGCTGAATTGGTTAACGCTATGGCGACT
TGGGCTATTAGTCCGCTGCTGGATGCGCTATCAGAGTTAACTTTTGACGAAACTGCGCCAGTGGCGATGACACTTTTGCCCCA
ATGGGCAGTAGTGCTGGCTTTTGAACTGGAAGGCCAGCAACTCAGTGGCGTGTTGTTTGATTGGCCTTTGGCGAGGTTAGCTG
AAACCTTATCTGATTGGCACCGCGAGCAGGTAGCCGAGCAAGATCTGTTGTGGCAGGCGGGATTAGTGGTGGGATGGAGTCG
ATT
>pstS_O527_Y11
TGGCTGGAATGCTGGATAACGGCCAGCCGGTGGTGGGCGATAATTTAGCTCAGGAACTGCAAAAGCGTATTCCGGTGATACA
AGCGCTGGTTAGGCAGTCCAAAGATATTCAGTTCCGTCAGATGAACATGCTGAATCAACAGTTTGAAGCATTGCGGCTACAG
AAAAAACGTCTGCAAATGAGAGGTGAGTTTGATGGCAAAGCGCAAGACCGCATTGAAGCCGAGTGGGGCGAACTCCAGCGC
AGCCATCAGGCGATGATGGAGAAATTACGCGGCTTGCAAATCGACCAAAGCCGTTACACTCTGCTGTTACGCGATATGAATG
GGCAAGTTCATCCACTGCCACTGAGCCAAATCAACCGTGCCTGGTACCCCAATGAGATGAGTCTTGGGCAGAAACTACGCCA
TTTCGTGGTACAAACACATAAGTTTCTGACCGATAACCCGCGTAACGCGAATACTGAAGGTGGGGTGTTCCCGGCTATTTTTG
GCACGGTG
>putative_toxin_subunit_of_TC_PAIYe_insecticidal_toxin_O527_Y11
GGGCTGATGATGAAATCCGCTATTTTGGCGAATATTGGCAAGGGGATGAGCGCGCTTATCATTTGAATAAAACACGGCTGAC
TTCATTCAATTCACTGACTGAATCAGATGATTTGCTGACGGAAATCACGCCAGAACAACACTATTGGTTATTCCGAGCCTTGA
AGGGCAGTTTACTGCGCAGCGAATTGTATGGCCTGGATAATAGTGGGCTGGATGAGGTGCCCTACACCACCAGTAGCGCGCG
TTATCAGGTGCGGCAAGTTCGGGCTTCGGCCATGCCAATTGTCATGCCGCAGGCACTGGAACAGTTAAGTTATCATTATGAGC

GCATCCCGCAAGACCCGCAGTACAGCCAACAGATTGTGATGCGCTGCAACCAGTATGGTCAGCCGCTGCAAACTGTGGCGGT
AAATTATCCGCGCCGCGCTCAAGACAGTGTTTCCCCATATTCATGGTTGGCGACGGCTAACTGGGCCAGCAGTTATGACGAAC
AGCAA
>hemolysin_O527_Y11
GATAAAGGCGCTATTACCTTAAATGCCAATAGCCATACCTTTGAAGCCGCAGTTGATAATGTGGTGCAACACACCAAAGAAA
CAACTGGCGGGGCGGATGTTCGAATCTATACCACCACCGGCAAAGATATTACGGTTGATGGTGAAGGCAAAGGAGGTTCCAA
AGATAGGCTGATAAAAGGTGATATTGCGCATACTGGCAATATGAATGCGGCAAATGGGGTCGACATTAACGTCAAAGAGAAT
GCCATTTATCAAGGCACCAATATCAATGCAGGGGAAGGTAAAATCACGGTCAATGCCGGCGGAGATATTCGCTTTGATCAGG
CCAGCAACCATACCAGTGAAAGCCACAATAATGTTGATGCCAATGCTAAATTGAACCTCGGCACTAATGCTAATAGTAAAGA
ATTTGGTGCTGGACTCGGCGGCGGCCACAGTAAAGGTGAAAGCAATACTTCAGTTGCTCAAGTGAGCCAACTGGGCGGCCAG
CAAGGTGT
>putative_protein_AatP_O527_Y11
ATGGGTATCTATATACTAGAGGCGATTATAAGTCTTAAGGAAAATATCAGGTATACCTTAACATTTATAGTTTTTTTATCACTC
AGCTTCCTTGGAGTAATCATCACTGATTCATTGATTTATAGTGTATCAATGCAGGCAGAAAAAGAACTAAAATCGAGAGGGA
ATAATATAATATCGATTGAGTTATATCAGAAAGGGAAAAAAGATAGTATAGCAAAGGTACTAAATGATTTCTATGCAAAGTT
GTCTTATTCACAACGTTCTTTTTTGAGTGGAGGAAATTCACCTTATTCTGATGAAACTATATCAGTCATGGCAACAGATAAGG
CCGGCATTAATCTTATAATGGGAGATGAGTTTGATGACTCTTTGTTTGAAGGTAACGTTGCTATTTATAACAACAATAATCAA
GACATTTCAGATAAGCCCAGAATGATTTTTTTTAACGGTGTTCCTTTTCATATTATAGGAGTAAAGAACAAATCTAAAGCAGA
ATT
>nrfD_protein_O527_Y11
ATGAGTCCTGCACTGAATACGCCATTTCATTTTGAGTCATTAGTCTGGGATTGGCCGATTGCTATCTATCTGTTTTTAATCGGC
GTCTCTGCCGGCATGGTGGTGGTGAGCCTGTTAGTTAAGCGGCGTGTCTTAGGCGGTGATGCTGCGCAAAGTGGCTTGTTGAA
ATCGACCGCGATTATCGCACCACTGGCGATTGTTCTGGGTTTGTTTATCTTGATTCTGCATCTGACCCGGCCGTGGACTTTCTG
GAAACTGATGTTTTTCTACAGCACCCGTTCGGTCATGTCGCTGGGAGTGATGTTATTTCAGATCTATATGGTGGTGTTATTAGT
TTGGCTGGCGGTGCTATTTCGTCATTGGATAACGGCATTATTACAGCCTTACCCTAAGCTGAGCTGGGTAAGTACGTTGCTGA
ATAAACTGGCCCCGATTGAACCTAAACTTGAGCCGCTGATGATGTTTTTGGCAGTGGCATTAGGGGCTTATACCGGCTTTCT
>antibiotic_acetyltransferase_O527_Y11
ATGAAAAATAAACACTGGTCACAAGTCGAGTATCTGCATCTTTCCGTCAAGAACCCGAATATTTTGATCAGAGGCCAACATA
GTTATTACAGTGACTGTTGGGATGATGGGTTTGAACACTCGGTTGTTCGTTATCTCCATGGCGACAGTGTTAGCCAGCAATGG
GAGCCGCTTGGGCATATTGACCAACTTATTATTGGTGATTATGTCTGTATTGGTGCGGAAGCTGTAATATTAATGGGTGGGAA
TCATAATCACTCCATTGATTTTATTAGTCTTTATCCATTTATGAGTATGGTGAAAAAATCCTATCAGTCCAGAGGGAATACCGT
ACTGAATGATGGGTGCTGGCTTGGAATGCGCTGTATGGTTATGCCTGGTGTTACGGTCGGTGAGGGCGCTGTTATTGCTGCGG
GGAGTATTGTCACAAAAGATATCCCTGCGTATGGCGTCGTCGGCGGCAACCCGGCTCGGGTCATAAAATACCGTTTCCCTGAA
GA
>putative_beta_fimbriae_protein_O527_Y11
ATGAAAAAGCAACTGGCTAAAATTACTGTACTGTCTTCTTTGATTTTTGGTATTAACGCTGCCAATGCAGAAGCACCTACTGC
GGAATTAAAAGTTATTGGTACTTTAACTGTGCCAAGCTGCACCGTGAATGCACCGGATGATGGTGTTTACGACTTTGGTAAAT
TATCTTCATCTTTGATTAAATCAGGCGCGACCACCACAGCACTGACAGCAATCACTAAAAACTGGACAGTAAACTGTGATGCG
GAAACTTATCTAAACTTTACACCAGTGGATAACCGCGCAGCGTCATCTGCCAACACTGGGGTAGCCACTAGCTTTGGTATGGG
TAATGTGAACGGCACCGGCAAAATCGGTTACTACACGGCACAAATCAAGAATGGTACCGTTGATGGTAAAGTATCTAACGTA
TTTTCTTCTGTGAACTCAACGTTCACCGCCACAACCGTTGCCAACCTGACCACGGGTATGCGTACCGGTTGGTCATCAGCTGCC
AA
>putative_iron_sulfur_cluster_protein_with_homology_to_NorA_O527_Y11
ATGGATTACCGCAATCAATCTCTGGGTACTTTGGCTATCGCCATTCCACGAGCAACAAAATTATTCCGTCAGCATCAGTTGGA
TTTCTGCTGTGGTGGCAAGCAAACACTGTTACGGGCCGCCACTAAGTTGAATCTGGATATTGATGAGCTGGAAGCGCAGCTCA
ATGCCTTGCAAGCCGAACCGCAGTCATCTGAGGATTGGCAACAGCAACCGCTGGCAAACATCATTGATTTTATTATCAGTCGC
TATCACAACCGCCATCGCGAACAACTGCCAGAATTGATTCTTATGGCTGAAAAAGTTGAACGCGTTCATGGTGAGAAACCAG
CATGCCCTCGTGGATTAGCAGCCGAACTAACAGCGATTCTAGAAGAGCTGACCCAGCATATGTACAAAGAGGAACAAATTCT
GTTTCCAATGATCAAACGCGGTATGGGGTCACAAGCCGGTGGGCCAATATTTGTGATGGAAGCTGAGCATGACGCGGTAGGA
CAACA
>putative_insecticidal_toxin_complex_TC_PAIYe_protein_O527_Y11
TTAGCGGCTCGTGGCTCACCATTAAAAAAGGTACTGAACAGGTGATTGTTAAGTCACTGGATTATTCCGCTGCCGGGCAAAAA
CTGCGGGAAGAAAGTGGCAATGGCGTGGTGACGGAATATCGCTATGAGGCTGAAACTCAACGCCTTATCGGCATTAAAACTA
CCCGTCCAGCTATCAAAGAGCGGCCAACCCTGTTGCAGGATTTGCGCTACGATTATGACCCGGTCGGTAATATTCTTGCCATC
CATAATGATGCCGAGGCCACCCGCTTCTTCCGCAATCAGAAAATCGTGCCGGAAACCACTTACCGCTATGATGCGTTGTATCA
ATTGACTGAAGCCACAGGGCGTGAATCCGACAGTAACCGCGCACAAAATGCCAGCCTCCCCGCCTTAAGTTCACTGACTGAC
GGCAATCAGTATGTCAATTACACCCGCCGTTACTCTTATGACCGCGCCGGTAACTTACTGAAAATTCAGCACTCCGGTGCCAG
CCAG
>rtxA_O527_Y11
GTAGAGTGAAAACCACGATGACCTATAAACTTGATGACTTGCCAAGGCAAATGAAGAATAAGCTGATTAACCGTTTGGCGAG
CGTTAATGCTGATACCACGTTGGGTGATATATTTGGTGTGGATTATGACAAGTCAGGCCGAATTATTTCTCGCACTGGTACGA
AAATTGATGAAGAGGCGGTACTTAAGGAGATGTTGGAAATTGTTGCTGAGTTTGGTGGTGAAAAACTGAAAGCAATTAGTAA
CCCAGAAAAATTGCTTGATGGCATAGAAGCTAATCTGAATATGGGGAGCAGTGCGATTACCTCATTTGCTCAGTCACATGGTT
TACAAGATAAAGCGCCTGACGAAACTGATAATGATTCCGTCACAGATATTTATACAAATTCTCCGGAACCTAAAAATGATAA
AGAATTTGGTTTTAACTCACTGAAAATCCCTAATTTATTTGCTGTAATATTTAATAAAGATCGGCAAACAGAATTAGAAGATA
TGGCT
>fimbrial_biogenesis_protein_O527_Y11
GTGGTGATCTCTATGTAACGGTAAAAGAAGCGGATGGCAGCGAACAACTGATTGTGGTGCCTTTTGCTTCCTTACCGGTATTA
CAGCGGGAAGGGCGACTGAAATATGCACTCACCGGTGGACAATATCGTTCCTACAACAGCCGTATCGAGAGAACTCCCTTTG
TGCAAGGCACGGGCATCTATGGATTGCCTTATGGCATAACAATTTATGGTGGTATTCAGGGCGCGAGTCGCTATCAATCTTTA
GCTTGCGGCCTGGGAGATAATCTGGGGACGTTGGGTGCGGTATCAGTGGATGTCACACATGCCCAAGCCACCATGCAAAACC
AACCACAGAGAAGTGGCCAATCTCTGCGGACTCGATATAGCAAGAATATAGCACAAACAGGAACCAACATTGCCATTGCGGG
TTACCGTTACGCAACCGCAGGTTATTACAACTTGCAAGAGACGCTGGATAGTTATCGTAAAGGTCAGAATACCGCCCTGGCA
GAGCGT
>hypothetical_CytC_peroxidase_Y11
GGTTATCCAGTGATAACACCGTATCTTGCGCCAGCTGCCATCAGCTTACAAAAGGTGGAGTCGATAATCTGGCGACCTCAACC
GGTGTTGATGGTCAGAAAGGTGGGATTAATGCACCGACAGTCTTCAACGCCGTATTCAATGTGAATCAATTCTGGGATGGCCG

CGCGGTTGATTTACAACAACAAGCCGGTGGTCCGCCACTGAATCCGGTAGAAATGGCGTCAGCATCGTGGGATGACATCATT
GGGAAACTGCAACAGGATGCCCAGTTAACACAAGAGTTTCTTCAGGTTTATCCGCAGGGCTATTCCGAACACAATATTACTGA
CGCTATTGCTGAATTTGAAAAAACGCTGACCACGCCAAATAGTCCCTTTGATCGCTATTTAAAAGGTGATGCCGATGCCCTCA
CCGCCAGCCAGAAAAACGGCCTGGCATTATTCCAGAAAAATAAATGCGACACCTGCCATGTGGGTAAAAATATTGGCGGGCA
GTCG
>negative_regulator_of_beta_lactamase_expression_O527_Y11
ACTGCTGCAATCTGGTTTTCTGGATACGGTGGAAAATCAAAAACACCACTGTGCTCAGTGGCCAAGTTGACAATTTCAATCCC
GATGGAAGTATCATTAATATTAGCGCGCCCCGCCCAACTGCTGACTCCTGCATGCCAAGCCCGCGCATTTTCATCCACCAGGT
TAAATATGCGTATGTTATTAAATCCCGCTTTCTTATAACTGTCATCATCAAGATCGGGAACTAAGTAATGAACACTAACATTA
TTGCCAGTTAGGGAATTGATGGAATTTTCAAAATTTTCGGCCGTATAGTGTAAGACCAAAAAGCGCACTCGGGAATTAAATG
ATTTGATAGCACGAAAATTATTATAGTCAATCATATACATATTGAACTCCTTGTAATTTAAATATCAGCTAATACATATCAAAT
AATGTTATTTATTGAGGATGAGATAATGCCAACTTTATAATGCTCTAAGCACGATAATCCCATGCTTATCTTTAACATATTGGA
G
>IS_Yen2A_B_O527_Y11
ACTGCTTCCACCGTAGGCACTCCTTCGTCTAACGCTATGGCTGCCGCGGTCACGCCTAAATCTTCACCATAATCAGCGATTAA
AGACAGCAACTGCACCATCGCTCTATCACCTTTCGGACGCTTGAGTAGATGTTGCTGGAGTTCTTTAACCGGTTTAGGTAAAT
CCCAATTAATAAAAGGTTCACCGTTTCTCAGGGCACCCGGTTTTCTTTTTAATGTGGACAGATAATGCCAGGGATTATAGGTT
GTCTGGTTTTTAATAAAACTGCGTTGATGTTGAGCGACACACTCGCTATCGACAATAATTTTAATTTCTGTTGCGGTCGTTTTT
AATGTCACTTTCTTACCCACTAAATTATTGGGAATGCTGTATTTATGCCCATCAAAACAGAATAAGGATAAGGAGTTGATTTG
GACTATTGCAAATTGCCCTCCGGTATAAGGGGAATAAGGAGCCAGCATAGTTCGCTCTACACAAAATAGAGTGTTAATGACT
TT
>flagellar_protein_O527_Y11
TACTTCCAAGCCCCGTAATGGTGATTGTTTTATCTTTCAATGCCAATAATTTGTGTAAGTTAGTCGCATTTCCTACGTTATCTTT
ATTATCAGAAAAACCAAGTTCACTTGCTTTAATATCAGTGACAGATAGCATGCCAGTTGTATTGGTAGCATCAAATGCAAATA
ATGCCTTGCCGGTAGCCCCGTTTAAATCATAACCCGCTGCTAACTGATTATTAACTTCAGTGGTTAATAACTCCGCCATTCCTT
GCACAATGCTTTCAGTCTGTTTTAATAAACCTTGCTCATAGTCATTAATAGAACCAAGACTCGCCCCACAGGACATATTTATTG
CAGTGCTGTTGCCGGCAAAATGAAGATCTATCTGCGGCTGTGCACTACCGGTATTTAGCGCTAATGTTGCCACCGTCTGGCCG
TTAATTAATGGCAAACCATTTTTCAGCGAAATATCATAGCCACCATCAGCACCTTTGGTGACGTGAATATCTACCATTTCG
>hypothetical_protein_O527_Y11
GTCTGTTGTATCTTTAGCCTTACGCAGCGAATTAACATAATCAAGATAAGTAAAGCAATCCAGGCCTCTGAAATCTATTACTA
ATTTTTCATGTTTTGTCGGTGACCCAATAAGTATATCTGCTTTATAGGGGGTCCCCAAAAACTCAGCGGAAACTTTATTATTCA
CTTCTCCCGGATGGCTATGACCCGTTTTTTTGACTTGAGTCTCTATAATGTAAGTGATTCGATTAGCTGTGTAATTATCAATAT
AGACTTGTTCATGAGCACCATTTTTCGCACCACAGCCAGCTAACACCGCAATTAAAACCACAGATAACGATTTGTGCATAAGT
CCCTTCTGCGTCTTATCAGGTTATTCCATTCAAGCTAAGAGCAGCCACCTTAATGTTTAAAATTTAAACAAGGCGTAAACAAA
AACCTATTGCTTAGAATTAGTTGTGAGTGTCGATAAAAACGCATTTTTTTGAGTCAAGAGAAGACAATCATGTCAATAATAAA
>hypothetical_Transferase_protein_O527_Y11
TACCGTCCGAAGGTAGAAGCTTAATCAAGTTTTCTATAGGAATCATTAGGTAACATCCTTATATTATTAACCCCCTACAGGGG
CCCCGATTAAACTCGATTTCCCTTCTAACTAAATTAATTAAGGCCACATATTAGAAATAGCTGACATACCCCATAAAGGTGGA
AATATCAGCTGTTCTCATTTTATTAAGTGTCTTTTTCCCAAGGGCATAGCTCGGTTAACGCGGTTTCCAGCGCGGAGATATCGT
CAGAGCTGAGCAATGGGATCAGCTTTTCAATTTCATCATCTGGTCGTTTATAAATCCGTAGATTTACGATCCTCGTGATGACCT
CTTCAGGGAAACGGTATTTTATGACCCGAGCCGGGTTGCCGCCGACGACGCCATACGCAGGGATATCTTTTGTGACAATACTC
CCCGCAACAATAACAGCGCCCTCACCGACCGTAACACCAGACATAACCATACAGCGCATTCCAAGCCAGCACCCATCATTCA
G
>hypothetical_chemotaxis_protein_O527_Y11
GGCAAGTTAATAGTGGCCAGTTAGCTATTCAAAGTGCAAATCTGCGAAAAGAGATTGAAAAGCTCATTCTTGATCAGGATTTT
ACTCAATTTATGGATGAATGGGTCGAAAGAAATAAGCAGCATCAACTTAAACAACTTAAAGATACTTCCGAGGCAGCACAAG
ATACTGTGACTAAATTAAGCGATAATATTAAACAGAATAGCATGTTACAAACTCAAATCGCCGGTTCATTGGTATAGTGAGGT
TATTATGAATATGACAGTTAAACATCAAATACAATCATCAGGGGACAATACATTATCTACAGAAAAAAACGGAGGTATGGGA
AATAATACTGATAATATCATCGCAGATCTGACTGATATTATTTTTAAAATGGTCGGATTAATGAAGCAAATGCGTGACTTACT
TAATGCTTATAATCAGAAGCAACAAATATTACGTTGGAATATACAAGTGAACGCATTGAGTAATAAACGTGAGGCGATCGAT
AAAAC
>hypothetical_T3SS_protein_O527_Y11
GGACTGCTTTTCTGCCTTGATCTGGTGCAACATCAATAACGCCAACATCTGATTGGCTTCCCCTTCTGAAAGGCCGCTATACAA
CTCCATACCACAGCCAGAAAGGAACATAACCACTAGCAATAGCGCCATCAAACGCTGCACCCGTTTCATTGCATATTGACCA
GTTTGTTGATGTTCTGCCCGACCACTCCGGCAACTTTAGCCGTTAAATCAATAGATAAGGTGGCGTGCATCCATTCAGACTGC
ACTGACAAAATTTGATCTGGAGATATCATTGGTGTGGCTGGCTCAACCTACTGCATTAACTGAGCAAACTTGGCGATGTGCTG
TGCATCGGCAGGTGGCTGAGAGGGGATATTCAACTTGGCGGCGGCGTTGATAGCATCAATTTCCATCGGATGACCTCACTTTA
ATAAATATATTTATCTTGCAGGTTGGCTTACATCTTGGGAGAACAAGAAATTCCTTGTTGCAGCAAAACCAATATTCCCTGAG
CC
>hypothetical_protein_O527_Y11
GCAGCATCCCAGCGTTCAAAACTTTGCCCTTGCGAAGGACAATCATCGATCCCCAAGCTATGATGTCCTTCATCCTCTGCGCC
TGAAATGCTCATGTCCAAGCTACTGATAAATTCTTCATACTCTTCGCTTTCCGCATCGCCCATATGGCTATCAGGCTGAAATAC
CAGCTTCCCATCAGCCGATAACCACATATTATCCTCAGTCCAAAAAGCCATATCTGGCTCTTTGCGGATATTAAAATTGTAGC
GTTGCCCTAATACATCCAGCTCTTTAGATAAACCATGACCAGATGTTTCCTCAGAGTGGTAAAGATGAGTTATAGGGTTTGTC
ACATCATCCAGTGACAAAGAGAAGTTCAAGCGGCTAGCTCTGGACATTGGCAGTGGGATTGCTGAGGATAAAAGATCAACGG
ATGAGGATTGGCTGCCAATTAAGTTTCTGGTCAGCTCATCAATAATATATTTTCTATTGTCTAAAAGAAGGTATTCAATTTTAT
T
>bacteriophage_V_tail_O527
AGCGCGACACGGCAATATGAAGCGCTGCCCACGCAAGGAACCTGCAACGGCGACCGGTTTTGTGCGCTGGGAGGGGGTCACC
AACGGTATTGAAGTGCCTGCCGGCAGAATTATTCAGCGCGATGATTTGCAGGAGTACACCACCACGGCGACGGCGACTGCTG
TTGCGGGGGTTCTGCGGGTACCGGTGATTTGTTCTGTTGCCGGTACTGTAGGTAATACAGATGATGGTATCGGTATGGTGCTG
ACTCAACCGATTAACGGTTTGCCATCATCTGCCGCCGCTGACAGCATCGAGGGCGGTACTGATGTGGAGCCAGTGGACGAGT
GGCGCGCACGAATAATCGAACGCTGGTATTACACCCCGCAGGGTGGCGCTGATGGCGATTACATTATCTGGGCTAAAGAGGT
ACCCGGCGTCACTCGTGCCTGGACTTATCGCCACTGGATGGGAACTGGCACGGTCGGTGTGATGGTGGCCAACAGCAATCTTG
AAAACC
>Glycosyltransferase_O527
ACCACTATGAAGCAGTGAAACATGATTTTGAGAGTTGGCTCCCAAAAATATCTAAACATAACGGTGTTGTGCTCTTCCATGAC

ACTTCTGTTGTAGCCGAGGGTTTTGGAGCTAAGCGTTATTGGAATGAGATAAAAGATAGTTATCCATCTTTTAGTTTTGTTCAT
TCTAATGGACTAGGCGTACTATTAGTCGGTAAAGAGCAGCCTGACGTATTGTTACAGTTATCGCATAGTGCTGAGTTATTAGA
TATAGTTCGTGCAATATACGAACTGTCTGGAAATATGTTTTTACATATGGCCAAAGAAAAATGGTGGAAAGCAGAAGCTGAT
AAACATGCCGCCTATGTATCCCATATAACAAGTGATGTTGGTCGCGGGCTTGCTGAAACTGTAAATATACATGTAGTTAACCA
TTTAAATTCGATATCGAGAGATAAAAGTGGTGGTATGATCGGTATAAAAAAAATAGCTAAAAAATTATTCAGAAAGCAAAAC
TAA
>hypothetical_transferase_protein_O527
ATGGGTGTTAATATATTAATACTAGCAGCGGGACAATCGAGTACTGAATCGCGCGAGGATAGCTATCCTTTATGTCTTACTGA
AACTGATGGTATGTCATTAATTGAACGTATCGTTTTAAACACAAAAAAAATCCCAGATTCCAAGTACTGTTTTGCTCTTCTTGA
TAAAGATGCAGAAAAATTTCACTTAGATAAAGTTGTTAAATTATTAGTCAAAGATAGTACTGTAATCAAGATTCCTGAATATA
CGAAAGGTTCGGCTTGTACTGCATTACTTGCAGCGAGTCAGTTGAATCAGCAGGATGAATTATTGATAATAAGTGCCAATGAA
TTAGTTAATATAGATTTTTCCAATGTTTTAAATGATTTTAACTTAAGAAAATTAGACGGTGGTACTCTTATATTCAAATCTGTT
CATCCGCGATATTCTTATGTAAGTCTTAATACGGATAGAATAGTGATTGAGGCGGCTCAACAGAAGCCAATAAGTCAGAATG
C
>uncharacterised_YheO_domain_protein_O527
ATGCTTAGCCAATATCAGCCCATTGTCGCTGCGATAGCCACACTTTTTTCTCCCTATGTGGAAGTGGCTCTCCACGATTTAGCG
AGTCAAAAGCTAGTTTATATTGCGAATAACTATTCACAGCGGGAATTAGGGGAAGACTCCAACCTCAGTGATCTGCAATTTGA
TCAGCACTCGCAGGTGATTGGCCCTTACCAAAAGCGCAACTGGGATGGTCGCCAACTGCGATCCATCAGCACCGTGCTGCGA
GATGACAATGGGCAAGCTGTCGGGCTGCTGTGTATCAATCTGGATATTACTGTTTTTGAAACTGCCAAAGCGGCACTGGATAT
CTTCCTCTCTGGTCATCAATTGCAACCGCAACCTACTGTATTATTTCAGGATGACTGGCAGGAGCGCATTAATACCTACATTCA
CCACTGGCTGCAACAACAACAACTCACCCTGGCGACACTGAACAGTATTAATCGCCGACAGTTAATTGAAGCGCTTTATCAGC
A
>Mu_like_phage_Gp27_O527
ATGAGCGATGAACGCCCAACTCGTGGCCGACCATCCAAGATTGATTTGCTACCTGAACCTATTCGCGATGCACTCCATCAGAT
GTTGCGCGATAAACGCTATACCCAGGAACAAATCAGGGAAGCGGTTAACGAGCTGATTGATGACGATGCACTGCCTGAAGAA
ATGCGGGTGAGCCGTACTGGGTTAAACCGCTATGCCAGCCGTATGGAGTCTATTGGATCACGTATCCGTCAGTCGCGGGAGAT
GGCTGATATCTGGGTAACAAAATTTGGTGAGACCCCCACCACTGATATCGGCAAGCTGTTGATGGAGTTTGTCAAAACGCTGG
CTTTTGAAACATCGATGAAGCTGGCTGAAGGTGAGGAACATGTAGAGCCAAAAGCCTTGGGGCAATTGGCTCTGGTGGCGCA
GCGCATTGAGCAGGCGCAGATGGTCAGCCACAAGCGCGAAAAAGAGATACGCCAGGCATTTGCTGAAGAGGTTGCCGCTCA
AACCGA
>putative_exported_protein_O527
GAAAATACCGGTAATTTTAGAGCTGAGGCCCGCAAAGTTGAAGGGCTGGCTGTAGTTGATCTTATTGAAAATACAGCAAATT
GTCGGCTGTTGTTGATTCCGGTAAAAGATGGTTCTATTCAAGTACACAGTAATGGCAGTTGGGGATGCCGTATTTCGATGCCC
AAAGATGTCTTTATTGATGGGCAGTATGTCCGGGCTGAAAAAGACCCACGTGAAATACCGACTTTACTTTCCATTGGTATTTT
CACTGATGCCGAGAATGATAAGGTTTTCCGTGAGTTAGTCGGTGAACATTATGCCCAATTTATTGCTTCAGCCAATGTCTATAT
CTATAGCCAAGACAGGGATAATCGAGGCGCAAAAGTATTATCTACGTGGGTGCGAGGGGCGGCAAATAAACGGGCATCAAT
AATTATGTATAACCGTAGCGGTATGATGTGGGCCGCATATGTTGCACCTGACAAGAATGGAAACCTGCGGGTTAATTATTTTA
CTAA
>putative_methyltransferase_O527
AGAATGCAAGTTGAGAAAACCATTCTCGATATGTGTTGCGGCTCTCGTATGTTCTGGTTCGACCGCACCAACCCTAATGTTGT
TTTTGGTGATATCCGTAGTGAGTCACACACTCTCTGTGATGGTCGCAGTCTTGAAGTTAAGCCGGATGTACTGCTTGACTTCCG
CGACCTGCCATTCAGCGACGAATCATTCAACATGGTTGTATTCGACCCACCACATCTGAGTCGCGCTGGTGAGAACGGCTGGC
AGTGCAAGAAATACGGCGTACTGAACAGAGATACATGGCGCGATGATTTAGCCGCTGCTTTCCGTGAAGCCTTTCGGGTGCT
GCGCCAGCATGGAACGCTGATATTCAAATGGAATGAAACGCAGATACCGACCCGCGAGATATTGGCCCTCACCGACCATAAA
CCGGCATTTGTCCACCCGTCTGGCGCCAGAGCAAAAACTCACTGGATATGCTTTATCAAAGATGGAGTAACAACTGATGAAT
AACA
>putative_predicted_metal_dependent_hydrolase_O527
ATGCCGGTCATGCAGATAGGTACGCTGCGCTTACAGATCAATCGTAAAGCGATTAAGAACTTGCATATCAGCGTCCTGCCGCC
TGACGGCAAAGTCCGCGTTTCAGCCCCGGAATACATGACAGAAACGGCGATACGAATGGCGGTCGCTTCCCGCTTTCGCTGG
ATCCGAAAACAACAGCAGGATTTTGCCAGACAGCCACGCCAGTCCGAGCGGGAGATGGTCAGCGGAGAATGCCACTATCTTT
GGGGTCGTAAATATCGCTTGTTTGTGATCGAGCGCCAGGGACGTCATGAGGTGAAAATTGCGGGAAATAATAAACTCCAGCT
TTATGTTCAGCCGGGAACCAGCAATGAAAATAAGGAGCTGGTATTAACCGACTTTTATCGGCATGAGCTGAAACGGCAGGCC
GCTAAATTACTGCCCGAATGGCAGGACAGAACCGGCGTTTCCCCGACGTTCTGGGGCATAAAAAAGATGAAAACCTTTTGGG
GCACCTG
>hypothetical_secretion_system_apparatus_protein_SsaL_O527
AGGAAAGCCAAGGGCTGGGGGCCATCAAGCAACTGTTCGCGCAAAGTGTTGAGCAAAATGAGCTGTCGGTCGCGGAGTGGTT
TGCTCAAGTGAGCCGCTGGCCGCAGCGGCAGCAACGAGTCCGGATATTGATGCGCGCAGTGGCGTTTGACCTGTCTTCTCAAC
CGCCACCAAAGCGTGGTGAACGTCTGGCCAGCACACTGCATCAGCTTCGCCGTTTGCTGATGTTTCTGGGGTTAGAAGACCAT
TGTTCCCATATGGGGAGAGCTTGTGCAGTAGAAGGGGATGTGATTTTGAGTGAAGTGCTGGCGGTGGTGGGGCAGCCCTGGT
TATTTAGTCACTGGTTGCAACCACGCATTGAGGCGATTTCAGGCCAGGATATCAATGCACATAGTAGATTTGTTCGCCGATTT
TATGAGTTATTCAACCTTATGCCGGTTGATTGCTTCAATGATCAAGAGCAGCAACAGCAAATTTTGGCAGTGTTGTTGGAGAT
ATAA
>transposase_O527
GTTCGTTGGCATAACGGTGAAATTCTGCGGCCGAAAACATGGATTTGGCAGGATATACGTACCCGTAAAATATTAGCCTGGC
GCACAGATGTAAGTGAGAACTCTGATTCAATTCGGCTTTCCCTTTCCGACTTGATAGAAAACTACGGTATACCGCAACAACTG
ACTATTGATAACACCCGCGCAGCGGCAAACAAGTGGCTTACCGGTGGTGTGCCTAACCGTTATCGATTCAAAGTTAAAGAAG
ATGACCCGATGGGGATCATCCCATTATTGGGTATCAAACTGCATTGGACTAGCGTCTTATTTGGCCGTGGTCATGGGCAGGCT
AAGCCGGTAGAACGTGCATTTTCTCATGGTGGCTTGGGTGAAACAGTCGATAAACATCCTTTATTGTCTGGCGCTCATACCGG
TGATAACCCAATGGACAAGCCCGATAATTACGGTTCACGGGCGGTAGAAGCTGACGTATTCCTGAAAGTCTTAGCTGAAGGG
ATCGC
>Type_II_restriction_enzyme_EcoRII_O527
GTGTCTGCTTTTCATAACTGGTTATCTGAAAAATTATCGGGCGATTGGTATATCTACATTAAACGCCTTTCCGGTAATGATACG
GGGGCAACAGGTGGACATCAGGCCGGCGTCTATATCCCGACGGAAGTGATTGAGAAGCAGTTCTCTCCTATTCTACGAACTG
ATGTGAGAAACCCTGACATTCTTTTGCCTGCCAGAATAAGTTCTCATGGAAATCTTGAGAGTGTAGTCCGGGCTATTTATTAC
AATAACCGTCATTTTGATGGTACACGTAATGAAAAGCGTATTACTCGTTGGGGTAAAGGTAGTCCGCTGTTAAATAAAGAAA
ATACTGGCGCACTGACTGTTTTCTCTTTTCATAGTCAGCCCGATAGCATCTGCGATTTCATTGACGTGTGGCTATGCAATTCCC

TGGCAGAGGAAGAACTGCTGGAATCCATGACGGGTGAGATCATCCCTGGCTTCTCAGTTTCTGGCCCATCTAATCAGGTTCTT
GG
>Type_I_site_specific_deoxyribonuclease_O527
ATGCAGGAAGCGGGCTATACCGATGAGGAAGTGGAAAGCATTCGCGCCGAAGTCGCTAAATTCGAGAGACTGCGCTACGAC
ATTAAACTGACCAGCGGCGATCTGCTGGACATGAAGCGCTTTGAGTCCGCCATGCGCCACCTGCTGGATATGTACGTTCGCGC
CGACGGTAGCGAAGTGCTTATGGATTTCGAAGAGTTAAGTCTGATTGAGCTGATTGTTGAAAAAGGAGCAGCAGCAACCGCT
GCTCTGCCGGAGGAGATCCGCAACAATCAGGATGCGATGGCGGAAACCATCGAAAACAACGTGCGTAAAACTATCGTGGAT
GAAAACCCGGTGAACCCAAAATACTACGGGCGCATGTTGCTTTTGCTGGATGAACTGATTACCCTTCGCCGCCAGAATGCGCT
TGATTATCAGCAGTACCTGGAGTGTATCCGGGATTTGAGCAAGCAGGTCATCCGCCCAGAACAAACCAACGGAGCATCCTAT
CCTGAAAC
>YE0694_putative_adhesin_O8
AAGTCACTTATCAGGCGACCCTCAAAGATGCTAACGATAACCTGCTGGGTGCCGGTATTCCGGTGCATTGGTCAGTCAATCGC
GACACTCTGATGAGTGGCAAGTTGATTAGCCTAACCAACAGTGCTGGCGTGGCAGAAGTTGAAATCAGCCGTGATTTGGCTG
GTGATGCTCTGGTCACTGCGGCGGTAGGCAATAACAGTTTGCAGGCTACGGCAGTTAAATTTATCAGTGGCGGAGTGGACAT
CAGCAAATCCAGCATGCAGTTACTGCAAGGTAATATCACTGCCGATAATCTCGATATTGCCACTATTCAGGTTGATATCCGTG
ATAGCAAAGGCAATCCGTTACCTAATCTGGCAAGTCAAATCACCACCTCACCAAAAAAGGGTGAGCATGGCTTGAAAATCGA
AACCATTGCTAACCCATCCGGCGACGGTTATCTGGTCAAGATGAAAGGTACTCAAGCCGGTAATCATACAGTGACCGTCTCTG
TGGCG
>fyuA_O8
GCGGCGTCTCTTCGGCGCAGGACTTCTATAACCCCGCCGTCACCCTGTATGTCGATGGCGTCCCTCAGCTTTCCACCAACACC
ATCCAGGCGCTTACCGATGTGCAAAGCGTGGAGTTGCTGCGTGGCCCACAGGGAACGTTATATGGCAAAAGCGCTCAGGGCG
GGATCATCAACATCGTCACCCAGCAGCCGGACAGCACGCCGCGCGGCTATATTGAAGGCGGCGTCAGTAGCCGCGACAGTTA
TCGAAGTAAGTTCAACCTGAGCGGCCCCATTCAGGATGGCCTGCTGTACGGCAGCGTCACCCTGTTACGCCAGGTTGATGACG
GCGACATGATTAACCCCGCGACGGGAAGCGATGATTTAGGCGGCACCCGCGCCAGCATAGGGAATGTGAAACTGCGTCTGGC
GCCGGACGATCAGCCCTGGGAAATGGGCTTTGCCGCCTCACGCGAATGTACCCGCGCCACCCAGGACGCCTATGTGGGATGG
AATGAT
>invA_O8_Y11_O527
GCACCATTACTGGTGGTTATCACCCCCTCTTGCTCGGCAACCGGCGTACTATCAATATTAGTCACTGTAAATTTCACCGTGATG
GGAGTCCTGCCATCAGCAAGCGCACCATCATCAGTCACTTCCGAGGTAATAACGAACTGAGGTTGCTGCACAATAACGGTCA
ACGTAGCTGGATTAGAACTATTGCCGTGGTTATCGATAGCCGTGGCGCTGAGGGTATAGGTATTGACCGCCGCTTCGCTCTCG
CCGGTTTTCCCCACAGTACGACTCGCTGGTGCCATCGGTTTATAAGGCGGTAAGACCACATTGTAATCTGTTGCACTGACGGG
TATTAATTTGCCGCCGGCAGCAACCCATTGCGCATCATTCCAGAGAATACGTTGTAGTGCAGATTGAGACTGTATTTGTGCAC
TAACGGAATAAACCTGCCCCGGCAGGCCGGAGAGTACGGCGGGTGAAAATGCCAATTTGATAGTATTCTGTTTTTGGTATTCC
AG
>Ail_O8_Y11_O527
ATGAAAAAGACATTACTAGCTAGTTCTCTAATAGCCTGTTTATCAATTGCGTCTGTTAATGTGTACGCTGCGAGTGAAAGTAG
TATTTCTATTGGTTATGCGCAAAGCCATGTAAAAGAAAATGGGGATACATTGGATAATGACCCTAAAGGTTTTAACCTGAAGT
ACCGTTATGAACTCGATGATAACTGGGGAGTAATAGGTTCGTTTGCTTATACTCATCAGGGATATGATTTCTTCTATGGCAGT
AATAAGTTTGGTCATGGTGATGTTGATTACTATTCAGTAACAATGGGGCCATCTTTCCGCATCAACGAATATGTTAGCCTTTAT
GGATTACTGGGGGCCGCTCATGGAAAGGTTAAGGCATCTGTATTTGATGAATCAATCAGTGCAAGTAAGACGTCAATGGCAT
ACGGGGCAGGGGTGCAATTCAACCCACTTCCAAATTTTGTCATTGACGCTTCATATGAATACTCCAAACTCGATAGCACAAAA
GT
>rcsA_O8_Y11_O527
GGTGATAACCCCAGAACCAAGCAATGAATCAACTTTGATGCCACTGCCATTAATTGATTTGCCGCTCAGTGCAGTATCCGTGA
TGGTGGCATTACCGCTGATACGTAGGCCTTCAGCCTGGATTGAACTGCCAGTCACATTAATACGGTTAATTTGATGCTCGGCA
CTGAACAAATGAACCCCAGAGCCATCAGTACTGCTTCCTTCAATCGTGACAGGGCCGCTACTATTATTAACCATACCGGAAAT
AGCCACCCCACTGCCAGCAACACTACTACCTTTAATTAGCGAGTCACCTTCACCATTTAATACAACTGTATTATTACCAATAG
GGAAATTAATACCGGCACCACTGGTTGATCGCCCTTCCATTGTTAAAGAACCATTATTATTAATAGTTAATGTTTTAGGGCTTA
GATAGCTGATACCATTTCCACCCGCAGTAATTTGTGAACCTGTATTTTCACCGAGAAGAGATAACTGGCTATCCTTGGCAATA
T
>lcrV_O8
TTACCTCGTGTCATCTAGCAGACGTTGCATTACTGAATCATATTTCTGAATGAAACGGTTCAGTGCTTCAATAGCTGAATTAAA
ACGTGATGTAATATCAGACAGCTGAGTTGTTTTTTGGCTAACCAAGTCGTTGAGCGGCCTGGACTTATCCGAGCAGGCGGTGG
CAAAGTGAGATAATTCATTATTATCTTTATTATAAGAGTATGAATTTTTCAGGTTACCCAACGCCCCGGTTCTTTTATTCTCAC
TTTCAAGAAAGTTCTTTATCGAGACTATTTTTTTTTCGGTCTCACCTTCCTTAATGGTGGTTTGAGACATTTTCTCAAGAATTTT
GTACTCTGCGCTGGCTTTAAAAATCTCTTCATCTGTATAACCATATAAATTTTTATCCATGAGATTAATTGATTTATCATGGAT
ATTTATGGTGCCACTATTCGACAGATACTTATTAATTTCAGCTTGAATAACTGAATAAATTTTTAATTCGGCGGTAAGCT
>yadA_O8
TTACCACTCGATATTAAATGATGCATTGTACATGACATCCGAGGAACCGGCATAAGCCACACCGGCTTTAAGTGCGACACTCT
CATTTACACGATAGCCAGAACCAATTGCTAATGCCTGACTAGAGCGATATCCCCCGACACCTGCAGTAAAGTTTACTTTCCCC
ACACCATATGGCTGGAACAAGCTGTTTAAAGCGGCTGAACTGGCTAAACCTTTGTCAACTCGTGTGTCAAGTTTATCTAAACG
GTTGTCAAGTTGACGGAATTTATGATCTGTGTATTGATTCGATTCACGGATTGCTTTCTTAGTCGAATTACTTACAGTCACATC
GGTATAGCTATTTGCAGTTTGTAGTGTGTGAGAAGACTTGCTGTCTGCATACACATTAGCTCTTGCTAACGTCTCAGCTGATTT
TTTATTTGTATGTTCTTCAGCAGTTTCTAAAGTTGTTCTAGCAACACTATTTGAGTGTTTTTTTGCCATATCCAAAGCATCGT

8. References

Achtman, Zurth, Morelli, Torrea, Guiyoule and Carniel (1999). Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proceedings of the National Academy of Sciences of the United States of America. **96:** 14043-14048.

Aili, Hallberg, Wolf-Watz and Rosqvist (2002). GAP activity of Yersinia YopE. Methods in enzymology. **358:** 359-370.

Aili, Isaksson, Carlsson, Wolf-Watz, Rosqvist and Francis (2008). Regulation of Yersinia Yop-effector delivery by translocated YopE. International journal of medical microbiology : IJMM. **298:** 183-192.

Aili, Isaksson, Hallberg, Wolf-Watz and Rosqvist (2006). Functional analysis of the YopE GTPase-activating protein (GAP) activity of Yersinia pseudotuberculosis. Cellular microbiology. **8:** 1020-1033.

Aleksic (1995). Occurrence of Y. enterocolitica antigens O:3, O:9 and O:8 in different Yersinia species, their corresponding H antigens and origin. Contributions to microbiology and immunology. **13:** 89-92.

Alonso, Bottini, Bruckner, Rahmouni, Williams, Schoenberger and Mustelin (2004). Lck dephosphorylation at Tyr-394 and inhibition of T cell antigen receptor signaling by Yersinia phosphatase YopH. The Journal of biological chemistry. **279:** 4922-4928.

Andersson, Carballeira, Magnusson, Persson, Stendahl, Wolf-Watz and Fallman (1996). YopH of Yersinia pseudotuberculosis interrupts early phosphotyrosine signalling associated with phagocytosis. Molecular microbiology. **20:** 1057-1069.

Andor, Trulzsch, Essler, Roggenkamp, Wiedemann, Heesemann and Aepfelbacher (2001). YopE of Yersinia, a GAP for Rho GTPases, selectively modulates Rac-dependent actin structures in endothelial cells. Cellular microbiology. **3:** 301-310.

Autenrieth, Reissbrodt, Saken, Berner, Vogel, Rabsch and Heesemann (1994). Desferrioxamine-promoted virulence of Yersinia enterocolitica in mice depends on both desferrioxamine type and mouse strain. The Journal of infectious diseases. **169:** 562-567.

Aziz, Bartels, Best, DeJongh, Disz, Edwards, Formsma, Gerdes, Glass, Kubal, Meyer, Olsen, Olson, Osterman, Overbeek, McNeil, Paarmann, Paczian, Parrello, Pusch, Reich, Stevens, Vassieva, Vonstein, Wilke and Zagnitko (2008). The RAST Server: rapid annotations using subsystems technology. BMC genomics. **9:** 75.

Bach, de Almeida and Carniel (2000). The Yersinia high-pathogenicity island is present in different members of the family Enterobacteriaceae. FEMS microbiology letters. **183:** 289-294.

Bachmann (1990). Linkage map of Escherichia coli K-12, edition 8. Microbiological reviews. **54:** 130-197.

Bakas, Ostolaza, Vaz and Goni (1996). Reversible adsorption and nonreversible insertion of Escherichia coli alpha-hemolysin into lipid bilayers. Biophysical journal. **71:** 1869-1876.

Balligand, Laroche and Cornelis (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. Infection and immunity. **48:** 782-786.

Barz, Abahji, Trulzsch and Heesemann (2000). The Yersinia Ser/Thr protein kinase YpkA/YopO directly interacts with the small GTPases RhoA and Rac-1. FEBS letters. **482:** 139-143.

Batzilla, Antonenka, Hoper, Heesemann and Rakin (2011a). Yersinia enterocolitica palearctica serobiotype O:3/4 - a successful group of emerging zoonotic pathogens. BMC genomics. **12:** 348.

Batzilla, Heesemann and Rakin (2011b). The pathogenic potential of Yersinia enterocolitica 1A. International journal of medical microbiology : IJMM. **301:** 556-561.

Batzilla, Hoper, Antonenka, Heesemann and Rakin (2011c). Complete Genome Sequence of Yersinia enterocolitica subsp. palearctica Serogroup O:3. Journal of bacteriology. **193:** 2067.

Bearson, Albrecht and Gunsalus (2002). Oxygen and nitrate-dependent regulation of dmsABC operon expression in Escherichia coli: sites for Fnr and NarL protein interactions. BMC microbiology. **2:** 13.

Beer and Miller (1992). Amino acid substitutions in naturally occurring variants of ail result in altered invasion activity. Journal of bacteriology. **174:** 1360-1369.

Ben-Efraim, Aronson and Bichowsky-Slomnicki (1961). NEW ANTIGENIC COMPONENT OF PASTEURELLA PESTIS FORMED UNDER SPECIFIED CONDITIONS OF pH AND TEMPERATURE. Journal of bacteriology. **81:** 704-714.

Benabdillah, Mota, Lutzelschwab, Demoinet and Cornelis (2004). Identification of a nuclear targeting signal in YopM from Yersinia spp. Microbial pathogenesis. **36:** 247-261.

Bengoechea, Pinta, Salminen, Oertelt, Holst, Radziejewska-Lebrecht, Piotrowska-Seget, Venho and Skurnik (2002). Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of Yersinia enterocolitica serotype O:8. Journal of bacteriology. **184:** 4277-4287.

Bent and Young (2010). Contribution of BlaA and BlaB beta-lactamases to antibiotic susceptibility of Yersinia enterocolitica biovar 1B. Antimicrobial agents and chemotherapy. **54:** 4000-4002.

Bhagat and Virdi (2011). The enigma of Yersinia enterocolitica biovar 1A. Critical reviews in microbiology. **37:** 25-39.

Biedzka-Sarek, Jarva, Hyytiainen, Meri and Skurnik (2008a). Characterization of complement factor H binding to Yersinia enterocolitica serotype O:3. Infection and immunity. **76:** 4100-4109.

Biedzka-Sarek, Salmenlinna, Gruber, Lupas, Meri and Skurnik (2008b). Functional mapping of YadA- and Ail-mediated binding of human factor H to Yersinia enterocolitica serotype O:3. Infection and immunity. **76:** 5016-5027.

Biedzka-Sarek, Venho and Skurnik (2005). Role of YadA, Ail, and Lipopolysaccharide in Serum Resistance of Yersinia enterocolitica Serotype O:3. Infection and immunity. **73:** 2232-2244.

Biraben (1975). Les hommes et la peste en France et dans les pays européens et méditerranéens. Paris: Mouton. **2 vols.**

Black, Montagna, Zitsmann and Bliska (1998). Identification of an amino-terminal substrate-binding domain in the Yersinia tyrosine phosphatase that is required for efficient recognition of focal adhesion targets. Molecular microbiology. **29:** 1263-1274.

Bliska and Falkow (1992). Bacterial resistance to complement killing mediated by the Ail protein of Yersinia enterocolitica. Proceedings of the National Academy of Sciences of the United States of America. **89:** 3561-3565.

Boardman and Satchell (2004). Vibrio cholerae strains with mutations in an atypical type I secretion system accumulate RTX toxin intracellularly. Journal of bacteriology. **186:** 8137-8143.

Bolin, Norlander and Wolf-Watz (1982). Temperature-inducible outer membrane protein of Yersinia pseudotuberculosis and Yersinia enterocolitica is associated with the virulence plasmid. Infection and immunity. **37:** 506-512.

Bolin and Wolf-Watz (1988). The plasmid-encoded Yop2b protein of Yersinia pseudotuberculosis is a virulence determinant regulated by calcium and temperature at the level of transcription. Molecular microbiology. **2:** 237-245.

Bonardi, Brindani, Pizzin, Lucidi, D'Incau, Liebana and Morabito (2003). Detection of Salmonella spp., Yersinia enterocolitica and verocytotoxin-producing Escherichia coli O157 in pigs at slaughter in Italy. International journal of food microbiology. **85:** 101-110.

Bottone (1997). Yersinia enterocolitica: the charisma continues. Clinical microbiology reviews. **10:** 257-276.

Bottone (1999). Yersinia enterocolitica: overview and epidemiologic correlates. Microbes and infection / Institut Pasteur. **1:** 323-333.

Bowen, Rocheleau, Blackburn, Andreev, Golubeva, Bhartia and ffrench-Constant (1998). Insecticidal toxins from the bacterium Photorhabdus luminescens. Science. **280:** 2129-2132.

Boyce, Evans, Evans and DuPont (1979). Production of heat-stable, methanol-soluble enterotoxin by Yersinia enterocolitica. Infection and immunity. **25:** 532-537.

Boyer and Roulland-Dussoix (1969). A complementation analysis of the restriction and modification of DNA in Escherichia coli. Journal of molecular biology. **41:** 459-472.

Bresolin, Morgan, Ilgen, Scherer and Fuchs (2006). Low temperature-induced insecticidal activity of Yersinia enterocolitica. Molecular microbiology. **59:** 503-512.

Bresolin, Trcek, Scherer and Fuchs (2008). Presence of a functional flagellar cluster Flag-2 and low-temperature expression of flagellar genes in Yersinia enterocolitica W22703. Microbiology. **154:** 196-206.

Brossolet and Mollaret (1994). Pourquoi la peste? Le rat, la puce et le bubon. Paris: Gallimard.

Brubaker (1991). Factors promoting acute and chronic diseases caused by yersiniae. Clinical microbiology reviews. **4:** 309-324.

Brubaker (2003). Interleukin-10 and inhibition of innate immunity to Yersiniae: roles of Yops and LcrV (V antigen). Infection and immunity. **71:** 3673-3681.

Bruckner, Rhamouni, Tautz, Denault, Alonso, Becattini, Salvesen and Mustelin (2005). Yersinia phosphatase induces mitochondrially dependent apoptosis of T cells. The Journal of biological chemistry. **280:** 10388-10394.

Bucher, Meyer, Grotzbach, Wacheck, Stolle and Fredriksson-Ahomaa (2008). Epidemiological data on pathogenic Yersinia enterocolitica in Southern Germany during 2000-2006. Foodborne pathogens and disease. **5:** 273-280.

Buchrieser, Brosch, Bach, Guiyoule and Carniel (1998). The high-pathogenicity island of Yersinia pseudotuberculosis can be inserted into any of the three chromosomal asn tRNA genes. Molecular microbiology. **30:** 965-978.

Burnens, Frey and Nicolet (1996). Association between clinical presentation, biogroups and virulence attributes of Yersinia enterocolitica strains in human diarrhoeal disease. Epidemiology and infection. **116:** 27-34.

Carniel (1999). The Yersinia high-pathogenicity island. International microbiology : the official journal of the Spanish Society for Microbiology. **2:** 161-167.

Carniel (2002). Plasmids and pathogenicity islands of Yersinia. Current topics in microbiology and immunology. **264:** 89-108.

Carniel, Guilvout and Prentice (1996). Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B Yersinia enterocolitica. Journal of bacteriology. **178:** 6743-6751.

Carver, Berriman, Tivey, Patel, Bohme, Barrell, Parkhill and Rajandream (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics. **24:** 2672-2676.

Carver, Thomson, Bleasby, Berriman and Parkhill (2009). DNAPlotter: circular and linear interactive genome visualization. Bioinformatics. **25:** 119-120.

Cascales, Buchanan, Duche, Kleanthous, Lloubes, Postle, Riley, Slatin and Cavard (2007). Colicin biology. Microbiology and molecular biology reviews : MMBR. **71:** 158-229.

Casutt-Meyer, Renzi, Schmaler, Jann, Amstutz and Cornelis (2010). Oligomeric coiled-coil adhesin YadA is a double-edged sword. PloS one. **5:** e15159.

Chandler and Parisi (1994). Radiological case of the month. Yersinia enterocolitica masquerading as appendicitis. Archives of pediatrics & adolescent medicine. **148:** 527-528.

Chang and Cohen (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. Journal of bacteriology. **134:** 1141-1156.

Chen, Cook, Stewart, Nagarajan, Sommer, Pop, Thomason, Thomason, Lentz, Nolan, Sozhamannan, Sulakvelidze, Mateczun, Du, Zwick and Read (2010). Genomic characterization of the Yersinia genus. Genome biology. **11:** R1.

Chouikha, Charrier, Filali, Derbise and Carniel (2010). Insights into the infective properties of YpfPhi, the Yersinia pestis filamentous phage. Virology. **407:** 43-52.

Clark, Hirst and Jepson (1998). M-cell surface beta1 integrin expression and invasin-mediated targeting of Yersinia pseudotuberculosis to mouse Peyer's patch M cells. Infection and immunity. **66:** 1237-1243.

Collyn, Billault, Mullet, Simonet and Marceau (2004). YAPI, a new Yersinia pseudotuberculosis pathogenicity island. Infection and immunity. **72:** 4784-4790.

Cordero, Kudryashov, Reisler and Satchell (2006). The Actin cross-linking domain of the Vibrio cholerae RTX toxin directly catalyzes the covalent cross-linking of actin. The Journal of biological chemistry. **281:** 32366-32374.

Cornelis (1998). The Yersinia Yop virulon, a bacterial system to subvert cells of the primary host defense. Folia microbiologica. **43:** 253-261.

Cornelis (2002a). The Yersinia Ysc-Yop 'type III' weaponry. Nature reviews. Molecular cell biology. **3:** 742-752.

Cornelis (2002b). The Yersinia Ysc-Yop virulence apparatus. International journal of medical microbiology : IJMM. **291:** 455-462.

Cornelis, Boland, Boyd, Geuijen, Iriarte, Neyt, Sory and Stainier (1998). The virulence plasmid of Yersinia, an antihost genome. Microbiology and molecular biology reviews : MMBR. **62:** 1315-1352.

Cornelis, Sluiters, Delor, Geib, Kaniga, Lambert de Rouvroit, Sory, Vanooteghem and Michiels (1991). ymoA, a Yersinia enterocolitica chromosomal gene modulating the expression of virulence functions. Molecular microbiology. **5:** 1023-1034.

Corpet (1988). Multiple sequence alignment with hierarchical clustering. Nucleic acids research. **16:** 10881-10890.

Cossart and Sansonetti (2004). Bacterial invasion: the paradigms of enteroinvasive pathogens. Science. **304:** 242-248.

Darling, Mau and Perna (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PloS one. **5:** e11147.

Datsenko and Wanner (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America. **97:** 6640-6645.

Delor and Cornelis (1992). Role of Yersinia enterocolitica Yst toxin in experimental infection of young rabbits. Infection and immunity. **60:** 4269-4277.

Delor, Kaeckenbeeck, Wauters and Cornelis (1990). Nucleotide sequence of yst, the Yersinia enterocolitica gene encoding the heat-stable enterotoxin, and prevalence of the gene among pathogenic and nonpathogenic yersiniae. Infection and immunity. **58:** 2983-2988.

Delpino, Marchesini, Estein, Comerci, Cassataro, Fossati and Baldi (2007). A bile salt hydrolase of Brucella abortus contributes to the establishment of a successful infection through the oral route in mice. Infection and immunity. **75:** 299-305.

Derbise, Chenal-Francisque, Pouillot, Fayolle, Prevost, Medigue, Hinnebusch and Carniel (2007). A horizontally acquired filamentous phage contributes to the pathogenicity of the plague bacillus. Molecular microbiology. **63:** 1145-1157.

Dersch and Isberg (2000). An immunoglobulin superfamily-like domain unique to the Yersinia pseudotuberculosis invasin protein is required for stimulation of bacterial uptake via integrin receptors. Infection and immunity. **68:** 2930-2938.

Dobrindt, Chowdary, Krumbholz and Hacker (2010a). Genome dynamics and its impact on evolution of Escherichia coli. Medical microbiology and immunology. **199:** 145-154.

Dobrindt, Hochhut, Hentschel and Hacker (2004). Genomic islands in pathogenic and environmental microorganisms. Nature reviews. Microbiology. **2:** 414-424.

Dobrindt, Zdziarski, Salvador and Hacker (2010b). Bacterial genome plasticity and its impact on adaptation during persistent infection. International journal of medical microbiology : IJMM. **300:** 363-366.

Dower, Miller and Ragsdale (1988). High efficiency transformation of E. coli by high voltage electroporation. Nucleic acids research. **16:** 6127-6145.

Dukuzumuremyi, Rosqvist, Hallberg, Akerstrom, Wolf-Watz and Schesser (2000). The Yersinia protein kinase A is a host factor inducible RhoA/Rac-binding virulence factor. The Journal of biological chemistry. **275:** 35281-35290.

Durbin (1998). Biological sequence analysis: probabilistic models of proteins and nucleic acids, Cambridge University Press.

El Tahir and Skurnik (2001). YadA, the multifaceted Yersinia adhesin. International journal of medical microbiology : IJMM. **291:** 209-218.

Ellison and Miller (2006). H-NS represses inv transcription in Yersinia enterocolitica through competition with RovA and interaction with YmoA. Journal of bacteriology. **188:** 5101-5112.

Ellison, Young, Nelson and Miller (2003). YmoA negatively regulates expression of invasin from Yersinia enterocolitica. Journal of bacteriology. **185:** 7153-7159.

Emody, Heesemann, Wolf-Watz, Skurnik, Kapperud, O'Toole and Wadstrom (1989). Binding to collagen by Yersinia enterocolitica and Yersinia pseudotuberculosis: evidence for yopA-mediated and chromosomally encoded mechanisms. Journal of bacteriology. **171:** 6674-6679.

Erfurth, Grobner, Kramer, Gunst, Soldanova, Schaller, Autenrieth and Borgmann (2004). Yersinia enterocolitica induces apoptosis and inhibits surface molecule expression and cytokine production in murine dendritic cells. Infection and immunity. **72:** 7045-7054.

Evdokimov, Anderson, Routzahn and Waugh (2001). Unusual molecular architecture of the Yersinia pestis cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. Journal of molecular biology. **312:** 807-821.

Evdokimov, Tropea, Routzahn and Waugh (2002). Crystal structure of the Yersinia pestis GTPase activator YopE. Protein science : a publication of the Protein Society. **11:** 401-408.

Feeley, Wells, Tsai and Puhr (1979). Detection of enterotoxigenic and invasive strains of Yersinia enterocolitica. Contributions to microbiology and immunology. **5:** 329-334.

Felmlee, Pellett, Lee and Welch (1985a). Escherichia coli hemolysin is released extracellularly without cleavage of a signal peptide. Journal of bacteriology. **163:** 88-93.

Felmlee, Pellett and Welch (1985b). Nucleotide sequence of an Escherichia coli chromosomal hemolysin. Journal of bacteriology. **163:** 94-105.

Felmlee and Welch (1988). Alterations of amino acid repeats in the Escherichia coli hemolysin affect cytolytic activity and secretion. Proceedings of the National Academy of Sciences of the United States of America. **85:** 5269-5273.

ffrench-Constant, Waterfield, Daborn, Joyce, Bennett, Au, Dowling, Boundy, Reynolds and Clarke (2003). Photorhabdus: towards a functional genomic analysis of a symbiont and pathogen. FEMS microbiology reviews. **26:** 433-456.

Forstner, Jabbal, Qureshi, Kells and Forstner (1979). The role of disulphide bonds in human intestinal mucin. The Biochemical journal. **181:** 725-732.

Foultier, Troisfontaines, Vertommen, Marenne, Rider, Parsot and Cornelis (2003). Identification of substrates and chaperone from the Yersinia enterocolitica 1B Ysa type III secretion system. Infection and immunity. **71:** 242-253.

Fredriksson-Ahomaa, Koch, Klemm, Bucher and Stolle (2004). Different genotypes of Yersinia enterocolitica 4/O:3 strains widely distributed in butcher shops in the Munich area. International journal of food microbiology. **95:** 89-94.

Fredriksson-Ahomaa, Niskanen, Bucher, Korte, Stolle and Korkeala (2003). Different Yersinia enterocolitica 4:O3 genotypes found in pig tonsils in Southern Germany and Finland. Systematic and applied microbiology. **26:** 132-137.

Fredriksson-Ahomaa, Stolle, Siitonen and Korkeala (2006). Sporadic human Yersinia enterocolitica infections caused by bioserotype 4/O : 3 originate mainly from pigs. Journal of medical microbiology. **55:** 747-749.

Frishman, Mokrejs, Kosykh, Kastenmuller, Kolesov, Zubrzycki, Gruber, Geier, Kaps, Albermann, Volz, Wagner, Fellenberg, Heumann and Mewes (2003). The PEDANT genome database. Nucleic acids research. **31:** 207-211.

Fuchs, Bresolin, Marcinowski, Schachtner and Scherer (2008). Insecticidal genes of Yersinia spp.: taxonomical distribution, contribution to toxicity towards Manduca sexta and Galleria mellonella, and evolution. BMC microbiology. **8:** 214.

Fullner and Mekalanos (2000). In vivo covalent cross-linking of cellular actin by the Vibrio cholerae RTX toxin. The EMBO journal. **19:** 5315-5323.

Gaede, Baumeister and Heesemann (1995). Decomplementation by cobra venom factor suppresses Yersinia-induced arthritis in rats. Infection and immunity. **63:** 3697-3701.

Gaede and Heesemann (1995). Arthritogenicity of genetically manipulated Yersinia enterocolitica serotype O8 for Lewis rats. Infection and immunity. **63:** 714-719.

Gaede, Mack and Heesemann (1992). Experimental Yersinia enterocolitica infection in rats: analysis of the immune response to plasmid-encoded antigens of arthritis-susceptible Lewis rats and arthritis-resistant Fischer rats. Medical microbiology and immunology. **181:** 165-172.

Galvan, Chen and Schifferli (2007). The Psa fimbriae of Yersinia pestis interact with phosphatidylcholine on alveolar epithelial cells and pulmonary surfactant. Infection and immunity. **75:** 1272-1279.

Galyov, Hakansson, Forsberg and Wolf-Watz (1993). A secreted protein kinase of Yersinia pseudotuberculosis is an indispensable virulence determinant. Nature. **361:** 730-732.

Garcia-Contreras, Zhang, Kim and Wood (2008). Protein translation and cell death: the role of rare tRNAs in biofilm formation and in activating dormant phage killer genes. PloS one. **3:** e2394.

Garrity (2004). Taxonomic Outline of the Procaryotes. Bergey's Manual of Systematic Bacteriology, Springer-Verlag, New York. **Second Edition., Release 5.0.**

Gayraud, Scavizzi, Mollaret, Guillevin and Hornstein (1993). Antibiotic treatment of Yersinia enterocolitica septicemia: a retrospective review of 43 cases. Clin Infect Dis. **17:** 405-410.

Gendlina, Held, Bartra, Gallis, Doneanu, Goodlett, Plano and Collins (2007). Identification and type III-dependent secretion of the Yersinia pestis insecticidal-like proteins. Molecular microbiology. **64:** 1214-1227.

Gierczynski, Jagielski and Rastawicki (2002). Molecular virulence attributes and occurrence of pYV-bearing strains among human clinical isolates of Yersinia enterocolitica in Poland. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. **21:** 158-159.

Golubov (2005). Comparative genome analysis of *Yersinia*. . Dissertation, LMU München: Fakultät für Biologie.

Golubov, Gierczynski, Heesemann and Rakin (2005). A novel insertion sequence element, IS Yen2, as an epidemiological marker for weakly pathogenic bioserotypes of Yersinia enterocolitica. International journal of medical microbiology : IJMM. **295:** 213-226.

Gonzalez, Lichtensteiger, Caughlan and Vimr (2002). Conserved filamentous prophage in Escherichia coli O18:K1:H7 and Yersinia pestis biovar orientalis. Journal of bacteriology. **184:** 6050-6055.

Graham, Smiley, Russell and Nairn (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. The Journal of general virology. **36:** 59-74.

Granfors (1998). Host-microbe interaction in reactive arthritis: does HLA-B27 have a direct effect? The Journal of rheumatology. **25:** 1659-1661.

Granfors, Merilahti-Palo, Luukkainen, Mottonen, Lahesmaa, Probst, Marker-Hermann and Toivanen (1998). Persistence of Yersinia antigens in peripheral blood cells from patients with Yersinia enterocolitica O:3 infection with or without reactive arthritis. Arthritis and rheumatism. **41:** 855-862.

Grant, Bennett-Wood and Robins-Browne (1998). Identification of virulence-associated characteristics in clinical isolates of Yersinia enterocolitica lacking classical virulence markers. Infection and immunity. **66:** 1113-1120.

Greenwood and Hooper (1990). Excretion of Yersinia spp. associated with consumption of pasteurized milk. Epidemiology and infection. **104:** 345-350.

Gripenberg-Lerche, Skurnik and Toivanen (1995). Relation of YadA to arthritogenicity of Yersinia enterocolitica: experimental studies in the rat. Contributions to microbiology and immunology. **13:** 188-190.

Gripenberg-Lerche, Skurnik, Zhang, Soderstrom and Toivanen (1994). Role of YadA in arthritogenicity of Yersinia enterocolitica serotype O:8: experimental studies with rats. Infection and immunity. **62:** 5568-5575.

Grosdent, Maridonneau-Parini, Sory and Cornelis (2002). Role of Yops and adhesins in resistance of Yersinia enterocolitica to phagocytosis. Infection and immunity. **70:** 4165-4176.

Groves, Rittinger, Amstutz, Berry, Holden, Cornelis and Caron (2010). Sequestering of Rac by the Yersinia effector YopO blocks Fcgamma receptor-mediated phagocytosis. The Journal of biological chemistry. **285:** 4087-4098.

Guan and Dixon (1990). Protein tyrosine phosphatase activity of an essential virulence determinant in Yersinia. Science. **249:** 553-556.

Guinet, Carniel and Leclercq (2011). Transfusion-Transmitted Yersinia enterocolitica Sepsis. Clin Infect Dis. **53:** 583-591.

Gutman, Ottesen, Quan, Noce and Katz (1973). An inter-familial outbreak of Yersinia enterocolitica enteritis. The New England journal of medicine. **288:** 1372-1377.

Hacker, Bender, Ott, Wingender, Lund, Marre and Goebel (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal Escherichia coli isolates. Microbial pathogenesis. **8:** 213-225.

Haensch, Bianucci, Signoli, Rajerison, Schultz, Kacki, Vermunt, Weston, Hurst, Achtman, Carniel and Bramanti (2010). Distinct clones of Yersinia pestis caused the black death. PLoS pathogens. **6**.

Hakansson, Galyov, Rosqvist and Wolf-Watz (1996). The Yersinia YpkA Ser/Thr kinase is translocated and subsequently targeted to the inner surface of the HeLa cell plasma membrane. Molecular microbiology. **20:** 593-603.

Haller, Carlson, Pederson and Pierson (2000). A chromosomally encoded type III secretion pathway in Yersinia enterocolitica is important in virulence. Molecular microbiology. **36:** 1436-1446.

Hamad and Nilles (2007). Roles of YopN, LcrG and LcrV in controlling Yops secretion by Yersinia pestis. Advances in experimental medicine and biology. **603:** 225-234.

Hamburger, Brown, Isberg and Bjorkman (1999). Crystal structure of invasin: a bacterial integrin-binding protein. Science. **286:** 291-295.

Hanahan (1983). Studies on transformation of Escherichia coli with plasmids. Journal of molecular biology. **166:** 557-580.

Haverly, Harrison and Dougherty (1996). Yersinia enterocolitica bacteremia associated with red blood cell transfusion. Archives of pathology & laboratory medicine. **120:** 499-500.

Heermann and Fuchs (2008). Comparative analysis of the Photorhabdus luminescens and the Yersinia enterocolitica genomes: uncovering candidate genes involved in insect pathogenicity. BMC genomics. **9:** 40.

Heesemann, Algermissen and Laufs (1984). Genetically manipulated virulence of Yersinia enterocolitica. Infection and immunity. **46:** 105-110.

Heesemann, Gross and Gruter (1987). Genetic manipulation of virulence of Yersinia enterocolitica. Contributions to microbiology and immunology. **9:** 312-316.

Heesemann and Grüter (1987). Genetic evidence that the outer membrane protein Yop1 of Yersinia enterocolitica mediates adherence and phagocytosis resistance to human epithelial cells. FEMS Microbiol. Lett. **40:** 37-41.

Heesemann, Hantke, Vocke, Saken, Rakin, Stojiljkovic and Berner (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. Molecular microbiology. **8:** 397-408.

Heesemann and Laufs (1983). Construction of a mobilizable Yersinia enterocolitica virulence plasmid. Journal of bacteriology. **155:** 761-767.

Henderson, Ng, Talukder and Thanassi (2011). Function of the usher N-terminus in catalysing pilus assembly. Molecular microbiology. **79:** 954-967.

Heroven and Dersch (2006). RovM, a novel LysR-type regulator of the virulence activator gene rovA, controls cell invasion, virulence and motility of Yersinia pseudotuberculosis. Molecular microbiology. **62:** 1469-1483.

Heroven, Nagel, Tran, Parr and Dersch (2004). RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and rovA expression in Yersinia pseudotuberculosis. Molecular microbiology. **53:** 871-888.

Heusipp, Young and Miller (2001). HreP, an in vivo-expressed protease of Yersinia enterocolitica, is a new member of the family of subtilisin/kexin-like proteases. Journal of bacteriology. **183:** 3556-3563.

Hoelen, Tjan, Schouten, Dujardin and van Zanten (2007). Severe Yersinia enterocolitica sepsis after blood transfusion. The Netherlands journal of medicine. **65:** 301-303.

Hoffman, Lindberg and Brubaker (1980). Structural studies of the O-specific side-chains of the lipopolysaccharide from Yersinia enterocolitica Ye 128. Carbohydrate research. **78:** 212-214.

Hoffmann, van Erp, Trulzsch and Heesemann (2004). Transcriptional responses of murine macrophages to infection with Yersinia enterocolitica. Cellular microbiology. **6:** 377-390.

Hoiczyk and Blobel (2001). Polymerization of a single protein of the pathogen Yersinia enterocolitica into needles punctures eukaryotic cells. Proceedings of the National Academy of Sciences of the United States of America. **98:** 4669-4674.

Hoiczyk, Roggenkamp, Reichenbecher, Lupas and Heesemann (2000). Structure and sequence analysis of Yersinia YadA and Moraxella UspAs reveal a novel class of adhesins. The EMBO journal. **19:** 5989-5999.

Howard, Gaunt, Hinds, Witney, Stabler and Wren (2006). Application of comparative phylogenomics to study the evolution of Yersinia enterocolitica and to identify genetic differences relating to pathogenicity. Journal of bacteriology. **188:** 3645-3653.

Hurst, Becher, Young, Nelson and Glare (2011). Yersinia entomophaga sp. nov., isolated from the New Zealand grass grub Costelytra zealandica. International journal of systematic and evolutionary microbiology. **61:** 844-849.

Hussain, Grove, Griffiths, Busby and Cole (1994). A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. Molecular microbiology. **12:** 153-163.

Hyman (1988). A new method of sequencing DNA. Analytical biochemistry. **174:** 423-436.

Iriarte and Cornelis (1995). MyfF, an element of the network regulating the synthesis of fibrillae in Yersinia enterocolitica. Journal of bacteriology. **177:** 738-744.

Iriarte and Cornelis (1996). Molecular determinants of Yersinia pathogenesis. Microbiologia. **12:** 267-280.

Iriarte and Cornelis (1998). YopT, a new Yersinia Yop effector protein, affects the cytoskeleton of host cells. Molecular microbiology. **29:** 915-929.

Isberg (1989a). Determinants for thermoinducible cell binding and plasmid-encoded cellular penetration detected in the absence of the Yersinia pseudotuberculosis invasin protein. Infection and immunity. **57:** 1998-2005.

Isberg (1989b). Mammalian cell adhesion functions and cellular penetration of enteropathogenic Yersinia species. Molecular microbiology. **3:** 1449-1453.

Isberg and Barnes (2000). Border patrols and secret passageways across the intestinal epithelium. Trends in microbiology. **8:** 291-293.

Isberg and Leong (1990). Multiple beta 1 chain integrins are receptors for invasin, a protein that promotes bacterial penetration into mammalian cells. Cell. **60:** 861-871.

Isberg, Voorhis and Falkow (1987). Identification of invasin: a protein that allows enteric bacteria to penetrate cultured mammalian cells. Cell. **50:** 769-778.

Juris, Rudolph, Huddler, Orth and Dixon (2000). A distinctive role for the Yersinia protein kinase: actin binding, kinase activation, and cytoskeleton disruption. Proceedings of the National Academy of Sciences of the United States of America. **97:** 9431-9436.

Kado and Liu (1981). Rapid procedure for detection and isolation of large and small plasmids. Journal of bacteriology. **145:** 1365-1373.

Kaniga, Delor and Cornelis (1991). A wide-host-range suicide vector for improving reverse genetics in gram-negative bacteria: inactivation of the blaA gene of Yersinia enterocolitica. Gene. **109:** 137-141.

Kasari, Kurg, Margus, Tenson and Kaldalu (2010). The Escherichia coli mqsR and ygiT genes encode a new toxin-antitoxin pair. Journal of bacteriology. **192:** 2908-2919.

Kirjavainen, Jarva, Biedzka-Sarek, Blom, Skurnik and Meri (2008). Yersinia enterocolitica serum resistance proteins YadA and ail bind the complement regulator C4b-binding protein. PLoS pathogens. **4:** e1000140.

Knapp, Hacker, Jarchau and Goebel (1986). Large, unstable inserts in the chromosome affect virulence properties of uropathogenic Escherichia coli O6 strain 536. Journal of bacteriology. **168:** 22-30.

Krall, Zhang and Barbieri (2004). Intracellular membrane localization of pseudomonas ExoS and Yersinia YopE in mammalian cells. The Journal of biological chemistry. **279:** 2747-2753.

Kudryashov, Cordero, Reisler and Satchell (2008a). Characterization of the enzymatic activity of the actin cross-linking domain from the Vibrio cholerae MARTX Vc toxin. The Journal of biological chemistry. **283:** 445-452.

Kudryashov, Durer, Ytterberg, Sawaya, Pashkov, Prochazkova, Yeates, Loo, Loo, Satchell and Reisler (2008b). Connecting actin monomers by iso-peptide bond is a toxicity mechanism of the Vibrio cholerae MARTX toxin. Proceedings of the National Academy of Sciences of the United States of America. **105:** 18537-18542.

Lachica and Zink (1984). Plasmid-associated cell surface charge and hydrophobicity of Yersinia enterocolitica. Infection and immunity. **44:** 540-543.

Laemmli (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. **227:** 680-685.

Lally, Hill, Kieba and Korostoff (1999). The interaction between RTX toxins and target cells. Trends in microbiology. **7:** 356-361.

Lally, Kieba, Sato, Green, Rosenbloom, Korostoff, Wang, Shenker, Ortlepp, Robinson and Billings (1997). RTX toxins recognize a beta2 integrin on the surface of human target cells. The Journal of biological chemistry. **272:** 30463-30469.

Lambert de Rouvroit, Sluiters and Cornelis (1992). Role of the transcriptional activator, VirF, and temperature in the expression of the pYV plasmid genes of Yersinia enterocolitica. Molecular microbiology. **6:** 395-409.

Laukkanen, Martinez, Siekkinen, Ranta, Maijala and Korkeala (2009). Contamination of carcasses with human pathogenic Yersinia enterocolitica 4/O:3 originates from pigs infected on farms. Foodborne pathogens and disease. **6:** 681-688.

Leclercq, Martin, Vergnes, Ounnoughene, Laran, Giraud and Carniel (2005). Fatal Yersinia enterocolitica biotype 4 serovar O:3 sepsis after red blood cell transfusion. Transfusion. **45:** 814-818.

Lee, Gerber, Lonsway, Smith, Carter, Puhr, Parrish, Sikes, Finton and Tauxe (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.

Lee, Lee, Kim, Kim, Oh, Kim, Kim and Choi (2008). Vibrio vulnificus rtxE is important for virulence, and its expression is induced by exposure to host cells. Infection and immunity. **76:** 1509-1517.

Lehr, Schutz, Oberhettinger, Ruiz-Perez, Donald, Palmer, Linke, Henderson and Autenrieth (2010). C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of Yersinia enterocolitica are decisive for its recognition and assembly by BamA. Molecular microbiology. **78:** 932-946.

Leung, Reisner and Straley (1990). YopM inhibits platelet aggregation and is necessary for virulence of Yersinia pestis in mice. Infection and immunity. **58:** 3262-3271.

Li, Rock and Nelson (2008). Identification and characterization of a repeat-in-toxin gene cluster in Vibrio anguillarum. Infection and immunity. **76:** 2620-2632.

Lin, Fullner, Clayton, Sexton, Rogers, Calia, Calderwood, Fraser and Mekalanos (1999). Identification of a vibrio cholerae RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. Proceedings of the National Academy of Sciences of the United States of America. **96:** 1071-1076.

Linhartova, Bumba, Masin, Basler, Osicka, Kamanova, Prochazkova, Adkins, Hejnova-Holubova, Sadilkova, Morova and Sebo (2010). RTX proteins: a highly diverse family secreted by a common mechanism. FEMS microbiology reviews. **34:** 1076-1112.

Liolios, Chen, Mavromatis, Tavernarakis, Hugenholtz, Markowitz and Kyrpides (2010). The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. Nucleic acids research. **38:** D346-354.

Liu, Chen, Galvan, Lasaro and Schifferli (2006). Effects of Psa and F1 on the adhesive and invasive interactions of Yersinia pestis with human respiratory tract epithelial cells. Infection and immunity. **74:** 5636-5644.

Locher, Lehnert, Krauss, Heesemann, Groll and Wilharm (2005). Crystal structure of the Yersinia enterocolitica type III secretion chaperone SycT. The Journal of biological chemistry. **280:** 31149-31155.

Lu, Yao, Meng, Cheng and Black (2002). Overexpression of apolipoprotein A-IV enhances lipid transport in newborn swine intestinal epithelial cells. The Journal of biological chemistry. **277:** 31929-31937.

Ludwig, Garcia, Bauer, Jarchau, Benz, Hoppe and Goebel (1996). Analysis of the in vivo activation of hemolysin (HlyA) from Escherichia coli. Journal of bacteriology. **178:** 5422-5430.

Ludwig, Jarchau, Benz and Goebel (1988). The repeat domain of Escherichia coli haemolysin (HlyA) is responsible for its Ca2+-dependent binding to erythrocytes. Molecular & general genetics : MGG. **214:** 553-561.

Mantle and Allen (1981). Isolation and characterization of the native glycoprotein from pig small-intestinal mucus. The Biochemical journal. **195:** 267-275.

Mantle, Pearson and Allen (1980). Pig gastric and small-intestinal mucus glycoproteins: proposed role in polymeric structure for protein joined by disulphide bridges. Biochemical Society transactions. **8:** 715-716.

Marra and Isberg (1997). Invasin-dependent and invasin-independent pathways for translocation of Yersinia pseudotuberculosis across the Peyer's patch intestinal epithelium. Infection and immunity. **65:** 3412-3421.

Marsal, Winblad and Wollheim (1981). Yersinia enterocolitica arthritis in southern Sweden: a four-year follow-up study. British medical journal. **283:** 101-103.

Martinez, Fredriksson-Ahomaa, Pallotti, Rosmini, Houf and Korkeala (2011). Variation in the prevalence of enteropathogenic Yersinia in slaughter pigs from Belgium, Italy, and Spain. Foodborne pathogens and disease. **8:** 445-450.

Masin, Basler, Knapp, El-Azami-El-Idrissi, Maier, Konopasek, Benz, Leclerc and Sebo (2005). Acylation of lysine 860 allows tight binding and cytotoxicity of Bordetella adenylate cyclase on CD11b-expressing cells. Biochemistry. **44:** 12759-12766.

Matsumoto and Young (2006). Proteomic and functional analysis of the suite of Ysp proteins exported by the Ysa type III secretion system of Yersinia enterocolitica Biovar 1B. Molecular microbiology. **59:** 689-706.

Matsumoto and Young (2009). Translocated effectors of Yersinia. Current opinion in microbiology. **12:** 94-100.

Maxson and Darwin (2004). Identification of inducers of the Yersinia enterocolitica phage shock protein system and comparison to the regulation of the RpoE and Cpx extracytoplasmic stress responses. Journal of bacteriology. **186:** 4199-4208.

McDonald, Vacratsis, Bliska and Dixon (2003). The yersinia virulence factor YopM forms a novel protein complex with two cellular kinases. The Journal of biological chemistry. **278:** 18514-18523.

McNally, Cheasty, Fearnley, Dalziel, Paiba, Manning and Newell (2004). Comparison of the biotypes of Yersinia enterocolitica isolated from pigs, cattle and sheep at slaughter and from humans with yersiniosis in Great Britain during 1999-2000. Letters in applied microbiology. **39:** 103-108.

Merhej, Adekambi, Pagnier, Raoult and Drancourt (2008). Yersinia massiliensis sp. nov., isolated from fresh water. International journal of systematic and evolutionary microbiology. **58:** 779-784.

Metzker (2005). Emerging technologies in DNA sequencing. Genome research. **15:** 1767-1776.

Michel-Briand and Baysse (2002). The pyocins of Pseudomonas aeruginosa. Biochimie. **84:** 499-510.

Miller, Bliska and Falkow (1990). Nucleotide sequence of the Yersinia enterocolitica ail gene and characterization of the Ail protein product. Journal of bacteriology. **172:** 1062-1069.

Miller and Falkow (1988). Evidence for two genetic loci in Yersinia enterocolitica that can promote invasion of epithelial cells. Infection and immunity. **56:** 1242-1248.

Miller, Farmer, Hill and Falkow (1989). The ail locus is found uniquely in Yersinia enterocolitica serotypes commonly associated with disease. Infection and immunity. **57:** 121-131.

Miller, Finlay and Falkow (1988). Factors essential for the penetration of mammalian cells by Yersinia. Current topics in microbiology and immunology. **138:** 15-39.

Mittal, Peak-Chew and McMahon (2006). Acetylation of MEK2 and I kappa B kinase (IKK) activation loop residues by YopJ inhibits signaling. Proceedings of the National Academy of Sciences of the United States of America. **103:** 18574-18579.

Moran (2002). Microbial minimalism: genome reduction in bacterial pathogens. Cell. **108:** 583-586.

Morris, Prado, Ferreccio, Robins-Browne, Bordun, Cayazzo, Kay and Levine (1991). Yersinia enterocolitica isolated from two cohorts of young children in Santiago, Chile: incidence of and lack of correlation between illness and proposed virulence factors. Journal of clinical microbiology. **29:** 2784-2788.

Mueller, Broz, Muller, Ringler, Erne-Brand, Sorg, Kuhn, Engel and Cornelis (2005). The V-antigen of Yersinia forms a distinct structure at the tip of injectisome needles. Science. **310:** 674-676.

Mukherjee, Keitany, Li, Wang, Ball, Goldsmith and Orth (2006). Yersinia YopJ acetylates and inhibits kinase activation by blocking phosphorylation. Science. **312:** 1211-1214.

Mulder, Michiels, Simonet, Sory and Cornelis (1989). Identification of additional virulence determinants on the pYV plasmid of Yersinia enterocolitica W227. Infection and immunity. **57:** 2534-2541.

Muller-Loennies, Rund, Ervela, Skurnik and Holst (1999). The structure of the carbohydrate backbone of the core-lipid A region of the lipopolysaccharide from a clinical isolate of Yersinia enterocolitica O:9. European journal of biochemistry / FEBS. **261:** 19-24.

Murros-Kontiainen, Fredriksson-Ahomaa, Korkeala, Johansson, Rahkila and Bjorkroth (2010a). Yersinia nurmii sp. nov. International journal of systematic and evolutionary microbiology.

Murros-Kontiainen, Johansson, Niskanen, Fredriksson-Ahomaa, Korkeala and Bjorkroth (2010b). Yersinia pekkanenii sp. nov. International journal of systematic and evolutionary microbiology.

Nagel, Lahrz and Dersch (2001). Environmental control of invasin expression in Yersinia pseudotuberculosis is mediated by regulation of RovA, a transcriptional activator of the SlyA/Hor family. Molecular microbiology. **41:** 1249-1269.

Navarro, Alto and Dixon (2005). Functions of the Yersinia effector proteins in inhibiting host immune responses. Current opinion in microbiology. **8:** 21-27.

Navarro, Koller, Nordfelth, Wolf-Watz, Taylor and Dixon (2007). Identification of a molecular target for the Yersinia protein kinase A. Molecular cell. **26:** 465-477.

Neubauer, Aleksic, Hensel, Finke and Meyer (2000). Yersinia enterocolitica 16S rRNA gene types belong to the same genospecies but form three homology groups. International journal of medical microbiology : IJMM. **290:** 61-64.

Nilles, Fields and Straley (1998). The V antigen of Yersinia pestis regulates Yop vectorial targeting as well as Yop secretion through effects on YopB and LcrG. Journal of bacteriology. **180:** 3410-3420.

Nishi, Sheikh, Mizuguchi, Luisi, Burland, Boutin, Rose, Blattner and Nataro (2003). The export of coat protein from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system. The Journal of biological chemistry. **278:** 45680-45689.

Nuccio and Baumler (2007). Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. Microbiology and molecular biology reviews : MMBR. **71:** 551-575.

Okamoto, Inoue, Shimizu, Hara and Miyama (1982). Further purification and characterization of heat-stable enterotoxin produced by Yersinia enterocolitica. Infection and immunity. **35:** 958-964.

Okamoto, Miyama, Takeda, Takeda and Miwatani (1983). Cross-neutralization of heat-stable enterotoxin activity of enterotoxigenic *Eschierichia coli* and of *Yersinia enterocolitica*. FEMS microbiology letters. **16:** 85-87.

Orth, Xu, Mudgett, Bao, Palmer, Bliska, Mangel, Staskawicz and Dixon (2000). Disruption of signaling by Yersinia effector YopJ, a ubiquitin-like protein protease. Science. **290:** 1594-1597.

Ostolaza, Bakas and Goni (1997). Balance of electrostatic and hydrophobic interactions in the lysis of model membranes by E. coli alpha-haemolysin. The Journal of membrane biology. **158:** 137-145.

Overbeek, Begley, Butler, Choudhuri, Chuang, Cohoon, de Crecy-Lagard, Diaz, Disz, Edwards, Fonstein, Frank, Gerdes, Glass, Goesmann, Hanson, Iwata-Reuyl, Jensen, Jamshidi, Krause, Kubal, Larsen, Linke, McHardy, Meyer, Neuweger, Olsen, Olson, Osterman, Portnoy, Pusch, Rodionov, Ruckert, Steiner, Stevens, Thiele, Vassieva, Ye, Zagnitko and Vonstein (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic acids research. **33:** 5691-5702.

Paauw, Leverstein-van Hall, van Kessel, Verhoef and Fluit (2009). Yersiniabactin reduces the respiratory oxidative stress response of innate immune cells. PloS one. **4:** e8240.

Pelludat, Hogardt and Heesemann (2002). Transfer of the core region genes of the Yersinia enterocolitica WA-C serotype O:8 high-pathogenicity island to Y. enterocolitica MRS40, a strain with low levels of pathogenicity, confers a yersiniabactin biosynthesis phenotype and enhanced mouse virulence. Infection and immunity. **70:** 1832-1841.

Pepe, Badger and Miller (1994). Growth phase and low pH affect the thermal regulation of the Yersinia enterocolitica inv gene. Molecular microbiology. **11:** 123-135.

Pepe, Wachtel, Wagar and Miller (1995). Pathogenesis of defined invasion mutants of Yersinia enterocolitica in a BALB/c mouse model of infection. Infection and immunity. **63:** 4837-4848.

Peres, Marches, Daigle, Nougayrede, Herault, Tasca, De Rycke and Oswald (1997). A new cytolethal distending toxin (CDT) from Escherichia coli producing CNF2 blocks HeLa cell division in G2/M phase. Molecular microbiology. **24:** 1095-1107.

Perry, Balbo, Jones, Fetherston and DeMoll (1999). Yersiniabactin from Yersinia pestis: biochemical characterization of the siderophore and its role in iron transport and regulation. Microbiology. **145 (Pt 5):** 1181-1190.

Perry and Fetherston (1997). Yersinia pestis--etiologic agent of plague. Clinical microbiology reviews. **10:** 35-66.

Phan, Lee, Cherry, Tropea, Burke and Waugh (2003). High-resolution structure of the Yersinia pestis protein tyrosine phosphatase YopH in complex with a phosphotyrosyl mimetic-containing hexapeptide. Biochemistry. **42:** 13113-13121.

Pierson and Falkow (1993). The ail gene of Yersinia enterocolitica has a role in the ability of the organism to survive serum killing. Infection and immunity. **61:** 1846-1852.

Pinheiro and Ellar (2007). Expression and insecticidal activity of Yersinia pseudotuberculosis and Photorhabdus luminescens toxin complex proteins. Cellular microbiology. **9:** 2372-2380.

Plano, Day and Ferracci (2001). Type III export: new uses for an old pathway. Molecular microbiology. **40:** 284-293.

Pommier, Mejean, Giordano and Iobbi-Nivol (1998). TorD, a cytoplasmic chaperone that interacts with the unfolded trimethylamine N-oxide reductase enzyme (TorA) in Escherichia coli. The Journal of biological chemistry. **273:** 16615-16620.

Portnoy and Martinez (1985). Role of a plasmid in the pathogenicity of Yersinia species. Current topics in microbiology and immunology. **118:** 29-51.

Posfai, Koob, Kirkpatrick and Blattner (1997). Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the Escherichia coli O157:H7 genome. Journal of bacteriology. **179:** 4426-4428.

Prehna, Ivanov, Bliska and Stebbins (2006). Yersinia virulence depends on mimicry of host Rho-family nucleotide dissociation inhibitors. Cell. **126:** 869-880.

Preston, Haubold and Rainey (1998). Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. Current opinion in microbiology. **1:** 589-597.

Prochazkova and Satchell (2008). Structure-function analysis of inositol hexakisphosphate-induced autoprocessing of the Vibrio cholerae multifunctional autoprocessing RTX toxin. The Journal of biological chemistry. **283:** 23656-23664.

Rabson, Hallett and Koornhof (1975). Generalized Yersinia enterocolitica infection. The Journal of infectious diseases. **131:** 447-451.

Rakin, Noelting, Schubert and Heesemann (1999). Common and specific characteristics of the high-pathogenicity island of Yersinia enterocolitica. Infection and immunity. **67:** 5265-5274.

Rakin, Urbitsch and Heesemann (1995). Evidence for two evolutionary lineages of highly pathogenic Yersinia species. Journal of bacteriology. **177:** 2292-2298.

Rao, Guandalini, Laird and Field (1979). Effects of heat-stable enterotoxin of Yersinia enterocolitica on ion transport and cyclic guanosine 3',5'-monophosphate metabolism in rabbit ileum. Infection and immunity. **26:** 875-878.

Ratnam, Mercer, Picco, Parsons and Butler (1982). A nosocomial outbreak of diarrheal disease due to Yersinia enterocolitica serotype 0:5, biotype 1. The Journal of infectious diseases. **145:** 242-247.

Reineke, Tenzer, Rupnik, Koschinski, Hasselmayer, Schrattenholz, Schild and von Eichel-Streiber (2007). Autocatalytic cleavage of Clostridium difficile toxin B. Nature. **446:** 415-419.

Revell and Miller (2000). A chromosomally encoded regulator is required for expression of the Yersinia enterocolitica inv gene and for virulence. Molecular microbiology. **35:** 677-685.

Revell and Miller (2001). Yersinia virulence: more than a plasmid. FEMS microbiology letters. **205:** 159-164.

Rhodes, Gray, Watson, Muratore, Kim, Hewlett and Grisham (2001). Structural consequences of divalent metal binding by the adenylyl cyclase toxin of Bordetella pertussis. Archives of biochemistry and biophysics. **395:** 169-176.

Robins-Browne and Prpic (1983). Desferrioxamine and systemic yersiniosis. Lancet. **2:** 1372.

Roggenkamp, Ackermann, Jacobi, Truelzsch, Hoffmann and Heesemann (2003). Molecular analysis of transport and oligomerization of the Yersinia enterocolitica adhesin YadA. Journal of bacteriology. **185:** 3735-3744.

Roggenkamp, Ruckdeschel, Leitritz, Schmitt and Heesemann (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. Infection and immunity. **64:** 2506-2514.

Ronaghi (2001). Pyrosequencing sheds light on DNA sequencing. Genome research. **11:** 3-11.

Ronaghi, Karamohamed, Pettersson, Uhlen and Nyren (1996). Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry. **242:** 84-89.

Roppenser, Roder, Hentschke, Ruckdeschel and Aepfelbacher (2009). Yersinia enterocolitica differentially modulates RhoG activity in host cells. J Cell Sci. **122:** 696-705.

Rose, Sebo, Bellalou and Ladant (1995). Interaction of calcium with Bordetella pertussis adenylate cyclase toxin. Characterization of multiple calcium-binding sites and calcium-induced conformational changes. The Journal of biological chemistry. **270:** 26370-26376.

Rosner, Stark and Werber (2010). Epidemiology of reported Yersinia enterocolitica infections in Germany, 2001-2008. BMC public health. **10:** 337.

Rosqvist, Magnusson and Wolf-Watz (1994). Target cell contact triggers expression and polarized transfer of Yersinia YopE cytotoxin into mammalian cells. The EMBO journal. **13:** 964-972.

Rosqvist, Persson, Hakansson, Nordfeldt and Wolf-Watz (1995). Translocation of the Yersinia YopE and YopH virulence proteins into target cells is mediated by YopB and YopD. Contributions to microbiology and immunology. **13:** 230-234.

Rosqvist, Skurnik and Wolf-Watz (1988). Increased virulence of Yersinia pseudotuberculosis by two independent mutations. Nature. **334:** 522-524.

Rossi, Fetherston, Letoffe, Carniel, Perry and Ghigo (2001). Identification and characterization of the hemophore-dependent heme acquisition system of Yersinia pestis. Infection and immunity. **69:** 6707-6717.

Roussos, Stambori, Aggelis, Kanavaki, Garzonis, Makarona, Kapralos, Karambela, Dalamaga and Ferti (2001). Transfusion-mediated Yersinia enterocolitica septicemia in an adult patient with beta-thalassemia. Scandinavian journal of infectious diseases. **33:** 859-860.

Ruckdeschel, Roggenkamp, Schubert and Heesemann (1996). Differential contribution of Yersinia enterocolitica virulence factors to evasion of microbicidal action of neutrophils. Infection and immunity. **64:** 724-733.

Ruter, Buss, Scharnert, Heusipp and Schmidt (2010). A newly identified bacterial cell-penetrating peptide that reduces the transcription of pro-inflammatory cytokines. J Cell Sci. **123:** 2190-2198.

Saken, Roggenkamp, Aleksic and Heesemann (1994). Characterisation of pathogenic Yersinia enterocolitica serogroups by pulsed-field gel electrophoresis of genomic NotI restriction fragments. Journal of medical microbiology. **41:** 329-338.

Sanger, Nicklen and Coulson (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America. **74:** 5463-5467.

Satchell (2007). MARTX, multifunctional autoprocessing repeats-in-toxin toxins. Infection and immunity. **75:** 5079-5084.

Sauvonnet, Pradet-Balade, Garcia-Sanz and Cornelis (2002). Regulation of mRNA expression in macrophages after Yersinia enterocolitica infection. Role of different Yop effectors. The Journal of biological chemistry. **277:** 25133-25142.

Scheftel (2002). Yersinia enterocolitica surveillance in Minnesota. International Conference on Emerging Infectious Diseases.

Scherer, Syverton and Gey (1953). Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. The Journal of experimental medicine. **97:** 695-710.

Schiellerup, Krogfelt and Locht (2008). A comparison of self-reported joint symptoms following infection with different enteric pathogens: effect of HLA-B27. The Journal of rheumatology. **35:** 480-487.

Schierack, Nordhoff, Pollmann, Weyrauch, Amasheh, Lodemann, Jores, Tachu, Kleta, Blikslager, Tedin and Wieler (2006). Characterization of a porcine intestinal epithelial cell line for in vitro studies of microbial pathogenesis in swine. Histochemistry and cell biology. **125:** 293-305.

Schubert, Cuenca, Fischer and Heesemann (2000a). High-pathogenicity island of Yersinia pestis in enterobacteriaceae isolated from blood cultures and urine samples: prevalence and functional expression. The Journal of infectious diseases. **182:** 1268-1271.

Schubert, Rakin, Fischer, Sorsa and Heesemann (1999). Characterization of the integration site of Yersinia high-pathogenicity island in Escherichia coli. FEMS microbiology letters. **179:** 409-414.

Schubert, Rakin and Heesemann (2004). The Yersinia high-pathogenicity island (HPI): evolutionary and functional aspects. International journal of medical microbiology : IJMM. **294:** 83-94.

Schubert, Sorsa, Cuenca, Fischer, Jacobi and Heesemann (2000b). HPI of high-virulent Yersinia is found in E. coli strains causing urinary tract infection. Structural, functional aspects, and distribution. Advances in experimental medicine and biology. **485:** 69-73.

Schuchmann, Gerbes, Heesemann and Sauter (1997). Multiple liver abscesses caused by Yersinia enterocolitica in a patient receiving long-term transfusion therapy for osteomyelosclerosis. Dig Dis Sci. **42:** 2501-2504.

Schutz, Weiss, Schindler, Hallstrom, Zipfel, Linke and Autenrieth (2010). Trimer stability of YadA is critical for virulence of Yersinia enterocolitica. Infection and immunity. **78:** 2677-2690.

Seoane and Garcia Lobo (2000). Identification of a streptogramin A acetyltransferase gene in the chromosome of Yersinia enterocolitica. Antimicrobial agents and chemotherapy. **44:** 905-909.

Seoane, Sanchez and Garcia-Lobo (2003). Tandem amplification of a 28-kilobase region from the Yersinia enterocolitica chromosome containing the blaA gene. Antimicrobial agents and chemotherapy. **47:** 682-688.

Shao and Dixon (2003). YopT is a cysteine protease cleaving Rho family GTPases. Advances in experimental medicine and biology. **529:** 79-84.

Shao, Merritt, Bao, Innes and Dixon (2002). A Yersinia effector and a Pseudomonas avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. Cell. **109:** 575-588.

Shao, Vacratsis, Bao, Bowers, Fierke and Dixon (2003). Biochemical characterization of the Yersinia YopT protease: cleavage site and recognition elements in Rho GTPases. Proceedings of the National Academy of Sciences of the United States of America. **100:** 904-909.

Sheahan, Cordero and Satchell (2007). Autoprocessing of the Vibrio cholerae RTX toxin by the cysteine protease domain. The EMBO journal. **26:** 2552-2561.

Sheahan and Satchell (2007). Inactivation of small Rho GTPases by the multifunctional RTX toxin from Vibrio cholerae. Cellular microbiology. **9:** 1324-1335.

Siebold, Flukiger, Beutler and Erni (2001). Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS). FEBS letters. **504:** 104-111.

Siguier, Perochon, Lestrade, Mahillon and Chandler (2006). ISfinder: the reference centre for bacterial insertion sequences. Nucleic acids research. **34:** D32-36.

Simon, Randolph, Irwin and Binkowski (1983). Stabilization of proteins by a bacteriophage T4 gene cloned in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America. **80:** 2059-2062.

Simonet, Riot, Fortineau and Berche (1996). Invasin production by Yersinia pestis is abolished by insertion of an IS200-like element within the inv gene. Infection and immunity. **64:** 375-379.

Sirisena and Skurnik (2003). Mutations in the genes for synthesis of the outer core region of the lipopolysaccharide of Yersinia enterocolitica O:3. Journal of applied microbiology. **94:** 686-692.

Skrzypek, Myers-Morales, Whiteheart and Straley (2003). Application of a Saccharomyces cerevisiae model to study requirements for trafficking of Yersinia pestis YopM in eucaryotic cells. Infection and immunity. **71:** 937-947.

Skurnik (1985). Expression of antigens encoded by the virulence plasmid of Yersinia enterocolitica under different growth conditions. Infection and immunity. **47:** 183-190.

Skurnik (2003). Molecular genetics, biochemistry and biological role of Yersinia lipopolysaccharide. Advances in experimental medicine and biology. **529:** 187-197.

Skurnik and Bengoechea (2003). The biosynthesis and biological role of lipopolysaccharide O-antigens of pathogenic Yersiniae. Carbohydrate research. **338:** 2521-2529.

Skurnik, Bolin, Heikkinen, Piha and Wolf-Watz (1984). Virulence plasmid-associated autoagglutination in Yersinia spp. Journal of bacteriology. **158:** 1033-1036.

Skurnik, Peippo and Ervela (2000). Characterization of the O-antigen gene clusters of Yersinia pseudotuberculosis and the cryptic O-antigen gene cluster of Yersinia pestis shows that the plague bacillus is most closely related to and has evolved from Y. pseudotuberculosis serotype O:1b. Molecular microbiology. **37:** 316-330.

Skurnik and Toivanen (1992). LcrF is the temperature-regulated activator of the yadA gene of Yersinia enterocolitica and Yersinia pseudotuberculosis. Journal of bacteriology. **174:** 2047-2051.

Skurnik, Venho, Bengoechea and Moriyon (1999). The lipopolysaccharide outer core of Yersinia enterocolitica serotype O:3 is required for virulence and plays a role in outer membrane integrity. Molecular microbiology. **31:** 1443-1462.

Skurnik, Venho, Toivanen and al-Hendy (1995). A novel locus of Yersinia enterocolitica serotype O:3 involved in lipopolysaccharide outer core biosynthesis. Molecular microbiology. **17:** 575-594.

Skurnik and Wolf-Watz (1989). Analysis of the yopA gene encoding the Yop1 virulence determinants of Yersinia spp. Molecular microbiology. **3:** 517-529.

Sonnenwirth (1970). Bacteremia with and without meningitis due to Yersinia enterocolitica, Edwardsiella tarda, Comamonas terrigena, and Pseudomonas maltophilia. Annals of the New York Academy of Sciences. **174:** 488-502.

Sory, Boland, Lambermont and Cornelis (1995). Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the cyaA gene fusion approach. Proceedings of the National Academy of Sciences of the United States of America. **92:** 11998-12002.

Sprague and Neubauer (2005). Yersinia aleksiciae sp. nov. International journal of systematic and evolutionary microbiology. **55:** 831-835.

Sprague, Scholz, Amann, Busse and Neubauer (2008). Yersinia similis sp. nov. International journal of systematic and evolutionary microbiology. **58:** 952-958.

Strauch, Kaspar, Schaudinn, Dersch, Madela, Gewinner, Hertwig, Wecke and Appel (2001). Characterization of enterocoliticin, a phage tail-like bacteriocin, and its effect on pathogenic Yersinia enterocolitica strains. Applied and environmental microbiology. **67:** 5634-5642.

Studier and Moffatt (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of molecular biology. **189:** 113-130.

Sun, Wu, Fedorov, Almo and Zhang (2003). Crystal structure of the Yersinia protein-tyrosine phosphatase YopH complexed with a specific small molecule inhibitor. The Journal of biological chemistry. **278:** 33392-33399.

Tahir, Kuusela and Skurnik (2000). Functional mapping of the Yersinia enterocolitica adhesin YadA. Identification Of eight NSVAIG - S motifs in the amino-terminal half of the protein involved in collagen binding. Molecular microbiology. **37:** 192-206.

Takao, Tominaga, Shimonishi, Hara, Inoue and Miyama (1984). Primary structure of heat-stable enterotoxin produced by Yersinia enterocolitica. Biochemical and biophysical research communications. **125:** 845-851.

Takao, Tominaga, Yoshimura, Shimonishi, Hara, Inoue and Miyama (1985). Isolation, primary structure and synthesis of heat-stable enterotoxin produced by Yersinia enterocolitica. European journal of biochemistry / FEBS. **152:** 199-206.

Takeuchi and Akira (2002). MyD88 as a bottle neck in Toll/IL-1 signaling. Current topics in microbiology and immunology. **270:** 155-167.

Tamm, Tarkkanen, Korhonen, Kuusela, Toivanen and Skurnik (1993). Hydrophobic domains affect the collagen-binding specificity and surface polymerization as well as the virulence potential of the YadA protein of Yersinia enterocolitica. Molecular microbiology. **10:** 995-1011.

Tchieu, Norris, Edwards and Saier (2001). The complete phosphotransferase system in Escherichia coli. J Mol Microbiol Biotechnol. **3:** 329-346.

Tennant, Grant and Robins-Browne (2003a). Pathogenicity of Yersinia enterocolitica biotype 1A. FEMS immunology and medical microbiology. **38:** 127-137.

Tennant, Skinner, Joe and Robins-Browne (2003b). Yersinia enterocolitica biotype 1A: not as harmless as you think. Advances in experimental medicine and biology. **529:** 125-128.

Tennant, Skinner, Joe and Robins-Browne (2005). Homologues of insecticidal toxin complex genes in Yersinia enterocolitica biotype 1A and their contribution to virulence. Infection and immunity. **73:** 6860-6867.

Thoerner, Bin Kingombe, Bogli-Stuber, Bissig-Choisat, Wassenaar, Frey and Jemmi (2003). PCR detection of virulence genes in Yersinia enterocolitica and Yersinia pseudotuberculosis and investigation of virulence gene distribution. Applied and environmental microbiology. **69:** 1810-1816.

Thomson, Howard, Wren, Holden, Crossman, Challis, Churcher, Mungall, Brooks, Chillingworth, Feltwell, Abdellah, Hauser, Jagels, Maddison, Moule, Sanders, Whitehead, Quail, Dougan, Parkhill and Prentice (2006). The complete genome sequence and comparative genome analysis of the high pathogenicity Yersinia enterocolitica strain 8081. PLoS genetics. **2:** e206.

Tomshich, Gorshkova, Elkin and Ovodov (1976). Lipopolysaccharide from Yersinia pseudotuberculosis, type 1 B. A structural study of O-specific chains. European journal of biochemistry / FEBS. **65:** 193-199.

Torres-Escobar, Juarez-Rodriguez and Curtiss (2010). Biogenesis of Yersinia pestis PsaA in recombinant attenuated Salmonella Typhimurium vaccine (RASV) strain. FEMS microbiology letters. **302:** 106-113.

Trasak, Zenner, Vogel, Yuksekdag, Rost, Haase, Fischer, Israel, Imhof, Linder, Schleicher and Aepfelbacher (2007). Yersinia protein kinase YopO is activated by a novel G-actin binding process. The Journal of biological chemistry. **282:** 2268-2277.

Trebesius, Harmsen, Rakin, Schmelz and Heesemann (1998). Development of rRNA-targeted PCR and in situ hybridization with fluorescently labelled oligonucleotides for detection of Yersinia species. Journal of clinical microbiology. **36:** 2557-2564.

Trosky, Liverman and Orth (2008). Yersinia outer proteins: Yops. Cellular microbiology. **10:** 557-565.

Tsubokura, Otsuki, Sato, Tanaka, Hongo, Fukushima, Maruyama and Inoue (1989). Special features of distribution of Yersinia pseudotuberculosis in Japan. Journal of clinical microbiology. **27:** 790-791.

Uliczka, Pisano, Schaake, Stolz, Rohde, Fruth, Strauch, Skurnik, Batzilla, Rakin, Heesemann and Dersch (2011). Unique cell adhesion and invasion properties of Yersinia enterocolitica O:3, the most frequent cause of human Yersiniosis. PLoS pathogens. **7:** e1002117.

Van Melderen (2010). Toxin-antitoxin systems: why so many, what for? Current opinion in microbiology. **13:** 781-785.

Vanden Bergh, Zecchinon, Fett and Desmecht (2008). Probing of Actinobacillus pleuropneumoniae ApxIIIA toxin-dependent cytotoxicity towards mammalian peripheral blood mononucleated cells. BMC research notes. **1:** 121.

Viboud and Bliska (2005). Yersinia outer proteins: role in modulation of host cell signaling responses and pathogenesis. Annual review of microbiology. **59:** 69-89.

Viboud, Mejia and Bliska (2006). Comparison of YopE and YopT activities in counteracting host signalling responses to Yersinia pseudotuberculosis infection. Cellular microbiology. **8:** 1504-1515.

Von Pawel-Rammingen, Telepnev, Schmidt, Aktories, Wolf-Watz and Rosqvist (2000). GAP activity of the Yersinia YopE cytotoxin specifically targets the Rho pathway: a mechanism for disruption of actin microfilament structure. Molecular microbiology. **36:** 737-748.

Waterfield, Bowen, Fetherston, Perry and ffrench-Constant (2001). The tc genes of Photorhabdus: a growing family. Trends in microbiology. **9:** 185-191.

Waterfield, Hares, Hinchliffe, Wren and ffrench-Constant (2007). The insect toxin complex of Yersinia. Advances in experimental medicine and biology. **603:** 247-257.

Waterman and Holden (2003). Functions and effectors of the Salmonella pathogenicity island 2 type III secretion system. Cellular microbiology. **5:** 501-511.

Wauters, Kandolo and Janssens (1987). Revised biogrouping scheme of Yersinia enterocolitica. Contributions to microbiology and immunology. **9:** 14-21.

Welch (2001). RTX toxin structure and function: a story of numerous anomalies and few analogies in toxin biology. Current topics in microbiology and immunology. **257:** 85-111.

Wiley, Nordfeldth, Rosenzweig, DaFonseca, Gustin, Wolf-Watz and Schesser (2006). The Ser/Thr kinase activity of the Yersinia protein kinase A (YpkA) is necessary for full virulence in the mouse, mollifying phagocytes, and disrupting the eukaryotic cytoskeleton. Microbial pathogenesis. **40:** 234-243.

Williams (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic acids research. **30:** 866-875.

Wollenhaupt (2007). Spondylitis ankylosans und reaktive Arthritiden. Medizinische Therapie. **2007 | 2008:** 584-591.

Wren (2003). The yersiniae--a model genus to study the rapid evolution of bacterial pathogens. Nature reviews. Microbiology. **1:** 55-64.

Yanisch-Perron, Vieira and Messing (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene. **33:** 103-119.

Yersin (1894a). La peste bubonique à Hong Kong. Archives de Médecine Navales**:** 256-261.

Yersin (1894b). La peste bubonique à Hong Kong. Ann Inst Pasteur Paris. **8:** 662-667.

Yotsu, Mii, Hayashi, Harada, Furukawa and Eto (2010). Erythema nodosum associated with Yersinia enterocolitica infection. The Journal of dermatology. **37:** 819-822.

Young (2007). The Ysa type 3 secretion system of Yersinia enterocolitica biovar 1B. Advances in experimental medicine and biology. **603:** 286-297.

Young, Miller, Falkow and Schoolnik (1990). Sequence, localization and function of the invasin protein of Yersinia enterocolitica. Molecular microbiology. **4:** 1119-1128.

Young and Young (2002). Evidence for targeting of Yop effectors by the chromosomally encoded Ysa type III secretion system of Yersinia enterocolitica. Journal of bacteriology. **184:** 5563-5571.

Zhang, Toivanen and Skurnik (1996). The gene cluster directing O-antigen biosynthesis in Yersinia enterocolitica serotype 0:8: identification of the genes for mannose and galactose biosynthesis and the gene for the O-antigen polymerase. Microbiology. **142 (Pt 2):** 277-288.

Zheng, Advani, Melefors, Glavas, Nordstrom, Ye, Engstrand and Andersson (2010). Titration-free massively parallel pyrosequencing using trace amounts of starting material. Nucleic acids research. **38:** e137.

Zhou, Monack, Kayagaki, Wertz, Yin, Wolf and Dixit (2005). Yersinia virulence factor YopJ acts as a deubiquitinase to inhibit NF-kappa B activation. The Journal of experimental medicine. **202:** 1327-1332.

Zumbihl, Aepfelbacher, Andor, Jacobi, Ruckdeschel, Rouot and Heesemann (1999). The cytotoxin YopT of Yersinia enterocolitica induces modification and cellular redistribution of the small GTP-binding protein RhoA. The Journal of biological chemistry. **274:** 29289-29293.

9. Acknowledgements

I am grateful to all people who have made this dissertation possible, who supported me with helpful advice, fruitful discussions and a pleasant atmosphere in the lab.

Most important, I would like to thank my advisor Dr. Alexander Rakin for the help and guidance he has given to me, and especially for supporting the publication of all my results.

I am also very grateful to Prof. Dr. Dr. Jürgen Heesemann for the opportunity to perform my work at the Max von Pettenkofer-Institute, for his helpful advice and discussions.

Dr. Uladzimir Antonenka introduced me to every speciality of my *Yersinia* and was my first address to discuss new ideas. Anette Schulz was a patient and helpful teacher improving my cell culture knowledge and always at my side for profound discussions. Likewise, I would like to thank Dr. Annika Schmid, Dr. Julia Niefnecker, Eva Schriefer, Dr. Christoph Guggenberger and all other *Yersinia* people for their great support and the pleasant time I have spent with them at bench and office.

In addition, I am very grateful to all my cooperation partners, including Professor Mikael Skurnik, Dr. Dirk Höper, Dr. Berthold Fartmann, Frank Uliczka and Thomas Rattei. They all supported my work with their knowledge of bioinformatics and O:3-specific lab methods. Without them, this work would have been less successful.

Beside the great support from advisors, colleagues and cooperation partners, this work got a lot of silent support from the private side.

Most important, I would like to thank my parents: they were always there to support and motivate me, and listened so often for hours to my stories from the lab or depressing experiment results they could not always comprehend in detail. Also, I am deeply grateful to my husband Christoph, who gave me the reason to go through these years of work, and who always supported me with scientific knowledge as well as his affinity and love to refill my energy.

Last but not least I would like to thank my friends, who had to wait so often for a sign of life during the stressful periods of this work. Especially Kerstin Hartwig, Dr. Kathrin Fischer and Bernadett Boeing always talked me out of bad lab day moods. I appreciate their patience so much.