

The repeat structure of two paralogous genes, *Yersinia ruckeri* Invasin (*yrInv*) and a “*Y. ruckeri* Invasin-like molecule”, (*yrIIm*) sheds light on the evolution of adhesive capacities of a fish pathogen

Agnieszka Wrobel^a, Claudio Ottoni^{a,b}, Jack C. Leo^a, Snorre Gulla^c, Dirk Linke^{a*}

^a Department of Biosciences, University of Oslo, 0316 Oslo, Norway

^b Centre for Ecological and Evolutionary Synthesis, University of Oslo, 0316, Norway

^c Fish Health Research group, Norwegian Veterinary Institute, 0454 Oslo, Norway

Corresponding author

* Corresponding author at: Department of Biosciences, University of Oslo, 0316 Oslo, Norway

E-mail address: dirk.linke@ibv.uio.no

Abstract

Inverse autotransporters comprise the recently identified type Ve secretion system and are exemplified by intimin from enterohaemorrhagic *Escherichia coli* and invasins from enteropathogenic *Yersiniae*. These proteins share a common domain architecture and promote bacterial adhesion to host cells. Here, we identified and characterized two putative inverse autotransporter genes in the fish pathogen *Yersinia ruckeri* NVH_1, namely *yrInv* (for *Y. ruckeri* invasins) and *yrIIm* (for *Y. ruckeri* invasins-like molecule). When trying to clone the highly repetitive genes for structural and functional studies, we experienced problems in obtaining PCR products. PCR failures and the highly repetitive nature of inverse autotransporters prompted us to sequence the genome of *Y. ruckeri* NVH_1 using PacBio sequencing, which produces some of the longest average read lengths available in the industry at this moment. According to our sequencing data, *YrIIm* is composed of 2603 amino acids (7812 bp) and has a molecular mass of 256.4 kDa. Based on the new genome information, we performed PCR analysis on four non-sequenced *Y. ruckeri* strains as well as the sequenced *Y.ruckeri* type strain. We found that the genes are variably present in the strains, and that the length of *yrIIm*, when present, also varies. In addition, the length of the gene product for all strains, including the type strain, was much longer than expected based on deposited sequences. The internal repeats of the *yrInv* gene product are highly diverged, but represent the same bacterial immunoglobulin-like domains as in *yrIIm*. Using qRT-PCR, we found that *yrIIm* and *yrInv* are differentially expressed under conditions relevant for pathogenesis. In addition, we compared the genomic context of both genes in the newly sequenced *Y.ruckeri* strain to all available PacBio-sequenced *Y. ruckeri* genomes, and found indications of recent events of horizontal gene transfer. Taken together, this study demonstrates and highlights the power of Single Molecule Real-Time technology for sequencing highly repetitive proteins, and sheds light on the genetic events that gave rise to these highly repetitive genes in a commercially important fish pathogen.

Keywords: inverse autotransporters, invasins, *Yersinia ruckeri*, PacBio sequencing, bacterial Ig-like domains

1. Introduction

Yersiniae are Gram-negative, rod-shaped, facultative anaerobes in the *Enterobacteriaceae* family. Eighteen different species have been described so far (Savin et al., 2014)(Reuter et al., 2014). Some members of the *Yersiniae* are pathogenic to humans and animals, in particular *Yersinia pestis*, the causative agent of plague (Rasmussen et al., 2015). *Y. ruckeri*, a fish pathogen, was initially isolated from rainbow trout in the Hagerman Valley of Idaho, USA in the 1950s (Rucker R, 1966). Since then, the bacterium has been found all over the world, including Canada, China, India, Australia, and Europe (Furones et al., 1993; Shaowu et al., 2013). The bacterium has also been associated with a broad range of non-fish hosts including muskrat (Stevenson and Daly, 1982), sea gulls and other birds (Willumsen, 1989), and humans (de Keukeleire et al., 2014).

Y. ruckeri is the causative agent of enteric redmouth disease (ERM), affecting mostly salmonids. ERM can lead to 70% mortality in infected fish farms. The typical features of the disease include hemorrhages of the mouth or tongue, exophthalmia, darkening of the skin and inflammation of the gut (Bullock, 1984; Busch, 1978). Exophthalmia, known as ‘pop-eye’, is an intraocular accumulation of

fluid which can result in rupture of the fish eye (Busch, 1978). The bacterium enters the host via the secondary gill lamellae and from there can spread to the blood and internal organs (Ohtani et al., 2014). Very little research has been done to investigate the virulence mechanisms of this pathogen. Several virulence factors have been identified, including the ruckerbactin iron uptake system, adhesins, a type III secretion system and exoproteins (Romalde and Toranzo, 1993). It is especially these extracellular proteins, including the iron-regulated haemolysin Yh1A (Fernández et al., 2007) and the metalloprotease Yrp1 (Fernandez et al., 2003; Secades and Guijarro, 1999) that have been reported to be toxic to fish, causing the typical clinical signs of ERM disease (Romalde and Toranzo, 1993).

Only a few studies have tried to identify potential adhesins in *Y. ruckeri* (Kawula et al., 1996). In contrast, many adhesins of the human pathogenic *Yersiniae* have been studied in detail (Chauhan et al., 2015). To address this issue, we searched the *Y. ruckeri* genome for proteins belonging to the intimin-invasin family of adhesins. Invasin (InvA), expressed by *Y. enterocolitica* and *Y. pseudotuberculosis*, plays a crucial role during the first stage of infection. InvA binds to $\beta 1$ integrins (Isberg and Leong, 1990) and is responsible for initial colonization and bacterial attachment (Simonet and Falkow, 1992). In addition to invasin, several others InvA-like autotransporters have been identified, including InvB (also called intimin family protein Ifp), InvC, InvD and the recently identified InvE in *Y. pseudotuberculosis*. Interestingly, *Y. pestis* possess an orthologue of InvC known as Iip (intimin/InvA-like protein) whereas Iip is disrupted by an IS285 insertion element (Chauhan et al., 2016; Pisano et al., 2012; Sadana et al., 2017; Schalk-hihi et al., 2011; Strong et al., 2011). These proteins belong to the recently described type Ve secretion system or inverse autotransporter family (Leo et al., 2012). The common feature of the inverse autotransporter family is the domain architecture. These proteins are composed of an N- signal peptide, followed by a short periplasmic region, a β -barrel domain and a C-terminal passenger (Leo et al., 2015b). The C-terminal passenger, also referred to as the extracellular domain, includes multiple repetitions of Ig-like domains, and is usually capped with a lectin domain that is presumed to host the adhesin function (Hamburger et al., 1999; Heinz et al., 2016; Tsai et al., 2010). Interestingly, the number of Ig-like domains, and thus the size of the passenger, is highly variable in inverse autotransporters (Heinz et al., 2016; Leo et al., 2012). In the current model of inverse autotransport, the protein is first transported into the periplasm by the Sec machinery via its N-terminal signal peptide (Leo et al., 2015b). Inverse autotransporters contain a transmembrane beta-barrel domain through which the C-terminal passenger is secreted into the extracellular space. According to this autotransport model, the outer membrane protein insertion depends on the BamA complex and the periplasmic chaperones SurA and Skp, whereas the chaperone/protease DegP plays a role in protein quality control (Leo et al., 2012; Oberhettinger et al., 2015, 2012). We identified two putative inverse autotransporter genes in the *Y. ruckeri* strain NVH_1 genome. We named the corresponding proteins YrInv (for *Y. ruckeri* invasin) (Leo et al., 2015a) and YrIIm (for *Y. ruckeri* invasin-like molecule)(this study). When trying to clone these genes, we experienced problems in obtaining PCR products, particularly regarding *yrIIm*. Our first attempts failed completely; only when we increased the extension times significantly did we obtain products. However, even with longer extension times, we observed multiple products that ran in a ladder-like fashion in agarose gels (Fig. S1). This suggested to us that the *yrIIm* gene was much longer than suggested by deposited sequences, and that the gene contains many repeated sequence elements that give rise to the observed ladder patterns by strand slippage or similar effects during the amplification reaction. Indeed, we had similar problems when amplifying *yrIIm* from DNA of the type strain *Y. ruckeri* ATCC29743. These observations, in combination with a lack of genomic information about *Y. ruckeri* NVH_1, prompted us to sequence the genome. We employed the third-generation Single Molecule Real-Time (SMRT) sequencing technology of Pacific Biosciences (PacBio; Menlo Park CA), which produces some of the longest average read lengths (> 10,000 bp) available in the industry at this moment. This sequencing technology allowed us to analyze the repetitive structure of the two inverse autotransporters of NVH_1. Here, we accurately characterize the genes, *yrInv* and *yrIIm*, and describe their expression under various conditions, and compare them to their orthologues across various *Y. ruckeri* strains.

2. Materials and methods

2.1. Bacterial strains and growth conditions

Y. ruckeri NVH_1 as well as *Y. ruckeri* 1435-94 and 1006-94 were kindly provided by a Norwegian Veterinary Institute in Oslo, Norway. *Y. ruckeri* NVH_1 was recovered from an outbreak of

clinical yersiniosis in farmed salmonid fish in Norway in the late 1980s (Wasteson U, Hvaal AB, Sørum H, Myhr E, 1989). Molecular characterization of the isolate by multilocus sequence typing showed that NVH_1 belongs to sequence type 3 (clonal complex 1) (Bastardo et al., 2012). *Y. ruckeri* OMBL4 and RS41 were provided by the University of Helsinki (Table 1). The *Escherichia coli* TOP10 strain was obtained from Invitrogen. All strains were grown in lysogeny broth (LB) containing 10g/l salt or LB agar plates (Bertani, 1951) at 28°C (for *Y. ruckeri*) or 37°C (for *E. coli*).

Table 1. *Y. ruckeri* strains used in this study

Strain	Host / Provenance	References	Source
NVH_1	Salmonid fish (Norway)	(Wasteson et al., 1989)	Duncan Colquhoun, Norwegian Veterinary Institute, Oslo
ATCC29473 (DSMZ18506)	Rainbow trout (Idaho, USA)	(Ross et al., 1966)	Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures
1006-94	Fish (Oslo, Norway)	(Mammeri et al., 2006)	Duncan Colquhoun, Norwegian Veterinary Institute, Oslo
1435-94	Fish (Oslo, Norway)	Unpublished	Duncan Colquhoun, Norwegian Veterinary Institute, Oslo
OMBL4	Whitefish (Finland)	(Reuter et al., 2014)	Mikael Skurnik, Haartman Institute, University of Helsinki
RS41	Rainbow trout (USA)	(Reuter et al., 2014)	Mikael Skurnik, Haartman Institute, University of Helsinki

2.2. DNA Sequencing and assembly

Initial database searches for invasin homologues were performed using pBLAST with the protein sequence of *Y. pseudotuberculosis* invasin (GenBank accession number: AAA27632.1) as a search model.

After failing to amplify and clone them for structural and functional analyses using standard molecular biology procedures, we decided to sequence the genome of NVH_1. Genomic DNA was isolated from 10 ml overnight culture of *Y. ruckeri* NVH_1 using a large genomic DNA extraction protocol (Sun et al., 2015). Genome sequencing was performed by the Norwegian Sequencing Centre (Oslo, Norway) using the Pacific Biosciences RS II platform. A DNA library was prepared following the Pacific Biosciences 20 kb library preparation protocol. The final library was selected based on a 8 kb cut-off using BluePippin (Sage Science). The library was sequenced using P6-C4 chemistry with 360 minutes movie time (time of fluorescent nucleotides incorporation). The generated reads were assembled using the Hierarchical Genome Assembly Process (HGAP) v3 (Pacific Biosciences, SMRT Analysis Software v2.3.0) comprising preassembly, assembly and consensus polishing.

2.3. Sequence Analysis and Annotation

The FASTA consensus sequence of the *Y. ruckeri* NVH_1 chromosomal genome generated through SMRT sequencing was uploaded into RAST (Rapid Annotation using Subsystem Technology) for automatic annotation of genomic features (Aziz et al., 2008). Visualization of annotations and translation of coding DNA sequences (CDSs) into proteins was done with Geneious (Kearse et al., 2012). The translated CDSs were screened for the presence of integral β -barrel domains with BOMP ('beta-barrel Outer Membrane protein Predictor') (Berven et al., 2004) and all CDSs containing a β -barrel were in turn screened for the presence of Ig-like domains with InterProScan (Quevillon et al., 2005). Repetitive regions were identified with RepeatScout (Price et al., 2005) and Tandem Repeat

Finder (Benson, 1999), and annotated with RepeatMasker (<http://www.repeatmasker.org>). Thereafter, Bedtools (Quinlan and Hall, 2010) was used to calculate the density of the repetitive regions within 500 bp-long genomic windows.

We compared the genome of *Y. ruckeri* NVH_1 (hereafter just NVH_1) with the genomes deposited in Genbank of *Y. ruckeri* Big Creek 74 (Big Creek 74) and *Y. ruckeri* CSF007-82 (CSF007-82) strains sequenced through PacBio SMRT technology, and of *Y. ruckeri* YRB (YRB) sequenced in Illumina and 454 platforms. The Illumina assembly of YRB, in contrast to the Illumina assembly of ATCC29473, was included in the genome comparison analysis due to the fact that YRB lacks both inverse autotransporters. This allowed us to make a proper genome comparison analysis despite the poor genome assembly. To be consistent with the annotation performed in NVH_1, the GenBank PacBio genomes of *Y. ruckeri* strains were re-annotated with RAST. Genome comparison was also done with three human pathogenic *Yersinia* strains, *Y. enterocolitica* 8081, *Y. pestis* CO92 and *Y. pseudotuberculosis* YPIII. Accession numbers and platforms used to sequence the genomes used in the comparative analysis are in Table 2. Locally collinear genomic blocks (LCBs) were identified by aligning the genomes of the three *Y. ruckeri* strains and of the three human pathogenic strains used in the comparative analysis to that of NVH_1 with Progressive Mauve (Darling et al., 2010), using default options and using the ‘seed family’ option to increase sensitivity. The output backbone file was then used to plot the LCBs in eight different Circos representations (Fig. 1, Fig. S2 and 3), where the genome of each strain in the comparative analysis to all others was used as reference.

Table 2. Type of sequencing technology used for genome sequencing in *Yersinia*

Species/Strain	Sequencing technology	GenBank
<i>Yersinia ruckeri</i> NVH_1	PacBio	This study (Accession to add)
<i>Yersinia ruckeri</i> ATCC29473	Illumina	GCA_000834255.1
<i>Yersinia ruckeri</i> 150	Illumina	GCF_001750505.1
<i>Yersinia ruckeri</i> CSF007-82	Illumina	GCF_000824965.1
<i>Yersinia ruckeri</i> CSF007-82	PacBio	LN681231
<i>Yersinia ruckeri</i> RS41	Illumina	GCF_001166725.1
<i>Yersinia ruckeri</i> OMBL4	Illumina	GCF_001172905.1
<i>Yersinia ruckeri</i> Big Creek 74	PacBio	NZ_CP011078.1
<i>Yersinia ruckeri</i> YRB	Illumina, 454	NZ_CP009539.1
<i>Yersinia enterocolitica</i> 8081	Sanger	NC_008800.1
<i>Yersinia pseudotuberculosis</i> ATCC6904	Illumina, 454	NZ_CP008943.1
<i>Yersinia pestis</i> CO92	Illumina	AL590842

2.4. Analysis of *Yersinia ruckeri* NVH_1 adhesins

2.4.1. In silico analysis of invasin (YrInv) and invasin like molecule (YrIIm)

N-terminal signal sequence predictions were made using Phobius (Käll et al., 2004). The domain prediction obtained with InterProScan was further revised and improved using HHpred (Homology detection & structure prediction by HMM-HMM comparison), available at the Max Planck Institute for Developmental Biology Toolkit (Söding et al., 2005), and Clustal Omega (Sievers et al., 2011; Thompson et al., 1994) (Fig. 2A, Fig S4-6).

2.4.2. Alignments of *yrInv* and *yrIIm* genomic regions

In order to investigate the degree of similarity and of structural variation of the larger genomic regions containing the two *yrInv* and *yrIIm* genes, we constructed an alignment of the genomic regions containing the identified adhesins and including up to six flanking open reading frames from the different *Y. ruckeri* strains sequenced and deposited in Genbank (Table 2), namely, sequences of NVH_1, Big Creek 74 and CSF007-82, ATCC29473, CSF, 150, YRB, RS41 and OMBL4 (the latter two available only for the *inv* genomic region) (Table 2.) The alignment was done with Geneious and refined manually, using CSF007-82 as reference for *yrIIm* and NVH_1 for *yrInv* (Fig. 2B-C).

2.5. PCR detection of *Yersinia ruckeri* inverse autotransporter adhesins

Bacterial genomic DNA was isolated from all available *Y. ruckeri* strains (Table 1) according to a large genomic DNA extraction protocol (Sun et al., 2015). Purified genomic DNA was used as the template in each PCR reaction (120 ng per 50µl reaction). It is important to note that the genomic DNA used in the PCR reactions was pure, intact, double stranded and of a high quality. This was a prerequisite for a successful and reliable PCR reaction. PCRs of full-length *yrInv* and full-length *yrIIm* were carried out according to our optimized protocol with *Taq* polymerase purchased from New England Biolabs (NEB). The cycling conditions were the following: initial denaturation at 95°C for 30 sec, denaturation at 95°C for 10 sec, annealing at 62°C for 30 sec, extension at 68°C for 10 min, final extension at 68°C for 5 min. PCR analysis of the translocation domain (β-barrel) for both *yrInv* and *yrIIm* were performed according to standard protocol with *Taq* polymerase (NEB). The cycling conditions were as follows: initial denaturation at 94°C for 3 min, denaturation at 94°C for 30 sec, annealing at 52°C for 30 sec, extension at 72°C for 60 sec, final extension 72°C for 5 min. PCR products were analyzed using 0.7% and 1,2% agarose gels for full-length PCR products and translocation domains, respectively. As a positive control for PCR reactions, 16S rDNA was amplified using universal primers unique for *Y. ruckeri* (Fig. 3) (LeJeune and Rurangirwa, 2000). Primers used in this study are listed in Table 3.

Table 3. PCR Primers and expected amplicon sizes for *Yersinia ruckeri* NVH_1

Target gene	Forward primer (5'-3')	Reverse primer (5'-3')	product size (bp)
<i>yrInv</i>	ATGGTAGGTCTCGAATGATGAATGAAA ATATCATCAGGCTTTTTTTT	ATGGTAGGTCTCAGCGCTTTATAATTGTTTTATACAA GTAACATGGC	2564
<i>yrIIm</i>	TAACGAGGGCAAAAAATGTCTACCTCC TTTACCCGAC	CGGGTGGCTCCAAGCGCTAAGACCCTGCCGACACAC AAC	7842
β-barrel	TTGAGCCAGTTCGGGAC	CTGATAATCCAACACTAAATGTGTG	735
<i>yrIIm</i>			
β-barrel	ATGGTAGGTCTCAGCGCTGTTGTTACGA TCAACCAAGTTATAAC	ATGGTAGGTCTCAGGCCACGGCTCAAGTGC GGCTAA AT	734
<i>yrInv</i>			
16S rDNA	CAGCGGAAAGTAGCTTG	TGTTCA GTGCTATTAACACTTAA	409

2.6. Quantitative reverse transcription PCR

For quantitative reverse transcription (qRT) PCR experiments, we tested the effect on adhesin gene expression of the following environmental factors: temperature (8°C, 28°C, 37°C), NaCl concentration (0 mM, 170 mM, 300 mM), oxygen availability (aerobic/anaerobic conditions), different media (Dulbecco's Modified Eagle Medium (DMEM), LB, minimal medium M9) and iron availability. For the iron availability analysis, we included 3 types of samples, containing 100 µM iron chelator (2, 2'-bipyridyl), 200 µM FeCl₂, or both. *Y. ruckeri* NVH_1 was grown in LB medium at 28°C and an optical density (OD₅₅₀) of the culture was adjusted to 1. Total RNA was extracted from three biological replicates using RiboPure™ RNA Purification Kit, yeast (ThermoFisher Scientific). Residual DNA was removed by DNase I treatment. RNA quality and concentration was measured by a Nandrop N-1000 spectrophotometer (ThermoFisher Scientific). First-strand cDNA reaction was performed using 1 µg of total RNA with the help of Moloney Murine Leukemia Virus reverse transcriptase (NEB) and random hexamer primers (ThermoFisher Scientific). In addition, a control reaction without reverse transcriptase was also included to check for genomic DNA contamination. After the reverse transcriptase reaction, cDNA was diluted tenfold in diethyl pyrocarbonate water and prepared for qRT-PCR reactions. Primers used in qRT-PCR reactions were designed using free online software from Integrated DNA Technologies. The primers used in the study are listed in the table below (Table 4). The amplification efficiency of primers was evaluated using a tenfold dilution series of the cDNA in the qRT-PCR reaction. The choice of reference genes (gyrase A, DNA polymerase I) used in this study was based on literature (Brudal et al., 2013; Rocha et al., 2015). qRT-PCR reactions were performed using SYBR Green Master mix (Roche, Basel, Switzerland) in a LightCycler® 96 Real Time PCR System (Roche,

Basel, Switzerland). Each reaction was performed in a 10 µl volume containing 2µl of cDNA, 5 µl of SYBR Green I Master mix, 1 µl of primer mix including forward and reverse primers (5µM). The qRT-PCR reaction was performed according to the manufacturer's instructions. At the end of each run, melting-curve analysis was also performed using the software provided with the machine to assess specificity of the assay. Controls lacking template cDNA or RNA that had not been reverse transcribed were included in each run. All reactions included three technical and three biological replicates (Fig. 4).

Table 4. Primers used for qRT-PCR.

Target gene	Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon size (bp)
<i>yrInv</i>	GTTACCGGTCTTACCTCAGTTAG	GAATGGTGTATAGGTTATCCCCG	136
<i>yrIlm</i>	CGTCAATGAGGACTTCCATCT	GAGGGTATTACGGCTGTCTTT	124
<i>gyrA</i>	ACCAGTAGCCGATCAATAAAGTC	AGGAAGATGTGGTTGTGACG	142
<i>polA</i>	TCCAACCTCGGTCAAACGTATC	GATTGAAATGCCGTTACTGCC	117

3. Results

3.1. An overview of a genome sequence of *Y. ruckeri* NVH_1

Here, we present the complete sequenced chromosomal genome of NVH_1. The sequenced NVH_1 genome contains a 3,7-Mb chromosome with a G+C content of 47.6%. The chromosome encodes 3,480 CDSs, 81 tRNA genes and 22 rRNA genes. The chromosomal genome of *Y.ruckeri* ATCC29473 type strain (NZ_KN150747.1) sequenced by Illumina appears to be shorter, with a lower number of rRNA and tRNA genes and CDS. This suggests multiple issues with completeness, gene annotation, and repeat detection. The general features of the four genomes annotated with RAST are summarized in the table below (Table 5). In addition to the chromosomal genome, we also noticed the presence of plasmids in the strain NVH_1. As our genes of interest are encoded on the chromosome, we ignored them for the purpose of the study. The presence of plasmids in *Y.ruckeri* ATCC29473 has been described by other authors (Garcia et al., 1998; Guilvout et al., 1988; Romalde et al., 1993). A large plasmid of approximately 75 MDa has received a lot of attention due to the fact that this plasmid might be correlated with the virulence of the pathogen. However, more research needs to be conducted in order to confirm this hypothesis.

Table 5. Summary of genomic features of selected *Y. ruckeri* strains

Feature	NVH_1 Pac Bio (this study)	ATCC29473 NZ_KN150747. 1*	BigCreek 74*	CSF007-82*
Genome size (bp)	3,766,700	3,672,847	3,699,725	3,799,036
GC content (%)	47.6	47.4	47.6	47.5
CDS number	3,480	3,420	3,387	3,515
CDS total length (bp)	2,958,357	2,907,687	3,107,346	3,189,951
CDS average length (bp)	850	850	917	907
CDS length/Genome (%)	79	79	84	84
tRNA	81	72	81	80
rRNA 5S-16S-23S	8-7-7	6-4-4	8-7-7	8-7-7

*For the purpose of our study, the genomes deposited in Genbank were re-annotated in RAST in order to use the same annotations of the *Y. ruckeri* genome sequenced in this study.

3.2. Whole-genome comparative analysis of *Y. ruckeri* strains

The initial screening with BOMP resulted in 65 CDSs containing a predicted β -barrel domain. The following prediction of domains with Interpro showed that only two of them contained Ig-like domains, the ‘invasin’ and ‘putative invasin precursor’, as reported in the RAST automatic annotation, corresponding respectively to *yrInv* and *yrIIm*. The *de novo* identification of repeat regions across the genome with RepeatScout showed a high density of repeats at the *yrIIm* locus. The analysis with Tandem Repeat Finder showed that *yrIIm* is characterized by 19 repeat units of 300 bp, each corresponding to a bacterial Ig-like domain (namely D1-D19). The DNA repeat units are characterized by a high GC content and are almost identical in sequence (97% identity across the 19 DNA repeat units). Interestingly, *yrIIm* is flanked by two transposases genes (Fig. 1, Fig. 2C). Such a highly homogeneous and repetitive structure was not observed for the *yrInv* gene, the product of which is characterized by three Ig-like domains, as revealed by the InterPro screening and manual revision (Fig 1, Fig 2A-B).

The comparative analysis of LCBs through Mauve between the chromosomes of Big Creek 74, CSF007-82 and YRB showed that *yrInv* is not present in either Big Creek 74 or YRB, whereas *yrIIm* is absent only in YRB (Fig.1). In CSF007-82, the *yrIIm* gene is duplicated (Fig. S2). *yrInv* and *yrIIm* are both absent in *Yersinia* spp. human pathogenic strains (Fig. S3).

3.2.1 Invasin-like fragment of *Y. ruckeri* YRB (YrbIIf)

Taking a closer look at the genome of YRB, we could identify a genomic region not present in Big Creek and NVH_1 that contains the coding sequences for a 257 amino- acid long ‘fibronectin type III repeat’ (an Ig-like fold) protein and a 276 amino-acid long ‘invasin-like element”, which we named Ig protein and invasin protein, respectively. The region is flanked by two transposases (Fig. 2D). InterPro screening, supported by homology prediction, showed that the invasin protein contains a N-terminal signal sequence (29 aa), followed by a short periplasmic region (77 aa) and a β -barrel domain (154 aa). Blast analysis found a 50% identical invasin protein in *Hafnia paralvei* (Gen Bank: WP_064573282.1, a potential human and animal pathogen within the *Enterobacteriaceae* family). Alignment analysis of YrInv from NVI_1 to the YRB invasin-like fragment showed only 12% identity, suggesting that the YRB fragment is not a product of a translocation event, but that the two regions and the contained genes are significantly different. The Ig protein also carries a short N-terminal signal peptide, and 2 Ig-like domains, 93 and 98 amino-acid long. This genomic region of YRB that contains the two genes could in principle be non-functional remnants of an inverse autotransporter gene, disrupted by mutations. However, the domain orientation is the opposite of what one would expect for an inverted autotransporter (i.e. the Ig-like domains precede the β -barrel domain), and both genes encode a signal peptide, suggesting that rather than being the decay products of a parental inverse autotransporter, these might be independent (and functional) genes. A possible explanation might be that these two genes constitute an inverse autotransporter ‘two-partner secretion system’ similar to ZirTS of *Salmonella* (Gal-Mor et al., 2008) (Fig. 2D).

3.3. Analysis of *Y. ruckeri* adhesins

3.3.1. Invasin (YrInv)

The YrInv protein consists of 842 amino acids with a molecular mass of 93.2 kDa and a pI of 5.44. Presence of a signal peptide confirmed that YrInv is an outer membrane protein like its well-studied homologues from other *Yersinia* species. YrInv consists of an N-terminal signal peptide, a periplasmic region, a transmembrane β -barrel domain and an extracellular region (Fig. 2A, Fig S5). The signal peptide has a length of 49 amino acids, making it unusually long; note that long signal peptides may have an important function in autotransporters in general (Bernstein, 2014). In the case of classical autotransporters, such long signal peptides function not only as targeting signal, but also prevent misfolding of the passenger in the periplasm (Szabady et al., 2005). We assume extended signal peptides play a similar role in the biogenesis of inverse autotransporters, but this has not been experimentally verified.

The periplasmic region of YrInv contains a lysin motif (LysM) minidomain, which we have previously shown to be involved in peptidoglycan binding and oligomerization (Leo et al., 2015). The translocation unit, based on secondary structure and topology prediction, is highly similar to that of other inverse autotransporters. In *Y. pseudotuberculosis* InvA, it comprises a 12-stranded β -barrel and a short

intrabarrel linker region (Fairman et al., 2012). The passenger consists of Ig-like domains capped by a C-type lectin like domain (CLTD). Three Ig-like domains were identified based on homology predictions and on comparison with *Y. enterocolitica* Inv, and designated as D0, D1 and D2 (Fig. S5). At the sequence level these three domains are highly diverged (Fig. S5). The remaining CLTD was named D3. Interestingly, a BLAST search using the YrInv sequence showed that it shares 43% identity over the whole sequence with a predicted inverse autotransporter from *Edwardsiella piscicida* C07-087 (GenBank: WP_069579258.1), an important, recently described fish pathogen. Invasin from *E. piscicida* C07-087, similarly to YrInv, consists of a β -barrel domain (56.2 % sequence identity to YrInv β -barrel) (residue 18-262) and 3 Ig-like domains (45.9 % sequence identity to YrInv Ig-like domains) (residues 272-557) capped by a CLTD (46.7 % sequence identity to YrInv CLTD) (residues 558-673) (Fig. S7).

3.3.2. Invasin like molecule (YrIIm)

The sequence of YrIIm is composed of 2603 amino acids (Fig. 2A.). It has a molecular mass of 256.4 kDa with a pI of 4.65. Subcellular localization prediction suggested that YrIIm is an outer membrane protein. Based on our secondary structure prediction, YrIIm, similarly to YrInv, consists of an N-terminal signal sequence (residues 1-39), a periplasmic domain, a transmembrane inverse autotransporter beta-barrel domain (residues 169-413) and a large extracellular region (residues 413 - 2603) (Fig. 2A). Like YrInv, the periplasmic region of YrIIm contains a LysM minidomain (Leo et al., 2015a). The large extracellular region consists of an array of 20 Ig-like domains capped with a predicted CLTD. 19 of the Ig-like domains are almost identical in sequence, suggesting recent expansion of this autotransporter passenger by duplication events (Fig. S4). Each Ig-like domain contains 100 amino acids with the exception of D0, which contains 101 amino acids. A BLAST search showed that YrIIm shares 59%, 53% and 51% identity over the whole sequence with orthologues from *Enterobacter cloacae* (Gen Bank: WP_023310354.1) and from human pathogens *Y. pseudotuberculosis* (Gen Bank: WP_042593065.1) and *Y. pestis* (Gen Bank: WP_011906434.1), respectively. *E. cloacae* is a member of *Enterobacteriaceae* family and it is an important opportunistic pathogen for humans. The invasin-like molecule from *E. cloacae* contains a transmembrane region, 69% identical in sequence to YrInv, whereas the extracellular region shares 58.8 % identity. The extracellular region is composed of 18 Ig-like domains which vary in size from 70 residues to 104 residues and it is capped by CLTD (Fig. S8).

3.3.3 Genomic context of *yrInv* and *yrIIm*

By focusing our comparative analysis more specifically to the genomic regions containing *yrInv* and *yrIIm*, we could also include assemblies and contigs from other, less complete *Y. ruckeri* strain genomes generated with either third generation- (SMRT, PacBio) or second generation (Illumina, 454) sequencing technologies, and investigate more thoroughly the structural variations within the targeted genomic regions. For *yrInv*, we aligned eight sequences spanning about 14 kb and containing up to ten CDSs from eight different strains. Overall, we observed very high similarity in the CDSs. *yrInv* and an upstream CDS are not present in Big Creek 74 and YRB (Fig. 2B). For *yrIIm*, we aligned seven sequences spanning almost 39 kb, containing up to 12 CDSs from six different *Y. ruckeri* strains. The region showed a high degree of structural variation and a variable number of DNA repeat units coding for the Ig-like domains in all the sequences analysed, regardless of the strain of origin and of the sequencing technology adopted (Fig. 2C). In the PacBio assembly of CSF00782 two copies of *yrIIm* were found, each flanked by transposases. The two copies of *yrIIm* contained a different number of Ig-like domains, 22 and 17, respectively. In stark contrast to the PacBio assembly, the Illumina-based assembly of the CSF00782 genome (Nelson et al., 2015) showed a single copy of *yrIIm* with only three Ig-like domains. The PacBio-based assembly of NVH_1 showed one copy of *yrIIm*, though with a higher number of Ig-like domains (20). Further variability was observed in the PacBio-assembly of Big Creek 74, which possesses 18 Ig-like domains. The Illumina assembly of ATCC29473 showed 7 Ig-like domains, in contrast to that of YRB which lacks *yrIIm* entirely, including the flanking transposases. In the strain 150 only one repeated Ig-like domain was identified in *yrIIm*.

3.4. Presence and absence of the adhesin genes *yrIIm* and *yrInv* in different *Yersinia ruckeri* strains

We wished to investigate the presence of the two inverse autotransporter genes in a collection of available *Y. ruckeri* strains (Table 1). The presence of a full-length *yrIIm* and a full-length *yrInv* was tested by PCR. High quality and purity of genomic DNA was a key point for reliable and repetitive PCR reactions. The genome sequence of NVH_1 was used as a reference sequence for primer design. The

regions used to design the primers were 100% conserved (when compared to all sequenced *Y. ruckeri* strains). *yrInv* is present in all the tested strains, apart from *Y. ruckeri* 1435-94, which did not give a PCR product. We did not detect any differences in *yrInv* size in different *Y. ruckeri* strains, suggesting identical repeat numbers (Fig. 3B). In addition, we obtained similar results when amplifying the *yrInv* translocation domain (β -barrel), which is the most conserved region in inverse autotransporters. All the tested *Y. ruckeri* isolates gave a 0.7 kb product, corresponding to the β -barrel sequence. The only exception was *Y. ruckeri* 1435-94, which did not give any product, confirming the complete absence of the gene (Fig. 3D). We continued to experience some problems with amplifying the *yrIIm* gene, probably due to a combination of gene length and repetitiveness. However, we were able to get a weak but reproducible product for five of the strains (Fig. 3A). Only *Yersinia ruckeri* 1006-94 failed to produce a product. Interestingly, the *yrIIm* PCR products differ in size, ranging from about 6 to 8 kb. We assume that the difference in the length of the *yrIIm* products in different strains is due to a different number of repeated Ig-like domains; similarly dramatic repeat variations have been shown in other classes of autotransporter adhesins, e.g. for *Bartonella* BadA (Riess et al., 2007). This is supported by our analysis of sequenced *Y. ruckeri* genomes, where the number of Ig-like domains varies considerably. The length of the *yrIIm* PCR product of ATCC29473 differs from the length of the *yrIIm* sequence deposited in the NCBI database. This suggests that Illumina assembly has limitations when analysing repetitive gene sequences. With its short read lengths ranging from 100 to 600 bp, this is prone to happen especially with long tandem repeats, as in the case of *yrIIm*. Therefore, we recommend aiming for PacBio sequencing when analysing genes with many internal repeats. Analysis of the *yrIIm* translocation domain by PCR showed no variability in PCR product length, except for *Y. ruckeri* 1006-94, which produced a fragment approximately twice the expected size (Fig. 3C). To determine what causes the discrepancy, we sequenced this PCR product. The resulting sequence corresponded to the gene for glycerol-3-phosphate acetyl-o-transferase (LN681231.1), suggesting that the unexpected PCR band was due to mispriming. Together with the finding that we could not amplify the full-length *yrIIm* from this strain, this suggests that 1006-94 does not contain *yrIIm*. As a positive control to make sure all strains represent *Y. ruckeri*, we amplified and sequenced 16S rDNA from genomic DNA of all *Y. ruckeri* isolates (Fig. 3E).

3.5 Gene expression analysis of *yrIIm* and *yrInv* under different conditions

The virulence of all three human *Yersinia* pathogens relies on number of adhesin molecules expressed on their surface. Some of them, such as *Yersinia* adhesin A (YadA), invasins (InvA), pH6 antigen (PsaA) of *Y. pestis* have been extensively studied (Chauhan et al., 2016). The expression levels of these virulence factors are modulated by environmental factors such as temperature (Isberg et al., 1988), osmolarity, oxygen presence or absence, and by various nutrients (Pepe et al., 1994). In order to investigate the effect of environmental factors on the expression levels of the two inverse autotransporters from *Y. ruckeri*, we performed qRT-PCR analysis. These studies can provide a valuable insight into their role in pathogenicity.

a) Effect of temperature

The expression of the *invA* gene from the human pathogenic *Yersinia* is regulated by temperature (Isberg et al., 1988). Therefore, we analyzed the effect of temperature on *yrInv* and *yrIIm* expression. Three different temperatures were selected for the analysis: 8°C, 28°C, and 37°C. These reflect the temperatures that the fish pathogen would encounter in the environment or in different hosts – which in extreme cases can include zoonotic infections of humans (de Keukeleire et al., 2014). Being psychrophilic, *Y. ruckeri* can easily adapt to and replicate at low temperatures, but grows optimally at 26°C to 28°C. According to our analysis, temperature does have a significant impact on *yrIIm* and *yrInv* expression (Fig. 4A). The transcription of both inverse autotransporters is upregulated at 37°C. However, the effect is stronger for *yrIIm*, with an increase of about three-fold compared with 28°C when we use *gyrA* as the reference gene. The expression level of both genes is unaffected at 8°C.

b) Effect of salt concentration

Like temperature, osmolarity is another environmental factor that can alter the transcription of virulence genes (Brzostek et al., 1999; Porter and Dorman, 1994). Changes in osmolarity affects

virulence in several bacterial pathogens including *Vibrio cholerae* (Miller and Mekalanos, 1988), *Salmonella typhimurium* (Galant and Iii, 1990), *Escherichia coli* (Mouriño et al., 1996), and *Pseudomonas aeruginosa* (Mittal et al., 2009). We analyzed the effect of salt concentration on the expression of *yrIIm* and *yrInv*. Osmolarity has an effect on the transcription of *yrIIm*, whereas expression of *yrInv* at hyperosmotic and hypoosmotic conditions does not show noticeable changes (Fig. 4B). The expression of *yrIIm* is upregulated 2.5 fold at 0 mM NaCl in comparison with 170mM NaCl when we used gyrase as the reference gene.

c) Effect of oxygen

Oxygen availability is another environmental factor that can influence the expression of virulence genes (Marteyn et al., 2010). We cultivated bacteria in LB medium with full aeration or oxygen limitation, provided by an anaerobic jar filled with Anaerocult® system and anaerobic indicator, at 28°C. The growth conditions used in this study aimed to compare the conditions encountered in *Y. ruckeri* life cycle such as in sediments, inside the host, where the conditions are anaerobic. Both genes are upregulated under anaerobic conditions, with *yrIIm* showing the larger effect (Fig. 4C). However, in our experiments, the stability of housekeeping genes (gyrase A, DNA polymerase I) varied between all biological replicates. To gain a full picture of the influence of oxygen on inverse autotransporter gene expression, more reference genes should be included.

d) Effect of the growth medium

The choice of growth medium is an important consideration when performing gene expression experiments; many medium ingredients can influence the expression of virulence genes. As an example, in *Salmonella typhimurium*, 287 genes were upregulated when culturing the bacteria in MOPS-based minimal medium compared to LB medium. The hypothesis is that the minimal medium mimics an environment where the genes in question are expressed in nature (Blair et al., 2013).

In order to understand the effect of the growth medium on the transcription of *yrInv* and *yrIIm* we performed qRT-PCR analysis. We cultured our bacterial cells in 3 different types of media: rich (but undefined) LB medium, DMEM and the minimal medium M9. We reasoned that DMEM would resemble conditions that the pathogen encounters when infecting a fish, whereas M9 would be more similar to the aqueous (sweetwater) environment of fish. Our results showed an increase in the transcription level of both *yrIIm* and *yrInv* in DMEM medium compared to transcription seen in LB medium. Transcription of *yrIIm* showed a 14-fold increase in DMEM, whereas a more dramatic 34-fold increase was seen for the transcription of *yrInv* when we used gyrase as a reference gene. When bacterial cells were grown in M9 medium, we found evidence for a 4-fold overproduction for *yrIIm* compared to LB medium, while *yrInv* expression was not significantly affected by these conditions when compared to gyrase as the reference gene (Fig. 4D).

e) Effect of iron availability

Iron is an essential nutrient and a limiting factor in bacterial infections. Iron influences bacterial metabolism, enzyme activity and host-pathogen interactions (Litwin and Calderwood, 1993; Trivier and Courcol, 1996). Therefore, the acquisition of iron by bacteria cells is important for their pathogenicity, and the absence of free iron is an important signal for the bacteria, suggesting to them that they are inside the host (Messenger and Barclay, 1983). In order to investigate the effect of iron limitation on the transcription of *yrIIm* and *yrInv*, NVH_1 was grown in LB medium supplemented with the iron chelator 2,2'-dipyridyl to ensure a lack of free iron in the medium. As controls in our analysis, we used 2 samples including only LB medium and LB supplemented with 100µM of 2,2'-dipyridyl and 200µM of FeCl₃ in the medium. Iron availability had an effect on *yrInv* expression, where lack of iron showed a marked reduction in mRNA levels (Fig. 4E). For *yrIIm*, the effect was not significant when comparing to both reference genes (Fig. 4E).

4. Discussion

Inverse autotransporters are prominent virulence factors of many organisms, including enterohaemorrhagic *E. coli* and the *Yersiniae* (Leo et al., 2015b). In the human pathogenic *Yersiniae*, InvA has been studied extensively, but several additional inverse autotransporters, including InvC, Iip and InvD/E have been identified (Chauhan et al., 2016; Ginocchio et al., 1992). Based on the available

genome sequences of *Y. ruckeri*, we tried to amplify the genes of interest for structural and functional studies. However, despite a significant effort, our PCR attempts failed. Therefore, the whole genome of the NVH_1 strain was sequenced using PacBio technology. This helped us to identify two inverse autotransporter genes, which we named *yrInv* and *yrIIm*. In addition to *yrInv* and *yrIIm*, we identified a third gene, an invasin-like fragment, and suggest that it is a non-functional inverse autotransporter gene that was disrupted. However, it is worth mentioning that the domain orientation has an opposite topology from the one that we would expect for an invasin homologue. Thus, the two genes could in principle also constitute a variant of a two-partner secretion system; these systems have received much attention in recent years due to their role in virulence of many Gram-negative bacteria (Gal-Mor et al., 2008; Henderson et al., 2000; Leo et al., 2012).

Our study highlights the problems associated with repetitive DNA sequences when using second generation sequencing technologies. Due to their short read length (ranging 100 to 600 bp), second-generation sequencing platforms often fail to assemble genomic regions characterized by a large number of repeat units, thus resulting in incorrectly assembled genomes. Though problematic for any repeated sequences, this is particularly prominent in the case of tandemly repeated protein domains with high sequence similarity, where the length of the repeated DNA sequences exceed that of the second generation sequencing read lengths. Third-generation sequencing platforms, such as the PacBio platform, offer a large improvement over second-generation platforms, including single-molecule templates, faster run time, lower cost per base and last but not least, significantly longer reads. These are key factors in overcoming challenges in assembling highly repetitive DNA regions.

Sequencing the NVH_1 genome using the PacBio platform, we were able to determine the correct number of repeated Ig-like domains in *yrIIm*. The number was verified by amplifying the full-length *yrIIm* gene by PCR, which gave a product size consistent with 20 repeated Ig-like domains.

The number of Ig-like domains in YrIIm differs between different strains. In strains sequenced by third-generation techniques, the number of Ig-like domains varies from 17 to 22. In the strain ATCC29473, the protein is shorter based on the obtained PCR product (6kb). However, according to the NCBI database, the deposited protein sequence of ATCC29473 should give a product corresponding to only 3909 bp. This discrepancy clearly illustrates the problem with using Illumina sequencing for analyzing repetitive sequences, and strongly suggests that multiple repeats are missing in the deposited genome. Moreover, these findings demonstrate that the *yrIIm* locus is genetically unstable, and that the number of Ig repeats can change relatively rapidly, presumably due to recombination between the near identical, repeated Ig-like domain sequences. Changes in protein length due to insertions or deletions of repeated elements has been observed in some autotransporters before, particularly in some trimeric (type Vc) autotransporters such as BadA from *Bartonella henselae* (Riess et al., 2004). The length of YadA from *Y. enterocolitica* also varies from strain to strain; here, the length of YadA correlates with the size of other surface structures, and thus the length variation plays a functional role (Mota et al., 2005). This may also be the case in *Y. ruckeri*.

yrIIm appears to be the product of horizontal gene transfer. The GC content of *yrIIm* is 62.3%, much higher than for the rest of the chromosome (47.6 %) – the insertion is thus readily visible in the map of the NVH_1 GC content (Fig. 1). The gene is flanked on both sides by transposase sequences, indicating that it can be readily mobilized. This is further demonstrated by the fact that some strains (YRB, 1006-94) appear to lack the gene entirely, whereas in CSF002-82, the *yrIIm* gene is duplicated. YrIIm is 59% identical with an inverse autotransporter of *Enterobacter cloacae* (EbIIm), an important nosocomial human pathogen. *E. cloacae* belongs to the *Enterobacteriaceae* family with the overall 54.79% of GC content of the chromosome (Ren et al., 2010). We assume that YrIIm and EbIIm are orthologous and hypothesize that *E. cloacae* could be the donor strain. The acquisition of *yrIIm* thus appears to be recent, and lack of selection pressure allows the locus to remain in flux, leading to variation in both the length of the protein and its presence in different strains.

In contrast, the *yrInv* locus seems to be more stable. There are no signs of recent genome changes at this locus, and most of the surrounding genes have housekeeping functions (Fig. 2B). Furthermore, in all the sequenced strains that contain *yrInv*, the structure of the gene is the same. This suggests that (if present) *yrInv* is under selection pressure. However, as this gene is absent from some strains, it is presumably not an essential virulence factor. Most described inverse autotransporters are virulence factors. A very similar gene to *yrInv* is present in *E. piscicida*. YrInv shares 43% identity over the whole sequence with invasin from *E. piscicida*, slightly more than with invasin from the human pathogen *Y.*

pseudotuberculosis (40% identity). As both *E. piscicida* and *Y. ruckeri* infect fish, presumably these orthologous proteins play a similar role in the infection process of both pathogens.

Interestingly, both YrIIm and YrInv are upregulated in DMEM medium. We assume this cell culture medium represents an environment similar to what *Y. ruckeri* would encounter in the fish host. As both proteins are presumed adhesins or invasins, they may play an important role in colonization of fish tissues. We are currently investigating the role that these proteins play in *Y. ruckeri* virulence.

Acknowledgements

We would like to thank the qRT-PCR facility at the Department of Biosciences, UiO, especially Dr. Mads Bengtsen for helping in the execution and analysis of the qRT-PCR experiments. The sequencing service which was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities. We wish to thank Prof Duncan Colquhoun (Norwegian Veterinary Institute) and Prof Mikael Skurnik (University of Helsinki) for providing *Y. ruckeri* strains. We also wish to thank Prof Henning Sørum (Norwegian University of Life Sciences) for his assistance with *Y. ruckeri* isolates.

Funding sources

This work by supported by Research Council of Norway Young Researcher grant 249793 (to JCL), and by Research Council of Norway FriMedBio grant 240483 (to DL).

Figure captions.

Fig. 1. Circular representation of the main features of the chromosomal genome of *Y. ruckeri* NVH_1 sequenced in this study. From the inner ring, each track represents 1) GC-skew, 2) GC content, 3) RNA, mobile elements and phage features as reported in colour legend, 4) DNA repeats content percentage, 5-6-7) Mauve locally collinear blocks (LCBs) shared with other *Y. ruckeri* strains as reported in colour legend, 8-9) Coding sequences in the forward and reverse strand, 10) other features related to flagella (yellow), secretion systems (red), pili and fimbriae (green), and other virulence factors (olive green).

Fig. 2. a) Domain structure of YrIIm and YrInv of *Y. ruckeri* NVH_1. All domains of YrIIm and YrInv are depicted by different colouring. The N-terminal signal peptide is shown in purple. The periplasmic region is shown in blue, the translocation domain is shown in yellow and the passenger or extracellular domain is shown in different shadows of red. The passenger consists of an array of tandemly repeated Ig-like domains (in red) or with Ig-like domains of low pairwise sequence similarity (in orange). The passenger is usually capped by a C-type lectin domain (CTLD, in green). The numbers indicated in the figure show the amino acid positions in the protein sequence. **b)** DNA alignment of a ~14 kb-long DNA region containing *yrInv* and flanking CDS in *Y. ruckeri* strains sequenced and deposited in Genbank (Table 2). Each CDS is indicated by a yellow arrow, with the percentage of sequence identity to NVH_1 reported inside the arrow. Regions corresponding to the structural domains of YrInv reported as arrows coloured as in a). **c)** DNA alignment of ~39 kb-long DNA region containing *yrIIm* and flanking CDS in *Y. ruckeri* strains sequenced and deposited in Genbank. Each CDS is indicated by a yellow arrow, with the percentage of sequence identity to CSF007-82 reported inside the arrow. Regions corresponding to the predicted structural domains in YrIIm are reported as in a). The dashed lines indicate gaps in the DNA alignment. The asterisk (*) in b) and c) indicates assemblies generated through PacBio SMRT sequencing. In strain 150 the grey box indicates a contig break in the assembly. **d)** Schematic representation of the DNA region in *Y. ruckeri* YRB containing an invasin-like fragment. Each CDS as predicted in RAST is indicated by a yellow arrow. Region corresponding to the predicted structural domains in invasin-like fragment are reported as in a).

Fig. 3. Analysis of presence and length of *yrIIm* and *yrInv* in different *Yersinia ruckeri* strains. PCR products of full-length *yrIIm* (A), full-length *yrInv* (B), translocation domains (C - *yrIIm*, D - *yrInv*) and 16S rRNA (E). All *Y.a ruckeri* isolates harbor a full-length *yrIIm* except *Y. ruckeri* 1006-94 (A - Note

that the length of the PCR bands differ in size). All *Y. ruckeri* strains harbor *yrInv* with the exception of *Y. ruckeri* 1435-94 (B – with no observable size differences). Detection of the β -barrel of *yrIIm* (735bp) (C) and the β -barrel of *yrInv* (734 bp) (D) confirms the presence or absence of the full-length genes.

Fig. 4. qRT-PCR analysis of *yrInv* and *yrIIm*

Effect of temperature (A), salt concentration (B), oxygen availability (C), medium selection (D) and iron availability (E) on the transcription of *yrInv* and *yrIIm* compared to gyrase A and DNA polymerase I. Histograms shows the relative expression levels represented as fold change. Relative transcription was evaluated by using 28°C (A), 170 mM NaCl (B), aerobic conditions (C), LB medium (D) as the reference. Columns show the data calculated from the average of three biological replicates, each performed in triplicate with the exception of panel c), where five biological replicates were used for calculations. Error bars indicate the standard error of the mean.

6. References

- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi:10.1186/1471-2164-9-75
- Bastardo, A., Ravelo, C., Romalde, J.L., 2012. Multilocus sequence typing reveals high genetic diversity and epidemic population structure for the fish pathogen *Yersinia ruckeri* 14, 1888–1897. doi:10.1111/j.1462-2920.2012.02735.x
- Benson, G., 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi:10.1093/nar/27.2.573
- Bernstein, H.D., 2014. NIH Public Access 4, 1–30. doi:10.1128/ecosalplus.4.3.6.Type
- Bertani, G., 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* 62, 293–300. doi:citeulike-article-id:149214
- Berven, F.S., Flikka, K., Jensen, H.B., Eidhammer, I., 2004. BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res.* 32, W394-9. doi:10.1093/nar/gkh351
- Blair, J.M.A., Richmond, G.E., Bailey, A.M., Ivens, A., Piddock, L.J. V., 2013. Choice of Bacterial Growth Medium Alters the Transcriptome and Phenotype of *Salmonella enterica* Serovar Typhimurium. *PLoS One* 8. doi:10.1371/journal.pone.0063912
- Brudal, E., Winther-Larsen, H.C., Colquhoun, D.J., Duodu, S., 2013. Evaluation of reference genes for reverse transcription quantitative PCR analyses of fish-pathogenic *Francisella* strains exposed to different growth conditions. *BMC Res. Notes* 6, 76. doi:10.1186/1756-0500-6-76
- Brzostek, K., Stencel, I., Hrebenda, J., 1999. Osmoregulation-dependent expression of *Yersinia enterocolitica* virulence factors. *Acta Microbiol Pol* 48, 31–37.
- Bullock, G.L., 1984. Enteric Redmouth Disease of Salmonids. *US Fish Wildl. Publ.*
- Busch, R.A., 1978. Enteric Red Mouth Disease. *Mar. Fish. Rev.* 40, 42–51.
- Chauhan, N., Wrobel, A., Skurnik, M., Leo, J.C., 2016. *Yersinia* adhesins: An arsenal for infection. *Proteomics - Clin. Appl.* doi:10.1002/prca.201600012
- Darling, A.E., Mau, B., Perna, N.T., 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5. doi:10.1371/journal.pone.0011147
- de Keukeleire, S., de Bel, a., Jansen, Y., Janssens, M., Wauters, G., Piérard, D., 2014. *Yersinia ruckeri*, an unusual microorganism isolated from a human wound infection. *New Microbes New Infect.* 2, 134–135. doi:10.1002/nmi2.56
- Fairman, J.W., Dautin, N., Wojtowicz, D., Liu, W., Noinaj, N., Barnard, T.J., Udho, E., Przytycka,

- T.M., Cherezov, V., Buchanan, S.K., 2012. Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic *E. coli* and enteropathogenic *Y. pseudotuberculosis*. *Structure* 20, 1233–43. doi:10.1016/j.str.2012.04.011
- Fernandez, L., Fernandez, L., Lopez, J.R., Lopez, J.R., Secades, P., Secades, P., Menendez, a, Menendez, a, Marquez, I., Marquez, I., Guijarro, J. a, Guijarro, J. a, 2003. In Vitro and In Vivo Studies of the Yrp1 Protease from. *Society* 69, 7328–7335. doi:10.1128/AEM.69.12.7328
- Fernández, L., Prieto, M., Guijarro, J.A., 2007. The iron- and temperature-regulated haemolysin YhIA is a virulence factor of *Yersinia ruckeri*. *Microbiology* 153, 483–489. doi:10.1099/mic.0.29284-0
- Furones, M.D., Rodgers, C.J., Munn, C.B., 1993. *Yersinia ruckeri*, the causal agent of enteric redmouth disease (ERM) in fish. *Annu. Rev. Fish Dis.* 3, 105–125. doi:10.1016/0959-8030(93)90031-6
- Gal-Mor, O., Gibson, D.L., Baluta, D., Vallance, B.A., Finlay, B.B., 2008. A novel secretion pathway of *Salmonella enterica* acts as an antivirulence modulator during salmonellosis. *PLoS Pathog.* 4. doi:10.1371/journal.ppat.1000036
- Galant, J.E., Iii, R.O.Y.C., 1990. Expression of *Salmonella typhimurium* genes Expression of *Salmonella typhimurium* Genes Required for Invasion Is Regulated by Changes in DNA Supercoiling 58, 1879–1885.
- Garcia, J.A., Dominguez, L., Larsen, J.L., Pedersen, K., 1998. Ribotyping and plasmid profiling of *Yersinia ruckeri*. *J. Appl. mMicrobiology* 85, 949–955.
- Ginocchio, C., Pace, J., Galan, J.E., 1992. Identification and molecular characterization of a *Salmonella typhimurium* gene involved in triggering the internalization of salmonellae into cultured epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5976–5980. doi:10.1073/pnas.89.13.5976
- Guilvout, I., Quilici, M.L., Rabot, S., Lesel, R., Mazigh, D., 1988. BamHI restriction endonuclease analysis of *Yersinia ruckeri* plasmids and their relatedness to the genus *Yersinia* 42- to 47-megadalton plasmid. *Appl. Environ. Microbiol.* 54, 2594–2597.
- Hamburger, Z.A., Brown, M.S., Isberg, R.R., Bjorkman, P.J., 1999. Integrin-Binding Protein Crystal Structure of Invasin: A Bacterial Crystal Structure of Invasin: A Bacterial Integrin-Binding Protein. *Science* (80-). 286, 291–295. doi:10.1126/science.286.5438.291
- Heinz, E., Stubenrauch, C.J., Grinter, R., Croft, N.P., Purcell, A.W., Strugnell, R. a., Dougan, G., Lithgow, T., 2016. Conserved features in the structure, mechanism and biogenesis of the inverse autotransporter protein family. *Genome Biol. Evol.* 8, evw112. doi:10.1093/gbe/evw112
- Henderson, I.R., Cappello, R., Nataro, J.P., 2000. Autotransporter proteins, evolution and redefining protein secretion: Response. *Trends Microbiol.* 8, 534–535. doi:10.1016/S0966-842X(00)01884-9
- <http://www.repeatmasker.org> [WWW Document], n.d. URL <http://www.repeatmasker.org>
- Isberg, R.R., Leong, J.M., 1990. Multiple β 1 chain integrins are receptors for invasin, a protein that promotes bacterial penetration into mammalian cells. *Cell* 60, 861–871. doi:10.1016/0092-8674(90)90099-Z
- Isberg, R.R., Swain, A., Falkow, S., 1988. Analysis of expression and thermoregulation of the *Yersinia pseudotuberculosis* *inv* gene with hybrid proteins. *Infect. Immun.* 56, 2133–2138.
- Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338, 1027–1036. doi:10.1016/j.jmb.2004.03.016
- Kawula, T.H., Lelivelt, M.J., Orndorff, P.E., 1996. Using a new inbred fish model and cultured fish tissue cells to study *Aeromonas hydrophila* and *Yersinia ruckeri* pathogenesis. *Microb. Pathog.* 20, 119–25. doi:10.1006/mpat.1996.0011
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A.A., Heled, J.,

2012. Geneious. Bioinformatics. doi:10.1093/bioinformatics/bts199

- LeJeune, J.T., Rurangirwa, F.R., 2000. Polymerase chain reaction for definitive identification of *Yersinia ruckeri*. J. Vet. Diagn. Invest. 12, 558–61. doi:10.1177/104063870001200611
- Leo, J.C., Grin, I., Linke, D., 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367, 1088–101. doi:10.1098/rstb.2011.0208
- Leo, J.C., Oberhettinger, P., Chaubey, M., Schütz, M., Kühner, D., Bertsche, U., Schwarz, H., Götz, F., Autenrieth, I.B., Coles, M., Linke, D., 2015a. The Intimin periplasmic domain mediates dimerisation and binding to peptidoglycan. Mol. Microbiol. 95, 80–100. doi:10.1111/mmi.12840
- Leo, J.C., Oberhettinger, P., Schütz, M., Linke, D., 2015b. The inverse autotransporter family: Intimin, invasins and related proteins. Int. J. Med. Microbiol. doi:10.1016/j.ijmm.2014.12.011
- Litwin, C.M., Calderwood, S.B., 1993. Role of iron in regulation of virulence genes. Clin. Microbiol. Rev. 6, 137–149. doi:10.1128/CMR.6.2.137.Updated
- Mammeri, H., Poirel, L., Nazik, H., Nordmann, P., 2006. Cloning and functional characterization of the ambler class C beta-lactamase of *Yersinia ruckeri*. FEMS Microbiol. Lett. 257, 57–62. doi:10.1111/j.1574-6968.2006.00148.x
- Marteyn, B., West, N.P., Browning, D.F., Cole, J. a, Shaw, J.G., Palm, F., Mounier, J., Prévost, M.-C., Sansonetti, P., Tang, C.M., 2010. Modulation of *Shigella* virulence in response to available oxygen in vivo. Nature 465, 355–358. doi:10.1038/nature08970
- Messenger, A.J., Barclay, R., 1983. Bacteria, iron and pathogenicity. Biochem. Educ. 11, 54–63. doi:10.1016/0307-4412(83)90043-2
- Miller, V.L., Mekalanos, J.J., 1988. A novel suicide vector and its use in construction of insertion mutations: Osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J. Bacteriol. 170, 2575–2583.
- Mittal, R., Sharma, S., Chhibber, S., Harjai, K., 2009. Effect of Osmolarity on Virulence of Uropathogenic *Pseudomonas Aeruginosa*. Am. J. Biomed. Sci. 1, 12–26. doi:10.5099/aj090100012
- Mota, L.J., Journet, L., Sorg, I., Agrain, C., Cornelis, G.R., 2005. Bacterial injectisomes: needle length does matter. Science 307, 1278. doi:10.1126/science.1107679
- Mouriño, M., Madrid, C., Balsalobre, C., Prenafeta, A., Muñoa, F., Blanco, J., Blanco, M., Blanco, J.E., Juárez, A., 1996. The Hha protein as a modulator of expression of virulence factors in *Escherichia coli*. Infect. Immun. 64, 2881–2884.
- Nelson, M.C., Lapatra, S.E., Welch, T.J., 2015. Complete genome sequence of *Yersinia ruckeri* strain CSF007-82, etiologic agent of red mouth disease in salmonid fish. Genome Announc. 3, 4–5. doi:10.1128/genomeA.01491-14.Copyright
- Oberhettinger, P., Leo, J.C., Linke, D., Autenrieth, I.B., Schütz, M.S., 2015. The inverse autotransporter intimin exports its passenger domain via a hairpin intermediate. J. Biol. Chem. 290, 1837–1849. doi:10.1074/jbc.M114.604769
- Oberhettinger, P., Schütz, M., Leo, J.C., Heinz, N., Berger, J., Autenrieth, I.B., Linke, D., 2012. Intimin and Invasin Export Their C-Terminus to the Bacterial Cell Surface Using an Inverse Mechanism Compared to Classical Autotransport. PLoS One 7. doi:10.1371/journal.pone.0047069
- Ohtani, M., Villumsen, K.R., Strøm, H.K., Raida, M.K., 2014. 3D visualization of the initial *Yersinia ruckeri* infection route in rainbow trout (*Oncorhynchus mykiss*) by optical projection tomography. PLoS One 9. doi:10.1371/journal.pone.0089672
- Pepe, J.C., Badger, J.L., Miller, V.L., 1994. Growth phase and low pH affect the thermal regulation of the *Yersinia enterocolitica* *inv* gene. Mol. Microbiol. 11, 123–135. doi:10.1111/j.1365-2958.1994.tb00295.x

- Pisano, F., Kochut, A., Uliczka, F., Geyer, R., Stolz, T., Thiermann, T., Rohde, M., Dersch, P., 2012. In vivo-induced inva-like autotransporters Ipf and invc of *Yersinia pseudotuberculosis* promote interactions with intestinal epithelial cells and contribute to virulence. *Infect. Immun.* 80, 1050–1064. doi:10.1128/IAI.05715-11
- Porter, M.E., Dorman, C.J., 1994. A role for H-NS in the thermo-osmotic regulation of virulence gene expression in *Shigella flexneri*. *J. Bacteriol.*
- Price, A.L., Jones, N.C., Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21. doi:10.1093/bioinformatics/bti1018
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., 2005. InterProScan: Protein domains identifier. *Nucleic Acids Res.* 33. doi:10.1093/nar/gki442
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033
- Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M.O., Nielsen, H.B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M.V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Ponten, T., Foley, R.A., Lahr, M.M., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell* 163, 571–582. doi:10.1016/j.cell.2015.10.009
- Ren, Y., Ren, Y., Zhou, Z., Guo, X., Li, Y., Feng, L., Wang, L., 2010. Complete genome sequence of *Enterobacter cloacae* subsp. *cloacae* type strain ATCC 13047. *J. Bacteriol.* 192, 2463–2464. doi:10.1128/JB.00067-10
- Reuter, S., Connor, T.R., Barquist, L., Walker, D., Feltwell, T., Harris, S.R., Fookes, M., Hall, M.E., Petty, N.K., Fuchs, T.M., Corander, J., Dufour, M., Ringwood, T., Savin, C., Bouchier, C., Martin, L., Miettinen, M., Shubin, M., Riehm, J.M., Laukkanen-Ninios, R., Sihvonen, L.M., Siitonen, A., Skurnik, M., Falcão, J.P., Fukushima, H., Scholz, H.C., Prentice, M.B., Wren, B.W., Parkhill, J., Carniel, E., Achtman, M., McNally, A., Thomson, N.R., 2014. Parallel independent evolution of pathogenicity within the genus *Yersinia*. *Proc. Natl. Acad. Sci.* 111, 6768–6773. doi:10.1073/pnas.1317161111
- Riess, T., Andersson, S.G.E., Lupas, A., Schaller, M., Schäfer, A., Kyme, P., Martin, J., Wälzlein, J.-H., Eehalt, U., Lindroos, H., Schirle, M., Nordheim, A., Autenrieth, I.B., Kempf, V. a J., 2004. Bartonella adhesin a mediates a proangiogenic host cell response. *J. Exp. Med.* 200, 1267–1278. doi:10.1084/jem.20040500
- Riess, T., Raddatz, G., Linke, D., Schäfer, A., Kempf, V.A.J., 2007. Analysis of Bartonella adhesin A expression reveals differences between various *B. henselae* strains. *Infect. Immun.* 75, 35–43. doi:10.1128/IAI.00963-06
- Rocha, D.J.P., Santos, C.S., Pacheco, L.G.C., 2015. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie Van Leeuwenhoek* 108, 685–93. doi:10.1007/s10482-015-0524-1
- Romalde, J.L., Magarinos, B., Barja, J.L., Toranzo, A.E., 1993. Antigenic and molecular characterization of *Yersinia ruckeri* proposal for a new intraspecies classification. *Syst. Appl. Microbiol.* 16.
- Romalde, J.L., Toranzo, A.E., 1993. Pathological activities of *Yersinia ruckeri*, the enteric redmouth (ERM) bacterium. *FEMS Microbiol. Lett.* 112, 291–9.
- Ross, A.J., Rucker, R.R., Ewing, W.H., 1966. Description of a Bacterium Associated with Redmouth Disease of Rainbow Trout (*Salmo gairdneri*). *Can. J. Microbiol.* 12, 763–770. doi:10.1139/m66-103
- Rucker R, 1966. Redmouth disease of rainbow trout (*Salmo gairdneri*). *Bull Off Int Epizoot.* 825–30.
- Sadana, P., Mönnich, M., Unverzagt, C., Scrima, A., 2017. Structure of the *Y. pseudotuberculosis* adhesin InvasinE. *Protein Sci.* 0. doi:10.1002/pro.3171

- Savin, C., Martin, L., Bouchier, C., Filali, S., Chenau, J.Ô., Zhou, Z., Becher, F., Fukushima, H., Thomson, N.R., Scholz, H.C., Carniel, E., 2014. The *Yersinia pseudotuberculosis* complex: Characterization and delineation of a new species, *Yersinia wautersii*. *Int. J. Med. Microbiol.* 304, 452–463. doi:10.1016/j.ijmm.2014.02.002
- Schalk-hihi, C., Schubert, C., Alexander, R., Bayoumy, S., Jose, C., Deckman, I., Desjarlais, R.L., Dzordzorme, K.C., Christopher, M., 2011. Crystal Structure of a Soluble Form of Human Monoglyceride Lipase in Complex with an Inhibitor at 1 . 35 Å Resolution Keywords. *Protein Sci.* 22, 1–50. doi:10.1002/pro.
- Secades, P., Guijarro, J. a., 1999. Purification and Characterization of an Extracellular Protease from the Fish pathogen *Yersinia ruckeri* and effect of culture condition on production. *Appl. Environ. Microbiol.* 65, 3969–3975.
- Shaowu, L., Di, W., Hongbai, L., Tongyan, L., 2013. Isolation of *Yersinia ruckeri* strain H01 from farm-raised Amur Sturgeon *Acipenser schrencki* in China. *J. Aquat. Anim. Health* 25, 9–14. doi:10.1080/08997659.2012.728169
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75
- Simonet, M., Falkow, S., 1992. Invasin expression in *Yersinia pseudotuberculosis*. *Infect. Immun.*
- Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33. doi:10.1093/nar/gki408
- Stevenson, R.M.W., Daly, J.G., 1982. BIOCHEMICAL AND SEROLOGICAL CHARACTERISTICS OF ONTARIO ISOLATES OF *YERSINIA-RUCKERI*. *Can. J. Fish. Aquat. Sci.* Can. J. Fish. Aquat. Sci, Can J Fish, Can J Fish Aquat Sci, Can J Fish. Aquat Sci, J. Can. Des Sci. Halieutiques Aquat. Can. J. Fish. Aquat. Sci. 39, 870–876. doi:10.1139/f82-118
- Strong, P.C.R., Hinchliffe, S.J., Patrick, H., Atkinson, S., Champion, O.L., Wren, B.W., 2011. Identification and characterisation of a novel adhesin Ifp in *Yersinia pseudotuberculosis*. *BMC Microbiol.* 11, 85. doi:10.1186/1471-2180-11-85
- Sun, Z., Harris, H.M.B., McCann, A., Guo, C., Argimón, S., Zhang, W., Yang, X., Jeffery, I.B., Cooney, J.C., Kagawa, T.F., Liu, W., Song, Y., Salvetti, E., Wrobel, A., Rasinkangas, P., Parkhill, J., Rea, M.C., O’Sullivan, O., Ritari, J., Douillard, F.P., Paul Ross, R., Yang, R., Briner, A.E., Felis, G.E., de Vos, W.M., Barrangou, R., Klaenhammer, T.R., Caufield, P.W., Cui, Y., Zhang, H., O’Toole, P.W., 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat. Commun.* 6, 8322. doi:10.1038/ncomms9322
- Szabady, R.L., Peterson, J.H., Skillman, K.M., Bernstein, H.D., 2005. An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proc. Natl. Acad. Sci. U. S. A.* 102, 221–226. doi:10.1073/pnas.0406055102
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi:10.1093/nar/22.22.4673
- Trivier, D., Courcol, R.J., 1996. Iron depletion and virulence in *Staphylococcus aureus*. *FEMS Microbiol. Lett.* doi:10.1016/0378-1097(96)00233-9
- Tsai, J.C., Yen, M.R., Castillo, R., Leyton, D.L., Henderson, I.R., Saier, M.H., 2010. The bacterial intimins and invasins: A large and novel family of secreted proteins. *PLoS One* 5. doi:10.1371/journal.pone.0014403
- Wasteson U, Hvaal AB, Sørnum H, Myhr E, F.K., 1989. Antibacterial Spectrum and some other Characteristics of an Antimicrobial Factor produced by *Yersinia ruckeri*. *Acta vet. scand.* 30,

Figure 1

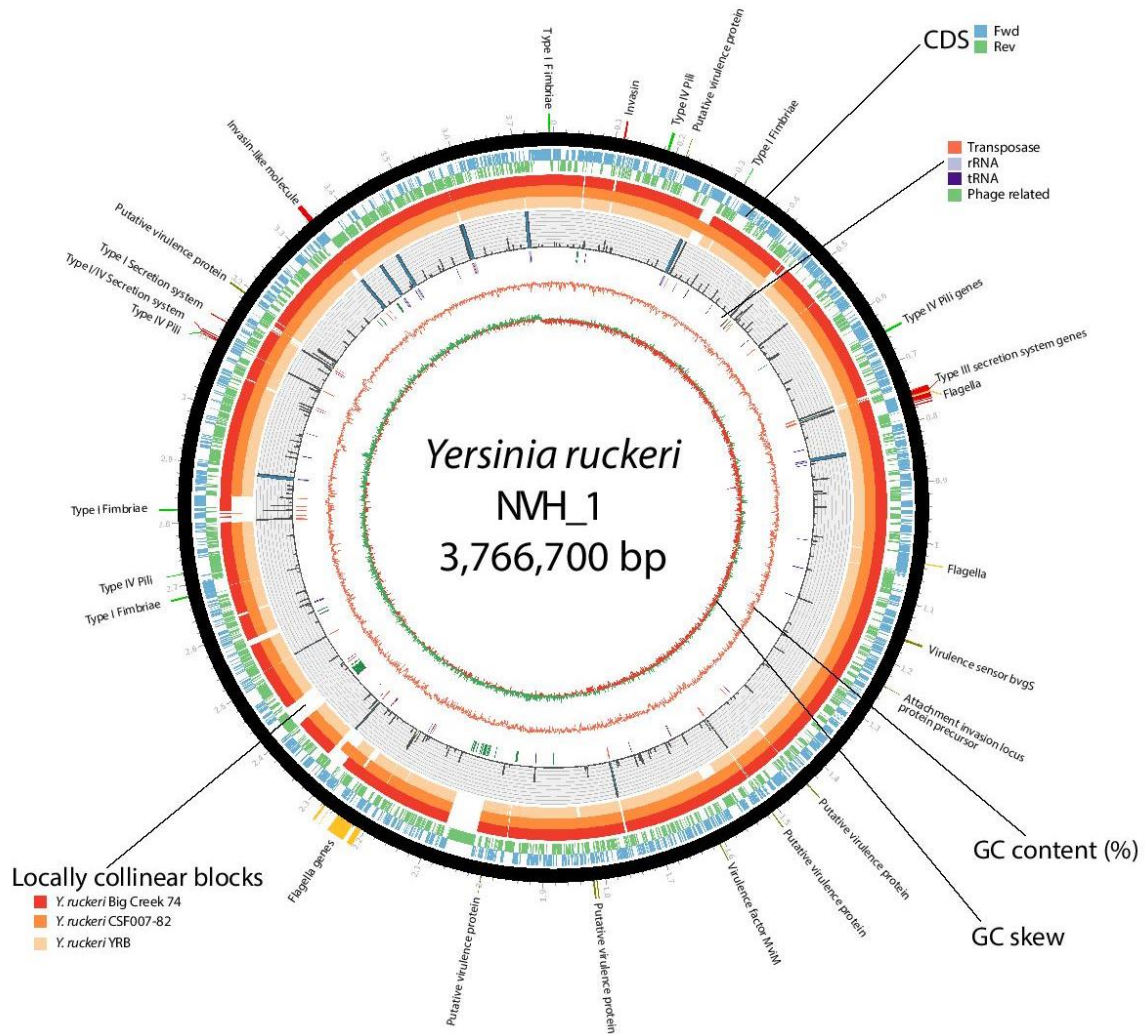
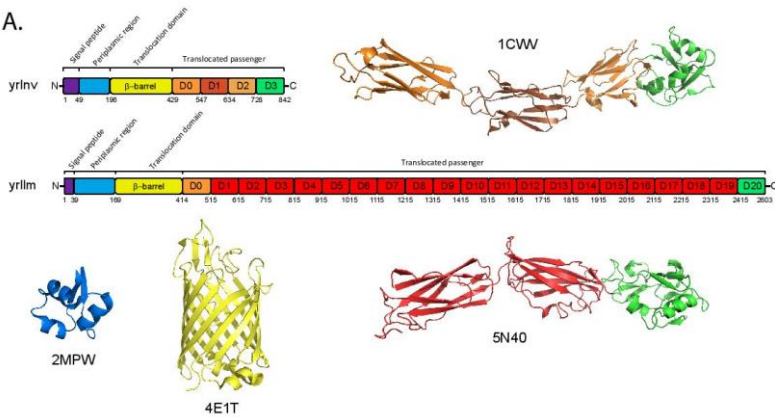
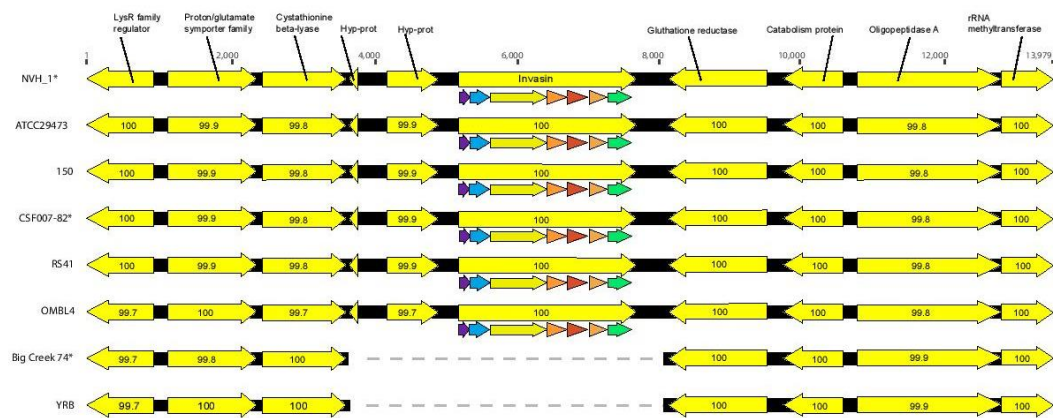


Figure 2

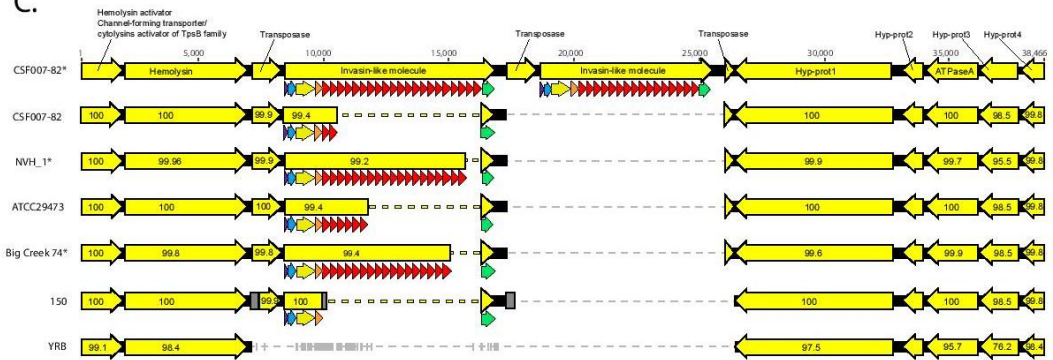
A.



B.



C.



D.

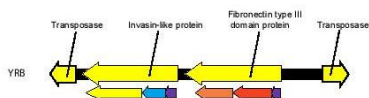


Figure 3

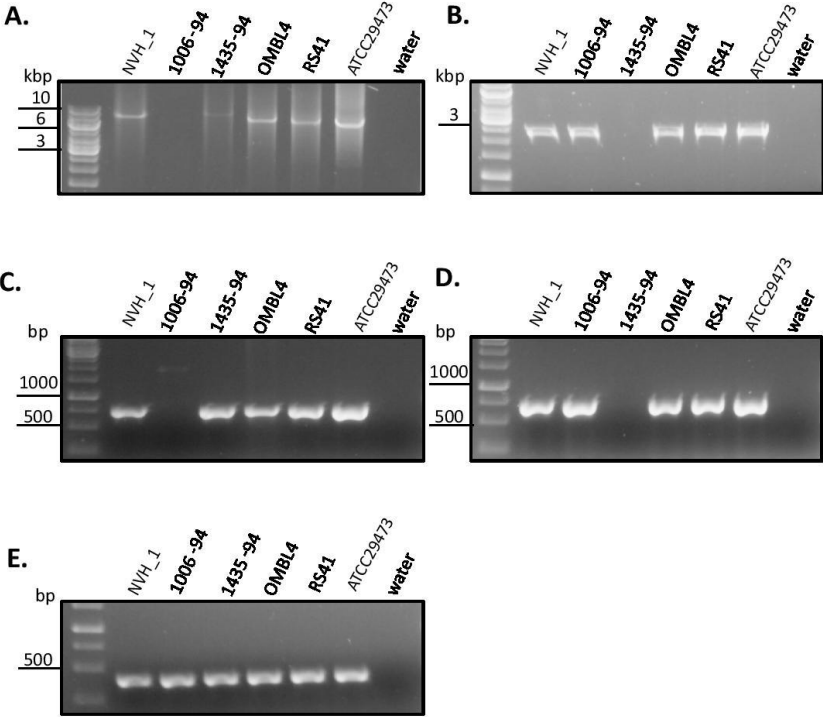
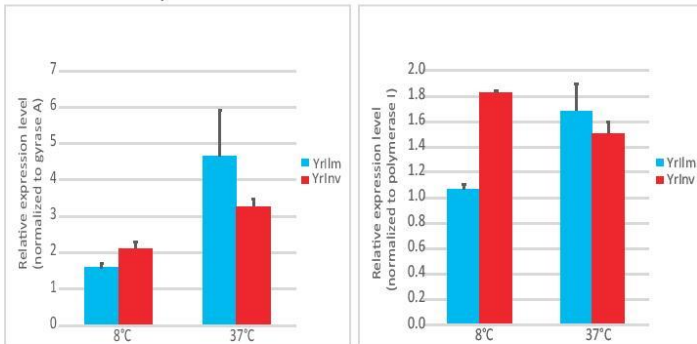
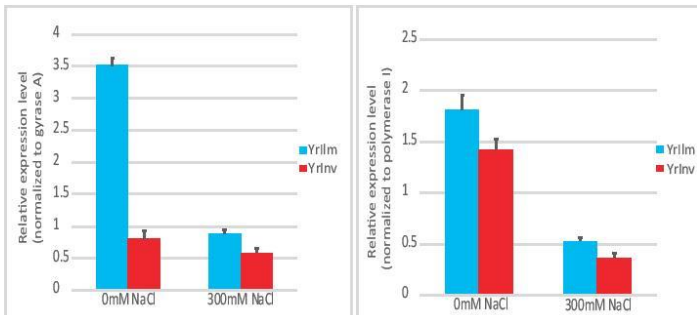


Figure 4

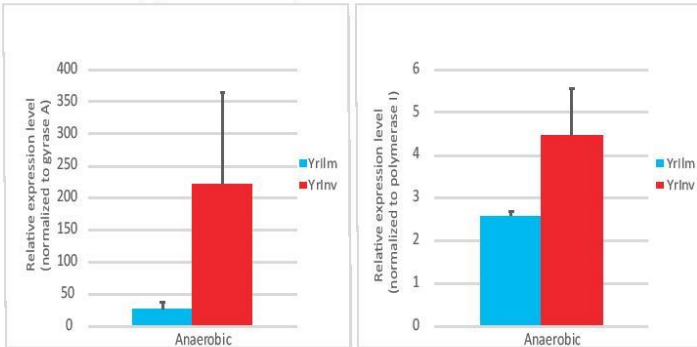
A. Effect of temperature



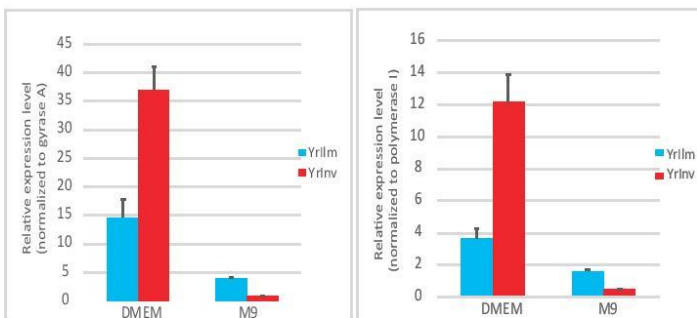
B. Effect of salt concentration



C. Effect of oxygen availability



D. Effect of medium selection



E. Effect of iron availability

