Generation of a CRISPR database for Yersinia pseudotuberculosis complex and role of CRISPR-based immunity in conjugation.

Katja A Koskela, Laura Mattinen, Laura Kalin-Mänttäri, Gilles Vergnaud, Olivier Gorgé, Simo Nikkari, Mikael Skurnik

environmental microbiology
environmental microbiology reports

Generation of a CRISPR database for Yersinia pseudotuberculosis complex and role of CRISPR based immunity in conjugation

Journal:	*Environmental Microbiology and Environmental Microbiology Reports*
Manuscript ID:	EMI-2015-0152
Manuscript Type:	EMI - Research article
Journal:	Environmental Microbiology
Date Submitted by the Author:	30-Jan-2015

Complete List of Authors:	Koskela, Katja; Centre for Military Medicine, CB Defence and Environmental Health Centre Mattinen, Laura; University of Helsinki, Department of Agricultural Sciences Nikkari, Simo; Centre for Military Medicine, CB Defence and Environmental Health Centre Kalin-Mänttäri, Laura; Centre for Military Medicine, CB Defence and Environmental Health Centre Vergnaud, Gilles; Univ Paris-Sud, Institut de Génétique et Microbiologie Gorgé, Olivier; Univ Paris-Sud, Institut de Génétique et Microbiologie Skurnik, Mikael; University of Helsinki Haartman Institute, Department of Bacteriology and Immunology

Keywords:	bacteria, genomics/functional genomics/comparative genomics, microbe:microbe interactions, microbial genetics

Generation of a CRISPR database for *Yersinia pseudotuberculosis* complex and role of CRISPR based immunity in conjugation

Katja A Koskela[1], Laura Mattinen[2], Laura Kalin-Mänttäri[1,2],Gilles Vergnaud[3,4,5], Olivier Gorgé[3,4,6], Simo Nikkari[1], and Mikael Skurnik[2,7]*

Centre for Military Medicine, Helsinki, Finland,[1] Department of Bacteriology and Immunology, Haartman Institute and Research Programs Unit, Immunobiology, University of Helsinki,[2] Univ Paris-Sud, Institut de Génétique et Microbiologie, UMR8621, Orsay, France,[3] CNRS, Orsay, France, [4] ENSTA ParisTech, Palaiseau, France,[5] DGA/MNRBC, Vert le Petit, France[6] and Helsinki University Central Hospital Laboratory Diagnostics, Helsinki, Finland[7]

***Address correspondence to:** Mikael Skurnik, Haartman Institute, PO Box 21, 00014 University of Helsinki, Finland; Phone: +358-9-191 26464; Fax: +358-9-191 26382; E-mail: mikael.skurnik@helsinki.fi

Abstract

The CRISPR/Cas system is used by bacteria and archaea against invading conjugative plasmids or bacteriophages. Central to this immunity system are genomic CRISPR loci that contain fragments of invading DNA. These are maintained as spacers in the CRISPR loci between direct repeats and the spacer composition in any bacterium reflects its evolutionary history. We analyzed the CRISPR locus sequences of 335 *Yersinia pseudotuberculosis* complex strains. Altogether 1902 different spacer sequences were identified and these were used to generate a database for the spacer sequences. Only ~10 % of the spacer sequences found matching sequences. In addition, surprisingly few spacers were shared by *Y. pestis* and *Y. pseudotuberculosis* strains. Interestingly 32 different protospacers were present in the conjugative plasmid pYptb32953. The corresponding spacers were identified from 35 different *Y. pseudotuberculosis* strains indicating that these strains had encountered pYptb32953 earlier. In conjugation experiments pYptb32953-specific spacers generally prevented conjugation with spacer-positive and spacer-free strains. However, some strains with one to four spacers were invaded by pYptb32953 and some spacer-free strains were fully resistant. Also some spacer-positive strains were intermediate resistant to conjugation. This suggests that one or more other defense systems are determining conjugation efficiency independent of the CRISPR/Cas system.

Introduction

Yersinia pseudotuberculosis is a gram-negative bacterium which causes disease in humans and animals. In humans *Y. pseudotuberculosis* is a cause of food-borne associated illness with symptoms of fever and abdominal pain, and sometimes diarrhea. In animals, it causes tuberculosis-like disease (Naktin & Beavis, 1999, Aleksic *et al.*, 1995, Tauxe, 2004). *Yersinia pestis*, the bacterium responsible for plague, evolved from its *Y. pseudotuberculosis* ancestor approximately 1,500-6,400 years ago (Morelli *et al.*, 2010, Achtman *et al.*, 1999, Cui *et al.*, 2013, Harbeck *et al.*, 2013). In a multilocus sequence typing (MLST) study *Y. pseudotuberculosis, Y. pestis* (representing a single ST), the recently described *Yersinia similis* (Sprague *et al.*, 2008) and a number of distinct strains, called the Korean group and recently named as *Yersinia wautersii* (Savin *et al.*, 2014), were collectively named as a *Y. pseudotuberculosis* complex (Laukkanen-Ninios *et al.*, 2011). Due to their close evolutionary

relationship, *Y. pseudotuberculosis* and *Y. pestis* are very similar and share ≥97% nucleotide sequence identity for most of the chromosomal genes depending on the *Y. pseudotuberculosis* strain in question. *Y. pseudotuberculosis* is commonly typed serologically based on the lipopolysaccharide O antigen. Some of the 15 known serotypes are divided into subtypes (O:1a, O:1b, O:1c, O:2a, O:2b, O:2c, O:4a, O:4b, O:5a, O:5b) resulting in a total of 21 serotypes (Bogdanovich *et al.*, 2003). *Y. pestis* does not express O-antigen due to pseudogenes in the O antigen biosynthetic genes, however, comparison of the *Y. pestis* O antigen gene cluster sequence with those of different *Y. pseudotuberculosis* serotype gene clusters suggested that *Y. pestis* evolved from a *Y. pseudotuberculosis* serotype O:1b strain (Skurnik *et al.*, 2000).

Therefore, differentiating and typing of these two species has been challenging (Chauvaux *et al.*, 2011). For instance, an earlier study has suggested ribotyping as one potential typing method, but even with this method differentiation was not accurate (Voskressenskaya *et al.*, 2005). Additionally, multilocus sequence typing (MLST) (Laukkanen-Ninios et al., 2011, Ch'ng *et al.*, 2011), 16S rRNA gene sequencing, and pulsed-field gel electrophoresis (Souza *et al.*, 2010) have been used for typing of *Y. pseudotuberculosis*. Some of these methods can identify and differentiate *Yersinia* species, but still typing of *Y. pseudotuberculosis* is challenging.

The CRISPR-Cas (Clustered regularly interspaced short palindromic repeat – CRISPR associated genes) system is a RNA-based immune system which regulates invasions of plasmids and viruses in bacteria and archaea. The functional mechanisms of CRISPR and its whole biological significance are still not fully known (Garneau *et al.*, 2010, Hale *et al.*, 2009, Barrangou *et al.*, 2007, Bolotin *et al.*, 2005, Pourcel *et al.*, 2005, Makarova *et al.*, 2011, Sorek *et al.*, 2013). CRISPRs are constructed from a chain of 21-47 bp repeated sequences (called direct repeats, DR) and in between DRs are unique spacer sequences. These spacers represent foreign DNA originating predominantly from bacteriophages and plasmids. A leader sequence is located at the 5'-end of the CRISPR and usually the *cas*-genes are located upstream of the leader of one of the CRISPR loci (Bolotin et al., 2005, Karginov & Hannon, 2010, Sontheimer & Marraffini, 2010, Pourcel et al., 2005). The three main types of CRISPR-Cas systems differ in the composition of *cas* genes and in the mechanisms of CRISPR RNA (crRNA) processing and interference (Makarova et al., 2011, Makarova *et al.*, 2013, Wiedenheft *et al.*, 2012). *Yersinia* contain the subtype I-F CRISPR-Cas system (Haft *et al.*, 2005, Makarova et al., 2011) and the

cas-genes are located upstream of the most ancestral spacers (Figure 1) of one of the three CRISPR loci present in *Yersinia*. The CRISPR locus and the *cas*-genes have the same transcription direction.

When a prokaryote comes into contact with foreign DNA, the host may integrate a fragment of this DNA, known as a protospacer, into the CRISPR locus as a new spacer. Earlier studies show that approximately 45 % of bacteria and nearly all of archaea contain a CRISPR-Cas system (Grissa *et al.*, 2007a, Pourcel & Drevet, 2013). The new spacers are acquired at the leader proximal end, such that leader distal spacers are older, thus often shared between more isolates (Pourcel et al., 2005, Barrangou et al., 2007). Due to their high diversity, the CRISPR sequences have been used for typing (Shariat & Dudley, 2014), for example, for *Mycobacterium tuberculosis* (Kamerbeek *et al.*, 1997), *Campylobacter* (Schouls *et al.*, 2003) *Streptococcus thermophilus* (Horvath *et al.*, 2008), *Escherichia coli* (Delannoy *et al.*, 2012a, Delannoy *et al.*, 2012b, Touchon *et al.*, 2011, Diez-Villasenor *et al.*, 2010), *Salmonella enterica* (Liu *et al.*, 2011, Fabre *et al.*, 2012) and also for *Erwinia amylovora* (McGhee & Sundin, 2012, Rezzonico *et al.*, 2011). In *Yersinia* there are three loci named as YP1, YP2 and YP3 of which the YP1-locus was initially used as a variable number tandem repeat (VNTR) marker (Le Flèche *et al.*, 2001, Pourcel et al., 2005).

In this study our aim was to generate a comprehensive *Y. pseudotuberculosis* complex database of CRISPR spacers, to use the database to distinguish between strains, and to compare these results with the 90 sequence types (ST) defined in the recent MLST study (Laukkanen-Ninios et al., 2011).

Materials and Methods

Bacterial isolates. A total of 76 *Y. pseudotuberculosis,* ten *Y. similis* and four *Y. wautersii* isolates from the Skurnik laboratory strain collection were analyzed in this study (Table S1). These isolates were selected to cover the largest possible geographic area, host range and to represent as many of the 21 serotypes as possible. Altogether 83 of the 90 STs (Laukkanen-Ninios et al., 2011) were represented, each with a single isolate except for ST3, ST14 and ST43 that were represented by three, four and three isolates, respectively. In addition, sequence data of 40 *Y. pseudotuberculosis* strains and 195 *Y. pestis* strains from earlier investigations were included ((Pourcel et al., 2005, Cui *et al.*, 2008) and Vergnaud&Gorgé, unpublished) as well as CRISPR loci sequences of published complete genome sequences of 4 *Y. pseudotuberculosis* and 6 *Y. pestis* strains (Table S1).

Culture conditions. The bacterial strains were grown in lysogeny broth (LB) (Bertani, 2004); the *Yersinia* strains at 20-22 °C (RT) unless otherwise mentioned and *E. coli* strains at 37 °C. LB supplemented with 1.5% Bacto Agar (LA) was used for solid cultures. *Yersinia* selective CIN-agar plates (Oxoid) were used in conjugation experiments. When required, appropriate antibiotics were added as follows: chloramphenicol (Clm) 20 µg/ml, nalidixic acid (Nal) 100 µg/ml, kanamycin (Kan) 100 µg/ml and diaminopimelic acid (Dap) 0.3 mM.

DNA extraction. Genomic DNA was isolated using the JetFlex DNA isolation kit (GENOMED GmdH, Löhne, Germany).

Sequencing of the CRISPR loci. The three CRISPR loci of *Y. pseudotuberculosis* complex were targeted by PCR based on previously published *Y. pestis* CRISPR primer sequences (Le Flèche et al., 2001, Pourcel et al., 2005). New primers were designed for CRISPR YP2 as the earlier published *Y. pestis* primers failed with many *Y. pseudotuberculosis* strains. The CRISPR loci-specific primers are presented in Table S2. PCR reactions were run in a final volume of 50 µl containing 50 pmol of each primer, 5 µl of 10 x Dynazyme II buffer, 200 µM of each dNTP, 1 U of Dynazyme II, and 150 ng of template DNA. PCR consisted of the following steps: 94°C for 3 min, 32 cycles of denaturation at 94°C for 40 s, annealing at 53°C for 40 s and extension at 72°C for 3 min, and final extension step at 72°C for 10min. PCR amplified fragments were visualized after agarose gel electrophoresis (1.2 % agarose) by ethidiumbromide staining. The DNA fragments were sent to Institute for Molecular Medicine Finland (FIMM) core facility for sequencing after Exonuclease I (Neo Lab) and Shrimp Alkaline Phosphatase (Promega) treatment. The fragments were sequenced using the Applied Biosystems Dye Terminator Kit (BigDye v.3.1) and ABI 3730xl DNA Analyzer. The CRISPR-loci specific primers were used as sequencing primers from the fragment ends and internal primers were designed for sequencing long PCR fragments.

Raw sequence read data was analyzed and assembled to contigs using either Sequencer 5.1 (Gene Codes Corporation) or the Staden Package (Staden, 1996). Before submitting the contig sequence data to CRISPRFinder tool at CRISPRs Web Server (http://crispr.u-psud.fr/) (Grissa *et al.*, 2007b, Grissa et al., 2007a) the data was combined with available CRISPR sequence data from *Y. pestis* and *Y. pseudotuberculosis* strains (Cui et al., 2008, Pourcel et al., 2005, Riehm *et al.*, 2012) (Table S1). The CRISPRFinder tool returned the recognized spacers with unique randomly selected identification

numbers. The genbank non-redundant (nr) nucleotide sequence database was searched for individual spacer sequences using the BLASTN tool (Altschul *et al.*, 1990).

Accession numbers. All the sequences reported in this article including the earlier published but not submitted sequence data (Cui et al., 2008, Pourcel et al., 2005, Riehm et al., 2012) were deposited to nucleotide sequence databases. The accession numbers for the YP1, YP2 and YP3 loci are listed in Table S1.

Construction of pYptb32953::*cat* and pTM100-CRISPR. Primers specific for pYptb32953 (Acc. no BX936400.1), the 27 kb cryptic plasmid of *Y. pseudotuberculosis* IP32953 (Table S2) were used to amplify a 797 bp fragment of pYptb32953 from a plasmid miniprep template. The PCR fragment was purified and digested with EcoRI followed by ligation with EcoRI-digested and SAP-treated suicide vector pSW23T (Demarre *et al.*, 2005). The ligation mixture was electroporated into *E. coli* strain ω7249 that is Kan^R (Babic *et al.*, 2008). Transformants carrying the correct insert were identified by PCR and the isolated plasmid named as pSW23T-pIP was further confirmed by restriction digestions. The suicide construct was introduced to *Y. pseudotuberculosis* IP32953 by conjugation from *E. coli* ω7249/pSW23T-pIP and Clm^R transconjugants were selected with LA-Clm plates where the donor was unable to grow due to its requirement for diaminopimelic acid (Dap). One of the Clm^R transconjugants was named as IP32953/pYptb32953::*cat* and used as a donor to introduce the tagged plasmid into *E. coli* strains PM191NaR, a Nal^R spontaneous derivative of PM191 (Meacock & Cohen, 1980) to obtain *E. coli* PM191NaR/pYptb32953::*cat,* and to strain ω7249 to obtain *E. coli* ω7249/pYptb32953::*cat*. pTM100-CRISPR was constructed by cloning a PCR-amplified 909-bp DNA-fragment of pYptb32953 (nucleotides 13,002-13,910; for PCR-primers, see Table S2) into EcoRV-site of pTM100 (Michiels & Cornelis, 1991). pTM100-CRISPR was electroporated into *E. coli* strain ω7249. pTM100-waaF (Noszczynska *et al.*, 2014) was used as spacer-free control plasmid in mobilization experiments.

Conjugation frequency assays. The *E. coli* ω7249/pYptb32953::*cat,* ω7249/pTM100-CRISPR and ω7249/pTM100-waaF strains were used as donor strains to determine conjugation/mobilization frequencies into a set of *Y. pseudotuberculosis* strains. The donor bacteria were grown in LB-Kan-Clm-Dap at 37ºC for 16 h, the culture was diluted 1:10 in fresh medium and incubated for an additional 3 h, washed and resuspended into PBS to OD_{600} of ~1.0. The recipient bacteria were grown in LB at 22ºC for 16 h, the culture was diluted 1:10 in fresh medium and incubated for an additional 3 h, washed and resuspended into PBS to OD_{600} of ~1.0. For each recipient strain, three parallel matings were prepared.

Equal amounts of the donor and recipient suspensions were mixed and 100 µl aliquots were pipetted in the middle of three parallel LA-plates supplemented with Dap but without antibiotics. The plates were incubated at 37 ºC for 16 h. The bacteria on the plates' surface were resuspended into 1 ml of PBS. 200 µl aliquots were recovered from each and diluted with PBS to OD_{600} of 0.2. The concentrations of donor, recipient and transconjugant bacteria in these mating mixtures were determined by pipetting 5 µl drops of 10^{0} – 10^{-8} diluted mixtures on LA-Kan-Clm-Dap plates (for donor counts), CIN plates (for total recipient counts) and CIN-Clm plates (for transconjugant counts). The donor plates were incubated for 24 h at 37ºC and the recipient and transconjugant plates at 22ºC for 48 h. The colonies in the last dilutions showing growth were counted. Conjugation frequencies were expressed as ratios between the transconjugant and recipient concentrations.

Results

The YP1 locus was amplified from 60 of the 90 Skurnik laboratory *Y. pseudotuberculosis* complex strains. Twenty *Y. pseudotuberculosis* strains and all ten *Y. similis* strains yielded no PCR product (Table S1). From the amplified YP1 locus fragments, five could be only partially sequenced and no sequence was obtained from 11 PCR products. The YP2 locus was amplified and sequenced from 61 strains, 19 strains yielded no PCR products and the PCR products of 4 strains could not be sequenced; in addition 6 strains yielded non-CRISPR sequences. The YP3 locus was amplified and sequenced from 81 strains. Five strains yielded no PCR products, two strains were partly sequenced and the PCR product from one strain could not be sequenced; in addition one strain yielded non-CRISPR sequence. We did not push to optimize the PCR-based sequencing approach as whole genome sequencing is a present day viable alternative. Typically highest numbers of spacers were found from the YP1 and YP3 loci (up to 50 different), while very few were in the YP2 locus.

Analysis of the YP1, YP2 and YP3 CRISPR loci in 335 Y. pseudotuberculosis complex strains

The above sequence data was complemented with the CRISPR loci sequences of 40 *Y. pseudotuberculosis* strains and 195 *Y. pestis* strains (Table S1) from earlier investigations ((Pourcel et al., 2005, Cui et al., 2008) and Vergnaud&Gorgé, unpublished). In addition, we extracted the CRISPR loci sequences from published complete genome sequences of 4 *Y. pseudotuberculosis* and 6 *Y. pestis* strains (Table S1). Then, the sequence data of altogether 335 *Y. pseudotuberculosis* complex strains was analysed by the CRISPRFinder tool. The DR consensus sequence identified from these sequences was identical to that of *Y. pestis* (Cui et al., 2008), i.e., 5'-

TTTCTAAGCTACCTGTGCGGCAGTGAAC-3'. Similar to Cui *et al.,* we identified a number of modified DRs with differences to consensus DR in various positions of the CRISPR loci (Table S3).

Altogether more than 6000 spacers with 1902 different spacer sequences were identified among the analysed sequence data (Table S4). The numbering in Table S4 is used to distuingish the spacers. Suprisingly little overlap of spacer distribution between the strains was noticed. 1153 spacers were unique to single strains (shown in Table S4), 311 were present in 2 strains, 143 in 3 strains, 77 in 4 strains, ca. 56 in 5 strains, ca. 47 in 6 strains and 25 in 7 strains. Those that were shared in ≥8 *Y. pseudotuberculosis* and *Y. similis* strains are shown in Table S5. Since these spacers did not give any significant hits in BLASTN search (Table S5) we at present have no clues of their origins except for spacer #7 that had similarity to *E. coli* plasmid sequence. This spacer was present in eight strains. The most common spacers (#1074, #1149, #507, #40) were always found close to the most ancestral, i.e., the leader distal, end of the CRISPR loci.

To visualize possible evolutionary relationships between the strains based on the organization of the spacer sequences, the *Y. pseudotuberculosis, Y. similis* and *Y. wautersii* strain-specific spacer patterns were manually aligned and the alignments are shown in Table S6. If all spacers present in the most recent common ancestor (MRCA) of *Y. pseudotuberculosis* complex had been subsequently maintained, present day strains should have the same root (most ancient) spacer. This is not the case, instead there were several root spacers both in the YP1 and YP3 loci. The most likely explanation is that older spacers were randomly lost. Furthermore, gaps had to be introduced to the spacer patterns to maximize their alignment. When the spacer pattern alignments were used for grouping of the strains we found that the phylogenies of YP1 and YP3 loci seem not to be congruent. In fact, the alignments indicated that the spacers had accumulated independently to these main storage loci. Table S6, parts A and B, show the strains sorted based on YP1 and YP3 alignments, respectively. As an example of this, the YP1 and YP3 alleles of selected strains are shown in Table 1. For instance, the ancient YP1 alleles 539.173.177.-- are associated either with ancient YP3 alleles 507.1238.-- or 1149.1332.-- or 1149.1199. (Table 1, top). Conversely, each of these is associated with two or more very different YP1 alleles (Table 1, bottom). This data strongly suggests that horizontal gene transfer (HGT) and recombination between the CRISPR loci has occurred within the *Y. pseudotuberculosis* complex, however, with the present data we cannot evaluate the full extent of such mosaicisms. Within the strain groups presented in Table S6 one can observe plenty of examples of possible recombination events and reassortments

leading to deletions of spacer(s). For example, in the spacer 539.173-rooted YP1 group (Table 1, top), the spacer block 539.173…187 is present in six strains, however, it is not identical in them as in some strains spacers from the middle are missing (eg. MW145-2, that strain is also missing the most ancient spacer #539). Strain Toyama60, on the other hand has gained spacer block 539.173-177, but not as the most ancient one but the block has recombined after the ancient spacers 104.801.802 (Table 1). On the other hand, in the spacer 1149-rooted YP3 alignments (Table 1, bottom) similar events can also be easily tracked. The clonal evolution of CRISPR loci observed in *Y. pestis* (Cui et al., 2008) may be an exception reminiscent of the situation observed with *Mycobacterium cannettii* and the *Mycobacterium tuberculosis* complex (Blouin *et al.*, 2014).

Among the 84 *Y. pseudotuberculosis* strains with YP1 sequences the most prevalent ancient or root spacer in the YP1 locus was #39 (present in 14 strains), followed by spacers #103, #40, #539, #403, #76 and #581 (present in 13, 9, 6, 4, 3 and 3 strains, respectively). Eight different ancient spacers were shared by two strains and 15 strains had unique ancient spacers. Sequence information for the YP1 locus was not obtained for 50 of the 124 *Y. pseudotuberculosis* and 10 *Y. similis* strains.

The YP2 locus of the strains carried generally 1 or 2 spacers, with only three exceptions in which the locus carried 6 or 8 spacers (Tables S6 and S7). By sequence comparison, a couple of different repeat variants and altogether 17 different spacers were detected in the YP2 locus. No spacers were present in the *Y. similis* YP2 locus. Figure 2 and Table S7 show the alignments of the *Y. pseudotuberculosis* YP2 locus sequences and their comparison to the *Y. pestis* CO92 YP2 locus. Here we exploited for the grouping of strains the CRISPR 5'- and 3'-flanking sequences obtained from the YP2 PCR fragments (Figure 2 and Table S7). Comparison of the YP2 3'-flanks revealed the presence or absence from the strains of five distinct sequence elements that we named as 3'A to 3'E (Figure 2 and Table S7). Interestingly, all five 3' elements are present only the *Y. pestis* and six *Y. pseudotuberculosis* strains (Figure 2). Most other strains were missing the 731-736 bp 3'-E and the 78 bp 3'-B element. The 3'-E element of strain CO92 includes the whole *ypo2574* gene encoding a putative membrane protein of the DUF1440 protein family. The 32 bp 3'D element was present in all strains. The 66 bp 3'-A and the 25 bp 3'-C elements were missing only from the *Y. similis* strains. Absence of the 3'-A element that is the YP2 locus CRISPR leader sequence might explain why *Y. similis* does not carry any spacers in the YP2 locus.

The YP3-locus spacer comparison is presented in Table S6 part B. Based on the identity of the most ancestral spacer the strains could be grouped into >10 groups. Spacers #507, #1149 and #1111 define the largest groups with 38, 31 and 12 strains, respectively. The other spacer groups #511, #1132, #1156, #1199, #1589, #1616, #1622 and #1853 included 2-5 strains each. The remaining strains had either a sporadic most ancestral spacer (7 strains) or we did not get a PCR-product or sequence from the locus (23 strains).

Interestingly, seventy-two spacers were present in two different CRISPR loci. In one instance, this peculiarity was observed in a single strain. Spacer #808 was found in both the YP1 and YP3 locus of strain No-151. Duplications occurred, for example spacer #257 was found in strains MW101-1 and TE-93081 as a tandem repeat duplicate in the YP1 locus. Another example is the spacer pair #1348.1349 that is present twice in strain DC356-2. Also spacer #1 is present twice in strain BB1152 (Table S6).

Relationships between spacer based grouping and sequence types

We next wanted to find out if the spacer based grouping was in line with the MLST study (Laukkanen-Ninios et al., 2011). Comparison of our results to the MLST minimal spanning tree of Laukkanen-Ninios et al (Laukkanen-Ninios et al., 2011) revealed that the CRISPR spacer-based grouping is not in synchrony with the MLST typing; at best weak correlation could be detected (Figures S1 and S2). However, in all spacer-based groups closest CRISPR types tended to belong to closely related sequence types. As an example one can take the YP3 spacer subgroup of #507-1350-rooted strains (Table S6 part B and Figure S2) that grouped in the MLST analysis with a maximum crosslink distance of 5. (Laukkanen-Ninios *et al.,* 2011.) Thus, the CRISPR loci are highly more differentiating than MLST as we found among the *Y. pseudotuberculosis* complex strains (excluding the *Y. pestis* strains) no CRISPR-identical strains. On the other hand, same sequence type strains tend to carry same spacers, for example, three of the four ST16 strains share 12 of the total of 17 different YP1 spacers (Table S6 part A). A similar situation could also be seen with the YP1 spacers of ST42, ST43 and ST19 strains and with YP3 spacers of ST42, ST43, ST9, ST48 and ST14 strains (Table S6 part B). Clearly, more complete CRISPR loci sequences of strains representing individual sequence types will be needed to get better picture of intra-ST CRISPR evolution, as previously done for *Y. pestis* which represents a single ST within the *Y. pseudotuberculosis* complex (Cui et al., 2008, Riehm et al., 2012).

Origin of spacers

BLASTN searches revealed that a number of spacer sequences showed similarity to various plasmid and bacteriophage sequences (Tables S8 and S9). It was interesting to notice that a few spacers (e.g. #585, #283, #82, #1206, #1154 and #1001) present in *Yersinia* species had matches to plasmids. Spacer #82 shows 97 % identity (one base pair difference) to *Y. enterocolitica* plasmid pYE854 and *Y. pseudotuberculosis* IP31758 59 kb plasmid. In Table S9 spacers that have similarities with different bacteriophage sequences are shown. There was good identity to e.g. *Enterobacteria, Erwinia, Escherichia, Salmonella* and *Burkholderia* phages. In Table S10 spacer sequences were compared to whole genome sequences. The bacterial species earlier seen in Table S9 (bacteriophage hits) can also be seen in this table. Spacer #1697 appears in many bacterial species simply because this spacer occurs in the highly conserved 16S ribosomal RNA gene. In many instances the spacers were located in prophage-like elements similar to *Yersinia* –specific spacers (see below).

***Yersinia* –specific spacers.** Some spacers were present in the genomes or plasmids of other *Yersinia* species. Altogether 40 spacers showed significant similarity to sequences present in *Y. enterocolitica* 8081 genome (Table 2). The 8081 genome carries four prophage-like elements (Thomson *et al.*, 2006) and 38 of the 40 spacer sequences were located within two of them, ϕYE98 (22 spacers) and ϕYE250 (16 spacers). A few strains carried two 8081 prophage-specific spacers.

Most spacers identified from *Y. pestis* were not shared with *Y. pseudotuberculosis.* The ones that are shared are shown in Tables S11 and S12. Table S13 is the conversion table for spacer nomenclature from previous studies to the present database (Cui et al., 2008, Riehm et al., 2012). Specifically, in the YP1 locus, spacers #403-405 were shared by a few *Y. pseudotuberculosis* strains, while in the YP3 locus spacer #507 seems to be present in all *Y. pestis* strains and is also common in *Y. pseudotuberculosis* (Tables S6 and S11). Interestingly, spacers #257, #1901 and #1902 are identical except for the one and two extra Gs present in the two latter ones, respectively (Table S4). Spacer #1901 is very common in the *Y. pestis* YP2 locus while #257 appears in YP1 locus in 11 strains of *Y. pseudotuberculosis* and in two of the strains it is present as duplicate. Futhermore, #257 appears in the YP3 locus in seven *Y. pseudotuberculosis* strains and twice in one of the strains. As mentioned before, *Y. pseudotuberculosis* strains rarely contained more than two spacers in YP2 loci. In contrast*, Y. pestis* YP2 locus usually carries three to six spacers or more.

Finally, to extend the spacer comparisons between *Y. pseudotuberculosis* and *Y. pestis*, all the spacer sequences present in *Y. pseudotuberculosis, Y. similis* and *Y. wautersii* were used to search the *Y. pestis*

genomes. Table S12 lists the 33 spacer sequences identified. The table shows seven hits to *Y. pestis* CRISPR elements, but also 14 hits to *Y. pestis* prophages.

pYptb32953. Altogether 32 unique spacers for the 27,702 bp cryptic plasmid pYptb32953 of *Y. pseudotuberculosis* IP32953 (Chain *et al.*, 2004) were identified in 34 strains. The distribution of the spacers along the plasmid sequence is shown in Table 3. No significant distribution bias can be detected. A majority (22) of the spacers map to the forward (+) and 10 map to the reverse (-) strand of the plasmid (Table 3). Spacer #1362 has two matches in the plasmid, one is a 100% match to nt 16931 (-) strand and the other a 31/32 (97%) match to nt 23321 (+) strand. Altogether 30 different strains had a spacer sequence with 100 % identity to the plasmid pYptb32953 sequence, and additionally a few spacers with some mismatches were identified. To see whether the spacer-carrying strains would reject the pYptb32953 plasmid and the spacers-free strains would accept it in conjugation experiments, we tagged the plasmid with a *cat*-gene (see M&M). We first demonstrated that pYptb32953 is indeed a self-conjugative plasmid as predicted based on its annotated sequence (Chain et al., 2004). The pYptb32953::*cat* transferred itself efficiently from IP32953 to *E. coli* strain PM191NaR (data not shown). The conjugation frequencies to 31 different *Y. pseudotuberculosis* strains were determined (Table 4). Examples of the conjugation experiments are shown in Figure 3. Among the strains we observed three levels of restriction to pYptb32953::*cat* conjugation: (i) non-resistant group to which the plasmid transferred without any apparent restriction and under the experimental conditions used 30 – 100 % of recipients were transformed, (ii) intermediate resistant group where 1 to 20% of recipients were transformed and (iii) fully resistant group with less than 0.01 % transformants. While 11 of 12 among the fully resistant strains carried a pYptb32953-specific spacer ten spacer-carrying strains were among the 13 strains in the non-resistant group. Among the 4 spacer-free strains tested 3 were non-resistant and one, YPII, was fully resistant (Table 4).

To find out whether the resistance differences in the spacer-carrying strains could be explained by the presence or absence of the type-IF specific PAM motif GG at the 3'-end of the protospacer (Wiedenheft *et al.*, 2011, Cady *et al.*, 2012, Mojica *et al.*, 2009), the pYptb32953 spacer-flanking sequences were analysed (Table 3). Altogether 23 of the 32 protospacers were flanked by the GG PAM motif and there was no correlation between the presence or absence of the PAM motif and resistance. For example the non-resistant strain J51 carries 4 spacers and two of the protospacers carry the GG PAM motif. Spacer #1632 containing the PAM motif is present in three *Y. similis* strains. One of the

strains is non-resistant and two are fully resistant. In addition, there are fully resistant strains that carry a spacer missing the PAM motif (Table 3).

To find out whether the CRISPR/Cas system in the *Y. pseudotuberculosis* strains is functional we constructed a pair of plasmids based on mobilizable plasmid pTM100 (Michiels & Cornelis, 1991). pTM100-CRISPR carries a 909 bp fragment of pYptb32953 (nucleotides 13002-13910) that contains six protospacers present as spacers in eight of the strains (Table 4). pYM100-waaF was used as a spacer-free control plasmid. The plasmids were mobilized into a set of six spacer-carrying and –free strains representing the non- and fully-resistant groups. A functional CRISPR/Cas system should restrict the mobilization of pTM100-CRIPSR but not that of pTM100-waaF into a spacer-carrying strain, and there should not be any differences in mobilization of either plasmid into a spacer-free strain. The results presented in Table 4 demonstrate that mobilization frequency of pTM100-CRIPSR to all five spacer-carrying strains was significantly lower than that of pTM100-waaF while no difference could be seen with spacer-free strain PB1. These results demonstrated that the CRISPR/Cas system in *Y. pseudotuberculosis* is functional.

As a single nucleotide change in the protospacer sequence in a phage genome may allow the phage to escape the CRISPR immunity (Levin *et al.*, 2013) we also checked for this possibility as some of the spacers had 1-4 mismatches with the protospacers (88-97% identity over 32 nt, Table 3). Also this seemed not to correlate with the resistance as spacer #1167 in strain 774 had 4 mismatches with the protospacer but the strain was fully resistant. Also #1579 in strain KP1244-2B had 1 mismatch but the strain was fully resistant.

Discussion

The three CRISPR loci of 335 *Y. pseudotuberculosis* complex strains were analysed. Altogether 1902 different spacers were found and surprisingly little overlap between the strains was observed. In spite of this, we noticed some correlation between the *Y. pseudotuberculosis* sequence types and CRISPR spacers. To visualize evolutionary relationships between the strains, we aligned the spacer profiles of the strains based on both the YP1 and YP3 spacers, but these aligments showed no congruence. This is a strong argument for the influence of HGT in shaping the genomes of *Y. pseudotuberculosis* and that specifically influences the YP1 and YP3 loci. This is supported by analogous reassortment of CRISPR loci in *Sulfolobus islandicus* (Held *et al.*, 2013) and in *E. coli* (Almendros *et al.*, 2013). On the other

hand, it also reflects the facts (i) that we intentionally selected the strains to represent as divergent collection of *Y. pseudotuberculosis* complex strains as possible, and (ii) that the number of the strains included in the study was still relatively small. Therefore, to draw meaningful evolutionary conclusions, CRISPR sequence data from larger number of strains is needed. The alignments using the YP2 loci demonstrated the high similarity between the *Y. similis* sequences and their distinct separation from other *Y. pseudotuberculosis* complex species (Table S7).

Earlier studies suggested that *Y. pseudotuberculosis* ST43 is the closest relative to *Y. pestis* (Laukkanen-Ninios et al., 2011, Riehm et al., 2012). Interestingly, the spacers of the six ST43 strains investigated here shared almost no spacers with *Y. pestis* (Table S11). It will be interesting to investigate more ST43 strains. The most ancestral *Y. pestis* YP1 and YP3 spacers are observed in *Y. pseudotuberculosis* ST14, ST16, ST41 and ST87 (Table S6) in a similar position but these STs are not close neighbors to *Y. pestis* or to each other (Laukkanen-Ninios et al., 2011). This indicates that these spacers were acquired well before *Y. pestis* speciation and were subsequently lost in most *Y. pseudotuberculosis* lineages or that CRISPR loci may be transferred horizontally.

Another peculiarity in our dataset was the observation that in some cases a spacer was found from two different loci in one strain. Furthermore certain spacers were shared between strains but occurred at different positions or even in different loci. This may be due to the fact that different strains have been invaded by the protospacer-carrying DNA in separate occasions.

Previous studies have shown that the CRISPR variability may be used for typing bacterial species, even though the CRISPR sequence diversity was not as wide as in the *Y. pseudotuberculosis* complex (Riehm et al., 2012). For example, Fabre and others concluded that the CRISPR spacer content in *Salmonella* correlated with MLST and serotyping results, and they indicated that CRISPR analysis may be a powerful tool for molecular typing of *Salmonella* isolates (Shariat *et al.*, 2013a, Shariat *et al.*, 2013b, Fabre et al., 2012). It was also shown that *E. coli* CRISPR typing combined with MLST analysis could differentiate strains from a single clonal group (Touchon et al., 2011). CRISPR has also proven to be a good typing tool for the clonal *Y. pestis* and hypothetical evolutionary models have been created based on the CRISPR spacer arrays (Cui et al., 2008). This has to be treated with utmost care as our present results revealed very big differences between the *Y. pseudotuberculosis* strains and indications that evolution within *Y. pseudotuberculosis* might not be clonal. This method may be very useful for forensic applications, however, this would require an extensive reference collection. We

show here that each ST would need to be considered almost as a single entity, as previously done for ST90 (*Y. pestis*).

The most common spacers had significant similarities mainly with *Yersinia* species. Comparison of spacer with plasmid sequences indicated one notable plasmid, pYptb32953, the cryptic 27,702 kb plasmid of *Y. pseudotuberculosis* strain IP32953 which had significant similarities with 32 spacers.

When the pYptb32953-specific spacers were identified from the 31 strains we set out to test whether the presence or absence of the spacers would influence the conjugation frequency of pYptb32953::*cat* to a strain. As spacer-negative control strains we selected 4 strains. Our hypothesis was that the plasmid would transform the spacer-negative strains but not the spacer-positive ones. The results shown in Table 4 were unexpected and demonstrated that bacteria are versatile. In addition, the finding that pYptb32953::*cat* conjugated efficiently to 10 spacer-positive strains raised the possibility that the CRISPR/Cas system in these strains would not be functional. The mobilization experiments carried out with the pTM100-waaF/pTM100-CRISPR plasmids, however, clearly demonstrated that the CRISPR/Cas system is functional also in these strains. Interestingly, we observed that the CRISPR/Cas system based resistance was not 100 % tight but could reduce the mobilization frequency to ca. 10 % of the spacer-free mobilization. We can make important conclusions from the results. Firstly, strain YPIII that lacks any pYptb32953-specific spacers was fully resistant to pYptb32953::*cat* transformation. The strain likely carries an efficient restriction-modification system or lacks a receptor for the pYptb32953 conjugation apparatus. Secondly, the presence of a group of spacer-positive strains that showed intermediate resistance to pYptb32953::*cat* transformation shows that the resistance can be leaky. We can speculate that one or more other defense systems in addition to the CRISPR/Cas system are required to achieve full conjugation resistance. Thus, it is likely that these systems are not present in the non-resistant and intermediate resistant groups. Further studies are warranted to elucidate the molecular mechanisms behind these phenomena.

When comparing the *Y. pseudotuberculosis, Y. similis* and *Y. wautersii* spacer sequences to the *Y. pestis* genomic sequences (Table S12), altogether 33 spacers showed significant similarity. Seven of the spacers were spacers in the CRISPR loci of *Y. pestis* and of the remaining 26 spacers 14 had hits in the prophage sequence in the *Y. pestis* genomes. Fewer spacer sequence hits were observed with *Y. pestis* strains 91001 and Antiqua reflecting the fact that they are missing certain prophages present in other *Y. pestis* strains. That has been described also earlier (Song *et al.*, 2004, Chain *et al.*, 2006).

We faced some difficulties in both PCR amplification and sequencing some of the CRISPR loci and to overcome this whole genome sequencing will be used in future. Whole genome sequencing will also open new possibilities to distinguish *Y. pseudotuberculosis* strains from each other.

Final conclusions

Our results suggest that *Y. pestis* after divergence from *Y. pseudotuberculosis* has lived protected or secluded life and it has not encountered many foreign transforming DNAs at least when measured with numbers of CRISPR spacers. Apparently there are rare instances in the *Y. pestis* life cycle where it is exposed to other bacteria or bacteriophages. This is realistic as the infected tissues in rodents and/or humans after flea bite are practically sterile, however, while we do not fully understand the life style of *Y. pestis* in the environment, the microbiota of the flea might be a likely source of foreign DNA. Therefore *Y. pestis* has a relatively low number of spacers compared to *Y. pseudotuberculosis.* The latter, on the contrary, is widely spread in nature and seems to have been highly exposed to various insulting genetic elements and this is visible in the high number of spacers present in a single strain, for example the strain YPIII has altogether 75 spacers.

Acknowledgements

This work is part of the European Biodefence project EDA B-1324-ESM4-GP. Scientific contribution to this work has been accomplished by representatives from France and Finland. Work by GV andOG was supported by the French ministry of defense. Anita Liljegren and Juha Laitinen are thanked for expert technical help. The authors declare no conflict of interest.

447 ***Figure legends***

448 **Figure 1.** The locations of the YP1, YP2 and YP3 CRISPR loci in the genome of *Y.*
449 *pseudotuberculosis* strain IP32953 (Accession number BX936398). The CRISPR associated *cas* and
450 *csy* genes are shown with gray shading and the variable CRISPR repeat sequences as striped arrows.
451 The new spacers are added between the leader sequence and the arrowhead of the CRISPR repeat
452 sequences. The locations of the leader sequence and of the primers used to amplify the loci are
453 indicated by a triangle and black arrows, respectively.

454 **Figure 2.** Schematic organization of the YP2 locus elements used for grouping of the strains. The list
455 of strains and the respective sequences are given in Table S7. The leader sequence element 3'-A is
456 indicated by the black box, the spacer-containing sequence by the striped arrow, the other sequence
457 elements by open boxes. The open reading frame within element 3'-E is indicated by the grey arrow.

458 **Figure 3.** Conjugation experiments showing the serial 10-fold dilution plating of 5 µl drops on
459 selective CIN-agar plates without or with Clm allowing the growth from the mating mixture of all the
460 recipient bacteria or only the transconjugant bacteria, respectively. PB1 and DC356-2 represent non-
461 resistant strains, Pa8728 represents intermediate resistant strains, and YPIII, CN3 and Toyama-60, the
462 fully resistant strains.

463 **Figure S1.** Comparison of YP1 spacer-based grouping of *Y. pseudotuberculosis* complex strains (see
464 Table S6) to MLST-based minimal spanning tree in Figure 1 of Laukkanen-Ninios *et al.* (Laukkanen-
465 Ninios et al., 2011). The groups are indicated by their root spacers and from them lines are drawn to the
466 sequence types the strains in the group represent.

467 **Figure S2.** Comparison of YP3 spacer-based grouping of *Y. pseudotuberculosis* complex strains (see
468 Table S6) to MLST-based minimal spanning tree in Figure 1 of Laukkanen-Ninios *et al.* (Laukkanen-
469 Ninios et al., 2011). The groups are indicated by their root spacers and from them lines are drawn to the
470 sequence types the strains in the group represent.

References

Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule & E. Carniel, (1999) *Yersinia pestis,* the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* **96**: 14043-14048.

Aleksic, S., J. Bockemühl & H.H. Wuthe, (1995) Epidemiology of *Y. pseudotuberculosis* in Germany, 1983-1993. In: Yersiniosis: Present and Future. G. Ravagnan & C. Chiesa (eds). Postfach/CH-4009 Basel/Switzerland: Karger, pp. 55-58.

Almendros, C., F.J. Mojica, C. Diez-Villasenor, N.M. Guzman & J. Garcia-Martinez, (2013) CRISPR-Cas functional module exchange in *Escherichia coli*. *MBio* **5**: e00767-00713.

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

Babic, A., A.M. Guerout & D. Mazel, (2008) Construction of an improved RP4 (RK2)-based conjugative system. *Res. Microbiol.* **159**: 545-549.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero & P. Horvath, (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Bertani, G., (2004) Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **186**: 595-600.

Blouin, Y., G. Cazajous, C. Dehan, C. Soler, R. Vong, M.O. Hassan, Y. Hauck, C. Boulais, D. Andriamanantena, C. Martinaud, E. Martin, C. Pourcel & G. Vergnaud, (2014) Progenitor "*Mycobacterium canettii*" clone responsible for lymph node tuberculosis epidemic, Djibouti. *Emerg Infect Dis* **20**: 21-28.

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

Bolotin, A., B. Quinquis, A. Sorokin & S.D. Ehrlich, (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Cady, K.C., J. Bondy-Denomy, G.E. Heussler, A.R. Davidson & G.A. O'Toole, (2012) The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.* **194**: 5728-5738.

Ch'ng, S.L., S. Octavia, Q. Xia, A. Duong, M.M. Tanaka, H. Fukushima & R. Lan, (2011) Population structure and evolution of pathogenicity of *Yersinia pseudotuberculosis*. *Appl. Environ. Microb.* **77**: 768-775.

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

Chain, P.S., P. Hu, S.A. Malfatti, L. Radnedge, F. Larimer, L.M. Vergez, P. Worsham, M.C. Chu & G.L. Andersen, (2006) Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J. Bacteriol.* **188**: 4453-4463.

Chauvaux, S., M.A. Dillies, M. Marceau, M.L. Rosso, S. Rousseau, I. Moszer, M. Simonet & E. Carniel, (2011) In silico comparison of Yersinia pestis and *Yersinia pseudotuberculosis* transcriptomes reveals a higher expression level of crucial virulence determinants in the plague bacillus. *Int. J. Med. Microbiol.* **301**: 105-116.

Cui, Y., Y. Li, O. Gorge, M.E. Platonov, Y. Yan, Z. Guo, C. Pourcel, S.V. Dentovskaya, S.V. Balakhonov, X. Wang, Y. Song, A.P. Anisimov, G. Vergnaud & R. Yang, (2008) Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE* **3**: e2652.

Cui, Y., C. Yu, Y. Yan, D. Li, Y. Li, T. Jombart, L.A. Weinert, Z. Wang, Z. Guo, L. Xu, Y. Zhang, H. Zheng, N. Qin, X. Xiao, M. Wu, X. Wang, D. Zhou, Z. Qi, Z. Du, H. Wu, X. Yang, H. Cao, H. Wang, J. Wang, S. Yao, A. Rakin, D. Falush, F. Balloux, M. Achtman, Y. Song & R. Yang, (2013) Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* **110**: 577-582.

Delannoy, S., L. Beutin, Y. Burgos & P. Fach, (2012a) Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains by use of the CRISPR locus as a target for a diagnostic real-time PCR. *J. Clin. Microbiol.* **50**: 3485-3492.

Delannoy, S., L. Beutin & P. Fach, (2012b) Use of clustered regularly interspaced short palindromic repeat sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR. *J. Clin. Microbiol.* **50**: 4035-4040.

Demarre, G., A.M. Guerout, C. Matsumoto-Mashimo, D.A. Rowe-Magnus, P. Marliere & D. Mazel, (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.* **156**: 245-255.

Diez-Villasenor, C., C. Almendros, J. Garcia-Martinez & F.J. Mojica, (2010) Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* **156**: 1351-1361.

Fabre, L., J. Zhang, G. Guigon, S. Le Hello, V. Guibert, M. Accou-Demartin, S. de Romans, C. Lim, C. Roux, V. Passet, L. Diancourt, M. Guibourdenche, S. Issenhuth-Jeanjean, M. Achtman, S. Brisse, C. Sola & F.X. Weill, (2012) CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PLoS ONE* **7**: e36995.

Garneau, J.E., M.E. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A.H. Magadan & S. Moineau, (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Grissa, I., G. Vergnaud & C. Pourcel, (2007a) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics* **8**: 172.

Grissa, I., G. Vergnaud & C. Pourcel, (2007b) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucl. Acids Res.* **35**: W52-57.

Haft, D.H., J. Selengut, E.F. Mongodin & K.E. Nelson, (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology* **1**: e60.

Hale, C.R., P. Zhao, S. Olson, M.O. Duff, B.R. Graveley, L. Wells, R.M. Terns & M.P. Terns, (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Harbeck, M., L. Seifert, S. Hansch, D.M. Wagner, D. Birdsell, K.L. Parise, I. Wiechmann, G. Grupe, A. Thomas, P. Keim, L. Zoller, B. Bramanti, J.M. Riehm & H.C. Scholz, (2013) *Yersinia pestis* DNA from skeletal remains from the 6(th) century AD reveals insights into Justinianic Plague. *PLoS Pathog* **9**: e1003349.

Held, N.L., A. Herrera & R.J. Whitaker, (2013) Reassortment of CRISPR repeat-spacer loci in *Sulfolobus islandicus*. *Environ Microbiol*.

Horvath, P., D.A. Romero, A.C. Coute-Monvoisin, M. Richards, H. Deveau, S. Moineau, P. Boyaval, C. Fremaux & R. Barrangou, (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**: 1401-1412.

Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal & J. van Embden, (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**: 907-914.

562 Karginov, F.V. & G.J. Hannon, (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea.
563 *Mol Cell* **37**: 7-19.
564 Laukkanen-Ninios, R., X. Didelot, K.A. Jolley, G. Morelli, V. Sangal, P. Kristo, C. Brehony, P.F. Imori, H.
565 Fukushima, A. Siitonen, G. Tseneva, E. Voskressenskaya, J.P. Falcao, H. Korkeala, M.C. Maiden, C.
566 Mazzoni, E. Carniel, M. Skurnik & M. Achtman, (2011) Population structure of the *Yersinia*
567 *pseudotuberculosis* complex according to multilocus sequence typing. *Environ Microbiol* **13**: 3114-3127.
568 Le Flèche, P., Y. Hauck, L. Onteniente, A. Prieur, F. Denoeud, V. Ramisse, P. Sylvestre, G. Benson, F. Ramisse &
569 G. Vergnaud, (2001) A tandem repeats database for bacterial genomes: application to the genotyping
570 of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol* **1**: 2.
571 Levin, B.R., S. Moineau, M. Bushman & R. Barrangou, (2013) The population and evolutionary dynamics of
572 phage and bacteria with CRISPR-mediated immunity. *PLoS Genet* **9**: e1003312.
573 Liu, F., R. Barrangou, P. Gerner-Smidt, E.M. Ribot, S.J. Knabel & E.G. Dudley, (2011) Novel virulence gene and
574 clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme
575 for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl. Environ. Microb.* **77**:
576 1946-1956.
577 Makarova, K.S., D.H. Haft, R. Barrangou, S.J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F.J. Mojica, Y.I.
578 Wolf, A.F. Yakunin, J. van der Oost & E.V. Koonin, (2011) Evolution and classification of the CRISPR-Cas
579 systems. *Nat Rev Microbiol* **9**: 467-477.
580 Makarova, K.S., Y.I. Wolf & E.V. Koonin, (2013) The basic building blocks and evolution of CRISPR-cas systems.
581 *Biochem Soc Trans* **41**: 1392-1400.
582 McGhee, G.C. & G.W. Sundin, (2012) *Erwinia amylovora* CRISPR elements provide new tools for evaluating
583 strain diversity and for microbial source tracking. *PLoS One* **7**: e41706.
584 Meacock, P.A. & S.N. Cohen, (1980) Partitioning of bacterial plasmids during cell division a cis-acting locus that
585 accomplishes stable plasmid inheritance. *Cell* **20**: 529-642.
586 Michiels, T. & G.R. Cornelis, (1991) Secretion of hybrid proteins by the *Yersinia* Yop export system. *J. Bacteriol.*
587 **173**: 1677-1685.
588 Mojica, F.J., C. Diez-Villasenor, J. Garcia-Martinez & C. Almendros, (2009) Short motif sequences determine the
589 targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.
590 Morelli, G., Y. Song, C.J. Mazzoni, M. Eppinger, P. Roumagnac, D.M. Wagner, M. Feldkamp, B. Kusecek, A.J.
591 Vogler, Y. Li, Y. Cui, N.R. Thomson, T. Jombart, R. Leblois, P. Lichtner, L. Rahalison, J.M. Petersen, F.
592 Balloux, P. Keim, T. Wirth, J. Ravel, R. Yang, E. Carniel & M. Achtman, (2010) *Yersinia pestis* genome
593 sequencing identifies patterns of global phylogenetic diversity. *Nature Genetics* **42**: 1140-1143.
594 Naktin, J. & K.G. Beavis, (1999) *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. *Clinics in laboratory*
595 *medicine* **19**: 523-536, vi.
596 Noszczynska, M., K. Kasperkiewicz, K.A. Duda, J. Podhorodecka, K. Rabsztyn, K. Gwizdala, A.S. Swierzko, J.
597 Radziejewska-Lebrecht, O. Holst & M. Skurnik, (2014) Serological characterization of the
598 enterobacterial common antigen substitution of the lipopolysaccharide of *Yersinia enterocolitica* O:3.
599 *Microbiology*.
600 Pourcel, C. & C. Drevet, (2013) Occurence, diversity of CRISPR-Cas systems and genotyping implications. In:
601 CRISPR-Cas systems. R. Barrangou & J. van der Oost (eds). Berlin Heidelberg: Springer-Verlag, pp. 33-
602 59.
603 Pourcel, C., G. Salvignol & G. Vergnaud, (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by
604 preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies.
605 *Microbiology* **151**: 653-663.
606 Rezzonico, F., T.H. Smits & B. Duffy, (2011) Diversity, evolution, and functionality of clustered regularly
607 interspaced short palindromic repeat (CRISPR) regions in the fire blight pathogen *Erwinia amylovora*.
608 *Appl. Environ. Microb.* **77**: 3819-3829.

Riehm, J.M., G. Vergnaud, D. Kiefer, T. Damdindorj, O. Dashdavaa, T. Khurelsukh, L. Zoller, R. Wolfel, P. Le Flèche & H.C. Scholz, (2012) *Yersinia pestis* lineages in Mongolia. *PLoS One* **7**: e30624.

Savin, C., L. Martin, C. Bouchier, S. Filali, J. Chenau, Z. Zhou, F. Becher, H. Fukushima, N.R. Thomson, H.C. Scholz & E. Carniel, (2014) The *Yersinia pseudotuberculosis* complex: Characterization and delineation of a new species, *Yersinia wautersii*. *Int. J. Med. Microbiol.* **304**: 452-463.

Schouls, L.M., S. Reulen, B. Duim, J.A. Wagenaar, R.J. Willems, K.E. Dingle, F.M. Colles & J.D. Van Embden, (2003) Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. *J. Clin. Microbiol.* **41**: 15-26.

Shariat, N. & E.G. Dudley, (2014) CRISPRs: molecular signatures used for pathogen subtyping. *Appl. Environ. Microb.* **80**: 430-439.

Shariat, N., M.K. Kirchner, C.H. Sandt, E. Trees, R. Barrangou & E.G. Dudley, (2013a) Subtyping of *Salmonella enterica* serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results. *J. Clin. Microbiol.* **51**: 2328-2336.

Shariat, N., C.H. Sandt, M.J. DiMarzio, R. Barrangou & E.G. Dudley, (2013b) CRISPR-MVLST subtyping of *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Heidelberg and application in identifying outbreak isolates. *BMC Microbiol* **13**: 254.

Skurnik, M., A. Peippo & E. Ervelä, (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. *Mol. Microbiol.* **37**: 316-330.

Song, Y., Z. Tong, J. Wang, L. Wang, Z. Guo, Y. Han, J. Zhang, D. Pei, D. Zhou, H. Qin, X. Pang, J. Zhai, M. Li, B. Cui, Z. Qi, L. Jin, R. Dai, F. Chen, S. Li, C. Ye, Z. Du, W. Lin, J. Yu, H. Yang, P. Huang & R. Yang, (2004) Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Research* **11**: 179-197.

Sontheimer, E.J. & L.A. Marraffini, (2010) Microbiology: slicer for DNA. *Nature* **468**: 45-46.

Sorek, R., C.M. Lawrence & B. Wiedenheft, (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**: 237-266.

Souza, R.A., A. Pitondo-Silva, D.P. Falcao & J.P. Falcao, (2010) Evaluation of four molecular typing methodologies as tools for determining taxonomy relations and for identifying species among *Yersinia* isolates. *J. Microbiol. Meth.* **82**: 141-150.

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

Staden, R., (1996) The Staden sequence analysis package. *Mol Biotechnol* **5**: 233-241.

Tauxe, R.V., (2004) Salad and pseudoappendicitis: *Yersinia pseudotuberculosis* as a foodborne pathogen. *J. Infect. Dis.* **189**: 761-763.

Thomson, N.R., S. Howard, B.W. Wren, M.T. Holden, L. Crossman, G.L. Challis, C. Churcher, K. Mungall, K. Brooks, T. Chillingworth, T. Feltwell, Z. Abdellah, H. Hauser, K. Jagels, M. Maddison, S. Moule, M. Sanders, S. Whitehead, M.A. Quail, G. Dougan, J. Parkhill & M.B. Prentice, (2006) The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genet* **2**: e206.

Touchon, M., S. Charpentier, O. Clermont, E.P. Rocha, E. Denamur & C. Branger, (2011) CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J. Bacteriol.* **193**: 2460-2467.

Wiedenheft, B., S.H. Sternberg & J.A. Doudna, (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338.

Wiedenheft, B., E. van Duijn, J.B. Bultema, S.P. Waghmare, K. Zhou, A. Barendregt, W. Westphal, A.J. Heck, E.J. Boekema, M.J. Dickman & J.A. Doudna, (2011) RNA-guided complex from a bacterial immune system

enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* **108**: 10092-10097.

Voskressenskaya, E., A. Leclercq, G. Tseneva & E. Carniel, (2005) Evaluation of ribotyping as a tool for molecular typing of *Yersinia pseudotuberculosis* strains of worldwide origin. *J. Clin. Microbiol.* **43**: 6155-6160.

Table 1. YP1 and YP3 spacer alignments of selected *Y. pseudotuberculosis* strains[a]

Strain	YP1 locus	YP3 locus
Toyama60	104.801.802.303.173-177.181-170.1842.230.334.217.130.171.283.172.526.173.1827.174.	1141.1136.1200.-----1202.331.1203.261.81.1204.1205.1773-1776.130.1777.361.
MW145-2	----------173-177.130. 178.----------17?.180.181-185.183.186.187.188.189.	107.1238-1241.1868.1810.848.71.1811.77.1847.712.340.80?.1812.1813-1818.171.
Pa3606	----------3?.173-177.nnn.----------.181-185.183.186.187.	107.1238-1241.1868.1810.848.71.1811.77.1847.712.340.80?.1812.
IP31758	----------33.173-177.1865.178.----140-143.1818.17?.180.181-185.183.186.187.361.	1141.1136.1200.1201.1202.331.1203.261.81.1204.1205.1466-1471.808.277.1472.833.1473-1477.
D426	----------3?.173-177.1865.178.----140-143.	1141.1332.----------1203.261.1333-1337.
Y718	----------33.173-177.1865.178.178.140-143.1818.17?.180.181-185.183.186.187.	1141.1332.----------1203.261.1333-1337.

Strain	YP3 locus	YP1 locus
RD20	1141.1136.1200.1201.1202.331.1203.261.81.1204.1205.1206.186.1207-1212.136.734.1213-1211.288.1338.133.1741.274.1818.266.	181.882.848.800.----852-854.883-886.859-877.128-146.
No-151	1141.1136.1200.1201.1202.331.1203.261.81.1204.1205.----------1465-1471.808.858.1644.	104.801.802.803.204.650.277.681-683.----------808-809.810-813.333.250.71.814-818.
IP31758	1141.1136.1200.1201.1202.331.1203.261.81.1204.1205.1466-1468.1468-1471.808.277.1472.833.1473-1477.	----------33.173-177.1865.178.----140-143.1818.17?.180.181-185.183.186.187.361.
Toyama60	1141.1136.1200.-----1202.331.1203.261.81.1204.1205.1773.1774.177?.177?.130.1777.361.	104.801.802.33.173-177.181-170.1842.230.334.217.130.171.283.172.526.173.1827.174.
D426	1141.1332.1203.261.1333-1337.	----------3?.173-177.1865.178.----140-143.
Gifu-liver	1141.1332.1203.261.1333-1337.331.	----------650.277.681-683.844-865.1820.870-877.1811.878-882.1101.883-887.1848.888-890.
Y718	1141.1332.1203.261.1333-1337.	----------33.173-177.1865.178.178.140-143.1818.17?.180.181-185.183.186.187.
Y732	1141.1332.1203.261.1333-1337.	----------1068-1069.248.248.1070.1071.
WP-931110	107.1238-1240.----------1244.-----1246-1241.1240-1252.1780.	33.41.77.78.86.88.81.87.80-82.88-92.---93.94.97.98.16.176-177.186.180.111.181.253.182-193.
Pa3606	107.1238-1240.1241.1242.1243.1244.1245.1246-1241.1868.1810.848.71.1811.77.1847.712.340.80?.1812.	33.173-177.nnn.----------181-185.1833.186.187.
MW145-2	107.1238-1240.1241.1242.1243.1244.1245.1246-1241.1868.1810.848.71.1811.77.1847.712.340.80?.1812.1813-1818.171.	----173-177.130.178.----------17?.180.181-185.1833.186.187.188.189.

[a] nnn, the middle part of the strain Pa3606 YP1 locus PCR fragment could not be completely sequenced.

Table 2. Spacers of *Y. pseudotuberculosis* complex with matching sequences (>88% identity in BLASTN search) to genomic sequence of *Y. enterocolitica* 8081 (GenBank accession no: AM286415). Strains in bold carry several spacers similar to *Y. enterocolitica* sequence. All spacers are 32 nt long, except #846 that is 33 nt long.

Spacer	Strain	YP locus	Location in 8081 genome	BLASTN search identity %
Spacers in prophage ϕYE98 (location in 8081 genome: 981223–1011295)				
1779	WP-931108	YP3	984218	88 (28/32)
1191	**79136**	YP3	984296	88 (28/32)
986	**WP-931110**	YP1	986574	97 (31/32)
1193	**79136**	YP3	987079	91 (29/32)
566	**2889, 2895, Y385, Y728, Y729**	YP1	987179	97 (31/32)
988	**WP-931110**	YP1	987234	97 (31/32)
244	PC95-219-1	YP1	988311	91 (29/32)
699	**H892-36-87**	YP1	991747	91 (29/32)
711	**H892-36-87**	YP1	991941	91 (29/32)
1471	IP31758, No-151, OK10700, **R30**	YP3	992073	100 (32/32)
1076	22917-2L, IP32952	YP2	993118	91 (29/32)
1203	8727-7, D426,D54,Gifu-liver	YP3	994228	94 (30/32)
1748	**R30**	YP3	994348	91 (29/32)
55	**BB28, Y74**	YP1	995024	94 (30/32)
86	**F-401-1,** Wla658, **WP-931110**	YP1	1001481	97 (31/32)
50	**BB28, Y74**	YP1	1004019	88 (28/32)
1629	**MW48, R103-2, R626R**	YP3	1004236	88 (28/32)
123	**H-3831**	YP1	1009835	100 (32/32)
79	**F-401-1**	YP1	1010599	91 (29/32)
647	**CN2**	YP1	1010603	91 (29/32)
1317	**CN3**-5end	YP3	1010802	88 (28/32)
846	PT245	YP1	1010919	94 (30/32)
Spacers in prophage ϕYE200 (location in 8081 genome: 1991720-2007210)				
885	**R104-2**	YP1	1993391	88 (28/32)
Spacers in prophage ϕYE250 (location in 8081 genome: 2503099-2554665)				
1315	**CN3**-5end	YP3	2527852	88 (28/32)
1681	Pa8728, d54	YP3	2533319	94 (30/32)
1270	B56, **No-21**	YP3	2533871	97 (31/32)
1256	AZ960106-1	YP3	2534586	100 (32/32)
1662	**MW48, R103-2, R626R**	YP3	2535837	94 (30/32)
1412	H146-84K, **R111, YPIII**	YP3	2538051	88 (28/32)
225	OK5466	YP1	2538088	94 (30/32)

570	**2889, 2895, Y385, 728, Y729**	YP1	2538983	97 (31/32)
1295	**CN2, R104-2**	YP3	2539577	100 (32/32)
121	**H-3831**	YP1	2540279	94 (30/32)
889	**R104-2**	YP1	2542146	100 (32/32)
1457	**H892-36-87**	YP3	2544550	88 (28/32)
918	**R111, YPIII**	YP1	2548392	94 (30/32)
821	**No-21**	YP1	2550610	91 (29/32)
695	**H892-36-87**	YP1	2552642	100 (32/32)
1234	93422-5end, **CN3**-5end	YP3	2552879	91 (29/32)
Other location *Ye2993 gltK*				
120	G5431	YP1	3264360	88 (28/32)

Table 3. The spacers specific for protospacers in pYptb32953, the cryptic 27,702 bp plasmid of *Y. pseudotuberculosis* strain IP32953 (Acc. no BX936400.1). Strains OK5466, J51, R103-1, S107, H-3831, No.21, DD110, Pa8728, MW101-1, and PC95-219-1 carry several different pYptb32953 spacers.

Spacer	Present in strains	Locus	Serotype	ST[a]	Location in plasmid (strand)	Protospacer sequence in pYptb32953 ± 6 nt. The GG PAM motif is grey-shaded.	Identity % (32 nt)	Conjugation resistance type (Table 4)
234	OK5466	YP1	O:5b	73	3697 (+)	tagtct/gatgggtctcaaaatacgcaccaaagggaacg/GGaaaa	100	Intermediate
790	J51	YP1	O:13	47	5093 (+)	gatttc/aacgaaaaaaacgccggtaatgcgtcgattgt/GGggac	100	Non-resistant
1285	BB28	YP3	O:2b	53	5312 (+)	gttaaa/agtggggaacctaccggatggaatccgtttcg/Gctgaa	100	Non-resistant
287	R103-1	YP1	O:3	58	5432 (+)	acccga/caggaaacccgcctcagtgacgccgttgatgc/tGttat	100	Non-resistant
	S107	YP1	O:5b	87				--
1178	79136	YP3	O:1b	88	5469 (+)	ctgtta/tgttggggcttgaccacgccagccgtgaccac/GGTatt	100	Non-resistant
789	J51	YP1	O:13	47	5751 (+)	tactgg/atggacggcgtttagtgctgtttattgatgag/Gtctgg	100	Non-resistant
1450	H-3831	YP3	O:4a	48	7359 (+)	agtctt/ttcgagtcaatcatgggaagactatctttatt/GGcagc	100	Fully resistant
1167	774	YP3	O:4a	32	8228 (-)	gtgagc/gcggttaaataccccccgcattagtaatgaa/GGtgat	88	Fully resistant
464	Y722	YP1	unknown	19	9676 (-)	gcgagt/ttatcggggtcggttgcatcactaatgacatt/GGaaat	100	--
	Vlassel	YP1	O:3	57				Fully resistant
	IP32802	YP1	unknown	19				--
	Y716	YP1	unknown	19				--
	H1960003	YP1	unknown	unknown				--
784	J51	YP1	O:13	47	10080 (-)	gaaaac/accaaggtagtgacataaccggcgcgagcatt/aattac	100	Non-resistant
1454	H-3831	YP3	O:4a	48	10082 (+)	ttggtg/ttttcggatgatgaggcggttattagcactga/GGtggg	100	Fully resistant
1264	B56	YP3	R	9	10680 (+)	tgatcc/agcatattaacccgacgatggtattacgctat/GGcgat	100	Fully resistant
	No.21	YP3	O:1a	86				Intermediate
556	22917/2L	YP1	O:5a	16	12868 (+)	cagcat/gagaactatgtgcatctgttttatccgtcaga/GGgtgg	100	Intermediate
1820	Gifu-liver	YP1	O:1b	22	12869 (+)	agcatg/agaactatgtgcatctgttttatccgtcagag/GGtgga	94	Intermediate
1632	R103-2 (*Y. sim*)	YP3	O:5b	45	13031 (+)	gttcag/gtcatgatgggttaatcacgtcaatgcacgct/GGcacc	100	Fully resistant
	MW48 (*Y. sim*)	YP3	O:9	80				Fully resistant
	R626R (*Y. sim*)	YP3	O:9	83				Non-resistant

1528	R103-1	YP3	O:3	58	13276 (+)	aaatga/aacatttaattagaccatgttgggtggctgtc/GGtttg	100	Non-resistant
	J92	YP3	O:13	82				Non-resistant
	S107	YP3	O:5b	87				--
134	H-3831	YP1	O:4a	48	13374 (+)	acaccg/acgacggttaacagcaccttgccccagtggca/GGaaaa	100	Fully resistant
1449	H-3831	YP3	O:4a	48	13453 (+)	catggt/caaagcgcatcacggattttcaggggataac/GGcagc	100	Fully resistant
1444	H-3831	YP3	O:4a	48	13545 (+)	gttgcg/agtaacagcagctcggcatggtttaacaccgc/GGcatg	94	Fully resistant
760	J51	YP1	O:13	47	13640 (+)	ttgttt/aacctgcaatacggttgatgtttatttgtctc/GGtcag	100	Non-resistant
133	H-3831	YP1	O:4a	48	14452 (-)	cacaga/ttttatttggtgatatttgaattgatcggcaa/GGcgta	100	Fully resistant
1154	H892-36/87	YP3	O:1a	12	15362 (+)	agttca/caaacaacattaaataatgctaataattatac/tGattc	100	Non-resistant
	CN3	YP3	O:14 (R)	17				Fully resistant
	2889	YP3	O:1b	43				--
	2895	YP3	O:1b	43				--
	Y734	YP3	unknown	43				--
1361	DD110	YP3	O:6	11	15405 (+)	caagag/tgagtaacattacaaatgtagctattgaagag/Gctaat	100	Fully resistant
	Pa8728	YP3	R	60				Intermediate
1579	KP1244-2B	YP3	O:2c	56	16774 (-)	tttttt/cattacctcattgatactcggaacttcatcaa/GGcagt	97	Fully resistant
1362	DD110	YP3	O:6	11	16931 (-)	cgttcg/cgggggtggcttgttggcctccccgcttcact/GGcttt	100	Fully resistant
1586	MW101-1	YP3	O:4b	28	17275 (-)	ggggat/atatccccccaaattaaccgcccactggggtg/GGcttt	100	Non-resistant
	PC95-219-1	YP3	R	84				Non-resistant
1585	MW101-1	YP3	O:4b	28	17545 (+)	cggagt/ggcgattgccgttgattggggtaactgcaagt/tatcgc	94	Non-resistant
	PC95-219-1	YP3	R	84				Non-resistant
218	WP-931109	YP1	O:15	22	21892 (-)	tgctct/tttcctcggtctggctggcggtgactgtcgcc/GGttgg	97	Intermediate
	OK5466	YP1	O:5b	73				Intermediate
1362	DD110	YP3	O:6	11	23321 (+)	gttctc/gggggtggcttgtcggcctccccgcttcactg/Gctttt	97	Fully resistant
283	Pa8728	YP1	R	60	24168 (-)	agtggg/tgatgtgcagcatgaaagctatattgcctcat/GGctta	100	Intermediate
232	OK5466	YP1	O:5b	73	24547 (-)	ctgtgt/ttaatgtccagcaaatagacggcttgccactag/aGaca	94	Intermediate
1407	No.21	YP3	O:1a	86	27085 (+)	tggggc/cgtagtgctgtttaaccgtttttgtggtcag/GGtatg	100	Intermediate
	H141-84	YP3	unknown	9				--
	Y717	YP3	unknown	9				--

[a]ST, Multilocus sequence type according to MLST Databases at the Unversity of Warwick, Warwick Medical School (http://mlst.ucc.ie/mlst/dbs/Ypseudotuberculosis/GetTableInfo_html) [Laukkanen-Ninios, 2011 #9224]

Table 4. Influence of pYptb32953-specific spacers on pYptb32953::*cat* conjugation and plasmid mobilization restriction by *Y. pseudotuberculosis* and *Y. similis* strains. Presented are for pYptb32953::*cat* conjugations transconjugant/recipient ratios of three parallel mating experiments. Variations between the ratios of the three matings are due to inaccuracy in the serial dilution drop method used for measuring bacterial concentrations in mating mixtures (see Figure 3). Mobilization of plasmids pTM100-waaF and pTM100-CRISPR was performed with six strains, highlighted in gray, and the mobilization frequencies are given as transconjugant/recipient percentages.

Strain	Spacer no	pYptb32953::*cat* conjugations			Mobilization frequency (%)	
		I	II	III	pTM100-waaF	pTM100-CRISPR
Non-resistant group (>0.2)						
PB1	-	0.6	0.5	1	~7	~2
Wla658	-	1.67	4.5	1.88		
DC356-2	-	1	0.82	0.9		
257	1115	1	1.67	2.5		
J51	760.784.789.790	0.35	0.33	1		
R103-1	287.1528	0.55	4.5	2.5	5-10	1
J92	1528	0.3	0.7	0.71		
79136	1178	2.33	2.33	2.8		
R626R (*Y. similis*)	1632	0.83	0.67	1	1	0.1
MW101-1	1585.1586	0.2	0.5	0.53		
PC95-219-1	1586	0.33	0.4	0.5		
H892-36/87	1154	1.67	1	1		
BB28	1285	1	1.25	1.33		
Intermediate resistant group (0.01 – 0.2)						
Gifu-Liver	1820	0.044	0.008	0.04		
22917/2L	556	0.1	0.08	0.013		
No.21	1264.1407	0.033	0.2	0.022		
Pa8728	283.1361	0.13	0.2	0.083		
WP-931109	218	0.13	0.2	0.13		
OK5466	218.232.234	0.067	0.05	0.17		
Fully resistant group (0 – 0.0099)						
YPIII	-	0	0.0063	0		
Vlassel	464	0.031	0.002	0.00067		
H-3831	133.134.1444.1449.1450.1454	0	0.004	0.004	1	<0.01
774	1167	0	0	0.001		
B56	1264	0.004	0.01	0.033		
R103-2 (*Y. similis*)	1632	0	0	0	1-5	0.1-0.5
MW48 (*Y. similis*)	1632	0.00005	0	0	0.1-1	0
WP-931205	1010	0	0	0		
Toyama-60	1775	0	0	0		
CN3	1154	0	0.0002	0		
DD110	1361.1362	0.048	0.01	0.0078		
KP1244-2B	1579	0.00017	0.00011	0.000083		

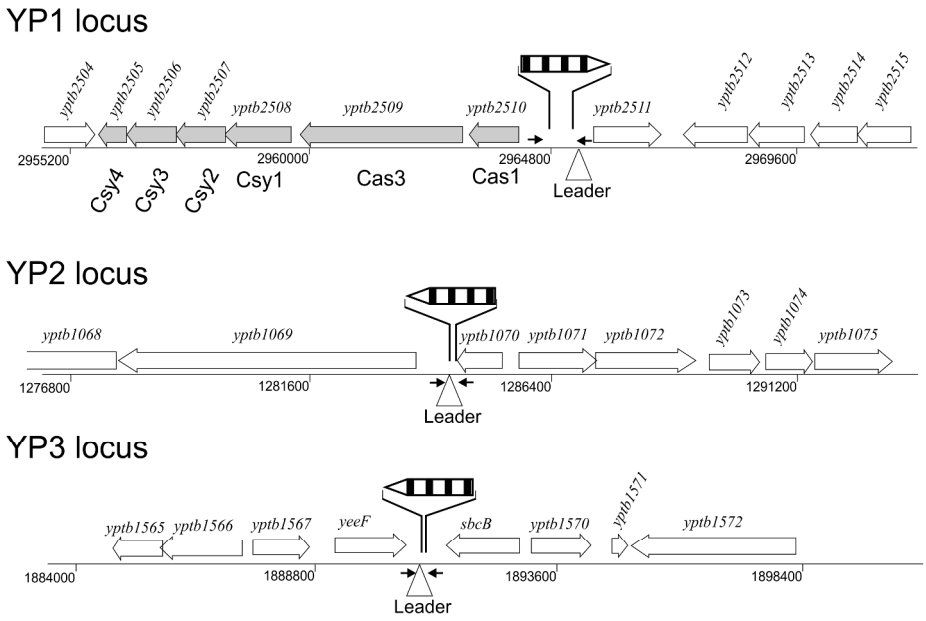

Figure 1. The locations of the YP1, YP2 and YP3 CRISPR loci in the genome of Y. pseudotuberculosis strain IP32953 (Accession number BX936398). The CRISPR associated cas and csy genes are shown with gray shading and the variable CRISPR repeat sequences as striped arrows. The new spacers are added between the leader sequence and the arrowhead of the CRISPR repeat sequences. The locations of the leader sequence and of the primers used to amplify the loci are indicated by a triangle and black arrows, respectively.
269x183mm (300 x 300 DPI)

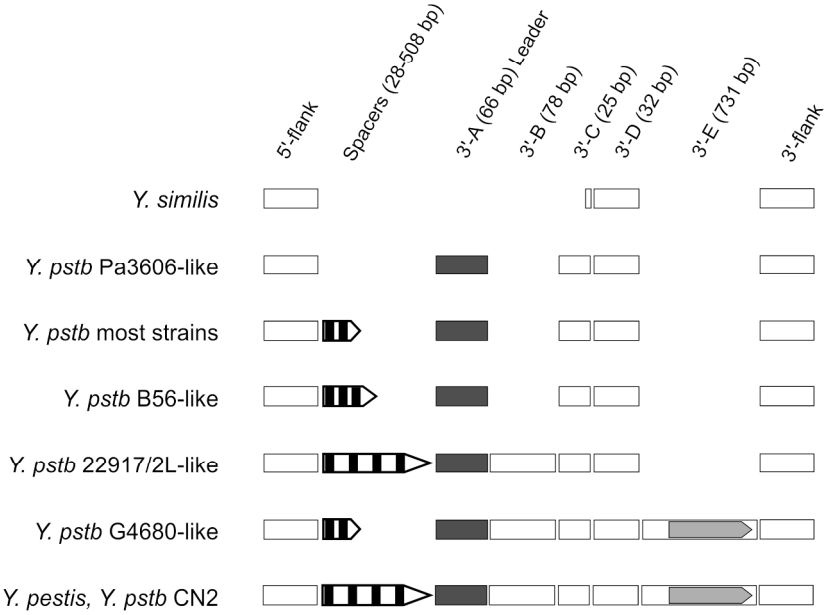

Figure 2. Schematic organization of the YP2 locus elements used for grouping of the strains. The list of strains and the respective sequences are given in Table S7. The leader sequence element 3'-A is indicated by the black box, the spacer-containing sequence by the striped arrow, the other sequence elements by open boxes. The open reading frame within element 3'-E is indicated by the grey arrow.
218x152mm (300 x 300 DPI)

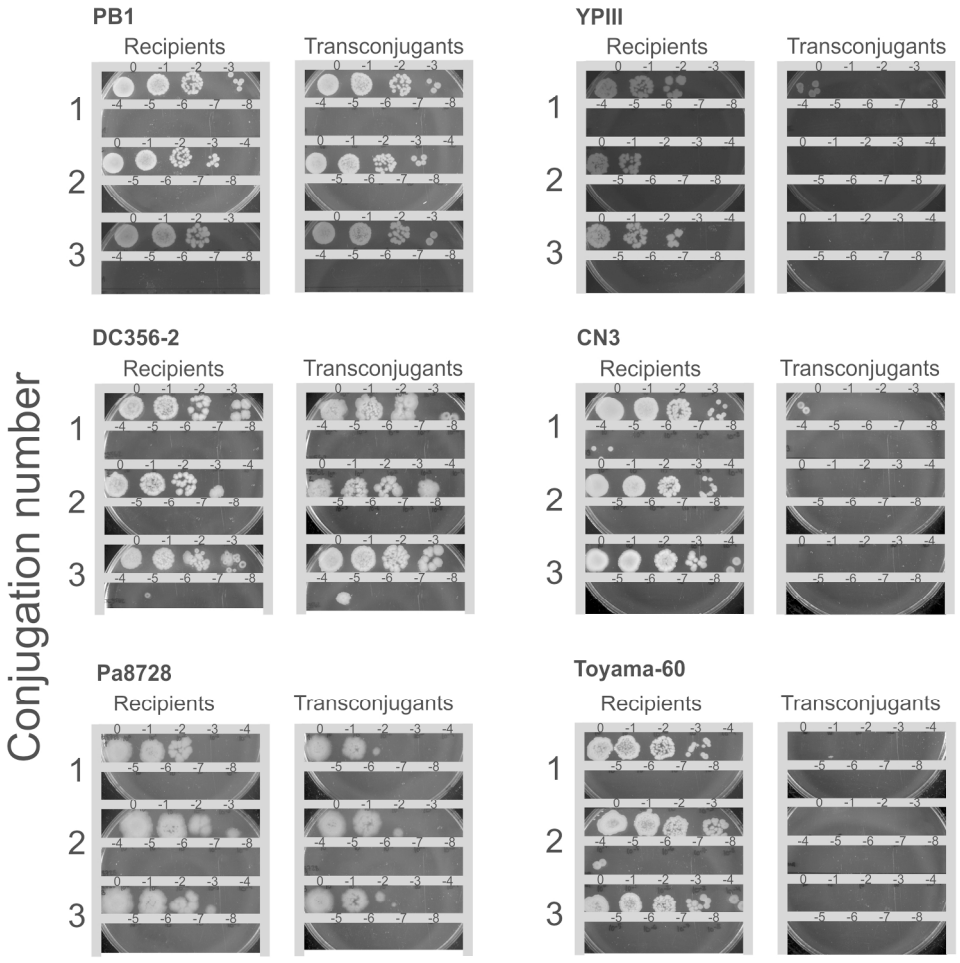

Figure 3. Conjugation experiments showing the serial 10-fold dilution plating of 5 µl drops on selective CIN-agar plates without or with Clm allowing the growth from the mating mixture of all the recipient bacteria or only the transconjugant bacteria, respectively. PB1 and DC356-2 represent non-resistant strains, Pa8728 represents intermediate resistant strains, and YPIII, CN3 and Toyama-60, the fully resistant strains.
195x195mm (300 x 300 DPI)