



**HAL**  
open science

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

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

### ► To cite this version:

Katja A Koskela, Laura Mattinen, Laura Kalin-Mänttari, Gilles Vergnaud, Olivier Gorgé, et al.. Generation of a CRISPR database for *Yersinia pseudotuberculosis* complex and role of CRISPR-based immunity in conjugation.. *Environmental Microbiology*, 2015, pp.12816. 10.1111/1462-2920.12816 . hal-01152593

**HAL Id: hal-01152593**

**<https://ensta-paris.hal.science/hal-01152593>**

Submitted on 19 May 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



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

Journal:	<i>Environmental Microbiology and Environmental Microbiology Reports</i>
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-Mänttari, 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

SCHOLARONE™  
Manuscripts

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

3

4 Katja A Koskela<sup>1</sup>, Laura Mattinen<sup>2</sup>, Laura Kalin-Mänttari<sup>1,2</sup>, Gilles Vergnaud<sup>3,4,5</sup>, Olivier Gorgé<sup>3,4,6</sup>, Simo  
5 Nikkari<sup>1</sup>, and Mikael Skurnik<sup>2,7\*</sup>

6

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

12

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

15

## 16 **Abstract**

17 The CRISPR/Cas system is used by bacteria and archaea against invading conjugative plasmids or  
18 bacteriophages. Central to this immunity system are genomic CRISPR loci that contain fragments of  
19 invading DNA. These are maintained as spacers in the CRISPR loci between direct repeats and the  
20 spacer composition in any bacterium reflects its evolutionary history. We analyzed the CRISPR locus  
21 sequences of 335 *Yersinia pseudotuberculosis* complex strains. Altogether 1902 different spacer  
22 sequences were identified and these were used to generate a database for the spacer sequences. Only  
23 ~10 % of the spacer sequences found matching sequences. In addition, surprisingly few spacers were  
24 shared by *Y. pestis* and *Y. pseudotuberculosis* strains. Interestingly 32 different protospacers were  
25 present in the conjugative plasmid pYptb32953. The corresponding spacers were identified from 35  
26 different *Y. pseudotuberculosis* strains indicating that these strains had encountered pYptb32953  
27 earlier. In conjugation experiments pYptb32953-specific spacers generally prevented conjugation with  
28 spacer-positive and spacer-free strains. However, some strains with one to four spacers were invaded  
29 by pYptb32953 and some spacer-free strains were fully resistant. Also some spacer-positive strains  
30 were intermediate resistant to conjugation. This suggests that one or more other defense systems are  
31 determining conjugation efficiency independent of the CRISPR/Cas system.

32

33

## 34 **Introduction**

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

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

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

61 The CRISPR-Cas (Clustered regularly interspaced short palindromic repeat – CRISPR associated  
62 genes) system is a RNA-based immune system which regulates invasions of plasmids and viruses in  
63 bacteria and archaea. The functional mechanisms of CRISPR and its whole biological significance are  
64 still not fully known (Garneau *et al.*, 2010, Hale *et al.*, 2009, Barrangou *et al.*, 2007, Bolotin *et al.*,  
65 2005, Pourcel *et al.*, 2005, Makarova *et al.*, 2011, Sorek *et al.*, 2013). CRISPRs are constructed from a  
66 chain of 21-47 bp repeated sequences (called direct repeats, DR) and in between DRs are unique spacer  
67 sequences. These spacers represent foreign DNA originating predominantly from bacteriophages and  
68 plasmids. A leader sequence is located at the 5'-end of the CRISPR and usually the *cas*-genes are  
69 located upstream of the leader of one of the CRISPR loci (Bolotin *et al.*, 2005, Karginov & Hannon,  
70 2010, Sontheimer & Marraffini, 2010, Pourcel *et al.*, 2005). The three main types of CRISPR-Cas  
71 systems differ in the composition of *cas* genes and in the mechanisms of CRISPR RNA (crRNA)  
72 processing and interference (Makarova *et al.*, 2011, Makarova *et al.*, 2013, Wiedenheft *et al.*, 2012).  
73 *Yersinia* contain the subtype I-F CRISPR-Cas system (Haft *et al.*, 2005, Makarova *et al.*, 2011) and the

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

76 When a prokaryote comes into contact with foreign DNA, the host may integrate a fragment of this  
77 DNA, known as a protospacer, into the CRISPR locus as a new spacer. Earlier studies show that  
78 approximately 45 % of bacteria and nearly all of archaea contain a CRISPR-Cas system (Grissa *et al.*,  
79 2007a, Pourcel & Drevet, 2013). The new spacers are acquired at the leader proximal end, such that  
80 leader distal spacers are older, thus often shared between more isolates (Pourcel *et al.*, 2005, Barrangou  
81 *et al.*, 2007). Due to their high diversity, the CRISPR sequences have been used for typing (Shariat &  
82 Dudley, 2014), for example, for *Mycobacterium tuberculosis* (Kamerbeek *et al.*, 1997), *Campylobacter*  
83 (Schouls *et al.*, 2003) *Streptococcus thermophilus* (Horvath *et al.*, 2008), *Escherichia coli* (Delannoy *et al.*,  
84 2012a, Delannoy *et al.*, 2012b, Touchon *et al.*, 2011, Diez-Villasenor *et al.*, 2010), *Salmonella*  
85 *enterica* (Liu *et al.*, 2011, Fabre *et al.*, 2012) and also for *Erwinia amylovora* (McGhee & Sundin,  
86 2012, Rezzonico *et al.*, 2011). In *Yersinia* there are three loci named as YP1, YP2 and YP3 of which  
87 the YP1-locus was initially used as a variable number tandem repeat (VNTR) marker (Le Flèche *et al.*,  
88 2001, Pourcel *et al.*, 2005).

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

## 92 **Materials and Methods**

93 **Bacterial isolates.** A total of 76 *Y. pseudotuberculosis*, ten *Y. similis* and four *Y. wautersii* isolates  
94 from the Skurnik laboratory strain collection were analyzed in this study (Table S1). These isolates  
95 were selected to cover the largest possible geographic area, host range and to represent as many of the  
96 21 serotypes as possible. Altogether 83 of the 90 STs (Laukkanen-Ninios *et al.*, 2011) were  
97 represented, each with a single isolate except for ST3, ST14 and ST43 that were represented by three,  
98 four and three isolates, respectively. In addition, sequence data of 40 *Y. pseudotuberculosis* strains and  
99 195 *Y. pestis* strains from earlier investigations were included ((Pourcel *et al.*, 2005, Cui *et al.*, 2008)  
100 and Vergnaud&Gorgé, unpublished) as well as CRISPR loci sequences of published complete genome  
101 sequences of 4 *Y. pseudotuberculosis* and 6 *Y. pestis* strains (Table S1).

102 **Culture conditions.** The bacterial strains were grown in lysogeny broth (LB) (Bertani, 2004); the  
103 *Yersinia* strains at 20-22 °C (RT) unless otherwise mentioned and *E. coli* strains at 37 °C. LB  
104 supplemented with 1.5% Bacto Agar (LA) was used for solid cultures. *Yersinia* selective CIN-agar  
105 plates (Oxoid) were used in conjugation experiments. When required, appropriate antibiotics were  
106 added as follows: chloramphenicol (Cm) 20 µg/ml, nalidixic acid (Nal) 100 µg/ml, kanamycin (Kan)  
107 100 µg/ml and diaminopimelic acid (Dap) 0.3 mM.

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

110 **Sequencing of the CRISPR loci.** The three CRISPR loci of *Y. pseudotuberculosis* complex were  
111 targeted by PCR based on previously published *Y. pestis* CRISPR primer sequences (Le Flèche et al.,  
112 2001, Pourcel et al., 2005). New primers were designed for CRISPR YP2 as the earlier published *Y.*  
113 *pestis* primers failed with many *Y. pseudotuberculosis* strains. The CRISPR loci-specific primers are  
114 presented in Table S2. PCR reactions were run in a final volume of 50 µl containing 50 pmol of each  
115 primer, 5 µl of 10 x Dynazyme II buffer, 200 µM of each dNTP, 1 U of Dynazyme II, and 150 ng of  
116 template DNA. PCR consisted of the following steps: 94°C for 3 min, 32 cycles of denaturation at  
117 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  
118 72°C for 10min. PCR amplified fragments were visualized after agarose gel electrophoresis (1.2 %  
119 agarose) by ethidiumbromide staining. The DNA fragments were sent to Institute for Molecular  
120 Medicine Finland (FIMM) core facility for sequencing after Exonuclease I (Neo Lab) and Shrimp  
121 Alkaline Phosphatase (Promega) treatment. The fragments were sequenced using the Applied  
122 Biosystems Dye Terminator Kit (BigDye v.3.1) and ABI 3730xl DNA Analyzer. The CRISPR-loci  
123 specific primers were used as sequencing primers from the fragment ends and internal primers were  
124 designed for sequencing long PCR fragments.

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

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

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

137 **Construction of pYptb32953::cat and pTM100-CRISPR.** Primers specific for pYptb32953 (Acc. no  
138 BX936400.1), the 27 kb cryptic plasmid of *Y. pseudotuberculosis* IP32953 (Table S2) were used to  
139 amplify a 797 bp fragment of pYptb32953 from a plasmid miniprep template. The PCR fragment was  
140 purified and digested with EcoRI followed by ligation with EcoRI-digested and SAP-treated suicide  
141 vector pSW23T (Demarre *et al.*, 2005). The ligation mixture was electroporated into *E. coli* strain  
142  $\omega$ 7249 that is Kan<sup>R</sup> (Babic *et al.*, 2008). Transformants carrying the correct insert were identified by  
143 PCR and the isolated plasmid named as pSW23T-pIP was further confirmed by restriction digestions.  
144 The suicide construct was introduced to *Y. pseudotuberculosis* IP32953 by conjugation from *E. coli*  
145  $\omega$ 7249/pSW23T-pIP and Clm<sup>R</sup> transconjugants were selected with LA-Clm plates where the donor was  
146 unable to grow due to its requirement for diaminopimelic acid (Dap). One of the Clm<sup>R</sup> transconjugants  
147 was named as IP32953/pYptb32953::cat and used as a donor to introduce the tagged plasmid into *E.*  
148 *coli* strains PM191NaR, a Nal<sup>R</sup> spontaneous derivative of PM191 (Meacock & Cohen, 1980) to obtain  
149 *E. coli* PM191NaR/pYptb32953::cat, and to strain  $\omega$ 7249 to obtain *E. coli*  $\omega$ 7249/pYptb32953::cat.  
150 pTM100-CRISPR was constructed by cloning a PCR-amplified 909-bp DNA-fragment of pYptb32953  
151 (nucleotides 13,002-13,910; for PCR-primers, see Table S2) into EcoRV-site of pTM100 (Michiels &  
152 Cornelis, 1991). pTM100-CRISPR was electroporated into *E. coli* strain  $\omega$ 7249. pTM100-waaF  
153 (Noszczyńska *et al.*, 2014) was used as spacer-free control plasmid in mobilization experiments.

154 **Conjugation frequency assays.** The *E. coli*  $\omega$ 7249/pYptb32953::cat,  $\omega$ 7249/pTM100-CRISPR and  
155  $\omega$ 7249/pTM100-waaF strains were used as donor strains to determine conjugation/mobilization  
156 frequencies into a set of *Y. pseudotuberculosis* strains. The donor bacteria were grown in LB-Kan-Clm-  
157 Dap at 37°C for 16 h, the culture was diluted 1:10 in fresh medium and incubated for an additional 3 h,  
158 washed and resuspended into PBS to OD<sub>600</sub> of ~1.0. The recipient bacteria were grown in LB at 22°C  
159 for 16 h, the culture was diluted 1:10 in fresh medium and incubated for an additional 3 h, washed and  
160 resuspended into PBS to OD<sub>600</sub> of ~1.0. For each recipient strain, three parallel matings were prepared.

161 Equal amounts of the donor and recipient suspensions were mixed and 100  $\mu$ l aliquots were pipetted in  
162 the middle of three parallel LA-plates supplemented with Dap but without antibiotics. The plates were  
163 incubated at 37 °C for 16 h. The bacteria on the plates' surface were resuspended into 1 ml of PBS. 200  
164  $\mu$ l aliquots were recovered from each and diluted with PBS to OD<sub>600</sub> of 0.2. The concentrations of  
165 donor, recipient and transconjugant bacteria in these mating mixtures were determined by pipetting 5  $\mu$ l  
166 drops of 10<sup>0</sup> – 10<sup>-8</sup> diluted mixtures on LA-Kan-Clm-Dap plates (for donor counts), CIN plates (for  
167 total recipient counts) and CIN-Clm plates (for transconjugant counts). The donor plates were  
168 incubated for 24 h at 37°C and the recipient and transconjugant plates at 22°C for 48 h. The colonies in  
169 the last dilutions showing growth were counted. Conjugation frequencies were expressed as ratios  
170 between the transconjugant and recipient concentrations.

## 171 **Results**

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

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

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

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

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

203 To visualize possible evolutionary relationships between the strains based on the organization of the  
204 spacer sequences, the *Y. pseudotuberculosis*, *Y. similis* and *Y. wautersii* strain-specific spacer patterns  
205 were manually aligned and the alignments are shown in Table S6. If all spacers present in the most  
206 recent common ancestor (MRCA) of *Y. pseudotuberculosis* complex had been subsequently  
207 maintained, present day strains should have the same root (most ancient) spacer. This is not the case,  
208 instead there were several root spacers both in the YP1 and YP3 loci. The most likely explanation is  
209 that older spacers were randomly lost. Furthermore, gaps had to be introduced to the spacer patterns to  
210 maximize their alignment. When the spacer pattern alignments were used for grouping of the strains we  
211 found that the phylogenies of YP1 and YP3 loci seem not to be congruent. In fact, the alignments  
212 indicated that the spacers had accumulated independently to these main storage loci. Table S6, parts A  
213 and B, show the strains sorted based on YP1 and YP3 alignments, respectively. As an example of this,  
214 the YP1 and YP3 alleles of selected strains are shown in Table 1. For instance, the ancient YP1 alleles  
215 539.173.177.-- are associated either with ancient YP3 alleles 507.1238.-- or 1149.1332.-- or 1149.1199.  
216 (Table 1, top). Conversely, each of these is associated with two or more very different YP1 alleles  
217 (Table 1, bottom). This data strongly suggests that horizontal gene transfer (HGT) and recombination  
218 between the CRISPR loci has occurred within the *Y. pseudotuberculosis* complex, however, with the  
219 present data we cannot evaluate the full extent of such mosaicisms. Within the strain groups presented  
220 in Table S6 one can observe plenty of examples of possible recombination events and reassortments

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

230 Among the 84 *Y. pseudotuberculosis* strains with YP1 sequences the most prevalent ancient or root  
231 spacer in the YP1 locus was #39 (present in 14 strains), followed by spacers #103, #40, #539, #403,  
232 #76 and #581 (present in 13, 9, 6, 4, 3 and 3 strains, respectively). Eight different ancient spacers were  
233 shared by two strains and 15 strains had unique ancient spacers. Sequence information for the YP1  
234 locus was not obtained for 50 of the 124 *Y. pseudotuberculosis* and 10 *Y. similis* strains.

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

250 The YP3-locus spacer comparison is presented in Table S6 part B. Based on the identity of the most  
251 ancestral spacer the strains could be grouped into >10 groups. Spacers #507, #1149 and #1111 define  
252 the largest groups with 38, 31 and 12 strains, respectively. The other spacer groups #511, #1132,  
253 #1156, #1199, #1589, #1616, #1622 and #1853 included 2-5 strains each. The remaining strains had  
254 either a sporadic most ancestral spacer (7 strains) or we did not get a PCR-product or sequence from  
255 the locus (23 strains).

256 Interestingly, seventy-two spacers were present in two different CRISPR loci. In one instance, this  
257 peculiarity was observed in a single strain. Spacer #808 was found in both the YP1 and YP3 locus of  
258 strain No-151. Duplications occurred, for example spacer #257 was found in strains MW101-1 and TE-  
259 93081 as a tandem repeat duplicate in the YP1 locus. Another example is the spacer pair #1348.1349  
260 that is present twice in strain DC356-2. Also spacer #1 is present twice in strain BB1152 (Table S6).

#### 261 ***Relationships between spacer based grouping and sequence types***

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

#### 278 ***Origin of spacers***

279 BLASTN searches revealed that a number of spacer sequences showed similarity to various plasmid  
280 and bacteriophage sequences (Tables S8 and S9). It was interesting to notice that a few spacers (e.g.  
281 #585, #283, #82, #1206, #1154 and #1001) present in *Yersinia* species had matches to plasmids. Spacer  
282 #82 shows 97 % identity (one base pair difference) to *Y. enterocolitica* plasmid pYE854 and *Y.*  
283 *pseudotuberculosis* IP31758 59 kb plasmid. In Table S9 spacers that have similarities with different  
284 bacteriophage sequences are shown. There was good identity to e.g. *Enterobacteria*, *Erwinia*,  
285 *Escherichia*, *Salmonella* and *Burkholderia* phages. In Table S10 spacer sequences were compared to  
286 whole genome sequences. The bacterial species earlier seen in Table S9 (bacteriophage hits) can also  
287 be seen in this table. Spacer #1697 appears in many bacterial species simply because this spacer occurs  
288 in the highly conserved 16S ribosomal RNA gene. In many instances the spacers were located in  
289 prophage-like elements similar to *Yersinia* –specific spacers (see below).

290 ***Yersinia* –specific spacers.** Some spacers were present in the genomes or plasmids of other *Yersinia*  
291 species. Altogether 40 spacers showed significant similarity to sequences present in *Y. enterocolitica*  
292 8081 genome (Table 2). The 8081 genome carries four prophage-like elements (Thomson *et al.*, 2006)  
293 and 38 of the 40 spacer sequences were located within two of them,  $\phi$ YE98 (22 spacers) and  $\phi$ YE250  
294 (16 spacers). A few strains carried two 8081 prophage-specific spacers.

295 Most spacers identified from *Y. pestis* were not shared with *Y. pseudotuberculosis*. The ones that are  
296 shared are shown in Tables S11 and S12. Table S13 is the conversion table for spacer nomenclature  
297 from previous studies to the present database (Cui *et al.*, 2008, Riehm *et al.*, 2012). Specifically, in the  
298 YP1 locus, spacers #403-405 were shared by a few *Y. pseudotuberculosis* strains, while in the YP3  
299 locus spacer #507 seems to be present in all *Y. pestis* strains and is also common in *Y.*  
300 *pseudotuberculosis* (Tables S6 and S11). Interestingly, spacers #257, #1901 and #1902 are identical  
301 except for the one and two extra Gs present in the two latter ones, respectively (Table S4). Spacer  
302 #1901 is very common in the *Y. pestis* YP2 locus while #257 appears in YP1 locus in 11 strains of *Y.*  
303 *pseudotuberculosis* and in two of the strains it is present as duplicate. Furthermore, #257 appears in the  
304 YP3 locus in seven *Y. pseudotuberculosis* strains and twice in one of the strains. As mentioned before,  
305 *Y. pseudotuberculosis* strains rarely contained more than two spacers in YP2 loci. In contrast, *Y. pestis*  
306 YP2 locus usually carries three to six spacers or more.

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

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

311 **pYptb32953.** Altogether 32 unique spacers for the 27,702 bp cryptic plasmid pYptb32953 of *Y.*  
312 *pseudotuberculosis* IP32953 (Chain *et al.*, 2004) were identified in 34 strains. The distribution of the  
313 spacers along the plasmid sequence is shown in Table 3. No significant distribution bias can be  
314 detected. A majority (22) of the spacers map to the forward (+) and 10 map to the reverse (-) strand of  
315 the plasmid (Table 3). Spacer #1362 has two matches in the plasmid, one is a 100% match to nt 16931  
316 (-) strand and the other a 31/32 (97%) match to nt 23321 (+) strand. Altogether 30 different strains had  
317 a spacer sequence with 100 % identity to the plasmid pYptb32953 sequence, and additionally a few  
318 spacers with some mismatches were identified. To see whether the spacer-carrying strains would reject  
319 the pYptb32953 plasmid and the spacers-free strains would accept it in conjugation experiments, we  
320 tagged the plasmid with a *cat*-gene (see M&M). We first demonstrated that pYptb32953 is indeed a  
321 self-conjugative plasmid as predicted based on its annotated sequence (Chain *et al.*, 2004). The  
322 pYptb32953::*cat* transferred itself efficiently from IP32953 to *E. coli* strain PM191NaR (data not  
323 shown). The conjugation frequencies to 31 different *Y. pseudotuberculosis* strains were determined  
324 (Table 4). Examples of the conjugation experiments are shown in Figure 3. Among the strains we  
325 observed three levels of restriction to pYptb32953::*cat* conjugation: (i) non-resistant group to which the  
326 plasmid transferred without any apparent restriction and under the experimental conditions used 30 –  
327 100 % of recipients were transformed, (ii) intermediate resistant group where 1 to 20% of recipients  
328 were transformed and (iii) fully resistant group with less than 0.01 % transformants. While 11 of 12  
329 among the fully resistant strains carried a pYptb32953-specific spacer ten spacer-carrying strains were  
330 among the 13 strains in the non-resistant group. Among the 4 spacer-free strains tested 3 were non-  
331 resistant and one, YPII, was fully resistant (Table 4).

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

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

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

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

## 359 **Discussion**

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

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

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

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

386 Previous studies have shown that the CRISPR variability may be used for typing bacterial species, even  
387 though the CRISPR sequence diversity was not as wide as in the *Y. pseudotuberculosis* complex  
388 (Riehm et al., 2012). For example, Fabre and others concluded that the CRISPR spacer content in  
389 *Salmonella* correlated with MLST and serotyping results, and they indicated that CRISPR analysis may  
390 be a powerful tool for molecular typing of *Salmonella* isolates (Shariat et al., 2013a, Shariat et al.,  
391 2013b, Fabre et al., 2012). It was also shown that *E. coli* CRISPR typing combined with MLST  
392 analysis could differentiate strains from a single clonal group (Touchon et al., 2011). CRISPR has also  
393 proven to be a good typing tool for the clonal *Y. pestis* and hypothetical evolutionary models have been  
394 created based on the CRISPR spacer arrays (Cui et al., 2008). This has to be treated with utmost care as  
395 our present results revealed very big differences between the *Y. pseudotuberculosis* strains and  
396 indications that evolution within *Y. pseudotuberculosis* might not be clonal. This method may be very  
397 useful for forensic applications, however, this would require an extensive reference collection. We

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

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

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

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

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

#### 431 ***Final conclusions***

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

#### 442 ***Acknowledgements***

443 This work is part of the European Biodefence project EDA B-1324-ESM4-GP. Scientific contribution  
444 to this work has been accomplished by representatives from France and Finland. Work by GV andOG  
445 was supported by the French ministry of defense. Anita Liljegren and Juha Laitinen are thanked for  
446 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  $\mu$ l 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.

471 **References**

472

473

- 474 Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule & E. Carniel, (1999) *Yersinia pestis*, the cause of  
475 plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* **96**:  
476 14043-14048.
- 477 Aleksic, S., J. Bockemühl & H.H. Wuthe, (1995) Epidemiology of *Y. pseudotuberculosis* in Germany, 1983-1993.  
478 In: *Yersiniosis: Present and Future*. G. Ravagnan & C. Chiesa (eds). Postfach/CH-4009 Basel/Switzerland:  
479 Karger, pp. 55-58.
- 480 Almendros, C., F.J. Mojica, C. Diez-Villasenor, N.M. Guzman & J. Garcia-Martinez, (2013) CRISPR-Cas functional  
481 module exchange in *Escherichia coli*. *MBio* **5**: e00767-00713.
- 482 Altschul, S.F., W. Gish, W. Miller, E.W. Myers & D.J. Lipman, (1990) Basic local alignment search tool. *J. Mol.*  
483 *Biol.* **215**: 403-410.
- 484 Babic, A., A.M. Guerout & D. Mazel, (2008) Construction of an improved RP4 (RK2)-based conjugative system.  
485 *Res. Microbiol.* **159**: 545-549.
- 486 Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero & P. Horvath, (2007)  
487 CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.
- 488 Bertani, G., (2004) Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.*  
489 **186**: 595-600.
- 490 Blouin, Y., G. Cazajous, C. Dehan, C. Soler, R. Vong, M.O. Hassan, Y. Hauck, C. Boulais, D. Andriamanantena, C.  
491 Martinaud, E. Martin, C. Pourcel & G. Vergnaud, (2014) Progenitor "*Mycobacterium canettii*" clone  
492 responsible for lymph node tuberculosis epidemic, Djibouti. *Emerg Infect Dis* **20**: 21-28.
- 493 Bogdanovich, T., E. Carniel, H. Fukushima & M. Skurnik, (2003) Use of O-antigen gene cluster-specific PCRs for  
494 the identification and O-genotyping of *Yersinia pseudotuberculosis* and *Yersinia pestis*. *J. Clin. Microbiol.*  
495 **41**: 5103-5112.
- 496 Bolotin, A., B. Quinquis, A. Sorokin & S.D. Ehrlich, (2005) Clustered regularly interspaced short palindrome  
497 repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.
- 498 Cady, K.C., J. Bondy-Denomy, G.E. Heussler, A.R. Davidson & G.A. O'Toole, (2012) The CRISPR/Cas adaptive  
499 immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered  
500 phages. *J. Bacteriol.* **194**: 5728-5738.
- 501 Ch'ng, S.L., S. Octavia, Q. Xia, A. Duong, M.M. Tanaka, H. Fukushima & R. Lan, (2011) Population structure and  
502 evolution of pathogenicity of *Yersinia pseudotuberculosis*. *Appl. Environ. Microb.* **77**: 768-775.
- 503 Chain, P.S., E. Carniel, F.W. Larimer, J. Lamerdin, P.O. Stoutland, W.M. Regala, A.M. Georgescu, L.M. Vergez,  
504 M.L. Land, V.L. Motin, R.R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V.  
505 Chenal-Francisque, B. Souza, D. Dacheux, J.M. Elliott, A. Derbise, L.J. Hauser & E. Garcia, (2004) Insights  
506 into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia*  
507 *pseudotuberculosis*. *Proc Natl Acad Sci U S A* **101**: 13826-13831.
- 508 Chain, P.S., P. Hu, S.A. Malfatti, L. Radnedge, F. Larimer, L.M. Vergez, P. Worsham, M.C. Chu & G.L. Andersen,  
509 (2006) Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene  
510 reduction in an emerging pathogen. *J. Bacteriol.* **188**: 4453-4463.
- 511 Chauvaux, S., M.A. Dillies, M. Marceau, M.L. Rosso, S. Rousseau, I. Moszer, M. Simonet & E. Carniel, (2011) In  
512 silico comparison of *Yersinia pestis* and *Yersinia pseudotuberculosis* transcriptomes reveals a higher  
513 expression level of crucial virulence determinants in the plague bacillus. *Int. J. Med. Microbiol.* **301**:  
514 105-116.

- 515 Cui, Y., Y. Li, O. Gorge, M.E. Platonov, Y. Yan, Z. Guo, C. Pourcel, S.V. Dentovskaya, S.V. Balakhonov, X. Wang, Y.  
516 Song, A.P. Anisimov, G. Vergnaud & R. Yang, (2008) Insight into microevolution of *Yersinia pestis* by  
517 clustered regularly interspaced short palindromic repeats. *PLoS ONE* **3**: e2652.
- 518 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.  
519 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,  
520 D. Falush, F. Balloux, M. Achtman, Y. Song & R. Yang, (2013) Historical variations in mutation rate in an  
521 epidemic pathogen, *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* **110**: 577-582.
- 522 Delannoy, S., L. Beutin, Y. Burgos & P. Fach, (2012a) Specific detection of enteroaggregative hemorrhagic  
523 *Escherichia coli* O104:H4 strains by use of the CRISPR locus as a target for a diagnostic real-time PCR. *J.*  
524 *Clin. Microbiol.* **50**: 3485-3492.
- 525 Delannoy, S., L. Beutin & P. Fach, (2012b) Use of clustered regularly interspaced short palindromic repeat  
526 sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of  
527 serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by real-time PCR.  
528 *J. Clin. Microbiol.* **50**: 4035-4040.
- 529 Demarre, G., A.M. Guerout, C. Matsumoto-Mashimo, D.A. Rowe-Magnus, P. Marliere & D. Mazel, (2005) A new  
530 family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid  
531 (IncPalpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.* **156**:  
532 245-255.
- 533 Diez-Villasenor, C., C. Almendros, J. Garcia-Martinez & F.J. Mojica, (2010) Diversity of CRISPR loci in *Escherichia*  
534 *coli*. *Microbiology* **156**: 1351-1361.
- 535 Fabre, L., J. Zhang, G. Guigon, S. Le Hello, V. Guibert, M. Accou-Demartin, S. de Romans, C. Lim, C. Roux, V.  
536 Passet, L. Diancourt, M. Guibourdenche, S. Issenhuth-Jeanjean, M. Achtman, S. Brisse, C. Sola & F.X.  
537 Weill, (2012) CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella*  
538 infections. *PLoS ONE* **7**: e36995.
- 539 Garneau, J.E., M.E. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A.H.  
540 Magadan & S. Moineau, (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and  
541 plasmid DNA. *Nature* **468**: 67-71.
- 542 Grissa, I., G. Vergnaud & C. Pourcel, (2007a) The CRISPRdb database and tools to display CRISPRs and to  
543 generate dictionaries of spacers and repeats. *BMC bioinformatics* **8**: 172.
- 544 Grissa, I., G. Vergnaud & C. Pourcel, (2007b) CRISPRFinder: a web tool to identify clustered regularly  
545 interspaced short palindromic repeats. *Nucl. Acids Res.* **35**: W52-57.
- 546 Haft, D.H., J. Selengut, E.F. Mongodin & K.E. Nelson, (2005) A guild of 45 CRISPR-associated (Cas) protein  
547 families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology* **1**:  
548 e60.
- 549 Hale, C.R., P. Zhao, S. Olson, M.O. Duff, B.R. Graveley, L. Wells, R.M. Terns & M.P. Terns, (2009) RNA-guided  
550 RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.
- 551 Harbeck, M., L. Seifert, S. Hansch, D.M. Wagner, D. Birdsell, K.L. Parise, I. Wiechmann, G. Grupe, A. Thomas, P.  
552 Keim, L. Zoller, B. Bramanti, J.M. Riehm & H.C. Scholz, (2013) *Yersinia pestis* DNA from skeletal remains  
553 from the 6(th) century AD reveals insights into Justinianic Plague. *PLoS Pathog* **9**: e1003349.
- 554 Held, N.L., A. Herrera & R.J. Whitaker, (2013) Reassortment of CRISPR repeat-spacer loci in *Sulfolobus*  
555 *islandicus*. *Environ Microbiol.*
- 556 Horvath, P., D.A. Romero, A.C. Coute-Monvoisin, M. Richards, H. Deveau, S. Moineau, P. Boyaval, C. Fremaux &  
557 R. Barrangou, (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J.*  
558 *Bacteriol.* **190**: 1401-1412.
- 559 Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen,  
560 R. Shaw, M. Goyal & J. van Embden, (1997) Simultaneous detection and strain differentiation of  
561 *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 Noszczyńska, M., K. Kasperkiewicz, K.A. Duda, J. Podchorodecka, 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) Occurrence, 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.

- 609 Riehm, J.M., G. Vergnaud, D. Kiefer, T. Damdindorj, O. Dashdavaa, T. Khurelsukh, L. Zoller, R. Wolfel, P. Le  
610 Flèche & H.C. Scholz, (2012) *Yersinia pestis* lineages in Mongolia. *PLoS One* **7**: e30624.
- 611 Savin, C., L. Martin, C. Bouchier, S. Filali, J. Chenau, Z. Zhou, F. Becher, H. Fukushima, N.R. Thomson, H.C. Scholz  
612 & E. Carniel, (2014) The *Yersinia pseudotuberculosis* complex: Characterization and delineation of a  
613 new species, *Yersinia wautersii*. *Int. J. Med. Microbiol.* **304**: 452-463.
- 614 Schouls, L.M., S. Reulen, B. Duim, J.A. Wagenaar, R.J. Willems, K.E. Dingle, F.M. Colles & J.D. Van Embden,  
615 (2003) Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism,  
616 multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and  
617 recombination. *J. Clin. Microbiol.* **41**: 15-26.
- 618 Shariat, N. & E.G. Dudley, (2014) CRISPRs: molecular signatures used for pathogen subtyping. *Appl. Environ.*  
619 *Microb.* **80**: 430-439.
- 620 Shariat, N., M.K. Kirchner, C.H. Sandt, E. Trees, R. Barrangou & E.G. Dudley, (2013a) Subtyping of *Salmonella*  
621 *enterica* serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship  
622 between CRISPR-MVLST and PFGE results. *J. Clin. Microbiol.* **51**: 2328-2336.
- 623 Shariat, N., C.H. Sandt, M.J. DiMarzio, R. Barrangou & E.G. Dudley, (2013b) CRISPR-MVLST subtyping of  
624 *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Heidelberg and application in  
625 identifying outbreak isolates. *BMC Microbiol* **13**: 254.
- 626 Skurnik, M., A. Peippo & E. Ervelä, (2000) Characterization of the O-antigen gene clusters of *Yersinia*  
627 *pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague  
628 bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. *Mol.*  
629 *Microbiol.* **37**: 316-330.
- 630 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,  
631 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  
632 genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Research* **11**:  
633 179-197.
- 634 Sontheimer, E.J. & L.A. Marraffini, (2010) Microbiology: slicer for DNA. *Nature* **468**: 45-46.
- 635 Sorek, R., C.M. Lawrence & B. Wiedenheft, (2013) CRISPR-mediated adaptive immune systems in bacteria and  
636 archaea. *Annu. Rev. Biochem.* **82**: 237-266.
- 637 Souza, R.A., A. Pitondo-Silva, D.P. Falcao & J.P. Falcao, (2010) Evaluation of four molecular typing  
638 methodologies as tools for determining taxonomy relations and for identifying species among *Yersinia*  
639 isolates. *J. Microbiol. Meth.* **82**: 141-150.
- 640 Sprague, L.D., H.C. Scholz, S. Amann, H.J. Busse & H. Neubauer, (2008) *Yersinia similis* sp. nov. *Int. J. Sys. Evol.*  
641 *Microbiol.* **58**: 952-958.
- 642 Staden, R., (1996) The Staden sequence analysis package. *Mol Biotechnol* **5**: 233-241.
- 643 Tauxe, R.V., (2004) Salad and pseudoappendicitis: *Yersinia pseudotuberculosis* as a foodborne pathogen. *J.*  
644 *Infect. Dis.* **189**: 761-763.
- 645 Thomson, N.R., S. Howard, B.W. Wren, M.T. Holden, L. Crossman, G.L. Challis, C. Churcher, K. Mungall, K.  
646 Brooks, T. Chillingworth, T. Feltwell, Z. Abdellah, H. Hauser, K. Jagels, M. Maddison, S. Moule, M.  
647 Sanders, S. Whitehead, M.A. Quail, G. Dougan, J. Parkhill & M.B. Prentice, (2006) The complete genome  
648 sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain  
649 8081. *PLoS Genet* **2**: e206.
- 650 Touchon, M., S. Charpentier, O. Clermont, E.P. Rocha, E. Denamur & C. Branger, (2011) CRISPR distribution  
651 within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J.*  
652 *Bacteriol.* **193**: 2460-2467.
- 653 Wiedenheft, B., S.H. Sternberg & J.A. Doudna, (2012) RNA-guided genetic silencing systems in bacteria and  
654 archaea. *Nature* **482**: 331-338.
- 655 Wiedenheft, B., E. van Duijn, J.B. Bultema, S.P. Waghmare, K. Zhou, A. Barendregt, W. Westphal, A.J. Heck, E.J.  
656 Boekema, M.J. Dickman & J.A. Doudna, (2011) RNA-guided complex from a bacterial immune system

657 enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* **108**: 10092-  
658 10097.

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

661

662

For Peer Review Only

Table 1. YP1 and YP3 spacer alignments of selected *Y. pseudotuberculosis* strains<sup>a</sup>

Strain	YP1 locus	YP3 locus
Toyama60	104.801.802.83.173-177.81-70.1842.20.334.27.130.71.263.172.688.73.1827.174.	1148.1118.1200.-----1202.331.1203.261.81.1204.1208.1773-1774.130.1777.361.
MW145-2	-----173-177.130.178.-----17.180.181-188.183.186.187.188.188.	07.1238-1240.1868.1810.848.71.1811.77.1847.712.340.80.1812.1813-1814.171.
Pa3606	-----83.173-177.nnn.-----181-188.183.186.187.	07.1238-1240.1868.1810.848.71.1811.77.1847.712.340.80.1812.
IP31758	-----83.173-177.1868.178.-----840-843.1818.17.180.181-188.183.186.187.861.	1148.1118.1200.1201.1202.331.1203.261.81.1204.1208.1466-1471.808.277.1472.813.1473-1477.
D426	-----83.173-177.1868.178.-----840-843.	1148.1332.-----1203.261.1333-1337.
Y718	-----83.173-177.1868.178.178.840-843.1818.17.180.181-188.183.186.187.	1148.1332.-----1203.261.1333-1337.
Strain	YP3 locus	YP1 locus
RD20	114.1118.1200.1201.1202.331.1203.261.81.1204.1208.1209.188.1207-1212.184.734.1213-1214.288.1338.133.171.1.274.181.266.	81.82.84.80.-----82-84.83-86.85-87.88-88.
No-151	114.1118.1200.1201.1202.331.1203.261.81.1204.1208.-----1466-1471.808.588.184.	104.801.802.803.804.850.277.831-833.-----808-809.810-813.333.20.71.814-814.
IP31758	114.1118.1200.1201.1202.331.1203.261.81.1204.1208.1466-1468.1466-1471.808.277.1472.813.1473-1477.	-----83.173-177.1868.178.-----840-843.1818.17.180.181-188.183.186.187.861.
Toyama60	114.1118.1200.-----1202.331.1203.261.81.1204.1208.1773.1774.1778.1777.130.1777.361.	104.801.802.83.173-177.81-70.1842.20.334.27.130.71.263.172.688.73.1827.174.
D426	114.1332.1203.261.1333-1337.	-----83.173-177.1868.178.-----840-843.
Gifu-liver	114.1332.1203.261.1333-1337.331.	-----850.277.831-833.834-838.1820.870-877.1811.878-882.1001.883-887.1848.888-890.
Y718	114.1332.1203.261.1333-1337.	-----83.173-177.1868.178.178.840-843.1818.17.180.181-188.183.186.187.
Y732	114.1332.1203.261.1333-1337.	-----1088-1089.248.248.1070.1071.
WP-931110	07.1238-1240.-----1244.-----1248-1248.1780-1282.1780.	81.87.77.78.88.88.87.80-82.88-92.-----83.84.878.88.88.87-87.888.880.111.861.263.882-882.
Pa3606	07.1238-1240.1241.1242.1243.1244.1244.1244.1248-1248.1868.1810.848.71.1811.77.1847.712.340.80.1812.	83.173-177.nnn.-----181-188.183.186.187.
MW145-2	07.1238-1240.1241.1242.1243.1244.1244.1248-1248.1868.1810.848.71.1811.77.1847.712.340.80.1812.1813-1814.171.	-----173-177.130.178.-----17.180.181-188.183.186.187.188.188.

<sup>a</sup> 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 %
<u>Spacers in prophage <math>\phi</math>YE98 (location in 8081 genome: 981223–1011295)</u>				
1779	WP-931108	YP3	984218	88 (28/32)
1191	<b>79136</b>	YP3	984296	88 (28/32)
986	<b>WP-931110</b>	YP1	986574	97 (31/32)
1193	<b>79136</b>	YP3	987079	91 (29/32)
566	<b>2889, 2895, Y385, Y728, Y729</b>	YP1	987179	97 (31/32)
988	<b>WP-931110</b>	YP1	987234	97 (31/32)
244	PC95-219-1	YP1	988311	91 (29/32)
699	<b>H892-36-87</b>	YP1	991747	91 (29/32)
711	<b>H892-36-87</b>	YP1	991941	91 (29/32)
1471	IP31758, No-151, OK10700, <b>R30</b>	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	<b>R30</b>	YP3	994348	91 (29/32)
55	<b>BB28, Y74</b>	YP1	995024	94 (30/32)
86	<b>F-401-1, Wla658, WP-931110</b>	YP1	1001481	97 (31/32)
50	<b>BB28, Y74</b>	YP1	1004019	88 (28/32)
1629	<b>MW48, R103-2, R626R</b>	YP3	1004236	88 (28/32)
123	<b>H-3831</b>	YP1	1009835	100 (32/32)
79	<b>F-401-1</b>	YP1	1010599	91 (29/32)
647	<b>CN2</b>	YP1	1010603	91 (29/32)
1317	<b>CN3-5end</b>	YP3	1010802	88 (28/32)
846	PT245	YP1	1010919	94 (30/32)
<u>Spacers in prophage <math>\phi</math>YE200 (location in 8081 genome: 1991720-2007210)</u>				
885	<b>R104-2</b>	YP1	1993391	88 (28/32)
<u>Spacers in prophage <math>\phi</math>YE250 (location in 8081 genome: 2503099-2554665)</u>				
1315	<b>CN3-5end</b>	YP3	2527852	88 (28/32)
1681	Pa8728, d54	YP3	2533319	94 (30/32)
1270	B56, <b>No-21</b>	YP3	2533871	97 (31/32)
1256	AZ960106-1	YP3	2534586	100 (32/32)
1662	<b>MW48, R103-2, R626R</b>	YP3	2535837	94 (30/32)
1412	H146-84K, <b>R111, YPIII</b>	YP3	2538051	88 (28/32)
225	OK5466	YP1	2538088	94 (30/32)

570	<b>2889, 2895, Y385, 728, Y729</b>	YP1	2538983	97 (31/32)
1295	<b>CN2, R104-2</b>	YP3	2539577	100 (32/32)
121	<b>H-3831</b>	YP1	2540279	94 (30/32)
889	<b>R104-2</b>	YP1	2542146	100 (32/32)
1457	<b>H892-36-87</b>	YP3	2544550	88 (28/32)
918	<b>R111, YPIII</b>	YP1	2548392	94 (30/32)
821	<b>No-21</b>	YP1	2550610	91 (29/32)
695	<b>H892-36-87</b>	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 <sup>a</sup>	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/gatgggtctcaaaatacgcaccaaaagggaaacg/ <b>GG</b> aaaa	100	Intermediate
790	J51	YP1	O:13	47	5093 (+)	gatttc/aacgaaaaaacgccggtaatgcgtcgattgt/ <b>GG</b> ggac	100	Non-resistant
1285	BB28	YP3	O:2b	53	5312 (+)	gttaaa/agtggggaacctaccggatggaatccgttttcg/ <b>G</b> ctgaa	100	Non-resistant
287	R103-1	YP1	O:3	58	5432 (+)	accoga/caggaaaccgcctcagtgacgccgttgatgc/ <b>tG</b> ttat	100	Non-resistant
	S107	YP1	O:5b	87				--
1178	79136	YP3	O:1b	88	5469 (+)	ctgtta/tgttggggcttgaccacgccagccgtgaccac/ <b>GG</b> Tatt	100	Non-resistant
789	J51	YP1	O:13	47	5751 (+)	tactgg/atggacggcgttttagtgctgtttattgatgag/ <b>G</b> tctgg	100	Non-resistant
1450	H-3831	YP3	O:4a	48	7359 (+)	agtctt/ttcgagtcfaatcatgggaagactatctttatt/ <b>GG</b> cagc	100	Fully resistant
1167	774	YP3	O:4a	32	8228 (-)	gtgagc/gcggttaataccccccgcattagtaaatgaa/ <b>GG</b> tgat	88	Fully resistant
464	Y722	YP1	unknown	19	9676 (-)	gcgagt/ttatcggggtcggttgcatcactaatgacatt/ <b>GG</b> aaat	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 (+)	ttgggtg/ttttcggatgatgaggcggttattagcactga/ <b>GG</b> tggg	100	Fully resistant
1264	B56	YP3	R	9	10680 (+)	tgatcc/agcatattaaccgcacgatggtattacgctat/ <b>GG</b> cgat	100	Fully resistant
	No.21	YP3	O:1a	86				Intermediate
556	22917/2L	YP1	O:5a	16	12868 (+)	cagcat/gagaactatgtgcatctgttttatccgtcaga/ <b>GG</b> gtgg	100	Intermediate
1820	Gifu-liver	YP1	O:1b	22	12869 (+)	agcatg/agaactatgtgcatctgttttatccgtcagag/ <b>GG</b> tgga	94	Intermediate
1632	R103-2 ( <i>Y. sim</i> )	YP3	O:5b	45	13031 (+)	gttcag/gtcatgatgggttaatacagtcfaatgacagct/ <b>GG</b> cacc	100	Fully resistant
	MW48 ( <i>Y. sim</i> )	YP3	O:9	80				Fully resistant
	R626R ( <i>Y. sim</i> )	YP3	O:9	83				Non-resistant

1528	R103-1	YP3	O:3	58	13276 (+)	aaatga/aacatttaattagaccatggtgggtggctgctc/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/acgacgggtaaacagcaccttgccccagtgcca/GGaaaa	100	Fully resistant
1449	H-3831	YP3	O:4a	48	13453 (+)	catggt/caaagcgcatcacggattttcagggggataac/GGcagc	100	Fully resistant
1444	H-3831	YP3	O:4a	48	13545 (+)	ggtgag/agtaacagcagctcggcatggtttaacaccgc/GGcatg	94	Fully resistant
760	J51	YP1	O:13	47	13640 (+)	ttgttt/aacctgcaatacggttgatgtttattttgtctc/GGtcag	100	Non-resistant
133	H-3831	YP1	O:4a	48	14452 (-)	cacaga/ttttatttgggtgatatttgaattgatcggcaa/GGcgta	100	Fully resistant
1154	H892-36/87	YP3	O:1a	12	15362 (+)	agttca/caaacaacattaataatgctaataattatac/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/tgagtaacattacaatgtagctattggaagag/Gctaata	100	Fully resistant
	Pa8728	YP3	R	60				Intermediate
1579	KP1244-2B	YP3	O:2c	56	16774 (-)	tttttt/cattacctcattgataactcggaacttcatcaa/GGcagt	97	Fully resistant
1362	DD110	YP3	O:6	11	16931 (-)	cgttcg/cgggggtggccttgttggcctccccgcttcaact/GGcttt	100	Fully resistant
1586	MW101-1	YP3	O:4b	28	17275 (-)	ggggat/atatcccccaaatcaaccgccactgggggtg/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/tttcctcggctctggctggcgggtgactgtcgcc/GGttgg	97	Intermediate
	OK5466	YP1	O:5b	73				Intermediate
1362	DD110	YP3	O:6	11	23321 (+)	gttctc/ggggggtggccttgtcggcctccccgcttcaactg/Gctttt	97	Fully resistant
283	Pa8728	YP1	R	60	24168 (-)	agtggg/tgatgtgcagcatgaaagctatatgtcctcat/GGctta	100	Intermediate
232	OK5466	YP1	O:5b	73	24547 (-)	ctgtgt/ttaatgtccagcaaatagacggccttgccactag/aGaca	94	Intermediate
1407	No.21	YP3	O:1a	86	27085 (+)	tggggc/cgtagtgtgtttaaccgttttttgtgtgctag/GGtatg	100	Intermediate
	H141-84	YP3	unknown	9				--
	Y717	YP3	unknown	9				--

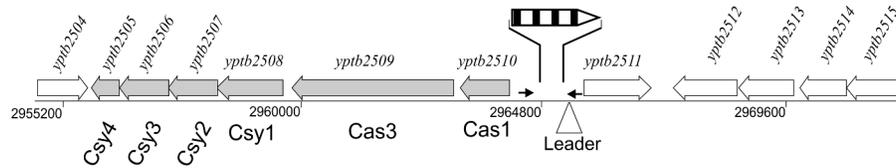
<sup>a</sup>ST, Multilocus sequence type according to MLST Databases at the University of Warwick, Warwick Medical School ([http://mlst.ucc.ie/mlst/dbs/Ypseudotuberculosis/GetTableInfo\\_html](http://mlst.ucc.ie/mlst/dbs/Ypseudotuberculosis/GetTableInfo_html)) [Laukkanen-Ninios, 2011 #9224]

For Peer Review Only

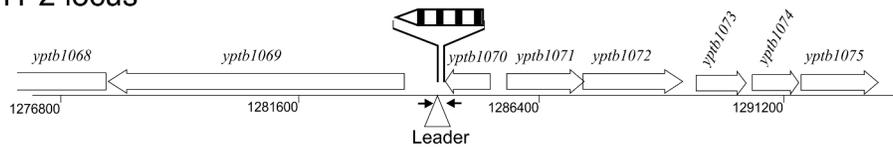
**Table 4.** Influence of pYptb32953-specific spacers on pYptb32953::cat conjugation and plasmid mobilization restriction by *Y. pseudotuberculosis* and *Y. similis* strains. Presented are for pYptb32953::cat conjugations transconjugant/recipient ratios of three parallel mating experiments. Variations 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
<b>Non-resistant group (&gt;0.2)</b>						
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 ( <i>Y. similis</i> )	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		
<b>Intermediate resistant group (0.01 – 0.2)</b>						
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		
<b>Fully resistant group (0 – 0.0099)</b>						
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 ( <i>Y. similis</i> )	1632	0	0	0	1-5	0.1-0.5
MW48 ( <i>Y. similis</i> )	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		

## YP1 locus



## YP2 locus



## YP3 locus

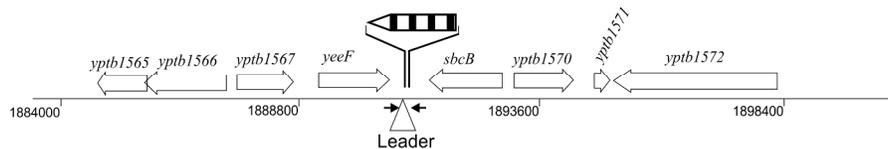


Figure 1. The locations of the YP1, YP2 and YP3 CRISPR loci in the genome of *Y. pseudotuberculosis* strain IP32953 (Accession number BX936398). The CRISPR associated *cas* and *csy* genes are shown with gray shading and the variable CRISPR repeat sequences as striped arrows. The new spacers are added 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)



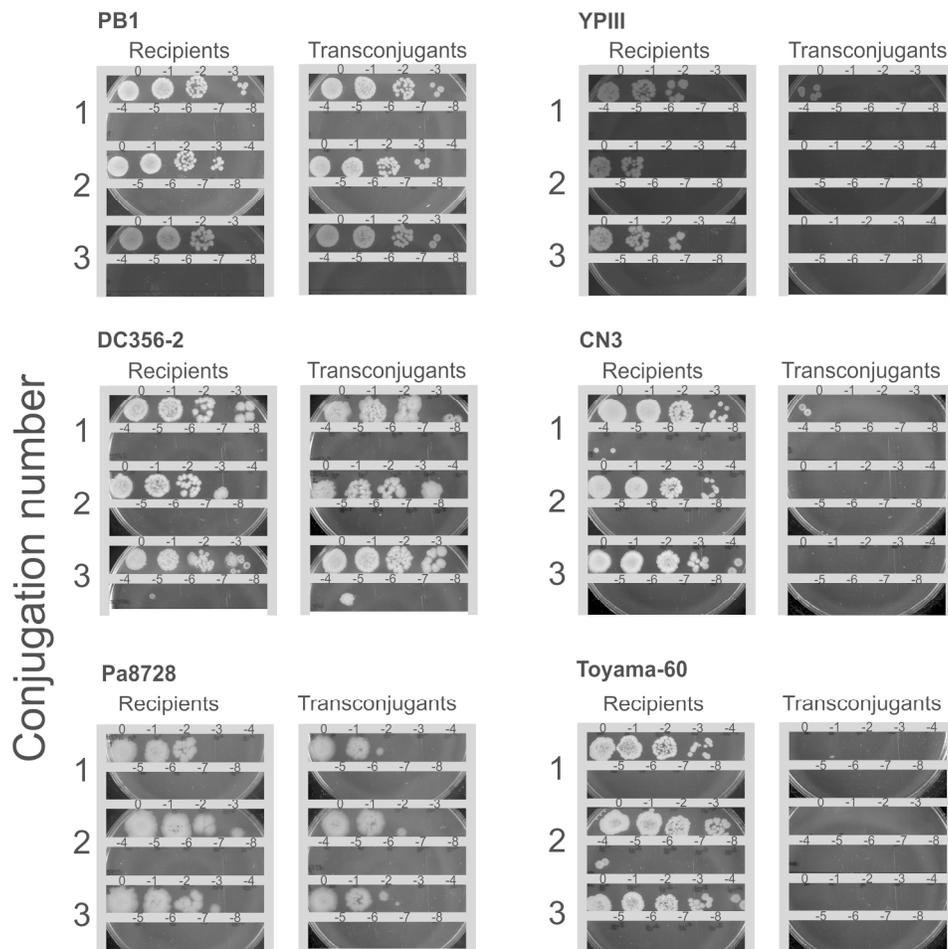


Figure 3. Conjugation experiments showing the serial 10-fold dilution plating of 5  $\mu$ l drops on selective CIN-agar plates without or with Clm allowing the growth from the mating mixture of all the recipient bacteria or only the transconjugant bacteria, respectively. PB1 and DC356-2 represent non-resistant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

