Originally published as:

Eva Schunder, Kerstin Rydzewski, Roland Grunow, Klaus Heuner, First indication for a functional CRISPR/Cas system in Francisella tularensis, International Journal of Medical Microbiology, Volume 303, Issue 2, March 2013, Pages 51-60.

DOI: 10.1016/j.ijmm.2012.11.004

First indication for a Functional CRISPR/Cas system in *Francisella tularensis*

Eva Schunder, Kerstin Rydzewski, Roland Grunow and Klaus Heuner*

Working group: Cellular Interactions of Bacterial Pathogens, Centre for Biological Security, Division 2 (ZBS2), Robert Koch-Institute, Berlin, Germany

Keywords: CRISPR, Cas, *Francisella tularensis*, bacteriophage
Running title: *Francisella tularensis* CRISPR/CAS

*Corresponding author:
Dr. Klaus Heuner
Cellular Interactions of Bacterial Pathogens
Centre for Biological Security (ZBS2)
Robert Koch-Institute
Nordufer 20
13353 Berlin
Phone: +49 30 18754-2226
E-mail: HeunerK@rki.de

Abstract

F. tularensis is a zoonotic agent and the subspecies *novicida* is proposed to be a water-associated bacterium. The intracellular pathogen *Francisella tularensis* causes tularemia in humans and is known for its potential to be used as biological threat. We analyzed the genome sequence of *F. tularensis* subsp. *novicida* U112 in silico for the presence of a putative functional CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated) system. CRISPR/Cas systems are known to encode an RNA-guided adaptive immunity-like system to protect bacteria against invading genetic elements, like bacteriophages and plasmids. In this work, we present a first indication that *F. tularensis* subsp. *novicida* encodes a functional CRISPR/Cas defence system. Additionally, we identified various spacer DNAs homologous to a putative phage present within the genome of *F. tularensis subsp. novicida*-like strain 3523. CRISPR/Cas is also present in *F. tularensis* subsp. *tularensis*, *holarctica* and *mediasiatica*, but these systems seem to be non-functional.

Introduction

Francisella tularensis is an intracellular pathogen that causes tularemia in humans and a wide range of animals (Ellis et al., 2002). Furthermore, *F. tularensis* is well known for its potential as a bacterial biological weapon (Dennis et al., 2001). Strains of *F. tularensis* subsp. (*Ft.*) *tularensis* can be lethal to humans and doses as low as 10-20 bacteria can be infective (Ellis et al., 2002). Transmission can mostly occur via aerosol ingestion or skin inoculation. *Ft. novicida* is less virulent and is thought to be an opportunistic pathogen. There are some reports of human infections by *Ft. novicida*, mostly these reports were from immunocompromised patients. The infection was localized and showed a relative mild illness (Clarridge et al., 1996; Hollis et al., 1989). However, *Ft. novicida* is thought to constitute an environmental lineage along with *F. philomiragia*. *F. philomiragia* has also been rarely associated with human disease in immunocompromized individuals (Hollis et al., 1989; Wenger et al., 1989).

Ft. novicida strain U112 is a water isolate from 1950 (Larson et al., 1955). There are further reports of several *Ft. novicida*-like bacteria isolated from humans (Birdsell et al., 2009; Clarridge et al., 1996; Hollis et al., 1989; Leelaporn et al., 2008; Whipp et al., 2003). The genome sequences of strain U112, the clinical isolate 3523 (Australian strain) and Fx1 (Texas strain) are available, as well as some *Ft. tularensis* and *F. philomiragia* strains (Barabote et al., 2009; Beckstrom-Sternberg et al., 2007; Champion et al., 2009; Chaudhuri et

al., 2007; Larsson et al., 2005; Modise et al., 2012; Nalbantoglu et al., 2010; Petrosino et al., 2006; Rohmer et al., 2007; Siddaramappa et al., 2011; Sjodin et al., 2012; Zeytun et al., 2012). *Ft. tularensis* is a highly infectious zoonotic agent. For *Ft. novicida* it is proposed to be a water-associated bacterium, but the environmental reservoir and the mechanism of transmission of *Ft. novicida* along vertebrate or invertebrate species is unknown. *Francisella* isolate 3523 is the first reported *Francisella* strain from the Southern Hemisphere, isolated from a patient infected by brackish water in the Northern Territory of Australia (Whipp et al., 2003).

Environmental bacteria easily get in contact with invading genetic elements, such as bacteriophages and plasmids. Bacteria have evolved an RNA-guided adaptive immunity mechanism to protect themselves against these invaders (Deveau et al., 2010; Horvath and Barrangou, 2010; Karginov and Hannon, 2010; Kunin et al., 2007; Makarova et al., 2011a; Makarova et al., 2011b; Sorek et al., 2008; van der Oost et al., 2009; Wiedenheft et al., 2012). The CRISPR/Cas system (Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR-associated) uses repeat/spacer-derived short CRISPR RNAs (crRNAs) to silence foreign nucleic acids in a sequence-specific manner. It targets the invading phages or plasmids in three steps: (1) adaptation of short pieces of DNA homologous to virus or plasmid sequences (spacers) into the CRISPR locus, (2) expression and processing of short guide crRNAs consisting of unique single repeat-spacer units and (3) interference with the alien nucleic acid by the activity of the proteins encoded by the *cas* genes. The hypothesis that the CRISPR/Cas system is necessary for the defence against invading DNA has been demonstrated in *Streptococcus thermophilus*. The integration of a short phage-specific sequence into the CRISPR locus of the strain conferred resistance to the cognate phage (Barrangou et al., 2007). The CRISPR/Cas system is divided into three distinct types (I, II and III). The signature genes for the three types are *cas3*, *cas9* and *cas10*, respectively (Makarova et al., 2011a; Makarova et al., 2011b).

In this paper we describe two different CRISPR/Cas systems identified within the genome of *Ft. novicida* U112 and the presence of these CRISPR systems in other *Francisella* strains. We present a first indication that the CRISPR/Cas systems of *Ft. novicida* encode functional defence systems by the identification of various spacer DNAs, homologous to a putative phage identified within the genome of the *Ft. novicida*-like strain 3523.

Materials and methods

CRISPR search in the genomes

Genome sequences used for *in silico* analysis are given in Table 1. CRISPR sequences in the genome of *Ft. novicida* U112 were identified by using the *CRISPRFinder* software (Grissa et al., 2007). Organization of the CRISPR/Cas systems of the strains investigated was analysed by using the software *ARTEMIS* (http://www.sanger.ac.uk/resources/software/artemis/). The CRISPR/Cas system (genes, proteins, direct repeats and spacer DNAs) of *Ft. novicida* U112 was used to identify homologous CRISPR/Cas systems within the analyzed *Francisella* genomes by using software *NBCI/BLAST/blastn* or *blastp* (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Results and discussion

Identification of two CRISPR/Cas systems in *F. tularensis* subsp. *novicida* U112

We used the "CRISPRFinder" software of the CRISPR database (CRISPRdp, http://crispr.u-psud.fr) for in silico analysis of the genome sequence of *Ft. novicida* U112 and identified four putative CRISPR systems. Two of them exhibit several direct repeats. For these two genomic regions, we also identified the presence of putative *cas* genes. Therefore, we further investigated the surrounding chromosomal DNA regions.

In Figure 1A, the CRISPR/Cas-1 system of *Ft. novicida* U112 is illustrated. The first gene FTN0757 encodes a Cas9 homolog. Cas9 represents the signature protein of the CasII-B type of *cas* genes (Makarova et al., 2011b). The *cas* gene order (*cas9-cas1-cas2-cas4*) is also similar to the CasII-B type of CRISPR systems (Makarova et al., 2011a; Makarova et al., 2011b). Cas1 and Cas2 proteins are involved in spacer DNA acquisition (see below), as well as Cas4, which is also thought to be involved in this process (Beloglazova et al., 2008; Makarova et al., 2011a; Makarova et al., 2011b; Wiedenheft et al., 2012). Downstream of *cas4*, a leader sequence of 359 bp, known to be necessary for spacer acquisition (Makarova et al., 2011a; Makarova et al., 2011b) and a direct repeat sequence are present in the CRISPR-1 system of *Ft. novicida* U112. It is proposed that the first repeat is the template for newly inserted repeats. In this process, the new repeat is replicated from repeat #1 (Yosef et al., 2012). The system of *Ft. novicida* U112 exhibits 13 spacer sequences of about 35 bp (34-37 bp) and all repeat sequences (37 bp in length) of the system are identical (Fig. 2A). This indicates that the CRISPR-1 system is functional, since spacer acquisition is present. The

spacer sequences are discussed later (see below). DNA sequences of the leader, the repeats and the spacers are given in Figure 2A.

In Figure 1B the CRISPR/Cas-2 system of *Ft. novicida* U112 is shown. As for the CRISPR/Cas-1 system, all putative necessary *cas* genes (FTN1397 [*cas9*-like], *cas1*, *cas2* and *cas4*), a leader sequence (139 bp), the first direct repeat (36 bp) and spacers of about 29 bp (26-32 bp) are present. The sequences of the leader and the spacers are given in Figure 2B. The repeats of the system are identical, with the exception of the last repeat (Fig. 2B). However, FTN1397 exhibits no Pfam_domain or motif, thus a putative Cas9-like function is proposed from the size and localisation of the gene. In additon, the *cas* gene order (*cas4-cas1-cas2*) is similar to the gene order found in type I-(A-D) CRISPR systems (Makarova et al., 2011a; Makarova et al., 2011b). Since *cas9* is thought to be the type signature gene, the *cas* type (I, II, or III) of the CRISPR-2 system of *Ft. novicida* U112 is unclear. However, since nine spacers are present (see also below), CRISPR/Cas-2 of *Ft. novicida* U112 also seems to encode a functional CRISPR system.

Presence and organization of CRISPR-1 and CRISPR-2 in other *Francisella* strains

We then analysed various available genome sequences of *Francisella* (see Tab. 1) for the presence of CRISPR-1 and CRISPR-2. The six investigated *Ft. novicida*-like strains are positive for CRISPR-1 and only the strains GA99-3548 and 3523 are negative for CRISPR-2. The genome sequences of *Ft. tularensis*, *Ft. holarctica* and *Ft. mediasiatica* exhibit a CRISPR-1, but no CRISPR-2 system (Tab. 1). Only one strain of *F. philomiragia* (ATCC25017) is postive for the CRISPR-2 system (Tab. 1). Then we had a closer look at the organization of the CRISPR-1 and -2 systems of the above mentioned strains.

The results for the CRISPR/Cas-1 system are shown in Figure 3. The CRISPR/Cas-1 system is localized directly downstream of the *fopA* gene. The outer membrane protein Fop A is known as a protective antigen for tularemia and the conserved *fopA* gene is used in diagnosis of tularemia using PCR and real-time TaqMan PCR assays (Fulop et al., 1996a; Fulop et al., 1996b; Hickey et al., 2011; Versage et al., 2003).

In all *Ft. novicida* strains investigated, the last direct repeat is followed by a gene homologous to FTN0761 of strain U112. All five strains exhibit a similar *cas*-gene organization with ~99% DNA identity to U112, with the exception of strain *Ft. novicida* 3523. In this strain, the leader sequence is different (90% DNA identity) and the direct repeat is not identical to that of strain U112 (Fig. 3 and 4). All strains exhibit different numbers of

spacer DNAs (Fig. 3). Interestingly, *Ft. novicida* strains Fx1 and GA99-3548 exhibit various spacer DNAs, although they possess a deletion within the *cas* gene region. Gene *cas2* is deleted and genes *cas1* and *cas4* are partially deleted. In strain GA99-3549 an IS element is integrated into the *cas4* gene. Therefore it is unclear, if the CRISPR system still encodes a functional CRISPR system. However, they may encode functional systems, since they exhibit various spacer DNAs and it was published that Cas9 seems to be sufficient for the generation of crRNA and cleaving of the target DNA (Barrangou et al., 2007; Deltcheva et al., 2011; Makarova et al., 2011a; Makarova et al., 2011b). In all other strains investigated, IS-Ftu2 was found downstream of the single (#1) direct repeat and these CRISPR systems do not exhibit any spacer DNAs, indicating that the identified CRISPR/Cas-1 systems of these strains are non-functional (Fig. 3). Nevertheless, the leader sequence is still present and identical to that of strain U112 (Fig. 4).

In the analysed *Ft. tularensis* strains, the *cas9* and *cas4* genes are present, but *cas1* and *cas2* genes are pseudogenes genereated by various point mutations (Fig. 3). In the three *Ft. holarctica* strains FSC022, FSC257 and LVS, the *cas9* gene is separated into three to four genes which encode putatively functional proteins. These strains also exhibit a complete *cas4* gene and various point mutations within the *cas1* and *cas2* genes. However, in strain LVS *cas2* seems to encode a putative functional protein. In *Ft. mediasiatica* FSC147, only *cas2* seems to encode a putative functional protein, all other *cas* genes are pseudogenes (Fig. 3). Within the genomes of the three *F. philomiragia* strains (ATCC25017, ATCC25015 and TX077308) investigated, no CRISPR-1 system could be identified and no CRISPR-like region could be identified downstream of *fopA* (*F. philomiragia* ATCC25017, see Fig. 3).

The results for the CRISPR/Cas-2 system are shown in Figure 5. The CRISPR/Cas-2 system is localized downstream of a gene (FTN1398) encoding an acetyl transferase (AcTF) and upstream of *arsR* (FTN1393) encoding a regulatory protein. All CRISPR/Cas-2 systems exhibit a similar *cas* gene organization, a leader region exhibiting 96-100% DNA identity to strain U112 and identical direct repeats, except the last direct repeat (Fig. 4B and Fig. 5). In contrast to CRSIPR-1, in CRISPR-2 the last direct repeat (#L) differs from the others, but is similar among the different CRISPR-2 systems analysed (Fig. 4B). This is surprising, since it is thought that all repeats are replicates of repeat #1 (Yosef et al., 2012). All CRISPR-2 systems identified seem to be functional, since they exhibit different amounts of spacer DNAs (5-18 copies). Like *cas9* in the CRISPR-1 system of the *Ft. holarctica* strains LVS and FSC257, the hypothetical signature protein of CRISPR-2 of *F. philomiragia* seems to be

separated into three open reading frames (*Fphi1650-52*) which seem to encode functional proteins, since the CRISPR-2 system of this strain exhibits five spacers (Fig. 5).

The CRISPR-2 system is present in all *Ft. novicida* strains investigated and is integrated within the same chromosomal region, with the exception of strain *Ft. novicida*-like 3523 (Fig. 5). In strain *Ft. novicida*-like 3523 there is no DNA insertion between the AcTF and the *asrR* gene. However, downstream of AcTF a #L repeat was detected and in GA99-3548 a #1 repeat-spacer-#L repeat structure is present, indicating the loss of a prior CRISPR/Cas system. CRISPR-2 is also present in *F. philomiragia* ATCC25017, but is located at another genomic region. In *F. philomiragia* ATCC25017, two *F. philomiragia* specific genes are located between AcTF and *argR* (Fig. 5). We were not able to detect a homolog of CRISPR-2 in *F. philomiragia* strains ATCC25015 and TX077308 as well as in any of the investigated available genome sequences of *Ft.* subsp. *tularensis, holarctica* or *mediasiatica* (Fig. 5 and Tab. 1). In *Ft. holarctica* LVS and *Ft. tularensis* SchuS4, an IS-Ftu1 element is integrated between repeat #1 and repeat #L, but no further elements of the CRISPR/Cas system are present (Fig. 5). In *Ft. mediasiatica*, a chromsosmal rearrangement has occured in the respective region, probably caused by a recombination event of two IS-Ftu1 elements. The loss of genes may be a process of genomic pathoadaptation in response to changing or new habitats of the bacteria (Champion et al., 2009; Ochman and Moran, 2001; Svensson et al., 2005). The CRISPR/Cas system may be more relevant for the water-associated species *Ft. novicida* and *F. philomiragia*, which gets in contact with a large number of different (invading) mobile genetic elements.

Analysis of the spacer DNAs of the CRISPR-1 and CRISPR-2 systems

The CRISPR/Cas system is an adaptive immunity-like system found in many bacteria (Makarova et al., 2011a; Makarova et al., 2011b). Bacteria are able to integrate short pieces of DNA, often homologous to virus or plasmid sequences, into the CRISPR locus (see introduction). The locus is then transcribed (pre-crRNA) and cleaved into short crRNAs. This step is catalyzed by the *cas* genes and directs interfering proteins to target nucleic acids matching the spacers. In *S. thermophilus* this system was shown to confer resistance of the bacteria to specific phages (Barrangou et al., 2007). The spacer DNA in general is very specific for each strain and therefore used for epidemiological subtyping (spoligotyping) (Ginevra et al., 2012; Price et al., 2007; Sorek et al., 2008). Since *Ft. novicida* and *F. philomiragia* are species known to be associated with aquatic habitats and therefore may get

into contact with various phages and plasmids, we had a closer look to the spacer DNA of the identified CRISPR systems.

The spacer DNAs of CRISPR-1 and CRISPR-2 are 34-38 and 26-33 bp in length, respectively. Results of "*Blastn*" searches with the spacer DNAs of the CRISPR-1 and CRISPR-2 systems are listed in Table 2. Interestingly, we found spacer DNAs with counter parts in a sequence of a phage minor tail protein (*Ft. novicida*-like 3523, CRISPR-1, 1. spacer), a *Vibrio phage* ICP1_2004_A sequence (*Ft. novicida* GA99, CRISPR-2, 2. spacer), of a plasmid of *Vibrio harveyi* (*Ft. novicida* U112, CRISPR-2, 8. spacer) and in *F. philomiragia* a spacer DNA sequence in antisense direction to an own gene (*Fphi1603*) (*F. philomiragia* 25017, CRISPR-2, 1. spacer). In one case, a part of spacer 11 of *Ft. novicida* U112 CRISPR-1 is also found (in antisense direction) in spacer 7 of *Ft. novicida* GA99-3549 (Tab. 2). This indicates that the CRISPR systems of these strains are functional, since the bacteria acquired spacers during their life-cycle, the predicted function of CRISPR/Cas systems. However, the most intriguing finding was the identification of altogether 12 spacer sequences, in both CRISPR systems of 5 different species, which all also were identified within the same chromosomal region of strain *Ft. novicida*-like 3523 (Tab. 2, underlined sequences and Fig. 6A, red arrows).

We therefore had a closer look to that DNA region of strain *Ft. novicida*-like 3523 and we identified a putative genomic island of about 34,259 bp exhibiting 48 open reading frames (ORFs) (Fig. 6A). Most of the encoded proteins are homologous to phage proteins (phage tail, phage major tail tube, tail sheath, base plate, tail fiber, phage terminase, coat protein, replication protein and regulatory proteins) and therefore, this genomic island may encode a *Ft. novicida* phage (Fig. 6B). The phage island is integrated into the $tRNA^{Val}$ (localized between gene *FN3523_0985* and *_0986*) and the corresponding *attL*-site (repeat-1, Fig. 6A) is localized between *FN3523_1033* and *FN3523_1034*. Gene *FN3523_1033* encodes a site-specific integrase with 31% identity to Lpc2818, a site-specific integrase of the genomic Island Trb-1 of *L. pneumophila* Corby. In *L. pneumophila* Trb-1 is also integrated into a tRNA ($tRNA^{Pro}$) gene and the integrase (Lpc2818) is responsible for the generation of a circular episomal form of Trb-1, which then can be transferred to other *Legionella* strains by conjugation (Glockner et al., 2008 and unpublished data). Therefore, it is likely that the phage of *Ft. novicida*-like 3523 may be excised from the genome in a circular form. In addition, gene *FN3523_1028* encodes a putative pro-phage repressor and gene *FN3523_1021* an phage anti-repressor (of the circular form). Due to the lack of chromosomal DNA of this strain, we could not prove this hypothesis so far, but it will be of interest to further analyse this putative

phage. Gene *FN3523_1034* encodes an HlyD-like protein, recently described to be part of a toxin encoding locus of *Ft. novicida*-like 3523 (Siddaramappa et al., 2011).

Surprisingly, this putative phage is found within the genome of a *Ft. novicida*-like isolate of Australia (Siddaramappa et al., 2011). However, the spacer DNA was identified in isolates of *Ft. novicida* U112 (Utah, USA), GA99-3548 (Louisiana, USA), GA99-3549 (California, USA) and in *Ft. novicida* Fx1 (Texas, USA). This indicates that this phage may be a general (ubiquitous) phage of *Ft. novicida* strains. Therefore, we called it a *Ft. novicida* phage. The spacers identified covered the 5`-part of the phage (Fig. 6A, red arrows). The complete region is shows low similarity to a lytic *E. coli* phage (vB_EcoM_ECO1230-10) isolated from diary farms manure systems (Santos and Bicalho, 2011).

Conclusion

In this work, we provide a first indication for functional CRIPSR/Cas systems in *F. tularensis* subsp. *novicida* and one strain of *F. philomiragia*, since various different spacer sequences are present within the identified systems. We also found a first hint for a general *Francisella* phage-targeting sequence within the analysed spacer DNAs and we identifed the corresponding putative phage within the genome sequence of *Ft. novicida*-like strain 3523. So far, other natural phages for *F. tularensis* are unknown. Interestingly, the evolutionarily oldest and water-associated *Ft. novicida* strains (Svensson et al., 2005) exhibit functional CRISPR/Cas systems, whereas the more virulent and evolutionary younger strains *Ft. tularensis*, *holarctica* and *mediasiatica* seem to carry inactive CRISPR/Cas systems, exhibiting various gene deletions or pseudogenes. Similar findings were also published for the whole genome sequences of these strains (Champion et al., 2009). This process may be an adaptation of the strains to new habitats (e.g., like eukaryotic hosts) (Dobrindt et al., 2010; Ochman and Moran, 2001).

For other bacteria, the spacer DNAs of the CRISPR/Cas systems are used for epidemiological subtyping of isolates (spoligotyping) (Ginevra et al., 2012; Price et al., 2007; Sorek et al., 2008). It would be interesting to analyse, if the CRISPR/Cas system is usable to generate a CRISPR/Cas spacer based spoligotyping for *Francisella*.

Acknowledgments

This work received financial support from the Robert Koch-Institute.

REFERENCES

Barabote, R.D., Xie, G., Brettin, T.S., Hinrichs, S.H., Fey, P.D., Jay, J.J., Engle, J.L., Godbole, S.D., Noronha, J.M., Scheuermann, R.H., Zhou, L.W., Lion, C., Dempsey, M.P., 2009. Complete genome sequence of *Francisella tularensis* subspecies *holarctica* FTNF002-00. PLoS One 4, e7041.

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

Beckstrom-Sternberg, S.M., Auerbach, R.K., Godbole, S., Pearson, J.V., Beckstrom-Sternberg, J.S., Deng, Z., Munk, C., Kubota, K., Zhou, Y., Bruce, D., Noronha, J., Scheuermann, R.H., Wang, A., Wei, X., Wang, J., Hao, J., Wagner, D.M., Brettin, T.S., Brown, N., Gilna, P., Keim, P.S., 2007. Complete genomic characterization of a pathogenic A.II strain of *Francisella tularensis* subspecies *tularensis*. PLoS One 2, e947.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E.V., Edwards, A.M., Savchenko, A., Yakunin, A.F., 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. 283, 20361-20371.

Birdsell, D.N., Stewart, T., Vogler, A.J., Lawaczeck, E., Diggs, A., Sylvester, T.L., Buchhagen, J.L., Auerbach, R.K., Keim, P., Wagner, D.M., 2009. *Francisella tularensis* subsp. *novicida* isolated from a human in Arizona. BMC Res. Notes 2, 223.

Champion, M.D., Zeng, Q., Nix, E.B., Nano, F.E., Keim, P., Kodira, C.D., Borowsky, M., Young, S., Koehrsen, M., Engels, R., Pearson, M., Howarth, C., Larson, L., White, J., Alvarado, L., Forsman, M., Bearden, S.W., Sjostedt, A., Titball, R., Michell, S.L., Birren, B., Galagan, J., 2009. Comparative genomic characterization of *Francisella tularensis* strains belonging to low and high virulence subspecies. PLoS Pathog. 5, e1000459.

Chaudhuri, R.R., Ren, C.P., Desmond, L., Vincent, G.A., Silman, N.J., Brehm, J.K., Elmore, M.J., Hudson, M.J., Forsman, M., Isherwood, K.E., Gurycova, D., Minton, N.P., Titball, R.W., Pallen, M.J., Vipond, R., 2007. Genome sequencing shows that European isolates of *Francisella tularensis* subspecies *tularensis* are almost identical to US laboratory strain Schu S4. PLoS One 2, e352.

Clarridge, J.E., 3rd, Raich, T.J., Sjosted, A., Sandstrom, G., Darouiche, R.O., Shawar, R.M., Georghiou, P.R., Osting, C., Vo, L., 1996. Characterization of two unusual clinically significant *Francisella* strains. J. Clin. Microbiol. 34, 1995-2000.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.

Dennis, D.T., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Friedlander, A.M., Hauer, J., Layton, M., Lillibridge, S.R., McDade, J.E., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T.M., Russell, P.K., Tonat, K., Working Group on Civilian, B., 2001. Tularemia as a biological weapon: medical and public health management. JAMA 285, 2763-2773.

Deveau, H., Garneau, J.E., Moineau, S., 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu. Rev. Microbiol. 64, 475-493.

Dobrindt, U., Chowdary, M.G., Krumbholz, G., Hacker, J., 2010. Genome dynamics and its impact on evolution of *Escherichia coli*. Med. Microbiol. Immunol. 199, 145-154.

Ellis, J., Oyston, P.C., Green, M., Titball, R.W., 2002. Tularemia. Clin. Microbiol. Rev. 15, 631-646.

Fulop, M., Leslie, D., Titball, R., 1996a. A rapid, highly sensitive method for the detection of *Francisella tularensis* in clinical samples using the polymerase chain reaction. Am. J. Trop. Med. Hyg. 54, 364-366.

Fulop, M., Manchee, R., Titball, R., 1996b. Role of two outer membrane antigens in the induction of protective immunity against *Francisella tularensis* strains of different virulence. FEMS Immunol. Med. Microbiol. 13, 245-247.

Ginevra, C., Jacotin, N., Diancourt, L., Guigon, G., Arquilliere, R., Meugnier, H., Descours, G., Vandenesch, F., Etienne, J., Lina, G., Caro, V., Jarraud, S., 2012. *Legionella pneumophila* sequence type 1/Paris pulsotype subtyping by spoligotyping. J. Clin. Microbiol. 50, 696-701.

Glockner, G., Albert-Weissenberger, C., Weinmann, E., Jacobi, S., Schunder, E., Steinert, M., Hacker, J., Heuner, K., 2008. Identification and characterization of a new conjugation/type IVA secretion system (trb/tra) of *Legionella pneumophila* Corby localized on two mobile genomic islands. Int. J. Med. Microbiol. 298, 411-428.

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

Hickey, A.J., Hazlett, K.R., Kirimanjeswara, G.S., Metzger, D.W., 2011. Identification of *Francisella tularensis* outer membrane protein A (FopA) as a protective antigen for tularemia. Vaccine 29, 6941-6947.

Hollis, D.G., Weaver, R.E., Steigerwalt, A.G., Wenger, J.D., Moss, C.W., Brenner, D.J., 1989. *Francisella philomiragia* comb. nov. (formerly *Yersinia philomiragia*) and *Francisella tularensis* biogroup *novicida* (formerly *Francisella novicida*) associated with human disease. J. Clin. Microbiol. 27, 1601-1608.

Horvath, P., Barrangou, R., 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167-170.

Karginov, F.V., Hannon, G.J., 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell 37, 7-19.

Kunin, V., Sorek, R., Hugenholtz, P., 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. 8, R61.

Larson, C.L., Wicht, W., Jellison, W.L., 1955. A new organism resembling *P. tularensis* isolated from water. Public Health Rep. 70, 253-258.

Larsson, P., Oyston, P.C., Chain, P., Chu, M.C., Duffield, M., Fuxelius, H.H., Garcia, E., Halltorp, G., Johansson, D., Isherwood, K.E., Karp, P.D., Larsson, E., Liu, Y., Michell, S., Prior, J., Prior, R., Malfatti, S., Sjostedt, A., Svensson, K., Thompson, N., Vergez, L., Wagg, J.K., Wren, B.W., Lindler, L.E., Andersson, S.G., Forsman, M., Titball, R.W., 2005. The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. Nat. Genet. 37, 153-159.

Leelaporn, A., Yongyod, S., Limsrivanichakorn, S., Yungyuen, T., Kiratisin, P., 2008. *Francisella novicida* bacteremia, Thailand. Emerg. Infect. Dis. 14, 1935-1937.

Makarova, K.S., Aravind, L., Wolf, Y.I., Koonin, E.V., 2011a. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol. Direct 6, 38.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E.V., 2011b. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467-477.

Modise, T., Ryder, C., Mane, S.P., Bandara, A.B., Jensen, R.V., Inzana, T.J., 2012. Genomic comparison between a virulent type A1 strain of *Francisella tularensis* and its attenuated O-antigen mutant. J. Bacteriol. 194, 2775-2776.

Nalbantoglu, U., Sayood, K., Dempsey, M.P., Iwen, P.C., Francesconi, S.C., Barabote, R.D., Xie, G., Brettin, T.S., Hinrichs, S.H., Fey, P.D., 2010. Large direct repeats flank genomic rearrangements between a new clinical isolate of *Francisella tularensis* subsp. *tularensis* A1 and Schu S4. PLoS One 5, e9007.

Ochman, H., Moran, N.A., 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292, 1096-1099.

Petrosino, J.F., Xiang, Q., Karpathy, S.E., Jiang, H., Yerrapragada, S., Liu, Y., Gioia, J., Hemphill, L., Gonzalez, A., Raghavan, T.M., Uzman, A., Fox, G.E., Highlander, S., Reichard, M., Morton, R.J., Clinkenbeard, K.D., Weinstock, G.M., 2006. Chromosome rearrangement and diversification of *Francisella tularensis* revealed by the type B (OSU18) genome sequence. J. Bacteriol. 188, 6977-6985.

Price, E.P., Smith, H., Huygens, F., Giffard, P.M., 2007. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. Appl. Environ. Microbiol. 73, 3431-3436.

Rohmer, L., Fong, C., Abmayr, S., Wasnick, M., Larson Freeman, T.J., Radey, M., Guina, T., Svensson, K., Hayden, H.S., Jacobs, M., Gallagher, L.A., Manoil, C., Ernst, R.K., Drees, B., Buckley, D., Haugen, E., Bovee, D., Zhou, Y., Chang, J., Levy, R., Lim, R., Gillett, W., Guenthener, D., Kang, A.,

Shaffer, S.A., Taylor, G., Chen, J., Gallis, B., D'Argenio, D.A., Forsman, M., Olson, M.V., Goodlett, D.R., Kaul, R., Miller, S.I., Brittnacher, M.J., 2007. Comparison of *Francisella tularensis* genomes reveals evolutionary events associated with the emergence of human pathogenic strains. Genome Biol. 8, R102.

Santos, T.M., Bicalho, R.C., 2011. Complete genome sequence of vB_EcoM_ECO1230-10: a coliphage with therapeutic potential for bovine metritis. Vet. Microbiol. 148, 267-275.

Siddaramappa, S., Challacombe, J.F., Petersen, J.M., Pillai, S., Hogg, G., Kuske, C.R., 2011. Common ancestry and novel genetic traits of *Francisella novicida*-like isolates from North America and Australia as revealed by comparative genomic analyses. Appl. Environ. Microbiol. 77, 5110-5122.

Sjodin, A., Svensson, K., Ohrman, C., Ahlinder, J., Lindgren, P., Duodo, S., Hnath, J., Burans, J.P., Johansson, A., Colquhoun, D.J., Larsson, P., Forsman, M., 2012. Genome characterisation of the genus *Francisella* reveals insight into similar evolutionary paths in pathogens of mammals and fish. BMC Genomics 13, 268.

Sorek, R., Kunin, V., Hugenholtz, P., 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nature Rev. Microbiol. 6, 181-186.

Svensson, K., Larsson, P., Johansson, D., Bystrom, M., Forsman, M., Johansson, A., 2005. Evolution of subspecies of *Francisella tularensis*. J. Bacteriol. 187, 3903-3908.

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., Brouns, S.J., 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci. 34, 401-407.

Versage, J.L., Severin, D.D., Chu, M.C., Petersen, J.M., 2003. Development of a multitarget real-time TaqMan PCR assay for enhanced detection of *Francisella tularensis* in complex specimens. J. Clin. Microbiol. 41, 5492-5499.

Wenger, J.D., Hollis, D.G., Weaver, R.E., Baker, C.N., Brown, G.R., Brenner, D.J., Broome, C.V., 1989. Infection caused by *Francisella philomiragia* (formerly *Yersinia philomiragia*). A newly recognized human pathogen. Ann. Intern. Med. 110, 888-892.

Whipp, M.J., Davis, J.M., Lum, G., de Boer, J., Zhou, Y., Bearden, S.W., Petersen, J.M., Chu, M.C., Hogg, G., 2003. Characterization of a novicida-like subspecies of *Francisella tularensis* isolated in Australia. J. Med. Microbiol. 52, 839-842.

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

Yosef, I., Goren, M.G., Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res. 40, 5569-5576.

Zeytun, A., Malfatti, S.A., Vergez, L.M., Shin, M., Garcia, E., Chain, P.S., 2012. Complete Genome Sequence of *Francisella philomiragia* ATCC 25017. J. Bacteriol. 194, 3266.

Figure Legends

Fig. 1. Organization of the CRISPR/Cas-1 (A) and CRISPR/Cas-2 (B) systems of *Ft. novicida* U112. The *cas* genes are indicated by grey arrows, gene numbers are given above and gene names and numbers of amino acids (aa) are given below. The leader (black line), the direct repeats (red arrows) and the spacers (blue arrows) are shown. The number of spacers are indicated above the blue arrows. #1, first repeat; #L, last repeat.

Fig. 2. Nucleotide sequence of the CRISPR-1 (Leader: 359 bp, DR: 37 bp; 13 spacers) (A) and CRISPR-2 (Leader: 139 bp, DR: 36 bp, 9 spacers) (B) region of *Ft. novicida* U112. The DNA sequence of the leader (green), the direct repeats (red), the spacers (black) and other chromosomal DNA (brown) are given.

Fig. 3. Organization and comparison of the CRISPR/Cas-1 systems of different *Francisella* strains. In all strains investigated, the system is integrated downstream of *fopA*. The *cas* genes are indicated by blue arrows, gene numbers are given within and gene names above the blue arrows. The leader (black line), the direct repeats (red arrows) and the spacers (brown arrows) are shown. The number of spacers are indicated below the brown arrows. DNA identity (%) of the *cas* gene encoding regions and the leader DNAs to the respective DNA region of *F. tularensis* subsp. *novicida* U112 (*Ft. nov* U112) is given. An IS element and *IS-Ftu2* are indicated by an orange and yellow arrow, respectively. Homologous genes are shaded in the same colour. Variations in the amino acid length of ORFs or the leader sequence are given in brackets. *Identical DNA sequences were identified in strains *Ft. tularensis* FSC198, NEO61598, TIGB03, TI0902 and MA00-2987. aa, amino acids; bp, base pairs; n.a, not annotated; ps, pseudogene; nov, *novicida*; hol, *holarctica*; med, *mediasiatica*; phi, *philomiragia*.

Fig. 4. DNA sequences of the direct repeats (DRs) of the CRISPR/Cas-1 and -2 systems. nov, *novicida*; tul, *tularensis*, hol, *holarctica*; med, *mediasiatica*, phi, *philomiragia*; GA99-48/-49, GA99-3548/49. #1, first repeat; 2., second repeat; #L, last repeat, (-), nucleotide not present; (**.)** , identical nucleotide; (*), strain does not exhibit a CRISPR/Cas-2 system.

Fig. 5. Organization and comparison of the CRISPR/Cas-2 systems of different *Francisella* strains. The *cas* genes are indicated by grey arrows, gene numbers are given within and gene

names above the arrows. The acetyl transferase (AcTF)-*arsR*-*naoX* gene cluster is shaded in blue and a *F. philomiragia* specific gene cluster in brown. Numbers of amino acids are given in brackets. The leader (black line), the direct repeats (red arrows) and the spacers (yellow arrows) are shown. The number of spacers are indicated below the yellow arrows. DNA identity (%) of the *cas* gene encoding regions and the leaders to the respective DNA region of *F. tularensis* subsp. *novicida* U112 (*Ft. nov* U112) are given. The IS element *IS-Ftu1* is indicated by green arrows. * Similar organization of the respective DNA region was observed for strains *Ft. tularensis* FSC198, NEO61598, TIGB03, TI0902 and *Ft. holarctica* OSU18. aa, amino acids; bp, base pairs; nov, *novicida*; hol, *holarctica*; med, *mediasiatica*, phi, *philomiragia*; GA-49, GA99-3549.

Fig. 6. Organization of the putative *Ft. novicida* phage (A) and details of all open reading frames (ORFs) present (B) identified within the genome sequence of *F. tularensis* subsp. *novicida*-like strain 3523. (A) The putative phage (34,259 bp, 48 ORFs) is integrated within the $tRNA^{Val}$ gene (yellow arrow). The 3`end of the $tRNA^{Val}$ gene, which is conform with repeat-1, is indicated as a yellow trapezium. Hypothetical proteins of the putative phage are shown in grey and genes with similarities to known phage proteins are shaded in blue. A putative site-specific integrase is indicated by an green arrow. Location of the spacer DNAs identified in the CRISPR/Cas-1 or -2 systems of analysed *Francisella* strains are indicated by red arrows. The sequence of the $tRNA^{Val}$ gene and of repeat-1 are shown below. (B) Gene designation, length of the proteins (aa) and putative function of all 48 open reading frames (ORFs) of the putative phage are given. black, hypothetical proteins; blue, proteins with similarities to known phage proteins; red, ORFs exhibiting nucleotide sequences identified as spacers within the CRISPR/Cas-1 or -2 systems of analysed *Francisella* strains; aa, amino acids; n.a, not annotated.

Table. 1 Genome sequences (species) used for CRISPR/Cas analysis

Species	Acession number	CRISPR-1	CRISPR-2
Ft. novicida U112	CP000439	+	+
Ft. novicida FTG	ABXZ01000001* ABXZ01000002*	+	+
Ft. novicida-like Fx1	CP002557	+	+
Ft. novicida-like GA99-3548	DS264589* DS264590*	+	-
Ft. novicida-like GA99-3549	DS264128* DS264129*	+	+
Ft. novicida-like 3523	CP002558	+	-
Ft. tularensis SchuS4	AJ749949	+	-
Ft. tularensis FSC198	AM286280	+	-
Ft. tularensis MA00-2987	DS990216*	+	**ni**
Ft. tularensis WY96-3418	CP000608	+	-
Ft. tularensis NE061598	CP001633.1	+	-
Ft. tularesnsis TIGB03	CP003048	+	-
Ft. tularensis TI0902	CP003049	+	-
Ft. holarctica LVS	NC_007880	+	-
Ft. holarctica FSC257	DS229056*	+	**ni**
Ft. holarctica FSC022	DS264138*	+	**ni**
Ft. holarctica FTNF002-00	CP000803	+	-
Ft. holarctica OSU18	CP000437	+	-
Ft. mediasiatica FSC147	CP000915	+	-
F. philomiragia ATCC25017	CP000937	-	+
F. philomiragia ATCC25015	DS99316*	-	-
F. philomiragia TX077308	CP002872	-	-
F. noatunensis subsp. *orientales*	CP003402	-	-

* contigs of shot gun sequences received from NCBI/GenBank/ "DBSource"
ni= not investigated

Table 2. Spacer sequences of CRISPR-1 (A) and CRIPR-2 (B) systems of various *Francisella* strains.

A. CRISPR-1	**Blastn (>25 bp in length)**
***Ft .novicida* U112 (34-37 bp)**	
1. caagattggtatatatcaactgctaatcactcgc	-
2. ggtaaaaccaaaaaccaaaagttaagatatctctt	-
3. tatgcggcttgtggttgtgtcattgcgaagactg	-
4. ttaagcggatcaatttcgattacatcttcaccagat	-
5. tttgtcaaatattggtcgagttatcttaaatagt	*Nostoc punctiforme* pNPUN03
6. actcctatatgaccatacaaaacctgcccattac	-
7. ttatttgccaacactcaaaagataggtaaccctatag	-
8. tagtttttattatctgtgctgttttctgtcattgc	-
9. gtatatatcagtagtttaaaagcattggctattact	-
10.gttaataatgctaaagaaccagctgaatgtgtaa	-
11.at**tcttgaattaagtgttgctg**ctgaggattaggca	**antisense of spacer 7 of *Ft. nov.* GA99-3549 (CRISPR-1)***
12.atataaatttggataattttaaggtattaaatcctt	-
13.atcacactataaaagtaatcaagcttgccattgt	-
***Ft. novicida* Fx1 (34-37 bp)**	
1. tctggatgtagtaaccaagtcttacgctcgataa	-
2. gttggatctgtctctctcattaatacaagattgtat	-
3. aataaagaaaatgtagacttattccaaaacgtta	-
4. agcttctcatatatgtatttactgtcgatttgat	-
5. ttttaataggtgatttgtgaattttgatgagttaaag	-
***Ft. novicida* FTG (34-38 bp)**	
1. aagtttgacttgttacccttttgaatttcatctcta	-
2. tcaatataccagctaggattactcgcataagcat	-
3. taatgcaaatttgtatagtctacttctgtagaataa	*Xenopus tropicales* mRNA, *Homo sapiens* DNA
4. **agtattttctcttgtcttgtaag**t**tctcttag**atag	**FN3523_0993**
5. taagaaacttcgtcggttcttgggtactatgtacta	-
6. agtattcttgacgcggctactaatacgtatcaaataaa	*Flavobacterium branchiophilum*
7. taaagcactctctaagatgacttcataattactac	-
8. aac**gatat**t**acacgcgaaagagctgcttttgaga**t	**FN3523_1002**
9. taagcaagactctcaaacagactttaaagcatcgttt	-
10.gctctatttaagtgttcgtgatactctgtttgtaa	-
11.acaagctaaagaaattattataattgcttactaca	-
12.tagcaagtattagtttacttcatttatttaaaat	-
13.agattatgtgtaagcaacttaattttaacgaacttc	-
14.atctaattggtaactcggaaagtataatcattg	-
15.acccaaataatctcgttatcatgaggctttatttg	-
16.tcctgaaatttccactggtctcccaagaaagcttg	-
17.ccgtatttagcctaatgttatctttaaatatttcaa	-
18.aagaattaaaacaagaaagcatgtaacaaacttagat	*M.truncatula* DNA
19.ctctttcttccccttttctgcttgtcgctcaataaa	-
20 ttattagaaaaaataaatgctggtatgttaaaattcg	*Homo sapiens* DNA
21.atcatagaagctaaatatagcggtaaaatgtttga	-
22.ctctttcttccccttttctgcttgtcgctcaataaa	-
***Ft. novicida* 3523 (34-37 bp)**	
1. ttgcca**gtataaccaccttcagcaaaacc**agtaaca	*Acinetobacter baumannii* **phage minor tail protein***
2. gccaaaaagactgagcctaaggctaaagttgagg	-
3. gtaataacagcctttgagcctttacagcctacaa	-
4. ccagcactatctgcaattgcaatgttgctttggt	-
5. attaaaaaagaaatatatgctgtatcgtatggca	-
6. catataacctgaattgtacccgtatcggcagcag	-

7. ttttttttatttgttttgcctgcaatgctcatagca	-
8. taacgcctcgaagttgaactttataacccaatatag	-
9. ttcatctaaccaatccccataatcgttccataattg	-
10.tcgacatactcttgatattcattatctggcttaa	-
11.aatatcgctagaaaacatattgttaagcatattat	-
12.ccgtttcaacaacaggcttgaagtttcaagcact	-
13.tcaccctcgccgccagtttgaccgcctccaccag	*Symbiobacterium thermophilum*
14.aattgtgaaatagataatatagcttctacattaga	*Mus musculus* BAC clone
15.ataaaaattaaataaaatatcttgaaatgcttatt	*Felis catus*
16.agagaatttgtatttagataccctaagattaagaaaa	-
17.tctggacgagtgttggttatagtcattgtcatacc	-

***Ft. novicida* GA99-3548 (35-37 bp)**

1. tcgacaggagtctgaccgtcaattaaagcagataatc	-
2. taggatttacgccagctaattcgttaactaggttca	-
3. gaactcgtcgtacatgatctccgccgccatgtgcag	-
4. cgctactataaataggggcggatataataaccgaa	-
5. ttcatcagcgaatcttttaaatttcttagatttgtca	-
6. gtacaaagagtttatttctcgtgttacttttttttatt	Mus musculus BAC clone
7. **gcccaaattttctg**c**tcaaa**t**tc**g**ccttttttgcg**	**FN3523_0999**
8. **tatcttcaa**t**ata**c**aaatcaatataagagttatct**	**FN3523_1006**
9. **ctat**a**ttcaaacttgcatatatttc**t**gata**gggct	**FN3523_1009**
10.attaattgcactagagaagcaagtccgaatctcggg	-

***Ft. novicida* GA99-3549 (34-36 bp)**

1. aagatagacagagctttattacaatcaagaaaagat	-
2. ccagttacgttatcgtctcgcactatcgcaggag	-
3. aatcatagagttaatttcgctcctgaccctgctg	-
4. tgattaataacacgataaacattacctttttgtttt	-
5. aatgattaaaatctacgaaactgatactcaaattc	-
6. agttgcctcgaatcgttccagcagatgtaaacat	-
7. aacccgcaa**cagcaacacttaattcaaga**gttgg	**antisense spacer 11 *Ft. nov* U112 (CRISPR-1)***
8. aaatctaaa**a**aatgtagattaataccaatacagcaa	B. subtilis plasmid

B. CRISPR-2	**Blastn (>22 bp in length)**

***Ft. novicida* U112 (26-32 bp)**

1. gagaagtcatttaataaggccactgttaaaa	-
2. gctactattcctgtgccttcagataattca	*Strongylocentrotus purpuratus*
3. gtctagagccttttgtattagtagccg	-
4. tagcgatttatgaaggtcatttttt	**intergenic region (*FN3523_1004-5*)**
5. agattaaaaggtaattctatcttgttgag	***FN3523_0998* (antisense)**
6. tacctagtagatacgcttactgataacaa	-
7. aaactttcatttatgatataaagtttttt	*Apis mellifera*, mRNA
8. tcaaaaggcaagagagacggaaataaatggac	*Sisymbrium irio*, *V. harveyi* plasmid
9. ttgtttgattgcttgcattgaaccttgaa	16S rRNA

***Ft. novicida* Fx1 (28-29 bp)**

1. gccatctcttcataataaccactccaaa	-
2. taacatatggactggcatattcttccaa	-
3. ctaacgtagcgtaataaccagtaagtagg	-
4. ataagggtacgttcggtcaaatagctact	-
5. aaacacttctgtcgttttttcaaacattt	-
6. atggattattacttaactggagtgtttac	***Fn 3523_0992***

***Ft. novicida* FTG (27-30 bp)**

1. tttttgtgtatgtaatgtagctttcttta	-
2. atggttgtctaggtagtgtaaattccat	-
3. aagcagctgcttatgttttgacaatcaaa	-
4. tcaataaaaaatgcccattgactaacgg	-
5. atattgttttaaatcgtcgctatcagcaaa	-
6. caaatatagctaacttttggaactcag	-
7. tctattatgtttcctgaaagagtagtttt	*Naegleria gruberi*, mRNA
8. gcca**caaatactacaaaaaataacttaa**	***FN3523_1005***
9. ctaggcttggaatatctgtcagagata	-

***Ft. novicida* GA99-3549 (28-33 bp)**

1. ggctttctttagcttcttctagagtttt	*C. perfringens purD*
2. atataactatcaccagcaatctgtacgag	***Vibrio* phage ICP1_2004_A**
3. taaaatggttgtttagttatgtcagaaat	Rattus norvegicus BAC clone
4. aatttacgccgtatttcatttcatctcc	*Drosophila melanogaster*
5. gcagagatcattaaaccaatagcacagc	-
6. tctgtagaaattcctagtttttgagatg	*Homo sapiens*
7. ggctcaagtaatatcaataatttagaat	-
8. atttatgtgatgactggtgttgctaatg	-
9. acaattatatccatcacataccctctaa	*Aciduliprofundum boonei*
10.tgtattttagataggtttcagcagaaat	-
11.agtgcttaagaaaatcgaaagtatttggg	-
12.tataatgaaaatacgattattacttagaag	-
13.attgtcaaa**acataagcagctgcttcaaatat**	***FN3523_1009***
14.gtcgctttatcatctttatactccttaa	*Lactobacillus ruminis*
15.tttatctgctctctggctactttcaacac	*Homo sapiens*
16.gcgtgggtgttgggctttttattgcgttat	-
17.**cttgcaaagt**g**tagtgtgttacacctt**a	***Faecalibacterium prausnitzii***
18.attgtcaaa**acataagcagctgcttcaaatatc**	***FN3523_1009***

***Ft. novicida* GA99-3548 (29 bp)**

1. **atggattattacttaa**c**tggagtgt**ttac	***FN3523_0992***

***F. philomiragia* (27-31 bp)**

1. atctcatttagaaattcaatttttaaattct	***Fphi1603*** **(antisense)**
2. aaccctaaagattctaacggtgttatt	-
3. gaaatggtggagtaatgccaagcaacg	-
4. ggactattcttttaatgtttgatacata	-
5. acttgatactatatgtacgaacatttagaa	*Acinetobacter baumannii*

Blastn results of selected sequences exhibiting ≤25 identical base pairs.
Sequences underlined were found within the nucleotide sequence of the *Ft. novicida* phage. *FN3523_* and *Fphi1603*, gene number of *Ft. novocida*-like strain 3523 and *F. philomiragia*, respectively.

Table. 1 Genome sequences (species) used for CRISPR/Cas analysis

Species	Acession number	CRISPR-1	CRISPR-2
Ft. novicida U112	CP000439	**+**	**+**
Ft. novicida FTG	ABXZ01000001* ABXZ01000002*	**+**	**+**
Ft. novicida-like Fx1	CP002557	**+**	**+**
Ft. novicida-like GA99-3548	DS264589* DS264590*	**+**	**-**
Ft. novicida-like GA99-3549	DS264128* DS264129*	**+**	**+**
Ft. novicida-like 3523	CP002558	**+**	**-**
Ft. tularensis SchuS4	AJ749949	**+**	**-**
Ft. tularensis FSC198	AM286280	**+**	**-**
Ft. tularensis MA00-2987	DS990216*	**+**	**ni**
Ft. tularensis WY96-3418	CP000608	**+**	**-**
Ft. tularensis NE061598	CP001633.1	**+**	**-**
Ft. tularesnsis TIGB03	CP003048	**+**	**-**
Ft. tularensis TI0902	CP003049	**+**	**-**
Ft. holarctica LVS	NC_007880	**+**	**-**
Ft. holarctica FSC257	DS229056*	**+**	**ni**
Ft. holarctica FSC022	DS264138*	**+**	**ni**
Ft. holarctica FTNF002-00	CP000803	**+**	**-**
Ft. holarctica OSU18	CP000437	**+**	**-**
Ft. mediasiatica FSC147	CP000915	**+**	**-**
F. philomiragia ATCC25017	CP000937	**-**	**+**
F. philomiragia ATCC25015	DS99316*	**-**	**-**
F. philomiragia TX077308	CP002872	**-**	**-**
F. noatunensis subsp. *orientales*	CP003402	**-**	**-**

* contigs of shot gun sequences received from NCBI/GenBank/ "DBSource"
ni= not investigated

Table 2. Spacer sequences of CRISPR-1 (A) and CRIPR-2 (B) systems of various *Francisella* strains.

A. CRISPR-1	**Blastn (>25 bp in length)**
***Ft .novicida* U112 (34-37 bp)**	
1. caagattggtatatatcaactgctaatcactcgc	-
2. ggtaaaaccaaaaaccaaaagttaagatatctctt	-
3. tatgcggcttgtggttgtgtcattgcgaagactg	-
4. ttaagcggatcaatttcgattacatcttcaccagat	-
5. tttgtcaaatattggtcgagttatcttaaatagt	*Nostoc punctiforme* pNPUN03
6. actcctatatgaccatacaaaacctgcccattac	-
7. ttatttgccaacactcaaaagataggtaaccctatag	-
8. tagtttttattatctgtgctgttttctgtcattgc	-
9. gtatatatcagtagtttaaaagcattggctattact	-
10.gttaataatgctaaagaaccagctgaatgtgtaa	-
11.at**tcttgaattaagtgttgctg**ctgaggattaggca	**antisense of spacer 7 of *Ft. nov.* GA99-3549 (CRISPR-1)***
12.atataaatttggataattttaaggtattaaatccttt	-
13.atcacactataaaagtaatcaagcttgccattgt	-
***Ft. novicida* Fx1 (34-37 bp)**	
1. tctggatgtagtaaccaagtcttacgctcgataa	-
2. gttggatctgtctctctcattaatacaagattgtat	-
3. aataaagaaaatgtagacttattccaaaacgtta	-
4. agcttctcatatatgtatttactgtcgatttgat	-
5. ttttaataggtgatttgtgaattttgatgagttaaag	-
***Ft. novicida* FTG (34-38 bp)**	
1. aagtttgacttgttacccttttgaatttcatctcta	-
2. tcaatataccagctaggattactcgcataagcat	-
3. taatgcaaatttgtatagtctacttctgtagaataa	*Xenopus tropicales* mRNA, *Homo sapiens* DNA
4. **agtattttctcttgtcttgtaag**t**tctcttag**atag	**FN3523_0993**
5. taagaaacttcgtcggttcttgggtactatgtacta	-
6. agtattcttgacgcggctactaatacgtatcaaataaa	*Flavobacterium branchiophilum*
7. taaagcactctctaagatgacttcataattactac	-
8. aac**gatat**t**acacgcgaaagagctgcttttgaga**t	**FN3523_1002**
9. taagcaagactctcaaacagactttaaagcatcgttt	-
10.gctctatttaagtgttcgtgatactctgtttgtaa	-
11.acaagctaaagaaattattataattgcttactaca	-
12.tagcaagtattagtttacttcatttatttaaaat	-
13.agattatgtgtaagcaacttaattttaacgaacttc	-
14.atctaattggtaactcggaaagtataatcattg	-
15.acccaaataatctcgttatcatgaggctttatttg	-
16.tcctgaaatttccactggtctcccaagaaagcttg	-
17.ccgtatttagcctaatgttatctttaaatatttcaa	-
18.aagaattaaaacaagaaagcatgtaacaaacttagat	*M.truncatula* DNA
19.ctctttcttccccttttctgcttgtcgctcaataaa	-
20 ttattagaaaaaataaatgctggtatgttaaaattcg	*Homo sapiens* DNA
21.atcatagaagctaaatatagcggtaaaatgtttga	-
22.ctctttcttccccttttctgcttgtcgctcaataaa	-
***Ft. novicida* 3523 (34-37 bp)**	
1. ttgcca**gtataaccaccttcagcaaaacc**agtaaca	*Acinetobacter baumannii* **phage minor tail protein***
2. gccaaaaagactgagcctaaggctaaagttgagg	-
3. gtaataacagcctttgagcctttacagcctacaa	-
4. ccagcactatctgcaattgcaatgttgctttggt	-
5. attaaaaaagaaatatatgctgtatcgtatggca	-
6. catataacctgaattgtacccgtatcggcagcag	-

7. ttttttttatttgttttgcctgcaatgctcatagca	-
8. taacgcctcgaagttgaactttataacccaatatag	-
9. ttcatctaaccaatccccataatcgttccataattg	-
10.tcgacatactcttgatattcattatctggcttaa	-
11.aatatcgctagaaaacatattgttaagcatattat	-
12.ccgtttcaacaacaggcttgaagtttcaagcact	-
13.tcaccctcgccgccagtttgaccgcctccaccag	*Symbiobacterium thermophilum*
14.aattgtgaaatagataatatagcttctacattaga	*Mus musculus* BAC clone
15.ataaaaattaaataaaatatcttgaaatgcttatt	*Felis catus*
16.agagaatttgtatttagataccctaagattaagaaaa	-
17.tctggacgagtgttggttatagtcattgtcatacc	-

***Ft. novicida* GA99-3548 (35-37 bp)**

1. tcgacaggagtctgaccgtcaattaaagcagataatc	-
2. taggatttacgccagctaattcgttaactaggttca	-
3. gaactcgtcgtacatgatctccgccgccatgtgcag	-
4. cgctactataaataggggcggatataataaccgaa	-
5. ttcatcagcgaatcttttaaatttcttagatttgtca	-
6. gtacaaagagtttatttctcgtgttacttttttatt	Mus musculus BAC clone
7. **gcccaaattttctgctcaaattcgccttttttgcg**	**FN3523_0999**
8. **tatcttcaatatacaaatcaatataagagttatct**	**FN3523_1006**
9. **ctatattcaaacttgcatatatttctgata**gggct	**FN3523_1009**
10.attaattgcactagagaagcaagtccgaatctcggg	-

***Ft. novicida* GA99-3549 (34-36 bp)**

1. aagatagacagagctttattacaatcaagaaaagat	-
2. ccagttacgttatcgtctcgcactatcgcaggag	-
3. aatcatagagttaatttcgctcctgaccctgctg	-
4. tgattaataacacgataaacattacctttttgtttt	-
5. aatgattaaaatctacgaaactgatactcaaattc	-
6. agttgcctcgaatcgttccagcagatgtaaacat	-
7. aacccgcaa**cagcaacacttaattcaaga**gttgg	**antisense spacer 11 *Ft. nov* U112 (CRISPR-1)***
8. aaatctaaa**a**aatgtagattaataccaatacagcaa	B. subtilis plasmid

B. CRISPR-2 **Blastn (>22 bp in length)**

***Ft. novicida* U112 (26-32 bp)**

1. gagaagtcatttaataaggccactgttaaaa	-
2. gctactattcctgtgccttcagataattca	*Strongylocentrotus purpuratus*
3. gtctagagccttttgtattagtagccg	-
4. tagcgatttatgaaggtcatttttt	**intergenic region (*FN3523_1004-5*)**
5. agattaaaaggtaattctatcttgttgag	***FN3523_0998* (antisense)**
6. tacctagtagatacgcttactgataacaa	-
7. aaactttcatttatgatataaagtttttt	*Apis mellifera*, mRNA
8. tcaaaaggcaagagagacggaaataaatggac	*Sisymbrium irio*, *V. harveyi* plasmid
9. ttgtttgattgcttgcattgaaccttgaa	16S rRNA

***Ft. novicida* Fx1 (28-29 bp)**

1. gccatctcttcataataaccactccaaa	-
2. taacatatggactggcatattcttccaa	-
3. ctaacgtagcgtaataaccagtaagtagg	-
4. ataagggtacgttcggtcaaatagctact	-
5. aaacacttctgtcgtttttttcaaacattt	-
6. atggattattacttaactggagtgtttac	***Fn 3523_0992***

***Ft. novicida* FTG (27-30 bp)**

1. tttttgtgtatgtaatgtagctttcttta	-
2. atggttgtctaggtagtgtaaattccat	-
3. aagcagctgcttatgttttgacaatcaaa	-
4. tcaataaaaaatgcccattgactaacgg	-
5. atattgttttaaatcgtcgctatcagcaaa	-
6. caaatatagctaacttttggaactcag	-
7. tctattatgtttcctgaaagagtagtttt	*Naegleria gruberi*, mRNA
8. gcca**caaatactacaaaaaataacttaa**	***FN3523_1005***
9. ctaggcttggaatatctgtcagagata	-

***Ft. novicida* GA99-3549 (28-33 bp)**

1. ggctttctttagcttcttctagagtttt	*C. perfringens purD*
2. atataactatcaccagcaatctgtacgag	***Vibrio* phage ICP1_2004_A**
3. taaaatggttgtttagttatgtcagaaat	Rattus norvegicus BAC clone
4. aatttacgccgtatttcatttcatctcc	*Drosophila melanogaster*
5. gcagagatcattaaaccaatagcacagc	-
6. tctgtagaaattcctagtttttgagatg	*Homo sapiens*
7. ggctcaagtaatatcaataatttagaat	-
8. atttatgtgatgactggtgttgctaatg	-
9. acaattatatccatcacataccctctaa	*Aciduliprofundum boonei*
10.tgtattttagataggtttcagcagaaat	-
11.agtgcttaagaaaatcgaaagtatttggg	-
12.tataatgaaaatacgattattacttagaag	-
13.attgtcaaa**acataagcagctgcttcaaatat**	***FN3523_1009***
14.gtcgctttatcatctttatactccttaa	*Lactobacillus ruminis*
15.tttatctgctctctggctactttcaacac	*Homo sapiens*
16.gcgtgggtgttgggctttttattgcgttat	-
17.**cttgcaaagt**g**tagtgtgttacacctt**a	***Faecalibacterium prausnitzii***
18.attgtcaaa**acataagcagctgcttcaaatatc**	***FN3523_1009***

***Ft. novicida* GA99-3548 (29 bp)**

1. **atggattattacttaa**c**tggagtgt**ttac	***FN3523_0992***

***F. philomiragia* (27-31 bp)**

1. atctcatttagaaattcaatttttaaattct	***Fphi1603*** **(antisense)**
2. aaccctaaagattctaacggtgttatt	-
3. gaaatggtggagtaatgccaagcaacg	-
4. ggactattcttttaatgtttgatacata	-
5. acttgatactatatgtacgaacatttagaa	*Acinetobacter baumannii*

---

*Blastn results of selected sequences exhibiting ≤ 25 identical base pairs
Sequences underlined were found within the nucleotide sequence of the *Ft. novicida* phage. *FN3523_* and *Fphi1603*, gene number of *Ft. novocida*-like strain 3523 and *F. philomiragia*, respectively.

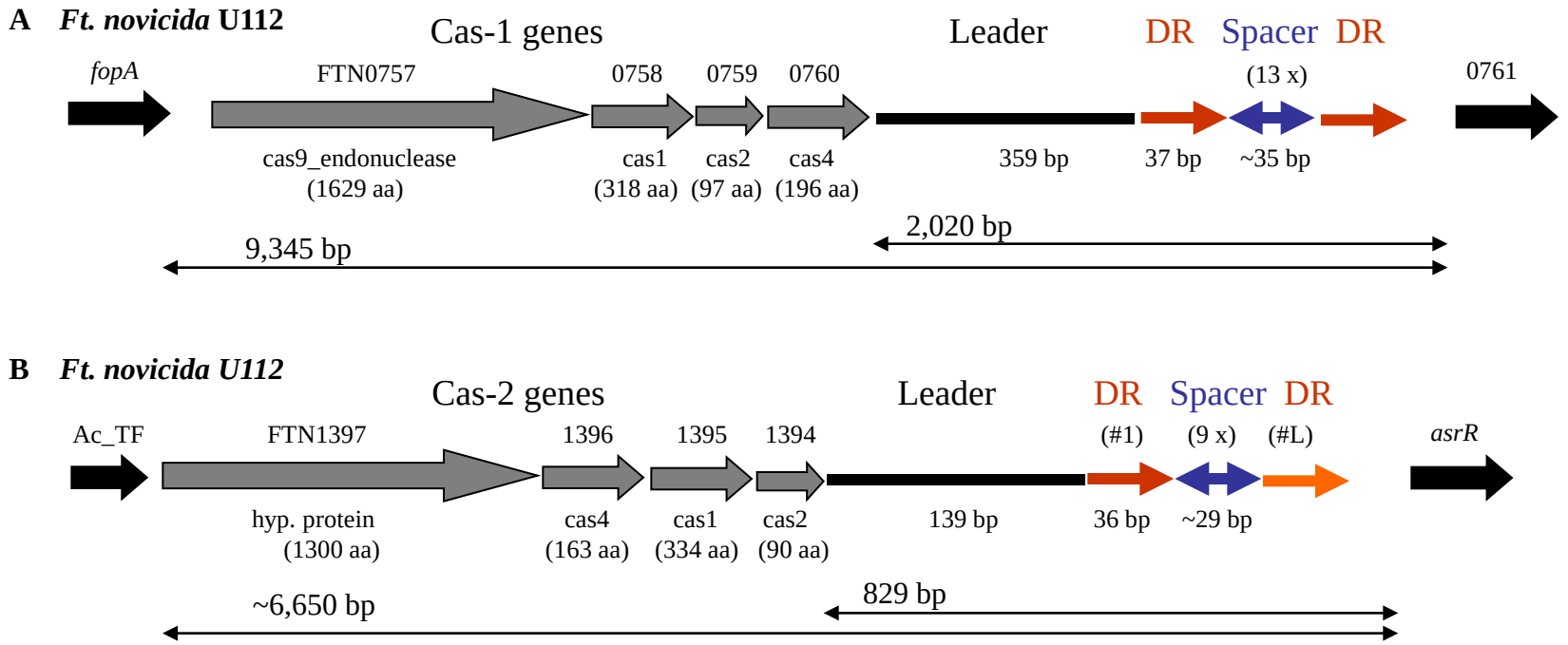

A
Ft. novicida U112
Cas-1 genes
Leader
DR
Spacer
DR
fopA
FTN0757
0758
0759
0760
(13 x)
0761
cas9_endonuclease
(1629 aa)
cas1
(318 aa)
cas2
(97 aa)
cas4
(196 aa)
359 bp
37 bp
~35 bp
2,020 bp
9,345 bp
B
Ft. novicida U112
Cas-2 genes
Leader
DR
Spacer
DR
Ac_TF
FTN1397
1396
1395
1394
(#1)
(9 x)
(#L)
asrR
hyp. protein
(1300 aa)
cas4
(163 aa)
cas1
(334 aa)
cas2
(90 aa)
139 bp
36 bp
~29 bp
~6,650 bp
829 bp

Fig. 1

A *Ft. novicida* U112 CRISPR-1

atgatcgttgatagtaaaacctctaattgccactgaaggtggaaatctcgcaaatatttagaaatctccatattaaatttctcagccatcttgatgtaatcgagctaagaagcgaacaagatatattaattgccaaacaaaaaaagccacttaggtggcaaatttttgctttttagttattcagacgtgtcaaacagaggtccgttcaaaatacttttaaatgattacagagcattaatttatttggtacatttataatttagatatttttttcgcaaaatgtctacgttttttttacaattagggtatttgcgaaaaaataccattttaaaaaactctcaaagccttgcaaacactaggtttgttttggctctaacagtagtttaccaaataattcagcaactgaaaccaagattggtatatatcaactgctaatcactcgcctaacagtagtttaccaaataattcagcaactgaaacggtaaaaccaaaaaccaaaagttaagatatctcttctaacagtagtttaccaaataattcagcaactgaaactatgcggcttgtggttgtgtcattgcgaagactgctaacagtagtttaccaaataattcagcaactgaaacttaagcggatcaatttcgattacatcttcaccagatctaacagtagtttaccaaataattcagcaactgaaacttgtcaaatattggtcgagttatcttaaatagtctaacagtagtttaccaaataattcagcaactgaaacactcctatatgaccatacaaaacctgcccattacctaacagtagtttaccaaataattcagcaactgaaacttatttgccaacactcaaaagataggtaaccctatagctaacagtagtttaccaaataattcagcaactgaaactagttttttattatctgtgctgttttctgtcattgcctaacagtagtttaccaaataattcagcaactgaaacgtatatatcagtagtttaaaagcattggctattactctaacagtagtttaccaaataattcagcaactgaaacgttaataatgctaaagaaccagctgaatgtgtaactaacagtagtttaccaaataattcagcaactgaaacattcttgaattaagtgttgctgctgaggattaggcactaacagtagtttaccaaataattcagcaactgaaacatataaatttggataattttaaggtattaaatccttctaacagtagtttaccaaataattcagcaactgaaacatcacactataaaagtaatcaagcttgccattgtctaacagtagtttaccaaataattcagcaactgaaacttaattaattttgtaaattattcctaaaagtttctagttggacacattgtatcaaataatataatcacgaaaacagttttacagcaaaatgttgat

B *Ft. novicida* U112 CRISPR-2

aaaaaattgcaaaccttagtctttatgttaaaataactactaagttcttagagatatttaaaaatatgactgttgttatatatcaaaatgctaaaaaaatcatagattttaggtctttttttgctgatttaggcaaaaacgggtctaagaactttaaataatttctactgttgtagatgagaagtcatttaataaggccactgttaaaagtctaagaactttaaataatttctactgttgtagatgctactattcctgtgccttcagataattcagtctaagaactttaaataatttctactgttgtagatgtctagagcctttgtattagtagccggtctaagaactttaaataatttctactgttgtagattagcgatttatgaaggtcatttttttgtctaagaactttaaataatttctactgttgtagatagattaaaaggtaattctatcttgttgaggtctaagaactttaaataatttctactgttgtagatacctagtagatacgcttactgataacaagtctaagaacttaaataatttctactgttgtagataaactttcatttatgatataaagtttttgtctaagaactttaaataatttctactgttgtagattcaaaaggcaagagagacggaaataaatggacgtctaagaactttaaataatttctactgttgtagatttgtttgattgcttgcattgaaccttgaagtctaagaactttaaataatttgtctgtatattattgatttctaaattagaatttttctaatatatcactacttgagaatttatagg

Fig. 2

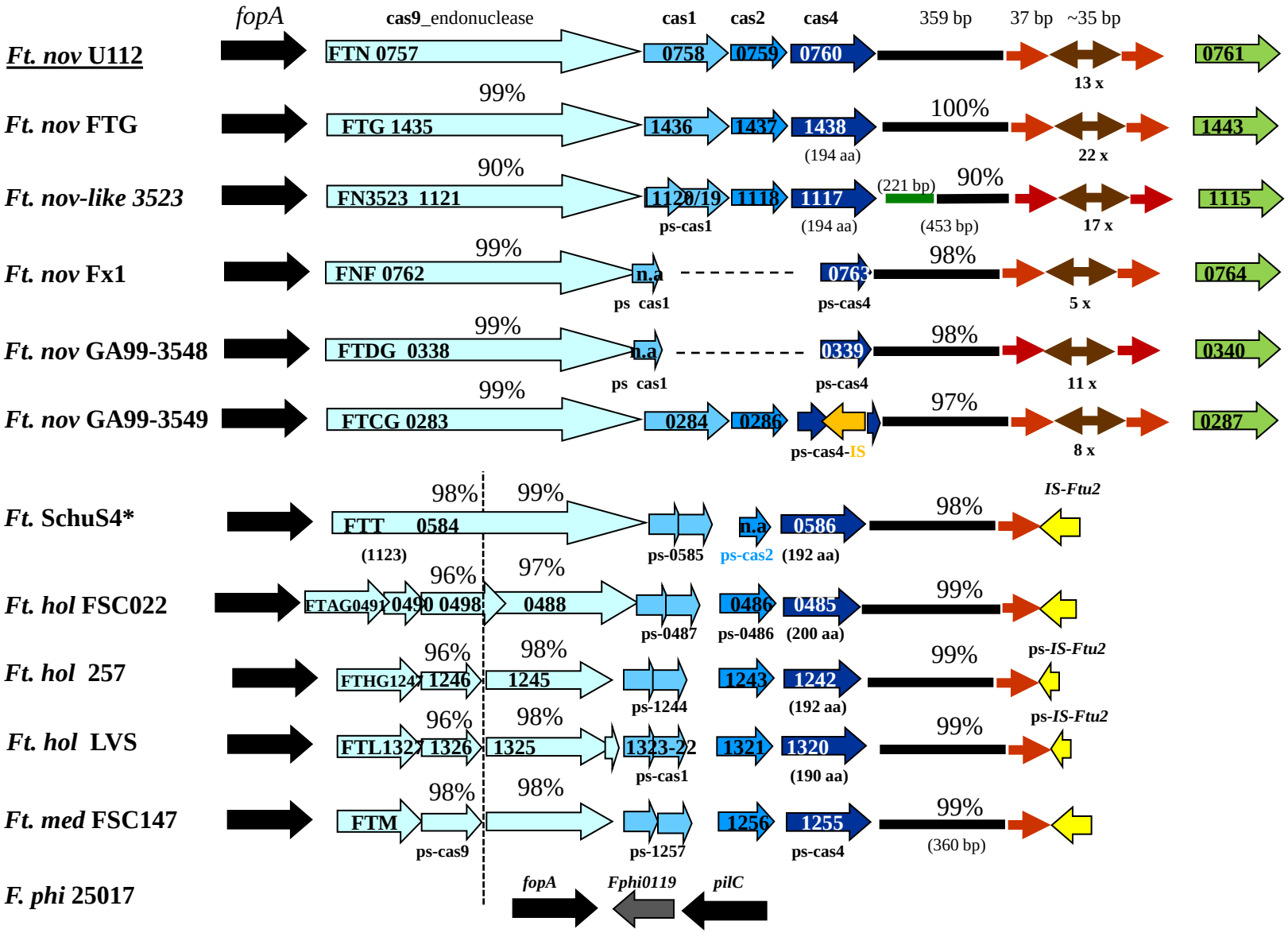

fopA
cas9_endonuclease
cas1
cas2
cas4
359 bp
37 bp
~35 bp
Ft. nov U112
FTN 0757
0758
0759
0760
0761
13 x
99%
Ft. nov FTG
FTG 1435
1436
1437
1438
100%
1443
(194 aa)
22 x
90%
Ft. nov-like 3523
FN3523 1121
1120/19
1118
1117
(221 bp)
90%
1115
ps-cas1
(194 aa)
(453 bp)
17 x
99%
Ft. nov Fx1
FNF 0762
n.a
98%
0764
ps cas1
ps-cas4
5 x
99%
Ft. nov GA99-3548
FTDG 0338
n.a
0339
98%
0340
ps cas1
ps-cas4
11 x
99%
Ft. nov GA99-3549
FTCG 0283
0284
0286
97%
0287
ps-cas4-IS
8 x
98%
99%
IS-Ftu2
Ft. SchuS4*
FTT 0584
n.a
0586
98%
(1123)
ps-0585
ps-cas2
(192 aa)
96%
97%
Ft. hol FSC022
FTAG0491
0490
0498
0488
0486
0485
99%
ps-0487
ps-0486
(200 aa)
96%
98%
ps-IS-Ftu2
Ft. hol 257
FTHG1247
1246
1245
1243
1242
99%
ps-1244
(192 aa)
96%
98%
ps-IS-Ftu2
Ft. hol LVS
FTL1327
1326
1325
1323-22
1321
1320
99%
ps-cas1
(190 aa)
98%
98%
Ft. med FSC147
FTM
1256
1255
99%
ps-cas9
ps-1257
ps-cas4
(360 bp)
F. phi 25017
fopA
Fphi0119
pilC

Fig. 3

A. CRISPR/Cas-1, DRs

```
Ft.nov U112      ctaacagtagtttaccaaataattcagcaactgaaac
Ft.nov FTG       ....................................
Ft.nov Fx1       ....................................
         (2.)    ....----------.....................
Ft.nov 3523      .......c...a..t........ct...t........
         (#L)    .......c...a..t........ct...t...t.t.a
Ft.nov GA-48     ....................................
Ft.nov GA-49     ....................................
Ft.tul MA00      ....................................
Ft.tul Schu      ....................................
Ft.hol FSC022    ....................................
Ft.hol 257       ....................................
Ft.hol  LVS      ....................................
Ft.med FSC147    a...................................
```

B. CRISPR/Cas-2, DRs

#1 repeat

```
Ft.nov U112      gtctaagaactttaaataatttctactgttgtagat
Ft.nov Fx1       ...................................
Ft.nov FTG       ...................................
Ft.nov GA-49     ...................................
Ft.nov GA-48*    ..a...c............................
Ft.tul Schu*     ..a...c............................
F.phi 25017      ...................................
```

#L repeat

```
Ft.nov U112      ......................g.ctgta.a.tat.
Ft.nov Fx1       ......................g.ct.ta.a.tat.
Ft.nov FTG       ......................gcctgta.a.tat.
Ft.nov GA-49     c.....................cctata.a.tat.
Ft.nov GA-48*    ......................g.ct.ta.a.tat.
Ft.nov 3523*     a.....c...............gc.aa.atatt..
Ft.tul Schu*     ......t...............g.ct.ta.a.tat.
F. phi 25017     ..........a.....a....c.gttaa.gtag.t.
```

Fig. 4

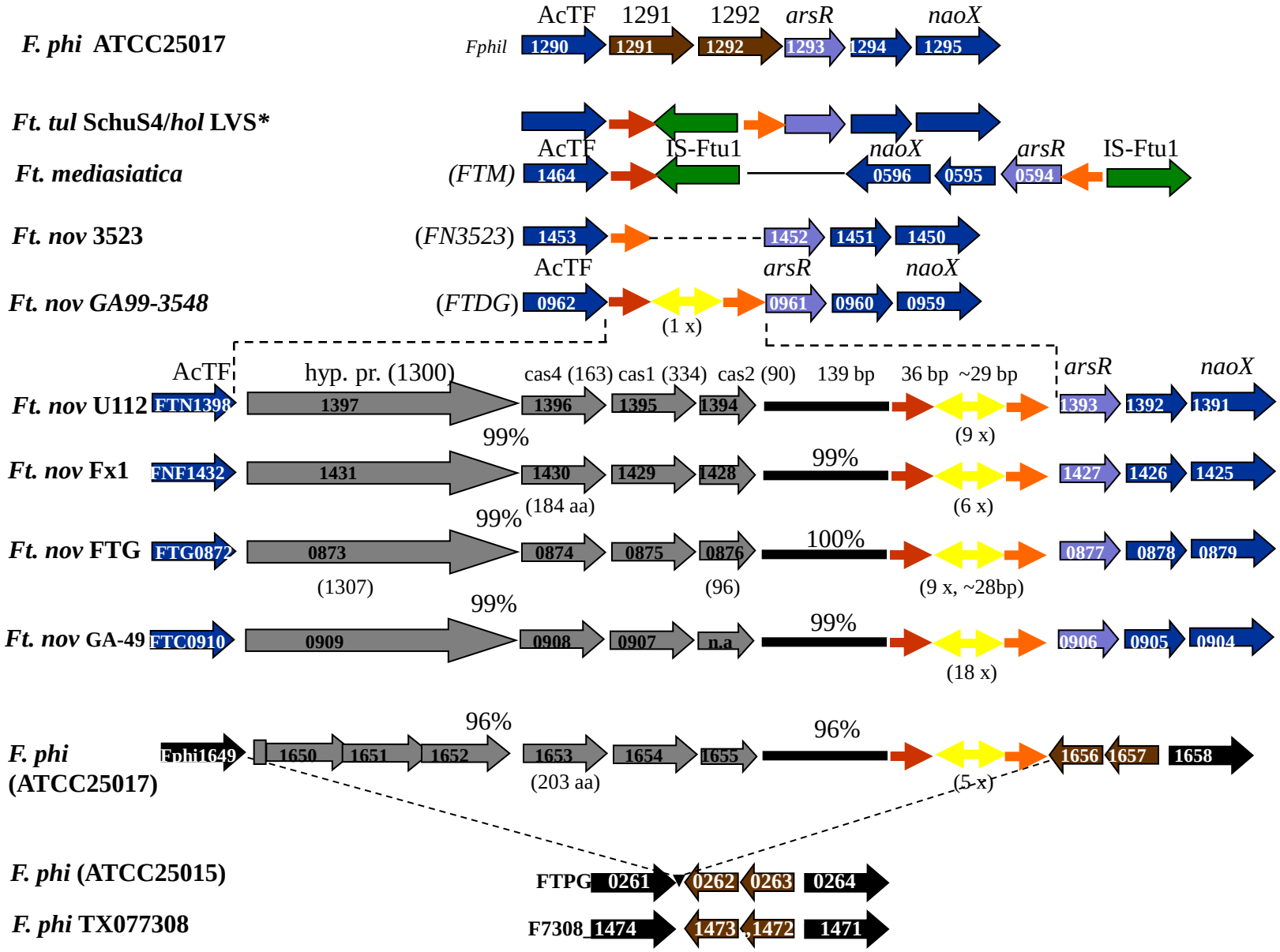
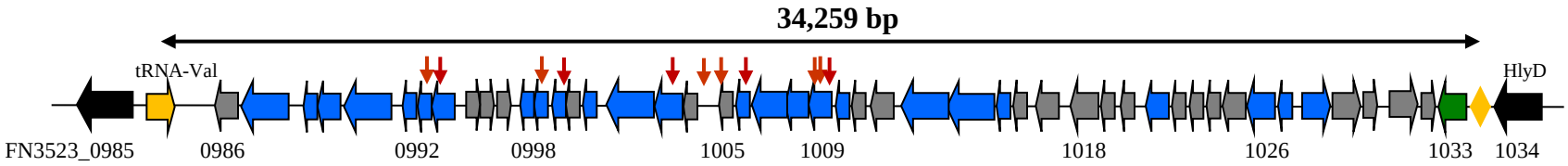

F. phi ATCC25017
AcTF
1291
1292
arsR
naoX
Fphil
1290
1291
1292
1293
1294
1295
Ft. tul SchuS4/hol LVS*
Ft. mediasiatica
AcTF
IS-Ftu1
naoX
arsR
IS-Ftu1
(FTM)
1464
0596
0595
0594
Ft. nov 3523
(FN3523)
1453
1452
1451
1450
Ft. nov GA99-3548
AcTF
arsR
naoX
(FTDG)
0962
0961
0960
0959
(1 x)
Ft. nov U112
AcTF
hyp. pr. (1300)
cas4 (163)
cas1 (334)
cas2 (90)
139 bp
36 bp
~29 bp
arsR
naoX
FTN1398
1397
1396
1395
1394
1393
1392
1391
(9 x)
Ft. nov Fx1
99%
99%
FNF1432
1431
1430
1429
1428
1427
1426
1425
(184 aa)
(6 x)
Ft. nov FTG
99%
100%
FTG0872
0873
0874
0875
0876
0877
0878
0879
(1307)
(96)
(9 x, ~28bp)
Ft. nov GA-49
99%
99%
FTC0910
0909
0908
0907
n.a
0906
0905
0904
(18 x)
F. phi
(ATCC25017)
96%
96%
Fphi1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
(203 aa)
(5 x)
F. phi (ATCC25015)
FTPG
0261
0262
0263
0264
F. phi TX077308
F7308
1474
1473
1472
1471

Fig. 5

A

***Ft. novicida* phage (48 ORFs)**



tRNA-Val ACTCGCATAGCTCAGTTGGTTAGAGTACTACCTTGACATGGTAGGGGTCACTGGTTCGAATCCAGTTGCGAGTACCA

repeat-1 CCTTGACATGGTAGGGGTCACTGGTTCGAATCCAGTTGCGAGTACCA

B

0986: 197 aa, 2x internal repeat, hypothetical protein
0987: 408 aa, phage_GPD, GpD, phage late control D protein
0988: 66 aa, phage_tail_X, GpX
0989: 138 aa, Phage_P2_GpU,
0990: 607 aa, phage tail tape measure protein
0991: 89 aa, Flu_Mu_gp41
0992: 159 aa, phage, major tail tube protein,
0993: 385 aa, tail sheath protein
0994: 106 aa, hypothetical protein
0995: 81 aa, hypothetical protein
0996: 66 aa, hypothetical protein
0997: 198 aa, phage_base_V, baseplate assembly GpV
0998: 163 aa, putative phage-related protein
0999: 166 aa, minor tail_Z
1000: 111 aa,
1001: 63 aa, putative phage-related protein
1002: 600 aa, Peptidase_U35 + major Phage_capsid; (E. phage vB_EcoM_ECO1230-10[~42.000 bp], dsDNA-virus)
1003: 473 aa, Phage_Portal_2, capsid protein, froming DNA-ejection hole (lambda familie)
1004: 88 aa, coiled-coil,
1005: 110 aa, conserved hypothetical protein
1006: 183 aa, putative phage protein
1007: 324 aa, phage tail fiber protein
1008: 195 aa, tail_P2_I, phage tail protein I
1009: 281 aa, baseplate_J
1010: 112 aa, Gpw_gp25, phage baseplate, lysozym activity
1011: 131 aa, 1x TM, coiled-coil
1012: 196 aa, 1x TM
1013: 602 aa, phage terminase GpA
1014: 671 aa, (phage related)
1015: 93 aa, virus coat protein
1016: 172 aa, FTG_0292
10017 : 387 aa, FTG_0291
1018: 422 aa, phage integrase,
1019: 64 aa,
1020: 72 aa,
(n.a) : 87 aa, SP, coilded coil
1021: 239 aa, phage rha gene, phage anti-repressor (circular form)
1022: 120 aa, YopX_SF,putative chaperone
1023: 100 aa, coiled-coil,
1024: 85 aa,
1025: 212 aa,
1026: 240 aa, Phg_2220_C SF, putative phage replication protein
1027: 56 aa, gp55, Cro EBPR podovirus 1
1028: 265 aa, H-T-H XRE, pro-phage repressor
1029: 293 aa,
1030: 112 aa,
1031: 264 aa, H-T-H,
1032: 130 aa,
1033: 375 aa, site-specific recombinase

Fig. 6