RESEARCH LETTER

Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*

Ekaterina Semenova[1], Maxim Nagornykh[1], Mikhail Pyatnitskiy[2], Irena I. Artamonova[3,4] & Konstantin Severinov[1,5,6]

[1]Department of Molecular Biology and Biochemistry, Waksman Institute, Rutgers, The State University, Piscataway, NJ, USA; [2]Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia; [3]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia; [4]Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia; [5]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia; and [6]Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia

Abstract

Clustered regularly interspaced short palindromic repeat (CRISPR) is a bacterial immunity system that requires a perfect sequence match between the CRISPR cassette spacer and a protospacer in invading DNA for exclusion of foreign genetic elements. CRISPR cassettes are hypervariable, possibly reflecting different exposure of strains of the same species to foreign genetic elements. Here, we determined CRISPR cassette sequences of two *Xanthomonas oryzae* strains and found that one of the strains remains sensitive to phage Xop411 despite carrying a cassette that has a spacer exactly matching a fragment of the Xop411 genome. To explain this apparent paradox, we identified *X. oryzae* CRISPR spacers of likely phage origin and defined a consensus sequence of a motif adjacent to *X. oryzae* phage protospacers. Our analysis revealed that the Xop411 protospacer that matches the CRISPR spacer has this motif mutated, which likely explains the phage's ability to infect its host. While similar observations were made previously with *Streptococcus thermophilus* and its phages, the conserved motif in *X. oryzae* phages is located on a protospacer side opposite to the *S. thermophilus* phages' motif. The results thus point to a considerable degree of variety of CRISPR-mediated phage resistance mechanisms in different bacteria.

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPRs) – short (21–47 bp) direct repeats interspaced with unrelated similarly sized nonrepetitive sequences (spacers) – were found in *c.* 40% of the analyzed eubacterial and most of the archaeal genomes (Nakata *et al.*, 1989; Mojica *et al.*, 1995; van Belkum *et al.*, 1998; Jansen *et al.*, 2002; Andersson & Banfield, 2008; Kunin *et al.*, 2008; Sorek *et al.*, 2008). CRISPR loci usually contain an adjacent nonrepetitive leader region and several CRISPR-associated (*cas*) genes (Haft *et al.*, 2005). Some CRISPR cassette spacers are similar to phage and plasmid sequences (Bolotin *et al.*, 2005). Makarova *et al.* (2006) suggested that CRISPRs and *cas* genes form a bacterial immune system that functions through an RNAi-like mechanism and is directed against extrachromosomal genetic elements.

In *Streptococcus thermophilus*, acquisition of a phage sequence (called a 'protospacer') as a leader-proximal CRISPR spacer leads to phage resistance (Barrangou *et al.*, 2007). *Streptococcus thermophilus* phages that overcame CRISPR-based resistance contain single-nucleotide substitutions in their protospacers. CRISPR-based immunity can also be overcome by changing sequences immediately adjacent to protospacers (Deveau *et al.*, 2008). In *S. thermophilus* phages, such sequences form a conserved 3–5-bp-long motif. Curiously, in *Escherichia coli*, there appears to be no requirement for a conserved motif adjacent to protospacers because apparently random (although appropriately sized) fragments of phage λ genome inserted as spacers in the CRISPR cassette were reported to lead to phage resistance, at least when positioned on a plasmid, transcribed by T7 RNA polymerase, and supplied by Cas proteins expressed from another plasmid (Brouns *et al.*, 2008).

Data from archaea (Lillestol *et al.*, 2006; Hale *et al.*, 2008) and *E. coli* (Brouns *et al.*, 2008) indicate that CRISPR cassettes are transcribed; the transcripts are processed to *c.* 60-base fragments that contain the entire spacer as well as portions of upstream and downstream repeats. Analysis of *cas* mutants suggests that in *E. coli*, processing of CRISPR cassette transcript is required for phage resistance (Brouns *et al.*, 2008).

While some bacteria lack CRISPRs, others are particularly well endowed. For example, the three completely sequenced strains of plant pathogen *Xanthomonas oryzae* have CRISPR cassettes containing 48 (strain MAFF 311018), 59 (strain KACC10331), and 75 (strain PXO99A) spacers (Salzberg *et al.*, 2008). Several *X. oryzae* phages have been completely sequenced (Yuzenkova *et al.*, 2003; Lee *et al.*, 2007), making *X. oryzae* and its phages a potentially useful model for studies of CRISPR function in phage resistance. In this work, we determined the sequences of CRISPR cassettes of two additional *X. oryzae* strains, demonstrated transcription and processing of CRISPR cassettes in both strains, and, through bioinformatics analysis, explained why one of these strains serves as a host for *X. oryzae* phage Xop411 despite harboring a spacer identical to a protospacer in the phage. Our results indicate that a conserved motif is present next to protospacer sequences from *X. oryzae* phages but is located on a different side (with respect to spacer orientation relative to CRISPR cassette transcription) than in *S. thermophilus* phage protospacers. This finding suggests a considerable mechanistic variety in the way CRISPR machineries function in different bacteria.

Materials and methods

Bacterial and phage growth conditions

Xanthomonas oryzae strain Xo604 was grown at 30 °C in TGSC medium (Yuzenkova *et al.*, 2003). *Xanthomonas oryzae* strain Xo21 was grown at 30 °C in TYG medium (Lee *et al.*, 2007). Exponentially growing *X. oryzae* cells were infected with phages at a multiplicity of infection of 5 as described (Semenova *et al.*, 2005).

CRISPR sequencing

The putative leader sequence of Xo604 and Xo21 was determined first. Leader fragments of 187 bp were amplified from genomic DNA using primers specific to *cas2* gene and CRISPR repeat (sequences of primers are available from the authors upon request) and sequences. Next, Xo604 and Xo21 CRISPR cassettes were amplified using a primer annealing in the leader sequence and another one specific to transposase gene flanking CRISPR cassette in three *X. oryzae* strains with known genome sequenced. PCR products (*c.* 5 kb for Xo21 and *c.* 2.5 kb for Xo604) were

sequenced using a primer-walking strategy. The sequences of CRISPR cassettes of Xo604 and Xo21 were deposited in GenBank (accession numbers: FJ668938 and FJ668939).

RNA isolation and Northern blot hybridization

Total RNA samples (*c.* 20 μg) isolated using TRIzole reagent (Invitrogen) were separated on 8 M urea – 8% polyacrylamide gels alongside ^{32}P-end radiolabeled RNA markers (Decade, Ambion), transferred to Hybond-XL membrane (GE Healthcare) and cross-linked. Membranes were hybridized at 42 °C with repeat-specific (probe 1: 5′-GTTTC AATCCACGCGCCCGTGAGGACGCGAC-3′) and spacer-specific (probe 2: 5′-TTCACACCATTTATAACGCCGGGGT TATTCTGATAT-3′; probe 3: 5′-ATATCAGAATAACCCCG GCGTTATAAATGGTGTGAA-3′) ^{32}P-end-labelled oligonucleotides in ExpressHyb hybridization solution (Clontech) according to manufacturer's protocol. After washing, hybridization products were revealed using PHOSPHORIMAGER (Molecular Dynamics).

Bioinformatics

Spacers of CRISPR cassettes from *X. oryzae* strains KACC10331 and MAFF 311018 were downloaded from CRISPRdb (Grissa *et al.*, 2007). CRISPR cassette from strain PXO99A was retrieved from GenBank contig PXO99A (accession number CP00967, Salzberg *et al.*, 2008) and annotated manually.

To extract *X. oryzae* phage protospacers, spacers from each of the five *X. oryzae* CRISPR cassettes were subjected to BLAST search against the viruses subsection (txid10239) of GenBank nr-collection using BLASTN program with *E*-value threshold set at 1.0 and all other parameters set by default. Nonphage BLAST hits were filtered out. Flanking sequences of 100 nt were obtained for each resulting hit; in order to reveal the true length of each protospacer, BLAST hits were extended, when necessary, in one or both directions to match the length of the spacer. In cases when several BLAST hits from related phage genomes were obtained for the same spacer, a consensus of derived protospacers and their flanking sequences was used as protospacer for this spacer. Motif logos were built using WEBLOGO (Crooks *et al.*, 2004).

Results

Sequencing and sequence analysis of *X. oryzae* CRISPR cassettes

Xanthomonas oryzae strain Xo604 is a host for a well-studied bacteriophage Xp10 (Yuzenkova *et al.*, 2003; Semenova *et al.*, 2005); Xo604 does not propagate Xop411, an Xp10 relative. Conversely, *X. oryzae* Xo21 is a host of Xop411 (Lee

et al., 2007), but does not propagate Xp10. The absence of productive infection of Xo604 by Xop411 (and Xo21 infection by Xp10) is not due to lack of absorption of either phage to nonpermissive bacteria (data not shown). We wondered whether host specificity could be caused by the function of CRISPR system in one or both hosts. To this end, we sequenced CRISPR cassettes in both strains. The Xo21 CRISPR cassette was found to contain 77 spacers; the Xo604 cassette contains 37 spacers. In Fig. 1, the results of comparison of Xo21 and Xo604 CRISPR cassettes with previously known cassettes from three other *X. oryzae* strains are schematically presented. As observed in other systems, spacers at one end of *X. oryzae* CRISPR cassettes (furthest from the leader and presumably corresponding to spacers acquired earlier in time) are more similar to each other than spacers at the other end of the cassette. The numbering system for spacers used here follows the one used previously (Salzberg *et al.*, 2008), with the 'oldest' spacer (rightmost in Fig. 1) assigned number 1. The first four spacers (S1–S4) are the same in all five strains. Spacers at the left are more divergent; this diversity may account for differential phage resistance/sensitivity of the strains. The repeats are identical in all five strains.

Analysis of *X. oryzae* CRISPR cassette spacers revealed the presence of multiple sequences that originate from known *X. oryzae* phages (Supporting Information, Table S1). One hundred different *X. oryzae* CRISPR spacers of phage origin matched fully or partially genomic sequences from a closely knit family of *Xanthomonas* phages that includes OP1, Xp10, and Xop411 (Lee *et al.*, 2007). Locations of sequences identical or similar to *X. oryzae* CRISPR spacers in the genome of Xp10, the best-studied phage of the family (Nechaev *et al.*, 2002; Yuzenkova *et al.*, 2003; Semenova *et al.*, 2005), are shown in Fig. 2. As can be seen, sequences matching CRISPR spacers are located in both early (29 sequences) and late (70 sequences) phage genes (the latter genes code for virion proteins). One spacer matches a

sequence in the intergenic region of the phage. For early genes, 20 out of 29 matching sequences are transcribed in the same direction as CRISPR cassette spacers (see below and Fig. 4). For late genes, the direction of transcription of 25 out of 70 matching sequences coincides with the direction of spacer transcription. Assuming that all CRISPR spacers of clearly phage origin had, at least at some point, produced phage resistance, the distribution of spacer-matching sequences in the Xp10 genome suggests that in *X. oryzae* CRISPR system restricts phage growth by targeting DNA, not RNA, in agreement with a recent observation made in *Staphylococcus epidermis* (Marraffini & Sontheimer, 2008).

Neither the Xo21 nor Xo604 CRISPR cassettes contained spacers fully identical to sequences from Xp10. Thus, the inability of Xp10 to infect Xo21 is unlikely to be caused by CRISPR-mediated resistance. Conversely, both Xo21 and Xo604 contained a spacer that was identical, over its entire length, to a sequence found in the Xop411 genome (Lee *et al.*, 2007). In Xo21, this spacer was the last (leftmost) one in the cassette (S77 in Fig. 1); in Xo604, this spacer occupied the third position from the end (S35). CRISPR cassettes from other *X. oryzae* strains did not contain such a spacer (Fig. 1). Control resequencing of the relevant genomic portion of Xop411 used in this work revealed no differences from the published sequence (Lee *et al.*, 2007) (data not shown). Thus, Xop411 infection of Xo21 proceeds normally, despite the fact that in this strain, a CRISPR spacer that is thought to contribute the most to phage resistance is identical to a protospacer in Xop411.

Analysis of *X. oryzae* CRISPR transcripts

Xop411 may be able to infect Xo21 because its CRISPR system is defective. To test this notion, we determined if processing of CRISPR cassette transcripts occurs in Xo21 and Xo604. To this end, total RNA was purified from
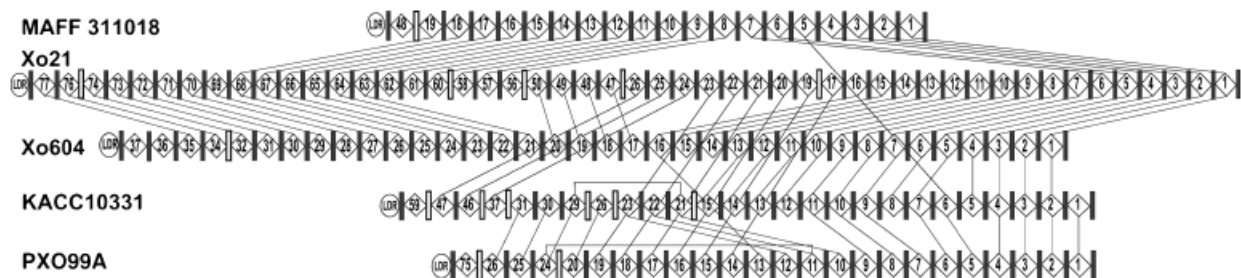

Fig. 1. The structure of CRISPR cassettes of *Xanthomonas oryzae*. CRISPR cassettes from five *X. oryzae* strains are schematically shown and identified by strain name at the left. On the left, leader sequences are shown by ovals labeled 'LDR.' Spacers are shown as diamonds numbered according to the nomenclature proposed in Makarova *et al* (2006). Spacer numbers match those in Table S1. Repeats are indicated as rectangles. Repeats indicated by dark fill separate adjacent spacers; repeats indicated as white rectangles separate nonadjacent spacers (note the 'missing' spacers are unique for individual cassettes and are not shown; the only unique spacers shown are terminal spacers adjacent to the leader). Identical spacers in different cassettes or within a same cassette are connected by lines.

uninfected Xo21 and Xo604, as well as Xo21 and Xo604 cells infected with Xp10 or Xop411 at a high multiplicity of infection and collected midway through the infection, and analyzed by Northern blotting with three different probes. Probe 1 was complementary to the CRISPR repeat and should therefore reveal CRISPR RNA transcribed in the rightward direction (Fig. 1). Probe 2 was complementary to Xo21 S77 and Xo604 S35 spacers and should also reveal rightward-transcribed RNA. This probe is also corresponds to a fragment of gene 21R of both Xp10 and Xop411 (Fig. 2). Gene 21R codes for peptidoglycan hydrolase and is expressed late in infection. Probe 3 was complementary to probe 2 and should reveal leftward-transcribed RNA. A defined radioactive band with an approximate size of 60 bases was observed with both repeat-specific (Fig. 3a) and

spacer-specific (Fig. 3b) probes designed to reveal rightward transcription. The robust signal obtained with both probes is in stark contrast with results obtained in *E. coli*, where steady-state levels of the spacer-specific transcript were found to be very low (Brouns *et al.*, 2008). The hybridization signal with repeat-specific probe 1 was much stronger than with probe 2, presumably reflecting the fact that the hybridized sequence is repeated multiple times in the genome. Probe 1 revealed several additional bands that likely correspond to incompletely processed ($120\,\text{nt} = 2 \times 60\,\text{nt}$; $180\,\text{nt} = 3 \times 60\,\text{nt}$, etc.) CRISPR transcripts. Probe 2 revealed, in addition to a 60-nt band, two bands of *c.* 55 and 50 nt, that did not hybridize with the repeat-specific probe. Spacer-specific probe 3 did not reveal any transcripts in uninfected cells (Fig. 3c).
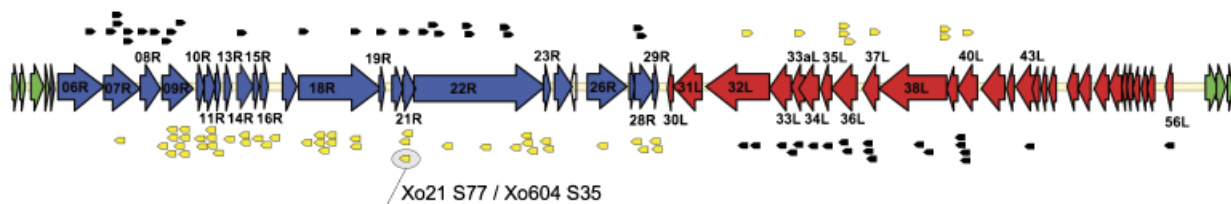

Fig. 2. Distribution of protospacers in the genome of *Xanthomonas oryzae* bacteriophage Xp10. The bar at the top schematically illustrates the 44 373-bp genome of bacteriophage Xp10. Temporal classes of genes (early, middle, and late) are indicated by red, green, and blue colors, respectively. Directions of arrows indicate direction of transcription. Above and below the Xp10 genome, positions of protospacers (sequences corresponding to spacers from known *X. oryzae* CRISPR cassettes) are shown. The position of protospacers above or below the Xp10 genome reflects their orientation (with respect to CRISPR spacers transcribed in the rightward direction, Fig. 1). Transcripts of spacers corresponding to protospacers shown at the top match the top strand of the phage genome; transcripts of spacers shown at the bottom match the bottom strand. The direction of transcription of protospacers shown in black matches the direction of transcription of corresponding spacers. Protospacers shown in yellow are transcribed in a direction opposite to the direction of transcription of the corresponding spacers. The protospacer corresponding to spacers S77 of Xo21 and S35 of Xo604 (used as a probe in Northern-blotting experiments) is highlighted by a gray oval.
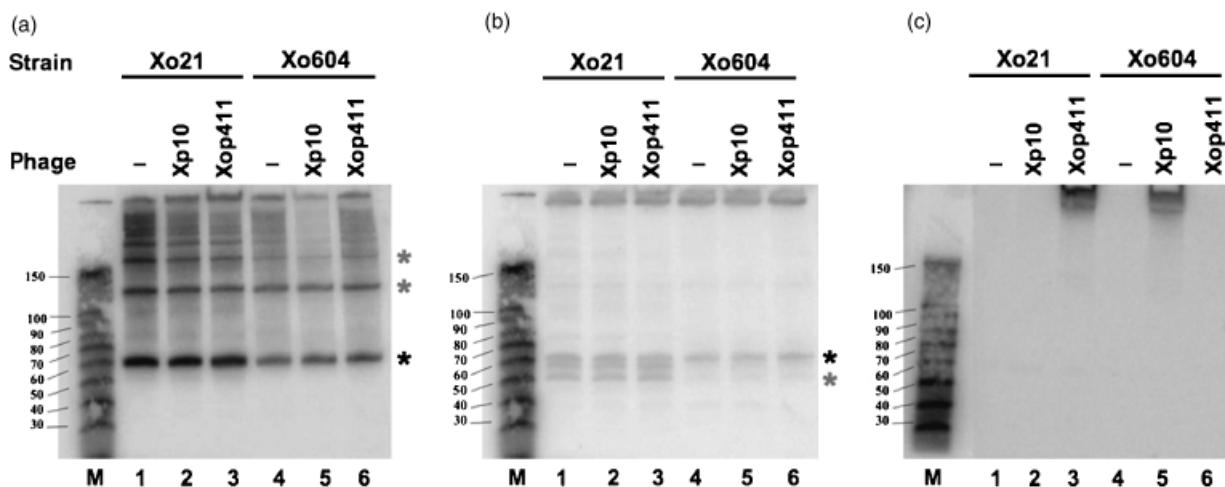

Fig. 3. Northern blot analysis of *Xanthomonas oryzae* CRISPR transcripts. Total RNA was isolated from exponentially growing *X. oryzae* Xo604 and Xo21 cells, as well as cells infected with Xop411 and Xp10. After blotting, RNA was hybridized to probe 1 (specific to CRISPR repeat and designed to reveal rightward CRISPR transcripts, panel A), probe 2 [specific to spacers S77/S35 of Xo21 and Xo604, correspondingly (Fig. 1), and designed to reveal rightward transcripts, panel B], and probe 3 (specific to spacers S77/S35 but designed to reveal leftward transcripts, panel C).

The abundance of processed CRISPR transcripts did not change upon phage infection, whether it was productive or nonproductive. With probe 3, a robust hybridization signal from high-molecular-weight RNA was observed in Xop411-infected Xo21 and Xp10-infected Xo604 samples (Fig. 3c, lanes 3 and 5, correspondingly), an expected result because these two phage–host pairs lead to productive infection and the probe used will hybridize to a transcript of phage gene 21R. The results demonstrate that CRISPR transcripts are processed in an expected way in both strains, and show that infection with a phage that contains a perfectly matching protospacer has no effect on the abundance of processed spacer transcripts. The results also show that the *X. oryzae* CRISPR cassettes are transcribed in one direction, from the most recently acquired spacers toward the oldest ones (from left to right in the scheme presented in Fig. 1).

Bioinformatic identification of a conserved sequence motif adjacent to *X. oryzae* phage protospacers

The data presented suggest that *X. oryzae* strain Xo21 possesses an apparently functional CRISPR system, which is, however, unable to interfere with Xop411 infection. We hypothesized that Xop411 productively infects Xo21 because the action of the host CRISPR system on this phage is prevented. Based on the analysis performed in *S. thermophilus*, this can happen if a short sequence located immediately adjacent to a protospacer targeted by CRISPR is altered (Deveau *et al.*, 2008).

A common motif adjacent to protospacers in *X. oryzae* phages was extracted as described in the Materials and methods. A detectable signal was present only on one end of the protospacers. Because the signal was virtually identical in each strain, a common signal logo using the entire set of *X. oryzae* phage protospacers was constructed. As can be seen, a trinucleotide motif TTC is present upstream of most protospacers (Fig. 4). While this work was in preparation, an identical motif was identified while studying the previously available *X. oryzae* CRISPR cassettes (Mojica *et al.*, 2009).

Among the 189 trinucleotides upstream of the protospacers in our collection, 100 are TTC whereas 73 differ from TTC in one position, forming 'inexact motifs.' In order to estimate the significance of the motif obtained, the occurrence of trinucleotides in phage DNA adjacent to protospacers was analyzed. Expected trinucleotide occurrence numbers calculated using corresponding single-nucleotide frequencies in the set of 100-nt regions flanking the protospacers were found to be 3.35 for TTC and 19.86 for inexact motifs. We conducted the exact binomial test to estimate the *P*-value of differences between observed and expected trinu-

Fig. 4. Logo of sequences upstream of *Xanthomonas oryzae* phage protospacers. A logo built using a common set of *X. oryzae* bacteriophage protospacers derived from five *X. oryzae* CRISPR cassettes is shown. Position numbered -1 is immediately adjacent to the protospacer; position numbered -6 is the furthest from it.

cleotide occurrence numbers ($< 10^{-16}$ and 10^{-15} for TTC and inexact motifs, correspondingly). We conclude that the TTC motif adjacent to protospacers from *X. oryzae* phages is highly nonrandom and is therefore likely required for CRISPR system function.
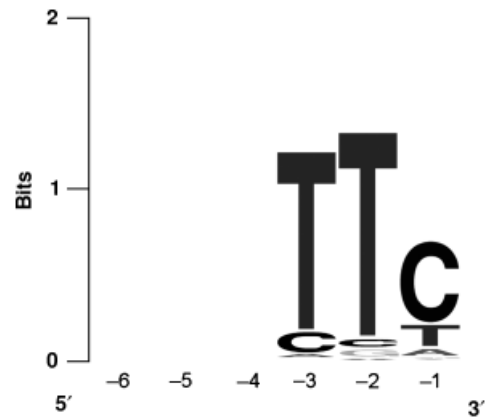
The conserved motif of the Xop411 protospacer matching the Xo21 spacer is mutated

We then wondered whether a site located upstream of Xop411 protospacer that matches spacers in Xo604 and Xo21 is intact. Inspection of the Xp411 genomic sequence revealed that the site was mutated (inexact motif TCC instead of the TTC consensus).

Discussion

In this work, we present complete sequences of CRISPR cassettes from two *X. oryzae* strains that exhibit differential sensitivity to two related bacteriophages. Statistical analysis of known *X. oryzae* CRISPR spacers and the extent of their matching to known phages infecting a given organism allows inferences about the total size of the 'phageome' (distinct phage genomes infecting a host). Of the 203 unique *X. oryzae* CRISPR spacers known currently, 139 can be mapped (with a small number of mismatches, *E*-value not exceeding 1) to protospacers in known *X. oryzae* phages. The 95% confidence interval for the fraction of *X. oryzae* spacers resulting from infection with these (or closely related) phages is 62–75% ($68.5 \pm 6.5\%$), implying that the total size of genomes of distinct phages capable of inducing CRISPR-associated immunity in *X. oryzae* is at most twice the total size of genomes of *X. oryzae* phages already known. In other

words, it appears that the variety of phages infecting *X. oryzae* (and affected by the CRISPR system of the host) is relatively small.

Sequence analysis revealed that *X. oryzae* strain Xo21, which serves as a host to phage Xop411, contains a CRISPR spacer that exactly matches a sequence in Xop411 genome, an unexpected result given the available evidence that spacers in corresponding positions of CRISPR cassettes from other bacteria contribute to phage resistance. Availability of large numbers of *X. oryzae* CRISPR cassette spacers for which corresponding *X. oryzae* phage protospacers could be identified allowed us to offer a plausible explanation for this apparent paradox. We propose that a single change converting a conserved TTC motif adjacent to a protospacer matching the host strain spacer to TCC explains the ability of Xop411 to plate on Xo21. Such a proposal is in line with earlier observation made with *S. thermophilus* phages, where it was shown that single-nucleotide changes in a conserved motif adjacent to protospacers render CRISPR-mediated exclusion ineffective (Deveau *et al.*, 2008). Experimental validation of our proposal requires the generation of an appropriate Xop411 mutant or the finding of a hypothetical parental phage that contains an intact adjacent site and is excluded by Xo21. Unproductive infections in two bacteria–phage pairs studied here (Xo604-Xop411 and Xo21-Xp10) must be unrelated to CRISPR function because the latter does not contain a matching spacer–protospacer pair, while the former, while containing a matching spacer–protospacer, has a mutated protospacer motif as the analysis of Xo21 infections demonstrates.

Horvath *et al.* (2008) identified two different motifs, each adjacent to protospacers corresponding to spacers from two of the three *S. thermophilus* CRISPR cassettes. The two motifs were unrelated, possibly reflecting the difference in CRISPR repeats in the two cassettes analyzed. The TTC motif revealed by our analysis is the only one present in *X. oryzae* phages. Closer inspection of motifs adjacent to *Streptococcus* and *Xanthomonas* protospacers revealed a principal difference that may reflect mechanistic differences in CRISPR system function in the two species. While no data on the transcription of the CRISPR cassette in *Streptococcus* is yet available, we assume that the transcription proceeds as in *X. oryzae*, from the leader sequence and newly introduced spacers toward older spacers. If this assumption is correct, then in *Streptococcus*, conserved motifs are located 'downstream' of protospacers (the direction here is determined by the CRISPR cassette transcription, with the leader sequence considered 'upstream') (Fig. 5). In contrast, in *X. oryzae* phages, the conserved motif is located 'upstream' (Fig. 5). In the case of *E. coli*, there appears to be no requirement for any conserved sequence around a protospacer for CRISPR-mediated phage immunity (Brouns *et al.*, 2008). Compara-
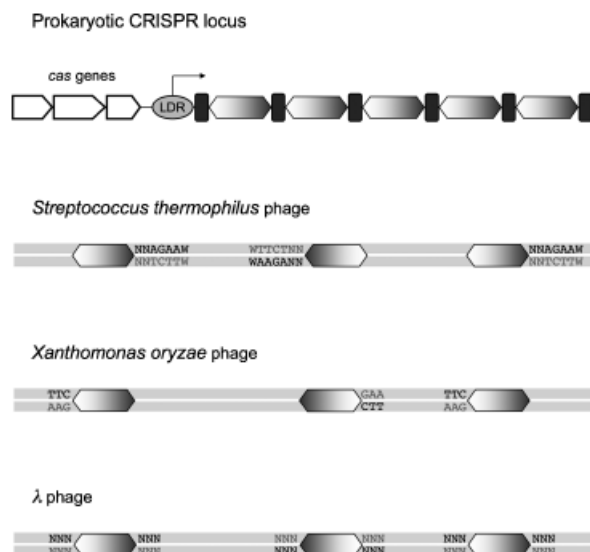

Fig. 5. Requirements for conserved motifs adjacent to protospacers for CRISPR function in different bacteria. At the top, a bacterial CRISPR locus consisting of a set of *cas* genes and a CRISPR cassette is shown with direction of transcription indicated. Below, locations of protospacers and adjacent conserved motifs in phages infecting *Streptococcus thermophilus*, *Xanthomonas oryzae*, and *Escherichia coli* are shown. Both strands of phage genomic DNA are schematically shown. Protospacers are shown as hexagons with the intensity of shading corresponding to the direction of transcription of the spacers in CRISPR cassettes (intense shading signifying the downstream end). In *S. thermophilus* phages, the conserved motif is located at the downstream end of the protospacers; in *X. oryzae* – at the upstream end. In the case of phage λ, there appears to be no requirement for a conserved motif.

tive analysis of CRISPR loci in various bacteria may help resolve this issue.

Acknowledgements

References

Andersson AF & Banfield JF (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**: 1047–1050.

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

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

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV & van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960–964.

Crooks GE, Hon G, Chandonia JM & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190.

Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P & Moineau S (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**: 1390–1400.

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

Haft DH, Selengut J, Mongodin EF & Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**: e60.

Hale C, Kleppe K, Terns RM & Terns MP (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14**: 2572–2579.

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

Jansen R, Embden JD, Gaastra W & Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565–1575.

Kunin V, He S, Warnecke F *et al*. (2008) A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res* **18**: 293–297.

Lee CN, Hu RM, Chow TY, Lin JW, Chen HY, Tseng YH & Weng SF (2007) Comparison of genomes of three *Xanthomonas oryzae* bacteriophages. *BMC Genomics* **8**: 442.

Lillestol RK, Redder P, Garrett RA & Brugger K (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59–72.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI & Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**: 7.

Marraffini LA & Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843–1845.

Mojica FJ, Ferrer C, Juez G & Rodriguez-Valera F (1995) Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol Microbiol* **17**: 85–93.

Mojica FJ, Díez-Villaseñor C, García-Martínez J & Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733–740.

Nakata A, Amemura M & Makino K (1989) Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J Bacteriol* **171**: 3553–3556.

Nechaev S, Yuzenkova Y, Niedziela-Majka A, Heyduk T & Severinov K (2002) A novel bacteriophage-encoded RNA polymerase binding protein inhibits transcription initiation and abolishes transcription termination by host RNA polymerase. *J Mol Biol* **320**: 11–22.

Salzberg SL, Sommer DD, Schatz MC *et al*. (2008) Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics* **9**: 204.

Semenova E, Djordjevic M, Shraiman B & Severinov K (2005) The tale of two RNA polymerases: transcription profiling and gene expression strategy of bacteriophage Xp10. *Mol Microbiol* **55**: 764–777.

Sorek R, Kunin V & Hugenholtz P (2008) CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181–186.

van Belkum A, Scherer S, van Alphen L & Verbrugh H (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol R* **62**: 275–293.

Yuzenkova J, Nechaev S, Berlin J, Rogulja D, Kuznedelov K, Inman R, Mushegian A & Severinov K (2003) Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage. *J Mol Biol* **330**: 735–748.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Protospacers of *Xanthomonas oryzae* phages.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.