CRISPR provides acquired resistance against viruses in prokaryotes

Rodolphe Barrangou[1], Christophe Fremaux[2], Hélène Deveau[3], Melissa Richards[1], Patrick Boyaval[2], Sylvain Moineau[3], Dennis A. Romero[1], Philippe Horvath[2]*

[1] Danisco USA Inc., 2802 Walton Commons West, Madison, WI 53718, USA

[2] Danisco France SAS, BP10, F-86220 Dangé-Saint-Romain, France

[3] Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Groupe de Recherche en Ecologie Buccale, Faculté de Médecine Dentaire, Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, G1K 7P4 Québec, Canada

* To whom correspondence should be addressed. E-mail: philippe.horvath@danisco.com

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPR) are a distinctive feature of the genomes of most Bacteria and Archaea and are thought to be involved in resistance to bacteriophage. We found that following viral challenge, bacteria integrated new spacers derived from phage genomic sequences. Removal or addition of particular spacers modified the phage-resistance phenotype of the cell. Thus, CRISPR, together with associated *cas* genes, provided resistance against phages, whereby specificity is determined by spacer/phage sequence similarity.

Bacteriophages are arguably the most abundant biological entity on the planet (*1*). Their ubiquitous distribution and abundance have an important impact on microbial ecology and the evolution of bacterial genomes (*2*). Consequently, bacteria have developed a variety of natural defense mechanisms that target diverse steps of the phage life cycle, notably blocking adsorption, preventing DNA injection, restricting the incoming DNA and abortive infection systems. These antiviral barriers can also be engineered and manipulated to better control phage populations (*2*, *3*).

Numerous bacteria have been selected by humans and used extensively for fermentation and biotechnology processes. Unfortunately, domesticated bacteria used in industrial applications are often susceptible to phage attack, including genera and species widely used as dairy cultures (*4*). Accordingly, the industry has devised various strategies to combat phage based on strain diversity, bacteriophage insensitive mutants, and plasmids bearing phage-resistance mechanisms.

Streptococcus thermophilus is a low G+C Gram-positive bacterium and a key species exploited in the formulation of dairy culture systems for the production of yogurt and cheese. Comparative genomics analyses of closely related *S. thermophilus* strains have previously revealed that genetic polymorphism primarily occurs at hypervariable loci, such as the *eps* and *rps* operons, as well as two clustered regularly interspaced short palindromic repeats (CRISPR) loci (*5*-*7*). CRISPR loci typically consist of several non-contiguous direct repeats separated by stretches of variable sequences called spacers, and are often times adjacent to *cas* genes (CRISPR-associated). Although the function of CRISPR loci has not been established biologically, *in silico* analyses of the spacers have revealed sequence homology with foreign elements, including bacteriophage and plasmid

sequences (*7-9*). Based exclusively on *in silico* analyses, several hypotheses have been put forward proposing roles for CRISPR and *cas* genes, that include providing immunity against foreign genetic elements via a mechanism based on RNA interference (*10*).

We analyzed the CRISPR sequences of various *S. thermophilus* strains, including closely related industrial strains and phage-resistant variants (Fig. S1). Differences in the number and type of spacers were observed primarily at the CRISPR1 locus. Notably, phage sensitivity appeared to be correlated with CRISPR1 spacer content. Specifically, spacer content was nearly identical between parental strains and phage-resistant derivatives, except for additional spacers present in the latter. These findings therefore suggest a potential relationship between the presence of additional spacers and the differences observed in the phage sensitivity of a given strain. This observation prompted us to investigate the origin and function of additional spacers present in phage-resistant mutants.

First, we tested the hypothesis that CRISPR loci are altered during the natural generation of phage-resistant mutants. A phage-host model system was selected, consisting of a phage-sensitive wild-type *S. thermophilus* strain widely used in the dairy industry, DGCC7710 (WT) and two distinct but closely related virulent bacteriophages isolated from industrial yogurt samples, phage 858 and phage 2972 (*11*). Nine phage-resistant mutants were generated independently by challenging the WT strain with phage 858, phage 2972 or simultaneously with both (*12*), and their CRISPR loci were analyzed. Differences were consistently observed at the CRISPR1 locus, where 1 to 4 additional spacers were inserted next to the 32 spacers present in the WT strain (Fig. 1). The addition of new spacers in response to phage infection seemed to be polarized towards

one end of the CRISPR1 locus. This is consistent with previous observations of spacer hypervariability at the leader end of the CRISPR locus in various strains (*9*, *13*). Sequence analysis of the additional spacers inserted in the CRISPR1 locus of the various phage-resistant mutants revealed similarity to sequences found within the genomes of the phages used in the challenge (Fig. 2, Fig. S2). Interestingly, similarities were observed throughout the phage genomes, in most functional modules, both on the coding and non-coding strands. No particular sequence, gene or functional group seemed to be targeted specifically. These results reveal that upon becoming resistant to bacteriophages, the CRISPR1 locus was modified by the integration of novel spacers apparently derived from phage DNA.

Surprisingly, we observed that some strains were resistant to both phages while others were resistant only to the phage used in the challenge (Fig. 1). The phage-resistance profile seemed correlated to the spacer content whereby strains with spacers showing 100% identity to sequences conserved in both phages were resistant to both phages, such as spacers S3, S6 and S7. In contrast, when nucleotide polymorphisms were observed between the spacer and the phage sequence (from 1 to 15 SNPs over 29 or 30 nucleotides), the spacer did not seem to provide resistance, such as spacers S1, S2, S4, S5 and S8 (Fig. 1, Fig. S2). Additionally, when several spacers were inserted (S9-S14), phage-resistance levels were higher. These findings indicate that the CRISPR1 locus is subject to dynamic and rapid evolutionary changes driven by phage exposure. Altogether, these results reveal that CRISPR loci can indeed be altered during the generation of phage-resistant mutants and establish a link between CRISPR content and phage

sensitivity, suggesting that the presence of a CRISPR spacer identical to a phage sequence provides resistance against phages containing this particular sequence.

To determine whether CRISPR spacer content defines phage resistance, we altered the CRISPR1 locus by adding and deleting spacers (*12*), and tested subsequent strain sensitivity to phages. All constructs were generated and integrated into the *S. thermophilus* chromosome using the system developed by Russell and Klaenhammer (*14*). We removed the spacers and repeats in the CRISPR1 locus of strain $WT_{\Phi858}^{+S1S2}$ and replaced them with a single repeat without any spacer (*12*). The resulting strain $WT_{\Phi858}^{+S1S2}\Delta$CRISPR1 was sensitive to phage 858, indicating that the phage resistance of the original phage-resistant mutant ($WT_{\Phi858}^{+S1S2}$) was probably linked to the presence of S1 and S2 (Fig. 3).

Further, to address the critical question of whether adding spacers provides novel phage resistance, we replaced the CRISPR1 locus of strain $WT_{\Phi2972}^{+S4}$ with a version only containing spacers S1 and S2 (*12*) and tested whether the phage sensitivity was affected. Remarkably, the resulting strain $WT_{\Phi2972}^{+S4}$::pS1S2 gained resistance to phage 858, suggesting that these two spacers have the ability to provide phage resistance *de novo* (Fig. 3). Altogether, these observed modifications establish the link between the CRISPR spacer content and phage resistance.

In the process of generating strain $WT_{\Phi858}^{+S1S2}\Delta$CRISPR1, we created $WT_{\Phi858}^{+S1S2}$::pR, a variant that contains the integration vector with a single repeat inserted between the *cas* genes and the native CRISPR1 locus (Fig. 3). Unexpectedly, strain $WT_{\Phi858}^{+S1S2}$::pR was sensitive to phage 858, although spacers S1 and S2 remained present on the chromosome (Fig. 3). Similarly, the $WT_{\Phi2972}^{+S4}$::pS1S2 construct lost the

resistance to phage 2972, although spacer S4 is present in the chromosome (Fig. 3). These results indicated that spacers alone did not provide resistance, and perhaps they have to be in a particular genetic context to be effective.

Although initial work suggested involvement in DNA repair (*15*), the current hypothesis is that *cas* genes (*5*, *16*) are involved in CRISPR-mediated immunity (*10*). Consequently, we inactivated two *cas* genes in strain $WT_{\Phi858}^{+S1S2}$ (*12*): *cas5* (COG3513) and *cas7,* which are equivalent to *str0657*/*stu0657* and *str0660*/*stu0660,* respectively (*6*, *7*). The *cas5* inactivation resulted in loss of the phage resistance (Fig. 3), and perhaps Cas5 acts as a nuclease, since it contains a HNH-type nuclease motif. In contrast, inactivating *cas7* did not alter the resistance to phage 858 (Fig. 3). Interestingly, we were repeatedly unable to generate CRISPR1 phage-resistant mutants from the *cas7* knockout, perhaps because Cas7 is involved in the synthesis and/or insertion of new spacers and additional repeats.

Upon testing sensitivity of the phage-resistant mutants, we found that plaque formation was dramatically reduced, but that a relatively small population of bacteriophage retained the ability to infect the mutants. We further analyzed phage variants derived from phage 858 that retained the ability to infect $WT_{\Phi858}^{+S1S2}$. In particular, we investigated the sequence of the genome region corresponding to additional spacers S1 and S2 in two virulent phage variants. In both cases, the genome sequence of the phage variant was mutated and two distinct single nucleotide polymorphisms were identified in the sequence corresponding to spacer S1 (Fig. S3).

Overall, prokaryotes appear to have evolved a nucleic-acid based “immunity” system whereby specificity is dictated by the CRISPR spacer content, while the

resistance is provided by the Cas enzymatic machinery. Additionally, we speculate that some of the *cas* genes not directly providing resistance are actually involved in the insertion of additional CRISPR spacers and repeats, as part of an adaptive "immune" response. Further studies are desired to better characterize the mechanism of action and identify the specific function of the various *cas* genes. This nucleic-acid based system contrasts with amino-acid based counterparts in Eukaryotes whereby adaptative immunity is not inheritable. The inheritable nature of CRISPR spacers supports the use of CRISPR loci as targets for evolutionary, typing and comparative genomic studies (*9*, *17*-*19*). Because this system is reactive to the phage environment, it likely plays a significant role in prokaryotic evolution and ecology and provides a historical perspective of phage exposure, as well as a predictive tool for phage sensitivity. The CRISPR / *cas* system may accordingly be exploited as a virus defense mechanism, and also potentially to reduce the dissemination of mobile genetic elements and the acquisition of undesirable traits such as antibiotic resistance genes and virulence markers. From a phage evolution perspective, the integrated phage sequences within CRISPR loci may also provide additional anchor points to facilitate recombination during subsequent phage infections, thus increasing the gene pool to which phages have access (*20*). Since CRISPR loci are found in the majority of bacterial genera, and are ubiquitous in Archaea (*5*, *13*, *21*), they will provide new insights in the relationship and co-directed evolution between Prokaryotes and their predators.

References and notes:

1. M. Breitbart, F. Rohwer, *Trends Microbiol.* **13,** 278 (2005).

2. S. Chibani-Chennoufi, A. Bruttin, M.-L. Dillmann, H. Brüssow, *J. Bacteriol.* **186,** 3677 (2004).

3. J. M. Sturino, T. R. Klaenhammer, *Nature Rev. Microbiol.* **4,** 395 (2006).

4. H. Brüssow, *Annu. Rev. Microbiol.* **55,** 283 (2001).

5. R. Jansen, J. D. A. van Embden, W. Gaastra, L. M. Schouls, *Mol. Microbiol.* **43,** 1565 (2002).

6. A. Bolotin *et al., Nature Biotechnol.* **22,** 1554 (2004).

7. A. Bolotin, B. Quinquis, A. Sorokin, S. D. Ehrlich, *Microbiology* **151,** 2551 (2005).

8. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, E. Soria, *J. Mol. Evol.* **60,** 174 (2005).

9. C. Pourcel, G. Salvignol, G. Vergnaud, *Microbiology* **151,** 653 (2005).

10. K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, E. V. Koonin, *Biol. Direct* **1,** 7 (2006).

11. C. Lévesque *et al., Appl. Environ. Microbiol.* **71,** 4057 (2005).

12. Information on materials and methods for the generation of phage-resistant mutants, engineering of CRISPR spacers and inactivation of *cas* genes is available on *Science* Online.

13. R. K. Lillestøl, P. Redder, R. A. Garrett, K. Brügger, *Archaea* **2,** 59 (2006).

14. W. M. Russell, T. R. Klaenhammer, *Appl. Environ. Microbiol.* **67,** 4361 (2001).

15. K. S. Makarova, L. Aravind, N. V. Grishin, I. B. Rogozin, E. V. Koonin, *Nucleic Acids Res.* **30,** 482 (2002).

16. D. H. Haft, J. Selengut, E. F. Mongodin, K. E. Nelson, *PloS Comput. Biol.* **1,** e60 (2005).

17. P. M. A. Groenen, A. E. Bunschoten, D. van Soolingen, J. D. A. van Embden, *Mol. Microbiol.* **10,** 1057 (1993).

18. E. F. Mongodin *et al., J. Bacteriol.* **187,** 4935 (2005).

19. R. T. DeBoy, E. F. Mongodin, J. B. Emerson, K. E. Nelson, *J. Bacteriol.* **188,** 2364 (2006).

20. R. W. Hendrix *et al., Proc. Natl. Acad. Sci. USA* **96,** 2192 (1999).

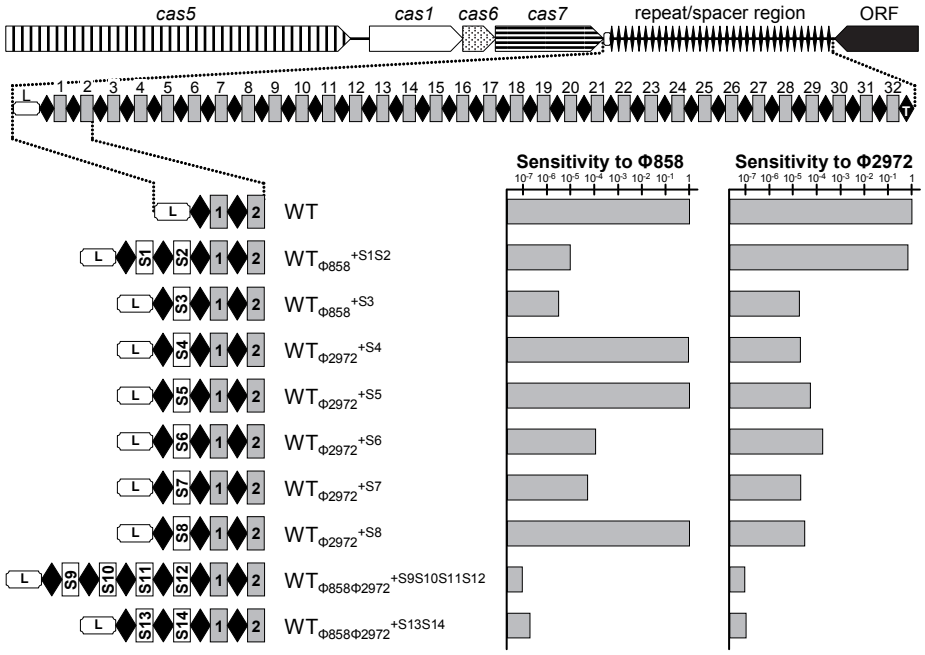
21. J. S. Godde, A. Bickerton, *J. Mol. Evol.* **62,** 718 (2006).

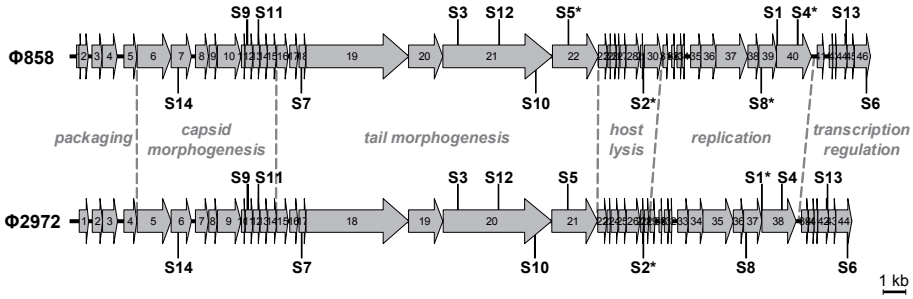
22. We thank L. Bayer, C. Vos and A.-C. Coûté-Monvoisin of Danisco Innovation as well as J. Labonté and D. Tremblay of U. Laval for technical support, and E. Bech Hansen for discussions and critical review of the manuscript. Also, we thank T. R. Klaenhammer for providing the integration system. This work was supported by funding from Danisco A/S. Also, S. M. would like to acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Program. Sequences were deposited in GenBank, accession numbers EF434458 to EF434504.

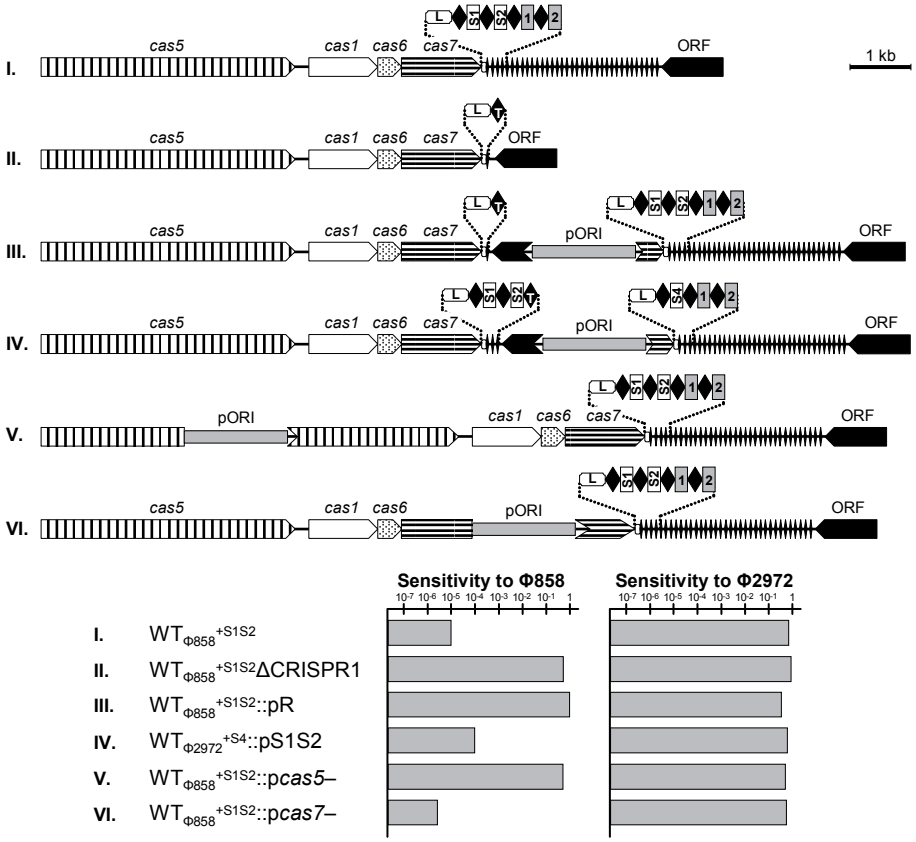
Fig. 1. *Streptococcus thermophilus* CRISPR1 locus overview, newly acquired spacers in phage-resistant mutants, and corresponding phage sensitivity. The CRISPR1 locus of DGCC7710 (WT) is at the top. The repeat/spacer region of WT is in the middle: repeats (black diamonds), spacers (numbered gray boxes), leader (L, white box) and terminal repeat (T, black diamond). At the bottom, the spacer content on the leader side of the locus in phage-resistant mutants is detailed on the left, with newly acquired spacers (white boxes, S1-S14). On the right, the sensitivity of each strain to phages 858 and 2972 is represented as a histogram of the efficiency of plaquing (EOP), which is the plaque count ratio of a mutant strain to that of the wild-type.

Fig. 2. *S. thermophilus* phage genome maps with the position of sequences similar to the acquired CRISPR1 spacers of the phage-resistant mutants. Spacers shown above and below the genome maps indicate that the spacer matches a sequence on the (+) and on the (-) strand, respectively. An asterisk indicates the existence of SNP between the spacer sequence and that of the phage genome (Fig. S1). The genome sequences of phage 2972 (accession number AY699705) and phage 858 are 93% identical.

Fig. 3. CRISPR spacer engineering, *cas* gene inactivation and corresponding phage sensitivity. I, mutant $WT_{\Phi858}^{+S1S2}$; II, mutant $WT_{\Phi858}^{+S1S2}\Delta$CRISPR1 where CRISPR1 was deleted; III, mutant $WT_{\Phi858}^{+S1S2}$::pR where CRISPR1 was displaced and replaced with a unique repeat; IV, $WT_{\Phi2972}^{+S4}$::pS1S2, mutant of strain $WT_{\Phi2972}^{+S4}$ where CRISPR1 was displaced and replaced with a version containing S1 and S2; V, $WT_{\Phi858}^{+S1S2}$::p*cas5*- with *cas5* inactivated; VI, $WT_{\Phi858}^{+S1S2}$::p*cas7*- with *cas7* inactivated. pORI indicates the integrated plasmid (*12*). The phage sensitivity of each strain to phages 858 and 2972 is represented at the bottom as a histogram of the efficiency of plaquing (EOP).


cas5
cas1
cas6
cas7
repeat/spacer region
ORF
L
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
T
Sensitivity to Φ858
Sensitivity to Φ2972
10-7 10-6 10-5 10-4 10-3 10-2 10-1 1
WT
WTΦ858+S1S2
WTΦ858+S3
WTΦ2972+S4
WTΦ2972+S5
WTΦ2972+S6
WTΦ2972+S7
WTΦ2972+S8
WTΦ858Φ2972+S9S10S11S12
WTΦ858Φ2972+S13S14


Φ858
S9 S11
S3 S12
S5*
S1 S4* S13
S14
S7
S10
S2*
S8*
S6
packaging
capsid morphogenesis
tail morphogenesis
host lysis
replication
transcription regulation
Φ2972
S9 S11
S3 S12
S5
S1* S4
S13
S14
S7
S10
S2*
S8
S6
1 kb


I.
II.
III.
IV.
V.
VI.
cas5
cas1
cas6
cas7
ORF
pORI
L
S1
S2
1
2
T
1 kb
Sensitivity to Φ858
Sensitivity to Φ2972
10-7 10-6 10-5 10-4 10-3 10-2 10-1 1
I. WTΦ858+S1S2
II. WTΦ858+S1S2ΔCRISPR1
III. WTΦ858+S1S2::pR
IV. WTΦ2972+S4::pS1S2
V. WTΦ858+S1S2::pcas5–
VI. WTΦ858+S1S2::pcas7–