The CRISPR System Protects Microbes against Phages, Plasmids

Palindromic DNA repeat sequences immunize microorganisms against phages and plasmids, while also directing their evolution

Rodolphe Barrangou and Philippe Horvath

Many scientists believe that phages are the most abundant life form on Earth. Although phages outnumber their bacterial prey 10-fold, bacteria persist, sometimes relying on clustered regularly interspaced short palindromic repeats (CRISPRs) of DNA sequence as a defense mechanism. CRISPRs, first recognized in *Escherichia coli* in 1987, are found within the genomes of about 40% of bacteria and 90% of archaea tested so far.

In recognizing these sequences, researchers speculated that CRISPRs facilitate replicon partitioning, repair DNA, or mediate immunity against extrachromosomal genetic elements via RNA interference. Eventually, several groups of researchers concluded that CRISPRs are part of a DNA-encoded defense system that provides bacteria and archaea with nucleic acid-based immunity against phages, plasmids, or other invading genetic elements.

Put another way, CRISPR is a recently recognized, molecular-level component of the ongoing struggle between viruses and hosts, and part of the dynamic process by which their genomes coevolve. Further, the CRISPR system applies selective pressure on phage populations, perhaps explaining the relatively high mutation rates of phage genomes in those populations where CRISPR loci are active.

**Components of CRISPR Systems**

CRISPR loci typically consist of several noncontiguous, partially palindromic DNA sequences that repeat and are separated by stretches of nonrepetitive spacer sequences (Fig. 1). CRISPR repeats vary between 21 and 48 bp, while spacers range from 26 to 72 bp. Repeat-spacer units can reach nearly 300, but loci tend to contain fewer than 50 units. Any particular bacterial or archael species may contain more than one CRISPR, and as many as 18 distinct CRISPRs are found within one genome. They are usually located on the chromosome, although CRISPRs occasionally appear on plasmids. CRISPR spacer sequences show homology to phages and plasmids.

Indeed, several phage genome sequences share 100% identity with CRISPR spacer sequences, suggesting that spacers originate from phage genomes, deriving from both the sense and antisense orientations, with a bias

**Summary**

- Clustered regularly interspaced short palindromic repeats (CRISPRs) are a diverse family of DNA repeat sequences that are widely distributed among archaea and bacteria.
- CRISPR serves as a cellular defense mechanism, providing microorganisms immunity against phages and plasmids in a sequence-specific manner.
- CRISPR spacers provide a record of interactions between host and viruses, providing a perspective of past exposure of the microorganism to phage that may be used for typing such cells.
- The CRISPR system helps to drive phage genome evolution.
towards sense. In general, the sequence of the repeat defines a CRISPR type. CRISPR arrays typically are located near cas (CRISPR-associated) genes, which consist of a large and polymorphic family, whose members encode proteins that carry nucleic acid-related domains such as nucleases, helicases, and nucleotide binding. Six “core” cas genes are known, including cas1 (COG1518), which appears to be a universal marker.

The CRISPR leader, which is upstream of the first CRISPR repeat, acts as a promoter. Several parameters help in defining and orienting the various elements that constitute a CRISPR locus, with the location and orientation of the cas genes among the most important. CRISPR repeats are usually colinear with the cas genes, while terminal repeats often degenerate at the trailer end.

### CRISPR “Defends” against Viruses and Plasmids

We are studying how CRISPR operates in *Streptococcus thermophilus* cells, providing them with resistance against phage. Specifically, when exposed to phage, these cells generate phage-resistant variants that incorporate one or multiple novel CRISPR repeat-spacer units at the leader end of one or more CRISPR loci (Fig. 1). Those novel CRISPR spacers appear to derive from the DNA of the invasive phage, with which they share sequences. Further, modulating the CRISPR-spacer affects phage sensitivity. Thus, adding a spacer provides resistance, whereas resistance is lost when spacers are removed.

CRISPR loci evolve in response to phages in *Leptospirillum*, according to a population dynamics study. Moreover, CRISPR is actively
Barrangou, a “Genomicist,” with a Passion for Bacteria and Archaea

Rodolphe Barrangou, whose passion is bacterial and archaeal genomes, holds no academic degree in microbiology. “I am a genomicist, but I would rather describe myself as a scientist,” he says. “Nowadays, most people have quite multidisciplinary backgrounds. In the area of genomics, many people are scientists of many trades.”

In terms of his scientific education, Barrangou is eclectic, having studied engineering and food science as well as genomics, and that background serves him well as a senior scientist at Danisco USA Inc., in Madison, Wis., an American affiliate of the Copenhagen, Denmark-based company, a global supplier of food ingredients and industrial bioproducts. The company uses natural materials to develop and produce ingredients for food and other domestic products, including ice cream, cheese, laundry detergent, and bread.

Barrangou focuses his research on clustered regularly interspaced short palindromic repeats, or CRISPR, a family of DNA repeats within genomes of bacteria and archaea. These DNA segments provide acquired immunity against invading genetic elements, such as phages and plasmids. “It is a novel and somewhat revolutionary immune system in microbes that allows them to build resistance against viruses in an adaptive, specific, and inheritable way,” he says. “It is quite surprising that this system has not been discovered until recently, and it is perhaps even more surprising that it was not first characterized in a common model organism.”

Barrangou shared a “eureka” moment over CRISPR with Philippe Horvath, his colleague at Danisco in France. That moment came, he says, when they both “realized that the key in CRISPR loci was the spacer sequence” as the single most significant milestone of his career.

Beyond CRISPR, Barrangou also relies on genome-based analyses to better understand several types of bacteria that are used to process foods or as dietary supplements. For example, he says we are “mining genomes of starter cultures to understand their ability to survive phage attack and provide acidification, desirable flavor, and texture for fermented dairy products such as yogurt and cheese.” The headquarters for the Danisco Cultures division, for which he works, is located in Paris. “Life in America is great, but different,” he says. “While the quality of life is much better here overall, I miss the food and the wine, and vacation here pales compared to the French benefits.”

Barrangou, 33, a native of France, lived in Paris until he was 23. Although married to an American, he retains his French citizenship and returns home once a year to visit family and friends.

involved in the coevolution of *Synechococcus* host and viral genomes. We find that *cas* genes provide CRISPR-encoded resistance, while the Cas enzymatic machinery appears to be functionally coupled with particular CRISPR repeats. We infer this link between *cas* and CRISPR repeats from the congruence between their sequence patterns. Moreover, the enzymatic machinery of a specific CRISPR locus does not work with the genetic content of another.

Phage resistance is lost when a *csn1*-like gene is inactivated. Further, inactivating *cas2/csn2*-like genes impairs the ability of the host to integrate novel CRISPR spacers following exposure to phage, showing that some *cas* genes provide CRISPR-encoded phage resistance while others are involved in acquiring novel spacers. Among them, the *E. coli* Cas3 is required for the phage resistance phenotype.

Some details as to how CRISPR confers phage resistance are emerging. For instance, transcribed CRISPR repeat-spacer arrays are processed into specific small RNA units that correspond to a spacer flanked by two partial repeats. After transcription, proteins in the CRISPR-associated complex for antiviral defense (Cascade) specifically cleave the CRISPR transcript within its repeat sequence to generate small CRISPR RNA molecules (crRNAs). CRISPR spacers match both sense and antisense phage DNA, a finding that is compatible with CRISPR targeting dsDNA. Moreover, assuming that the target DNA is recognized through sequence identity with crRNAs and then cleaved by a Cas endonuclease, the CRISPR/Cas system could be considered a variable geometry restriction system with broad and adaptive specificity.

While many CRISPR studies focus on phage resistance mechanics, others emphasize the
similarities between CRISPR spacers and phage sequences. However, it is important not to underestimate the importance of CRISPR activity against plasmids. Although the ubiquitous and predatory nature of phages may explain the overwhelming representation of phage sequences in CRISPR spacers, CRISPR also can dramatically interfere with plasmids. For example, in Staphylococcus epidermidis, CRISPR blocks both plasmid transfer and conjugation—in this case, directly targeting DNA. CRISPR/Cas systems in other species of bacteria or archaea may target messenger RNA in addition to, or instead of, DNA.

**CRISPR Loci Are Subject to Dynamic Flux**

In model phage-host systems as well as in microbial-viral communities, CRISPR can change rapidly, producing hypervariability regions that set apart closely related strains from one another, including *S. thermophilus*, *Leptospirillum* from acid mine drainage, and *Synechococcus* in microbial mats. Indeed, phage-driven dynamics induce such changes within CRISPR loci.

While CRISPR sequences primarily add novel spacers at the leader end of each locus following exposure to phages, internal spacers sometimes are deleted (Fig. 2). Such deletions, likely via homologous recombination between CRISPR repeats, help to limit the size of CRISPR loci. Deleting spacers at the trailer end, which target phages that were encountered in the past, would allow conservation of spacers targeting viruses currently circulating. Also, proximal spacers are probably transcribed at a higher frequency than are distal spacers. Maintaining relatively small numbers of repeat-spacer units helps to limit transcript lengths.

Even though some CRISPR arrays contain

---

His father is an attorney, his mother a homemaker. He is the oldest of four and the only scientist in his immediate family; his siblings all went to business school.

“As I look back, I was perhaps always a scientist, being curious and observant,” he says. “While I was always fascinated by science as an area of legitimate and sound knowledge, I don’t think I realized I was a scientist until college,” he says. He received his B.S. in biological engineering at the Université de Technologie de Compiègne in France in 2000, and his M.S. in food science at North Carolina (NC) State University in Raleigh at the end of that same year. He earned his doctorate, also at NC State, in 2004. “I actually spent the last [undergraduate] year studying abroad at NC State where I concurrently started an M.S.—and delayed my graduation date as much as I could to enjoy my stay there,” he says.

When Barrangou was 16, he attended a seminar by Pierre-Gilles de Gennes, who won the Nobel Prize in physics in 1991. Barrangou was struck by how de Gennes conveyed complex matters very simply in his talk, but could not do the same when answering questions. “He provided very complicated answers to what I thought were simple questions,” Barrangou says. “It was my first true introduction to what I call the ‘multiple layers of scientific knowledge,’ whereby any question can be answered using a nearly infinite level of details and complexity, depending on the audience.”

Barrangou prefers working in the industrial sector instead of at a university, noting the company allows him enough scientific freedom “to explore uncharted territories, yet enough restraint to ensure we focus on relevant and applicable scientific endeavors,” he says. “[Craig] Venter is a role model for many in the industry, and genomics overall, who has compellingly shown that one need not be in an academic institution to do the best science.”

In his spare time, Barrangou enjoys playing basketball and golf and running, including occasional half-marathons. He and his wife have a young son and daughter, plus two dogs. He is teaching his California-born wife and his children to speak French. He also sees some scientific curiosity emerging from his offspring. “I am absolutely fascinated by the fact that my children, although quite young, are obviously scientists,” he says. “I am most amazed by their ability to replicate behaviors they observe from their parents or pets. Toddlers master the true trial and error approach—and can tell the difference between a successful experiment and a failure.”

Marlene Cimons

Marlene Cimons is a freelance writer in Bethesda, Md.
nearly 300 spacers, they tend to be much smaller. In lactic acid bacteria, for example, they average 20 spacers per locus. CRISPR size does not correlate with activity. Thus, for instance, the CRISPR loci in *S. thermophilus* remain relatively short but are highly active. Meanwhile, CRISPR spacers in *Sulfolobus islandicus* are having a dramatic impact on phage populations, with spacer content being closely related to local phage genomic sequences.

Perhaps because they effectively defend so many different kinds of cells against invading genetic elements, CRISPR sequences can transfer efficiently between distant types of organisms. This capacity also helps to explain the low phylogenetic congruence between CRISPR elements and the organisms in which they are found. Plasmids, megaplasmids, and even prophages may mediate horizontal transfer of CRISPR loci. Also, insertion sequence elements on both sides of CRISPR-cas regions might indicate their higher propensity for lateral gene transfer, as do differences in GC content between CRISPR loci and other components of the chromosome.

**Figure 2**
Various evolutionary paths of CRISPR loci, including acquisition of novel repeat-spacer units, mutation of the CRISPR repeat sequence, and deletion of consecutive repeat-spacer units.

**Figure 3**
Various strategies to evade CRISPR-encoded immunity, including mutation in the proto-spacer (top left), insertion/deletion in the proto-spacer (top right), mutation within the CRISPR motif (bottom left), and large-scale deletion (bottom right). In each bubble, the crRNA is shown at the top, and the corresponding targeted DNA is shown at the bottom, where the star represents the CRISPR motif. The wild-type proto-spacer is shown in the bubble in the center.

**Phage Can Adapt to CRISPR-Encoded Resistance**

Although CRISPR sites provide high levels of phage resistance, some phages continue to infect host cells that supposedly are immune. CRISPR-imposed selective pressure can lead to specific mutations in the proto-spacer sequence, the phage sequence corresponding to a CRISPR spacer (Fig. 3). In fact, a single nucleotide change can allow a phage to circumvent CRISPR-encoded immunity. In addition, phages may alter targeted sequences via deletions. These mutations often affect the amino acid sequence, perhaps indicating a fitness cost associated with circumventing CRISPR. However, in the short term, acquiring novel CRISPR spacers does not seem to cost the host in terms of fitness. Extensive genome recombina-
tion events in environmental phage populations may also reflect the impact of CRISPR.

Phage may also circumvent the CRISPR when they develop mutations in the direct vicinity of the proto-spacer. A careful analysis of phage sequences adjacent to proto-spacers revealed conserved sequences, called CRISPR motifs. The CRISPR motif, which is located in the direct vicinity of the proto-spacer (typically fewer than 10 nucleotides outside the sequence), seems to be critical to CRISPR-encoded resistance. When the CRISPR motif mutates, phage can escape, strongly suggesting that the CRISPR motif is involved in the resistance rather than the acquisition of novel spacers. Although apparently randomly located on phage genomes, the CRISPR motif plays a key role in selecting functional spacers. While the CRISPR defense system targets no particular sequence, gene, functional group, or DNA strand, CRISPR spacers are not chosen randomly. Indeed, different lineages may independently acquire the same CRISPR spacer.

By our tally, 72% of proto-spacers are located on the sense strand in phage genomes. That figure is remarkably similar to the 75% of CRISPR motifs that are located on the sense strand in phage genomes of *S. thermophilus*. Since mutations in the CRISPR motif enable phages to escape CRISPR-mediated resistance, that motif likely is directly involved in the defense system, perhaps as a recognition sequence for nucleic acid cleavage.

**Putting CRISPR Systems to Work**

CRISPR loci provide avenues to useful applications that include strain typing based on hyper-variability, predicting and modulating phage resistance, and controlling the dissemination of genetic elements.

CRISPR spacer content and hypervariability, even across very closely related strains, provide a genetic basis for strain typing. For example, early in the 1990s, this property was applied to mycobacteria via DNA-DNA hybridizations through a technique called spoligo-typing. And, because CRISPR spacers arise from successive exposures to phages and plasmids, they provide an ecological and epidemiological record with which to study particular strains.

Natural spacers or synthetic sequences can be integrated into CRISPR loci, providing them with a genetic signature that investigators can use to tag, detect, or track microbial strains of interest. For industrial applications, CRISPR may also be exploited as a defense mechanism against phage infections, either via selection of variants with efficient spacers or via genetic engineering of spacers corresponding to highly conserved sequences. We have generated numerous phage resistant strains naturally through iterative rounds of exposure and selections, yielding strains with novel spacers that confer broad and deep phage resistance while retaining other critical traits (Fig. 4).

Because CRISPR loci can impede transfers of particular nucleic acid sequences into a microbial cell, they can be genetically engineered to preclude the transfer of particular genetic elements, such as antibiotic resistance markers or other potentially harmful genes. Such loci may also be engineered to limit the spread of insertion sequences and transposons.

---

**FIGURE 4**

Iterative build-up of CRISPR-encoded phage resistance via consecutive additions of novel spacers, which yields an increase of the level of resistance (Lor), as well as a reduction of the spectrum of sensitivity (Sos) to phages. Accordingly, variants that have iteratively acquired novel spacers have broader and deeper phage resistance spectra.
ACKNOWLEDGMENTS

We acknowledge our colleagues and collaborators Egon Bech Hansen, Patrick Boyaval, Christophe Fremaux, Dennis Romero, Anne-Claire Coute-Monvoisin, Josiane Garneau, and Jessica Labonte for their support and many scientific contributions. We also thank Sylvain Moineau and Jill Banfield for their insights and expertise. Thanks to Kelly Paralis Keenan for masterful assistance with the artwork. This work was supported by Danisco.

SUGGESTED READING


