



Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array

Sandra C. Garrett*

Department of Genetics and Genome Sciences, Institute for Systems Genomics, UConn Health, Farmington, CT, United States

OPEN ACCESS

Edited by:

Muhammad Kamruzzaman,
Westmead Institute for Medical
Research, Australia

Reviewed by:

Malcolm F. White,
University of St Andrews,
United Kingdom
Alexander P. Hynes,
McMaster University, Canada

*Correspondence:

Sandra C. Garrett
garrett@uchc.edu;
sandrakoralgarett@gmail.com

Specialty section:

This article was submitted to
Microbial Physiology and Metabolism,
a section of the journal
Frontiers in Microbiology

Received: 04 February 2021

Accepted: 12 March 2021

Published: 01 April 2021

Citation:

Garrett SC (2021) Pruning
and Tending Immune Memories:
Spacer Dynamics in the CRISPR
Array. *Front. Microbiol.* 12:664299.
doi: 10.3389/fmicb.2021.664299

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated genes) is a type of prokaryotic immune system that is unique in its ability to provide sequence-specific adaptive protection, which can be updated in response to new threats. CRISPR-Cas does this by storing fragments of DNA from invading genetic elements in an array interspersed with short repeats. The CRISPR array can be continuously updated through integration of new DNA fragments (termed spacers) at one end, but over time existing spacers become obsolete. To optimize immunity, spacer uptake, residency, and loss must be regulated. This mini-review summarizes what is known about how spacers are organized, maintained, and lost from CRISPR arrays.

Keywords: adaptation, spacer acquisition, repeat, array, CRISPR, spacer deletion

INTRODUCTION

Prokaryotes have evolved a diverse repertoire of tools to restrict the proliferation of deleterious mobile genetic elements (Koonin et al., 2017). Uniquely among these tools, CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated genes) provides sequence-specific protection that can be updated in the face of novel threats, making it an adaptive immune system. CRISPR stores sequence information about potentially parasitic or harmful mobile genetic elements in an array (Barrangou et al., 2007) and uses that information to carry out targeted degradation of DNA or RNA, depending upon CRISPR type (Makarova et al., 2020). CRISPR-Cas systems are diverse and have been classified into two classes, six distinct types (I–VI), and at least 33 subtypes (Makarova et al., 2020), but certain characteristics are shared. All CRISPR arrays contain a series of direct repeats separated by short sequences called “spacers” which match DNA from previously encountered invaders (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). An upstream leader sequence regulates transcription of the array and mediates addition of new spacers (Jansen et al., 2002; Yosef et al., 2012; Wei et al., 2015; Alkhnabashi et al., 2016). In addition to the CRISPR array, there are usually nearby genes encoding CRISPR-associated (Cas) proteins, including nucleases.

After transcription, CRISPR array RNAs are processed into short guide RNAs (crRNAs) which associate with Cas nucleases to form a crRNA-guided effector complex (Hille et al., 2018). The crRNA base pairs with its complementary sequence in the target DNA or RNA (termed the “protospacer” since it corresponds to the invader nucleic acid that was originally captured and stored as a spacer) and leads to its degradation (interference). For DNA-targeting CRISPR systems, there must be a short activating sequence next to the target (called the Protospacer Adjacent Motif or PAM) for efficient interference (Deveau et al., 2008; Mojica et al., 2009; Shah et al., 2013). New

spacers are added to the array in a process called adaptation, wherein two proteins, Cas1 and Cas2, integrate fragments of DNA (McGinn and Marraffini, 2019) to produce new immune memories.

While the field has made great gains in understanding interference and adaptation in a wide range of organisms, many questions remain. For one, how are the individual immune memories in this heritable and adaptable system maintained over time? New spacers are continuously added in response to novel threats, but most arrays are less than 30 spacers long, suggesting that some immune memories are purged—which ones and how? This review will examine what we have learned about the dynamics of CRISPR arrays, with a focus on how immune memories (the spacers) are organized, maintained, and lost.

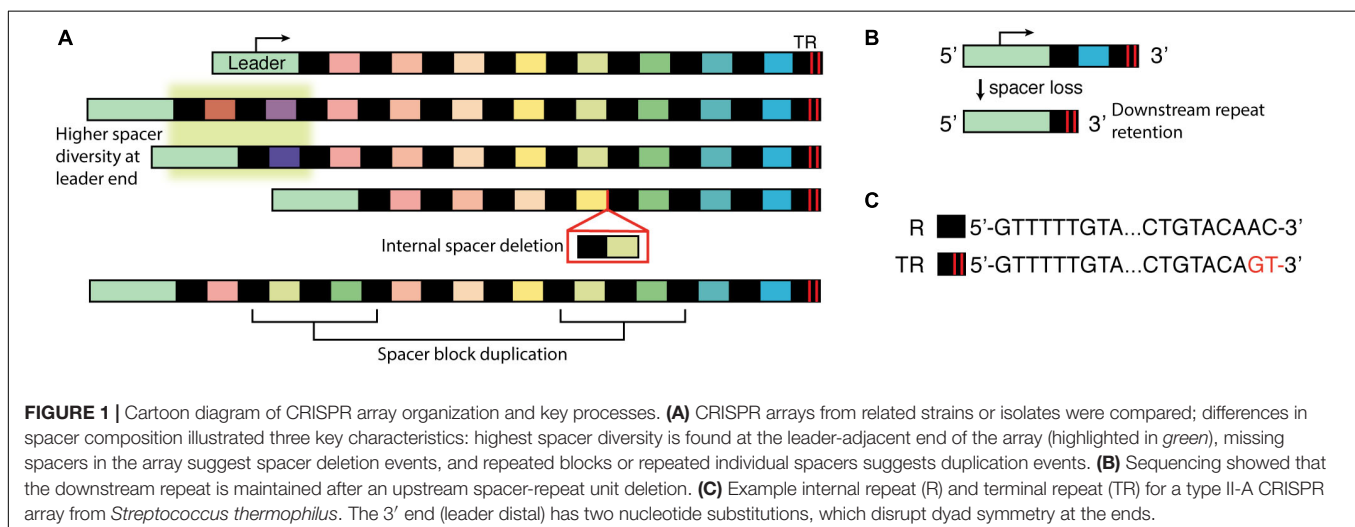
CRISPR Arrays Are Uniquely Organized Sequence Storage Banks

The most notable component of CRISPR-Cas systems is the repeat-spacer array, and the unusual structure of these elements was the first component of CRISPR-Cas to capture researchers attention as they studied nearby genes in *Escherichia coli* (Ishino et al., 1987; Nakata et al., 1989). Other types of repeats had been described in prokaryotic genomes, but in these new elements they found a novel layout: about a dozen direct repeats with loose dyad symmetry were arranged in a regularly spaced array (Figure 1A). The repeats were identical (or nearly identical) in sequence and length, while the intervening spacers had a common length but seemingly random sequence. The authors searched for and found the repeats in genomes of two other species of gram-negative bacteria and other groups found similar repeats in a range of bacteria and archaea (Groenen et al., 1993; Mojica et al., 1993; Mojica et al., 1995; Masepohl et al., 1996; Hoe et al., 1999). The broad distribution and surprisingly well-conserved layout suggested an important functional role for CRISPR arrays (Mojica et al., 2000; Jansen et al., 2002). That role was uncovered through a key observation about spacers: their sequences often matched DNA of mobile genetic elements like plasmids, phages,

and prophages. Thus the CRISPR array appeared to be part of an immune system, with the spacer sequences acting as immune memories (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). This immune function was then confirmed directly. In a lab setting, cultures of a CRISPR-endowed strain of *Streptococcus thermophilus* were almost entirely killed off by lytic phage, but the small number of survivors (bacteriophage-insensitive mutants, BIM) all had at least one new spacer which matched the phage genome (Barrangou et al., 2007).

These and other studies showed that CRISPR-Cas could function as an immune system, and they also began to reveal general characteristics of how new spacers were acquired and stored. First, while studying bacteriophage-insensitive mutants it was noted that new spacers were added to one end of the array (Barrangou et al., 2007); this end contained the “leader,” a 200–300 bp stretch of non-coding DNA (Jansen et al., 2002), which was later shown to regulate spacer uptake and array transcription (Pougach et al., 2010; Yosef et al., 2012). During uptake of a new spacer, the repeat was duplicated, so that an entire spacer-repeat unit was added (Barrangou et al., 2007). Later work showed that new spacer-repeat units could occasionally be added to the interior of the array, termed “ectopic” integration. In rare examples where ectopic integrations appear to outnumber leader-adjacent events, mutations in the leader were found and thought to cause the atypical localization (McGinn and Marraffini, 2016). Recently, ectopic integrations were reported in type II systems of *S. thermophilus* in the absence of leader mutations. While most (83%) integrations were leader-adjacent, the minority of ectopic events show that polarity is typical but not always absolute (Achigar et al., 2021).

Polarity of spacer uptake was not unique to lab-cultured organisms. Whenever CRISPR arrays from related strains were compared, a common pattern emerged: the greatest diversity of spacers was observed near the leader, with many of those spacers being unique to one strain or another, while the distal end of the array tended to have a series of spacers that was shared by many strains (Figure 1A; Pourcel et al., 2005; Lillestol et al., 2006; Horvath et al., 2008; Held et al., 2010; Lopez-Sanchez et al., 2012;



Lier et al., 2015; Rao et al., 2016). This pattern supported a polar and sequential process of spacer addition, with recent events near the leader and ancestral events at the distal end. A careful comparison of *Sulfolobus islandicus* isolates from a single hot spring lent particular support for this model: when two arrays shared non-identical spacers that likely arose from the same viral invader, the spacers were often in the same relative position within their respective arrays. The spacers' positions appeared to serve as a time stamp for the moment when the virus appeared in the spring and was captured into the CRISPR arrays (Held et al., 2010).

Second, it became clear that spacer-repeat units could be duplicated or deleted from the array. These changes were often observed in the middle of the array, while the distal end (termed “trailer” or “anchor” end) was typically conserved (Pourcel et al., 2005; Lillestol et al., 2006; Horvath et al., 2008; Lopez-Sanchez et al., 2012; Weinberger et al., 2012; Lam and Ye, 2019; Deecker and Ensminger, 2020). Evidence of losses or duplications was first inferred by comparing arrays from related strains; arrays that differed only by the absence of one or more contiguous spacers were thought to be the result of deletions (Figure 1A; Pourcel et al., 2005; Lillestol et al., 2006; Held et al., 2010; Gudbergdottir et al., 2011; Lopez-Sanchez et al., 2012; Achigar et al., 2017). Spacer deletion was also sometimes detected while sequencing bacteriophage survivors (Deveau et al., 2008; Achigar et al., 2017). A minority of survivors both lost a contiguous block of existing spacers and added a new spacer against the experimental phage, leading some authors to suggest that “spacer deletion may occur concomitantly with the addition of new spacers” (Deveau et al., 2008). Repeated blocks of spacers were presumed to be duplications rather than independent adaptation events (Bolotin et al., 2005; Lillestol et al., 2006; Held et al., 2010; Lopez-Sanchez et al., 2012; Lier et al., 2015; Stout et al., 2018). While a second encounter with an old invader could conceivably lead to uptake of the same spacer twice, an identical series of spacers is unlikely.

Spacer deletion was particularly apparent in experiments wherein cultures were subjected to a selective pressure that favored failure of interference. For example, Jiang et al. (2013) introduced a conjugative plasmid encoding antibiotic resistance into *Staphylococcus epidermidis* RP62a, which had a type III-A CRISPR system and a spacer targeting the plasmid. When cultures were grown in the presence of antibiotics, interference against the plasmid resulted in 3–4 orders of magnitude fewer transconjugants as compared to controls. However, a few transconjugants were isolated and 13% of these had lost the plasmid-targeting spacer from their array. Other transconjugants had different mutations, all of which would disrupt CRISPR interference and thus allow the plasmid to persist and provide antibiotic resistance. Additional experiments suggested that these mutations arose spontaneously in the population rather than being induced by the selective pressure, and authors estimated that such mutations occurred in roughly one of 10^3 or 10^4 cells (Jiang et al., 2013).

In similar experiments with different organisms, spontaneous deletion of the targeting spacer was responsible for a larger share of escapees, in some cases occurring in more than 80% of the sequenced isolates (Gudbergdottir et al., 2011;

Lopez-Sanchez et al., 2012; Citorik et al., 2014; Rao et al., 2017; Stout et al., 2018; Canez et al., 2019). Deletion often included blocks of spacer-repeat units rather than only the targeting spacer (Gudbergdottir et al., 2011; Lopez-Sanchez et al., 2012; Stout et al., 2018) and it was also sometimes associated with duplications of other non-targeting spacers (Lopez-Sanchez et al., 2012). As in the *S. epidermidis* work, rearrangements of the array were found even in the absence of selective pressure: for a strain of *Legionella pneumophila* bearing an engineered short array, roughly one of every 1,000–2,000 cells underwent a spontaneous spacer-repeat deletion (Rao et al., 2017). Sequencing revealed that the boundaries of the downstream repeat were maintained after the deletion, leading the authors to hypothesize that homologous recombination between repeats underlies array rearrangements (Figure 1B; Gudbergdottir et al., 2011; Rao et al., 2017).

Mechanisms and Functions for Polarized Spacer Uptake

Polarized, leader-end addition of spacers was a reproducible observation and mechanisms soon emerged to show how it occurs. Cas1 and Cas2 are necessary for spacer uptake (Yosef et al., 2012) and are associated with all adaptation-active systems (Makarova et al., 2020). These two proteins are necessary and sufficient for *in vitro* integration (Nunez et al., 2015). In some organisms, Cas1 and Cas2 strongly favor integration at the leader-adjacent repeat and this bias is mediated by sequences in the leader (Wei et al., 2015; McGinn and Marraffini, 2016; Wright and Doudna, 2016; Xiao et al., 2017; Kim et al., 2019). But for other systems, *in vitro* experiments show that Cas1 and Cas2 alone will carry out integrations at other repeats in the array and even at repeat-like sequences outside the array (Nunez et al., 2015; Grainy et al., 2019). These same systems show polarized integration *in vivo* (Datsenko et al., 2012; Yosef et al., 2012; Shiimori et al., 2018), suggesting that additional factors can guide the reaction. In type I systems of *E. coli* and other bacteria, a protein called integration host factor (IHF) ensures polarization by binding to Cas1 and the leader (Nunez et al., 2016; Fagerlund et al., 2017; Wright et al., 2017). Other protein factors likely play a similar polarizing role in other organisms and await characterization (Rollie et al., 2018).

Several studies show that leader-adjacent integration is likely necessary for optimal immune function. A new spacer arises from contemporary mobile genetic elements, which likely represent the most current and therefore pressing threats for host cells. In addition, new spacers should be free of mismatches that accumulate for older spacers as their targets develop escape mutations (Deveau et al., 2008; Semenova et al., 2011; Cady et al., 2012; Rao et al., 2016). From that we could expect that leader-adjacent spacers would be prioritized for defense, and these spacers do indeed produce more robust interference (McGinn and Marraffini, 2016; Rao et al., 2016; Deecker and Ensminger, 2020). The mechanism underlying this difference is not entirely clear. Leader-adjacent spacers (and the crRNAs they encode) may be better expressed or more efficiently processed than downstream spacers. RNA sequencing data show more abundant crRNAs in the leader half of the array for many

CRISPR loci (Elmore et al., 2013; Carte et al., 2014; McGinn and Marraffini, 2016). Another possibility rests on the idea that individual crRNAs are essentially in competition to form a crRNP effector complex with less numerous Cas nucleases. As the first to be transcribed, leader-adjacent spacers may have a head start in a race to associate with Cas proteins and could suffer the least from the “dilution” effect of multiple spacers (Martynov et al., 2017). Experiments with a constructed mini-array lent support to the general idea of competition: a truncated array was created in a strain of *Legionella pneumoniae* by deleting all but the leader-adjacent spacer and its upstream and downstream repeats. This mini-array strain showed about 100-fold more plasmid targeting than the wildtype strain, which has 42 additional spacers downstream (Rao et al., 2017). Though the sequence and position was identical for the first spacer, loss of additional spacers dramatically increased its effectiveness.

CRISPR Arrays Vary in Length

In addition to influencing the polarity of spacer uptake, the dilution effect may also represent a functional constraint on the overall length of CRISPR arrays. Adaptation without spacer loss would presumably lead to ever-longer arrays, but among genomes sequenced so far, extremely long arrays are relatively rare. Arrays with greater than 100 spacers are observed; *Haliangium ochraceum* is a notable example, with a single array of 587 spacers and two other arrays measuring 189 and 36 spacers, respectively (Ivanova et al., 2010; Pourcel et al., 2020). However, a typical array contains fewer than 50 spacers in bacteria and fewer than 100 in archaea (Horvath et al., 2008; Mangericao et al., 2016; Pourcel et al., 2020). Array length does not appear to be limited by genome size (Pourcel et al., 2020) nor by cell resources: experimentally, lengthening an array by several spacers did not reduce fitness (Vale et al., 2015). Also, many genomes harbor more than one CRISPR system (up to 37 have been observed, in a species of *Actinoalloteichus*), and presumably the energy demands for a single 500 spacer array are similar to those for ten 50 spacer arrays. On the other hand, *cas* genes can have a fitness cost (Vale et al., 2015), so the observation that many organisms have evolved multiple short arrays suggests that array length is not limited by energetic costs of carrying extra spacers.

One hypothesis to explain array length patterns is that array size represents a tradeoff between the dilution effect described above and maintaining immunity to a range of potential threats, i.e., depth of immunity (Bradde et al., 2020). In turn, depth of immunity is balanced against the need to update the array frequently enough to contend with novel threats but not so frequently that the cell risks toxic auto-immunity (Stern et al., 2010; Vercoe et al., 2013; Weissman et al., 2018). Organisms may deal with dueling constraints by having multiple arrays, each with a different length and optimized depth of immunity (Weissman et al., 2018). This would imply that arrays can have different rates of both spacer uptake and loss. Regarding uptake, evidence already exists that adaptation efficiency varies among systems and can also change in response to certain cues like cell density (Hoyland-Kroghsbo et al., 2017; see Sternberg et al. (2016) for a general review of adaptation). Data on spacer loss is sparser, but at least one report suggests that the frequency of spacer loss can

differ between systems in the same organism. Specifically, when otherwise identical plasmids with either a type I or type II mini-array were grown in *E. coli*, frequent spacer loss was observed for type I but not type II (Canez et al., 2019).

Spacer Turnover Is Not Strictly Chronological

As a spacer's residence time in an array increases, and it loses relevance, position, and sequence identity to its targets, we might expect selective pressures to no longer favor its maintenance. In that context, deletion events could be a useful means for shedding older spacers. However, multiple observations suggest that old spacers are not purged in a chronological manner and that mismatched or inefficient spacers may prove useful. A minority of older spacers can maintain identity to their protospacer targets, possibly due to stable or cyclical exposure to phages (Sun et al., 2016). We also now know that relatively ineffective spacers can participate in immunity through the process of primed adaptation. In short, priming occurs when a crRNP effector complex recognizes a protospacer target and then stimulates new spacer uptake using DNA located near that target (Datsenko et al., 2012; Swarts et al., 2012). Priming is observed even when interference is relatively inefficient, like when the protospacer does not have a canonical PAM or when there are mismatches between the crRNA and the protospacer, particularly in the “seed” region adjacent to the PAM (Semenova et al., 2011; Wiedenheft et al., 2011; Fineran et al., 2014; Li et al., 2014; Richter et al., 2014; Semenova et al., 2016; Garrett et al., 2020). Since primed adaptation tolerates these changes, a spacer that might otherwise be obsolete can contribute to CRISPR immunity by updating the CRISPR array. Experimentally, spacers in the middle of an array (*L. pneumophila*) were shown to give relatively inefficient interference but still effectively support priming (Deecker and Ensminger, 2020). Thus turnover of older spacers may not always appear steady or strictly chronological.

Many studies that demonstrated the polarity of spacer acquisition also described the relative stability of the array's trailer end (Lopez-Sanchez et al., 2012; Weinberger et al., 2012; Lam and Ye, 2019). Assuming sequential spacer uptake, we would expect these terminal spacers to be the oldest and thus the most likely to have lost protective potential. Phylogenetic relationships inferred from multilocus sequence typing supported the idea that terminal spacers are indeed ancestral (Lopez-Sanchez et al., 2012). These spacers should be lost if shedding is chronological, yet they are apparently deleted far less frequently than newer spacers toward the middle of the array. Therefore trailer end spacers may be maintained for reasons unrelated to their value in interference. The stability may be a simple outcome of fewer opportunities for recombination: an internal spacer can be lost through recombination involving any upstream or downstream repeat, but the last spacer would only be lost if recombination occurred at the terminal repeat. Terminal repeats may also be stabilized due to polymorphisms: in many systems the repeat sequences are identical throughout the array except at the end (Jansen et al., 2002; Bolotin et al., 2005; Pourcel et al., 2005; Horvath et al., 2008; Lopez-Sanchez et al., 2012;

Deecker and Ensminger, 2020; Refregier et al., 2020). Specifically, nucleotide substitutions or deletions are often found in the 3' end of the terminal repeat (Figure 1C). Since identical repeats are most amenable to homologous recombination (Treangen et al., 2009), a trailer repeat without polymorphisms could potentially undergo recombination with the leader-adjacent repeat and eliminate the entire array, leaving only a copy of itself. Terminal repeat polymorphisms may thereby tend to reduce the likelihood of array collapse.

What would array collapse mean for immunity in the CRISPR locus? Experiments have confirmed that naïve adaptation can occur with a solitary leader-adjacent repeat (Yosef et al., 2012; Wei et al., 2015), which suggests that arrays could potentially be repopulated following a collapse, at least in laboratory conditions. Deecker and Ensminger (2020) also found evidence that “array collapse and repopulate” events occur naturally. First they showed that priming *in trans* could replenish a collapsed array in the lab: their strain of *L. pneumophila* naturally contains a type I-F system on both its chromosome and its endogenous plasmid. The chromosomal array was mutated to only contain the terminal repeat, mimicking a collapsed array. They transformed in a plasmid targeted by a spacer in the endogenous plasmid array and observed primed adaptation into the chromosomal collapsed array. The authors noted that patterns of repeat polymorphisms among naturally occurring *L. pneumophila* isolates looked like the replenished arrays they had created in the lab, suggesting this happens in nature. However, it remains unclear whether replenishment of a collapsed array is a universal phenomenon. If collapse is not well tolerated, terminal repeat polymorphisms may be functionally important in preventing it. On the other hand, if a system can readily bounce back from array collapse, terminal repeat polymorphisms may simply represent spontaneous mutations that persist because they are resistant to loss through recombination. Interestingly, many terminal repeat polymorphisms are nucleotide substitutions or truncations in the 3' end, which partially disrupt the loose dyad symmetry of repeat ends (Jansen et al., 2002; Bolotin et al., 2005; Pourcel et al., 2005; Horvath et al., 2008; Lopez-Sanchez et al., 2012; Deecker and Ensminger, 2020). Since dyad symmetry is a frequently observed feature of repeats, one could speculate that loss of dyad symmetry helps stabilize terminal repeats.

From the studies discussed above, we can conclude that the trailer end of the array typically does not obey a pattern of chronological turnover. In an extreme example, spacer turnover across the entire array also bucks chronological turnover, even over the course of thousands of years. Savitskaya et al. (2017) acquired an intestinal microbiome sample from a well-preserved mammoth calf that was frozen for 42,000 years and they captured *E. coli* type I repeat-spacer amplicons by PCR. Reads primarily yielded data about individual spacers but a subset of reads were long enough to span two or three repeat-spacer units and these provided additional information about spacer order in the ancient arrays. Ancient spacers and spacer combinations were then compared to over 1,700 modern *E. coli* type I-E arrays from public databases. About 20% of the ancient spacers matched a modern spacer, and surprisingly those matches

were positioned all over the modern arrays rather than being concentrated in the distal end. Trends for the two and three-spacer data were similar. This striking example demonstrated that for some systems, spacer order may not recapitulate a timeline of spacer acquisition.

Mechanisms for Spacer-Repeat Rearrangement

Repeats have often been associated with genome plasticity, and rearrangement of a repeat element like the CRISPR array is consistent with those observations. Repeats can undergo recombination through two general mechanisms: RecA-dependent homologous recombination and RecA-independent mechanisms like replication misalignment (slippage or slipped-strand mispairing) (Bzymek and Lovett, 2001; Treangen et al., 2009). In homologous recombination, RecA protein plays a key role as it binds and coats ssDNA and promotes strand exchange and annealing once it has found a region with sufficient sequence identity. The branched heteroduplex is then extended and resolved by, for example, RuvABC complex (Kowalczykowski, 2015). RecA-independent mechanisms also rely on homology but there are multiple distinguishing characteristics. First, RecA-independent mechanisms have shorter homology length requirements and are thought to be the primary source of recombination in prokaryotes for repeats that are less than about 200 bp (Bi and Liu, 1994; Lovett, 2004). Second, RecA-independent recombination is thought to involve the replication fork: in replication misalignment (the most well-described form of RepA-independent recombination), direct repeats mispair during replication, giving rise to duplications and deletions (Lovett, 2004). The frequency of this type of recombination increases with repeat length and identity and decreases as the spacing between repeats grows (Lovett, 2004). Disruption of replication can further promote misalignments and increase deletions (Michel, 2000).

Given that CRISPR repeats are short and closely spaced, a RecA-independent mechanism like replication misalignment could underlie array deletions and duplications. RecA is not necessary for adaptation (Ivancic-Bace et al., 2015; Radovic et al., 2018). While there is not yet direct evidence for or against a role for RecA in CRISPR array rearrangements, a report about recombination in a CRISPR-derived system hints that it is not necessary. Ding et al. (2020) sought to improve the performance of dual guide RNA plasmids in CRISPR-based genome editing applications. The plasmids were designed to express two separate guide RNAs: each 20 bp guide spacer had an identical promoter (35 bp) upstream and identical “scaffold” (82 bp) downstream (the scaffold included 12 bp corresponding to the 5' end of the CRISPR repeat, a 4 bp linker, and “tracr”, the trans-activating crisper RNA important for forming a mature guide RNA). They found that the plasmid was extremely unstable: 73% had mutations, mostly deletions that excised either one of the two promoter-spacer-scaffold units. Changing the promoter to reduce the extent of homology did not eliminate deletions. The group also observed no reduction

in deletion frequency using strains with deleted or inactive RecA, showing that the process was not RecA dependent. Ultimately the group found that inverting one of the promoter-spacer-scaffold units was necessary to stop the deletion events. Interestingly, the authors found that growth and transformation conditions also influenced the frequency of deletions. Using electroporation instead of heat shock and culturing in rich growth medium both reduced (but not eliminated) deletions. They hypothesized that nutrient deprivation and DNA damage slow replication and thereby promote deletion through a replication mispairing mechanism on the lagging strand. These findings might not be directly applicable to native CRISPR arrays (for example, the homologous region in their engineered plasmid was longer—82 bp when different promoters were used), but they imply that natural spacer deletions could also be RecA-independent, possibly occurring through misalignment between repeats during replication (Figure 2).

DISCUSSION

Bringing together these different observations and experimental results, we can speculate on a general model for spacer dynamics in a CRISPR array: new spacers are added at the leader end of the array at some basic frequency, which varies among species, systems, and conditions. Rearrangement of the array is ongoing at some level, though the particular frequency is also variable among species and CRISPR-Cas classes and it may be modulated by as-of-yet unidentified factors and conditions. These rearrangements can lead to both deletions and duplications, and the interplay between spacer addition and loss determines array length and underpins the balance of immune depth, immune novelty, and crRNA dilution for that array. The terminal spacer-repeat unit rarely participates in rearrangements, potentially because of polymorphisms, so the array is maintained and the last spacer-repeat unit is stable. Together with adaptation events, rearrangements present immunogenic diversity on which selection can act. In most circumstances the dominant array form persists for generations, but the system is poised for change should conditions shift.

There is much more to learn about the dynamics and outcomes of spacer turnover. For one, it will be interesting

to know how common array rearrangements are in different natural populations. Often these events were only detected because of a strong selective pressure against interference—how frequent are they in a native array under neutral conditions? Is the frequency consistent or does it vary with or independently from adaptation frequency? Since evidence suggests that the arrangements may not be equally common for all systems (Canez et al., 2019), it will be worthwhile to explore their frequency in multiple species and conditions. Long-read sequencing approaches may be particularly suitable for these experiments since they can capture the spacer composition of an entire array without the ambiguities inherent to assembled short reads.

Second, we have much to learn about the mechanism by which spacers are duplicated or deleted. The nature of CRISPR repeats and patterns of spacer loss are suggestive of rearrangement by recombination, but direct data are needed. For example, do deletions and duplications arise from RecA-independent mechanisms like misalignment in the replication fork? This would be supported by the results from the dual guide RNA plasmid experiments described above (Ding et al., 2020), and if experimentally confirmed, it could have interesting implications for immune diversity. If, for example, deletions primarily occur on the lagging strand during replication, we would expect them to be passed along to only one of two daughter cells. Since autoimmunity is thought to represent a fitness cost associated with CRISPR-Cas (Stern et al., 2010; Vercoe et al., 2013), replication fork deletion of new spacers may present a way to hedge against toxic self-targeting adaptation. If replication fork misalignment does underlie array rearrangements, are there factors or conditions that promote or inhibit the process and do they regulate spacer maintenance? And looking beyond the replication misalignment model, are there other enzymes or processes that can lead to spacer deletions or duplications? As we have learned more about adaptation and interference, points where the processes are modulated have been uncovered. Similarly, answering these and other questions about array dynamics may also help us uncover novel mechanisms that govern how existing spacers are managed to optimize immunity.

AUTHOR CONTRIBUTIONS

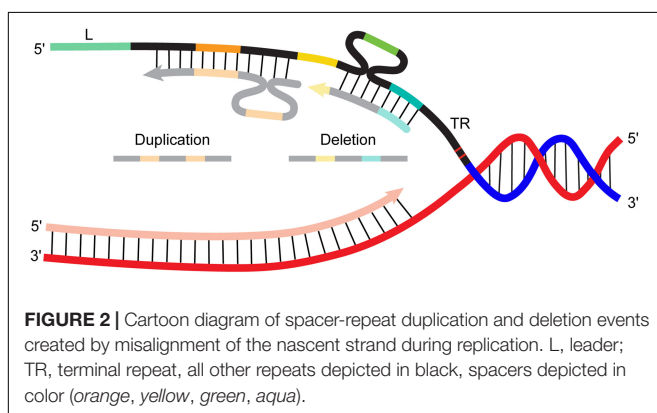
SG devised and wrote the manuscript.

FUNDING

This work was supported by the National Institutes of Health (R35GM118140).

ACKNOWLEDGMENTS

I thank Brenton Graveley, members of the Graveley lab, and Michael Terns for helpful discussion during preparation of this work.



REFERENCES

- Achigar, R., Magadan, A. H., Tremblay, D. M., Julia Pianzzola, M., and Moineau, S. (2017). Phage-host interactions in *Streptococcus thermophilus*: genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. *Sci. Rep.* 7:43438.
- Achigar, R., Scarrone, M., Rousseau, G. M., Philippe, C., Machado, F., Duvos, V., et al. (2021). Ectopic spacer acquisition in *Streptococcus thermophilus* CRISPR3 array. *Microorganisms* 9:512. doi: 10.3390/microorganisms9030512
- Alkhnbashi, O. S., Shah, S. A., Garrett, R. A., Saunders, S. J., Costa, F., and Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics* 32, i576–i585.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140
- Bi, X., and Liu, L. F. (1994). *recA*-independent and *recA*-dependent intramolecular plasmid recombination. Differential homology requirement and distance effect. *J. Mol. Biol.* 235, 414–423. doi: 10.1006/jmbi.1994.1002
- Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561. doi: 10.1099/mic.0.28048-0
- Bradde, S., Nourmohammad, A., Goyal, S., and Balasubramanian, V. (2020). The size of the immune repertoire of bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 117, 5144–5151. doi: 10.1073/pnas.1903666117
- Bzymek, M., and Lovett, S. T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8319–8325. doi: 10.1073/pnas.111008398
- Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R., and O'Toole, G. A. (2012). The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.* 194, 5728–5738. doi: 10.1128/jb.01184-12
- Canez, C., Selle, K., Goh, Y. J., and Barrangou, R. (2019). Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*. *FEMS Microbiol. Lett.* 366:fnz105.
- Carte, J., Christopher, R. T., Smith, J. T., Olson, S., Barrangou, R., Moineau, S., et al. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Mol. Microbiol.* 93, 98–112. doi: 10.1111/mmi.12644
- Citorik, R. J., Mimee, M., and Lu, T. K. (2014). Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* 32, 1141–1145. doi: 10.1038/nbt.3011
- Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* 3:945.
- Deecker, S. R., and Ensminger, A. W. (2020). Type I-F CRISPR-Cas distribution and array dynamics in *Legionella pneumophila*. *G3 (Bethesda)* 10, 1039–1050. doi: 10.1534/g3.119.400813
- Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., et al. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390–1400. doi: 10.1128/jb.01412-07
- Ding, T., Huang, C., Liang, Z., Ma, X., Wang, N., and Huo, Y. X. (2020). Reversed paired-gRNA plasmid cloning strategy for efficient genome editing in *Escherichia coli*. *Microb. Cell Fact.* 19:63.
- Elmore, J. R., Yokooji, Y., Sato, T., Olson, S., Glover, C. V. III, Graveley, B. R., et al. (2013). Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. *RNA Biol.* 10, 828–840. doi: 10.4161/rna.24084
- Fagerlund, R. D., Wilkinson, M. E., Klykov, O., Barendregt, A., Pearce, F. G., Kieper, S. N., et al. (2017). Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proc. Natl. Acad. Sci. U.S.A.* 114, E5122–E5128.
- Fineran, P. C., Gerritzen, M. J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S. A., et al. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1629–E1638.
- Garrett, S., Shimori, M., Watts, E. A., Clark, L., Graveley, B. R., and Terns, M. P. (2020). Primed CRISPR DNA uptake in *Pyrococcus furiosus*. *Nucleic Acids Res.* 48, 6120–6135. doi: 10.1093/nar/gkaa381
- Grainy, J., Garrett, S., Graveley, B. R., and PTerns, M. (2019). CRISPR repeat sequences and relative spacing specify DNA integration by *Pyrococcus furiosus* Cas1 and Cas2. *Nucleic Acids Res.* 47, 7518–7531. doi: 10.1093/nar/gkz548
- Groenen, P. M., Bunschoten, A. E., van Soolingen, D., and van Embden, J. D. (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* 10, 1057–1065. doi: 10.1111/j.1365-2958.1993.tb00976.x
- Gudbergdottir, S., Deng, L., Chen, Z., Jensen, J. V., Jensen, L. R., She, Q., et al. (2011). Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79, 35–49. doi: 10.1111/j.1365-2958.2010.07452.x
- Held, N. L., Herrera, A., Cadillo-Quiroz, H., and Whitaker, R. J. (2010). CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS One* 5:e12988. doi: 10.1371/journal.pone.0012988
- Hille, F., Richter, H., Wong, S. P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The biology of CRISPR-Cas: backward and forward. *Cell* 172, 1239–1259. doi: 10.1016/j.cell.2017.11.032
- Hoe, N., Nakashima, K., Grigsby, D., Pan, X., Dou, S. J., Naidich, S., et al. (1999). Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerg. Infect. Dis.* 5, 254–263. doi: 10.3201/eid0502.990210
- Horvath, P., Romero, D. A., Coute-Monvoisin, A. C., Richards, M., Deveau, H., Moineau, S., et al. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1401–1412. doi: 10.1128/jb.01415-07
- Hoyland-Kroghsbo, N. M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., et al. (2017). Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc. Natl. Acad. Sci. U.S.A.* 114, 131–135.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433. doi: 10.1128/jb.169.12.5429-5433.1987
- Ivancic-Bace, I., Cass, S. D., Wearne, S. J., and Bolt, E. L. (2015). Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res.* 43, 10821–10830. doi: 10.1093/nar/gkv1213
- Ivanova, N., Daum, C., Lang, E., Abt, B., Kopitz, M., Saunders, E., et al. (2010). Complete genome sequence of *Haliangium ochraceum* type strain (SMP-2). *Stand. Genomic Sci.* 2, 96–106. doi: 10.4056/sgs.69.1277
- Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x
- Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B. R., and Marraffini, L. A. (2013). Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet.* 9:e1003844. doi: 10.1371/journal.pgen.1003844
- Kim, J. G., Garrett, S., Wei, Y., Graveley, B. R., and Terns, M. P. (2019). CRISPR DNA elements controlling site-specific spacer integration and proper repeat length by a Type II CRISPR-Cas system. *Nucleic Acids Res.* 47, 8632–8648. doi: 10.1093/nar/gkz677
- Koonin, E. V., Makarova, K. S., and Wolf, Y. I. (2017). Evolutionary genomics of defense systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* 71, 233–261. doi: 10.1146/annurev-micro-090816-093830
- Kowalczykowski, S. C. (2015). An overview of the molecular mechanisms of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* 7:a016410. doi: 10.1101/cshperspect.a016410
- Lam, T. J., and Ye, Y. (2019). Long reads reveal the diversification and dynamics of CRISPR reservoir in microbiomes. *BMC Genomics* 20:567. doi: 10.1186/s12864-019-5922-8
- Li, M., Wang, R., and Xiang, H. (2014). *Haloarcula hispanica* CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res.* 42, 7226–7235. doi: 10.1093/nar/gku389
- Lier, C., Baticle, E., Horvath, P., Haguenoer, E., Valentin, A. S., Glaser, P., et al. (2015). Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. *Front. Genet.* 6:214. doi: 10.3389/fgene.2015.00214
- Lillestol, R. K., Redder, P., Garrett, R. A., and Brugger, K. (2006). A putative viral defense mechanism in archaeal cells. *Archaea* 2, 59–72. doi: 10.1155/2006/542818
- Lopez-Sanchez, M. J., Sauvage, E., Da Cunha, V., Clermont, D., Ratsima Hariniaina, E., Gonzalez-Zorn, B., et al. (2012). The highly dynamic CRISPR1

- system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol. Microbiol.* 85, 1057–1071. doi: 10.1111/j.1365-2958.2012.08172.x
- Lovett, S. T. (2004). Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 52, 1243–1253. doi: 10.1111/j.1365-2958.2004.04076.x
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi: 10.1038/s41579-019-0299-x
- Mangericao, T. C., Peng, Z., and Zhang, X. (2016). Computational prediction of CRISPR cassettes in gut metagenome samples from Chinese type-2 diabetic patients and healthy controls. *BMC Syst. Biol.* 10(Suppl. 1):5. doi: 10.1186/s12918-015-0248-x
- Martynov, A., Severinov, K., and Ispolatov, I. (2017). Optimal number of spacers in CRISPR arrays. *PLoS Comput. Biol.* 13:e1005891. doi: 10.1371/journal.pcbi.1005891
- Masepohl, B., Gorlitz, K., and Bohme, H. (1996). Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Biochim. Biophys. Acta* 1307, 26–30. doi: 10.1016/0167-4781(96)00040-1
- McGinn, J., and Marraffini, L. A. (2016). CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. *Mol. Cell* 64, 616–623. doi: 10.1016/j.molcel.2016.08.038
- McGinn, J., and Marraffini, L. A. (2019). Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* 17, 7–12. doi: 10.1038/s41579-018-0071-7
- Michel, B. (2000). Replication fork arrest and DNA recombination. *Trends Biochem. Sci.* 25, 173–178. doi: 10.1016/s0968-0004(00)01560-7
- Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182. doi: 10.1007/s00239-004-0046-3
- Mojica, F. J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* 36, 244–246. doi: 10.1046/j.1365-2958.2000.01838.x
- Mojica, F. J., Ferrer, C., Juez, G., and Rodriguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloflex mediterranei* and *Haloflex volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* 17, 85–93. doi: 10.1111/j.1365-2958.1995.mmi_17010085.x
- Mojica, F. J., Juez, G., and Rodriguez-Valera, F. (1993). Transcription at different salinities of *Haloflex mediterranei* sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* 9, 613–621. doi: 10.1111/j.1365-2958.1993.tb01721.x
- Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* 155, 733–740. doi: 10.1099/mic.0.023960-0
- Nakata, A., Amemura, M., and Makino, K. (1989). Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* 171, 3553–3556. doi: 10.1128/jb.171.6.3553-3556.1989
- Nunez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., and Doudna, J. A. (2016). CRISPR immunological memory requires a host factor for specificity. *Mol. Cell* 62, 824–833. doi: 10.1016/j.molcel.2016.04.027
- Nunez, J. K., Lee, A. S., Engelman, A., and Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193–198. doi: 10.1038/nature14237
- Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K. A., Djordjevic, M., Wanner, B. L., et al. (2010). Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* 77, 1367–1379. doi: 10.1111/j.1365-2958.2010.07265.x
- Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653–663. doi: 10.1099/mic.0.27437-0
- Pourcel, C., Touchon, M., Villeriot, N., Vernadet, J. P., Couvin, D., Toffano-Nioche, C., et al. (2020). CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic Acids Res.* 48, D535–D544.
- Radovcic, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E. L., and Ivancic-Bace, I. (2018). CRISPR-Cas adaptation in *Escherichia coli* requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5' ssDNA exonucleases. *Nucleic Acids Res.* 46, 10173–10183.
- Rao, C., Chin, D., and Ensminger, A. W. (2017). Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss. *RNA* 23, 1525–1538. doi: 10.1261/rna.062083.117
- Rao, C., Guyard, C., Pelaz, C., Wasserscheid, J., Bondy-Denomy, J., Dewar, K., et al. (2016). Active and adaptive *Legionella* CRISPR-Cas reveals a recurrent challenge to the pathogen. *Cell. Microbiol.* 18, 1319–1338. doi: 10.1111/cmi.12586
- Refregier, G., Sola, C., and Guyeux, C. (2020). Unexpected diversity of CRISPR unveils some evolutionary patterns of repeated sequences in *Mycobacterium tuberculosis*. *BMC Genomics* 21:841. doi: 10.1186/s12864-020-07178-6
- Richter, C., Dy, R. L., McKenzie, R. E., Watson, B. N., Taylor, C., Chang, J. T., et al. (2014). Priming in the type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* 42, 8516–8526. doi: 10.1093/nar/gku527
- Rollie, C., Graham, S., Rouillon, C., and White, M. F. (2018). Pre-spacer processing and specific integration in a type I-A CRISPR system. *Nucleic Acids Res.* 46, 1007–1020. doi: 10.1093/nar/gkx1232
- Savitskaya, E., Lopatina, A., Medvedeva, S., Kapustin, M., Shmakov, S., Tikhonov, A., et al. (2017). Dynamics of *Escherichia coli* type I-E CRISPR spacers over 42 000 years. *Mol. Ecol.* 26, 2019–2026. doi: 10.1111/mec.13961
- Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10098–10103.
- Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K. A., et al. (2016). Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7626–7631. doi: 10.1073/pnas.1602639113
- Shah, S. A., Erdmann, S., Mojica, F. J., and Garrett, R. A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* 10, 891–899. doi: 10.4161/rna.23764
- Shiimori, M., Garrett, S. C., Graveley, B. R., and Terns, M. P. (2018). Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Mol. Cell* 70, 814–824.e816.
- Stern, A., Keren, L., Wurtzel, O., Amitai, G., and Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26, 335–340. doi: 10.1016/j.tig.2010.05.008
- Sternberg, S. H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas systems. *Mol. Cell* 61, 797–808.
- Stout, E. A., Sanozky-Dawes, R., Goh, Y. J., Crawley, A. B., Klaenhammer, T. R., and Barrangou, R. (2018). Deletion-based escape of CRISPR-Cas9 targeting in *Lactobacillus gasseri*. *Microbiology (Reading)* 164, 1098–1111. doi: 10.1099/mic.0.000689
- Sun, C. L., Thomas, B. C., Barrangou, R., and Banfield, J. F. (2016). Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *ISME J.* 10, 858–870. doi: 10.1038/ismej.2015.162
- Swarts, D. C., Mosterd, C., van Passel, M. W., and Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One* 7:e35888. doi: 10.1371/journal.pone.0035888
- Treangen, T. J., Abraham, A. L., Touchon, M., and Rocha, E. P. (2009). Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.* 33, 539–571. doi: 10.1111/j.1574-6976.2009.00169.x
- Vale, P. F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., and Gandon, S. (2015). Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proc. Biol. Sci.* 282:20151270. doi: 10.1098/rspb.2015.1270
- Vercoe, R. B., Chang, J. T., Dy, R. L., Taylor, C., Gristwood, T., Clulow, J. S., et al. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.* 9:e1003454. doi: 10.1371/journal.pgen.1003454
- Wei, Y., Chesne, M. T., Terns, R. M., and Terns, M. P. (2015). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res.* 43, 1749–1758. doi: 10.1093/nar/gku1407

- Weinberger, A. D., Sun, C. L., Plucinski, M. M., Deneff, V. J., Thomas, B. C., Horvath, P., et al. (2012). Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* 8:e1002475. doi: 10.1371/journal.pcbi.1002475
- Weissman, J. L., Fagan, W. F., and Johnson, P. L. F. (2018). Selective maintenance of multiple CRISPR arrays across prokaryotes. *CRISPR J.* 1, 405–413. doi: 10.1089/crispr.2018.0034
- Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10092–10097. doi: 10.1073/pnas.1102716108
- Wright, A. V., and Doudna, J. A. (2016). Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.* 23, 876–883. doi: 10.1038/nsmb.3289
- Wright, A. V., Liu, J. J., Knott, G. J., Doxzen, K. W., Nogales, E., and Doudna, J. A. (2017). Structures of the CRISPR genome integration complex. *Science* 357, 1113–1118. doi: 10.1126/science.aao0679
- Xiao, Y., Ng, S., Nam, K. H., and Ke, A. (2017). How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature* 550, 137–141. doi: 10.1038/nature24020
- Yosef, I., Goren, M. G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 40, 5569–5576. doi: 10.1093/nar/gks216

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Garrett. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.