Print ISSN 2319-2003 | Online ISSN 2279-0780

IJBCP International Journal of Basic & Clinical Pharmacology

doi: http://dx.doi.org/10.18203/2319-2003.ijbcp20150361

Review Article

Clustered regularly interspaced short palindromic repeats cas systems: a comprehensive review

Shahin Mahmud[1], Jahir Ahmed[1], Md. Abdul Aziz[1], Mirza Rokibul Hasan[1], Shoaib Mahmud Shaon[2], Md. Naieem Al-Hasan Bhuiyan[2], Md. Farzanoor Rahman[1], Hasibul Haque Rakib[1], Md. Shariful Islam[1]*

[1]Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University, Tangail-1902, Bangladesh, [2]Department of Biochemistry and Molecular Biology, Jahangirnagar University, Savar-1342, Bangladesh

Received: 18 June 2015
Accepted: 05 July 2015

***Correspondence to:**
Md. Shariful Islam,
Email: sharifbge@gmail.com

ABSTRACT

The clustered regularly interspaced short palindromic repeats (CRISPR) system was recently identified as a bacterial defense mechanism against phages and plasmids. The CRISPR system is composed of DNA arrays containing short sequences identical to those present in phages and plasmids. These short DNAs are transcribed and processed by CRISPR associated proteins that also guide other CRISPR proteins to target the invading DNA. Only a few of the CRISPR components have been characterized to date, and their mechanism of action is still largely unknown. Phage defense mechanisms probably have co-evolved against the CRISPR system, but none has yet been found. We propose to identify phage genes that counteract the CRISPR system.

Keywords: Clustered regularly interspaced short palindromic repeats system, Plasmid, DNA, Short sequences

INTRODUCTION

Bacteriophages constitute the most populous life-forms on Earth.[1] In sea water, an environment in which phage abundance has been extensively studied, it has been estimated that there is 5-10 phage for every bacterial cell.[2] Despite being outnumbered by the phage, bacteria proliferate and avoid extinction by using a battery of innate phage resistance mechanisms such as restriction enzymes and abortive infection.[3] In this Progress article, we describe the CRISPR system, a recently discovered defense mechanism, which is remarkable because it confers acquired phage resistance in Bacteria and Archaea. A hallmark of this system is arrayed of short direct repeats interspersed by non-repetitive spacer sequences, the so-called clustered regularly interspaced short palindromic repeats (CRISPR). Additional components of the system include CRISPR associated (CAS) genes and a leader sequence. The recently discovered CRISPR-Cas adaptive immune system is present in almost all archaea and many bacteria. It consists of cassettes of CRISPR repeats that incorporate spacers homologous to fragments of viral or plasmid genomes that are employed as guide RNAs in the immune response, along with numerous CRISPR-associated (*cas*) genes that encode proteins possessing diverse, only partially characterized activities required for the action of the system. Here, we investigate

the evolution of the *cas* genes and show that they evolve under purifying selection that is typically much weaker than the median strength of purifying selection affecting genes in the respective genomes. The exceptions are the *cas1* and *cas2* genes that typically evolve at levels of purifying selection close to the genomic median. Thus, although these genes are implicated in the acquisition of spacers from alien genomes, they do not appear to be directly involved in an arms race between bacterial and archaeal hosts and infectious agents. These genes might possess functions distinct from and additional to their role in the CRISPR-Cas mediated immune response. Taken together with evidence of the frequent horizontal transfer of *cas* genes reported previously and with the wide-spread micro scale recombination within these genes detected in this work, these findings reveal the highly dynamic evolution of *cas* genes. This conclusion is in line with the involvement of CRISPR-Cas in antiviral immunity that is likely to entail a co-evolutionary arms race with rapidly evolving viruses. However, we failed to detect evidence of strong positive selection in any of the *cas* genes. A novel family of repeats, the CRISPR, were discovered in almost all archaeal genomes including about 40% of sequenced bacterial genomes. CRISPRs are composed of short direct repeats (24-47 bp) that are interspaced by non-repetitive spacer sequences (26-72 bp).[4] The number of repeats per cluster varies from 2 to 250 and genomes can have 1-18 CRISPR loci. Since the spacer sequences were often found to match short genomic regions of viruses and plasmids, it was proposed that CRISPR spacers mediate immunity against infection by extrachromosomal elements. Other CRISPR-associated genes (*Cas* genes) encode the CAS proteins that add new spacer-repeat pairs, process the CRISPR transcript and cleave the recognized foreign DNA. Experimental evidence in support of this was recently provided in two bacterial systems. However, in archaea the CRISPR system appears to be much more complex than in bacteria, and it has been suggested that it may play a more regulatory role in archaea. In subsequent years similar CRISPR arrays were found in *Mycobacterium tuberculosis*,[5] *Haloferax mediterranei*[6] *Methanocaldococcus jannaschii*[7] *Thermotoga maritime*[8] and other bacteria and archaea. The accumulation of sequenced microbial genomes allowed genome-wide computational searches for CRISPRs and the most recent computational analyses revealed that CRISPRs are found in ~40% of bacterial and ~90% of archaeal genomes sequenced to date.[9-11]

STRUCTURAL FEATURES OF CRISPR SYSTEMS

CRISPR arrays and *Cas* genes (which together forming the "CRISPR system") vary greatly among microbial species. The direct repeat sequences frequently diverge between species[12,13] and an extreme sequence divergence is also observed in the CAS genes.[14] The size of the repeat can vary between 24bp and 47bp, with spacer sizes of 26-72bp.[11] The number of repeats per array can vary from two to 249 in *Verminephrobacter eiseniae*[11] and while many genomes contain a single CRISPR locus, the number of loci in *M. jannaschii* reaches 18.[7] Finally, while in some CRISPR systems only six or fewer CAS genes were identified, others involve more than 20.[15] Despite this great diversity most CRISPR systems have some conserved characteristics.

Repeats

In a single array, repeats are almost always identical, with respect to size and sequence.[12] Despite being divergent between species, repeats can be clustered based on sequence similarity into at least 12 major groups.[10] Some of the larger groups contain a short (5bp-7bp) palindrome, and hence the word "palindromic" in the CRISPR acronym.[12] These palindromes were inferred as contributing to an RNA stem-loop secondary structure of the repeat,[10] a hypothesis supported both by compensatory mutations existing in the repeats to maintain the stem structure, and by observations that the repeat-spacer array is transcribed into RNA[10,16-18] For other repeat groups evidence for RNA secondary structures is lacking. Apart from the structural feature, many repeats have a conserved 3' terminus of GAAA (C/G). Both the structural features and the conserved 3' motif were suggested to act as binding sites for one or more of the CRISPR associated proteins (Figure 1).[10]

Spacers

In any CRISPR system spacers are generally unique, with a few exceptions thought to result from segmental duplications.[11] Similarity searches of various CRISPRs consistently showed that many spacers frequently match (with high sequence identity) to phages and other extrachromosomal element.[14,16,19-21] Mojica and coworkers have studied 4500 spacers from 67 microbial strains; 88 (2%) of them had similarity with known sequences, with more than 50% of these similar to a sequence found within a known phage and 10% within a plasmid.[19] Comparable numbers were reported in a separate study where 2156 spacers were examined.[20] The observation that only 2% of all spacers match any known sequence presumably reflect the general under-sampling of phage sequence space and is in agreement with recent estimates of huge untapped

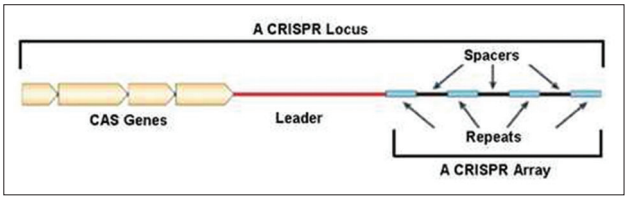

Figure 1: Simplified diagram of a clustered regularly interspaced short palindromic repeats (CRISPR) locus. The three major components of a CRISPR locus are shown: *cas* genes, a leader sequence, and a repeat-spacer array.

phage environmental diversity.[17] Indeed, in lactic acid bacteria such as *S. thermophilus*, for which more than a dozen phage genomes have been isolated and sequenced, ~40% of the spacers had a homologue, matching either phage (75%) or plasmid (20%) sequences[21] Spacers seem to be evenly distributed across the phage genomes and derive both from the sense (coding) and antisense (non-coding) orientations[4,16,19,20] although one report suggested a preference towards spacers derived from one strand of the phage.[21] Two recent studies have reported on a short motif present in phage genomes 1-2 nucleotides downstream to spacer-matching sequences.[22,23] This motif was hypothesized to be important for recognition, or cleavage, of phage sequences by the CRISPR system. The recognition motif can vary between CRISPR systems, being AGAA and GGNG for spacers found in CRISPR1 and CRISPR3 loci of *S. thermophilus*, respectively.

Leader

A sequence of up to 550bp is located 5' to most CRISPR loci, directly adjoining the first repeat.[12,16] This common sequence was denoted the "leader" and is usually ATrich.[12] Similar to the repeats, leaders lack an open reading frame and are generally not conserved between species; however, when several CRISPR loci are found in the same chromosome their leaders can be conserved.[6,17,18] When a new repeat-spacer unit is added to the CRISPR array, it almost always occurs between the leader and the previous unit, suggesting that the leader might function as a recognition sequence for the addition of new spacers.[4,20] The leader was also suggested to act as the promoter of the transcribed CRISPR array, as it is found directly upstream of the first repeat.[13,24]

CAS genes

Two recent studies have characterized the large set of gene families that are associated with CRISPR arrays[14,15] so in this review only the general features of these genes are discussed. CRISPR systems have been divided into 7 or 8 subtypes: each subtype contains 2-6 different subtype-specific CAS (CRISPR-associated) genes. In addition, six core CAS genes (*cas1-6*) are found associated with multiple subtypes, although the identity of *cas5* and *cas6* was not agreed upon.[14,15] The *cas1* gene (COG1518; TIGR00287) is especially noteworthy as it serves as a universal marker of the CRISPR system (found linked to all CRISPR systems except for that of *Pyrococcus abyssii*). Additional genes that are more loosely associated with CRISPRs, such a members of the Repeat Associated Mysterious Protein (RAMP)[15,17] superfamily that occur only in genomes that contain CRISPR systems, but not necessarily nearby the CRISPR, were also characterized. Specific functional domains identified in Cas proteins include endonuclease and exonuclease domains, helicases, RNA- and DNA- binding domains, and domains involved in transcription regulation.[11,14,15,25]

CRISPRS MECHANISM

An important clue came from the observation of bacterial mutants that have acquired phage resistance. Among them were some that have added a new repeat-spacer pair at the leader end of the array. Each of these new spacer sequences matches some section of the infecting phage genome (called the *protospacer*). Since those added spacers are necessary for the newly acquired phage resistance, the CRISPR system is a blatant smoking gun. As noted in one of the 2010 reviews: the mechanism by which CRISPR[26] provides resistance against foreign genetic elements is not fully characterized. Some mechanisms have been ruled out: CRISPR defense does not block phage adsorption or DNA injection, does not involve a restriction-modification system, and is not an abortive infection mechanism. Most likely, the degradation of the targeted nucleic acid by a CRISPR endonuclease is the key (Figure 2).

Here are some of the known details. The entire CRISPR array is transcribed constitutively as a single, long RNA that is then cut at a specific site in each repeat to yield the mature CRISPR RNAs (crRNAs). Each crRNA contains one entire spacer sequence plus recognizable "handles" provided by short 3' and 5' flanking regions derived from the adjacent repeats. The precise cutting is done by a complex of the CAS proteins called Cascade. Some of these proteins remain bound to the crRNAs to form the active defense agents. The spacer sequence provided by the crRNA is thought to recognize and guide the complex to the specific target sequence; then one (or more) of the proteins with nuclease activity cuts the invading nucleic acid. Since all the spacers are constitutively transcribed and processed into active defense agents, the host is continuously on guard against all the corresponding phages and plasmids.

When this story first came to light, piece by piece, there were speculations that the CRISPR mechanism might be analogous to RNA interference in eukaryotes. Indeed, there are analogous steps involved in the formation of the active defense complexes in both systems. Spacers complementary to either the coding or non-coding strand of phage λ confer resistance. Other experiments are demonstrating this used the CRISPR locus of a clinical isolate of Staphylococcus

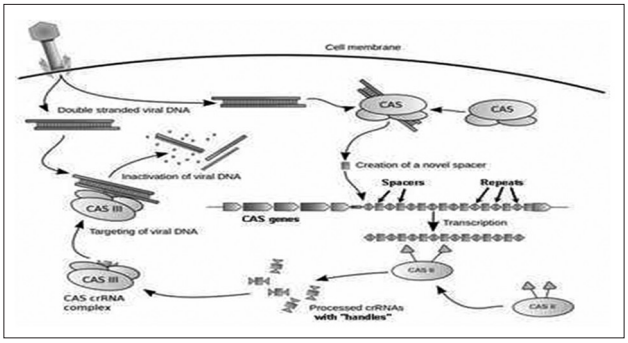

Figure 2: The clustered regularly interspaced short palindromic repeats defense mechanism.

epidermidis and a plasmid with an intron-containing gene. Here, the researchers experimentally introduced spacers that matched either the plasmid's gene sequence (including the intron) or the corresponding spliced mRNA sequence (lacking the intron). Only spacers against the intron-containing DNA sequence were effective.

The biological function of the CRISPR-cas system

It is truly a testament to post-genomic era research that the essence of the CRISPR-cas system was first discovered purely by computational sequence analysis and that the hypotheses generated through these efforts only later received remarkable experimental support. However, the crucial observation was that some were clearly derived from extrachromosomal DNA elements.[19-21] Mojica and colleagues found that 88 out of 4500 spacers from a broad range of bacteria and archaea matched to known sequences, with most being similar to bacteriophage and plasmids.[20] Remarkably, species containing identified spacer elements were immune to the corresponding foreign invaders or had no prophage remnants as evidence of prior infections. In contrast, closely related CRISPR-negative species were susceptible. In this model, to acquire resistance, new spacer information must be incorporated into the CRISPR locus. One source of that information might be the elements themselves, leading to the notion that the content of the locus can also serve as a record of infections from which a host had recovered. The distal end of the cluster contains "older" spacers, those that are shared among strains.[6,19,27] "Newer," strain specific spacers accumulate next to the leader sequence at the proximal end of the cluster. Clusters that lack a leader sequence do not appear to incorporate new spacers, suggesting that they are inactive remnants[16] a role for the leader sequence in cluster evolution, adaptation, or function, and points to an orchestrated mechanism of polarized cluster growth. In addition to increases in cluster content, spacers also appeared to be lost by internal deletions of one or more repeat units.

During spacer acquisition, sequence elements from invading nucleic acids become incorporated at the leader-proximal end of the CRISPR locus. In the processing stage, the locus is transcribed and processed into mature crRNA/psiRNAs containing an 8nt repeat tag and a single spacer unit. During the effector stage, the mature crRNAs in complex with associated cas proteins leads to degradation of complementary nucleic acids. In the CRISPR system, this is accomplished by pirating nucleic acids from incoming pathogens and incorporating them into a programmable silencing locus. This initial step involves cas7 and, likely, cas1. Potential spacer sequences are selected by the presence of a short proto-spacer adjacent motif, PAM. Precisely how the CRISPR system distinguishes foreign from domestic nucleic acids is unclear, but it seems unlikely to depend upon the selective presence of protospacer adjacent motifs (PAMs) in invader versus host genomes (Figure 3).

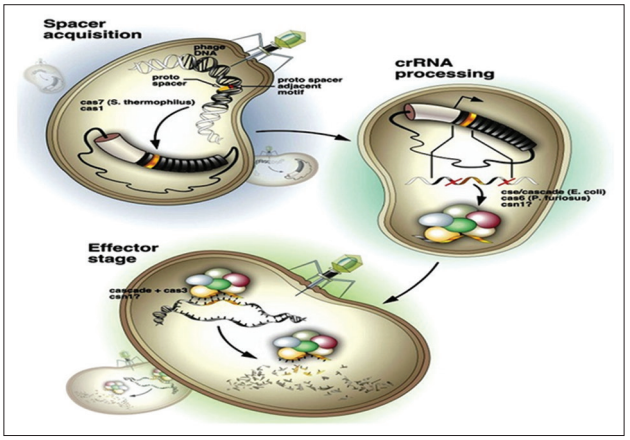

Figure 3: An overall model of clustered regularly interspaced short palindromic repeats/cas activity.

Possible roles for cas1 in CRISPR immunity for invasive DNA, chopped and in the CRISPR

To maintain genetic identity, bacteria must prevent the acquisition of foreign genetic material often carried by phages and conjugative plasmids. Clustered, regularly interspaced, short palindromic repeats (CRISPR) systems have been recently characterized as a programmable barrier to phage infection[4] and conjugation.[28] CRISPR interference is assisted by a set of CRISPR associated (Cas) proteins that are encoded by the *cas* genes usually found immediately adjacent to the repeats. Cas proteins can be classified into 45 different types, but their precise biochemical functions are largely unknown.[15] Only one protein, Cas1, has orthologs in all CRISPR loci. CRISPR loci include a set of repeats (white boxes) separated by spacers (colored boxes) with sequences that match phage and plasmid DNA that often invade bacteria. These clusters are preceded by a "leader" sequence and are usually flanked by *cas* genes. When a phage that does not carry a sequence identical to any CRISPR spacers infects bacteria, most cells succumb to the infection. At low frequency, however, a fragment of the phage DNA can become incorporated as a new repeat-spacer unit at one end of the CRISPR locus (top panel). Cas1 is a likely candidate in the adaptation process,[29] and demonstrate DNase activity of Cas1 consistent with this possibility. The new spacer sequence is then included among the crRNAs generated by the CRISPR locus, leading to immunity against subsequent infections by the phage (bottom panel) (Figure 4).

CRISPR-based adaptive immune systems

Small RNA-based defense systems that provide adaptive, heritable immunity against viruses, plasmids, and other mobile genetic elements have recently been discovered in archaea and bacteria. The RNA and protein components of these immune systems arise from the CRISPR (clustered regularly interspaced short palindromic repeat) and Cas (CRISPR-associated) genes, respectively.[30] The CRISPR

Cas pathway functions in three phases are an adaptation of CRISPRs to invaders, crRNA biogenesis, and invader silencing. The most organisms have only some of these six genes, and only cas1 and cas2 appear to be universal. In a given organism, the core cas genes are supplemented by one or more of the nine sets of subtype-specific cas genes.

The CRISPR Cas invader defense pathway

In the adaptation phase, a short fragment of foreign DNA (protospacer) is acquired from the invader and integrated into the host CRISPR locus adjacent to the leader. PAMs are found near invader sequences selected for CRISPR integration. The CRISPR locus consists of short direct repeat sequences (black) that separate similarly sized, invader derived sequences (multiple colors). In the biogenesis phase of the pathway, CRISPR locus transcripts are processed to release individual mature crRNAs (each targeting a different sequence). Mature crRNAs typically retain some of the repeat sequence, which is thought to provide a recognizable signature of the crRNAs. In the silencing phase, crRNA Cas protein effector complexes recognize foreign DNA or RNA through base pairing of the crRNA. The Cmr and Csn systems affect cleavage of target RNA and DNA, respectively. PAMs provide important auxiliary signals for the recognition of invaders for some DNA-targeting systems (Figure 5).[30]

The combinations of Cas proteins create diverse CRISPR Cas systems

Cas1-6 are core Cas proteins found in many and diverse organisms. In addition, there are eight primary modules of subtype-specific Cas proteins (consisting of two to six proteins each), and the auxiliary Cmr module. A typical CRISPR Cas systemis composed of the nearly universal Cas1 and Cas2 proteins (yellow), a specific combination of the other core Cas proteins (green) and a set of subtype-specific Cas proteins (blue). A given organism may possess more than one CRISPR Cas system, and may also have the Cmr module (purple) (Figure 6).[30]

Self-targeting by CRISPR: gene regulation or autoimmunity

The prokaryotic immune system known as CRISPR is based on small RNAs ("spacers") that restrict phage and plasmid infection. It has been hypothesized that CRISPRs can also regulate self-gene expression by utilizing spacers that target self-genes. By analyzing CRISPRs from 330 organisms, we found that one in every 250 spacers is self-targeting, and that such self-targeting occurs in 18% of all CRISPR-bearing organisms. However, complete lack of conservation across species, combined with abundance of degraded repeats near self-targeting spacers, suggests that self-targeting is a form of autoimmunity rather than a regulatory mechanism.[31]

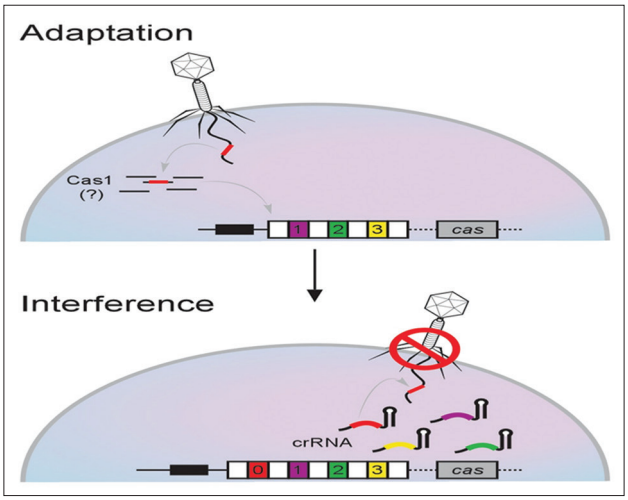

Figure 4: Possible roles for Cas1 in clustered regularly interspaced short palindromic repeats immunity.

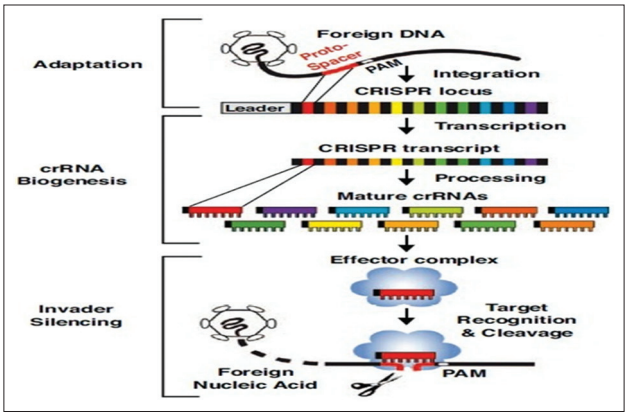

Figure 5: Overview of the clustered regularly interspaced short palindromic repeats Cas invader defense pathway.

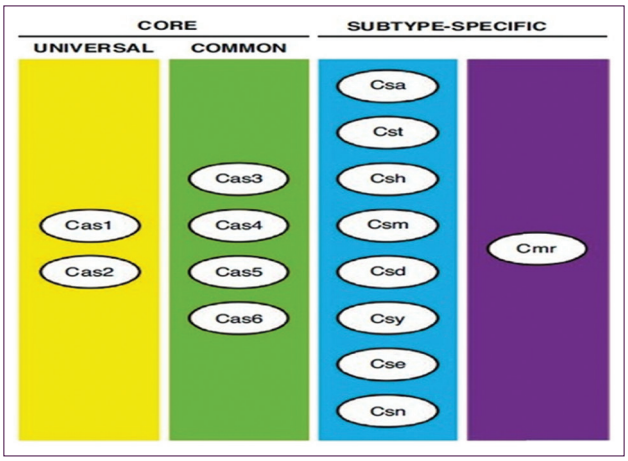

Figure 6: Combinations of Cas proteins create diverse CRISPR Cas systems.

Self-targeting CRISPR spacers

To identify potential self-targeting spacers, 23550 spacers from 330 CRISPR-encoding organisms were scanned for

an exact full match between the spacer and a portion of the endogenous genomic sequence that is not part of a CRISPR array (termed target, or self proto spacer). Our results reveal that 100 of 23550 spacers (0.4%) are self-targeting. However, encoding a self-targeting spacer is not a rare phenomenon: 59 of 330 (18%) CRISPR-encoding organisms possess at least one array with at least one self-targeting spacer. These spacers are widely distributed over diverse phylogenetic lineages and are dispersed throughout different arrays in each organism.[31]

CRISPR system elements involved in the recognition and integration of new spacers

The conservation of the orientation of spacers in the CRISPR loci with respect to the PAM suggests the recognition of some sequence in the integration site that, given the preference for the leader proximal end of the array, would involve nucleotides in that region, including the CRISPR unit. Assuming a common mechanism for uptake of new spacers, the detection of potential proto-spacer carriers should be unrelated to any peculiarity of the genetic element. The only feature shared by the putative spacer donors identified to date is that all are DNA molecules, which at some stage will be present in double-stranded form in the receptor cell. How the CRISPR machinery recognizes, such invaders is an essential question to be elucidated. Related to this, the fact that proto-spacers are found, on the basis of their associated PAM locations, in either the sense or antisense strand excludes a recognition of the spacer precursors on transcript RNA molecules, in support of dsDNA as the donor. This is better understood when considering pairs of proto-spacers with overlapping complementary sequences just one matching the transcript. For new CRISPR formation, ssRNA could be the precursor of a double-stranded donor molecule. This would imply either indiscriminate duplication of the foreign RNA (rather anti-economical for the cell) or, alternatively, that a signal different from the PAM be recognized in those to be duplicated. In addition, reverse transcription (RT) would be required for generating the spacer DNA. However, only a few CRISPR-harbouring strains have putative CRISPR-linked RT genes (Figure 7).[32]

PAMs determine the spacer orientation. Spacers 1 and 2 derive from oppositely oriented sequences with respect to each other but in the same orientation with respect to the PAM. The recognition of the same motif sequence requires the availability of both strands.

THE CRISPR/CAS REGULATORY SYSTEM

One of the basic postulates of evolutionary theory is that functional elements undergo purifying selection, leading to their conservation across different organisms. Returning to the superficial analogy with eukaryotic RNAi, miRNAs are among the most highly conserved non-coding elements in mammalian genomes. Hence, an essential requirement for

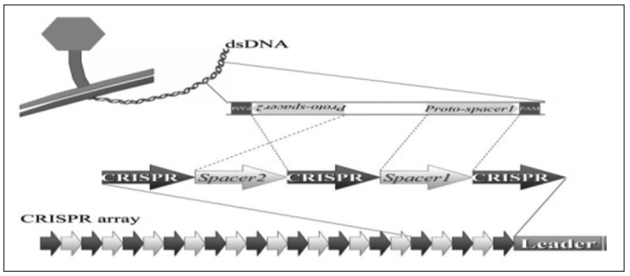

Figure 7: Acquisition of new spacers.

CRISPR to function as an established regulatory system is the evolutionary conservation of the self-targeting spacers across several species. Considering the possibility that CRISPR regulation might occur via partial base pairing (as in eukaryotic miRNAs or in the RAMP CRISPR-associated module). Once again, both partial and fully matching endogenous spacers showed no signs of conservation - in other words, they were present in only one organism. The each pair of self-targeting spacer and target exists only in one organism. The initial insertion of a self-targeting spacer conferred an evolutionary advantage to the organism and had it acquired a functional role in gene regulation, purifying selection would have led to its perpetuation.[31]

PHAGE EVADE MECHANISM

The bacterial CRISPR defense is very finicky and will cleave foreign DNA only if it contains a sequence that exactly matches a spacer. Thus, phage evasion is only a single base change away. That phage do indeed escape this way has been shown by challenging cultures with a phage for which they already have a spacer. One finds a small number of phages that are able to successfully multiply. A close look at the genome sequence of these phages typically reveals a single nucleotide change in the protospacer. (The archaeal system is less exacting, requiring more than one nucleotide change in that region for escape.). Although protospacers are found on both coding and non-coding strands throughout the phage genome, their locations are not entirely random. They are always just a few base pairs from a short motif that is recognized by the CRISPR system. Phages without that motif in their genome are immune. Another evasion trick used by viruses and other mobile elements is to insert themselves in one of the CAS genes or otherwise interfere with the operation of the CRISPR. This might be one of the factors that make it worthwhile for hosts to carry more than one CRISPR locus. (The current record is 18 loci, accounting for 1% of the genome in the archaeon *M. jannaschii*.) CRISPR-associated (Cas) proteins play an important role in the initial recognition of phage genetic material and incorporating these proto-spacers in the CRISPR array. Once incorporated, these spacers are then transcribed and used as templates to target homologous phage sequences within the bacterial cell, again mediated by Cas protein complexes. The bottom left of the figure illustrates the specificity of CRISPR-mediated resistance. While resistance to a specific

phage genotype can be acquired by incorporating a spacer derived from that genotype, point mutations in the phage (represented by the black squares) are sufficient to evade resistance. Hosts are only able to resist this mutant by incorporating a phage-derived spacer containing the new mutation. This could lead to an ongoing "arms race," with hosts incorporating more spacers in response to increasing phage mutation accumulation. The order in which these spacers are incorporated also provides a sequential record of past phage infections (Figure 8).[26,33,34]

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. 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 recombination 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 protospacer. 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, the 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 (Figure 9).[35]

CRISPRs differ from most other repeats described herein that these small sequences are part of a complex genetic arrangement. This consists of an array of palindromic DRs of approximately 28-49 bp. Linked with each repeat are variable spacer sequences that are fragments of foreign DNA (phage or plasmid DNA), or in some cases, host DNA. An array of protein genes termed CRISPR-associated (cas) genes are also closely associated with the palindromic repeat/spacer units. CRISPRs function as regulatory complexes. Recently, there has been great interest in the genetic and molecular characteristics of CRISPRs and for several reasons. First, the CRISPR system can function as a bacterial and archeal immune system, whereby CRISPR defends the organism from invading viral or plasmid DNA.[36] In addition, the mechanism of action of CRISPR systems has similarities to eukaryotic piwi-interacting RNAs (piRNA) mechanism of RNA-based immune system that inhibits mobile elements in germ line cells.[33,37] Finally, this genetic element offers an example of a type of Lamarckian inheritance in prokaryotes.[38] The CRISPR DNA complex was first found in *Escherichia coli*,[4] although much of its characterization and functions have only been elucidated recently, approximately during the past 10 years. Here, we provide a short description of the molecular/genetics aspects of CRISPR functions as

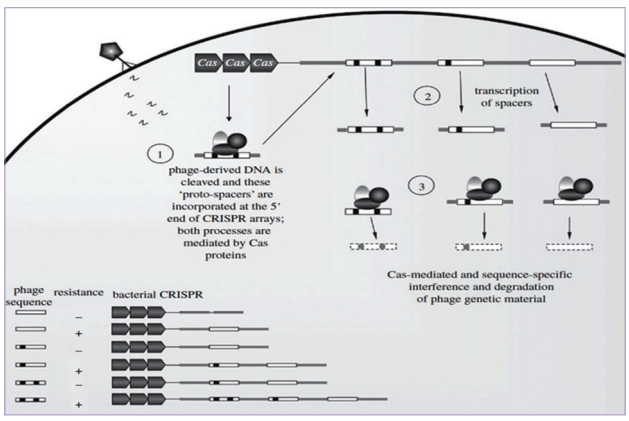

Figure 8: A schematic representation of CRISPR-mediated phage resistance.

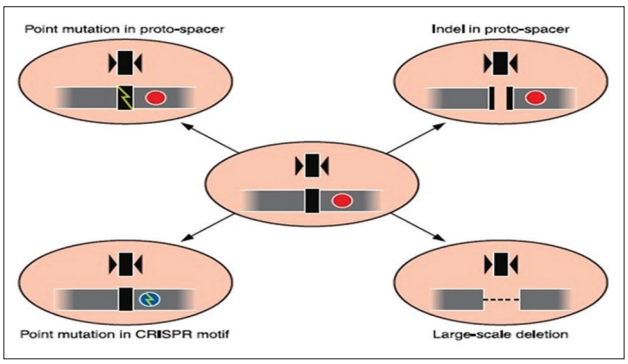

Figure 9: Various strategies to evade clustered regularly interspaced short palindromic repeats (CRISPR)-encoded immunity, including mutation in the proto spacer (top left), insertion/deletion in the proto-spacer (top right), mutation within the CRISPR motif (bottom left), and large-scale deletion (bottom 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.

they relate to immunity to invading or self-DNA. There are basically three stages in the molecular and genetic processes of CRISPR function. During the acquisition stage, CRISPRs can capture fragments of foreign DNA from virus or plasmid sequences when challenged with the foreign DNA.

A short segment (approximately 25-70 bp) of the foreign DNA, called a proto-spacer is inserted into the CRISPR locus of the host DNA between two palindromic repeat sequences. How the cell recognizes the short foreign DNA is unclear, but inserted foreign DNAs that have a small sequence (approximately a few nucleotides) adjacent to the spacer may be a recognition site. This small sequence is termed a proto-spacer-adjacent motif sequence.[39] Two Cas proteins may be involved in the acquisition process. Additional spacers are then added to form an array of spacer-palindromic sequence repeats. It is not known if the palindromic repeat sequences serve as Cas protein recognition sites for integration of DNA fragments into the CRISPR complex. In the second stage, the CRISPR complex is transcribed, and cas genes are transcribed and translated. In *E. coli*, the large precursor CRISPR transcript is processed by a ribonucleoprotein complex termed Cascade (CRISPR-associated complex for antiviral defense).

A Cas-specific endonuclease processes the RNA via cleavage at the base of the repeat stem loop sequence, and with additional trimming, the mature RNA is formed.[40] After processing, the Cascade complex retains RNA transcripts of foreign spacer DNA and stem-loop repeat sequences and bound Cas proteins. In the third stage, Cascade binds one strand of the target DNA via complementary base-pairing between spacer RNA and target DNA to form an RNA/DNA heteroduplex duplex. The target DNA strand is subsequently cleaved. Cas3 protein, which has endonuclease properties may be the major protein associated with target DNA inactivation in *E. coli*.[40] This molecular process that results in defense against invading DNA was shown to be present in organisms that include Streptococcus, Staphylococcus, and *E. coli* species. However, CRISPRs can display different roles in different microorganisms, and spacer DNA may consist of a fragment of a host protein gene. 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. CRISPR spacer content and hyper variability, 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 spoligotyping. 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 (Figure 10).[31]

CURRENT AND FUTURE APPLICATIONS

Current implementation of CRISPR Cas systems to develop phage resistance in dairy starter cultures has already shown that CRISPR can be leveraged industrially. Current analyses of CRISPR polymorphism in pathogenic species will determine how relevant these loci may be for epidemiological surveys, clinical analyses, and food safety. To develop the technique, the team used *Neisseria meningitides* bacteria and human pluripotent stem cells. *N. meningitides*, the bacteria responsible for causing meningitis in humans, possess specific CRISPR that produce the desired protein needed to cleave damaged DNA strands. CRISPR are loci with direct repeats in a genetic sequence that act as an immune system in prokaryotes. CRISPR associated proteins, Cas, defend the bacterial cell by splitting sections of intruding viral DNA. Cas9 is an enzyme capable of pairing with an RNA transcript to target specific DNA sequences in invading viruses and cut them up, thus preventing harm to the bacterial cell. The researchers have used this understanding to devise a method for directing Cas9 to cleave a specific damaged or mutated DNA site in human cells and then remove, repair or replace the sequence by an RNA-guided mechanism. Zhonggang Hou, the lead author on the study, said that his team used the CRISPR system in the bacteria to produce Cas9 and specifically cleave the desired DNA sites targeted for repair. "The technique is called gene repair by CRISPR. This is done by introducing a targeted DNA double strand break in the region of interest on the DNA," said Hou in an email. "This technique can potentially correct genetic mutations in the genome." The genomic engineering was performed on human pluripotent stem cells, which proliferate indefinitely and can differentiate into any cell type. By extracting the

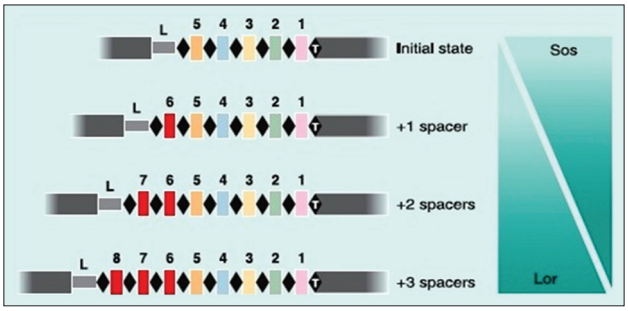

Figure 10: Iterative build-up of clustered regularly interspaced short palindromic repeats -encoded phage resistance via consecutive additions of novel spacers, which yields an increase of the level of resistance (Lor), as well as a reduction of the spectrum of sensitivity (Sos) to phages. Accordingly, variants that have iteratively acquired novel spacers have broader and deeper phage resistance spectra.

stem cells from skin or blood, they were reprogrammed to become specialized cells whose genome the researchers could manipulate. Fairly recent gene-correcting methods, such as zinc finger nucleases and transcription activator like effector nucleases, require engineered proteins that take several weeks or months to generate. The RNA coupled with Cas9 gene repair takes only 1-3 days to produce, offering a leap in current gene therapy and disease protocols.[41]

CONCLUSION

Overall, the dynamic nature of CRISPR loci is potentially valuable for typing and comparative analyses of strains and microbial populations. Given that some loci are relatively active while others bear lower levels of polymorphism, the potential of a given CRISPR locus for typing and epidemiological studies has to be assessed on a case-by-case basis. Since CRISPRs are widely distributed in *Bacteria* and *Archaea* and actively involved in an adaptive immune system against foreign genetic elements, as well as intrinsic chromosomal elements, they provide critical insights into the relationships between prokaryotes and their environments, notably the coevolution of host and viral genomes. Until now the CRISPR system has been heralded as an exceptional form of defense against foreign invaders. CRISPR Cas immune systems play a globally important biological role in host–parasite interactions and collectively shape the evolution and ecology of prokaryotes and viruses. The early studies have revealed that there is a diverse series of CRISPR Cas pathways that function through distinct components and mechanisms, which are dispersed throughout archaea and bacteria. Much of our still very limited knowledge has come from studies with a small set of model organisms that collectively do not encompass the known CRISPR–Cas modules, and further investigation in other organisms will help address this gap. In the near future, the concerted effort of numerous research groups is expected to provide answers to fundamental questions such as how novel protospacers are acquired from invaders and integrated into CRISPRs, what constitutes functional crRNAs and how they are generated, and how silencing is achieved for each of the CRISPR Cas pathways, and should illuminate the molecular mechanisms governing the astonishing CRISPR Cas-mediated prokaryotic immune pathways.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 2005;13(6):278-84.
2. Wommack KE, Colwell RR. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev. 2000;64(1):69-114.
3. Sturino JM, Klaenhammer TR. Engineered bacteriophage-defence systems in bioprocessing. Nat Rev Microbiol. 2006;4(5):395-404.
4. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315(5819):1709-12.
5. Hermans PW, van Soolingen D, Bik EM, de Haas PE, Dale JW, van Embden JD. Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. Infect Immun. 1991;59(8):2695-705.
6. Mojica FJ, Ferrer C, Juez G, Rodríguez-Valera F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. Mol Microbiol. 1995;17(1):85-93.
7. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science. 1996;273(5278):1058-73.
8. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, et al. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature. 1999;399(6734):323-9.
9. Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol. 2000;36(1):244-6.
10. Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. 2007;8(4):61.
11. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics. 2007;8:172.
12. Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol. 2002;43(6):1565-75.
13. Tang TH, Polacek N, Zywicki M, Huber H, Brugger K, Garrett R, et al. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus. Mol Microbiol. 2005;55(2):469-81.
14. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational 10 analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct. 2006;1:7.
15. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (CAS) protein families and multiple CRISPR/CAS subtypes exist in prokaryotic genomes. PLoS Comput Biol. 2005;1(6):60.
16. Lillestøl RK, Redder P, Garrett RA, Brügger K. A putative viral defence mechanism in archaeal cells. Archaea. 2006;2(1):59-72.
17. Edwards RA, Rohwer F. Viral metagenomics. Nat Rev Microbiol. 2005;3:504-10.
18. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, et al. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature. 1997;390(6658):364-70.
19. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol. 2005;60(2):174-82.
20. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake

of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 2005;151:653-63.
21. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151:2551-61.
22. Viswanathan P, Murphy K, Julien B, Garza AG, Kroos L. Regulation of dev, an operon that includes genes essential for Myxococcus xanthus development and CRISPR-associated genes and repeats. J Bacteriol. 2007;189(10):3738-50.
23. Edgar RC. PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics. 2007;8:18.
24. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, et al. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. Proc Natl Acad Sci U S A. 2002;99(11):7536-41.
25. Ebihara A, Yao M, Masui R, Tanaka I, Yokoyama S, Kuramitsu S. Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain. Protein Sci. 2006;15(6):1494-9.
26. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010;327(5962):167-70.
27. Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol. 2008;190(4):1401-12.
28. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008;322(5909):1843-5.
29. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure. 2009;17(6):904-12.
30. Terns MP, Terns RM. CRISPR-based adaptive immune systems. Curr Opin Microbiol. 2011;14:321-7.
31. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet. 2010;26(8):335-40.
32. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct. 2006;1:7.
33. Karginov FV, Hannon GJ. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell. 2010 15;37:7-19.
34. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 2010;11:181-90.
35. Barrangou R, Horvath P. The CRISPR system protects microbes against phages, Plasmids. Microbe. 2009;4:5.
36. Al-Attar S, Westra ER, van der Oost J, Brouns SJ. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biol Chem. 2011;392(4):277-89.
37. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 2010;11(3):181-90.
38. Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 2008;36(21):6688-719.
39. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology. 2009;155:733-40.
40. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008;321(5891):960-4.
41. Shindgikar P. Meningitis bacterial defence mechanism finds applications in future of regenerative medicine. October 1, 2013.

Cite this article as: Mahmud S, Ahmed J, Aziz MA, Hasan MR, Shaon SM, Bhuiyan MNA, Rahman F, Rakib HH, Islam MS. Clustered regularly interspaced short palindromic repeats cas systems: a comprehensive review. Int J Basic Clin Pharmacol 2015;4:613-22.