Title

Diversity, evolution, and therapeutic applications of small RNAs in prokaryotic and eukaryotic immune systems

Journal

Physics of Life Reviews, 11(1)

ISSN

1571-0645

Authors

Cooper, Edwin L
Overstreet, Nicola

Publication Date

2014-03-01

DOI

10.1016/j.plrev.2013.11.002

Peer reviewed

Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 11 (2014) 113–134

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Review

Diversity, evolution, and therapeutic applications of small RNAs in prokaryotic and eukaryotic immune systems

Edwin L. Cooper *, Nicola Overstreet

Laboratory of Comparative Immunology, Department of Neurobiology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095-1763, United States

Received 14 September 2013; accepted 5 November 2013

Available online 8 November 2013

Communicated by M. Frank-Kamenetskii

Abstract

Recent evidence supports that prokaryotes exhibit adaptive immunity in the form of CRISPR (Clustered Regularly Interspersed Short Palindromic Repeats) and Cas (CRISPR associated proteins). The CRISPR–Cas system confers resistance to exogenous genetic elements such as phages and plasmids by allowing for the recognition and silencing of these genetic elements. Moreover, CRISPR–Cas serves as a memory of past exposures. This suggests that the evolution of the immune system has counterparts among the prokaryotes, not exclusively among eukaryotes. Mathematical models have been proposed which simulate the evolutionary patterns of CRISPR, however large gaps in our understanding of CRISPR–Cas function and evolution still exist. The CRISPR–Cas system is analogous to small RNAs involved in resistance mechanisms throughout the tree of life, and a deeper understanding of the evolution of small RNA pathways is necessary before the relationship between these convergent systems is to be determined. Presented in this review are novel RNAi therapies based on CRISPR–Cas analogs and the potential for future therapies based on CRISPR–Cas system components.

Contents

* Corresponding author.

1571-0645/$ – see front matter
http://dx.doi.org/10.1016/j.plrev.2013.11.002

1. Introduction

1.1. Prokaryotes and eukaryotic immune systems

Living organisms have evolved into three domains: Bacteria, Archaea, and Eukaryota. Two of these three domains – the Bacteria and the Archaea – consist of generally unicellular organisms that lack a cell nucleus and other membrane-bound organelles. These organisms are known as prokaryotes. On the other hand, the eukaryotes of domain Eukaryota contain DNA and metabolic sites enclosed within membranes.

All large, complex organisms, such as plants and animals, are eukaryotes, and can sport varied and multifaceted immune systems. Traditionally, the immune system can be divided into two main systems: the innate and adaptive. The distinguishing characteristics of the two systems contrast greatly. The innate immune system is natural, non-specific, non-anticipatory, non-clonal, germ line, and does not retain memory of past exposures. This type of immune system is found throughout the animal and plant kingdoms. The adaptive immune system is acquired, specific, anticipatory, clonal, somatic, and allows for memory of past exposures [27]. Differences in animal evolution and morphology have led to associations with either or both of these systems. Primarily, vertebrates employ some combination of the two systems, while invertebrates tend only to have access to the innate immune system [25,27]. Invertebrates do not possess a T-cell-like recognition system, whereas vertebrates have a highly specialized recognition system against non-self components [28]. Often when both systems are present they are highly interrelated [26,27]. Innate mechanisms can be viewed as a part of a continuum of adaptive immune system responses [29].

Until recently, eukaryotes appeared to have been the sole winners of the evolutionary lottery for defense against genomic parasites, however this is undoubtedly not the case. The restriction–modification system (RM system) is used by bacteria as protection from foreign DNA. RM systems inhibit the growth of viruses through sequence-specific restriction enzymes. Around 25% of known bacteria possess RM systems [137]. Prokaryotes not only have innate immune defenses such as restriction–modification systems and the alteration of bacteriophage receptors, but also have working adaptive immune systems.

1.2. CRISPR–Cas overview

CRISPR–Cas systems are the RNA-based immune systems of prokaryotes. CRISPR–Cas can establish resistance against assaults by plasmids or viruses. The system captures short DNA sequences of the DNA of the genetic attacker, incorporates these sequences within the CRISPR loci in the host genomes, and produces small RNAs (CRISPR (cr)RNAs) that direct Cas proteins to recognize and either silence the expression of or destroy the invasive genetic material. Almost all archaea and around half of bacteria have CRISPR–Cas systems [122].

1.3. Discovery of CRISPR–Cas

In 1987, Ishino and colleagues created the first characterization of CRISPR while studying the alkaline phosphatase isozyme conversion gene *iap* in *Escherichia coli*. They described fourteen repeats of twenty-nine base pairs that were intermingled with around thirty spacer sequences that did not repeat [76]. Similar CRISPR arrangements were discovered shortly after this initial study. For example, in 1991 Hermans and colleagues found that *Mycobacterium tuberculosis* has repeats of an IS3-like element [72].

In 1995, Mojica and colleagues noticed long segments of short repeats in *Archaea Haloferax mediterranei* and *Haloferax volcanii* [103]. And five years later, after a computational analysis of known microbial genomes, the group characterized the clustered repeated short elements they had seen, calling them Short Regularly Spaced Repeats (SRSRs) [104]. In 2005, Mojica and co-workers were also among three research groups – also including the teams of Bolotin and Purcel – who first proposed that the spacer sequences are made up of generic material derived from plasmid or phage DNA, and that CRISPRs function as a form of immunity against these attackers [105].

More recently, Barrangou and colleagues confirmed this concept, by showing that CRISPR protects cells from destruction by phages, and modifications to CRISPR sequences changes phage resistance. Additionally, the specificity of the resistance was demonstrated to be directly related to the sequence of the spacers, which are derived from DNA strands of phage open reading frames [7]. This was further supported by Deveau and colleagues, through studies of phage responses in *Streptococcus thermophilus* which demonstrated that CRISPR confers phage resistance [36].

2. CRISPR–Cas elements

2.1. The CRISPR locus

The characteristic element of CRISPR–Cas systems is CRISPR loci. The locus contains a series of identical repeat sequences that are around 25 to 40 nucleotides long. In Archaea, the locus can contain up to 100 of these repeat units [52]. There are twelve different families of CRISPR repeats, and their family assignment is based on predicted secondary structure, size, and sequence similarity. The repeats are interspersed with spacers, of a similar size, around 35 to 45 nucleotides long. Spacer sequences are derived from invading genetic material, and are incorporated into the CRISPR locus. The number of spacers found in any CRISPR locus varies widely, and could be anywhere from 2 spacers to several hundred spacers, depending on the species of organism and the CRISPR–Cas variant. The upstream end of the CRISPR locus contains a leader segment where the promoters for CRISPR locus transcription are located. The leader sequence is also implicated in the adaption phase of CRISPR functioning [122].

2.2. Cas proteins

Cas proteins are an integral part of prokaryotic immune systems. More than 45 families of Cas genes have been discovered, but an organism generally only possesses between 4 and 20 Cas genes [37]. Cas1 and Cas2 are the only

Table 1
Summary of Cas core protein qualities.

Cas protein	Important characteristics	Known/suspected functions	References
Cas1	• Best conserved Cas gene across both Bacteria and Archaea • High isoelectric point • Metal-dependent activity	• Endonuclease activity • Suspected to process invader DNA into fragments	[135]
Cas2	• Small nuclease (80–120 amino acids) • Characteristic structural motifs including a polar amino acid following an N-terminal beta sheet • Second-best conserved Cas gene	• Metal-dependent endoribonuclease activity	[37,115]
Cas3	• Contains motif found in helicases (DEAD/DEAH motif) • Often attached to a nuclease	• Helicase, possible role in all stages of CRISPR–Cas system, unknown function • Predicted involvement in the silencing phase of defense	[37]
Cas4	• Cysteine residues at the C terminus • Similar structure to RecB exonuclease family	• Exonuclease, unknown function	[37]
Cas5	• Around 250 animo acids long • Possess a moderately conserved N-terminal, but not C-terminal, domain across CRISPR–Cas variants	• RNA-binding protein, unknown function • Predicted involvement in the Csn complex of crRNA processing	[37]
Cas6	• At times is the furthest of the Cas genes from the CRISPR locus • Has features of RNA-binding proteins – two ferredoxin-like folds	• Processes the CRISPR precursor crRNAs to release invader-specific crRNAs	[16,37]

Cas genes that are universal to all CRISPR–Cas systems, although Cas3 through Cas6 are also very widespread [63,122]. The core Cas genes are known to be adjacent to the CRISPR locus, and a summary of their known characteristics and functions can be found in Table 1.

The core Cas genes (Cas1–Cas6) are not the only Cas genes at work in any CRISPR system. Nine subsets of Cas genes supplement the action of the core genes. Each subset is composed of two to six genes. The names of the first eight subsets are based on an organism in which the subset is the only addition to the core genes. Their names are as follows: Aeropyrum (csa), Desulfovibrio (csd), Escherichia (cse), Haloarcula (csh), Mycobacterium (csm), Neisseria (csn), Thermotoga (cst), and Yersinia (csy). The ninth subset, RAMP (cmr), is not found in any organisms without the presence of an additional subset [122]. The Csa, Cst, Csh, and Csm subtype systems are common in Archaea, and all of the Cas systems are found in bacteria except for Csa. The broad range of variation in which Cas genes are present in a given CRISPR–Cas system suggests that a variety of pathways are used to carry out CRISPR–Cas immune functions, although the differences in the utility of these systems as a result of their unique pathways is unknown [64].

3. CRISPR–Cas mechanisms

3.1. Integration of spacers: adaption

The addition of new spacer sequences to the CRISPR locus – adaption – is the first stage in CRISPR immune systems. A protospacer, a short fragment of foreign DNA, is taken from an invading genetic parasite when the sequence

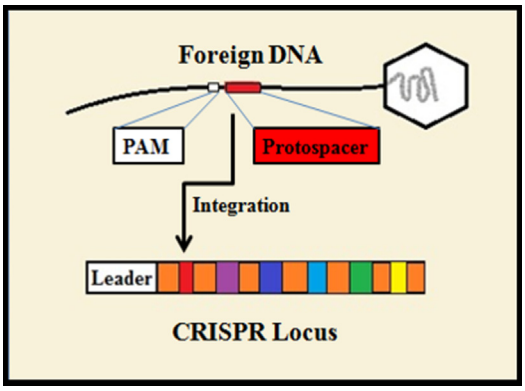

Fig. 1. Overview of Adaption – protospacer integration into the CRISPR locus. The CRISPR locus, composed of repeat sequences (orange) which are intermingled with spacers, integrates a new spacer from a foreign protospacer next to a PAM (protospacer adjacent motif) upstream (in type I and III systems) in the locus next to the leader sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is situated in the foreign DNA nearby to a PAM sequence (protospacer adjacent motif). PAMs are crucial for the recognition of invading genetic elements. The PAM motif is a short segment around 2 to 4 base pairs upstream from the protospacer. The PAM sequence that the CRISPR–Cas system targets varies based on the type of CRISPR–Cas system present in the cell [106]. Cas1 is implicated in the cleavage of invader DNA to form the fragment that will be incorporated into the CRISPR locus; however, Cas1 cleaves dsRNA (double-stranded RNA) without the presence of PAMs, so the mechanism of protospacer cleavage is still unknown. It has been suggested that Cas1 couples with other Cas subset and non-Cas proteins to perform this function [135].

Analysis of spacer units in *Sulfolobus solfataricus* revealed that spacers are added upstream in the CRISPR locus after the leader sequence. Because adaption only occurs upstream, the CRISPR locus is composed of spacers that represent a roughly chronological record of past invaders. The integration of the spacers can only occur when certain Cas proteins are attached to the leader [91]. A summary of the adaption stage can be seen in Fig. 1.

The CRISPR locus is dynamic not only because of the addition of new spacers, but also because of the exchange or loss of spacer sequences. Downstream locations on the CRISPR locus cannot be changed via the sole addition of an external protospacer, and although it is uncertain exactly how spacers can be substituted for one another, some studies suggest that this occurs due to internal recombination events [119].

3.2. *crRNA biogenesis*

The second step in the CRISPR defense pathway involves the creation of crRNA (CRISPR RNA) from the CRISPR locus. The CRISPR locus is transcribed beginning in the leader sequence, which houses four promoters [122]. Although generally the transcription of the CRISPR locus occurs continuously at low levels, in *Thermus thermophilus*, the presence of foreign DNA can cause increased levels of expression of CRISPR–Cas elements [1]. The initial transcription of the CRISPR locus produces a long crRNA precursor that can contain an RNA transcript of the entire locus [12,101]. The pre-crRNA is processed into shorter fragments through cuts in the sequence which corresponds to the repeat sequence between spacers [122]. The result of these cuts is finished crRNA products (see Fig. 2).

The pre-crRNA is processed through the endonuclease activity of either Cas6 or cse3. In *P. furiosus*, the cleavage of the pre-crRNA is performed by Cas6 [16]. However, in *E. Coli*, the cascade complex, a protein complex composed of cse1, cse2, cse3, cse4, and cse5 (also known as CasA, B, C, D, and E), cleaves the pre-crRNA inside of the transcript resultant from the repeat sequence [1]. Cse3 cleaves at the 5′ end of the new crRNA, and the 3′ end is then processed via an unknown mechanism. The cascade complex retains the crRNA once it has been fully processed and thus is implicated in the next phase of the CRISPR–Cas system: invader silencing [136].

Surprisingly, Cas6 does not have any known homologs in the cascade complex subtype which uses cse3 to cleave the transcript, even though their function in the cleavage of the pre-crRNA is highly analogous. Cse3 and Cas6 do however contain similar structures and RNA-recognition motifs, even though they lack sequence homology [16,63,126].

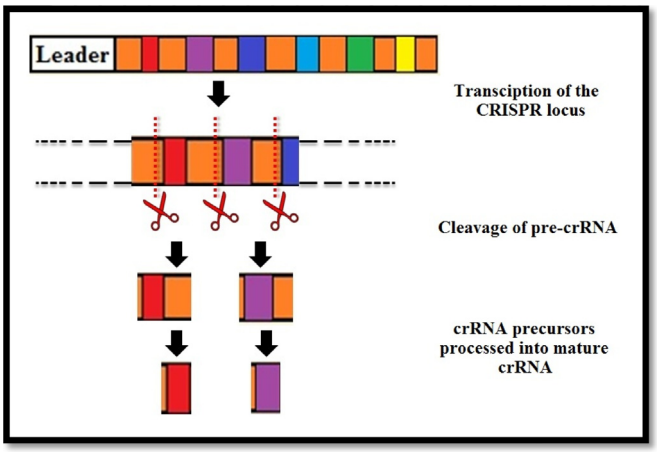

Fig. 2. The progression of crRNA biogenesis, which includes the initial transcription of the CRISPR locus and processing of crRNA precursors.

In bacteria containing Csn subtype (Type II) CRISPR systems, an RNA transcribed from nearby to the CRISPR locus anneals to the repeat sequence of the pre-crRNA. The pre-crRNA is then processed by RNase III and possibly Csn1 (Cas5) [35].

The crRNA produced contains an RNA transcript of a complete spacer sequence and a partial leading repeat. In many organisms, the repeat portion that remains in the mature crRNA is an 8 nucleotide-long tag at the 5′ end of the crRNA [16,122]. However, unusual crRNA transcripts have been noted in antisense products in some species including *S. acidocaldarius* [79,91].

3.3. Invader silencing

For most CRISPR–Cas systems, foreign DNA is the primary target [91]. After crRNA has been fully processed, the small RNA is primed for defense against genomic parasites through interference with the expression of the foreign genes or the destruction of foreign elements.

In order to effectively interfere with the expression of foreign DNA, the crRNA associates with Cas proteins, forming an effector complex. In *E. coli*, it is believed that this effector complex is formed when cascade recruits Cas3 [136]. The complex recognizes the foreign DNA through base-paring with the crRNA. A few mismatches between the crRNA and the target protospacer DNA do not inhibit interference [59]. Although it was proposed that the PAM motif recognized by the CRISPR–Cas system, which was initially used to recognize protospacers for spacer insertion, is the marker which the complex uses to identify foreign DNA, it has been shown in *E. coli* that only 6 of the 7 nucleotides of the PAM need to be matched to the crRNA for the effector complex to cleave the DNA [117]. The complex might rely on Csn1 (Cas5) in order to cleave the foreign DNA after it is recognized [51].

In the case of the invasion of RNA, another complex, the Cmr complex, cleaves the foreign RNA. Seven different cmr proteins are incorporated into complex along with crRNA, and catalyze the cleavage of the foreign RNA about 14 nucleotides from their 3′ end [115]. Although the guiding crRNA has the potential to be cleaved during this process, often one crRNA can lead to the degradation of several foreign RNAs [142]. Around 30% of bacterial CRISPR systems and 70% of archaeal systems possess the Cmr complex [119].

For other CRISPR–Cas systems, the mechanism of silencing is still unknown. It is uncertain as to whether the silencing of exogenous elements involves cleavage or other methods. However, it has been predicted that Cas3 proteins are involved in silencing in many subsystems. Cas3 can associate with COG2254 family proteins. One such COG2254 protein, the Cas gene SSO200, is an endonuclease that shows no preference for the cleavage of dsRNA over dsDNA, and may partner with Cas3 protein SSO1999 in order to silence foreign invaders [64,69].

3.4. Memory of past exposures

In a modified *E. coli* CRISPR–Cas system, spacers serve as a memory of past exposures, in a way that surpasses their use as a genetic history of previous invasions. When point mutations occur in the PAM or protospacer segment of

a genetic invader to which the CRISPR system has already been exposed, crRNA silencing is inhibited. However, the old CRISPR spacers are not rendered useless. In a streamlined, genetically-altered CRISPR–Cas system in *E. coli*, it was found that when crRNA partially matches foreign DNA and silencing is impaired, the acquisition of protospacers is triggered, allowing the CRISPR–Cas system to quickly incorporate new spacers [32]. Cas1 and Cas2 are required for the adaption process, and thus are also necessary for the acquisition of new protospacers from old, but newly mutated, invaders [135]. It has been suggested that these elements, along with cascade, scan the DNA after a partial crRNA match and then select new protospacers. Because the impaired silencing of foreign genes causes increased protospacer acquisition, the modified CRISPR-Cas system was able to adapt to mutations in the genome of previously encountered foreign genomes, and prevent losing immunity to these invaders [32].

4. Evolution of CRISPR–Cas systems

4.1. CRISPR–Cas and Lamarckian evolution

CRISPR–Cas systems are an example of Lamarckian evolution. Lamarckian evolution is named after and based on the work of French naturalist Jean-Baptiste Lamarck. In Lamarckian evolution, variation arises in response to environmental stimuli and is directed towards an adaptation to that stimulus. Darwinian evolution, in contrast, was proposed by English naturalist Charles Darwin, and involves evolution dependent on random variation, which lacks a specific direction [84].

The criteria for Lamarckian evolution are as follows: genomic changes are caused by the environment, the changes induced are specific to relevant genes, and the changes produce a phenotype that acts as an adaptation to the initial environmental stimuli. The CRISPR-Cas system meets all three of these requirements. The integration of spacer elements derived from the environment of the cell act in a pathway that responds to this environment [83].

While CRISPR–Cas systems do meet all requirements for Lamarckian evolution, they are not exempt from Darwinian evolutionary pressures. The spacers undergo Darwinian selection, and there may be strong selection for spacers which are especially effective [45].

4.2. Co-evolving CRISPR–Cas elements

CRISPR–Cas systems involve three elements that have coevolved. These elements are the CRISPR leader sequence, Cas genes, and the CRISPR repeat sequence. Because the function of these three elements is so intimately intertwined in defense against genomic parasites, their evolution was likely concurrent. This is supported by the fact that different CRISPR–Cas system subtypes consistently include distinct leader sequences, subsystem Cas proteins, and repeat sequence types unique to the CRISPR–Cas subtype [86,119,122]. These elements evolved together.

However, the Cmr complex, involved in the silencing phase of CRISPR–Cas defense, seems to have evolved separately from other cas proteins, which suggests that the DNA-targeting portion of CRISPR–Cas evolved separately from the RNA-targeting portion, even though the two are functionally dependent on one another [98,119].

4.3. Diversification of CRISPR forms

The CRISPR locus is extremely diverse across different species of prokaryotes. There is great variation in the size of the locus, which can be as short as a single spacer sandwiched by two repeats or as long as 375 spacers [90,119]. Over time, the CRISPR locus can mutate by undergoing sequence deletions, duplications, or recombinations.

There are several examples of CRISPR locus sequence deletions. The general trend is that deletions occur more frequently towards the downstream end of the CRISPR locus. This region contains older spacers and repeats, since new spacers are acquired upstream adjacent to the leader sequence. It is likely that deletion in the middle or downstream sections of the CRISPR locus do not impact the immune defense of the cell as much as an upstream deletion would, because these older spacers are not as relevant to the current or recent invaders of the cell [37,91]. Van Embden and company discovered that the CRISPR locus, also known as the DR (direct repeat) region, in strains of *M. tuberculosis* showed multiple deletions – but no genetic shuffling – between strains [127].

The duplication of spacers and repeats within the CRISPR locus has been documented in the cases of methanoarchaea and mycobacteria [127]. An alignment of the CRISPR loci of three *S. solfataricus* strains revealed that in

addition to some deletion events and different spacers being incorporated into the locus after the strains deviated from one another, there was one implied instance of recombination. In the P1 strain, six repeat-spacer pairs from CRISPR locus A were deleted and replaced by four repeat-spacer pairs from CRISPR locus B, in what may have been a single event of recombination [91].

The CRISPR locus could also be altered through the insertion of transposons, mobile genetic elements. However, there appears to be strong selective pressure against the insertion of these elements [90]. The CRISPR locus is rarely targeted by integrative genetic elements, even in species in which genetic shuffling caused by these mobile sequences is common. Nonetheless, transpositions do on occasion occur, and interrupt the usual spacer-repeat sequence of the CRISPR locus [74,119].

4.4. Horizontal CRISPR–Cas element transfer

Horizontal gene transfer is the transfer of genes between organisms by mechanisms other than reproduction. Horizontal gene transfer allows genes to pass between distinct strains of bacteria, and plays a major role in the evolution of prokaryotes. Horizontal gene transfer is an example of quasi-Lamarckian evolution. While it represents a change in the organism genome caused by the external environment of the cell, the changes induced are not specific to genes relevant to the environment, and the genes transferred may or may not address the environment of the organism [83].

CRISPR loci have been transmitted between different, and sometimes distantly related, organisms. This horizontal transfer of CRISPR loci has been documented in many cases. For example, transfer of the CRISPR locus and Cas genes has been implicated between lactic acid bacteria and crenarchaea [74,91,119].

The mechanisms by which CRISPR loci can be transferred are largely unknown, but the transfer of loci via plasmids, phage, and other mobile elements is suspected. Some CRISPR loci are considered too large to be transferred by any mechanism but conjugation. During conjugation, a prokaryotic cell transfers genetic material to another cell through a direct linkage between the two cells [91]. However, it is likely that unknown transfer mechanisms are also at work.

4.5. Mathematical models of CRISPR–Cas evolution

The CRISPR–Cas system has caught the eye of many groups interested in creating models of its evolution. A Matlab simulation created by Childs and colleagues was used to test how CRISPR–Cas immunity can cause diversification of both the host prokaryote and the invading virus, and implicated that the host and virus coevolve into new stains. The most successful CRISPR-Cas genotypes showed similar phenotypes in a given environment. It is likely that several dominant CRISPR–Cas forms can exist in a population, which suggests that host-virus interactions help to maintain the coexistence of multiple CRISPR forms [23].

He and Deem generated a population dynamics model which suggested that the upstream end of the CRISPR loci would be more diverse than the downstream end [70]. In light of evidence that new spacers are acquired next to the leader sequence at the 5′ end of the locus [122], this model appears to be accurate.

Haeter and colleagues created a simple mathematical model which predicted that even when the amount of viruses that the CRISPR–Cas system must face is beyond the genetic capacity of the CRISPR–Cas system, the CRISPR–Cas system and viruses are still able to coexist. Haeter suggests that this is because the species are interdependent, even when there is significant stress on the CRISPR–Cas system [62].

4.6. Origin of CRISPR–Cas systems

CRISPR–Cas systems in archaea and bacteria have always seemed closely related. Comparisons of Cas gene sequences, adaptation mechanisms, and crRNA processing mechanisms have demonstrated that even though Bacteria and Archaea represent two entirely different domains of life, and Archaea is more closely related to eukaryotes than bacteria, the CRISPR–Cas systems found in these two domains are indeed very similar [63,64,98].

However, notable differences in the two systems are present. In Archaea, CRISPR loci are more abundant. They also are generally more complex and longer [57,106]. The repeat sequences of Archaea and Bacteria are also quite different; there is not much overlap between the sequences of families of CRISPR loci from these two domains. Additionally, while the repeats found in bacterial CRISPR loci allow for the formation of hairpin loops in their transcript,

archaeal repeats generally do not create transcripts with this secondary structure [45]. Most likely, adaptations specific to each domain have occurred during the course of evolution.

One environmental pressure that could account for the difference in the abundance of CRISPR loci between Bacteria and Archaea is the presence of a heated environment. Around 90% of thermophilic archaea possess some form of CRISPR–Cas system, while only around 50% of bacteria possess CRISPR–Cas. Weinberger and company proposed that this because high temperature environments are more difficult to survive in, making it less likely that a mutation in a virus will be favorable enough for that virus to survive. This leads to decreased rates of mutation in organisms in high temperature environments. Theoretical models predicted that a lower rate of mutation is conducive to CRISPR–Cas immunity, because rapidly mutating viruses evolve too quickly for the acquisition of spacers for each virus to be an efficient form of defense. Because of this, it is possible that bacteria lost their CRISPR–Cas systems as they were overwhelmed by viral adaptability and were no longer useful to the cell [133].

However, the previously mentioned simulation by Haeter and colleagues is in disagreement with this hypothesis, because in Haeter's simulation, even in environments where viral diversity overwhelmed the ability of CRISPR–Cas to acquire spacers, CRISPR–Cas was able to coexist with the viruses present [70]. In order to determine which proposed model is most reflective of reality, experiments in which CRISPR–Cas presence and CRISPR locus length is studied in comparison to the number of virus strains in the environment are necessary. In addition, studies that support that the rate of viral mutation is decreased in heated environments are also necessary. The results of such experiments could help to better determine the interaction between CRISPR–Cas and viruses in the environment, and in turn suggest the evolutionary history of CRISPR–Cas.

A comparison of Cas gene families by Makarova and colleagues led to the suggestion that the first evolved Cas proteins were the CasA through CasE, which make up the cascade effector complex. This is supported by the observation that systems which contain only the cascade complex, in absence of any other CRISPR modules, exist. On the other hand, there are no known systems that involve Cas1, Cas2, and the CRISPR locus without the presence of effector complexes. This may be because the function of Cas1 and 2 in the acquisition of spacers is dependent on the presence of the effector complex, and suggests that CRISPR–Cas systems could have evolved from an ancestor which contained only a variant of the cascade complex, which did not need Cas1, Cas2, and the CRISPR locus in order to degrade invading DNA molecules [97].

Because CRISPR–Cas systems could be selected for in conditions of high temperature, it is possible that the first CRISPR–Cas system arose from a cascade-only ancestor in a thermophilic archaeon. While some researchers agree that the CRISPR–Cas system originated in thermophilic Archaea and subsequently was transferred horizontally among prokaryotes [97], others disagree because of the difficulty associated with the horizontal transfer of genes between domains [119]. One alternative to origination in thermophilic Archaea is the origination of CRISPR–Cas before bacteria and archaea had diverged. However, in this scenario the evolution of Bacterial and Archaeal CRISPR–Cas systems would still require a small amount of interdomain gene transfer in order to align with current Cas gene phylogenies [119].

5. Eukaryotic analogs: miRNA, siRNA, and piRNA

CRISPR–Cas systems feature small RNAs which silence foreign genetic elements as an immune defense strategy against foreign invaders. Analogous systems can be seen throughout the diversity of life, as small, silencing RNAs are also intimately tied to immune function in domain Eukaryota. Eukaryotic immune systems involve micro(mi)RNA, small-interfering(si)RNA, and piwi-interacting(pi)RNA.

5.1. miRNA in immune defense

5.1.1. Overview of miRNA formation and function

MicroRNAs (miRNAs) are small RNAs that regulate gene expression by targeting messenger(m)RNAs in order to degrade them or repress their translation. A mature miRNA is generally around 21 to 25 nucleotides long. MiRNA precursors are transcribed from non-coding genes within the cell, and the initial transcription of these genes results in the creation of primary miRNA (pri-miRNA) transcripts [133]. Pri-miRNA is around 70 to 100 nucleotides long, and can contain one or more miRNA precursors (pre-miRNA). Pre-miRNA is structured in a characteristic hairpin shape. Initial miRNA transcripts are processed by RNAses Drosha and Dicer. The number of nucleotides conserved

between the hairpin form and mature miRNA form varies between miRNAs. In general, the loop of the hairpin is poorly conserved [124,139].

Mature miRNA can silence the expression of genes through the repression or cleavage of mRNA. During mRNA repression, the miRNA base-pairs to the target mRNA, which prevents the mRNA from being translated into a protein by ribosomes. Although multiple complementary base-pairs are necessary to successfully bind and prohibit translation, the miRNA does not need to be perfectly complementary to the target mRNA and cause degradation through regular mRNA decay machinery [41,124].

MiRNA is associated with an effector complex, the RNA-induced silencing complex (RISC). RISC contains Argonaute (AGO) proteins, and catalyzes the cleavage of mRNA [55]. In order for silencing via cleavage to occur, the miRNA must be fully complementary to the mRNA at the site of cleavage [66]. However, a large amount of complementary base-pairing does not necessarily mean that the mRNA will be degraded [21].

There is a clear parallel between miRNA and CRISPR-Cas systems; both involve the biogenesis of small RNAs, which are then used in conjunction with effector complexes to cleave complementary targets.

5.1.2. Function in mammalian immune systems

MiRNA can regulate the gene expression of not only the cells of the organism in which it is produced, but also the expression of foreign genetic elements. In humans, around 1000 unique miRNAs regulate up to 30% of the human genome encoding for proteins [81,97]. However, aside from this self-targeting, human miRNA can also target pathogenic viruses. It is estimated that most, if not all, mammalian viral genomes can be recognized by one or more human miRNAs, which would allow specific silencing of the viral genome to occur [132].

Several human miRNAs have been found to repress the replication of the pathogenic virus hepatitis B (HBV) [94]. Additionally, at least three human miRNAs target the influenza virus. The human miRNAs which target HIV-1 include miR-26a, miR-125b, miR-150, miR-223, and miR-382. Other pathogens for which human miRNAs have been found are Epstein Barr virus (EBV), coxsackie virus, human papilloma virus (HPV), and Kaposi's sarcoma herpes virus (KSHV) [75,77]. While this targeting could be chance, it appears that miRNAs are an important part of human antiviral defenses, and target and prohibit the expression of viral genes [77].

5.1.3. Evolution of miRNA

New miRNAs form when RNA produced from a non-coding intron randomly folds into a short hairpin structure. Other ways that new miRNAs can arise include the modification or duplication of pre-existing miRNAs [108]. A rapid increase in the diversity of miRNAs in existing genomes is often correlated to a burst of new morphological diversity, and thus miRNA may be crucial to the development of complex morphologies because of its regulatory activity [71,110,134].

MiRNA that has very recently evolved undergoes mutations at a rate that is similar to other introns, and can often be lost from the genome. However archaic miRNAs often have an extremely slow rate of mutation, which suggests that the miRNA undergoes purifying selection. Once an miRNA has gained a function, it rarely is lost from the genome [108,110]. Because older miRNAs evolve at a very low rate, and as a result are well conserved, they can be used as excellent phylogenetic markers [134].

5.1.4. Implications of CRISPR–Cas and miRNA similarities

The primary role of miRNA in eukaryotic species is self-regulation. MiRNA serves as a post-transcriptional control system that may be as important as transcription factors in determining whether or not a gene is expressed. MiRNA in eukaryotes is often self-encoded. The miRNA targets the cell's own genes in order to regulate all aspects of cell activity.

Although CRISPR–Cas has not traditionally been viewed as a regulatory system, the similarities between the self-targeting portions of miRNA and CRISPR loci lead to the conclusion that the development of a regulatory system from CRISPR–Cas may be possible. There are self-targeting spacers present in around 18% of organisms that possess a CRISPR–Cas system, however there is controversy over whether these spacers are autoimmune, and harm the organism, or are indeed regulatory. Some studies assert that the lack of conservation across CRISPR-Cas loci, which would be expected in a regulatory system and is seen in miRNA, leads to the conclusion that the self-targeting spacers are purely accidental and are harmful to the cell [120].

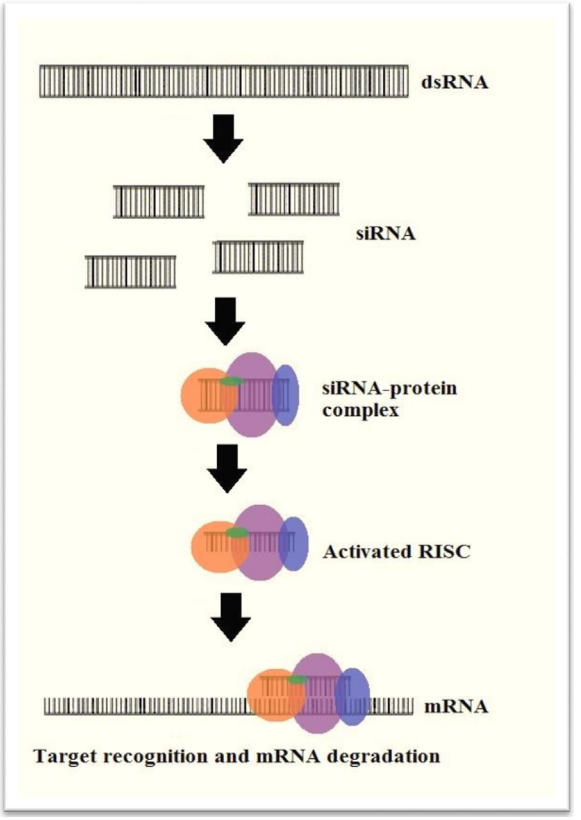

Fig. 3. The formation of siRNA-associated RISC from dsRNA.

However, research into non-immune functions of the CRISPR–Cas system revealed that CRISPR–Cas systems play a role in the behavior of at least two types of bacteria, which suggests that a regulatory component derived from the CRISPR–Cas system may have already arisen [4]. Although the mechanism of this regulation is unknown, both *Pseudomonas aeruginosa* and *M. Xanthus* show evidence of CRISPR-mediated regulation. In *P. aeruginosa*, biofilm formation is controlled by complex signaling which may be connected to CRISPR-Cas activation. In *M. Xanthus*, the formation of fruiting bodies is also potentially correlated to the activation of CRISPR–Cas. It may be the case that the initial phases of the CRISPR–Cas system are related to these regulatory functions [4].

5.2. SiRNA in immune defense

5.2.1. Overview of siRNA formation and function

SiRNA is non-coding, double-stranded RNA around 21 to 23 nucleotides long, which causes the degradation of mRNAs by directing the cleavage of these mRNAs. The process of siRNA silencing begins with exposure to dsRNA, including RNA viruses or transgenes. The initial RNA recognized is cleaved by the RNAse III Dicer, which breaks the RNA into duplexes that are around 21 to 24 nucleotides long. These duplexes are known as siRNA [8,67]. The activity of Dicer and its homologs is dependent on ATP, and involves an RNA helicase, dsRNA-binding proteins, and an argonaute(AGO)-like protein [8,121]. The siRNAs are then denatured from their duplex state and associated with RISC [68]. The complex uses single-stranded siRNA to recognize mRNA complements, and cause the degradation of these mRNAs (see Fig. 3) [8,44,68].

SiRNA and miRNA are very similar, as both cleave target mRNA after forming effector complexes with Argonaute (AGO). However, there are a few notable distinctions between miRNA and siRNA. MiRNA biogenesis requires the RNAses Drosha and Dicer, but siRNA processing depends solely on Dicer or Dicer-like RNAses [77]. Also, while miRNA often does not need to be highly complementary to repress translation, siRNA is extremely sequence specific [77].

5.2.2. Function in plant immune systems

The role of gene silencing in plant immune systems was first discovered by Lindbo and colleagues in 1993 [92]. In a process first known as PTSG, post-transcriptional gene silencing, plants combat foreign genetic elements by degrading the invasive genes through siRNAs.

The idea that viral infection in plants triggers the silencing of viral RNAs has been supported by several lines of experimentation. In 1999, Hamilton and Baulcombe investigated whether virus-specific small RNAs could be found in a virus-infected plant. RNAs 25 nucleotides long which were complementary to the potato virus X (PVX) were detected 4 days after infection of the plant *N. benthamiana*, and had a continued presence in the plant for at least another 6 days. It was concluded that these 25-nucleotide RNAs could not be produced after a degradation of the viral RNA, because the RNAs were antisense [67]. Additionally, plants that have compromised RNA silencing capacities are more susceptible to viral infection than plants with intact systems. For example, *Arabidopsis thaliana* contains the genes *SGS2*/*SDE1*, *SGS3*, *SDE3* and *AGO1*, which are all involved in RNA silencing [128]. *Arabidopsis* mutants defective in any of these genes were hypersensitive to CMV (Cucumber mosaic virus), but were equally susceptible to five other viruses as non-mutant plants [10,31,40,107]. However, tobacco and *Arabidopsis* plants mutant for a gene related to *SGS2*/*SDE1*, which, unlike *SGS2*/*SDE1*, is induced by viral infection, were more susceptible to both tobacco mosaic virus and PVX [3,40,138,140].

When dsRNA is detected in plants, Dicer-like (DCL) proteins, homologous to Dicer, are directed to cleave the dsRNA into 21–24 nucleotide duplexes. DCLs are endonucleases which behave mechanistically like Dicer, cleaving dsRNA into characteristic siRNAs [2,3,93]. In *Arabidopsis thaliana*, four DCLs have been identified. Each DCL has a different function in the creation of siRNAs. For example, while DCL3 produces 24-nucleotide products which target transposons and retroelements, DCL4 is the primary sensor which produces 21-nucleotide viral-targeting siRNAs. DCL1 acts in conjunction with DCL2, and if DCL4 is compromised, DCL2 can produce 22-nucleotide products which also can target viral mRNA [34,53]. The role of DCL2 and DCL4 in antiviral defense has been demonstrated for Cauliflower mosaic virus (CaMV) and CMV [38,102]. The loss of both DCL4 and DCL2 leads to the production of 24-nucleotide viral siRNA by DCL3 [50]. Based on *Arabidopsis* research, it is believed that other plants also possess several DCLs. DCLs associate with dsRNA-binding proteins (DRB) which assist in the production of siRNAs [73]. In *Arabidopsis*, DRB4 operates with DCL4 to produce siRNAs from viral RNA [30]. The interaction of CDL4 and DRB4 in viral siRNA production has been shown in experiments both *in vitro* and *in vivo* [3].

The siRNA produced in response to viral infection in plants is also dependent on RNA-dependent RNA polymerases (RDRs). *Arabidopsis* encodes six different RDR proteins. RDR1 assists in the formation of CMV – and Tobacco mosaic virus (TMV) – derived siRNAs [38,111,130]. Tobacco rattle virus (TRV)-derived siRNAs are formed via the activity of RDR1, RDR2, and RDR6 in conjunction with one another [50]. CMV accumulates in mutants lacking RDR6 [130]. Garcia-Ruiz and colleagues demonstrated that plants mutant for RDR1, RDR2, or RDR6 genes are more susceptible to primary infection. Both RDR1 and RDR6 were necessary, but individually insufficient, to prevent systemic infection. This supports that these three RDR proteins act in coordination to restrict infection. However, the mechanism of action for RDRs in preventing systemic infection is still unknown [50]. It is proposed that RDR6 associates with the coiled-coil protein SGS3 in order to amplify siRNA after initial AGO–siRNA targeting of the viral genome, creating secondary siRNAs. The initial AGO–siRNA targeting events could involve the cleavage of viral RNA as well as the association of AGO–siRNA complexes with viral RNA without cleavage. It is possible that AGO serves as a marker for the recruitment of RDR1 and RDR6 [50].

The *Arabidopsis* genome contains 10 AGO genes, and each functions to interact with specific short RNAs. A mutant virus whose siRNA suppressor activity had been silenced accumulated to higher levels in plants lacking AGO1 than in non-mutant plants, which supports that AGO1 plays a role in viral RNA silencing [50]. Wang and colleagues demonstrated that AGO4, AGO6, and AGO9 do not play a critical role in the silencing of CMV by the viral siRNAs, since mutants for these proteins did not show a greater susceptibility to CMV infection. However, CMV did accumulate at higher levels in *ago1 ago2* double mutants, as compared with *ago1* or *ago2* single mutants, which indicates that AGO1 and AGO2 likely cooperate in viral silencing. Additionally, CMV accumulated higher levels in the *ago1* and *ago2* double mutants than in mutants lacking RDR6 and SGS3. Since RDR6 and SGS3 are necessary to form secondary siRNAs, this implies that AGO1 and AGO2 are required for the antiviral activities of both viral primary and secondary siRNAs [131].

One shortfall in the research that has been conducted into siRNA antiviral defenses in plants is that the focus of research has largely remained almost exclusively on *Arabidopsis*. Although the *Arabidopsis* genome is conducive

to these studies, research into the siRNA pathways of diverse species are necessary in order to discover silencing mechanisms which may be different from those seen in *Arabidopsis* [3].

5.2.3. Function in invertebrate immune systems

SiRNA has been shown to provide antiviral immunity in a variety of invertebrates, including insects, crustaceans, and chelicerates. SiRNA machinery is well-documented in the nematode *C. elegans*, because siRNA was first characterized in this organism [56]. In 1998, Andrew Fire, Craig C. Mello, and colleagues found that in *C. elegans*, dsRNA was substantially more effective at producing interference than an individual strand of RNA. DsRNA was shown to cause strong specific interference after a few molecules of dsRNA were injected into a cell. This was a surprise to researchers, and a clear characterization of siRNA silencing. In 2006, Fire and Mello shared the Nobel Prize in Physiology or Medicine for their work on RNA interference [48]. Research into *C. elegans* siRNA has since progressed far beyond this point. Two *C. elegans* RDRs genes – *ego-1* and *rrf-1* – are implicated in the accumulation of antiviral siRNAs *in vivo*. A complex containing the proteins Mut-7 and Rde-2/Mut-8 is likely to be an intermediate between the initial Dicer complex and the RISC effector. Over 33 genes have been identified in the siRNA pathway in *C. elegans* [56].

In 2002, Caplen and co-workers presented evidence that an RNAi-like defense against Dengue virus (DENV) can be triggered in mosquito cells through the introduction of dsRNA corresponding to a portion of the DENV-1 or DENV-2 genome. The study demonstrated the silencing of DENV-1 replication after the infection of *Aedes albopictus* C6/36 cells with DENV-1 specific dsRNA [14].

In *Drosophila melanogaster* fruit flies, siRNA was detected after the injection of the flies with flock house virus (FHV). The siRNA corresponded specifically to FHV RNA1, supporting that siRNA opposes the accumulation of FHV RNAs. Additionally, siRNA quantities detected post-infection in *Drosophila* S2 cells has been shown to correlate with the quantity of dsRNA introduced to the cells [49,89]. It is postulated Dicer-2 and R2D2 are bound to the double-stranded siRNAs, to direct them to the RISC effector [125]. Of the many proteins involved in RISC, only a few have been characterized functionally, and those with known functions were identified after the purification of the RISC from *Drosophila melanogaster* S2 cells extracts. The proteins identified in the RNA-induced silencing complex of *D. melanogaster* include AGO2, the *Drosophila* Fragile X-Related protein, the Vasa Intronic Gene (VIG), and the micrococcal nuclease family member Tudor-SN [17–19,125].

In the marine shrimp *Litopenaeus vannamei*, sequence-specific dsRNA-mediated antiviral immunity was demonstrated by Robalino and colleagues. Sequence-dependent antiviral responses and gene silencing were triggered by the injection of long dsRNAs *in vivo*. This supports that marine shrimp possess not only innate immunity to viruses, but also an adaptive siRNA-mediated mechanism of antiviral immunity [112].

A comparative genomics study by Kurscheid and colleagues screened 13,643 *R. microplus* tick genome reads against model organisms *D. melanogaster* and *C. elegans* to identify proteins related to siRNA silencing pathways in ticks. Their analysis identified 31 proteins including a tick Dicer-like protein, RISC associated proteins such as AGO2, an RNA-dependent RNA polymerase, and 23 homologues which are implicated in dsRNA uptake and processing. This study also provides evidence that even though much of the basic gene silencing machinery is well-conserved evolutionarily, some components differ, which indicates that siRNA pathways might vary significantly between different classes of arthropod [87].

5.3. piRNA in immune defense

5.3.1. Overview of piRNA

The piRNA (piwi-interacting RNA) system was first discovered independently by several groups in 2006 [5,54,58,88]. PiRNAs are 26 to 30 nucleotide RNAs which associate with a Piwi protein. Piwi proteins are members of the Argonaute protein family. The Piwi subfamily is involved in germline-specific events such as stem cell maintenance and meiosis, and is responsible for maintaining incomplete differentiation in stem cells and stable rates of cell division in germ line cells. Piwi proteins are well-conserved, and are present in all studied animals [60].

Similarly, piRNAs have been found in invertebrates and vertebrates. PiRNAs are the small RNA guides of Piwi proteins, and do not have any distinct characteristics beyond a uridine at their 5′ end and a high likelihood of an adenosine as the tenth nucleotide [60]. While other small RNAs like miRNAs and siRNAs are created from double-stranded

hairpin RNA or other dsRNAs after cleavage by the Dicer enzyme, it is believed that piRNA is produced from long polycistronic RNA transcripts via an unknown Dicer-independent mechanism.

There are hundreds of thousands of piRNAs, which are formed from clustered, discrete loci in genomes which are riddled with transposable elements and repeat-sequence elements [119]. PiRNAs are distinctively localized, and are predominantly grouped into clusters of around 20–90 kilonucleotides, from which many small RNAs are transcribed [54].

PiRNAs appear to be formed by the seemingly near-random cleavage of the initial piRNA transcript [9]. Even though piRNAs begin with a uridine in almost all cases and thus there is a known bias for where the initial transcript is cleaved, the determination of which sequences become piRNAs is still believed to be largely random.

PiRNAs function to silence endogenous selfish genetic elements such as retrotransposons and repetitive sequences in germ cells. This ensures the stability of the germ cell genome. Because the majority of piRNAs are complementary to transposon sequences, it has been suggested that transposons are the primary piRNA target [85,100].

In 2007, Brennecke and colleagues proposed a Ping-Pong mechanism for the biogenesis of piRNA, based on the observation of a 10 nucleotide overlap between sense and antisense piRNAs. In this model, an antisense piRNA, which was derived from a piRNA locus and associated in a complex with a Piwi protein, would recognize and cause the cleavage of a sense transposon transcript. The cleavage would occur between nucleotides 10 and 11 of the antisense piRNA. The RNA resulting from the cleavage event would then associate with a second Piwi protein and ultimately become a new piRNA after 3′ processing. This mechanism explains the lack of uridine bias at the 5′ end of sense-strand piRNAs in AGO3 complexes, and would depend on the transcription of both sense and antisense transposon transcripts [11].

5.3.2. PiRNA evolution

The piRNA system is ancient, as are Piwi proteins. However, no homologs of Piwi have been found outside of the kingdom Animalia. While the locations of piRNA cluster loci are generally conserved between closely related species, the piRNA sequences themselves differ vastly, for example between rats and mice, *C. elegans* and *C. briggsae*, or *Drosophila melanogaster* and *D. simulans*. Therefore, it would appear that piRNAs evolve extremely rapidly [6,85,100,114].

In a recent study of human piRNAs, Lukic and Chen examined population differences in piRNAs, in light of the fact that piRNAs are not conserved outside of species. The study revealed strong statistical evidence that piRNAs are under selective constraint in three African populations. This finding is consistent with an independent study showing that African populations have higher rates of transposon insertion than do other populations. The study also suggested that reverse transcriptase might have an endogenous role specific to humans and therefore be protected from piRNA-mediated repression [47,85,96].

Assis and Kondrashov studied the evolution of 140 rodent piRNA clusters, of which 103 were purely non-coding. These clusters were compared between rats and mice. Fourteen of the studied clusters originated after the divergence of rats and mice as species. A high rate of piRNA cluster duplication was discovered, and not a single piRNA cluster was seen to be deleted, which suggests positive selection for higher expression level of piRNAs [6].

In 2009, Lu and Clark created a computer model to simulate the co-evolution of piRNA clusters and the mobile transposons that piRNA targets. The simulation was optimized for *Drosophila melanogaster*, and demonstrated that retrotransposons that generate piRNAs are advantageous, and retrotransposons that are repressed by piRNAs are likely to appear more frequently in a genome. These results were corroborated with a study of 9 *D. melanogaster* strains. This suggests that piRNAs can increase the fitness of individuals, while at the same time allowing retrotransposons to accumulate in a population by protecting the host from the deleterious effects of retrotransposition [95].

5.4. PiRNA and CRISPR–Cas comparisons

5.4.1. Similarities to CRISPR–Cas systems

CRISPR–Cas systems in prokaryotes and small RNA silencing systems in eukaryotes all bear similarities in that they function to interfere with gene expression; however, the closest analog to the CRISPR–Cas system is the piRNA system. PiRNA clusters, full of repeat-sequence elements and transposons, are very similar to CRISPR loci. PiRNA clusters incorporate transposon sequences which expands the catalogue of genetic elements that the cluster can silence. The CRISPR–Cas system also integrates new sequences to be repressed. Both systems are examples of Lamarckian

evolution, in which environmental factors directly cause heritable genetic changes. CRISPR loci and piRNA clusters are initially are transcribed into very long preliminary transcripts, which are then processed into useful small RNAs. PiRNAs associate with Piwi to form a silencing complex, similarly to how crRNAs associate with Cas/Cmr proteins to form silencing complexes. Similarities have even been found between Cas and piRNA-associated proteins at the tertiary level of protein structure [85,119].

5.4.2. Differences in CRISPR–Cas and piRNA

Despite the apparent functional similarities between CRISPR–Cas and piRNA, no homologous proteins have been found between the two systems at a sequence level. While proteins containing Piwi-like domains exist among prokaryotes, they are not believed to be involved in the CRISPR–Cas system or in any RNA interference pathway [119].

Additionally, the two systems have generally distinct targets. There have been some reports of CRISPR loci which contain matches to transposons [91,98,106,119]. However, these transposons were found in virus or plasmid genomes, and likely were not targeted for their identity as transposons, but simply because they were foreign genetic elements preceded by PAMs [85,119].

There are also differences in the arrangement of piRNA clusters and CRISPR loci. For CRISPR loci, spacers are incorporated at the leader sequence, causing a distinct and linear organization of spacers. However, when new transposons are acquired by the piRNA cluster, new inserts either tansposition into the cluster randomly or with a preference for the loci, but do not integrate in any specific logical sequence. Because of the randomness associated with new additions to piRNA clusters, adaption into pRNA clusters is less efficient than adaption into CRISPR loci [80].

PiRNAs play a more obvious role in self-regulation than crRNAs. PiRNAs function to control the gene expression of the host, through mRNA regulation, histone modification direction, and chromatin structures [80]. PiRNA performs these functions independently of its role in transposon silencing. The use of the CRISPR–Cas system as a potential form of endogenous gene regulation is still poorly understood. Although piRNA is certain to be subject to selective pressures similar to those undergone by miRNA, this may not be the case for CRISPR–Cas systems [22].

5.4.3. piRNA and CRISPR–Cas evolution

Because of the aforementioned differences between the CRISPR–Cas system and the piRNA system, their respective rates of evolution may be very different. Three key factors which bear importance for the evolution of these systems are their effective population sizes, generation turn-over rates, and rates of mutation. The effective population size of a cluster of transposable elements is the number of copies of the transposable element per haploid genome in the host species multiplied by the population size of the host species [42,85]. The number of copies is taken as an average, and can vary widely between loci and host species.

In the CRISPR–Cas system, the foreign elements which contribute spacers to the CRISPR loci are not confined to replicate along with the host, and thus generally have a larger population size than the CRISPR host. The effective population size of phages is thus larger than that of the CRISPR host. Although in general the generation times of phages and bacteria may be similar, phages have the potential for faster generation times, and have higher rates of mutation [43]. All of these factors contribute to more efficient natural selection for phage than for the CRISPR–Cas system.

In the piRNA system, the piRNA clusters are replicated along with the host genome, and have roughly the same rate of mutation as the rest of the host genome. Transposon elements, like phages, have a higher effective population size than the host species, and have the same generation turn-over as the host. The primary difference between the evolution of elements involved in the piRNA and CRISPR systems is the mutation rate, because phage genomes undergo mutation much more rapidly than transposons [85].

5.4.4. Gaps in evolutionary understanding

To date, homology has not been discovered between the CRISPR–Cas system and piRNA, the eukaryotic system most analogous to it. However, the functional similarity between these two distinct systems is undeniable. Still, some researchers maintain that the evolution of CRISPR–Cas occurred entirely independently of eukaryotic analogs [118]. The lack of evidence for homology strengthens this view. And yet other researchers believe that over time, homologous structures will be found that indicate very distant phylogenetic relationships between these Lamarckian systems. The mechanisms of CRISPR–Cas and piRNA systems are still poorly understood, and it is acknowledged that as the structures of their components are slowly revealed, homology may be found [119].

Ultimately, the evolutionary properties of small RNAs are also not yet clear enough in order to definitely determine the relationship between prokaryotic CRISPR–Cas systems and eukaryotic piRNA. The rate of evolution of CRISPR loci at the sequence level is still unknown, as is the rate of piRNA evolution. The amount of the CRISPR locus and piRNA sequence that is affected by external selective pressures is also uncertain.

Beyond the specific unknowns associated with the evolutionary dynamics of CRISPR–Cas and piRNA, there are broader questions left unanswered in the evolutionary dynamics of immune systems. For example, how the interaction of the CRISPR–Cas system with other prokaryotic defenses can affect the evolution of these systems has not been elucidated. In-depth, quantitative models of piRNA and CRISPR–Cas evolution are necessary in order to explain potential relationships between the two systems [85].

Although the relationship between prokaryotic and eukaryotic small RNA defense systems remains unknown, the involvement of small RNAs in immune functions is amazingly widespread. Small RNA molecular immunity is spread throughout the diversity of life, from single-celled bacteria, to plants, to invertebrate animals, to humans, which suggests not only that defense against genetic invaders has been selected for in all domains, but that small RNAs are an adaptation crucial to all immune systems.

6. Therapeutic value of small RNAs

6.1. Overview of RNAi therapies

RNA interference (RNAi) is silencing of gene expression through sequence-specific dsRNA. RNAi has been used in research to quickly and effectively silence genes, even in a tissue-specific manner, which has proved in some instances to be a tool more valuable than complete gene knock-outs. The potential for RNAi as an agent of genetic therapies has been studied in cases of cancers, genetic disorders, and viral infections [129]. The small RNA systems being used in RNAi therapies include siRNA and mimics of endogenous miRNAs. While siRNAs produce the efficient silencing of one specific gene, miRNA mimics tend to suppress the expression of more than one gene, but in a way such that expression is silenced to a lesser degree [13,109]. While canonical mechanisms of RNAi include miRNA, siRNA, and piRNA pathways, CRISPR–Cas can also be viewed as an RNAi system. It seems fitting to attempt to augment the human immune system by exploiting a system of immune defense that plays such a large role in the immunity of all forms of life.

6.2. Obstacles in therapeutic RNAi

RNAi has the potential to be used as a tool for the design of drugs based on small RNAs which can silence the expression of genes which lead to disease. However, there have been significant obstacles to overcome in making RNAi-based drug therapies a reality. RNAi can cause expression suppression in genes homologous to the desired gene, and produce effects which were not the result of the silencing of the target gene. In early tests of siRNA therapies in small animals, the effects produced by the therapies may not have been from the target knock-down at all, but from the off-target effects [82,113]. Additionally, RNA sensors in the innate immune system can prohibit effective RNAi by triggering the release of interferons. However, these obstacles are being overcome with the addition of chemical modifications to the RNA which have the ability to disrupt off-target effects without disrupting the silencing of the target gene [109].

The primary obstacle which RNAi therapies still must overcome is the delivery of small RNAs into specific tissues and cells. The delivery of small RNAs is a problem for all forms of RNAi therapy. When siRNA is injected into the bloodstream, they are quickly cleared from the blood by the kidneys and extracellular RNAses. Chemical modifications can help to increase the size of the siRNA so that it is above the weight threshold for kidney filtration, and also reduce its susceptibility to RNAses. However, entering the cell from the bloodstream is still a hurdle. Because of their negative charge and high molecular weight, siRNAs cannot cross the plasma membrane [82,113]. Cells can uptake siRNAs through endocytosis, but then the siRNA is trapped within an endosome, and cannot produce gene knock-downs. For example, while macrophages are able to bring the constructed siRNAs into their cell body through endocytosis, the majority of the siRNA remains trapped in its endosome, and therefore the knock-down effect of the siRNA is drastically diminished [33].

Effective siRNA delivery systems will need to allow the siRNA to successfully cross the cell or endosomal membrane. Whether or not new delivery systems will be effective depends on the quantitative of siRNA that reaches the cytosol of the target cell and how long it stays actively incorporated into RISC [109].

6.3. Targets of current research

Clinical trials have produced promising results for several siRNA drugs, including ALN-RSV01, TD101, ALN-VSP02, and CALAA-01, although there were many more RNAi therapy candidates for clinical development in 2012. ALN-RSV01, which combats Respiratory Syncytial Virus (RSV) after lung transplant, has completed phase 2 trials. TD101 targets keratin 6a N171K mutant mRNA, in order to fight pachyonychia congenital (PC1), an autosomal dominant skin disease [109]. Additionally, drugs have been presented with the aim of fighting age-related macular edema, diabetic molecular edema, hepatitis C virus, glaucoma, optic atrophy, amyloidosis, cytomegalovirus, human immunodeficiency virus, human papillomavirus, and many different forms of cancer [13,61,82,109,123]. For example, ALN-VSP02 aims to treat solid liver tumors, EZN-2968 aims to treat lymphoma, advanced solid tumors, and liver metastases, and the FANG vaccine aims to treat ovarian cancer, colon cancer, and other solid tumors. CALAA-01 also aims to treat cancer [42].

Huntington's disease (HD), an autosomal dominant neurodegenerative disease which causes movement disorder and dementia, is also a potential target for RNAi therapies. SiRNAs have been used to treat HD in pre-clinical trials studying HD equivalents in animals and cells from HD patients. An siRNA knock-out of the mutated HD transgene significantly inhibits neurodegeneration, improves motor control, and extends survival of mice with HD. SiRNA could prove to be a promising therapy for HD, given that the siRNA can be made selective enough to target the mutant gene without preventing the expression of the its normal, wild-type allele counterpart [141].

While all of the aforementioned therapies show the potential for the development of RNAi into powerful medicine, they also have highlighted the continued need for better siRNA delivery systems. Aptamer-siRNA chimeras and transferring-decorated nanoparticles are being studied as potential delivery augmenters, and stable nucleic acid lipid particles (SNALPs) and RNAi/oligonucleotide nanoparticle delivery (RONDEL) are both currently being used to deliver siRNA [123]. As new delivery systems develop, the efficacy of RNAi therapies will increase [13,33].

6.4. Potential for therapies based on the CRISPR–Cas system

The creation of novel RNAi therapies to treat various human diseases has thus far been limited to the utilization of siRNA and miRNA interference pathways. However, new research by the several different teams supports that not only is the artificial construction of RNAi systems which utilize CRISPR–Cas components possible, but it seems to be a promising new direction for next generation targetable cleavage reagents.

In order to figure out how to design crRNAs against a chosen transcript, Hale and colleagues determined the components necessary for crRNA/Cmr-mediated target recognition in *Pyrococcus furiosus*. Deep sequencing of the crRNA associated with Cmr complex proteins led to the discovery of two predominant characteristics universal to *P. furiosus* crRNAs. The first lies at the 5′ end of the crRNA: an 8 nucleotide leading tag left on the crRNA after cleavage of the pre-crNRA transcript within the repeat sequence. The loss of this tag or the mutation of even two of the nucleotides within the tag leads to a complete loss of target crRNA-mediated cleavage. The second key feature is the guide sequence, derived from a CRISPR locus spacer. Cmr complex crRNAs are consistently either 39 or 45 nucleotides in length. Guide sequences in *P. furiosus* CRISPRs are generally 37 nucleotides long, and including the 8 nucleotide tag, are 45 nucleotides long. In contrast, the 39-nt species still possesses the 8 nucleotide tag, but lacks 6 nucleotides of the guide sequence at the 3′ end [65].

After determining the defining characteristics necessary for functional Cmr-associating crRNA, whether or not such a crRNA could be constructed to target a chosen transcript for cleavage was examined. *In vitro* assays were used to combine the six proteins of the Cmr complex with crRNAs complementary to target mRNA, and these crRNA constructs yielded the predicted cleavage products. In order to silence a target prokaryotic gene in a species already producing the Cmr complex, it is thus only necessary to create a plasmid which encodes for the desired guide sequence and the tag sequence. For a prokaryote that lacks Cmr genes, these can easily be introduced in a single operon along with the engineered crRNA. This opens up potential avenues for microorganism customization and research into gene functions [65].

Experimentation into the use of CRISPR–Cas elements in genetic engineering did not stop with the utility of crRNA constructs in artificial prokaryote gene silencing. Jinek and colleagues used spacers 20–30 nucleotides in size to demonstrate the efficiency and specificity of cleavage by Cas9–RNA complexes. A minimum of 16 base pairs is required for cleavage by Cas9, and Cas9 recognizes two to three nucleotides of the PAM sequence. This information was used to produce an RNA molecule that works with Cas9 to cleave RNA. The authors proposed that this system has the capacity to replace the widely used zinc-finger nuclease (ZFN) and transcription activator-like effector nuclease (TALEN) systems currently used to cleave targeted genes in eukaryotes [78].

For the constructed crRNA–Cas9 system to work in organisms other than prokaryotes, Cas9 and the crRNA would need to be expressed by the target cells or organism. It is possible that both could be introduced into cells via vector after their formation *in vitro* in conjunction with appropriate promoters. The targeting RNA is highly specific to the desired target, as a match of only 16 nucleotides is necessary to cause recognition and subsequent silencing, which is more specific than current ZFN and TALEN systems [46,78,99].

However, both ZFN and TALEN are composed of transcription factors that occur naturally in eukaryotes, while CRISPR–Cas components normally are not usually expressed in these cells [15]. It is possible that Cas9 would not efficiently function in the context of a genetic environment composed of chromatin. It is also possible that the DNA–RNA hybrid that would be necessary for cleavage will not be stable enough. Only attempts at introducing the crRNA–Cas9 system to eukaryotic cells would be able determine the validity of these concerns [78].

Fortunately, experimentation by Cong and colleagues confirmed that the editing of genetic sites within the mammalian genome is possible using Cas9. These studies focused on two separate type II CRISPR–Cas systems, and demonstrated that Cas9 can be used to cause the specific cleavage of loci in both rat and human genomes. Additionally, multiple crRNA constructs can be introduced into eukaryotic cells within a single CRISPR array, facilitating the simple simultaneous editing of multiple sites of interest. These studies also showed that the CRISPR–Cas system can take advantage of recombination machinery to replace a DNA locus with an engineered DNA cassette [24].

Cas9 was also explored as a human genetic engineering tool by Mali and colleagues. The CRISPR system was used to repair a defective integrated GFP construct in human cells at 8% efficiency, as demonstrated by the detection of fluorescence by flow cytometry. The custom guide sequence (gRNA) mimicking crRNA was introduced along with Cas9 into mammalian cells, and showed targeting rates of 10 to 25% in 293T cells, 13 to 8% in K562 cells, and 2 to 4% in induced pluripotent stem cells, for the AAVS1 locus. Mali and colleagues also demonstrated that the gRNA–Cas9 system can edit multiple gene sites at once. Additionally, the authors computed a resource of unique gRNAs which could be used to target around 40.5% of human exons, thereby establishing a broad RNA-guided editing tool for human genome engineering [99].

In April of 2013, DiCarlo and colleagues studies the mutagenic capabilities of CRISPR–Cas, as demonstrated by the CAN1 mutagenesis experiments, and highlighted the potential for the use of CRISPR–Cas to make targeted gene knock-outs. Their studies showed that if a mutagenic donor DNA to knock-out the target gene is provided for homologous recombination, the efficiency of the Cas9 system can be dramatically increased [39].

In the same month, Chang and colleagues demonstrated that RNA-guided Cas9 nuclease efficiently facilitates genome editing in both mammalian cells and zebrafish embryos. Over 35% of site-specific mutations were found when specific Cas/gRNA was used to target genes in living zebrafish embryos. The study also successfully achieved site-specific insertion of mloxP sequence induced by Cas9/gRNA system in zebrafish embryos, and showed that the Cas9/gRNA system has the potential to be a simple and efficient reverse genetic tool for zebrafish and other model organisms [20].

The use of CRISPR–Cas components in human genetic engineering is more scalable and affordable than the use of zinc fingers or TALEN systems. Perhaps most importantly, it is also far easier, as the design of crRNAs is simpler than the design of ZFNs or TALENs [65]. Some authors speculate that as the new Cas9/gRNA system improves, it will very rapidly become the nuclease of choice for genetic engineering [116]. It opens new doors for previously difficult multiplex genome editing in mammalian cells, which will have significant implications across not only the field of biotechnology, but also medicine. The ability to successfully and efficiently edit the human genome could lead to novel gene therapies for a variety of diseases, and the magnitude of the discovery of CRISPR–Cas system applications will be revealed as research into this new tool progresses.

References

[1] Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, et al. Transcription profile of Thermus thermophilus CRISPR systems after phage infection. J Mol Biol 2010;395:270–81.
[2] Aliyari R, Ding S. RNA-based viral immunity initiated by the Dicer family of host immune receptors. Immunol Rev 2009;227:176–88.
[3] Alvarado V, Scholthof H. Plant responses against invasive nucleic acids: RNA silencing and its suppression by plant viral pathogens. Semin Cell Dev Biol 2009;20:1032–40.
[4] Amitai G, Sorek R. Roles of CRISPR in regulation of physiological processes. In: Barrangou R, van der Oost J, editors. CRISPR–Cas systems. Berlin Heidelberg: Springer; 2013. p. 251–66.
[5] Aravin A, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, et al. The small RNA profile during Drosophila melanogaster development. Dev Cell 2003;5:337–50.
[6] Assis R, Kondrashov A. Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. Proc Natl Acad Sci USA 2009;106:7079–82.
[7] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007;315:1709–12.
[8] Bernstein E, Caudy A, Hammond S, Hannon G. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001;409:363–6.
[9] Betel D, Sheridan R, Marks D, Sander C. Computational analysis of mouse piRNA sequence and biogenesis. PLoS Comput Biol 2007;3:e222.
[10] Boutet S, Vazquez F, Liu J, Béclin C, Fagard M, et al. Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr Biol 2003;13:843–8.
[11] Brennecke J, Aravin A, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 2007;128:1089–103.
[12] Brouns S, Jore M, Lundgren M, Westra E, Slijkhuis R, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 2008;321:960–4.
[13] Burnett J, Rossi J. RNA-based therapeutics: current progress and future prospects. Chem Biol 2012;19:60–71.
[14] Caplen N. Inhibition of viral gene expression and replication in mosquito cells by dsRNA-triggered RNA interference. Mol Ther 2002;6:243–51.
[15] Carroll D. A CRISPR approach to gene targeting. Mol Ther 2009;20:1658–60.
[16] Carte J, Wang R, Li H, Terns R, Terns M. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev 2008;22:3489–96.
[17] Caudy A, Myers M, Hannon G, Hammond S. Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev 2002;16:2491–6.
[18] Caudy A, Ketting R, Hammond S, Denli A, Bathoorn A, Tops B, et al. A micrococcal nuclease homologue in RNAi effector complexes. Nature 2003;425:411–4.
[19] Caudy A, Hannon G. Induction and biochemical purification of RNA-induced silencing complex from drosophila S2 cells. In: Gott J, editor. RNA interference, editing, and modification, vol. 265. Humana Press; 2004. p. 59–72.
[20] Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X, et al. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res 2013;23:465–72.
[21] Chen X. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 2004;303:2022–5.
[22] Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet 2007;8:93–103.
[23] Childs L, Held N, Young M, Whitaker R, Weitz J. Multi-scale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. Evolution 2012;66:2015–29.
[24] Cong L, Ran F, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339:819–23.
[25] Cooper EL. Evolution of immune systems from self/not self to danger to artificial immune systems (AIS). Phys Life Rev 2010;7:55–78.
[26] Cooper EL. Self/not self, innate immunity, danger, cancer potential. Phys Life Rev 2010;7:85–6.
[27] Cooper EL. Adaptive immunity from prokaryotes to eukaryotes: broader inclusions due to less exclusivity?. In: Kanwar J, editor. Recent advances in immunology to target cancer, inflammation and infections. InTech; 2012. p. 495–520.
[28] Cooper EL, Rinkevich B, Uhlenbruck G, Valembois P. Invertebrate immunity: another viewpoint. Scand J Immunol 1992;35:247–66.
[29] Cooper EL, Kauschke E, Cossarizza A. Digging for innate immunity since Darwin and Metchnikoff. BioEssays 2002;24:319–33.
[30] Curtin S, Watson J, Smith N, Eamens A, Blanchard C, Waterhouse P. The roles of plant dsRNA-binding proteins in RNAi-like pathways. FEBS Lett 2008;582:2753–60.
[31] Dalmay T, Horsefield R, Braunstein T, Baulcombe D. SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis. EMBO J 2001;20:2069–78.
[32] Datsenko K, Pougach K, Tikhonov A, Wanner B, Severinov K, et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun 2012;3:945.
[33] de Fougerolles A, Vornlocher H, Maraganore J, Lieberman J. Interfering with disease: a progress report on siRNA-based therapeutics. Nat Rev Drug Discov 2007;6:443–53.
[34] Deleris A, Gallego-Bartolome J, Bao J, Kasschau K, Carrington J, Voinnet O. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science 2006;313:68–71.
[35] Deltcheva E, Chylinski K, Sharma C, Gonzales K, Chao Y, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011;471:602–7.

[36] Deveau H, Barrangou R, Garneau J, Labonte J, Fremaux C, et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 2008;190:1390–400.
[37] Deveau H, Garneau J, Moineau S. CRISPR/Cas system and its role in phage–bacteria interactions. Annu Rev Microbiol 2010;64:475–93.
[38] Diaz-Pendon J, Li F, Li W, Ding S. Suppression of antiviral silencing by cucumber mosaic virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. Plant Cell 2007;19:2053–63.
[39] Dicarlo J, Norville J, Mali P, Rios X, Aach J, Church G. Genome engineering in Saccharomyces cerevisiae using CRISPR–Cas systems. Nucleic Acids Res 2013;41:4336–43.
[40] Ding S, Li H, Lu R, Li F, Li W. RNA silencing: a conserved antiviral immunity of plants and animals. Virus Res 2004;102:109–15.
[41] Doench J, Sharp P. Specificity of microRNA target selection in translational repression. Genes Dev 2004;18:504–11.
[42] Dolgin E, Charlesworth B, Cutter A. Population frequencies of transposable elements in selfing and outcrossing Caenorhabditis nematodes. Genet Res (Camb) 2008;90:317–29.
[43] Duffy S, Shackelton L, Holmes E. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 2008;9:267–76.
[44] Elbashir S. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev 2001;15:188–200.
[45] Erdmann S, Garrett R. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol 2012;85:1044–56.
[46] Esvelt K, Wang H. Genome-scale engineering for systems and synthetic biology. Mol Syst Biol 2013;9:641.
[47] Ewing A, Kazazian H. Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. Genome Res 2011;21:985–90.
[48] Fire A, Xu S, Montgomery M, Kostas S, Driver S, Mello C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 1998;391:806–11.
[49] Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann J, Imler J. Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nat Immunol 2006;7:590–7.
[50] Garcia-Ruiz H, Takeda A, Chapman E, Sullivan C, Fahlgren N, Brempelis K, et al. Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. Plant Cell 2010;22:481–96.
[51] Garneau J, Dupuis M, Villion M, Romero D, Barrangou R, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 2010;468:67–71.
[52] Garrett R, Vestergaard G, Shah S. Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol 2011;19:549–56.
[53] Gasciolli V, Mallory A, Bartel D, Vaucheret H. Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr Biol 2005;15:1494–500.
[54] Girard A, Sachidanandam R, Hannon G, Carmell M. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 2006;442:199–202.
[55] Gregory R, Chendrimada T, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell 2005;123:631–40.
[56] Grishok A. RNAi mechanisms in Caenorhabditis elegans. FEBS Lett 2005;579:5932–9.
[57] Grissa I, Vergnaud G, Pourcel C. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 2008;36:W145–8.
[58] Grivna S, Pyhtila B, Lin H. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. Proc Natl Acad Sci USA 2006;103:13415–20.
[59] Gudbergsdottir S, Deng L, Chen Z, Jensen J, Jensen L, et al. Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol Microbiol 2011;79:35–49.
[60] Gunawardane L, Saito K, Nishida K, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science 2007;315:1587–90.
[61] Haasnoot J, Westerhout E, Berkhout B. RNA interference against viruses: strike and counterstrike. Nat Biotechnol 2007;25:1435–43.
[62] Haerter J, Trusina A, Sneppen K. Targeted bacterial immunity buffers phage diversity. J Virol 2011;85:10554–60.
[63] Haft D, Selengut J, Mongodin E, Nelson K. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 2005;1:60.
[64] Hale C, Zhao P, Olson S, Duff M, Graveley BR, et al. RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. Cell 2009;139:945–56.
[65] Hale C, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, et al. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol Cell 2012;45:292–302.
[66] Haley B, Zamore P. Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol 2004;11:599–606.
[67] Hamilton A, Baulcombe D. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 1999;286:950–2.
[68] Hammond S, Caudy A, Hannon G. Post-transcriptional gene silencing by double-stranded RNA. Nat Rev Genet 2001;2:110–9.
[69] Han D, Krauss G. Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2. FEBS Lett 2009;583:771–6.
[70] He J, Deem MW. Heterogeneous diversity of spacers within CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). Phys Rev Lett 2010;105:128102.
[71] Heimberg A, Sempere L, Moy V, Donoghue P, Peterson K. MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci USA 2008;105:2946–50.
[72] Hermans P, et al. 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;59:2695–705.
[73] Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, Murai Y, et al. Specific interactions between dicer-like proteins and HYL1/DRB- family dsRNA-binding proteins in Arabidopsis thaliana. Plant Mol Biol 2005;57:173–88.

[74] Horvath P, Romero D, Coute-Monvoisin A, Richards M, Deveau H, et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol 2008;190:1401–12.
[75] Houzet L, Jeang K. MicroRNAs and human retroviruses. Biochim Biophys Acta 2011;1809:686–93.
[76] Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol 1987;169:5429–33.
[77] Jeang K. RNAi in the regulation of mammalian viral infections. BMC Biol 2012;10:58.
[78] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012;337:816–21.
[79] Karginov F, Hannon G. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell 2010;37:7–19.
[80] Khurana J, Wang J, Xu J, Koppetsch B, Thomson T, Nowosielska A, et al. Adaptation to P element transposon invasion in Drosophila melanogaster. Cell 2011;147:1551–63.
[81] Kim V, Han J, Siomi M. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 2009;10:126–39.
[82] Kleinman M, Yamada K, Takeda A, Chandrasekaran V, Nozaki M, Baffi J, et al. Sequence- and target-independent angiogenesis suppression by siRNA via TLR3. Nature 2008;452:591–7.
[83] Koonin E. The CRISPR–Cas system of antivirus immunity in prokaryotes: The showcase of a bona fide Lamarckian mechanism. The logic of chance: the nature and origin of biological evolution. FT Press; 2011.
[84] Koonin E, Wolf Y. Is evolution Darwinian or/and Lamarckian? Biol Direct 2009;4:42.
[85] Kumar M, Chen K. Evolution of animal Piwi-interacting RNAs and prokaryotic CRISPRs. Brief Funct Genom 2012;11:277–88.
[86] Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol 2007;8:R61.
[87] Kurscheid S, Lew-Tabor A, Rodriguez Valle M, Bruyeres A, Doogan V, Munderloh U, et al. Evidence of a tick RNAi pathway by comparative genomics and reverse genetics screen of targets with known loss-of-function phenotypes in Drosophila. BMC Mol Biol 2009;10:26.
[88] Lau N, Seto A, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel D, et al. Characterization of the piRNA complex from rat testes. Science 2006;313:363–7.
[89] Li H, Li W, Ding S. Induction and suppression of RNA silencing by an animal virus. Science 2002;296:1319–21.
[90] Lillestøl R, Redder P, Garrett R, Brügger K. A putative viral defence mechanism in archaeal cells. Archaea 2006;2:59–72.
[91] Lillestol R, Shah S, Brugger K, Redder P, Phan H, et al. CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol Microbiol 2009;72:259–72.
[92] Lindbo J, Silva-Rosales L, Proebsting W, Dougherty W. Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. Plant Cell 1993;5:1749–59.
[93] Liu Q, Feng Y, Zhu Z. Dicer-like (DCL) proteins in plants. Funct Itegr Genomics 2009;9:277–86.
[94] Liu W, Yeh S, Chen P. Role of microRNAs in hepatitis B virus replication and pathogenesis. Biochim Biophys Acta 2011;1809:678–85.
[95] Lu J, Clark A. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. Genome Res 2010;20:212–27.
[96] Lukic S, Chen K. Human piRNAs are under selection in Africans and repress transposable elements. Mol Biol Evol 2011;28:3061–7.
[97] Makarova K, Aravind L, Wolf Y, Koonin E. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR–Cas systems. Biol Direct 2011;6:38.
[98] Makarova K, Grishin N, Shabalina S, Wolf Y, Koonin E. 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.
[99] Mali P, Yang L, Esvelt K, Aach J, Guell M, DiCarlo J, et al. RNA-guided human genome engineering via Cas9. Science 2013;339:823–6.
[100] Malone C, Hannon G. Small RNAs as guardians of the genome. Cell 2009;136:656–68.
[101] Marraffini L, Sontheimer E. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 2008;322:1843–5.
[102] Moissiard G, Voinnet O. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis dicer-like proteins. Proc Natl Acad Sci USA 2006;103:19593–8.
[103] Mojica F, et al. 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;17:85–93.
[104] Mojica F, Diez-Villasenor 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;36:244–6.
[105] Mojica F, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 2005;60:174–82.
[106] Mojica F, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 2009;155:733–40.
[107] Morel J, Godon C, Mourrain P, Beclin C, Boutet S, et al. Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. Plant Cell 2002;14:629–39.
[108] Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in Drosophila species. Genome Biol Evol 2010;2:180–9.
[109] Peer D, Lieberman J. Special delivery: targeted therapy with small RNAs. Gene Ther 2011;18:1127–33.
[110] Peterson K, Dietrich M, McPeek M. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. BioEssays 2009;31:736–47.
[111] Qi X, Bao F, Xie Z. Small RNA deep sequencing reveals role for Arabidopsis thaliana RNA-dependent RNA polymerases in viral siRNA biogenesis. PLoS ONE 2009;4:e4971.

[112] Robalino J, Bartlett T, Shepard E, Prior S, Jaramillo G, Scura E, et al. Double-stranded RNA induces sequence-specific antiviral silencing in addition to nonspecific immunity in a marine shrimp: convergence of RNA interference and innate immunity in the invertebrate antiviral response? J Virol 2005;79:13561–71.
[113] Robbins M, Judge A, Ambegia E, Choi C, Yaworski E, Palmer L, et al. Misinterpreting the therapeutic effects of small interfering RNA caused by immune stimulation. Hum Gene Ther 2008;19:991–9.
[114] Ruby J, Jan C, Player C, Axtell M, Lee W, Nusbaum C, et al. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell 2006;127:1193–207.
[115] Samai P, Smith P, Shuman S. Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. Acta Crystallographica Section F Struct Biol Cryst Commun 2010;66:1552–6.
[116] Segal D. Bacteria herald a new era of gene editing. Elife 2013;2:e00563.
[117] Semenova E, Jore M, Datsenko K, Semenova A, Westra E, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci USA 2011;108:10098–103.
[118] Shabalina S, Koonin E. Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol 2008;23:578–87.
[119] Shah S, Garrett R. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol 2011;162:27–38.
[120] Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet 2010;26:335–40.
[121] Tabara H, Yigit E, Siomi H, Mello C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. Cell 2002;109:861–71.
[122] Terns M, Terns R. CRISPR-based adaptive immune systems. Curr Opin Microbiol 2011;14:321–7.
[123] Uchino K, Ochiya T, Takeshita F. RNAi therapeutics and applications of MicroRNAs in cancer treatment. Jpn J Clin Oncol 2013.
[124] Valencia-Sanchez M, Liu J, Hannon G, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev 2006;20:515–24.
[125] van den Berg A, Mols J, Han J. RISC-target interaction: cleavage and translational suppression. Biochim Biophys Acta 2008;1779:668–77.
[126] van der Oost J, Jore M, Westra E, Lundgren M, Brouns S. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci 2009;34:401–7.
[127] van Embden J, van Gorkom T, Kremer K, Jansen R, van Der Zeijst B, et al. Genetic variation and evolutionary origin of the direct repeat locus of Mycobacterium tuberculosis complex bacteria. J Bacteriol 2000;182:2393–401.
[128] Vance V, Vaucheret H. RNA silencing in plants-defense and counterdefense. Science 2001;292:2277–80.
[129] Wall N, Shi Y. Small RNA: can RNA interference be exploited for therapy?. Lancet 2003;362:1401–3.
[130] Wang X, Wu Q, Ito T, Cillo F, Li W, Chen X, et al. RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. Proc Natl Acad Sci USA 2010;107:484–9.
[131] Wang X, Jovel J, Udomporn P, Wang Y, Wu Q, Li W, et al. The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in Arabidopsis thaliana. Plant Cell 2011;23:1625–38.
[132] Watanabe Y, Kishi A, Yachie N, Kanai A, Tomita M. Computational analysis of microRNA-mediated antiviral defense in humans. FEBS Lett 2007;581:4603–10.
[133] Weinberger A, Wolf Y, Lobkovsky A, Gilmore M, Koonin E. Viral diversity threshold for adaptive immunity in prokaryotes. MBio 2012;3:e00456-12.
[134] Wheeler B, Heimberg A, Moy V, Sperling E, Holstein T, Heber S, et al. The deep evolution of metazoan microRNAs. Evol Dev 2009;11:50–68.
[135] Wiedenheft B, Zhou K, Jinek M, Coyle S, Ma W, et al. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 2009;17:904–12.
[136] Wiedenheft B, Lander G, Zhou K, Jore M, Brouns S, et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 2011;477:486–9.
[137] Wilson G. Restriction and modification systems. Annu Rev Genet 1991;25:585–627.
[138] Xie Z, Fan B, Chen C, Chen Z. An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense. Proc Natl Acad Sci USA 2001;98:6516–21.
[139] Yi R, Qin Y, Macara I, Cullen B. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev 2003;17:3011–6.
[140] Yu D, Fan B, MacFarlane S, Chen Z. Analysis of the involvement of an inducible Arabidopsis RNA-dependent RNA polymerase in antiviral defense. Mol Plant-Microbe Interact 2003;16:206–16.
[141] Zhang Y, Friedlander R. Using non-coding small RNAs to develop therapies for Huntington's disease. Gene Ther 2011;18:1139–49.
[142] Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, et al. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell 2012;45:303–13.