Update

Genome Analysis

Self-targeting by CRISPR: gene regulation or autoimmunity?

Adi Stern[1*], Leeat Keren[2*], Omri Wurtzel[1], Gil Amitai[1] and Rotem Sorek[1]

[1] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel
[2] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel

The recently discovered prokaryotic immune system known as CRISPR (clustered regularly interspaced short palindromic repeats) is based on small RNAs ('spacers') that restrict phage and plasmid infection. It has been hypothesized that CRISPRs can also regulate self gene expression by utilizing spacers that target self genes. By analyzing CRISPRs from 330 organisms we found that one in every 250 spacers is self-targeting, and that such self-targeting occurs in 18% of all CRISPR-bearing organisms. However, complete lack of conservation across species, combined with abundance of degraded repeats near self-targeting spacers, suggests that self-targeting is a form of autoimmunity rather than a regulatory mechanism. We propose that accidental incorporation of self nucleic acids by CRISPR can incur an autoimmune fitness cost, and this could explain the abundance of degraded CRISPR systems across prokaryotes.

CRISPR/Cas, an acquired anti-viral system in prokaryotes

CRISPR loci are found in nearly all of archaeal and about 40% of sequenced bacterial genomes. CRISPR loci, together with their CRISPR-associated (*cas*) genes, have recently been shown to constitute a defense system that restricts propagation of incoming viruses and plasmids [1,2]. CRISPR arrays are composed of short repeat sequences separated by similarly sized hypervariable 'spacer' sequences, flanked on one side by an AT-rich sequence called the leader. The discovery that CRISPR spacers often match DNA from foreign elements led to the realization that they represent a 'memory of past genetic aggressions' [3–5]. Step by step, the mechanism underlying CRISPR defense has begun to be unraveled, yet our understanding of this system is far from complete. It has been revealed that the CRISPR locus is transcribed into a single RNA transcript, which is then further cleaved by the Cas proteins to generate smaller CRISPR RNA (crRNA) units, each including one targeting spacer [6]. These units then interfere with incoming foreign genetic material by complementary base-pairing with the foreign nucleic acid [7–10]. CRISPR systems have been divided into different clusters based on their repeat sequences [11], and these correlate with different subtypes of *cas* genes [12]. It was shown that Cas subtypes *mtube*, *ecoli* and *nmeni* are likely to target DNA [2,5,6,13], whereas the Cas module RAMP was shown to target RNA [14].

Although CRISPR/Cas was initially prophesized to be analogous to eukaryotic RNA interference (RNAi) [15], it is now becoming clear that there are key differences between RNAi and CRISPR [10]. Nevertheless, the conceptual similarities between these two systems allow us to use our broader understanding of RNAi to guide the study of the CRISPR system [8]. Eukaryotic RNAi systems are divided into two branches: the antiviral branch that targets viruses and transposons for degradation, and the regulatory branch that utilizes microRNAs (miRNAs) for translational repression of target mRNA molecules via partial base pairing [16]. Previous limited searches revealed CRISPR spacers targeting chromosomal genes [4,5,17–19] and, based on the conceptual analogy between RNAi and CRISPR, it was therefore hypothesized that the CRISPR system in prokaryotes could also participate in gene regulation [9].

We have explored this possibility by studying self-targeting CRISPR spacers from all known CRISPR arrays [20] in all currently sequenced prokaryotic genomes (Table S1 in supplementary material online). Unexpectedly, our results point to a different explanation for self-targeting by CRISPR: leaky incorporation of self nucleic-acids leading to autoimmunity. We further explore this new concept of CRISPR-based autoimmunity from an evolutionary angle, as well as its consequences and fitness costs on CRISPR-bearing organisms.

Self-targeting CRISPR spacers

To identify potential self-targeting spacers, 23 550 spacers from 330 CRISPR-encoding organisms were scanned for an exact full match between the spacer and a portion of the endogenous genomic sequence that is not part of a CRISPR array (termed target, or self proto-spacer). Our results reveal that 100 of 23 550 spacers (0.4%) are self-targeting (Table S2). However, encoding a self-targeting spacer is not a rare phenomenon: 59 of 330 (18%) CRISPR-encoding organisms possess at least one array with at least one self-targeting spacer. These spacers are widely distributed over diverse phylogenetic lineages (Figure S1), and are dispersed throughout different arrays in each organism.

Is CRISPR/Cas a regulatory system?

One of the basic postulates of evolutionary theory is that functional elements undergo purifying selection, leading to their conservation across different organisms [21]. Returning to the superficial analogy with eukaryotic RNAi,

Corresponding author: Sorek, R. (rotem.sorek@weizmann.ac.il).
* These authors contributed equally to this work.

miRNAs are among the most highly conserved non-coding elements in mammalian genomes [22]. Hence, an essential requirement for CRISPR to function as an established regulatory system is the evolutionary conservation of the self-targeting spacers across several species. To test for such conservation we compared the sequences of all CRISPR self-targeting spacers. However, we did not find a conserved self-targeting spacer in even one single case.

Considering the possibility that CRISPR regulation might occur via partial base pairing (as in eukaryotic miRNAs or in the RAMP CRISPR-associated module [14]), we also examined spacers with partial or inexact matches to endogenous DNA. Once again, both partial and fully matching endogenous spacers showed no signs of conservation – in other words, they were present in only one organism (apart from rare cases where self-targeting spacers were present in very closely related strains; see supplementary material). To summarize, our results showed that each pair of self-targeting spacer and target exists only in one organism. This lack of conservation casts doubts on the hypothesis that the self-targeting spacers we detected regulate self genes: had the initial insertion of a self-targeting spacer conferred an evolutionary advantage to the organism, and had it acquired a functional role in gene regulation, purifying selection would have led to its perpetuation.

Self-targeting spacers frequently target non-mobilome genes

We returned to examine the targets of the 100 full-match self-targeting spacers. About half of the self-targets were found to reside within elements of putative exogenous origin such as proviral sequences, transposon sequences, and established native plasmids. The existence of an exact match to a proviral sequence indicates that this virus once infected the organism, yet managed to escape CRISPR degradation. However, it is also possible that CRISPR has a role in preventing the induction of these latent viruses, and in general has a role in preventing the expansion of mobile elements.

Nevertheless, this role cannot explain all the self-targeting spacers detected: 53 spacers from 39 different organisms were found to target genes that are unlikely to be from a mobile origin, based on their putative gene function and on their gene neighborhood. Examples include spacers targeting 16S RNA, DNA polymerase I, tRNA synthetases, and others (Table 1). If so, what might explain the existence of a CRISPR spacer against a cellular gene?

Negative effects of self-targeting spacers

One possible explanation for the acquisition of self-targeting spacers is that they represent accidents of the CRISPR insertion mechanism, potentially leading to deleterious effects on the cell. Although the average size of a CRISPR array with self-targeting spacers is 30 spacers, we found that 37% of all self-targeting spacers are located at the first or second positions in the array (near the leader sequence) – a four-fold enrichment compared to all spacers in our dataset ($P < 10^{-13}$; Fisher exact test; Table S1). Because the addition of new spacers takes place in a polarized fashion proximal to the leader end of the CRISPR array, it appears that self-targeting spacers represent recent acquisition events by the CRISPR array. This implies that self-targeting spacers survive only a short time, and are thus not selectively neutral, and could instead be deleterious to the organism.

Based on these results we hypothesized that, following the integration of a self-targeting spacer, the CRISPR/Cas system must become inactivated in order to survive. For instance, *Lactobacillus acidophilus* NCFM harbors a self-targeting spacer against 16S ribosomal RNA, which could have a high negative cost if functional. However, this organism appears to have lost all *cas* genes. Thus, the negative effect of autoimmune self-targeting spacers might explain the abundance of highly degraded CRISPR systems that contain *cas* pseudogenes [12,23] (Table 1).

We next sought to determine whether the self-targeting spacer itself could become inactivated without affecting the entire array. Studies have shown that the repeats are target sites for multiple Cas proteins, and participate in crRNA maturation and function [6,14,24]. We found that the two repeats flanking the self-targeting spacer are twice as likely to harbor mutations from the consensus repeat sequence, as compared to a background of all CRISPR spacers ($P < 0.005$, Fisher exact test; Table 1). Such mutations could potentially affect the maturation of the self-targeting spacer while leaving the rest of the array functional. Self-targeting spacers flanked by mutated repeats were found throughout the array (and not specifically at the beginning of the array), suggesting that such mutations allow the self-targeting spacer to perpetuate without any negative effect on the organism (Figure S2).

It was recently shown that CRISPR has a unique mechanism that avoids targeting the locus encoding the CRISPR itself [25]. Base-pairing between three specific bases of the upstream repeat sequence and the crRNA results in protection from CRISPR degradation. We set out to test whether targets are protected in this way from their cognate spacers. Because it is possible that different CRISPR systems have slightly different types of protection, we used a sliding window to scan whether three bases upstream or downstream from a non-mobile target match those of the repeat sequence. Our results showed that 14% and 17% of the targets match three base-pairs upstream or downstream of the repeat, respectively. Notably, these numbers do not deviate significantly from the expected number under a random binomial distribution (see supplementary material), and such base-pairing could therefore be a random property of the targets.

We next tested whether targets display proto-spacer adjacent motifs (PAMs). These motifs were first identified experimentally in *Streptococcus thermophilus* [18,26], and were later identified in a widespread computational analysis as recurring sequences adjacent to the target/proto-spacer [27]. PAMs were suggested to take part in the acquisition and/or the interference stages for some CRISPR/Cas subtypes because mutation at these sequences allowed viral escape [18,26,28]. Interestingly, PAMs appear to be unimportant for CRISPR interference in arrays associated with subtype Mtube [25] or with the Cas module RAMP [14]. Furthermore, PAMs might also be

Table 1. A list of organisms bearing self-targeting spacers against non-mobile elements

Organism	Target gene description	Putative protection from autoimmunity[a]	Associated Cas subtype
Actinobacillus pleuropneumoniae sv 3 JL03	Hydrogenase maturation factor	4	Ypest
Bifidobacterium adolescentis ATCC 15703	3-Phosphoshikimate 1-carboxyvinyltransferase	4	Dvulg
Bifidobacterium longum DJO10A	*fhu* operon transcription regulator	4	Nmeni
Campylobacter hominis ATCC BAA-381	Preprotein translocase subunit SecA		unknown
Chlorobium limicola DSM 245	Succinyl-CoA synthetase, beta subunit		Csx.RAMP
Chlorobium phaeobacteroides DSM 266	Amidophosphoribosyltransferase		unknown
Clostridium botulinum E3 Alaska E43	Amino acid carrier protein AlsT	1	RAMP[b]
Clostridium tetani E88	Putative S-layer		Hmari
Clostridium tetani E88	Stage IV sporulation protein A	4	Csx.tneap
Clostridium tetani E88	DNA mismatch repair protein	4	Csx.tneap
Clostridium tetani E88	Phosphoribosylformylglycinamidine synthase II	4	Csx.tneap
Clostridium tetani E88	Putative sporulation sigma-E factor processing	4	Csx.tneap
Enterobacter sp. 638	FdrA family protein	4	Ypest
Erwinia carotovora atroseptica SCRI1043	Putative plasmid transfer protein	2,4	Ypest
Flavobacterium psychrophilum JIP02/86	Hypothetical protein		Nmeni
Francisella philomiragia philomiragia ATCC 25017	Hypothetical protein		unknown
Frankia alni ACN14a	Putative acyl-CoA dehydrogenase	4	Ecoli
Frankia sp. CcI3	Amino acid adenylation		Dvulg
Geobacter uraniumreducens Rf4	2-methylisocitrate lyase	2	unknown
Lactobacillus acidophilus NCFM	16S ribosomal RNA	1,2,4	Ecoli[b]
Lactobacillus acidophilus NCFM	Peptide-binding protein	1,2,4	Ecoli[b]
Lactobacillus acidophilus NCFM	Dipeptidase	1,4	Ecoli[b]
Lactobacillus delbrueckii bulgaricus	Hypothetical protein	4	Ecoli
Methanospirillum hungatei JF1	Dihydroorotate dehydrogenase 1B		Csc.apern
Methylobacillus flagellatus KT	Glycosyl transferase, group 1	4	Dvulg
Mycoplasma arthritidis 158L31	Phenylalanyl-tRNA synthetase beta subunit		unknown
Mycoplasma gallisepticum R	Hypothetical protein		Nmeni
Mycoplasma gallisepticum R	ParC/GyrA; topoisomerase IV subunit A		Nmeni
Mycoplasma mobile 163K	Putative glycosyltransferase		Nmeni
Myxococcus xanthus DK 1622	Putative lipoprotein	2,4	Dvulg
Nocardia farcinica IFM 10152	Putative non-ribosomal peptide synthetase	2,4	Ecoli
Nocardia farcinica IFM 10152	Putative transcriptional regulator	4	Ecoli
Pelobacter carbinolicus DSM 2380	Histidyl-tRNA synthetase	2	Ecoli
Pelobacter propionicus DSM 2379	Hydrophobe/amphiphile efflux-1 (HAE1) family	2,4	Ecoli
Pelobacter propionicus DSM 2379	DNA topoisomerase I	2,4	Ecoli
Pelobacter propionicus DSM 2379	RND efflux system outer membrane lipoprotein	2,4	Ecoli
Porphyromonas gingivalis W83	Saccharopine dehydrogenase[c]	4	unknown
Prosthecochloris aestuarii SK413	Chromosome segregation protein SMC	2,4	Ecoli
Pyrococcus horikoshii OT3	Hypothetical protein		unknown
Roseiflexus castenholzii DSM 13941	PAS domain-containing protein	4	Mtube
Salinispora arenicola CNS-205	Hypothetical protein	4	Ecoli
Streptococcus agalactiae 2603V/R	Uracil permease	4	Nmeni
Streptococcus agalactiae A909	DNA polymerase I	4	Nmeni
Streptomyces avermitilis MA-4680	Multidrug resistance efflux protein	2,4	Ecoli
Streptomyces griseus griseus NBRC 13350	PII uridylyl-transferase	2	Ecoli
Sulfolobus tokodaii 7	Hypothetical protein	3	Apern
Treponema denticola ATCC 35405	Hypothetical protein	4	Nmeni
Yersinia pestis Antiqua	Phosphotransferase enzyme II, A component	4	Ypest
Clostridium tetani E88	Intergenic region	3	Hmari
Granulibacter bethesdensis CGDNIH1	Intergenic region	4	Ecoli
Lactobacillus acidophilus NCFM	Intergenic region	1,2	Ecoli[b]
Porphyromonas gingivalis W83	Intergenic region	3	RAMP

[a]Evidence suggesting that the target is protected from autoimmunity is marked by 1–4: (1) lack of *cas* genes, (2) mutated adjacent repeats, (3) extended base-pairing with the upstream flanking repeat, and (4) absence of PAM as defined by Mojica *et al.* [27] for some of the repeat clusters. Full details on all self-targeting spacers can be found in Table S1.

[b]For organisms with no *cas* genes, the subtype identity of the inferred lost *cas* operon was deduced based on the homologous array in a related strain, and on the identity of the repeats.

[c]Two distinct spacers exist that target this gene (see also Table S1).

less important for CRISPR interference with subtype Ecoli, based on the ability of spacers lacking a matching PAM to restrict phage infection [6]. It is, however, important to note that the role of PAM in CRISPR interference has not yet been determined conclusively for each of the Cas subtypes.

When testing whether targets display a PAM we observed an intriguing pattern. First, for the targets associated with Cas subtypes where PAM appears to be less importance for interference (as defined above), 14 of 21 (67%) targets were putatively protected by flanking mutated repeats, by base-pairing with the repeat, or by loss of the *cas* operon. By contrast, in CRISPR types that require PAM for interference, such putative protection from autoimmunity was only observed in 2 of 17 (12%) targets ($P < 0.001$; Fisher exact test). Notably, in all these

targets the PAM sequence was absent from the target. To summarize, our results tentatively suggest that when PAM is necessary for CRISPR interference, absence of a PAM sequence can protect the organism from autoimmunity, whereas when PAM is unnecessary for interference, mutations in other elements must occur to protect from the deleterious effects of CRISPR autoimmunity.

Autoimmunity in bacteria?

All in all, our analysis shows that the self-targeting CRISPR spacers are not evolutionary conserved, and that their occurrence is frequently associated with partial or full degradation of CRISPR activity. We therefore conclude that the self-targeting spacers most probably have not been selected to take part in non-transient endogenous gene regulation. We propose a model whereby CRISPR self-targeting spacers result from leaky incorporation of self nucleic acids into CRISPR arrays (Figure 1), and this could lead to a negative fitness cost to the organism. Our results suggest that some CRISPR subtypes are more prone to such leaky incorporation than others (see supplementary material online). The rate of incorporation of self-targeting spacers is at least 0.2% (Table S2). This is probably an underestimate because the calculation does not take into account highly deleterious self-targeting spacers that were immediately cleared by purifying selection or were counter-selected at the population level.

One can envisage several different mechanisms by which such leaky incorporation might occur: viruses, plasmids or transposons could harbor genes from previous rounds of infection (as occurs during lateral gene transfer) [29], and this could lead to CRISPR recognizing these genes as foreign DNA. Alternatively, faulty incorporation of self nucleic acids could occur simply because of CRISPR 'errors'. Notably, irrespective of the mechanism of acquisition, in the absence of protection it is expected that harboring a self-targeting spacer will incur a fitness cost to the CRISPR-bearing organism. This cost can be high or low, depending among other factors on the level of transcription of the self-targeting CRISPR spacer, on the mode of operation of the CRISPR array (i.e. targeting of DNA or RNA), and on the identity of the targeted gene. If this cost is relatively high, the self-targeting spacer, the targeted gene, or even the entire CRISPR/*cas* locus will be prone to

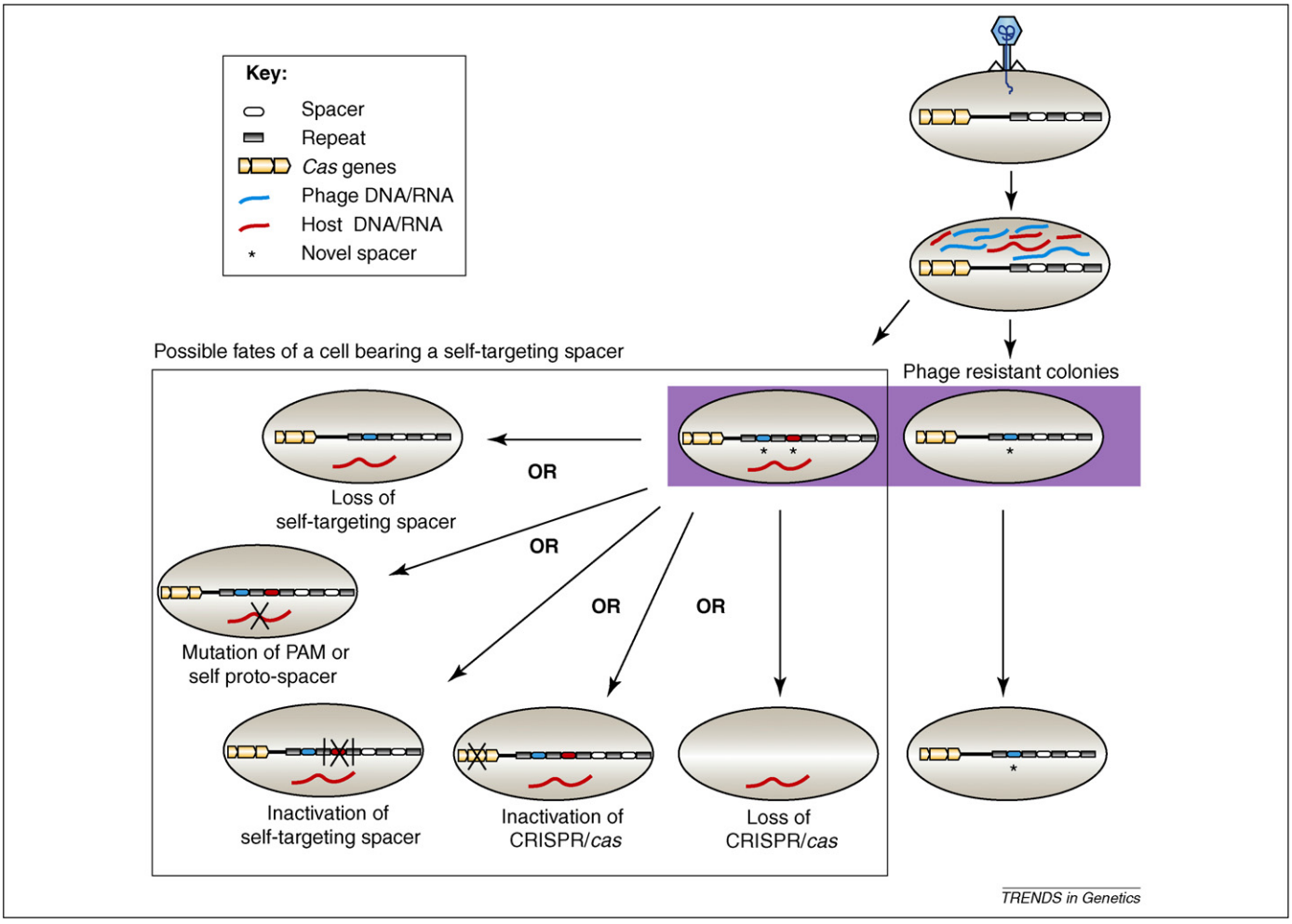

Figure 1. A model for CRISPR autoimmunity via leaky incorporation of self-DNA/RNA and its possible outcomes. During infection, genetic material in the cell will include both viral and host nucleic acids. CRISPR activity leads to insertion of new spacers, most probably derived from foreign genetic material. Virus-resistant colonies (boxed in purple) will include one or more spacers against the virus, but could also include spacers against the endogenous host DNA. Some of these colonies will not survive due to autoimmunity. However, if the fitness cost of autoimmunity is low enough, these colonies could thrive due to the beneficial existence of the anti-viral spacer. To prevent the negative effects associated with autoimmunity, inactivation of the CRISPR/*cas* locus can occur in a variety of ways, including mutation or loss of the *cas* genes, or mutations of the leader sequence. Inactivation of the self-targeting spacer could also occur by several mechanisms: mutation of the adjacent repeats, mutation of the spacer sequence, mutation of the targeted gene, or mutation of the PAM.

be lost, inactivated, or mutated. This scenario could become more likely in a virus-free environment. Interestingly, a recent study in strains of *Escherichia coli* supports the notion of such a fitness cost, showing that in strains where a spacer targeted one of the endogenous *cas* genes, the *cas* operon was lost [30].

Although the autoimmunity model is yet to be experimentally substantiated, this model might explain why a system that is so valuable for combating foreign invaders is present in only ~40% of bacteria. Together with lateral gene transfer [12,31], the model also explains the checkered pattern of existence or non-existence of CRISPR among closely related species, and the abundant occurrences of degraded CRISPR arrays (Figure 1). Nonetheless, we cannot rule out that other models exist to explain the existence of self-targeting spacers, where the spacers perform some type of transient regulation. For example, an alternative model is that CRISPR targets endogenous host genes that contribute to virus replication. Experimental validation for this model would include deleting the self-targeting spacer and testing the fitness cost to the organism with and without viral infection.

We note that our results do not preclude the notion of alternative forms of CRISPR taking part in gene regulation (see supplementary material), because our analysis focused on all identified 'typical' CRISPR structures that have multiple tandem repeats and spacers. If CRISPR had indeed evolved to perform gene regulation, the structure of the system would possibly have been altered, and could thus differ from the CRISPR system as we know it today. Such a system might be composed of only one spacer flanked by partial repeats. Such altered, regulatory CRISPR systems are yet to be discovered.

Concluding remarks

Until now the CRISPR system has been heralded as an exceptional form of defense against foreign invaders, with apparently no fitness cost to the host. However, we detected CRISPR spacers that match cellular genes with important housekeeping roles. This targeting is completely non-conserved, and thus it is proposed here to be a flaw in the CRISPR mechanism. If indeed this self-targeting induces autoimmunity, this is a striking example of the Achilles' heel of the CRISPR system.

Acknowledgements

We thank Uri Gophna, Debbie Lindell, Itay Mayrose, Sarit Edelheit, Hila Shimon, and Hila Sberro for stimulating discussions. R.S. is a European Molecular Biology Organisation (EMBO) Young Investigator. He was supported, in part, by the Israel Science Foundation Focal Initiatives in Research in Science and Technology (FIRST) program (grant 1615/09), National Institutes of Health grant R01AI082376-01, the Wolfson Family Trust miRNA research program, the Minerva foundation, and the Yeda-Sela Center for basic research. A.S. was supported by the Clore postdoctoral fellowship. O.W. was supported by the Kahn Center for Systems Biology of the Human Cell, and is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2010.05.008.

References

1 Lillestol, R.K. *et al.* (2006) A putative viral defence mechanism in archaeal cells. *Archaea* 2, 59–72
2 Barrangou, R. *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712
3 Pourcel, C. *et al.* (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653–663
4 Mojica, F.J. *et al.* (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182
5 Bolotin, A. *et al.* (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561
6 Brouns, S.J. *et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964
7 van der Oost, J. *et al.* (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* 34, 401–407
8 Karginov, F.V. and Hannon, G.J. (2010) The CRISPR System: small RNA-guided defense in bacteria and archaea. *Mol. Cell* 37, 7–19
9 Sorek, R. *et al.* (2008) CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186
10 Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170
11 Kunin, V. *et al.* (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8, R61
12 Haft, D.H. *et al.* (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 1, e60
13 Marraffini, L.A. and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843
14 Hale, C.R. *et al.* (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945–956
15 Makarova, K.S. *et al.* (2006) 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* 1, 7
16 He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 522–531
17 Horvath, P. *et al.* (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.* 131, 62–70
18 Horvath, P. *et al.* (2008) Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *J. Bacteriol.* 190, 1401–1412
19 Shah, S.A. *et al.* (2009) Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem. Soc. Trans.* 37, 23–28
20 Grissa, I. *et al.* (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8, 172
21 Kimura, M. (1983) *Neutral Theory of Molecular Evolution*, Cambridge University Press
22 Pang, K.C. *et al.* (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* 22, 1–5
23 van der Ploeg, J.R. (2009) Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* 155, 1966–1976
24 Hale, C. *et al.* (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14, 2572–2579
25 Marraffini, L.A. and Sontheimer, E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463, 568–571
26 Deveau, H. *et al.* (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390–1400
27 Mojica, F.J. *et al.* (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740

28 Semenova, E. *et al.* (2009) Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiol. Lett.* 296, 110–116
29 Partridge, S.R. *et al.* (2009) Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol. Rev.* 33, 757–784
30 Diez-Villasenor, C. *et al.* (2010) Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* 156, 1351–1361
31 Godde, J.S. and Bickerton, A. (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* 62, 718–729

0168-9525/$ – see front matter
doi:10.1016/j.tig.2010.05.008 Trends in Genetics 26 (2010) 335–340

Genome Analysis

Repeats, longevity and the sources of mtDNA deletions: evidence from 'deletional spectra'

Xinhong Guo[1,7*], Konstantin Yu. Popadin[2*], Natalya Markuzon[3], Yuriy L. Orlov[4], Yevgenya Kraytsberg[1], Kim J. Krishnan[5], Gábor Zsurka[6], Douglas M. Turnbull[5], Wolfram S. Kunz[6] and Konstantin Khrapko[1]

[1] Harvard Medical School/Beth Israel Deaconess Medical Center, Boston, MA USA
[2] Russian Academy of Sciences Institute for Information Transmission Problems and Moscow State University, Moscow, Russia
[3] Draper Laboratory, Cambridge MA, USA
[4] Genome Institute of Singapore, Singapore
[5] The Institute of Ageing and Health and Newcastle University, Newcastle upon Tyne, UK
[6] Department of Epileptology and Life&Brain Center, University Bonn Medical Center, Bonn, Germany
[7] Hunan University, Changsha, Hunan, People's Republic of China

Perfect direct repeats and, in particular, the prominent 13 bp repeat, are thought to cause mitochondrial DNA (mtDNA) deletions, which have been associated with the aging process. Accordingly, individuals lacking the 13 bp repeat are highly prevalent among centenarians and overall number of perfect repeats in mammalian mitochondrial genomes negatively correlates with species' longevity. However, detailed examination of the distribution of mtDNA deletions challenges a special role of the 13 bp repeat in generating mtDNA deletions. Instead, deletions appear to depend on long and stable, albeit imperfect, duplexes between distant mtDNA segments. Furthermore, significant dissimilarities in breakpoint distributions suggest that multiple mechanisms are involved in creating mtDNA deletions.

Direct repeats in the mitochondrial genome and longevity

The premise that accumulation of mtDNA mutations [1] and, in particular, of large-scale deletions in mtDNA is one of the possible causes of aging has received substantial support from biochemical and longevity studies [2–5]. mtDNA deletions are usually flanked by direct repeats, implying that these repeats are involved in the generation of deletions. Recombination [6,7], slip-replication [8], and double-strand break repair [9] have been suggested as potential alternative mechanisms involving direct repeats. In corroboration of the connection between mtDNA deletions and aging, the number of direct repeats in mtDNA of various mammal species is inversely correlated with longevity [10,11]. Of particular interest is the so-called 'common deletion' [6], the deletion most frequently detected in humans, which is flanked by a prominent 13 bp perfect direct repeat. Interestingly, carriers of the well-studied D4a mitochondrial haplogroup, who are significantly enriched among Japanese centenarians [12], lack the 13 bp direct repeat in their mtDNA, and thus presumably lack the common deletion, and this seems to support the premise that deletions are involved in the aging process [13]. It should be noted, however, that although the 'common' deletion is the most abundant mtDNA deletion, it typically constitutes no more than 10% of all deletions in aging tissues [5]. Therefore D4a individuals would have at most 10% fewer deletions, which perhaps is too moderate a change to affect longevity. There is another possibility, however [13]. According to an elegant hypothesis of Samuels, Schon and Chinnery, the 13 bp repeat might be responsible for the formation of nearly all mtDNA deletions, not just the common deletion [14]. Thus, absence of this repeat could result in a reduction of overall deletion burden and, if deletions indeed are involved in the aging process, this might constitute a realistic cause of exceptional D4a longevity. The importance of this question prompted us to test the Samuels, Schon, Chinnery hypothesis.

mtDNA deletions, in general, are not related to the 13 bp repeat

The Samuels, Schon, Chinnery hypothesis rests on the observation that the distribution of deletion breakpoints across the mitochondrial genome consists of two broad peaks centered around the 5' and 3' arms of the 13 bp

Corresponding author: Khrapko, K. (kkhrapko@gmail.com).
* X.G. and K.P. contributed equally to this work.