RESEARCH HIGHLIGHT

New CRISPR-Cas systems discovered

Cell Research (2017) 27:313-314. doi:10.1038/cr.2017.21; published online 21 February 2017

In bacteria and archaea, CRISPR-Cas adaptive immune systems utilize RNA-guided endonucleases to defend against invasion by foreign nucleic acids of bacteriophage, virus and plasmid origin. In a recent paper published in *Nature*, Burstein *et al.* identified the first Cas9 protein in uncultivated archaea and two novel CRISPR-CasX and CRISPR-CasY systems in uncultivated bacteria by capitalizing on analysis of terabase-scale metagenomic datasets from natural uncultivated organisms.

Bacteria and archaea have developed RNA-guided adaptive defense systems to protect themselves against invaders such as foreign nucleic acids, which are known as CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) systems. The CRISPR-Cas immunity pathway can be divided into three stages: spacer acquisition, crRNA (CRISPR RNA) biogenesis and target interference. The divergent Cas proteins and crRNA are encoded by the CRISPR locus, which is composed of variable *cas* genes and a CRISPR array comprising short direct repeats separated by spacers. Based on the architecture of the interference modules, the CRISPR-Cas systems can be divided into two classes that can be further subdivided into 6 types and 16 subtypes: Class 1 systems (Type I, III and IV), which employ a multi-subunit effector complex termed Cascade (CRISPR-associated complex for antiviral defense), are formed by multiple Cas proteins and crRNA. By contrast, Class 2 systems (Type II, V and VI), which employ crRNA and a single large multi-domain effector protein, can in

certain cases also require tracrRNA (trans-activating crRNA), as observed for Type II and Type V-B systems [1, 2].

The three steps of CRISPR immunity are mediated by combinations of distinct Cas proteins [1, 2]. Generally, Cas1 and Cas2 proteins, supplemented by Cas4 (Type II-B and Cas4 homolog in Type II-C), Csn2 (Type II-A) and Cas9 (Type II-A), are involved in the spacer acquisition step required for processing and integrating the fragments of foreign DNA (protospacers) into the CRISPR array as new spacers. During the expression and processing step of crRNA, either Cas6 endoribonuclease in Cascade (Type I and Type III), or Cas9 in cooperation with tracrRNA and RNAase III (Type II), or Cpf1 (Type V-A) [3], or C2c2 (Type VI) [4, 5] processes the pre-crRNAs into mature crRNAs. In the interference step, the crRNA-effector complexes recognize and bind with target DNA or RNA containing sequences complementary to the spacer sequence of the crRNA, with recognition also extending to the PAM (protospacer adjacent motif) in target DNA or PFS (protospacer flanking site) in target RNA (except that Type III system lacks a PAM). Cleavage of target DNA in multi-subunit Class 1 systems is achieved by either HD nuclease and helicase domains of Cas3 (Type I) or through cooperation of Cas7 and HD nuclease domain of Cas10 (Type III). Cleavage of target DNA in single effector Class 2 systems is achieved either by RuvC and HNH domains of Cas9 (Type II) or RuvC domain of Cpf1/C2c1 (Type V-A/B). In addition, cleavage of target RNA in single effector Class 2 systems is achieved by a pair of HPEN domains

of C2c2 (Type VI). To date, two class 2 CRISPR-Cas effectors, Cas9 (Type II) and Cpf1 (Type V-A) have been successfully applied to genome engineering in a variety of cell types and organisms [2, 6, 7].

All these studies of CRISPR-Cas systems have been based on cultivated organisms, while little is known about related adaptive immune systems in uncultivated microbes. Now, in a recent paper published in *Nature*, Burstein *et al.* [8] identified novel CRISPR-Cas systems in uncultivated archaea and bacteria. Previous studies show that archaea lacks RNAase III, the nuclease responsible for the processing of pre-crRNA, indicative of the possibility that the Type II CRISPR-Cas system is perhaps absent in archaea [9]. The authors analyzed acid-mine drainage (AMD) metagenomic datasets and identified the first Type II Cas9 proteins in two nanoarchaea (designated ARMAN-1 and ARMAN-4). By reconstructing alternative ARMAN-1 CRISPR-Cas arrays from AMD samples, Burstein *et al.* found the spacer content to be hypervariable, indicating that the ARMAN-1 CRISPR-Cas system is active in these samples. Besides, they identified a 3'-NGG PAM in this system. The ARMAN-4 CRISPR-Cas system lacks the typical CRISPR array and *casI* gene, which implies that this system may function via a single spacer. To better understand archaeal Cas9 proteins, Burstein *et al.* identified putative tracrRNA for both ARMAN-1 and ARMAN-4 CRISPR-Cas systems and checked for cleavage activity. However, the ARMAN-1 and ARMAN-4 Cas9 proteins show no detectable cleavage

activity in both *in vitro* biochemical assays and *in vivo* *E.coli* targeting assays. The authors proposed that either post-translational modifications or co-factor participation might be required for activation of these Cas9 proteins.

Burstein *et al.* also identified two new CRISPR-Cas systems in bacteria, designated CRISPR-CasX and CRISPR-CasY. In the CRISPR-CasX system, the CRISPR array encodes *cas1*, *cas2*, *cas4*, and *casx* genes. Interestingly, they identified that the CasX protein contains a RuvC domain in the C-terminus and lacks the HNH domain, features similar to the Type V system. The domain organization of the other regions of CasX remains currently unknown. They also revealed that CasX recognized a 5'-TTCN PAM and employs both crRNA and tracrRNA for dual-RNA-guided DNA cleavage. The CRISPR-CasY system shows unique features that are different from any other reported CRISPR-Cas systems. In the CRISPR-CasY system, most of the CRISPR arrays contain 17-19 nt spacers, which are shorter than those observed for all other reported systems. Burstein *et al.* also analyzed the domain organization of six CasY proteins. Four CasY proteins share similarity in the RuvC domain to C2c3 whereas two CasY proteins share no detectable similarity to any other known proteins. In addition, they also identified a 5'-TA PAM and showed that the CRISPR-CasY systems exhibit RNA-guided DNA cleavage activity. However, whether CasY proteins utilize tracrRNA for cleavage remains to be determined.

The work by Burstein *et al.* discussed

above resulted in the identification of three new CRISPR-Cas systems that exhibit unique and distinct features relative to earlier reported CRISPR-Cas systems. Their findings expand on the diversity of reported CRISPR-Cas systems and also raise many questions requiring further investigation. How do archaeal CRISPR-Cas systems process the pre-crRNA to mature crRNA without RNAase III? What is the missing component(s) that accounts for the inactivity of these archaeal Cas9 proteins? In this regard, recent studies have reported several inhibitors for Class 1 Type I-F (Csy Cascade) [10, 11], and Class 2 Type II (Cas9) [12, 13] and Type V-B (Cas13b) systems, as well as an activator for Type V-B system (Cas13b) [14], suggesting that there might also be certain regulator(s) or cofactor(s) required for cleavage in archaeal CRISPR-Cas systems. Alternatively, auto-inhibitory processes may control the cleavage activity of these archaeal Cas9 proteins. Given that the *cas1* gene is missing in ARMAN-4 CRISPR-Cas system, does this system use an alternative approach to achieve spacer acquisition? Is there a Cas1 homolog or some as yet unknown protein function involved in spacer acquisition? The functional implications of the short spacers in CRISPR-CasY systems also remain to be uncovered. Given the novel domain architectures of these Cas proteins, gaps remain in our understanding of the mechanisms underlying PAM and target DNA recognition and pre-crRNA processing, challenges which will require further studies. The phylogenies of these systems also remain to be established and will

require characterization. Finally, a lot of basic and translational research remains to be done to enhance the capabilities of these systems to serve as powerful and cutting-edge tools for genome editing and other related biotechnological and biomedical applications.

Hui Yang¹, Dinshaw J Patel¹

¹Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, 10065, USA

Correspondence: Hui Yang^a, Dinshaw J Patel^b

^aE-mail: yangh3@mskcc.org

^bTel: +1-212-639-7207; Fax: +1-212-715-3066

E-mail: pateld@mskcc.org

References

- 1 Makarova KS, Wolf YI, Alkhnbashi OS, *et al.* *Nat Rev Microbiol* 2015; **13**:722-736.
- 2 Wright AV, Nunez JK, Doudna JA. *Cell* 2016; **164**:29-44.
- 3 Fonfara I, Richter H, Bratovic M, *et al.* *Nature* 2016; **532**:517-521.
- 4 East-Seletsky A, O'Connell MR, Knight SC, *et al.* *Nature* 2016; **538**:270-273.
- 5 Abudayyeh OO, Gootenberg JS, Konermann S, *et al.* *Science* 2016; **353**:aaf5573.
- 6 Sternberg SH, Doudna JA. *Mol Cell* 2015; **58**:568-574.
- 7 Zetsche B, Gootenberg JS, Abudayyeh OO, *et al.* *Cell* 2015; **163**:759-771.
- 8 Burstein D, Harrington LB, Strutt SC, *et al.* *Nature* 2016; **542**:237-241.
- 9 Deltcheva E, Chylinski K, Sharma CM, *et al.* *Nature* 2011; **471**:602-607.
- 10 Bondy-Denomy J, Garcia B, Strum S, *et al.* *Nature* 2015; **526**:136-139.
- 11 Pawluk A, Staals RH, Taylor C, *et al.* *Nat Microbiol* 2016; **1**:16085.
- 12 Rauch BJ, Silvis MR, Hultquist JF, *et al.* *Cell* 2017; **168**:150-158.
- 13 Pawluk A, Amrani N, Zhang Y, *et al.* *Cell* 2016; **167**:1829-1838.
- 14 Smargon AA, Cox DB, Pyzocha NK, *et al.* *Mol Cell* 2017 Jan 4. doi:10.1016/j.molcel.2016.12.023