JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CRISPR/CAS TECHNOLOGY: PRINCIPLE AND CLASSIFICATION

Sowmya K L[1], Dr. Ramalingappa B[2]

1. Research Scholar, Department of Microbiology

2. Professor, Department of Microbiology

Department of Microbiology,

Davangere University, Shivagangothri,

Davangere-577007, Karnataka, India

E-mail: swomyakl456@gmail.com

Abstract: CRISPR is an indispensable tool in biological research. Recent advances in genome engineering technologies based on the CRISPR- associated RNA- guided endonuclease Cas 9 are enabling the systematic interrogation of mammalian genome function. Analogous to the search function in modern word processors, Cas9 can be guided to specific locations within complex genomes by a short RNA search string. Using this system, DNA sequences within the endogenous genome and their functional outputs are now easily edited or modulated in virtually any organism of choice. Cas9-mediated genetic perturbation is simple and scalable. Once known as the bacterial immune system against invading viruses, the programmable capacity of the Cas9 enzyme is now revolutionizing diverse fields of medical research, biotechnology, and agricultural. In this review, we describe the development and applications of Cas9 for a variety of research or translational applications, highlighting the potential of this technology for cancer treatment as well as future and prospective of CRISPR.

KEY WORDS: CRISPR-Cas system , CAS gene nomenclature. Type I,II,III and IV.

1. Introduction

CRISPR/CAS9 is a gene editing tool that can manipulate gene expression in plants, humans and animals. Discovery of the Clustered Regularly-Interspaced Short Palindromic Repeats(CRISPR), their function as part of an adaptive prokaryotic immune system *(CRISPR*- associated system, Cas), and subsequent development into a genomic editing tool, hasrevolutionized the field of molecular biology. Much of this enthusiasm centers on the clinical potential of CRISPR/Cas9 for treating human disease and editing the human genome. However, the simplicity and specificity with which CRISPR/Cas9 can edit DNAis changing the pace of biological research in many areas, including identifying and understanding mechanisms of genetic diseases, validating disease targets, developing animal disease models, facilitating genetic engineering in plants, and allowing for more thorough epigenetic studies. This broad impact of the CRISPR/Cas9 gene editing tool has led to over 6000

research publications since its development five years ago (Timlin *et al*., 2018).

The action of the CRISPR-Cas system is usually divided into three stages: (1) adaptation or spacer integration, (2) processing of the primary transcript of the CRISPR locus (pre-crRNA) and maturation of the crRNA which includes the spacer and variable regions corresponding to 5′ and 3′ fragments of CRISPR repeats, and (3) DNA (or RNA) interference. Two proteins, Cas1 and Cas2, that are present in the great majority of the known CRISPR-Cas systems are sufficient for the insertion of spacers into the CRISPR cassettes. These two proteins form a complex that is required for this adaptation process; the endonuclease activity of Cas1 is required for spacer integration whereas Cas2 appears to perform a nonenzymatic function. The Cas1-Cas2 complex represents the highly conserved "information processing" module of CRISPR-Cas that appears to be quasi-autonomous from the rest of the system. The second stage, the processing of pre-crRNA into the guide crRNAs, is performed either by a dedicated RNA endonuclease complex or via an alternative mechanism that involves bacterial RNase III and an additional RNA species. The mature crRNA is bound by one (type II) or several (types I and III) Cas proteins that form the effector complex, which targets the cognate DNA or RNA. The effector complex of type I systems is known as Cascade (CRISPR-associated complex for antiviral defense). (Koonin et al., 2018).

2. The principle of the CRISPR genome engineering tool

Over the past decades, genome editing technologies have been composed of zinc-finger nucleases (ZFNs) and transcriptional activator-like effector nucleases (TALENs), empowering scientific results at both the basic and clinical level. Despite the advances that have been reported in the field of genomic engineering, the use of ZNFs or TALENs nucleases is associated with several obstacles. For example, the design for genomic engineering techniques remains complex, and therefore, these techniques cannot modulate the expression of multiple target genes. The principle in using ZNFs and TALENs is protein-based and the associated toxicity is very high, thus prompting researchers to uncover a novel genome engineering tool (Bratovic *et al*.,2020). A novel RNA-guided endonuclease-relied genome editing technology that was termed the clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9) system, markedly altered the landscape of genomic engineering. The story began with the study of the immune system in bacteria and archaea in an attempt to elucidate the mechanisms through which these organisms combat viral infection. In native context, it was found that CRISPR in combination with Cas protein provide bacteria with immunity against infections. Specifically, it was shown that the role of repeats was to recognize mobile genetic elements (MGEs), and thus it was possible to cut them into small sequences and integrate them as spacers into the genome of bacteria. That approach was based on the microbial immune system that used RNA-guided nuclease to recognize and cleave foreign genetic elements (Xin *et al*.,2019). In 2012, an adaptation of the prokaryotic immune system in mammalian cells as a gene editing tool was simultaneously reported for the first time by four different research groups, causing a certain debate regarding the intellectual rights of this innovative technique. The newly engineered CRISPR system consisted of two components: A chimeric single-guide RNA(sgRNA) that provided target specificity and Cas9 that acted as a helicase and anuclease in order to unwind and cut the target DNA. In this system, the only restriction for the targeting of a specific locus was the protospacer adjacent motif (PAM) sequence ('NGG' in the case of SpCas9). The CRISPR system was further simplified, based on its ability to interfere with and participate in bacterial adaptive immunity, comprising Cas nuclease and single-guide RNA (sgRNA). In general, the CRISPR system main mechanism of action is mediated by the Cas nuclease, which interacts with DNA and generates double strand breaks (DSBs) in the DNA sequence, and also matches the broken genomic region with a sgRNA. The sgRNA is a chimeric RNA, which consists of programmable CRISPR RNA (crRNA) and a trans-activating RNA (tracrRNA). Specifically, the CRISPR-Cas system includes a cluster of proteins, categorized into Class 1 (Types I,III and IV) and Class 2 (Types II, V, VI),

all of which constitute specific RNA-guided DNA endonuclease proteins (Cas). Cas proteins are driven by RNA and not by other proteins,to recognize the desired DNA sequence. The Class 2 subtype of the CRISPR system, which generally exploits Cas9 nuclease, is usually selected. The 100 bp sgRNA forms complementary bonds with the target DNA sequence of 17-20 nucleotides, via Watson-Crick base pairing, and the rRNA is the component which Cas9 nuclease binds to. Specifically, the sgRNA recognizes the target sequence, which is located upstream of the triplicate sequence named PAM, given that the PAM motif recruits Cas9 nuclease atsite of DNA cleavage. Of note, the PAM sequence plays the determinant role in recognizing the correct DNA sequence and in preventing the direction of RNA to self- targets and non-specific sequences. This is possible as repeats of the CRISPR system do not involve PAM and the orientation of Cas9 depends on the PAM sequence. Overall,the genomic sequence of 14 nucleotides defines the target at which Cas9 nuclease exerts its effects. More specifically, this sequence is composed of 12 nucleotides of sgRNA in conjunction with two nucleotides of protospacer adjacent motif. Notably, there is a wide range of PAM sequences depending on their origin. In the case of Cas9 derived from Streptococcus pyogenes, the motif of the PAM sequence may be composed of any base, followed by two additional guanine bases (Zeng *et al*., 2020).

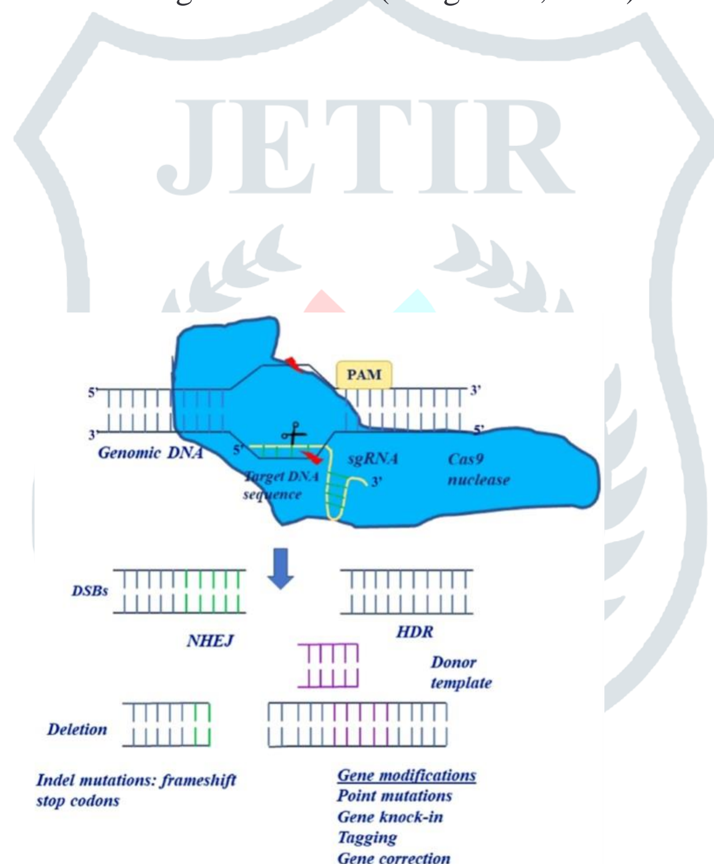

Figure 2.1 CRISPR system mechanism of action. The main action of the CRISPR system is mediated by the Cas nuclease. This nuclease is recruited to DNA by the orientation of the PAM motif and generates double-strand breaks in DNA sequence, matching the broken genomic region with a single guide RNA. Following this, non-homologous end joining or homologous mediated repair mechanisms are conducted to restore the nucleotide sequence induced by double-strand breaks, causing the anticipated genomic alterations. CRISPR, clustered regularly interspaced short palindromic repeats; PAM, protospacer adjacent motif; DSBs, double-strand breaks; NHEJ, non-homologous end joining; HDR, homology directed repair.

3. Classification of CRISPR-Cas Systems

Considering technical problems with the sensitivity and selectivity of Cas proteinfamily profiles and uncertainties of Cas1 phylogeny and CRISPR-Cas subtype classification described below, fully

automated identification of CRISPR-Cas subtypes in general is not currently feasible. The best approach to ensure the correct classificationis to combine several sources of information such as Cas1 phylogeny, identification, and annotation of as many Cas proteins as possible in the locus in question and, for type II systems, identification of the trans-activating crRNA (tracrRNA) genes. Table 1 provides a description of the key features of each subtype that help in the classificationof the CRISPR-Cas systems.

Extra caution should be exercised when introducing new gene names and new subtypes, because, due to the often extreme sequence divergence of the Cas proteins, the similarity with already defined genes and subtypes can easily be overlooked. Furthermore, the abundance of associated genes that are likely to represent(quasi)independent immunity mechanism and are only loosely linked to CRISPR-Cas loci requires extra evidence to assign new Cas names. Extra caution should be exercised when introducing new gene names and new subtypes, because, due to the often extreme sequence divergence of the Cas proteins, the similarity with already defined genes and subtypes can easily be overlooked. Furthermore, the abundance of associated genes that are likely to represent (quasi) independent immunity mechanismand are only loosely linked to CRISPR-Cas loci requires extra evidence to assign new Cas names. (Koonin *et al.,* 2018).

3.1 Major Types of CRISPR-Cas Systems, Their Subtypes, and Cas Gene Nomenclature

The top level of the current CRISPR-Cas classification hierarchy includes the three major types (I, II, and III) and the less common but clearly distinct type IV. The distinction between the CRISPR-Cas types is based on the respective signature genes and the typical organization of the respective loci.

3.1.1 Type I CRISPR-Cas Systems

All type I loci contain the signature gene cas3 which encodes a large protein witha helicase possessing a single-stranded DNA (ssDNA)-stimulated ATPase activity coupled to unwinding of DNA-DNA and RNA-DNA duplexes. Often, but not always, the helicase domain is fused to an HD family domain which has an endonuclease activity and is involved in the cleavage of the targeted DNA. Exonuclease (3′–5′) activity on single-stranded DNAs and RNAs has also been reported for the HD domain. The HD domain is located at the N-terminus of Cas3 proteins or is encoded by a separate gene within the same locus as cas3 helicase. In the latter case, the helicase is denoted cas3′ and the HD nuclease is denoted cas3″. In type I-F systems, cas3 is additionally fused to the cas2 gene (Bolotin *et al*., 2005).

Usually type I systems are encoded by a single operon containing the cas1 and cas2 genes, genes for the subunits of the Cascade or effector complex, including large subunit, small subunit (often fused to the large subunit), cas5 and cas7 genes, and cas6gene that is directly responsible for pre-crRNA transcript processing. Each gene in the type I system operons is usually present in a single copy. Several exceptions foreffector complex organization are described in Table 1 and below in the text. Type I systems are currently divided into six subtypes, I-A to I-F, each of which has its own signature gene and distinct features of operon organization (Table 3.1). Unlike other subtypes, I-E and I-F lack the cas4 gene. These subtypes are related according to the Cas1 phylogeny. Subtypes I-A, I-B, I-C, I-E, and I-F mostly correspond to the originally proposed ones, with the exception of Hmari and Tneap subtypes that were combined into subtype I-B. Recently, other diverged variants of several subtypes have beenidentified; these, however, share several features with the existing subtypes and

thus still could be described within existing classification, e.g., several type I-C variants and aderived type I-F variant. In addition, the number and diversity of stand-alone (not associated with cas1 cas2 gene pair) effector complexes are growing. These “solo” effector complexes are often present on plasmids and/or associated with transposon- related genes, such as TniQ/TnsD, a DNA-binding protein required for transposition. Many such cases are derivatives of subtype I-F and some others (e.g., Ava_3490- Ava_3493 Anabaena variabilis ATCC 29413, with genes encoding Cas6, Cas8, Cas5, Cas7) are derivatives of subtype I-C. If a system includes a derived variant of a known Cas protein family, this family might have an optional suffix indicating the subtype to which this protein belongs (e.g., Cas6f is a highly derived member of Cas6 superfamily specific for subtype I-F). Notably, the phylogenetic tree of the type I signature protein Cas3 seems to accurately reflect the subtype classification. (Deltcheva *et al.,*2018).

3.1.2 Type II CRISPR-Cas Systems

The signature gene for type II CRISPR-Cas systems is cas9, which encodes a multidomain protein that combines all the functions of effector complexes and the target DNA cleavage and is essential for the maturation of the crRNA. The type II systems are also known as the “HNH” systems, Streptococcus-like or Nmeni subtype. Every CRISPR-Cas locus of this subtype, in addition to the cas9 gene, also contains theubiquitous cas1 and cas2 genes. In addition to these three protein-coding genes, the vast majority of type II loci also encompass one or two genes for tracrRNA, an RNA thatis partially homologous to the cognate CRISPR. These systems use cellular (notencoded within the CRISPR- Cas loci) RNase III and tracrRNA for the processing of pre-crRNA. The large Cas9 protein (~800–1,400 amino acids) contains two nuclease domains, namely the RuvC-like nuclease (RNase H fold) and the HNH (McrA-like)nuclease domain that is located in the middle of the protein. Both nucleases are required for target DNA cleavage (Koonin *et al*., 2018).

Type II CRISPR-Cas systems are currently classified into three subtypes (II-A, II-B,and II-C), two of which were introduced in the updated classification and one was proposed recently on the basis of the distinct operon organization. Type II-A systems encompass an additional gene, known as csn2 (Fig. 1), which is considered a signature gene for this subtype. The Csn2 protein is not required for interference but apparently has an unclear role in spacer integration. The Csn2 proteins form homotetrameric ringsthat bind linear double-stranded DNA through the central hole. This protein has been shown to adopt a highly derived P-loop ATPase fold in which the ATP- binding site appears to be inactivated. Several highly diverged Csn2 subfamilies have been identified,in particular short and long forms for which structures and biochemical characterizationare available. Type II-B systems do not encode the csn2 gene but possess a distinct fourth gene that belongs to the Cas4 family which is also associated with subtypes I-A to I-D (but not I-E and I-F). The Cas4proteins possess 5′-single-stranded DNA exonuclease activity and belong to the PDEDxK family of nucleases. The actual role of the Cas4 proteins in the CRISPR-Cas systems remains unknown. The recently proposedtype II-C CRISPR-Cas systems possess only three protein-coding genes (cas1, cas2, andcas9) and are common in sequenced bacterial genomes. Recently, type II systems havebeen developed into a powerful genome editing and engineering tool with a major biotechnological potential (Brueghton *et al.,* 2018).

3.1.3 Type III CRISPR- Cas Systems

All type III systems possess the signature gene cas10 which encodes a multidomain protein containing a palm domain similar to that in cyclases and polymerases of the PolB family. Thus, this protein originally was predicted to be a polymerase (Akinsheye *et al*.,2011). Recently, the structure of Cas10 has been solved and four distinct domainshave been identified: the N-terminal cyclase-like domain that adopts the same RRM foldas the palm domain but is not predicted to possess enzymatic activity, a helical domaincontaining the Zn-binding treble clef motif, the palm domain that retains the catalytic residues and is predicted to be active, and the C-terminal alpha helical domain resembling the thumb domain of A-family DNA polymerase and Cmr5, a small alpha helical protein present in some of the type III CRISPR-Cas systems. Cas10 is the large subunit of effector complexes of type III systems. Each type III locus also encodes othersubunits of effector complexes such as one gene for the small subunit, one gene for a Cas5 group RAMP protein, and usually several genes for RAMP proteins of the Cas7 group. Often Cas10 is fused to an HD family nuclease domain that is distinct from the HD domains of type I CRISPR-Cas systems and, unlike the latter, contains a circular permutation of the conserved motifs. Type III CRISPR-Cas systems often do not encodetheir own cas1 and cas2 genes but use crRNAs produced from CRISPR arrays associated with type I or type II systems. Nevertheless, in many genomes that lack typeI and type II systems, cas1, cas2, and cas6 genes are linked to a type III system that accordingly is predicted to be fully functional. Currently, there are two subtypes within type III, III-A (former Mtube subtype or Csm module), and III-B (former RAMP module orCmr module), which are clearly related but could be distinguished by the presence of distinct genes for small subunits of effector complexes, csm2 and cmr5, respectively. The subtype III-A loci often possess cas1, cas2, and cas6 and have been shown to target DNA, whereas most of the III-B systems lack these genes and therefore depend on other CRISPR-Cas systems present in the same genome. The type III-B CRISPR-Cassystems have been shown to target RNA. All type III systems possess the signature gene cas10 which encodes a multidomain protein containing a palm domain similar to that in cyclases and polymerases of the PolB family. Thus, this protein originally was predicted to be a polymerase. Recently, the structure of Cas10 has been solved and fourdistinct domains have been identified: the N-terminal cyclase-like domain that adopts the same RRM fold as the palm domain but is not predicted to possess enzymatic activity, a helical domain containing the Zn-binding treble clef motif, the palm domain that retains the catalytic residues and is predicted to be active, and the C-terminal alpha helical domain resembling the thumb domain of A-family DNA polymerase and Cmr5, asmall alpha helical protein present in some of the type III CRISPR-Cas systems. Cas10 isthe large subunit of effector complexes of type III systems. Each type III locus also encodes other subunits of effector complexes such as one gene for the small subunit, one gene for a Cas5 group RAMP protein, and usually several genes for RAMP proteins of the Cas7 group. Often Cas10 is fused to an HD family nuclease domain that is distinct from the HD domains of type I CRISPR-Cas systems and, unlike the latter, contains a circular permutation of the conserved motifs. Type III CRISPR-Cas systems often do not encode their own cas1 and cas2 genes but use crRNAs produced from CRISPR arrays associated with type I or type II systems. Nevertheless, in many genomes that lack type I and type II systems, cas1, cas2, and cas6 genes are linked to atype III system that accordingly is predicted to be fully functional. Currently, there are two subtypes within type III, III-A (former Mtube subtype or Csm module), and III-B (former RAMP module or Cmr module), which are clearly related but could be distinguished by the presence of distinct genes for small subunits of effector complexes, csm2 and cmr5, respectively. The subtype III-A loci often possess cas1, cas2, and cas6 and have been shown to target DNA, whereas most of the III-B systemslack these genes and therefore depend on other CRISPR-Cas systems present in the same genome. The type III-B CRISPR-Cas systems have been shown to target RNA (Bibikova *et al.,* 2003).

3.1.4 Type IV CRISPR- Cas Systems

Type IV CRISPR-Cas systems, found in several bacterial genomes. Similar to subtype III-A, this system lacks cas1 and cas2 genes and often is not associated with CRISPR arrays. Moreover, in many bacteria, this is the only CRISPR-Cas system, with no CRISPR cassette detectablein the genome. The type IV systems possess an effector complex that consists of a highly reduced large subunit (csf1), two genes for RAMP proteins of the Cas5 (csf3) andCas7 (csf2) groups, and, in some cases, a gene for a predicted small subunit. The csf1 gene could be considered a signature gene for this system. There are two distinct subtypes of type IV systems, one of which contains a DinG family helicase csf4, whereas the second subtype lacks DinG but typically contains a gene for a small alpha helical protein, presumably a small subunit. Type IV CRISPR-Cas systems could be mobile modules that, similar to type III systems, could utilize crRNA from different CRISPR arrays once these become available (Cohen *et al*.,2019).

The classification of CRISPR-Cas systems outlined above more or less adequately covers the representation of these systems in sequenced bacterial and archaeal genomes. However, considering the rapid evolution of CRISPR-Cas, these variants might represent only the proverbial a tip of the iceberg with respect to the true diversity of prokaryotic adaptive immunity. Based on some marginal similarities, these loci could be tentatively assigned to type I and type III, respectively; however, they do not contain anysignature genes described above that would allow one to classify them into any known subtype. Similarly, classification of certain type I systems, is hampered by the apparent absence of signature genes of the known type I subtypes. Accumulation of such "unclassifiable" variants raises the possibility that the current principles of CRISPR- Cas system classification might have to be reconsidered to take into account the challenge of the ever-increasing diversity. (Koonin *et al.,* 2018).

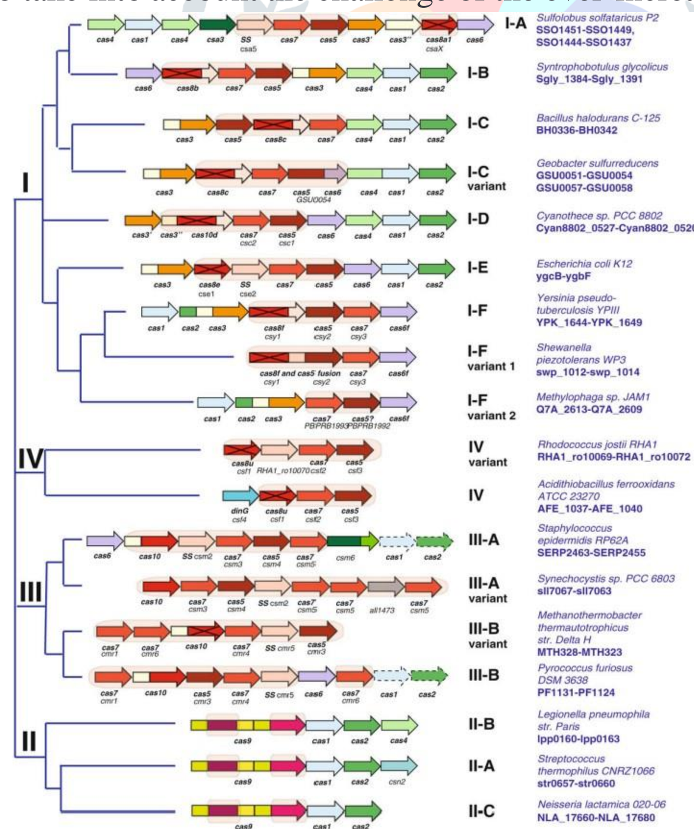

Figure 3.1 Classification and organization of CRISPR-Cas systems. Typical operon organization is shownfor each CRISPR-Cas subtype. For each CRISPR-Cas subtype, a representative genome and the respectivegene locus tag names are indicated. Homologous genes are color coded and identified by a family name.Names follow the classification from [5]. See also details in [30]. Names in bold are proposed systematicnames; "legacy names" are in regular font. Abbreviations: LS large subunit (including subfamilies of Cas10, Cas8, Cse1, Csy1), SS small subunit (including Cmr2, Cmr5, Cse2). Genes coding for inactivated large subunits

are indicated by crosses. Genes and domain components for effector complexes arehighlighted by pink background.

CONCLUSIONS

The CRISPR-Cas system is a unique technology for gene editing. Studies summarized in this review represent only the first steps in the CRISPR-Cas era of genetic engineering. The number and diversity of known CRISPR-Cas systems have substantially increased in recent years. Here we provide an updated evolutionary classification of CRISPR-Cas systems and Cas genes.

References

1. Abbott TR, Dhamdhere G, Liu Y, Lin X, Goudy L, Zeng L.2020. Development of CRISPR as an antiviral strategy to combat SARS-CoV-2 and influenza. *Cell*.
2. Akinsheye I, Alsultan A, Solovieff N, Ngo D, Baldwin CT, Sebastiani P. 2011. Fetal hemoglobin in sickle cell anemia. *Blood*.
3. Amer MH. Gene therapy for cancer: present status and future perspective.2014. *Mol Cell Ther*.
4. Anders C, Bargsten K, Jinek M. 2016. Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. *Mol Cell*.
5. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA.Nature.
6. Basak A, Sankaran VG. 2016.Regulation of the fetal hemoglobin silencing factor BCL11A. *Ann NY Acad Sci*.
7. Bibikova M, Beumer K, Trautman JK, Carroll D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science*.
8. Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici M. 1995. T lymphocyte-directed gene therapy for ADASCID: Initial trial results after 4 years. *Science*.
9. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*.
10. Bordignon C, Notarangelo LD, Nobili N, Ferrari G, Casorati G, Panina P. Gene therapy in peripheral blood lymphocytes and bone marrow for ADA- immunodeficient patients. *Science*.
11. Bratovic M, Fonfara I, Chylinski K, Gálvez EJC, Sullivan TJ, Boerno S. 2020. Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat Chem Biol*.
12. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*.
13. Broughton JP, Deng X, Yu G, Fasching CL, Servellita V, Singh J. 2020. CRISPR–Cas12-based detection of SARS-CoV-2. *Nat Biotechnol*.
14. Bruegmann T, Deecke K, Fladung M. 2019. Evaluating the efficiency of gRNAs in CRISPR/Cas9 mediated genome editing in poplars. *Int J Mol Sci*.

15. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G. 2018. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol*.

16. Cavazzana-Calvo M, Hacein-Bey S, De Saint Basile G, Gross F, Yvon E, Nusbaum P. 2000. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science*.

17. Chandrasegaran S, Carroll D. 2016. Origins of programmable nucleases for genome engineering. *J Mol Biol.*

18. Chatterjee P, Jakimo N, Jacobson JM. 2018. Minimal PAM specificity of a highly similarSpCas9 ortholog. *Sci Adv*.

19. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB. 2017. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*.

20. Cohen J. 2019. Did CRISPR help—or harm—the first-ever gene-edited babies? *Science*.

21. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*.

22. Cong L, Zhang F.2017. Genome engineering using CRISPR-Cas9 system. *MolBiol*.

23. Cui Y, Xu J, Cheng M, Liao X, Peng S. 2018. Review of CRISPR/Cas9 sgRNA design tools. *Interdiscipl Sci Comput Life Sci.*

24. Cullot G, Boutin J, Toutain J, Prat F, Pennamen P, Rooryck C. 2019. CRISPR- Cas9genome editing induces megabase-scale chromosomal truncations. *Nat Commun.*

25. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*.

26. Humbert O, Davis L, Maizels N. 2012.Targeted gene therapies: tools, applications, optimization. *Crit Rev Biochem Mol Biol*.

27. Kulcsár PI, Tálas A, Tóth E, Nyeste A, Ligeti Z, Welker Z. 2020. Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9 with 5′G-extended sgRNAs. *Nat Commun*.

28. Prakash V, Moore M, Yanez-Munoz RJ. 2016. Current progress in therapeutic gene editing for monogenic diseases. *Mol Ther*.

29. Vakulskas CA, Dever DP, Rettig GR, Turk R, Jacobi AM, Collingwood MA. 2018. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enablesefficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med*.

30. Xin H, Wan T, Ping Y.2019. Off-targeting of base editors: BE3 but not ABE induces substantial off-target single nucleotide variants. *Sig Transduct Target Ther*.

31. Ye L, Wang J, Beyer AI, Teque F, Cradick TJ, Qi Z. 2014. Seamless modification ofwild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. *Proc Natl Acad Sci USA*.

32. Zhai P, Ding Y, Wu X, Long J, Zhong Y, Li Y. 2020. The epidemiology, diagnosis and treatment of COVID-19. *Int J Antimicrobial Agents*.