



THE CRISPR-CAS SYSTEM: ITS ORIGIN, FUNCTIONS AND APPLICATIONS IN BIOTECHNOLOGY: A REVIEW

Tukur, A.M.

Department of Microbiology, Bayero University Kano, Nigeria.

Correspondence Author: gabidah@yahoo.com; +2348143410062

ABSTRACT

The CRISPR- Cas system is a genome editing system seen in prokaryotic immune system. Bacteria and archaea protect itself against invading viruses and plasmid by targeting RNA or DNA of the invading element predominantly using this gene-editing tool. The CRISPR-Cas defense mechanism is carried out in three stages; adaptation stage where the spacers are inserted into the CRISPR locus, the expression stage where crRNA is formed by transcription of the CRISPR loci and the interference stage where the invading element is destroyed by the crRNA and cas proteins. The CRISPR-cas has been involved in many other functions apart from the immune defense they include; DNA repair, regulation of virulence, genome evolution, inhibit biofilm formation etc. The application of CRISPR-cas system include genome engineering, agriculture to efficiently target and mutate plants, improve crop yield and crop resistance, in medicine to eradicate genetic diseases. However, ethical considerations are a major setback of CRISPR-Cas application especially in medicine. CRISPR-Cas has been used in variety of species including cultured human cell, rice, drosophila and mice.

Keywords: CRISPR, Cas proteins, genome, editing and immune system

INTRODUCTION

Bacteria and archaea have been found to adapt to harsh and unfavorable environments and survive competition of nutrients and other resources. In addition, bacteria and archaea are capable of withstanding predation from other living organisms and viruses. This has been made possible because of the bacteria's ability to modify and expand their genetic content, ability to defend itself from virus invasion and control its genetic homeostasis. Bacteria usually eubacteria try to acquire useful genetic material from their external environment usually plasmid and transposons and they get rid of their redundant gene in other to control the size of their genome. Clustered regulated interspaced short palindromic repeat (CRISPR) together with Cas protein have been recently found to provide archaea and bacteria immunity against viruses and plasmids by targeting their DNA and RNA (Barrangou, 2015).

The prevalence of CRISPR loci has been found to be 84% in archaea and 45% in bacteria. The CRISPR, which consist of an array of short repeated sequence (24-47 bp) are separated by spacers in a unique sequence can be found on both chromosomal and plasmid DNA. CRISPR is part of the antiviral system whereby spacers are obtained from DNA or RNA of the plasmid or virus and are use as a recognition element that

binds to the virus genome in other to destroy the virus or confers immunity (Rathet *et al.*, 2015).

CRISPR-CAS SYSTEMS

The CRISPR-Cas system is a genome editing system first observed in prokaryotic RNA-guided immune system. CRISPR repeats were first observed in *E. coli* genome that is found to be an unusual repeat locus (Harrison *et al.*, 2014 and Rathet *et al.*, 2015). It was later found out that bacteriophage and its genetic element resemble that of the CRISPR loci and this leads to investigation that the spacers are derived from the viral genome. The proto-spacers from the invading viral genome or plasmid genome is copied as the spacers and are inserted in repetitive sequence in the CRISPR locus, which forms short CRISPR RNAs (crRNAs) by transcription. The Cas proteins are then guided by the crRNAs to the target sequence (Harrison *et al.*, 2014).

Three major types of CRISPR –Cas system; type I, type II, type III and eleven subtypes of CRISPR-Cas system have been found (Harrison *et al.*, 2014; Chylinski, 2014; Barrangou, 2015). The CRISPR-Cas system type I and II, consist of the proto-spacer adjacent motif (PAM) (a 2-6 base pairs DNA sequence) that cleaves target DNA and also enables the Cas system to distinguish foreign phage or plasmid genome from the host CRISPR locus of the host genome

that does not have an adjacent motif (Harrison *et al.*, 2014). All the three types of the CRISPR-Cas system contain Cas1 and Cas2 genes that integrate and or cleave genetic element into the CRISPR region. The Cas1 gene is a metal dependent DNA-specific endonuclease that produces double stranded DNA (dsDNA). The Cas2 gene is also a metal dependent endonuclease that is involved in spacer acquisition. However, each CRISPR-Cas system has a specific Cas gene. The type I CRISPR-Cas system has Cas3 gene as its specific gene that is responsible for helicase and DNAse activities. The Cas9 gene is specific for type II CRISPR-Cas system, a protein that is responsible for RuvC-like and HNH nuclease domain. The type III system consists of Cas10 gene that entails nucleic acid polymerase and nucleotide cyclase. The RAMPs (Repeat associated mysterious proteins), which is also a Cas protein, found in all types of CRISPR-Cas systems and involved in processing the pre-crRNA transcription. Cas proteins are diverse in the sense that they do not resemble each other; studies have shown that over 30 Cas families were studied but no similarity was found (Makarova *et al.*, 2011). Type II Cas systems have the most minimal number of Cas genes and require the least number of components: a CRISPR repeat spacer array, three to four Cas genes, trans-activating CRISPR-RNA (tracrRNA) and endogenous RNase III (Chylinski *et al.*, 2014). Type II CRISPR-Cas9 systems are considered the easiest and most promising for use in genetic engineering (Harrison *et al.*, 2014).

THE ORIGIN OF CRISPR-CAS SYSTEM

A series of discoveries occurred that allowed researchers to first uncover how the CRISPR-Cas system functions, determine the Cas9 endonuclease capabilities, and then figure out how to reapply the system in humans. The first recorded recognition of repeat CRISPR sequences occurred in 1987 in *E. coli* when studying the involvement of iap enzyme in isoenzyme conversion of alkaline phosphatase by a group of scientists from Japan (Hsu *et al.*, 2014; Ishino *et al.*, 1987; Fineran and Charpentier, 2012). They noticed that repeat sequences were arranged in an array of spacer sequences in the genome of *E. coli* before anyone had discovered the function of these genes (Ishino *et al.*, 1987). In fact, it wasn't until the year 2000 that these

sequences were classified as a family of clustered repeat elements and Mojica and Jansen coined the acronym "CRISPR" in 2000 (Hsu *et al.*, 2014).

Two more key discoveries happened before genome editing of CRISPR-Cas type II system was approached. First study showed that Cas9 (previously called Cas5) has endonuclease activity and is able to cleave both plasmid and phage DNA that are in the proto-spacer region. The second discovery uncovered a significant step in the maturation of crRNA (Hsu *et al.*, 2014). Researchers uncovered abundantly transcribed RNA that had 24-25 nucleotides complementary to all pre-crRNA that they called the trans-activating CRISPR-RNA (tracrRNA) and noted that pre-crRNA maturation requires binding to tracrRNA, and the presence of RNase III and Cas9 (Deltcheva *et al.*, 2011).

With all of the components for the CRISPR-Cas9 system uncovered, scientists began to consider the applicability of Cas9 for genetic engineering (Hsu *et al.*, 2014). On the path to harnessing the type II CRISPR system as a gene-editing tool, Sapranaukas and colleagues (2011) transferred a CRISPR locus from *S. thermophilus* into *E. coli*. They reported that it is possible to express a functional CRISPR-Cas9 system from one bacteria strain to prevent plasmid transformation in different strain.

One more essential article was published before CRISPR-Cas9 was used to target DNA in mammalian cells (Hsu *et al.*, 2014). Jinek and others (2012) reported the ability to program Cas9 for site-specific DNA cleavage using a single chimeric guideRNA instead of the biological tracrRNA and crRNA components. The chimeric guideRNA imitated the respective roles of tracrRNA and crRNA in facilitating DNA cleavage and guiding Cas9 to the target DNA (Jinek *et al.*, 2012). The use of an engineered chimeric RNA as a new programmable RNA-guided endonuclease could be used and provided great promise for future uses in gene-targeting and gene-editing applications (Jinek *et al.*, 2012). One year after this critical paper was published, as predicted, two different studies showed that type II CRISPR systems could be used to edit mammalian genomes (Hsu *et al.*, 2014). The figure below shows the time line of CRISPR-Cas discovery and experiment regarding it by several scientists.

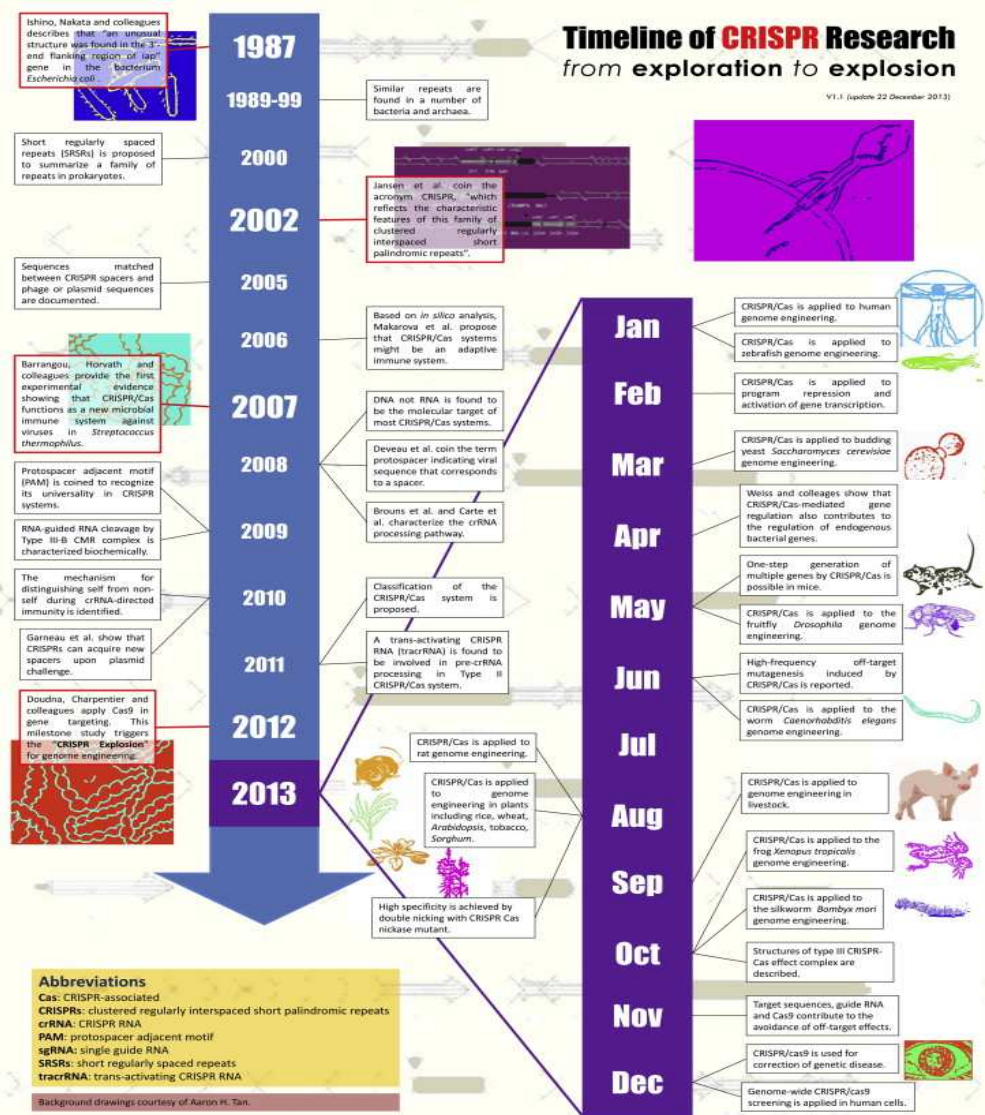


Figure 1: The history of CRISPR-Cas system and experiments regarding it by several scientists since its discovery in 1987. Adopted from Bassett, A. R. and Liu, J. (2014).

FUNCTIONS OF CRISPR-CAS SYSTEM

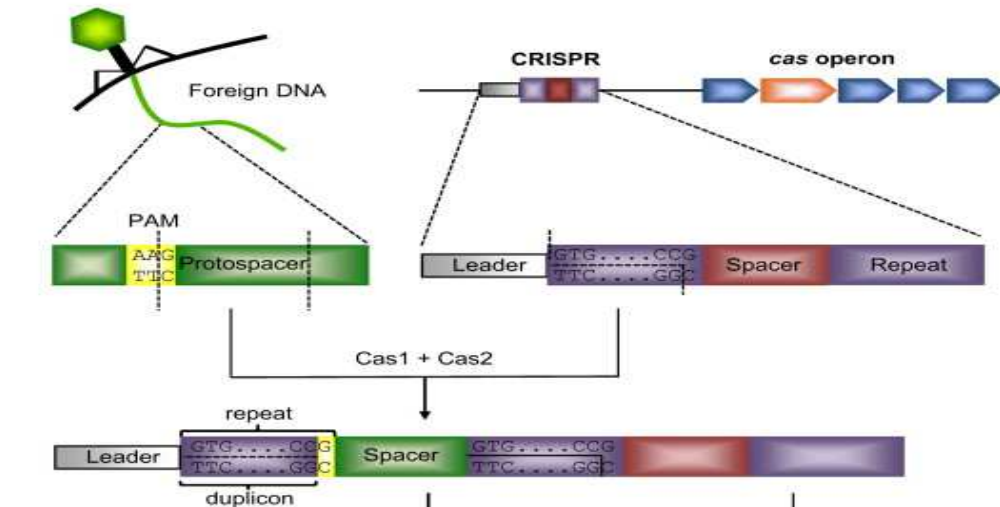
The CRISPR-Cas as an Adaptive Immunity in Prokaryotes

The CRISPR-Cas system serve as an immune defense mechanism predominant in bacteria and archaea to protect itself against invading viruses and plasmids by targeting RNA or DNA of the invading element (Rathet *al.*, 2015). The CRISPR-Cas defense mechanism are in three stages; adaptation stage, whereby the spacers are inserted into the CRISPR locus, expression stage where crRNA is formed by transcription of the CRISPR loci and the interference stage where the virus or plasmid is detected and destroy by the crRNA and Cas proteins (Jiang and Doudna, 2015; Rathet *al.*, 2015).

Adaptation stage

In this stage the genetic memory is established, which is very important for expression and interference stage in other to counteract the re-invasion of foreign nucleic acid and the new spacers are inserted into the CRISPR loci. There are two types of spacer acquisition; the naïve in which this is the first time the virus attacks the bacteria or archaea cell and prime in which there have been previous encounters with the intruding genetic element (Figure 2). After the acquisition, the protospacer and spacer material are selected which is then followed by the insertion of the spacer into the CRISPR loci and repeats are then synthesized (Figure 2). The Cas1 and Cas2 proteins are the main key factors in spacer integration (Fineran and Charpentier, 2012; Rathet *al.*, 2015).

A naïve acquisition



B priming acquisition

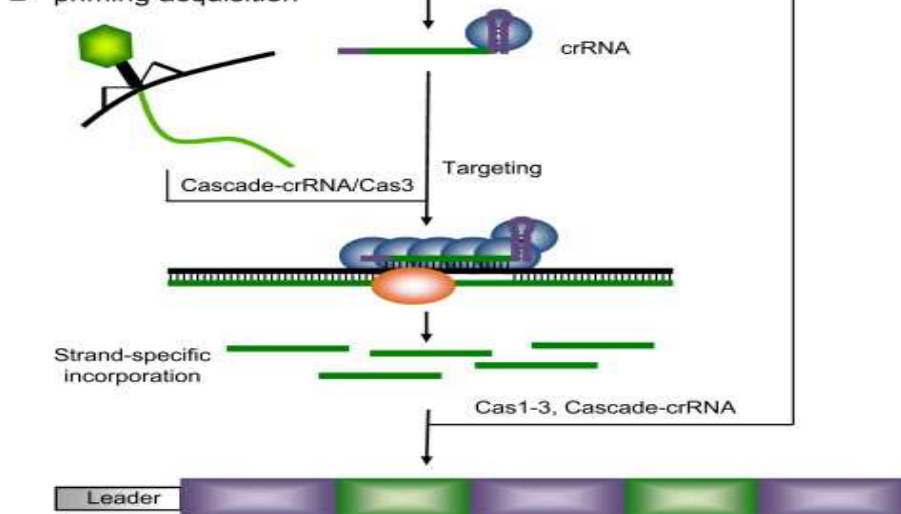


Figure 2: Mechanism of spacer acquisition during adaptation in CRISPR-Cas system. Adopted from Fineran and Charpentier (2012).

Expression stage

This is also known as the CRISPR-RNA (crRNA) biogenesis whereby the CRISPR locus is transcribed into precursor RNA (pre-RNA) and then into crRNA by endonucleolytic cleavage. The crRNA is composed of the single spacer that is surrounded by CRISPR repeat sequence (Jiang and Doudna, 2015). The crRNA biogenesis is similar in most organisms, however; type specific differences have been observed in some organisms (Rath *et al.*, 2015).

Interference stage

In the interference stage, the CRISPR-RNA is bound to the CRISPR-assisted (Cas) protein to locate the protospacer in the invading plasmid or virus and destroy the DNA thereby preventing the propagation of the virus or plasmid. The destruction is usually performed by the specific Cas nucleases (Jiang and Doudna, 2015; Rath *et al.*, 2015). In CRISPR-Cas type II system, tracrRNA and crRNA together with Cas9 are required for the interference stage. The tracrRNA binds to Cas9 and crRNA targets and destroy the invading genetic element (Rath *et al.*, 2015).

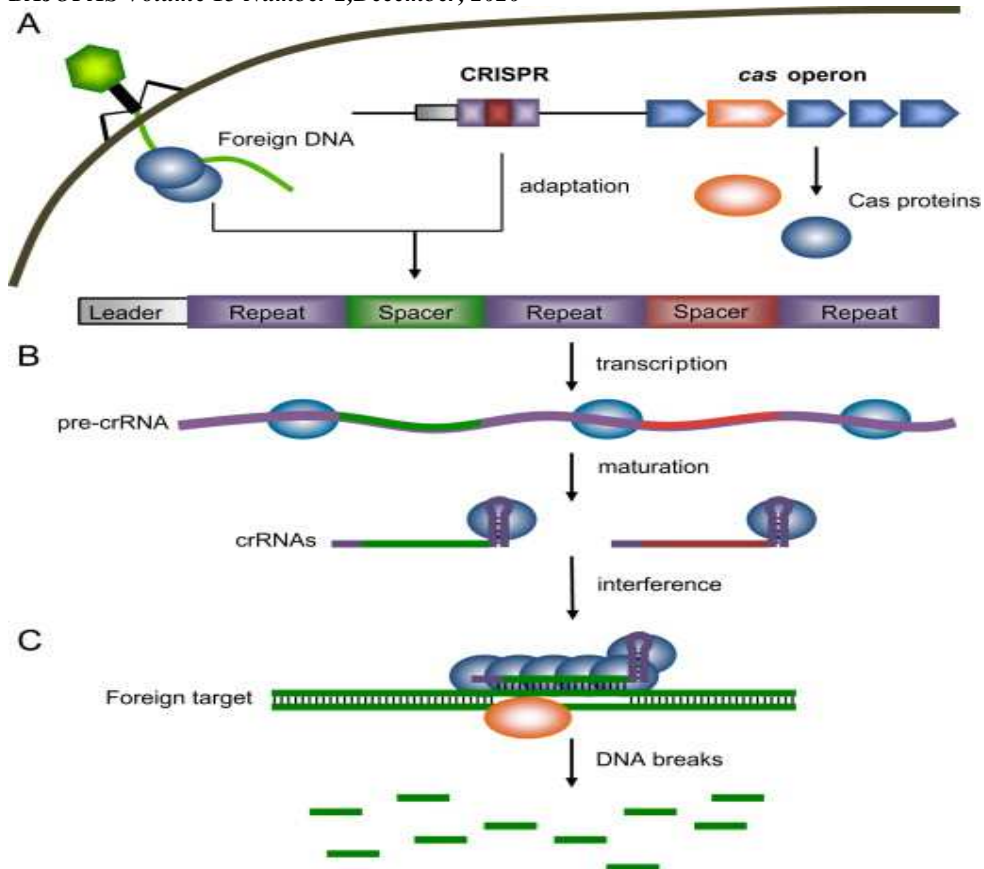


Figure 3 CRISPR-Cas adaptive immune mechanism in prokaryotes illustrating the adaptation stage (a), expression stage (b) and interference stage (c). Adopted from Fineran and Charpentier(2012).

CRISPR-Cas Involvement in other Functions

CRISPR-Cas has been found to perform other functions in prokaryotes other than providing immunity against invading viruses and other genetic element. CRISPR-Cas system is found to carry out other functions within the cell including DNA repair, regulation of virulence and genome evolution. CRISPR-Cas has been found to control defective protein in *E. coli*. In *Francisella novicida*, CRISPR-Cas9 promotes the repression of endogenous lipoprotein gene. it has also been observed that *Pseudomonas aeruginosa* CRISPR inhibits the formation of biofilm. However, these functions are still under investigation to determine whether they are independent function or they are related to CRISPR-Cas predominant immune function (Rathet *al.*, 2015; Barrangou, 2015).

APPLICATIONS OF CRISPR-Cas SYSTEM

Genome Engineering

The majority of CRISPR-Cas system application is in the field of genome engineering because of its ability to produce double stranded break (DSB) at specific region within the genome. The DSB, is repaired by homologous recombination (HR) or non-homologous end joining (NHEJ)

because, HR and NHEJ can be use to promote mutation and manipulate the genome (Bassett and Lui, 2014). CRISPR-Cas9 system was used to detect mutation in *Drosophila melanogaster* in which injection of the Cas9 mRNA into *Drosophila* embryo induces mutation of the desired gene, 88% of the injected flies were found to express those genes and mutation were transmitted through germline and stable line was established (Bassett *et al.*, 2013).

CRISPR-Cas system effectively use for multiplex genome engineering (Lui and Fan, 2014). The CRISPR-Cas9 system is capable of multiplexing with multiple guideRNA targeting different sites and one Cas9 protein; multiple sites can be altered at the same times. This is useful when one guideRNA is not enough to disrupt a gene (Auer and Del Bene, 2014) or when modifying two or more genes at the same time (Zhou *et al.*, 2014).

Using two different guideRNAs to target the same gene has been shown to increase mutagenic frequencies. On the other hand, targeting more than one gene at the same time is also useful because it can decrease the time and money required to produce animal models with different mutations.

Multiplexing offers a unique ability of CRISPR-Cas9 that has not been successfully utilized using ZFNs or TALENs due to the complexity required to design a protein for each targeted gene and the variable efficacy at each gene (Zhou *et al.*, 2014)

Multiplexing with CRISPR-Cas9 was performed shortly after the system was first established for use in eukaryotes. One of the first multiplexing experiments was completed in mice with two guideRNAs targeting the Tet1 and Tet2 genes (Wang *et al.*, 2013). The two guideRNAs were co-injected with Cas9 to yield biallelic mutations of both genes at 80% efficiency (Wang *et al.*, 2013). A similar experiment in rats resulted in bi-allelic mutagenesis of three different genes at a rate of 60%. As many as six genes have been targeted at one time with ten guideRNAs, resulting in multiplexed gene alteration. Multiplexing with CRISPR-Cas9 was also used to generate mouse models with mutations in multiple immune response genes (Zhou *et al.*, 2014).

In an experiment, five different guideRNA were injected with Cas9 into 9 mice. The results showed that 22.2% had four mutations, 33.4% had three mutations, and 44.4% had two mutations. No mice had five different genes altered. The multiplexing efficiency increased to 85% when multiple guideRNA targeting the same genes were used at the same time (Zhou *et al.*, 2014).

Agriculture

Genome editing using CRISPR-Cas system has been found to be effective in eukaryotes. CRISPR-Cas has been reported to efficiently target and mutate plants. Gene editing is of very important in agriculture as it improves crop yield and crop resistance. Before the discovery of CRISPR-Cas systems, zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been used in gene editing, however, they are tedious and time consuming unlike CRISPR-Cas that is easier and cheap that requires a pair of oligonucleotide to construct plasmid (Xie and Yang, 2013).

Xie and Yang (2013) demonstrated the use CRISPR-Cas system as gene editing tool in rice protoplast resulted in 60% of GFP expression in 18 hours after transformation and 90% after 36-74 hours after transformation indicating the expression of cas9 nuclease. Therefore CRISPR-Cas can be adopted as gene editing and

REFERENCES

Auer, T. O. and Del Bene, F. (2014). CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods* **69**, 142-150.

targeting tool for characterization of plant genes and modification of crops to improve resistance and crop yield.

Medicine

Hepatitis B virus (HBV) has remained one of the killer diseases even though vaccination towards the disease is fully taken. About 4% of the world's population remains infected with HBV. All treatments of HBV leads to inhibition of HBV leads to inhibition of HBV reverse transcriptase which does not lead to elimination of the disease because the episomal DNA (cccDNA) is very stable during replication (Kennedy *et al.*, 2015). Kennedy and his colleagues combine together *Streptococcus pyogenes* Cas9 (spy Cas9) with HBV specific sgRNA and induced it into the hepatoma-infected cell using a lentiviral vector system. Three target sites were selected in the HBV genome. They found out that there was a strong inhibition of viral DNA after the introduction of the Cas/sgDNA combination in all the three target sites. Therefore, CRISPR-Cas system can directly target HBV cccDNA in infected liver. However, this experiment was carried out outside the human body, there is need to experiment it in animal model to ensure whether it will be applicable efficiently in curing the disease (whether Cas9/sgRNA are safe to inhibit the HBV cccDNA pool) (Kennedy *et al.*, 2015).

However, ethical considerations are major setback of CRISPR-Cas application in medicine. Chinese scientists have reported to edit non-viable human embryo genome using CRISPR-Cas system in other to modify the gene that causes β -thalassemia. But controversies regarding this breakthrough as some regard it as unethical because it causes modification of germline and is heritable which might have effect on the future generation, while others regard it as a breakthrough in eradicating genetic disease (Cyranoski and Reardon, 2015).

CONCLUSION

CRISPR-Cas since its discovery has been found to be applicable in biotechnology and other aspects of science and medicine since then, the system has been used in a variety of species including, cultured human cell, *Drosophila*, mice, rice, etc. From the rapid research and discovery to its already vast applications, one can start to see how truly exciting this new technology is.

Barrangou, R. (2015). The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Current Opinion in Immunology*. **32**, 36-41.

- Bassett, A. R., Tibbit, C., Ponting, C. P. and Liu, J. L. (2013). Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Reports* **4**, 220-228.
- Bassett, A. R. and Liu, J. (2014). CRISPR/Cas9 and genome editing in *Drosophila*. *Journal of Genetics and Genomics* **41**, 7-19.
- Chylinski, K., Makarova, K. S., Charpentier, E. and Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research* **42**, 6091-6105.
- Cyranoski, D. and Reardon, S. (2015). Chinese scientists genetically modify human embryos. *Nature* **520**, 7549.
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. and Charpentier, E. (2011). CRISPR-RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607.
- Fineran, P. C. and Charpentier, E. (2012). Memory of viral infections by CRISPR-cas adaptive immune systems: Acquisition of new information. *Virology* **434**, 202-209.
- Harrison, M. M., Jenkins, B. V., O'Connor-Giles, K. M. and Wildonger, J. (2014). A CRISPR view of development. *Genes and development* **28**, 1859-1872.
- Ishino, Y., Shinagawa, H., Amemura, M. and Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology* **169**, 5429-5433.
- Jiang, F. and Doudna, J. A. (2015). The structural biology of CRISPR-cas systems. *Current Opinion in Structural Biology* **30**, 100-111.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821.
- Kennedy, E. M., Bassit, L. C., Mueller, H., Kornepati, A. V. R., Bogerd, H. P., Nie, T., Chatterjee, P., Javanbakht, H., Schinazi, R. F. and Cullen, B. R. (2015). Suppression of hepatitis B virus DNA accumulation in chronically infected cells using a bacterial CRISPR/Cas RNA-guided DNA endonuclease. *Virology* **476**, 196-205.
- Liu, L. and Fan, X. D. (2014). CRISPR-cas system: A powerful tool for genome engineering. *Plant Molecular Biology* **85**, 209-218.
- Makarova, K. S., Aravind, L., Wolf, Y. I. and Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct* **6**, 1-27.
- Rath, D., Amlinger, L., Rath, A. and Lundgren, M. The CRISPR-cas immune system: Biology, mechanisms and applications. *Biochimie* **117**, 119-128.
- Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research* **39**, 9275-9282.
- Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. and Jaenisch, R. (2013). One-Step generation of mice carrying mutations in multiple genes by CRISPR/Cas-Mediated genome engineering. *Cell* **153**, 910-918.
- Xie, K. and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR-Cas system. *Molecular Plant* **6**, 1975-1983.
- Zhou, J., Shen, B., Zhang, W., Wang, J., Yang, J., Chen, L., Zhang, N., Zhu, K., Xu, J., Hu, B. et al. (2014). One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. *International Journal of Biochemistry and Cell Biology* **46**, 49-55.