The Birth of CRISPR/Cas9

Bianca Bolocan[1*], Jacob Odell[2]

[1] Issaquah High School; [2] Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca NY, 14853
*Corresponding author: bibibolocan@gmail.com(425-371-9597)
Keywords: CRISPR, Cas9, Cas13, genetic engineering, CRISPR/Cas9, CRISPR/Cas13, History of CRISPR

Abstract

CRISPR/Cas9 is a very important component of genetic engineering, originating from the immune system of bacteria. It is far more efficient than any other method discovered as of 2023. Despite this, however, many people interested in this tool are not aware of how it was discovered and adapted. There are several contributors who helped build CRISPR/Cas9 into what it is today, from Francisco Mojica, who discovered the system, to Le Cong, who identified its purpose, and finally, Feng Zhang, who adapted the system for humans. This paper will discuss the history of CRISPR, from its discovery to its uses today, along with how it continues to be developed.

Introduction

Almost every living thing in the world will, at some point, be attacked by a virus. They exist in every part of the world and can attack many living things, including microbes. Much like us, microbes have evolved an adaptive immune system to fend off these viruses. For example, there is an immune system called CRISPR that protects bacteria from their own form of virus. CRISPR (standing for clustered regularly interspaced short palindromic repeats) is a complex system that relies on palindromic repeats and arrays of enemy DNA to protect a prokaryote from invading viruses that deposit their genetic material within the bacteria (Jinek et al, 2012). An overview of this system is presented in Figure 1, and is described in detail below. This system contains a crRNA (CRISPR RNA) composed of spacers, segments of DNA derived from the genetic material of a bacterium-infecting virus called a bacteriophage (or phage). crRNA is encoded on a piece of DNA within the bacterial genome called a CRISPR locus where the spacers are separated by a series of palindromic repeats, and the crRNA is transcribed from only the spacers so that it can match complementarily with the base pairs of any DNA complementary to it (Cong et al, 2013). Part of this system includes a complementary tracrRNA (Trans-activating CRISPR RNA) that aids in the cutting process and the folding of early crRNA (pre-crRNA) into crRNA. Finally, the type II CRISPR system utilizes a protein called Cas9, that cuts the phage DNA. In the wild, the crRNA is compared by a Cas9 protein to the DNA found inside a bacterium until a match is found, signaling that the DNA is from a phage attempting to multiply itself, rather than the bacterium's own DNA (Jinek et al, 2012). When a match to one of these phage segments is found, the Cas9 uses nuclease domains to cut the selected piece, which destroys the genetic material of the virus, rendering it harmless (Jinek et al, 2012) (Cong et al, 2013).

Figure 1: Class 2, Type II CRISPR-Cas9 System from Streptococcus thermophilus. Published originally in Lander et al, 2016. "Type II systems are the simplest of the three types of CRISPR systems and have been the basis for genome editing technology. **"(A)** The locus

contains a CRISPR array, four protein-coding genes (cas9, cas1, cas2, and cns2) and the tracrRNA. The CRISPR array contains repeat regions (black diamonds) separated by spacer regions (colored rectangles) derived from phage and other invading genetic elements. The cas9 gene encodes a nuclease that confers immunity by cutting invading DNA that matches existing spacers, while the cas1, cas2, and cns2 genes encode proteins that function in the acquisition of new spacers from invading DNA. **(B)** The CRISPR array and the tracrRNA are transcribed, giving rise to a long pre-crRNA and a tracrRNA. **(C)** These two RNAs hybridize via complementary sequences and are processed to shorter forms by Cas9 and RNase III. **(D)** The resulting complex (Cas9 + tracrRNA + crRNA) then begins searching for the DNA sequences that match the spacer sequence (shown in red). Binding to the target site also requires the presence of the protospacer adjacent motif (PAM), which functions as a molecular handle for Cas9 to grab on to. **(E)** Once Cas9 binds to a target site with a match between the crRNA and the target DNA, it cleaves the DNA three bases upstream of the PAM site. Cas9 contains two endonuclease domains, HNH and RuvC, which cleave, respectively, the complementary and non-complementary strands of the target DNA, creating blunt ends."

Scientists have managed to "hack" this system by creating a chimeric gRNA, which is a fusion of crRNA and tracrRNA created in the lab, that can then be inserted into a cell. To generate a knockout (de-activation of a specific gene) this gRNA and Cas9 are added to a cell. If a targeted gene edit is desired instead, a "template piece" of DNA will be added as well. After entering the cell, the gRNA will be compared to the DNA by the Cas9 of the cell in question until a match is discovered. Once this match is found, a protospacer adjacent motif (PAM), that was found beforehand as the Cas9 scanned the DNA, is used as an anchor point for the Cas9 to use its nuclease domains to sever it (Lander et al, 2016). Once the cut is made 3 base pairs upstream of the PAM, the cell will repair that cut DNA. Cells will usually end up making a mistake if they repair DNA via non-homologous end joining, which is a process in which the cell will attempt to ligate two pieces together, leading to mutations that may add or remove certain base pairs. If the cell successfully repairs the break, the CRISPR-Cas9 complex will cut it again, and the cycle repeats until the cell makes a mistake, resulting in new base pairs appearing or old ones not being copied (Cong et al, 2013). This mutation disrupts the normal translation of the protein, making it impossible to create, and the knockout is complete. Another DNA repair pathway a cell can take is homologous recombination, which uses a template piece much like the one added in a targeted gene edit, and uses it to repair the broken strand. The cell may use the lab-grown template to put the DNA back together, which may result in a lab-grown edit of any type desired by the researcher (Cong et al, 2013). This way, researchers can precisely deliver a genome with any mutation needed, or even repair existing mutations, allowing for a multitude of edits within living cells. This paper will discuss how this process was discovered and identified, how scientists started to use it for gene editing, and how it's used today (Lander et al, 2016).

Discovery of CRISPR

The first to discover CRISPR was Francisco Mojica. The first cell he discovered the system in was *Haloferax mediterranei*, a microbe living in the saltwater marshes of Santa Pola. Within this organism, he discovered a segment of DNA that stood out from the rest: it contained 30-base sequences of DNA that were roughly palindromic, separated by 36-base pairs that

resembled nothing discovered before (Mojica et al., 1993). As his career went on, he discovered such loci in 20 other species (Mojica et al., 2000) and, as other scientists began to discover such loci, CRISPR became a focus of the genetic engineering community. Many experts, including him, suspected CRISPR was important for something; these similarities were not likely to be coincidental.

In 1990, another important tool was created: the BLAST gene database. Standing for Basic Local Alignment Search Tool, it is a program that uses algorithms to find similarities between sample pieces of DNA, RNA, or protein, and information from a database full of protein and nucleotide sequences, meant to help find regions of similarity between sequences: exactly what Mojica needed. In hopes of discovering the source of the repeat DNA he discovered earlier, he types in each base pair of the *H. mediterranei* locus and recovers interesting results: the sequence matched a phage infecting *E. Coli*, suggesting for the first time that CRISPR may help a bacterium fend off phage infections (Mojica et al., 2005). At the same time, CRISPR also caught the eye of the French Direction Generale de l'Armement. Gilles Verngaud, a researcher there, was assigned to attempt to make sense of what the function of CRISPR was. Verngaud found the peculiar genetic structure in the genome of *Y. pestis* and found many of the spacers corresponded to a *Y. Pestis* phage, signaling its existence as "a record of past aggressions" (Pourcel et al., 2005). Yet another researcher, Alexander Bolotin, was the first to theorize on how CRISPR might confer immunity. Bolotin proposed that the transcripts worked via anti-sense RNA inhibition: he believed that the CRISPR arrays were transcribed to RNA that was then compared to any RNA in the prokaryote until a match was found between it and virus RNA. The crRNA would then bind itself complementarily to the target virus RNA, promoting its degradation and inhibiting the virus' ability to make more protein and reproduce (Bolotin et al., 2005). However, he would prove to be wrong.

Table 1. Classification and Examples of CRISPR Systems

Class	Type	Subtype	Hallmarks	Example effector	Example organism	Studies Cited
Class 1	Type I		multisubunit effector complex; Cas3	Cascade	*E. coli*	Brouns et al., 2008
	Type III	III-A	multisubunit effector complex; Csm effector module; DNA targeting	Cas10-Csm	*S. epidermidis*	Marraffini and Sontheimer, 2008
		III-B	multisubunit effector complex; Cmr effector module; RNA targeting	Cmr	*P. furiosus*	Hale et al., 2009
Class 2	Type II		single protein effector; tracrRNA	Cas9	*S. thermophilus*	Bolotin et al., 2005; Barrangou et al., 2007; Sapranauskas et al., 2011; Gasiunas et al., 2012
					S. pyogenes	Deltcheva et al., 2011; Jinek et al., 2012; Cong et al., 2013; Mali et al., 2013
	Type V		single protein effector; single-RNA guided	Cpf1	*F. novicida*	Zetsche et al., 2015

Adapted from Lander et al, 2016

Despite many ideas about how CRISPR worked, as of 2012, nobody knew how CRISPR worked. To solve this issue, Emanuelle Charpentier's lab performed a study. They used a type-2 CRISPR system (defined above in Table 1) *in vitro* to figure out its parts. They then painstakingly remove certain parts of the system, such as the Cas9 or the tracrRNA, to figure out the necessary components for efficient CRISPR cutting and narrow it down to a pre-crRNA

or crRNA, a tracrRNA, and Cas9 as physical components, alongside the PAM. On top of that, they find what nuclease domains of Cas9 are used to generate cutting: HNH cuts the complementary strand, and RuvC cuts the non-complementary strand, which can be seen in Figure 2 (Jinek et al, 2012). With CRISPR's purpose in mind, they started experimenting. They created truncated forms of tracrRNA and found that only nucleotides 23 to 48 are required, and only a 13-base-pair match between the crRNA and the target DNA is required for Cas9 to cleave target DNA efficiently. Last but not least, they created a chimeric RNA: a combination of a crRNA and a tracrRNA, designed to bind complementarily to a specific sequence, in this case, a gene encoding the green fluorescent protein. Five different chimeric gRNAs were used to test this hypothesis and they all successfully led to Cas9 cutting the DNA template, suggesting for the first time that gene editing might be possible via CRISPR. However, since the reaction happened *in vitro*, more experiments were necessary to prove gene editing via CRISPR was feasible, and to adapt the system for use in living organisms (Jinek et al, 2012).

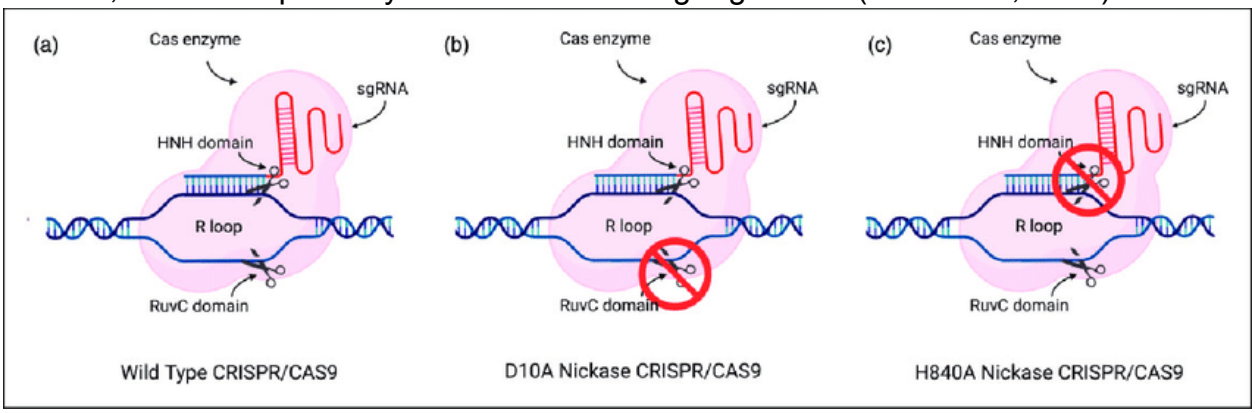

Figure 2: Wild type and nickase CRISPR/Cas9. Adapted from Caso et al, 2021. "**(a)** In the wild-type CRISPR/Cas9 system, the binding of the sgRNA to the target strand results in the formation of an R loop, and the two nuclease domains (HNH and RuvC) of the Cas9 enzyme are responsible for the cleavage of each single strand of DNA. **(b)** An engineered CRISPR/Cas9 nickase harboring a D10A mutation that inactivates the RuvC domain, resulting only in the cleavage of the target strand of DNA (a single 'nick') via the still active HNH domain. **(c)** An engineered CRISPR/Cas9 nickase harboring an H840A mutation that inactivates the HNH domain, resulting only in the cleavage of the non-target strand of DNA (a single 'nick') via the still active RuvC domain." The use of two nickases makes the CRISPR-Cas9 system much more accurate because it requires both nickases to bind to the correct sequence for a successful cut. The likelihood of both nickases binding to the same incorrect sequence is incredibly low.

CRISPR in the Lab

Before CRISPR's discovery, other methods of gene editing were devised. One of the methods of gene editing used zinc finger nucleases. They are man-made combinations of zinc and protein that bind onto a specific DNA sequence and snip in two places like Cas9 (Durai et al, 2005). However, they had a high risk of causing off-target edits. A new technique of 2011-2013 started with the discovery of an important part of plant bacteria: TALE (trans-activator-like effector nucleases), a trans-activating protein that recognizes DNA, binds to it, and cuts it in a somewhat similar fashion to Cas9, which can be seen in Figure 3 (Boch et al

2009). They were first repurposed for mammals by Feng Zhang, creating full-length domains to be transplanted into a desirable cell in a similar manner to CRISPR/Cas9 (Zhang et al, 2011). These processes, however, were very expensive and time-consuming to make compared to CRISPR. Zhang still wanted a more efficient solution. He studied the Cas9 of *S. thermopilus* and set out to create a better version. He found that it tended to spread unevenly in the nucleus, so he replaced it with *S. pyogenes* Cas9, which solved the issue. Alongside that, he tested out isoforms of tracrRNA to find one stable in humans. Zhang's lab would also go on to add a nuclear localization signal to Cas9 to allow for the use of CRISPR within a mammal. His lab also decided to drive the expression of crRNA and tracrRNA using a U6 promoter, avoiding the addition of a poly-A tail (a measure that is meant to prevent RNA degradation), allowing the RNAs to fold correctly (Cong et al, 2013). In the end, he created a complex with 3 components: Cas9 from *S. pyogenes* or *S. thermophilus*, the tracrRNA, and a CRISPR array. With this complex, he showed it was possible to edit DNA in human and rat cells via CRISPR effectively. The solution would then be perfected by the labs of both Feng Zhang and Jennifer Duonda until it was proven effective.

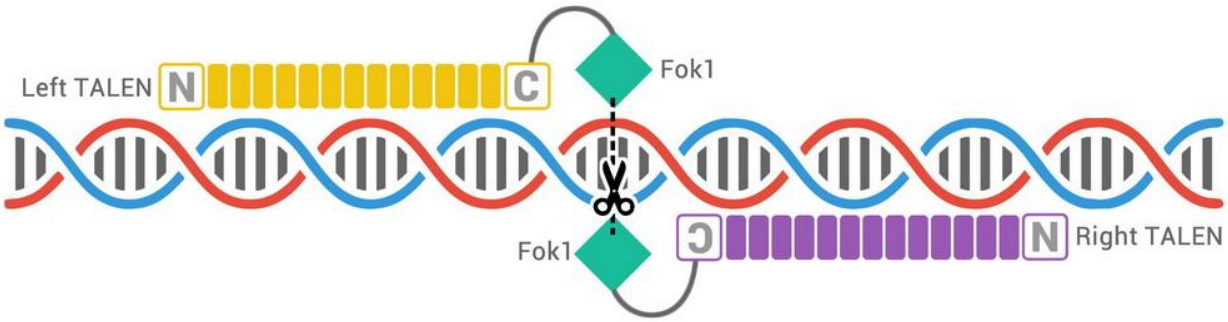

Figure 3. TALENs. Adapted from ISAAA Inc. TALE-nuclease fusions cut the DNA, making a precise double-strand break. This break initiates the plant cell's existing DNA repair mechanisms which may remove DNA bases around the cut or incorporate new DNA at the position of the break. The outcome is either small insertions or deletions that create sequence diversity, as often found in nature, leading to new functions or inactivating existing ones. Additionally, new genes can be inserted.2,3 These molecular scissors provide incredibly precise means to fine-tune traits in a targeted, predictable fashion anywhere in the genome.

Zhang would go on to show how versatile CRISPR could be. It could make complex models of inherited diseases, off-target cutting was found to be very low, and further research could only improve CRISPR's accuracy. Following the submission of his paper, he would start working closely with the non-profit Addgene to distribute Zhang's reagents, helping to raise awareness about CRISPR and make this system of gene editing more accessible (Cong et al, 2013). CRISPR would go on to raise many eyebrows, including getting the attention of George Church, a collaborator of Zhang's. Church would go on to test crRNA and tracrRNA for himself, eventually creating "gene drives", a means of spreading a gene rapidly through a population. This would bring hope toward ending diseases like malaria, help endangered species adapt to the changing ecosystem through means such as higher temperature tolerance, make certain preventative treatments such as vaccines obsolete by modifying human populations, and even weaken or destroy invasive species through population control genes. Gene drives work by cutting a chromosome at a specific site in a chromosome, allowing for the insertion of a driven

gene that will affect all offspring of an animal with that gene drive edit, as seen in Figure 4(Garrood et al, 2022) (Bier et al, 2021).

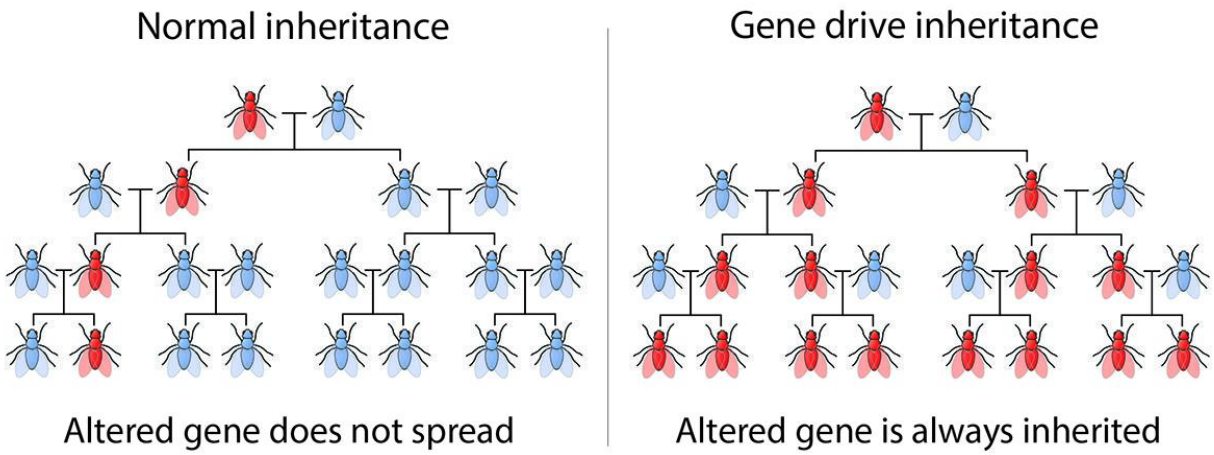

Figure 4. Gene Drive Inheritance. Image modified from Mariuswalter via Wikimedia Commons. "With normal inheritance, an altered gene could be removed from the population easily. But with gene drive inheritance, an altered gene is always inherited, so it quickly spreads through a population." In the context of malaria, for example, the driven gene could confer transgenic mosquitos resistance to malaria, preventing human infections.

CRISPR Today

CRISPR's contributions are nowhere near done. While the CRISPR-Cas9 complex is well-studied, for example, it is not the only one. The CRISPR-Cas13 complex, for example, works on RNA, not DNA. Huawei Tong would go on to study this complex and learn that wild-type Cas13 is very error-prone, destroying any RNA in its vicinity rather than one specific type. The team would then go on to create Cas13d, a higher-fidelity version of the wild-type Cas13, before refining it into hfCas13d and Cas13X by further editing its amino acid sequence, chemically turning the two into different compounds from the wild-type Cas13 (Tong et al, 2023). In an experiment, hfCas13d would prove its superiority over Cas13d when inserted into mice in an albinism experiment, where the Cas13 was directed to cut up RNA relating to pigmentation. While the mice with Cas13d died after a few weeks, the mice with hfCas13d would live to adulthood. This CRISPR-Cas13 complex has the potential to selectively target harmful RNA transcripts, which may be more favorable than the CRISPR-Cas9 complex which can create irreversible damage by editing rarely-transcribed, long-lasting DNA instead of RNA that is widely duplicated and has a short lifespan. This would make the CRISPR-Cas13 complex more viable for treating diseases like sickle cell anemia, where only part of the red blood cells may be affected. Another use for it is for regulating the production of certain overproduced proteins, such as the overabundance of huntingtin that causes Huntington's disease (Tong et al, 2023).

Alongside this, many medical studies are currently underway. In Guangzhou, China, testing has begun, comparing CRISPR's and TALEN's efficacy against HPV, which can remain in a host's body for a long time. Topical gels with both gene editing systems are applied to the affected area and compared with a control group that got no medical attention (ID NCT03057912). Another study currently going on in Palo Alto aims to use CRISPR to cure

sickle cell anemia with one IV infusion of a CRISPR-Cas9 complex. The complex will be administered directly to patients, editing the BCL11A locus, which is responsible for red blood cell production, theoretically resulting in the destruction of the gene creating sickle-shaped cells (ID NCT03745287). On top of that, not all uses of CRISPR may be related to medicine. Daniel B Sulis and his colleagues created gene edits in trees to help sustainable fiber production. They edited poplar trees to give them a greater growth rate and greater genetic diversity (Sulis et al, 2023).

Conclusion

From here, CRISPR can take on many shapes. It will likely achieve its medical purpose, slowly perfecting Cas9 and gRNAs to severely lower the probability of off-target edits and make it easier to get into a desired cell. Cas13 may become a common cure for diseases stemming from the overproduction of protein, while Cas9 could be used for the production of faulty protein. Farther down the line, these complexes could result in the end of genetic disease, the stabilization of ecosystems, and even purposes farther from medicine, such as creating humans with increased physical and mental capabilities, and even aesthetic modifications to living things, making the creation of “living art” a common pastime for people living decades or even centuries from now. However, this could result in the editing of superficial traits, discrimination against people with genetic edits, and even the creation of biological weapons. Currently, edits on a human being are banned by the World Health Organization. It is important to be both prudent and curious with CRISPR/Cas9, as with any other new technology.

References

1. Bier, E. (2021). Gene drives gaining speed. *Nature Reviews Genetics*, *23*(1), 5–22. https://doi.org/10.1038/s41576-021-00386-0
2. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of tal-type III effectors. *Science*, *326*(5959), 1509–1512. https://doi.org/10.1126/science.1178811
3. Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0
4. Caso, F., & Davies, B. (2021). Base editing and prime editing in Laboratory Animals. *Laboratory Animals*, *56*(1), 35–49. https://doi.org/10.1177/0023677221993895
5. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143
6. Durai, S. (2005). Zinc finger nucleases: Custom-designed molecular scissors for Genome Engineering of plant and mammalian cells. *Nucleic Acids Research*, *33*(18), 5978–5990. https://doi.org/10.1093/nar/gki912
7. Garrood, W. T., Cuber, P., Willis, K., Bernardini, F., Page, N. M., & Haghighat-Khah, R. E. (2022). Driving down malaria transmission with engineered gene drives. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.891218
8. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

9. Lander, E. S. (2016). The heroes of CRISPR. *Cell*, *164*(1–2), 18–28. https://doi.org/10.1016/j.cell.2015.12.041
10. Mojica, F. J., Juez, G., & Rodriguez-Valera, F. (1993). Transcription at different salinities of *haloferax mediterranei* sequences adjacent to partially modified *pst*I sites. *Molecular Microbiology*, *9*(3), 613–621. https://doi.org/10.1111/j.1365-2958.1993.tb01721.x
11. Mojica, F.J.M., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the *archaea haloferax mediterranei* and *haloferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology*, *17*(1), 85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x
12. Mojica, Francisco J.M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3
13. Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653–663. https://doi.org/10.1099/mic.0.27437-0
14. Sulis, D. B., Jiang, X., Yang, C., Marques, B. M., Matthews, M. L., Miller, Z., Lan, K., Cofre-Vega, C., Liu, B., Sun, R., Sederoff, H., Bing, R. G., Sun, X., Williams, C. M., Jameel, H., Phillips, R., Chang, H., Peszlen, I., Huang, Y.-Y., … Wang, J. P. (2023). Multiplex CRISPR editing of wood for Sustainable Fiber Production. *Science*, *381*(6654), 216–221. https://doi.org/10.1126/science.add4514
15. Tong, H., Huang, J., Xiao, Q., He, B., Dong, X., Liu, Y., Yang, X., Han, D., Wang, Z., Wang, X., Ying, W., Zhang, R., Wei, Y., Xu, C., Zhou, Y., Li, Y., Cai, M., Wang, Q., Xue, M., … Yang, H. (2022). High-fidelity CAS13 variants for targeted RNA degradation with minimal collateral effects. *Nature Biotechnology*, *41*(1), 108–119. https://doi.org/10.1038/s41587-022-01419-7
16. Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M., & Arlotta, P. (2011). Efficient construction of sequence-specific tal effectors for modulating mammalian transcription. *Nature Biotechnology*, *29*(2), 149–153. https://doi.org/10.1038/nbt.1775
17. clinicaltrials.gov ID NCT03057912, https://clinicaltrials.gov/study/NCT03057912?distance=50&cond=NCT03057912&rank=1
18. clinicaltrials.gov ID NCT03745287, https://clinicaltrials.gov/study/NCT03745287?distance=50&cond=NCT03745287%20&rank=1