


# Studying the Features of CRISPR Loci in *Stenotrophomonas*

Hengxia Fu<sup>1</sup>, Mengmeng Zhang<sup>1</sup>, Jiangnan Qin<sup>3</sup>,  
Wei Luo<sup>2</sup>, Bin Wang<sup>1</sup>

<sup>1</sup>Department of Microbiology, Medical College of Qingdao University, Qingdao 266071, Shandong, China

<sup>2</sup>Qingdao Municipal Hospital, Qingdao 266071, Shandong, China

<sup>3</sup>401 Hospital of PLA, Qingdao 266071, Shandong, China

**Abstract:** To study the genetic structure of the *Stenotrophomonas* CRISPR-Cas system using bioinformatics methods. **Methods** The sequence information of all *Stenotrophomonas* strains published in the CRISPRdb database was collected, and the CRISPR locus was analyzed using the CRISPRFinder software; All spacers were searched by BLAST platform in the PubMed database to find homologous sequences, and then the relationship between the number of spacer sequences and the number of phages was statistically analyzed. **Results** According to statistics, 15 confirmed CRISPR structures and 132 questionable CRISPRs were found in 26 strains of *Stenotrophomonas*, and the repeat sequences of CRISPR structures in different strains were more conservative. Only 1.3% of spacers were homologous with the sequences of known bacteriophages or plasmid in NCBI database. **Conclusion** The targeted genes of the spacer sequences are mainly from the genome of the bacteria, indicating that the evolution of the *Stenotrophomonas* CRISPR is related to other bacterial genes. In addition, the negative correlation between the spacer sequence and the number of phages indicates that CRISPR can prevent phage invasion. Analysis of the structure of CRISPR loci in the genome of *Stenotrophomonas* laid the foundation for further study of drug resistance and genomic stability.

**Keywords:** *Stenotrophomonas*, CRISPR-Cas Systems, Repeat Sequence, Spacer Sequence

## 1. INTRODUCTION

*Stenotrophomonas spp.* is a kind of Gram-negative non-fermenting bacteria, of which *S. maltophilia* is the most common bacteria in hospital infection (Alavi et al., 2014). The genus is parasitic in humans and animals and is widely distributed in both natural and hospital environments. This organism can produce a variety of biologically active substances and have strong metabolic activity and adaptability (Ribitsch et al., 2012). In addition, *S. maltophilia* is a common opportunistic pathogen causing nosocomial infections. In recent years, the clinical isolation rate of *S. maltophilia* has been increasing due to the wide application of various broad-spectrum antibacterial drugs and invasive diagnostic techniques. The bacteria are resistance to most antibacterial drugs, and have increasingly attracted the attention of researchers in related fields at domestic and foreign (Brooke 2012).

Clustered regular interspaced short palindromic repeats (CRISPR) is an adaptive immune system for foreign genes that is widely found in archaea and bacterial genomes. The CRISPR system consists of CRISPR-associated gene (*cas* gene), a leader sequence (LS) and a CRISPR locus. The CRISPR locus are mainly composed of different numbers of direct repeats

(DRs) and spacer sequences (Karimi et al., 2018). According to the composition of CAS protein and the structure of CRISPR locus, the CRISPR system can be divided into 5 types and 16 subtypes (Makarova et al., 2011). The CRISPR system is equivalent to the bacterial acquired immune system, which can block foreign DNA such as phage and plasmid from invading bacteria and avoid interference from foreign DNA (Mali et al., 2013). The CRISPR system has attracted widespread attention due to its unique structure and function (Wright et al., 2016).

In the laboratory study of the genetically engineered technology of *S. maltophilia*, it has been found that it is difficult to transfer foreign DNA from different sources into the strain, the success rate was relatively low, and the construction of recombinant expression plasmids is easily lost. Therefore, it is speculated that the resistance to foreign plasmids may be related to the CRISPR system. At present, the research on CRISPR mainly focuses on *E. coli* (Rahmatabadi et al., 2016), *S. aureus* (Zhao et al., 2018), and *Salmonella* (Shariat et al., 2015). The research on the bacterium of the genus *Stenotrophomonas* has not been reported. Therefore, the bioinformatics method was used to analyze the

This article is published under the terms of the Creative Commons Attribution License 4.0

Author(s) retain the copyright of this article. Publication rights with Alkhaer Publications.

Published at: <http://www.ijsciences.com/pub/issue/2019-03/>

DOI: 10.18483/ijSci.1978; Online ISSN: 2305-3925; Print ISSN: 2410-4477



Hengxia Fu (Correspondence)



gene composition of the CRISPR system of the genus *Stenotrophomonas*.

## 2. MATERIALS AND METHODS

### 2.1 Strain Selection

The CRISPR information of the 13 strains of *Stenotrophomonas* that have been published in the CRISPRdb database (<http://crispr.u-psud.fr/>) (as of December 25, 2018) was collected (Grissa et al., 2007). The genome sequences of 13 strains of *Stenotrophomonas* were downloaded from the National Center for Biotechnology Information (NCBI) network database (<http://www.ncbi.nlm.nih.GOV/>). The GenBank numbers are shown in Table 1.

### 2.2 analysis of CRISPR locus

For published genome sequences in CRISPR database, CRISPR loci were available through the database directly. Alternatively, genome sequences downloaded from GenBank at the NCBI, the detection of CRISPR loci in the downloaded strains were achieved using CRISPRFinder (<http://crispr.u-psud.fr/>).

### 2.3 Analysis of repeat sequences and spacer sequences

The repeat sequences in the obtained CRISPR loci were aligned by ClustalW, then repeat sequences were input into MEGA7.0 software for comparison analysis, and a phylogenetic tree was constructed. RNA secondary structure and minimum free energy (MFE) prediction by RNAfold with default arguments, and the conservation of repeats was represented by WebLogo. Blast analysis was performed on the spacer sequences in the locus to find foreign gene elements that is homologous to the sequence.

### 2.4 Statistical analysis of spacer sequences and phage sequences

The phage sequences in the *Stenotrophomonas* genome were predicted by Prophinder (<http://aclame.ulb.ac.be/>), and the correlation analysis between the spacer and the number of phage sequences was performed using the Spearman rank correlation test.

## 3. RESULTS

### 3.1 Distribution of CRISPR arrays

In the 26 strains that were analyzed, a total of 15 confirmed CRISPR arrays and 132 questionable structures were identified. For 15 confirmed CRISPR arrays, the number of confirmed CRISPR arrays identified in the *S. maltophilia* EPM1 (GCF\_000344215) genome is the highest. The second is *S. Maltophilia* (GCF\_000742995), there are 3 confirmed CRISPR arrays, the rest of the strain genome contains only one confirmed locus. Most of these CRISPR-cas locus were located on their

chromosome. 4 strains with more complete CRISPR arrays information, the structure of the CRISPR arrays are shown in Figure 1. The CRISPR system of *S. Acidaminiphila* JCM 13310, *S. Nitritireducens* DSM12575 and *Stenotrophomonas* sp. Leaf70 has the same structure and belongs to the I-F type, *S. ginsengisoli* DSM24757 belongs to the I-C type.

### 3.2 Repeat Sequence

#### 3.2.1 General overview of Repeat Sequence

A total of 147 CRISPR arrays were found in the 26 strains, of which 15 were confirmed CRISPR arrays, and the rest were questionable structures. Each CRISPR arrays contained more than 2 repeats. The repeat sequence is 25-32bp in length, while 25 bp is more common among all the repeats (Table 1). Sequence alignments show that the repeat sequences are highly conserved in the same CRISPR array. The same repeats are also distributed among different strains and they are conserved. However, there are also differences in repeat sequences in the same strains, which are reflected in partial base mutations and differences in CRISPR structures (Table 1).

#### 3.2.2 RNA secondary structures and MFE of repeats

We predicted the sequence stability and conservation of DR by RNA secondary structure and MFE values. The secondary structures as well as the MFE of 8 repeats were predicted by RNA fold, and found that all of the repeats could formed a conserved RNA secondary structure that resembles a dumbbell shape (Fig.2). The RNA secondary structure of the stem length is generally 4, 5, 6, 7 and 8bp. In addition to the stem length, the length of the repeat sequence and the "GC" content can also affect the stability of the secondary structure, repeats with higher "GC" content are more stable with the same length. In summary, DR sequences with lower MFE values are more stable than DR sequences with high MFE values. The formation of the RNA secondary structure of the Repeat sequence suggests that it may mediate the interaction of invasive DNA or RNA with the CAS-encoded protein. Conservative analysis of the repeat sequences was performed by Weblogo and the results are shown in Fig. 3. The results show that different *Stenotrophomonas* are intertwined and cannot be separated from each other, so this sequence cannot be used for classification and identification.

### 3.3 spacer sequences

There are 376 spacer sequences in 15 confirmed CRISPR arrays, ranging in length from 18 to 60 bp (Table.1). Although the spacer sequences are different in different CRISPRs in the same species, but at *S. Acidaminiphila* JCM 13310, *S. Nitritireducens* DSM12575 and *Stenotrophomonas* sp. Leaf70 the middle length of the spacer sequence is almost at 32 bp,

*S. maltophilia* Ab55555 has the shortest spacer, only 18 bp. Homologous analysis of 376 spacers by NCBI BLAST revealed that most of the spacers matched the chromosome regions of other bacteria, only 1.3% (5/376) matched the plasmid or viral sequences.

### 3.4 Correlation analysis between the spacer sequence and the number of phages

13 strains of *Stenotrophomonas* downloaded from NCBI and 3 strains of *S. maltophilia* JV3, *S. maltophilia* K279a, *S. maltophilia* R5513 with complete genome sequence were predicted by Prophinder for phage sequence. It was found that not every *Stenotrophomonas* genome contains a phage, for example, up to five of *S. Maltophilia* AU1209 and *S. Maltophilia* Ab55555, *S. Maltophilia* K279a and *S. maltophilia* 13637. Using the Spearman rank correlation coefficient, it was found that there was a negative correlation between the number of spacer sequences and the number of phage ( $\rho = -0.538$ ,  $P = 0.039$ ). The negative correlation between the number of spacer sequences and the number of phages indicates that as the number of spacer sequences increases, the number of phage cells gradually decreases.

## 4. DISCUSSION

CRISPR-Cas system is an acquired immunity system that provides defense against invading genetic elements, such as bacteriophages and plasmids (Redman et al., 2016; Barrangou et al., 2007; Marraffini et al., 2008). The CRISPR-Cas immunity system functions in three steps: adaptation, expression, and interference (Jiang et al., 2013). Most of the prokaryotic CRISPR loci are located on their chromosomes. The main reason for this phenomenon is that the CRISPR system can provide clearance of exogenous phage and plasmid through the targeted interference mechanism of RNA. This also makes prokaryotes very adaptable in the face of the environment. In recent years, the gene editing technology based on CRISPR has been widely used in the fields of genetic engineering and genetic diseases because of its simple operation and high efficiency, and it has become a new research hotspot (Reed 2017). Therefore, the research and analysis of the CRISPR locus structure of bacteria is of great significance for the study of its function and the development and application of gene editing tools.

CRISPR was first discovered in *E. coli*, and many studies have shown that this system exists in 90% of archaea and 40% of bacteria (Hille et al., 2018). *Stenotrophomonas* is not only common environmental bacteria, but also an important conditional pathogen in nosocomial infections. Recently, due to its increasing drug resistance and morbidity, it has attracted people's attention (Hauben et al., 1999). In the process of using

the genetic engineering method to study the bacteria, it was found that the plasmid transformation, homologous recombination, modification of the bacteria has lower success rates than most other bacteria. It is suggested that *Stenotrophomonas* can resist the invasion of foreign DNA such as phage and plasmid, which may be related to the immune function of CRISPR system. Therefore, this study used bioinformatics methods to analyze the genetic structure of the CRISPR system in *Stenotrophomonas*.

In this study, 15 confirmed and 132 questionable CRISPRs were identified by analyzing the CRISPR locus of 26 strains of *Stenotrophomonas* published by the CRISPRdb database and NCBI. It is slightly lower than the record percentage of CRISPR-containing genomes in the *P. aeruginosa*, which is 61.05% (van, Belkum et al., 2015). The system exists in all 26 strains, and some of the same CRISPRs can be distributed among different strains, which are conservative, suggesting that the strains evolved in the same direction due to the same foreign DNA invasion. In addition, the variability of the CRISPR structure may reveal that different strains have different resistance to foreign DNA interference, and also suggest the diversity of foreign DNA invasion.

Repeat sequences are also conserved in different strains, suggesting that different strains may form the same defense system during genetic evolution. Evolutionary analysis of the *Stenotrophomonas* repeats revealed that the same repeats were distributed among the different strains, suggesting that CRISPR is mainly transmitted between strains and independently evolved through a horizontal transfer mechanism. One study has implicated that stem-loop structures of some repeats may serve to facilitate recognition and mediated contact between the spacer-targeted foreign RNA or DNA and Cas-encoded proteins (Kunin et al., 2007). MFE of *Stenotrophomonas* is lower than that of other bacteria (Yang et al., 2015), and the bacteria have longer "stems". therefore, they may form more stable secondary structures. Studies have shown that the stability of RNA secondary structure may affect the function of the CRISPR locus. From Figure 3 we can see that the repeats of *Stenotrophomonas* are not conserved.

The average length of the spacer sequence in the *Stenotrophomonas* confirmed CRISPR-cas loci is the most common at 30bp, similar to most other bacteria (Barrangou et al., 2014). The correlation analysis between the spacer sequence and the number of phages showed that there was a negative correlation between the number of spacer sequences and the number of phages, which is consistent with the conclusion of Nozawa et al. on *Streptococcus pyogenes*, that is, the greater the number of CRISPR locus spacer sequences

(Nozawa et al. 2011). the smaller the number of phages in the genome, suggesting that CRISPR limits the invasion of phage to *Stenotrophomonas*.

In this study, the bioinformatics analysis of the CRISPR gene structure in the genome of *Stenotrophomonas* contributed to the study of drug resistance and pathogenicity and CRISPR-mediated gene knockout technology, which laid the foundation

for future research and development. Since the genome sequence of the *Stenotrophomonas* strains that have been published for analysis in the GenBank database is relatively small, and some of the information on the full gene annotation is not comprehensive, it is not possible to comprehensively analyze the *cas* gene in the CRISPR system. Therefore, this result has yet to be further improved.

## 5 TABLE AND FIGURES

Table.1 Genome sequence information of *Stenotrophomonas*

Stains	CRISPR ID	Number of CRISPR	CRISPR length	Number of spacers	Spacer length	DR length
<i>S.acidaminiphila</i> GCF_001314305	NZ_CP012900_1	1,1	300	5	30,33,33,30,30	25
<i>S.rhizophila</i> GCF_001704155	NZ_CP016294	0,2				
<i>S.maltophilia</i> EPM1 GCF_000344215	NZ_CM001824	4,15	190	3,1,1,1	30,60,55,47	25
<i>Stenotrophomonas</i> sp. LM091 GCF_001806305	NZ_CP017483.1	0,1				
<i>S.maltophilia</i> GCF_000742995	NZ_CP008838	2,4	89,146	2,1	37,53	25,26
<i>S.maltophilia</i> D457	NC_017671.1	0,7				
<i>S.maltophilia</i> GCF_001274595	NZ_CP011010	1	105	1	54	26
<i>S.maltophilia</i> strain ISMMS2 GCF_001274655.1	NZ_CP011306.1	0,3				
<i>S. maltophilia</i> GCF_001274655	NZ_CP011305_2	0,1				
<i>S. maltophilia</i> 13637 GCF_000742995.1	NZ_CP008838_8	1,9	104	1	53	26
<i>S. maltophilia</i> GCF_001274675	NZ_CP011306	0,1				
<i>S. maltophilia</i> ATCC19867 GCF_000382065.1	NZ_KB907511.1	0,2				
<i>S. maltophilia</i> JV3 GCF_000223885.1	NC_015947.1	0,1				
<i>S. maltophilia</i> AU12-09 GCF_000346445.1	NZ_APIT00000000.1	0,6				
<i>S. maltophilia</i> K279a GCF_000072485.1	AM743169.1	0,14				
<i>S. maltophilia</i> Ab55555 GI: 401461733	NZ_JH791780.1	1	146	3	18	23
<i>S. maltophilia</i> R551-3 GCF_000020665.1	NC_011071.1	0,6				
<i>S. acidaminiphila</i> JCM13310 GCF_001431595.1	NZ_LDJO01000043.1	1	2009	33	32	28
<i>S. nitritireducens</i> GCF_001700965	NZ_CP016756	0,12				
<i>S. acidaminiphila</i> ZAC14D2_NAIMI4_2 GCF_001314305.1	NZ_CP012900.1	1,1	300	5	30	25
<i>S. maltophilia</i> GCF_002025605	NZ_CP018756	0,8				
<i>S. ginsengisoli</i> strain DSM24757 GCF_001431485.1	NZ_LDJM01000011.1	1	5541	83	34	32
<i>Stenotrophomonas</i> sp. KCTC12332 GCF_001562215	NZ_CP014274.1	0,23				
<i>S. nitritireducens</i> DSM12575 GCF_001431425.1	NZ_LDJG01000008.1	1,1	6031	100	32	28
<i>Stenotrophomonas</i> sp. YM1 GCF_001562215.1	GI: 1011253290	0,23				
<i>Stenotrophomonas</i> sp. Leaf70 GCF_001422025.1	NZ_LMLT01000010.1	1,1	6389	106	32	28

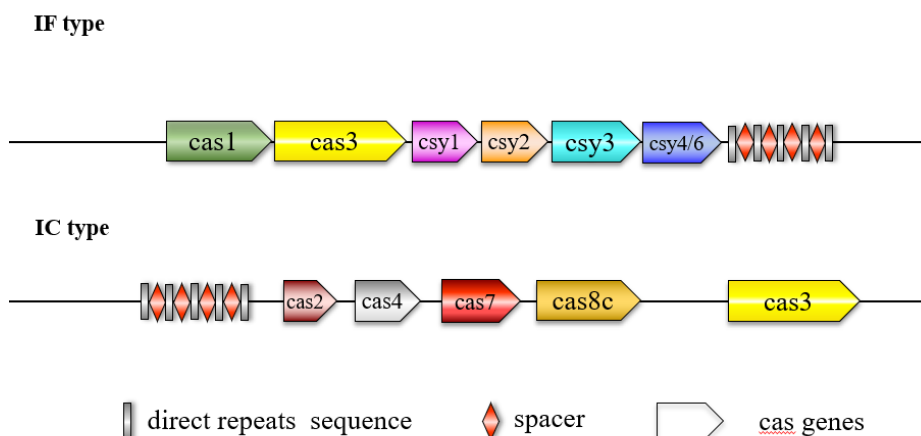


Fig.1 Structure diagram of CRISPR locus structure in *Stenotrophomonas*

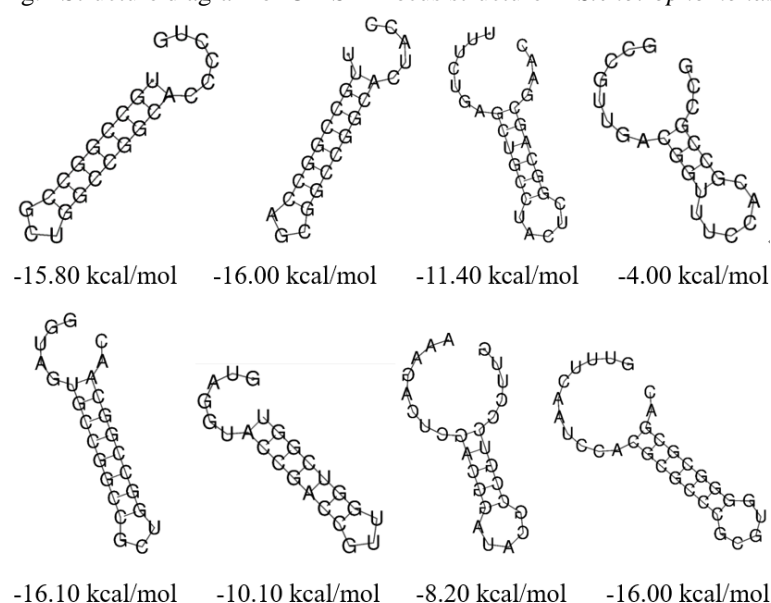


Fig.2. The RNA secondary structures and the MFE of the repeat sequences.

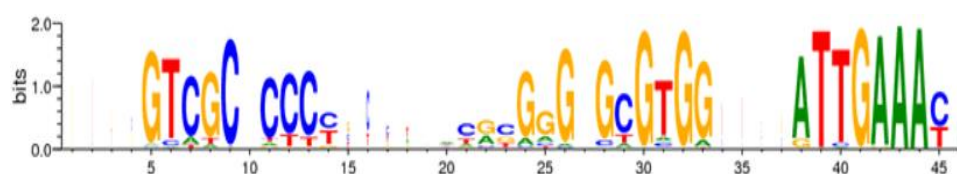


Fig.3. Repeat sequences family conservation analysis.

### Conflicts of interest

The authors declare that they have no conflict of interest.

### REFERENCES

- Alavi, P., Starcher, M.R., Thallinger, G.G. (2014). *Stenotrophomonas* comparative genomics reveals genes and functions that differentiate beneficial and pathogenic bacteria. *BMC Genomics*, 15(1): 482-496. doi: 10.1186/1471-2164-15-482.
- Ribitsch, D., Heumann, S., Karl, W. (2012). Extracellular serine proteases from *Stenotrophomonas maltophilia*: Screening, isolation and heterologous expression in *E. coli*. *J Biotechnol*, 157(1):140-7. DOI: 10.1016/j.jbiotec.2011.09.025
- Brooke, J.S. (2012). *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin Microbiol Rev*, 25(1):2-41. doi: 10.1128/CMR.00019-11.
- Karimi, Z., Ahmadi, A., Najafi, A., et al. (2018). Bacterial CRISPR Regions: General Features and their Potential for Epidemiological Molecular Typing Studies. *Open. Microbiol. J.* 12: 59-70. DOI: 10.2174/1874285801812010059

5. Makarova, K.S., Aravind, L., Wolf, Y.I. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct.* 6:38. DOI: 10.1186/1745-6150-6-38
6. Mali, P., Yang, L., Esvelt, K.M., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339: 823-826. DOI: 10.1126/science.1232033
7. Wright, A.V., Nuñez, J.K., Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell.* 164(1-2):29-44. DOI: 10.1016/j.cell.2015.12.035.
8. Rahmatabadi, S.S., Nezafat, N., Negahdaripour, M., et al. (2016). Studying the features of 57 confirmed CRISPR loci in 29 strains of *Escherichia coli*. *J. Basic. Microbiol.* 56: 645-653. DOI: 10.1002/jobm.201500707.
9. Zhao, X., Yu, Z., Xu, Z. (2018). Study the Features of 57 Confirmed CRISPR Loci in 38 Strains of *Staphylococcus aureus*. *Front Microbiol.* 9:1591. DOI: 10.3389/fmicb.2018.01591.
10. Shariat, N., Timme, R.E., Pettengill, J.B. (2015). Characterization and evolution of *Salmonella* CRISPR-Cas systems. *Microbiology.* 161(2):374-386. doi: 10.1099/mic.0.000005.
11. Grissa, I., Vergnaud, G., Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC. Bioinformatics.* 8 :172. DOI: 10.1186/1471-2105-8-172
12. Redman, M., King, A., Watson, C. (2016). What is CRISPR/Cas9? *Arch Dis Child Educ Pract Ed.* 101(4):213-5.
13. Redman, M., King, A., Watson, C. (2016). What is CRISPR/Cas9? *Arch Dis Child Educ Pract Ed.* 101(4):213-5.
14. Barrangou, R., Fremaux, C., Deveau, H. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 315(5819):1709-12. DOI: 10.1126/science.1138140
15. Marraffini, L.A., Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in *staphylococci* by targeting DNA. *Science.* 322 (5909): 1843-5. DOI: 10.1126/science.1165771
16. Jiang, W., Maniv, I., Arain, F. (2013). Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet.* 9(9): e1003844. DOI: 10.1371/journal.pgen.1003844
17. Reed, F.A. (2017). CRISPR/Cas9 Gene Drive: Growing Pains for a New Technology. *Genetics.* 205(3):1037-1039. doi: 10.1534/genetics.116.198887.
18. Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell.* 172 (6): 1239-1259. doi: 10.1016/j.cell.2017.11.032.
19. Hauben, L., Vauterin, L., Moore, E., Hoste, B., Swings, J. (1999). Genomic diversity of the genus *Stenotrophomonas*. *Int J Syst Bacteriol.* 49 (4): 1749-60. DOI: 10.1099/00207713-49-4-1749
20. van, Belkum, A., Soriaga, L.B., LaFave, M.C. (2015). Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant *Pseudomonas aeruginosa*. *MBio.* 6: e01796-1815. DOI: 10.1128/mBio.01796-15
21. Kunin, V., Sorek, R., Hugenholtz, P. (2007). Evolutionary conservation of Sequence and secondary structures in CRISPR repeats, *Genome. Biol.* 8: R61. DOI: 10.1186/gb-2007-8-4-r61
22. Yang, S., Liu, J., Shao, F. (2015). Analysis of the features of 45 identified CRISPR loci in 32 *Staphylococcus aureus*. *Biochem. Biophys. Res. Commun.* 464 :894-900. doi: 10.1016/j.bbrc.2015.07.062.
23. Barrangou, R., Marraffini, L.A. (2014). CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular Cell.* 54 (2): 234-44. doi: 10.1016/j.molcel.2014.03.011.
24. Nozawa, T., Furukawa, N., Aikawa, C. (2011). CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. *PLoS One.* 6(5): e19543. doi: 10.1371/journal.pone.0019543.