

## Bioinformatic analysis of *Listeria monocytogenes* CRISPR

Daofeng Qu<sup>1</sup>, Yang Shen<sup>1</sup>, Lin Xu<sup>1</sup>, Ziyang Xiang<sup>1</sup>, Lu Zhang<sup>1</sup>, Yuewen Chen<sup>1</sup>, Dongping Huang<sup>1</sup>, Anna Tong<sup>1</sup>, Liying Zhang<sup>1</sup> and Jianzhong Han<sup>1</sup>

<sup>1</sup>Food Safety Key Laboratory of Zhejiang Province, School of Food Science and Biotechnology, Zhejiang Gongshang University, Hangzhou 310035, PR China

Correspondence to: Jianzhong Han, email: jianzhonghan1@163.com

Keywords: *L. monocytogenes*; CRISPR; repeat; Cas; plasmid

Received: July 22, 2017

Accepted: December 24, 2017

Published: January 03, 2018

Copyright: Qu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

***Listeria monocytogenes* is a leading causes of death from food-borne pathogens. Bioinformatics approach was applied to investigate the features of *L. monocytogenes* CRISPR structure and the relationship between CRISPR and plasmid transposase content. Among 93 *L. monocytogenes* genomes, 95 confirmed CRISPR structure loci were identified and classified into 5 groups based on repeat size. RNA secondary structure and minimum free energy indicated that the secondary structure of Group 5 (36 bp) was more stable than other groups. Type I-B or II-A Cas genes were found in 36 strains, and the CRISPR-Cas system of type I-B was more conserved than type II-A. Furthermore, CRISPR loci affected the enzyme transposase content of *L. monocytogenes* plasmid. This study examined the diversity of the CRISPR-Cas system in *L. monocytogenes*, classified CRISPR structure and repeats, and demonstrated the influence of the CRISPR-Cas system on the number of transposase in plasmid.**

### INTRODUCTION

The gram-positive bacteria *L. monocytogenes* is a foodborne pathogen with high mortality rates. *L. monocytogenes* is a significant challenge in food production due to its ability to survive under conditions of salinity, alkalinity and temperature stress [1]. Extra chromosomal plasmids are relatively small compared to the bacterial chromosome and often harbor antibiotic resistant genes. Plasmids participate in the spread of antibiotic resistance genes through horizontal gene transfer, which can lead to environmental pressure adaptations such as enhancing virulence or resistance [2].

Clustered, regularly interspaced, short palindromic repeats (CRISPRs) encoded by CRISPR-associated (Cas) genes have been identified in many bacteria, and CRISPR-Cas systems provide an adaptive immune response to genetic elements such as plasmids, phages, insertion sequences, transposons, and integrons [3, 4]. The CRISPR structure has three major features: a set of Cas genes, an AT-rich leader sequence, and palindromic direct repeats

separated by variable sequences called spacers [5]. The repeats are highly conserved, always contain palindromic motifs, and may constitute RNA secondary structure [6]. Unique spacer sequences are usually derived from mobile genetic elements such as plasmids and phages [7] while Cas genes are often adjacent to the CRISPR loci. Two CRISPR loci were recently identified in the *L. monocytogenes* genome [8], and are associated with type I-B and type II-A Cas genes [9].

We analyzed 93 *L. monocytogenes* genomic nucleotide sequences from the NCBI database. The structural characteristics of CRISPR-Cas were investigated by bioinformatic method. CRISPR loci were categorized based on the size of repeats and the structure of Cas genes. We investigated the plasmid genetic content of transposase in *L. monocytogenes* strains with different CRISPR loci. This study demonstrated the diversity of CRISPR-Cas system in *L. monocytogenes* strains, identified the features of CRISPR structure and repeat classification, and elucidated the influence of the CRISPR-Cas system on the number of transposase in plasmid.

## RESULTS

### CRISPR loci of *L. monocytogenes*

We selected all publicly available *L. monocytogenes* complete genomes from the NCBI database. Only 24 strains do not contain CRISPR loci, accounting for 25.8%. The other 69 strains (74.2%) possessed between 1 and 3 CRISPR loci. We selected confirmed chromosomal CRISPR sequences for further investigation. According to CRISPRdb and Guo et al. [10], confirmed CRISPR should contain at least two different spacers. A total of 95 confirmed CRISPR loci were detected among 93 *L. monocytogenes* genomes. These loci were classified into 5 groups according to direct repeat length. The number of spacers ranged from 3 to 58, and the number of direct repeats was between 4 and 59 (typically 28, 29, or 36 bp, Supplementary Table 1).

### Direct repeats of RNA secondary structure

Since the direct repeat length of CRISPR loci is similar within each locus, we selected the same length direct repeat sequences for multiple sequence alignment analysis. Based on the alignment, 97 CRISPR loci in 69 *L. monocytogenes* strains were assigned to 5 groups with the same direct repeat length (Table 1). The direct repeat length was between 23 and 36 bp, and typically 28, 29 or 36 bp, accounting for 24.74%, 43.3%, and 22.68%, respectively. We utilized WebLogo to analyze the representative repeats of the same size direct repeat to better understand the features (Figure 1A). Previous studies have suggested that CRISPR repeats may form stable hairpin-like secondary structures due to the partially palindromic nature [11, 12]. The RNA secondary structure and minimum free energy (MFE) were detected for representative direct repeat sequences of each group through the RNAfold Web Server (Figure 1B). In all groups except Group 5, RNA secondary structure was composed of two rings at each end and a stem in the middle. The stem length in Group 5 was 10 bp, while the length was 4 and 6 bp in other groups. The MFE of Group 5 ( $\Delta G = -6.70$  kcal/mol) was less than other groups ( $P < 0.05$ ), indicating a more stable RNA secondary structure than those of other groups due to the great number of base pairs in the stem.

### Structural features of *L. monocytogenes* CRISPR/Cas

Previous studies have suggested that *L. monocytogenes* CRISPR loci are associated with type I-B or type II-A Cas genes [13]. We searched for Cas genes from 10,000 bp upstream to 10,000 bp downstream the CRISPR loci in the NCBI database. Two CRISPR-Cas types were found in 36 strains (Supplementary Table 2). For CRISPR-Cas type I, the architecture is conserved

with four Cas genes (*csn2*, *cas2*, *cas1*, *cas9*) located downstream of the repeat-spacer region. In contrast, the content and organization for CRISPR-Cas type I vary, with 6-8 Cas genes (*cas2*, *cas1*, *cas4*, *cas3*, *cas5*, *cas7*, *cas8b1*, *cas6*) located downstream of the repeat-spacer region. To better understand the features of the CRISPR-Cas system, 5 representative strains (*L. monocytogenes* HCC23, *L. monocytogenes* Finland 1998, *L. monocytogenes* 10-092876-0055 LM4, *L. monocytogenes* 10-092876-1763 LM10, *L. monocytogenes* J01611) were chosen for further study. Although Cas gene sequence similarity is high within CRISPR-Cas type I, the gene organization is different (Figure 2). Interestingly, *cas2* is the only conserved gene among the five loci.

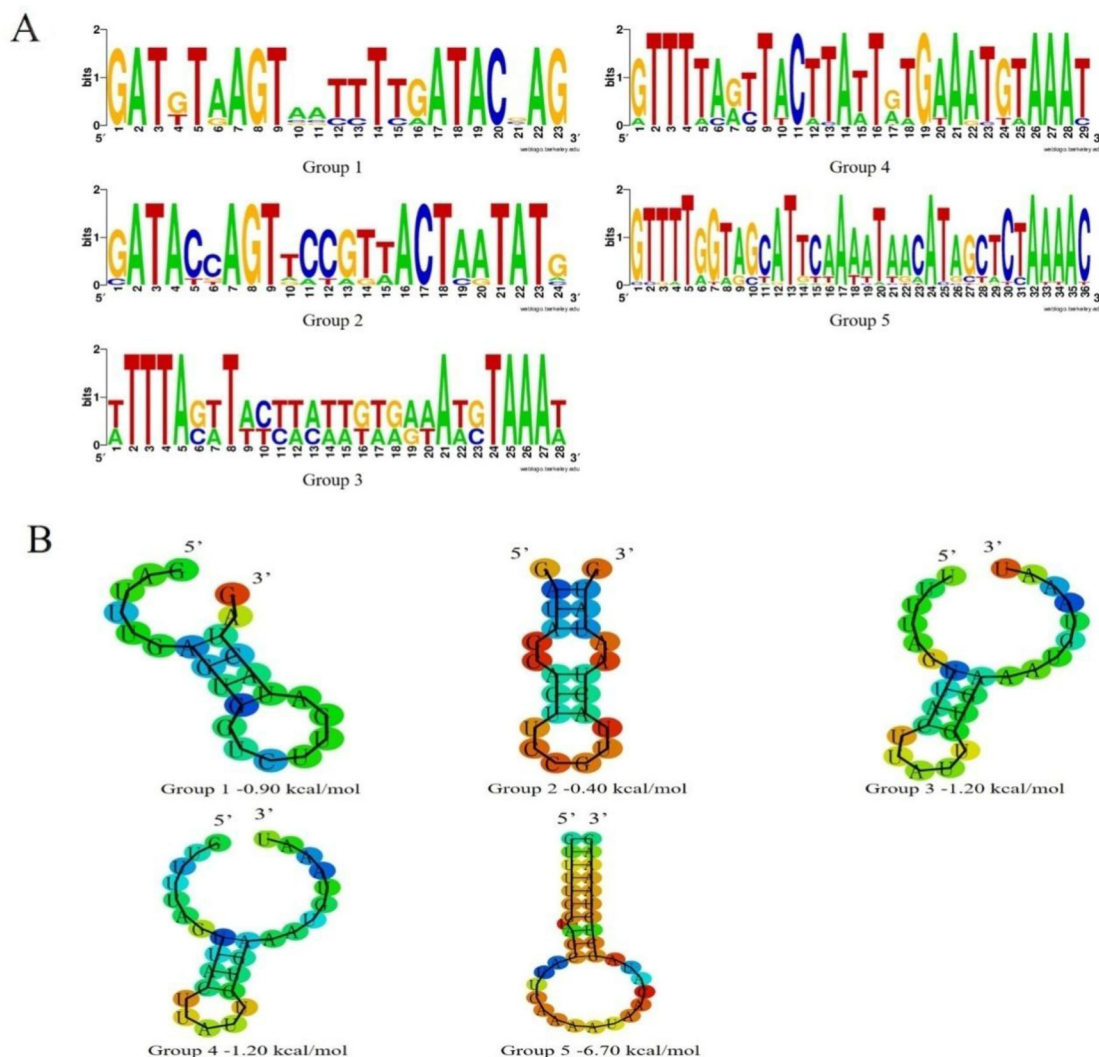
Since we observed a correlation between *cas2* genes and CRISPR repeats in *L. monocytogenes*, we investigated whether there was a relationship between CRISPR repeats and *cas2* genes across bacterial strains. Across a variety of strains, the clustering of the typical CRISPR repeats was similar to that of the *cas2* genes and consistent with previous observations by Horvath et al. [14]. Comparative analysis of the evolutionary trees revealed similar clustering patterns, with different clusters for two CRISPR-Cas types. Sequence alignments are provided in the supplemental material (Supplementary Table 3). Although the trees were based on widely different element sizes (the direct repeat size varied between 29 and 36 bp, while *cas2* varied between 279 and 342 bp), the congruence between them is relatively high (Figure 3). This observation suggests coevolution of *cas2* genes and CRISPR repeats, indicating a potential functional link.

### The relationship between spacers and repeats

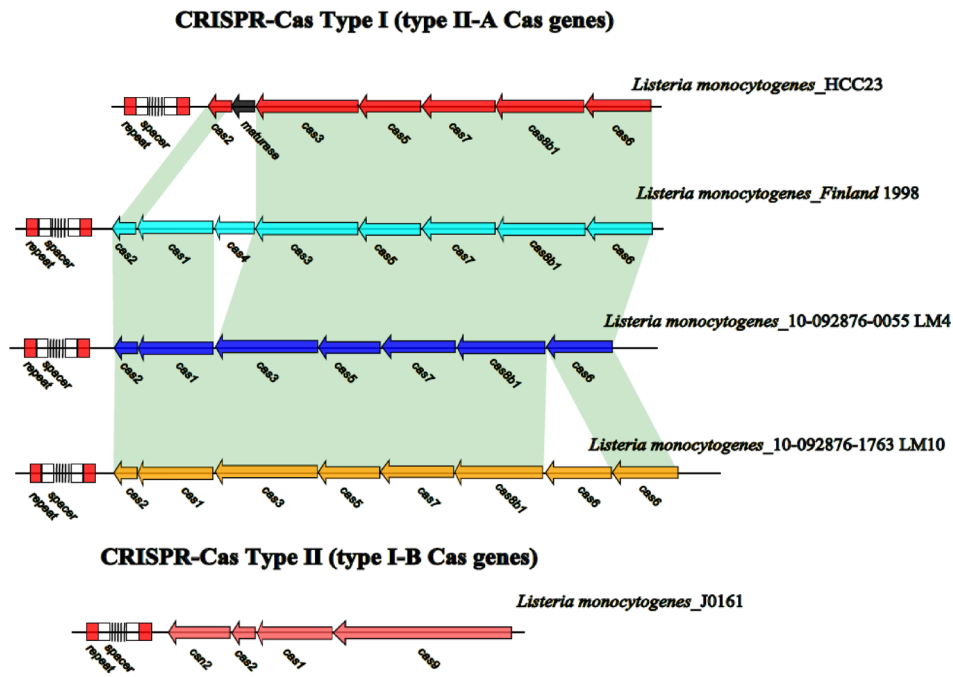
A total of 1417 spacers (5 for Group 1, 24 for Group 2, 166 for Group 3, 657 for Group 4, and 565 for Group 5) were found with 5, 6, 20, 310, and 221 unique spacers in Groups 1, 2, 3, 4, and 5, respectively. Besides Group 1, the degree of polymorphism regarding unique spacers was the highest in Group 4 ( $P < 0.05$ ) (Table 2). Polymorphisms were also observed regarding spacer size. Analysis of the spacer size distribution indicated that variability was greatest in Group 1 ( $P < 0.05$ ). The spacer size was 55 bp and 54 bp for Group 1 and Group 2, respectively. The typical spacer size was 36 bp, ranging from 35 to 43 bp in Group 3 and 30 to 49 bp in Group 4, compared to a typical spacer size of 30 bp ranging from 29 bp to 36 bp in Group 5. The proportions of spacers of typical size were 54% (90 of 166), 36% (239 of 657), and 93% (525 of 565), for Group 3, Group 4, and Group 5, respectively (Figure 4A-4E). The spacer length has been shown to influence the activity of CRISPR loci [15]. Our data indicate a negative correlation between the size of repeat and spacer (Figure 4F). We further hypothesize that repeats are related to spacer size and change the activity of CRISPR loci, but this requires further investigation.

**Table 1: Grouping of direct repeats based on the number, length, and percentage of CRISPR loci in each group**

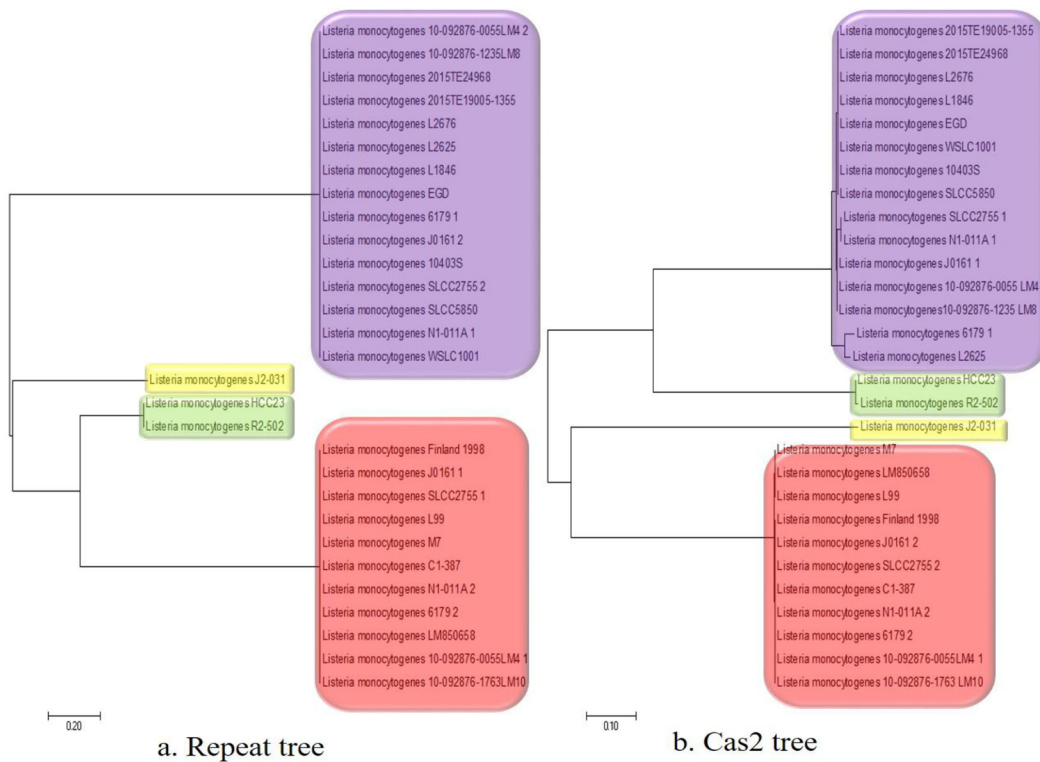
Group	Number of CRISPR	Typical DR consensus	DR length (bp)	percentage (%)
1	1	GATGTAAGTAATTTTAATACGAG	23	1.03
2	8	GATACCAGTCCGTTACTAATATG	24	8.25
3	24	TTTTAGTTACTTATTGTGAAATGTAAT	28	24.74
4	42	GTTTTAGTTACTTATTGTGAAATGTAAT	29	43.3
5	22	GTTTTGGTAGCATTCAAATAACATAGCTCTAAAAC	36	22.68



**Figure 1:** The WebLogo-generated (A) and RNA secondary structure of repeats of five groups (B). WebLogogenerated the typical sequence from multi sequence alignment analysis. Group 1 contains 6 typical sequences; Group 2 contains 8 typical sequences; Group 3 contains 24 typical sequences; Group 4 contains 42 typical sequences; Group 5 contains 22 typical sequences. The sequence of secondary structure was the typical sequence from multi sequence alignment analysis. The numbers indicate MFE: structures with a lower MFE are more stable than those with a higher MFE value.



**Figure 2:** *L. monocytogenes* CRISPR-Cas loci. *L. monocytogenes* have two CRISPR loci, CRISPR-Cas type I type II-A Cas genes and CRISPR-Cas type II type I-B Cas genes, both encoded on the antisense strand. There are 6-8 Cas genes located upstream of CRISPR-Cas type I and 4 Cas genes located upstream of CRISPR-Cas type II, indicated with boxed arrows. Shaded regions denote regions of homology (>95% nucleotide identity).



**Figure 3:** The evolutionary tree of repeats and *cas2*. (A) The evolutionary tree of repeats. (B) The evolutionary tree of *cas2*. The repeats and *cas2* genes have 29 strains, respectively. Strains located in one group indicate most evolutionary similarity. The evolutionary distance scale of repeats and *cas2* is 0.20 and 0.10, respectively. Boxes with different colors represent different groups.

**Table 2: Analysis of the number and percentage of unique CRISPR spacers in different strains**

Group	Number of spacers	Unique spacers	Percentage (%)
1	5	5	100.0
2	24	6	25.0
3	166	20	12.0
4	657	310	47.1
5	565	221	39.1
Total No. of spacers	1417	562	39.7

### The transposase of plasmid

Previous studies have suggested that CRISPR-Cas systems provide an adaptive immune response to bacteriophages and plasmids [11]. So we analyzed the characteristics and structure of 5 *L. monocytogenes* strains (J1962, HB5622, 2015TE, 6179, J1-208) plasmid sequence to find the relationship between CRISPR loci and plasmid. The J1962 genome contains no CRISPR loci, the HB5622 chromosome sequence contains one locus without a Cas gene, the 2015TE chromosome sequence contains one locus with atype I Cas gene, the 6179 chromosome sequence contains two loci with type I and type II Cas genes, and the J1-208 plasmid contains two CRISPR loci without a Cas gene. The complete sequence of the J1962 plasmid and J1-208 plasmid have circularly closed DNA sequences and contain 64 and 75 total predicted open reading frames, respectively (Figure 5A and 5B). The modular structure of each plasmid is regarded as the backbone with the insertion of multiple separate accessory modules. Linear comparison of sequence plasmids indicated the transposase content in 5 plasmids (Figure 5C). The length of the J1-208 plasmid containing two CRISPR loci is 77.83 kb, but transposase constitutes 1.0%. The length of the HB5612 plasmid, which is similar to J1-208, is 77.11 kb, but transposase constitutes 11.3%. The sizes of J1926, 2015TE and 6179 plasmid sequences are smaller, and transposase constitutes 15.7%, 14.3%, and 6.9%, respectively (Figure 5D). The percentage of transposase in the J1-208 plasmid is significantly lower than the others ( $P < 0.05$ ) and we postulate that this is related to the presence of CRISPR on the plasmid.

### DISCUSSION

We provided thorough sequence analysis and characterization of the CRISPR-Cas system in *L. monocytogenes*. Some *L. monocytogenes* CRISPRs have been previously identified [16], and investigated CRISPR diversity in *L. monocytogenes* strains of different lineages to estimate the potential practicability of a CRISPR-based approach in resolving this species' biodiversity. Bioinformatic analysis of distributions and features of CRISPR in our study may elucidate its function in *L. monocytogenes*. Confirmed CRISPR loci from the CRISPRdb contain at least two unique spacers, while questionable CRISPR only contain one unique

spacer [17]. We selected confirmed chromosomal CRISPR sequences and defined a total of 95 confirmed CRISPR loci within 93 genomes.

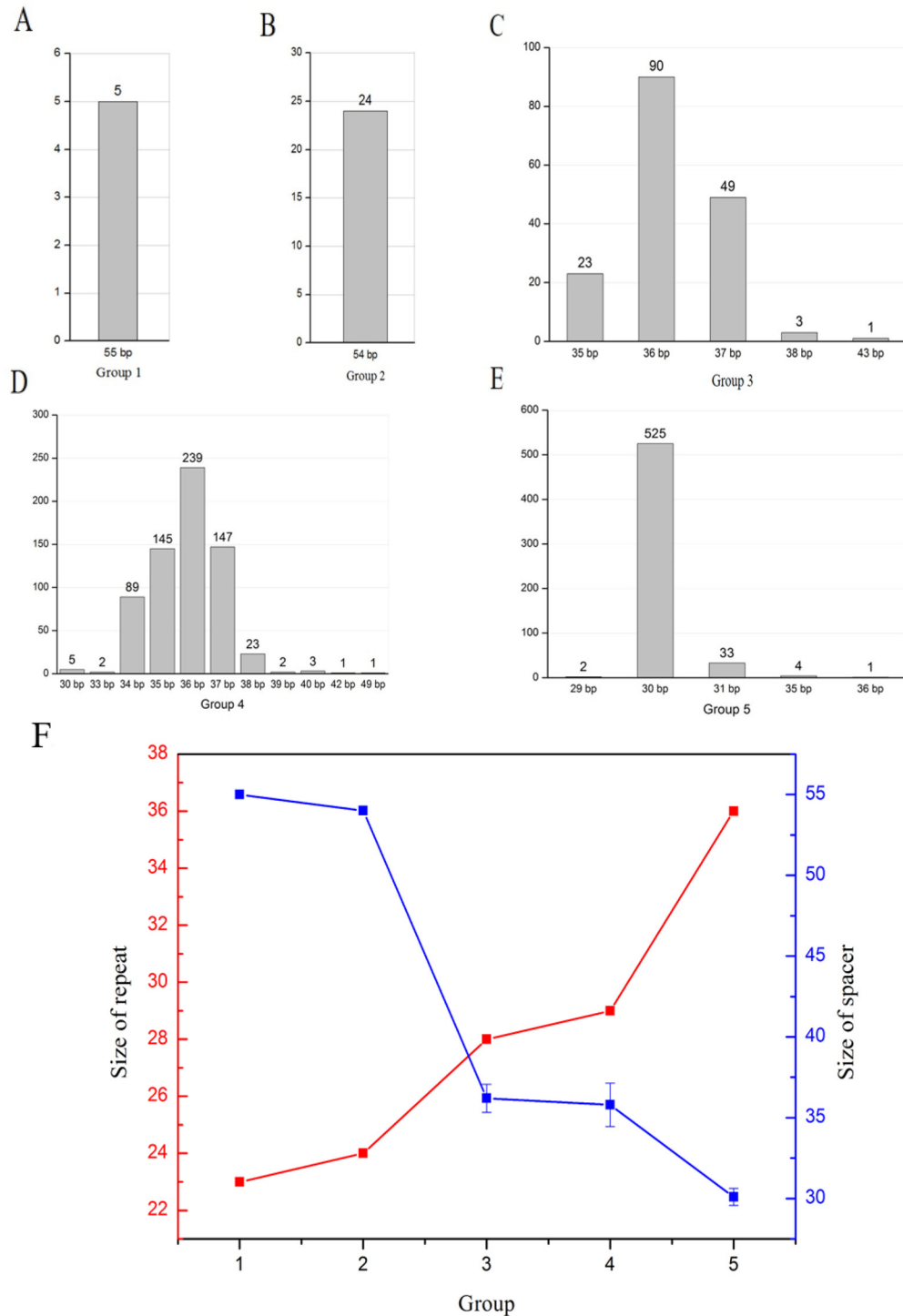
We found that there can be one or several modified nucleotides in the same size repeats of different CRISPR loci through multiple sequence alignment analysis, but they are frequently conserved. Therefore, we grouped CRISPR loci into 5 groups according to repeat size. The RNA secondary structure and MFE of the direct repeats were also investigated. Since the repeats can undergo polymorphism, particularly in the terminal repeat [18], we analyzed the secondary structure of typical repeat sequences of each group. The low MFE of direct repeats in 69 strains indicated the formation of a stable RNA secondary structure [19]. Our data indicated that the RNA secondary structure of repeats in Group 5, the longest repeat size, is most stable ( $P < 0.05$ ). Therefore, we postulated that longer repeats have a more stable secondary structure because there are more nucleotide base pairs. Previous studies indicated that stem-loop structures of some direct repeats facilitate contact between foreign RNA or the DNA targeting spacer and Cas-encoded proteins [20]. Moreover, the stability of RNA secondary structures may strengthen the function of CRISPR loci.

The two types of Cas genes located in the vicinity of CRISPR loci were identified from the NCBI database and were consistent with previous studies [18]. A total of 35 bacteria strains contained Cas genes near CRISPR loci. The architecture of CRISPR-Cas type I can undergo polymorphism and contains more Cas genes than CRISPR-Cas type II. Despite the architectural differences of these CRISPR-Cas systems, *cas2* was ubiquitous. We analyzed the relationship between *cas2* and repeats of different CRISPR loci. Interestingly, our data suggested the potential co-evolution of *cas2* genes and CRISPR repeats, indicating a potential functional link between them [21].

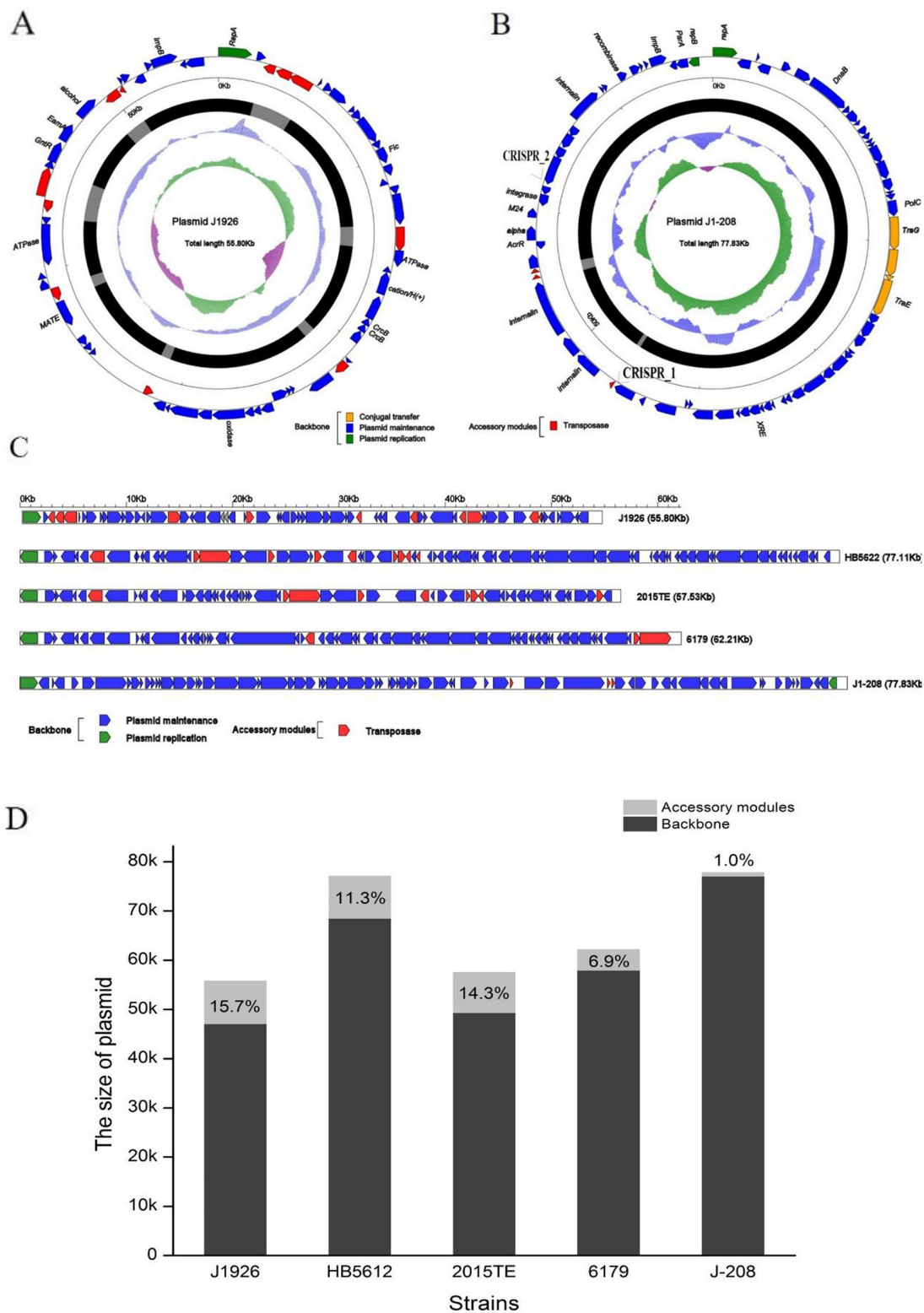
Spacers are located in the CRISPR locus near the leader sequence. Spacer diversity is well studied in strains of *Salmonella* and *E. coli* [22, 23], and the length and sequences of spacers affect the activity of CRISPR systems in bacteria. Di *et al.* determined that CRISPR loci containing more spacers with a length of 30 bp were more active than those containing fewer spacers with a length of 36 bp, which indicated a link between spacer number and length on CRISPR loci activity [2]. We identified 1417

spacers, and the data indicated a negative correlation between repeat size and spacer number. We posited that the relationship between repeats and spacer size mitigates

the activity of CRISPR loci and that the repeat-spacer unit length is genetically regulated, but these theories require further study.



**Figure 4: CRISPR spacer size variability.** The relationship between the size of repeat and spacer among five groups: (A) Group 1 spacers; (B) Group 2 spacers; (C) Group 3 spacers; (D) Group 4 spacers; (E) Group 5 spacers. The x-axis represents the size of a CRISPR spacer, in nucleotides. The y-axis represents the number of CRISPR spacer sequences of a given size. (F). The x-axis represents the groups. The left (red) y-axis represents the size of CRISPR repeat. The right (blue) y-axis represents the size of CRISPR spacer. The size of repeat and spacer were negatively correlated.



**Figure 5:** Schematic maps of (A) plasmid J1926 and (B) plasmid J1-208. Arrows denote genes and are colored based on gene function classification. The innermost circle presents GC-Skew  $[(G-C)/(G+C)]$  with a window size of 500 bp and a step size of 20 bp. The blue circle presents GC content. Shown also are backbone and accessory module region. (C) Linear comparison of sequence plasmid. Arrows denote genes and are colored based on gene function classification. The blue arrows represent plasmid maintenance; the green arrows represent plasmid replication; the red arrows represent transposase. Percentage of accessory modules among five strains of plasmid (D). The gray box represents the accessory modules. The black box represents the backbone.

Transposons are highly evolved forms of smaller moveable DNA segments, termed insertion sequences, that are on the order of 700–1500 bp in size and encode a specific recombinase (transposase) to facilitate movement [24]. We analyzed the characteristics and structure of plasmid sequences in 5 *L. monocytogenes* strains to identify the relationship between CRISPR loci and plasmid transposase. We found that the transposase percentage of the J1-208 plasmid containing two CRISPR loci was lower than the others ( $P < 0.05$ ), suggesting a relationship to the CRISPR on the plasmid. This study only identified one plasmid with CRISPR loci; further study requires more CRISPR loci-containing plasmid.

## MATERIALS AND METHODS

### Sequence collection

We analyzed 93 publicly available *Listeria monocytogenes* complete genomes from National Center for Biotechnology Information (NCBI) nucleotide database (<https://www.ncbi.nlm.nih.gov/genome/>) with default parameters. CRISPR gene signatures were searched in the CRISPRdb (<http://crispr.i2bc.paris-saclay.fr/crispr/>), we obtained the flanking sequences and repeat sequences of *Listeria monocytogenes* CRISPR. The conserved sequences upstream of the first repeat and downstream of the last repeat were obtained by multiple sequence alignment. These sequences are regarded as the specific gene signature for *L. monocytogenes* CRISPR [25]. The conserved sequences were utilized to search the arrays of 93 publicly available *L. monocytogenes* genomes.

### Analysis method

We downloaded the sequence of CRISPR loci from 10,000 bp upstream to 10,000 bp downstream from the CRISPRdb [26], which contains CRISPR arrays. CRISPR finder Program Online allowed us to acquire the numbers and sequences of repeats and spacers of CRISPRs [27]. The typical repeats of CRISPR were analyzed through multiple sequence alignment using Cluster X software, and the alignments of these repeats were visualized with WebLogo (<http://weblogo.berkeley.edu/logo.cgi>). The grouping of CRISPR sequences was performed based on the distance between the repeats of CRISPR loci for each group. Secondary structure prediction and minimum free energy (MFE) of the repeats in each group were determined by RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) [28]. We searched Cas genes from 10,000 bp upstream to 10,000 bp downstream the CRISPR loci in the NCBI database (<http://crispr.u-psud.fr/crispr/BLAST/CRISPRsBlast.php>).

## Data validation

CRISPRfinder allowed us to acquire the basic characteristics of CRISPR with the last update on 2017/1/2. The database contains 231 and 6600 analyzed genomes and 890 and 8732 CRISPRs for archaea and bacteria, respectively. The RNAfold web server performed secondary structure prediction and MFE of the CRISPR repeats with current limits of 7500 nt for section function calculations and 10,000 nt for MFE-only predictions. Publicly available complete sequences of plasmids, phages, and microbial genomes were obtained from the BLAST database. The CRISPRTarget databases provided GenBank-Phage, RefSeq-Plasmid, and RefSeq-Microbial and RefSeq-viral, and the cutoff score was the default parameter value [29].

## Author contributions

Jianzhong Han, Daofeng Qu, and Yang Shen designed experimental procedures. Daofeng Qu, Yang Shen, Lin Xu, Liying Zhang, Ziyang Xiang, Dongping Huang, Yuewen Chen, Anna Tong, and Jianzhong Han collected and analyzed the data. Jianzhong Han, Daofeng Qu, and Yang Shen wrote this manuscript.

## CONFLICTS OF INTEREST

The authors declare no competing financial interests.

## FUNDING

This work was supported by Zhejiang science and technology project (2017C32036) and Zhejiang Provincial Collaborative Innovation Center of Food Safety and Nutrition (2017SICR118).

## REFERENCES

1. Dreyer M, Aguilar-Bultet L, Rupp S, Guldemann C, Stephan R, Schock A, Otter A, Schupbach G, Brisse S, Lecuit M, Frey J, Oevermann A. *Listeria monocytogenes* sequence type 1 is predominant in ruminant rhombencephalitis. *Sci Rep.* 2016; 6:36419. <https://doi.org/10.1038/srep36419>.
2. Partridge SR. Analysis of antibiotic resistance regions in gram-negative bacteria. *FEMS Microbiol Rev.* 2011; 35:820–55. <https://doi.org/10.1111/j.1574-6976.2011.00277.x>.
3. Reeks J, Naismith JH, White MF. CRISPR interference: a structural perspective. *Biochem J.* 2013; 453:155–66. <https://doi.org/10.1042/BJ20130316>.
4. Touchon M, Rocha EP. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. *PLoS One.* 2010; 5:e11126. <https://doi.org/10.1371/journal.pone.0011126>.



5. Lin TL, Pan YJ, Hsieh PF, Hsu CR, Wu MC, Wang JT. Imipenem represses CRISPR-Cas interference of DNA acquisition through H-NS stimulation in *klebsiella pneumoniae*. *Sci Rep*. 2016; 6. <https://doi.org/10.1038/srep31644>.
6. Liu X, Homma A, Sayadi J, Yang S, Ohashi J, Takumi T. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Sci Rep*. 2016; 6:19675. <https://doi.org/10.1038/srep19675>.
7. Cady KC, White AS, Hammond JH, Abendroth MD, Karthikeyan RS, Lalitha P, Zegans ME, O'Toole GA. Prevalence, conservation and functional analysis of yersinia and escherichia CRISPR regions in clinical pseudomonas aeruginosa isolates. *Microbiology*. 2011; 157:430–7. <https://doi.org/10.1099/mic.0.045732-0>.
8. Lorenz R, Bernhart SH, Honer Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. *Algorithms Mol Biol*. 2011; 6:26. <https://doi.org/10.1186/1748-7188-6-26>.
9. Chen X, Lu X, Shu N, Wang S, Wang J, Wang D, Guo L, Ye W. Targeted mutagenesis in cotton (*gossypium hirsutum* L.) using the CRISPR/Cas9 system. *Sci Rep*. 2017; 7:44304. <https://doi.org/10.1038/srep44304>.
10. Guo X, Wang Y, Duan G, Xue Z, Wang L, Wang P, Qiu S, Xi Y, Yang H. Detection and analysis of CRISPRs of shigella. *Curr Microbiol*. 2015; 70:85–90. <https://doi.org/10.1007/s00284-014-0683-8>.
11. Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet*. 2011; 45:273–97. <https://doi.org/10.1146/annurev-genet-110410-132430>.
12. Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol*. 2007; 8:R61. <https://doi.org/10.1186/gb-2007-8-4-r61>.
13. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 2012; 482:331–8. <https://doi.org/10.1038/nature10886>.
14. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R. Diversity, activity, and evolution of CRISPR loci in *streptococcus thermophilus*. *J Bacteriol*. 2007; 190:1401–12. <https://doi.org/10.1128/jb.01415-07>.
15. Di H, Ye L, Yan H, Meng H, Yamasak S, Shi L. Comparative analysis of CRISPR loci in different *Listeria monocytogenes* lineages. *Biochem Biophys Res Commun*. 2014; 454:399–403. <https://doi.org/10.1016/j.bbrc.2014.10.018>.
16. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of forty-five CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*. 2005; 1:e60. <https://doi.org/10.1371/journal.pcbi.0010060.eor>.
17. Sesto N, Touchon M, Andrade JM, Kondo J, Rocha EP, Arraiano CM, Archambaud C, Westhof E, Romby P, Cossart P. A PNPase dependent CRISPR System in *Listeria*. *PLoS Genet*. 2014; 10:e1004065. <https://doi.org/10.1371/journal.pgen.1004065>.
18. Wang P, Zhang B, Duan G, Wang Y, Hong L, Wang L, Guo X, Xi Y, Yang H. Bioinformatics analyses of shigella CRISPR structure and spacer classification. *World J Microbiol Biotechnol*. 2016; 32:38. <https://doi.org/10.1007/s11274-015-2002-3>.
19. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*. 1981; 9:133–48.
20. Zeng H, Zhang J, Li C, Xie T, Ling N, Wu Q, Ye Y. Corrigendum: the driving force of prophages and CRISPR-Cas system in the evolution of *cronobacter sakazakii*. *Sci Rep*. 2017; 7:46783. <https://doi.org/10.1038/srep46783>.
21. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *clostridium*. *Sci Rep*. 2016; 6:25666. <https://doi.org/10.1038/srep25666>.
22. Diez-Villasenor C, Almendros C, Garcia-Martinez J, Mojica FJ. Diversity of CRISPR loci in *escherichia coli*. *Microbiology*. 2010; 156:1351–61. <https://doi.org/10.1099/mic.0.036046-0>.
23. Medina-Aparicio L, Rebollar-Flores JE, Beltran-Luviano AA, Vazquez A, Gutierrez-Rios RM, Olvera L, Calva E, Hernandez-Lucas I. CRISPR-Cas system presents multiple transcriptional units including antisense RNAs that are expressed in minimal medium and upregulated by pH in *salmonella enterica* serovar typhi. *Microbiology*. 2017; 163:253–65. <https://doi.org/10.1099/mic.0.000414>.
24. Clewell DB. Antibiotic resistance plasmids in bacteria. 2014. <https://doi.org/10.1002/9780470015902.a0001491.pub3>.
25. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*. 2011; 9:467–77. <https://doi.org/10.1038/nrmicro2577>.
26. Rahmatabadi SS, Nezafat N, Negahdaripour M, Hajighahramani N, Morowvat MH, Ghasemi Y. Studying the features of 57 confirmed CRISPR loci in 29 strains of *escherichia coli*. *J Basic Microbiol*. 2016; 56:645–53. <https://doi.org/10.1002/jobm.201500707>.
27. Ge R, Mai G, Wang P, Zhou M, Luo Y, Cai Y, Zhou F. CRISPRdigger: detecting CRISPRs with better direct repeat annotations. *Sci Rep*. 2016; 6:32942. <https://doi.org/10.1038/srep32942>.
28. Taylor WR, Hamilton RS. Exploring RNA conformational space under sparse distance restraints. *Sci Rep*. 2017; 7:44074. <https://doi.org/10.1038/srep44074>.
29. Biswas A, Gagnon JN, Brouns SJ, Fineran PC, Brown CM. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol*. 2013; 10:817–27. <https://doi.org/10.4161/ra.24046>.