Sequence analysis

CRISPR-RT: a web application for designing CRISPR-C2c2 crRNA with improved target specificity

Houxiang Zhu¹, Emily Richmond² and Chun Liang¹,²,*

¹Department of Biology and ²Department of Computer Science and Software Engineering, Miami University, Oxford, OH 45056, USA

*To whom correspondence should be addressed.

Associate Editor: Ivo Hofacker

Received on March 7, 2017; revised on August 18, 2017; editorial decision on September 12, 2017; accepted on September 13, 2017

Abstract

Summary: CRISPR-Cas systems have been successfully applied in genome editing. Recently, the CRISPR-C2c2 system has been reported as a tool for RNA editing. Here we describe CRISPR-RT (CRISPR RNA-Targeting), the first web application to help biologists design crRNAs with improved target specificity for the CRISPR-C2c2 system. CRISPR-RT allows users to set up a wide range of parameters, making it highly flexible for current and future research in CRISPR-based RNA editing. CRISPR-RT covers major model organisms and can be easily extended to cover other species. CRISPR-RT will empower researchers in RNA editing.

Availability and implementation: Freely available at http://bioinfolab.miamioh.edu/CRISPR-RT.

Contact: liangc@miamioh.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

During the past several years, CRISPR-Cas systems have been successfully applied in genome editing, but no system has been reported for RNA editing. Therefore, new CRISPR-Cas systems that regulate RNA activities are necessary for studying the roles of RNA molecules. Recently, the CRISPR-C2c2 system has been demonstrated as a tool for RNA targeting (Abudayyeh et al., 2016). CRISPR-C2c2 was discovered in 21 bacterial genomes and belongs to the Type VI of Class 2 CRISPR systems (Shmakov et al., 2015). Researchers have characterized the CRISPR-C2c2 system from the bacteria Leptotrichia shahii (Abudayyeh et al., 2016). The L. shahii C2c2 locus is simply organized, including C2c2, Cas1, Cas2 and a CRISPR array (Supplementary Fig. S1). C2c2, which contains two higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains, mainly functions as a sole effector protein mediating single-strand RNA cleavage (Abudayyeh et al., 2016). Similar to other Class 2 systems, the CRISPR array of CRISPR-C2c2 is first transcribed into pre-crRNA (Shmakov et al., 2015). Differently, the pre-crRNA is processed by C2c2 into mature crRNAs without attaching to trans-activating crRNAs (East-Seletsky et al., 2016). The mature crRNA binds to C2c2 and guides it to target a specific single-strand RNA. C2c2 combined with a 22–28 length of the target complementarity region of crRNA would effectively mediate cleavage, and the secondary structure of the crRNA is also required for RNA cleavage (Abudayyeh et al., 2016). The seed region is located in the center of the crRNA-target duplex, where it is more sensitive to mismatches than the non-seed region (Abudayyeh et al., 2016). Single mismatch can be fully tolerated by C2c2, but if double mismatches are located in the seed region, C2c2 is unable to cleave the single-strand RNA; C2c2 can even tolerate three consecutive mismatches in the non-seed region (Abudayyeh et al., 2016). Whether C2c2 can tolerate gaps remains unknown, which might be explored in near future. The CRISPR-C2c2 system prefers H (A, U or C) for the 30 protospacer flanking site (PFS) sequence of one single base length to mediate single-strand RNA cleavage (Abudayyeh et al., 2016). CRISPR-C2c2 has already been successfully used for specific RNA knockdown in Escherichia coli (Abudayyeh et al., 2016). Researchers found that C2c2 cleaves the targeted single-strand RNA in addition to collateral RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016), which has been applied for RNA
detection in human total RNAs (East-Seletsky et al., 2016) and viral strains detection (Gootenberg et al., 2017). The inactive dC2c2, just like dCas9 (Gao et al., 2016; Gilbert et al., 2013; Zetsche et al., 2015), also has many potential applications as an RNA-binding protein, such as bringing effectors to specific RNAs to regulate their translation and tracking specific RNAs by fluorescent tag (Abudayyeh et al., 2016). Therefore, the CRISPR-C2c2 system has been viewed as a powerfully programmable tool for RNA editing (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Nainar et al., 2016; Puchta, 2017; Wang and Qi, 2016). However, until now, there is no available public software for designing crRNAs of the CRISPR-C2c2 system.

We have developed CRISPR-RT (CRISPR RNA-Targeting), a web application to help biologists design the crRNA for the CRISPR-C2c2 system. To maximize the flexibility for current and future research in CRISPR-based RNA editing, CRISPR-RT allows users to set up a wide range of parameters, such as length of the target complementarity region of crRNA, length of the seed region, the PFS, and the number of mismatches or gaps tolerated by off targets. After setting up the required parameters, CRISPR-RT will find target candidates from the input RNA sequence and employ a rigorous alignment algorithm to search on- and off-target sites for each target candidate within the reference transcriptome. The results are displayed in highly interactive graphical interfaces. Users can rank target candidates by the total number of target sites in the reference transcriptome, which help them choose the target candidate based on the minimum effect of off targets. In addition, users are able to validate the on- and off-target sites in the background of annotated genome and transcript features by data visualization through JBrowse (Skinner et al., 2009).

2 Implementation

2.1 Graphic input interface

Supplementary Figure S2 shows the CRISPR-C2c2 setting page. First, users can input an RNA/cDNA sequence that they want to target in FASTA format. They can also use an example sequence by clicking the ‘Example Sequence’ button. Second, users select a reference transcriptome. They can also click the ‘custom’ link to upload a custom reference transcriptome in FASTA format. Third, in terms of our current understanding of the CRISPR-C2c2 system architecture (Supplementary Fig. S1), users can set up the PFS sequence and crRNA requirements for the CRISPR-C2c2 system properly. The on- and off-target PFS sequence can be set, respectively. Users then set up the length of the target complementarity region of crRNA and the length of the seed region. The seed region is located in the center of the crRNA-target duplex, and its length should not be greater than the length of the target complementarity region of crRNA. Fourth, users can choose an off-target setting (‘Basic settings’ or ‘Specific settings’). For ‘Basic settings’, the number of mismatches or gaps tolerated by off targets and by the seed region can be set, respectively. The number of consecutive mismatches or gaps in the seed or non-seed region tolerated by off targets can also be configured. For ‘Specific settings’, users can set up more detailed parameters in the seed and non-seed regions separately, such as the number of mismatches in the seed or non-seed region tolerated by off targets. Users can also set up the search sensitivity of Bowtie2 (Langmead and Salzberg, 2012), which is related to the alignment options setting of Bowtie2. Higher sensitivity setting causes alignments to be more sensitive, but it usually results in a longer search time. After setting up all the parameters, users click the ‘Find targets!’ button, which will run programs in the background to get the results. If users want to retrieve a recent job, they can click ‘Retrieve Jobs’ on the left-side menu to enter the ‘Job ID’, which is generated in the result page. The results will be kept in our server for only one week and will be deleted automatically afterwards.

2.2 Graphic output interface

Supplementary Figure S3A shows all the target candidates and relevant information for one RNA query. Users can view the input sequence by clicking the ‘Input Sequence Viewer’ button. In the sequence viewer, users can search and highlight any subsequence such as the target candidate sequence (Supplementary Fig. S3B). Users can also download the target candidates’ file by clicking the ‘Download’ link. In the table, the protospacer and PFS of each target candidate are labeled by different colors. The corresponding crRNA of each target candidate can be accessed by clicking ‘crRNA’; a graph will appear to help users design their own crRNAs (Supplementary Fig. S4). The table also displays the start position, end position and GC content of each target candidate sequence. The last two columns of the table show the numbers of targeted transcripts and genes, respectively, for each target candidate. By clicking table headers, users can rank all target candidates based on the number of target sites including on- and off-target sites. Target candidates with fewer number of target sites have higher target specificity. If the number of target sites is 1, the corresponding table cell will be highlighted with green background color to indicate that the target candidate is highly specific. By clicking the number of targeted transcripts, users can view the detailed information of targeted transcripts for each target candidate (Supplementary Fig. S3C). Because CRISPR-RT has converted the transcriptome mapping result to a genome mapping result, the information of targeted transcripts is displayed in genomic context with genomic coordinates and gene annotations, including the transcript isoform, mapped gene, chromosome, start position and strand. Users can click the ‘Transcript ID’ link or ‘Gene ID’ link of each target site to view the detailed description of transcript or gene where the target site located. The number of mismatches or gaps in each target site is also shown in the table. To visualize and manually validate the targeted transcripts, users can click the ‘JBrowse’ link to visualize each target site in the background of genome and transcript features annotated by Ensembl or Phytozome (Supplementary Fig. S3D). The implementation methods of CRISPR-RT are provided in the Supplementary Methods.

Funding

This work was supported by Office for the Advancement of Research and Scholarship (OARs); and Biology Department, Miami University, Oxford, Ohio, USA.

Conflict of Interest: none declared.

References


