CRISPy-web
An online resource to design sgRNAs for CRISPR applications

Blin, Kai; Pedersen, Lasse Ebdrup; Weber, Tilmann; Lee, Sang Yup

Published in:
Synthetic and Systems Biotechnology

Link to article, DOI:
10.1016/j.synbio.2016.01.003

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
CRISPy-web: An online resource to design sgRNAs for CRISPR applications

Kai Blin a, Lasse Ebdrup Pedersen a, Tilmann Weber a,*, Sang Yup Lee a, b, **

a The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Alle 6, 2970 Hørsholm, Denmark
b Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea

A R T I C L E   I N F O

Article history:
Received 17 October 2015
Received in revised form 5 January 2016
Accepted 10 January 2016
Available online

Keywords:
CRISPR
Cas9
sgRNA
Guide RNA
Genome editing
Genome engineering

A B S T R A C T

CRISPR/Cas9-based genome editing has been one of the major achievements of molecular biology, allowing the targeted engineering of a wide range of genomes. The system originally evolved in prokaryotes as an adaptive immune system against bacteriophage infections. It now sees widespread application in genome engineering workflows, especially using the Streptococcus pyogenes endonuclease Cas9. To utilize Cas9, so-called single guide RNAs (sgRNAs) need to be designed for each target gene. While there are many tools available to design sgRNAs for the popular model organisms, only few tools that allow designing sgRNAs for non-model organisms exist. Here, we present CRISPy-web (http://crispy.secondarymetabolites.org/), an easy to use web tool based on CRISPy to design sgRNAs for any user-provided microbial genome. CRISPy-web allows researchers to interactively select a region of their genome of interest to scan for possible sgRNAs. After checks for potential off-target matches, the resulting sgRNA sequences are displayed graphically and can be exported to text files. All steps and information are accessible from a web browser without the requirement to install and use command line scripts.

© 2016 The authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

“Clustered regularly interspaced short palindromic repeats” (CRISPR) and their associated RNA-guided endonucleases are bacterial adaptive immune systems protecting the bacteria from infections with bacteriophages.1 The biotechnological application of this system is currently revolutionizing molecular biology and provides new opportunities for synthetic biology applications.2 The CRISPR system allows targeted genome engineering of bacteria,3–5 and also eukaryotes including yeast,6–7 plants,8 human cell lines9 and many more. Although several alternative CRISPR systems have recently been described (e.g., References 10–13), most CRISPR systems for genome engineering are based on the Streptococcus pyogenes type II CRISPR/Cas9 system. Cas9 is an RNA-guided endonuclease, which introduces double strand DNA breaks at positions that are complementary to a crRNA sequence that binds to Cas9 in a duplex with a tracrRNA. It has also been demonstrated that Cas9 accepts artificially fused crRNA:tracrRNA-hybrids termed “single guide RNAs” (sgRNAs) to direct it to the target cut sites.14 Thus CRISPR can be used as an “in vivo programmable restriction enzyme,” which cuts the target DNA within an exactly defined target sequence determined by the sgRNA.

One essential step in the design of CRISPR experiments is to identify suitable sgRNA sequences within the target gene(s), which have to fulfill certain conditions: (i) the 20 bp target sequence has to be directly upstream of a “protopspacer adjacent motif” (PAM), for the S. pyogenes PAM, this motif is “NGG”, (ii) it is desirable that the sequence of the target motif is unique within the genome of the organisms to prevent off-target activity, i.e. the cleavage of the chromosome at wrong positions.

Many programs and web-servers exist to assist biologists in finding such CRISPR/Cas9 target sites. However, most of these tools (e.g. CCTop,15 CHOPCHOP16 CRISPR Design,17 WU-CRISPR,18 WGE CRISPR Finder,19 and CRISPY CHO20) are limited to a narrow set of – mostly eukaryotic – target genomes of model organisms. Only very few tools (e.g. sgRNAcas921) allow running Cas9 target searches on

http://dx.doi.org/10.1016/j.synbio.2016.01.003
2405-805X/© 2016 The authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
A. Overview sgRNAs in *Streptomyces coelicolor* A3(2) actinorhodin biosynthetic gene cluster

Fig. 1. Example output of a CRISPy-web run to identify sgRNAs in the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor* A3(2) (NCBI GenBank ID: NC_003888.3).

(A) Overview of sgRNAs identified in the complete actinorhodin gene cluster. (B) Zoom view of SCO5087 (actIORF1, KS-a). (C) List of sgRNAs selected for export.
user-specified genomes. Unfortunately, to the best of our knowledge, none of these generally applicable tools are provided as user-friendly web services, which can be used by non-computer scientists.

Here, we therefore present a web server implementation of the sgRNA prediction software CRISPy (http://crispy.secondarymetabolites.org/) that was originally designed for CRISPR-based engineering of Chinese hamster ovary (CHO) cells. It is implemented as a standalone web application for Cas9 target prediction that can design sgRNA for any gene/region of interest in user-submitted genome data. In addition, CRISPy-web tightly connects with the popular natural product prediction tool antiSMASH to assist researchers working on microbial secondary metabolites in leveraging the CRISPR/Cas9 system. Using antiSMASH or working on natural products of course is not a prerequisite for using CRISPy-web.

2. Design and implementation

CRISPy-web consists of multiple decoupled service components tied together by a message queue. The main services are the CRISPy worker service, the web application programming interface (API) server and the web user interface (UI). Sequences to be scanned for Cas9 target sites can be uploaded directly (in GenBank format) or transferred from antiSMASH by specifying the antiSMASH job ID. In the next version of antiSMASH, a direct link to assist researchers working on microbial secondary metabolites gene cluster locations using Biopython. Two steps are performed by the worker service: preparing a sequence for analysis and then scanning and ranking Cas9 target sites. In the preparation step, the input files are parsed and – if the files were transferred from antiSMASH – annotated with secondary metabolite gene cluster locations using Biopython.

Based on the CRISPy algorithm initially used in Ronda et al., the scanning step searches for the ‘NGG’ PAM sequence on both the forward and the reverse strands of the selected region of interest, counting the number of direct off-target hits for sequences with the same 13bp core sequence upstream of the PAM. It then calculates the number of off-target hits containing 0, 1, 2, or 3 bp mismatches on the whole genome for the 13 bp sequence by considering all possible combinations.

After collecting all possible Cas9 targets, sites that have direct off-target matches, sites with off-target matches with a single base-pair mutation, or sites with more than ten off-target matches with a two base-pair mutation are discarded. All these hits are stored in a GenBank file. To further reduce the number of sites displayed on the user interface, the targets are now sorted by 2bp mismatch off-targets, 3 bp mismatch off-targets and coordinates on the genome, and only the first 1000 targets are displayed to the user.

2.2. Web application programing interface server

The web API server provides a Representational State Transfer (REST) interface that both the CRISPy-web UI and other tools can use to submit jobs to and obtain results from the CRISPy worker service. Data are transferred using the JavaScript Object Notation (JSON) format. The API server is implemented in Python using the Flask web framework and a Redis server for the message queue.

2.3. Web user interface

Built around the AngularJS framework and D3 for cluster visualization, the web UI is the primary method of interacting with CRISPy-web. Communicating with the web API server via asynchronous JavaScript (AJAX) calls, the web UI leads the user through the process of identifying Cas9 target sites in regions of interest.

3. Example usage: scanning for sgRNAs in the actinorhodin gene cluster of Streptomyces coelicolor

To scan for sgRNAs in S. coelicolor, the user first needs to upload the corresponding genome file in GenBank format. Such a file can easily be obtained from NCBI with the RefSeq ID NC_003888. After a short preliminary scan of the genome, CRISPy-web displays the target selection page. The target selection page gives a short overview of the uploaded genome, such as organism, description, sequence length and the number of annotated genes. Targets can be selected as a nucleotide range or by providing a locus tag, gene name, or protein-ID if annotated in the genome. If available, annotated gene clusters will also be shown in a list and can be selected as regions to scan. On S. coelicolor, the actinorhodin cluster is located from 5,509,800 to 5,552,424 bp on the genome, so that is the range to select as target region. After a click on “Find targets”, the scan for sgRNAs starts.

Depending on the size of the selected region and the overall genome size, this sgRNA scan will take several minutes. After the scan is completed, a page graphically displays the scanned region. Genes are shown as gray arrows, sgRNAs on the forward strand are displayed as red boxes above the genes, sgRNAs on the reverse strand are displayed below the genes. A table below the graphical overview gives details on the identified sgRNAs, ranked by the least number of potential off-target matches. Hovering over the table row of an sgRNA hit highlights the corresponding sgRNA box in the graphics, so you can quickly locate the hit on the region.

As there are a lot of hits for this region, it is possible to zoom in on a particular gene of interest by clicking on the gene arrow and selecting “show results for this gene only”. Because the whole region already was scanned, this operation is instantaneous. On both the overview and on the zoomed view, sgRNAs you want to save for download can be added to the download basket by clicking the corresponding table row. The download basket button on the upper right will track the number of selected sgRNAs.

A click on the download basket icon will display the download overview page. There, the selected sgRNAs can be verified again. A click on the “Download CSV file” button will start the download of all selected sgRNAs in comma-separated value format. Files in this format can easily be used in a spreadsheet application, processed for downstream computational analysis or oligonucleotide orders.

Acknowledgment

This work was funded by a grant of the Novo Nordisk Foundation.

References


