



PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing

Achal Rastogi, Omer Murik, Chris Bowler, Leila Tirichine

► To cite this version:

Achal Rastogi, Omer Murik, Chris Bowler, Leila Tirichine. PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. BMC Bioinformatics, 2016, 17 (1), pp.261. 10.1186/s12859-016-1143-1 . inserm-02059918

HAL Id: inserm-02059918

<https://inserm.hal.science/inserm-02059918>

Submitted on 7 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SOFTWARE

Open Access



PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing

Achal Rastogi, Omer Murik, Chris Bowler and Leila Tirichine*

Abstract

Background: With the emerging interest in phytoplankton research, the need to establish genetic tools for the functional characterization of genes is indispensable. The CRISPR/Cas9 system is now well recognized as an efficient and accurate reverse genetic tool for genome editing. Several computational tools have been published allowing researchers to find candidate target sequences for the engineering of the CRISPR vectors, while searching possible off-targets for the predicted candidates. These tools provide built-in genome databases of common model organisms that are used for CRISPR target prediction. Although their predictions are highly sensitive, the applicability to non-model genomes, most notably protists, makes their design inadequate. This motivated us to design a new CRISPR target finding tool, PhytoCRISP-Ex. Our software offers CRISPR target predictions using an extended list of phytoplankton genomes and also delivers a user-friendly standalone application that can be used for any genome.

Results: The software attempts to integrate, for the first time, most available phytoplankton genomes information and provide a web-based platform for Cas9 target prediction within them with high sensitivity. By offering a standalone version, PhytoCRISP-Ex maintains an independence to be used with any organism and widens its applicability in high throughput pipelines. PhytoCRISP-Ex out pars all the existing tools by computing the availability of restriction sites over the most probable Cas9 cleavage sites, which can be ideal for mutant screens.

Conclusions: PhytoCRISP-Ex is a simple, fast and accurate web interface with 13 pre-indexed and presently updating phytoplankton genomes. The software was also designed as a UNIX-based standalone application that allows the user to search for target sequences in the genomes of a variety of other species.

Keywords: CRISPR, Cas9, Protists, Genome editing, Eukaryotes

Background

Phytoplankton are microalgae that form an essential constituent of the marine food chain. Though microscopic and mostly uncharacterized, these minute organisms have tremendously showcased themselves as potential research models [22, 20]. Recent large-scale sampling to understand the morphological and genetic diversity of this hidden community [3, 12, 21] has already established the foundation for further molecular studies. The successful achievement of this exploration also reflects the

interest of research communities towards the functional characterization of phytoplankton in the near future.

Clustered regularly interspaced short palindromic repeats CRISPR/CAS systems have recently emerged as a simple and accurate tool for genome editing [13], and show facile editing in numerous organisms including bacteria [10], yeast [4], plants [6], human [14] and other animals [2, 5, 7, 9, 17, 23, 24]. Common designed CRISPR systems consist of expression of a Cas9 nuclease or nickase and a single guide RNA (sgRNA). The latter includes a 20-bp target sequence used to target the Cas9/RNA complex to the desired chromosomal location. By modifying only these 20-bp, the targeting of the whole CRISPR/Cas9 complex will change and thus the DNA cleavage site. A well designed target sequence must not only optimally bind to its desired target, but

* Correspondence: leila.tirichine@ens.fr

Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, PSL Research University, CNRS UMR 8197, INSERM U1024, 46 rue d'Ulm, F-75005 Paris, France

must also not target any other site in the genome being edited, to avoid undesired off-target mutations [11]. Several parameters were shown to contribute to a better targeting of the target sequence, and the general form of G-N19-NGG is widely accepted, although others were also suggested [1].

Because target site choice is a key point for promoting successful editing, several computational tools have been designed aiming to automate this procedure, all offering candidate target sites for a given input sequence/gene and potential off-target sites for a given background genome. This search is usually restricted for genomes of selected model organisms such as human, mice, fruit-fly, *C. elegans*, and yeast, making them irrelevant for researchers aiming to use CRISPR on other, less common, organisms. We designed PhytoCRISP-Ex, a user friendly web interface, as well as a stand-alone software, which predicts potential target sites for CRISPR/CAS projects. The off-target analysis can be performed against indexed genomes from 13 algae (diatoms, green algae, haptophytes, etc.), or any user defined genome or transcriptome, assembled or not, making it useful for designing CRISPR projects for many communities.

Algorithm

PhytoCRISP-Ex is a rational and flexible tool for finding Cas9 target sites with low/no off-target potential. Given a DNA query (e.g., gene) sequence, the pipeline first fetches all the possible regions of length 23, structured as [5'-G/N(19 bps)PAM-3']/[3'-PAM(19 bps)C/N-5']. These regions are then evaluated and filtered to show low/no off-target activity across the whole genome (see Fig. 1a for general workflow). The choice of PAM sequence and CRISPR target start base is kept flexible. PAM sequence can be selected from NGG or NAG and CRISPR start base can be chosen as G or any (N) base. These options are also implemented in the standalone version of the software and can be used as arguments.

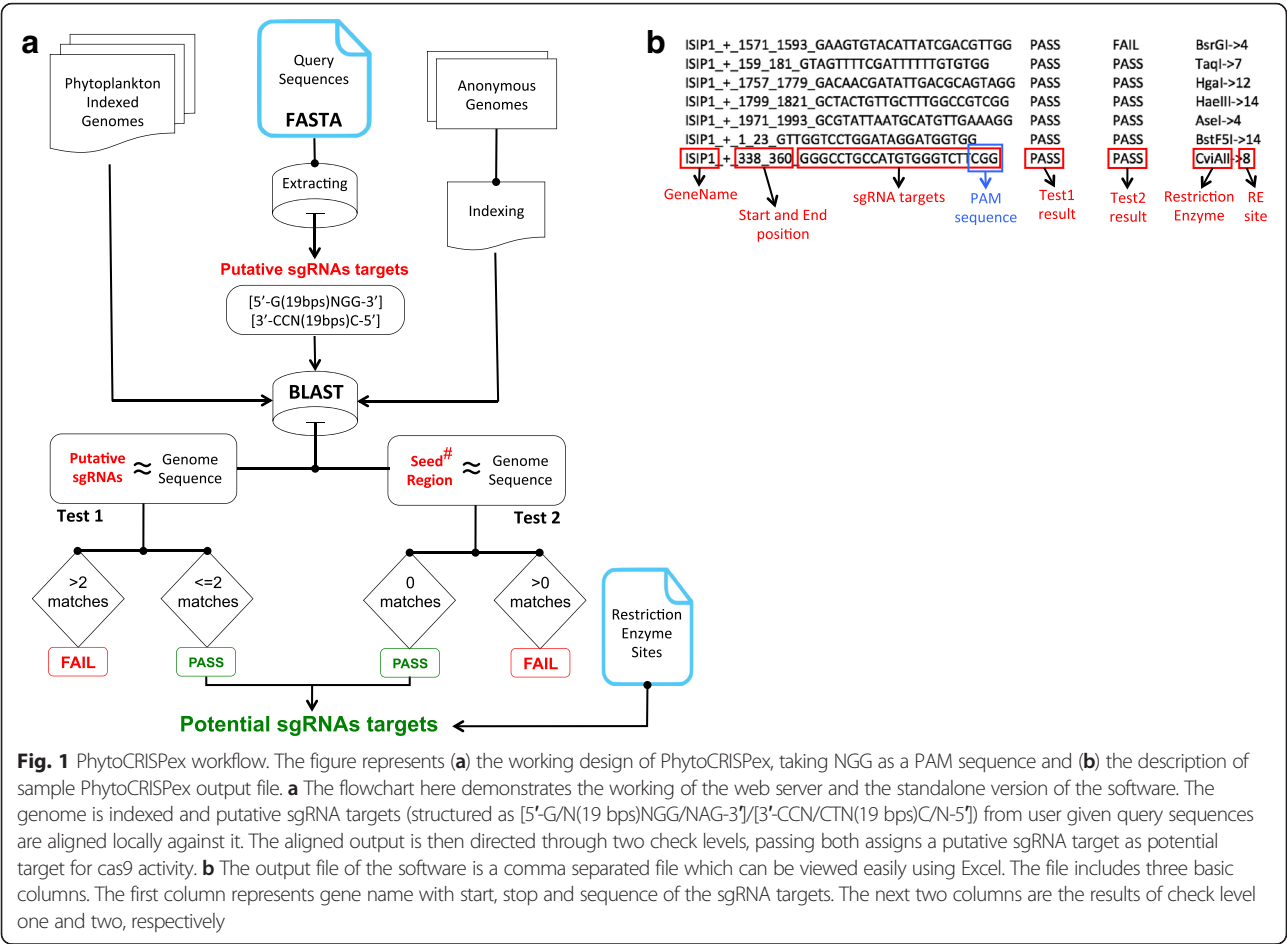
There are two levels of filters and passing both designates the region as a potential sgRNA for Cas9 activity. The first level accepts the target sequence and tags it "PASS" if it has less than or equal to 2 base-pair mismatch with the closest off-target in the genome. The second filter, accepts the target if the seed region (last 15 bases including the PAM sequence) has 0 mismatch with off-targets anywhere in the genome. The pipeline reports the instances which are accepted by both or either one of the filters, but the ones which passes both the filters are designated as potential Cas9 targets.

Once the potential sgRNAs are filtered, they are then checked for the presence of none, one or more restriction sites corresponding to pre-selected and most common restriction enzymes (Additional file 1: Figure S1). Cas9 enzyme cleaves both DNA strands ~3–4 bases upstream of

the PAM [19, 25]. Therefore along with the presence of restriction sites on the entire target sequence, PhytoCRISP-Ex reports specifically, if any, the restriction sites overlapping three and four bases upstream to the PAM sequence. Embedding a restriction site in the target sequence will help screening for mutants using PCR after a digest with the appropriate restriction enzyme. Presence and absence of the restriction sites along with its position on the sgRNA are reported in the output file (Fig. 1b). The list of restriction enzymes (Additional file 2: Table S1) can be updated and used when using the standalone version of the software. The PhytoCRISP-Ex pipeline is mounted with the index of 13 genomes and is accessible via web-server. An off-line standalone software is also developed which gives liberty to its users to use it for any genome, assembled or un-assembled. The standalone package also includes a few example files and a README file to help users install and execute PhytoCRISP-Ex. The standalone version can be found under "Download" tab at <http://www.phytocrispex.biologie.ens.fr/CRISP-Ex/>.

Results

Several computational tools have been published aiming to identify candidates for CRISPR/CAS targeting. In order to validate the novelty and sustainability of PhytoCRISP-Ex, we compared it with 8 other previously published tools. Our criteria for comparison included first, the possibility to use the tool as a browser-based or stand-alone application, so to meet the need of a wide range of users. Second, whether it provides the flexibility to perform off-target analysis on any or a restricted list of genomes. The third criteria compares the possibility to perform restriction site mutant screening analysis. The summary of this comparison is presented in Table 1 and shows PhytoCRISP-Ex as a robust and adequately designed application. Further, we evaluated our tool against the whole exome of the model diatom *Phaeodactylum tricornutum* (http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index/) and *Thalassiosira pseudonana* (<http://genome.jgi.doe.gov/Thaps3/Thaps3.home.html>). The statistics revealed that ~94 % and ~95 % of the genes in *P. tricornutum* and *T. pseudonana*, respectively, have at least 1 potential guide RNA to be used as a Cas9 target against these genes. Among the latter, most of the genes have high percent efficiency in terms of constituting mostly potential targets among total predicted Cas9 targets (Additional file 1: Figure S1). These findings suggest that potentially almost any gene in these two species can be targeted with high probability for generating a single specific mutation. Out of the total, ~89 % genes in both the species (*P. tricornutum* and *T. pseudonana*) possess potential targets with restriction sites over the probable Cas9 cleavage site. Therefore, choosing such targets might help in fast screening of the mutants. PhytoCRISP-Ex is a



CRISPR/Cas9 target extraction web-package, built to extract targets using 13 model phytoplankton genomes. The algorithm has also been designed as a UNIX based standalone package which provides flexibility to its users to use it on other non-model genomes. The simple design of PhytoCRISP-Ex allows its use by end-users with moderate or no software programming background.

Conclusions

With the persuasive interest of scientific community towards phytoplankton research, the need of establishing genetic transformation tools for plankton species is thriving. PhytoCRISP-Ex provides a reliable and first ever application for predicting CRISPR/Cas9 targets within various plankton genomes. PhytoCRISP-Ex is also equipped with an easy to

Table 1 PhytoCRISP-Ex vs others

Softwares	Browser-based application	Stand-alone application	Restriction screening	Background Genome flexibility	Reference
PhytoCRISP-Ex	✓	✓	✓	✓	Current study
CRISPR MultiTargeter	✓	X	X	X	[18]
CasFinder	X	✓	X	X	[1]
CHOPCHOP	✓	X	X	X	[15]
CRISPRdirect	✓	X	X	X	[16]
E-CRISP	✓	X	X	X	[8]
sgRNAcas9	X	✓	X	✓	[27]
CasOT	X	✓	X	✓	[26]
CRISPRseek	X	✓	✓	✓	[28]

Comparison between PhytoCRISP-Ex and several previously published CRISPR target analysis tools

use, yet powerful, standalone version which gives its user the flexibility to use it on any genome. Many such and other unique features make the software more advance and appropriate for the use by a broad research community.

Additional files

Additional file 1: Figure S1. PhytoCRISP-Ex efficacy. The scatter plot depicts that most of the predicted Cas9 targets per gene in *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, respectively, are potential candidates (passing both PhytoCRISP-Ex filter). X-axis represents the percent efficiency of each gene in terms of having high number of potential Cas9 targets compared to the total number of targets. Y-axis represents the number of all potential targets per gene. (PDF 182 kb)

Additional file 2: Table S1. The list of restriction enzymes being used by PhytoCRISP-Ex. (XLSX 13 kb)

Abbreviations

CAS, CRISPR associated protein; CRISPR, Clustered regularly-interspaced short palindromic repeats; DNA, Deoxyribo nucleic acid; PAM, Protospacer adjacent motif; PCR, Polymerase chain reaction; RNA, Ribo nucleic acid.

Acknowledgements

We thank Pierre Vincens for his help in mounting PhytoCRISP-Ex over the web-server. AR was supported by the Labex Memolife International doctoral program.

Availability of data and materials

Project name: PhytoCRISP-Ex

Project home page: <http://www.phytocrispex.biologie.ens.fr/CRISP-Ex/>

Compatible browsers: Mozilla, Internet Explorer, Chrome, etc.

Operating system(s): UNIX

Programming language: Shell, Perl, HTML

License: NA

Any restriction to use by non-academics: None

Authors' contributions

AR, OM and LT conceived and designed the structure of the software. AR wrote the software. AR, OM CB and LT wrote the manuscript. LT coordinated the study. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Received: 15 January 2016 Accepted: 22 June 2016

Published online: 01 July 2016

References

- Aach JEA. CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes. *bioRxiv*. (2014);005074. doi:<http://dx.doi.org/10.1101/005074>.
- Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods*. 2015;12:823–6.
- de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, Lara E, Berner C, Le Bescot N, Probert I, et al. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science*. 2015;348:1261605.
- DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res*. 2013;41:4336–43.
- Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 2016;34:184–91.
- Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, Cao F, Zhu S, Zhang F, Mao Y, et al. Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res*. 2013;23:1229–32.
- Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods*. 2013;10:741–3.
- Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. *Nat Methods*. 2014;11:122–3.
- Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT, and Yeh JR. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One*. 2013;8, e68708.
- Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31:233–9.
- Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol*. 2014;32:677–83.
- Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, et al. Ocean plankton. Determinants of community structure in the global plankton interactome. *Science*. 2015; 348:1262073.
- Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods*. 2013;10:957–63.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, and Church GM. RNA-guided human genome engineering via Cas9. *Science*. 2013;339:823–6.
- Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res*. 2014;42:W401–7.
- Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*. 2015;31: 1120–3.
- Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L, Kang Y, Zhao X, Si W, Li W, et al. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*. 2014;156:836–43.
- Prykhodzhiy SV, Rajan V, Gaston D, Berman JN. CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. *PLoS One*. 2015;10, e0119372.
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8:2281–308.
- Rastogi A, Lin X, Lombard B, Loew D, Tirichine L. Probing the evolutionary history of epigenetic mechanisms: What can we learn from marine diatoms. *AIMS Genet*. 2015;2:173–91.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al. Ocean plankton. Structure and function of the global ocean microbiome. *Science*. 2015;348: 1261359.
- Tirichine L, Bowler C. Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *Plant J*. 2011;66:45–57.
- Tzur YB, Friedland AE, Nadarajan S, Church GM, Calarco JA, Colaiacovo MP. Heritable custom genomic modifications in *Caenorhabditis elegans* via a CRISPR-Cas9 system. *Genetics*. 2013;195:1181–5.
- Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, and Jaenisch R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*. 2013;153:910–8.
- Wu X, Kriz AJ, Sharp PA. Target specificity of the CRISPR-Cas9 system. *Quant Biol*. 2014;2:59–70.
- Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, Zhang B. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*. 2014.
- Xie S, Shen B, Zhang C, Huang X, Zhang Y. sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS One*. 2014;9:e100448.
- Zhu LJ, Holmes BR, Aronin N, Brodsky MH. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One*. 2014;9, e108424.