GM Crops & Food, 6:266–276, 2015
Published with license by Taylor & Francis Group, LLC
ISSN: 2164-5698 print / 2164-5701 online
DOI: 10.1080/21645698.2015.1137690

A quick guide to CRISPR sgRNA design tools

Vincent A Brazelton, Jr,[1,2,†] Scott Zarecor,[3,†] David A Wright,[3] Yuan Wang,[4,5] Jie Liu,[4,5] Keting Chen,[4,5] Bing Yang,[3] and Carolyn J Lawrence-Dill[1,2,3,4,*]

[1]Interdepartmental Genetics and Genomics Program; Iowa State University; Ames, IA USA
[2]Department of Agronomy; Iowa State University; Ames, IA USA
[3]Department of Genetics; Development and Cell Biology; Iowa State University; Ames, IA USA
[4]Interdepartmental Bioinformatics and Computational Biology Program; Iowa State University; Ames, IA USA;
[5]Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology; Iowa State University; Ames, IA USA

ABSTRACT. Targeted genome editing is now possible in nearly any organism and is widely acknowledged as a biotech game-changer. Among available gene editing techniques, the CRISPR-Cas9 system is the current favorite because it has been shown to work in many species, does not necessarily result in the addition of foreign DNA at the target site, and follows a set of simple design rules for target selection. Use of the CRISPR-Cas9 system is facilitated by the availability of an array of CRISPR design tools that vary in design specifications and parameter choices, available genomes, graphical visualization, and downstream analysis functionality. To help researchers choose a tool that best suits their specific research needs, we review the functionality of various CRISPR design tools including our own, the CRISPR Genome Analysis Tool (CGAT; http://cropbioengineering.iastate.edu/cgat).

INTRODUCTION

Early in the 20th century Muller showed that X-rays cause genetic mutations in Drosophila (Muller, 1927). Likewise, Stadler showed the mutational effects of X-rays on barley and maize (Stadler, 1928; Stadler 1944) which paved the way for researchers to broadly use mutagens such as X-rays and chemical agents to induce random genetic changes. However,

*Correspondence to: Carolyn J Lawrence-Dill; Email: triffid@iastate.edu
Received August 4, 2015; Revised December 18, 2015; Accepted December 23, 2015.
[†]Contributed equally to the development of this project and manuscript.

those methods yielded many mutations that had to be sorted out over generations to isolate the one responsible for causing changes to specific phenotypes/traits of interest. More recently, basic research to understand the processes underlying natural chromosomal recombination, microbial immune and virulence responses, and DNA binding domains has led to discoveries that made possible the development of *targeted* genome editing techniques that pair sequence-specific DNA binding proteins with enzymes that cleave DNA (reviewed in Wright et al., 2014). Development of these methods led to the realization that a RNA directed bacterial immune system could also be developed into an effective genome editing tool. Now three major systems for genome editing exist: Zinc Finger Nucleases (ZFNs), TAL Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)/CRISPR associated proteins 9 (CRISPR/Cas9; reviewed in Peng et al., 2014).

Zinc finger proteins are classified into distinct families based on specific structural motifs. Shared among all are DNA binding domains along with one or more zinc ion(s) that serve to stabilize the fold (Klug, 2010). Early NMR spectroscopy experiments revealed that the Cys2His2 zinc finger binding domain in the *Xenopus* transcription factor IIA is comprised of a 30 amino acid repeat sequence with conserved $\beta\beta\alpha$ secondary structure (Ruiz i Altaba et al., 1987). This architecture allows amino acids on the surface of the α-helix to interact with specific major groove nucleotides, thus conferring specificity for particular double-stranded DNA sequences (Beerli and Barbas, 2002; Gaj et al., 2013; Lee et al., 1989). It was later found that by changing amino acids in the α-helix, DNA binding specificity and affinity could be altered. Engineered zinc fingers were combined with the DNA cleavage domain of FokI, a type II restriction endonuclease, to form ZFNs, which allow for specific targeted double-strand breaks in DNA.. Induction of DNA damage triggers the cellular repair pathway via error-prone non-homologous end joining or template mediated homology directed repair thus giving limited control over the repair process in a targeted manner (Lieber, 2010). Non-homologous end joining can create loss-of-function mutations due to insertions, deletions, or rearrangements whereas homology directed repair can create a precise mutation in the presence of a specific DNA template (Bogdanove, 2014; Lieber, 2010)

Transcription activator-like effector (TALE; also called TAL effector) proteins are major components of the type III secretion system conferring pathogenicity in the Gram negative bacteria *Xanthomonas* (White et al., 2009; Boch and Bonas 2010). Of the more than 30 families of bacterial effector proteins, TALEs are unique in their ability to distinguish specific DNA sequences via a central repetitive 34 amino acid DNA binding motif (Boch et al., 2009; Moscou and Bogdanove, 2009). The repeat variable di-amino acids (RVDs) at positions 12 and 13 determine overall specificity and affinity for specific nucleotides in a target sequence. When coupled with the nuclease domain of FokI, TALE nucleases (TALENs) emerged as a novel genome-editing tool (Christian et al., 2010; Li et al., 2011).

ZFNs are known to cleave at off-target sites. This hampers their use and has been shown to cause cellular toxicity (Gaj et al., 2013; Jiang et al., 2013a). ZFNs are also difficult (and costly) to design and construct with variable rates of success (reviewed in (Gaj et al., 2013; Jiang et al., 2013a). Compared to ZFNs, TALEN assisted genome editing has significantly reduced toxicity due to off-target effects; however, construct design complexity due to specific requirements in base composition coupled with a lack of support for the TALEN lentiviral delivery systems (reviewed in (Gaj et al., 2013; Holkers et al., 2013) have held back broad adoption and use of TALENs (Sander and Joung, 2014).

The difficulties of both ZFN and TALEN techniques lie in designing and validating proteins that recognize specific DNA sequences. In contrast, the CRISPR system is RNA-mediated. The natural CRISPR system is a defense mechanism that provides bacterial adaptive immunity to a wide range of potential

pathogens (Barrangou et al., 2007; Rath et al., 2015). There are three major classes (types I, II, III) and ten subclasses of CRISPRs based on the specific CRISPR-associated (Cas) proteins and non-coding RNA species involved (Carte et al., 2014; Makarova et al., 2011). The type II CRISPR-Cas9 system has been co-opted for genome editing.

The native CRISPR-Cas9 system (**Fig. 1**) is comprised of three distinct architectural components: a small non-coding transactivating CRISPR RNA (tracrRNA), an operon that encodes the Cas proteins, and a repeat array encompassing crRNA units comprised of a 5' 20-nucleotide targeting sequence and a 19-22 nucleotide repeat sequence (referred to as spacers; Deltcheva et al., 2011). Multiple studies suggest that Cas9 endonuclease activity requires a highly conserved 3' three nucleotide protospacer adjacent motif (PAM) directly preceding the target sequence (Jiang et al., 2013b; Zhang et al., 2014). PAM sequence composition is highly diverse depending on the CRISPR type/subtype, with NGG representing the most effective trinucleotide for the CRISPR-Cas9 system of *Streptococcus pyogenes* (Zhang et al., 2014).

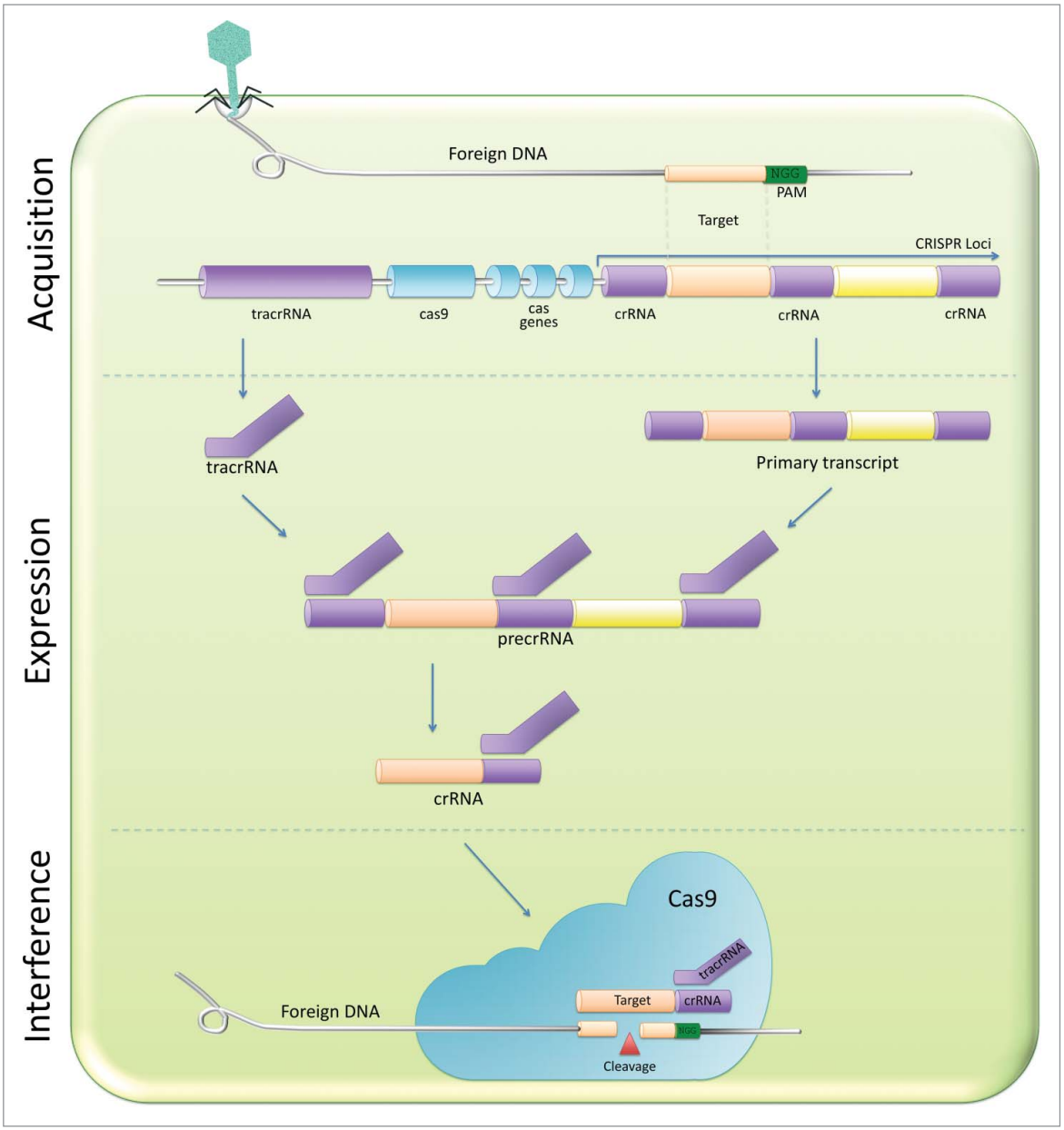
The native CRISPR-Cas9 genome editing mechanism is broken into 3 processes: acquisition, expression, and interference (Carte et al., 2014; Makarova et al., 2011). Upon host infection, exogenous genetic elements are incorporated into the CRISPR locus (acquisition phase). These repeat sequences are then transcribed into noncoding precursor CRISPR RNAs (pre-crRNAs; expression phase). The Cas9 nuclease uses these guide RNA sequences to cleave invading plasmids or phage molecules including any double stranded DNA matching the CRISPR RNAs (interference). Double strand DNA breaks are repaired via non-homologous end joining or homology directed repair *in vivo*, frequently leading to errors or elimination of invading DNA.

To simplify the system for targeted mutation, researchers combined the endogenous tracrRNA and crRNA to produce effective single guide RNA (sgRNA) constructs with unique restriction sites for targeting oligo insertion. The broad applicability of CRISPR to gene editing in diverse species coupled with simple design rules has resulted in the development of myriad bioinformatics tools that aim to identify potential sgRNA target sites in genomes of interest. Although multiple CRISPR sequence design tools already exist, they are not all the same. Some are user friendly, others are more difficult to use. Some are available via web servers, others are not available online. Many perform only a few steps in a full computational analysis and design pipeline, and deliver results that are voluminous with no mechanism to sort. In addition, the genomes available for use within many tools are limited, and very few tools have been subjected to peer-review. To help researchers choose a tool that best suits their specific research needs, we compared the functionality of various CRISPR design software including our own, CGAT the CRISPR Genome Analysis Tool.

CRISPR COMPUTATIONAL RESOURCES COMPARISON

Of the available CRISPR resources we evaluated (see **Table 1**), there are two major classes: those that enable researchers to query experimentally validated sgRNAs for which genetic stocks are available, and those that predict potential CRISPR targets in a given sequence. At the time of this writing, the only resource we find that is in the former category is CrisprGE, though we anticipate that other species will develop such resources in the very near future. CrisprGE is a high-quality, curated database that contains thousands of sgRNAs for hundreds of constructs and their available germplasm resources. To locate resources of interest, tools that enable browse and search functionality are available from the website at http://crdd.osdd.net/servers/crisprge/. In contrast to this resource, other tools predict which sites within a given DNA sequence are amenable to CRISPR-based editing. For the remainder of this discussion, we focus on tools that can be used to predict potential CRISPR targets given an input sequence.

FIGURE 1. The CRISPR-Cas adaptive immune system. Three processes underlie the system, acquisition, expression, and interference. Foreign DNA is shown entering the cell. During acquisition, target DNA (beige; next to the PAM sequence shown in green) is incorporated into the CRISPR locus. Expression involves transcribing target DNA into noncoding pre-crRNAs to which tracrRNAs attach. During interference the Cas9 endonuclease uses these sequences to target foreign DNA for cleavage. (Color figure available online.)



Multiple computational tools are available to aid in the prediction and design of CRISPR sgRNA constructs to target specific genomic loci. For all tools compared in this analysis, the ability to predict sgRNAs in any user-submitted DNA sequence is possible, enabling researchers to design CRISPR sgRNAs for: various versions of genome

Table 1. CRISPR tool and resources examined

Tool Name	Species	Publication	Web Address
s	vertebrates, invertebrates, plants	Ma et al., 2013	http://cas9.cbi.pku.edu.cn/
CCTop	vertebrates, invertebrates, plants,	Stemmer et al., 2015	http://crispr.cos.uni-heidelberg.de/
CGAT	Plants	This paper	http://cbc.gdcb.iastate.edu/cgat/
CHOPCHOP	vertebrates, invertebrates, plants	Montague et al., 2014	https://chopchop.rc.fas.harvard.edu/
COSMID	vertebrates, invertebrates	Cradick et al., 2014	https://crispr.bme.gatech.edu/
CRISPR design	vertebrates, invertebrates, arabidopsis	N/A	http://crispr.mit.edu/
CRISPRdirect	vertebrates, invertebrates, fungi	Naito et al., 2014	http://crispr.dbcls.jp/
Crispr Finder	Vertebrates invertebrates fungi	Grissa et al., 2007	http://crispr.u-psud.fr/Server/
CrisprGE*	various: plants, animals, fungi, prokaryotes, protists	Kaur et al., 2015	http://crdd.osdd.net/servers/crisprge/
CRISPR Multitargeter	vertebrates, invertebrates, plants	Prykhozhij et al., 2015	http://www.multicrispr.net/
Crispr-P	Plants	Lei et al., 2014	http://cbi.hzau.edu.cn/crispr/
CRISPRseek	vertebrates, invertebrates, fungi, plants, protists	Zhu et al., 2014	http://www.bioconductor.org/packages/release/bioc/html/CRISPRseek.html
CROP-IT	vertebrates: mouse and human	Singh et al., 2015	http://cheetah.bioch.virginia.edu/AdliLab/CROP-IT/homepage.html
E-crisp	vertebrates, invertebrates, plants, fungi, protists	Heigwer et al., 2014	http://www.e-crisp.org/E-CRISP/
flyCRISPR	invertebrates	Gratz et al., 2014	http://flycrispr.molbio.wisc.edu/
GT-SCAN	vertebrates, invertebrates, plants, fungi	O'Brien and Bailey, 2014	http://flycrispr.molbio.wisc.edu/
sgRNAcas9	vertebrates, invertebrates	Xie et al., 2014	http://www.biootools.com/col.jsp?id=140
SSFinder	N/A	Upadhyay and Sharma, 2014	https://code.google.com/p/ssfinder/

*queries sgRNA sequences against experimentally validated sgRNAs for which genetic stocks are available.

assemblies, non-model species of interest, and diverse alleles of genes of interest. For some tools, a database of sequences is preloaded, enabling the user not only to specify a gene of interest within a sequenced reference genome, but also to optionally search the rest of the genome for off-target sites that could be recognized by sgRNAs.

In **Table 2**, sgRNA design tools are compared based on whether they are available online via web server, allow the user to search for matching sgRNAs by gene name, provide options to use alternate PAM sequences, provide options to predict off-targets (by genomic sequence similarity), sort and/or rank lists of identified targets, and aggregate all analyses within a single, all-in-one pipeline. Here we specifically highlight the functionality of 17 CRISPR design tools and report on their comparative functionality (**Table 2**). Note that the tools compared here are limited to non-commercial software, though the commercial tools to enable sgRNA design are very much in keeping with functionality described here.

CRISPRseek, sgRNAcas9 and SSFinder are only available as stand-alone systems and require installation and configuration. CRISPR target sequences are identified and evaluated based on user input. These tools are best suited for users with some technical expertise.

Beyond databases of validated CRISPR constructs and tools that must be downloaded and

Table 2. Comparison of CRISPR tool functionalities

Tool Name	Web Server	Search by Gene Name	Alternate PAM Sequence	Predicts Off-targets	Ranks Output	All in One Tool
Cas9-Design	✓	×	×	✓	×	✓
CCTop	✓	×	✓	✓	×	✓
CGAT	✓	✓	×	✓	✓	✓
CHOPCHOP	✓	✓	✓	✓	✓	✓
COSMID	✓	×	✓	✓	✓	✓
CRISPR design	✓	×	✓	✓	✓	✓
CRISPRdirect	✓	✓	✓	✓	✓	✓
Crispr Finder	✓	×	×	✓	×	×
CRISPR Multitargeter	✓	×	✓	✓	×	×
Crispr-P	✓	✓	✓	✓	✓	✓
CRISPRseek	×	×	✓	✓	×	✓
CROP-IT	✓	✓	✓	✓	✓	×
E-crisp	✓	×	✓	✓	×	×
flyCRISPR	✓	×	✓	×	×	✓
GT-SCAN	✓	×	✓	✓	×	✓
sgRNAcas9	×	×	✓	✓	×	×
SSFinder	×	×	×	×	×	×

installed, myriad online tools exist that allow users to quickly parse an input to predict putative CRISPR targets. Tools in this category tend to allow the greatest amount of user flexibility in terms of sgRNA design criteria. As the CRISPR system continues to improve, specifications such as the ability to search non-canonical PAM sequences, an option to designate promoter-specific bases preceding the seed sequence, and improved prioritization for potential targets will provide the greatest expansion in utility across a multitude of genomes and cell types.

A major concern with targeted nuclease technology is the potential for off-target cleavage and associated toxicity. With this in mind, many tools check the rest of a genome for additional matches to predicted target sequences. Even more sophisticated tools produce a ranked output of CRISPR targets by interpreting off-target scores as a function of the overall sgRNA score.

Only CGAT, Crispr-P, CHOPCHOP and CRISPRdirect offer access online, enable search by gene name, predict off-targets, enable ranking of identified targets, and contain all of these functionalities within a single pipeline. Here we describe the functionality of CGAT and demonstrate its capability as a specific example that shows how such tools work.

MATERIALS AND METHODS

CGAT is built upon a variety of technologies. PostgreSQL 9.3 (http://www.postgresql.org/) is the relational database system (RDBMS). For data retrieval, CGAT makes use of PostgreSQL's procedural language extensibility with portions of the database query logic written in PL/Python (http://www.postgresql.org/docs/9.3/static/plpython.html). The current version of the parser that processes genomic FASTA-formatted files into relational database tables is written in the Go programming language (version 1.4.2) (https://golang.org/).

The website itself is written in Python 2.7.x using the 1.8.2 version of the Django framework (https://www.djangoproject.com/). Finally, the client-side functionality of the tool is written in Javascript using the 1.3.9 version of the AngularJS framework (https://angularjs.org/).

Code is available online at https://github.com/ISU-Crop-Bioengineering-Consortium/crispr. While the above technology stack is relatively stable, version numbers of discrete pieces of the stack are likely to change as CGAT and the individual technologies on which it is built mature over time.

At this time, the current genome assemblies for maize, soy, rice, Chlamydomonas, peanut, and sorghum are available for gene model-

specific query from within the CGAT tool. The breadth of the list will be increased over time, with a plan to automatically populate the available genomes from long-lived resources (e.g., EnsembPlants; Kersey et al., 2016). Because the CGAT tool accepts DNA sequence as input directly, DNA sequence from any organism can be evaluated for sites amenable to CRISPR design, but only those with genomes loaded into the database can be evaluated for potential off-target sites.

RESULTS

In overview, the CGAT tool works in two steps. In step one, CRISPR targets are identified in a user-specified sequence of interest with the sequence being pasted into a text field or selected from a list of gene/gene model names from the species of interest. In the second step, potential off-targets are identified. These two functionalities encompass the following steps:

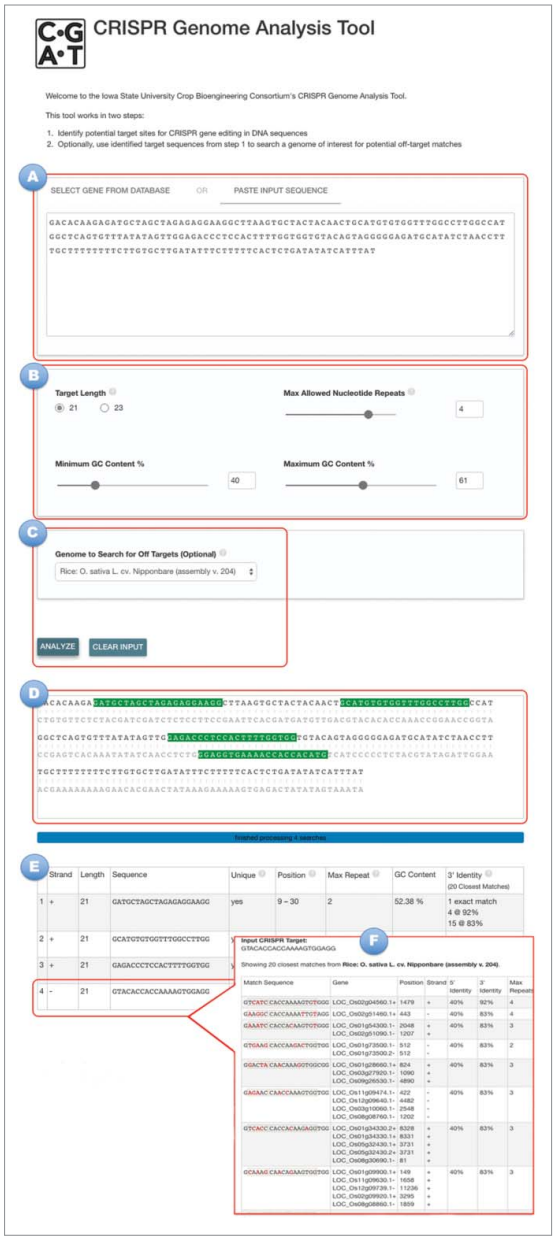
1. For each genome available to search above, the genome sequence has been parsed in advance for valid CRISPR target sequences. All found target sequences were exported to a SQL database along with some relevant metadata. Additionally, the transcript data for each gene has also been stored in the SQL DB for easy retrieval when a user opts to select the input sequence from a specific gene.
2. In the tool interface, Javascript is used to parse both the input sequence and its complement for valid CRISPR targets based on the user-provided search parameters (i.e., Target Length, GC Content and Allowed Nucleotide Repeats). The results are rendered in the browser and, for each found target sequence, a request is sent to the webserver to search the specified genome database for potential off-target matches.
3. For each request sent from the web browser to the webserver in the previous step, the server queries the database for the target genome with the user-provided search parameters.
4. Search results are filtered and sorted primarily by an identity score between an input subsequence (bases 6–18 for 21 base sequences or bases 6–20 for 23 base sequences) and the corresponding subsequences stored in the database. Additional sorting is performed based on an identity score between the subsequence at bases 2-5 of the input sequence and the corresponding subsequences in the database.
5. Finally, the webserver returns the search results to the browser, which updates the existing table. Clicking any table row reveals more details about the result.

OsSWEET11 Example

The SWEET gene family of sugar transporters has been shown to play a vital role in multiple plant growth and developmental processes, including seed nutrition. They are also responsible for host recognition and subsequent sugar acquisition by the bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* - the causal agent of rice bacterial blight (Chen et al., 2010; Boch et al., 2014). Jiang et al. demonstrated that efficient Cas9-mediated modification of the *OsSWEET11* promoter decreased pathogen-host interaction in rice (2013b). Here we search *japonica* rice (*Oryza sativa* L. cv. Nipponbare) for the same target as a representative usage example for CGAT.

As shown in **Figure 2**, the sequence for the *OsSWEET11* gene promoter (GenBank: CM000145.1 nucleotides 25503600-25503800) was used as input. CGAT default parameters were set to identify CRISPR targets of at least 21 nucleotides. The results table highlights potential CRISPR target regions in green. The *OsSWEET11* CRISPR target exploited by Jiang et al. (2013b) to induce a mutation that increased host resistance to bacterial blight is the last in the group (i.e., sequence 5'-GTACACCAC-CAAAAGTGGAGG-3'). Next, the targets were used to query for off-target matches genome-wide. No off-target 100% identical to the Jiang et al. target was identified in the rice genome.

FIGURE 2. CGAT example functionality using *OsSWEET11*. (A) Paste into the box a sequence (or select a sequence from the database). (B) Specify design parameters including target length, the maximum number of tandemly repeated nucleotides, and minimum/maximum GC content (which has been shown to correlate with sgRNA efficiency; Ren et al. 2014). (C) Select a genome to query for potential off-target recognition and hit the 'Analyze' button. (D) Evaluate and prioritize targets using sequence identity as well as (E) off-target sequence identity. (Color figure available online.)



CONCLUSIONS AND FUTURE WORK

The CRISPR-Cas adaptive immune system continues to show increased potential as an excellent tool for genome editing. This obvious and general use across the life sciences has sparked the rapid production of bioinformatics tools to predict and analyze target sequences across a multitude of genomes. In this review, we compared functionality among a list of CRISPR prediction software and described in detail how to use CGAT.

To enable generalized bioinformatics support of the CRISPR-Cas9 system, emerging CRISPR sequence analysis tools are anticipated to provide functionality beyond guide RNA design and off-target identification. Improvements that would simplify the process include: direct access to public sequence databases such as ENSEMBL and Genbank at NCBI, the addition of integrated tools to simplify cloning vector design, and identification of restriction enzyme cut sites within target sequences to simplify screening putative transformants by restriction digest of PCR products. Additionally, reporting whether off-target matches represent duplicate genes and/or gene family members would be a useful feature.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

ACKNOWLEDGMENTS

We thank Darwin Campbell for help with computer administration and CGAT icon creation. Reviewer comments and suggestions were helpful and are greatly appreciated.

FUNDING

Efforts by SZ were supported by the Iowa State University's Crop Bioengineering Consortium, a Presidential Initiative (described at http://cropbioengineering.iastate.edu/). VAB was partially supported by the Iowa State University Graduate Minority Assistantship Program.

AUTHOR CONTRIBUTIONS

VAB: Developed design criteria and usage examples. Contributed heavily to writing the manuscript.

SZ: Developed design criteria, coded the CGAT tool, and contributed to writing the manuscript.

DW: Advised CGAT design and contributed to writing the manuscript.

YW, JL, and KC: Created a working CGAT proof-of-concept and approved the manuscript.

BY: Conceived of the tool, contributed to CGAT design, and contributed to writing the manuscript.

CJLD: Guided CGAT development and contributed heavily to writing the manuscript.

REFERENCES

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007; 315(5819):1709-12; PMID:17379808; http://dx.doi.org/10.1126/science.1138140

Beerli RR, Barbas CF. Engineering polydactyl zinc-finger transcription factors. Nat Biotech 2002; 20(2):135-141; http://dx.doi.org/10.1038/nbt0202-135

Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. Ann Rev Phytopathol 2010; 48:419-36; http://dx.doi.org/10.1146/annurev-phyto-080508-081936

Boch J, Bonas U, Lahaye T. TAL effectors–pathogen strategies and plant resistance engineering. New Phytol 2014; 204, 823-32; PMID:25539004; http://dx.doi.org/10.1111/nph.13015

Boch J, Scholze H, Schornack S Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 2009; 326(5959):1509-12; http://dx.doi.org/10.1126/science.1178811

Bogdanove, AJ. Principles and applications of TAL effectors for plant physiology and metabolism. Curr Opin Plant Biol 2014; 19:99-104; PMID:24907530; http://dx.doi.org/10.1016/j.pbi.2014.05.007

Carte J, Christopher RT, Smith JT, Olson S, Barrangou R, Moineau S, Terns MP. The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in Streptococcus thermophilus. Mol Microbiol 2014; 93

(1):98-112; PMID:24811454; http://dx.doi.org/10.1111/mmi.12644

Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, Frommer WB. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature 2010; 468(7323):527-32; PMID:21107422; http://dx.doi.org/10.1038/nature09606

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 2010; 186:757-61.; PMID:20660643; http://dx.doi.org/10.1534/genetics.110.120717

Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. Mol Therapy Nucleic Acids 2014; 3: e214; http://dx.doi.org/10.1038/mtna.2014.64

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011; 471 (7340):602-607; PMID:21455174; http://dx.doi.org/10.1038/nature09886

Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotech 2013; 31(7):397-405; http://dx.doi.org/10.1016/j.tibtech.2013.04.004

Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics 2014; 196 (4):961-71; PMID:24478335; http://dx.doi.org/10.1534/genetics.113.160713

Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 2007; 35(Web Server issue):W52-7; PMID:17537822; http://dx.doi.org/10.1093/nar/gkm360

Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. Nat Methods 2014; 11 (2):122-3; PMID:24481216; http://dx.doi.org/10.1038/nmeth.2812

Holkers M, Maggio I, Liu J, Janssen JM, Miselli F, Mussolino C, Recchia A, Cathomen T, Gonçalves MAFV. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucleic Acids Res 2013; 41(5):e63; PMID:23275534

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotech 2013a; 31(3):233-9

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 2013b; 41(20): e188; PMID:23999092

Kaur K, Tandon H, Gupta AK, Kumar M. CrisprGE: a central hub of CRISPR/Cas-based genome editing. Database: The J Biol Databases Curation 2015; bav055

Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res 2016; 44(D1)

Klug A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. Annu Rev Biochem 2010; 79:213-31; PMID:20192761

Lee MS, Gippert GP, Soman KV, Case DA, Wright PE. Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science 1989; 245 (4918):635-637; PMID:2503871

Lei Y, Lu L, Liu H-Y, Li S, Xing F, Chen LL. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol Plant 2014; 7(9):1494-6; PMID:24719468

Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, Yang B. TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucl Acids Res 2011; 39:359-72.; PMID:20699274

Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Ann Rev Biochem 2010; 79:181-211; PMID:20192759

Ma M, Ye AY, Zheng W, Kong L. A Guide RNA Sequence Design Platform for the CRISPR/Cas9 system for model organism genomes. Biomed Res Int 2013; 2013:270805

Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Koonin EV. Evolution and classification of the CRISPR-Cas systems. Nat Rev 2011; 9(6):467-77

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 (W1):W401-7; PMID:24861617

Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science 2009; 326 (5959):1501.; PMID:19933106

Muller HJ. Artificial transmutation of the gene. Science 1927; 66(1699):84-87; PMID:17802387

Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics 2014; 31 (7):1120-3; PMID:25414360; http://dx.doi.org/10.1093/bioinformatics/btu743

O'Brien A, Bailey TL. GT-Scan: identifying unique genomic targets. Bioinformatics 2014; 30 (18):2673-5; http://dx.doi.org/10.1093/bioinformatics/btu354

Peng Y, Clark KJ, Campbell JM, Panetta MR, Guo Y, Ekker SC. Making designer mutants in model organisms. Development 2014; 141(21):4042-4054; PMID:25336735; http://dx.doi.org/10.1242/dev.102186

Prykhozhij 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(3): e0119372; PMID:25742428; http://dx.doi.org/10.1371/journal.pone.0119372

Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie 2015; 117:119-28; PMID:25868999

Ren X, Yang Z, Xy J, Sun J, Mao D, Hu Y, Yang SJ, Qiao HH, Wang X, Hu Q, et al. Enhanced specificiy and efficiency of the CRISPR/Cas9 systen with optimized sgRNA parameters in Drosophila. Cell Rep 2014 9 (3):1151-1162; PMID:25437567; http://dx.doi.org/10.1016/j.celrep.2014.09.044

Ruiz i Altaba A, Perry-O'Keefe H, Melton DA. Xfin: an embryonic gene encoding a multifingered protein in Xenopus. EMBO J 1987; 6(10):3065-70.; PMID:2826129

Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotech 2014; 32(4):347-55; http://dx.doi.org/10.1038/nbt.2842

Singh R, Kuscu C, Quinlan A, Qi Y, Adli M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. Nucleic Acids Res 2015; 43(18):e118

Stadler LJ. Mutations in barley induced by X rays and radium. Science 1928; 68:186-7; PMID:17774921; http://dx.doi.org/10.1126/science.68.1756.186

Stadler LJ. The Effect of X-rays upon Dominant Mutation in Maize. Proc Natl Acad Sci U S A 1944; 30 (6):123-8.; PMID:16588634; http://dx.doi.org/10.1073/pnas.30.6.123

Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PloS One 2015; 10(4):e0124633; PMID:25909470; http://dx.doi.org/10.1371/journal.pone.0124633

Upadhyay SK, Sharma S. SSFinder: High throughput CRISPR-Cas target sites prediction tool. Biomed Res Int 2014; 2014:742482

White FF, Potnis N, Jones JB, Koebnik R. The type III effectors of Xanthomonas. Mol Plant Pathol 2009; 10 (6)749-66; PMID:19849782; http://dx.doi.org/10.1111/j.1364-3703.2009.00590.x

Wright DA, LI T, Yang B, Spalding MH. TALEN-mediated genome editing: prospects and perspectives. Biochem J 2014; 462(1):15-24; PMID:25057889; http://dx.doi.org/10.1042/BJ20140295

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(6):e100448; PMID:24956386; http://dx.doi.org/10.1371/journal.pone.0100448

Zhang Y, Ge X, Yang F, Zhang L, Zheng J, Tan X, Jin Z-B, Qu J, Gu F. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Scientific Reports 2014; 4:5405; PMID:24956376

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(9):e108424; PMID:25247697; http://dx.doi.org/10.1371/journal.pone.0108424