Flexible guide-RNA design for CRISPR applications using Protospacer Workbench

To the Editor:

The CRISPR-Cas9 endonuclease system permits genome editing at nucleotide resolution, but ensuring high efficiency while minimizing off-target cleavage requires careful design of the single guide RNA (sgRNA). Currently most researchers use online tools (e.g., E-CRISP[1], CRISPR[2], ZiFiT[3]) for sgRNA design. However, these tools have several drawbacks: the number of target-sgRNA mismatches allowed is limited; they are unable to detect every potential off-target site; they are generally limited in flexibility and throughput; and they are available for only a few organisms. Offline solutions such as sgRNAcas9 (ref. 4) provide greater flexibility and speed but require knowledge of programming and/or command-line interfaces. Neither online nor offline software facilitates sharing of sgRNA libraries, their putative off-target sites or the statistical reports that lead to successful designs.

Protospacer Workbench, an offline software for rapid, flexible design of Cas9 sgRNA, addresses all of these issues (**Table 1**). It combines a graphical user interface with a file-based database and third-party sequence mapping tools to maximize flexibility and information retrieval when designing sgRNAs (**Fig. 1**). Important design statistics are calculated, such as the off-target and Cas9-activity scores developed by Hsu *et al.*[2] and Doench *et al.*[5], respectively. Protospacer allows researchers to build, analyze and share their own databases of CRISPR targets, facilitating the development of custom sgRNA libraries and the transfer of CRISPR technology to new organisms. Protospacer can manage multiple databases for laboratories interested in several strains or organisms.

Protospacer databases can be created from any FASTA file and annotated using either the GTF or GFF format. Each database begins as a simple catalog of the sequence and annotation data and grows through use. Protospacer's flexible user interface is structured around four main tasks: target finding, broad target filtering, selection of candidate targets, and sgRNA design and analysis. Target finding allows searching in one or more regions of interest, defined by genomic ranges, distance to a point of interest, gene identifier, feature attribute or sequence similarity. Potential sgRNA targets found within regions of interest can be filtered by nucleotide content, by the Doench-Root activity[5] score or by uniqueness in the genome. One or more of the remaining targets may be marked for further analysis, plotting, saving and annotation. The workbench also supports connection to the Broad Institute's Integrative Genomics Viewer (IGV)[6], which allows targets to be viewed in the context of arbitrary genome browser tracks such as genomic annotations, amino acid sequences and high-throughput sequencing data. We provide online video tutorials and written documentation to familiarize users with the Protospacer interface. The software, tutorials, mailing list and databases for several organisms are located at http://www.protospacer.com/. The Protospacer

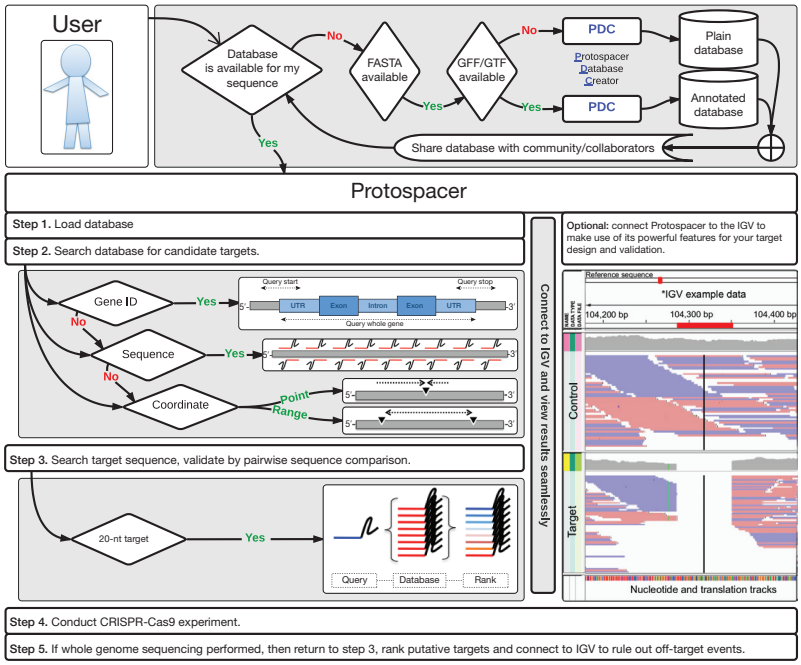


Figure 1 Typical Protospacer Workbench users first obtain a useful Protospacer database from www.protospacer.com or collaborators, or build a custom database. Step 1: the database is loaded into the Protospacer system. Step 2: targets are selected by gene, exon/intron/UTR, sequence or genomic coordinate. Targets may then be filtered by nucleotide content, their Doench-Root activity score and/or their uniqueness in the genome. To aid in target selection, targets may be viewed in the context of other genomic annotations or amino acid sequences by connecting to the integrated genomics viewer (IGV) browser (optional). Step 3: candidate sgRNAs are generated and analyzed, including analysis of potential off-target sites. Finally, Protospacer Workbench allows researchers to rank predicted off-target sites for any further experimental follow-up. *IGV example data provided by Ghorbal *et al.*[7].

Table 1 Feature comparison of Protospacer Workbench with six online and offline solutions

		Protospacer WB	E-CRISPR	CRISPR (MIT)	ZiFiT	SSFinder	sgRNAcas9	Cas-offinder
gRNA design	Single target design	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Paired target design	Yes (very fast)	Yes (fast)	Yes (slow)	Yes (very fast)	No	No	No
	Target-site ranking	Yes	Yes	Yes	No	No	No	No
	Off-target prevention	Yes	Yes	Yes	No	Excludes duplicates	Yes	Yes
	Off-target validation	Yes (IGV+HTSeq)	No	No	No	No	No	No
	Predict cleavage efficiency	Yes (Doench score)	No	No	No	No	No	No
	Search by sequence similarity	Yes	No	No	No	No	No	No
	Search by gene ID	Yes (any)	Yes (ENSEMBLE)	No	No	No	No	No
	Search by genomic coordinates	Yes	No	No	No	No	No	No
	Considers NAG and NGG off-target sites	Both	Both	NGG only	NGG only	NGG only	NGG only	Both
	Filters (more refined searches)	Yes	Yes	No	No	No	No	No
	Documentation	Video tutorial + text	Text (good)	Text (good)	Text (poor)	Text (poor)	Text (good)	Text (good)
	Dedicated website	Yes	Yes	Yes	Yes	No	No	Yes
Input	Database of organism	Required	Required	Required	NA	NA	NA	NA
	Sequence of interest	Optional	Optional	Required	Required	Required	Required	Required
	Gene ID	Optional	Optional	NA	NA	NA	NA	NA
Software design	Platform	Mac OS X	All	All	All	All	All	All OpenCL
	32/64 bit	64 bit	NA	NA	NA	Both	Both	Both
	Online/offline	Offline	Online	Online	Online	Offline	Offline	Offline
	Target audience	Biologist or informatician	Biologist	Biologist	Biologist	Informatician	Informatician	Informatician
	Programming knowledge	Not required	Not required	Not required	Not required	Required	Required	Required
	Flexibility of design	++++	+++	++	+	++	+++	+++
	Graphical user interface	Yes	Yes	Yes	Yes	No	No	No
	Technical ability required	Easy to advanced	Easy	Easy	Easy	Advanced	Advanced	Advanced
	Aid in transfer of CRISPR technology to new organisms	Yes	No	No	No	Yes	Yes	Yes
	Installation	Easy	NA	NA	NA	Technical	Technical	Technical
Limitations	Genome size	Very large	Very large	Very large	Very small	Small	Very large	Very large
	Custom genome/sequence	Yes	No	No	Yes	Yes	Yes	Yes
	Infrastructure/maintenance	Small	Large	Large	Large	Small	Small	Small
	Access to information for downstream analysis	High	Medium	Medium	Low	Low	High	High
	Genomes available	Unlimited: dependent on community and author	Limited: dependent on author	Limited: dependent on author	Unlimited short sequences	Limited: small genomes	Unlimited: dependent on community and author	Unlimited: dependent on community and author
	Speed of single search	++++	+++	+	+++	++	++++	++++
	Speed of design (multiple searches)	+++	++	+	++	+	+	+
	gRNA design	N(20)NGG	N(20)NRG	N(20)NGG	N(20)NGG	N(20)NGG	N(20)NGG	N(20)NNN

Workbench has been launched on the Mac OS X operating system with opt-in queues for Windows and Linux users who wish to take part in early-access testing.

ACKNOWLEDGMENTS

The authors would like to thank M. Ghorbal for extensive testing and recommendations. This work was supported by the French Parasitology consortium ParaFrap (ANR-11-LABX0024) ERC Advanced grant (PlasmoEscape 250320).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Cameron Ross MacPherson & Artur Scherf

[1]Biology of Host-Parasite Interactions Unit, Institut Pasteur, Paris, France. [2]CNRS, ERL 9195, Paris, France. [3]INSERM, Unit U1201, Paris, France.

e-mail: cameron.crm@gmail.com or artur.scherf@pasteur.fr

Published online 29 June 2015; doi:10.1038/nbt.3291

1. Heigwer, F., Kerr, G. & Boutros, M. *Nat. Methods* **11**, 122–123 (2014).
2. Hsu, P.D. *et al. Nat. Biotechnol.* **31**, 827–832 (2013).
3. Sander, J.D., Zaback, P.Z., Joung, J.K., Voytas, D.F. & Dobbs, D. *Nucleic Acids Res.* **35**, W599–W605 (2007).
4. Xie, S., Shen, B., Zhang, C., Huang, X. & Zhang, Y. *PLoS ONE* **9**, e100448 (2014).
5. Doench, J.G. *et al. Nat. Biotechnol.* **32**, 1262–1267 (2014).
6. Robinson, J.T. *et al. Nat. Biotechnol.* **29**, 24–26 (2011).
7. Ghorbal, M. *et al. Nat. Biotechnol.* **32**, 819–821 (2014).