# CROPSR: An Automated Platform for Complex-Genome Wide CRISPR gRNA Design and Validation

Hans Müller Paul,[1,2,3]* (hmpaul2@illinois.edu), Jacob Heldenbrand,[1,4] and **Matthew Hudson**[1,2,3,4]

[1]University of Illinois at Urbana-Champaign, Urbana, Ill.; [2]DOE Center for Advanced Bioenergy and Bioproducts Innovation, Urbana, Ill.; [3]Carl R. Woese Institute for Genomic Biology, Urbana, Ill., [4]National Center for Supercomputing Applications, Urbana, Ill.

https://cabbi.bio/research/feedstocks-theme/

**Project Goals: RNA guided technologies such as CRISPR/Cas9 are a revolutionary tool for crop engineering and synthetic biology. To take advantage of this technology, scientists use online guide RNA design aid software; generally designed based on human and mouse genomes. Differences in genomic architecture between these genomes and the complex, paleopolyploid genomes of bioenergy and bioproduct crops are impacting the utility of the technology. Our goal is to develop a unified software platform to direct edits to specific genes or groups of genes using the CRISPR system. Experimental validation is key to improving both guide RNAs and crop traits, but polymerase chain reaction (PCR) validation can be challenging in complex intergenic regions or genes with multiple copies. The software will design guide RNAs genome-wide, evaluate them based on experimental data including plant genomes, and provide the optimal guides along with unique primers for PCR validation.**

Recent advances in genome editing technology such as the CRISPR/Cas9 system have changed the pace and face of functional genomics research. This RNA guided technology is a groundbreaking innovation and an excellent tool for plant biotechnology due to its efficiency, accuracy, and ease of use (1, 2). The key components can often be assembled in a single plasmid, facilitating a quick turnaround time in the experimental procedure. Currently, the default pipeline includes the scanning of the target gene sequences for protospacer adjacent motif (PAM) sites, where the Cas9 enzyme cuts in the DNA, designing and validating the target specificity of the guide RNA, introducing the gRNA and the Cas9 in to the plant and then validating the mutations with PCR and sequencing. Currently, these design steps are mostly done manually by bench scientists, using various software products to help design the gRNA and the PCR primers. This process is labor intensive and can introduce inconsistency in experimental efficiency. We have designed and are building a suite of software that will maximize success probability by controlling all the parameters of CRISPR/CAS9 genome editing construct design and validation, from target selection to accurate validation and interpretation of edits. Our approach consists of firstly pre-computing all potential target PAM sites in the genome, designing a guide RNA for each site, and storing them in a data frame, where we can easily update each entry to add information. We then evaluate each gRNA for their on-site and off-site activities utilizing a machine-learning approach that combines support vector machine (SVM) with L1 logistic regression. The currently accepted standard methods utilize human and mouse genomes for the

training and evaluation steps (3-5), whereas we have also introduced data from the rice genome (6) to improve the predictive power of the model in energy crops such as *Sorghum bicolor*. The scored gRNAs are then associated with the genomic features present in their genomic positions, allowing for easy selection by the user. Our tool also provides PCR primers for validation with uniqueness in the genome, confirmed by Burrows-Wheeler alignment. Our highly parallelized software leverages the power of high-performance computing to provide the user with a wide array of options to choose for each target, allowing them to proceed to the next stages of genome editing experiments with increased confidence of success.

## References


1. L. Arora, A. Narula, Gene Editing and Crop Improvement Using CRISPR-Cas9 System. *Front Plant Sci* **8**, 1932 (2017).

2. X. Liu, S. Wu, J. Xu, C. Sui, J. Wei, Application of CRISPR/Cas9 in plant biology. *Acta Pharm Sin B* **7**, 292-302 (2017).

3. J. G. Doench *et al.*, Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**, 1262-1267 (2014).

4. J. G. Doench *et al.*, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).

5. J. Listgarten *et al.*, Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature Biomedical Engineering* **2**, 38-47 (2018).

6. Y. Lu *et al.*, Genome-wide Targeted Mutagenesis in Rice Using the CRISPR/Cas9 System. *Mol Plant* **10**, 1242-1245 (2017).



*This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*