HARNESSING THE DIVERSITY OF CAS9 ORTHOLOGS FOR GENOME EDITING

Tomas Urbaitis[1,3], Giedrius Gasiunas[1], Joshua K. Young[2], Monika Jasnauskaite[1], Mantvyda Grusyte[1], Sushmitha Paulraj[2], Jennifer L. Curcuru[4], Megumu Mabuchi[4], Ryan T. Fuchs[4], Ezra Schildkraut[4], G. Brett Robb[4] and Virginijus Siksnys[1,3]

[1] CasZyme, Vilnius, LT-10257, Lithuania
[2] Department of Molecular Engineering, Corteva Agriscience™ Agriculture Division of DowDuPont™, Johnston, IA 50131, USA
[3] Institute of Biotechnology, Vilnius University, Vilnius, LT-10257, Lithuania
[4] New England Biolabs, Ipswich, MA 01938, USA
tomas.urbaitis@gmc.vu.lt

The Cas9 protein from CRISPR (Clustered Regularly Interspaced Palindromic Repeats)-Cas (CRISPR Associated) bacterial defense systems has been adopted as a robust and multifaceted genome editing tool. The Cas9 RNA guided DNA endonuclease can be directed to cleave, nick or bind a specific site in the chromosomal DNA just by changing the guide RNA sequence. Cas9-based tools have been used to edit genomic DNA, modulate gene expression, visualize genomic loci in cells and deaminate nucleotide bases. However, for Cas9 to bind a given target, a short nucleotide sequence motif, termed PAM, is required. This PAM constraint as well as insufficient specificity are major obstacles for Cas9 genome editing. Thus, analysis of natural Cas9 orthologs could offer an increased diversity of PAM sequences and biochemical properties which may be beneficial to genome editing applications.

Cas9 nucleases are abundant in microbes. To explore this large uncharacterized diversity of Cas9 orthologs, we established a phylogeny-guided bioinformatic selection approach and developed biochemical screens based on cell-free recombinant protein expression and interrogation of plasmid libraries containing randomized PAM sequences for the rapid characterization of novel Cas9 proteins and identification of PAM requirements. Guide RNAs for each Cas9 ortholog were designed *in silico* by identifying putative tracrRNA (trans-activating CRISPR RNA) coding regions in respective native loci. The examined set revealed nucleases that exhibit a wide range of distinctive T-, A-, C- and G-rich PAM preferences, ranging from two to more than four nucleotides, as well as generate staggered-end breaks or require longer spacers to function robustly. Our results indicate that the natural diversity of Cas9 orthologs provides a source of various PAM recognition sequences and other potentially desirable properties that may be used to expand the genome editing toolbox.