

## **Enhancement of the Random Barcode Transposon Sequencing Applications within KBase**

<sup>2</sup>Omree Gal-Oz ([ogaloz@lbl.gov](mailto:ogaloz@lbl.gov)), <sup>2</sup>Adam Arkin, and <sup>1,3</sup>Ryan Gill

<sup>1</sup>Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO;

<sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>3</sup>Danish Technical Institute, Copenhagen, Denmark.

<http://kbase.us/>

### **Project Goals:**

**One of the central challenges in the synthetic design of microbes for operation in variable environments is that productivity and fitness can be negatively affected as environments vary. A central hypothesis of this project is that it is possible to build synthetic circuits in cells that deploy certain genes at the right times to respond to environmental change and optimize fitness and productivity across variable environments. We have been adding computational tooling into the DOE Systems Biology Knowledgebase to support this design workflow<sup>1</sup>. Previously, we have prototyped the use of a circuit design system developed by the Voigt lab at MIT called Cello that, if you know which genes you wish to express when, will design a minimal genetic circuit to meet performance goals. To support such designs, it is necessary to be able to assess which genes of an organism are likely to tune both fitness and productivity in each environment. Recently, a number of different high-throughput genetic techniques have emerged to efficiently enable this assessment. The Arkin lab has developed RB-TNSEQ as an approach to this problem<sup>2</sup>. This version of tagged transposon mutagenesis has been used across diverse bacteria (and some fungi) and has been used to identify critical genes that could be engineered to improve production goals. The experimental protocols and reagents have been published and distributed widely to the community. However, the software to quality assess, analyze and produce the final prioritized list of genes important for a given phenotype is complex and difficult to use. Our goal this period has been to lower the barrier to its use by simplifying, enhancing, and documenting the workflow within KBase.**

The current implementation of RB-TNSEQ on KBase allows users to step through mapping their RB-TNSEQ library, assessing it for bias, and analyzing fitness/phenotype differences across conditions. This involves nine applications listed in tables below which are partially in beta and nearly ready for release. We have created a comprehensive user manual, a well-annotated demonstrative public Narrative, and an instructional video to help users learn the system. We expect full release of this system by the end of Q2, 2022.

The following page consists of three tables which group the applications by function. “Mapping Applications” refers to the applications that should be run using the sequencing reads that are acquired after running transposon insertion. “Fitness Assessment Applications” refers to applications that should be run using the sequencing reads that are acquired when counting the barcodes which have been previously decided upon. “Complimentary Applications” refers to applications that assist with the workflow. Within all categories, the apps that are not necessary for the entire workflow to be completed are labeled with a star.

## Mapping Applications

<b>RBtNSeq MapTables*</b>	Map DNA reads to insertion locations within the genome
<b>RBtNSeq Maps to Pool</b>	Create insertion locations (as above) and create a pool of curated insertions. Returns a list of barcodes to be used downstream.
<b>RBtNSeq Mutant Pool Visualizer*</b>	Run the data visualization component of ‘RBtNSeq Maps to Pool’ - allows for users to visualize <b>completed</b> data after edits

## Fitness Assessment Applications

<b>RBtNSeq Reads to Pool Counts</b>	Count barcodes from reads representing growth in conditions
<b>RBtNSeq BarSeqR</b>	Compute Fitness Results from counted barcodes files
<b>RBtNSeq Fitness Visualizer*</b>	Run the data visualization component of ‘RBtNSeq BarSeqR’

## Complimentary Applications

<b>RBtNSeq Find TnSeq Model*</b>	To be run before mapping applications. This finds the most likely TnSeq Transposon Model used in a previous sequencing run.
<b>RBtNSeq Genome to Gene Table</b>	Convert KBase Genome Objects to Gene Tables (which are similar to .gff files)
<b>RBtNSeq Download Tables*</b>	Download RBtNSeq data objects from KBase as tables (TSV files)

\* - auxiliary application (i.e. not a necessary application to produce fitness results)

## References

1. Wetmore, K.M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**, e00306–e00315 (2015).
2. Arkin, A., Cottingham, R., Henry, C. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* **36**, 566–569 (2018).

*This research on Design and Engineering of Synthetic Control Architectures is supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231 and DE-SC0018368*