**Integrating read-based microbiome taxonomy classification tools into KBase**

Mark Flynn (mflynn@lanl.gov),[1*] Chien-Chi Lo,[1] Paul Li,[1] Karen Davenport,[1] Bin Hu,[1]and Patrick S. G. Chain [1]

[1]Los Alamos National Laboratory, Los Alamos, NM

http://bioedge.lanl.gov
https://edgebioinformatics.org

**Project Goals: Our goal is to add a new capability, read-based taxonomy, to KBase. These classification tools will allow KBase users to analyze bacterial-fungal interactions by determining which organisms are present in a microbiome sample.**

Interactions between bacteria and fungi play a critical role in soil ecosystems. To facilitate the study of these interactions we have added a new capability to KBase, read-based taxonomic classification. Currently KBase has only one taxonomy classifier, the amino acid-based Kaiju algorithm. While it is highly sensitive and able to detect all organisms in a sample that are in its database, it has very low specificity which results in many false positive classifications. To help overcome this problem, we have added four taxonomic classifiers to KBase: Centrifuge, Kraken2, as well as two profilers developed by our team at LANL: GOTTCHA2 and PanGIA. These nucleotide-based classifiers each make different trade-offs between sensitivity and specificity. GOTTCHA2 uses unique signatures within genomes to significantly reduce the number of false positives while maintaining high sensitivity. PanGIA strikes a balance between sensitivity and specificity. Centrifuge and Kraken2 have higher sensitivity and are considerably faster than GOTTCHA2 and PanGIA but have lower specificity. We chose to add all four classifiers to KBase since there may be circumstances where several tools in conjunction would provide a more comprehensive analysis.
We are in the process of creating an app that allows the generation of a custom database of unique genomic signatures from any user-supplied set of references. This app will be used to power the dynamic updating of a database of unique fungal sequence signatures as the JGI adds additional fungal genome assemblies to their publicly available datasets.
Adding these new capabilities will allow users to make use of KBase's flexible platform to: 1) perform enhanced metagenomic data analytics; and 2) and determine unique genomic signatures at each taxonomic rank for fungal genomes (similar to what exists for bacterial genomes) to allow unambiguous assignment and to reduce false discovery.

*This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59C*