

CosmosID

CosmosID allows you to run your samples using CosmosID's curated genome databases and high performance algorithms to provide rapid, accurate, and actionable microbial identification at the species, subspecies, and/or strain level.

CosmosID identifies multiple microorganisms in a whole genome shotgun metagenomic sample using statistical and computational methods, with no prior assumptions as to what is present in the sample. CosmosID uses raw, unassembled reads as input and matches the sequence against our reference databases of known bacteria, viruses, fungi, protists, antibiotic resistance genes, and virulence factor genes.

For use with whole genome shotgun samples only, not 16S.

Signing In

In the CLC genomics workbench, click **Edit -> Preferences**

Scroll down and enter the email address and password provided to you by Qiagen for your CosmosID account.

Analyzing Samples

For paired end samples, use CLC to merge the samples before uploading.

Click **Analyze Sample** and select samples for analysis.

After your samples are uploaded, analysis begins automatically. In the Dashboard you will see the samples you have recently uploaded and their status.

Retrieving Analysis Results

To view results, click **Retrieve Analysis**. Click **Expand** to view the list of results from each database. The 'Status' will say 'Running' if the sample is still undergoing analysis. 'Success' indicates that the sample ran successfully and 'Failed' indicates a problem. Contact support@cosmosid.com for help if needed.

Select the checkboxes next to the results you would like to view and click **Next**. Select **Open** or **Save** and click **Finish**. Once the table is saved or opened, click 'Show Analyses Workbook' (the icon on the left-bottom) to open the table. Interpretation of the table is explained below.

To switch between databases use 'Choose Database:' drop down menu. To switch between filter sets use 'Choose Filter:' drop down menu. CosmosID provides an option to show both filtered and unfiltered results. The filtering threshold is based on internal statistical scores determined by analyzing a large number of diverse metagenomes. Organisms that are identified with 'filtered' are most likely to be present in the sample. Organisms below the filtering threshold are shown as 'unfiltered'; those organisms need further validation to determine if they are actually present in the sample - either by deeper sequencing of the sample followed by re-analysis or by orthogonal validation using targeted PCR or other methods.

Converting to MGM Abundance Table

For utilizing the analysis and visualization tools in MGM plugin, CosmosID provides a feature to convert the result table into MGM abundance table. Click **Convert to MGM Abundance Table** to open the converting wizard window. Following the steps, choosing the database and values, and click **Finish**.

Getting Token Information

To determine how many tokens are available, click **Get Token Information**.

Interpretation of Results

Table:

Name: The organism (most likely strain) identified in the queried sample. Identification of a strain conveys the natural concept of strain, that is, a strain will always have a like counterpart in nature and potentially is a sub-sample derived from a strain found in nature. Therefore, identification of a given strain indicates either presence of itself or its sub-strain sharing highest genomic similarity to it. Branch calls: Sometimes in the results you will see the term 'Branch' used to describe an organism. The CosmosID database is organized as a phylogenetic tree and the term 'Branch' is used to indicate that there are still nodes or leafs below the point where an identification was made.

Taxonomy Id: The NCBI taxonomy id. Clicking on it takes you to the NCBI entry for that organism.

Frequency: The number of unique kmer occurrences in the queried sample. This is roughly equivalent to the number of reads that hit the organism.

Unique Matches Percent: The number of different kmers found in the sample that are unique to the organism identified divided by the number of pre-calculated total possible unique kmers in the reference database.

Total Matches Percent: The shared plus unique matches divided by the pre-calculated shared plus unique matches possible in the reference database.

Relative Abundance: Relative abundance is calculated based on the number of organism specific kmers and their observed frequency in the sample and then normalized to represent the abundance of each organism.

Problems or Questions

Email us at support@cosmosid.com