MUSiCC User Manual

MUSiCC is a toolkit for correcting biases in gene abundance measurements derived from shotgun metagenomic sequencing, and is available as an online tool and as a Python module. MUSiCC is developed by the Borenstein group at the University of Washington and is available online at:

http://elbo.gs.washington.edu/software_musicc.html.

MUSiCC>Web

MUSiCC>Web allows researchers to normalize and correct their metagenomic gene abundance measurements online, and requires only a web browser. The MUSiCC>Web application is built on top of the MUSiCC>Python module.

Overview of use

To normalize and correct the abundance measurements of genes using MUSiCC>Web, the user simply uploads their gene abundance file to the MUSiCC server by filling out the form shown in Figure 1 a). After checking the integrity of the file, MUSiCC>Web normalizes and corrects the measured abundance. A link to the file containing the corrected abundances is displayed along with statistics of the data and correction procedures, as shown in Figure 1 b).

(A)

(B)

Fig. 1: The MUSiCC>Web interface. (A) The user is first presented with a form to upload the abundance file and select the analysis options. (B) The gene abundance file is corrected and available for download, along with statistics collected during the correction.

Setting up the analysis

Selecting the abundance file to correct

Choose File / Browse: Select the abundance file on your local drive. In the drop-down box below, choose the appropriate file type.

MUSiCC>Web supports abundance files formatted as plain text with various delimiters. Gene abundance files have one line per gene, with the gene ID appearing as a row header, followed by all the abundance measurements corresponding to different samples separated by a space, comma or tab. Each line must use the same separating character. The first line of the file must contain column headers with names for each sample, with the first cell in the line containing any string (e.g., KO).

Analysis Options

Once you have specified the abundance file, the next step is to choose what types of corrections should be applied to the data:

Correct inter-sample variation: Selecting this will perform the normalization step of MUSiCC, correcting inter-sample variation.

Correct intra-sample variation: Selecting this will perform the intra-sample correction, with the generic model learned from the Human microbiome project (HMP) stool samples.

Citing information

If you use the MUSiCC software, please cite the following paper:

MUSiCC: A marker genes based framework for metagenomic normalization and accurate profiling of gene abundances in the microbiome. Ohad Manor and Elhanan Borenstein. *Submitted*