Welcome to the interactive metagenome stability diagram guide!

The accompanying literature for this tool can be found here:

https://www.biorxiv.org/content/10.1101/2025.01.16.633425v1

The Data:

All of the data in this tool is sourced from publicly available metagenomes in the Joint Genome Institute (JGI) Integrated Microbial Genomes & Microbiomes (IMG) database. While using this tool does not require you to be signed into a JGI IMG account (free of cost), following any web links to metagenome data will require either signing in to IMG or continuing as a guest and navigating to the metagenome on your own with the IMG Taxon ID that this tool provides. We recommend signing into IMG when prompted for a more smooth user experience.

The Plot:
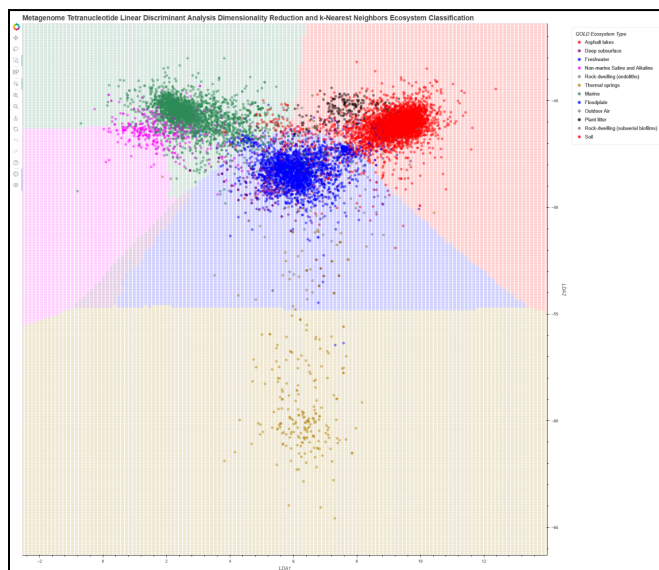
Now, navigating the tool. The main plot on the right side of the tool is where tetranucleotide frequencies from metagenome filtered sequencing reads have been reduced to two dimensional coordinates with linear discriminant analysis (LDA). Each circular point in this plot is a metagenome, colored by its corresponding Genomes Online Database (GOLD) Ecosystem Type classification. The colored background regions are k-nearest neighbor (KNN) Ecosystem classification phases. Metagenome data points are plotted into the KNN Ecosystem phase where k-nearest metagenome data points have the corresponding GOLD Ecosystem Type classification. The KNN Ecosystem classification phases and their color scheme can be found in the Annotation Panel to the left of the plot.

The Plot Toolbar:

The plot toolbar can be found on the left side of the plot. Some tools are selected by default, which is denoted by a blue line highlight just to the right of the tool symbol. The toolbar starts with the upper left most symbol for the plotting software Bokeh, which takes you to their webpage:

The first plot tool is "Pan" which allows you to click and drag your mouse to slide the plot in any direction that you want:

The second plot tool is "Lasso Select" which allows you to select a group of metagenome data points by clicking and dragging to draw an enclosed shape around them. The selected metagenomes will be highlighted and information about the selection will be shown in the Annotation Panel on the left:

The next plot tool is "Box Zoom" which allows you to click and drag to draw a box on the plot to zoom into. The edges of the drawn box will become the new axes of the plot. This plot tool is selected by default:

The next plot tool is "Wheel Zoom" which allows you to zoom the plot view in and out with the scroll wheel on your mouse:

The next plot tool is "Tap" which allows you to select single data points or data points that are overlapping. The selected metagenomes will be highlighted and information about the selection will be shown in the Annotation Panel on the left. This plot tool is selected by default:

The next two plot tools are "Zoom In" and "Zoom Out" which allow you to zoom in and out of your current plot view:

The next plot tool is "Save" which allows you to save your current plot view as an image:

⤓

The next three plot tools are "Reset," "Undo" and "Redo" which allow you to reset the plot view to default, undo a plot action, or redo a plot action, respectively:

↻

↩

↪

The next plot tool is a question mark button that will take you to a Bokeh webpage to learn more about how the plot tools work, including some of the code behind them:

⊙

The last two plot tools are "Crosshair" and "Hover" which allow you to toggle on/off the cursor crosshairs that line up to the LDA1 and LDA2 axes, and the information windows that appear when hovering the cursor over a data point. These two tools are selected by default:

⊕

💬

The Annotation Panel:

The Annotation Panel on the left side of the tool has three sections, separated by black horizontal lines. The top section activates the Protein Annotation View. By entering a protein annotation ID in the format of one of the examples given on the page (COG, KO, Pfam, or EC) into the text box, the plot will overlay gene annotation relative abundances onto the metagenomes in the plot. To accommodate ubiquitously high or low abundances, the heat map color scheme of the protein annotation relative abundance values can be adjusted with the Heat Color Scale slider, after an ID has been entered. To reset the view back to the default ecosystem classifications, enter the work "Ecosystem" into the text box and hit enter.

The middle section is the plotted KNN Metagenome Ecosystem Classification selector, which allows you to toggle on/off the metagenome data points that are plotted in each of the respective phases. By default, all phases are toggled on.

The bottom section is the Selection Table, which displays the IMG Taxon IDs, GOLD Ecosystem Subtypes, GOLD Ecosystem Specifics, and the Mapped Value of the selected data points, whether in the ecosystem or protein annotation view. The IMG Taxon IDs listed for the selected metagenome data points are links to their IMG data page, for a more detailed description of the dataset. If no data points are selected, a button labeled "Download All" will download a data table for all metagenomes in the plot view. If a selection of metagenome data points has been made, then a button labeled "Download Selection" will download a data table for the selected metagenomes.

If you want to hide the Annotation Panel, you can click the "Collapse Annotation Panel" button. To have the Annotation Panel return, you can click the "Expand Annotation Panel" button. You may have to wait a couple seconds before the panel toggle completes.

Have fun, and if you have any questions contact Matthew Kellom at mkellom (at) lbl (dot) gov