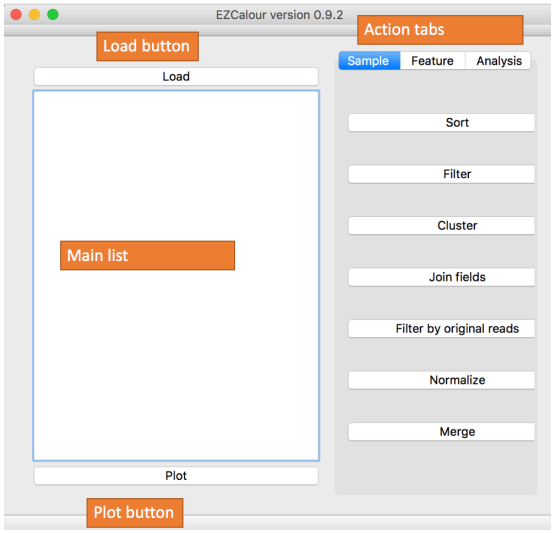
Using EZCalour

EZCalour is a point and click GUI for the Calour microbiome analysis package.

EZCalour can be used to read, process, and plot interactive heatmaps from microbiome experiments.

General concepts



Each dataset is called an experiment. All experiments are displayed in the main list in the EZCalour window. Following each processing step, a new experiment is created. Right clicking on an experiment enables deleting from memory and saving of the experiments.

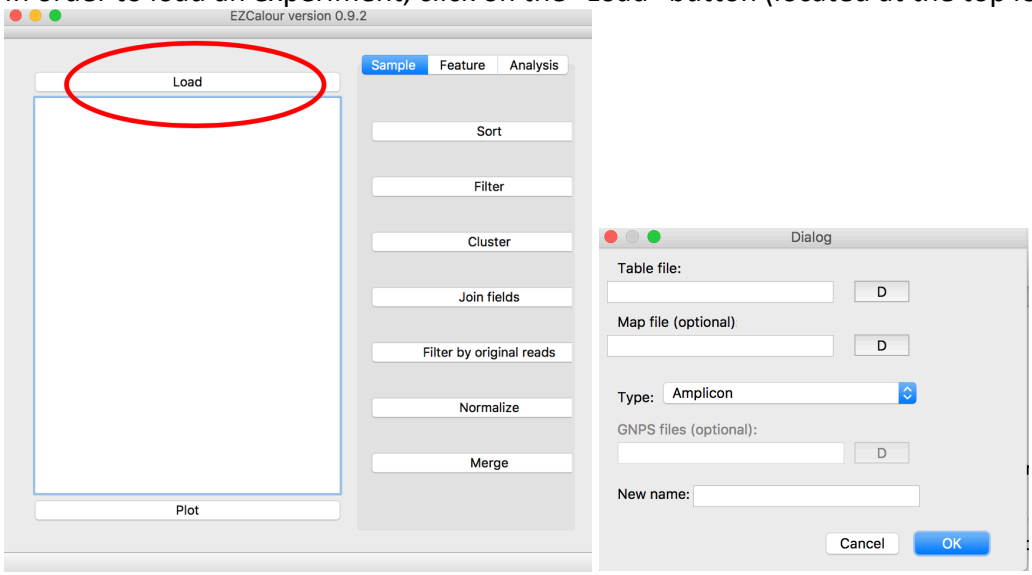
On the right-hand side, there are three action tabs - for processing samples, features and analysis. Commands from a given tab relate to the appropriate axis (i.e. Filter from the sample tab filters samples, whereas Filter from the features tab filters features). Commands work on the selected experiment from the main list.

In order the plot the interactive heatmap of an experiment, it needs to be selected from the main list, and then press the "Plot" button (located at the lower left side)

Loading data

EZCalour works with microbiome BIOM tables, metabolomics MZMine2 tables, or any CSV text file.
Besides the main table, ezcalour can also load a tab-separated mapping file, containing information about each sample.

In order to load an experiment, click on the "Load" button (located at the top left side).



Mandatory fields:

- "Table file" : name of the biom or mzmine2 table (can click the "D" button for GUI file selection)
- "Type": the type of the table file:
    - "Amplicon" for a microbiome biom table. When loading, the table is normalized by TSS to 10000 reads/sample. Samples with <1000 reads are dropped.
    - "MZMine2" for an MZMine2 metabolomics table
    - "TSV" for a general tab separated table (Each sample is a column, each feature is a row)

Optional fields:

- "Map file" : name of the sample TSV mapping file
- "GNPS file" : For mass-spec, the per-metabolite info file (see here)
- "New name" : the name for the experiment in the main list (defaults to the table file name)

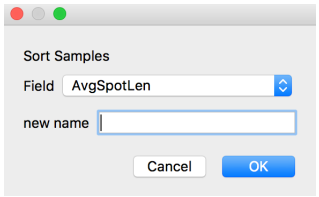
Processing data

Sample tab

Buttons from this tab affect the samples of the experiment. Button actions are performed on the selected experiment from the main list. Each action generates a new experiment in the main list.

Sort

Sort the samples according to the selected field.
Sorting is conservative, meaning samples with same value in the field retain the previous order. In order to sort by two fields (i.e. "Disease" and "Day" within each disease), sort first by the second field (i.e. "Day") and then by the first (i.e. "Disease").



Mandatory fields:

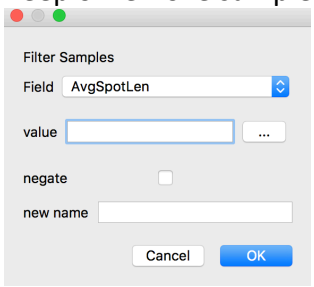
- "Field": select the sample metadata field to sort by

Optional fields:

- "New name" : the name for the experiment in the main list (defaults to the table file name)

Filter

Keep or remove samples with specified mapping file field values



Mandatory fields:

- "Field": select the sample metadata field to sort by
- "Value": the values to filter for the field.

Optional fields:

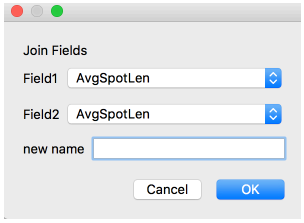
- "neagte": if checked, remove samples with the selected values, otherwise keep samples with selected values
- "New name" : the name for the experiment in the main list (defaults to the table file name)

Cluster

Cluster the samples by putting similar samples next to each other

Join Fields

Create a new metadata field by joining the values of two fields.



Mandatory fields:

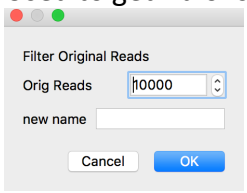
- "Field1", “Field2”: The two fields to join (new field will be field1-field2).

Optional fields:

- "New name" : the name for the experiment in the main list.

Filter by original reads

Throw away samples with < threshold original reads (before normalization).
Used to get rid of samples with a small number of reads



Mandatory fields:

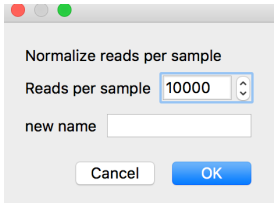
- "Orig. reads”: the minimal number of reads in the sample in order to keep

Optional fields:

- "New name" : the name for the experiment in the main list.

Normalize

TSS (total sum scaling) normalization of the reads per sample.
When loading a microbiome table, samples are automatically scaled to 10000 reads/sample.
NOTE: This is not rarefaction. So features can have fractional reads.



Mandatory fields:

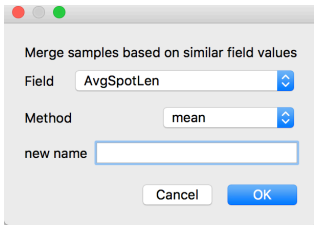
- "Reads per sample”: The sum of reads per sample to normalize to

Optional fields:

- "New name" : the name for the experiment in the main list.

Merge

merge samples having the same value in the selected field.
Samples can be merged using mean, sum or randomly choosing one of the samples with the value.



Mandatory fields:

- "Field": The field to merge samples sharing the same value
- "Method": how to merge the samples with same value. Options are:
    - "mean": new sample contains the mean of each of the features from all the samples with the value
    - "random": new sample contains a randomly chosen sample with the value
    - "sum": new sample contains the sum of each of the features from all the samples with the value

Optional fields:

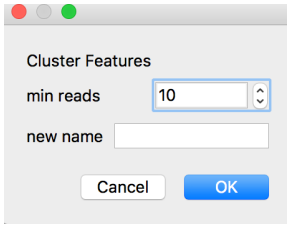
- "New name" : the name for the experiment in the main list.

Feature tab

Buttons from this tab affect the features of the experiment. Button actions are performed on the selected experiment from the main list. Each action generates a new experiment in the main list.

Cluster

Cluster the features by putting similar behaving features next to each other.
Each feature is normalized to mean=0, std=1 and then Euclidian distance is used for clusterin, so similar behaving features over the samples will be close to each other, without dependence on the absolute level of the features.
Can also remove low abundance features (to speed up the clustering).



Mandatory fields:

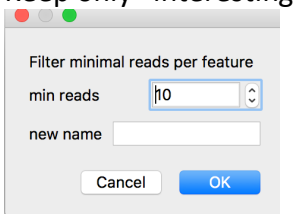
- "min reads": Keep only features with total reads over all samples >= min reads.

Optional fields:

- "New name" : the name for the experiment in the main list.

Filter min reads

Keep only "interesting features" that have enough total reads (over all samples)



Mandatory fields:

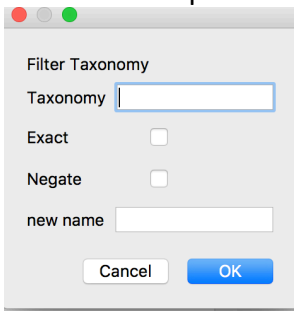
- "min reads": Keep only features with total reads over all samples >= min reads.

Optional fields:

- "New name" : the name for the experiment in the main list.

Filter taxonomy

Keep (or remove) features matching a given taxonomy string.
NOTE: This requires the biom table to contain taxonomy information.



Mandatory fields:

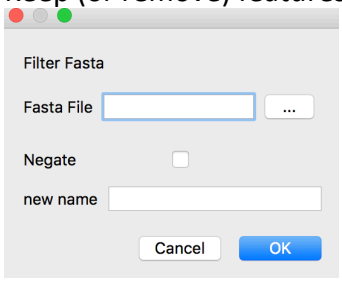
- "Taxonomy": A partial or complete taxonomy string. Needs to match the taxonomy embedded in the biom table.

Optional fields:

- "Exact": Check to filter only taxonomy strings fully matching the string. Uncheck to allow partial matches (i.e. "roteo" will match "p__Proteobacteria")
- "Negate": Check to remove matching features, uncheck to keep matching features.
- "New name" : the name for the experiment in the main list.

Filter fasta

Keep (or remove) features appearing in a fasta file



Mandatory fields:

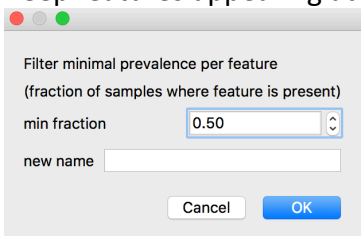
- "Fasta file": Name of the fasta file containing the feature IDs (usually sequences or qiime2 hashes).

Optional fields:

- "Negate": Check to remove matching features, uncheck to keep matching features.
- "New name" : the name for the experiment in the main list.

Filter prevalence

Keep features appearing at least in a given fraction of the samples (common features)



Mandatory fields:

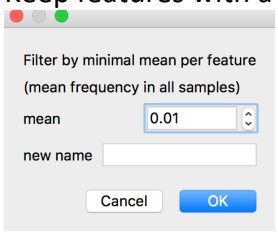
- "min fraction": the minimal fraction of the samples where the feature is present.

Optional fields:

- "New name" : the name for the experiment in the main list.

Filter mean

Keep features with a large enough mean frequency (over all samples)



Mandatory fields:

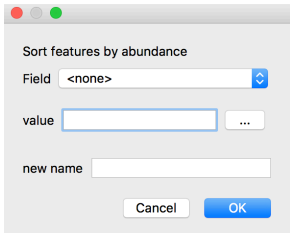
- "mean": the minimal mean frequency (over all samples) for features to be kept.

Optional fields:

- "New name" : the name for the experiment in the main list.

Sort abundance

Order the features based on their mean frequency over (an optional subset of) samples.

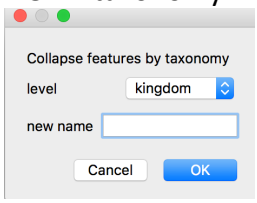


Optional fields:

- "Field": <none> to sort based on frequency over all samples, otherwise sort based on frequency only based on samples matching "value" in "Field".
- "value": The value to use for the sample subset.
- "New name" : the name for the experiment in the main list.

Collapse taxonomy

Merge features based on their taxonomy (summing the frequencies of all features with the same taxonomy).
NOTE: taxonomy information must be embedded in the biom table.



Mandatory fields:

- "level": The taxonomic level to merge by (i.e. Phyla, Genus, etc.)

Optional fields:

- "New name" : the name for the experiment in the main list.