FUNGuild Manual: V1.0

Introduction

FUNGuild is a two-component system that includes an online community annotated database and a python script that assigns functional guilds to operational taxonomic unit (OTU) obtained from next generation sequencing. The script uses the taxonomic string in the user's OTU table, and search against the online FUNGuild database containing taxonomic key words and functional metadata. The output is the original OTU table with functional metadata appended. Users also have options to output only assigned or unassigned OTUs.

General usage

`python FUNGuild_1.0.py [-h] -otu <path to OTU file> [-m] [-u]`

`-h`: Open the in-script help document

`-otu`: Specify the path to the OTU file. This is the only required parameter for this script. It is required that the OTU table contains a column named "taxonomy". This column should contain taxonomic strings from a reference database (we highly recommend using UNITE (https://unite.ut.ee/)). The script will try to detect the delimiter in the OTU table, but tab or comma delimited file is preferred. The output file will have a suffix of '_function.txt'. For example, if the input OTU table is 'user_otu_table.txt', then the new OTU table contains functional information will be 'user_otu_table.function.txt'. The output file will contain the exact OTU table as the input file, only with the metadata column appended.

`-m`: Ask the script to output another file that contains only OTUs that has been assigned a function. The output file will have suffix '_assigned.txt'.

`-u`: Ask the script to output another file that contains all OTUs that failed to be assigned a function. The output file will have suffix '_unassigned.txt'.

Notes

- Users will need internet access to reach the database.
- All metadata assigned by this script will appear on the last several column of the file.
- The OTUs in the output files will be sorted by the total number OTUs per row.
- The script will overwrite any existing file in its working folder.

Interpretation and use of output results

The result will be the original OTU table with appended metadata columns appended at the end. It is up to the user to interpret and decide which portion of the data should be used for downstream analyses.

Important items to consider:

1. OTU genus identification – The FUNGuild database is based on the *genus* level of classification. Therefore, users should only accept guild matches that are at or above the level of genus. Each genus will have a different OTU cut-off at the genus level, anywhere from a conservative 93% to 75% in ITS. These wildly different numbers are due to both genetic variability (rate of evolution)

and taxonomic arbitrarity in classification at the genus level. Therefore, we suggest that the user pick a genus cut-off value that is most appropriate for their study system.

2. Barcoding gene – Linked to point #1, a 93% genus cut-off in ITS means a different thing than in the LSU gene. Users should be aware of this and make the appropriate genus level cut-off for their gene of interest.

3. Confidence – The confidence value is only a guide. Users should decide whether the confidence level for each OTU is appropriate for their guild in association with a certain environment. We suggest that users generally new to studying fungi only accept guild assignments that are highly probable, so as not to over-interpret their data ecologically.

Definitions in FUNGuild metadata columns

- **Trophic Mode**: This column is broken down into three categories: Pathotroph (gaining nutrients by harming host cells), Symbiotroph (gaining nutrients by exchanging some benefit with host cells), and Saprotroph (gaining nutrients by breaking down dead host cells)
- **Guild**: This column is probably the most useful breakdown of the trophic modes and requires the most analysis to determine if the OTU belonging to this guild should be kept for downstream analyses.
- **Confidence**: This column weighs the likelihood in which an OTU belongs to a guild - Possible, Probable, and Highly Probable. For more reliability, we recommend using OTUs that belong to guilds that are ranked as Probable and Highly Probable. Guilds with Possible probability should be closely examine and are not appropriate to use in many cases.
- **Growth Morphology**: Growth forms such as yeast, facultative yeasts, and thallus are noted.
- **Trait**: Other traits that might be relevant to the genus or species in question.
- **Notes**: Any special notes that the contributor feels is necessary about the entry.
- **Citation/Source**: Each entry contains at least one reference or source where the information originated. They typically are from primary research papers.

Database annotation

The database is open for annotation. For instructions, please refer to the README file on the GitHub repository: https://github.com/UMNFuN/FUNGuild/blob/master/README.md

Citation

Cite FungGuild as:

Nguyen NH, Song Z, Bates ST, Branco S, Leho Tedersoo, Menke J, Schilling J, Kennedy P. FUNGuild: an open-annotation database for parsing high-throughput fungal molecular data by ecological guild. Submitted to Fungal Ecology.