TaXon Table Tools (v 1.0.9 uw) user manual

Table of content

1. Introduction

2. Installation

2.1 Requirements

- The GUI in some cases only works on an HD screen (1920 x 1080 pixel)
- python3.6 or python3.7 (Tutorial is shown for 3.7, but works as well for 3.6)
- python3x-tk (needs to be installed separately on Ubuntu)
- Python depencencies:
  pySimpleGUI, pandas, numpy, matplotlib, matplotlib-venn, xlrd, openpyxl, xlsxwriter
- Optional: Krona tools (https://github.com/marbl/Krona/wiki)
  Krona allows hierarchical data to be explored with zooming, multi-layered pie charts. These interactive charts can be automatically generated with TaXonTableTools. This function requires Krona tools to be installed (it's currently not officially supported on Windows).

Terminal usage

- Ubuntu/Mac: Terminal
- Windows: Power shell terminal

$ = command to type
>> = expected output of the command

2.2 Ubuntu

Open a new terminal

Install python3.7 on your machine (https://www.python.org)
Verify the installation with:

$ python3.7 --version
>> Python 3.7.5

Install pip for your python version (https://pypi.org/project/pip/)
Verify the installation with:

$ python3.7 -m pip --version
*>> pip 20.0.2 from */python3.6/site-packages/pip (python 3.7)*

Change directory to the TaxonTableTools folder

$ cd /your_directory/TaxonTableTools/

Install the required dependencies with:

$ python3.7 -m pip install -r basics.deps

Install the additional dependencies with:

$ python3.7 -m pip install -r addons.deps

Ubuntu sometimes requires an additional installation of tkinter:

$ sudo apt-get install python3.7-tk

Start TaXonTableTools by calling it from the command line:

$ python3.7 TaXon_Table_Tools.py

2.3 Mac

Open a new terminal

Install python3.7 on your machine (https://www.python.org)
Verify the installation with:

$ python3.7 --version

>> Python 3.7.5

Install pip for your python version (https://pypi.org/project/pip/)
Verify the installation with:

$ python3.7 -m pip --version

*>> pip 20.0.2 from */python3.6/site-packages/pip (python 3.7)*

Change directory to the TaxonTableTools folder

$ cd /your_directory/TaxonTableTools/

Install the required dependencies with:

$ python3.7 -m pip install -r basics.deps

Install the additional dependencies with:

$ python3.7 -m pip install -r addons.deps

Start TaXonTableTools by calling it from the command line:

$ python3.7 TaXon_Table_Tools.py

2.4 Windows

Open a new Windows power shell terminal

Install python3.7 on your machine (https://www.python.org)
Verify the installation with:

$ py --version

>> Python 3.7.5

Install pip for your python version (https://pypi.org/project/pip/)
Verify the installation with:

$ py -m pip --version

*>> pip 20.0.2 from */python3.6/site-packages/pip (python 3.7)*

Install the required dependencies with:

$ py -m pip install -r basics.deps

Install the additional dependencies with:

$ python3.7 -m pip install -r addons.deps

Change directory to the TaxonTableTools folder

$ cd ./your_path/TaxonTableTools/

Start TaXonTableTools by calling it from the command line:

$ py TaXon_Table_Tools.py

2.5 Krona tools

Install Krona tools (https://github.com/marbl/Krona/wiki).
Check your Krona tools installation by running:

$ ktImportText

>> KronaTools 2.7.1 - ktImportText

3. Getting started

3.1 Projects

Start TaXonTableTools by calling it from the command line:
Ubuntu/Mac:

$ python3.7 TaXon_Table_Tools.py

Windows:

$py TaXon_Table_Tools.py

The starting screen appears and will ask for creating a new project or load an existing project.

> Type in your a project name.
or
> Load an existing project folder.
or
> Leave blank to create a “Default_project” folder.

A new project folder has been created! You can find your results in the respective directories under “Projects/your_project/”.

3.2 Sample IDs

A specific sample ID format is required for a couple of tools to work properly. Five different categories can be read from this format. Thus, a number of 5 underscores (which delimit the information) is mandatory. The required looks as follows.

Project_Site_Sample_Date_Replicate
(e.g. Dessau_MuldeOH_5A_180419_a)

Project = name of the project
Site= name of the site
Sample= sample ID
Date = date when samples were taken
Replicate = _a and _b (PCR replicates) or _A and _B (extraction replicates)
>> Note: currently only 2 sample replicates can be merged
>> see 4.2.1 for more details

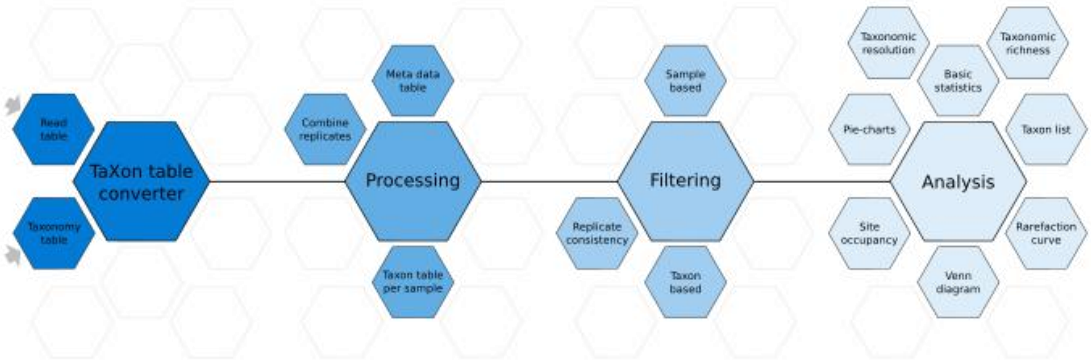
If the samples are not named accordingly, replicates cannot be merged. All other functions will still work. Only the creation of meta data tables requires user input when the format demands are not met.

3.3 Plots

Occasionally the labels will not be displayed correctly or colors might no be pleasing to the user. Thus, all plots created with TTT will be saved in .pdf format and are

compatible for any downstream adjustments with most common image manipulation software. Since the plots are vector based graphics, colors can easily be adjusted, labels renamed and vectors be moved. TTT was written to create plots in a general manner, which can subsequently be individualized by the user.

4. Tools



4.1 Data conversion

4.1.1 TaXon table converter

TaXonTableTools requires two files as input format: a taxonomy table and a read table, which are then merged to a single file. The demanded file format for both tables is .xlsx (Microsoft Excel 2007). All downstream tables will also be saved in .xlsx file, which can easily be opened with excel or any equivalent program.

Read table format:

The read table can e.g. be generated by the JAMP pipeline (https://github.com/VascoElbrecht/JAMP). The format requires the first column to be the OTU names ("OTUs"), followed by the included samples (see 3.2 for more information) and the sequence as last column. Notice to remove the last row "below 0.01p", the "sort" column and rename "IDs" to "OTUs" and "sequ" to "Sequences", when using the JAMP pipeline.

OTUs	Sample_1	Sample_2	(...)	Sample_n	Sequences
OTU_1	88817	56644	...	67544	ATGCTAA...
OTU_2	6384	18919	...	21877	ATGGTAT...
OTU_3	655	0	...	0	ATGCTTT...
(...)	...	...	...	...	...

OTU_n	0	73	...	87	ATGCTAG...

Taxonomy table format:

The taxonomy table can e.g. be generated by the BOLDigger tool (Buchner et al., unpublished). The format requires the first column to be the OTU IDs ("OTUs"), followed by taxonomy ("Phylum", "Class", "Order", "Family", "Genus" and "Species"), the similarity to the reference sequence ("Similarity") and the status of the reference sequence ("Status"). The excel sheet has to be named "JAMP results" (this will change in the future). Species names are recommended to be written as two words, including genus and epithet to avoid epithet duplicates. The status is automatically derived by the BOLDigger tool, but can manually be added as e.g. "public". Same accounts for the Similarity.

OTUs	Phylum	Class	Order	Family	Genus	Species	Similarity	Status
OTU_1	...	...	...	...	...	Genus epithet	100	public
OTU_2	...	...	...	...	...	...	100	public
OTU_3	...	...	...	...	...	...	99	public
(...)	...	...	...	...	...	...	...	...
OTU_n	...	...	...	...	...	...	98	public

JAMP results

Merging the read table and the taxonomy table creates a new file in the so called "TaXon table" format, which is also an .xlsx file. This table format consists of all information of both the taxonomy table and the read table and will be used as standard input format for all subsequent steps. The newly created TaXon tables are found in the "Projects/your_project/TaXon_tables" directory.

4.2 Table processing

4.2.1 Combine replicates

4.2.2 Taxon table per sample

4.2.3 Meta data table

4.3 Filtering

4.3.1 Taxon-based filtering

4.3.2 Sample-based filtering

4.3.3 Replicate consistency filter

4.4 Analyses

4.4.1 Basic statistics

Executing the basic statistics tool will create an overview of the basic read and taxonomy table statistics that can be gathered from the TaXon table. The overall number of samples and OTUs is extracted along the number of taxa per taxonomic level (from Phylum to Species level). All available database states of the reference sequence are counted. The minimum, average and maximum length (in bp) of the sequences, as well as the average and total number of reads for each sample is calculated. The results are printed on the screen and written to a xlsx-file in the “Projects/your_project/Basic_stats” directory.

4.4.2 Taxonomic resolution

Taxonomy resolution highly depends on the used reference database. OTUs with low similarity towards the reference database are recommended to be reported at a higher taxonomic level to prevent false positive results. Hence, this tool first plots the taxonomic resolution per taxonomic level (Phylum to Species), by counting the number of OTUs that are reported for the respective taxonomic level. In a second plot the total number OTUs per taxonomic level is given. The results are written to a pdf-file in the “Projects/your_project/Taxonomic_resolution_plots” directory.

4.4.3 Taxonomic richness

This tool will plot the number of taxa per taxonomic level (Phylum to Species) and usually shows a steep increase towards the lower taxonomy. The results are written to a pdf-file in the “Projects/your_project/Taxonomic_richness_plots” directory.

4.4.4 OTU abundance pie-charts

For each taxonomic level (Phylum to Species) a pie-chart depicting the relative OTU abundances will be created. These plots show the relative number of reported OTUs per taxon and not the read abundances (refer to 4.4.7 for read-based plots). Two version of each plot are created, one including and one excluding the “nan” status, since the proportion of unassigned taxonomy usually increases with higher taxonomic level and thus can render the plots unreadable. The results are written to a pdf-file in the “Projects/your_project/Pie_charts” directory.

4.4.5 Venn diagram

Venn diagrams are a common way of comparing different sets. This tool uses venn diagrams to depict the taxonomy overlap (Phylum to Species, respectively) of all samples in two TaXon tables. Here, each table is handled as one set. The easiest way to compare specific sample sets is to create new TaXon tables using the “sample-based filter” tool (see 4.3.2) or the “taxon table per sample” tool (see 4.2.2). When executed, this tool scales the overlap of shared taxa and the non-shared taxa intro three distinctive areas. The respective number of taxa is given for each area. Taxon names are written to a seperate xlsx-file. The final venn diagrams can be found in pdf-format along the xlsx-file in the “Projects/your_project/Venn_diagrams” directory.

4.4.6 Rarefaction curve

4.4.7 Read proportions

4.4.8 Site occupancy

4.5 Taxon list

4.5.1 Additional information

4.5.2 Taxon list

4.5 Troubeshooting