eDNAnalyzer MANUAL

eDNAnalyzer is a user-friendly and open-access computational tool developed to process and filter taxonomic assignment data from metabarcoding studies, particularly derived from environmental DNA (eDNA) and invertebrate-derived DNA (iDNA) approaches.

How to cite eDNAnalyzer? Olimpio, L.W.G.F.; Gestich, C.C.; Saranholi, B.H.; Galetti Jr, P.M.; Freitas, P.D. 2025. eDNAnalyzer: a fast and user-friendly computational tool for processing massive taxonomic assignment data derived from eDNA and iDNA metabarcoding (doi:).

How to access the eDNAnalyzer? You can find the eDNAnalyzer tool by accessing the GitHub repository at https://github.com/Leo-9821/eDNAnalyzer. The executable file (.exe) is available for Windows®, while the Python source code (.py files) can be used to run the program on Linux® or macOS®, in this case download the “main.py” and “metabar.py” files and “img” folder to the same directory in your computer, and make sure to install pandas, Pillow and openpyxl libraries.

How does eDNAnalyzer work? The software provides two main functions: “Threshold application” and “Results consolidation”, each requiring a specific input file. Examples of input and output files are available in the repository https://github.com/Leo-9821/eDNAnalyzer/tree/master/example_files.

An overview of the general operation of the program, showing its interfaces, is provided at the end of this manual, along with more detailed initial instructions for performing the two main functions of eDNAnalyzer: “threshold application” and “results consolidation”.

Choosing the option **Threshold Application**

This option processes data from prior taxonomic assignments by calculating the total number of reads per OTU per sequencing sample, filtering out OTUs according to an adopted threshold value (default ≥ 0.05), and generating outputs following selected filters.

Process Overview:

1. Calculate the total number of reads per OTU per sequencing sample, based on previous taxonomic assignments provided (input file).
2. Filter OTUs according to an adopted cutoff value (default: ≥ 0.05%).
3. Generate files (output files) containing the total number of reads per OTU, the OTUs that did not pass by the threshold value and the OTUs that passed.

Input File: a file in .xlsx or .csv format, (e.g., example_input.xlsx available at https://github.com/Leo-9821/eDNAnalyzer/blob/master/example_files/eng/xlsx/1_threshold_application/input/input_example.xlsx), containing the taxonomic assignment data, must be provided with the following required columns:

- sequencing_sample (identification of sequencing samples)
- area_sampler (indicating the collection area and DNA sampler - e.g., water sample, soil sample, invertebrate used as sampler, etc. - and ensure that both areas and samplers are unambiguously identified to avoid errors during separation. Use an underscore to separate the area from the sampler: area_sampler)
- point (indicating the collection point)
- n_reads (number of reads per OTU)
- taxon (taxon determined from taxonomic assignment)

Extra columns containing additional information can also be included in the table, as shown in Table 1.

Table 1: Summary table of an input file showing the general information to perform the threshold application step.

sequencing_sample	barcode	tag	area_sampler	point	aliquot	otu	n_reads	%_id	taxon
P08	12SrRNA	TA	P1_MQ	1	1	1	233	100	*Nycticorax nycticorax*
P08	12SrRNA	TB	P2_MC	2	1	1	50	99.26	*Bos taurus*
P08	12SrRNA	TB	P2_MC	2	1	2	41	99.25	*Canis lupus*
P08	12SrRNA	TB	P2_MC	2	1	4	3	100	*Sus scrofa*
P08	12SrRNA	TC	P2_MC	2	1	2	20	99.26	*Canis lupus familiaris*
P08	12SrRNA	TC	P2_MC	2	1	3	38	100	*Canis lupus familiaris*
P08	12SrRNA	TC	P2_MC	2	1	4	524	100	*Canis lupus familiaris*
P08	12SrRNA	TC	P2_MC	2	1	6	7	100	*Canis lupus*
P08	12SrRNA	TC	P2_MC	2	1	7	2	99.26	*Equus caballus*
P08	12SrRNA	TC	P2_MC	2	1	8	4	100	*Leopardus pardalis*
P08	12SrRNA	TC	P2_MC	2	1	9	2	100	*Puma concolor*
P08	12SrRNA	TF	P2_MC	2	1	1	6	100	*plant junction region*
P08	12SrRNA	TG	P2_MC	2	1	1	17	99.29	*Canis lupus*
P08	12SrRNA	TH	P3_MC	3	1	1	109	100	*Bos taurus*
P08	12SrRNA	TH	P3_MC	3	1	2	52	100	*Canis lupus familiaris*
P08	12SrRNA	TH	P3_MC	3	1	3	9	100	*Pecari tajacu*
P08	12SrRNA	TH	P3_MC	3	1	4	1859	100	*Rhea americana*
P08	12SrRNA	TH	P3_MC	3	1	5	7	100	*Didelphis albiventris*
P08	12SrRNA	TH	P3_MC	3	1	6	4	98.56	*Gallus gallus*
P08	12SrRNA	TH	P3_MC	3	1	7	6	97.81	*Gymnogyps californianus*
P08	12SrRNA	TH	P3_MC	3	1	9	4	100	*Streptopelia decaocto*
P08	12SrRNA	TI	P3_MQ	3	1	1	14	100	*Bos taurus*
P08	12SrRNA	TI	P3_MQ	3	1	2	127	100	*Equus caballus*
P08	12SrRNA	TI	P3_MQ	3	1	3	28	100	*Sus scrofa*
P08	12SrRNA	TI	P3_MQ	3	1	4	27	100	*Tamandua tetradactyla*
P08	12SrRNA	TI	P3_MQ	3	1	5	8	100	*Hydrochoerus hydrochaeris*
P08	12SrRNA	TI	P3_MQ	3	1	6	25	100	*Bubulcus ibis*
P08	12SrRNA	TI	P3_MQ	3	1	7	1013	100	*Gallus gallus*
P08	12SrRNA	TI	P3_MQ	3	1	8	4	100	*Gallus gallus*
P08	12SrRNA	TI	P3_MQ	3	1	9	2	100	*Nycticorax nycticorax*
P08	12SrRNA	TK	P3_MC	3	1	1	33	100	*Zaedyus pichiy*
P08	12SrRNA	TK	P3_MC	3	1	2	26	97.04	*Euphractus sexcinctus*
P08	12SrRNA	TL	P3_MC	3	1	1	1111	100	*Canis aureus*
P08	12SrRNA	TL	P3_MC	3	1	2	481	100	*Zaedyus pichiy*

Output Files: All tables can be saved in .xlsx and .csv format. See an example of a file provided in this step in Table 2.

Table 2: Summary table of an output file showing the data processed after running the threshold application step.

sequencing_sample	barcode	tag	area_sampler	point	aliquot	otu	n_reads	%_id	taxon	final_otu_curated
P08	12SrRNA	TA	P1_MQ	1	1	1	233	100	*Nycticorax nycticorax*	
P08	12SrRNA	TB	P2_MC	2	1	1	50	99.26	*Bos taurus*	
P08	12SrRNA	TB	P2_MC	2	1	2	41	99.25	*Canis lupus*	
P08	12SrRNA	TC	P2_MC	2	1	2	20	99.26	*Canis lupus familiaris*	
P08	12SrRNA	TC	P2_MC	2	1	3	38	100	*Canis lupus familiaris*	
P08	12SrRNA	TC	P2_MC	2	1	4	524	100	*Canis lupus familiaris*	
P08	12SrRNA	TG	P2_MC	2	1	1	17	99.29	*Canis lupus*	
P08	12SrRNA	TH	P3_MC	3	1	1	109	100	*Bos taurus*	
P08	12SrRNA	TH	P3_MC	3	1	2	52	100	*Canis lupus familiaris*	
P08	12SrRNA	TH	P3_MC	3	1	4	1859	100	*Rhea americana*	
P08	12SrRNA	TI	P3_MQ	3	1	1	14	100	*Bos taurus*	
P08	12SrRNA	TI	P3_MQ	3	1	2	127	100	*Equus caballus*	
P08	12SrRNA	TI	P3_MQ	3	1	3	28	100	*Sus scrofa*	
P08	12SrRNA	TI	P3_MQ	3	1	4	27	100	*Tamandua tetradactyla*	
P08	12SrRNA	TI	P3_MQ	3	1	6	25	100	*Bubulcus ibis*	
P08	12SrRNA	TI	P3_MQ	3	1	7	1013	100	*Gallus gallus*	
P08	12SrRNA	TK	P3_MC	3	1	1	33	100	*Zaedyus pichiy*	
P08	12SrRNA	TK	P3_MC	3	1	2	26	97.04	*Euphractus sexcinctus*	
P08	12SrRNA	TL	P3_MC	3	1	1	1111	100	*Canis aureus*	
P08	12SrRNA	TL	P3_MC	3	1	2	481	100	*Zaedyus pichiy*	

Choosing the option **Results Consolidation**

This process will provide tables with the results, lists of species with their number of detections and reads. The tables can be generated by separating the lists by sampler, area, or both.

Process Overview:

1. Edit the table containing the OTUs of interest by completing the final_otu_curated column with the selected taxonomic assignment after manual curation. This step aims to review and refine taxonomic identification to correct possible inconsistencies in assignment; for this, data on species distribution and field observations must be incorporated for better accuracy.
2. Input the curated table (input file) into eDNAnalyzer, and the program will process the data and then return the information of interest according to the filters selected by the user (e.g.; sampler, area, or sampler and area).
3. Generate files with consolidated results (output files) in .xlsx and .csv formats. When necessary, .csv files in a .zip file will be provided.

Input File: The input file (Table 3) for the “results consolidation” step is a table (in .xlsx or .csv format) containing the final OTU data from the previous threshold application step and the curated OTU list.

Table 3. Summary table of an input file showing the general information to perform the consolidation step.

sequencing_sample	barcode	tag	area_sampler	point	aliquot	otu	n_reads	%_id	taxon	final_otu_curated
P08	12SrRNA	TA	P1_MQ	1	1	1	233	100	*Nycticorax nycticorax*	*Nycticorax nycticorax*
P08	12SrRNA	TB	P2_MC	2	1	1	50	99.26	*Bos taurus*	*Bos taurus*
P08	12SrRNA	TB	P2_MC	2	1	2	41	99.25	*Canis lupus*	*Canis lupus familiaris*
P08	12SrRNA	TC	P2_MC	2	1	2	20	99.26	*Canis lupus familiaris*	*Canis lupus familiaris*
P08	12SrRNA	TC	P2_MC	2	1	3	38	100	*Canis lupus familiaris*	*Canis lupus familiaris*
P08	12SrRNA	TC	P2_MC	2	1	4	524	100	*Canis lupus familiaris*	*Canis lupus familiaris*
P08	12SrRNA	TG	P2_MC	2	1	1	17	99.29	*Canis lupus*	*Canis lupus familiaris*
P08	12SrRNA	TH	P3_MC	3	1	1	109	100	*Bos taurus*	*Bos taurus*
P08	12SrRNA	TH	P3_MC	3	1	2	52	100	*Canis lupus familiaris*	*Canis lupus familiaris*
P08	12SrRNA	TH	P3_MC	3	1	4	1859	100	*Rhea americana*	*Rhea americana*
P08	12SrRNA	TI	P3_MQ	3	1	1	14	100	*Bos taurus*	*Bos taurus*
P08	12SrRNA	TI	P3_MQ	3	1	2	127	100	*Equus caballus*	*Equus caballus*
P08	12SrRNA	TI	P3_MQ	3	1	3	28	100	*Sus scrofa*	*Sus scrofa*
P08	12SrRNA	TI	P3_MQ	3	1	4	27	100	*Tamandua tetradactyla*	*Tamandua tetradactyla*
P08	12SrRNA	TI	P3_MQ	3	1	6	25	100	*Bubulcus ibis*	*Bubulcus ibis*
P08	12SrRNA	TI	P3_MQ	3	1	7	1013	100	*Gallus gallus*	*Gallus gallus*
P08	12SrRNA	TK	P3_MC	3	1	1	33	100	*Zaedyus pichiy*	*Euphractus sexcinctus*
P08	12SrRNA	TK	P3_MC	3	1	2	26	97.04	*Euphractus sexcinctus*	*Euphractus sexcinctus*
P08	12SrRNA	TL	P3_MC	3	1	1	1111	100	*Canis aureus*	*Canis lupus familiaris*
P08	12SrRNA	TL	P3_MC	3	1	2	481	100	*Zaedyus pichiy*	*Euphractus sexcinctus*

Output File: all tables can be saved in .xlsx and .csv formats. See an example of a file provided in this step in Table 4.

Table 4. Summary table of an output file showing the consolidated results after running the consolidation step. Note that the taxon list is presented according to the filter selected in different guides of the datasheet.

	A	B	C	D
1		**Taxon**	**Reads**	**Detections in P6**
2	**0**	*Tapirus terrestris*	4195	1
3	**1**	*Canis lupus familiaris*	1269	1
4	**2**	*Pitheciidae*	736	1
5	**3**	*Coendou insidiosus*	451	1
6	**4**	*Callicebus nigrifrons*	313	1
7	**5**	*Bos taurus*	212	1
8	**6**	*Lycalopex vettulus*	202	1
9	**7**	*Rattus rattus*	177	1
10	**8**	*Puma concolor*	164	1
11	**9**	*Callithrix penicillata*	118	1
12	**10**	*Columba livia*	113	1
13	**11**	*Gallus gallus*	98	1
14	**12**	*Sus scrofa*	94	1
15	**13**	*Cingulata*	88	1
16	**14**	*Pecari tajacu*	65	1
17	**15**	*Thraupis sayaca*	65	1
18	**16**	*Callithrix sp.*	63	1
19	**17**	*Euphractus sexcinctus*	48	1
20	**18**	*Cairina moschata*	39	1
21	**19**	*Hydrochaeris hydrochaeris*	38	1
22	**20**	*Canidae*	36	1
23	**21**	*Dasypus novemcinctus*	32	1
24	**22**	*Chrysocyon brachyurus*	19	1
25	**23**	*Nymphicus hollandicus*	18	1

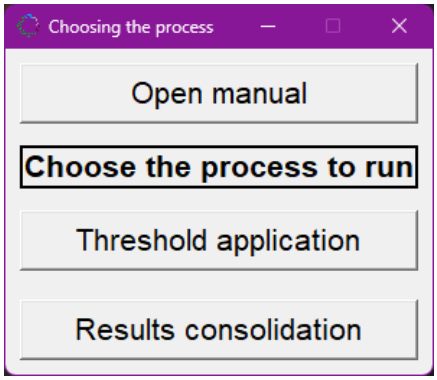
P6 | P4 | P7 | P3 | P8 | P2

General Operation of the eDNAnalyzer Program

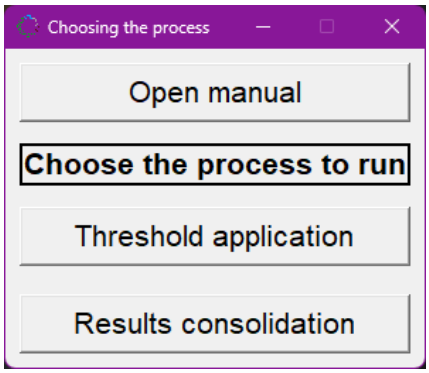
After accessing the program at https://github.com/Leo-9821/eDNAnalyzer, follow the steps by clicking on the options available.

Step 1: Select the language

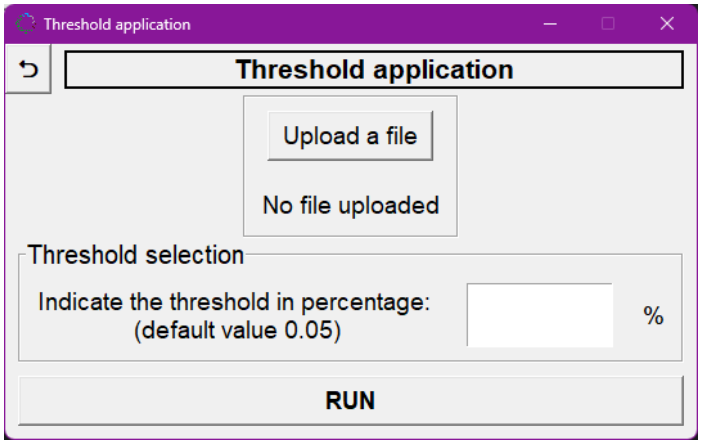
Step 2: Read the manual and then choose one of the two processing options



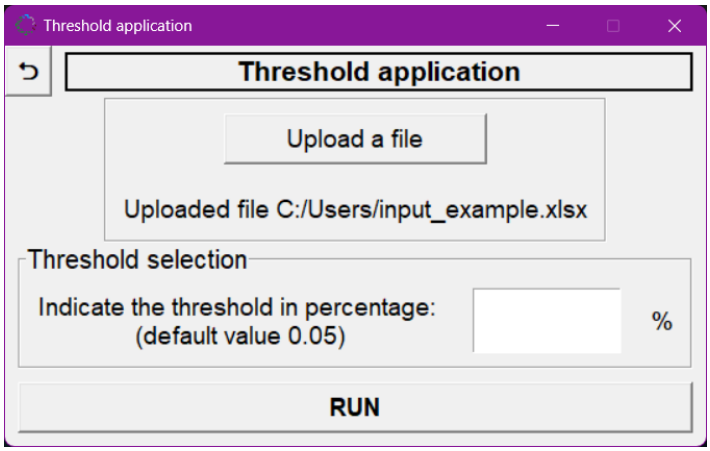
Step 3: Choose the option threshold application



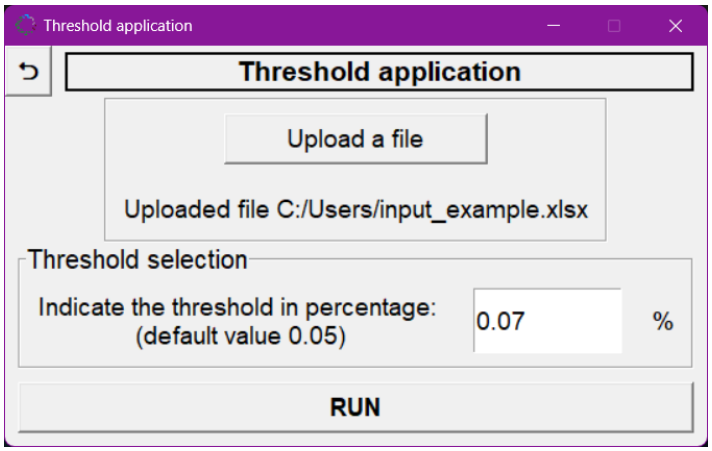
Step 4: Upload a file containing the taxonomic assignment data



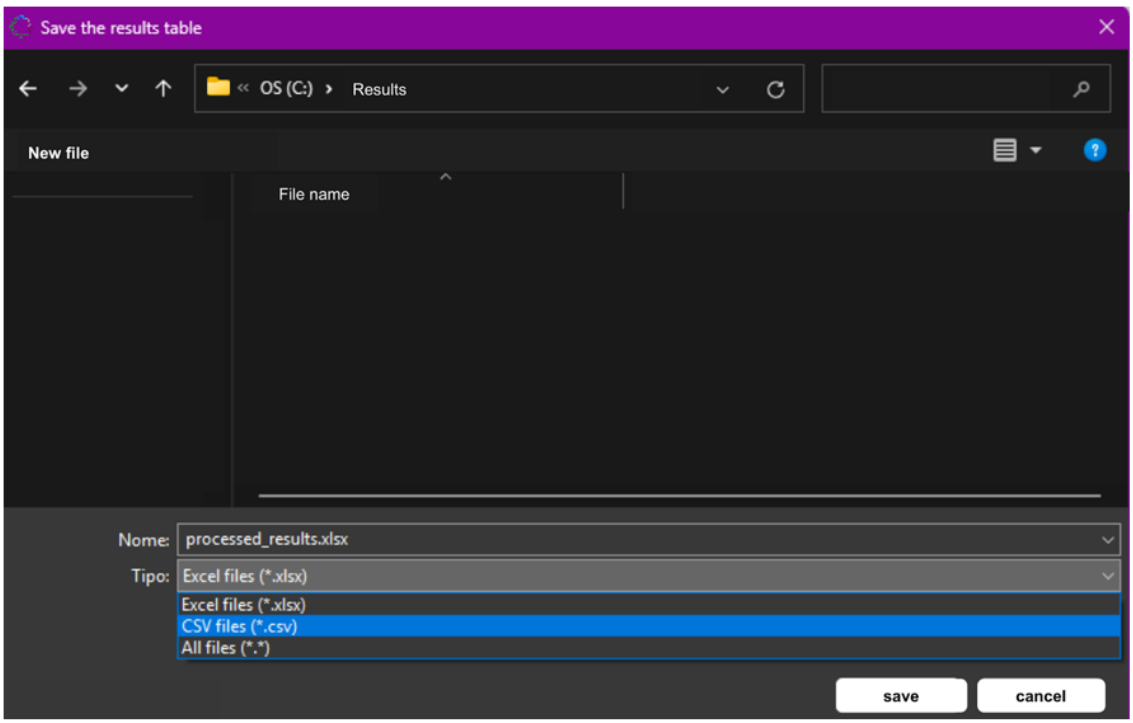
Step 5: Check the provided folder directory to ensure that the file was uploaded correctly.



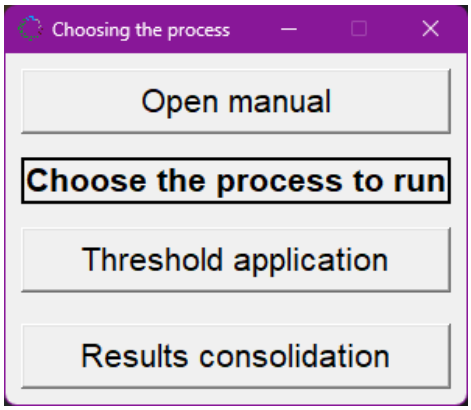
Step 6: Insert a threshold value in percentage or keep empty to use the default (0.05%) and run the Threshold application step.



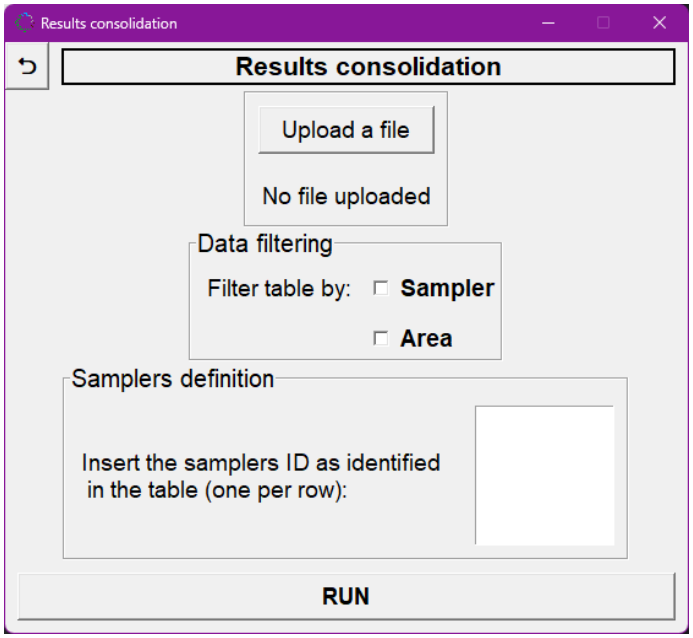
Step 7: After running the threshold application, three output files will be generated, each containing, respectively: (i) the calculated threshold value per sequencing sample, (ii) the OTUs with number of reads below the adopted threshold value, and (iii) the OTUs with number of reads equal to or greater than the adopted threshold. Save at least this last file in .xlsx or .csv formats, to use it for manual curation and then as an input file for data filtering in the results consolidation step.



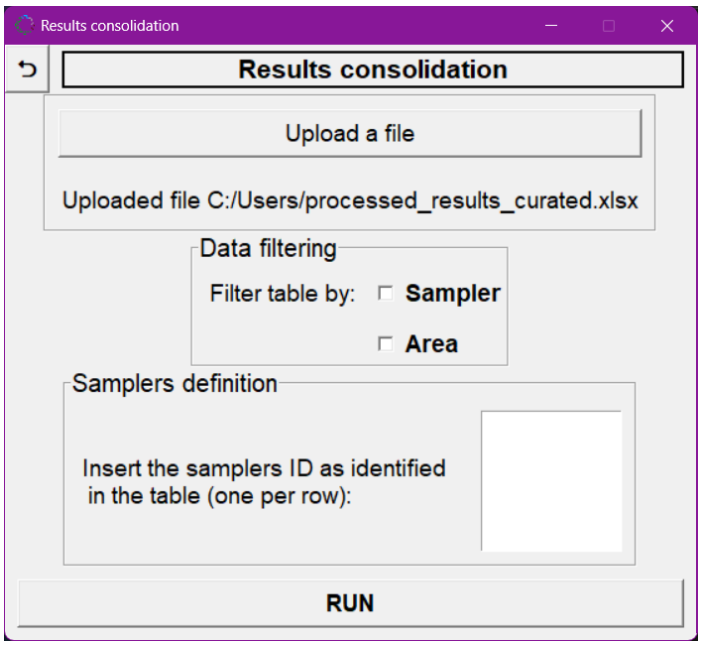
Step 8: Choose the option results consolidation.



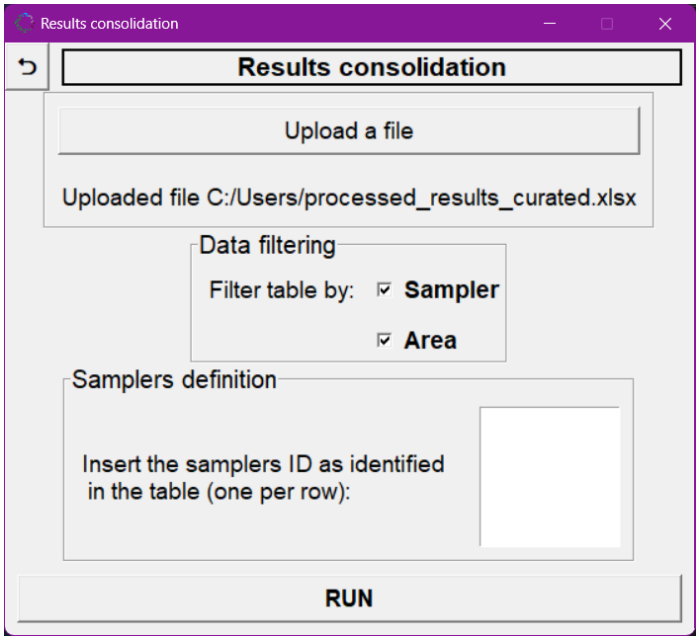
Step 9: Upload a file containing a curated table after running the threshold application step



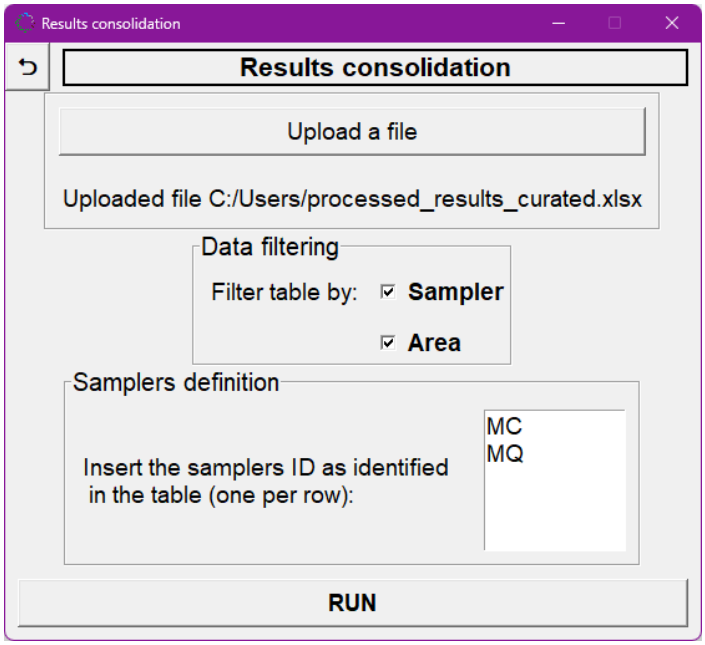
Step 11: Check the provided folder directory to ensure that the file was uploaded correctly.



Step 12: Select the options to filter the data by sampler and/or area



Step 12: If you selected the "sampler" option, type the sampler IDs, separating them from each other, by pressing the key “enter” from your keyboard. If you did not select the "Sampler" option in the previous step, leave this field blank.



Step 13: Save the tables generated with the data selected according to the chosen filters. To compress all the generated tables in .csv format into a single .zip file, select the ZIP option. Thus, you will save the consolidated results with the final list of taxa, and information about the number of reads that each taxon presented and the number of times it was detected.