



MANUAL eDNAAnalyzer

eDNAAnalyzer é uma ferramenta computacional de fácil uso e acesso aberto, desenvolvida para processar e filtrar dados de atribuição taxonômica provenientes de estudos de metabarcoding, especialmente aqueles derivados de DNA ambiental (eDNA) e DNA derivado de invertebrados (iDNA).

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

Como acessar o eDNAAnalyzer? Você pode encontrar a ferramenta eDNAAnalyzer acessando o repositório no GitHub em <https://github.com/Leo-9821/eDNAAnalyzer>. O arquivo executável (.exe) está disponível para Windows®, enquanto o código-fonte em Python (.py) pode ser usado para executar o programa em Linux® ou macOS®. Neste caso, baixe os arquivos "main.py" e "metabar.py" e a pasta "img" para o mesmo diretório em seu computador e certifique-se de instalar as bibliotecas pandas, Pillow e openpyxl.

Como o eDNAAnalyzer funciona? O software oferece duas funções principais: "Aplicação do *Threshold*" e "Consolidação de Resultados", cada uma exigindo um arquivo de entrada específico. Exemplos de arquivos de entrada e saída estão disponíveis no repositório https://github.com/Leo-9821/eDNAAnalyzer/tree/master/example_files.

Uma visão do [funcionamento geral](#) do programa, é fornecida ao final deste manual, juntamente com instruções detalhadas para realizar as duas funções principais: "[Aplicação do Threshold](#)" e "[Consolidação dos Resultados](#)".

Escolhendo a opção **Aplicação do Threshold**

Esta opção processa dados de atribuições taxonômicas prévias, calculando o número total de reads por OTU por amostra de sequenciamento, filtrando OTUs de acordo com um valor do *threshold* adotado (padrão $\geq 0,05\%$) e gerando saídas após a aplicação dos filtros selecionados.

Visão geral do processo:

1. Calcular o número total de reads por OTU por amostra de sequenciamento, com base nas atribuições taxonômicas fornecidas ([arquivo de entrada](#)).
2. Filtrar OTUs de acordo com um valor de corte adotado (padrão: $\geq 0,05\%$).
3. Gera arquivos ([arquivos de saída](#)) contendo o número total de reads por OTU, as OTUs que não passaram no limiar e as OTUs que passaram.

Arquivo de entrada: Um arquivo em formato .xlsx ou .csv (exemplo: input_exemplo.xlsx disponível em https://github.com/Leo-9821/eDNAalyzer/blob/master/example_files/pt-br/xlsx/1_aplicacao_threshold/input/input_exemplo.xlsx), contendo os dados de atribuição taxonômica, deve ser fornecido com as seguintes colunas obrigatórias:

- amostra_sequenciamento (identificação das amostras de sequenciamento)
- area_amostrador (indicando a área de coleta e o amostrador de DNA – por exemplo, amostra de água, solo, invertebrado utilizado como amostrador – e garanta que áreas e amostradores sejam identificados sem ambiguidade para evitar erros. Use um *underscore* para separar a área do coletor: area_amostrador)
- ponto (indicando o ponto de coleta)
- n_reads (número de *reads* por OTU)
- taxon (táxon determinado a partir da atribuição taxonômica)

Colunas adicionais com informações extras também podem ser incluídas, conforme mostrado na Tabela 1.

Tabela 1: Tabela resumo de um arquivo de entrada mostrando as informações gerais para a etapa de aplicação do *threshold*.

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

Arquivos de saída: Todas as tabelas podem ser salvas em formato .xlsx e .csv. Veja um exemplo de arquivo gerado nesta etapa na Tabela 2.

Tabela 2: Tabela resumo de um arquivo de saída mostrando os dados processados após a execução da etapa de aplicação do *threshold*.

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

Escolhendo a opção **Consolidação dos Resultados**

Este processo fornece tabelas com os resultados, listas de espécies com seus números de detecções e *reads*. As tabelas podem ser geradas separando as listas por amostrador, área ou ambos.

Visão geral do processo:

1. Edite a tabela contendo as OTUs de interesse, preenchendo a coluna *otu_final_curada* com a atribuição taxonômica selecionada após curadoria manual. Esta etapa visa revisar e refinar a identificação taxonômica para corrigir possíveis inconsistências, incorporando dados de distribuição de espécies e observações de campo para maior precisão.
2. Insira a tabela curada ([arquivo de entrada](#)) no eDNAAnalyzer. O programa processará os dados e retornará as informações de interesse de acordo com os filtros selecionados pelo usuário (por exemplo: amostrador, área ou ambos).
3. Gera arquivos com resultados consolidados ([arquivos de saída](#)) em formatos .xlsx e .csv. Quando necessário, arquivos .csv em arquivo .zip serão fornecidos.

Arquivo de entrada: A tabela de entrada (Tabela 3) para a etapa de consolidação de resultados é uma tabela (em formato .xlsx ou .csv) contendo os dados de OTU final da etapa anterior de aplicação de *threshold* e a lista de OTU curada.

Tabela 3. Tabela resumo de um arquivo de entrada mostrando as informações gerais para a etapa de consolidação..

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

Arquivos de saída: Todas as tabelas podem ser salvas em formatos .xlsx e .csv. Veja um exemplo de arquivo gerado nesta etapa na Tabela 4.

Tabela 4. Tabela resumo de um arquivo de saída mostrando os resultados consolidados após a execução da etapa de consolidação. Observe que a lista de táxons é apresentada de acordo com o filtro selecionado em diferentes guias da planilha.

	A	B	C	D
1		Taxon	Reads	Detecções em 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	Dasyopus novemcinctus	32	1
24	22	Chrysocyon brachyurus	19	1
25	23	Nymphicus hollandicus	18	1

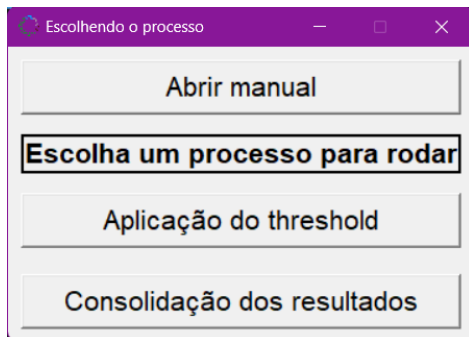
Funcionamento Geral do Programa eDNAlyzer

Após acessar o programa em <https://github.com/Leo-9821/eDNAlyzer>, siga as etapas clicando nas opções disponíveis.

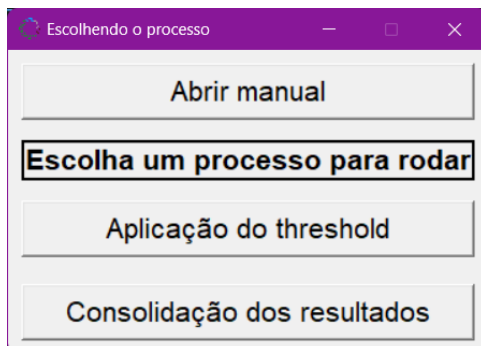
Etapa 1: Selecione o idioma



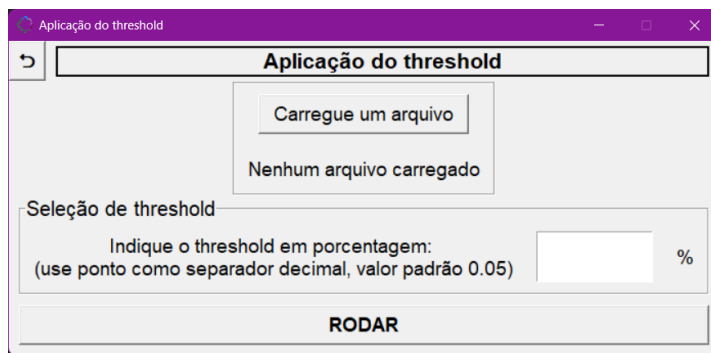
Etapa 2: Leia o manual e escolha uma das duas opções de processamento



Etapa 3: Escolha a opção "Aplicação do Threshold".

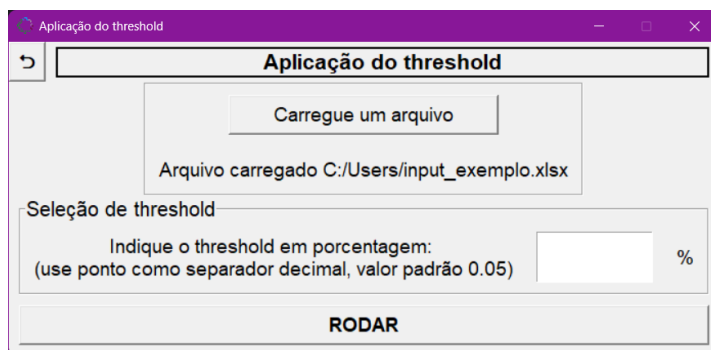


Etapa 4: Carregue um arquivo contendo os dados de atribuição taxonômica



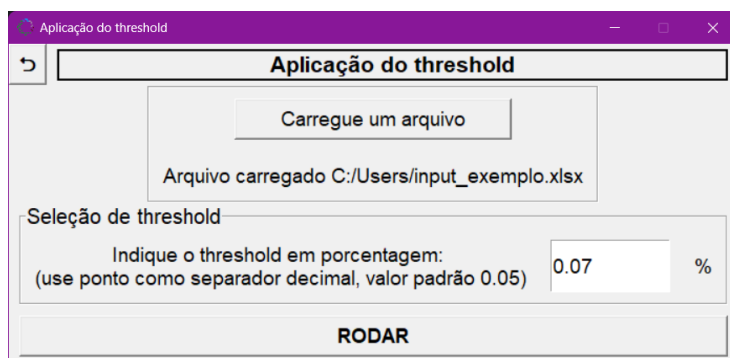
The screenshot shows a window titled "Aplicação do threshold". It has a tab labeled "Aplicação do threshold". Inside, there is a button labeled "Carregue um arquivo". Below it, the text "Nenhum arquivo carregado" is displayed. There is a section titled "Seleção de threshold" with a label "Indique o threshold em porcentagem: (use ponto como separador decimal, valor padrão 0.05)" and an empty text input field followed by a "%" symbol. At the bottom, there is a button labeled "RODAR".

Etapa 5: Verifique o diretório fornecido para garantir que o arquivo foi carregado corretamente.



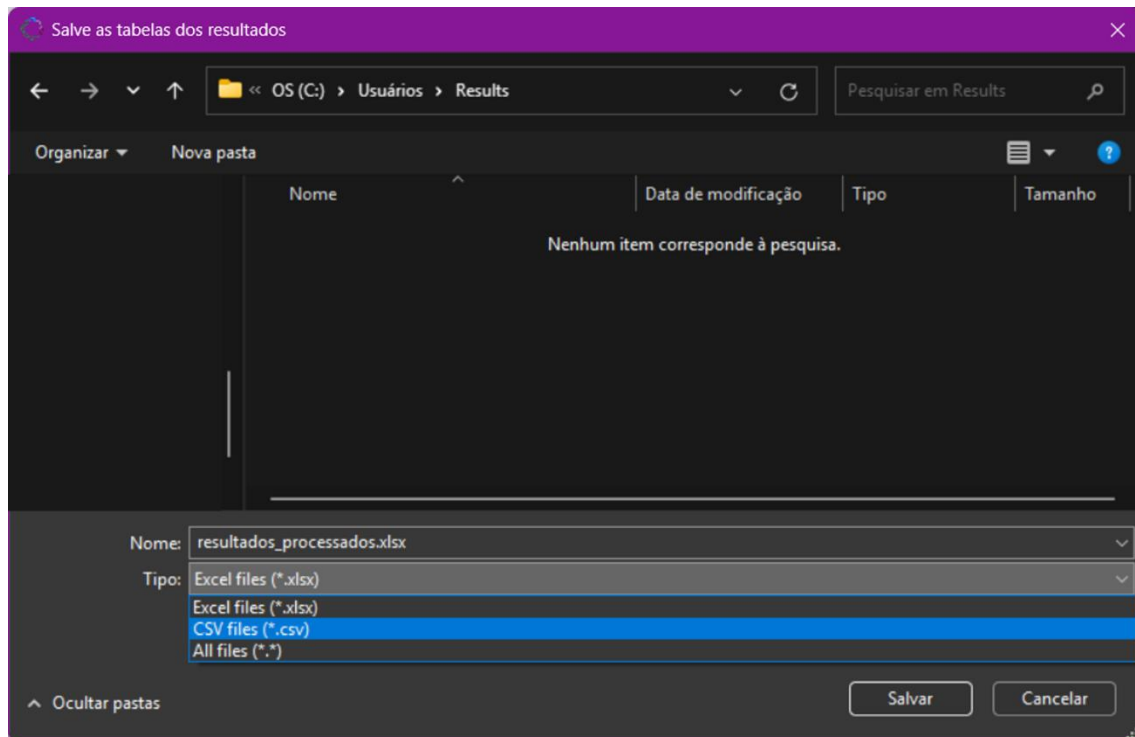
The screenshot shows the same window as before, but now the text "Arquivo carregado C:/Users/input_exemplo.xlsx" is displayed below the "Carregue um arquivo" button. The "Seleção de threshold" section and the "RODAR" button remain the same.

Etapa 6: Insira um valor do *threshold* em porcentagem ou deixe vazio para usar o padrão (0,05%) e execute a etapa de Aplicação do *threshold*.

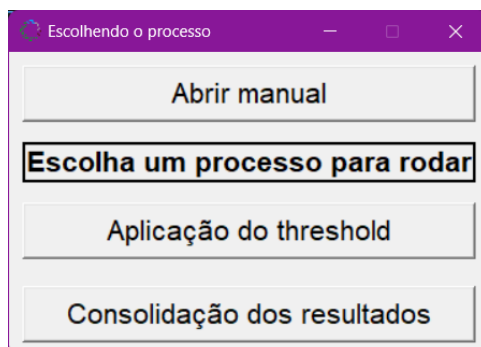


The screenshot shows the window with the file loaded. In the "Seleção de threshold" section, the text input field now contains the value "0.07". The "RODAR" button is still at the bottom.

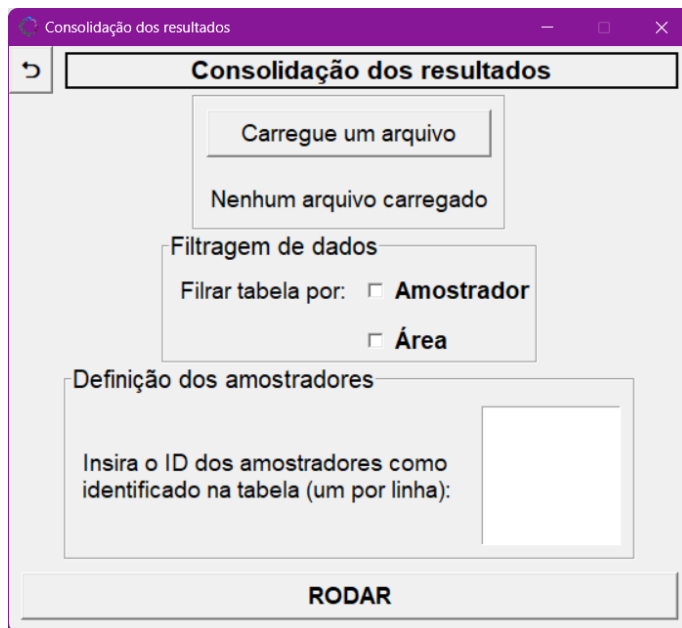
Etapa 7: pós a execução, três arquivos de saída serão gerados, contendo respectivamente: (i) o valor do *threshold* calculado por amostra de sequenciamento, (ii) as OTUs com reads abaixo do limiar e (iii) as OTUs com reads iguais ou acima do limiar. Salve pelo menos o último arquivo em .xlsx ou .csv para usar como entrada na etapa de consolidação.



Etapa 8: Escolha a opção "Consolidação de Resultados".

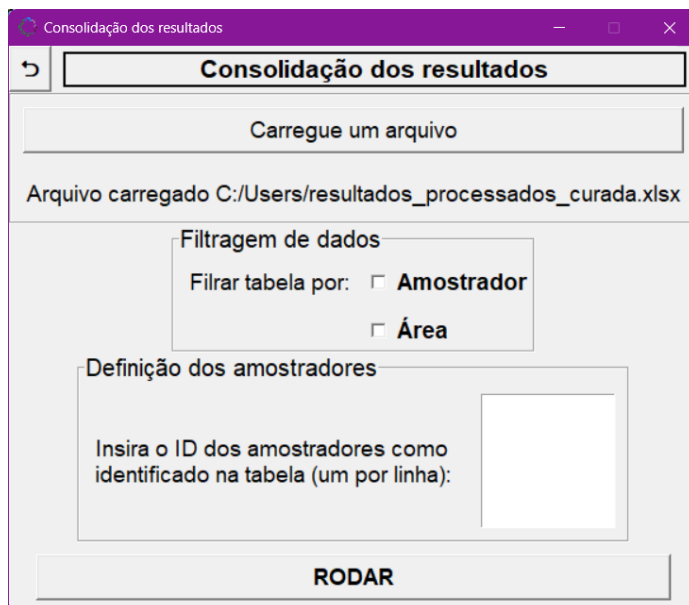


Etapa 9: Carregue um arquivo contendo a tabela curada após a aplicação do *threshold*.



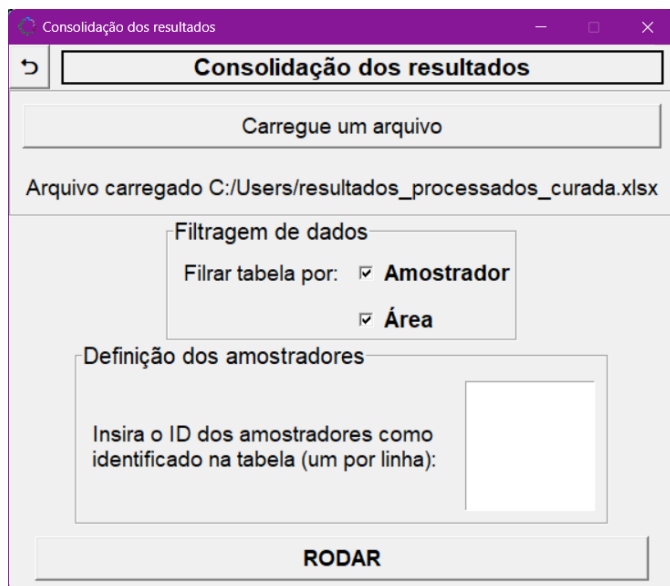
The screenshot shows a window titled "Consolidação dos resultados" with a purple header bar. Below the title bar is a navigation bar with a back arrow and the title "Consolidação dos resultados". The main area contains several sections: a "Carregue um arquivo" button, a status message "Nenhum arquivo carregado", a "Filtragem de dados" section with a label "Filtrar tabela por:" and two radio buttons, "Amostrador" (selected) and "Área", and a "Definição dos amostradores" section with a label "Insira o ID dos amostradores como identificado na tabela (um por linha):" and a text input field. At the bottom is a large "RODAR" button.

Etapa 11: Verifique o diretório para confirmar o carregamento correto do arquivo.



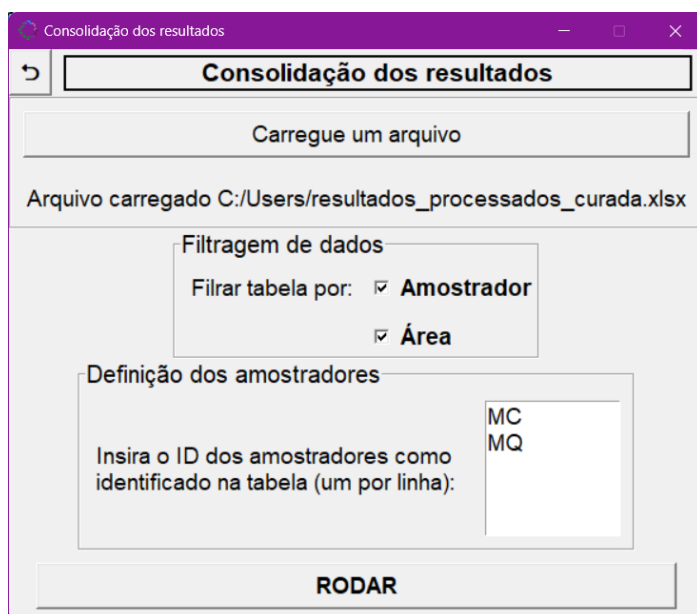
The screenshot shows the same "Consolidação dos resultados" window, but now the status message indicates "Arquivo carregado C:/Users/resultados_processados_curada.xlsx". The "Carregue um arquivo" button is now disabled. The "Filtragem de dados" section and "Definição dos amostradores" section remain the same. The "RODAR" button is still at the bottom.

Etapa 12: Selecione as opções para filtrar os dados por coletor e/ou área.



The screenshot shows a window titled "Consolidação dos resultados". At the top, there is a button "Carregue um arquivo". Below it, the text "Arquivo carregado C:/Users/resultados_processados_curada.xlsx" is displayed. The "Filtragem de dados" section contains the label "Filtrar tabela por:" followed by two checked checkboxes: "Amostrador" and "Área". Below this is the "Definição dos amostradores" section, which includes the instruction "Insira o ID dos amostradores como identificado na tabela (um por linha):" and an empty text input field. At the bottom of the window is a large button labeled "RODAR".

Etapa 12: Se selecionou a opção "Amostrador", digite os IDs dos coletores, separando-os com a tecla "Enter". Caso contrário, deixe o campo em branco.



This screenshot is similar to the previous one, but the text input field in the "Definição dos amostradores" section now contains the text "MC" followed by "MQ" on a new line, demonstrating the use of the "Enter" key to separate collector IDs.

Etapa 13: Salve as tabelas geradas com os dados filtrados. Para compactar todas as tabelas em .csv em um único arquivo .zip, selecione a opção ZIP. Assim, você salvará os resultados consolidados com a lista final de táxons e as informações sobre o número de *reads* que cada táxon apresentou e o número de vezes que foi detectado.

