

# ロングリードデータを使用した菌叢解析

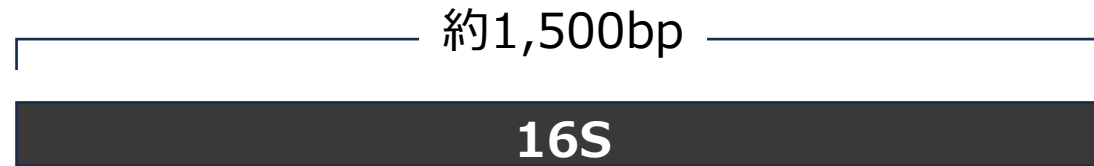
フィルジェン株式会社 バイオインフォマティクス部  
(biosupport@filgen.jp)



菌叢解析に使用される16S遺伝子の全長が約1,500bpなのに対し、一般的なNGSのリード長は数百bpであるため、一部の領域のみを利用している。



情報量が限られるため、種レベルの分類には不十分な場合がある



一方でロングリードシーケンサーでは、数kbのリードも得られるため、全長を使用可能できる。

しかしながら、ロングリードではエラー率が比較的高いという別の問題が存在する。

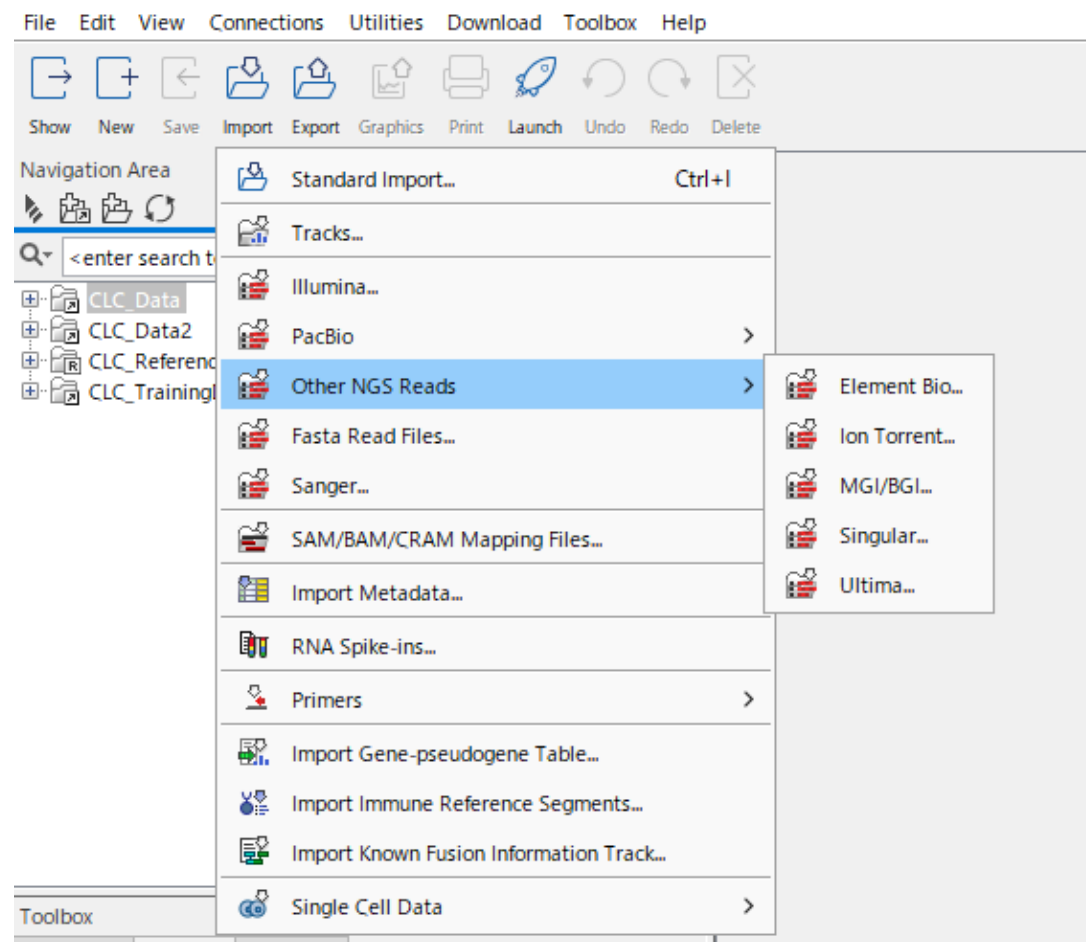
↓

既存のアルゴリズムをそのままロングリードデータに適用することができない。

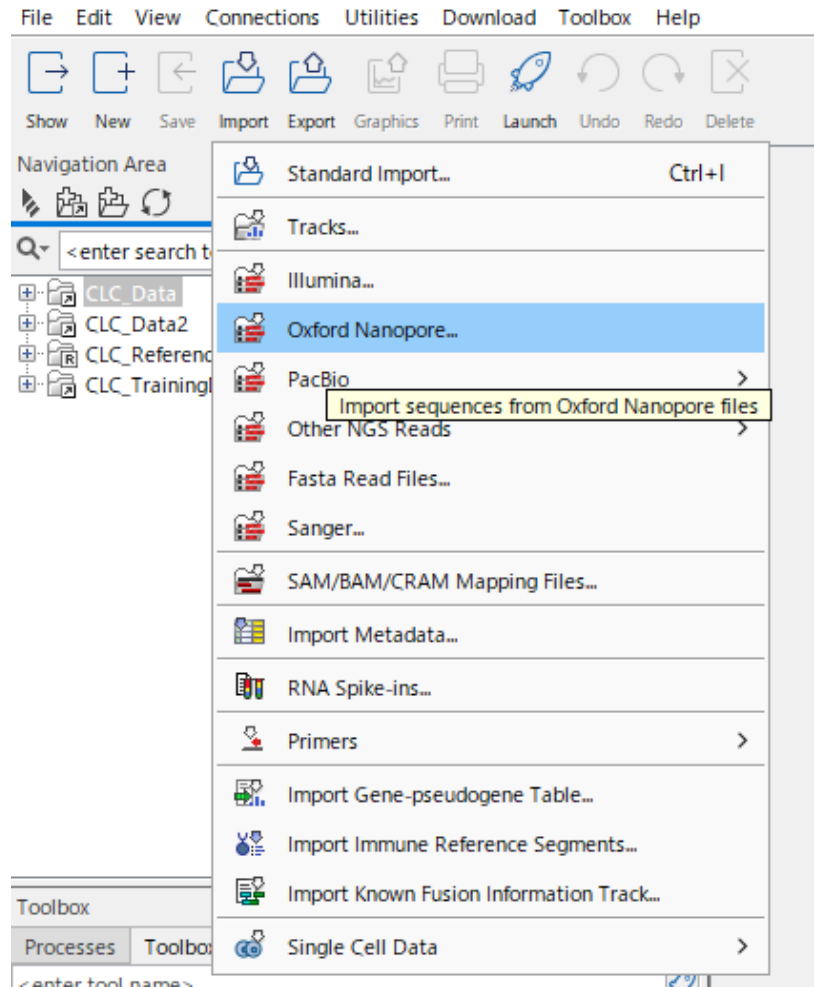
Curry, K.D., Wang, Q., Nute, M.G. *et al.* **Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data.** *Nat Methods* **19**, 845–853 (2022).  
<https://doi.org/10.1038/s41592-022-01520-4>

“Emu”というロングリード用の分類群識別アルゴリズムが2022年に公開された。

CLC Genomic WorkbenchのプラグインCLC Microbial Genomics Moduleにも、Emuに基づくロングリード用の分類群識別ツールが搭載された（ver.24以降）。



デフォルトの状態では、ロングリード（Nanopore）のインポーターが利用できないので、プラグインLong Read Support（無償）をインストールする必要がある。

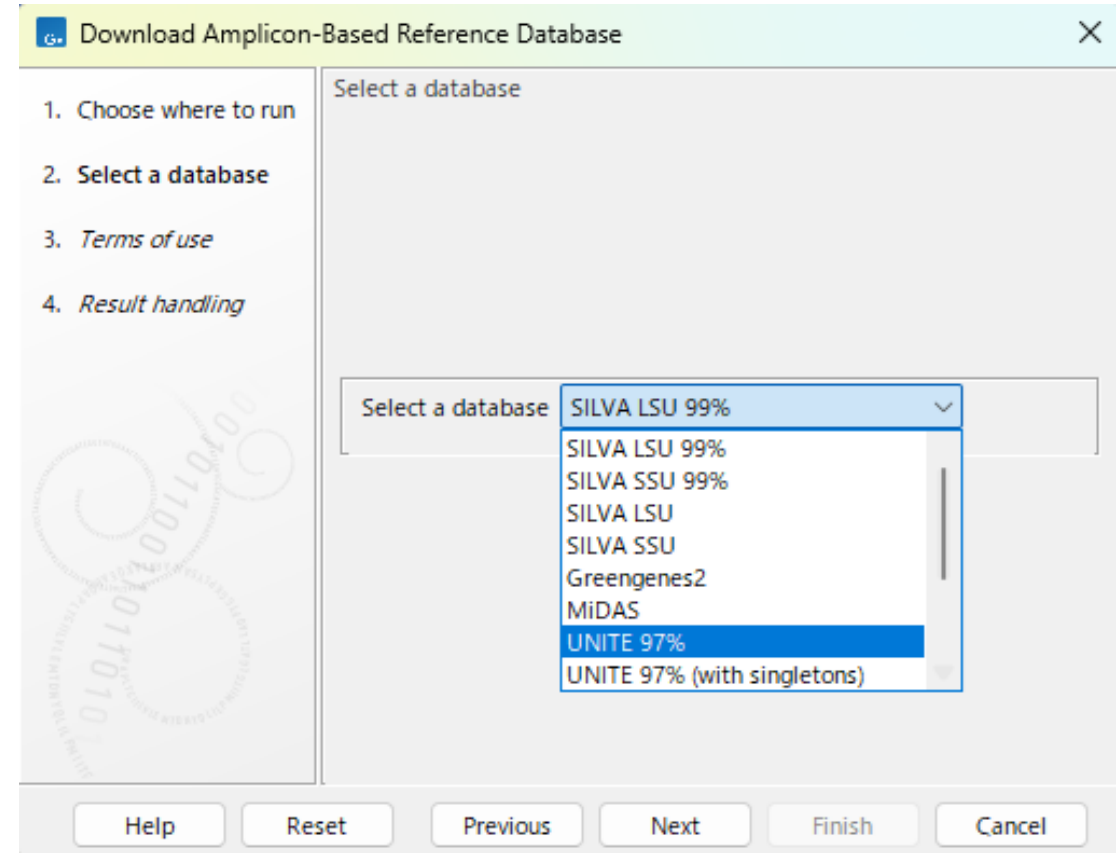
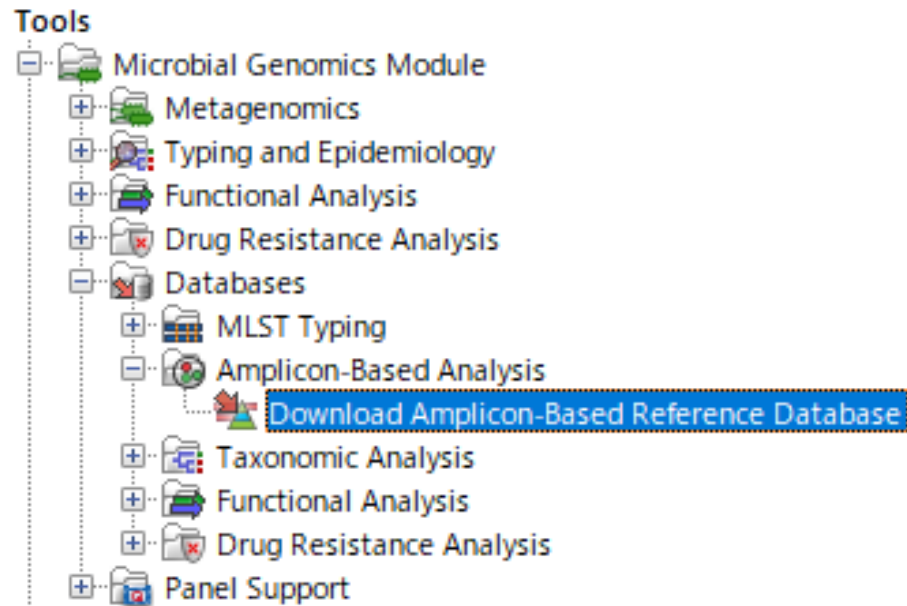


プラグインをインストールすると、Nanoporeのインポーターが使用可能になる。

## Tools

- Microbial Genomics Module
- Long Read Support
  - Correct Long Reads
  - De Novo Assemble Long Reads
  - Map Long Reads to Reference
  - Polish with Reads
  - RNA-Seq Analysis for Long Reads
  - Structural Variant Caller for Long Reads

プラグインのインストール後は、ロングリード用のツールがいくつか使用可能になる。  
ロングリードのマッピング、トランスクリプトーム解析、ハイブリッドアセンブリ等が可能。



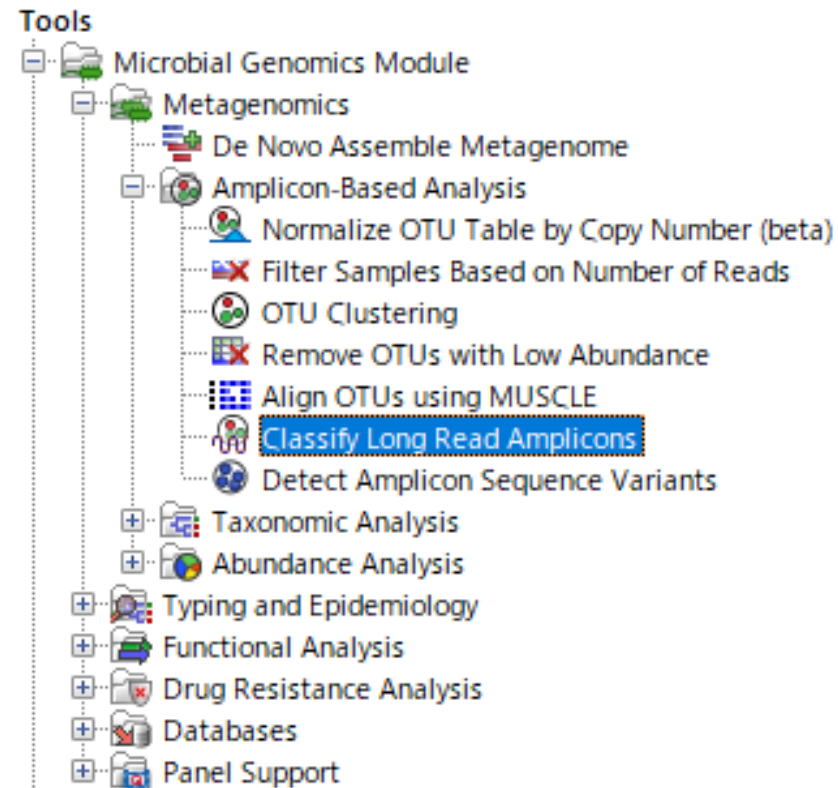
事前にリファレンスのデータベースをダウンロードする必要がある。  
(ショートリードの場合と同様)



- ERR3363792
- ERR3363793
- ERR3363794
- ERR3363781
- ERR3363782
- ERR3363783
- ERR3363804
- ERR3363805
- ERR3363806


	A	B	
1	Name	Month	
2	ERR336781	April	
3	ERR336782	April	
4	ERR336783	April	
5	ERR336792	June	
6	ERR336793	June	
7	ERR336794	June	
8	ERR336804	August	
9	ERR336805	August	
10	ERR336806	August	
11			

インポートするリード名と、それぞれが属するグループの情報を含んだエクセルファイルを用意しておく。



今回主に使用するツールは、Microbial Genomics Moduleのフォルダに入っている“Classify Long Read Amplicons”ツールです。

## ショートリード用のツール

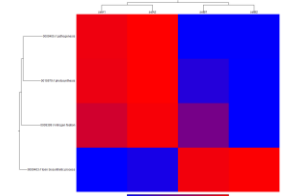
-  OTU Clustering
-  Detect Amplicon Sequence Variants

## ロングリード用のツール

-  Classify Long Read Amplicons

### ERR3363783 (abundance table)

Name	Taxonomy	Combine...	ERR336...	E
Limnohabitan	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Limno...	13462	4865	
Comamonadace...	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae	8933	4178	
Rhodofera	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Rhodo...	2836	1626	
Cytophagaceae	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Cytophagaceae	17734	1453	
Stramenopiles	Bacteria; Cyanobacteria; Chloroplast; Stramenopiles	20545	908	
Arcobacter	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Campylobacteracea...	9300	717	
Sphingobacteri...	Bacteria; Bacteroidetes; Sphingobacterii; Sphingobacteriales	10636	674	
Rhodocyclaceae	Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae	690	506	
Curvus	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Limno...	1679	458	
Polaromonas	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Polaro...	696	453	
Hydrogenophaga	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Hydro...	1233	427	
Oxalobacteraceae	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae	1144	215	
Flavobacterium	Bacteria; Bacteroidetes; Flavobacterii; Flavobacteriales; Flavobacteriaceae; Flavobacteri...	1504	153	
[Cerasicocceae]	Bacteria; Verrucomicrobia; Opitutae; [Cerasicoccales]; [Cerasicocceae]	4506	138	



ショートリード用のツールと同様のフォーマットで結果が出力される。



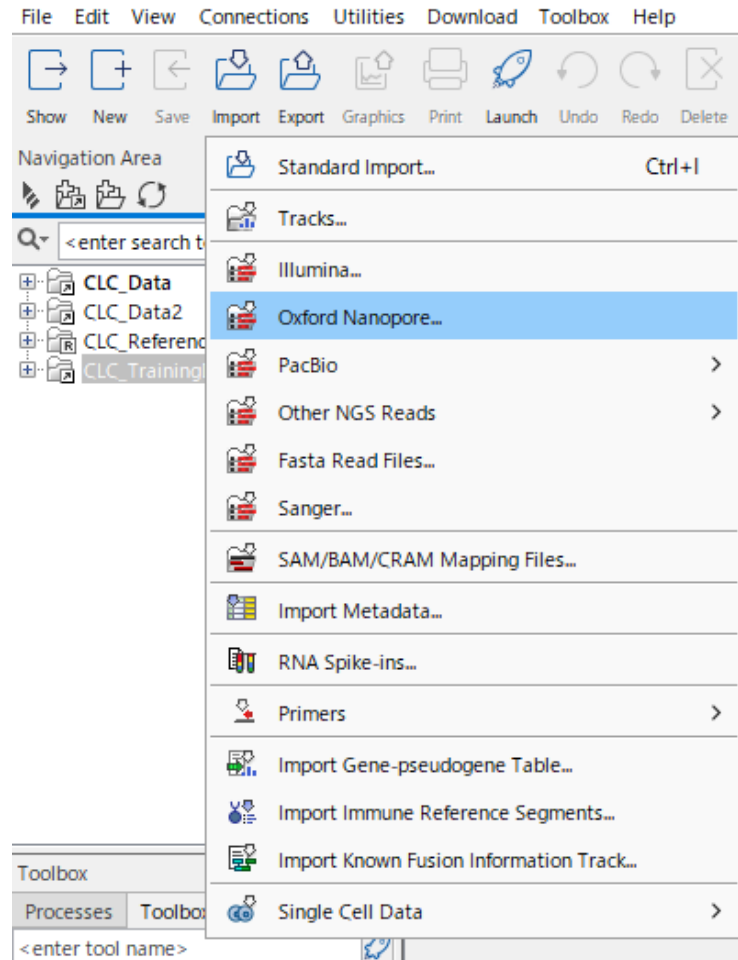
この後は、これまでと同様に図表の作成（サンバースト図・ヒートマップ）が可能

ロングリード（Nanopore）のインポート

サンプルごとの組成解析

サンプル間の組成比較

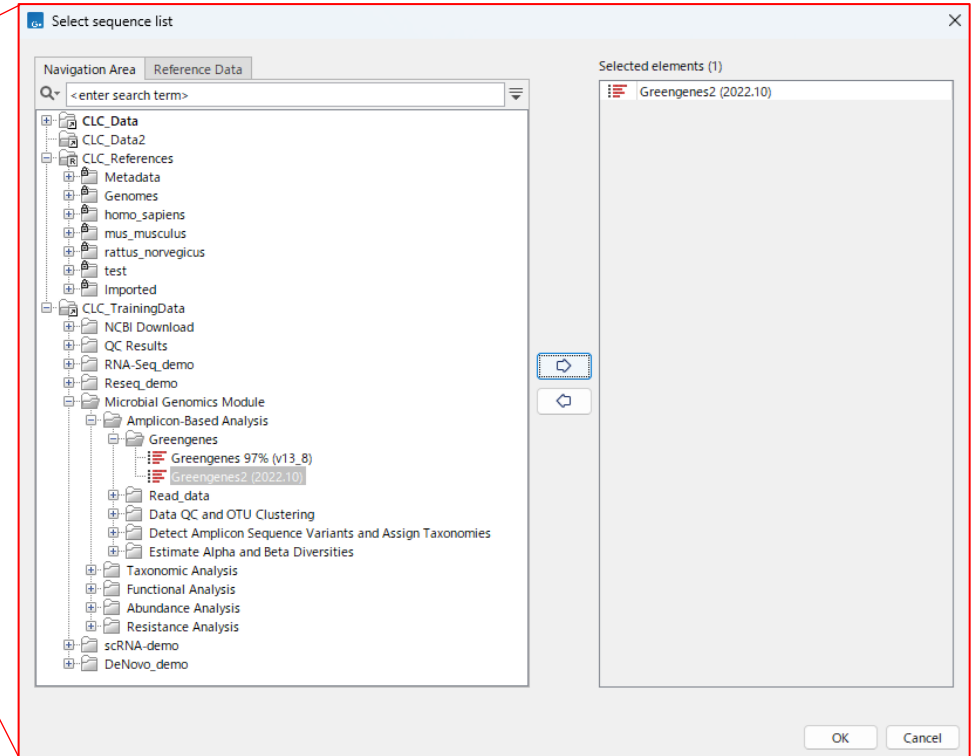
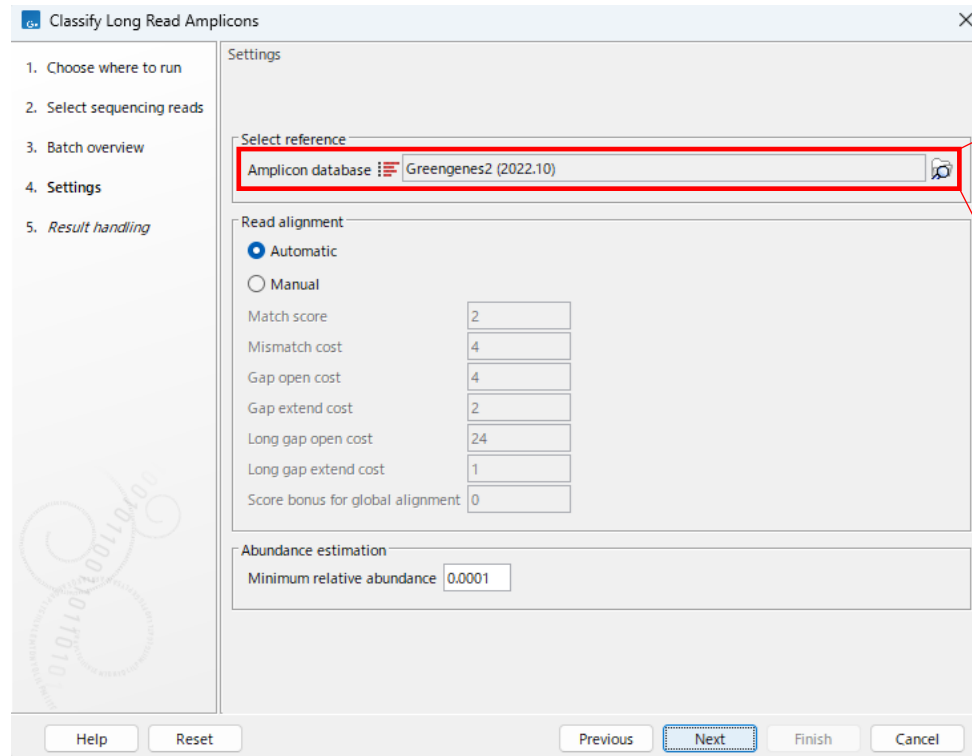
図表の作成



ImportメニューのNanoporeのインポーターを利用してインポート

The image shows a software interface for 'Classify Long Read Amplicons'. On the left, a 'Tools' tree is visible with 'Classify Long Read Amplicons' highlighted. A large grey arrow points from this tool to a screenshot of the tool's execution window. The window is titled 'Classify Long Read Amplicons' and has a close button (X) in the top right. It contains a navigation pane on the left with steps: 1. Choose where to run, 2. Select sequencing reads, 3. Settings, and 4. Result handling. The main area is titled 'Select sequencing reads' and features a search bar with the placeholder '<enter search term>'. Below the search bar is a list of sequencing reads, including 'test1', 'SRA MetadataTable-1', 'SRR10391187', 'SRA MetadataTable-2', 'ERR3363804', 'SRA MetadataTable-3', 'ERR3363792', 'SRA MetadataTable-4', 'ERR3363781', 'SRA MetadataTable-5', 'ERR3363782', 'SRA MetadataTable-6', 'ERR3363783', 'SRA MetadataTable-7', 'ERR3363793', 'ERR3363793 (sampled)', 'SRA MetadataTable-8', 'ERR3363794', 'SRA MetadataTable-9', 'ERR3363805', 'ERR3363805 (sampled)', 'SRA MetadataTable-10', 'ERR3363806', and 'ERR3363806 (sampled)'. At the bottom of this list is a 'CLC Data2' folder icon. To the right of the list is a 'Selected elements (9)' panel containing a list of read IDs: ERR3363792, ERR3363781, ERR3363782, ERR3363783, ERR3363793 (sampled), ERR3363805 (sampled), ERR3363806 (sampled), ERR3363794, and ERR3363804. Below the list are two arrow buttons. At the bottom of the window, there is a 'Batch' checkbox which is checked, and a row of buttons: Help, Reset, Previous, Next, Finish, and Cancel.

Classify Long Read Ampliconツールを起動し、インポートしたロングリードデータを選択  
(各リードが別々のサンプルに由来する場合はbatchをチェック)



Amplicon databaseでリファレンスとして使用するデータベースを選択

# サンプルごとの組成解析

CLC Genomics Workbench 24.0.1 - <New Workspace> - Evaluation 123 days remaining

Navigation Area: ERR3363783 (abundance table) X

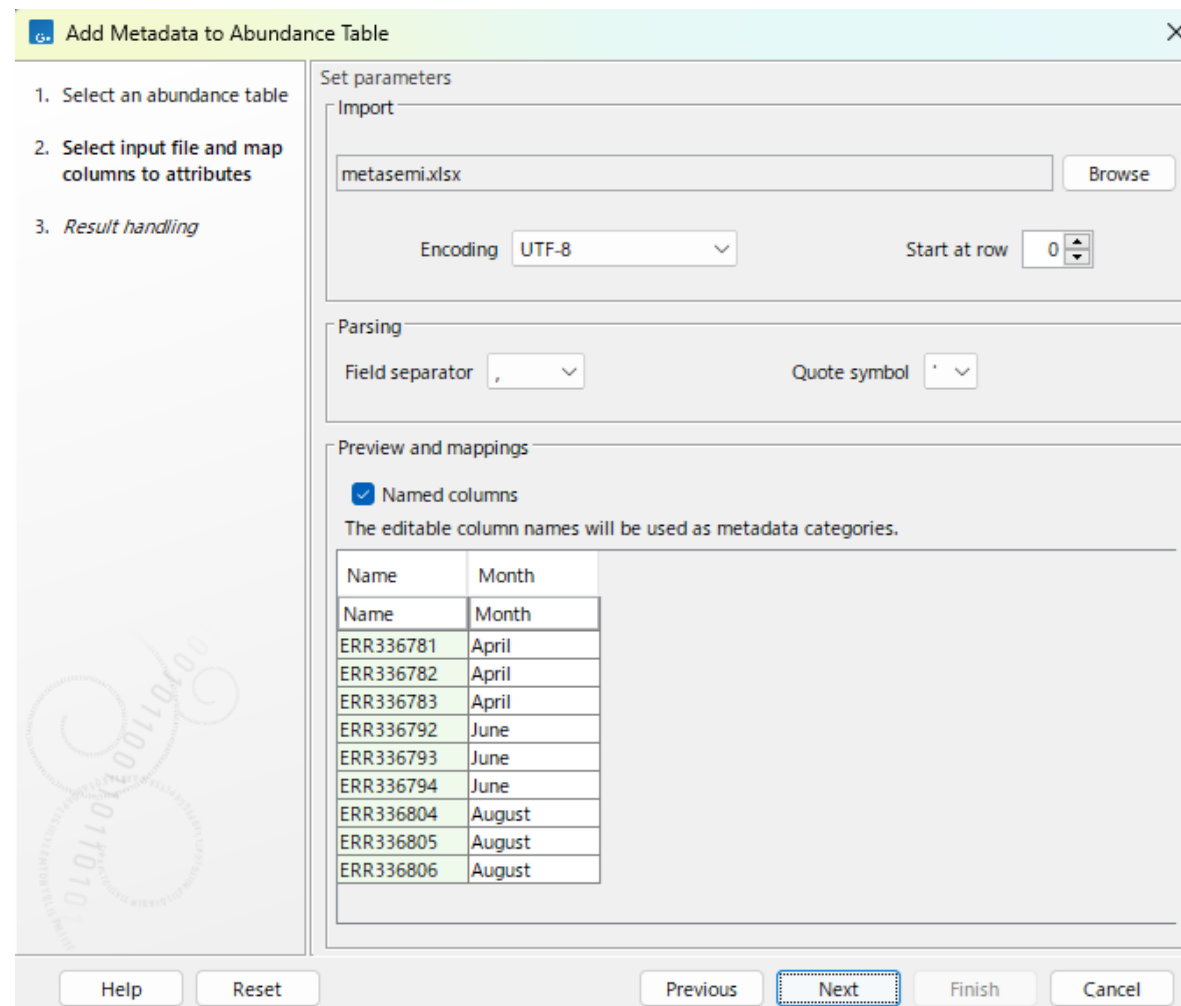
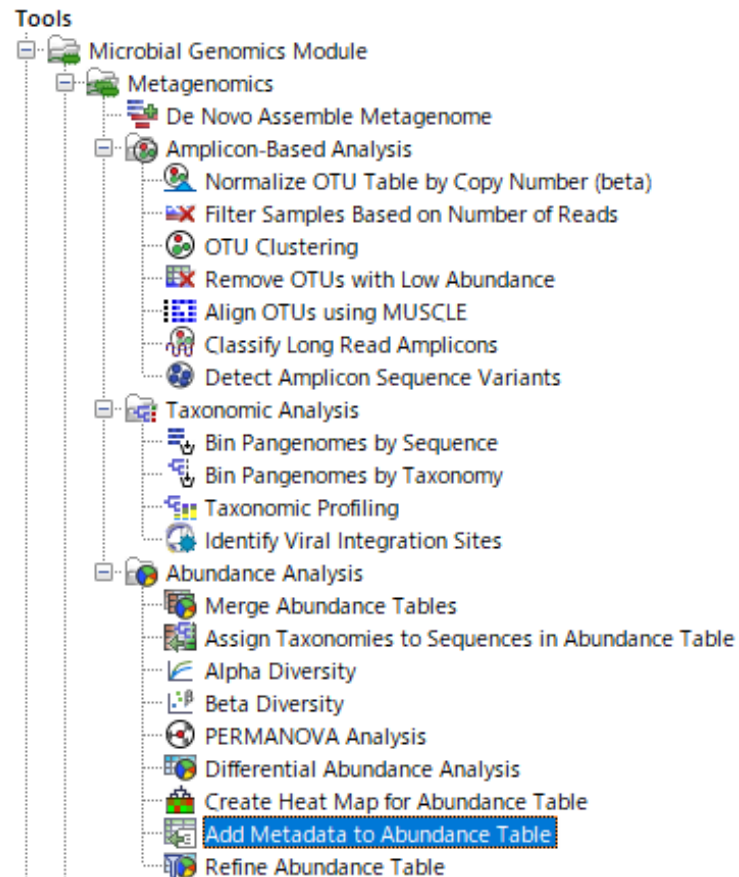
Name	Taxonomy	Combined Abundance	ERR3363783 Abundance
Stramenopiles	Bacteria; Cyanobacteria; Chloroplast; Stramenopiles	55	55
Commonadaceae	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae	43	43
Limnochabans	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Limnochabans	38	38
Sphingobacteriales	Bacteria; Bacteroidetes; Sphingobacteriales; Sphingobacteriales	37	37
Cytophagaceae	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Cytophagaceae	37	37
Arcobacter	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Campylobacteraceae; Arcobacter	26	26
[Cerasioccaceae]	Bacteria; Vermicomicrobia; Opisthokonta; Cerasioccaceae; Cerasioccaceae	19	19
Rhodoferrax	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Rhodoferrax	16	16
Rhodobacter	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Rhodobacter	11	11
Salmonella	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella	10	10
Enterobacteriaceae	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae	9	9
Sphingomonadales	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales	8	8
Hydrogenophaga	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Hydrogenophaga	7	7
Sonnei	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella; Sonnei	6	6
Flavobacterium	Bacteria; Bacteroidetes; Flavobacteriales; Flavobacteriales; Flavobacterium	6	6
Seeligii	Bacteria; Firmicutes; Bacilli; Bacillales; Listeriaceae; Listeria; Seeligii	6	6
Pseudomonas	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	6	6
Curvus	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Limnochabans; Curvus	5	5
Rhodocytaeae	Bacteria; Proteobacteria; Betaproteobacteria; Rhodocytales; Rhodocytaeae	5	5
Aureus	Bacteria; Firmicutes; Bacilli; Bacillales; Staphylococcaceae; Staphylococcus; Aureus	5	5
Bacillus	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus	5	5
Polaromonas	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Polaromonas	5	5
Nitrospira	Bacteria; Nitrospirae; Nitrospira; Nitrospirales; Nitrospiraceae; Nitrospira	4	4
Dechloromonas	Bacteria; Proteobacteria; Betaproteobacteria; Rhodocytales; Rhodocytaeae; Dechloromonas	4	4
Coli	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia; Coli	4	4
Kujiense	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Sulfuricum; Kujiense	3	3
Novosphingobium	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Novosphingobium	3	3
Zoogaea	Bacteria; Proteobacteria; Betaproteobacteria; Rhodocytales; Rhodocytaeae; Zoogaea	3	3
Sulfurspirillum	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Campylobacteraceae; Sulfurspirillum	3	3
SC-184	Bacteria; Proteobacteria; Betaproteobacteria; SC-184	3	3
Xanthomonadaceae	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae	3	3
Lachnospiraceae	Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae	3	3
Rickettsiales	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales	3	3
Armatimonadaceae	Bacteria; Armatimonadetes; Armatimonadales; Armatimonadaceae	3	3
Sphingomonadaceae	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae	3	3
Oxalobacteraceae	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae	3	3
Enterica	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella; Enterica	3	3
Cryaeisphilus	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Campylobacteraceae; Arcobacter; Cryaeisphilus	3	3
Polynucleobacter	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; Polynucleobacter	2	2
Bacillaceae	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae	2	2
Pseudomonadaceae	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae	2	2
Methylophilaceae	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae	2	2
Alcaligenaceae	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae	2	2
RS62	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; RS62	2	2
Paracoccus	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Paracoccus	2	2
Enterococcus	Bacteria; Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus	2	2
Lactobacillus	Bacteria; Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus	2	2
Fluvicola	Bacteria; Bacteroidetes; Flavobacteriales; Flavobacteriales; Cryomorphaceae; Fluvicola	2	2
Dok59	Bacteria; Proteobacteria; Betaproteobacteria; Rhodocytales; Rhodocytaeae; Dok59	2	2
Halicomonobacter	Bacteria; Bacteroidetes; [Saprosiriales]; [Saprosiriales]; Saprosiraceae; Halicomonobacter	2	2

Visualization icons:

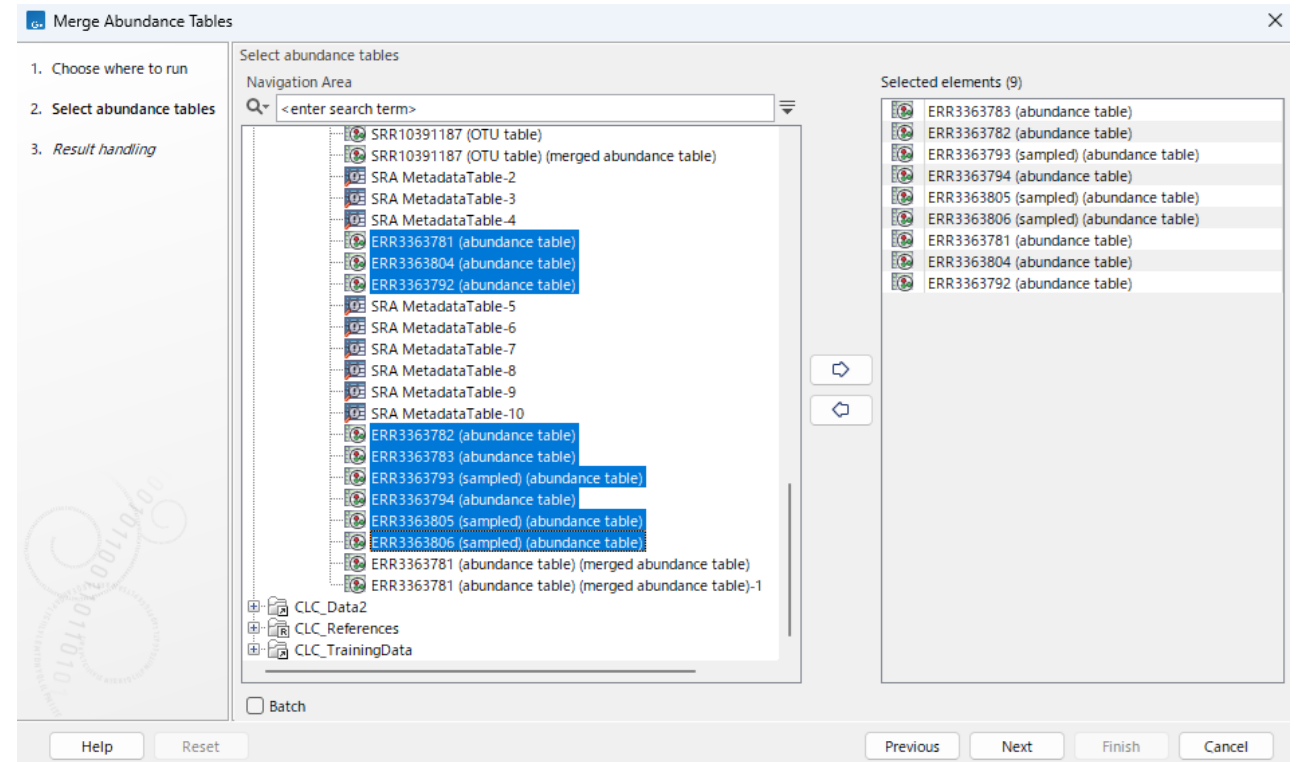
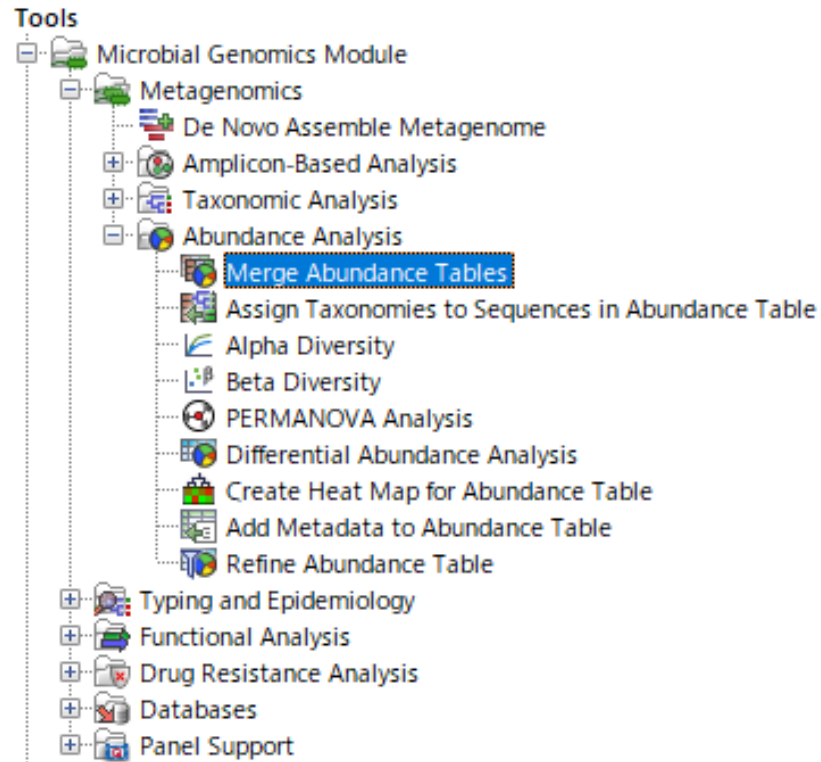
ERR3363783 (abundance table)

下のアイコン列から、積み上げ棒グラフやサンバー-スト図を表示させることが可能



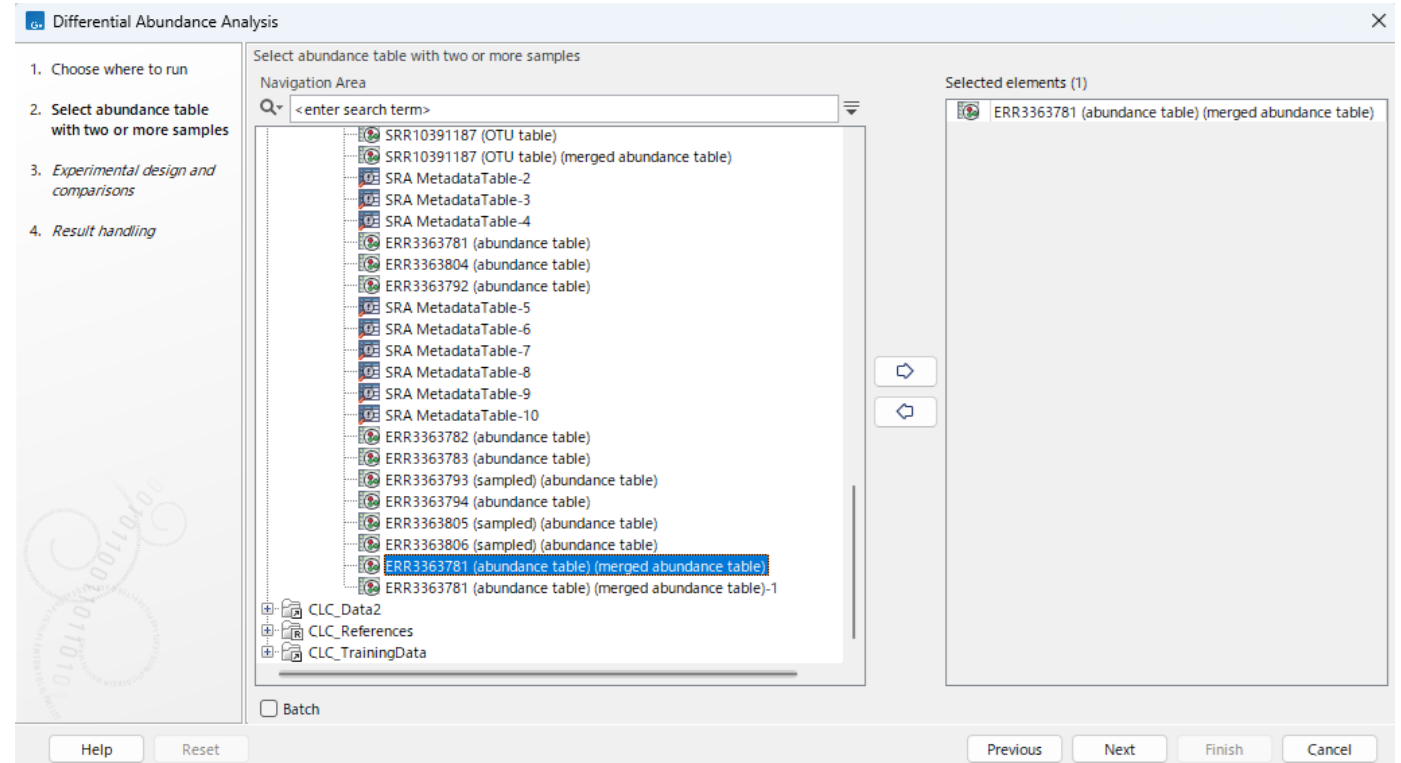
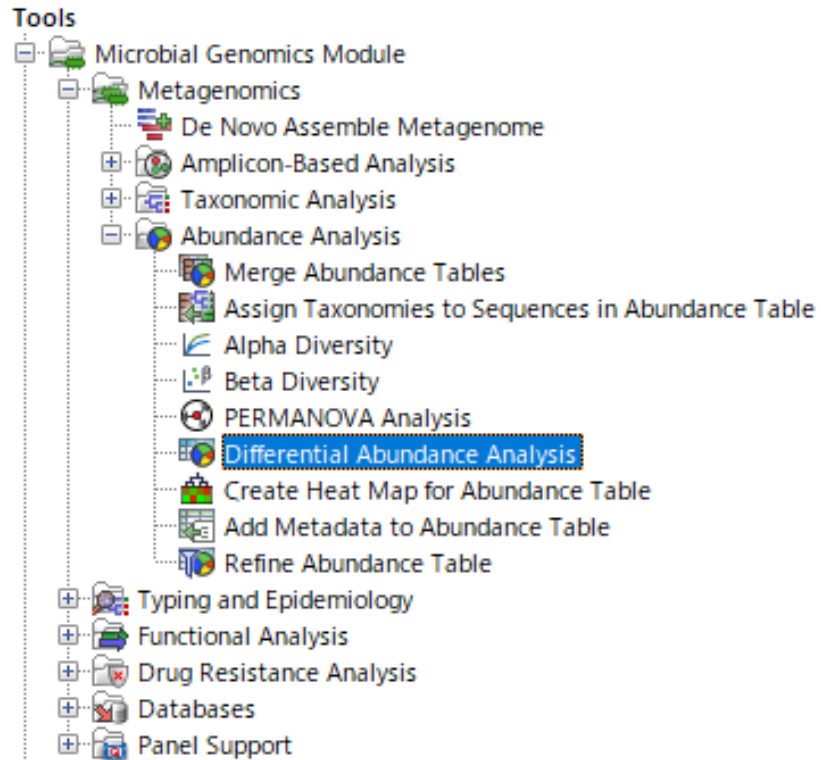


Add Metadata to Abundance Tableツールで、組成表にメタデータの情報进行付加



Merge Abundance Tablesを使用し、サンプルごとの組成表をすべて選択

# サンプル間の組成比較



Differential Abundance Analysisツールを起動し、マージした組成表を選択

Differential Abundance Analysis

1. Choose where to run  
2. Select abundance table with two or more samples  
3. Experimental design and comparisons  
4. Result handling

Experimental design and comparisons

Experimental design

Test differential abundance due to: Sample Name

Correct for: (Nothing selected)

Comparisons

Across groups (ANOVA-like)  
 All group pairs  
 Against control group

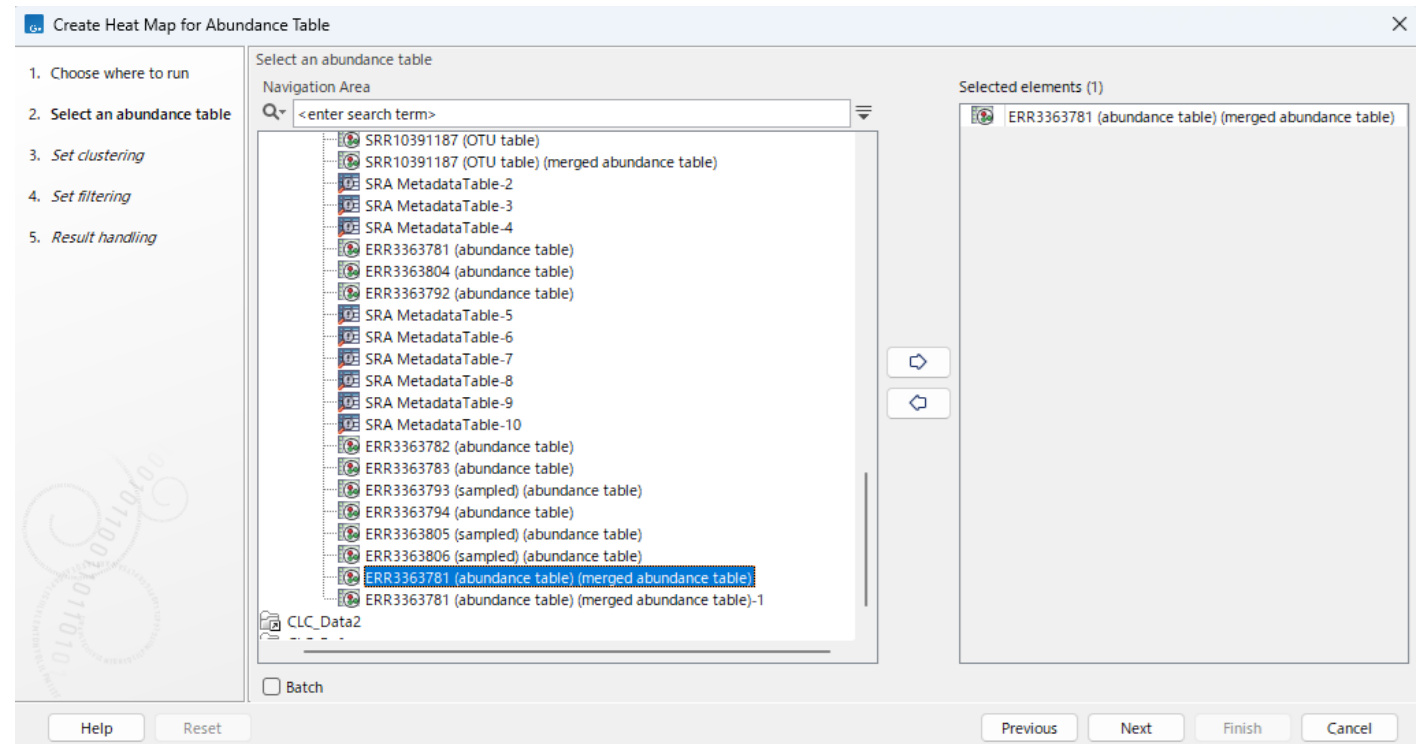
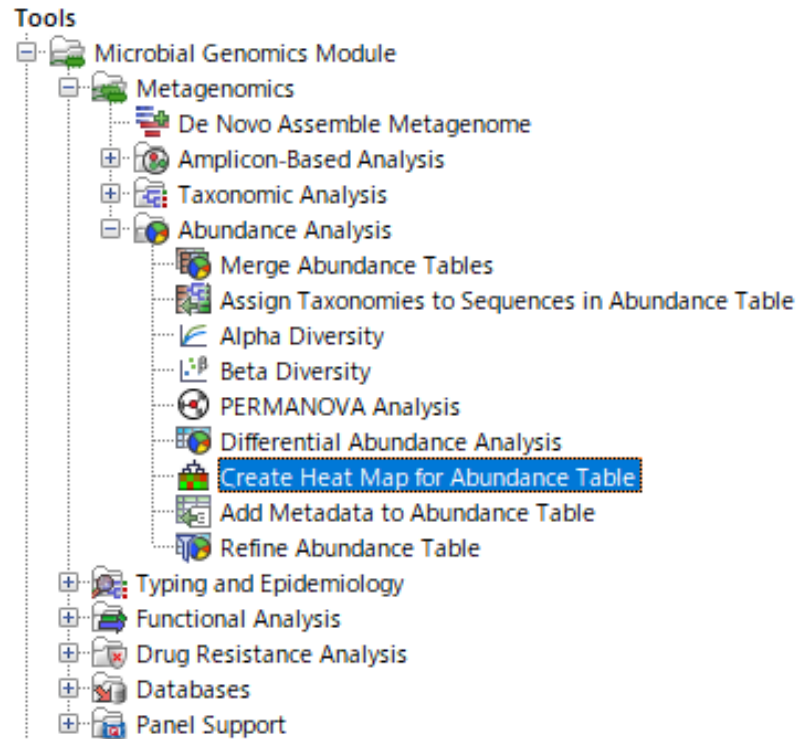
Among comparison groups: Selected 9 elements.

Control group: 9

Attention: No replicates for group: 9, 10, 11, 20, 21, 22, 32, 33, 34. The results may be unreliable. Proceed with care.

Help Reset Previous Next Finish Cancel

比較に用いるキーおよび比較方法を選択  
(組成表とメタデータが紐づいている場合は、群間比較等が可能)



Create Heat Map for Abundance Tableツールを起動し、マージした組成表を選択

お問い合わせ先：フィルジエン株式会社

TEL: 052-624-4388 (9:00～17:00)

FAX: 052-624-4389

E-mail: [biosupport@filgen.jp](mailto:biosupport@filgen.jp)