

CLC Genomics Workbench Premiumを利用した ASV解析とPERMANOVA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)



次世代シーケンス解析用ソフトウェア

◆ リシーケンシング解析

- ・ リファレンスゲノムへのマッピング
- ・ 変異検出

◆ トランスクリプトミクス解析

- ・ RNA-seq解析
- ・ small RNA解析

◆ エピゲノミクス解析

- ・ ChIP-seq解析
- ・ バイサルファイトシーケンス解析

◆ De Novo シーケンス解析

- ・ De Novo Assembly
- ・ BLAST解析

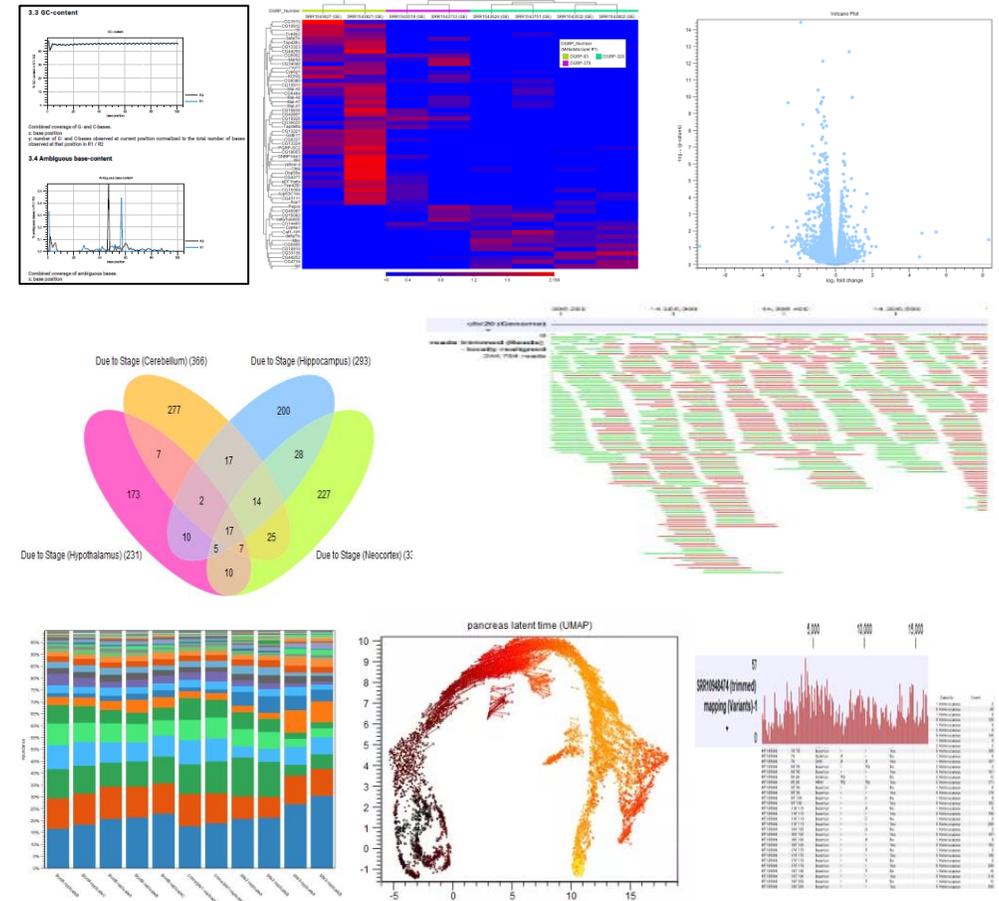
◆ 菌叢解析

◆ ゲノムフィニッシング解析

◆ シングルセル解析

◆ 超高速変異解析

} Premium版 限定機能



データのダウンロード

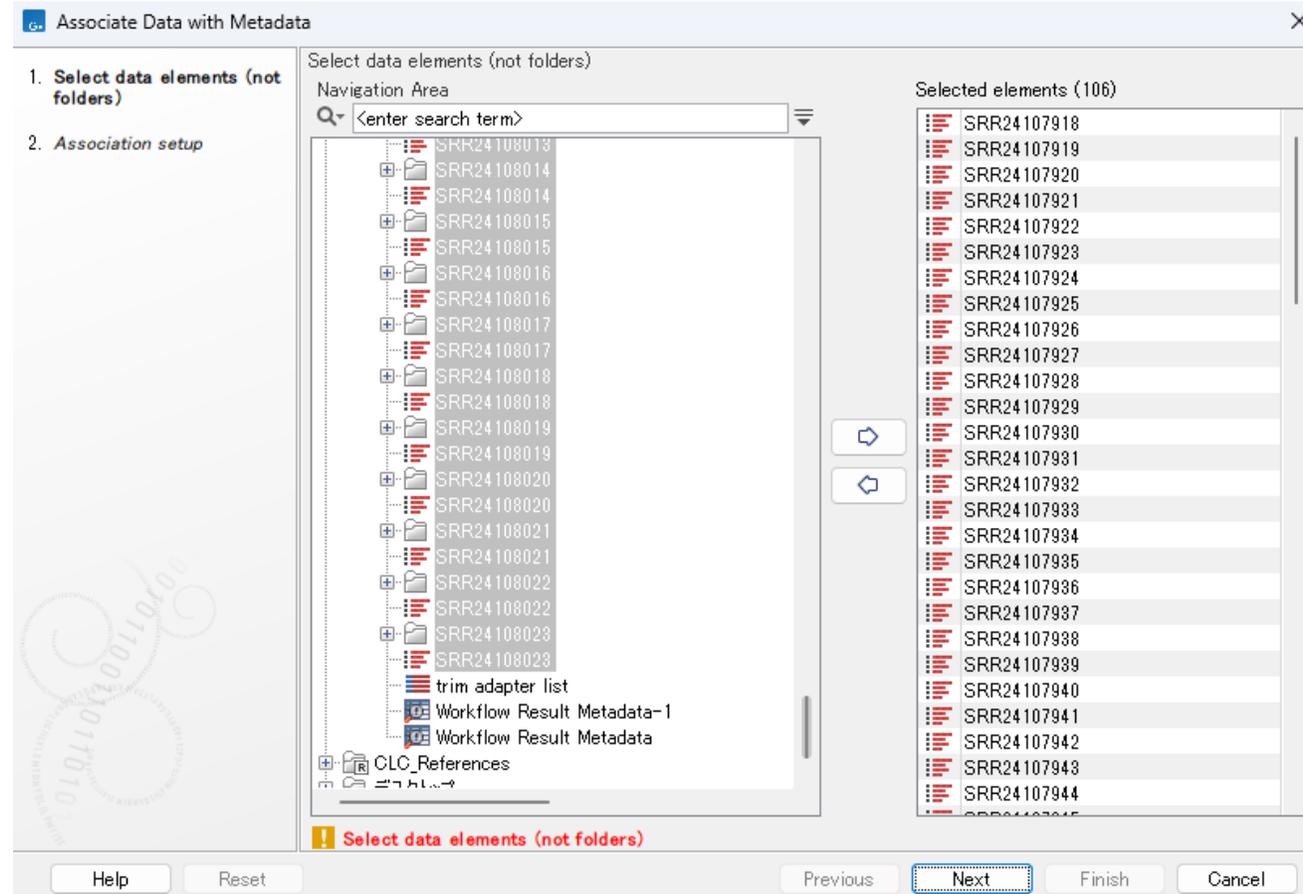
#	Run accession	Study accession	Scientific name	Download size (MB)	Paired	Read orientation	Average length	Spots	PubMed
1	SRR24107957	SRP431526	soil metagenome	25	Yes	Unknown	501	83,284	
2	SRR24107956	SRP431526	soil metagenome	30	Yes	Unknown	501	121,705	
3	SRR24107955	SRP431526	soil metagenome	31	Yes	Unknown	501	103,719	
4	SRR24107954	SRP431526	soil metagenome	25	Yes	Unknown	501	88,741	
5	SRR24107953	SRP431526	soil metagenome	36	Yes	Unknown	501	125,053	
6	SRR24107952	SRP431526	soil metagenome	29	Yes	Unknown	501	100,492	
7	SRR24107951	SRP431526	soil metagenome	36	Yes	Unknown	501	123,179	
8	SRR24107950	SRP431526	soil metagenome	28	Yes	Unknown	501	95,749	
9	SRR24107949	SRP431526	soil metagenome	49	Yes	Unknown	501	164,280	
10	SRR24107948	SRP431526	soil metagenome	32	Yes	Unknown	501	109,212	
11	SRR24107957	SRP431526	soil metagenome	30	Yes	Unknown	501	101,582	
12	SRR24107956	SRP431526	soil metagenome	29	Yes	Unknown	501	97,627	
13	SRR24107955	SRP431526	soil metagenome	26	Yes	Unknown	501	80,775	
14	SRR24107954	SRP431526	soil metagenome	20	Yes	Unknown	501	98,091	
15	SRR24107953	SRP431526	soil metagenome	22	Yes	Unknown	501	74,967	
16	SRR24107952	SRP431526	soil metagenome	19	Yes	Unknown	501	66,747	
17	SRR24107951	SRP431526	soil metagenome	20	Yes	Unknown	501	66,670	
18	SRR24107950	SRP431526	soil metagenome	22	Yes	Unknown	501	76,602	
19	SRR24107949	SRP431526	soil metagenome	19	Yes	Unknown	501	54,519	
20	SRR24107948	SRP431526	soil metagenome	31	Yes	Unknown	501	107,277	
21	SRR24107947	SRP431526	soil metagenome	27	Yes	Unknown	501	92,744	
22	SRR24107946	SRP431526	soil metagenome	23	Yes	Unknown	501	77,484	
23	SRR24107945	SRP431526	soil metagenome	29	Yes	Unknown	501	99,005	
24	SRR24107944	SRP431526	soil metagenome	30	Yes	Unknown	501	105,199	
25	SRR24107943	SRP431526	soil metagenome	20	Yes	Unknown	501	71,088	
26	SRR24107942	SRP431526	soil metagenome	25	Yes	Unknown	501	86,591	
27	SRR24107940	SRP431526	soil metagenome	25	Yes	Unknown	501	81,187	
28	SRR24107941	SRP431526	soil metagenome	21	Yes	Unknown	501	73,919	
29	SRR24107939	SRP431526	soil metagenome	27	Yes	Unknown	501	94,781	
30	SRR24107938	SRP431526	soil metagenome	31	Yes	Unknown	501	106,126	
31	SRR24107937	SRP431526	soil metagenome	24	Yes	Unknown	501	83,706	
32	SRR24107936	SRP431526	soil metagenome	29	Yes	Unknown	501	99,006	
33	SRR24107935	SRP431526	soil metagenome	26	Yes	Unknown	501	89,062	
34	SRR24107933	SRP431526	soil metagenome	28	Yes	Unknown	501	96,219	
35	SRR24107934	SRP431526	soil metagenome	38	Yes	Unknown	501	129,045	
36	SRR24107932	SRP431526	soil metagenome	36	Yes	Unknown	501	121,925	
37	SRR24107931	SRP431526	soil metagenome	24	Yes	Unknown	501	85,605	
38	SRR24107930	SRP431526	soil metagenome	30	Yes	Unknown	501	101,079	
39	SRR24107929	SRP431526	soil metagenome	29	Yes	Unknown	501	99,762	
40	SRR24107928	SRP431526	soil metagenome	30	Yes	Unknown	501	104,160	
41	SRR24107927	SRP431526	soil metagenome	27	Yes	Unknown	501	94,382	
42	SRR24107926	SRP431526	soil metagenome	30	Yes	Unknown	501	103,341	
43	SRR24107925	SRP431526	soil metagenome	34	Yes	Unknown	501	111,886	
44	SRR24107922	SRP431526	soil metagenome	19	Yes	Unknown	501	62,936	
45	SRR24107924	SRP431526	soil metagenome	24	Yes	Unknown	501	81,467	
46	SRR24107923	SRP431526	soil metagenome	1	Yes	Unknown	500	3,997	
47	SRR24107921	SRP431526	soil metagenome	37	Yes	Unknown	501	127,661	
48	SRR24107920	SRP431526	soil metagenome	31	Yes	Unknown	501	108,403	
49	SRR24107919	SRP431526	soil metagenome	29	Yes	Unknown	501	80,652	

SRAダウンロード機能を使用し、配列データとメタデータをダウンロードします。
今回は異なる環境中の土壌に由来するメタゲノムデータを利用し、解析を行います。

配列とメタデータの紐づけ

The screenshot displays the CLC Genomics Workbench 23.0.5 interface. The main window shows a metadata table with 106 rows and 10 columns. The columns are: Run Accession, Experiment Accession, Project Accession, Sample Accession, Submission Acc., BioProject, BioSample, Center, and Center Proj... Design. The table contains data for various samples, including Run Accession (e.g., SRR24107947), Experiment Accession (e.g., SRX19907706), Project Accession (e.g., SRP431526), Sample Accession (e.g., SRS17261418), Submission Acc. (e.g., SRA1617461), BioProject (e.g., PRJNA953437), BioSample (e.g., SAMN04116318), Center (e.g., Arizona State University), and Center Proj... Design (e.g., DNA was extracted from 250). The interface includes a navigation area on the left with a search bar and a list of files, a toolbar at the bottom, and a metadata table settings panel on the right. The settings panel shows options for column width and show column (associations).

ダウンロードしたメタデータを開き、“Associate Data”をクリックします。



配列データを指定します。

Associate Data with Metadata

1. Select data elements (not folders)

2. **Association setup**

Association setup

Matching scheme

Exact - data element names must match a key exactly to be associated

Prefix - data elements with names starting with a matching key will be associated

Suffix - data elements with names ending with a matching key will be associated

Role assignment

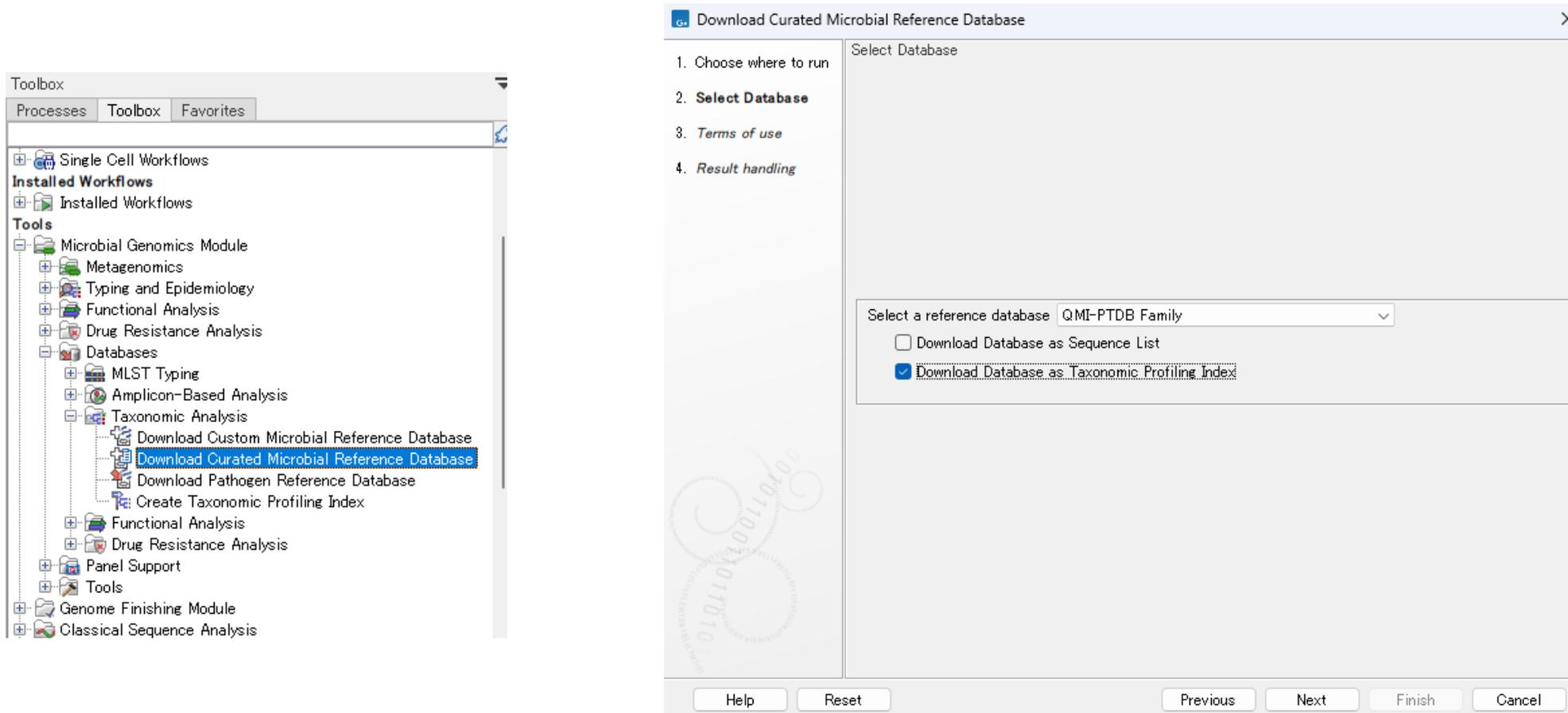
Role to assign: Sample data

Data association preview

Key	Element
SRR24107918	SRR24107918
SRR24107919	SRR24107919
SRR24107920	SRR24107920
SRR24107921	SRR24107921
SRR24107922	SRR24107922
SRR24107923	SRR24107923
SRR24107924	SRR24107924
SRR24107925	SRR24107925
SRR24107926	SRR24107926
SRR24107927	SRR24107927
SRR24107928	SRR24107928
SRR24107929	SRR24107929
SRR24107930	SRR24107930
SRR24107931	SRR24107931
SRR24107932	SRR24107932
SRR24107933	SRR24107933
SRR24107934	SRR24107934

Help Reset Previous Next Finish Cancel

メタデータ中のキーカラムと配列データ名を、
完全一致、前方一致、後方一致のいずれかで紐づけます。



分類に使用するデータベースをダウンロードします。
“Download Database as Taxonomic Profiling Index”にチェックを入れます。

アダプター（プライマー）リストの作成

The screenshot displays the CLC Genomics Workbench 23.0.5 interface. The main window shows a metadata table with 106 rows and columns for Run Accession, Experiment Accession, Project Accession, Sample Accession, Submission Accession, BioProject, BioSample, Center, and Center Project. The table contains data for various runs, including SRR24107941 through SRR24107922. The interface includes a menu bar (File, Edit, View, Connections, Utilities, Download, Toolbox, Help), a toolbar, a left sidebar with navigation options (Enzyme List, Folder, Sequence, Track List, Trim Adapter List, Motif List, Metadata Table, Workflow, Location, Merged abundance table with taxonomies, SRA Metadata Table), and a right sidebar with column settings (Show column, Column width, Manual, Select All, Deselect All, Show column (associations)).

つぎに、配列から除去するアダプター（プライマー）のリストを作成します。

アダプター（プライマー）リストの作成

CLC Genomics Workbench 23.0.5 - <New Workspace> - Evaluation 72 days remaining

File Edit View Connections Utilities Download Toolbox Help

Browser

Navigation Area Remote Files

Center search term

Template Workflows

Preparing Raw Data

Basic Workflow Designs

Biomedical Workflows

Genome Finishing Workflows

Microbial Workflows

Single Cell Workflows

Installed Workflows

Installed Workflows

Tools

Microarray

Genome Finishing Module

Classical Sequence Analysis

Molecular Biology Tools

BLAST

Prepare Sequencing Data

Quality Control

Resequencing Analysis

Single Cell Analysis

RNA-Seq and Small RNA Analysis

Microarray Analysis

Epigenomics Analysis

De Novo Sequencing

Idle

Rows: 2

Name	Sequence	Reads	When an adapter is found	For reads without adapters
66f	AATGATACGGCGACACCGAGATCTACACTCTTTCCCTACACGAGGCTCTCCOGATCT	All	Trim 5' end	Keep the read
86f	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC	All	Trim 3' end	Keep the read

Trim Adapter List Editor Settings

Column width Automot...

Show column

Name

Sequence

Reads

When an adapter is found

For reads without adapters

Select All

Deselect All

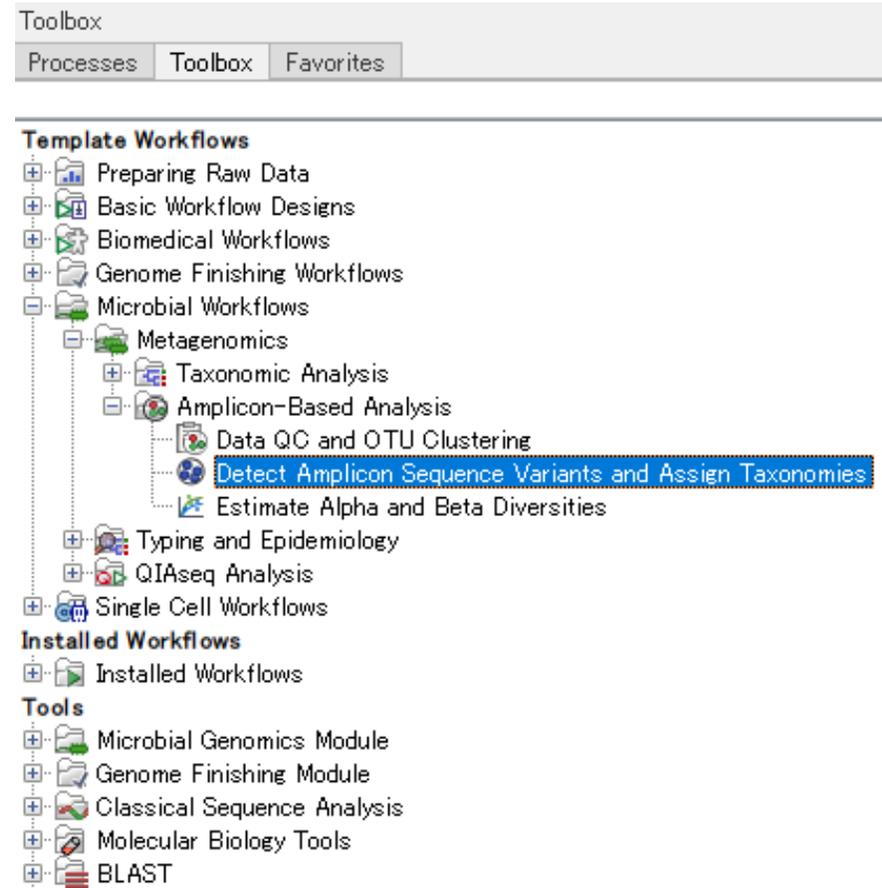
Add Row Edit Row Delete Row

Help Save View...

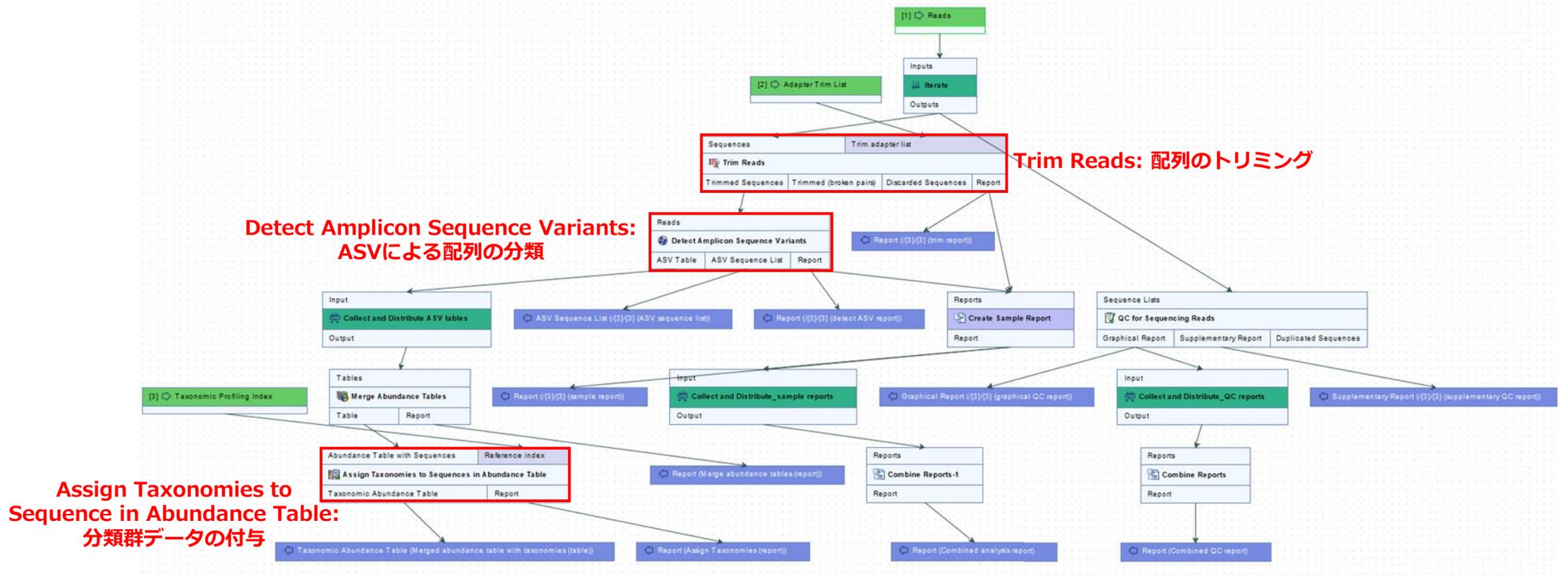
1 element(s) are selected

“Add Row”をクリックし、プライマーの配列および
プライマーの位置（5'側、3'側）を一つずつ登録します。

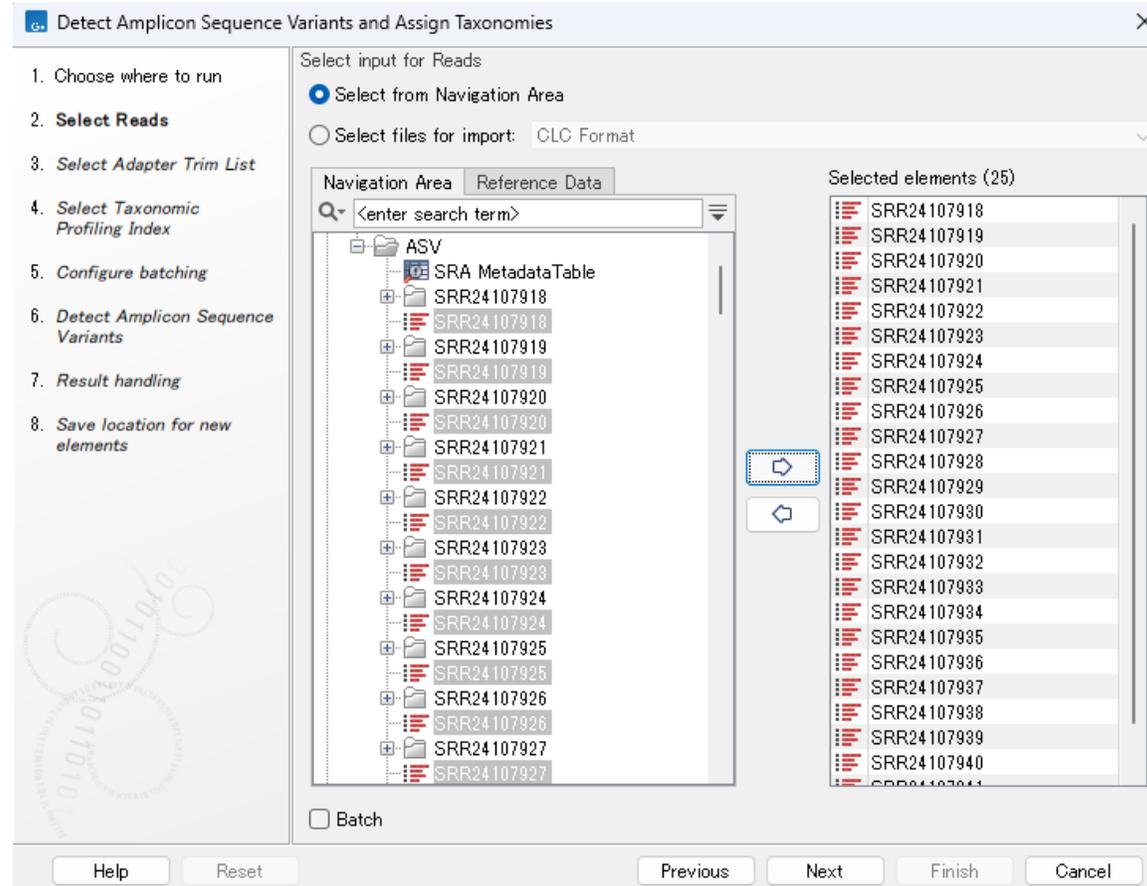
これで準備は終了です。



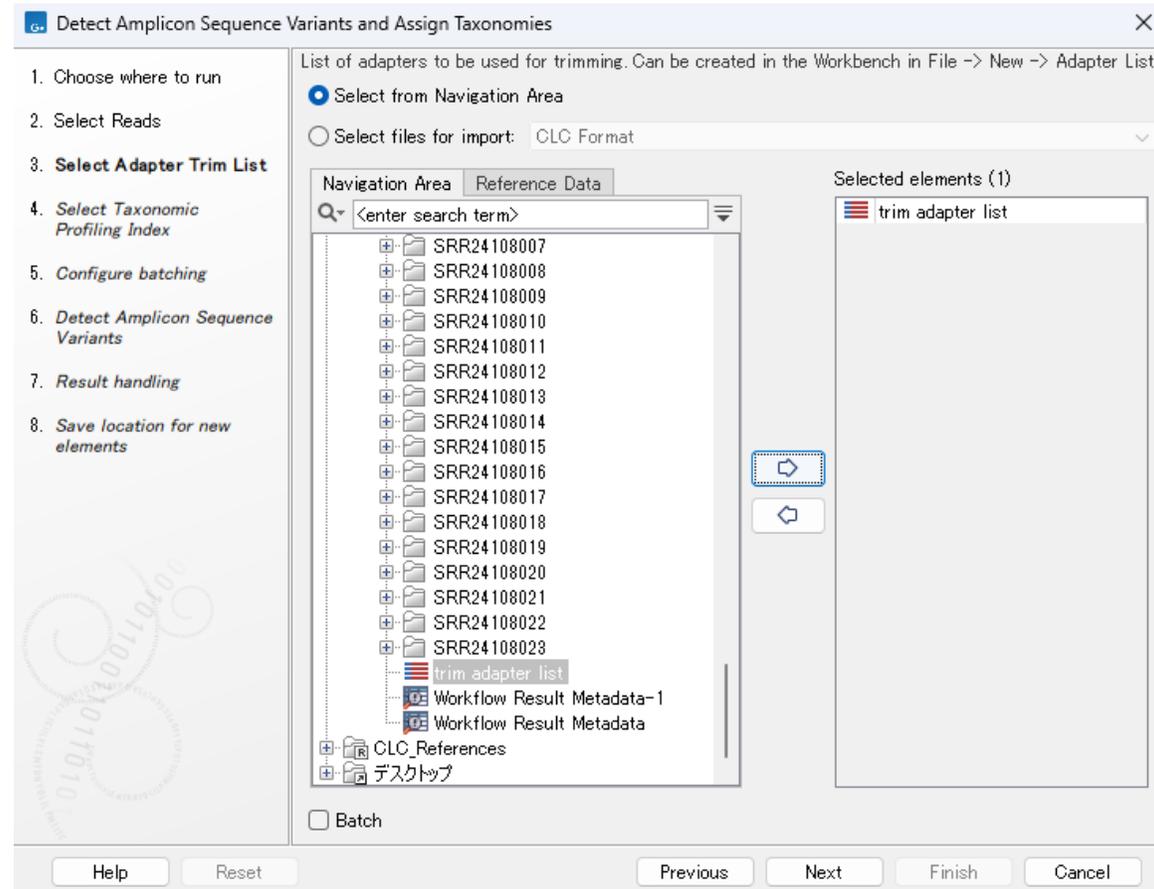
今回は、上記ワークフローを利用し、解析を行います。



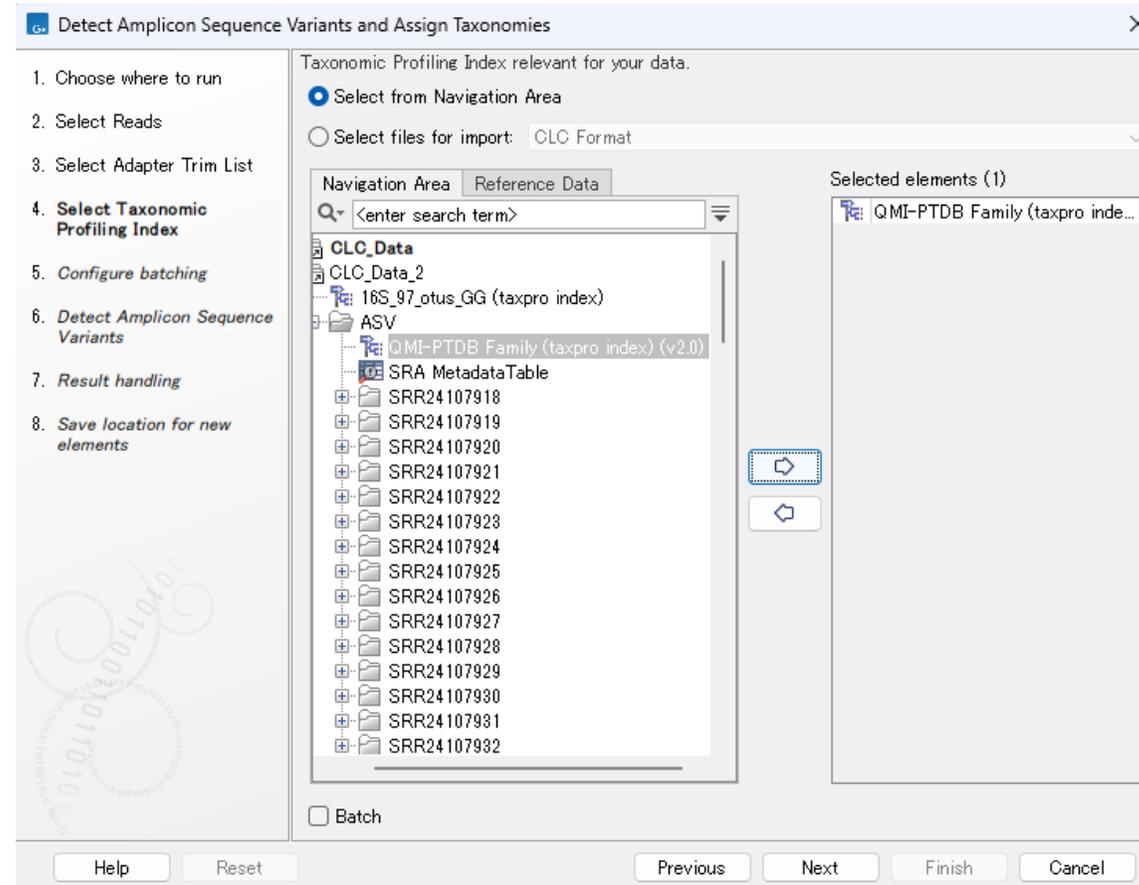
配列のトリミング、QC、ASV解析、分類群データの付与などをまとめて実行できます。



ワークフローを起動し、配列データをすべて指定します。



先ほど用意したアダプターリストを指定します。



ダウンロードしたデータベース（インデックス形式）を指定します。

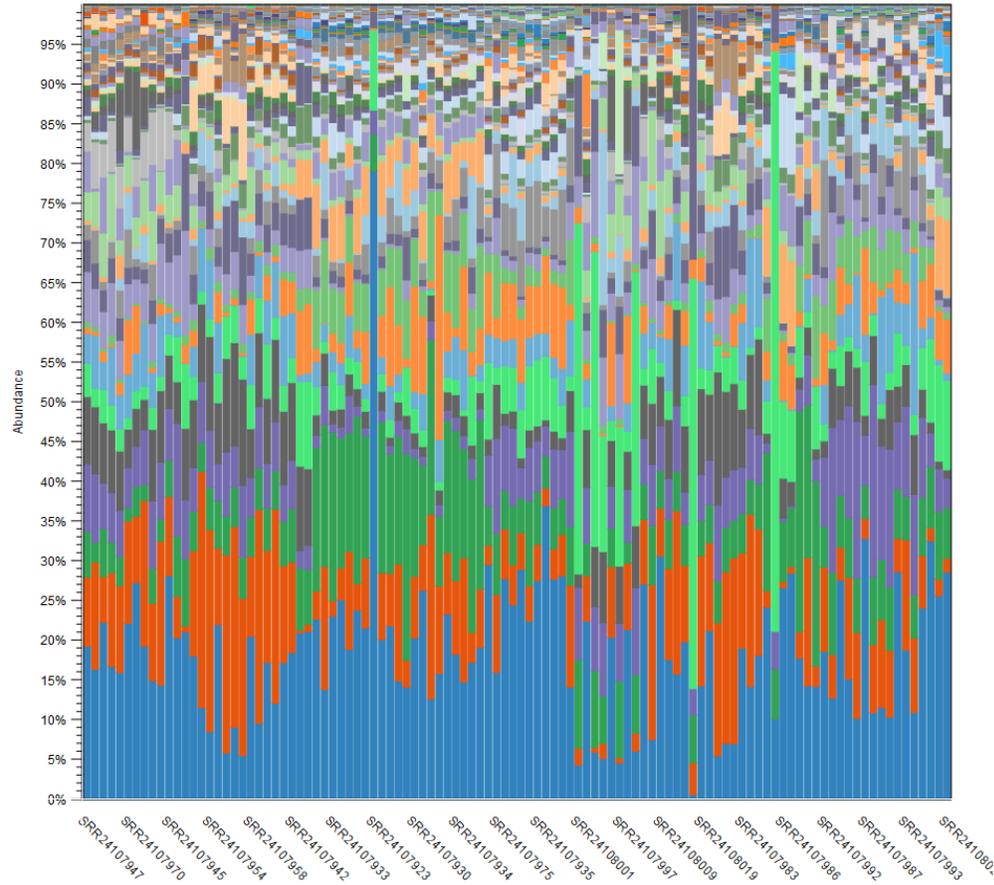
ワークフローの出力

Merged abundance table with taxonomies (table)

The screenshot displays a software interface for handling biological data. At the top, a status bar indicates 'Rows: 57,990'. Below this is a table with two columns: 'Name' and 'Taxonomy'. The 'Name' column contains long alphanumeric strings, while the 'Taxonomy' column contains hierarchical biological classifications such as 'Bacteria; Acidobacteriota; Acidobacteriae; Bryobacteriales; Bryobacteraceae; Bryobacter; Bryobacter aggregatus'. To the right of the table is a 'Table settings' panel with a 'Column width' dropdown set to 'Manual' and a 'Show column' list. The 'Show column' list includes options for 'ID', 'Name', 'Taxonomy', 'Combined Abundance', 'Min', 'Max', 'Mean', 'Median', 'Std', and a series of 'SRR24' abundance values (e.g., SRR24107947 Abundance). At the bottom of the interface, there is a toolbar with buttons for 'Create Abundance Subtable', 'Create Normalized Abundance Subtable', 'Create Sequence Sublist', and 'Extract Reads from Selection'.

データに含まれる分類群に関するリストが作成されます。

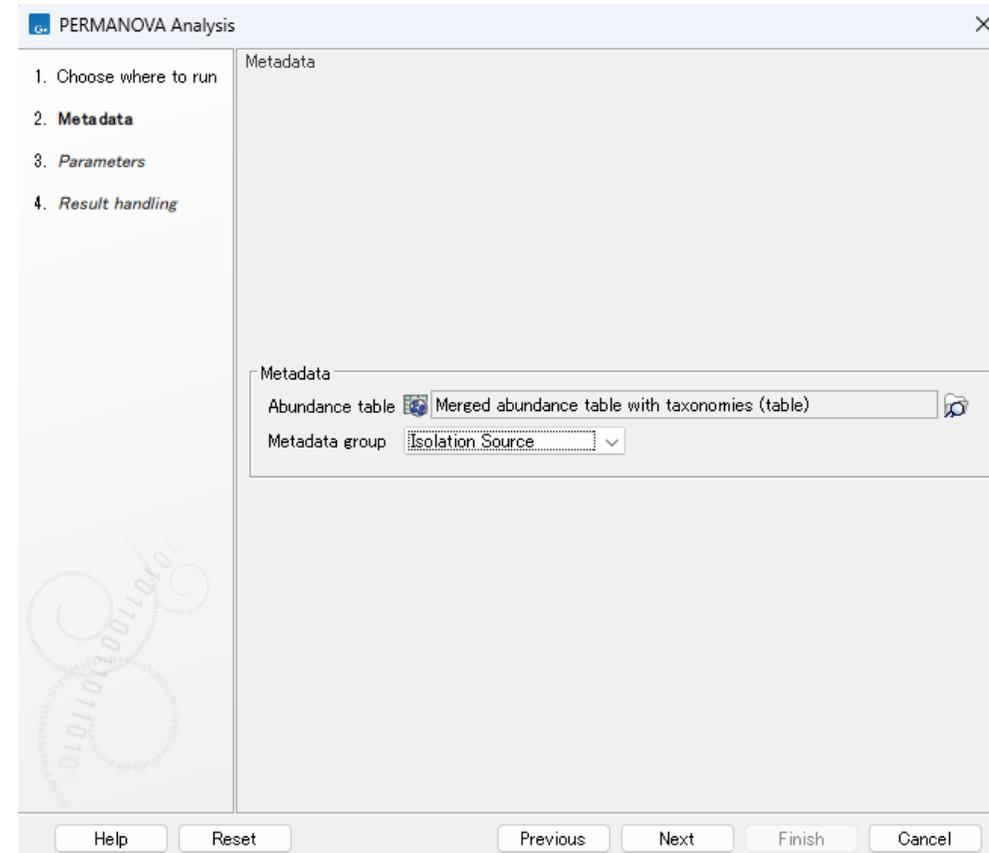
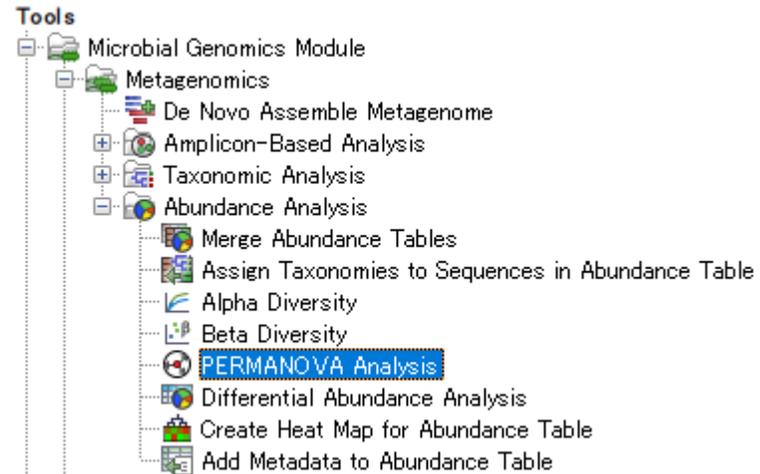
網レベルのAbundance



門レベルのAbundance



このデータから、バーチャートやサンバースト図の表示も可能です。
バーチャートでは、分類群のレベルによって表示を切り替えることも可能です。



PERMANOVA Analysisツールで先ほどのデータを選択し、比較を行う基準を指定します。
今回はサンプルが由来する環境（Isolation Source）間での比較を行います。

PERMANOVA Analysis

1. Choose where to run
2. Metadata
3. **Parameters**
4. Result handling

Parameters

Beta diversity measures

- Bray-Curtis
- Jaccard
- Euclidean

Phylogenetic diversity

Phylogenetic tree: Merged abundance table with taxonomies (table) (Filtered) alignment_tree

- Unweighted UniFrac
- Weighted UniFrac
- Weighted UniFrac not normalized
- D_0 UniFrac
- D_0.5 UniFrac

PERMANOVA parameters

Number of permutations: 99,999

Help Reset Previous Next Finish Cancel

計算方法および系統樹データを指定します。
系統樹データは、先のワークフローで出力されるデータに含まれています。

PERMANOVA解析の結果

1 PERMANOVA analysis (Bray-Curtis)

Variable	Groups	Pseudo-f statistic	p-value
Isolation Source	Desert Soil, forest soil, swamp soil, Riparian soil	24.96215	0.00001

Group 1	Group 2	Pseudo-f statistic	p-value	p-value (Bonferroni)
Desert Soil	forest soil	21.62626	0.00001	0.00006
Desert Soil	swamp soil	24.95269	0.00001	0.00006
forest soil	swamp soil	26.27344	0.00001	0.00006
Desert Soil	Riparian soil	24.33763	0.00001	0.00006
forest soil	Riparian soil	25.90952	0.00001	0.00006
swamp soil	Riparian soil	26.31809	0.00001	0.00006

2 PERMANOVA analysis (Jaccard)

Variable	Groups	Pseudo-f statistic	p-value
Isolation Source	Desert Soil, forest soil, swamp soil, Riparian soil	13.74230	0.00001

Group 1	Group 2	Pseudo-f statistic	p-value	p-value (Bonferroni)
Desert Soil	forest soil	12.14557	0.00001	0.00006
Desert Soil	swamp soil	13.31787	0.00001	0.00006
forest soil	swamp soil	14.44737	0.00001	0.00006
Desert Soil	Riparian soil	13.32328	0.00001	0.00006
forest soil	Riparian soil	14.55890	0.00001	0.00006
swamp soil	Riparian soil	14.51638	0.00001	0.00006

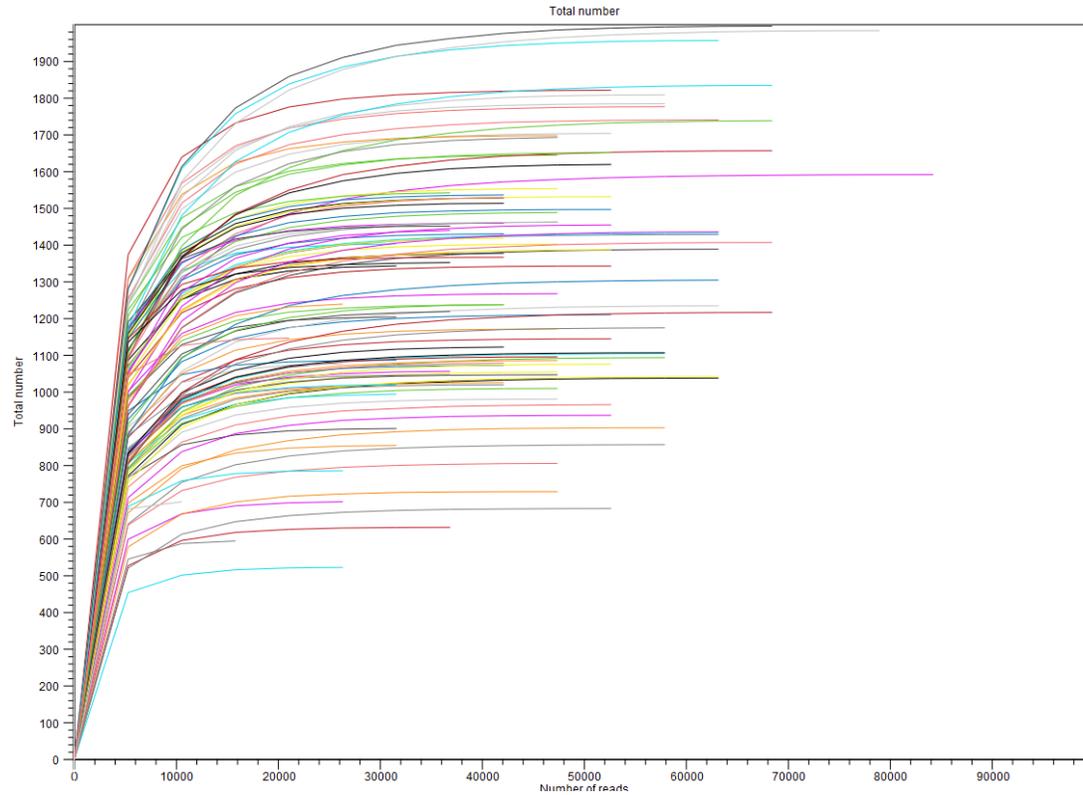
3 PERMANOVA analysis (Unweighted UniFrac)

Variable	Groups	Pseudo-f statistic	p-value
Isolation Source	Desert Soil, forest soil, swamp soil, Riparian soil	74.60272	0.00001

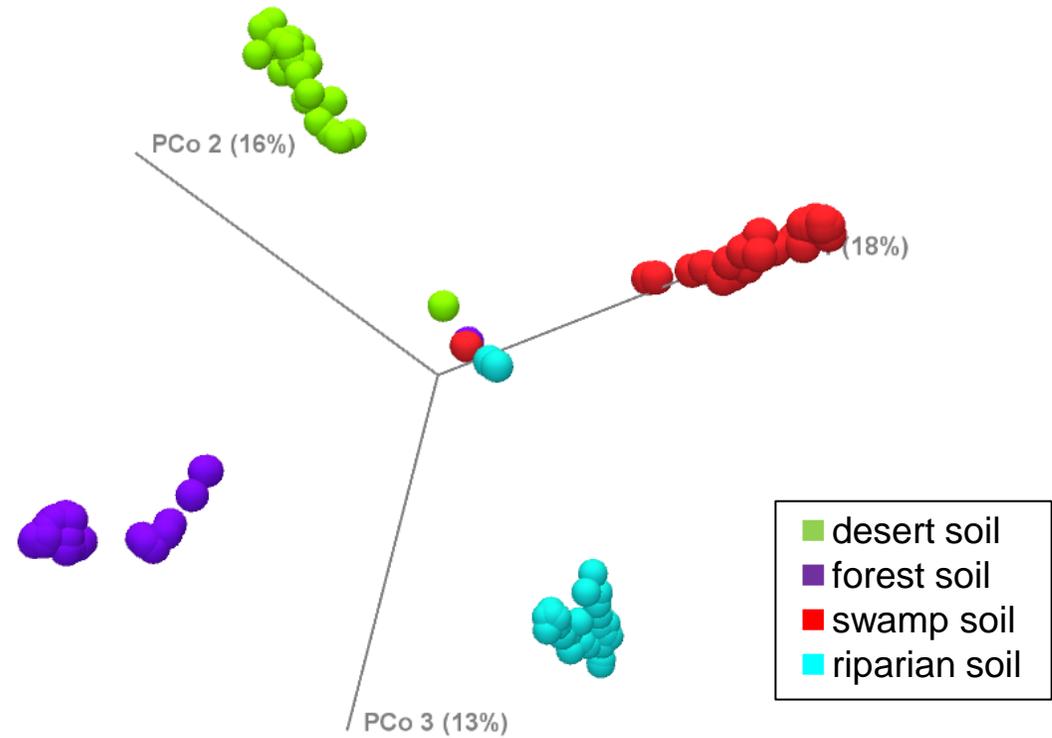
Group 1	Group 2	Pseudo-f statistic	p-value	p-value (Bonferroni)
Desert Soil	forest soil	22.44454	0.00001	0.00006
Desert Soil	swamp soil	63.63533	0.00001	0.00006
forest soil	swamp soil	60.87901	0.00001	0.00006
Desert Soil	Riparian soil	88.31102	0.00001	0.00006
forest soil	Riparian soil	110.04247	0.00001	0.00006
swamp soil	Riparian soil	101.71093	0.00001	0.00006

統計検定の結果が、レポート形式で出力されます。

アルファ多様性



ベータ多様性



また、ワークフローで出力されたデータを使用し、アルファ多様性、ベータ多様性に関するプロットもすぐに作成可能です。

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