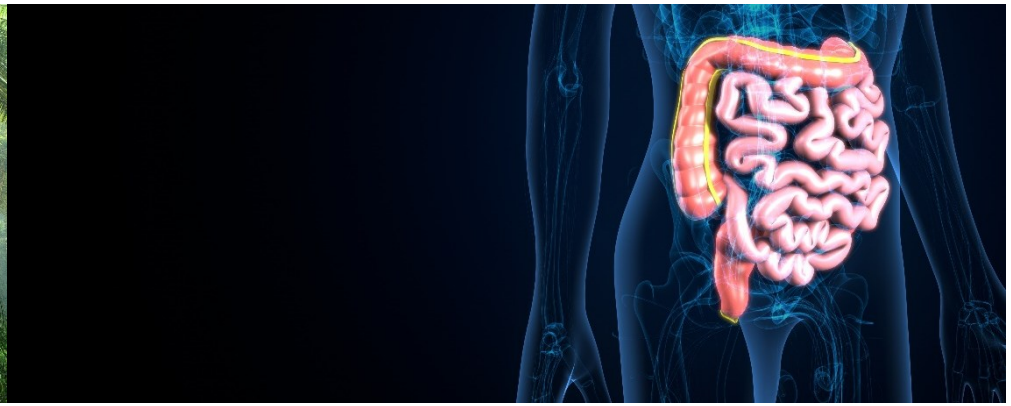
生体試料と環境試料の菌叢解析

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)

2024.06

メタゲノム解析のアプローチ



環境（土壌、水など）

生体試料（口腔、腸内、皮膚など）

環境間 or 生体試料間の菌叢の違いは？

アンプリコン シーケンス (16S/ITS)

- シーケンスコスト低
- ゲノム内に複数の rRNA 遺伝子コピーを持っているため、種の定量結果が不鮮明

全ゲノム シーケンス (WGS)

- シーケンスコスト高
- アンプリコンシーケンスより高い精度、菌叢のもつ機能推定も可能

解析ワークフロー

QC・トリミング

- NGSより出力された生データが良好か、下流分析に影響する問題がないか確認。
  de novo アセンブリに有用な高品質なリードデータが得られる。
  FastQCとTrimmomaticツールを統合

宿主ゲノムの除去

- 宿主ゲノムがデータのノイズとなる実験の場合は必要に応じて行う宿主ゲノムを除去してクリーンなデータを作成する。
  Bowie2、Samtools、および Qualimapツールを統合

Taxonomic Classification

- DNA ショートリードデータに対しに分類学的ラベルを割り当てることで菌叢解析を行う。
  WGS および 16S/ITS アンプリコンデータに対応。
  Kraken2ツールを統合

Differential Abundance Analysis

- 統計的にサンプルまたは条件間の菌叢の違いを検出。

16S,全ゲノムを同じ手順で解析できます。

Omicsbox

- 解析環境の構築や高スペックPCや導入が必要
- コマンドライン型のツール：パラメータ設定の操作が煩雑

OmicsBoxの菌叢解析

- メーカーのサーバーで高速計算　高価なPCの購入は不要
- マウス操作で簡単に解析
- 16S,全ゲノムを同じ手順で解析できます。

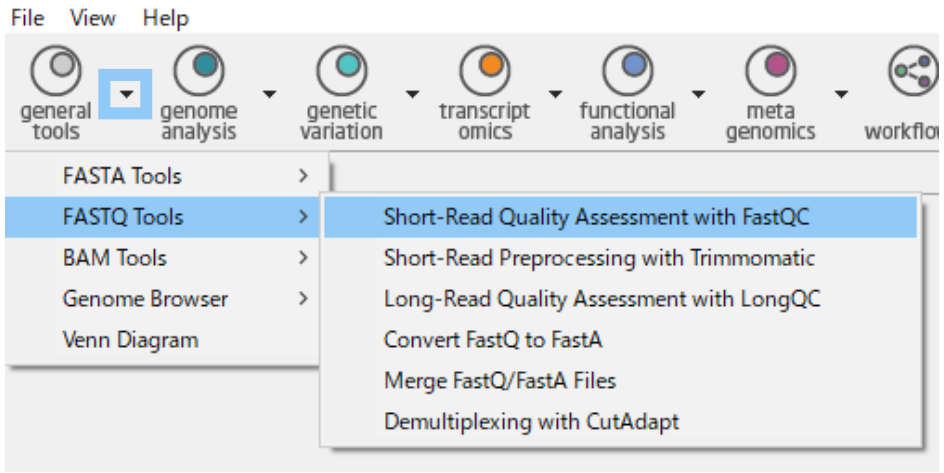
必要なファイル

Fastq

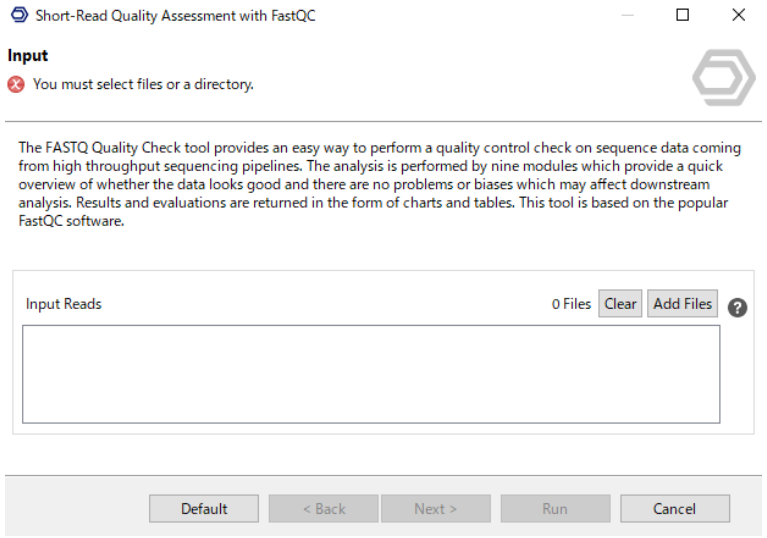
シーケンサーから出力された生データ
もしくは受託サービスで得られたクリーンリードデータ

QC

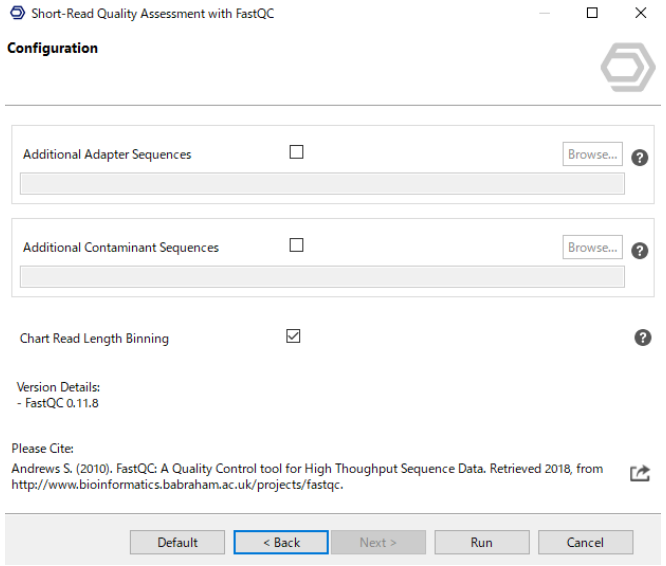
・データが良好か、下流分析に影響する問題がないか確認



画面上部のアイコンからQCツールを起動



全てのサンプルをこちらに指定



任意でアダプタリストを追加

結果

Welcome Message | FASTQ Quality Check (Dataset) | FASTQ Quality Check (ERR1948631_1.fastq) | FASTQ Quality Check (clean_ERR1948631_1.fq) | *Chart: Adapter Conten

FASTQ Quality Check

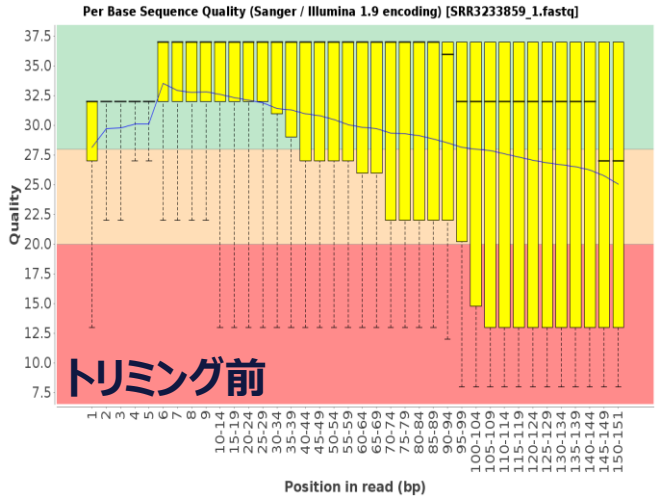
Name: Dataset

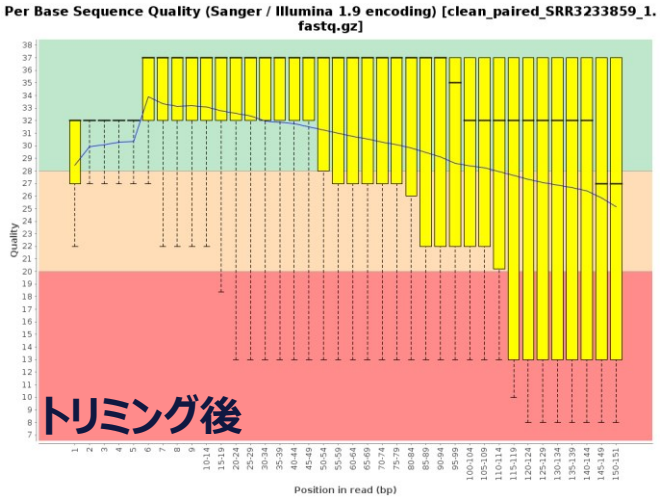
Overall Results

Name	Per Base Sequence Quality	Per Sequence Quality Scores	Per Base Sequence Content	Per Sequence GC Content	Per Base N Content
ERR1948631_1.fastq	PASS	PASS	FAIL	PASS	PASS
clean_ERR1948631_1.fq	PASS	PASS	FAIL	PASS	PASS

Name	Sequence Length Distribution	Adapter Content	Overrepresented Sequences	Sequence Duplication Levels	Report
ERR1948631_1.fastq	PASS	FAIL	WARNING	FAIL	
clean_ERR1948631_1.fq	WARNING	PASS	WARNING	FAIL	

The FASTQ quality check task is performed by nine analysis modules. The table above provides a quick evaluation of whether the results of each module seem entirely normal (pass), sightly abnormal (warning) or very unusual (fail). Note that these evaluations must be taken in the context of what is expected from the library. For example, some experiments may be expected to produce libraries which are biased in particular ways. Therefore, the summary evaluations should be treated as pointers that guide the preprocessing of the libraries.





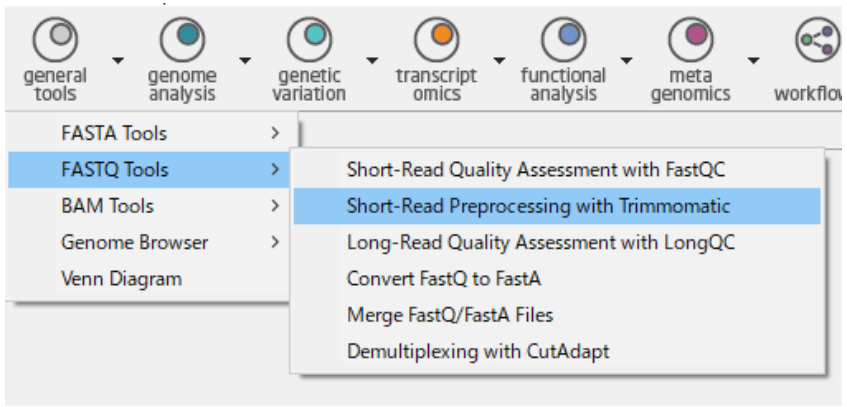
✓ 解析が終了するとレポートが作成

正常（PASS）
わずかに異常（WARNING）
異常（FAIL）

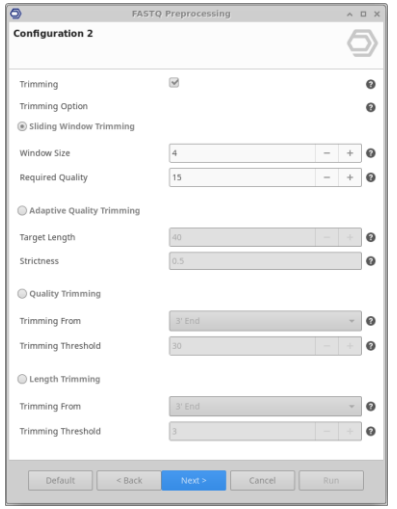
シーケンスデータの品質をすばやく評価

レポートのアイコンをクリック→さらに詳細な結果を見ることが可能

トリミング

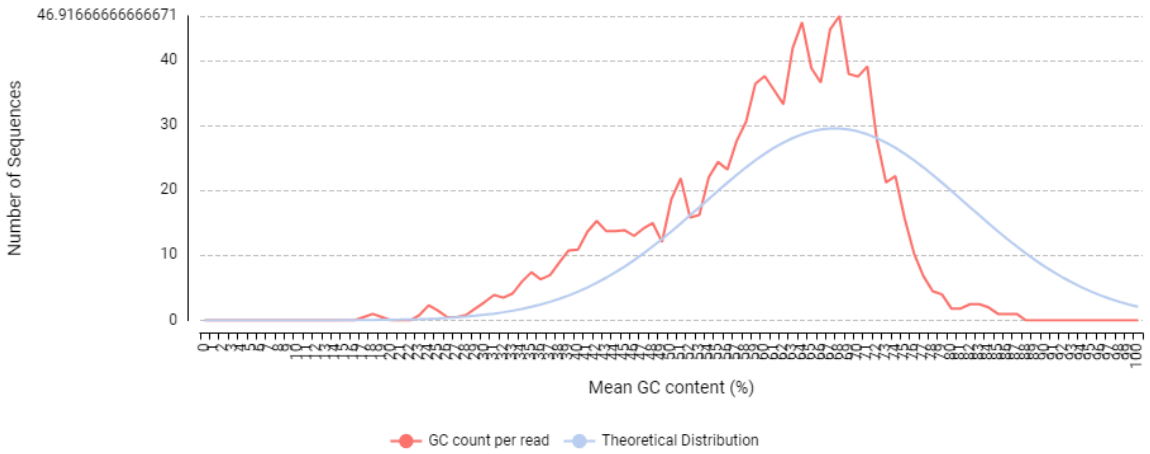


QC後必要に応じてトリミングを行う



様々なトリミング方法
トリミング後再度QCを行う

【補足】必ずしもすべてのQC項目を正常（PASS）にする必要はありません。



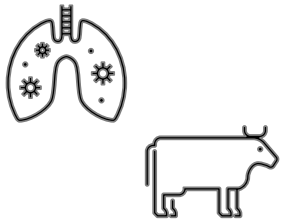
FalseやWarningと判定されていても、それを説明できる理由があれば、適切なリードであると判断することが多いです。

例）

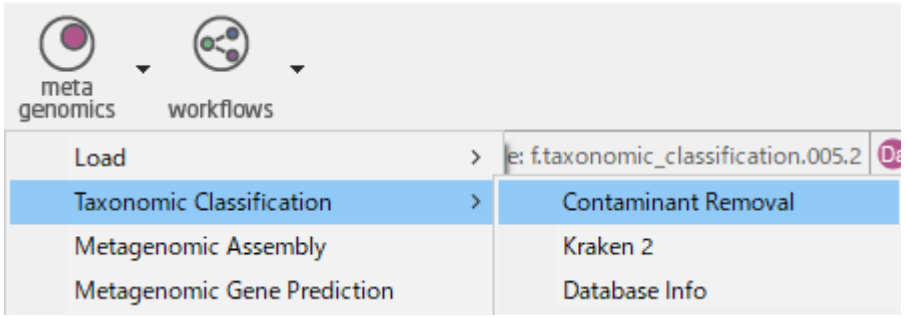
左の結果ではGC ContentがFalseですが、

- メタゲノムでは多様な種が含まれるため正規分布から外れた形状は想定されます。

宿主ゲノムの除去

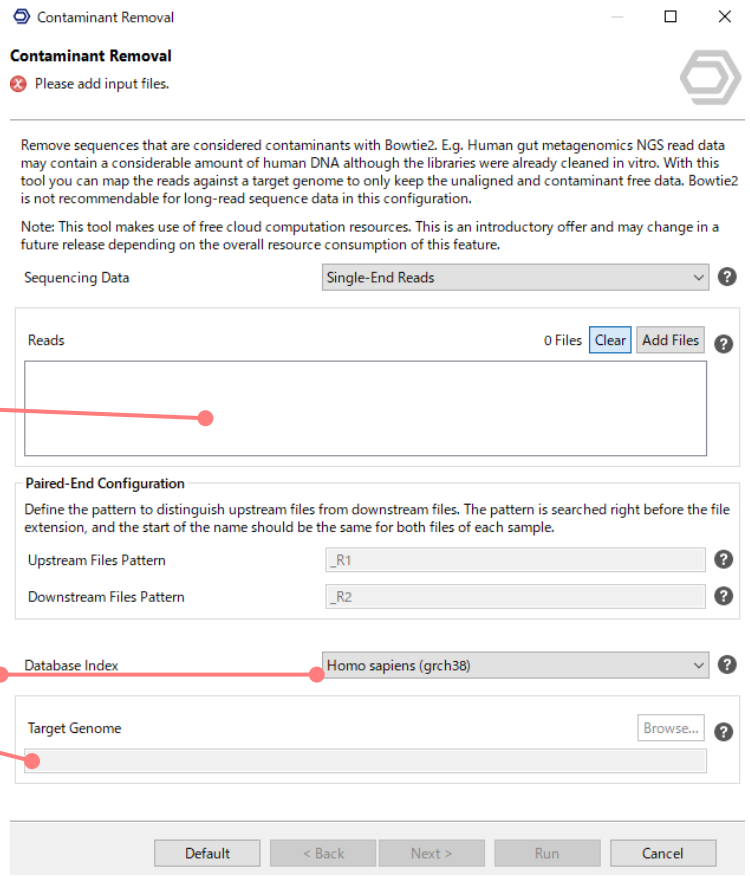


宿主DNAを含むリードは Kraken2では分類できず、データセットのノイズが増加する原因となることがあります。
リードデータを宿主ゲノムにマッピングすることで、未分類のリードの数を減らすことができます。
このプロセスを繰り返して、(系統発生的に近いターゲットゲノムなどを使用して) データを改良することができます。



①Contaminant Removalを選択

②リードデータを入力



③宿主ゲノムの生物種を指定
リファレンスゲノムデータを
入力することも可能

④保存先の指定

宿主ゲノムの除去

Alignment Overview

Global

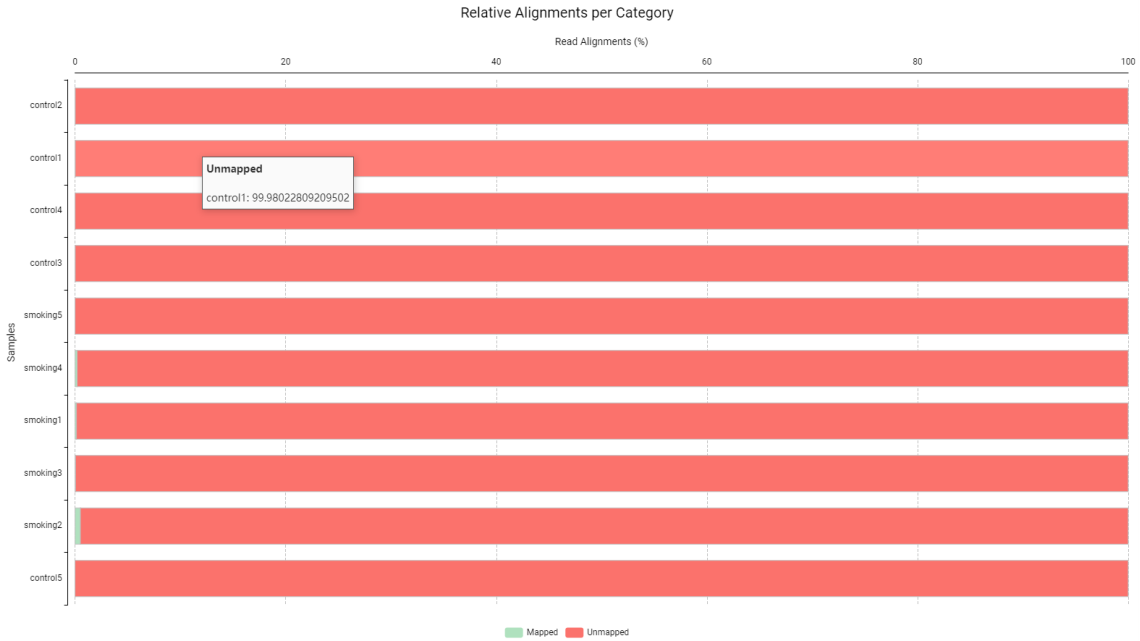
Sample	Total Alignments	Mapped / Contaminants	Unmapped / Contaminant-Free	Duplicated Reads (estimated)	Duplication Rate
control2	185,179	35 / 0.019%	185,144 / 99.981%	1 / 0.001%	2.94
control1	237,711	47 / 0.02%	237,664 / 99.98%	3 / 0.001%	6.82
control4	237,480	16 / 0.007%	237,464 / 99.993%	0	0
control3	132,728	23 / 0.017%	132,705 / 99.983%	0	0
smoking5	33,077	1 / 0.003%	33,076 / 99.997%	0	0
smoking4	9,810	20 / 0.204%	9,790 / 99.796%	3 / 0.031%	17.65
smoking1	29,645	41 / 0.138%	29,604 / 99.862%	14 / 0.047%	29.63
smoking3	25,220	10 / 0.04%	25,210 / 99.96%	0	0
smoking2	8,297	42 / 0.506%	8,255 / 99.494%	3 / 0.036%	2.56
control5	234,046	1 / 0%	234,045 / 100%	0	0

ACTG Content

Sample	A's	C's	T's	G's	N's	GC (%)
control2	1,158 / 27.697%	954 / 22.818%	1,097 / 26.238%	972 / 23.248%	0	46.07
control1	1,607 / 25.919%	1,412 / 22.774%	1,678 / 27.065%	1,503 / 24.242%	0	47.02
control4	614 / 28.088%	530 / 24.245%	495 / 22.644%	547 / 25.023%	0	49.27
control3	615 / 25.445%	494 / 20.439%	704 / 29.127%	604 / 24.99%	0	45.43
smoking5	51 / 27.273%	51 / 27.273%	34 / 18.182%	51 / 27.273%	0	54.55
smoking4	811 / 26.4%	709 / 23.079%	719 / 23.405%	833 / 27.116%	0	50.2
smoking1	1,685 / 25.871%	1,641 / 25.196%	1,532 / 23.522%	1,655 / 25.411%	0	50.61
smoking3	343 / 24.189%	368 / 25.952%	351 / 24.753%	356 / 25.106%	0	51.06
smoking2	1,614 / 23.655%	1,730 / 25.355%	1,716 / 25.15%	1,763 / 25.839%	0	51.19
control5	20 / 25.974%	19 / 24.675%	11 / 14.286%	27 / 35.065%	0	59.74

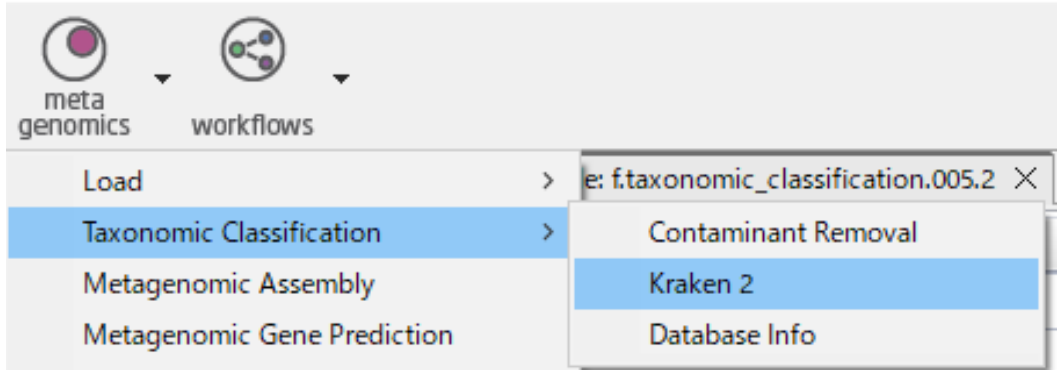
Coverage

Sample	Mean	Standard Deviation
control2	0X	0.001X
control1	0X	0.002X

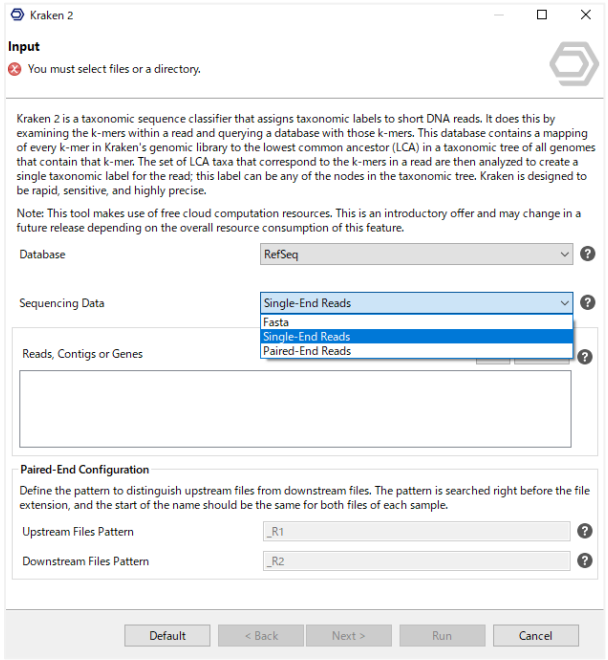


宿主ゲノムが除去されたリードデータの他
リード中に含まれた宿主ゲノムの割合などを含むレポートやグラフが作成される

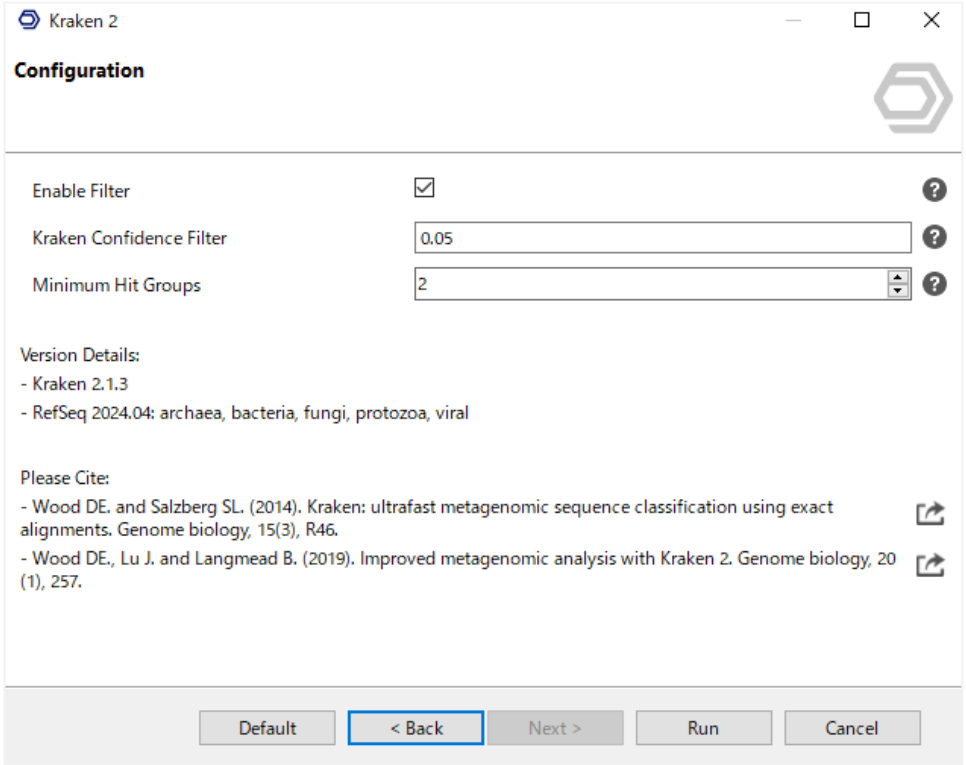
菌叢解析



①Kraken2を選択



②リードデータを入力



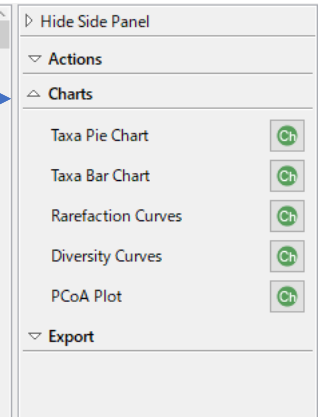
③任意で設定（セミナーではデフォルト設定）
Runボタンをクリックすると解析が開始される

複雑な設定はなく解析を実行できます。

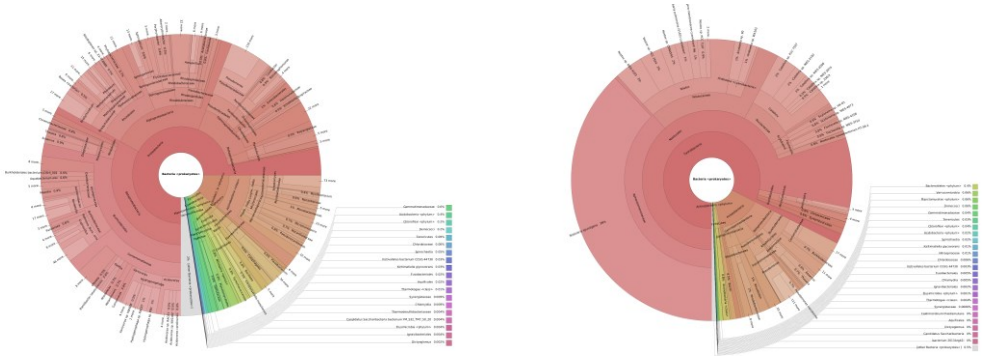
菌叢解析

Rank	TaxId	Scientific Name	f.VMB2	f.VMB3
unclassified	0	Unknown		144289
species	2654248	Mesorhizobium sp. INR15		1
species	2588711	Pelagovum pacificum		1
species	1916956	Synechococcus sp. SynAce01		0
species	2654249	Mesorhizobium sp. NBSH29		0
species	1720344	Psychrobacter sp. AntiMn-1	0	0
superkingdom	2	Bacteria <bacteria>	37059	284656
genus	131079	Limnobacter	0	0
genus	6	Azorhizobium	0	0
species	7	Azorhizobium caulinodans	0	0
species	9	Buchnera aphidicola	0	1
species	1179672	Flavobacterium sp. KBS0721	0	1

サイドパネルより様々な図を
簡単に作成することができます。

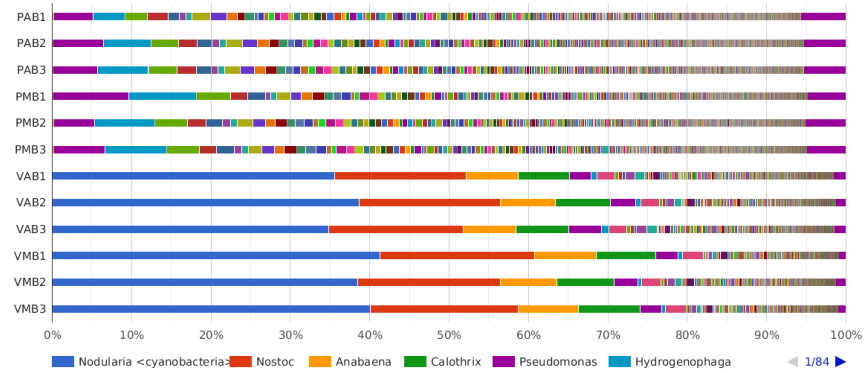


菌種同定に関する結果の表が作成されます。



Kronaチャート

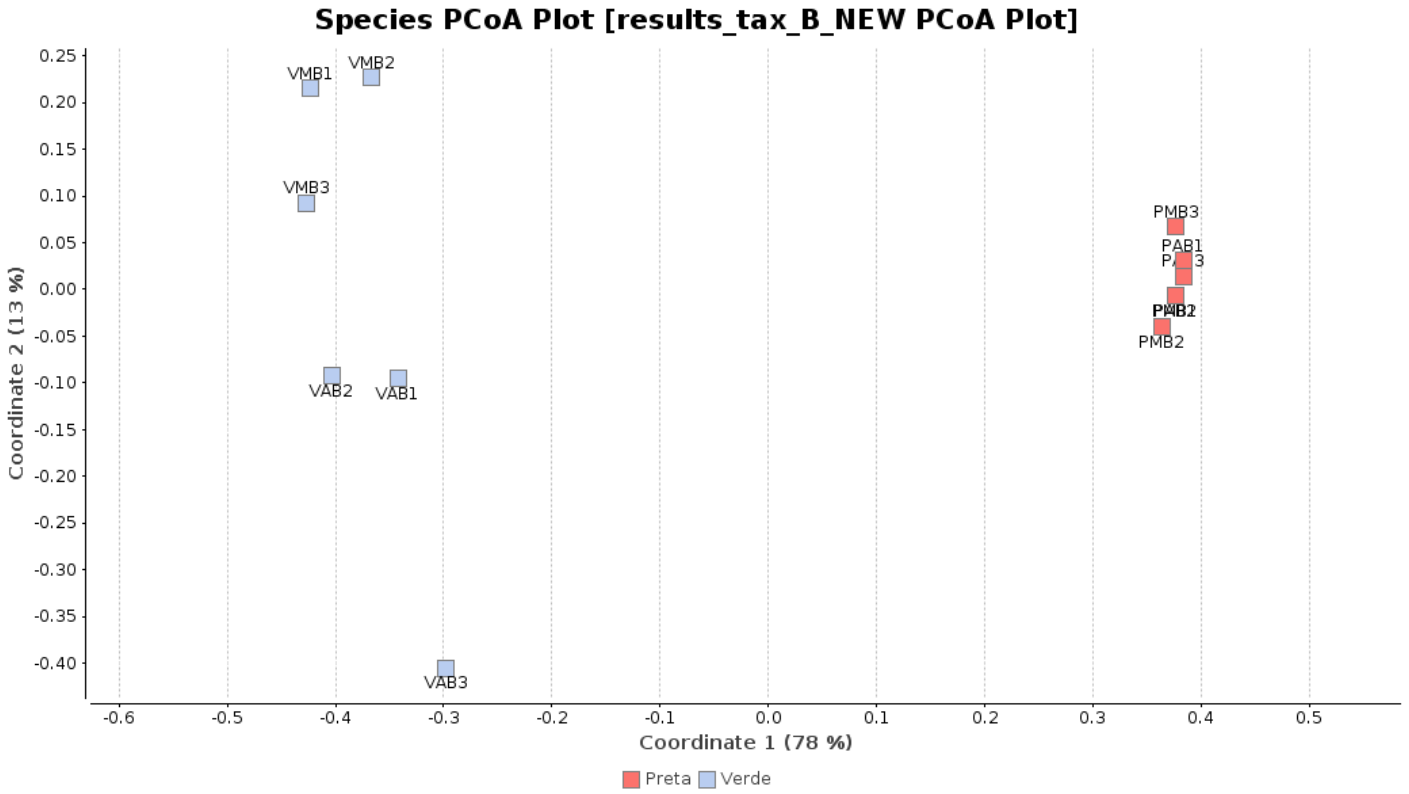
このグラフを使用すると、サンプルを相互に比較しやすくなります。



積み上げ棒グラフ

各サンプルの界から種までの菌種同定を棒グラフで見やすく表示できます。

菌叢解析



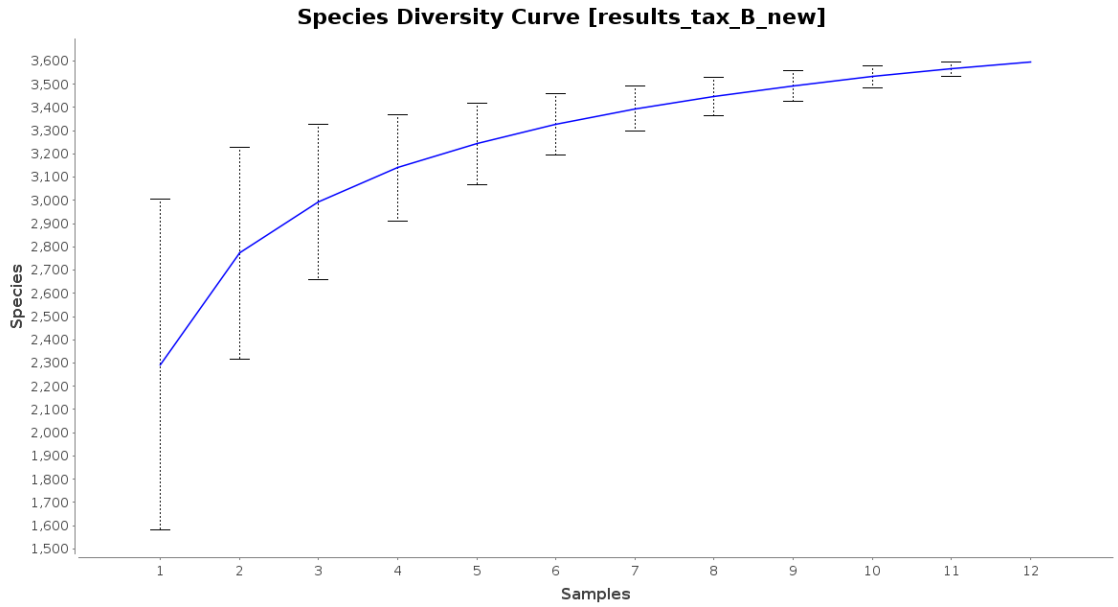
PCoAプロット

サンプル間の類似パターンを距離で表すことができます。

菌叢解析

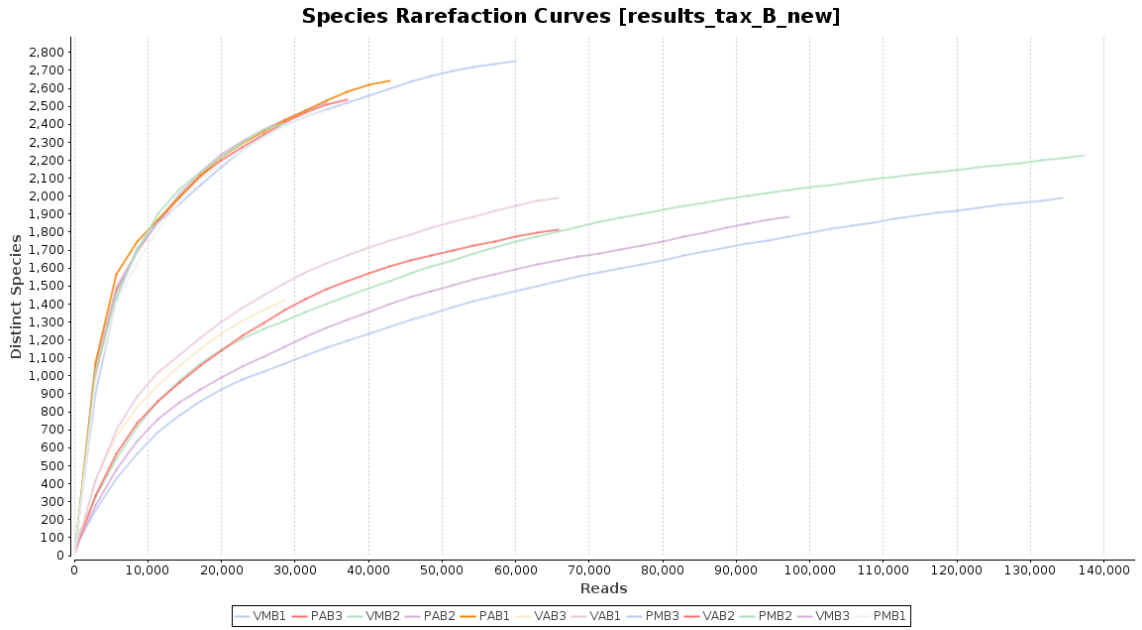
【Diversity Curve】

シーケンスのカバレッジが十分に深いかどうかを判断することができます。

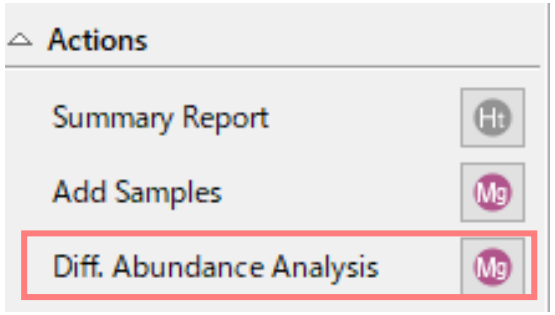


【Rarefaction Curves 】

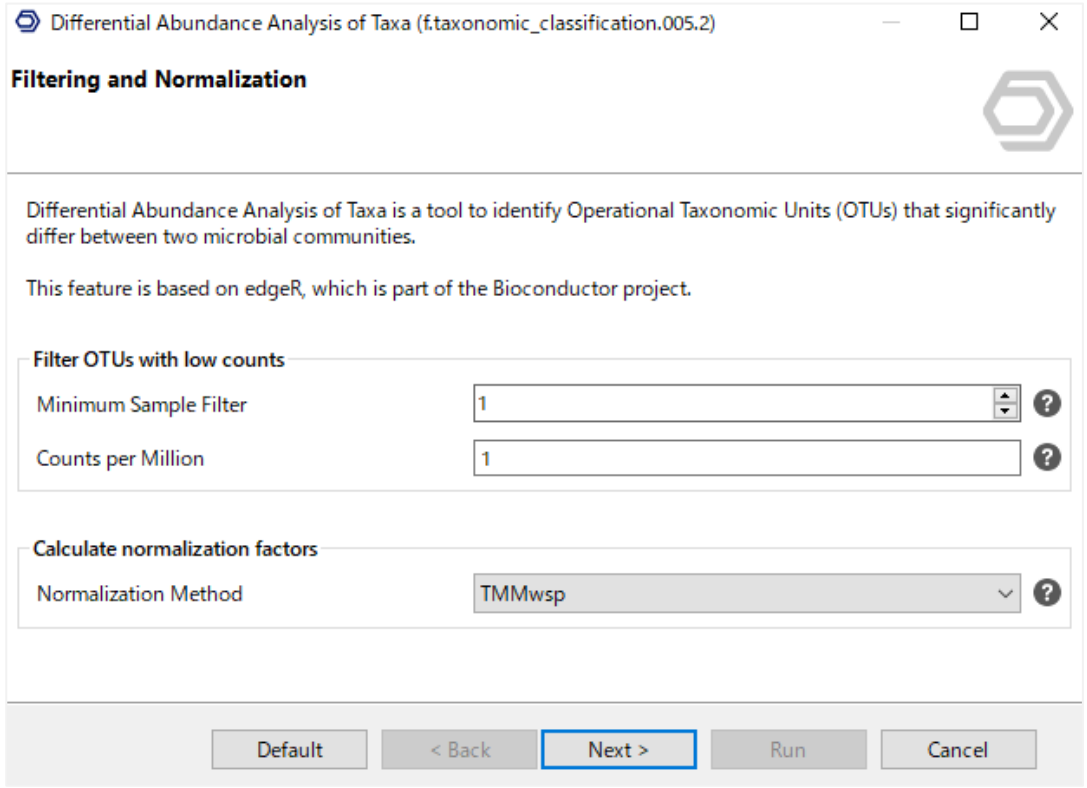
追加のサンプルをデータセットに含めることの
微生物の多様性における利点を評価することが可能です。



比較解析

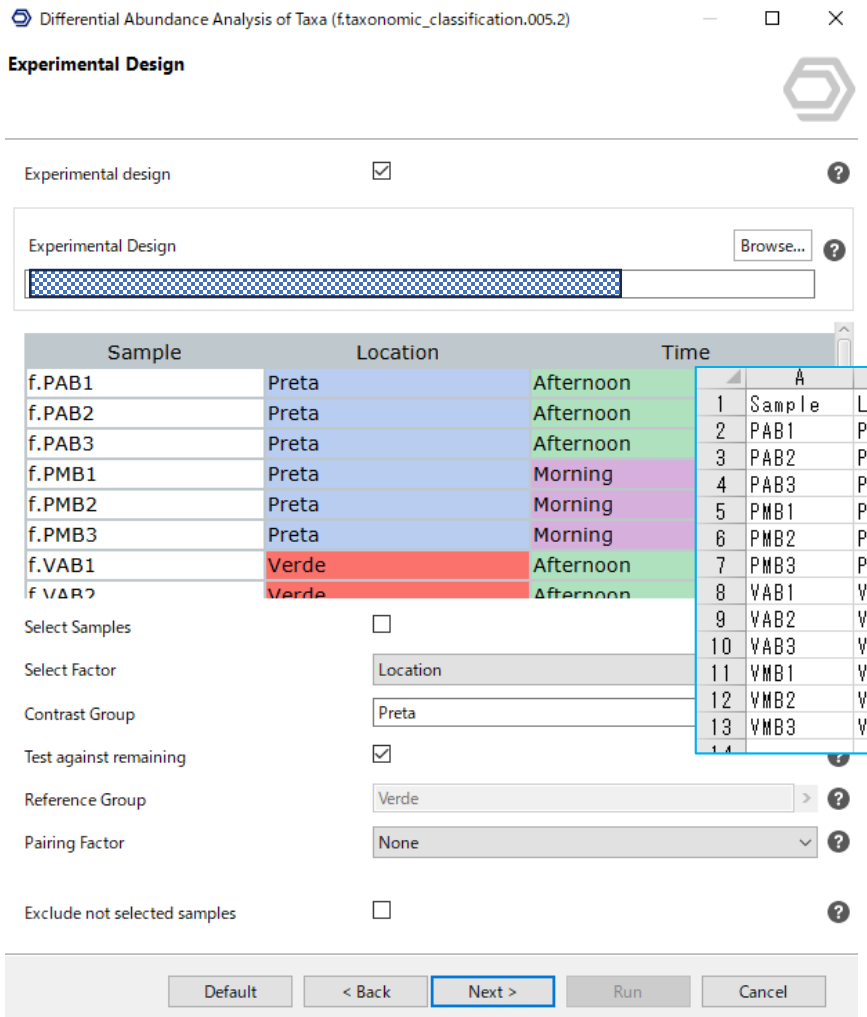


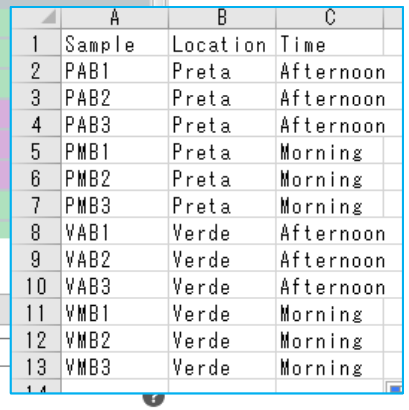
①先ほどの結果のサイドパネルから
Differential Abundance Analysisを選択



②任意でフィルターを調整

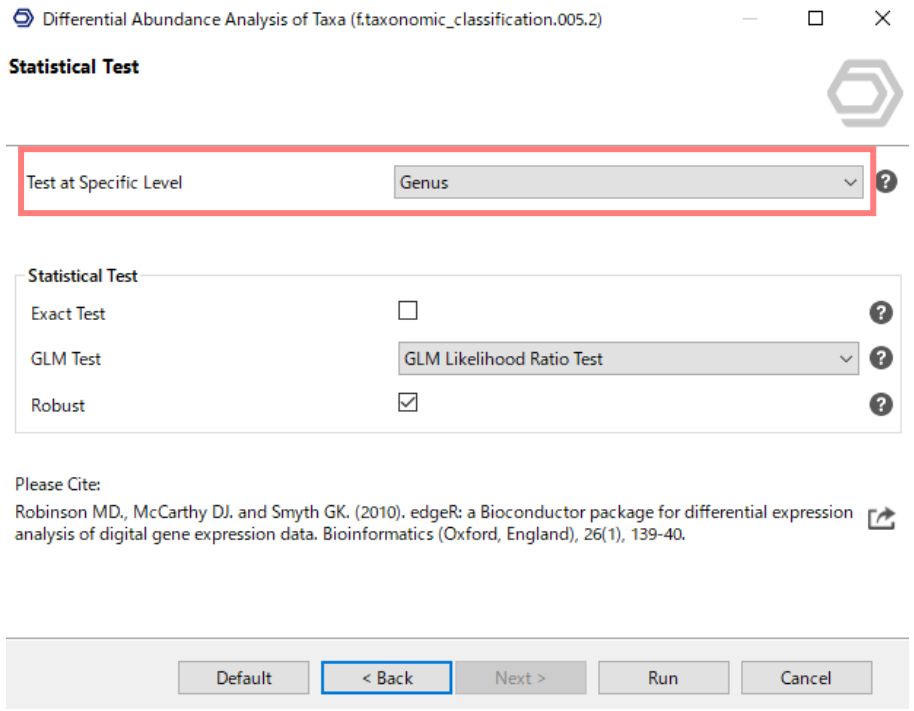
比較解析





	A	B	C
1	Sample	Location	Time
2	PAB1	Preta	Afternoon
3	PAB2	Preta	Afternoon
4	PAB3	Preta	Afternoon
5	PMB1	Preta	Morning
6	PMB2	Preta	Morning
7	PMB3	Preta	Morning
8	VAB1	Verde	Afternoon
9	VAB2	Verde	Afternoon
10	VAB3	Verde	Afternoon
11	VMB1	Verde	Morning
12	VMB2	Verde	Morning
13	VMB3	Verde	Morning

実験デザインファイルは各サンプルの属性をエクセルなどでまとめタブ区切りテキストファイルとして出力しておきます。



③実験デザインファイルを入力後
コントロール群、テスト群を指定

④どの分類階級に対して比較するか指定
Runボタンをクリックすると解析が開始される

比較解析

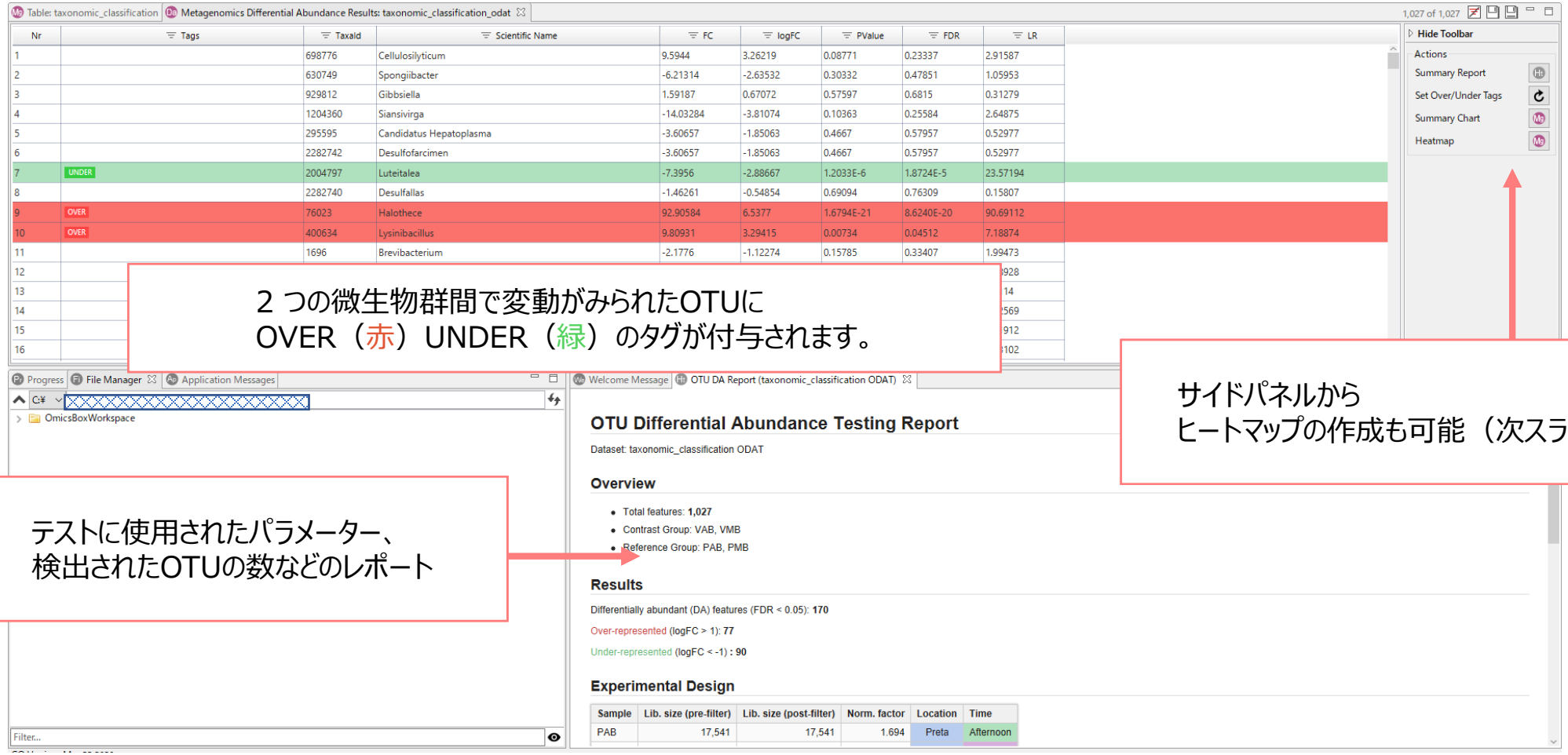
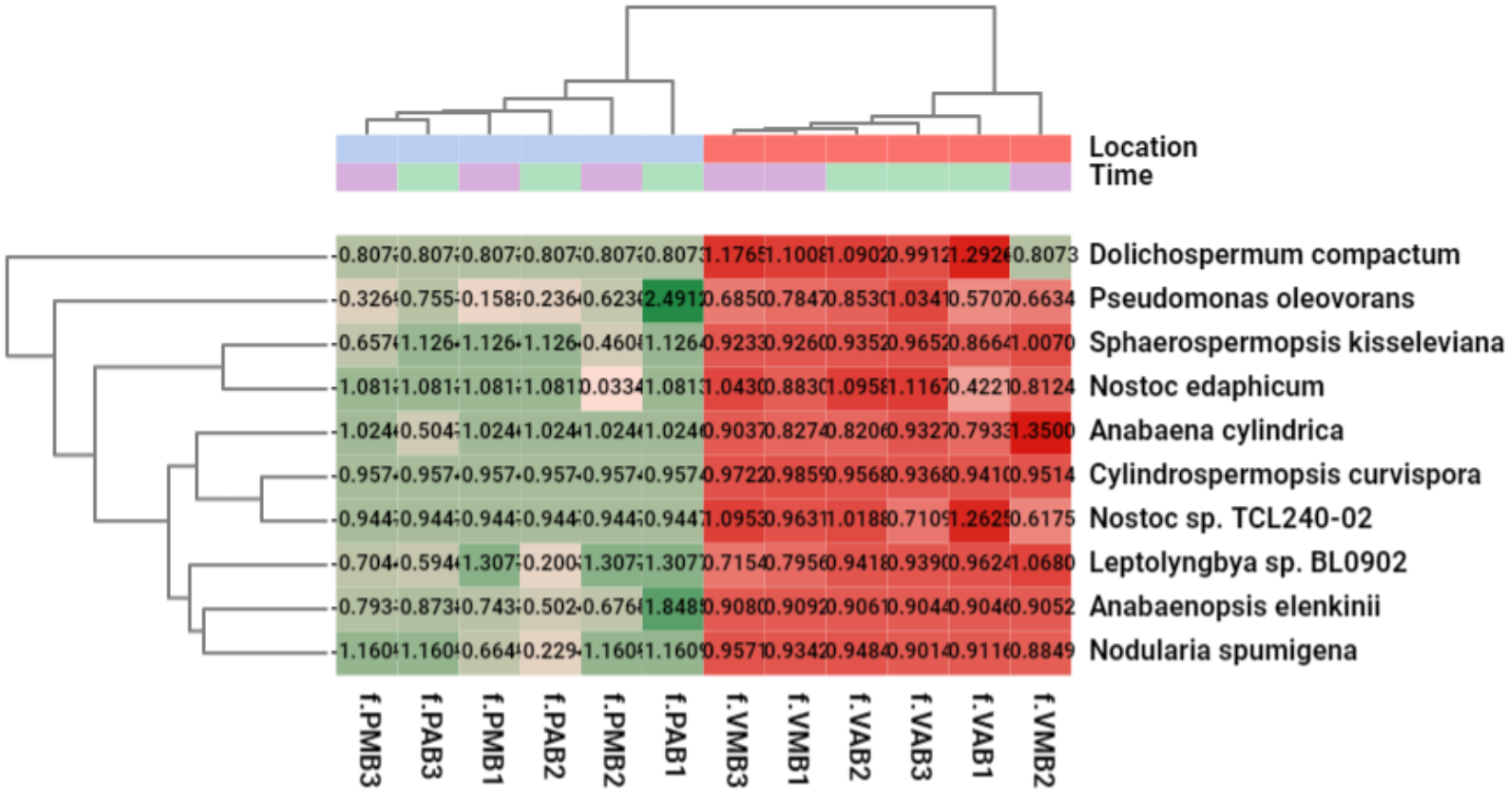


Table: taxonomic_classification
Metagenomics Differential Abundance Results: taxonomic_classification_odat
1,027 of 1,027
Nr | Tags | TaxaId | Scientific Name | FC | logFC | PValue | FDR | LR
1 | 698776 | Cellulosilyticum | 9.5944 | 3.26219 | 0.08771 | 0.23337 | 2.91587
2 | 630749 | Spongiibacter | -6.21314 | -2.63532 | 0.30332 | 0.47851 | 1.05953
3 | 929812 | Gibbsiella | 1.59187 | 0.67072 | 0.57597 | 0.6815 | 0.31279
4 | 1204360 | Siansivirga | -14.03284 | -3.81074 | 0.10363 | 0.25584 | 2.64875
5 | 295595 | Candidatus Hepatoplasma | -3.60657 | -1.85063 | 0.4667 | 0.57957 | 0.52977
6 | 2282742 | Desulfofarcimen | -3.60657 | -1.85063 | 0.4667 | 0.57957 | 0.52977
7 | UNDER | 2004797 | Luteitalea | -7.3956 | -2.88667 | 1.2033E-6 | 1.8724E-5 | 23.57194
8 | 2282740 | Desulfallas | -1.46261 | -0.54854 | 0.69094 | 0.76309 | 0.15807
9 | OVER | 76023 | Halothece | 92.90584 | 6.5377 | 1.6794E-21 | 8.6240E-20 | 90.69112
10 | OVER | 400634 | Lysinibacillus | 9.80931 | 3.29415 | 0.00734 | 0.04512 | 7.18874
11 | 1696 | Brevibacterium | -2.1776 | -1.12274 | 0.15785 | 0.33407 | 1.99473
Hide Toolbar
Actions
Summary Report
Set Over/Under Tags
Summary Chart
Heatmap
2 つの微生物群間で変動がみられたOTUに
OVER（赤）UNDER（緑）のタグが付与されます。
サイドパネルから
ヒートマップの作成も可能（次スライド）
Progress
File Manager
Application Messages
C:¥
OmicsBoxWorkspace
Filter...
Welcome Message
OTU DA Report (taxonomic_classification ODAT)
OTU Differential Abundance Testing Report
Dataset: taxonomic_classification ODAT
Overview
Total features: 1,027
Contrast Group: VAB, VMB
Reference Group: PAB, PMB
Results
Differentially abundant (DA) features (FDR < 0.05): 170
Over-represented (logFC > 1): 77
Under-represented (logFC < -1) : 90
Experimental Design
Sample | Lib. size (pre-filter) | Lib. size (post-filter) | Norm. factor | Location | Time
PAB | 17,541 | 17,541 | 1.694 | Preta | Afternoon
テストに使用されたパラメーター、
検出されたOTUの数などのレポート

菌叢解析



サイドパネルからヒートマップを作成できます。

全ゲノムメタゲノム解析の機能推定

QCトリミング

16SrRNA・全ゲノム

菌叢解析

比較解析

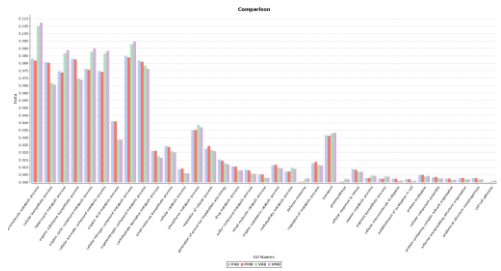
全ゲノム

メタゲノムアゼンブリ

メタゲノム遺伝子予測

機能アノテーション

比較解析

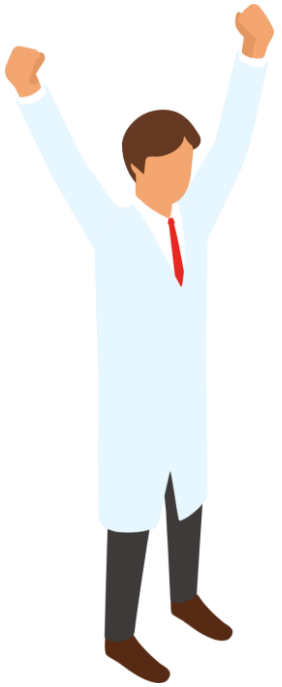


サンプル群を比較して、どのような機能アノテーションに違いがあるか作図できます。

Omicsboxでは菌叢解析だけでなく、ショットガンメタゲノムシーケンス（全ゲノム）を使用した機能推定解析もサポート

OmicsBox の菌叢解析

- 16S,全ゲノムを同じ手順で解析できます。
- 初心者でも解析できるインターフェース
- 7日間無料のデモライセンス→ 詳細(PDF)



Information

弊社HP紹介ページ

お問い合わせ先：フィルジェン株式会社

TEL 052-624-4388 (9:00～17：00)

FAX　052-624-4389

E-mail: biosupport@filgen.jp