

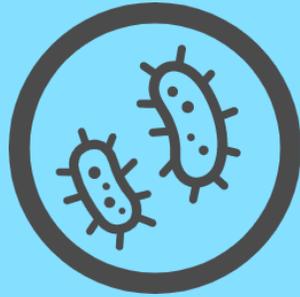


# 2つの湖の微生物叢を比較する (ショットガンメタゲノム解析)

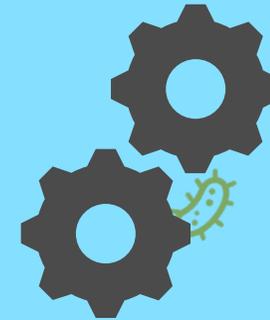
～OmicsBoxを使用した解析機能のご紹介～

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)

ショットガンメタゲノム解析は全ゲノムを使用するため  
より精度の高い菌種組成の解明だけでなく、菌叢の持つ機能を解明することも可能です。



サンプルにどのような微生物叢がみられるか？

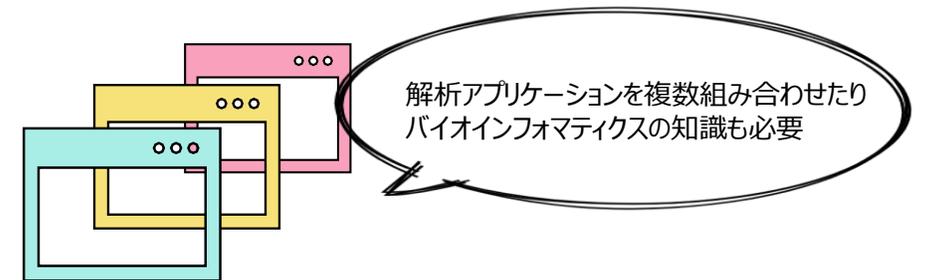


どのような機能を持つ微生物がいるか？

その反面・・・



計算量の多いため高度なスペックのPCが必要



解析アプリケーションを複数組み合わせたり  
バイオインフォマティクスの知識も必要

特に機能特性評価は複雑な作業手間と時間がかかる



## ✓非モデル生物に対応

OmicsBoxはリファレンスゲノムがないデータでも解析が実行可能です。  
農学系のユーザーに適した解析ツールが搭載されています。

## ✓バイオインフォマティクスの知識は不要

実績は高いがコマンドライン型であったりOSに制限がある  
オープンソースソフトウェアを多数組み込みマウス操作で簡単に  
解析できるようしたのがOmicsBoxの特徴の1つです。

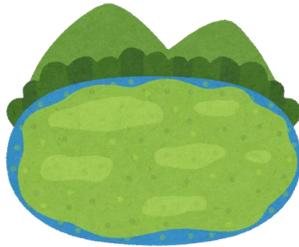
## ✓高価で高スペックなPCは不要

解析や計算は、統合させたウェブサイトや  
Biobam社のクラウドを通して行われるため、  
安定したインターネット接続があれば解析が可能です。



ソーダ湖は、アフリカ、ヨーロッパ、アジアなどで見られる特別な生態系です。これらの湖は、炭酸ナトリウムのレベルが高く、塩分とpHが高くなっています。

ブラジルにある2つのソーダ湖



Verde湖  
シアノバクテリアの発生の記録アリ



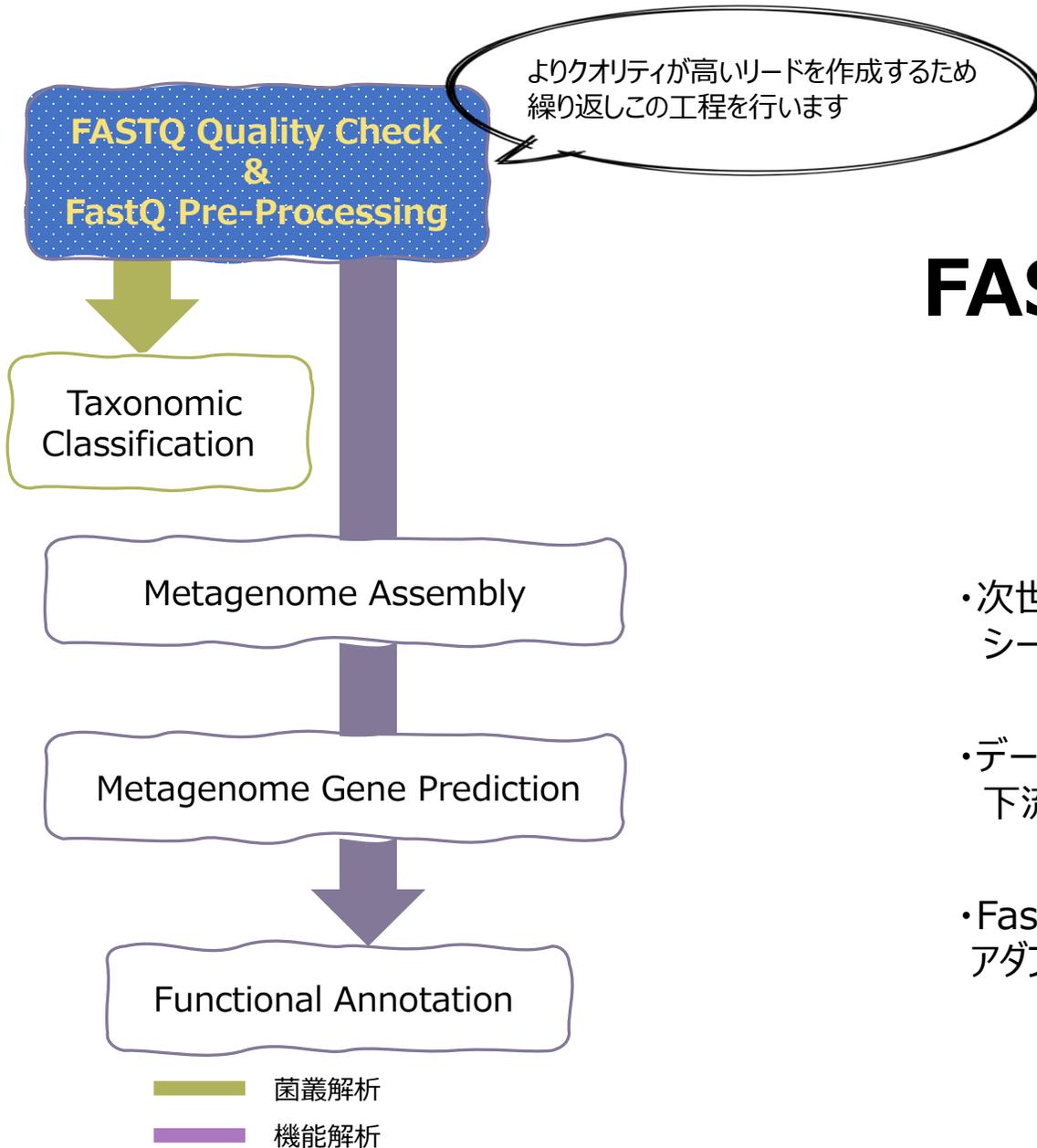
Preta湖  
シアノバクテリアの発生記録ナシ

湖の間には非常に異なる微生物群集と機能的組成物が存在することが期待できます。

## 目的

- PretaとVerdeの間の細菌の多様性を明らかにする。
- これらの微生物群集の機能的な遺伝的特性を識別する。





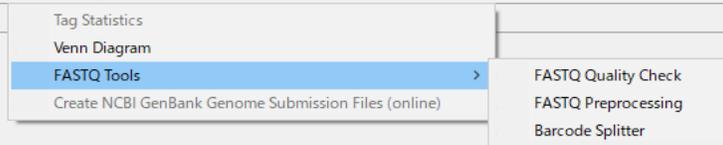
## FASTQ Quality Check & Pre-Processing



- 次世代シーケンサーより出力されたシーケンスデータのクオリティチェックを簡単に実行することができます。
- データが良好に見えるかどうか、下流分析に影響する可能性のある問題を簡単に概観できます。
- FastaQ Quality Checkで確認後アダプター除去や様々なトリミングができます。

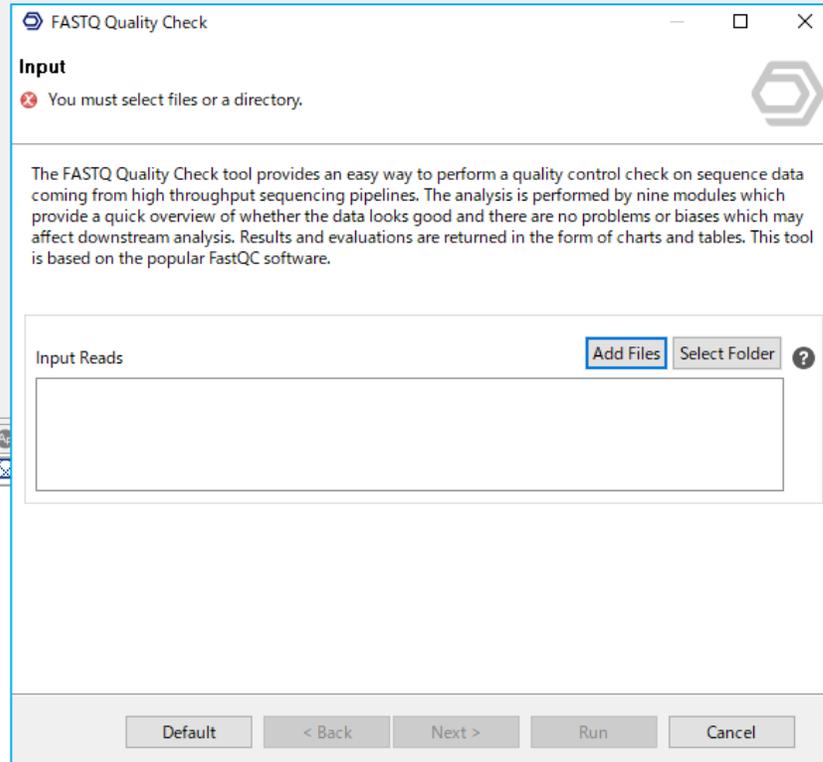
# 解析に使用するデータの品質を確認・トリミングする

File View Help



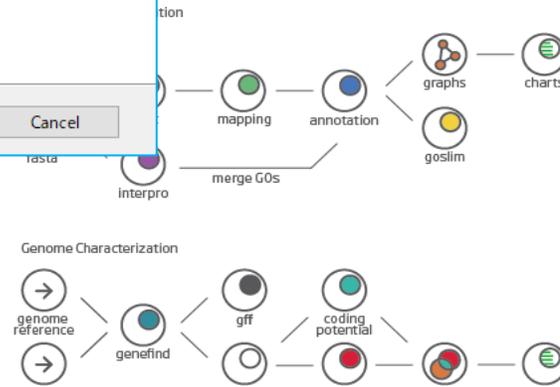
✓行いたい解析は画面上部のアイコンから選択

✓ DNA、RNA、メタゲノムなど各カテゴリーに機能が集約されているので簡単に探すことができます。



✓クオリティチェック・トリミングはどちらも一度に複数サンプルを指定して解析可能

... shows some of the many analysis scenarios available in OmicsBox. Combinations are possible to adjust best to your needs. Hover over the icons to find out more about each step.



Version 1.2

## Metagenomics Module

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

## Transcriptomics Module

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads
- New RNA-Seq De Novo Assembly Option: SuperTranscripts
- Export Normalised Counts of TimeCourse Results
- Export Normalised Counts of Pairwise Differential Expression Without

# 解析に使用するデータの品質を確認・トリミングする

FASTQ Quality Check: Raw\_read\_QC

| Nr | Tags   | Name                  | File Type         | Encoding            | Total Reads | Poor Quality ... | Read Length | %GC  |
|----|--------|-----------------------|-------------------|---------------------|-------------|------------------|-------------|------|
| 1  | REVIEW | ERR1948631_2.fastq.gz | Conventional b... | Sanger / Illumin... | 27602050    | 0                | 125         | 46.0 |
| 2  | REVIEW | ERR1948631_1.fastq.gz | Conventional b... | Sanger / Illumin... | 27602050    | 0                | 125         | 46.0 |
| 3  | REVIEW | ERR1948632_1.fastq.gz | Conventional b... | Sanger / Illumin... | 36384057    | 0                | 125         | 46.0 |
| 4  | REVIEW | ERR1948633_1.fastq.gz | Conventional b... | Sanger / Illumin... | 27772613    | 0                | 125         | 45.0 |
| 5  | REVIEW | ERR1948633_2.fastq.gz | Conventional b... | Sanger / Illumin... | 27772613    | 0                | 125         | 45.0 |
| 6  | REVIEW | ERR1948632_2.fastq.gz | Conventional b... | Sanger / Illumin... | 36384057    | 0                | 125         | 46.0 |
| 7  | REVIEW | ERR1948634_1.fastq.gz | Conventional b... | Sanger / Illumin... | 28646629    | 0                | 125         | 46.0 |
| 8  | REVIEW | ERR1948635_1.fastq.gz | Conventional b... | Sanger / Illumin... | 23536629    | 0                | 125         | 46.0 |
| 9  | REVIEW | ERR1948634_2.fastq.gz | Conventional b... | Sanger / Illumin... | 28646629    | 0                | 125         | 46.0 |
| 10 | REVIEW | ERR1948635_2.fastq.gz | Conventional b... | Sanger / Illumin... | 23536629    | 0                | 125         | 46.0 |
| 11 | REVIEW | ERR1948636_1.fastq.gz | Conventional b... | Sanger / Illumin... | 29046629    | 0                | 125         | 46.0 |
| 12 | REVIEW | ERR1948636_2.fastq.gz | Conventional b... | Sanger / Illumin... | 29046629    | 0                | 125         | 46.0 |

Hide Toolbar  
Actions  
Summary Report

✓解析が終了すると結果が表示されます。

サイドパネルよりレポートを作成できます。

Progress File Manager Application Messages

OmicsBoxWorkspace

Welcome Message

OmicsBox Example Workflows

The following schemes shows some of the many analysis scenarios available in OmicsBox. Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.

Gene Ontology Annotation

fasta → blast → mapping → annotation → graphs → charts  
interpro → merge GOs → annotation → goslim

Genome Characterization

genome reference → genefind → gff → coding potential → charts

OmicsBox Release Notes

Version 1.2

Metagenomics Module

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

Transcriptomics Module

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads
- New RNA-Seq De Novo Assembly Option: SuperTranscripts
- Export Normalised Counts of TimeCourse Results
- Export Normalised Counts of Pairwise Differential Expression Without

✓レポートが作成され各項目について

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できます。

# 解析に使用するデータの品質を確認・トリミングする

FASTQ Quality Check: after

| Nr | Tags   | Name           | File Type     | Encoding         | Total Reads | Poor Qualit... | Read Length |     |
|----|--------|----------------|---------------|------------------|-------------|----------------|-------------|-----|
| 1  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 431121      | 0              | 80-589      | 59. |
| 2  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 489008      | 0              | 80-589      | 59. |
| 3  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 583583      | 0              | 80-589      | 59. |
| 4  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 360399      | 0              | 80-591      | 58. |
| 5  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 355127      | 0              | 80-589      | 58. |
| 6  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 310547      | 0              | 80-571      | 48. |
| 7  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 679214      | 0              | 80-566      | 48. |

FASTQ Quality Check (after)

## FASTQ Quality Check

Name: after

### Overall Results

| Name                           | Per Base Sequence Quality | Per Sequence Quality Scores | Per Base Sequence Content | Per Sequence GC Content | Per Base N Content |
|--------------------------------|---------------------------|-----------------------------|---------------------------|-------------------------|--------------------|
| clean_allcombinedPAB2.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPAB3.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPAB1.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPMB2.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPMB1.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedVAB1.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |
| clean_allcombinedVAB2.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |
| clean_allcombinedPMB3.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |
| clean_allcombinedVAB3.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |
| clean_allcombinedVMB1.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |

各チェック項目を図にて表示することも可能です。

Per Base Sequence Content [clean\_allcombinedPAB2.fastq.gz]

Sequence Length Distribution [clean\_allcombinedPMB3.fastq.gz]

GO Version: Mar 19 2019

# 解析に使用するデータの品質を確認・トリミングする

FASTQ Quality Check: after

| Nr | Tags   | Name           | File Type     | Encoding         | Total Reads | Poor Qualit... | Read Length |     |
|----|--------|----------------|---------------|------------------|-------------|----------------|-------------|-----|
| 1  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 431121      | 0              | 80-589      | 59. |
| 2  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 489008      | 0              | 80-589      | 59. |
| 3  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 583583      | 0              | 80-589      | 59. |
| 4  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 360399      | 0              | 80-591      | 58. |
| 5  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 355127      | 0              | 80-589      | 58. |
| 6  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 310547      | 0              | 80-571      | 48. |
| 7  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 334454      | 0              | 80-568      | 49. |
| 8  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 752149      | 0              | 80-574      | 57. |
| 9  | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 151633      | 0              | 80-551      | 49. |
| 10 | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 606928      | 0              | 80-576      | 46. |
| 11 | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 474245      | 0              | 80-562      | 47. |
| 12 | REVIEW | clean_allco... | Convention... | Sanger / Illu... | 679214      | 0              | 80-566      | 48. |

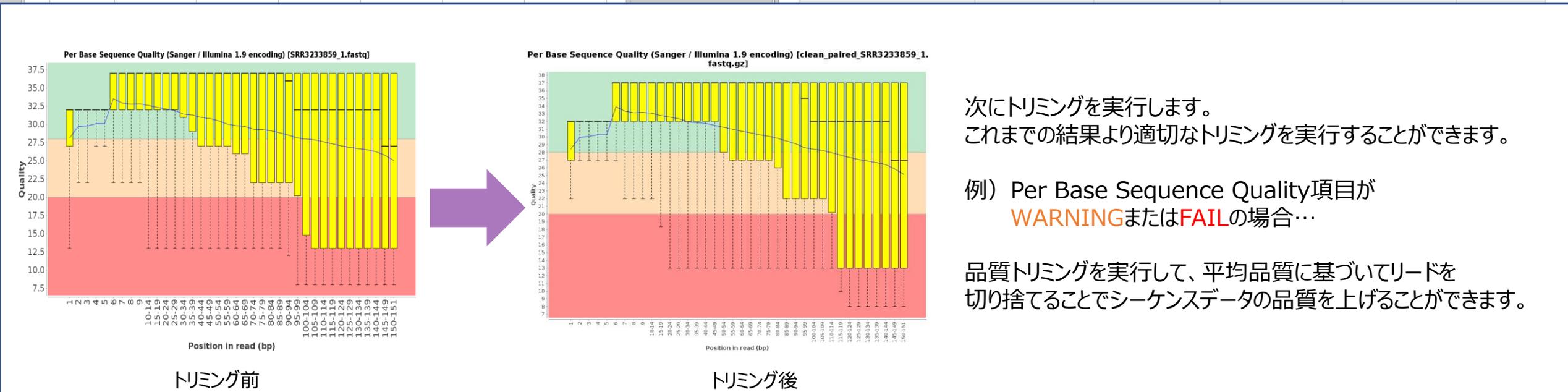
FASTQ Quality Check (after)

## FASTQ Quality Check

Name: after

### Overall Results

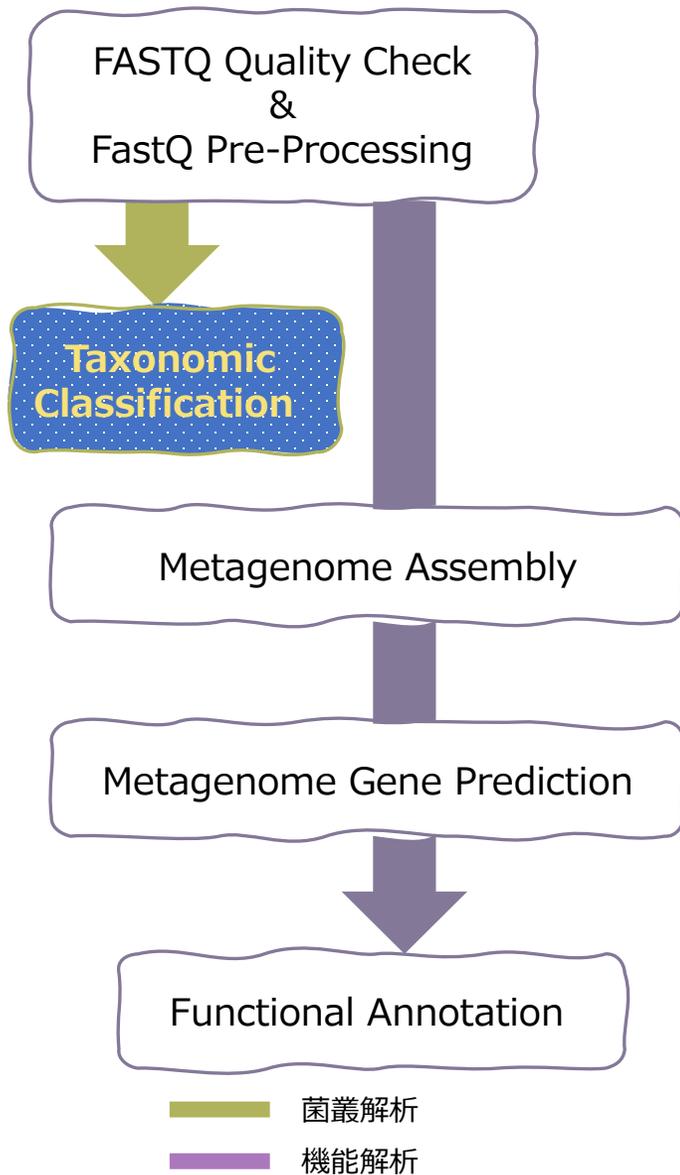
| Name                           | Per Base Sequence Quality | Per Sequence Quality Scores | Per Base Sequence Content | Per Sequence GC Content | Per Base N Content |
|--------------------------------|---------------------------|-----------------------------|---------------------------|-------------------------|--------------------|
| clean_allcombinedPAB2.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPAB3.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPAB1.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPMB2.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedPMB1.fastq.gz | PASS                      | PASS                        | WARNING                   | FAIL                    | PASS               |
| clean_allcombinedVAB1.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |
| clean_allcombinedVAB2.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |
| clean_allcombinedPMB3.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |
| clean_allcombinedVAB3.fastq.gz | PASS                      | PASS                        | FAIL                      | FAIL                    | PASS               |



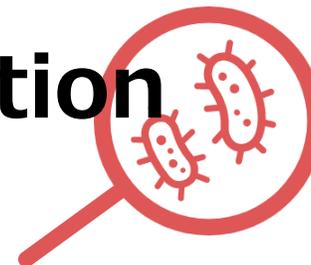
次にトリミングを実行します。  
これまでの結果より適切なトリミングを実行することができます。

例) Per Base Sequence Quality項目が  
**WARNING**または**FAIL**の場合...

品質トリミングを実行して、平均品質に基づいてリードを  
切り捨てることでシーケンスデータの品質を上げることができます。



## Taxonomic Classification



- 微生物ゲノムのデータベースに対してシーケンスを照合して、各シーケンスの分類を識別します。
- このツールはKrakenに基づいておりメタゲノミクスWGSだけでなく、16S / ITSアンプリコンデータを扱う場合にも使用できます。

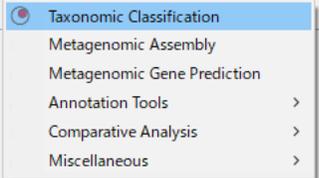
本ツールでPretaとVerdeの間の細菌の多様性を明らかにします。

File View Help



分析アイコン「Metagenomics」>  
「Taxonomic Classification」より実行できます。

Start typing to search actions...



**Taxonomic Classification**

**Input**  
You must select files or a directory.

Kraken is a taxonomic sequence classifier that assigns taxonomic labels to short DNA reads. It does this by examining the k-mers within a read and querying a database with those k-mers. This database contains a mapping of every k-mer in Kraken's genomic library to the lowest common ancestor (LCA) in a taxonomic tree of all genomes that contain that k-mer. The set of LCA taxa that correspond to the k-mers in a read are then analyzed to create a single taxonomic label for the read; this label can be any of the nodes in the taxonomic tree. Kraken is designed to be rapid, sensitive, and highly precise.

Sequencing Data: Single-End Reads

Reads, Contigs or Genes  
Add Files Select Folder

- F:\salmonella\ERR274480\_1.fastq.gz
- F:\salmonella\ERR274480\_2.fastq.gz
- F:\salmonella\ERR277203\_1.fastq.gz
- F:\salmonella\ERR277203\_2.fastq.gz
- F:\salmonella\ERR277204\_1.fastq.gz

**Paired-End Configuration**  
Define the pattern to distinguish upstream files from downstream files. The pattern is searched right before the file extension, and the start of the name should be the same for both files of each sample.

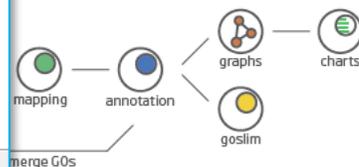
Upstream Files Pattern: \_1  
Downstream Files Pattern: \_2

Default < Back Next > Run Cancel

シーケンスデータの種類と  
トリミング後のシーケンスデータを選択します。

Workflow

Shows some of the many analysis scenarios available in OmicsBox. You can adjust the parameters as possible to adjust best to your needs. Hover over the icons to get more information about each step.



OmicsBox Release Notes

Version 1.2

**Metagenomics Module**

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

**Transcriptomics Module**

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads

File View Help



Start typing to search actions...

### Taxonomic Classification

**Configuration**

Do not save read names

Contaminant Index: None

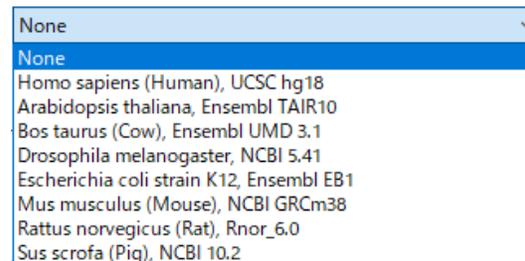
Version Details:  
- Kraken 1.0 with DB 2019.06 (archaea, bacteria, fungi, human, protozoa, viral)  
- Bowtie2 2.3.5.1

Please Cite:  
- Wood DE. and Salzberg SL. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, 15(3), R46.  
- Langmead B. and Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357-9.

Buttons: Default, < Back, Next >, Run, Cancel

メタゲノムシーケンスデータは、多くの場合、宿主生物によって「汚染」されています

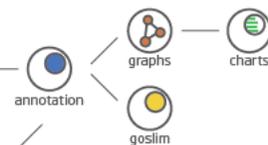
例) ヒトの腸のゲノムシーケンスプロジェクトには、  
Homo sapiensからのリードが含まれます。



宿主生物に由来する可能性のあるリードを除外できます。

Workflow

of the  
ossible  
about e



#### Metagenomics Module

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

#### Transcriptomics Module

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads

Genome Characterization



# 菌叢解析

File View Help



Start typing to search actions...

Table: taxonomic\_classification

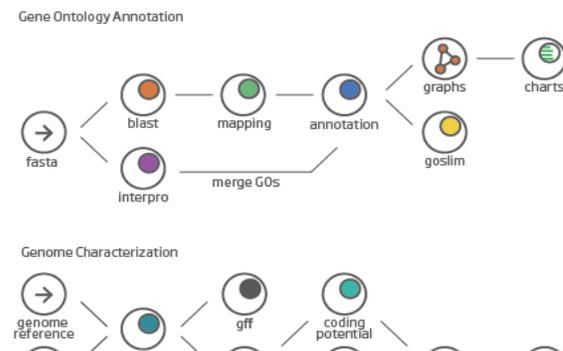
5,211 of 5,211

| Rank          | TaxID   | Scientific Name                   | PAB   | PMB   | VAB   | VMB   |
|---------------|---------|-----------------------------------|-------|-------|-------|-------|
| unclassified  | 0       | Unknown                           | 46816 | 51267 | 29049 | 42399 |
| below species | 1089553 | Thermacetogenium phaeum DSM 12270 | 0     | 0     | 1     | 1     |
| species       | 1916956 | Synechococcus sp. SynAce01        | 11    |       |       |       |
| species       | 1990687 | Rhodococcus sp. S2-17             | 3     |       |       |       |
| species       | 1720344 | Psychrobacter sp. AntiMn-1        | 0     |       |       |       |
| superkingdom  | 2       | Bacteria <bacteria>               | 20883 |       |       |       |
| genus         | 6       | Azorhizobium                      | 15    |       |       |       |
| species       | 7       | Azorhizobium caulinodans          | 15    |       |       |       |
| species       | 9       | Buchnera aphidicola               | 1     | 0     | 0     | 0     |
| genus         | 10      | Cellvibrio                        | 5     | 4     | 0     | 0     |
| species       | 270351  | Methylobacterium aquaticum        | 16    | 12    | 0     | 3     |
| species       | 11      | Cellulomonas gilvus               | 2     | 3     | 1     | 0     |
| species       | 712710  | Tannerella sp. oral taxon HOT-286 | 1     | 2     | 0     | 1     |
| genus         | 13      | Dictyoglomus                      | 0     | 1     | 0     | 0     |
| genus         | 65551   | Meiothermus                       | 5     | 3     | 1     | 2     |
| genus         | 16      | Methylophilus                     | 1     | 2     | 0     | 0     |
| species       | 2072590 | Massilia armeniaca                | 11    | 13    | 5     | 4     |
| below species | 1146883 | Blastococcus saxosidens DD2       |       | 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。

✓ 結果の表が作成され、門から種までの菌種同定がされます。

Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.



## OmicsBox Release Notes

### Version 1.2

#### Metagenomics Module

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

#### Transcriptomics Module

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads
- New RNA-Seq De Novo Assembly Option: SuperTranscripts
- Export Normalised Counts of TimeCourse Results

# 菌叢解析

File View Help



Start typing to search actions...

Table: taxonomic\_classification

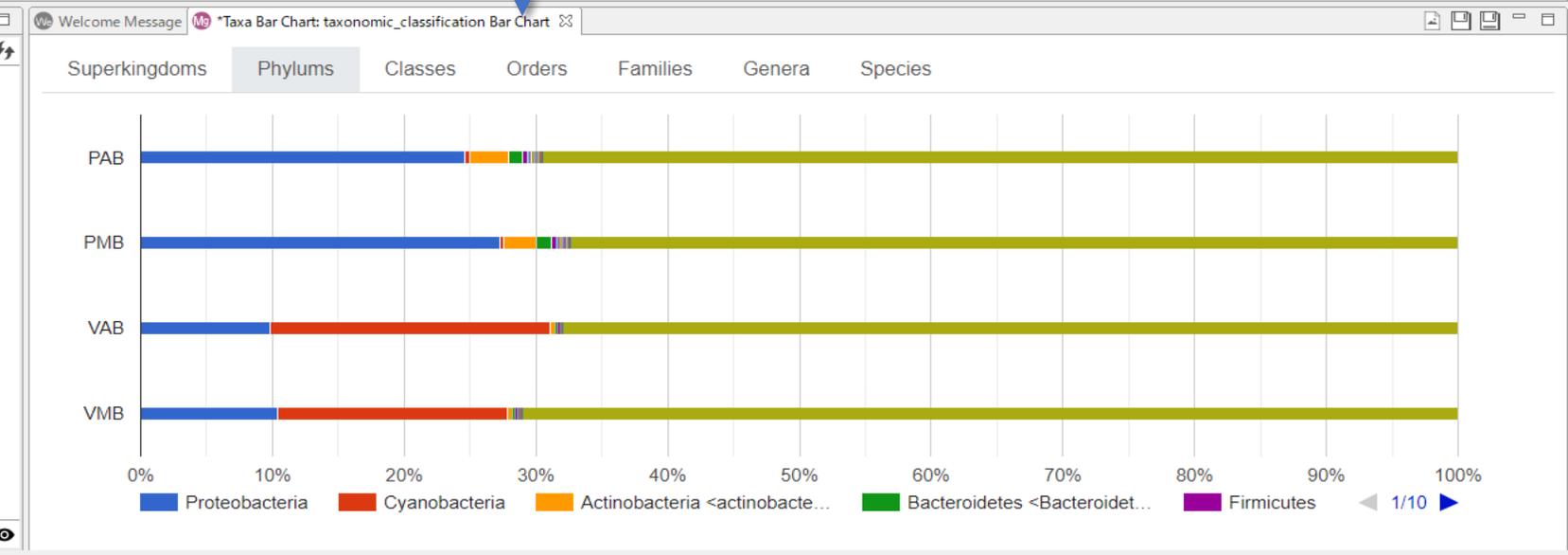
| Rank          | TaxID   | Scientific Name                   | PAB   | PMB   | VAB   | VMB   |
|---------------|---------|-----------------------------------|-------|-------|-------|-------|
| unclassified  | 0       | Unknown                           | 46816 | 51267 | 29049 | 42399 |
| below species | 1089553 | Thermacetogenium phaeum DSM 12270 | 0     | 0     | 1     | 1     |
| species       | 1916956 | Synechococcus sp. SynAce01        | 11    | 7     | 1     | 0     |
| species       | 1990687 | Rhodococcus sp. S2-17             |       |       |       |       |
| species       | 1720344 | Psychrobacter sp. AntiMn-1        |       |       |       |       |
| superkingdom  | 2       | Bacteria <bacteria>               |       |       |       | 67    |
| genus         | 6       | Azorhizobium                      |       |       |       |       |
| species       | 7       | Azorhizobium caulinodans          |       |       |       |       |
| species       | 9       | Buchnera aphidicola               |       |       |       |       |
| genus         | 10      | Cellvibrio                        |       |       |       |       |
| species       | 270351  | Methylobacterium aquaticum        |       |       |       |       |
| species       | 11      | Cellulomonas gilvus               |       |       |       |       |
| species       | 712710  | Tannerella sp. oral taxon HOT-286 |       |       |       |       |
| genus         | 13      | Dictyoglomus                      | 0     | 1     | 0     | 0     |
| genus         | 65551   | Meiothermus                       | 5     | 3     | 1     | 2     |
| genus         | 16      | Methylophilus                     | 1     | 2     | 0     | 0     |
| species       | 2072590 | Massilia armeniaca                | 11    | 13    | 5     | 4     |

図やレポートはこちらに表示されます。  
分類棒グラフでは門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。

Progress File Manager Application Messages

C:\> OmicsBoxWorkspace

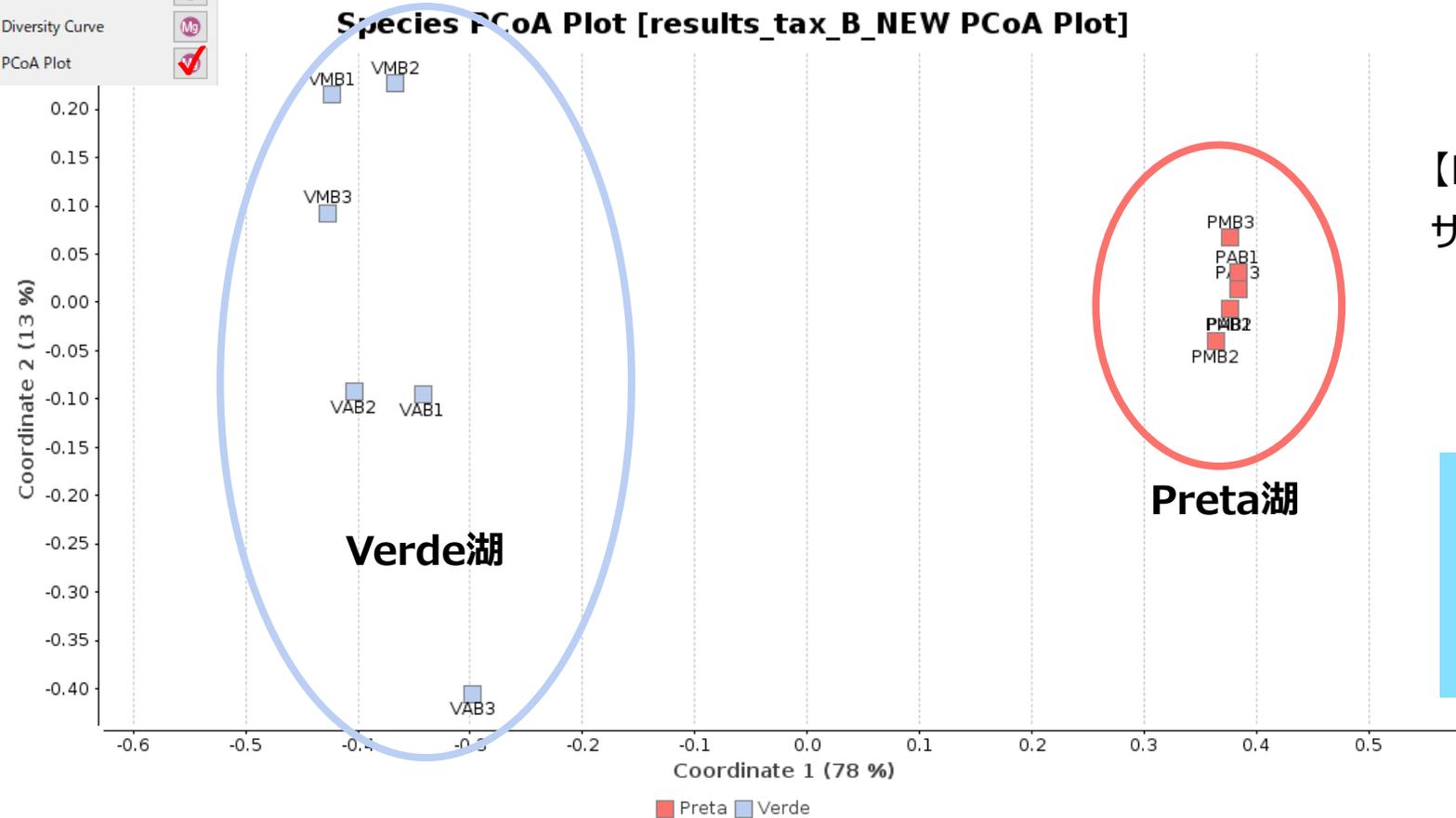
Filter...



Hide Toolbar

Actions

- Summary Report
- Taxa Bar Chart
- Taxa Pie Chart
- Rarefaction Curves
- Diversity Curve
- PCoA Plot



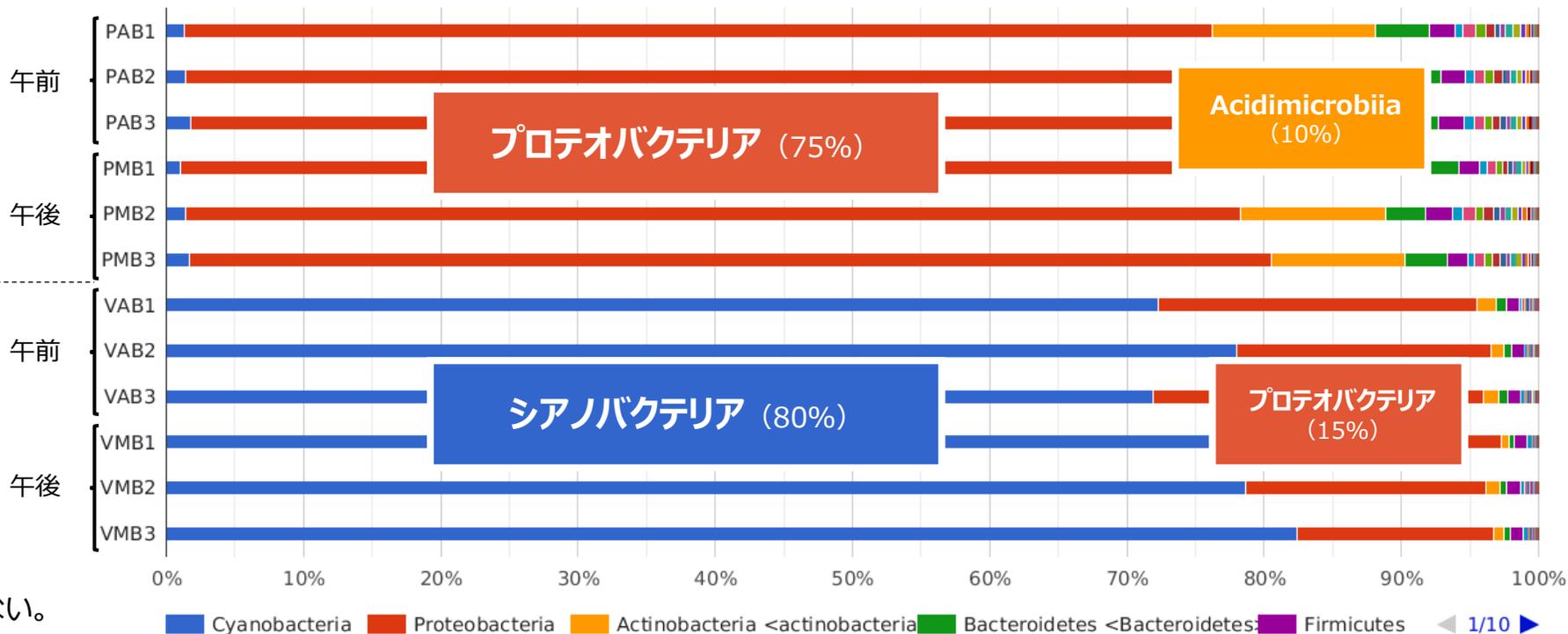
## 【PCoAプロット】

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。

✓ Verde湖とPreta湖の分類構成間に大きな違いがあることを示唆しています

## 【分類学的棒グラフ・門】

Preta湖



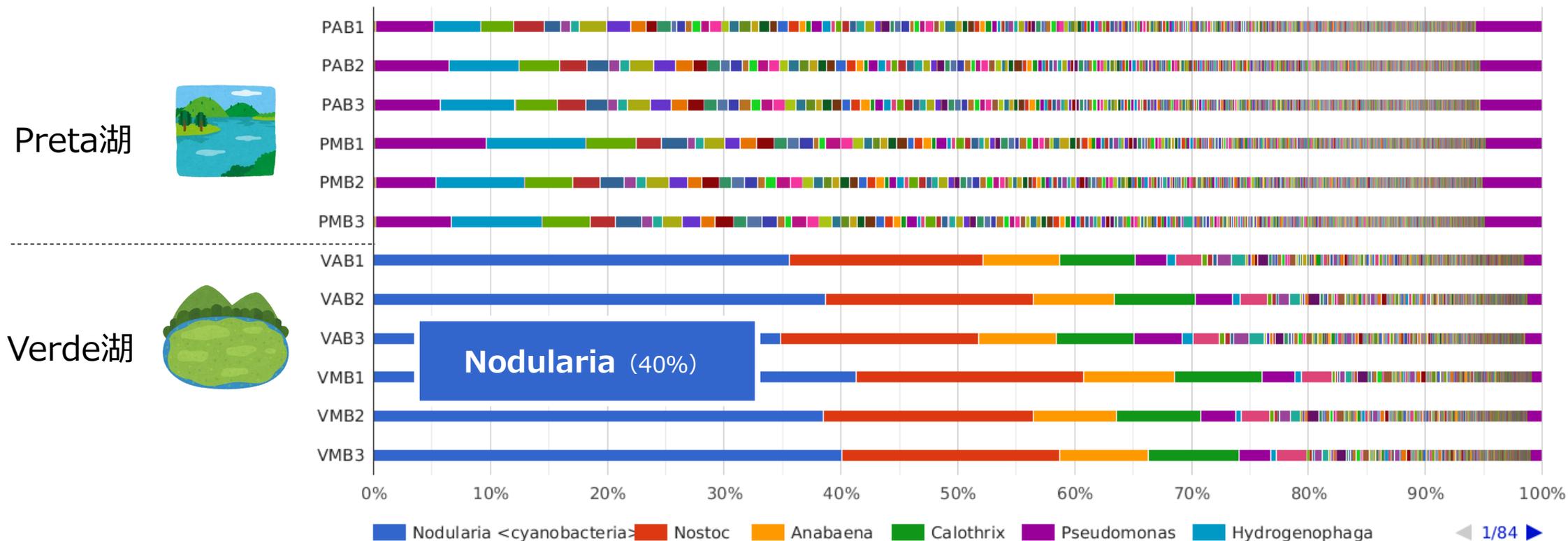
Verde湖



サンプリング時間の間に湖の間に大きな違いはない。

✓シアノバクテリアの割合は、Verde湖見つけた割合と比較してPreta湖では非常に低かった（それぞれ1%と80%）

## 【分類学的棒グラフ・属】

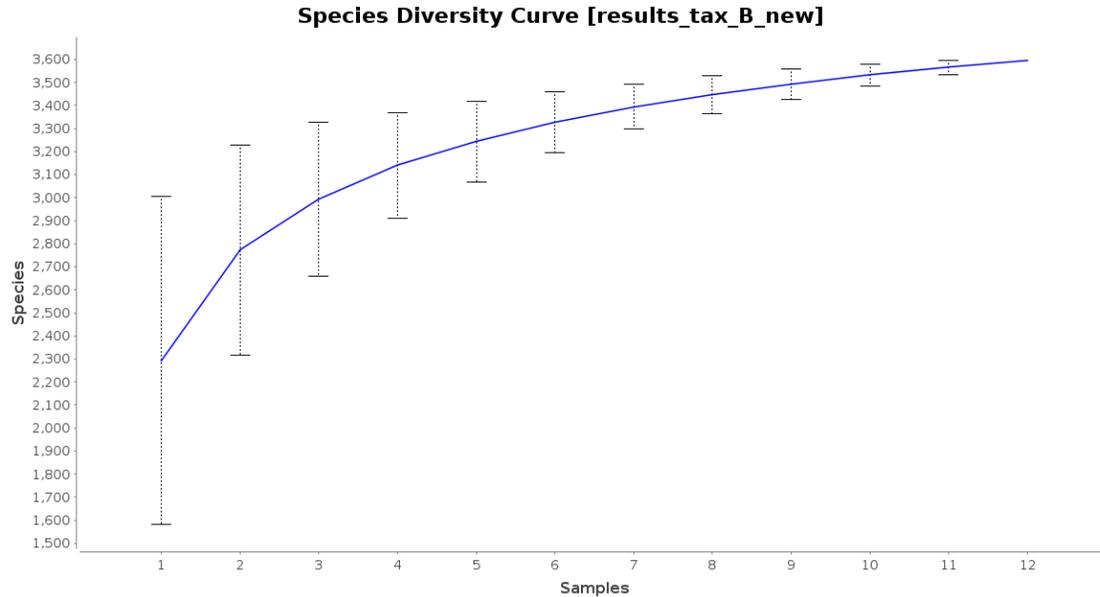


Nodulariaは糸状窒素固定グループで、一般的に塩水システムでの藻類ブルームに関連しています。

他のシアノバクテリア属はNostoc、Calothrix、Anabaena、Fischerellaで、それらのどれもPreta湖で豊かさは見られなかった。

## 【Diversity Curve】

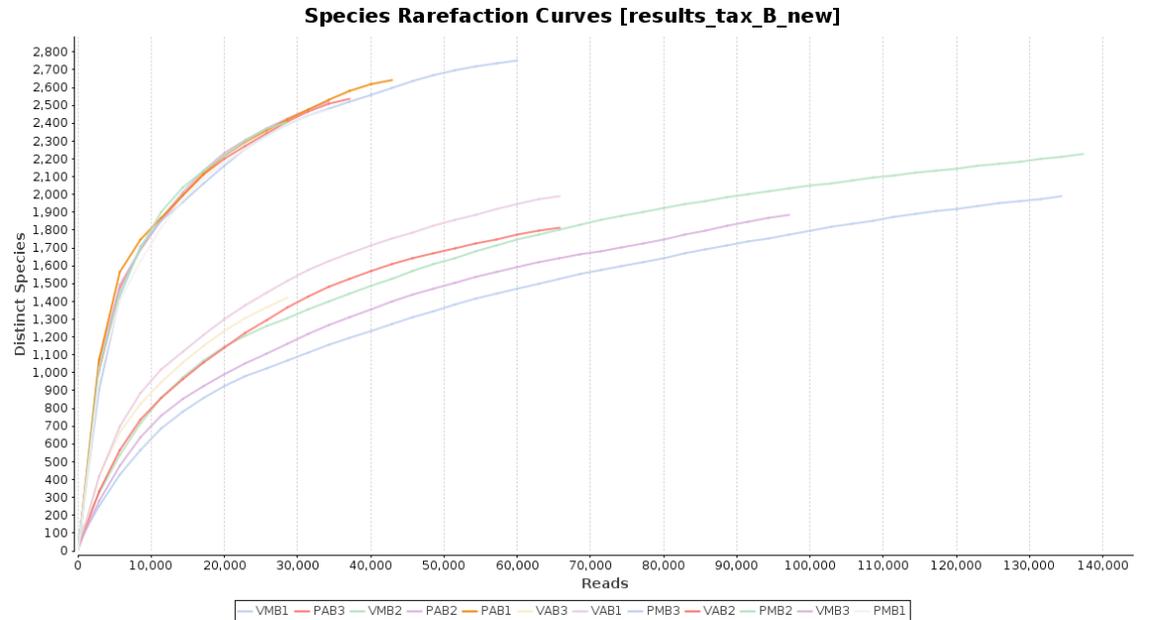
シーケンスのカバレッジが十分に深いかどうかを判断することができます。



データセットにサンプルを追加しても効果が向上しないことを示唆した。

## 【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。



Preta湖の環境に存在するすべての種をカバーするにはシーケンスの深さが十分ではなかったことを示唆している。

# 菌叢解析

The screenshot displays the OmicsBox software interface. At the top, there is a menu bar with 'File', 'View', and 'Help'. Below it is a toolbar with icons for 'start', 'workflows', 'genome browser', 'functional analysis', 'genome analysis', 'transcript omics', 'meta genomics', and 'general tools'. A search bar on the right says 'Start typing to search actions...'. The main workspace shows a 'Table: taxonomic\_classification' with columns for Rank, TaxID, and taxonomic names. A dropdown menu is open under 'Taxonomic Classification', with 'Comparative Analysis' selected, and a sub-menu showing 'OTU Differential Abundance Testing' highlighted. A dialog box titled 'OTU Differential Abundance Testing (taxonomic\_classification)' is open, showing 'Filtering and Normalization' options: 'Minimum Sample Filter' set to 1, 'Counts per Million' set to 1, and 'Normalization Method' set to 'TMMwsp'. A text box explains that the tool is based on edgeR. At the bottom, there are buttons for 'Default', '< Back', 'Next >', 'Run', and 'Cancel'. On the right side, there are panels for 'Diversity Curve', 'PCoA Plot', and 'Release Notes'.

| Rank          | TaxID   | Taxonomic Name             |
|---------------|---------|----------------------------|
| unclassified  | 0       | Unknown                    |
| below species | 1089553 | Thermacetogenium phaeum DS |
| species       | 1916956 | Synechococcus sp. SynAce01 |
| species       | 1990687 | Rhodococcus sp. S2-17      |
| species       | 1720344 | Psychrobacter sp. AntiMn-1 |
| superkingdom  | 2       | Bacteria <bacteria>        |
| genus         | 6       | Azodibzobium               |

|       | PAB | PMB | VAB | VMB |
|-------|-----|-----|-----|-----|
| 46816 |     |     |     |     |
| 20883 |     |     |     |     |
| 15    | 11  | 2   |     | 1   |
| 15    | 11  | 2   |     | 1   |
| 1     | 0   | 0   |     | 0   |
| 5     | 4   | 0   |     | 0   |
| 16    | 12  | 0   |     | 3   |
| 2     | 3   | 1   |     | 0   |
| 1     | 2   | 0   |     | 1   |
| 0     | 1   | 0   |     | 0   |
| 5     | 3   | 1   |     | 2   |
| 1     | 2   | 0   |     | 0   |
| 11    | 13  | 5   |     | 4   |
| 7     | 7   | 0   |     | 1   |

OTU Differential Abundance Testing (taxonomic\_classification)

### Filtering and Normalization

The Taxonomic Differential Abundance Testing is a tool to identify Operational Taxonomic Units (OTUs) that significantly differ between two microbial communities.

This feature is based on edgeR, which belongs to the Bioconductor project.

**Filter OTUs with low counts**

Minimum Sample Filter: 1

Counts per Million: 1

**Calculate normalization factors**

Normalization Method: TMMwsp

Buttons: Default, < Back, Next >, Run, Cancel

✓ OmicsBoxには2つの微生物群の間で著しく異なるOUTを識別する統計解析ツールを搭載しています。

✓ フィルターの条件と統計計算方法を指定

Version 1.2

#### Metagenomics Module

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

#### Transcriptomics Module

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads

File View Help



Start typing to search actions...

Table: taxonomic\_classification

| Rank          | Count |
|---------------|-------|
| unclassified  | 0     |
| below species | 1089  |
| species       | 1916  |
| species       | 1990  |
| species       | 1720  |
| superkingdom  | 2     |
| genus         | 6     |
| species       | 7     |
| species       | 9     |
| genus         | 10    |
| species       | 2703  |
| species       | 11    |
| species       | 7127  |
| genus         | 13    |
| genus         | 6555  |
| genus         | 16    |
| species       | 2072  |
| below species | 1146  |

### OTU Differential Abundance Testing (taxonomic\_classification)

**Experimental Design**

File does not exist

Without experimental design

Experimental Design  Browse...

---

Select Samples

Select Factor

Contrast Group

Test against remaining

Reference Group

Pairing Factor

Default < Back Next > Run Cancel

|       | PAB | PMB   | VAB   | VMB   |
|-------|-----|-------|-------|-------|
| 46816 |     | 51267 | 29049 | 42399 |
| 0     |     | 0     | 1     | 1     |
| 11    |     | 7     | 1     | 0     |
| 3     |     |       |       |       |
| 0     |     |       |       |       |
| 2088  |     |       |       |       |
| 15    |     |       |       |       |
| 15    |     |       |       |       |
| 1     |     |       |       |       |
| 5     |     |       |       |       |
| 16    |     |       |       |       |
| 2     |     |       |       |       |
| 1     |     |       |       |       |
| 0     |     |       |       |       |
| 5     |     |       |       |       |
| 1     |     |       |       |       |
| 11    |     |       |       |       |
| 7     |     |       |       |       |

実験デザインを選択します。  
 デザインのデータはタブ区切りのファイルを別途用意します。

|    | 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   |

Verde湖のサンプルをContrastグループとして、  
 Preta湖のサンプルをReferenceグループとして解析

- Hide Toolbar
- Actions
  - Summary Report
  - Taxa Bar Chart
  - Taxa Pie Chart
  - Refraction Curves
  - Diversity Curve
  - CoA Plot

- Transcriptomics Module
- Completeness Assessment with BUSCO
  - Predict Coding Regions with TransDecoder
  - Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
  - New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads
  - New RNA-Seq De Novo Assembly Option: SuperTranscripts
  - Export Normalised Counts of TimeCourse Results

File View Help

start workflows genome browser functional analysis genome analysis transcript omics meta genomics general tools

Start typing to search actions...

Table: taxonomic\_classification 5,211 of 5,211

| Rank          | TaxID   | Scientific Name                   | PAB   | PMB   | VAB   | VMB   |
|---------------|---------|-----------------------------------|-------|-------|-------|-------|
| unclassified  | 0       | Unknown                           | 46816 | 51267 | 29049 | 42399 |
| below species | 1089553 | Thermacetogenium phaeum DSM 12270 | 0     | 0     | 1     | 1     |
| species       |         |                                   | 11    | 7     | 1     | 0     |
| species       |         |                                   | 3     | 3     | 1     | 1     |
| species       |         |                                   | 0     | 0     | 0     | 1     |
| superkingdom  |         |                                   | 20883 | 25108 | 13805 | 17367 |
| genus         |         |                                   | 15    |       |       |       |
| species       |         |                                   | 15    |       |       |       |
| species       |         |                                   | 1     |       |       |       |
| genus         |         |                                   | 5     |       |       |       |
| species       |         |                                   | 16    |       |       |       |
| species       |         |                                   | 2     |       |       |       |
| species       |         |                                   | 1     |       |       |       |
| genus         |         |                                   | 0     |       |       |       |
| genus         |         |                                   | 5     |       |       |       |
| genus         |         |                                   | 1     |       |       |       |
| species       |         |                                   | 11    |       |       |       |
| below species |         |                                   | 7     |       |       |       |

OTU Differential Abundance Testing (taxonomic\_classification)

### Statistical Test

Test at Specific Level: Genus

Statistical Test:

- Exact Test:
- GLM Test: GLM Likelihood Ratio Test
- Robust:

Please Cite:  
Robinson MD., McCarthy DJ. and Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139-40.

Default < Back Next > Run Cancel

特定の分類群（種、属、家族など）の結果のみを考慮することができます。

今回「属」に指定

豊富にあるOTUを検出するために使用する統計テストを選択できます。

Workflows

OmicsBox Release Notes

Version 1.2

**Metagenomics Module**

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

**Transcriptomics Module**

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads
- New RNA-Seq De Novo Assembly Option: SuperTranscripts
- Export Normalised Counts of TimeCourse Results

Filter...

GO Version: Mar 23 2020

# 菌叢解析

File View Help

start workflows genome browser functional analysis genome analysis transcript omics meta genomics general tools

Start typing to search actions...

Table: taxonomic\_classification | Metagenomics Differential Abundance Results: taxonomic\_classification\_odat

| Nr | Tags  | Taxald  | Scientific Name         |          |          |            |            |          |
|----|-------|---------|-------------------------|----------|----------|------------|------------|----------|
| 1  |       | 698776  | Cellulosilyticum        |          |          |            |            |          |
| 2  |       | 630749  | Spongiibacter           |          |          |            |            |          |
| 3  |       | 929812  | Gibbsiella              |          |          |            |            |          |
| 4  |       | 1204360 | Siansivirga             |          |          |            |            |          |
| 5  |       | 295595  | Candidatus Hepatoplasma |          |          |            |            |          |
| 6  |       | 2282742 | Desulfofarcimen         |          |          |            |            |          |
| 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   | Halotheca               | 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  |
| 12 |       | 444459  | Flagellimonas           | -8.81971 | -3.14073 | 0.20743    | 0.39265    | 1.58928  |
| 13 |       | 1938335 | Sulfuritortus           | -1.03529 | -0.05004 | 0.91497    | 0.9378     |          |
| 14 |       | 1609958 | Confluentimicrobium     | -1.10646 | -0.14596 | 0.87267    | 0.90711    |          |
| 15 |       | 35719   | Thielavia               | 3.03675  | 1.60253  | 0.15533    | 0.33407    |          |
| 16 |       | 32207   | Rothia <high GC Gram+>  | -1.19092 | -0.25207 | 0.86021    | 0.8978     |          |

1,027 of 1,027

Hide Toolbar

Actions

- Summary Report
- Set Over/Under Tags
- Summary Chart
- Heatmap

特定のOTUが過剰に表現されている場合はOVER（赤）  
表現されていない場合はUNDER-（緑）で表示されます。

サイドパネルから  
図の作成も可能（次スライド）

Progress File Manager Application Messages

OmicsBoxWorkspace

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

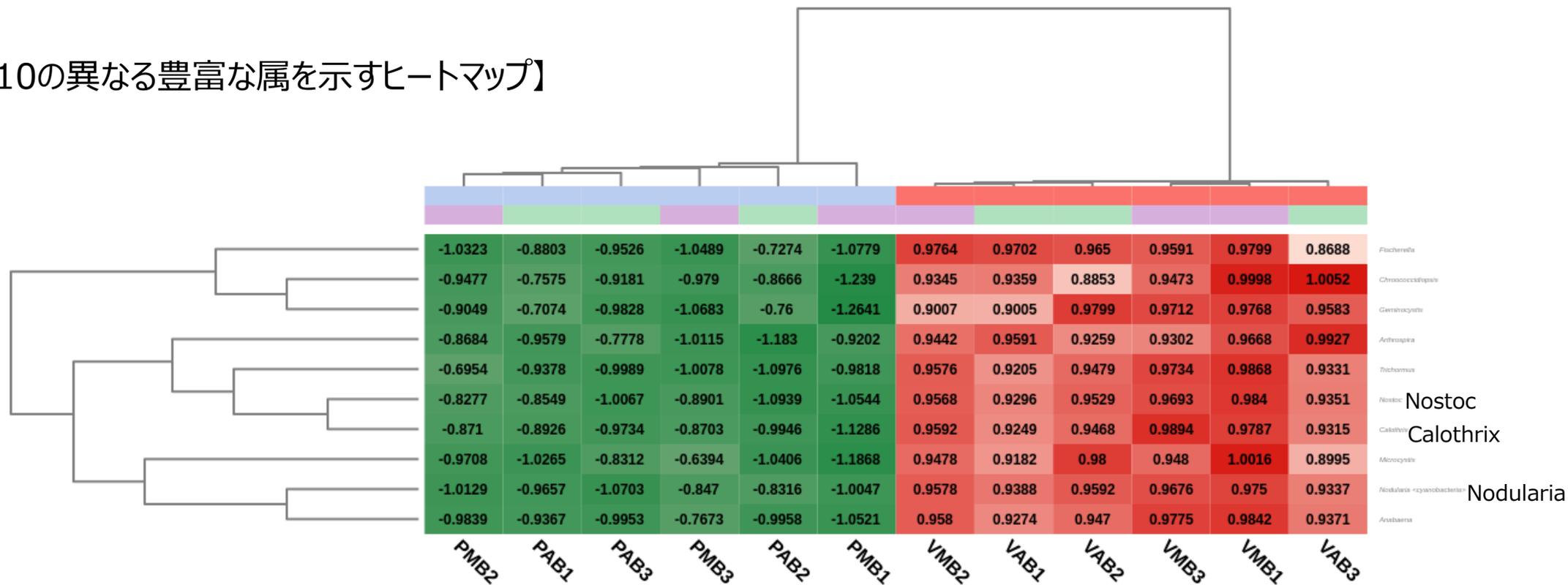
Under-represented (logFC < -1): 90

#### 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

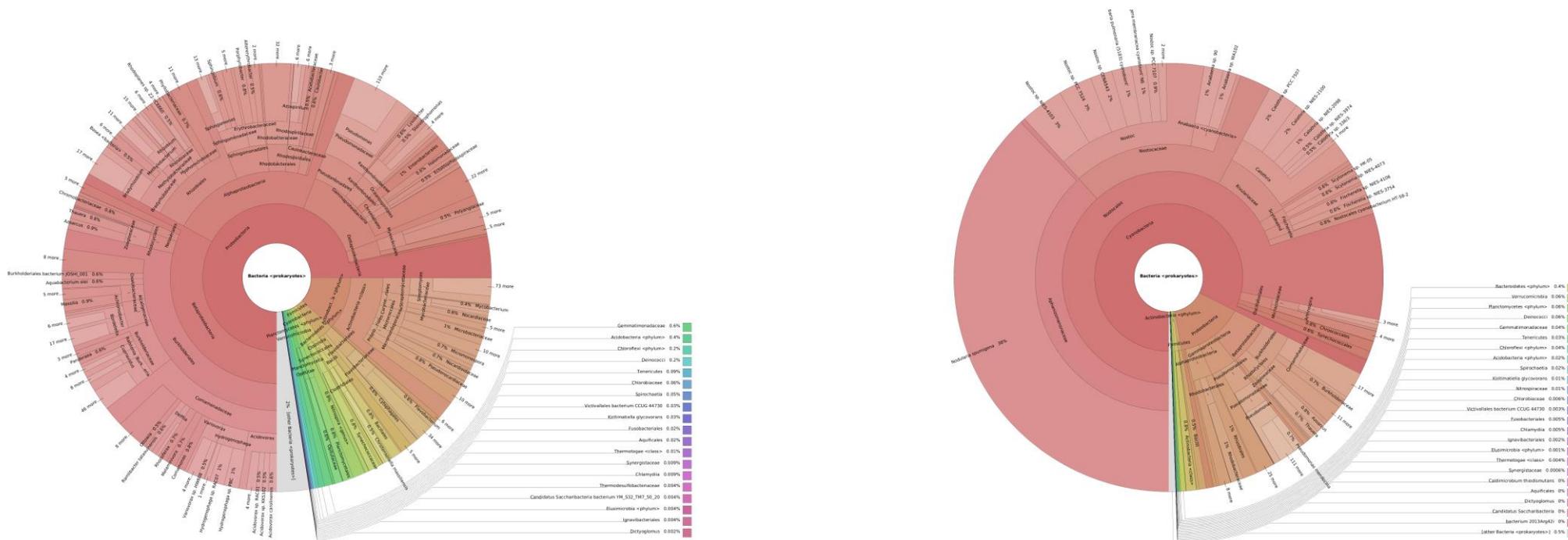
【 TOP10の異なる豊富な属を示すヒートマップ】



✓Verde湖で最も豊富な属がシアノバクテリア（Nostoc、Nodularia、Calothrixなど）に対応することを示した。

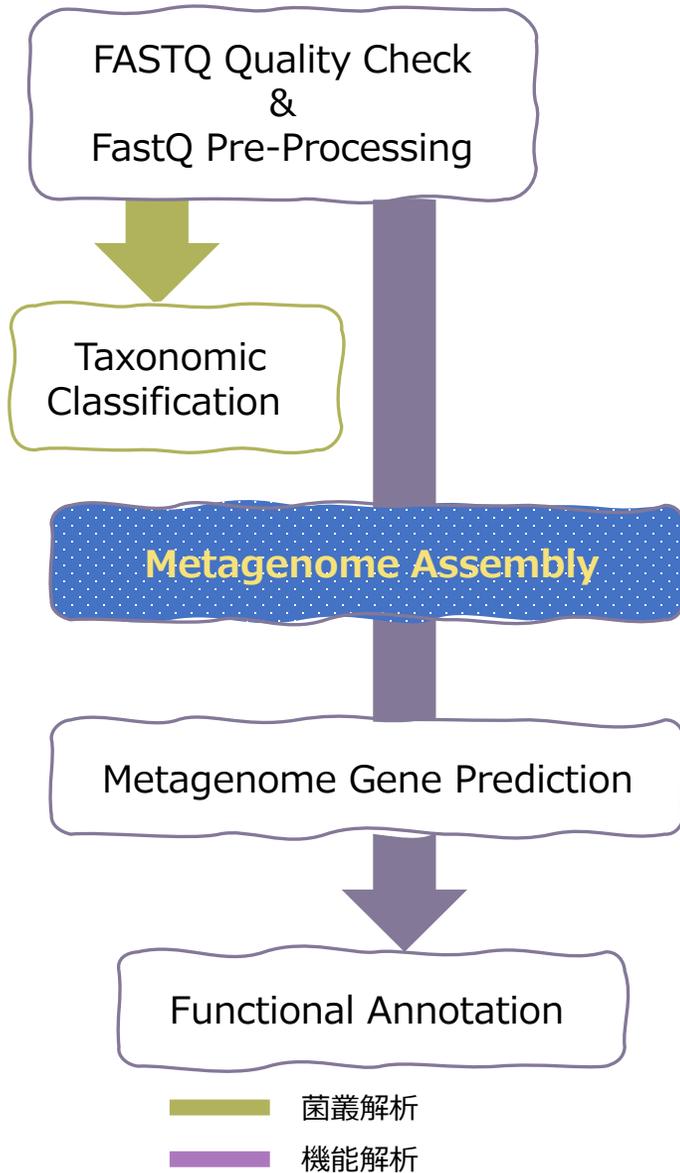
PCoAプロットと同様、単一の湖内におけるサンプリング時間に明確な違いがないことを示した。

## 【Kronaチャート】



PAB (左) とVAB (右) のKronaチャートの比較。

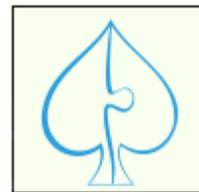
OmicBoxを使用すると、さまざまな分類が階層的に表示されるKronaチャート作成することができます。このグラフを使用すると、サンプルを相互に比較しやすくなります。



## Metagenome Assembly



- 生データを機能特性評価するには配列が短すぎます。  
したがって、遺伝子予測と機能アノテーションの前のステップとして、メタゲノムアセンブリツールを使用します。
- 本ツールは「metaSPAdes」と「MEGAHIT」の2種類のアセンブラーを選択して解析を行います。



# 機能解析

File View Help



Start typing to search actions...

- Taxonomic Classification
- Metagenomic Assembly
- Metagenomic Gene Prediction
- Annotation Tools >
- Comparative Analysis >
- Miscellaneous >

分析アイコン「Metagenomics」>  
「Taxonomic Classification」より実行できます。

Metagenomic Assembly

### Metagenomic Assembly

**● MEGAHIT**  
MEGAHIT is an NGS de novo assembler for assembling large and complex metagenomics data in a time- and cost-efficient manner via a succinct de Bruijn graph data structure. It was created in the same research group that was involved in the development of SOAPdenovo and SOAPdenovo2 and may be seen as the successor of these tools.

**○ meta-SPAdes**  
SPAdes - St. Petersburg genome assembler - is an assembly toolkit containing various assembly pipelines based on de Bruijn graph. metaSPAdes addresses various challenges of metagenomic assembly by capitalizing on computational ideas that proved to be useful in assemblies of single cells and highly polymorphic diploid genomes.

Default < Back **Next >** Run Cancel

アSEMBラーの選択  
今回はMEGAHITを使用

MEGAHIT

**Input**  
You must select files or a directory.

MEGAHIT is an NGS de novo assembler for assembling large and complex metagenomics data in a time- and cost-efficient manner via a succinct de Bruijn graph data structure. It was created in the same research group that was involved in the development of SOAPdenovo and SOAPdenovo2 and may be seen as the successor of these tools.

Sequencing Data: Single-End Reads

Reads: Add Files Select Folder

**Paired-End Configuration**  
Define the pattern to distinguish upstream files from downstream files. The pattern is searched right before the file extension, and the start of the name should be the same for both files of each sample.

Upstream Files Pattern: \_1  
Downstream Files Pattern: \_2

Default < Back **Next >** Run Cancel

シーケンスデータの種類と  
トリミング後のシーケンスデータを選択します。

MEGAHIT

### Configuration

**Basic**

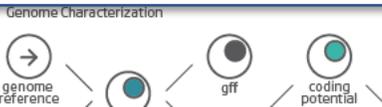
Minimum Multiplicity: 2  
K-mer Sizes: 21,29,39,59,79,99,119,141

**Advanced**

No Mercy K-mers:   
Bubble Level: High (2)  
Bubble Merge Level L: 20  
Bubble Merge Level S: 0.95  
Prune Level: High (2)  
Prune Depth: 2  
Low Local Ratio: 0.2  
Max Tip Length: 2  
Disable Local Assembly:

Default < Back **Next >** Run Cancel

しきい値などの調整を必要に応じて行います。



- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads
- New RNA-Seq De Novo Assembly Option: SuperTranscripts

File View Help



Start typing to search actions...

MEGAHIT

### Output

Contigs Fasta

Version Details:  
MEGAHIT v1.2.8

Please Cite:  
- Li D., Liu CM., Luo R., Sadakane K. and Lam TW. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)*, 31(10), 1674-6.  
- Li D., Luo R., Liu CM., Leung CM., Ting HF., Sadakane K., Yamashita H. and Lam TW. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods (San Diego, Calif.)*, 102, 3-11.

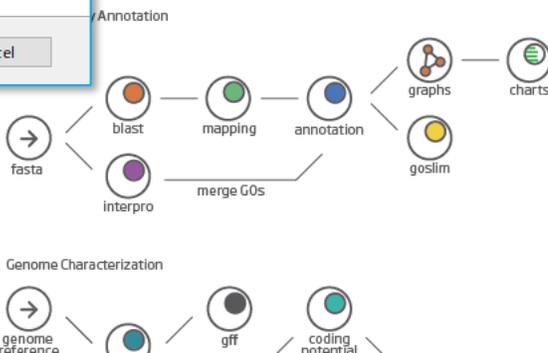
結果のアセンブルデータの保存先を指定します。

Progress File Manager

C¥    
 > OmicsBoxWorkspace

## Example Workflows

g schemes shows some of the many analysis scenarios available in OmicsBox. Different combinations are possible to adjust best to your needs. Hover over the elements to find out more about each step.



## OmicsBox Release Notes

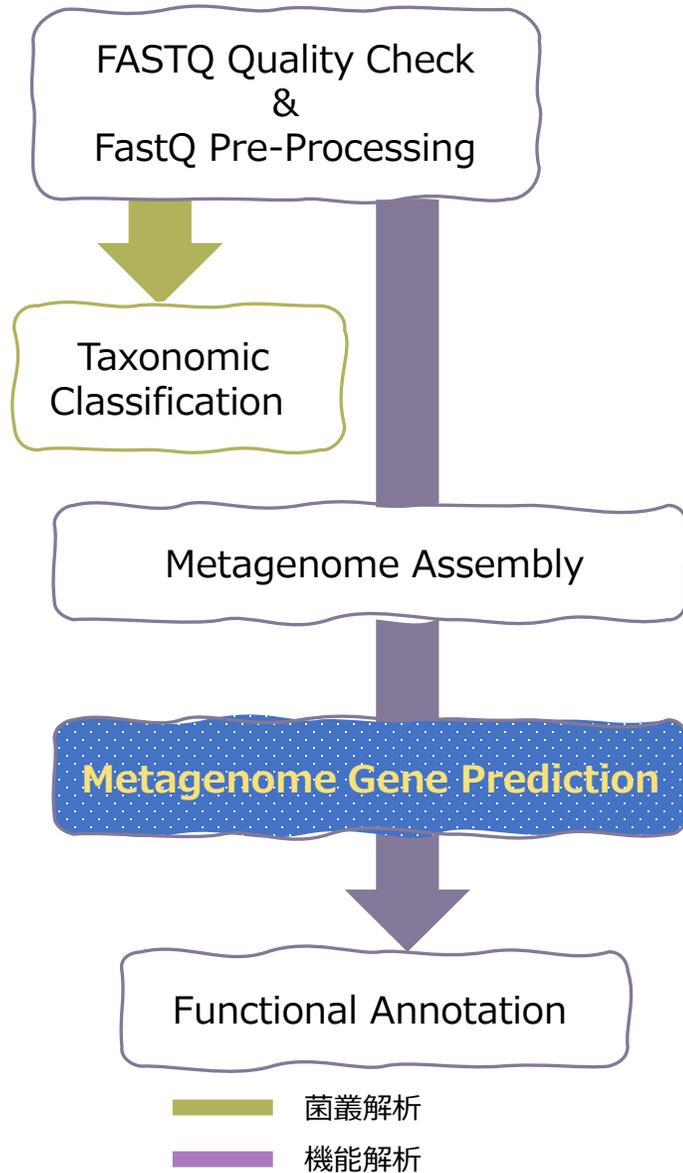
### Version 1.2

#### Metagenomics Module

- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

#### Transcriptomics Module

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced Junctions, Export Unmapped Reads
- New RNA-Seq De Novo Assembly Option: SuperTranscripts



## Metagenome Gene Prediction

- ゲノムの原核生物遺伝子またはORF領域を予測するツールです。
- 「FragGeneScan」または「Prodigal」のアルゴリズムを使用して、可能性のある遺伝子やタンパク質を特定して抽出します。

FGS

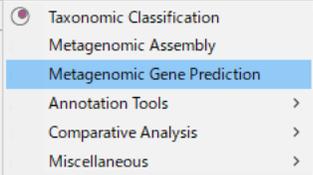
prodigal

# 機能解析

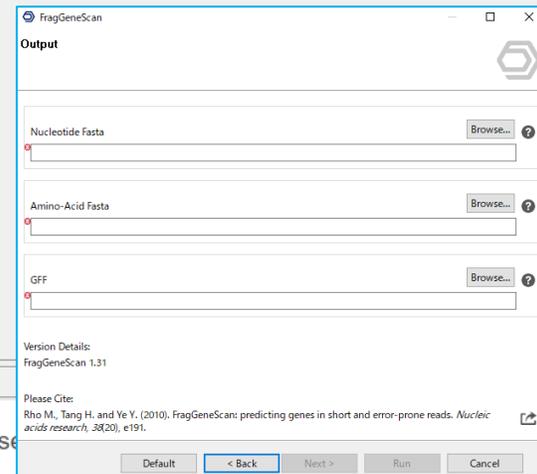
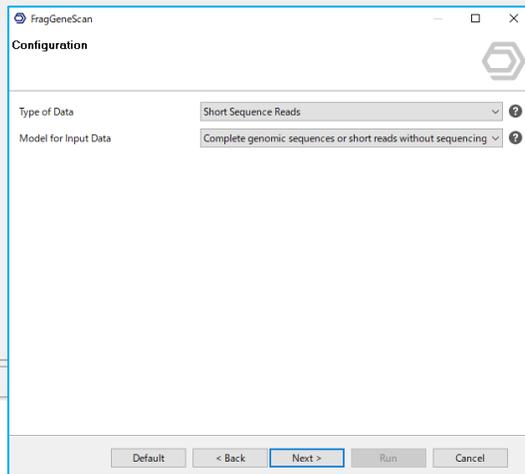
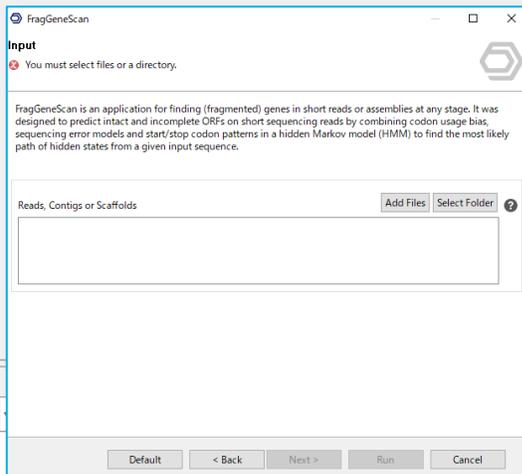
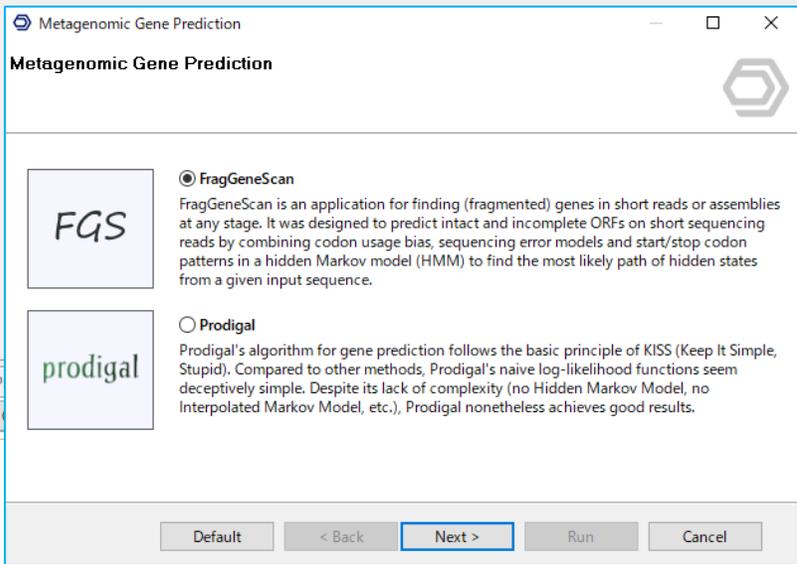
File View Help



Start typing to search actions...



分析アイコン「Metagenomics」>  
「Metagenome Gene Prediction」より実行できます。



The following schemes shows some of the many analysis scenarios available in OmicsBox. Many different combinations are possible to adjust best to your needs. Use the different elements to find out more about each step.

アSEMBルされたデータを  
ここで選択します。

データのタイプやモデルを  
選択します。

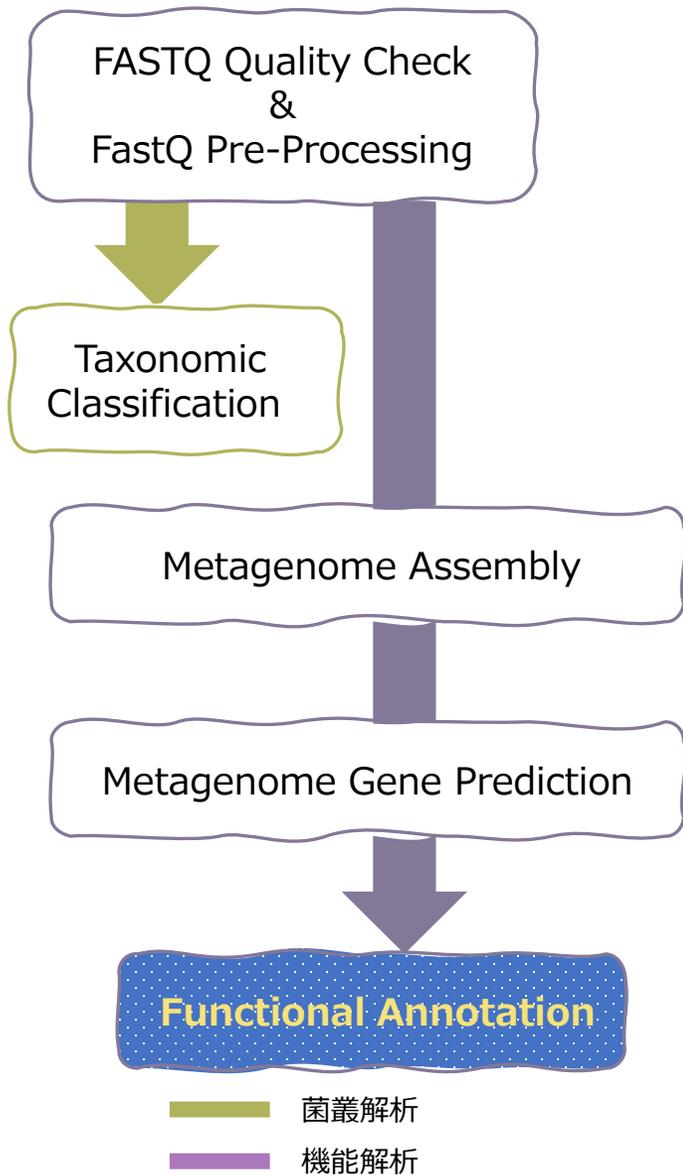
各出力データの保存先を  
指定します。  
上の2つはmulti fastaとして  
出力されます。

アルゴリズムの選択  
今回はFragGeneScanを使用

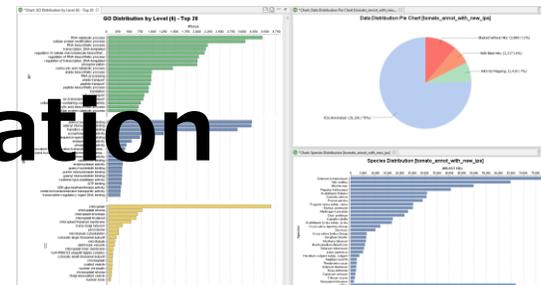


|                     |                              | <b>PAB</b> | <b>PMB</b> | <b>VAB</b> | <b>VMB</b> |
|---------------------|------------------------------|------------|------------|------------|------------|
| <b>Assembly</b>     | <b>N. contigs</b>            | 13363      | 16607      | 7399       | 8727       |
|                     | <b>N50</b>                   | 691        | 717        | 884        | 936        |
| <b>Gene Finding</b> | <b>Predicted genes</b>       | 71557      | 76640      | 42880      | 59839      |
|                     | <b>Avg. gene length (nt)</b> | 389.79     | 404.66     | 394.01     | 381.19     |

ここまでで配列を組み立て各アセンブリ内の遺伝子と遺伝子フラグメントを検索しました。



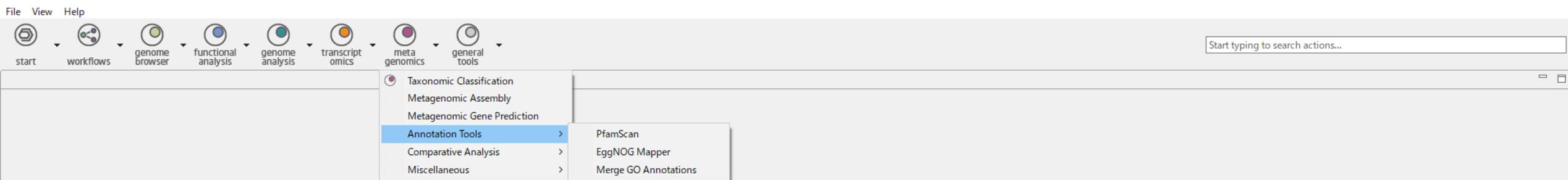
# Functional Annotation



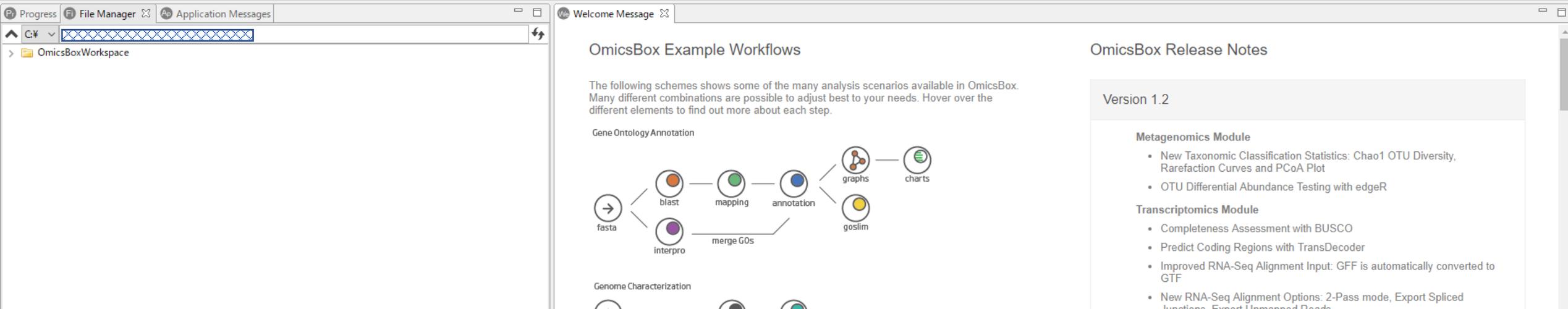
・予測された遺伝子に機能情報を付与します。

・メタゲノム解析では3つのプログラムでアノテーション情報を付与することが可能です。

1. Egnog-mapperによるeggNOGベースのオーソロジー割り当てを利用した新規配列（遺伝子またはタンパク質）の高速機能アノテーション
2. Pfamによるタンパク質ドメイン検索のデータベースです。
3. Blast2GO方法論を使用した高品質なアノテーション情報付与



Eggnog-mapperとPfamによるアノテーション情報を付与は  
分析アイコン「Metagenomics」>「Annotation Tools」より実行できます。  
遺伝子予測されたmulti fastaデータを入力することで解析を実行できます。



## 【EggNOG-mapperとPfam共通】

| Type      | Query ID  | Gene Name | EggNOG Description   | E-Val     | Bit Score | Best Taxa | EC Codes                | # | GOs                                       | GO Names  | KEGG KO  | KEGG P... |
|-----------|-----------|-----------|--|-----------|-----------|-----------|-------------------------|---|---|---|----------|-----------|
| CCOG      | k141_9_1  |           | anaphase-promoting complex binding   | 1.80E-4   | 51.6      | Bacteria  |                         | 0 |   |   |          |           |
| CCOG ENOG | k141_11_1 | rpoC2     | DNA-dependent RNA polymerase catalyzes the transcription of DNA into RN...       | 2.00E-... | 134.4     | Nosto...  | EC:2.7.7.6              | 1 | C:GO:000428                               | C:DNA-directed RNA polymerase complex                             | K03046   | map0...   |
| CCOG ENOG | k141_13_1 | acsA      | Catalyzes the conversion of acetate into acetyl-CoA (AcCoA), an essential int... | 8.80E-... | 159.1     | Hydro...  | EC:6.2.1.1; EC:6.2.1.16 | 0 |   |   | K0189... | map0...   |
| CCOG ENOG | k141_16_1 |           | Trypsin-like peptidase domain  | 6.80E-5   | 52        | Nosto...  |                         | 0 |   |   |          |           |
| CCOG ENOG | k141_19_1 |           | N-6 DNA Methylase  | 2.60E-... | 134       | Cyano...  |                         | 0 |   |   |          |           |
| CCOG ENOG | k141_22_1 | nqrF      | NQR complex catalyzes the reduction of ubiquinone-1 to ubiquinol by two ...      | 6.80E-... | 246.5     | Ruege...  | EC:1.6.5.8              | 6 | F:GO:0071949; P:GO:0055114; F:GO:00515... | Poxidation-reduction process; F:AD binding; F:NADH dehyd...       | K00351   |           |
| CCOG ENOG | k141_23_1 |           | Repeats in polycystic kidney disease 1 (PKD1) and other proteins                 | 9.50E-1   | 40        | Sphin...  |                         | 0 |   |   |          |           |
| ENOG      | k141_28_1 |           |  | 2.50E-... | 147.9     | Nosto...  |                         | 0 |   |   |          |           |
| CCOG ENOG | k141_30_1 |           |  | 4.30E-... | 160.3     | Bact...   |                         | 2 | C:GO:0030116; F:GO:0005100; P:GO:004...   | Ferredoxin-like oxidoreductase; Oxidoreductase (NADPH oxidase)... |          |           |

| Type   | Query ID    | Frame | HMM Acc | HMM Name        | Bit Score | E-Value  | # | GOs  | GO Names   |
|--------|-------------|-------|---------|-----------------|-----------|----------|---|--|--|
| FAMILY | k141_4628_1 |       | PF03235 | DUF262          | 30.8      | 3.00E-7  | 0 | -  | -  |
| DOMAIN | k141_4647_1 |       | PF02041 | Auxin_BP        | 11.8      | 1.40E-1  | 1 | F:GO:0010011   | Fauxin binding   |
| DOMAIN | k141_4666_1 |       | PF16770 | RTT107_BRCT_5   | 11.2      | 2.50E-1  | 0 | -  | -  |
| MOTIF  | k141_4684_1 |       | PF07595 | Planc_extracel  | 10.6      | 3.70E-1  | 0 | -  | -  |
| DOMAIN | k141_4690_1 |       | PF04972 | BON             | 55.5      | 4.80E-15 | 0 | -  | -  |
| DOMAIN | k141_4690_1 |       | PF08302 | tRNA_lig_CPD    | 2.6       | 7.00E1   | 3 | F:GO:0005524; P:GO:0006388; F:GO:0003972               | P:tRNA splicing, via endonucleolytic cleavage and ligation; F:RNA ligase (ATP) activity; F:ATP binding         |
| DOMAIN | k141_4690_1 |       | PF08302 | tRNA_lig_CPD    | 11.4      | 1.50E-1  | 3 | F:GO:0005524; P:GO:0006388; F:GO:0003972               | P:tRNA splicing, via endonucleolytic cleavage and ligation; F:RNA ligase (ATP) activity; F:ATP binding         |
| DOMAIN | k141_4690_1 |       | PF04972 | BON             | 53.6      | 1.90E-14 | 0 | -  | -  |
| DOMAIN | k141_4690_1 |       | PF08302 | tRNA_lig_CPD    | 3.3       | 4.50E1   | 3 | F:GO:0005524; P:GO:0006388; F:GO:0003972               | P:tRNA splicing, via endonucleolytic cleavage and ligation; F:RNA ligase (ATP) activity; F:ATP binding         |
| DOMAIN | k141_4690_1 |       | PF04972 | BON             | 66.5      | 1.80E-18 | 0 | -  | -  |
| DOMAIN | k141_4690_1 |       | PF00403 | HMA             | 14.4      | 3.70E-2  | 2 | P:GO:0030001; F:GO:0046872                             | F:metal ion binding; P:metal ion transport   |
| FAMILY | k141_4712_1 |       | PF15842 | DUF4718         | 14        | 3.00E-2  | 0 | -  | -  |
| FAMILY | k141_4712_1 |       | PF01868 | UPF0086         | 20.1      | 4.20E-4  | 4 | F:GO:0003723; P:GO:0006396; C:GO:0030677; F:GO:0004540 | F:ribonuclease activity; F:RNA binding; C:ribonuclease P complex; P:RNA processing                             |
| FAMILY | k141_4722_1 |       | PF08223 | PaaX_C          | 13.7      | 5.40E-2  | 0 | -  | -  |
| FAMILY | k141_4735_1 |       | PF06374 | NDUF_C2         | 14.2      | 3.30E-2  | 3 | C:GO:0005743; P:GO:0006120; F:GO:0008137               | P:mitochondrial electron transport, NADH to ubiquinone; C:mitochondrial inner membrane; F:NADH dehydrogenas... |
| FAMILY | k141_4735_1 |       | PF00535 | Glycos_transf_2 | 31.9      | 9.60E-8  | 0 | -  | -  |
| FAMILY | k141_4759_1 |       | PF00534 | Glycos_transf_1 | 34        | 1.90E-8  | 0 | -  | -  |
| FAMILY | k141_4786_2 |       | PF17528 | DUF5449         | 6.2       | 7.30E0   | 0 | -  | -  |

結果として、  
検索されたGOデータが表示されます。

**GO:0005524**

**Name**  
ATP binding

**MF**

**Definition**  
Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator.

ATP binding

Term Information

Accession: GO:0005524  
Name: ATP binding  
Ontology: molecular\_function  
Synonyms: None  
Alternative IDs: None  
Definition: Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator. Source: ISBN:0198506732  
Comment: None  
History: See term history for GO:0005524 at QuickGO  
Subset: None  
Related: [Link](#) to all genes and gene products annotated to ATP binding  
[Link](#) to all direct and indirect annotations to ATP binding  
[Link](#) to all direct and indirect annotations download (limited to first 10,000) for ATP binding

Annotations: Graph Views | Inferred Tree View | Neighborhood | Mappings

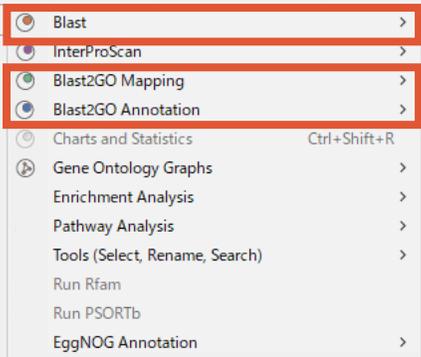
Filter results

Total annotations: 47307, showing 1-10  
Results count: 10

| Gene/product | Gene/product name       | Annotation qualifier | GO class   | Annotation extension      | Contributor | Organism             | Evidence                            | Evidence with        | PANTHER family | Type    | isoform | Reference     | Date     |
|--------------|-------------------------|----------------------|------------|---------------------------|-------------|----------------------|-------------------------------------|----------------------|----------------|---------|---------|---------------|----------|
| AGAD3RVUS    | Uncharacterized protein | ATP binding          | GO_Central | Pan troglodytes           | IBA         | CGD:CAL800197369     | pos9 activating factor bap7-related | PANTHER:PTN000027630 |                | protein |         | PMID:21873035 | 20170228 |
| HSP70A       | Heat shock protein 70A  | ATP binding          | GO_Central | Chlamydomonas reinhardtii | IBA         | PANTHER:PTN000321897 | heat shock protein 70-class b       | PANTHER:PTN000321897 |                | protein |         | PMID:21873035 | 20170831 |

個々のシーケンスデータにおける  
アノテーションの詳細の閲覧も可能で、  
リンクアウトからアノテーション付きGOに関する  
詳細情報を確認できます。

File View Help



3ステップで簡単に解析することができます。

Blast2GO方法論によるアノテーション情報を付与は分析アイコン「functional analysis」より実行できます。

Blast2GO方法論は7000件以上の研究引用の実績があります。詳しくは以前のセミナー資料をご参照ください。

フィルジエン WEBセミナー

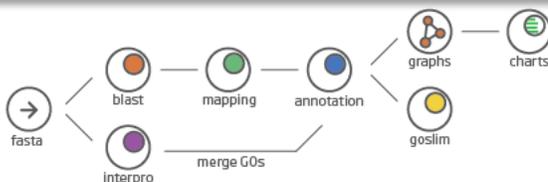
## 高速BLAST・アノテーション解析

～OmicsBoxを使用した解析機能のご紹介～



[詳しくはこちら](#)

フィルジエン Webセミナー（2019年4月18日）



Genome Characterization

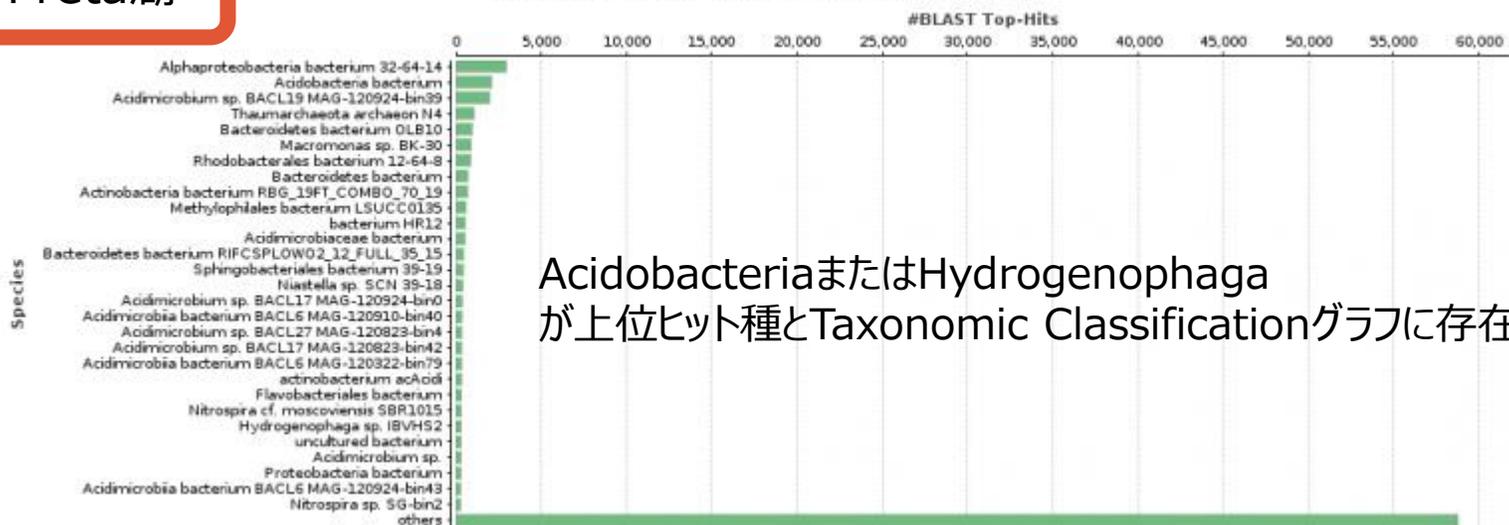
- New Taxonomic Classification Statistics: Chao1 OTU Diversity, Rarefaction Curves and PCoA Plot
- OTU Differential Abundance Testing with edgeR

### Transcriptomics Module

- Completeness Assessment with BUSCO
- Predict Coding Regions with TransDecoder
- Improved RNA-Seq Alignment Input: GFF is automatically converted to GTF
- New RNA-Seq Alignment Options: 2-Pass mode, Export Spliced

## Preta湖

Top-Hit Species Distribution [pab\_proteins]



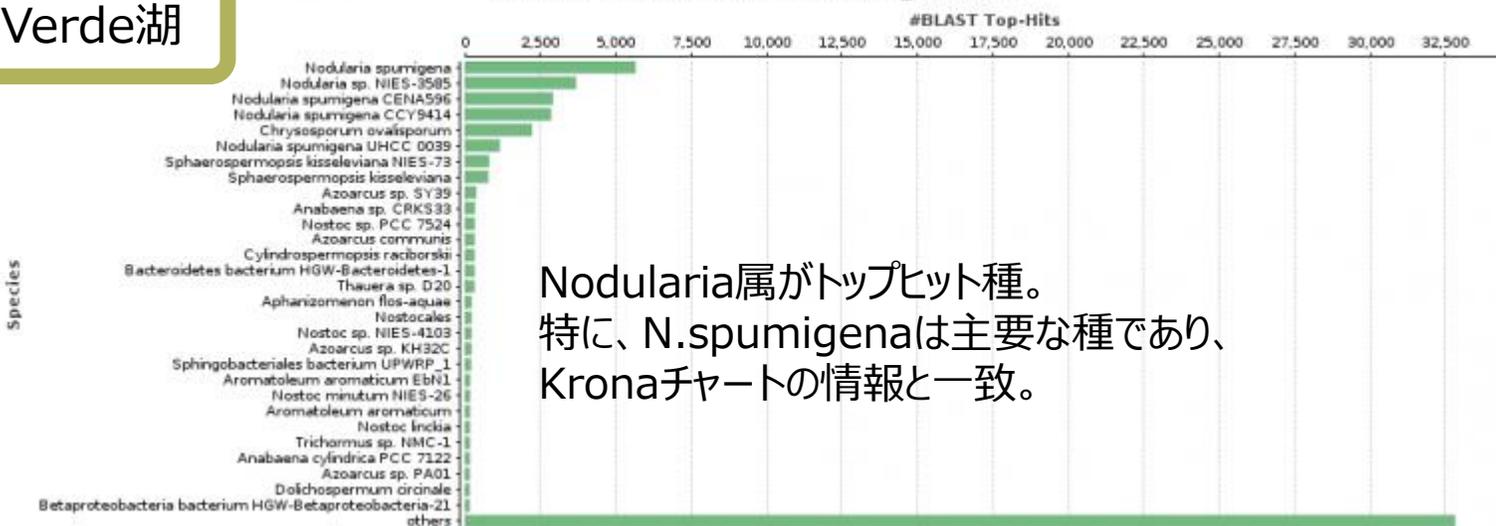
AcidobacteriaまたはHydrogenophaga  
が上位ヒット種とTaxonomic Classification  
グラフに存在

Top-Hit Species分布図を作成できます。

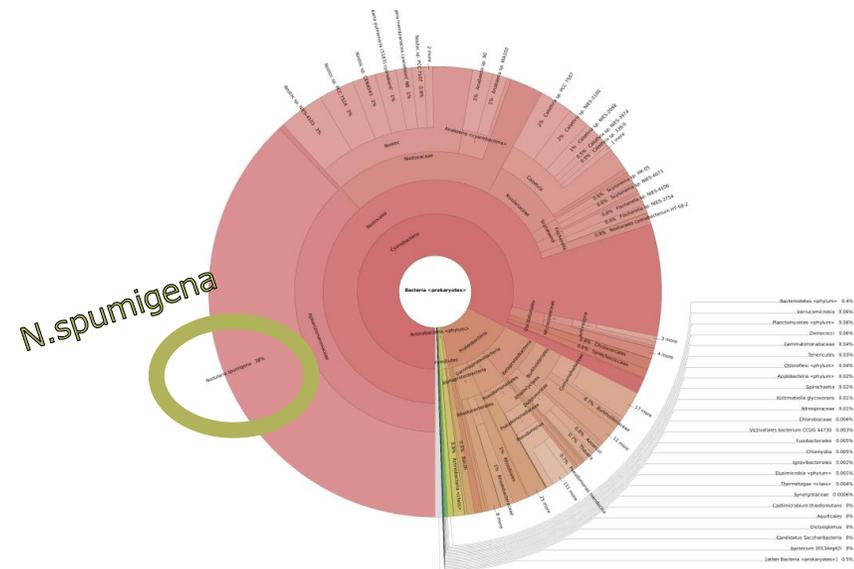
→Taxonomic Classification (種)  
の結果と非常に似ていた。

## Verde湖

Top-Hit Species Distribution [vab\_proteins]



Nodularia属がトップヒット種。  
特に、N.spumigenaは主要な種であり、  
Kronaチャートの情報と一致。



Verde湖Kronaチャート

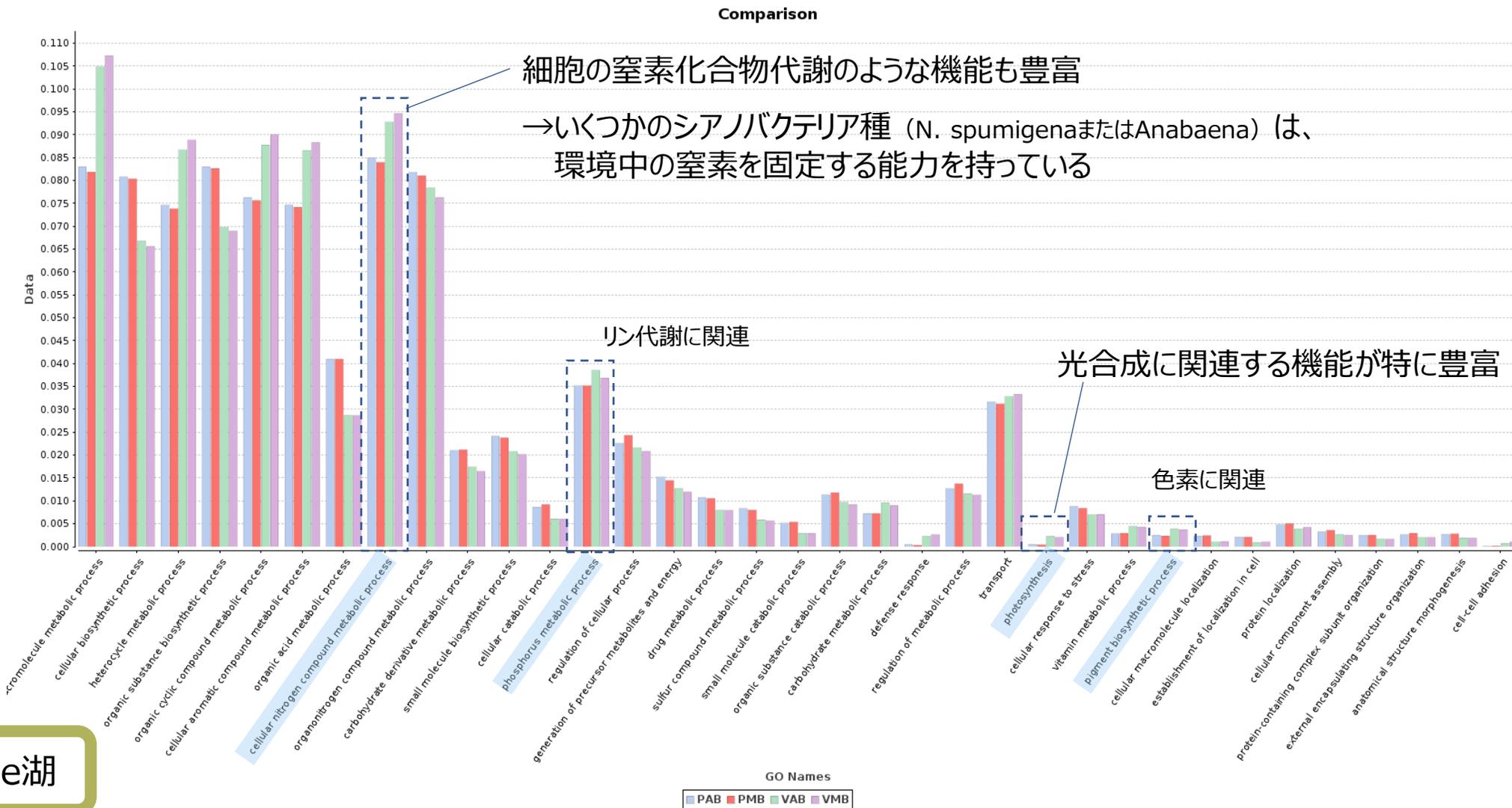
# 機能解析

各アノテーションアプローチは異なるGOtermを生成するため、各サンプルのすべての情報をマージしました。このステップによりサンプルごとに1つ、4つのアノテーションができました。

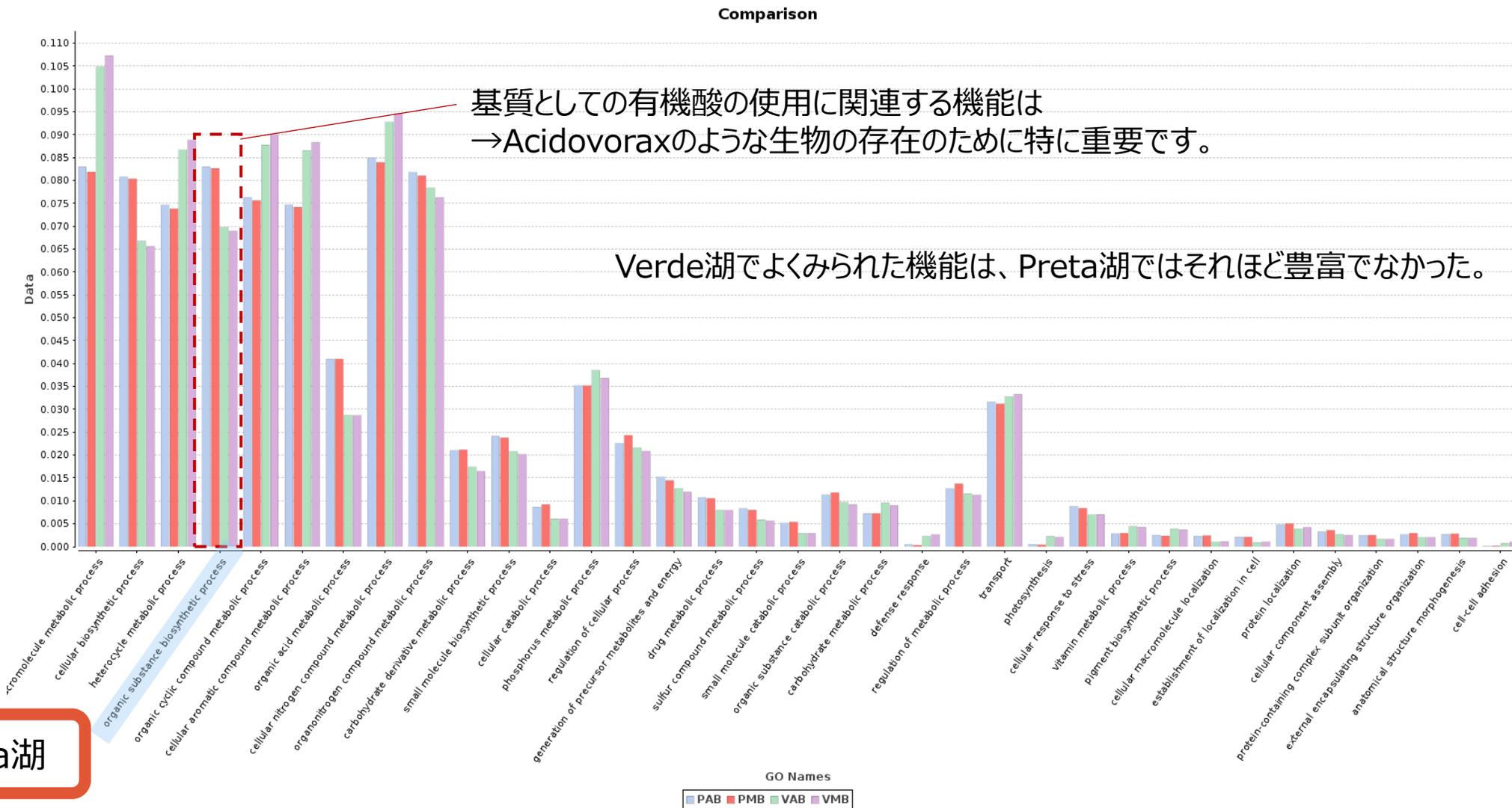
|                        |                      | PAB           | PMB           | VAB           | VMB           |
|------------------------|----------------------|---------------|---------------|---------------|---------------|
| <b>Predicted genes</b> |                      | 71557         | 76640         | 42880         | 59839         |
| <b>EggNOG</b>          | <b>Ann. genes</b>    | 18075         | 19318         | 9333          | 14379         |
|                        | <b>% annot.</b>      | 25.26%        | 25.21%        | 21.77%        | 24.03%        |
| <b>PfamScan</b>        | <b>F. structures</b> | 31651         | 36285         | 15195         | 34736         |
|                        | <b>% charact.</b>    | 44.23%        | 47.34%        | 35.44%        | 58.04%        |
| <b>Blast2GO</b>        | <b>Ann. genes</b>    | 48061         | 55372         | 27291         | 35725         |
|                        | <b>% annot.</b>      | 67.16%        | 72.25%        | 63.65%        | 59.70%        |
| <b>FINAL</b>           | <b>Ann. genes</b>    | <b>66139</b>  | <b>74694</b>  | <b>36628</b>  | <b>50123</b>  |
|                        | <b>% annot.</b>      | <b>92.43%</b> | <b>97.46%</b> | <b>85.42%</b> | <b>83.76%</b> |

✓高品質なアノテーション情報の付与

最後に、各湖の機能構成を調べて比較するための最終的な注釈を含むサンプル比較チャート  
 (「Metagenomics」>「Comparative Analysis」>「Sample Comparison Chart」) を取得しました。



最後に、各湖の機能構成を調べて比較するための最終的な注釈を含むサンプル比較チャート（「Metagenomics」>「Comparative Analysis」>「Sample Comparison Chart」）を取得しました。





- OmicsBoxを使用したメタゲノム解析により、湖に生息する微生物群集を特定し、機能的な遺伝的潜在能力を測定することができました。
- Preta湖とVerde湖の分類学的および機能的な構成は、主にVerde湖で発生するシアノバクテリアのブルームのため、非常に異なっていました。
- Verde湖の一部の生物は、光合成と窒素固定プロセスを実行する可能性がありました。これらの機能は、Preta湖では重要ではありません。

# 手間を省き 高品質なアノテーション

✓実績の高いオープンソースソフトウェアを簡単に使用することができます

6.7M  
INPUT  
READS

250,000  
PREDICTED  
GENES

90%  
HIGH-QUALITY  
ANNOTATED

# 高速計算

✓インターネット接続さえあれば高価なPCは必要ありません

6時間  
処理時間

同時に解析！

- 分類学的分類：サンプルあたり15分
- メタゲノムの組み立て：サンプルあたり5分
- 遺伝子予測：サンプルあたり10分
- 機能アノテーション
- PfamScan：サンプルあたり30分
- EggNOG-Mapper：サンプルあたり2時間
- NRに対するBlastP：サンプルあたり1～5.5時間

(合計150万クラウドユニット)



## Functional Analysis



### 機能アノテーションツール

- BlastとInterProの高速解析
- 機能アノテーション情報付与
- エンリッチメント解析

## Genome Analysis



### 新規ゲノムの配列決定

- De-Novo Assembly
- Repeat Masking
- 真核生物 原核生物のORF領域の遺伝子予測

## Transcriptomics



### RNA-seqデータ解析

- De-Novo Assembly
- 発現値定量（モデル生物・非モデル生物対応）
- 発現変動遺伝子の同定

## Metagenomics



### メタゲノム解析 (16S・WGS)

- Taxonomic Classification
- OTU Abundances Table
- Metagenomic Assembly/遺伝子予測

## Information

[弊社HPでの紹介ページ](#)

[カタログ](#)

OmicBoxは、1週間使用可能なデモライセンスがあります。  
ご希望の場合はbiosupport@filgen.jpまでお問い合わせください。

**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](mailto:biosupport@filgen.jp)