

2025年2月26日(水)15:00 – 17:00

TARA FD Seminar

第1回インシリコバイオロジーデータ解析講習会

細菌叢解析入門

本講習会では、事前に登録が必要です。
まだ登録されていない方は、以下のURLか
QRコードから登録してください。
ご協力をお願いいたします。

<https://forms.office.com/r/3dmYH1DfAt>



賴本 隼汰 (Shunta Yorimoto), Ph.D.

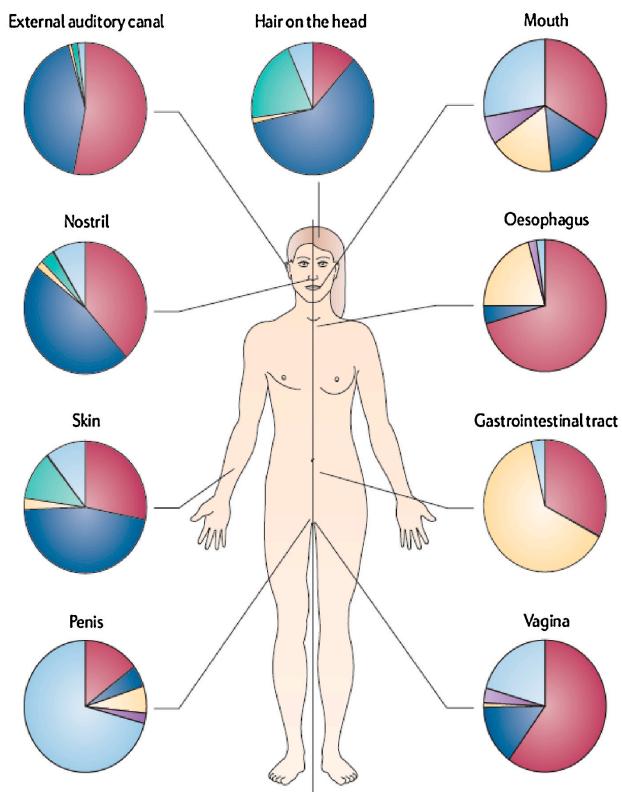
筑波大学 TARAセンター 生命情報統合部門

微生物叢 microbiota

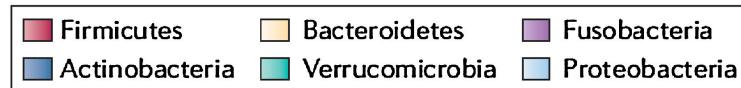
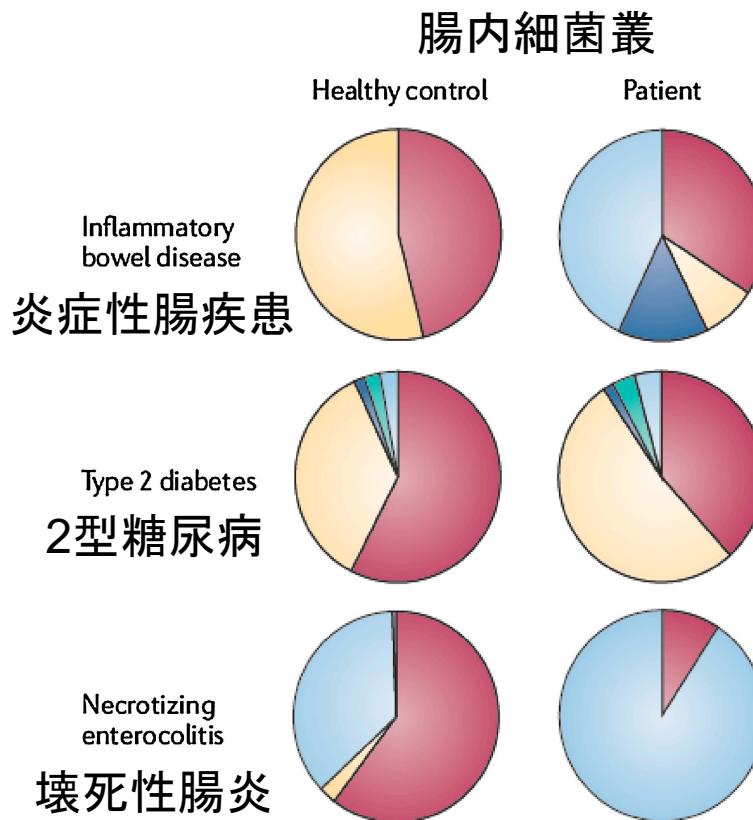
- 微生物叢 (microbiota)
特定の環境中に存在する微生物集団の総体
細菌、古細菌、真菌、原生生物、微細藻類、(ウイルス)が含まれる
- マイクロバイオーム (microbiome, microbio + -ome)
微生物集団に含まれる遺伝情報の総体

Human Microbiome Project

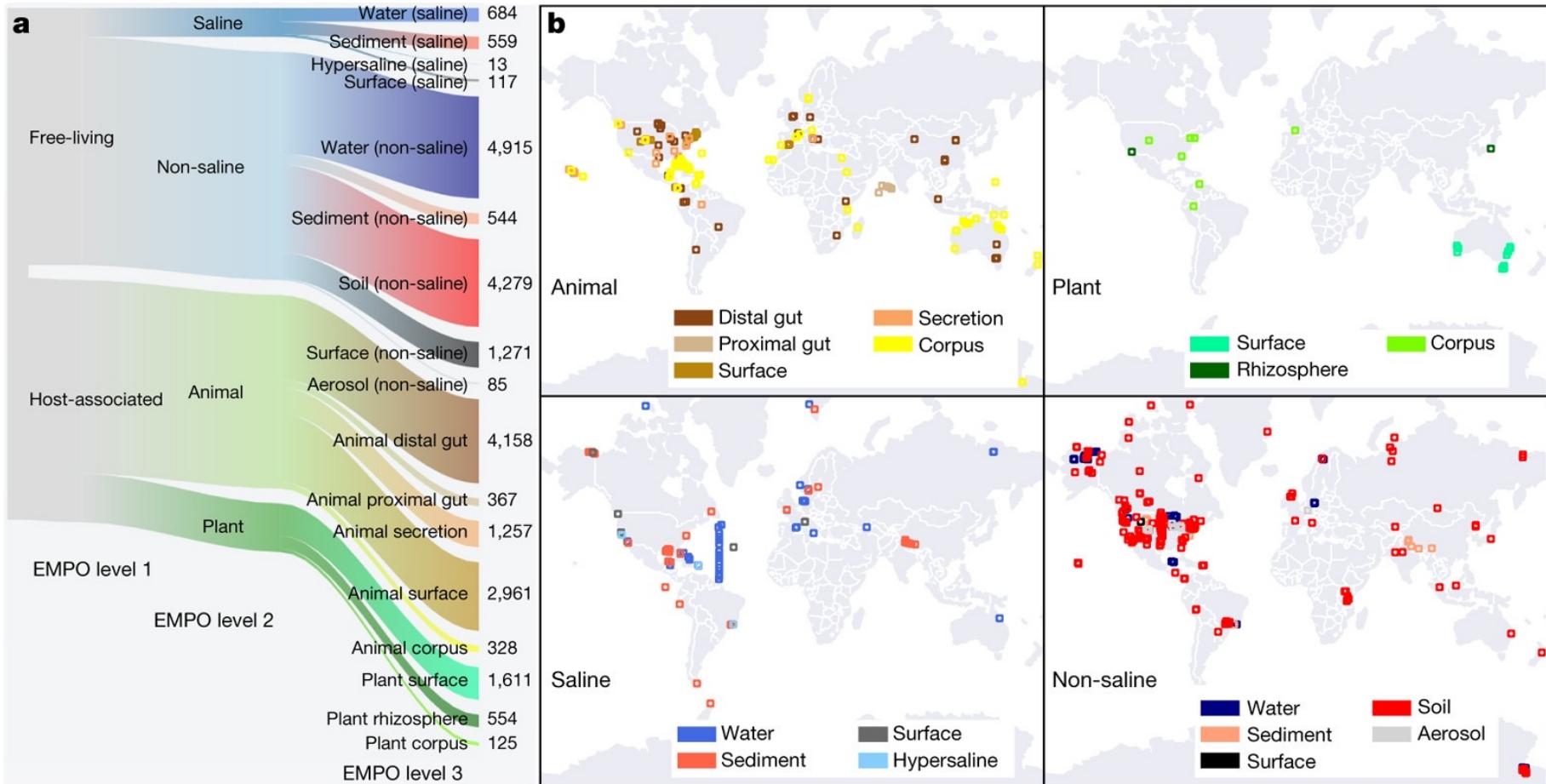
A



B



Earth Microbiome Project



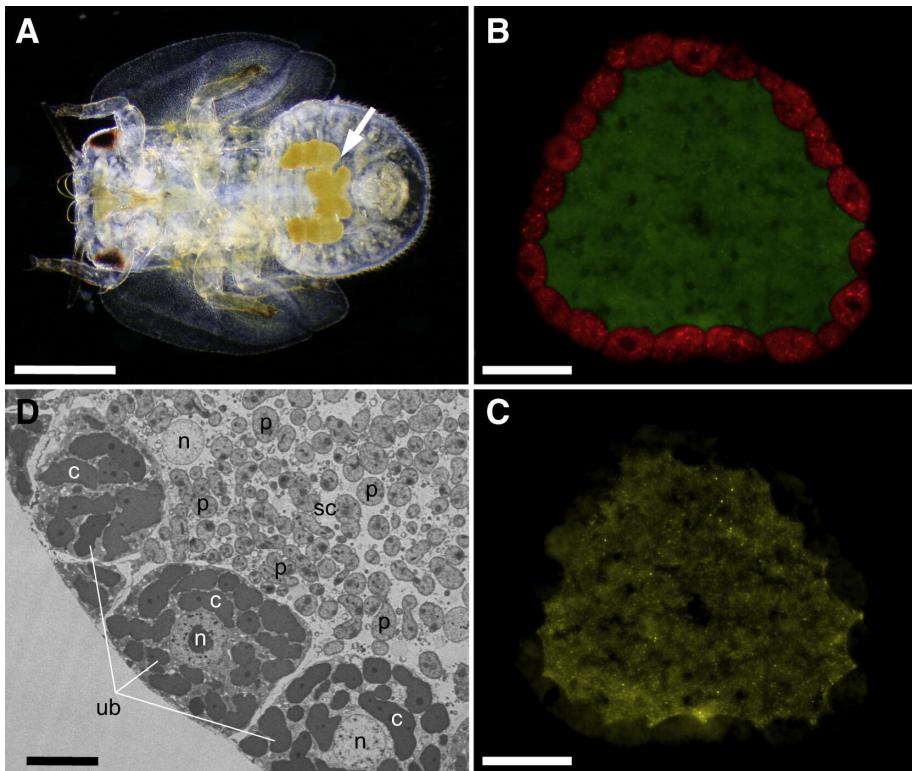
23,828 samples from 7 continents, 43 countries, and 17 environments

Thompson et al., 2017

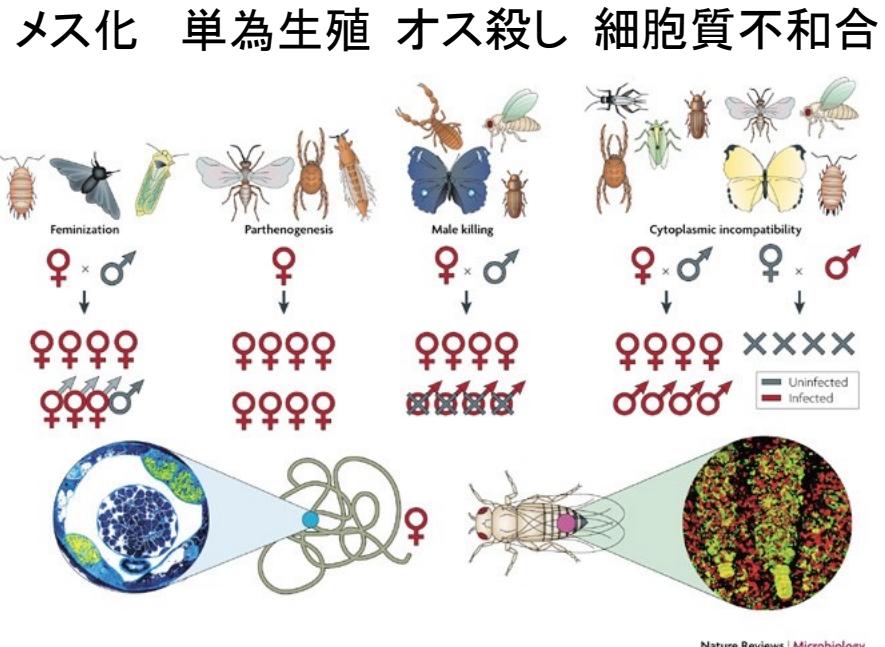
共生微生物

- 栄養の供給
- 毒産生による防御

- 宿主の性・生殖操作



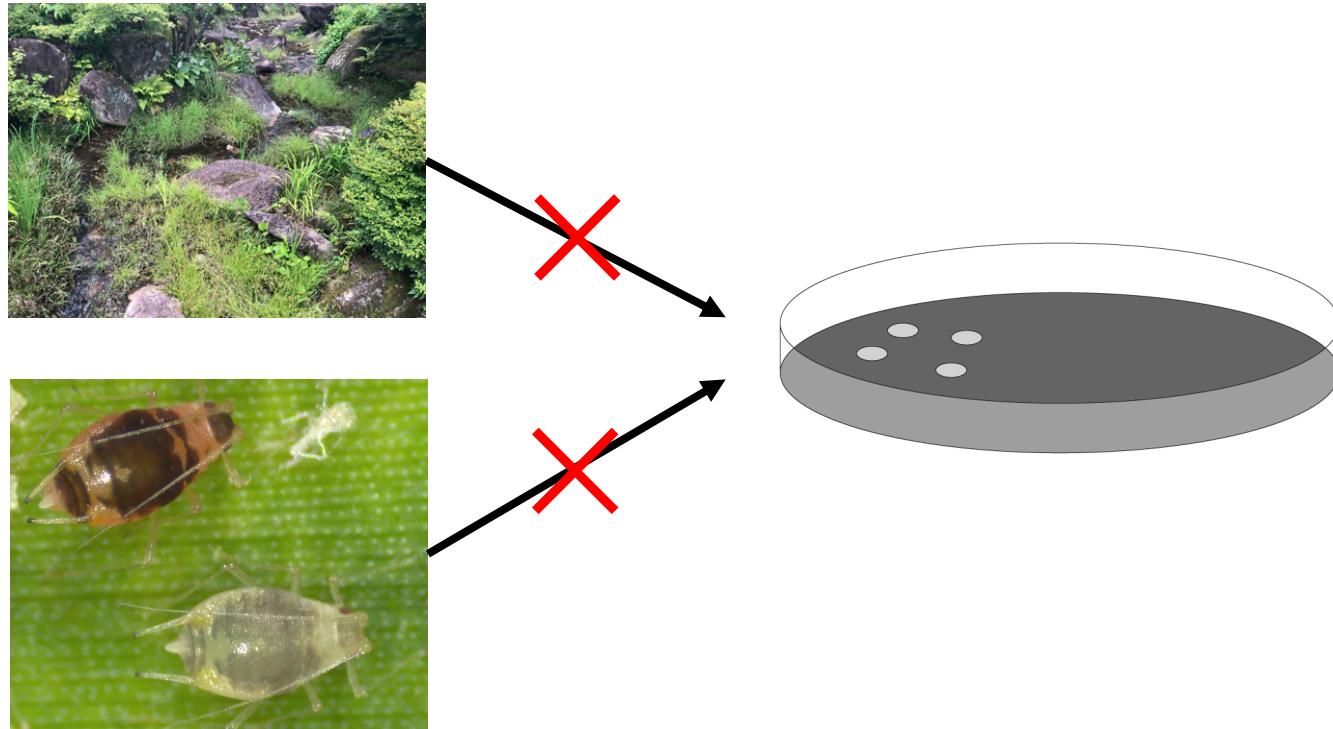
Nakabachi et al., 2013



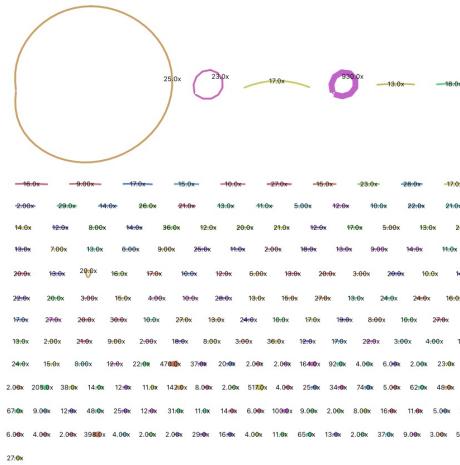
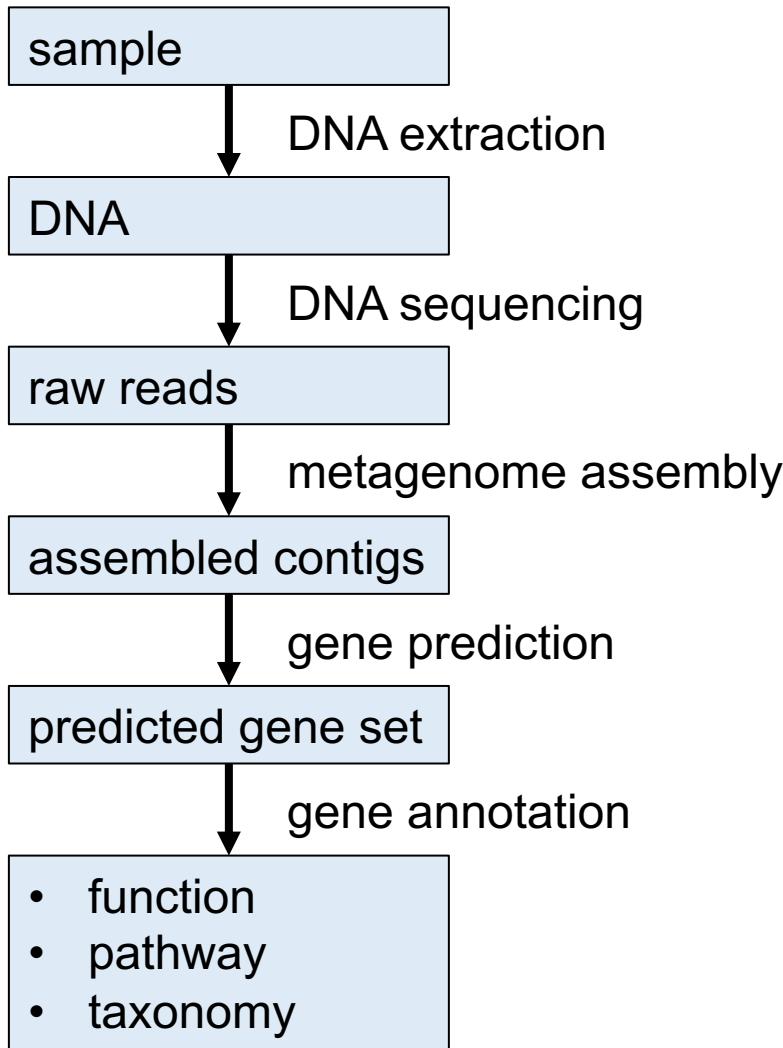
Werren et al., 2008

ほとんどの微生物は培養できない

- 栄養条件(炭素、窒素、ミネラル、ビタミン、アミノ酸、塩基類)
- 生育条件(温度、酸素、pH、塩濃度、浸透圧、圧力、光)
- 培養方法(純粋培養、集積培養、液体・個体培地、etc...) 坂本、2008



Metagenome 解析



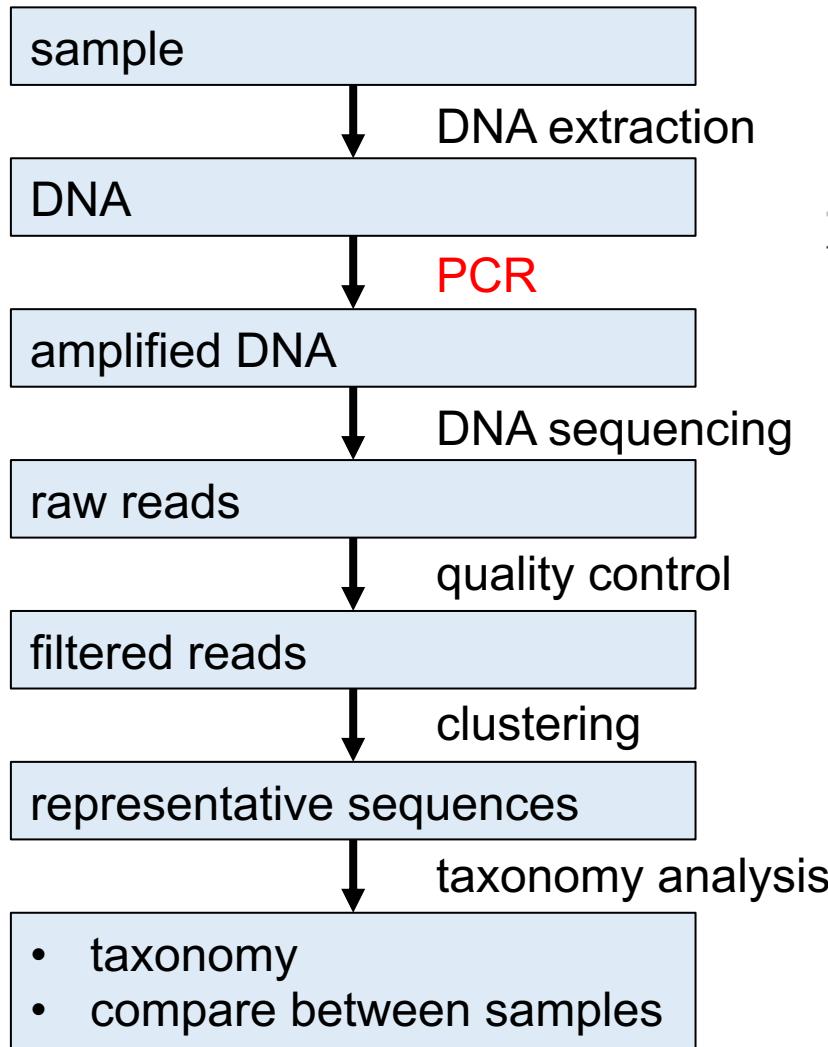
利点 :

- 系統組成と機能的な情報が得られる
- 宿主を含むhologenomeの再構築が可能
- 新規遺伝子を発見する可能性

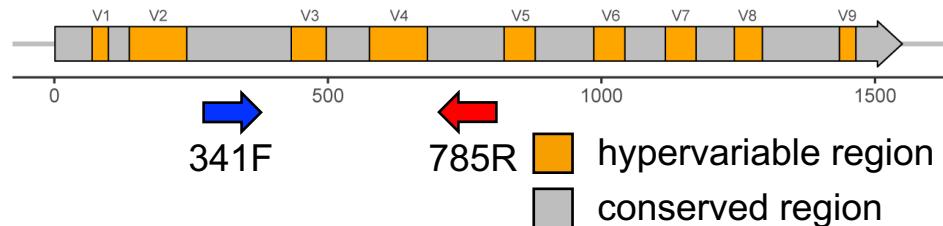
欠点 :

- シーケンスコストが高い
- 宿主DNAの混入の影響
- 低存在量の種の検出が困難
- 解析が複雑で専門知識が必要

16S ribosomal RNA gene amplicon sequencing 解析



16S rRNA gene, conserved in most bacteria and archaea



利点:

- ・ シーケンスコストが安い
- ・ 少量のDNAサンプルでも実施可能
- ・ 他検体の比較に適している
- ・ 解析パイプラインが確立されている

欠点:

- ・ PCRバイアス
- ・ コピー数、倍数化による定量性の問題
- ・ 種レベル以下の分解能に限界

Amplicon primersの設計

JOURNAL ARTICLE

Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies ⓘ

Anna Klindworth, Elmar Pruesse, Timmy Schweer, Jörg Peplies, Christian Quast, Matthias Horn,
Frank Oliver Glöckner ✎

Nucleic Acids Research, Volume 41, Issue 1, 1 January 2013, Page e1,
<https://doi.org/10.1093/nar/gks808>

Published: 28 August 2012 Article history ▾

ex) 16S rRNA V3–V4領域 : 341F–785R

Forward:

5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-CCTACGGGNGGCWGCAG

degenerate primers

Reverse:

5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-GACTACHVGGGTATCTAATCC

5' [Illumina adapter overhang sequence]-[locus-specific sequence]

下線: IUPAC ambiguity codes

https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

真核生物でもamplicon sequencingは可能である

Amplicon primersの条件

- 解析対象となる系統群ですべての種が保持している領域
- よく保存されている領域と変異の入りやすい領域が存在
- シーケンス可能な増幅長

真核生物でよく使われている対象領域

- 18S rRNA gene
- 28S rRNA gene
- MT-CO1 (mitochondrial cytochrome c oxydase subunit 1)
- MT-CO2
- ITS (Internal transcribed spacer)

18S rDNA – ITS – 5.8S rDNA – ITS 28S rDNA

(注: 菌種により長さは著しく異なるため、増幅できないものもある)

全長16S rRNA gene sequencingにより種レベルの同定が可能になった

Research | [Open access](#) | Published: 16 February 2024

Optimized bacterial community characterization through full-length 16S rRNA gene sequencing utilizing MinION nanopore technology

[Alessandro Bertolo](#), [Ezra Valido](#) & [Jivko Stoyanov](#)✉

BMC Microbiology 24, Article number: 58 (2024) | [Cite this article](#)

7288 Accesses | 8 Altmetric | [Metrics](#)

Research | [Open access](#) | Published: 25 March 2024

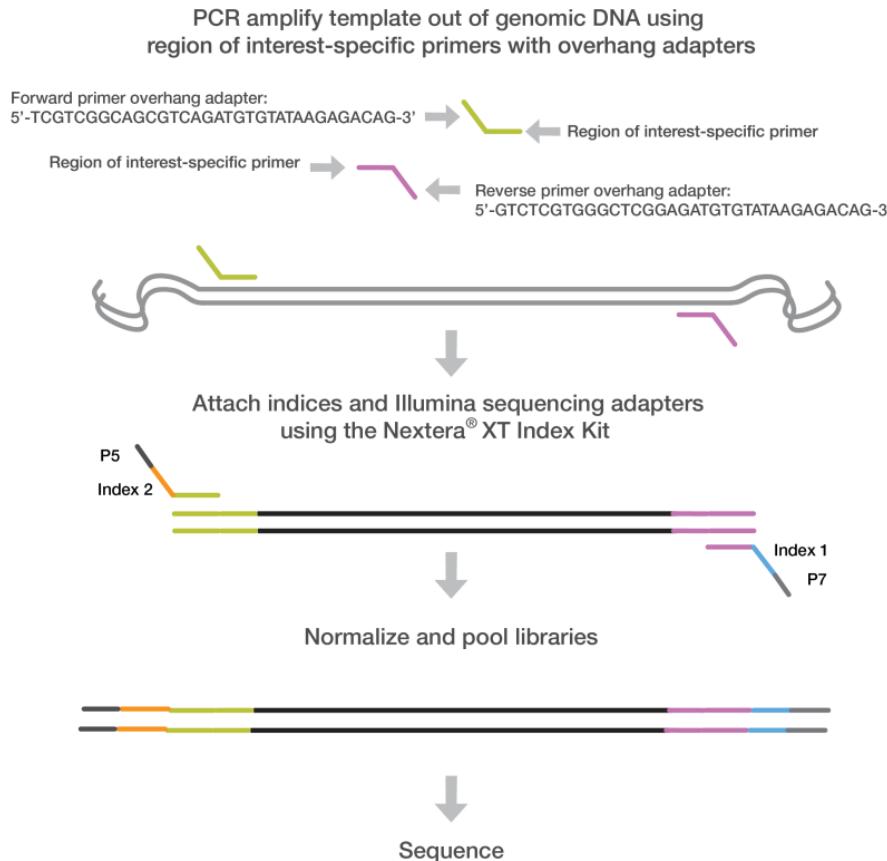
Full-length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human microbiome samples

[Elena Buetas](#), [Marta Jordán-López](#), [Andrés López-Roldán](#), [Giuseppe D'Auria](#), [Llucia Martínez-Priego](#), [Griselda De Marco](#), [Miguel Carda-Diéguex](#)✉ & [Alex Mira](#)

BMC Genomics 25, Article number: 310 (2024) | [Cite this article](#)

7703 Accesses | 6 Altmetric | [Metrics](#)

Primer trimming and quality filtering



https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

FASTQ format

@M01037...

CCTACGGGNGGCWGCAG...

+

>AA1>111ADD10A0FGBG...

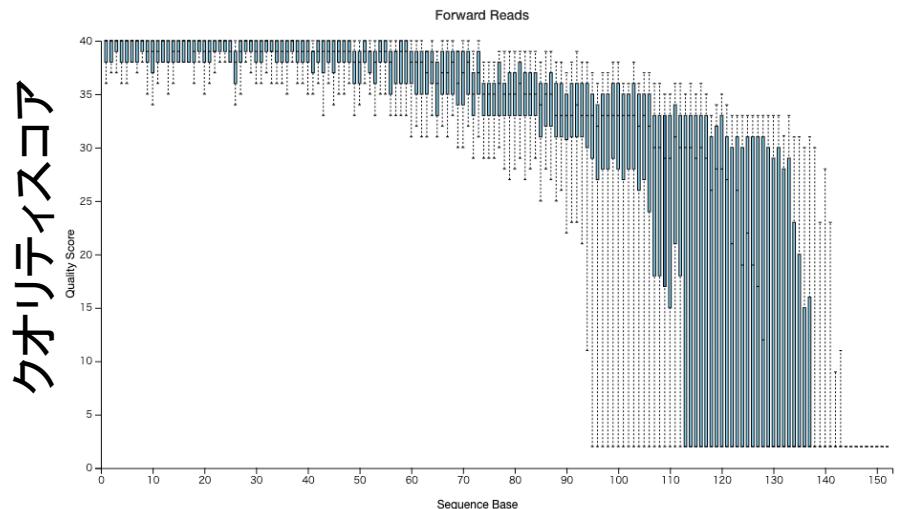
CCTACGGGNGGCWGCAG

@配列ID

塩基配列

+

クオリティスコア



塩基配列

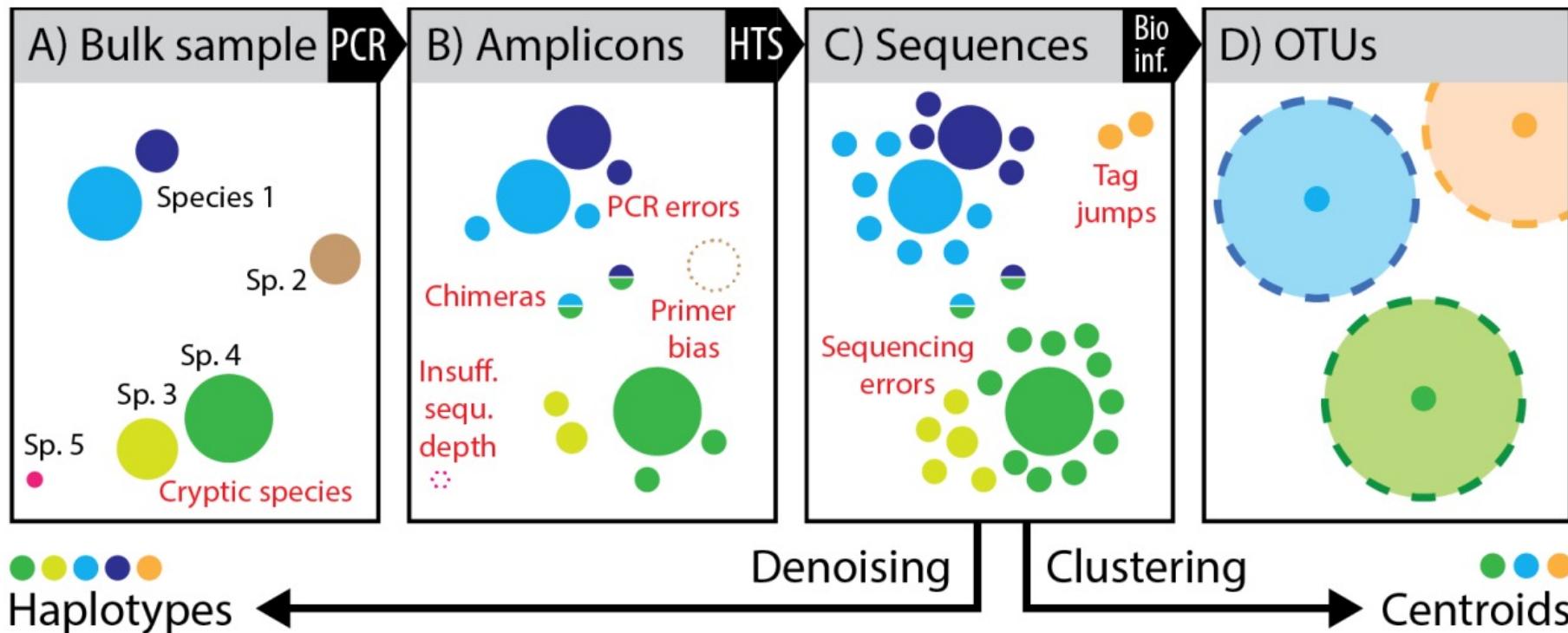
<https://docs.qiime2.org/2024.10/tutorials/moving-pictures/>

Operational Taxonomic Unit (OTU)/Amplicon Sequence Variant (ASV)

OTU: 特定の配列類似度(一般的に97%)でクラスタリングされた配列群

ASV: エラーを統計的に処理した生物学的に真の塩基配列バリエント

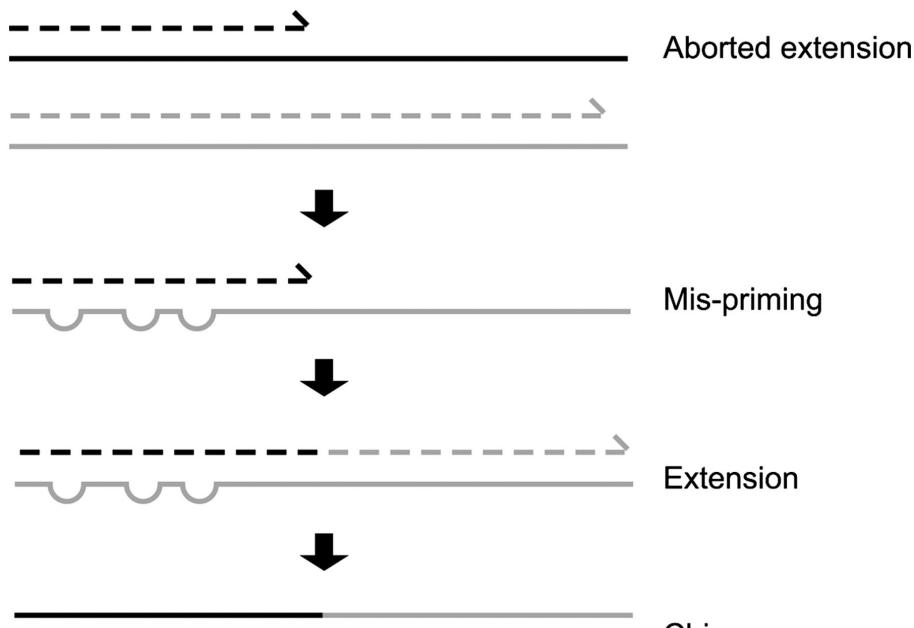
昔はよくOTUが使われていたが、NGSが当たり前の現在ではASVが主流



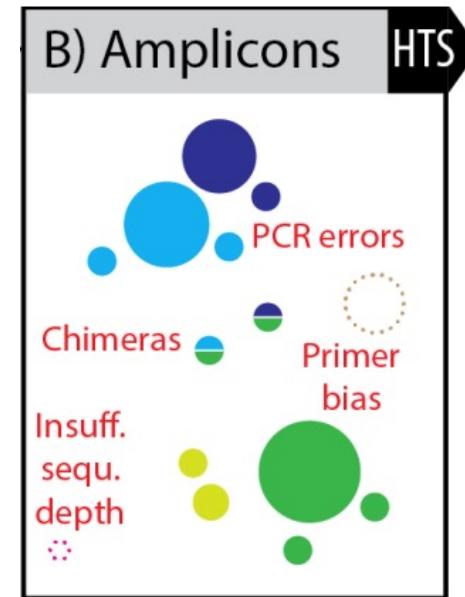
Elbrecht et al., 2018

Removal of chimeric sequences

- PCRの過程で異なる生物由来の配列が結合することがある
- 16S rRNA geneは配列が似ているのでキメラ配列が生じやすい
- 存在量の多いASVを親配列とみなし、検査対象の配列が2つの異なる親配列から再構成可能かを確認する。

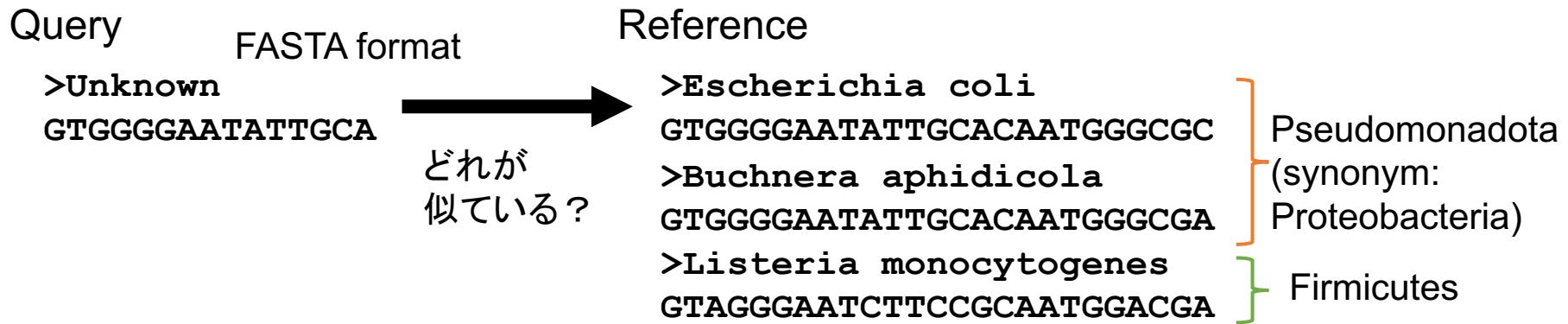


Haas et al., 2011



Elbrecht et al., 2018

Taxonomy assignment and reference database

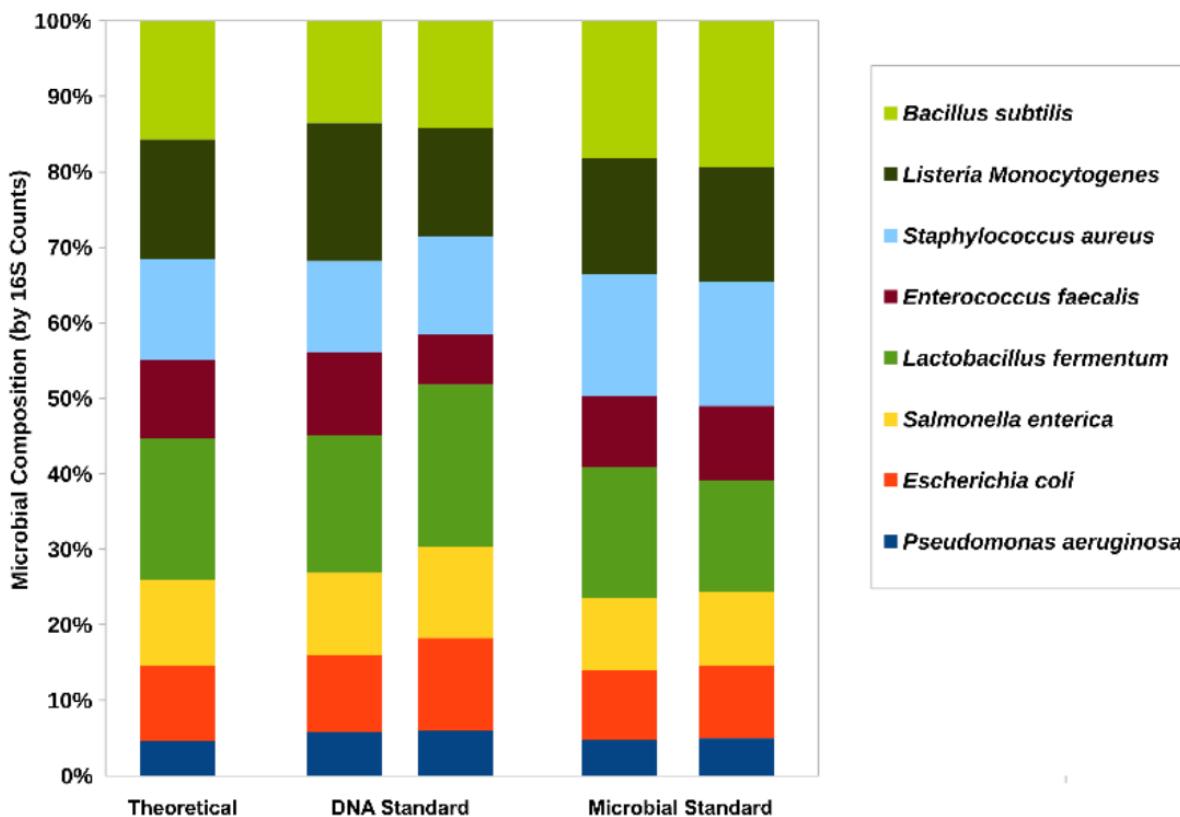


16S rRNA gene reference database

- SILVA <https://www.arb-silva.de/>
定期的な更新により最新の分類情報を提供する
SILVA論文 (Quast et al., 2012) の被引用数は28,030 (google scholar)
Last update: 2024年7月
- Greengenes2 <https://greengenes2.ucsd.edu/>
Greengenesは2017年以降更新が停止していたが、最近Greengenes2を公開
Greengenes論文 (DeSantis et al., 2006) の被引用数は12,058 (google scholar)
Last update: 2024年9月

ZymoBIOMICS Microbial Community DNA Standard as a positive control

ZymoBIOMICS Microbial Community DNA Standards:
純粋培養された8種の細菌と2種の真菌から抽出したDNAを
既知の比率で混合調製した参考サンプル



https://files.zymoresearch.com/protocols/_d6400_quick-16s_ngs_library_prep_kit.pdf

細菌叢解析ツール

- QIIME2 <https://qiime2.org/>
現在最も広く使用されている解析プラットフォームの一つ
プラグインとして機能拡張することができ、例えば内部でDADA2を利用できる
系統解析、多様性解析、統計解析、可視化などさまざまな用途に使用できる
Python based
QIIME2論文 (Bolyen et al., 2019) の被引用数は17,550 (google scholar)
- DADA2 <https://benjneb.github.io/dada2/index.html>
ASVを実装したのが画期的
使い方もシンプル(R Bioconductor packageとして提供されている)
DADA2論文 (Callahan et al., 2016) の被引用数は26,565 (google scholar)
- Mothur <https://mothur.org/>
古くからある解析プラットフォーム
C and C++ based
Mothur論文 (Schloss et al., 2009) の被引用数は21,727 (google scholar)



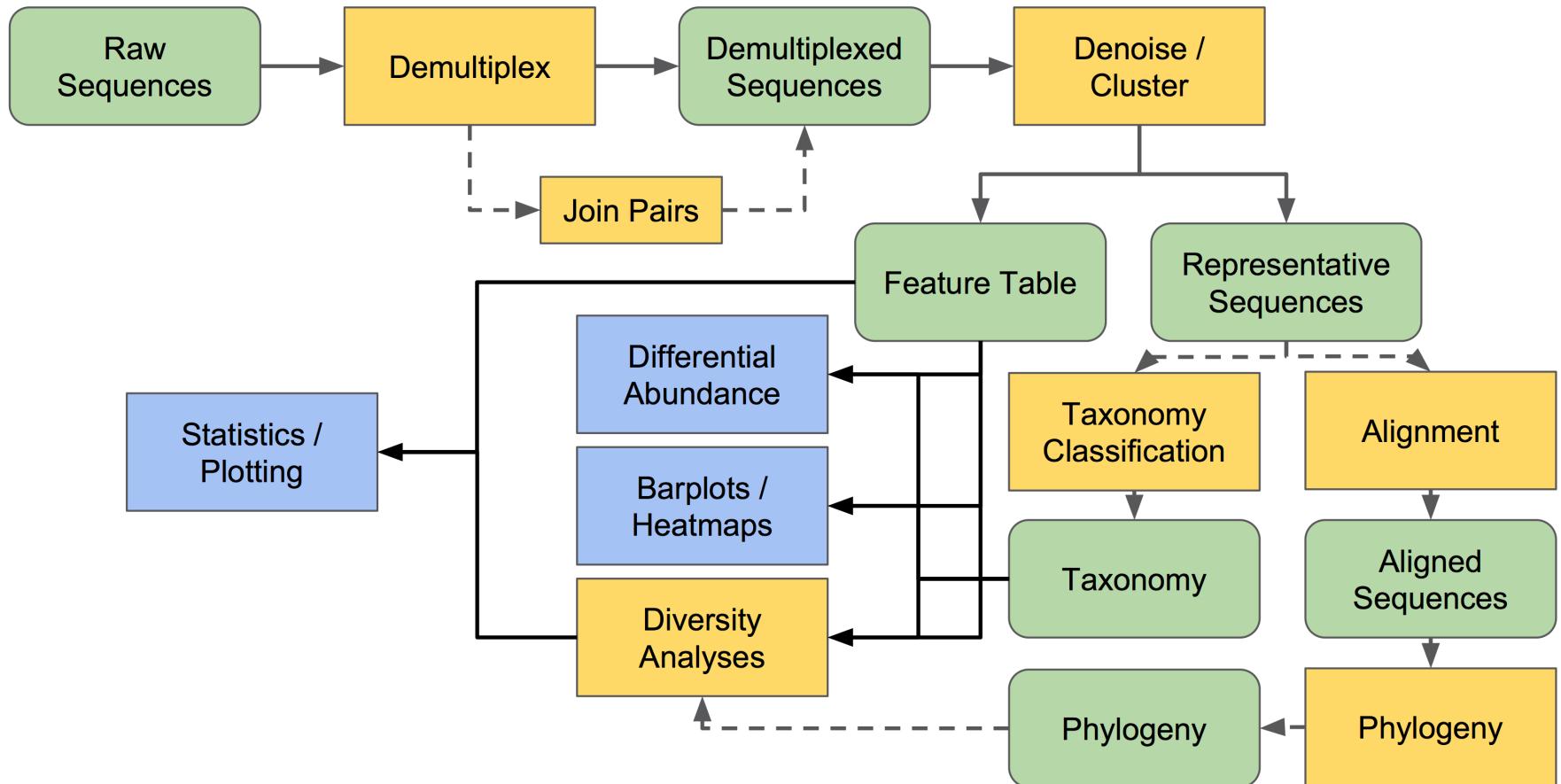
[QIIME 2™](#) (pronounced “chime two”) is a microbiome multi-omics bioinformatics and data science platform that is trusted, free, open source, extensible, and community developed and supported.

Take a look through the thousands of published research articles, hundreds of patent applications, and tens of clinical trial records that reference the project.

Topic		Community-Contributions		
	Replies	Views	Activity	
↳ About the Community Contributions category	0	1.1k	about 7 years ago	
q2-kmerizer: a QIIME 2 plugin for k-mer-based diversity analysis	1	23	4 days ago	
Calculating dispersion using q2-convexhull longitudinal-beta-pcoa	0	29	10 days ago	
Processing, filtering, and evaluating the SILVA database (and other reference sequence data) with RESCRIPT taxonomy·feature-classifier·rescript·plugin-development·tutorials	16	56.3k	17 days ago	
Artifact Cache Basic Instructions	0	163	9 months ago	
PRONAME: Enhancing Taxonomic Accuracy with Nanopore Long-Read Metabarcoding and QIIME2	0	126	about 1 month ago	
How to train a GTDB SSU classifier using RESCRIPT taxonomy·feature-classifier·rescript·tutorial·gtdb-ssu	1	1.9k	about 1 month ago	
Q2-ITSxpress: A tutorial on a QIIME 2 plugin to trim ITS sequences its·tutorial·fungi	24	21.8k	about 2 months ago	

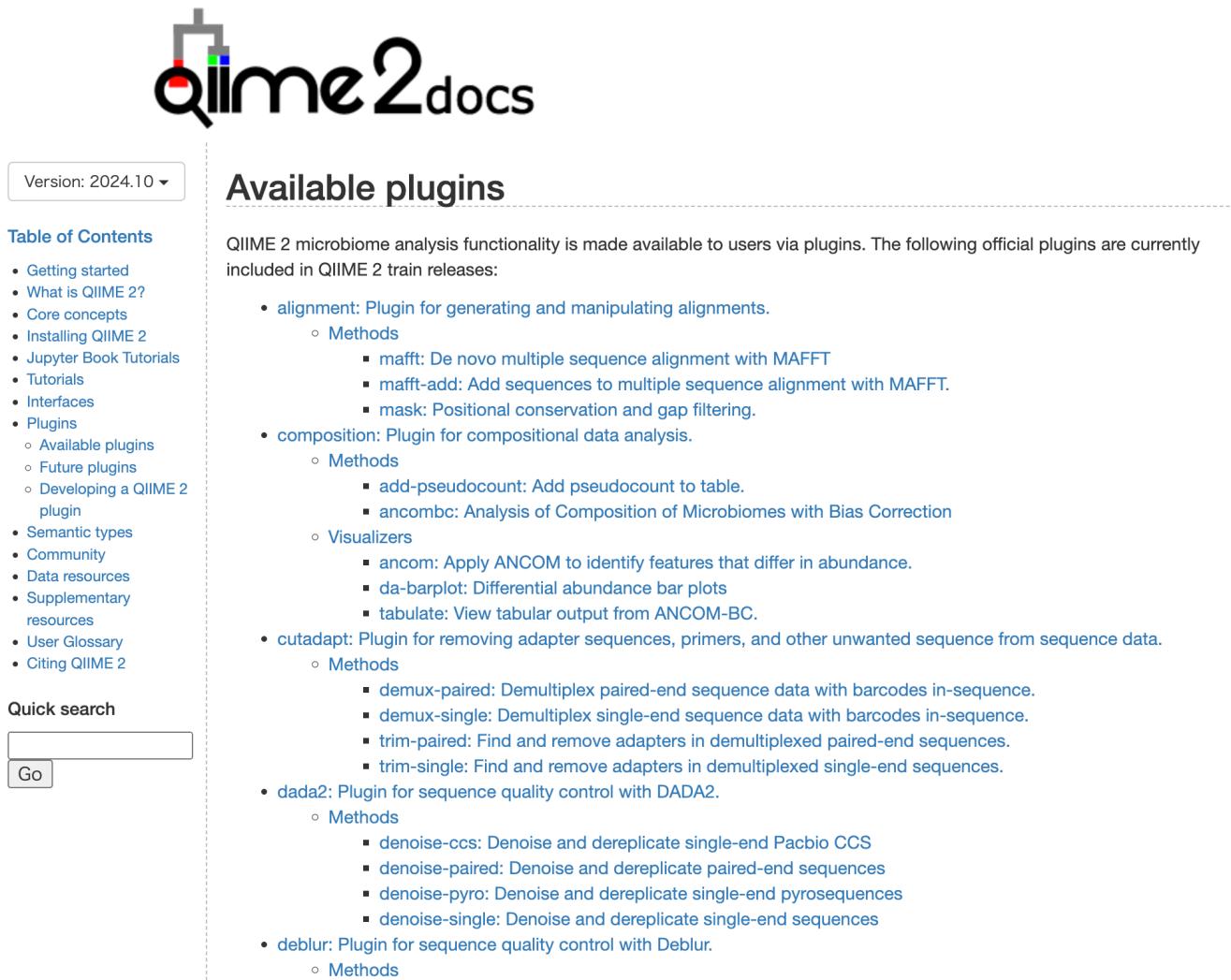
<https://qiime2.org/>

Conceptual overview of QIIME2



<https://docs.qiime2.org/2024.10/tutorials/overview/>

QIIME2はさまざまな機能拡張が可能



The screenshot shows the QIIME2 documentation website. At the top, there's a navigation bar with a search icon, a user icon, and a "Logout" button. Below the header, the main content area has a large title "Available plugins". To the left, there's a sidebar with a "Table of Contents" and a "Quick search" bar.

Table of Contents

- Getting started
- What is QIIME 2?
- Core concepts
- Installing QIIME 2
- Jupyter Book Tutorials
- Tutorials
- Interfaces
- Plugins
 - Available plugins
 - Future plugins
 - Developing a QIIME 2 plugin
- Semantic types
- Community
- Data resources
- Supplementary resources
- User Glossary
- Citing QIIME 2

Quick search

Version: 2024.10 ▾

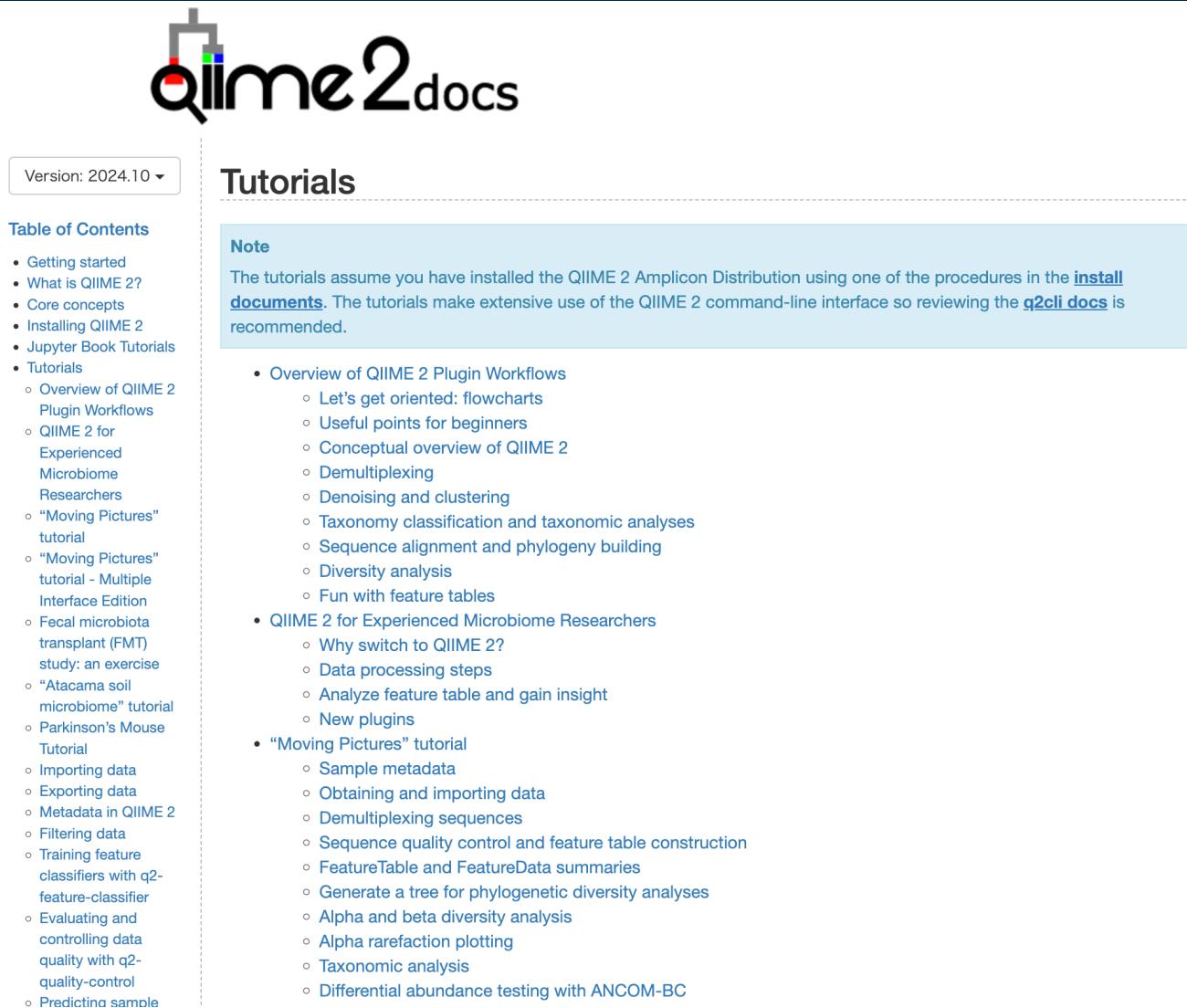
Available plugins

QIIME 2 microbiome analysis functionality is made available to users via plugins. The following official plugins are currently included in QIIME 2 train releases:

- alignment: Plugin for generating and manipulating alignments.
 - Methods
 - mafft: De novo multiple sequence alignment with MAFFT
 - mafft-add: Add sequences to multiple sequence alignment with MAFFT.
 - mask: Positional conservation and gap filtering.
- composition: Plugin for compositional data analysis.
 - Methods
 - add-pseudocount: Add pseudocount to table.
 - ancombc: Analysis of Composition of Microbiomes with Bias Correction
 - Visualizers
 - ancom: Apply ANCOM to identify features that differ in abundance.
 - da-barplot: Differential abundance bar plots
 - tabulate: View tabular output from ANCOM-BC.
- cutadapt: Plugin for removing adapter sequences, primers, and other unwanted sequence from sequence data.
 - Methods
 - demux-paired: Demultiplex paired-end sequence data with barcodes in-sequence.
 - demux-single: Demultiplex single-end sequence data with barcodes in-sequence.
 - trim-paired: Find and remove adapters in demultiplexed paired-end sequences.
 - trim-single: Find and remove adapters in demultiplexed single-end sequences.
- dada2: Plugin for sequence quality control with DADA2.
 - Methods
 - denoise-ccs: Denoise and derePLICATE single-end Pacbio CCS
 - denoise-paired: Denoise and derePLICATE paired-end sequences
 - denoise-pyro: Denoise and derePLICATE single-end pyrosequences
 - denoise-single: Denoise and derePLICATE single-end sequences
- deblur: Plugin for sequence quality control with Deblur.
 - Methods

<https://docs.qiime2.org/2024.10/plugins/available/>

QIIME2は公式ドキュメントが充実している



The screenshot shows the QIIME2 documentation website. At the top left is the QIIME2 logo with the suffix "docs". Below it is a navigation bar with a dropdown menu set to "Version: 2024.10". The main content area has a header "Tutorials". A "Note" section contains a message about the tutorials assuming installed Amplicon Distribution and recommending review of g2cli docs. The "Table of Contents" lists various tutorial categories and sub-topics, many of which are linked in blue.

Version: 2024.10 ▾

Tutorials

Note

The tutorials assume you have installed the QIIME 2 Amplicon Distribution using one of the procedures in the [install documents](#). The tutorials make extensive use of the QIIME 2 command-line interface so reviewing the [g2cli docs](#) is recommended.

Table of Contents

- Getting started
- What is QIIME 2?
- Core concepts
- Installing QIIME 2
- Jupyter Book Tutorials
- Tutorials
 - Overview of QIIME 2 Plugin Workflows
 - Let's get oriented: flowcharts
 - Useful points for beginners
 - Conceptual overview of QIIME 2
 - Demultiplexing
 - Denoising and clustering
 - Taxonomy classification and taxonomic analyses
 - Sequence alignment and phylogeny building
 - Diversity analysis
 - Fun with feature tables
 - QIIME 2 for Experienced Microbiome Researchers
 - Why switch to QIIME 2?
 - Data processing steps
 - Analyze feature table and gain insight
 - New plugins
 - “Moving Pictures” tutorial
 - Sample metadata
 - Obtaining and importing data
 - Demultiplexing sequences
 - Sequence quality control and feature table construction
 - FeatureTable and FeatureData summaries
 - Generate a tree for phylogenetic diversity analyses
 - Alpha and beta diversity analysis
 - Alpha rarefaction plotting
 - Taxonomic analysis
 - Differential abundance testing with ANCOM-BC
- Predicting sample

QIIME2 Viewでデータの可視化が可能



This interface can view .qza and .qzv files directly in your browser without uploading to a server. [Click here to learn more.](#)

Drag and drop or click here

to view a QIIME 2 Artifact or Visualization (.qza/.qzv) from your computer.

You can also provide a link to a [file on Dropbox](#), a [file on Zenodo](#), or a [file from the web](#).

<https://view.qiime2.org/>

QIIME2 Viewはクライアントサイドのアプリケーションとして動作する、つまりユーザーのローカルマシン(ブラウザ)上で.qzaや.qzvファイルが処理されるため、データがインターネット上のサーバーに送信されることはなく、未公開データの機密性を保持できる。また、QIIME2をインストールしていない共同研究者とも容易に共有することができる。

<https://view.qiime2.org/about/>

Number of sample metadata columns provided: 8

Download

[SVG \(bars\)](#) [SVG \(legend\)](#) [CSV](#)

Taxonomic Level

Level 6

Color Palette [i](#)

schemeAccent

Sort Samples By [+/-](#)

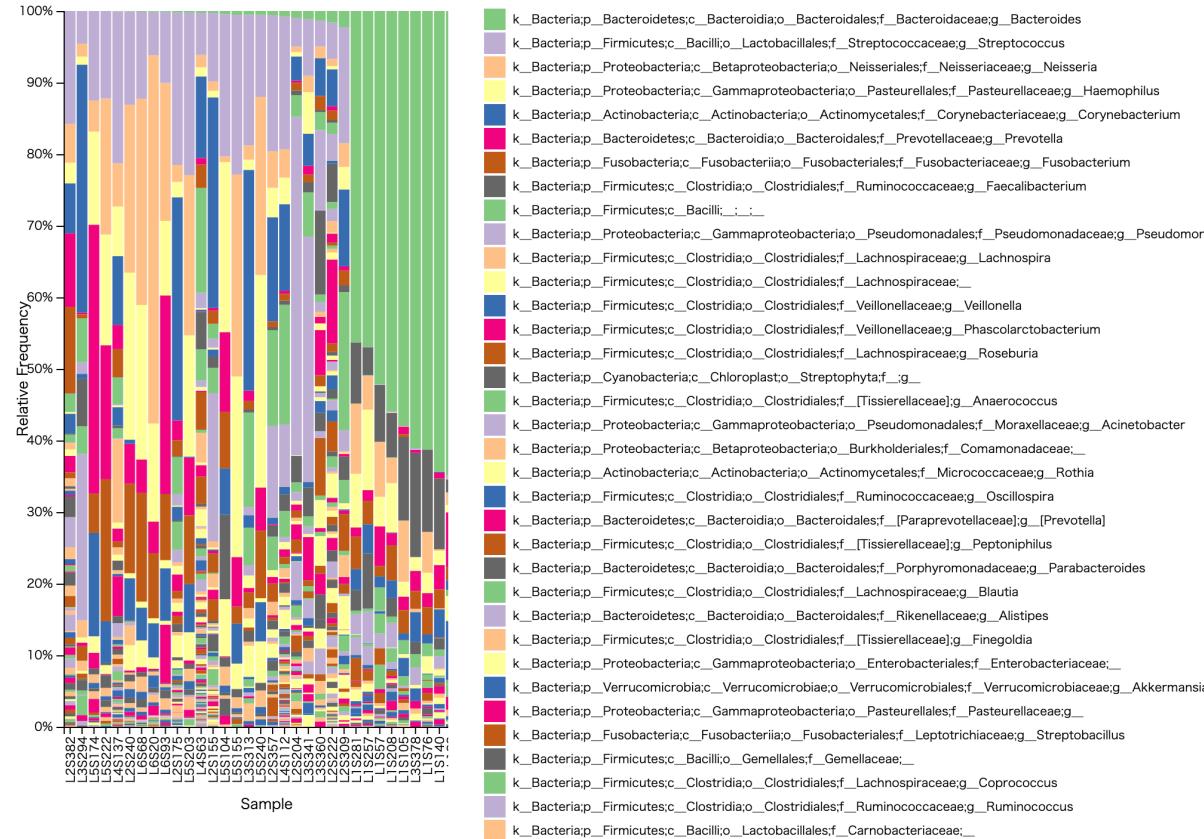
k_Bacteria;p_Bacteroidetes;c_Bacteroides

Ascending

Bar Width

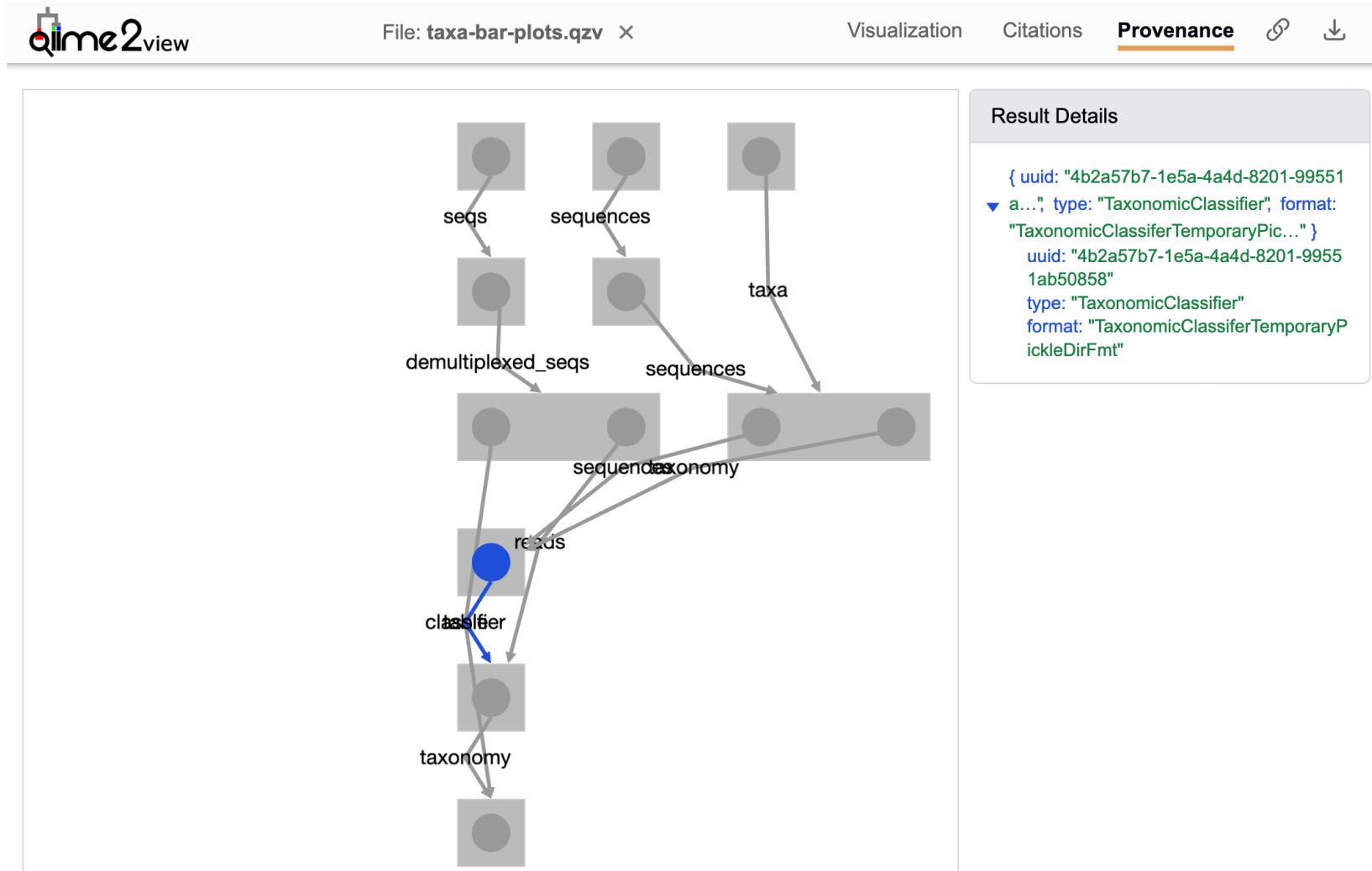


Hover over the plot to learn more



<https://view.qiime2.org/visualization/?src=https://docs.qiime2.org/2024.2/data/tutorials/moving-pictures/taxa-bar-plots.qzv>

QIIME2 Viewで解析過程を追跡できる



<https://view.qiime2.org/visualization/?src=https://docs.qiime2.org/2024.2/data/tutorials/moving-pictures/taxa-bar-plots.qzv>

QIIME2 Viewで引用すべき論文リストが閲覧できる

"16S rRNA gene amplicon sequencing analysis was performed using QIIME2 (Bolyen et al., 2019)."だけでは不十分。QIIME2のバージョン、実行したコマンドと関連論文、設定したパラメータ値を具体的に記載する必要がある。さらに、解析に使用したコード全体をGitHubなどで公開することが推奨される。

The screenshot shows the QIIME2 View application interface. At the top, there's a navigation bar with tabs: 'File: taxa-bar-plots.qzv' (highlighted), 'Visualization', 'Citations' (which is the active tab), 'Provenance', and download icons. Below the navigation bar, the main content area has two sections:

- Details of taxa-bar-plots.qzv**: This section displays metadata for the file:

```
name: "taxa-bar-plots.qzv"
uuid: "c65bc82d-6ed3-4a63-accf-12803fb3eca4"
type: "Visualization"
format: null
```
- Citations**: This section shows citation details. It includes a dropdown for 'Citation Format' set to 'BibTex' and a 'Download' button. The BibTeX code is shown below:

```
@article{framework|qiime2:2024.2.0|,
  author = {Bolyen, Evan and Rideout, Jai Ram and Dillon, Matthew R. and Bokulich, Nicholas A. and Abnet, Christian C. and Al-Ghalith, Ghada and McDonald, Daniel and Clemente, Jose C and Kuczynski, Justin and Stombaugh, Jesse and Wendel, Doug and Knight, Rob},
  doi = {10.1038/s41587-019-0209-9},
  issn = {1546-1696},
  journal = {Nature Biotechnology},
  number = {8},
  pages = {852-857},
  title = {Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2},
  url = {https://doi.org/10.1038/s41587-019-0209-9},
  volume = {37},
  year = {2019}
}

@article{view|types:2024.2.0|BIOMV210DirFmt|,
  author = {McDonald, Daniel and Clemente, Jose C and Kuczynski, Justin and Rideout, Jai Ram and Stombaugh, Jesse and Wendel, Doug and Knight, Rob},
  doi = {10.1186/2047-217X-1-7},
  journal = {GigaScience},
```

<https://view.qiime2.org/visualization/?src=https://docs.qiime2.org/2024.2/data/tutorials/moving-pictures/taxa-bar-plots.qzv>