

ポスター番号11

微生物研究管理プラットフォームを活用した 住環境マイクロバイオームのメタ解析

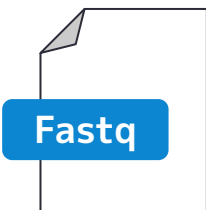
2022.10.05
株式会社BIOTA
伊藤光平



Licensed under a Creative Commons 表示 4.0 国際ライセンス
(c)2022 伊藤光平 (株式会社BIOTA)



マイクロバイオーマデータの統合解析における問題



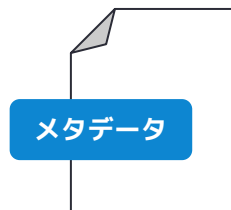
生データにアクセスできない

シーケンスデータ (Fastq等)や実験手法などの生データにおける問題点。

- ・ 公開されていない
- ・ 全データにアクセス不能



先行研究の生データを容易に取得できない。



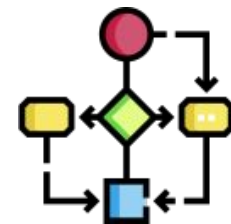
メタデータが整理されていない

サンプルに付与するメタデータにおける問題点。

- ・ 統一規格が使われていない
- ・ 記述内容やフォーマットがバラバラ



複数のメタデータを統合した解析が容易に実施できない。



ワークフロー・解析パラメーターの情報が確認できない

データ取得のために実施された処理が不明である。

- ・ 実験手法
- ・ ワークフロー
- ・ 解析パラメータ



解析して得られた結果が、技術的差異か生物学的な差異かを識別できない。

統合解析における問題

Microbiota of the indoor environment: a meta-analysis (2015)

概要

複数の屋内環境のマイクロバイオーム研究論文のデータを用いて、
屋内環境におけるマイクロバイオームの系統組成や多様性、発生源における一貫したパターンや、それらの違いを生み出す環境的要因を探索。

手法

収集：16S rRNAアンプリコンシーケンスで屋内環境を調べた16論文からデータを取得。

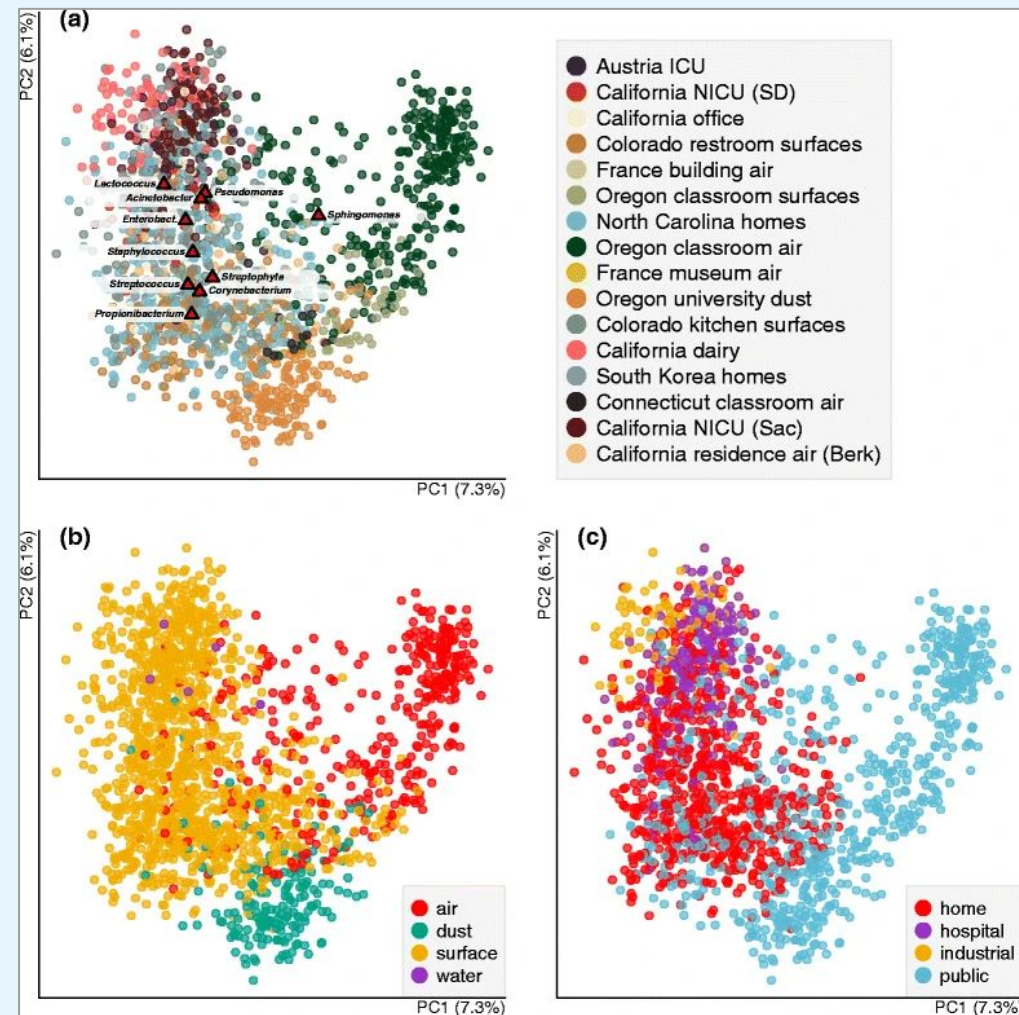
処理：全サンプルに対してQIIME2にて同様の前処理を実行。

解析：系統組成や多様性の解析を実施。

結果

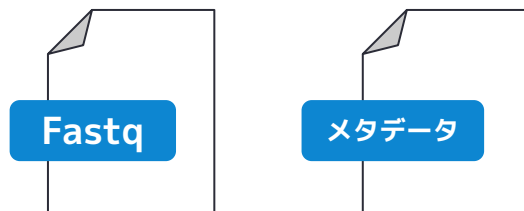
マイクロバイオームを分ける大きな要因は個々のプロトコルだった。

屋内環境は多様な系統が混在するマイクロバイオームのため、分類学的指標より、系統発生の指標（UniFrac 距離）を用いることでプロトコルによるばらつきを一部除外できた。



QIITAの特徴

多様なシーケンスデータと規格に準拠したメタデータで作られたStudyを組み合わせることで同一ワークフローで統合解析できる。



複数の生データとメタデータを統合管理

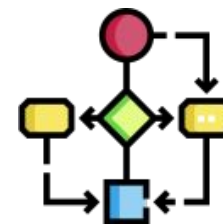
1. 研究者が、シーケンスデータ (Fastq等)とメタデータをQIITAに登録する。
2. 管理者の審査によってMIXS規格に準拠したメタデータのみ公開される。
3. 同タイミングでEuropean Nucleotide Archive (ENA) にも登録可能。

複数のシーケンスデータ
が登録可能で、登録した
プロジェクトは、Study
として管理される。

Study	Preparation Information					Study	Preparation Information			
	Sample Information	16S	Metabolomics	WGS	ITS		Sample Information	16S	18S	WGS
Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	
Sample	Sample			Sample		Sample	Sample			
Sample	Sample	Sample		Sample	Sample	Sample		Sample		
Sample	Sample			Sample	Sample	Sample	Sample		Sample	
...	...	...	...	...	...	...	...	...	...	
Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	

Study	Preparation Information	
	Sample Information	16S
Sample	Sample	Sample
Sample	Sample	Sample
Sample	Sample	Sample
Sample	Sample	Sample
Sample	Sample	Sample
...	...	...
Sample	Sample	Sample

... as many as available



デフォルトワークフローによる統合解析

研究者は自分のStudyを他のStudyをワークフロー（同一の解析手法・パラメーター）にて自動で解析できる。

16S、18Sアンプリコンシーケンスデータの場合

1. リードのトリミング (90bp, 100bp, 150bp)
2. Deblur またはclosed OTU pickingの実施
3. Feature テーブルを出力

マルチオミクスに対応したワークフローがある

メタトランスクリプトーム、メタゲノム、単離株など

QIITAからのデータ収集

7つのStudyから、1,787検体を収集。

選定基準

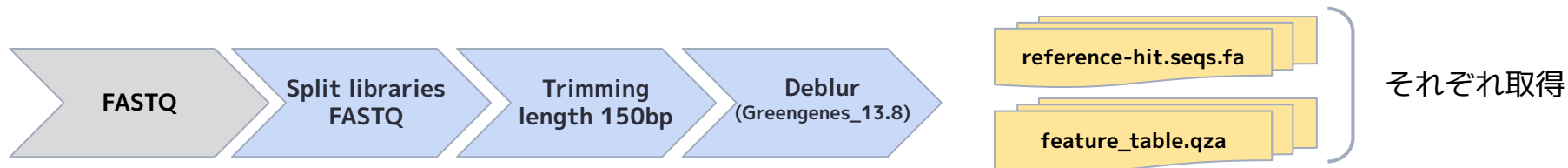
- ・ [QIITA](#)にて公開
fastq、メタデータが登録済み
- ・ レビュー済み論文
bioRxiv等は除外
- ・ 増幅領域 / プライマー
V4領域 / 515F/806R

論文名	採取場所	データ数	QIITA ID
Farm-like indoor microbiota in non-farm homes protects children from asthma development	住居	367	11010
Home Life: Factors Structuring the Bacterial Diversity Found within and between Homes	住居	180	1802
Geography and Location Are the Primary Drivers of Office Microbiome Composition	オフィス	736	10423
Surface microbes in the neonatal intensive care unit: changes with routine cleaning and over time	病院	147	1798
SARS-CoV-2 detection status associates with bacterial community composition in patients and the hospital environment	病院	160	13092
Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure	緑地	45	13064
Urban stress is associated with variation in microbial species composition, but not richness, in Manhattan	緑地	152	1674

解析の流れ

1. データ取得

[QIITA](#) から論文に対応するプロジェクトを検索。
[Default Workflows](#)にて、同一ツール・パラメータで処理されたファイルを取得。



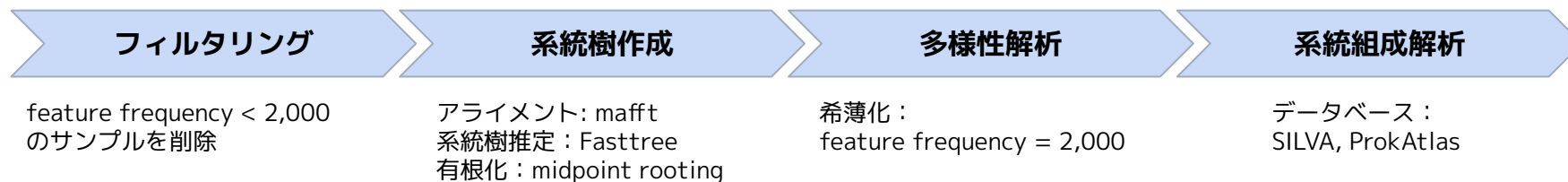
2. メタデータ追加

Environment Ontology (ENVO)に加えて、今回の解析用に以下の項目を作成。

area_category urban, rural	in_out inside, outside	space_category home, office, hospital, green space	surface_category air, surface, dust, soil, leaf	title 論文ごと
--------------------------------------	----------------------------------	--	---	----------------------

3. 解析と図表作成

QIIME2 2021.2にて解析後、R 4.0.4のggplot2にて図表を描画。



住環境の系統組成の比較

住環境ごとの系統組成

手法

サンプルを住環境ごとに分けて、リード数の中央値の上限 (“median-ceiling”) を算出しグループ化。Naive Bayes法にてSILVAデータベースにアサインメントし、系統組成テーブルを作成。門レベルでの相対存在量を積み上げ棒グラフとして表示。相対存在量が高い順に17種を表示し、残りは “Remainder” として表示。

結果

すべての環境で上位の細菌門4つは共通しており、*Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Bacteroidota* であった。

住環境のマイクロバイオームの生息源予測

手法

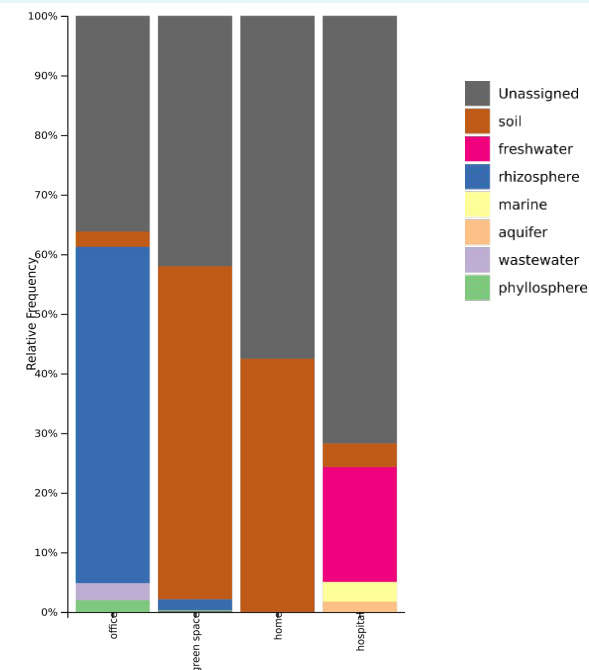
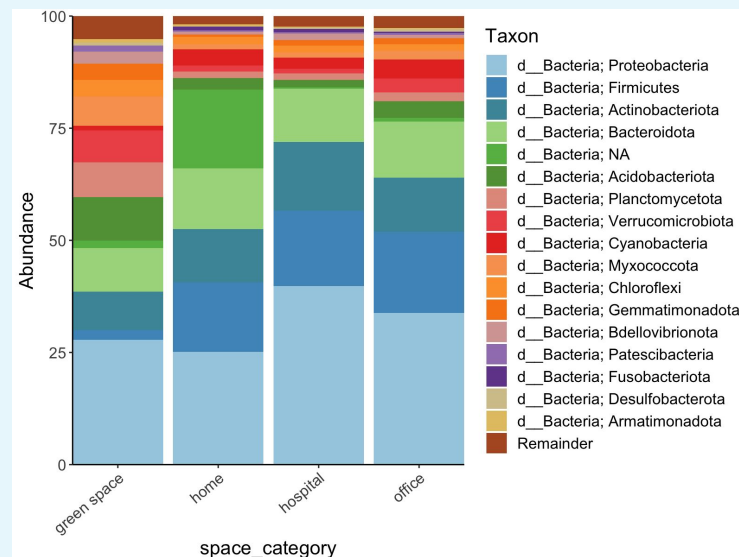
[ProkAtlas](#)から配列データ (fasta) をダウンロードし、データベースを作成。QIIME2 上の [classify-consensus-blast](#) にてアサインメント。門レベルでの相対存在量を積み上げ棒グラフとして表示。

結果

半数ほどの配列が未分類 “Unassigned” になった。残りの配列は土壌、葉圏、水、海洋が生息源として予測された。

入力として用いた配列 (90bp) が短かったため、分類精度が低下したと考えられる。

150bpで実施した際は、未分類 “Unassigned” の割当てが低下し、糞便などのヒト由来の生息源が予測されるようになった (データ未掲載)。



住環境の多様性の比較

住環境ごとのアルファ多様性の差異

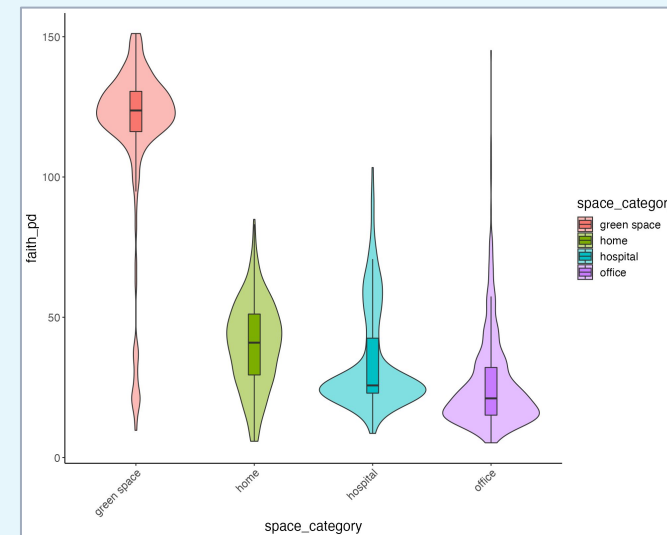
手法

リード単位でFaithの系統的多様性を計算し、箱ひげ図として描画（右上図）。

結果

系統的多様性は住環境ごとに異なり、緑地サンプルと室内環境（家、オフィス、病院）では多様性に大きな差があった。

(Kruskal-Wallis検定、BH法でFDR調整)



住環境ごとのベータ多様性の差異

手法

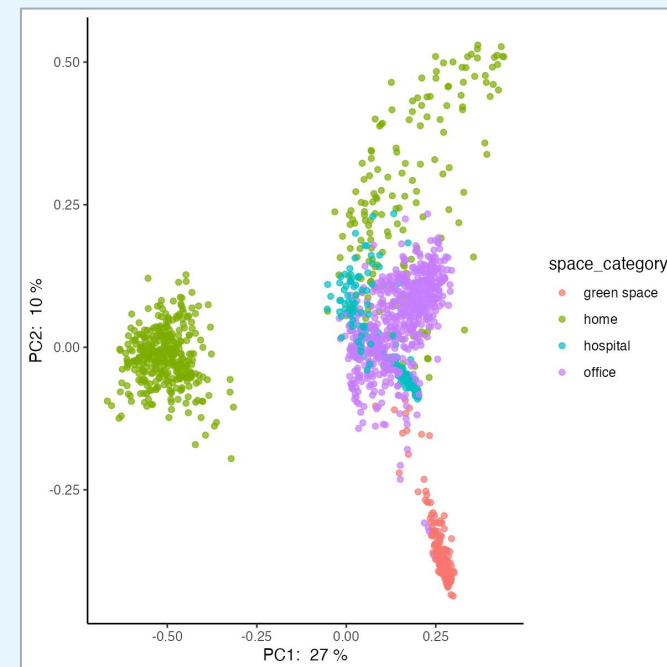
UniFrac 距離 (Weighted) を計算し、散布図として描画（右下図）。

結果

群集構造は住環境ごとに異なった。

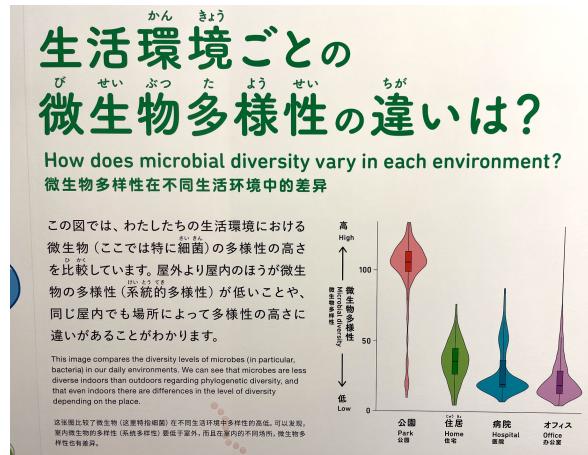
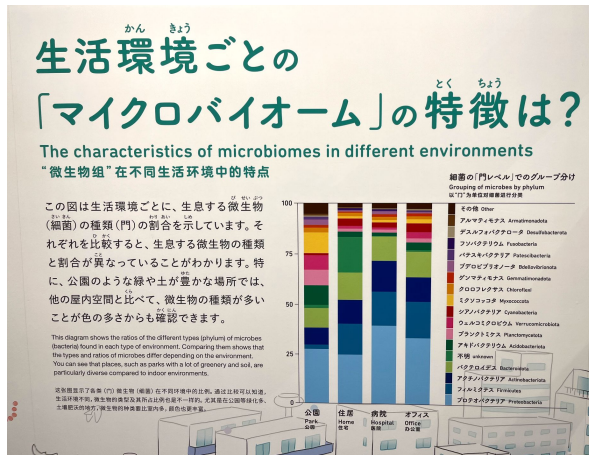
家のサンプルがもっともばらつき、緑地サンプルは室内環境サンプルとは異なる群集構造を持っていた。

(Kruskal-Wallis検定、BH法でFDR調整)



解析結果の展示による 科学コミュニケーション

日本科学未来館で監修をつとめる常設展示にて、得られた解析結果を用いて、住環境ごとのマイクロバイオームの特徴を説明するパネルを作成した。



実際の展示パネル

メタ解析を通して、住環境マイクロバイオームの特徴や差異が
バイアスフリーで可視化されただけでなく、
科学コミュニケーションを通して来場者への理解促進につながった。

