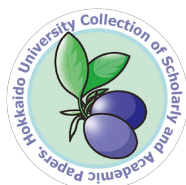




Title	細菌叢解析に関する神話と誤解 : 神話の解読の開始
Author(s)	板垣, 竜樹; 長谷部, 晃
Citation	北海道歯学雑誌, 45, 10-11
Issue Date	2024-09-15
Doc URL	https://hdl.handle.net/2115/93119
Type	journal article
File Information	45_03.pdf



特 集

細菌叢解析に関する神話と誤解：神話の解読の開始

Myths and misconceptions about microbiome analysis: Starting to Decode the Myths

北海道大学大学院歯学研究院・歯学院 口腔病態学分野¹⁾ 微生物学教室
北海道大学大学院歯学研究院・歯学院 口腔病態学分野²⁾ 口腔診断内科学教室
板垣 竜樹^{1,2)}, 長谷部 晃¹⁾

はじめに

過去20年にわたり、細菌叢に関する研究が頻繁に行われるようになった¹⁾。近年、口腔細菌叢や皮膚細菌叢、さらに皮膚真菌叢も注目されてきている。しかしながら、細菌叢解析におけるレビューや論文には科学的な根拠が不十分であるにもかかわらず、正しいと信じられ広まった情報が多い¹⁾。細菌叢解析は、16SリボソームRNA (16S rRNA) 遺伝子アンプリコンシーケンシングとショットガンメタゲノムシーケンシングによる細菌分類が一般的である²⁾。これらは配列のバリエーションと相対量から、菌叢を構成する菌種組成を得ることができる²⁾。

本稿では、新技術 Barcoding Bacteria for Identification and Quantification (BarBIQ) を紹介し、さらに細菌叢解析に関する組成データの解析手法について解説する。

細菌叢：遺伝情報の抽出と分類

細菌叢解析は、それぞれの細菌に特徴的な 16S rRNA をコードする遺伝子である 16S rDNA の配列を調べ、菌種を同定する方法で、これにより試料中の微生物の相対存在量を網羅的に解析することが可能となった²⁾。しかし、細菌はゲノム内に 16S rRNA 遺伝子の異なるコピー数 (1 ~ 15 コピー) を持ち、多くの細菌のコピー数は不明であるため、細菌細胞の数を測定して比較することは困難であった²⁾。さらにゲノム内に複数の 16S rRNA 配列タイプを持つ細菌が存在するため、検出された異なる 16S rRNA 配列が同じ細菌細胞に由来するかどうかを特定するのは困難であった²⁾。また、従来の方法における配列エラーや増幅エラーが存在し、これらエラーを除去しない限り未知の細菌から新たに同定された 16S rRNA 配列の存在を確認するのは困難である²⁾。

近年は誤分類が少ないとされるアンプリコン配列バリエーション (ASV) などが用いられることもあるが、BarBIQ と名

付けられた新技術で、より正確な分類が可能となった^{2,3)}。

BarBIQ 法

BarBIQ 法では、細菌を一つ一つ液滴に封じ込め、それらを区別するための DNA 配列を細菌 DNA に付与しながら、単一塩基の精度で 16S rRNA 配列に基づいて DNA を増幅し、単一の細菌細胞を分類群に分類する。さらに、細菌叢内の各細胞ベースの操作分類単位 (cOTU) の細胞数を計数することで、16S rRNA の数ではなく細胞の数の計数を実現した^{2,3)}。

組成データ (Compositional data)

組成データとは、構成成分が正の値で (したがってゼロはデータに含まれない)、その総和が一定の定数となるデータであり、データの信頼性による重み付けのない標準化されたデータである。もっともなじみ深い組成データの一つは百分率で表されたデータであるが、細菌叢解析における組成データは10,000リードあたりなどで読み込まれた操作分類単位 (OTU) とするデータであることが多い。組成データを解析する時は、データの信頼性、つまり、(1/標準誤差) で重み付けして解析するのが常套手段であるが、複数の構成成分を持つ組成データのみではこの重み付けはできない。組成データの強みは試料からランダムサンプリングを行うと毎回同じ組成が得られるという点であるが、偏ってサンプリングを行うと組成は大きく変化しうる。組成データの特徴は、実数に適用する和・差・積・商の四則演算が通常の定義と異なることであり、同じ構成成分の組成データ同士の比較しかできない^{4,5)}。組成データの空間は単体空間 (Simplex space) と呼ばれる三角形を一般化した空間であり、実数データ (basis) におけるユークリッド空間 (Euclidean space) とは異なる^{4,5)}。組成データは、実数のデータと異なり、相対的変動情報しか保有していない^{4,5)}。

相対的情報とは、組成そのものを指すが、構成成分のいずれかによって比をとり、相対的変動情報の抽出を行う必要がある。この相対的変動情報を用いた具体的な解析の例としては、細菌叢解析において Firmicutes/Bacteroidetes ratio を使用する比解析が一般的である。比解析において比をとる成分は任意であるが、絶対量が不変な成分 (unchanging component)、つまり実数データで同じ数の成分で比をとるときのみ絶対的変動情報を推定できるという利点がある⁴⁾。しかし組成データから実数データにおける絶対量が不変な成分を特定するのは困難であり、必ずしもそのような成分があるとは限らない^{4,5)}。組成データのみで、実数データにおいて最も変動が小さい成分の推定は可能^{4,5)}だが、その成分が不変な成分である保証はないため、この解析は相対的変動情報として見かけ上の変化を見ている可能性は否定できない。

次元の圧縮：主座標分析と主成分分析

高次元データの圧縮を行う数学的手法は主座標分析 (principal coordinates analysis: PCoA) や主成分分析 (principal component analysis: PCA) などがある。主座標分析と主成分分析は全く異なる方法で次元の圧縮を行う。一般的に行われている主座標分析による β 多様性解析では、距離行列に UniFrac 距離や Bray-Curtis 距離などを多く用いられているが、これらはユークリッド空間における距離であるから、組成データである細菌叢解析に用いるのは誤りである。つまり、サンプル間における細菌の存在量の順序が定まっていないことが理由で主座標分析の結果は単なる数学的に無意味な計算となる。さらに帰無仮説のうちの一つに細菌同士が独立であると誤った仮定により、小さな p 値を得ることが多い⁵⁾。一方で主成分分析は、お互いに独立である n 例の被験者を n 次元ベクトル空間の座標軸にし、お互いに相関がある p 種類の観測項目 (細菌種) を p 次元部分空間の座標軸にする⁵⁾。これにより、積み上げ棒グラフの相対的変動情報の次元を削減でき、主成分スコアとその散布図は相対的な多様性の指標に相当する。

従来の細菌叢解析の望ましい研究デザインの例

組成データは構成成分が異なれば、比較するまでもなく「別物」である。つまり、従来の技術による細菌叢のデータは組成データであることから、主成分分析を行い主成分スコアで相対的な多様性を示すことができるが、細菌叢の構成成分は宿主によりそれぞれ異なるので、他者と細菌叢の組成を比較すべきではない。

望ましい試験デザインの一つは、同一個体内の細菌叢に関して構成成分 (細菌種) が同じであることを仮定し、比解析によってその細菌叢の経時変化を探索する試験である⁵⁾。

しかし、不変成分は特定されていないため、絶対的変動情報の推定が行える保証はなく、相対的変動情報による見かけ上の変化を解析していることもある点に注意が必要である^{4,5)}。

おわりに

細菌叢解析についての研究が普及したが、組成データにおける正しい解析の普及は、不十分である。細菌の分類操作単位 (OTU) は、サンプルごとに標準化され、サンプル内で存在量の順序はあるが、サンプル間で存在量の順序はない。サンプル間で存在量の順序が成立しないため、順位を使用するノンパラメトリック検定などはできない。相対的変動情報は見かけ上の変化を示し、科学的意義は少ない。新技術の開発により、組成データから解放され、絶対的変動情報 (実数データ) による解析が可能となり、細菌叢における真実が発見されるものと期待される。

謝 辞

本稿を終えるにあたり、本研究に助言いただきました杉本典夫様 (杉本解析サービス) に厚くお礼申し上げます。

参 考 文 献

- 1) Walker AW, Hoyles L: Human microbiome myths and misconceptions. *Nature Microbiology*, 8: 1392-1396, 2023.
- 2) Jin J, Yamamoto R, Takeuchi T, Cui G, Miyauchi E, Hojo N, Ikuta K, Ohno H, Shiroguchi K: High-throughput identification and quantification of single bacterial cells in the microbiota. *Nature Communications*, 13: 863, 2022.
- 3) Jin J, Yamamoto R, Shiroguchi K: High-throughput identification and quantification of bacterial cells in the microbiota based on 16S rRNA sequencing with single-base accuracy using BarBIQ. *Nature Protocols*, 19: 207-239, 2024.
- 4) Ohta T, Arai H, Noda A: Identification of the Unchanging Reference Component of Compositional Data from the Properties of the Coefficient of Variation. *Mathematical Geosciences*, 43: 421-434, 2011.
- 5) Itagaki T, Sakata K-i, Hasebe A, Kitagawa Y. Diversity of the Japanese Gut Microbiome Analysis: Relative Approach Using Principal Component Analysis. in Preprints (Preprints, 2024).