



Mock community と日本人腸内細菌叢データを用いた メタプロテオーム解析手法の評価

Evaluation of Metaproteomic Analysis Methods Using Mock Community
and Japanese Gut Microbiota Data

福本亮太^{1,2}, 小寺義男¹, 河野信²

1.北里大学理学部, 2.北里大学未来工学部

1 研究背景

メタプロテオーム解析は、微生物群集の機能的多様性と動態を理解するための手法である。従来のプロテオーム解析は単一種の生物を対象にするが、メタプロテオーム解析は土壌や水、腸内フローラなど複数の微生物が混在する環境試料を対象とする。これにより、環境中の微生物の活動や相互作用を包括的に解析でき、エコシステムの健康状態や変化、病気のメカニズムなどの理解が進む。環境保全、医学、バイオテクノロジーなど多岐にわたる分野で応用が期待されている。しかし、膨大な量のデータの解析によるデータの複雑性やその解析に対する解釈の困難さがメタプロテオーム解析を実行する上で障壁となっている。

2 研究目的

本研究ではmock communityデータを使用し、人工的なコミュニティ内に含まれる生物種と、プロテオーム解析結果により同定された生物種を比較することで、メタプロテオームの解析方法を検討する。また、この解析手法を基に日本人の腸内細菌プロテオームデータを用いて、被験者の分類を行う。

3 研究方法・結果

① Mock communityを用いたデータ解析

28~32種類の細菌、古細菌、真核生物、バクテリオファージが含まれる人工コミュニティのデータ (PXD006118) を利用した。本実験では、解析ソフトとしてMetaLab MAGを、リファレンスデータベースとしてMGnify human_gut v2.0を用いて、全スペクトル数とスペクトルの同定率、ペプチド、タンパク質の同定数、さらに人工的に混合した生物種数と同定された生物種数の違いを解析した。また、COGを用いてタンパク質の機能について解析した。

表1 各mock communityデータの解析結果

sample	配合した生物種	MS.MS	MS.MS.Identified	同定率	peptide同定数	protein同定数	class同定数
community1	28	578002	67948	11.76	17281	4356	18
community2	28	557921	66974	12.00	18920	4651	24
community3	28	564968	49475	8.76	14653	3939	19
community4	28	559316	50207	8.98	14295	3745	18
community5	28	577123	87036	15.08	22163	5172	27
community6	28	556677	96088	17.26	30183	6634	23
community7	28	554350	73551	13.27	23715	5650	20
community8	28	546424	76780	14.05	22776	5309	24
community9	32	525443	107306	20.42	27871	5629	19
community10	32	567609	105214	18.54	33106	6199	18
community11	32	543125	92847	17.09	27105	5417	17
community12	32	538024	102773	19.10	27988	5373	18

Salmonella enterica (24.0%)

腸内細菌科のグラム陰性桿菌であり、食中毒の原因となる病原菌の一種。

表2 community1に含まれる28生物種

Pseudomonas denitrificans ATCC 13867
Bacillus subtilis 168
Paracoccus denitrificans ATCC 17741
Roseobacter sp. AK189
Pseudomonas pseudoalcaligenes KF 707
Chromobacterium violaceum CV026
Agrobacterium tumefaciens NTL4
Stenotrophomonas maltophilia Ssp1E02
Cupriavidus metallireducens CH34; (DSM 2839; LMG 1195; CIP 107179)
Pseudomonas fluorescens ATCC 13525
Burkholderia xenovorans LB400
Staphylococcus aureus subsp. *aureus* Rosenbach, Smith [NCTC 10399], ATCC 13709
Staphylococcus aureus subsp. *aureus* Rosenbach, Seattle 1945; ATCC 25923
Alteromonas macleodii ATCC 27126
Salmonella enterica typhimurium LT2 with Flac+ Plasmid, SAZ197
Escherichia coli K12 with Flac+ Plasmid
Salmonella enterica typhimurium LT2
Salmonella enterica typhimurium LT2-H88, Rough mutant of LT2
Thermus Thermophilus HB27
Chlamydomonas reinhardtii
Phage M13
Phage F2
Phage P22 (HT105)
Phage F0
Phage ES18 (H1)
Rhizobium leguminosarum bv. *viciae* 3841
Rhizobium leguminosarum bv. *viciae* VF39
Nitrososphaera viennensis

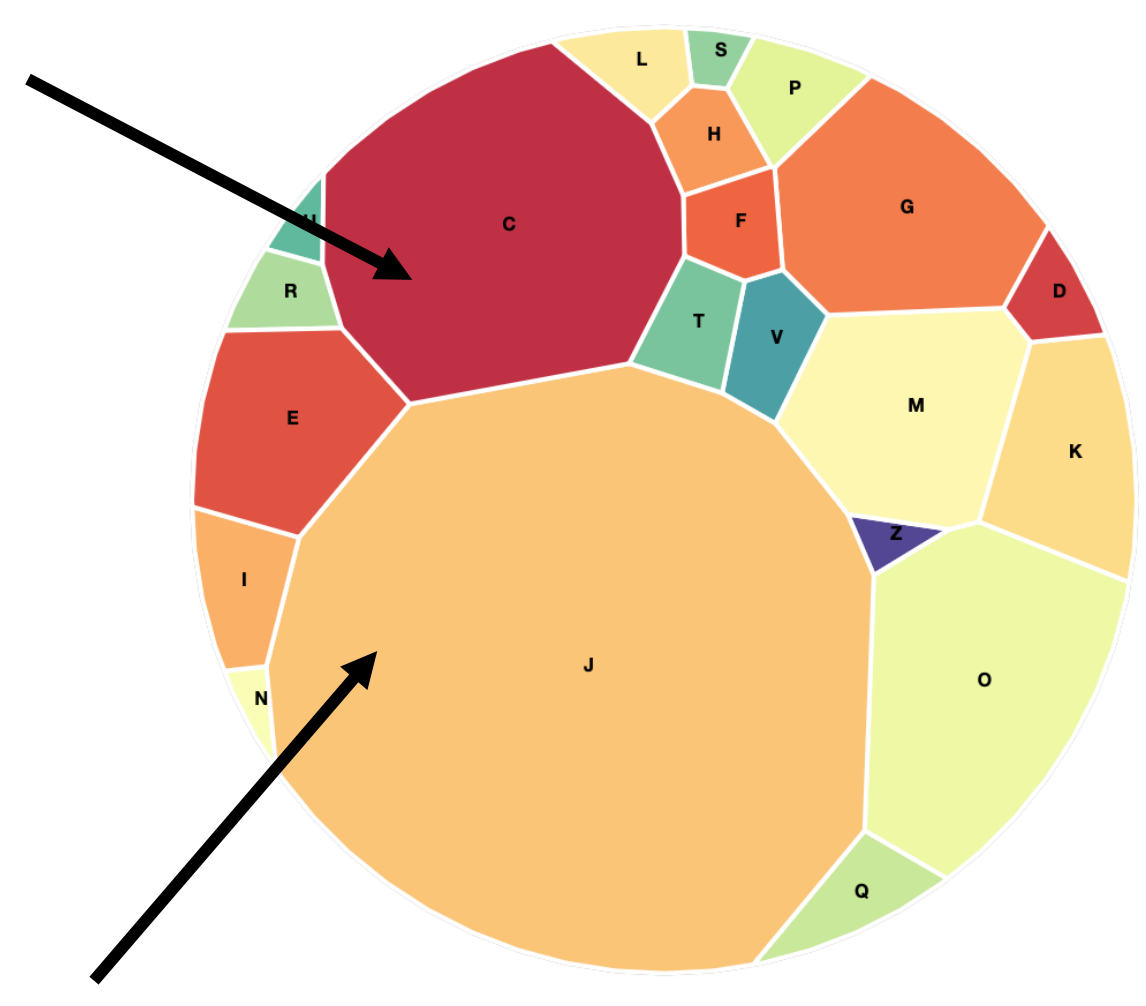
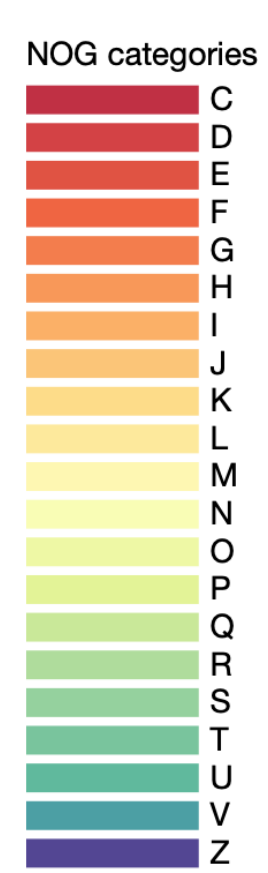


Pseudomonas B luteola (13.4%)

グラム陰性の非発酵性好気性桿菌で、土壌、水、植物など多様な環境に広く存在する細菌。

図1 community1 の生物種同定結果

C:ENERGY PRODUCTION AND CONVERSION



J:TRANSLATION, RIBOSOMAL STRUCTURE AND BIOGENESIS

図2 community1 のタンパク質の機能分類

② 日本人腸内細菌叢を用いたデータ解析

日本人の腸内細菌叢のデータはPXD010754を使用し、mock communityと同じ方法で解析を行った。本実験では50人分のデータを5人ずつまとめて解析を行い、全スペクトル数とスペクトルの同定率、ペプチド、タンパク質の同定数や同定された生物種について解析した。

また、解析により得られた生物種をphylumレベルでクラスタリング解析を行った。

表3 各日本人腸内細菌叢データの解析結果

sample	MS.MS	MS.MS.Identified	同定率	peptide同定数	protein同定数
no.1-5	893233	203251	22.75	104291	14207
no.6-10	907341	211834	23.35	96672	48129
no.11-15	880007	143752	16.34	77375	11620
no.16-20	881666	206869	23.46	86732	12189
no.21-25	867091	60705	7.00	19170	5017
no.26-30	869013	131247	15.10	46481	10195
no.31-35	674704	39019	5.78	16283	4045
no.36-40	690983	57170	8.27	20594	5175
no.41-45	852065	58614	6.88	22664	5548
no.46-50	842137	52963	6.29	21488	5471

PCA of Gut Microbiome Data

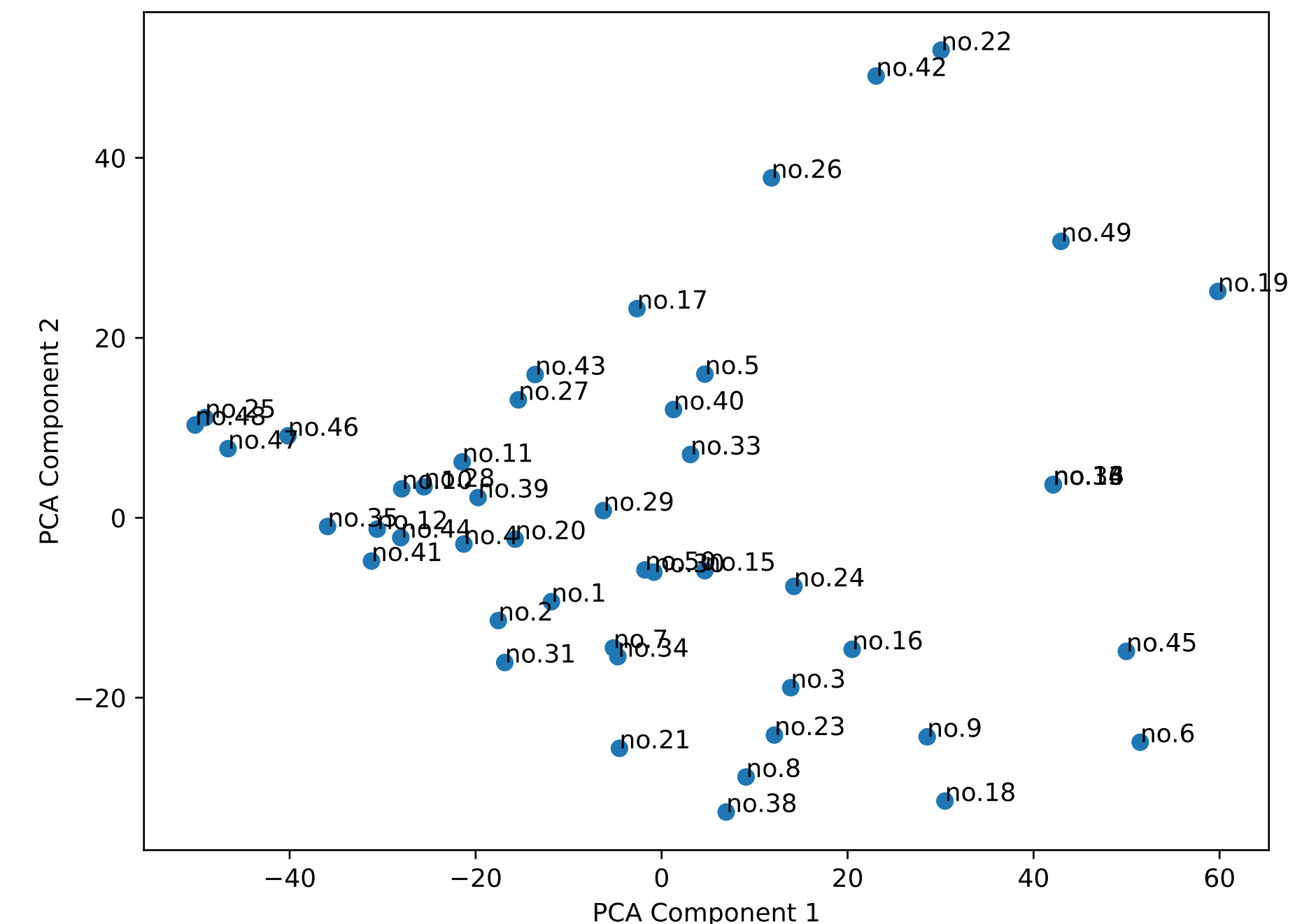


図4 被験者50人の主成分分析

Dendrogram of Gut Microbiome Data

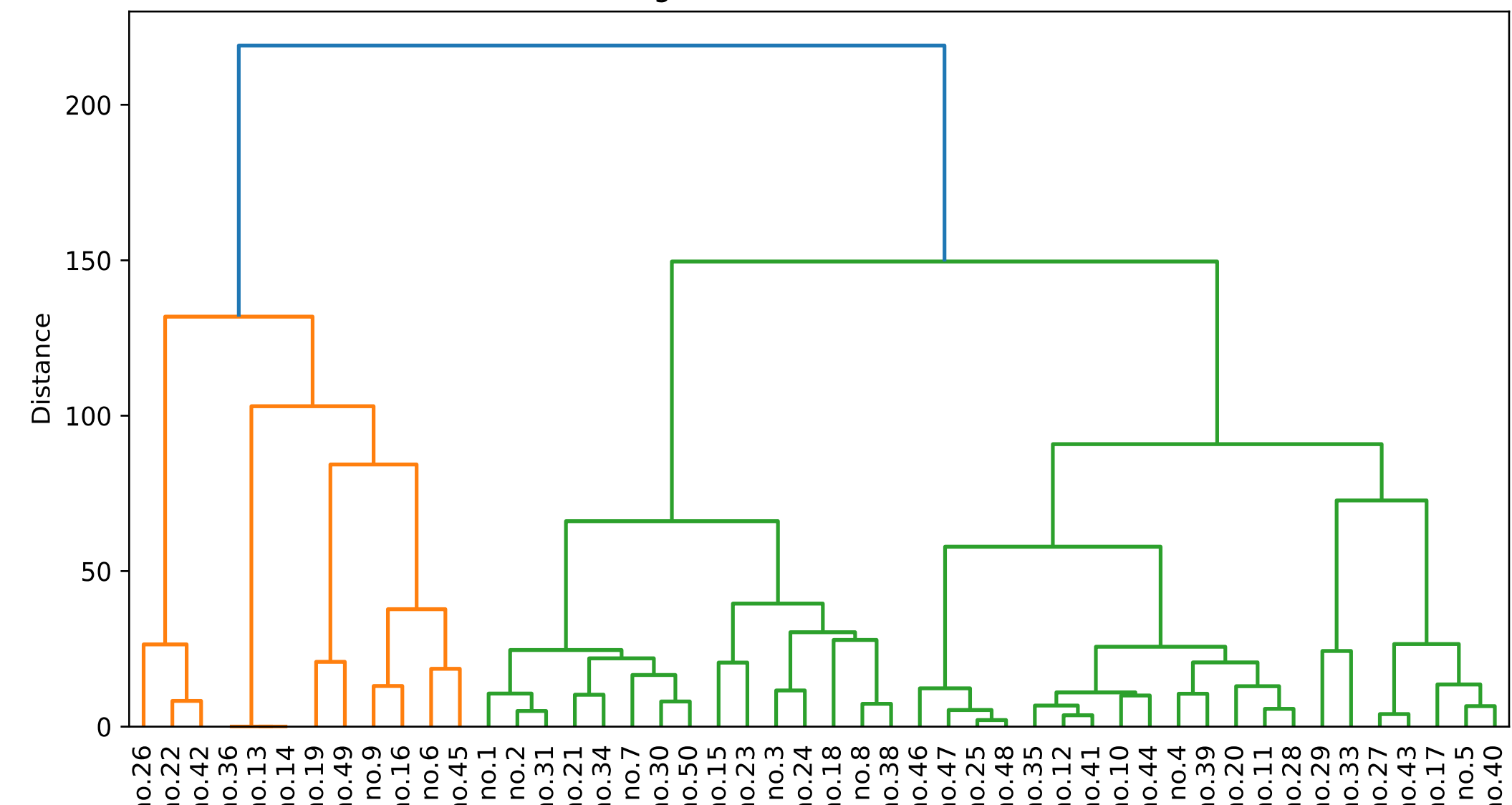


図5 被験者50人をクラスタリングした樹形図

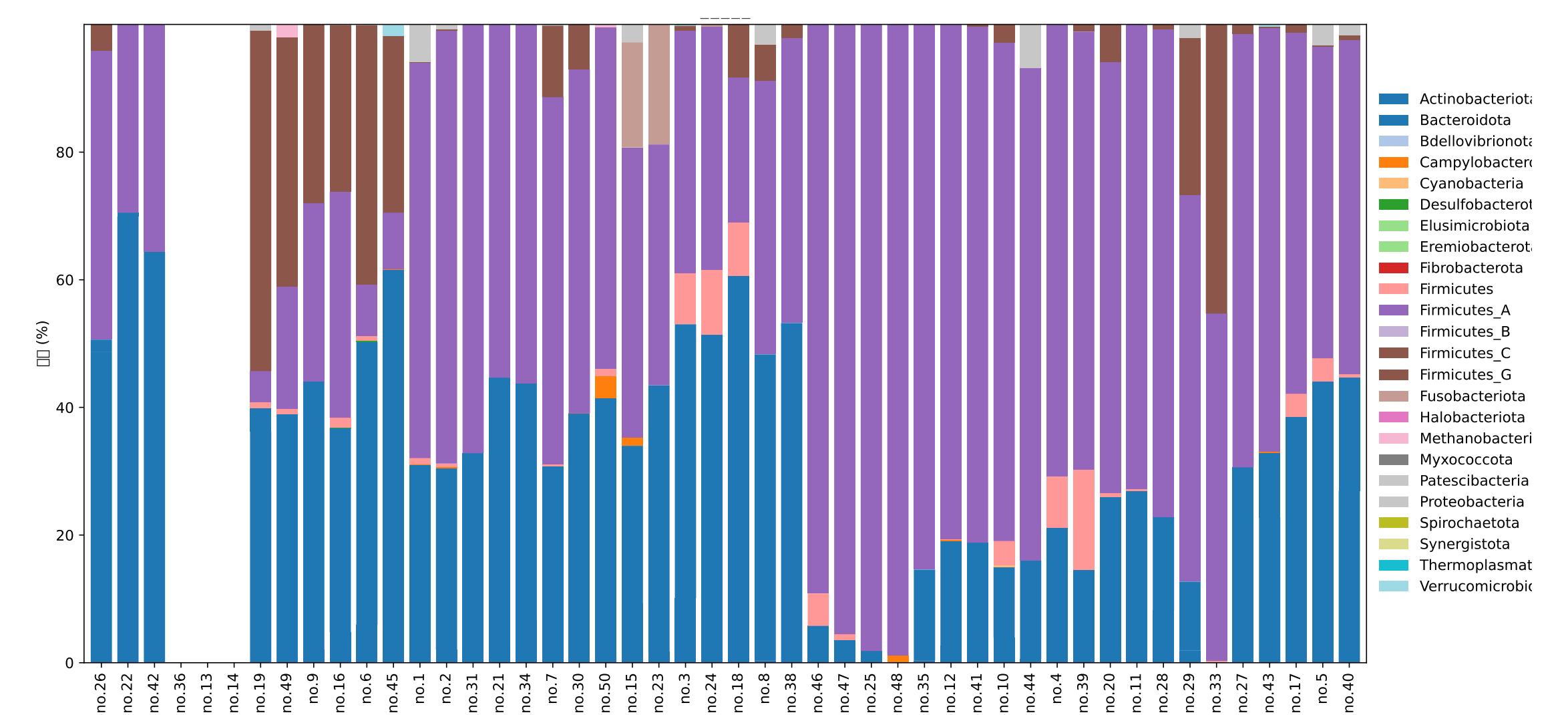


図6 被験者50人のphylumレベルでの細菌割合

5 考察・展望

- Mock communityを用いたデータ解析では同定率が約10%弱~20%ほどであったが、実際に混合した *Salmonella enterica* や *Pseudomonas* などを同定できていた。また、classレベルでの生物種の同定数は混合生物種数とほぼ同等であったため、classレベルで見ると良いと考えられる。
- 日本人の腸内細菌プロテオームデータでは同定率が約5~20%強であった。同定された細菌の割合を基に50人分のデータをクラスタリングして、被験者をグループ分けすることができた。今後は、健康状態との対応付けを行う予定である。
- 今回はリファレンスデータベースとして、欧米人が中心の MGnify を利用した。これの代わりに日本人の腸内細菌メタゲノムをベースとする microbiome databuh を利用することで、解析の精度がさらに高まるものと予測できる。さらに、より高精度にメタゲノム解析が可能なロングリードシーケンサを利用したリファレンスデータベースを使うことで、解析の精度がさらに高まるものと期待される。