

令和元年5月29日現在

機関番号：32665

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K07205

研究課題名(和文) 16S rRNA遺伝子に依存しないNグラム配列に基づく細菌叢解析法の開発

研究課題名(英文) Phylogenetic Analysis Based on N-gram Independent of 16 rRNA gene sequences

研究代表者

中野 善夫 (NAKANO, Yoshio)

日本大学・歯学部・教授

研究者番号：80253459

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：種ごとの連続塩基配列出現頻度に基づいて細菌種の距離を求める方法、従来の16S rRNA配列によって細菌種の距離を求める方法に加え、遺伝子の配列順序がどれくらい似ているかを細菌種のゲノム構成の距離とする方法で細菌の系統解析を行ったところ、従来の方法では16S rRNA配列では解析が困難であったYersinia属細菌、およびEscherichia coli/Shigella属細菌の系統解析が、5塩基連続配列の出現頻度に基づく方法で精度良く系統解析が行えた。一方、菌叢解析に基づき口腔内細菌叢から口臭の有無を深層学習を用いて予測したところ、97%以上の精度で予測できることが明らかになった。

研究成果の学術的意義や社会的意義

本研究で得られた結果から、5塩基連続配列の出現頻度に基づく系統解析は、これまでの手法では困難であった系統解析を可能にすると期待できる。この方法は特に近縁種の解析に有効だったが、遺伝子の連続配置の出現頻度に基づく方法が進化的に離れた種や属のあいだでの解析に効果を発揮する可能性がある。今まで主流だった方法は、特定の遺伝子あるいは複数の遺伝子の相同性によって解析するもので、ゲノム全体の比較によって解析を行なうものがほとんどなかったため、本研究の成果から新たな視点での系統解析が発展すると期待できる。さらに、このような手法に基づく菌叢解析の応用例も示すことができた。

研究成果の概要(英文)：We demonstrated genome-wide comparisons based on n-gram (pentagram) profiles and construction of phylogenetic trees. Pentagram profiles consisting of 512 degenerated patterns of five nucleotides were calculated from 2,602 sequences of bacterial genomes. Pentanucleotide frequency analysis was used to separate species that are difficult to distinguish based on 16S rRNA gene sequences, and the results showed clear separation of Yersinia pestis from Yersinia pseudotuberculosis, Shigella from Escherichia coli. In addition, a discrimination classifier model was constructed by profiling 16S rRNA-based operational taxonomic units (OTUs) and calculating their relative abundance in saliva samples from 90 subjects. Our deep learning model achieved a predictive accuracy of 97%, compared to the 79% obtained with a support vector machine. This approach is expected to be useful in screening the saliva for prediction of oral malodour before visits to specialist clinics.

研究分野：口腔細菌学

キーワード：口腔細菌叢 機械学習 深層学習 系統解析 口臭

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

生物種の塩基配列の並び方の組み合わせには、それぞれ偏りがある。これまで、連続配列の出現頻度を利用した解析では、自己組織化マップと組み合わせたメタゲノム解析などが報告されてきたが、自己組織化マップ以外のパターン認識や機械学習等の手法と組み合わせた報告がほとんどない。本研究では以下の2点に着目することにした。

- (1) 細菌叢の構成分析:環境由来の試料中の菌種の割合・分布を分析するためには、16S rRNA 遺伝子配列による解析が行われているが、以下のような問題や不便があると指摘されている。一つの菌が複数の 16S rRNA 遺伝子を保有している種によってその数が異なること、PCR による増幅が必要であるために増幅効率に偏りがある場合が多いこと、極めて相同性の高い配列の増幅を繰り返すため PCR 増幅時にキメラ分子が生成してしまうこと、精度の高い分析には 500 塩基以上の正確な塩基配列が必要であること、数は少ないが 16S rRNA 遺伝子にも水平伝播の例が報告されていること等である。これを、連続配列出現頻度によってこれらの問題を回避する解析方法を確立できれば、短時間に多数の試料の菌叢解析を高精度に実行することが可能になる。
- (2) 細菌の系統解析:系統解析にも主に 16S rRNA 遺伝子配列が利用されているが、上記の問題の他に、近縁種の識別が難しい場合もある。*E. coli* と *Shigella* 属細菌がその例として知られている。他の遺伝子を指標とする場合は、指標遺伝子の選択と評価が容易ではないが、ゲノム全体の連続塩基配列出現頻度による比較ならば、どの種であっても等しく同じ指標で比較することが可能となる。本項においても、連続塩基数と特徴配列の抽出で精度の高い解析が可能になると予想していた。研究開始当初までの予備的な解析では、16S rRNA 遺伝子で系統解析が難しい *Yersinia* 属細菌で良好な結果が得られていた。

2. 研究の目的

現在、細菌叢の構成菌種分析や系統解析では 16S rRNA 遺伝子の相同性に基づく方法が一般的に広く利用されているが、いくつかの問題が指摘されている。本研究では、これらの問題の影響を受けない菌叢構成種を計算する方法を開発するために、生物は遺伝子の塩基配列に種固有の「癖」を持っていて連続する塩基配列の出現頻度 (n グラム) に偏りがあることに注目した。環境試料由来の DNA 中の 5~8 塩基程度の連続塩基配列の出現頻度を数えあげ、出現頻度の組合せから最適解を計算する方法を開発する。この特定の遺伝子の配列に依存しない方法により、(1) 遺伝子のコピー数の違いの影響を受けず、(2) PCR 増幅時のキメラや読み取りエラーの影響が極めて少なく、(3) 遺伝子の選択と有効性の検証という手間が不要で、(4) 100 塩基断片のような短い配列の蓄積で計算ができ、(5) 遺伝子の水平伝播の影響がほとんどない解析法を開発する。また、n グラム出現頻度で細菌ゲノムの特徴を捕えることができるのであれば、それらの蓄積である細菌叢の特徴も塩基の n グラム出現頻度で特徴づけることが可能であるはずである。本研究では、口腔内細菌叢のメタゲノム解析により n グラム出現頻度から口臭の有無が機械学習によって予測できるかどうかを検討することとした。

3. 研究の方法

- (1) 種ごとの連続塩基配列出現頻度をゲノム配列が既知の種において集計した。すなわち、例えば 5 塩基連続配列であれば、AAAAA、AAAAC、AAAAG.....TTTTT までの 1024 の組合せとなる。ここで遺伝子の方向による配列の偏りを排除するために、相補配列を

統合して、例えば AAAAA と TTTTT を同一配列として集計すれば、512 種の組合せとなる。これで、一つの種に対して 512 項のベクトルができる。得られたベクトルの距離、あるいは角度を、それぞれの種の間距離として計算する。その計算は Euclidean 距離と Ward アルゴリズムで求めた。比較する 16S rRNA 配列に基づく系統樹は当該遺伝子全長を用いて、Neighbor-joining method にて解析した。

(2) 細菌系統解析

上記の方法で細菌種の距離を求める方法、従来の 16S rRNA 配列によって細菌種の距離を求める方法に加え、遺伝子の配列順序がどれくらい似ているかを細菌種のゲノム構成の距離とする方法も検討した。すなわち、細菌種のゲノム上の遺伝子配列順にアミノ酸配列を並べ、一方で解析対象のすべての細菌種のゲノム上にコードされている遺伝子から求められるアミノ酸配列に基づき一次構造の相同性から同一あるいは近縁の遺伝子産物を決める。次いで、ゲノム上の遺伝子の並びからどのような遺伝子が隣り合っているかを一次構造の相同性により求めた遺伝子機能で集計する。1 が塩基の penta-gram であるのに対して、これは遺伝子の bi-gram である。

(3) 細菌叢に基づく口臭予測

口臭をもつ人から唾液を採取し、唾液中の細菌叢を 16S rRNA 配列と penta-gram 配列出現頻度で解析した。福岡歯科大学の口臭外来を来院した患者 90 名の呼気と口腔内細菌叢を測定・解析して、メチルメルカプタン濃度 0.3 ppm 以上を口臭ありとして集計した。45 人ずつの口臭ありとなしのグループにわけ、深層学習によって細菌叢から口臭の有無を予測できるかどうかを検証した。90 人から 1 人を無作為に選び、残り 89 人で学習し、対象の 1 人の細菌叢データから口臭の有無を予測する手順を 90 回繰り返した。このあと、penta-gram 配列出現頻度で口臭の有無を予測する予定であったが間に合わなかった。ただ、本項は当初の予定から先の追加部分である。

4. 研究成果

(1) 系統解析

16S rRNA 配列での系統解析が難しい *E. coli/Shigella* の解析を penta-gram の出現頻度で解析する方法、16S rRNA 配列に基づく方法、および遺伝子の bi-gram に基づく方法にしたがって行ったところ、図 1 に示すように、最も精度高く *E. coli* と *Shigalla* 属細菌を分けることができたのは penta-gram の出現頻度に基づく方法で、次に遺伝子の bi-gram に基づく方法だった。

Yersinia 属細菌も 16S rRNA 配列による解析が困難な種として知られている。この属には 3 つの種が含まれているが、それが本方法によりはっきり系統樹上で分けられた (図 2)。 *Y. pestis* と *Y. pseudotuberculosis* の 16S rRNA 塩基配列はまったく同一である。したがって、16S rRNA の配列に基づいた系統解析で両種を解析することはできない。しかし、今回の研究成果から、これらの種を含む 3 つの種を 5 塩基連続配列の出現頻度に基づく明確に分けることができた。 *Y. pestis* は 1,500–20,000 年前に *Y. pseudotuberculosis* から分岐したと考えられている。このことから、今回の結果は 5 連続塩基配列の解析によってこれら 2 つの種を明確に識別できたことが示されているといえよう。

5 塩基連続配列による解析と遺伝子の bi-gram による解析方法は、16S rRNA 配列による系統解析とは異なり、ゲノム全体の比較ができる。他に、いくつかの house keeping gene を選んでその有無によって系統解析を行う multilocus sequence analysis (MLSA) based が

知られているが、それも特定の数種の遺伝子によって解析する方法で、本方法のようなゲノム全体を比較するものではない。この解析方法で今まで系統解析が困難であったゲノムでも系統解析が容易に行えるようになったといえる。

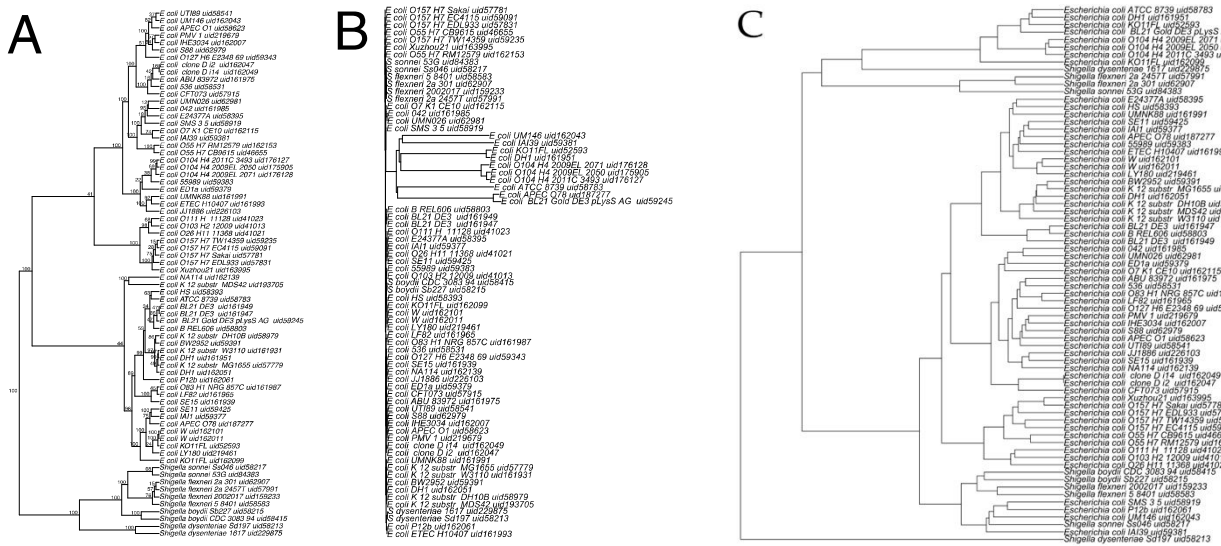


図1 . *E. coli* と *Shigella* 属細菌の系統樹解析。A: 5塩基連続配列出現頻度に基づく系統樹 B: 16S rRNA 配列に基づく系統樹 C: 遺伝子の bi-gram に基づく系統解析

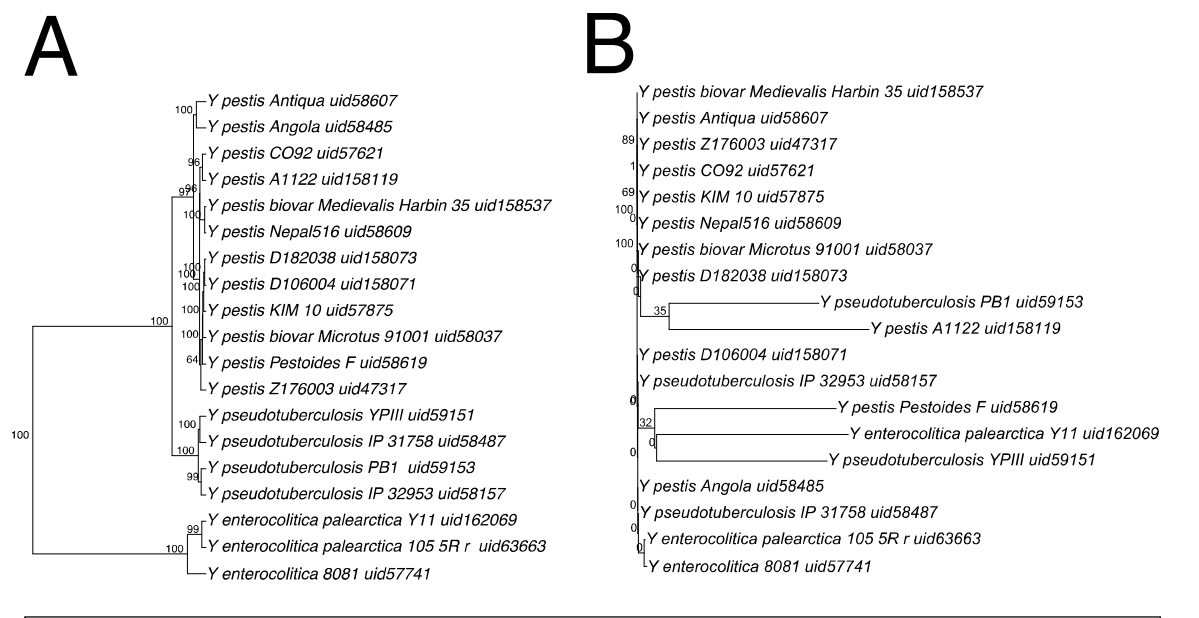


図2 . *Yersinia* 属細菌の系統樹解析。A: 5塩基連続配列出現頻度に基づく系統樹 B: 16S rRNA 配列に基づく系統樹

(2) 細菌叢に基づく口臭予測

口臭外来を受診した90名の唾液から口腔内細菌叢を解析した。まず、16S rRNA 遺伝子配列解析によりOUT単位で集計し、その結果からサポートベクターマシン(SVM)と深層学習で口臭の有無を予測できるかを検証した。その結果、表1に示したように、深層学

習を用いた場合に感度 100%、特異度 93.3%、精度 96.7%で口臭を予測することが可能であった。

	感度 (%)	特異度 (%)	精度 (%)
SVM	77.8	80.0	78.9
Deep Learning	100	93.3	96.7

また、口臭のある群とない群で細菌叢内の菌種を推定してみると、図3に示すような程度の傾向が認められた。しかしながら、特定の数種の菌の有無でただちに口臭の有無が判定できるわけではなく、このことから本研究のような菌叢全体を比較する機械学習による手法が必要であり、また有効であることが示された。

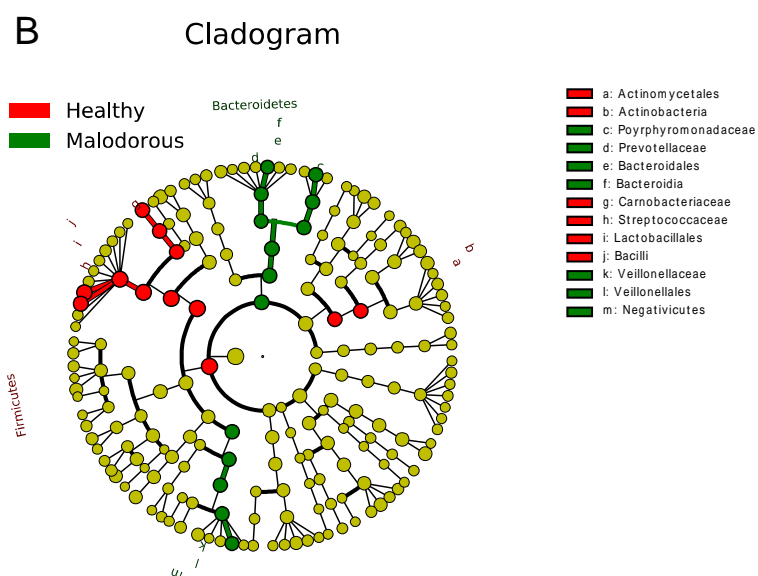


図3 . 口臭のある口腔内細菌に多い菌 (緑) と健康な口腔細菌叢に多い菌 (赤) を系統樹上で示したグラフ。

5 . 主な発表論文等

〔雑誌論文〕(計2件)

1. N. Suzuki, Y. Nakano, T. Watanabe, M. Yoneda, T. Hirofuji, and T. Hanioka. 2018.
Two mechanisms of oral malodor inhibition by zinc ions.
J. Appl. Oral Sci. 26: e20170161. DOI: 10.1590/1678-7757-2017-0161
査読あり
2. Y. Nakano, N. Suzuki, and F. Kuwata. 2018.
Predicting oral malodour based on the microbiota in saliva samples using a deep learning approach.
BMC Oral Health, 18: 128. DOI: 10.1186/s12903-018-0591-6
査読あり

〔学会発表〕(計7件)

1. 渡邊妙子、吉田尚恵、武田伊織、小野洋一、中野善夫、山岸賢司：ライフサイエンスデ

データベースを活用したバイオインフォマティクス研究．トーゴの日シンポジウム
2018（2018年）

2. 谷口奈央、中野善夫、桑田文幸、埴岡隆：金属イオンによる消臭メカニズムの解明．第
60回歯科基礎医学会学術大会（2018年）
3. 中野善夫：1 クラス・サポートベクターマシンを用いた外来性遺伝子の探索．第91回
日本細菌学会総会（2018年）
4. 中野善夫、谷口奈央、桑田文幸：深層学習による口腔内菌叢解析に基づく口臭の判別 2017
年度生命科学系学会合同年次大会(2017年)
5. 中野善夫、谷口奈央、桑田文幸、埴岡隆：細菌叢解析に基づく機械学習による口臭の判
別．第59回歯科基礎医学会(2017年)
6. 中野善夫：口腔内細菌叢の5連続塩基出現頻度に基づく解析法．第90回日本細菌学会
(2017年)
7. 中野善夫、谷口奈央、桑田文幸：連続塩基出現頻度に基づいた菌叢構成種解析．第40
回日本分子生物学会(2016年)

〔図書〕(計0件)

〔産業財産権〕

出願状況(計1件)

名称：口臭判定装置及びプログラム

発明者：中野善夫

権利者：学校法人日本大学

種類：特許

番号：特開2019-030275

出願年：2017年

国内外の別：国内

取得状況(計0件)

〔その他〕

ホームページ等

6．研究組織

(1)研究分担者

研究分担者氏名：谷口 奈央

ローマ字氏名：TANIGUCHI, Nao

所属研究機関名：福岡歯科大学

部局名：口腔歯学部

職名：准教授

研究者番号(8桁)：60372885

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。