

平成 30 年 6 月 6 日現在

機関番号：63801

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K07214

研究課題名(和文) 細菌群集の機能遺伝子組成の環境に依存した収斂現象の解明

研究課題名(英文) Environment-dependent convergence of functional composition in microbial communities

研究代表者

森 宙史 (Mori, Hiroshi)

国立遺伝学研究所・生命情報研究センター・助教

研究者番号：40610837

交付決定額(研究期間全体)：(直接経費) 3,900,000円

研究成果の概要(和文)：環境が類似した細菌群集間で、系統組成が大きく異なっているが遺伝子機能組成が類似する「細菌群集の収斂現象」について、公共のメタゲノム配列データを収集・解析することで、その存在を明らかにするとともに、収斂現象が観察されやすい環境の特徴を明らかにすることができた。また、収斂現象を引き起こす要因として、当初は遺伝子の水平伝播を仮定していたが、その影響は限定的であり、自然選択等の他の要因が主として考えられることを示唆することができた。

研究成果の概要(英文)：We have developed 16S rRNA gene amplicon and metagenomic sequence analysis pipeline MeGAP2. Using MeGAP2, we have inferred taxonomic composition and functional composition of each 16S rRNA gene amplicon and metagenomic sequencing sample in public INSDC DRA database. In addition, we have annotated the environment information of each sample by using Metagenome/Microbes Environmental Ontology. By comparing taxonomic composition and functional composition, we have identified more than 200 sample pairs of possibly causing environment-dependent convergence of functional composition in microbial communities.

研究分野：微生物生態学・ゲノムインフォマティクス

キーワード：細菌群集 メタゲノム バイオインフォマティクス データベース

## 1. 研究開始当初の背景

細菌は地球上のいたるところに多数の種からなる群集を形成して生息している。細菌はクローン増殖するが、親細胞から子細胞への遺伝子の垂直伝播に加えて、個体間の遺伝情報の伝達方法として、プラスミドや細菌におけるウイルスであるファージなどの転移因子による、異なる種間での遺伝情報の水平伝播が普遍的に存在する。隣接した数十個の遺伝子が一度に種を越えて伝播する水平伝播は、複数の酵素が担う化学物質の代謝経路全体を一度に獲得できる可能性がある。我々の先行研究では、数千サンプルの既存のメタゲノム解析データの遺伝子機能組成と系統組成の比較解析を行った結果から、環境が類似した細菌群集間で、系統組成が大きく異なっているが遺伝子機能組成が類似する「細菌群集の収斂現象」の存在が示唆された。これより、環境が類似した細菌群集間で系統組成が大きく異なっているにもかかわらず、遺伝子の水平伝播と自然選択によって類似した機能遺伝子組成を持った細菌群集が形成される場合があると仮説を立てた。

## 2. 研究の目的

本研究では、既存のメタゲノム配列データを収集・解析した後、サンプル間で詳細に比較解析し、収斂現象が観察されるサンプルペアを抽出する。抽出したメタゲノムデータの詳細な比較解析により、収斂現象が観察される細菌群集で頻度が高く特徴的な機能遺伝子を持つ細菌の系統を推定すると共に、それら機能遺伝子の周囲にファージやプラスミドに関連する遺伝子があるか否かなどを解析することによって、収斂現象が起こる機構が水平伝播によるものか否かを検証し、最近群集において収斂現象が起こる機構の解明を目的とした。

## 3. 研究の方法

### (1)メタ 16S・メタゲノム解析パイプラインの開発

公共の塩基配列データベース中に存在する数万サンプルのメタ 16S・メタゲノム配列データは、シーケンサー由来の単なる配列データ(リードデータ)であり、個々の配列にコードされている遺伝子の機能や由来した系統の情報はアノテーションされていない。それらのリードデータからサンプルごとに系統組成と遺伝子機能組成を高速に推定しサンプル間で比較可能にするために、遺伝研のスーパーコンピュータ上で実行可能な高速なメタ 16S・メタゲノム解析パイプライン MeGAP2 を開発した。

### (2)メタ 16S・メタゲノムデータの収集・解析

公共の塩基配列データベースである INSDC DDBJ Read Archive (DRA) から、環境データ等、配列に付随するデータであるメタデータに対するテキスト検索により、メタ 16S・メ

タゲノムサンプルの配列データとメタデータをサンプルごとに取得した。取得した配列データは、サンプルごとに MeGAP2 で系統組成と遺伝子機能組成を推定した。メタデータについては、そのサンプルが由来した環境についての情報を、我々が開発している Metagenome/Microbes Environment Ontology (MEO) (<https://bioportal.bioontology.org/ontologies/MEO>) 及び Metagenome Sample Vocabulary (MSV) (<https://bioportal.bioontology.org/ontologies/MSV>) を用いて研究協力者と共にマニュアルでアノテーションし、環境に関する記述をサンプル間で統一して整理した。

### (3)収斂現象が観察された細菌群集の詳細な比較解析

(2)のメタゲノムサンプルの遺伝子機能組成推定結果から、ほとんどのサンプルで存在量が普遍的に多い遺伝子を、サンプル間の変動係数及び平均存在量をもとに除外した。また、一部のメタトランスクリプトームデータのように遺伝子機能組成が極端に偏っているサンプルを除外した。その後、系統組成及び KEGG Orthology (KO) を用いた遺伝子機能組成それぞれにおいてサンプル間で相関係数を計算し、両結果を統合した上で、サンプル間で系統組成が大きく異なっているが遺伝子機能組成が類似している、つまりは収斂現象が起こっていると考えられたサンプルペアを抽出した。抽出したサンプルペアが由来した環境を ME0 のアノテーション情報をもとに集計し、収斂現象が起こっていると考えられたサンプルが由来した環境の特徴を解析した。また、それらのサンプルで相対的に多かった遺伝子機能を環境ごとに集計し、遺伝子機能組成としての特徴を解析した。収斂現象が起こっていると考えられたサンプルが相対的に多かった環境に注目し、それらの環境由来のメタゲノムデータについて、サンプルごとに MEGAHIT でメタゲノムアセンブルを行い、優占する系統のゲノム中の遺伝子の並びを推定した。

## 4. 研究成果

### (1)メタ 16S・メタゲノム解析パイプラインの開発

本研究で構築した MeGAP2 は、配列のクオリティフィルタリング、メタ 16S・メタゲノムデータの自動判別、系統組成推定、メタゲノムリードからのタンパク質コーディング遺伝子の予測と機能推定の 4 ステップから構成される (図 1)。MeGAP2 では、メタ 16S・メタゲノムデータから高速に属レベルの系統組成を推定するために GPU を用いた配列相同性検索を行う VITCOMIC2 を使用し (Mori H. et al. 2018)、メタゲノムのリードデータから予測したアミノ酸配列から高速に各サンプルの遺伝子機能組成の推定を行うために、配列相同性検索ツールとして GHOSTX を

使用した (Suzuki S. et al. 2014)。遺伝子機能と配列情報のデータベースとしては、KO 及び KEGG Genes を用いた。MeGAP2 を用いることで、数万サンプルのメタ 16S・メタゲノム配列データから現実的な時間で系統組成と遺伝子機能組成をサンプルごとに推定することができるようになった。

(2)メタ 16S・メタゲノムデータの収集・解析  
 INSDC DRA から、約 6 万サンプルのメタ 16S・メタゲノム配列データとメタデータを取得した。取得した配列データについて MeGAP2 を用いて遺伝研スパコン上で解析を行った。MeGAP2 による解析の際には、paired-end の場合は片側(R1 側)のみ解析に使用し、また、系統組成及び遺伝子機能組成推定の際の正確性の問題から、リード長が 100 base 未満のデータは解析から除外した。MeGAP2 を用いることによって、メタ 16S・メタゲノム約 6 万サンプルの系統組成と、メタゲノム約 4 千サンプルの遺伝子機能組成を推定することができた。それらのメタ 16S・メタゲノムサンプルについて、MEO 及び MSV を用いた環境情報のアノテーションを行い、それらのサンプルがサンプリングされた環境情報を整理した。系統組成・遺伝子機能組成・環境情報のサンプルごとのデータは、NBDC が運用する Life Science Database Archive の MicrobeDB.jp のページ中に存在する FTP サイトから、RDF 形式でダウンロードすることが可能である (<https://dbarchive.biosciencedbc.jp/jp/microbedb/desc.html>)。

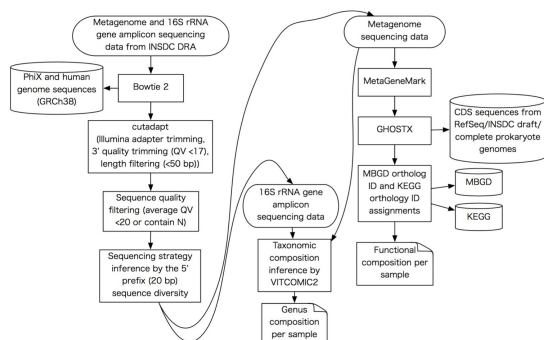


図 1. 本研究で開発したメタ 16S・メタゲノム解析パイプライン MeGAP2 のワークフロー

(3)収斂現象が観察された細菌群集の詳細な比較解析

系統組成データと遺伝子機能組成データの両方が存在したメタゲノムデータ約 4 千サンプルから、遺伝子機能組成が極端に偏っている一部のメタトランスクリプトーム等のサンプル約 200 サンプルを除外した。さらに、*rpoB* やリボソームタンパク質遺伝子等の必須遺伝子のように、サンプル間で相対頻度がほとんど変わらない遺伝子を解析対象から除外した。残ったサンプルについて、遺伝子

機能組成及び系統組成それぞれでサンプル間の相関係数を計算し、系統組成の相関係数の絶対値は低いながら遺伝子機能組成の絶対値は高いサンプルペアを抽出した。抽出されたペアは、相関係数の閾値にも依存するが、概ね 300 から 700 サンプルペアであった。それらの、収斂現象が起きていると考えられるサンプルペアについて、MEO による環境情報のアノテーション結果を整理したところ、大まかには、土壌等の群集組成が多様な環境よりも、宿主に共生する細菌群集や極限環境等、中程度以下の系統多様性の群集に多い傾向が見られた。実際、サンゴに共生する細菌群集で、地理的に離れていて系統組成が異なっているにもかかわらず、遺伝子機能組成が類似する例は報告されており (Thomas T. et al. 2016)、宿主と共生している細菌群集では、収斂現象が起こりやすいと考えられる。なお、遺伝子機能組成は、本研究では KO を主に利用したが、トランスポーター等、より詳細にアノテーションを行うと、KO が同じでも基質が異なる等、どのレベルで遺伝子機能を整理するかによって、収斂現象が起こっているとして抽出されてくるサンプルペアの数は大きく変わることも分かった。また、収斂現象が起こっていると推定されたサンプルペアで、サンプルごとにメタゲノムアセンブルを行った結果から、Scaffold 中の遺伝子のシnten構造をサンプルペア間で比較したところ、遺伝子機能組成が類似したサンプル間でも、シnten構造は大きく異なる場合が多かった。これより、水平伝播等で同一の遺伝子クラスターやオペロンが系統的に離れた細菌間で移動していることが、収斂現象が観察される原因であるケースは少ないことが示唆された。

収斂現象が観察される原因としては、水平伝播以外の自然選択等の要因が主に考えられることが本研究から示唆されたが、研究を進める過程で、実験的な機能解析が詳細に行われている大腸菌や枯草菌等のモデル微生物とは遠く離れているが、環境中では優占している系統群に対して、断片的な配列データからどの程度詳しい遺伝子機能を推定できるかという問題点に直面した。今後、本研究で得られた収斂現象が起こっている可能性が高いサンプルペアの情報と、上記問題を解決するような高度なバイオインフォマティクスの解析技術や高精度なリファレンスデータベースが組み合わされることによって、なぜある環境で特定の遺伝子機能組成を持った群集が形成されるのか、等の研究へと発展できると期待される。

#### 引用文献

1. Mori H. et al. VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing. BMC Systems Biology, 2018, 12:30.

2. Suzuki S. et al. GHOSTX: An improved sequence homology search algorithm using a query suffix array and a database suffix array, PLOS One, 2014, 9:e103833.
3. Thomas T. et al. Diversity, structure and convergent evolution of the global sponge microbiome, Nature Communications, 2016, 7:11870.

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

1. Mori H, Maruyama T, Yano M, Yamada T, Kurokawa K, VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing. BMC Systems Biology, 2018 12:30, 10.1186/s12918-018-0545-2, 査読あり.

〔学会発表〕(計9件)

1. 森宙史, 黒川顕, 微生物群集の系統組成推定ツール VITCOMIC2 の MicrobeDB.jp への応用, 第12回日本ゲノム微生物学会年会, 京都大学桂キャンパス, 3月5-7日, 2018, ポスター.
2. Mori H, Developing tools and database for microbial community analysis, マイクロバイオーム研究開発の最前線, 日本橋ライフサイエンスハブ, January 25, 2018, 招待講演.
3. Mori H, Maruyama T, Yano M, Yamada T, Kurokawa K, VITCOMIC2: Visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing, GIW 2017, Seoul, October 31-November 3, 2017, oral.
4. Mori H, Maruyama T, Yano M, Yamada T, Kurokawa K, VITCOMIC2: visualization of the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing, ProkaGENOMICS 2017, Gottingen, September 19-22, 2017, poster.
5. Mori H, Maruyama T, Yano M, Yamada T, Kurokawa K, VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing data, NIG International Symposium 2017 Commemorating the 30th Anniversary of DDBJ, Mishima, May 27-29, 2017, poster.
6. 森宙史, 微生物群集の系統組成推定・描画ツール VITCOMIC2 の開発と応用, NGS 現場の会第五回研究会, 仙台国際センター展示

棟, 5月22-24日, 2017, 口頭.

7. 森宙史, 丸山貴之, 矢野雅大, 黒川顕, 大量データに対応した微生物群集の系統組成推定・描画ツール VITCOMIC2, 第11回日本ゲノム微生物学会年会, 慶應義塾大学湘南藤沢キャンパス, 3月3-4日, 2017, ポスター.
8. 森宙史, 微生物統合データベース MicrobeDB.jp 2.0 のデータ統合の実際, 第5回生命医薬情報学連合大会, 東京国際交流館プラザ平成, 9月29日, 2016, 口頭.
9. 森宙史, 藤澤貴智, 千葉啓和, 山本希, 内山郁夫, 菅原秀明, 中村保一, 黒川顕, MicrobeDB.jp プロジェクトチーム, 微生物統合データベース MicrobeDB.jp の検索システムの高度化と新解析パイプライン, 第10回日本ゲノム微生物学会年会, 東京工業大学大岡山キャンパス, 3月4-5日, 2016, 口頭.

〔図書〕(計3件)

1. 森宙史, 細菌のゲノム進化と群集ダイナミクス, 生体の科学, 医学書院, 2017, 68, pp. 155-159.
2. 森宙史, 黒川顕, メタゲノム解析・ホロゲノム解析, 化学療法の領域, 医薬ジャーナル社, 2017, 33, pp. 111-117.
3. 東光一, 森宙史, 黒川顕, 今すぐ始める! メタゲノム解析 実験プロトコル, 羊土社, 2016, pp. 15-21.

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

特に無し

#### 6. 研究組織

(1) 研究代表者

森 宙史 (MORI, Hiroshi)

国立遺伝学研究所・生命情報研究センター・助教

研究者番号: 40610837