

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 3 日現在

機関番号：12608

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24770015

研究課題名(和文) 群集としての協調代謝を明らかにするための土壌細菌群集解析

研究課題名(英文) Soil microbial community analysis to identify syntrophic relationships between microbes

研究代表者

森 宙史 (Mori, Hiroshi)

東京工業大学・生命理工学研究科・助教

研究者番号：40610837

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)：芳香族化合物を添加した土壌の細菌群集について時系列でメタゲノム解析を行ったデータを独自の解析手法で解析することで、芳香族化合物の分解に伴いどのような系統および遺伝子が、互いに相関して時系列で群集中における頻度が変動するのかの概要を記述することが出来た。
また、様々な遺伝子の配列が混ざったメタゲノム配列データ中から高精度な系統組成推定を行う際に利用可能な、16S rRNA遺伝子配列のみを抽出して系統推定を行い、そのサンプルの系統組成を推定すると共にサンプル間での系統組成の類似性や各サンプルで特徴的な系統を抽出することが可能なツールVITCOMIC2を開発した。

研究成果の概要(英文)：We conducted a bioinformatics analysis of a time-series metagenomic sequencing data of a soil microbial community with the aromatic compounds pollution. Developing and using the bioinformatics method to efficiently identify fluctuating genes and taxa in time-series metagenomic data, we identified a catalog of fluctuating taxa and genes during aromatic compounds degradation in the soil. In addition, we developed a tool VITCOMIC2 which can (i) identify 16S rRNA gene sequences in metagenome sequence data with minimum false positive rate, (ii) assign taxonomy of the sequences, and (iii) calculate taxonomic composition of the sample.

研究分野：微生物生態学・バイオインフォマティクス

キーワード：細菌群集 代謝 メタゲノム バイオインフォマティクス

1. 研究開始当初の背景

多種多様な細菌が集まって形成された細菌群集は、細菌の多くが未だに培養困難であるため、培養法に基づいた細菌学の各種解析手法が適用できず、未だその生態および機能については不明な部分が多い。特に、土壌は非常に多様な細菌から群集が構成されており、複雑な群集中で異なる系統の細菌間がどのように相互作用して生息しているのかを明らかにすることは、細菌の群集中での生態を理解する上で重要である。しかしながら、細菌間では捕食-被食の関係が稀にしか存在しないため、細菌群集における種間相互作用の実体が何であるのかについて、我々はほとんど知識を持っていない。

本研究では、細菌群集において個々の細菌がつながり群集として機能するために重要な種間相互作用の一つとして、複雑な化合物が添加された場合にまずそれらの分解に特化した少数の種の細菌が分解を行い、分解の過程で生じた中間代謝産物の一部を、他の種の細菌が取り込み利用する、化合物を介した複数種の細菌間の相互作用である「協調代謝」が存在すると仮定した。

2. 研究の目的

本研究では、芳香族化合物を添加した土壌由来の細菌群集から時系列で DNA・RNA のサンプルを取得し、メタゲノム・メタトランスクリプトーム解析を行うことによって、細菌がそれらの芳香族化合物を代謝する過程で生じた中間代謝産物の一部を、他の種の細菌が取り込み利用する「協調代謝」について、その種間相互作用の詳細を解明することを目的とした。

3. 研究の方法

(1) 土壌からの DNA・RNA の抽出条件の検討

特に RNA については、土壌からのバイアスが少ない RNA 抽出法のコンセンサスが研究者コミュニティ内で確立されていなかったため、土壌からの DNA・RNA 抽出条件の検討を行った。単一の畑の土壌から、DNA については MoBio Power Soil DNA Isolation kit を用いて複数回独立に抽出を行い、RNA については MoBio Power Soil Total RNA Isolation kit を用いて複数回独立に抽出を行った。得られた単一土壌由来の複数の DNA および RNA サンプルについて、Illumina MiSeq シーケンサーによるメタゲノムシーケンシングおよびメタトランスクリプトームシーケンシングを行った。得られたシーケンスデータについて、サンプルごとに IDBA-UD によるアセンブルと MetaGeneMark による遺伝子予測、All vs All BLASTP と OrthoMCL によるサンプル間の Homolog 解析を行い、サンプル間でどの程度 Homolog を共有しているかを Jaccard Index や Yue and Clayton theta index などの統計量を用いて解析し、土壌からの DNA・RNA の抽出条件の検討を行った。

(2) 芳香族添加土壌のメタゲノム解析

著者らの先行研究において行った、数年間肥料や農薬などが使用されていない畑の土壌を採取した後、25 度の実験室内で静置し、3 クロロ安息香酸、ピフェニル、カルバゾール、フェナントレンの 4 種類の芳香族化合物を添加する直前、添加して 1 週間後、3 週間後、6 週間後、12 週間後、24 週間後の計 6 タイムポイントの土壌細菌群集についての Illumina GA IIx によるメタゲノムシーケンスデータについて、最新の version の Reference 配列データベース (KEGG) を用いて BLASTX によるリードの遺伝子機能のアサインメントを行った。また、後述する (3) の手法でメタゲノムシーケンスデータ中から 16S rRNA 遺伝子由来リードを抽出し、系統アサインメントを行った。得られた系統組成および遺伝子機能組成のデータを基に、時系列での存在量の変動パターンの類似性による遺伝子機能および系統の、PAM 法および階層的クラスタリング法によるクラスタリングを行った。それらのクラスタリング結果を基に、どの系統間が相互作用しているか、およびその相互作用にはどのような機能遺伝子が関わっているのかを推定した。

(3) メタゲノムシーケンスデータからの高精度な系統推定ツール VITCOMIC2 の開発

細菌群集における細菌間の相互作用の実体を明らかにする上で、メタゲノムシーケンスデータから高精度に系統組成を推定することは非常に重要となる。タンパク質コーディング遺伝子については、細菌の全系統に占めるゲノムが解読された系統の割合が未だ僅かであるため、メタゲノムシーケンスデータからの系統組成の高精度な推定に用いるのは不向きである。そこで、本研究ではメタゲノムシーケンスデータから、Reference 配列データベースが非常に充実している 16S rRNA 遺伝子配列のみを抽出すると同時に各配列の系統推定を行い、そのサンプルの高精度な系統組成を推定するツール、VITCOMIC2 を開発した。次ページの図 1 に上記の解析を行う VITCOMIC2 の具体的なワークフローを示す。

4. 研究成果

(1) 土壌からの DNA・RNA の抽出条件の検討

土壌のメタゲノム・メタトランスクリプトーム解析についての様々な研究で比較的良好に用いられている、MoBio Power Soil DNA Isolation kit と MoBio Power Soil Total RNA Isolation kit を用いて単一土壌から複数回独立に DNA および RNA を抽出してメタゲノム・メタトランスクリプトームシーケンシングを行った後、得られたシーケンスデータからゲノムまたは cDNA 配列を再構築し、予測されたタンパク質コーディング遺伝子配列群について、サンプル間でどの程度共有されているかを評価した。サンプル間の再現性を

見る指標としては、Homolog の種類の一致率のみ評価する Jaccard index と、各 Homolog の存在量の類似性を評価する Yue and Clayton theta index の 2 種類の統計指数を用いて評価した。結果として、メタゲノムデータについては両統計指数がサンプル間で高い再現性があることを示す値となったが、メタトランスクリプトームデータについては、両統計指数がサンプル間で再現性があまり高くはないことを示す値となった。土壌中には腐食酸などの PCR 阻害物質が多数存在しており、その影響で RNA を cDNA に逆転写する際の RT-PCR にバイアスがかかった可能性もあるが、一般に同じ土壌由来でも、土壌中の微細環境は非常に多様であることが報告されており (Vos M et al. 2013)、本研究における抽出前の土壌の攪拌が不十分であったために、土壌粒子中に RNA が吸着されてしまった可能性や、土壌中の RNase などの RNA 分解物質などが RNA を分解してしまっただけの可能性など様々な RNA 抽出・精製効率に影響するパラメータがサンプル間で無視できないレベルで異なってしまった可能性などが原因として考えられる。

結論としては、本研究で用いることを計画していた土壌と RNA 抽出法の組み合わせでは、本研究の目的を達成することが出来る有意義なメタトランスクリプトームデータを得ることは困難であると判断されたため、本研究とは別の目的で 4 種類の芳香族化合物を土壌に加え、6 タイムポイントでサンプリングを行いメタゲノムシーケンシングを行った、著者らの先行研究のデータを新しい視点・データベース・解析手法によって再解析することで、本研究の目的を達成する戦略に切り替えた。

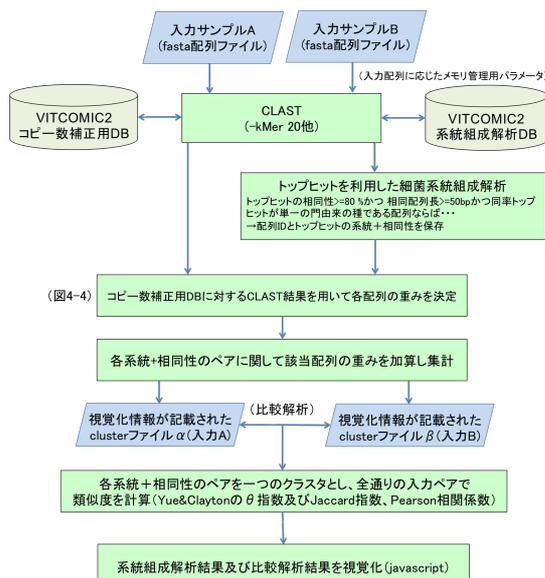


図 1. VITCOMIC2 のワークフロー

(2) 芳香族添加土壌のメタゲノム解析

6 タイムポイントのメタゲノム解析データを用いた、時系列での存在量の変動パターンの類似性による遺伝子機能および系統の階層的クラスタリングの結果の要点は以下である。

- 芳香族化合物を添加する前の土壌で多かった *Acidobacteria* や *Verrucomicrobia* などの系統群の多くは、化合物添加後 24 週目になっても存在量は元に戻っておらず、これらの芳香族化合物添加が土壌の細菌群集へ与える影響は少なくとも約半年間は続いた。
- フェナントレンなどの多環芳香族化合物を分解することが既に報告されている *Mycobacterium* や *Cupriavidus* などは、フェナントレンなどが分解され始める 6 週目に存在量が増加し、それらの化合物がほぼ存在しなくなる 24 週目には存在量が化合物添加前と同等のレベルに戻った。
- ほとんど機能がわかっていない *Dongia* や TM7 などのいくつかの系統が、同様の 6 週目のタイミングで存在量が増加し、その多くは 24 週目には存在量が化合物添加前と同等のレベルに戻るが、*Dongia* などは 24 週目になっても存在量は多少 6 週目よりは減るものの多いままであった。
- 芳香族化合物の分解に関与していると考えられる *Dongia* や TM7 などの系統群は、それらの系統に類似した系統由来のタンパク質コーディング遺伝子の Reference 配列データの不足によって、どのような遺伝子機能を所持しているかはほとんどわからなかった。

これらの結果から、今回 6 週目に存在量が増加していた *Dongia* や TM7 などの系統群については細菌群集中における機能の詳細はわからないものの、既に添加した芳香族化合物を分解することが報告されている *Mycobacterium* や *Cupriavidus* などと同様に、添加した芳香族化合物の分解に関与していると考えられる。しかしながら、*Dongia* などの一部の系統については添加した芳香族化合物が検出されなくなった 24 週目になっても存在量がそれほど減少しなかったことから、*Mycobacterium* や *Cupriavidus* などとは生態的地位が異なることが推測される。土壌由来の *Dongia* や TM7 については近縁系統のゲノム解読が行われておらず、かつ本研究で用いた土壌の細菌群集は非常に多様であるため、メタゲノムシーケンスデータ中から、それらの系統由来の配列のみを抽出し、その系統が持つ遺伝子機能を推定することは困難である。また、これらの系統は近縁系統も含め未だ安定的な培養法が確立されていないため、菌株保存機関等から菌株を取得してゲノムをシーケンスし、メタゲノムの

Reference 配列データとして用いることも困難である。これらの系統の芳香族添加土壤中での生態的地位を明らかにするためには、今後は例えば、芳香族化合物を加えた土壤中から菌体を単離し、シングルセルゲノムシーケンシングを行うなどの戦略を取る必要があると考えられる。

(3)メタゲノムシーケンスデータからの高精度な系統推定ツール VITCOMIC2 の開発

メタゲノムシーケンスデータからの 16S rRNA 遺伝子の抽出については、未培養の細菌の 16S rRNA 遺伝子配列も含んだデータベースである SILVA や RDP、Greengenes などに対して、BLAST で配列類似性検索を行う手法が一般的である。しかしながら、それらのデータベース中には tRNA 遺伝子や ITS 領域、23S rRNA 遺伝子などの配列の断片も多数混入しており、それらの配列は 16S rRNA 遺伝子とは進化速度が異なったりするため、16S rRNA 遺伝子配列以外にそれらの配列も系統組成の推定に用いてしまうと、実際の群集とは異なった系統組成を推定してしまう危険性がある。

そこで、VITCOMIC2 では、それらのデータベースの一つである RDP から配列データを取得した後、著者らの先行研究で構築した、完全ゲノム解読済みの細菌から抽出した高精度な 16S rRNA 遺伝子配列のデータベース (Mori et al. 2010) を Reference 配列データとして、BLAST による RDP 由来の配列の配列類似性検索を行い、RDP 由来の配列中に含まれる 16S rRNA 遺伝子配列以外の配列を除去し、未培養の細菌の 16S rRNA 遺伝子配列を豊富に含んだ高精度な 16S rRNA 遺伝子配列データベースを構築した。

この Reference 配列データベース相手に著者らが開発した GPU を用いた高速塩基配列類似性検索ソフト CLAST を用いて (Yano et al. 2014)、メタゲノムシーケンスデータを検索することで、16S rRNA 遺伝子由来リードのみを抽出すると共に系統推定を行い、図 2 のような円形のプロットとして系統組成を配列間の進化距離も考慮しながら視覚化することが出来るようになった。VITCOMIC2 を用いることによって、様々なメタゲノムシーケンスデータから、16S rRNA 遺伝子配列のみを取得し、系統アサインメントを行い、そのサンプルの高精度な系統組成を推定することが可能になった。本研究で開発した VITCOMIC2 は、今後メタゲノム解析を利用した様々な研究において、重要なツールになると期待できる。

<引用文献>

1. Vos M. et al., Micro-scale determinants of bacterial diversity in soil. FEMS Microbiol. Rev., 2013, 37:936-954.
2. Mori H. et al., VITCOMIC:

visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences, BMC Bioinformatics, 2010, 11:332.

3. Yano M., Mori H., et al. CLAST: CUDA implemented large-scale alignment search tool. BMC Bioinformatics, 2014, 15:406.

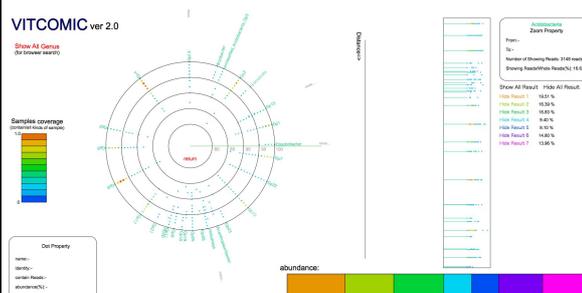


図 2. VITCOMIC2 における細菌群集の系統組成の表現

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 3 件)

- Wang T, Mori H, Zhang C, Kurokawa K, Xing X, Yamada T, DomSign: a top-down annotation pipeline to enlarge enzyme space in the protein universe. BMC Bioinformatics 2015 16:96, 10.1186/s12859-015-0499-y, 査読あり.
- Yano M, Mori H, Akiyama Y, Yamada T, Kurokawa K, CLAST: CUDA implemented large-scale alignment search tool. BMC Bioinformatics 2014 15:406, 10.1186/s12859-014-0406-y, 査読あり.
- Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, Nagata Y, Fujiyama A, Tsuda M, Kurokawa K. Design and Experimental Application of a Novel Non-Degenerate Universal Primer Set that Amplifies Prokaryotic 16S rRNA Genes with a Low Possibility to Amplify Eukaryotic rRNA Genes, DNA Research, 2014 21:217-227, 10.1093/dnares/dst052, 査読あり.

[学会発表](計 3 件)

- Mori H, Maruyama T, Yano M, Yamada T, Kurokawa K, VITCOMIC2 analysis system: visualization tool for taxonomic composition of microbial communities, International Symposium on Genome Science 2015, Hitotsubashi Hall, Tokyo, Japan, January 20-21, 2015, poster.
- Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, Nagata Y, Fujiyama

A, Tsuda M, Kurokawa K. Design and experimental validation of a novel non-degenerate universal primer set for 16S rRNA gene amplicon sequencing with a low possibility to amplify eukaryotic rRNA genes, 2nd Thunen Symposium on Soil Metagenomics, Thunen Institute, Braunschweig, Germany, P5, Dec. 11-13, 2013, poster.

森宙史, メタゲノム解析を用いた細菌群集動態の解明, 第 195 回信州大学理学部生物科学科教室セミナー, 信州大学理学部 10 番講義室, 2012 年 11 月 3 日, 招待講演.

〔図書〕(計 1 件)

山本希, 森宙史, 山田拓司, 黒川顕, メタゲノミクスの現状と未来, 生命のビッグデータ利用の最前線, シーエムシー出版, pp48-57, 総ページ数 231 ページ, 2014 年.

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

6 . 研究組織

(1)研究代表者

森 宙史 (Mori Hiroshi)

東京工業大学・大学院生命理工学研究科・助教

研究者番号 : 4 0 6 1 0 8 3 7