

平成 30 年 6 月 21 日現在

機関番号：82626

研究種目：挑戦的萌芽研究

研究期間：2016～2017

課題番号：16K14646

研究課題名(和文) エマルジョンを利用したメタゲノミクスのための次世代型ゲノム再構築技術の開発

研究課題名(英文) Development of new genome binning techniques using fusion PCR for genome-centric metagenomics.

研究代表者

関口 勇地 (Sekiguchi, Uji)

国立研究開発法人産業技術総合研究所・生命工学領域・研究グループ長

研究者番号：20313570

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：地球上には多様な未培養微生物群の遺伝情報にアクセスするため、メタゲノム情報から個々の微生物群のゲノム情報を抽出、再構築するためのビンニング技術の開発が進められている。しかしながら、それらビンニング技術はメタゲノム情報を所得した後計算機上でゲノム断片を再分類するため、それに付随する様々な問題が存在する。その問題を回避するため、本研究課題ではメタゲノム情報を得る際、各塩基配列情報に対し、それが由来する微生物の系統分類情報を物理的に付加する新しい技術を提供することを目的とする。その方法を可能とするため、同じ細胞に由来する16S rRNA遺伝子と他のゲノム断片を融合し解析する技術を開発した。

研究成果の概要(英文)：Genome-centric metagenomics is a key technique in state-of-the-art microbiome research. This method involves reconstruction of genome sequences from metagenomic reads using advanced bioinformatics algorithms and has been successfully applied for recovering tens to hundreds of population genome (genome bins, GBs). However, methodological advances remain needed to improve our ability to specifically recover GBs associated with populations of interest, identified using e.g., high-throughput 16S rRNA gene amplicon sequencing (16S-seq). Current techniques for genome binning are based mainly on genome signatures such as tetranucleotide frequencies and abundance, without taken into account information of the origin of the reads from specific phylotypes. We developed a new workflow that combines phylotype signatures with traditional metagenome binning strategies, which can significantly improve performance of genome-centric metagenomics.

研究分野：環境メタゲノミクス

キーワード：メタゲノム 難培養微生物 未培養微生物 微生物ゲノム

1. 研究開始当初の背景

近年、DNA シークエンシング技術の急速な発展により、特定の環境に存在する複合微生物群に由来する膨大なメタゲノム情報を容易かつ低コストで入手できるようになってきた。また、最近の生物情報学の技術的進展により、メタゲノム情報から複合微生物群を構成する個々の微生物集団のゲノム（ポピュレーションゲノム）を抽出（ピニング）、再構築することが容易になりつつある。最新の生物情報学技術を利用することで、優占度の低い微生物種を含めた全ゲノム配列を高い精度で再構築することが可能になっている。しかしながら、これらの方法では、コピー数の変動が大きなゲノム断片を正しくピニングすることや、得られたシーケンシングリードのアセンブル時のミスアセンブル（特に異なる微生物種からのゲノム断片が融合したキメラ配列など）を回避することが難しく、技術的な改善が必要となっている。メタゲノム配列断片を次世代シーケンサでバラバラの断片情報として取得する前に、それぞれのゲノム断片にそれが由来する微生物の系統分類情報を物理的に付加する技術が開発できれば、計算機上でのピニングに付随する問題を解決できると申請者は考えた。

2. 研究の目的

上記の着想を具体化するため、図1にある分子生物学的技術を考案した。提案技術では、ゲノム断片時に原核微生物の系統分類において広く使われる 16S rRNA 遺伝子の一部を付加し、16S rRNA 遺伝子と他のゲノム断片が融合したキメラ DNA を作製する。その断片の両末端の配列を次世代シーケンサで取得し、その際得られた 16S rRNA 遺伝子配列に基づきシーケンシングリードを分類する。その後、個別の rRNA 遺伝子配列から得られたゲノム断片を個別にアセンブルすることにより、個々のゲノムを再構築する工程を有する。本研究課題では、本技術の実現可能性を評価するとともに、開発技術を利用したゲノムピニングを実環境試料に適用し、未培養微生物群のゲノム解読に利用しその有効性を検証することを目指した。

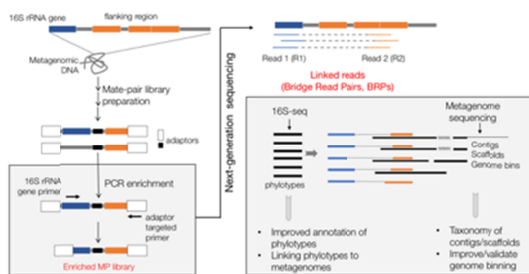


図1 提案技術の概要

3. 研究の方法

我々のアプローチは、メタゲノム断片が同じ分子または細胞に由来する 16S rRNA 遺伝子断片でタグ付けされる、ゲノム断片ペアの形成が重要な要素である。本課題を達成するため、まずはメイトペア（MP）シーケンシングライブラリの作製、その後ユニバーサル 16S rRNA 遺伝子プライマーを用いて 16S rRNA 遺伝子タグを含むリードペアを濃縮する PCR を実施した（図1）。これにより得られた配列ライブラリーは、リード1が 16S rRNA 遺伝子断片を表し、リード2がゲノムの隣接領域に由来する DNA 断片（bridging read pair、BRP）が高度に濃縮されていると考えられる。本研究では、（1）ライブラリ作成技術の最適化と検証、（2）読み取りデータの分析のためのバイオインフォマティクスワークフローの開発、（3）環境微生物の解析への適用を試みた。

4. 研究成果

[4.1] 16S rRNA 遺伝子と他のゲノム断片との融合技術の確立

本技術の実現可能性を検証するため、モデルとして大腸菌のゲノム DNA を用いて Nextra メイトペアライブラリを利用した検証を実施した。Illumina MiSeq シークエンシングプラットフォームを使用し、最適なアニリング温度を用いた 2 ステップ PCR を使用した検証を実施したところ、ユニバーサル 16S rRNA 遺伝子プライマー-515F および 806R をそれぞれ用いた場合、想定される断片を含む配列（BRP）が全配列の約 5% および 25% まで濃縮できることが示された（図2、上パネル）。これは、未濃縮ライブラリーと比較して 100 倍を超える濃縮率を示し、BRP ライブラリを構築するための PCR 濃縮の有効性を実証した。大腸菌参照ゲノム配列への BRP のリード2のマッピング結果は、予想された 7 つの rRNA オペロンそれぞれの上流および下流部位を正しく解読していた（図2、下パネル）。

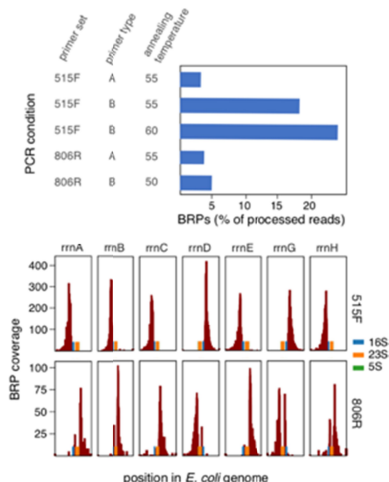


図2 BRP ライブラリ作製結果の一例

The top panel shows the proportion of processed read pairs identified as BRPs for different PCR conditions and two primer sets. The bottom panel shows the coverage of read 2 of recovered BRPs across the *E. coli* genome.

[4.2] 16S rRNA 遺伝子配列に基づきゲノム断片をグループ化する解析スクリプトの作成

BRP ライブラリから生成されたシーケンシングデータの解析を想定し、統合された計算ワークフローを開発した(図3)。SortMeRNA [Bioinformatics 28 (24): 3211-7]を用いてBRPを回収することにより、リードプロセッシング(アダプタトリミング、プライマートリミングおよび品質管理を含む)を自動化した。BRPの同定に続いて、回収されたBRPのリード1は、BLAST検索によるGreengenes 16S rRNA 遺伝子データベースとの比較によってその系統分類群を推定し、BBmapを用いてメタゲノムコンティグにマッピングするように設計した。両方の出力ファイルは、Rのカスタムスクリプトにより処理されるよう設計した。

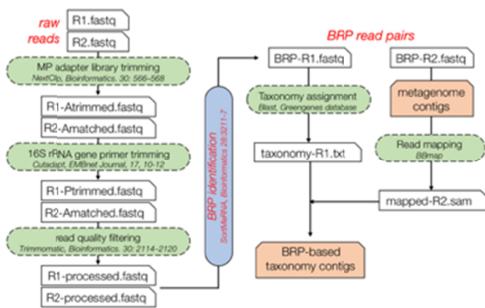


図3 ゲノム断片をグループ化する解析スクリプトの概要

[4.3] 実環境試料に対する開発技術の適用

上記技術の有用性を実証するため、排水処理嫌気性汚泥とマウス腸管マイクロバイオーム(糞便)の2種類の試料に適用した。嫌気性汚泥サンプルでは、使用されたプライマーセット(図4、左パネル)に応じ、BRPが約15~20%を含む高度に濃縮されたBRPライブラリを生成できることを検証した。このサンプルから得られたメタゲノムからのピニングの結果、多数のポピュレーションゲノムが取得されている(Sekiguchi *et al.*, PeerJ 3: e740)。その内、高品質(n=64)ポピュレーションゲノムセットに対するリード2のマッピングを実施したところ、プライマー-515Fでは24のポピュレーションゲノム、806Rでは25のポピュレーションゲノムが少なくとも3つのBRPリードによりその系統分類群の推定が可能であった(図4、右のパネル)。10種類のポピュレーションゲノムについては、両方のプライマーセットのBRPが回収され、両方のプライマーを用いた系統

分類が共に完全に一致しており、本発明者らの技術の精度が確認された。従来のゲノムピニングでは、16S rRNA 遺伝子が多コピー遺伝子であるため、ピニングされたポピュレーションゲノムに含まれないことがほとんどであるが、本手法によりポピュレーションゲノムの系統分群を正しく評価することが可能となることが示された。

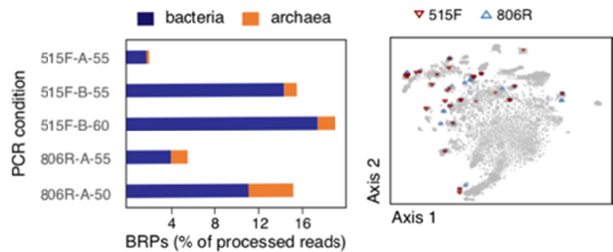


図4 BRP 集積技術による嫌気性汚泥由来メタゲノムのピニングの高度化

The left panel illustrates the efficacy of BRP enrichment using different PCR conditions. The right panel highlights identification of BRPs across the high-quality GBs extracted by genome-centric metagenomics (Sekiguchi *et al.*, 2015). Data points represents individual scaffolds/contigs plotted using VizBin [Microbiome 3(1): 1] based on tetranucleotide frequency and are scaled based on size. Red and blue contigs are supported by at least 3 BRPs for libraries generated with primer sets 515F and 806R, respectively.

次に、本技術をマウス糞便メタゲノムに適用して、特徴的なポピュレーションゲノム、例えば16S-seq分析によって同定されたポピュレーションの推定に本技術を利用した。具体的には、*Bacteroidales*の科レベルの未培養分類群であるS24-7に属するポピュレーションゲノムを抽出することを試みた。本分類群はマウスの腸内では高頻度に見出されるが、この研究の時点ではほとんどその生理学的な機能等は知られていない系統分類群である。メタゲノムシーケンシング(NextSeq 500プラットフォーム)および従来のゲノムピニング(MetaBATを使用)により、110種類の単一コピー遺伝子により60%以上の完全性、15%未満のコンタミネーションを有すると推定される合計62の高品質ポピュレーションゲノムが得られた。そのうち、36(58%)のポピュレーションゲノムには16S rRNA 遺伝子が含まれていなかった。この系について、プライマーセット515Fおよび806Rについてそれぞれ11(28%)および13(33%)のポピュレーションゲノムに16S rRNA 遺伝子タグ(BRP)を関連づけることに成功した。さらに、BRPを使用して目標とするS24-7科に属する4つのポピュレーションゲノムを取得することができた(図5)。

全体に、これらのデータは提案された方法

論の有効性を実証し、ポピュレーションゲノムのさらなる精密化に有効な方法であることが示された。

生命工学領域・主任研究員
研究者番号：00598485

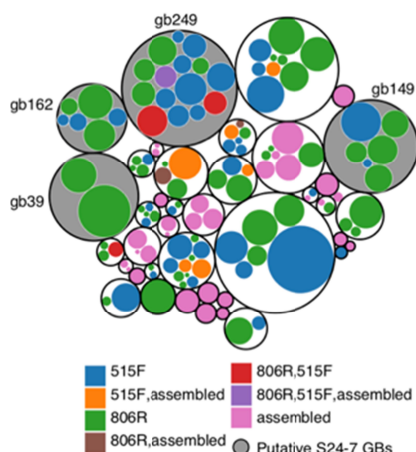


図5 BRP 集積技術によるマウス糞便由来メタゲノムのピニングの高度化

Shown is a packed circle graph illustrating the types of recovered 16S rRNA genes across GBs (*'515F'* and *'806R'* are based on BRPs and assembled indicate 16S rRNA present in the *'assembled'* contigs themselves). GBs lacking 16S rRNA gene are not depicted and for each GB only contigs with a 16S rRNA gene are included. Circles/contigs are sized based on their length and colored based on the types of 16S rRNA genes recovered. Solid black lines enclose individual GBs. GBs identified as putative S24-7 are colored in grey.

なお、エマルジョンを用いて1細胞を1エマルジョンに閉じ込め、その中で16S rRNA 遺伝子配列とそれ以外のランダムなゲノム断片を融合し、*rrn* オペロンの上流および下部位以外の領域のゲノム断片に系統分類情報を付加する試みを同時並行し実施したが、想定される結果が得られず、現時点では技術的に困難であることが判明した。現在、それらの技術的隘路を解消するための研究開発を進展中である。

5. 主な発表論文等

無し

出願状況 (計0件)

6. 研究組織

(1) 研究代表者

関口勇地 (SEKIGUCHI, Yuji)
国立研究開発法人産業技術総合研究所・
生命工学領域・研究グループ長
研究者番号：20313570

(2) 研究分担者

ディーター トゥールース (TOURLOUSSE, Dieter)
国立研究開発法人産業技術総合研究所・