

平成 30 年 6 月 7 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2016～2017

課題番号：16K21144

研究課題名(和文)可視化メタゲノム解析法の開発

研究課題名(英文)Development of targeted genome sequencing in microbiome

研究代表者

按田 瑞恵 (Anda, Mizue)

東京大学・大学院理学系研究科(理学部)・特任研究員

研究者番号：60759455

交付決定額(研究期間全体):(直接経費) 3,200,000円

研究成果の概要(和文):本研究課題では、微生物叢を構成する個々の微生物の探索にFACSを駆使して、標的微生物を可視化、分取し、その核酸を増幅後、次世代シーケンサーで解析する方法論の構築に取り組んだ。全ゲノム増幅産物からのゲノム構築では、キメラDNAが問題となるために従来ショートリードシーケンサーのみが使用されてきたが、本研究ではロングリードシーケンサーを併用することで、高精度なゲノム構築が可能となることを明らかにした。

研究成果の概要(英文):The aim of this study was to develop a methodology for targeted genome sequencing in microbiome by using fluorescent activated cell sorting. This study showed that hybrid assembly of short and long reads improves genome assembly derived from whole genome amplification product.

研究分野:ゲノム生物学

キーワード:全ゲノム増幅 FISH FACS ロングリードシーケンス キメラDNA

## 1. 研究開始当初の背景

メタゲノム解析による微生物探索が大気中、海洋、土壌、ヒト体内問わず、あらゆる環境で行われており、これらの解析結果が続々とデータベースに登録されている。しかし、メタゲノム解析は集団としての変化を観察したに過ぎず、実際に変化のキープレイヤーとなる細菌は何かと問われると答えはない状況である。細菌叢中のキープレイヤーを見出すためには数百種類、何百兆とある集団中からターゲット集団を選択し、さらには個々のゲノムまで抽出して追加解析を行う必要があるものの、こうした方法論は未だに確立されていない。

ターゲット集団からゲノムを構築するために、Fluorescence Activated Cell Sorting (FACS) を用いて可視化した細菌細胞を分取し、全ゲノム増幅を行う方法があるものの、次の問題点が知られていた。(1) FACS ではエアロゾルが発生するため、感染サンプルからの病原体の同定には使用できない。(2) 全ゲノム増幅過程では 10-22 kb に 1 箇所割合でキメラ DNA が生成されることが知られており、ミスアセンブリを含む断片化したゲノムしか構築できない。

## 2. 研究の目的

本研究課題では迅速に腸内細菌叢中のキープレイヤーを見出すために、腸内細菌叢のターゲット集団を可視化し、(メタ)ゲノム解析を行う方法論の確立を目指す。具体的には主に以下2つの技術課題を設定する。

(1) ターゲット集団の可視化および FACS によるセルソーティング法の研究

(2) ターゲット集団のゲノム解析法の確立

特に(2)に関しては、全ゲノム増幅によるキメラ DNA がロングリードシークエンスに与える影響を評価し、ショートリードとロングリードのハイブリッドアセンブリを行うことで、高精度なゲノム構築が可能となるか調べた。

## 3. 研究の方法

(1) まず2種の非病原性細菌を用いて可視化とソーティングの条件検討を行い、続いて

サンプルを供試した。可視化するために、標的細菌の 16S rRNA を特異的に認識するプローブを用いた Fluorescent in situ hybridization (FISH) を行なった。通常の FACS だけでなく、感染検体もサンプルとして使用することを視野に入れて閉鎖系 FACS も使用した。分取率の評価は qPCR で行った。

(2) 全ゲノム増幅産物と通常のゲノム DNA をロングリードシークエンサー (PacBio RSII) でシークエンスし、全ゲノム増幅がリードやコンティグに与える影響を評価した。続いて、ショートリードシークエンサー (MiSeq) のリードとロングリードとを併用することで、高精度なゲノム構築が可能となるか評価した。

評価には、*Vibrio parahaemolyticus* のゲノムを用いた。本菌は、3.3 Mb と 1.9 Mb からなる二つの染色体を有し、rRNA オペロンが 10 あるいは 1 コピー存在するため、バクテリアゲノムの中ではゲノム構築が難しい細菌種として選択した。

全ゲノム増幅反応の鋳型濃度を変えることで、増幅率が  $5 \times 10^3$ - $10^6$  となる全ゲノム増幅産物 4 種を得た。続いて全ゲノム増幅産物と通常のゲノム DNA をロングリードシークエンサー (PacBio RSII) およびショートリードシークエンサー (MiSeq) でシークエンスし、HGAP による *de novo* アセンブリ、あるいは HybridSPAdes などハイブリッドアセンブリを行なった。

キメラあるいはミスアセンブリは、リファレンスゲノムに対してリード、Preassembled Read、コンティグをそれぞれマッピングし、部分的にヒットするものを抽出して、ヒットしない部分を再度マッピングすることで評価した。リードは BLASR および BridgeMapper、Preassembled read は BLAST、コンティグは BLAST および QUAST を用いた。

## 4. 研究成果

(1) 標的集団の FACS による可視化及びセルソーティング法を構築するため、2 種の非病原性細菌を用いた FACS の条件検討を行った。具体的には、腸内細菌叢で優先することの知

られる2種の非病原性細菌(一方をFISH法で染色、もう一方を無染色)の混合物を2種のFACS(開放系及び閉鎖系流路)を用いて分取し、特異的なプライマーを用いたqPCRで分取率を調べた。その結果、開放系FACSでは高精度で濃縮できたが、閉鎖系FACSでは無染色の株も分取されてしまい、全く濃縮できなかった。続いて、FISHで染色した腸管液からの標的細菌の分取を試みたところ、いずれも開放系FACSでは同等の分取率であったことから、開放系FACSによる分取条件を設定できた。

閉鎖系FACSで分取条件を設定できなかった理由としては、通常のFISH法では蛍光輝度が低く、用いた閉鎖系FACSの検出器では標的細菌由来の微弱な蛍光を検出することができないためと考えられた。開放系FACSはエアロゾルを生じるためにそのままでは病原体の分取はできない。新規病原体の同定を行うには、安全キャビネット内に設置した開放系FACSの使用するか、より輝度の高い方法で標的細菌を染めて閉鎖系FACSで分取をするか、他法を用いた分取が必要となる。

(2) 通常のゲノムDNAからゲノム構築したところ、2つの環状となるコンティグから成る完全ゲノムを構築できた。しかし、全ゲノム増幅産物を供試したところ、N50が58-248 kbとなる54-191個のコンティグが得られ、増幅率が増えるほどコンティグ数が増加し、N50が低下することが判明した。続いて、ロングリード、Preassembled read、コンティグにおけるキメラおよびミスアセンブリを評価した。リードでは7.3-12.7 kbに一箇所の割合でキメラが検出され、増幅率が高いほどキメラ頻度が高まった。Preassembled readからは61.8-148.7 kb、コンティグからは93.1-587.7 kbに一箇所の割合でミスアセンブリが検出された。これらの結果から、既報と同等の頻度でリードからキメラが検出されるものの、アセンブリ過程で大半のキメラが除去されること、しかしロングリードだけでは完全にキメラを除去できないことが判明した。

Preassembled readやコンティグにおけるミスアセンブリは、rRNAオペロンに隣接し、

キメラを反映しない場合が観察された。したがって、Preassembled readが短いゆえにミスアセンブリが起りやすくなったと考えられる。rRNA遺伝子を含むリードの大半はPreassembled readを構築する過程で失われていることから、全ゲノム増幅産物からPacBioだけでゲノム構築するには、このアセンブリ過程を見直す必要があると考えられた。

ショートリードとロングリードを併用することで、全ゲノム増幅産物から高精度のゲノムを構築できる方法を探索した。まず、ショートリードだけでは104個のコンティグ(1 kb以上)が構築され、ミスアセンブリが4つ含まれていた。ハイブリッドアセンブリを行なったところ、コンティグ数が14-21個、ミスアセンブリ数が1-3個となった。以上から、ショートリードシーケンサーのみを使用した場合よりもロングリードシーケンサーを併用することでコンティグ数およびミスアセンブリ数を減らすことができ、より精度の高いゲノムを構築できることが判明した。

PacBio RSIIのライブラリ調整に必要なDNA量は、ショートリードよりも多い(ug v. s. ngオーダー)。そこで、『全ゲノム増幅産物由来のロングリード』を『通常のゲノム由来のショートリード』で修正し、両者をアセンブリすることで、高精度なゲノム構築ができるか試した。この結果、ショートリードだけでは46個のコンティグが構築されたのに対し、ロングリードを併用することで5-10個のコンティグまで減らすことができた。ショートリード単独よりミスアセンブリが1-3個増加したものの、第二染色体に相当する環状コンティグが構築される場合も観察された。以上から、ngオーダーしか入手できないゲノムDNAに関しては、本法の適用によってレプリコン情報が得られる程度までゲノム構築の精度を高める可能性がある。

5. 主な発表論文等  
該当なし。

6. 研究組織

(1)研究代表者

按田 瑞恵 (ANDA, Mizue)

東京大学・大学院理学系研究科・特任研究  
員

研究者番号 : 60759455