

平成28年度 委託研究開発成果報告書

I. 基本情報

事業名： (日本語) 革新的先端研究開発支援事業
(英語) Advanced Research and Development Programs for Medical Innovation,

研究開発課題名： (日本語)
メタゲノムアセンブリに基づくメタトランスクリプトーム解析手法の構築とコモンマーモセットメタトランスクリプトーム地図の作成
(英語)

Development of metatranscriptome analysis method based on megagenome assembly and its application to metatranscriptome map of common marmoset

研究開発担当者 (日本語) 慶應義塾大学工学部 教授 榊原康文
所属 役職 氏名： (英語) Keio University, Professor, Yasubumi Sakakibara

実施期間： 平成28年10月1日 ～ 平成29年3月31日

II. 成果の概要（総括研究報告）

・ 和文

榎原康文教授（慶應義塾大学理工学部）のグループは、メタゲノムアセンブリに基づくメタトランスクリプトーム解析手法の構築とコモンマーモセットメタトランスクリプトーム地図の作成を目的として、腸内の部位ごとの細菌叢を理解するために、ヒトの前臨床実験動物で霊長類に属するコモンマーモセットの腸内細菌叢のメタゲノム解析を行った。

健康個体のコモンマーモセットの解剖により、糞便、直腸および横行結腸の内容物を取得し、3種類の部位ごとの細菌叢を回収した。回収された細菌叢をビーズによる菌体破碎を用いて DNA 抽出をした。得られた DNA を 500bp に断片化し、Illumina MiSeq によるペアエンドシーケンスを行った。シーケンスにより得られたリードデータを当研究室で開発されたメタゲノム専用アセンブラである MetaVelvet-SL を用いてアセンブリし、遺伝子予測を行った。MetaVelvet-SL とは、メタゲノム専用アセンブラの第一バージョンである MetaVelvet に機械学習の要素を適用して精度の改良を行ったアセンブラである。MetaVelvet-SL は、菌種が混合したメタゲノムシーケンスデータから菌種別のアセンブリを行うために識別モデルの事前学習を必要とするが、16S rRNA 解析による菌種組成解析結果を元にして識別モデルを学習した。

また、糞便および直腸の DNA については、V3-V4 領域をターゲットにして、Illumin MiSeq による 16S rRNA ペアエンドシーケンスを行った。シーケンスされたリードデータはペアエンドの重複を利用してアセンブリを行い、16S rRNA 配列のデータベースである 16S Microbial を用いて菌種の相同性検索を行った。

ショットガンメタゲノムのシーケンスによりアセンブルされたスキファールド配列の合計長は、糞便、直腸、横行結腸の細菌叢に対して、それぞれ 103Mbp, 63Mbp, 115Mbp, またアセンブリの指標である N50 長は 2.5Kbp, 2.3Kbp, 2.0Kbp となった。次にアセンブルされたスキファールド配列に対して、メタゲノム専用遺伝子予測プログラム MetaGene を用いて遺伝子コード領域の予測を行った。その結果、糞便では 91,503 個、直腸では 143,308 個、横行結腸では 157,961 個の遺伝子が予測された。これらの予測遺伝子に対して相同性検索を行うことにより菌種の情報が付加されたものは、糞便では 63,119 遺伝子、直腸では 41,868 遺伝子、横行結腸では 73,317 遺伝子であった。結果として、これら相同性検索によって菌種の付加されなかった予測遺伝子の数は、直腸では実に 10 万遺伝子にのぼり、これらは新規菌種の遺伝子と考えられる。さらに、COG を用いた遺伝子機能解析を行い、その上位 10 機能について各部位ごとに比較したところ、部位ごとの大きな差は見られないという予想外の結果が得られた。

また、16S rRNA 解析について菌種組成解析を行ったところ、大腸の部位によって、菌種組成が変化することが明らかとなった。具体的には、糞便ではグルコースを代謝して乳酸や酢酸およびギ酸を作り出す *Collinsella tanakaei* がもっとも多く存在し、直腸サンプルについては、宿主の食物繊維の摂取により、血糖値のコントロールをするとされる *Prevotella copri* がもっとも多く存在した。とくに、*Prevotella copri* の組成が部位間で大きく変化していた。

· 英文

To understand the bacterial flora in the intestine with the aim of constructing a metatranscriptome analysis method based on the metagenomic assembly and creating a common marmoset metatranscriptome map, we performed metagenomic analysis of stool and two gut locations: rectal and transverse colon in common marmoset, as a target preclinical primate model of human.

We collected gut luminal samples of rectal and transverse colon and a fecal sample from common marmoset. We extract metagenomic DNA samples by beads processing. We fragmented the DNA samples into 500bp DNA pieces, and sequenced the fragments by Illumina MiSeq.

The read data obtained by sequencing was assembled using MetaVelvet-SL, an assembler specialized for metagenomes developed in our laboratory, and gene prediction was performed. MetaVelvet-SL is an assembler that applied machine learning elements to MetaVelvet, which is the first version of the assembler specialized for metagenome, to improve its accuracy. MetaVelvet-SL requires preliminary learning of the discrimination model in order to assemble the metagenome from the metagenomic sequence data in which bacterial species are mixed.

We also performed 16S rRNA gene amplicon sequencing of V3-V4 region of 16S rRNA by Illumina MiSeq, using fecal and rectal samples. The sequenced read data was assembled using duplication of pair end and homology search of bacterial species was carried out using 16S Microbial which is the database of 16S rRNA sequences.

The total length of scaffold sequences assembled by MetaVelvet-SL was 103 Mbp, 63 Mbp, 115 Mbp for the bacterial flora of feces, rectum and transverse colon respectively, the N 50 length as an index of the assembly quality was 2.5 Kbp, 2.3 Kbp, 2.0 Kbp. For the assembled scaffold sequence, the gene coding region was predicted using the metagenomic dedicated gene prediction program MetaGene, which estimated 91,503 genes in a fecal sample, 143,308 in rectal sample, and 157,961 in transverse colon sample. Homology search annotated 63,119 genes in fecal sample, 41,868 in rectal sample, and 73,317 in transverse colon sample. As a result, the number of predicted genes to which bacterial species have not been annotated by these homology searches actually reaches 100,000 genes in the rectum, and these are thought to be genes of novel bacterial species. Furthermore, the gene function analysis using COG was carried out and the top ten functions were compared for each part, and as a result, unexpected results were obtained that no significant difference was found for each part.

As a result of analyzing bacterial species composition for 16S rRNA analysis, it was revealed that the taxonomic composition of the bacterial species varies depending on the site of the large intestine. In particular, the composition of *Prevotella copri* was changing greatly between sites.

III. 成果の外部への発表

(1) 学会誌・雑誌等における論文一覧（国内誌0件、国際誌0件）

(2) 学会・シンポジウム等における口頭・ポスター発表

1. コモンマーセットの腸内細菌叢メタゲノム解析，口頭，林樹永，小湊みのり，長谷純崇，井上貴史，佐々木えりか，榊原康文，第11回日本ゲノム微生物学会年会，2017/3/2，国内.

(3) 「国民との科学・技術対話社会」に対する取り組み
該当なし.

(4) 特許出願
該当なし.