

## 微生物ゲノムシーケンシング体制の活用による微生物システム解明への基盤構築

●服部 正平<sup>1)</sup> ◆藤 英博<sup>2)</sup> ◆高見 英人<sup>3)</sup>

1) 東京大学大学院新領域創成科学研究科 2) (独) 理化学研究所 3) (独) 海洋研究開発機構

### <研究の目的と進め方>

本研究は、「比較ゲノム」および「応用ゲノム」領域との連携等により、シーケンシング解析体制を駆使して、自然環境下での微生物集団を含めた微生物ゲノム解析を実行し、微生物ゲノム解析支援及び微生物システム解明のための新たな研究基盤の構築を目的とする。

基本的に2つのタイプの研究を進める。ひとつは「比較ゲノム」及び「応用ゲノム」領域から取り上げられるさまざまな微生物を対象とした解析支援である。これに関しては高速廉価なゲノム解析システムを駆使して当該テーマの推進を計る。ふたつ目は自然環境下の微生物集団のゲノム解析（メタゲノム解析）である。本研究では、ヒトの健康と病気/感染症と密接に関係するヒト腸内細菌叢のメタゲノム解析をおもなテーマとする。数百種類の細菌から構成される複雑かつ多様なヒト腸内細菌叢を解明するため、まず細菌叢サンプルの収集法、細菌叢の溶菌法、ゲノムDNAの純化法/ライブラリー作成法、シーケンシングの情報学的解析法等の基本的なメタゲノム解析法を確立する。ついで、確立した解析手法を用いて、ヒト腸内細菌叢、さらにはさまざまな環境細菌叢のメタゲノム解析を進める。

本研究によって、ヒト腸内細菌叢等の環境生息細菌叢における菌種組成、遺伝子組成、機能的特徴等の実体および細菌—細菌—宿主間相互作用等の包括的な微生物生命システムの解明をめざす。また、本研究では、細菌叢の大半を占める難培養性未知細菌、それらがコードする新規遺伝子や代謝物の発見等、学術及び産業上有用なバイオ資源発掘の基盤になると期待される。

### <2007年度の研究の当初計画>

<研究支援> 昨年にひきつづき、「応用ゲノム」および「比較ゲノム」領域との連携と他機関との共同研究として、ヒト常在菌を含めた個別細菌ゲノム（クロストリジウム属4株、バクテロイデス属2株、ビフィドバクテリウム属2株、ルミノコッカス属1株、ロールレラ属1株、マウスのSFB1株など）のフィニッシングを進める。このほかに、*Micromonospora*（放線菌）1株、トレボネマ（ウシ病原菌）1株などのゲノムシーケンシングを開始する。

<細菌叢メタゲノム解析> 昨年度までに、さまざまな年齢層の成人及び乳児（生後3ヶ月以降）を含む健康な13名の腸内細菌叢よりゲノムDNAを調製し、それらのショットガンシーケンシングを行い、得られたメタ配列データの情報学的解析から、遺伝子予測、個々遺伝子の同源性検索による遺伝子 annotation と組成解析、クラスタリングによるCOG解析、個人間や各種環境間における遺伝子同源性解析とCOG頻度比較等の解析を通して、腸内細菌叢ゲノムの多様性、個人間相違、動的变化等に関して研究してきた。

2007度においては、1) 誕生後～3ヶ月以内の乳児ふん便（複

数）から分離した腸内細菌叢のメタゲノム解析を行い、これまで得られていた3ヶ月以降のデータと連続させて、誕生直後から離乳直前までの腸内細菌叢の変化を解析する。2) メタ配列データからその優先菌株のゲノム配列を再構築し、培養によって分離した株との比較を行う。とくに、2人の乳児で50%以上を占めていた *Bifidobacterium breve* および *Bifidobacterium longum* のゲノムシーケンシングの再構築を進める。

### <2007年度の成果>

<研究支援> 「比較ゲノム」および「応用ゲノム」領域との連携によって、2007年度ではヒト常在菌を含めた個別細菌ゲノム（クロストリジウム属4株、バクテロイデス属2株、ビフィドバクテリウム属2株、ルミノコッカス属1株、ロールレラ属1株、マウスのSFB1株など）のフィニッシングを進めた。また、*Micromonospora*（放線菌）1株、トレボネマ（ウシ病原菌）1株のドラフトデータを生産した。このほか、上記ビフィドバクテリウム属の2株に加えて、その近縁種である（腸内以外の膈や口腔由来の）6菌種（*Parascardovia denticolens*, *Scardovia inopinata*, *Gardnerella vaginalis*, *Bifidobacterium dentium*, *Bifidobacterium infantis*, *Bifidobacterium scardovii*）のゲノム解析を「応用ゲノム」との共同で開始した。これらのシーケンシングの目的は、ヒト常在菌の優占菌種であり善玉菌として知られるビフィドバクテリウム属及び近縁種の大規模比較ゲノム解析にある。

<細菌叢メタゲノム解析> ヒト腸内細菌叢のメタゲノム解析で得られた結果を論文報告した（黒川顕他：DNA Res. 14, 169-181 (2007).）。発表した内容をまとめると、(1) 家族を含めた健康人の13サンプル（年齢が3ヶ月から40歳代の男女）について、サンプルあたり約8万リード（塩基数にして約55Mb）を生産し、トータルで約727Mbのメタ配列データを得た。(2) サンプルごとのアセンブリによって、15～50Mbのnon-redundant配列を得た（13サンプルのトータルで479Mb）。(3) 得られたnon-redundant配列からの20,063～67,740個の遺伝子/サンプル（計66万個の遺伝子/13サンプル）を同定した。遺伝子予測のプログラムとして野口、高木らが開発したMetaGeneを用いた。(4) 各遺伝子のアミノ酸配列を閾値（E値=-8）で公的データバンクに対する同源性検索及びクラスタリングを行い、全遺伝子の約3/4が既知遺伝子と有意な配列類似度を有し、1,617～2,921COGs/サンプル（3,268COGs/13サンプル）を得た。腸内細菌叢の遺伝子と既知遺伝子との配列類似度の分布は、70%前後及び95%以上の2カ所に高いピークがあった。95%以上の類似度もつ遺伝子は既知大腸菌などのヒト常在菌由来の遺伝子にヒットしたものであった。一方、70%前後の類似度を示す遺伝子は公的データバンク中の病原菌や環境菌等の遺伝子とベストヒットしていた。この結果は、大部分の腸内細菌遺伝子は、他環境由来の細菌種とは進化的に別

系統の細菌種に由来することを示唆する。(5) 上記の相同性検索から、162,647 個の新規遺伝子候補 (全遺伝子の約 1/4 に相当) を得た。これらと他環境細菌叢 (海や土壌) 由来の新規遺伝子とのクラスタリング解析から、647 個の腸内細菌叢由来の遺伝子だけから構成されるクラスター (5 ~ 48 遺伝子) を見いだした。これらは腸内細菌叢特異的であり、これらの機能解析は腸内細菌叢の機能解明に重要である。(6) 各個人の全遺伝子同士の配列類似度解析から、個人間の腸内細菌叢の関係を調べた。その結果、大人及び離乳後の子供 (9 サンプル) は互いに似た 1 つのグループ (大人タイプ) を形成するが、各離乳前乳児 (4 サンプル、乳児タイプ) ではそれらの間及びそれらと大人/子供との間での配列類似度が明らかに低くなっていた。すなわち、腸内細菌叢の遺伝子及び菌種組成が離乳前後において大きく変化することが明らかとなった。また、親子間や家族内サンプルが他人よりも近い関係にあることや男女の差を示すデータは得られなかった。これらの知見及び上記した既知遺伝子との配列類似度分布は、腸内細菌叢の形成機構や由来を解明する上で新たな視点となる。(7) 接合型トランスポゾン Tn1549 に関連した 5,325 個の遺伝子群が高頻度に存在することを見いだした。これら遺伝子は腸内細菌叢特異的で他の環境由来の細菌ゲノムやメタゲノムデータには存在しない。この結果は、腸内環境が遺伝子の伝達や分散などの水平伝播の場であることを裏付けており、接合型トランスポゾンが大きくそのプロセスに関与していることを示唆している。(8) メタ配列データの細菌種への帰属と組成解析をそれらの既知細菌ゲノムへのマッピングによって調べた。その結果 (95% $\geq$ , 150bp $\geq$  の条件)、離乳前乳児のメタ配列データの 50% 程度がマップされる一方、大人/子供のメタ配列データの 80% 以上はマップされなかった。離乳前乳児では、既知菌種である大腸菌やビフィドバクテリウムなどが優占菌種であったため、マップされたリードの割合が高くなったと解釈できる。一方、大人/子供においては、それらを構成する常在菌細菌のゲノム情報がきわめて少ないことを示しており、これらの単離とそのゲノムシーケンスの必要性を物語っている。

#### <国内外での成果の位置づけ>

今回の論文発表は本特定ゲノム 4 領域に属する班員を中心に構成された日本ヒト常在細菌叢メタゲノムコンソーシアム (HMGJ: Human Metagenome Consortium Japan) がまとめたものであり、これまでに発表された腸内細菌叢のメタゲノム解析論文中、世界最大の配列データ量と解析量となっている。この論文発表及びデータリリースは国内の新聞にも掲載され、また国外からの共同研究の申し込みが既にいくつか来ている (新規遺伝子産物の立体構造解析、我々のメタ配列データを用いた解析パイプラインの構築、野口&高木が開発した遺伝子予測プログラム MetaGene の使用等)。一方、ヒト常在菌の国際的な研究の流れは、米国のヒトマイクロバイオーム計画 (HMP) 及びフランスを中心とした EU による MetaHIT 計画が開始され、日本の HMGJ、米国の HMP、EU の MetaHIT、さらには中国、オーストラリア、カナダ、シンガポールからなる国際コンソーシアム (IHMC; International Human Microbiome Consortium) も 2007 年 12 月に設立された。IHMC では細菌叢のサンプリング条件やデータ解析の標準化/統一化が今後議論される方向にある。データの互換性は国際コンソーシアムにおける人種や食生活の相違等を解析する上で絶対条件であることは言うまでもない。こういった流れの中での HMGJ による先駆的な今回の論文発表はタイムリーであったと考えられ

る。つまり、HMGJ が今回用いた方法論や解析手法が世界標準となるよう IHMC の中で示して行く必要がある。しかしながら、米国 HMP の予算 (5 年間) は HMGJ の 100 倍以上であり、IHMC では数百名の健康及び病態細菌叢のメタゲノム解析及び 1,000 個の常在菌個別ゲノムシーケンスが計画されている。このように、今後ヒト常在菌研究が世界的に進む中で、研究費を含めて日本の研究をどのように位置づけ、国際的にプレゼンスを示すかを真摯に検討する必要がある。

#### <達成できなかったこと、予想外の困難、その理由>

2007 年度においては、誕生後 ~ 3 ヶ月以内の離乳前乳児の腸内細菌叢のメタゲノム解析を進める予定であったが、乳児サンプルの取得が思うようにいかず断念するに至った。このような早期の乳児サンプルの取得は両親へのインフォームドコンセント等、誕生予定よりも前からの周到な準備をする必要があると痛感した。今後も、本特定内で、早期乳児のサンプル取得と解析をできるかぎり進めたいと考えている。

#### <今後の課題>

支援研究である個別細菌ゲノムのシーケンスについては、新たに導入された ロッシュ 454 FLX シークエンサーをルーチン稼働できるように技術確立をはかる。実際の細菌ゲノムのシーケンスは完成版が要求されることから、従来の ABI3730 シークエンサーとの併用による全体的なプロセスのスピードアップとコストダウンをめざす。また解析ターゲットとして、応用/比較ゲノム班と共同して、ヒト常在菌とその近縁常在菌種や近縁病原菌との比較解析を行い、常在性の意味をゲノムレベルで解明する。前述した多種類のビフィドバクテリウム属細菌種の配列決定と比較解析はその一つと位置付ける。また、リソースが整備されつつある院内及び市中感染の原因菌である多種類の薬剤耐性黄色ブドウ球菌株のゲノムシーケンスを進め、市中と院内感染菌の違いやとびひ等の病態の違いをゲノム配列レベルで解明する。

ヒト腸内細菌叢に関しては、前述のメタ配列データのマッピングから、シーケンスされた常在菌が著しく少ないことが判明した。そのため、健康人サンプルの 16S 解析を進め、この 16S 情報より、優占菌種を特定し、それらのいくつかの菌種の個別ゲノムシーケンスを行う。また、潰瘍性大腸炎などの炎症性腸疾患患者からのサンプルを収集し、そのメタゲノム解析を開始する。

#### <成果公表リスト>

1) 論文/プロシーディング (査読付きのものに限る)

1. 0801221008

Morita, H. Kuwahara, T. Ohshima, K. Sasamoto, H. Itoh, K. Hattori, M. Hayashi, T. and Takami, H.: An improved isolation method for metagenomic analysis of the microbial flora of the human intestine. *Microbes Environ.* 22, 214-222 (2007).

2. 0801251406

Kurokawa, K. Itoh, T. Kuwahara, T. Oshima, K. Toh, H. Toyoda, A. Takami, H. Morita, H. Sharma, VK. Srivastava, TP. Taylor, TD. Noguchi, H. Mori, H. Ogura, Y. Ehrlich, DS. Itoh, K. Takagi, T. Sakaki, Y. Hayashi, T. and Hattori, M.: Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* 14, 169-181 (2007).