

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¹⁾ ◇藤 英博²⁾ ◇高見 英人³⁾

1) 東京大学大学院新領域創成科学研究科 2) 理化学研究所基幹研究所 3) 海洋研究開発機構海洋・極限環境生物圏領域

<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、シーケンス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、腸内細菌叢をはじめとした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、環境棲息細菌叢の実体や細菌—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、得られる新規遺伝子や代謝物等は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。

<研究開始時の研究計画>

本研究では、2つの実際的な活動を進める。1つは「比較ゲノム」及び「応用ゲノム」領域の微生物研究者との連携によるゲノムシーケンシングの支援である。これらの連携研究では研究競争を優位に進めるためにより高速で低コストかつ高精度なデータ生産ラインと迅速なプライマリー遺伝子機能注釈システムを確立する。支援内容についてはこれら領域と密接な連携のもとに総括班DNAシーケンシングセンター委員会において適切かつ柔軟に対処する。本研究グループは今までにショットガンライブラリー作製、鋳型DNA調製、シーケンス反応、シーケンサー稼動、アセンブリ、フィニッシングなどのゲノムシーケンスの全技術及び情報学的解析に実績をもつ。

2つ目はヒト常在細菌叢や環境微生物などの難培養微生物を対象としたゲノムシーケンシングをベースとした新たな微生物研究の基盤構築である。前述したように、この実行には上記の従来システムの効率化に加えて、極微量ゲノムDNAのゲノムシーケンシング法の開発（1ピコグラム以下のゲノムDNAからのシーケンスデータ取得）、データアセンブリ工程の高精度化（1鋳型複数リードとアセンブラーを装備したデータ処理システム開発）などが必須である。そのため、本研究では申請者全員が共同して、1～2年内に手法を確立しこれらのゲノムシーケンシングをいち早く実行する計画である。

また本研究では腸内、口腔、皮膚などの常在細菌叢のゲノム解析を実行するが、そのサンプル採取が必要であるため領域外グ

ループとの共同研究も頻繁に行う。それぞれの細菌叢をベースに医学や感染症研究をしている細菌学者（腸内細菌叢の分類、健康状態との関連など）、免疫学者（生活習慣病、日和見感染症など）、歯科細菌学者（歯周病など）などが相手となり、細菌叢ゲノムDNAの純化などのサンプル採取法の確立を共同研究する。この共同研究を通じて各細菌叢の実体や特徴を解明しこれら分野にゲノム研究の基盤を創成する。なお上記全ての研究活動の遂行には代表者である服部が常時加わり全体を総括する。

本研究では期間中にトータルで約200～300Mbのゲノムシーケンシングを目標とする。そのため実験補助スタッフ（5～6名）の謝金、次世代シーケンサーなどの機器類を導入する設備拡充費（レンタル契約）、シーケンス反応などの消耗品費用が大部分の経費の内訳となる。また外国旅費については、本研究分野の著しい進展に関する情報収集および本研究のプレゼンスを示す積極的な国際学会参加が目的である。

<研究期間の成果>

1.微生物ゲノム解析

本研究ではゲノム4領域内での共同／支援研究、本計画研究、領域外共同研究によって微生物のゲノム解析を進めた。現時点で77種類の細菌ゲノム解析を完了し、その解読ゲノムの総塩基数は約235Mbとなった（表1）。その内訳は、病原細菌（31株）常在細菌（27株）、昆虫共生細菌（6株）、産業有用菌（6株）、土壌や海洋棲息の環境細菌（7株）である。この解読塩基数は本研究開始時の計画／目標であった最低値（上述）を越えており、目標は達成できたと考える。この他、本研究ではプラスミド、ファージ、細菌叢16Sリボソーム遺伝子等のシーケンス（表2）及び次世代シーケンサーを用いた真核微生物類（35～125Mb）のドラフトゲノム解析（表3）も行った。これらの解析量から、高速かつ低コストでのシーケンシング技術の微生物ゲノム研究への活用、並びに次世代シーケンサーをいち早く活用できる体制を本研究機関内にほぼ確立できたと結論できる。

各種細菌種の個別ゲノム解析では、昆虫共生細菌種のゲノム研究の2報がScience誌に発表された。ひとつはキジラミに共生するカルソネラのゲノムサイズがわずか160kbで、そこにコードされる遺伝子が182個であることを発見した論文である。これまで、昆虫共生細菌のゲノムは進化の中で縮小化する傾向にあり、その最小サイズは500kb程度であろうと考えられていた。しかし、本研究のカルソネラはそれを大幅に下回る驚くべき極小ゲノムを持つことが明らかになった。カルソネラはほとんどのessential遺伝子を欠失する一方で、アミノ酸合成系を維持する等、宿主との絶対共生が示唆された。現在、カルソネラが棲息するキジラミ菌細胞の転写物解析を行っており、その共生システム解明を進めている。2つ目はシロアリに共生するある種の原生動物の共生細菌のゲノム解析である。この研究から、このシロアリ共生細菌は空気中の窒素からアミノ酸を合成する経路をもつことを発見し、宿主シロアリに不足する窒素源を供給する共生システ

ムが世界で初めて明らかになった。

本研究の計画研究の一環として、26株のヒト常在菌個別ゲノム解析を進めた。その多くは腸内由来である。本解析では、シーケンズがされていないJCM登録株とその既知菌種の16S配列と97%以下の16S配列類似度をもつ分離株を選択的にシーケンズした。これらの常在菌ゲノム情報は腸内細菌叢のメタゲノムデータの解析におけるリファレンスゲノムとして有用である（後述）。

表1. 細菌ゲノム個別解析

	微生物	サイズ (Mb)	種類
1	Adlercreutzia equolifaciens DSM19450T	3.0	C
2	Aeromonas hydrophila	4.7	P
3	Bacillus cereus NC7401	5.2	P
4	Bacteroides massiliensis A_03	4.8	C
5	Bacteroides sp. A_01	5.0	C
6	Bifidobacterium angulatum JCM7096	2.0	C
7	Bifidobacterium bifidum JCM1255	2.2	C
8	Bifidobacterium breve JCM1192	2.3	C
9	Bifidobacterium catenulatum JCM1194	2.1	C
10	Bifidobacterium dentium JCM1195	2.4	C
11	Bifidobacterium longum ATCC15707	3.2	C
12	Bifidobacterium longum infantis 157F-NC	2.4	C
13	Bifidobacterium longum infantis ATCC15697	2.8	C
14	Bifidobacterium pseudocatenulatum JCM1200	2.5	C
15	Bifidobacterium scardovii JCM12489	2.6	C
16	Bifidobacterium sp.JCM16039	2.3	C
17	Bifidobacterium sp. JCM15439	2.3	C
18	Burkholderia multivorans ATCC17616	6.8	P
19	Carsonella ruddii PV	0.2	S
20	Chlamydomydia caviae	1.2	P
21	Chlamydomydia felis Fe/C-56	1.2	P
22	Clostridium acetobutylicum	4.0	I
23	Clostridium chauvoei	3.0	P
24	Clostridium innocuum JCM1292	3.0	C
25	Clostridium paraputrificum JCM1293	3.5	C
26	Clostridium ramosum JCM1298	3.2	C
27	Clostridium septicum	3.0	P
28	Corynebacterium macginleyi TBS-13	2.7	C
29	Cyanothece sp. TU126	6.0	E
30	Cycloclasticus sp. SUSI-4	2.3	E
31	Diaphorina citri 2次共生体	4.6	S
32	Eggerthella sp. YY7918	3.1	C
33	Escherichia coli O103 strain12009	5.4	P
34	Escherichia coli O111 strain11128	5.4	P
35	Escherichia coli O26 strain11044	5.7	P
36	Escherichia coli SE11	4.9	C
37	Escherichia coli SE15	4.7	C
38	Fingoldia magna ATCC29328	1.8	P
39	Gardnerella vaginalis JCM11026T	1.7	P
40	Halomonas elongata	4.1	P
41	Helicobacter pylori F16	1.6	P
42	Helicobacter pylori F30	1.6	P
43	Helicobacter pylori F32	1.6	P
44	Helicobacter pylori F57	1.6	P
45	Ishikawaella sp1	0.8	S
46	Ishikawaella sp2	0.8	S
47	Lactobacillus casei(zeae) ATCC393	2.9	I
48	Lactobacillus fermentum IFO3956	2.1	E
49	Lactobacillus paracasei JCM8130T	3.0	I
50	Lactobacillus reuteri JCM1112	2.0	C
51	Lactobacillus rhamnosus ATCC53103	3.0	I
52	Lactococcus garviae ATCC49156	2.0	E

53	Lactococcus garviae Lg2	2.0	E
54	Lactococcus lastis IO-1	2.4	E
55	Methanocaldococcus gasagricola RMAS	1.8	E
56	Micromonospora olivasterospor	8.0	I
57	Orientia tsutsugamushi IKEDA	2.0	P
58	Parascardovia denticolens JCM12538T	1.9	P
59	Porphyromonas gingivalis ATCC33277	2.4	P
60	Raoultella ornithinolytica AA097	5.5	C
61	Rickettsia japonica	1.0	P
62	Ruminococcus gnavus TBH11	4.0	C
63	Scardovia inopinata JCM12537T	1.8	C
64	Segmented Filamentous Bacteria (Mouse)	1.9	C
65	Serratia marcescens SM28	5.2	P
66	Sodalis glossinidius str.morsitans	4.2	S
67	Staphylococcus aureus TY34	2.9	P
68	Staphylococcus aureus TY825	2.9	P
69	Staphylococcus aureus TY114	2.9	P
70	Staphylococcus aureus ATCC49775	2.9	P
71	Staphylococcus pseudointermedius	2.4	P
72	Staphylococcus hyicus	2.5	P
73	Staphylococcus saprophyticus ATCC15305	2.5	P
74	Streptococcus mutans NN2025	2.4	P
75	Streptomyces griseus IFO13350	8.5	I
76	Treponema phagedenis	3.5	P
77	Wigglesworthia morsitans	0.7	S

合計 234.5

C: 常在菌; P: 病原菌; S: 共生細菌; I: 産業有用; E: 環境菌

表2. プラスミド/ファージ等

名前	ゲノムサイズまたはリード数
Lettuce chloroplast	150kb
Poplar chloroplast	150kb
Phage 8325-4/v469	50kb
Staphylococcus aureus phage	50kb
Staphylococcus hyicus phage	50kb
Escherichia coli plasmid	100kb
Legionella dumoffii plasmid	70kb
Providencia alcalifaciens plasmid	100kb
Sphingomonas 3 plasmids	5000リード
Staphylococcus hyicus plasmid	45kb
トルエン分解 プラスミド	150kb
地殻深部の化石DNA 18S	500リード
イネ共生細菌叢 16S	7000リード
水田土壌細菌叢 16S	18000リード
ヒト全エキソン (Exonome)	1000 Mb (454)

表3. 真核微生物

真核微生物	サイズ (Mb)	シーケンサー
Emericella varicolor	35	454
Phomopsis sp.	50	454
Aspergillus oryzae	40	SOLiD
Chlamydomonas reinhardtii	125	454

2. メタゲノム解析

メタゲノム解析は本特定が我が国で最初の試みであった。本特定の開始とともに、2005年に本特定の微生物学、バイオインフォマティクスの班員と特定外の免疫学や腸内細菌学の研究者からなるコンソーシアム (Human Metagenome Consortium Japan: HMGJ) を立ち上げ、ヒト腸内細菌叢のメタゲノム解析を開始した。

この成果は2007年に論文発表することができた (論文14)。

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。以上を表 4 に示す。

表 4. 13 名の日本人腸内細菌叢のサンプル、シーケンシング、遺伝子同定のまとめ

サンプル名	年齢	塩基数 (Mb)	重複のない塩基数 (Mb)	同定遺伝子数	既知遺伝子数	新規遺伝子数
In-A	45	52.5	29.9	38778	30210	8568
In-B	6 ヶ月	62.8	14.9	20063	15127	4936
In-D	35	55.1	49.5	67740	49079	18661
In-E	3 ヶ月	56.8	28.1	37652	28513	9139
In-M	4 ヶ月	57.8	26.4	34330	27050	7280
In-R	24	55.4	46.8	63356	46104	17252
F1-S	30	53.6	38.9	54151	40771	13380
F1-T	28	55.4	44.3	65156	47955	17201
F1-U	7 ヶ月	53.9	25.8	35260	28711	6549
F2-V	37	55.9	47.0	66461	49955	16506
F2-W	36	54.9	41.0	57213	43625	13588
F2-X	3	56.6	40.1	57446	42452	14994
F2-Y	1.5	56.3	46.3	64942	50349	14593
計		726.9	478.8	662548	499901	162647

(4) 各遺伝子のアミノ酸配列を閾値 (E 値 = 8) で公的データベースに対する相同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、1,617～2,921 COGs/サンプル (3,268 COGs/13 サンプル) を得た。(5) 残りの 162,647 個 (全遺伝子の約 1/4 に相当) は既知遺伝子と有意な配列類似度を示さず新規遺伝子候補である。これら新規遺伝子候補と他環境細菌叢 (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。これらの遺伝子ファミリーは腸内細菌叢特異的であり、腸内細菌叢の機能解明に重要であると考えられる。(6) ヒト腸内細菌叢に特異的に増幅している遺伝子群 (COGs) を腸内環境以外の各種環境細菌の遺伝子データベースとの比較から探索した。その結果、ヒト腸内細菌叢で有意に enrich している 315 個の COGs を同定した (大人で 237 個、乳児で 135 個、両者に共通する COGs が 58 個)。これらの COGs のうち炭水化物の代謝や輸送に関わる遺伝子が大きな割合を占めていた。大人と離乳前乳児における炭水化物の代謝と輸送に関わる遺伝子を詳細に比較した結果、大人では多糖類の分解に関わる遺伝子群が、乳児では単糖類の取り込みに関わる遺伝子群がそれぞれ有意に enrich していることがわかった。この結果は腸内細菌叢の機能組成は食事成分に大きく依存することを強く示唆した。(7) 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。これらの結果は腸内細菌が個人に特異的であることを示唆し、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。(8) 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た。

本論文では、菌種組成解析を各メタゲノム配列リードの既知細菌ゲノム (リファレンスゲノム) へのマッピングによって試みた。90% \geq 、100bp \geq の条件下のマッピングでは、メタゲノム配列リードの 80% 以上はマップされなかった。このことはメタゲノムデータの多くが未知ヒト常在菌に由来していることを示唆し、それらのゲノム情報がきわめて少ないことを意味した。そのため、ヒト常在菌の個別ゲノム解析を進めた (上述)。ヒト常在菌の個別ゲノム解析は国際 HMP でも大きな目標のひとつになっている。現時点において、我々の 26 株も含めて約 300 株の常在菌ゲノムがコンソーシアムにリリースされている。これらと DDBJ 等の公的 DNA データバンクに登録されている全細菌ゲノムをリファレンス (約 1200 ゲノム) として、454 で得られた複数のヒト腸内細菌叢メタゲノムリード (表 5) をマッピング解析したところ、平均して約 50% のリードがマップされ、上記した 2007 年での約 20% のマッピング率よりも菌種帰属できるリード数が増大した。このマッピング法は同属の異種株の識別も可能にするもので、PCR 時におけるバイアスとキメラ配列の生成で問題となっている従来の 16S 配列をベースにした方法よりもより高精度で定量的な菌種組成解析法になると考えられる。HMP では今後数年以内に難培養性細菌種も含めた 1,000 株のヒト常在菌個別ゲノムシーケンスを計画しており、将来的には 90% 以上のメタゲノムリードの菌種解析が可能になると見込まれる。

なお、本研究では、細菌叢の構成細菌種のほぼすべてを溶菌する方法も開発した (論文 16)。本方法は従来のリゾチームに加えてアクロモペプチダーゼというグラム陽性菌の溶菌に効果的な酵素を用いることによって達成できた。これまでいくつかの細菌叢の溶菌法が開発されているが、その評価はほとんどなされておらず、細菌叢メタゲノム解析を進める上できわめて大事な研究である。本方法はこれまでのビーズ粉砕法にくらべて高分子量かつ高収率に DNA を回収でき、古細菌等の溶菌が困難な菌種もメタゲノムデータとして得ることができる精度の高い手法であると考えられる。

ヒト腸内細菌叢以外に、工場排水中のアンモニアを窒素に変換する嫌氣的アンモニア酸化細菌叢 (Anammox)、原油流出海洋表面における細菌叢、イネに共生する細菌叢、並びに次世代シーケンサー 454 を用いた炎症性腸疾患等の病態腸内細菌叢のメタゲノム解析も実行した。現在これらのデータ解析を進めている。

表5. メタゲノム解析

微生物叢	シーケンス量 (Mb)	シーケンサー
13名日本人腸内細菌叢	727	3730xl
Anammox Bacteria	80	3730xl
海洋表面細菌群	40	3730xl
イネ共生細菌叢	700	454
健康・病態ヒト腸内細菌叢	1800	454

3. 次世代シーケンサーを用いた微生物ゲノム解析

本特定で購入したロッッシュ社 454 シーケンサーを種々の配列決定に用いて、その技術確立もめざしてこれまでに 30 回以上のラン (最新鋭バージョンである 454FLX Titanium) を行った。このうち 25 回のランの結果を仕様値に対する相対値として表 6 にまとめた。本機種種の 1 ランにおける仕様値はトータルリード数: 100 万、トータル塩基数: 400Mb、リード長: 400 塩基であり、表 6 には各ランにおける仕様値に対する相対値を示す。この 25 ランでは、平均リード数: 1,036,000 リード (1.036)、平均塩基数: 384.3Mb (0.9608)、平均リード長: 370 塩基 (0.9248) となり、ほぼ仕様値に近いデータが生産されている。Run3 や 18 のように仕様値の 1.5 倍のデータが生産できる場合もあれば、Run14 と 15 のような仕様値の半分以下のデータ生産もあった。これらの原因を現在追跡中であるが、Run15 は平均 GC 含量が 70% を越える DNA であり、454 による高 GC 含量の DNA のシーケンスプロトコールは改良の必要があると考えている。

表 6. 454 FLX Titanium のシーケンス評価 (仕様値に対する相対値)

Run	Total read#	Total length	Read length
1	0.61	0.61	1.00
2	1.30	1.11	0.85
3	1.42	1.53	1.08
4	1.07	0.70	0.66
5	0.85	0.76	0.90
6	1.17	1.08	0.93
7	1.20	1.19	1.00
8	1.13	1.21	1.08
9	1.09	1.08	0.99
10	0.99	0.70	0.71
11	1.10	1.11	1.02
12	1.01	0.69	0.68
13	1.21	1.18	0.97
14	1.19	0.43	0.36
15	0.50	0.35	0.70
16	0.97	0.98	1.01
17	1.07	1.00	0.94
18	1.42	1.42	1.00
19	0.64	0.62	0.96
20	1.13	1.25	1.10
21	0.82	0.90	1.10
22	1.05	1.13	1.08
23	1.12	1.19	1.06
24	1.03	0.91	0.89
25	0.82	0.87	1.07

Total read#: 1M reads=1.0; Total length: 400Mb=1.0; Read length: 400 bases=1.0

454 で得られるゲノム配列の quality についてもいくつかの検討を行った。既に完成版である大腸菌 SE11 株 (4.9Mb) を 454 でシーケンスし、454 リード専用の Newbler でのアセンブリデー

タを完成版配列と比較した。その結果、base call のエラー部位は 207 カ所 (0.0042%) みつかり、その大部分は同じ塩基が 7 個以上並んだホモ塩基部位における 1 塩基欠失または挿入であった。また、5 カ所のミスアセンブリの箇所があった。これらの結果から、454 の配列精度は 99.99% 以上であり、ドラフト配列として十分に使用可能であることを実証した。しかしながら、1 塩基の欠失挿入は遺伝子の同定に不適であり、これらのエラー部位を発見する技術開発が必要であると考えている。また、ミスアセンブリはリード長が ABI37303xl (平均リード長: 700 塩基以上) の約半分であることに起因しており、より長いリード長を生産する 454 の改良型へのバージョンアップが待たれる。

454 を用いたメタゲノム解析で判明したこととして、454 の全リードの中には 10-20% の artificial reads が含まれる。この artificial reads は同じ塩基で始まる同一配列をもったリードのことであり、2 リード以上 (最大で数百リード) の同一配列をもつリードが数百種類存在する。メタゲノム解析はランダムなショットガンシーケンスであるので、このような同一配列が 2 から 100 リード以上生産されることはあり得ない。これらが生産する原因は、1 分子の DNA に対して 1 個以上の固定ビーズが鋳型 DNA の調製における emulsion PCR 中に存在するためである。通常、鋳型 DNA とビーズは 1 : 1 の比で emulsion PCR されるのだが、その比がビーズ過剰になったとき artificial reads が生産される。つまり、同一配列の DNA が複数のビーズに固定され、これらのビーズが独立したポアの中でシーケンスされるため、artificial reads ができる。artificial reads はメタゲノム解析では容易に検出・除去でき、また、アセンブリにも影響を及ぼさないが、たとえば、16S リボソーム配列のようなランダムでないシーケンスングでは artificial reads と正しい reads を区別することができず、定量性を求める菌種組成の解析にはまったく使用できないという大きな欠点をもつことがわかった。この DNA とビーズの混合比はマニュアル操作であり、今後改良する必要が多分にある。これまでに 454 を用いた 16S 配列解析による従来にない網羅性での菌種組成解析の研究がいくつか論文発表されているが、この artificial reads のために定量性は大きく欠けていると推察される。

<国内外での成果の位置づけ>

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) による日本人 13 名の腸内細菌叢メタゲノム解析の論文発表 (論文 14) では、発表時点で世界最大量の配列データ量を公的 DNA データバンクにリリースした。この論文発表は国内の新聞にも報道され、国外においてはこのデータセットを用いた新規遺伝子産物の立体構造解析、メタゲノム解析パイプラインの構築、遺伝子予測プログラムの開発等の多くの研究に有用されている。また、研究代表者は、国内で開催された国内及び国際シンポジウムや米国の環境微生物研究者の定期ミーティングである Metagenomics 2007、フランス、韓国で開催された腸内細菌研究に関するシンポジウムの invited speaker として招待された。さらに、本特定の生命システム情報班員である野口らが開発した遺伝子予測プログラム MetaGene は国際標準として高い評価を現在受けている。最近における米国、EU、日本などの 10 カ国が参加する国際ヒトマイクロバイオーム計画 (HMP) の開始にあるように、ヒト常在菌ゲノム研究は生命科学研究の中で世界的な注目を浴びてきており、この世界の流れの中で上記の Kurokawa 論文は先行研究としてきわめてタイムリーであった。

この他、世界初のシロアリ共生細菌 Termite Group 1 に属する菌種と世界最小ゲノム (160kb) をもつキジラミ共生細菌カルソ

ネラのゲノム解析論文はともに Science 誌に記載され、国内外の一般新聞や Nature News 等にも取り上げられた (論文 8、11、18)。また、これらの昆虫共生細菌は培養できないため、ごく微量 DNA (数百分子) からのゲノム解読の技術開発を行うことができた。

<達成できなかったこと、予想外の困難、その理由>

- ① 本研究では腸内以外に、口腔や皮膚などの常在細菌叢のゲノム解析も実行する計画であったが、そのほとんどが実行できなかった。この原因はサンプリング体制の不備にあった。口腔細菌については、最終年度になって歯科臨床医と共同したサンプリング体制を確立できたが、解析の実行までは至らなかった。
- ② 同様に、炎症性腸疾患などの病態腸内細菌叢の解析を進める計画であったが、思った以上にサンプリング数が集まらず今日に至った。早期のサンプリング体制の確立が必要であった。現在は約 10 サンプルが集まっており、そのメタゲノム解析を開始したところである。

<今後の課題、展望>

ヒト常在菌の研究は、米国のヒトマイクロバイオーム計画及びフランスを中心とした EU による MetaHIT 計画が 2008 年より開始され、公式に国家プロジェクトとして世界的に推進されている。さらに、この米国と EU のメンバーに、日本 (代表研究者と数名の HMGJ メンバー)、中国、オーストラリア、カナダ、シンガポール、アイルランド、韓国が加わった国際コンソーシアム (IHMC; International Human Microbiome Consortium) も設立された。IHMC は数百名の健康及び病態細菌叢のメタゲノム解析及び難培養性菌種も含めた 1,000 株の常在菌個別ゲノムシーケンスを計画している。代表者は IHMC の運営委員として細菌叢のサンプリング条件やデータ解析の標準化/統一化に関与している。データの互換性は人種や食生活の相違、病態細菌叢の特徴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。本研究においても次世代シーケンサー 454 を用いたより網羅性の高いメタゲノム解析にくわえて、これまでに約 30 株のヒト腸内常在菌のゲノムシーケンスを行ってきた (上述)。IHMC 全体では約 300 株の常在菌ゲノムのシーケンスデータが現時点でリリースされている。これらの個別ゲノムデータはリファレンスゲノムとしてメタゲノムショットガンリードの直接マッピングによるより定量性の高い菌種組成解析を可能にする。すなわち、このような国際的な研究戦略に対応したヒト常在菌研究を推進する必要がある。

<研究期間の全成果公表リスト>

1) 論文

1. 0911222315
Morita H, Toh H, Oshima K et al: Complete genome sequence of probiotic *Lactobacillus rhamnosus* ATCC 53103. J. Bacteriol. 191, 7630-7631 (2009).
2. 0911221909
Ishii S, Yamamoto M, Kikuchi M et al: Microbial populations responding to denitrification-inductive conditions in rice paddy soil, as revealed by comparative 16S rRNA gene analysis. Appl. Environ. Microbiol., 75, 7070-7078 (2009).
3. 0911221829
Nakao Y, Kanamori T, Itoh T et al: Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res.,16,115-129 (2009).

4. 0911221811
Nieuwenhuis EE, Matsumoto T, Lindenberg D et al: CD1d-dependent regulation of bacterial colonization in the intestine of mice. J.Clin. Invest., 119,1241-1250 (2009).
5. 0903101453
Hattori M and Taylor TD: The human intestinal microbiome: A new frontier of human biology. DNA Res., 16, 1-12 (2009).
6. 0901142352
Oshima K, Toh H, Ogura Y et al: Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. DNA Res., 15, 375-386 (2008).
7. 0901150042
Kumagai H, Utsunomiya S, Nakamura S et al: Large-scale microfabricated channel plates for high-throughput, fully automated DNA sequencing. Electrophoresis 29, 4723-4732 (2008).
8. 0805032052
Hongoh Y, Sharma VK, Prakash T et al: Genome of an endosymbiont coupling N₂ fixation to cellulolysis within protist cells in termite gut. Science, 322, 1108-1109 (2008).
9. 0911222029
Naito M, Hirakawa H, Yamashita A et al: Determination of the genome sequence of *Porphyromonas gingivalis* strain ATCC 33277 and genomic comparison with strain W83 revealed extensive genome rearrangements in *P. gingivalis*. DNA Res., 15, 215-225 (2008).
10. 0901142348
Morita H, Toh H, Fukuda S et al: Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. DNA Res., 15, 151-161 (2008).
11. 0805032052
Hongoh Y, Sharma VK, Srivastava TP et al: Complete genome of the uncultured 'Termite Group I' bacteria in a single host protist cell. Proc. Natl. Acad. Sci. USA. 105, 5555-5560 (2008).
12. 0911222016
Ohnishi Y, Ishikawa J, Hara H et al: The genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. J. Bacteriol. 190, 4050-4060 (2008).
13. 0911221958
Goto T, Yamashita A, Hirakawa H et al: Complete genome sequence of *Finegoldia magna*, an anaerobic opportunistic pathogen. DNA Res.,15,39-47 (2008).
14. 0801251406
Kurokawa K, Itoh T, Kuwahara T et al: Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res.,14,169-181 (2007).
15. 0911230115
Qin T, Hirakawa H, Iida KI et al: Complete nucleotide sequence of pLD-TEX-KL, a 66-kb plasmid of *Legionella dumoffii* TEX-KL strain. Plasmid, 58, 261-268 (2007).
16. 0911221939
Morita H, Kuwahara T, Ohshima K et al: An improved isolation method for metagenomic analysis of the microbial flora of the human intestine. Microbes Environ. 22, 214-222

- (2007).
17. 0911230143
Miura F, Kawaguchi N, Sese J et al: A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc. Natl. Acad. Sci. USA 103, 17846-17851 (2006).
 18. 0701280312
Nakabachi A, Yamashita A, Toh H et al: The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science, 314, 267 (2006).
 19. 0911230135
Kanamoto H, Yamashita A, Asao H et al: Efficient and stable transformation of *Lactuca sativa* L. cv. Cisco (lettuce) Plastids. Transgenic Res. 15, 205-217 (2006).
 20. 0605311556
Izawa M, Kitamura N, Odake N et al: A rapid and simple transcriptional sequencing method for GC-rich DNA regions. Jpn J Vet Res. 53, 159-68 (2006).
 21. 0701280255
Azuma Y, Hirakawa H, Yamashita A et al: Complete genome sequence of the zoonosis agent *Chlamydomonas reinhardtii*. DNA Res., 13, 15-23 (2006).
 22. 0605311525
Toh H, Weiss BL, Perkin SAH et al: Massive genome erosion and functional adaptations provide insights into the commensalistic lifestyle of the tsetse endosymbiont, *Sodalis glossinidius*. Genome Res., 16, 149-156 (2006).
 23. 0605311605
Kuroda M, Yamashita A, Hirakawa H et al: Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection. Proc. Natl. Acad. Sci. USA 102, 13272-13277 (2005).
- 2) 学会発表／招待講演
- 0912011338,0912011332,0912011329, 0912011326, 0912011323, 0912011300,0912011257,0912011252, 0912011250, 0912011248, 0912011245,0912011242,0912011238,0912011235,0912011232, 0912011227,0912011209,0912011204, 0912011159, 0912011152
- 3) 図書
- 0911231711,0911231706,0911231703,0911231659,0911231654, 0911231651,0911230453,0911230451,0911230448,0911230446, 0911230443,0911230441,0911230437,0911230429,0911230427, 0911230425,0911230422,0911230420,0911230412,0911230406, 0911230402,0911230359,0911230357,0911230354,0911230350, 0911230344,
- 4) データベース/ソフトウェア
- 0911231813 (ヒト腸内細菌叢メタゲノムデータ)
<http://metagenome.jp/microbes/>
- 5) 新聞発表、その他
- 0911231646 (読売、NY Times 他), 0911231638 (読売), 0911231635 (毎日), 0911231607 (日経), 0911231604 (日経), 0911231600 (朝日), 0911231555 (日刊工業), 0911231549 (毎日, Nature news 他), 0911231544 (朝日), 0911231717 (NHKサイエンスゼロ)