

ヒト腸内細菌叢のゲノムシーケンス

服部 正平^{1*}, 林 哲也², 黒川 顕³, 伊藤 武彦⁴, 桑原 知巳⁵^{1*}東京大学大学院・新領域創成科学研究科,²宮崎大学・フロンティア科学実験総合センター,³奈良先端科学技術大学院大学・情報科学研究科,⁴(株)三菱総合研究所・先端科学研究センター,⁵徳島大学大学院・ヘルスバイオサイエンス研究部Masahira HATTORI,^{1*} Tetsuya HAYASHI,² Ken KUROKAWA,³ Takehiko ITOH⁴ and Tomomi KUWAHARA⁵^{1*}Graduate School of Frontier Sciences, The University of Tokyo²Frontier Science Research Center, University of Miyazaki³Graduate School of Information Science, Nara Institute of Science and Technology⁴Research Center for Advanced Science and Technology, Mitsubishi Research Institute⁵Institute of Health Biosciences, The University of Tokushima Graduate School

要旨 メタゲノム解析は培養ステップを経ないで自然環境下に棲息する細菌集団(叢)のゲノム塩基配列を直接決定し、情報学的解析からその細菌叢がもつ生体システムを解明する新たなゲノム解析手法である。これにより、従来では解析困難であった難培養性細菌が大部分を占めるさまざまな環境細菌叢の機能実体を包括的に知ることが可能となった。著者らはメタゲノム解析手法を用いてヒト腸内細菌叢の解析を行った。

キーワード: メタゲノム; 腸内細菌; シークエンス

1. はじめに

今日までに、細菌の生態系や多様性の解明、医療や産業に利用できる有用酵素や代謝物の探索等を目的にして、数千種の細菌が環境中のさまざまな細菌集団(叢)から分離培養され、千種以上の細菌のゲノム解析等、個々細菌のもつ性質や機能の研究が行われてきた。しかし、現場環境を技術的に実験室で再現できないことや、細菌間の共生関係の実体が不明であるなどの理由から、これら分離できる細菌は地球に生息する細菌全体の1%未満と言われている。すなわち、これまで研究されてきた細菌種の大半は培養可能な菌種に限定され、大部分を占める難培養性の細菌種に関する知見は皆無に等しいと言っても過言ではない。そのため、これら難培養性細菌の同定やその遺伝子解析には、PCRや活性相補実験等を用いた環境細菌叢DNAからの16S rDNAや特定遺伝子の直接クローニング等が行われている。しかしながら、これらの方法は、既知遺伝子に類似した一部の遺伝子の

単離には効果的だが、細菌叢に存在する多様性が膨大未知な遺伝子全体をカバーすることはできない。また、16S rDNA配列情報は菌種の同定には有効だが、細菌の性質や機能の解明には直結しない。

このような難培養性細菌を含む環境細菌叢の全体像を解明する方法として、培養を経ないで環境細菌叢DNAのシーケンス情報を直接かつ網羅的に獲得するメタゲノム解析が開発された(たとえば、総説6, 7)。メタゲノム解析によって片寄りのない大量の遺伝子情報が得られるため、99%以上を占める未知細菌の正体とともに、それらが数種から数万種で構成する自然界細菌叢の生態や多様性等を解明する糸口となる。そして、これら細菌叢の大半が未知細菌である理由から、メタゲノム解析によって発見が予想される新規な細菌、遺伝子、代謝反応、代謝物質の数はこれまでの数百倍になることが必然見込まれる。これらは医療、エネルギー、食糧、環境等の幅広い産業分野において、これまでの限られた微生物資源をはるかに凌駕した多種多様なバイオ資源になると期待できる。

2. メタゲノム解析

メタゲノム解析の基本プロセスは、同一環境下に棲息

2007年6月4日受付

* 〒277-8561 柏市柏の葉5-1-5 5-1-5 Kashiwanoha, Kashiwa 277-8561, Japan

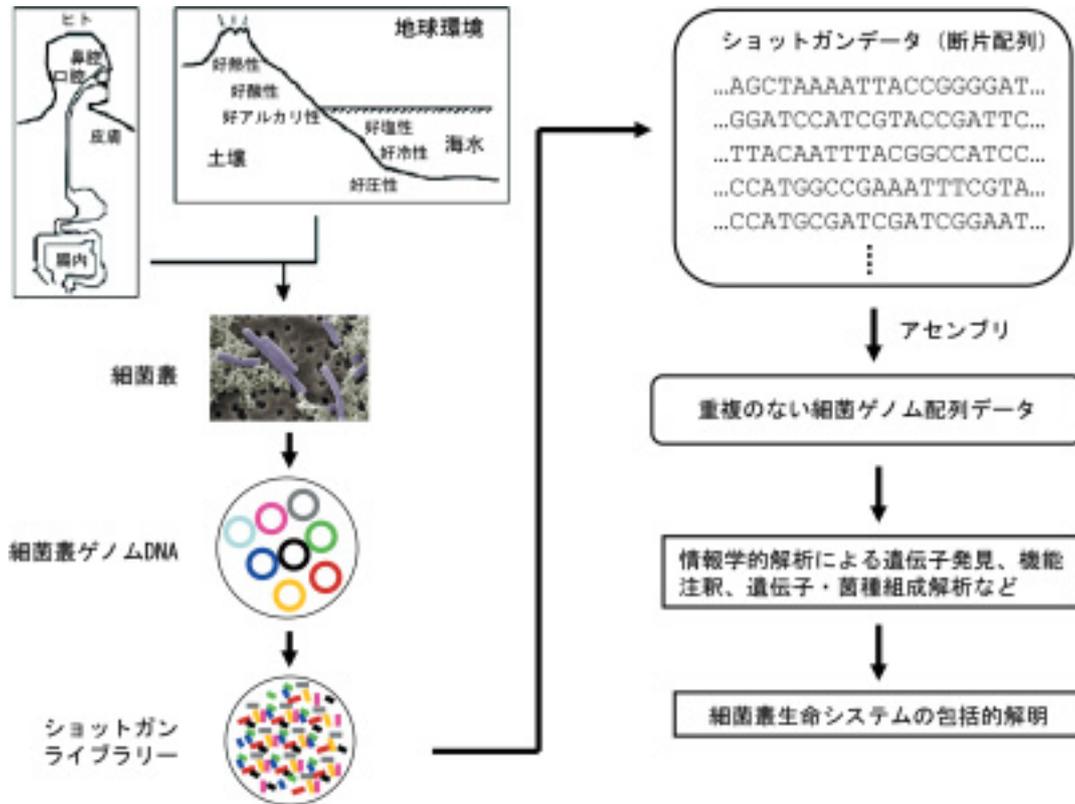


図1. 自然環境細菌叢のメタゲノム解析.

している細菌叢の全DNA（構成細菌種ゲノムの混合物）の調製，ショットガンライブラリーの作成，ショットガンシーケンシング，シーケンシングデータの情報学的解析（遺伝子アノテーション等）からなる（図1）．これによって，そこに棲息している各菌の培養性に関係なく，それらがコードする遺伝子などのゲノム情報をバイアスなく取得できる．これまでに，2004年に発表された鉱山からの強酸性排水中の細菌叢（18）をはじめとして，海洋細菌群（14, 19），肥沃な農場土壌細菌叢等（16）等，現在では70以上のプロジェクトが世界中で進められている（5）．

3. ヒト腸内細菌叢のメタゲノム解析

著者らは2005年度より，さまざまな年齢層の成人及び離乳前後の幼児を含む健康な個人の腸内細菌叢の解析（遺伝子予測，個々遺伝子の相同性検索による機能注釈，クラスタリングによるCOG（Cluster of Orthologous Group）解析，個人間や各種環境間の比較メタゲノム等の解析など）を通して，腸内細菌叢の多様性，個人間相違，動的变化等について調べてきた．以下に結果の一部を紹介する．

3.1. 細菌叢ゲノムDNAのシーケンシング

腸内細菌叢メタゲノムのシーケンシングは，(1) ふん便からの腸内細菌（原核生物）叢の分離，(2) 細菌叢からのゲノムDNAの調製，(3) ショットガンライブラリー作製，(4) 鋳型DNAの調製とシーケンシングの工程からなる．高純度の細菌叢の単離は，ふん便をPBS等のバッファーに懸濁し，ナイロンフィルターを用いて濾過することによって行う（この操作により夾雑物として考えられるヒト細胞はほぼ完全に除去できる）．細菌叢はリゾチーム／SDS等を用いた方法で溶菌し，高分子量のゲノムDNAを得る．ショットガンライブラリーは，高分子量DNAをヒドロシェアによって断片化し，3 kb前後のDNAをアガロースゲル電気泳動より分取し，プラスミドベクターにクローニングすることによって得る．鋳型DNAはショットガンライブラリーから任意にピックアップした大腸菌コロニーから，PCRによってインサートDNAを増幅することによって調製する．ついで，得られたPCR産物の一部をABIまたはGE Health社のダイターミネーターシーケンシングキット（サイクルシーケンシング法）によってシーケンシングする．なお，1つの鋳型DNAについてその両端をシーケンシングする．上記の工程によって，90%以上の成功

表1. サンプル, シークエンシング, 遺伝子同定 (結果の一部)

サンプル	年齢	リード数	総塩基数 (bp)	コンティグ数	シングルトン数	決定サイズ (Mb)	遺伝子数
個人	6ヶ月	80617	62792581	1721	8481	14.88	20063
個人	35	84237	55137918	7613	36312	49.55	67740
個人	3ヶ月	80852	56781600	4819	16838	28.07	37652
個人	24	85787	55404826	8935	36524	46.79	63356
家族	37	80772	55926002	7919	38442	47.02	66461
	36	79163	54885684	6778	30550	40.97	57213
	3	80858	56587120	5032	34252	40.05	57446
	1.5	79754	56276047	9159	32461	46.31	64942

率で細菌由来の配列データを得ている。今回の解析では、家族を含めた健康人のサンプル（年齢：3ヶ月から40歳の成人男女）について、各サンプルあたり8万リード、トータルで100万リード以上の配列データを生産し、計約727 Mbのメタ配列データ (phred \geq 15) を得た。

3. 2. 配列データのアセンブリと遺伝子の同定

配列データを個人ごとに phred-phrap-consed または PCAP ソフトウェアを用いてアセンブリを行い、それぞれの重複のない配列データを得た。得られるコンティグやシングルトン（アセンブリしないリード）の数は、サンプルごとに大きく変動し、サンプルごとの最終的な重複のない配列データの塩基数は約15 Mbから50 Mbとなった。これは、各サンプル中の優先菌種の割合を反映しているからだと考えられる。つまり、同じゲノムをもつ菌が優先的にたくさん存在する場合、重複しない配列のサイズは小さくなる。得られた重複のない配列データからの遺伝子同定には、HMMアルゴリズムをベースにした MetaGene というメタゲノム配列用に開発されたソフトウェア (12) を用いた。このプログラムによって、各サンプルあたり2~6.8万個のたんぱく質をコードした遺伝子 (ORF \geq 20残基) を同定した。このサンプル間の遺伝子数の変動は、上記した重複のない配列データのサイズとおおむね比例関係にある。同定された遺伝子の平均サイズは535 bpであり、通常の個別細菌ゲノムから同定される遺伝子の平均サイズ（たとえば、大腸菌 K12株の平均遺伝子サイズは950 bp）よりも短い。シングルトンだけから同定された遺伝子の平均サイズが417 bpであることから、メタデータ中の平均遺伝子サイズが短い理由は、シングルトンに含まれる断片化ORFsに大きく依っているものと考えられる。またこのことは、断片化ORFsのいくつかが同じ遺伝子に由来することも示唆しており、今回同定された遺伝子数は、実際の遺伝子数よりも多く見積もっている可能性がある。MetaGeneを用いてサルガッソーのメタデータから遺伝

子を同定したところ、報告されている遺伝子数の約9%多い数字が得られている。表1にサンプル、シーケンシング、アセンブリ、遺伝子同定についての結果の一部を示す。

4. 腸内細菌叢ゲノムの情報学的解析

図2に著者らが確立した基本的な腸内細菌叢遺伝子解析パイプラインを示す。腸内細菌叢メタデータ中に同定された遺伝子セットについて、公的データバンク中に登録されている細菌遺伝子に対する相同性検索 (BLASTP解析) を行った。用いたデータバンク中の細菌ゲノムは、NCBIによって全遺伝子のCOGが特定されている343個の細菌ゲノムである。

4. 1. 腸内細菌叢遺伝子の機能注釈

公的データバンクへのBLASTPの結果、腸内細菌叢の全遺伝子の48%が有為な相同性をもって3,268個のCOGに特定された。各サンプルあたりでは、離乳前乳児では1,617~2,857個のCOGが、大人及び離乳後の子供では、2,355~2,921個のCOGがそれぞれ特定された。この結果は、離乳前の腸内細菌叢の遺伝子セットが、大人/子供に比べて機能的に変動または多様化しており、一方、大人/子供のそれは互いにある程度似ていることを示唆している。さらに、各サンプルのCOG数は他の環境細菌叢であるサルガッソー (5,184 COGs) の海洋細菌、土壌細菌叢 (4,423 COGs) よりも少なく、それぞれのシーケンス量の違いはあるものの、腸内環境がこれらの他環境に比べて、よりシンプルな生体反応系であることも示唆している。

同定した腸内細菌叢遺伝子と既知遺伝子との間のアミノ酸配列の類似度は、大腸菌、*Bacteroides*, *Bifidobacterium*等のヒト腸内由来の細菌遺伝子との相同性はあきらかに高い (大部分は90%以上) が、これら以外の細菌種 (たとえば、*Bacillus*や*Clostridium*) の遺伝子とベストヒットする腸内細菌叢遺伝子の類似度

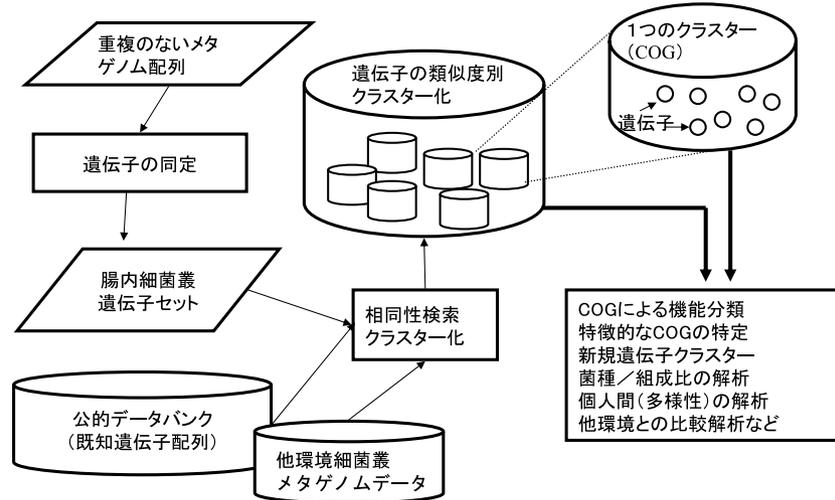


図2. 腸内細菌叢メタゲノムデータの情報学的解析パイプライン.

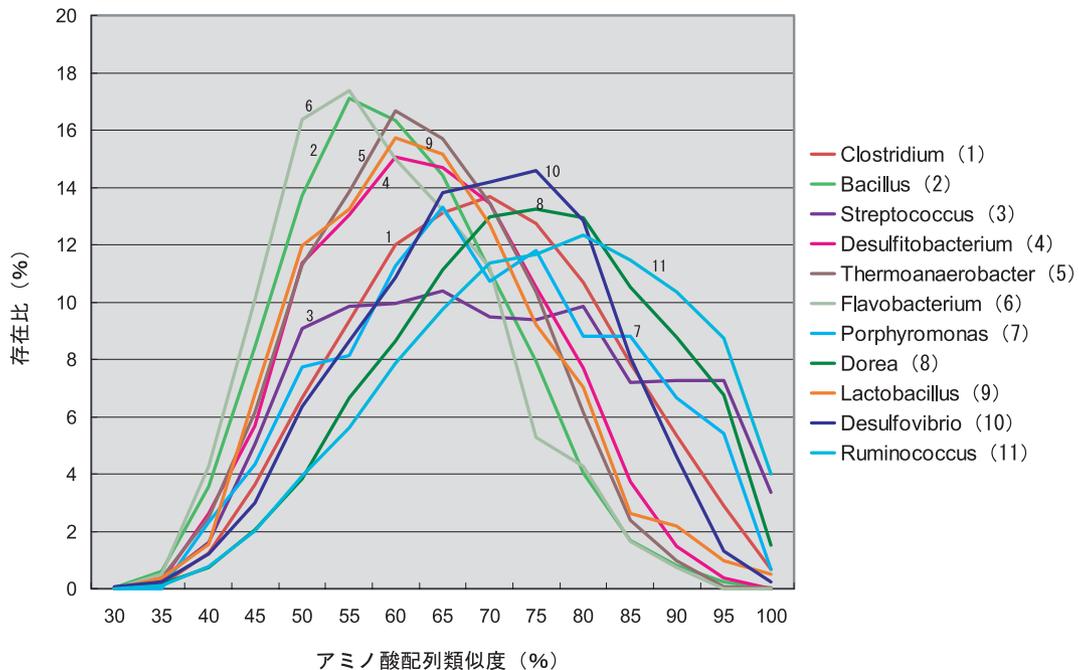


図3. リストした菌種 (右カラム) の遺伝子と腸内細菌叢遺伝子とのアミノ酸配列類似度分布.

は50～70%をピークとした分布を示す (図3)。これらの結果は、ヒト腸内細菌遺伝子は他環境由来の既知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。

4. 2. 個人サンプル間の配列類似度

個人サンプル間の遺伝子セットの配列同一性を調べると、上記したCOG解析と一致して、大人/子供は互いに似た配列をもって1つのグループを形成するが、乳児

間の類似度は大人/子供サンプル間の類似度よりも明らかに低いことがわかった。また、この解析からは、家族内のサンプルが他人とのサンプルに比べてより近い関係にあるという結果や、男女間で何らかの相違を示す結果は得られなかった。つまり、同一の食事をともにする機会の多い家族や遺伝的に近縁の親子/兄弟が、良く似た腸内細菌叢を大部分において共有するという証拠は得られなかったことになる。同様に、同一性間での共通性も見られなかった。これらの知見及び図3に示した結果

は、腸内フローラの形成機構や由来を解明する上で新たな視点となる。

4. 3. 腸内細菌叢に特徴的な遺伝子

各個人の腸内細菌叢の全遺伝子セットの特徴を解析するために、243個の既知細菌ゲノムの遺伝子セットよりも有意にメタデータ中で頻度高く存在 (enrich) している遺伝子 (COG) セットの探索を行った。これら243個のゲノムには、大部分が同じ遺伝子で構成される同種の異株のゲノム及びあきらかにヒト腸内由来のゲノムは含まれない。解析の結果、今回解析した全サンプルにおいて“炭水化物の輸送と代謝”に関係したCOGが243個の他環境菌遺伝子セットよりも有意に多く、一方で、“脂質の輸送と代謝”に関係したCOGが有意に減少していることがわかった。“炭水化物の輸送と代謝”が多いことはアメリカ人 (大人) 2名のメタデータの解析と一致する (4)。また、大人/子供のメタデータでは、“防御機構”がenrichし、“細胞運動性”、“二次代謝物の合成や輸送”、“翻訳後のたんぱく質修飾やシャペロン”に関係した遺伝子が減少していた。

結果的に、大人/子供の腸内細菌叢で共通してenrichしているCOGが237個見いだされた。これらのCOGのうち188個 (約80%) はアメリカ人の腸内細菌叢サンプル (4) においても同様にenrichしていた。一方、これらのCOGのうちのわずか5~10%だけが他の環境細菌叢においてenrichしていた。これらのことから、大人/子供に共通したCOGの大部分はきわめて腸内特異的にenrichしている遺伝子群であると考えられる。このうち、53 COGs (24%) は“炭水化物の輸送と代謝”に関係したものである。それらは、植物由来の多糖類やヒト宿主の細胞に由来する proteoglycans または glycoconjugates を加水分解する glycosyl hydrolases、また、多糖類の分解によって生成する単糖類や二糖類の代謝に関係する L-fucose isomerase, L-arabinose isomerase, galactokinase などである。このほかに、ある種のペプチダーゼもenrichしていた。一方で、脂肪酸代謝に関わる遺伝子の大部分は減少していた、これらのことは、腸内細菌叢はもっぱら食物中の未分解多糖類とペプチドをおもなエネルギー源としていることを示唆する。

ABC-type antimicrobial peptide transport system や ABC-type multidrug transport system のような抗菌性ペプチド及び多剤性の薬剤排出ポンプに関係する遺伝子もenrichしていた。宿主の腸管細胞や多くの細菌は β -defensins のような抗菌性ペプチドを生産することが知られており、これは病原菌などによる外的攻撃からの防御機構の一つと考えられている。つまり、腸内常在菌は

自身でこれらの排出ポンプを多く備えることによって、抗菌性ペプチドや薬への耐性を増大させ、腸管内での生存能力を高めているものと考えられる。

Mismatch repair ATPase (MutS family), DNA primase, DNA-damage-inducible protein J などの“DNAの修復”プロセスに含まれるCOGもenrichしていた。これらのことから、腸管には食物由来及び宿主細胞と腸内細菌によって生成するニトロソ化合物、インドールなどの環状アミン類や二次胆汁酸などのDNAにダメージを与える物質が存在し、腸内細菌のゲノムが予想以上に変異や切断などのダメージを受けていると考えられる。このほかに、大人/子供の腸内細菌叢遺伝子では flagella や chemotaxis に関する遺伝子が著しく少ないことがわかった。このことは、腸管では活発な細胞運動性や走化性等の機能をもった細菌種が他の環境細菌叢に比べて少ないことを示唆する。

離乳前乳児の腸内細菌叢の遺伝子セットは、大人/子供のそれらと大きくその組成を異にし、また互いの共通性も低い。しかし、これら離乳前の乳児間にも共通してenrichしている136個のCOGが存在する。136個のCOGのうち47個 (35%) は“炭水化物の輸送と代謝”に関係したCOGであり、大人/子供でもenrichしていた glycosyl hydrolases が含まれ、さらに植物由来の多糖類を分解する pullulanase, arabinogalactan endo-1,4- β -galactosidase, endopolygalacturonase 等もenrichしていた。これらの酵素はおそらく母乳に含まれるオリゴ多糖類や宿主由来のムチンのような proteoglycans の分解に働いているものと考えられる。つまり、離乳前の腸内細菌叢には、既に食物に多く含まれる多糖類を分解する能力がある程度備わっていることになる。

離乳前乳児にもっとも特徴的なことは、さまざまなトランスポーターが顕著にenrichしていることである。とくに phosphotransferase systems (PTSs) はラクトースや母乳に含まれるさまざまな低分子糖類の取り込みに働いていると考えられる。また、29個のCOG (22%) は、アミノ酸、長鎖脂肪酸、ヌクレオチド、ビタミン等の母乳に含まれている栄養物質を取り込むトランスポーター関係の遺伝子であった。なお、大人/子供でenrichしていた“防御機構”と“DNAの修復”に関係するCOGは離乳前乳児ではenrichしていない。

上述した腸内細菌叢に特異的にenrichしたCOGsには、実際の機能がはっきりしないCOG (カテゴリーRとS) に属するものがかなり含まれている。さらに、既知のCOGに相同性をもたない、つまり腸内細菌叢の遺伝子だけで構成される新規な647種類のクラスターがみつかった。これらは1クラスターあたり5個以上の遺伝

子からなり、最大のクラスターは48個の遺伝子で構成されていた。これらは他環境細菌叢のメタゲノムデータ(サルガッソーや鯨骨、土壌等)にも存在せず、きわめて腸内細菌叢に特異的な遺伝子群と考えられ、今後の腸内細菌叢の機能を解明する上で重要な解析ターゲットとなる。

このほか、接合型トランスポゾンに属する遺伝子がすべてのサンプルで共通して顕著にenrichしていることを発見した。このような遺伝子セットは報告されているいくつかのヒト由来細菌ゲノム(たとえば文献3)だけにしか存在せず、腸内細菌叢内での外来遺伝子を運ぶベクター(運び屋遺伝子群)の役割をもつ機能が含まれると考えられる。つまり、高い細菌密度をもつ腸内は「遺伝子の水平伝播」に適した環境であると考えられてきたが(9)、接合型トランスポゾンの遺伝子セットの増幅は、腸内が細菌細胞間の接触を介した遺伝子伝搬や交換の「場」であることを裏打ちしたデータと言える。

5. メタデータから菌種の特定

メタデータ中に同定された遺伝子のアミノ酸配列類似度を指標としたBLASTPによる属レベルでの菌種組成の解析を試みた。そのために、GenBankの重複のないアミノ酸配列データベース(version 26 May 2007)に47種類の未発表細菌ゲノムからMetaGeneで予測した遺伝子セットを加えたインハウスの重複のないデータセットを構築した。47種類の細菌ゲノムデータには、おもにワシントン大学、サンガー研究所と著者らのグループで決定されたヒト常在菌が多く含まれる(表2)。公的データバンクに登録されている細菌遺伝子の大部分が他の環境由来であるため、これらヒト由来の常在菌ゲノム情報を加えることによって菌種同定の精度を上げる意味がある。

菌種を特定する検索では、検索する腸内細菌叢の各遺伝子のアミノ酸配列がベストヒットするデータセット中の遺伝子由来する菌種をその遺伝子の菌種と定義した。アミノ酸配列の類似度の閾値を90%以上としたときの結果を図4に示す。大人/子供の各サンプルにおいて、全遺伝子の17%~43%が35~65属/個人に特定され合計121属が特定された。一方、離乳前乳児の各サンプルでは、それぞれ35~55%の遺伝子が31~61属/個人に特定され合計84属が特定された。

大人/子供では、*Bacteroides*, *Firmicutes*門の*Eubacterium*, *Ruminococcus*, *Clostridium*, *Bifidobacterium*がおもな属となっている。離乳前乳児では、*Bifidobacterium*または大腸菌、*Raoultella*, *Klebsiella*等の*Enterobacteriaceae*が主な構成菌種とし

て特定された。これらの結果は、これまでの培養法でわかっていた菌種組成とおおむね一致する(1)。つまり、離乳前後において腸内細菌叢の組成が大きく変化し、大人及び離乳後乳児の組成は個人間で互いに共通する傾向にあるが、離乳前幼児間の組成は大きく異なる傾向にある。しかしこの解析条件においては、全遺伝子の約1/3の遺伝子しか属特定ができず、残りの2/3がどの菌種に由来するのか不明となっている。

メタデータをより精度高く菌種に特定する方法として、各ショットガンリードのDNA配列を細菌ゲノムへマッピング(150 bp以上の配列が95%以上の配列類似度をもつリードのゲノムへの張り付け)を行った。その結果、大人/子供では全リードの15%~37%が、離乳前乳児ではその42%~70%が既知ゲノムにマップされた。離乳前乳児でのマッピングされたリード数の割合が高い理由は、これらのサンプル中では大腸菌や*Bifidobacterium*等の既にシーケンスされた菌種が優先菌種となっているためである。しかし、上述したアミノ酸配列を指標とした場合と同様に、大人/子供の場合では全リードの約2/3は特定の菌種にマップされなかった。すなわち、これらの結果はヒト常在菌の大部分が未だ分離またはシーケンスされていないことを意味する。

このショットガンリードのマッピングは、同じ属の異なった種への特定を可能にする。たとえば、*Bacteroides*では現在8種類の株(種)が(発表済みと未発表を含めて)ゲノムシーケンスされており、これらゲノムに対するショットガンリードをマッピングした結果の一部を図5に示す。この結果から、個人における*Bacteroides*属の種レベルでの組成比、*Bacteroides*という1つの属のトータルの組成比、個人間での*Bacteroides*属菌種の組成比の違い等を知ることができる。また、たとえば*Bifidobacterium*属のゲノムに対して、同じマッピングを同時に行くと、*Bifidobacterium*における種レベルでの組成比とともに、*Bacteroides*と*Bifidobacterium*の異なった属間や種間の組成比を知ることができる。この解析方法は、菌種ごとに独立した培養条件を用いる従来の解析法と異なるし、16Sデータをベースにした方法(2)における配列類似度の閾値の取り方によるあいまいさ(同属または同種の区別をそれぞれ97%, 98%, 99%の配列類似度で区切ることによって生じる結果の違い)を伴わない。このマッピング解析では、70種類の細菌に対して有為にショットガンリードがマップされた。また、全リードの0.1%以上がマップされる菌種の数、大人/子供の各サンプルで18~23菌種、離乳前乳児の各サンプルで8~19菌種となっ

表2. ヒト常在菌を含む未発表ゲノムシーケンス

菌名	ゲノムサイズ ^a (Mb)	解析機関 ^b	データの状況	予測遺伝子数 ^c
<i>Bacteroides</i> WH2	7.12	WUSTL	ドラフト	5294
<i>Bacteroides caccae</i>	4.59	WUSTL	ドラフト	3631
<i>Bacteroides capillosus</i>	4.25	WUSTL	ドラフト	3936
<i>Bacteroides distasonis</i>	5.00	WUSTL	ドラフト	4149
<i>Bacteroides ovatus</i>	6.47	WUSTL	ドラフト	4832
<i>Bacteroides</i> sp. A01	6.42	HMGJ	ドラフト	5167
<i>Bacteroides</i> sp. A03	5.91	HMGJ	ドラフト	4862
<i>Bacteroides thetaiotaomicron</i> 3731	7.08	WUSTL	ドラフト	5464
<i>Bacteroides thetaiotaomicron</i> 7330	6.77	WUSTL	ドラフト	5409
<i>Bacteroides uniformis</i>	4.68	WUSTL	ドラフト	3651
<i>Bacteroides vulgatus</i>	5.42	WUSTL	ドラフト	4377
<i>Bifidobacterium adolescentis</i>	2.62	WUSTL	ドラフト	2496
<i>Bifidobacterium bifidum</i> JCM1255	2.21	HMGJ	完成	1814
<i>Bifidobacterium breve</i> JCM1192	2.28	HMGJ	ドラフト	1963
<i>Bifidobacterium catenulatum</i> JCM1194	2.16	HMGJ	ドラフト	1808
<i>Citrobacter diversus</i>	4.69	WUSTL	ドラフト	4194
<i>Clostridium clostridioforme</i> JCM1291	6.78	HMGJ	ドラフト	6075
<i>Clostridium paraputrificum</i> JCM1293	6.06	HMGJ	ドラフト	4845
<i>Clostridium ramosum</i> JCM1298	5.20	HMGJ	ドラフト	5050
<i>Clostridium symbiosum</i>	4.95	WUSTL	ドラフト	4511
<i>Collinsella aerofaciens</i>	2.47	WUSTL	ドラフト	2232
<i>Dorea longicatena</i>	2.90	WUSTL	ドラフト	2682
<i>Enterobacter sakazakii</i>	4.45	WUSTL	ドラフト	4247
<i>Escherichia coli</i> SE11	5.16	HMGJ	完成	4603
<i>Escherichia coli</i> SE15	4.84	HMGJ	完成	4270
<i>Eubacterium eligens</i>	2.82	WUSTL	ドラフト	2597
<i>Eubacterium rectale</i>	3.51	WUSTL	ドラフト	3290
<i>Eubacterium ventriosum</i>	2.85	WUSTL	ドラフト	2519
<i>Finnegoldia magna</i>	1.99	HMGJ	完成	1836
<i>Klebsiella oxytoca</i>	5.29	WUSTL	ドラフト	5381
<i>Klebsiella pneumoniae</i>	5.69	WUSTL	完成	5187
<i>Kuenenia stuttgartiensis</i>	4.22	Public	ドラフト	3258
<i>Lactobacillus fermentum</i>	2.10	HMGJ	完成	2001
<i>Lactobacillus reuteri</i>	2.04	HMGJ	完成	1980
<i>Lactobacillus rhamnosus</i> GG	3.01	HMGJ	完成	2680
<i>Methanobrevibacter smithii</i>	1.91	Public	ドラフト	1867
<i>Parabacteroides merdae</i>	4.42	WUSTL	ドラフト	3445
<i>Porphyromonas gingivalis</i>	2.35	HMGJ	完成	1781
<i>Raoultella ornithinolytica</i>	5.99	HMGJ	ドラフト	5328
<i>Rice Cluster I archaea</i>	3.18	Public	完成	3085
<i>Ruminococcus gnavus</i>	3.51	WUSTL	ドラフト	3388
<i>Ruminococcus obeum</i>	3.63	WUSTL	ドラフト	3469
<i>Ruminococcus torques</i>	2.74	WUSTL	ドラフト	2377
<i>Salmonella enterica</i> serovar <i>Arizonae</i>	4.54	WUSTL	ドラフト	4073
<i>Salmonella enterica</i> serovar <i>Diarizonae</i>	4.90	WUSTL	ドラフト	4541
<i>Salmonella enterica</i> serovar <i>Paratyphi B</i>	4.79	WUSTL	ドラフト	4426
<i>Vibrio harveyi</i>	6.04	WUSTL	ドラフト	5543

^aドラフト配列のゲノムサイズはコンティグとシングルトンの総サイズの和。 ^bWUSTL (Washington University Genome Sequencing Center), Sanger (the Sanger Institute), HMGJ (Human Metagenome Consortium Japan). ^cMetaGeneプログラムでの予測。

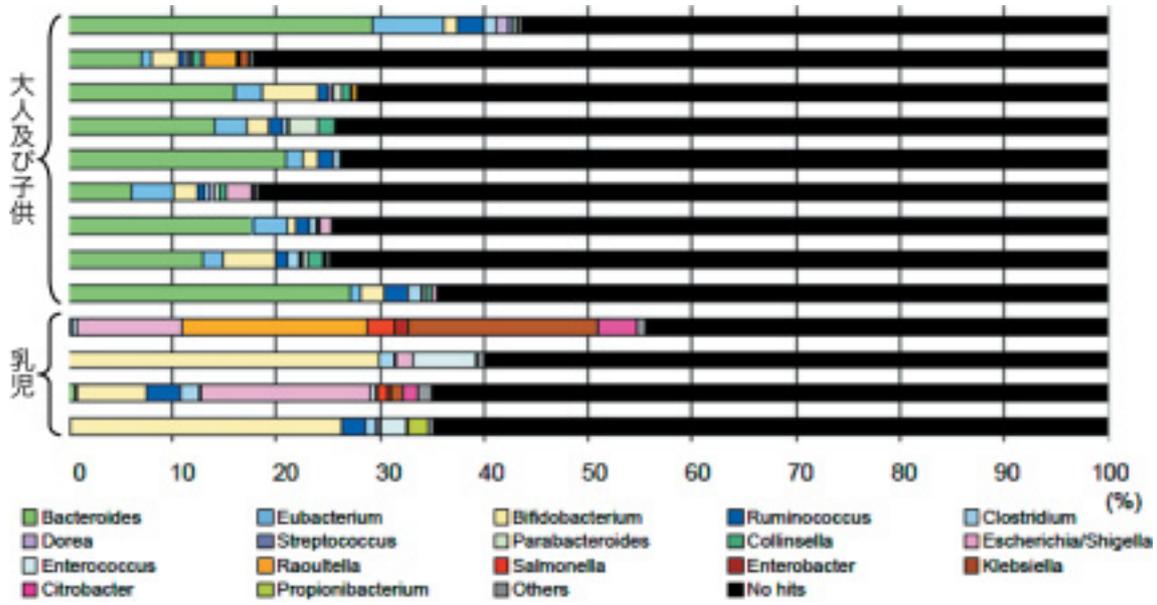


図4. 腸内細菌叢遺伝子のアミノ酸配列類似度を指標とした菌種組成の解析.

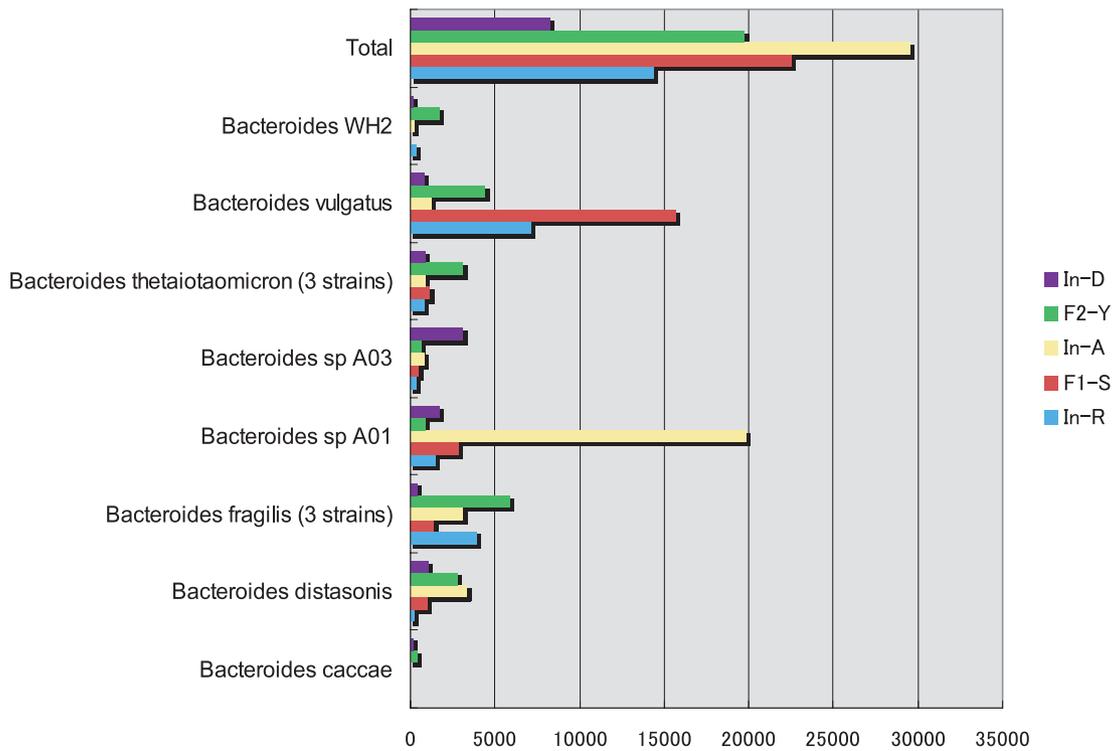


図5. メタデータ (ショットガンリード) の既知8 *Bacteroides* speciesゲノムへのマッピング. 5サンプル (右カラム) について配列類似度 $\geq 95\%$ 及び相同塩基数 ≥ 150 bpで解析した.

た。ただし、この方法はゲノム情報のない未知細菌種に対しては応用できないこと、細菌叢中に少なくとも0.1%以上存在する菌種に対してのみ有効であることなど、より正確な結果をより多くの細菌種について得るに

は難培養性細菌を含めた多くのヒト常在菌ゲノムの解読を進める必要がある。その一環として表2に一部リストしたようなヒト常在菌 (ただし培養可能な菌種) のゲノムシーケンスが現在国際的に進められている。

6. 腸内細菌叢メタゲノム解析の国際動向

ここに紹介した健康なヒト腸内細菌叢のメタゲノム解析は、質量ともこれまでになく大規模なものである。今日までに報告されているヒト腸内細菌叢を対象としたメタゲノム解析は2報ある。1報は2人の大人サンプルの解析から、腸内細菌叢では宿主ヒトに欠損するグリカン、アミノ酸、ビタミンなどの代謝に関わる遺伝子が著しくenrichしていることを示し、総合的なヒトの代謝機能が宿主のヒト自身と腸内細菌叢の間の相互扶助的関係で成り立っていることを報告している(4)。もう1報は肥満マウスのメタゲノムデータをベースにして肥満が宿主の遺伝的機能だけでなく腸内細菌叢の代謝機能と関連していることを示したものである(17)。いずれの論文も2006年のScience誌とNature誌に発表されており、ヒト常在菌の生体機能解明をめざした研究の重要性が国際的に再認識されていることを物語っている。一方、今回の著者らの解析は、遺伝子/COG組成、菌種組成、個人間の組成の違い、他環境細菌叢との比較、大人と幼児間の違いや共通性など、幅広い年齢のサンプルや大量のメタゲノムデータを用いて、健康なヒト腸内細菌叢の基盤的で広範囲な実体の解明を成し遂げていると言える。

しかしながら、最近ではフランスやイギリスを中心としたEU諸国、アメリカ、中国、シンガポールなどがヒト常在菌の研究（健康及び病態のメタゲノム解析、常在菌の個別ゲノム解析、細菌叢の機能解析等）の国レベルでの推進を公表または具体的に実施しており、本研究分野の研究競争は国際的に激化の様相を呈して来た。とくにアメリカでは、「ヒト・動物・環境中の細菌叢ゲノム解読はヒトゲノム解読と同じように重要である」という認識のもと、ヒト常在細菌叢を含めたさまざまな環境細菌叢のメタゲノム解析の重要性が唱えられている(13)。その中で、ヒト常在菌に関してはHuman Microbiome Project (HMP)として常在菌と病気の関連解明をゴールとした大規模なメタゲノム解析の開始が2007年5月に

承認された。また、これに先んじて開催されたNHGRI (NIHヒトゲノム研究所)主催のHuman Microbiome Project Workshop (2007年4月)には、著者らを含めた日米欧中のおもな研究者が参加し、HMPはヒトゲノム計画と同様に、参加国の協力体制のもと国際プロジェクトとして進められることが合意された。

このような環境細菌叢をターゲットとした国際的なゲノム研究の動向には、ここ数年のうちに著しく進歩してきたシーケンス技術が背景にある。現在、ヒト個人のゲノムを1000ドルで解析できるシーケンス技術の開発がアメリカを中心に急がれている(8, 11, 15)。この理由は、ヒトの病気と関連する遺伝子やSNPs (一塩基多型)などの発見には、個人のゲノムを端から端までシーケンスすることが必要となっているからである。また、そのためには多くの病態や健康人のゲノムをシーケンスする必要があり、そのスピードアップと低コスト化が必須の克服すべきテーマとなっている。

既に現行のキャピラリー型シーケンサー (ABI 3730xl やMegaBACE4500)とは異なった原理をベースにした次世代シーケンサーが数タイプ開発されている(表3)。これらは数十万から数千万の配列データを一度に生産できる超並列型シーケンサーであり(10)、ランあたり現行の数百から一千倍の塩基数生産能力を有する。とくに、Roche-454及びIllumina-Solexaのシーケンサーは既に主なシーケンスセンターでルーチン稼働している。これらのシーケンサーを用いたときのコストは現行の1/10から1/100であり、現行の1万分の1が目的である1000ドルにはまだほど遠いが、数Mbの1個の細菌ゲノムのシーケンスコストは既に1000ドル程度になっている。また、ここに挙げた次世代型シーケンス技術には、現行プロセスにある大腸菌へのクローニング操作を必要とせず、それによって生ずる配列データのバイアスがないことも特徴である。

表3. 現行シーケンサーと次世代シーケンサー

シーケンサー	総塩基数/ラン	塩基数/リード×リード数/ラン
<現行>		
ABI 3730xl	0.1 Mb	900 bp/リード×96リード
MegaBACE4500	0.3 Mb	800 bp/リード×384リード
<次世代>		
Roche 454 G20	30 Mb	100 bp/リード×30万リード
Roche 454 FLX	100 Mb	250 bp/リード×40万リード
Illumina-Solexa	1000 Mb	25 bp/リード×4000万リード
ABI SOLiD	1000 Mb	25 bp/リード×4000万リード

7. おわりに

メタゲノム解析によってこれまで不可能であった知見を多数得ることができ、概ねこの基本的な進め方／技術は確立でき、その有効性が証明できたと考えている。しかしながら、計算機解析のスピードアップなどの技術的改良点、得られた遺伝子や菌種組成から細菌叢のもつ機能に関するデータマイニングの手法開発や情報量の不足など、今後追求すべき基本的な課題が浮かび上がってきた。たとえば、得られた遺伝子セット (COGs) からその機能を情報学的に的確に推定する (論文情報を活用した) データベースの構築が大事になると考えている。

メタゲノムをどういった観点から情報学的に解析し、どういったフォーマットでその結果をアウトプットし、得られた結果から何を抽出しどう解釈するのかは研究レベルの高度化に重要である。そのため、充実したコンピュータシステムの装備とともに生物学と情報学の知識と技術を十分に有するバイオインフォマティクス研究者の養成または参加が必須である。上述したシーケンス技術のさらなる進歩を考えると、今後もこれまでとは桁違いのメタゲノムデータの生産が予想され、バイオインフォマティクスの高度化は将来におけるメタゲノム解析の推進に大きな鍵を握る。

ヒト腸内細菌叢に関しては、健康で幅広い年齢層サンプルからの基本的データを得たので、今後は感染症を含めたさまざまな病態、生活習慣、プロバイオティクスとの関連など、より有益な成果をめざしたテーマを進める計画である。さらに、常在菌ゲノムの個別解析を進め、メタゲノムデータのより効率的で高度な解析結果の取得に繋げたい。一方で、*in vivo*での腸内細菌叢機能の解明はきわめて初歩的な段階にある。メタゲノム及び個別の常在菌ゲノム情報を活用するために、よりシステムティックな機能解析実験系 (たとえば、無菌マウスの利用) の確立は急務であると考えられる。これらの研究の推進は国際的なHMPへの日本の貢献と重なる。

最後に、ここで述べた解析結果は著者らを含めた文部科学省特定領域研究「基盤ゲノム」、「応用ゲノム」、「比較ゲノム」の班員を中心として構成されているヒト常在菌叢ゲノム研究コンソーシアム (Human MetaGenome Consortium Japan; HMGJ) の成果の一部である。HMGJメンバーである大島健志朗 (北里大学博士課程学生)、藤英博 (理研GSC)、豊田敦 (理研GSC)、高見英人 (海洋研究開発機構)、森田英利 (麻布大学)、伊藤喜久治 (東京大学大学院農学生命科学研究科)、榊佳之 (理研GSC)、Taylor Todd (理研GSC)、Sharma Vineet

K. (理研GSC)、Srivastava Tulika P. (理研GSC)、野口英樹 (東京大学大学院新領域創成科学研究科)、高木利久 (東京大学大学院新領域創成科学研究科) の各氏は本研究の共同研究者であることを付記しておく。

文 献

- (1) Benno Y, Endo K, Mizutani T, Namba Y, Komori T, Mitsuoka T. 1989. Comparison of fecal microflora of elderly persons in rural and urban areas of Japan. *Appl Environ Microbiol* **55**: 1100-1105.
- (2) Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial flora. *Science* **308**: 1635-1638.
- (3) Garnier F, Taourit S, Glaser P, Courvalin P, Galimand M. 2000. Characterization of transposon Tn1549, conferring VanB-type resistance in *Enterococcus* spp. *Microbiology* **146**: 1481-1489.
- (4) Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. 2006. Metagenomic analysis of the human distal gut microbiome. *Science* **312**: 1355-1359.
- (5) GOLD (Genomes OnLine Database v 2.0): <http://www.genomesonline.org/>
- (6) Green S, Rubin EM. 2005. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* **6**: 805-814.
- (7) Handelsman J. 2004. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**: 669-685.
- (8) 服部正平. 2006. 第二次大規模ゲノムシーケンスの予兆. *実験医学* **24**: 3003-3010.
- (9) Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**: 837-848.
- (10) Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
- (11) Metzker ML. 2005. Emerging technologies in DNA sequencing. *Genome Res* **15**: 1767-1776.
- (12) Noguchi H, Park J, Takagi T. 2006. MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34**: 5623-5630.
- (13) Pennisi E. 2007. Massive microbial sequence project proposed. *Science* **315**: 1781.
- (14) Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson

- H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers Y-H, Falcón LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Birmingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Neelson K, Friedman R, Frazier M, Venter JC. 2007. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern tropical Pacific. *PLoS Biol* **5**: e77, 398–431.
- (15) Shendure J, Mitra RD, Church GM. 2004. Advanced sequencing technologies: Methods and goals. *Nature Rev Genet* **5**: 335–344.
- (16) Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. 2005. Comparative metagenomics of microbial communities. *Science* **308**: 554–557.
- (17) Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.
- (18) Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.
- (19) Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Neelson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers Y-H, Smith HO. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.