メタゲノム解析の現状と将来 知識データベースの開発

Current and future trends in metagenomics
Development of knowledge bases

森 宙史¹ | 山田 拓司¹ | 黒川 顕¹

MORI Hiroshi¹; YAMADA Takuji¹; KUROKAWA Ken¹

1 東京工業大学大学院生命理工学研究科 (〒226-8501 神奈川県横浜市緑区長津田町4259) E-mail: ken@bio.titech.ac.jp

1 Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology (4259 Nagatsuta-cho Midori-ku Yokohama-shi, Kanagawa 226-8501)

原稿受理 (2012-04-06)

情報管理 55(3), 167-175, doi: 10.1241/johokanri.55.167 (<http://dx.doi.org/10.1241/johokanri.55.167>)

著者抄録

地球上のあらゆる環境に存在する細菌は、巨大なバイオマスとして物質循環の主役を担っており、地球環境の恒常性の維持に欠くことができない存在となっている。しかし、従来の研究手法では、例えば、どのような種がどれくらい存在し何をしているのか、という基本的な疑問にすら答えることが困難であった。DNAを解読する技術の急速な発展を背景に、環境中の細菌由来の膨大なDNA情報をコンピューターにより解析するメタゲノム解析手法を用いることで、細菌群集と環境との関係が明らかになりつつある。本稿では、メタゲノム解析について解説するとともに、加速度的に膨れ上がるメタゲノム情報から新たな知識を発見するための知識データベースの開発、さらには融合研究領域である今後のメタゲノム研究分野の展開について述べる。

キーワード

メタゲノム, 細菌群集, データ駆動型研究, 新型シーケンサー, バイオインフォマティクス, 知識データベース, セマンティックWeb, オントロジー

1. はじめに

細菌は、海洋や河川、土壌や大気中、ヒトや動物の腸内、皮膚、口腔内、さらには大気圏、深海、地殻内や地球深部の熱水など、他の生物が生存し得ない極限環境にも存在が確認されており、その環境に特化した多様な細菌が群集を形成し棲息している。環境中の細菌は、個体間、種間の相互作用のみならず群集全体として環境と相互作用し、環境の恒常性

の維持に欠くことができない存在となっている。環境の根幹を形成する細菌群集の総体としての生命システムや、環境との相互作用を明らかにするためには、環境中の細菌群集についての詳細な理解が必須となる。しかしながら、環境中に存在する細菌のほとんどが培養困難であるため、培養技術に立脚したこれまでの細菌学的手法では、環境中の細菌に関して得られる知識が極めて偏倚かつ限定的で、群集を構成する種の組成や、生命システムとしての機能、

環境との相互作用などについては未解明なままである。近年、細菌群集を単離せずそのまま解析の対象とし、環境と生命との相互作用などを明らかにするメタゲノム解析と呼ばれる新たな解析手法が急速に発展している。

本稿では、ヒトを解析対象としたヒトメタゲノムおよび環境を解析対象とした環境メタゲノムを紹介するとともに、これら分野の歴史、現状と今後の発展に不可欠なデータベースの開発に関して紹介する。

2. メタゲノミクス

細菌群集由来の膨大な遺伝子で満たされた環境は、巨大な「遺伝子プール」と言っても過言ではない。環境中の遺伝子プールを対象とする研究分野をメタゲノミクスと言う。ここでは、メタゲノム解析の歴史的経緯およびヒトメタゲノム解析、環境メタゲノム解析に関して具体的な研究成果を含め紹介する。

2.1 メタゲノム解析とは

メタゲノム解析は、環境中の細菌群集からDNAを丸ごと抽出し、遺伝子プールを構成するゲノム配列を徹底的に解読（シーケンシング）し、群集構造を明らかにするとともに、遺伝子プールの変動や環境との相互作用を解明することを可能にした解析手法である。メタゲノム解析を実施するためには、大量のシーケンシングならびにそれら膨大な情報を処理するコンピューター、いわゆるバイオインフォマティクスと呼ばれる生命情報解析技術が必須となる。古くは1998年に「メタゲノム」の概念が提唱され¹⁾、2000年にDeLongらによって具体的な解析手法に関して提案がなされた²⁾。メタゲノム解析を一躍メジャーにしたのは、10億塩基にもおよぶシーケンスのメタゲノム解析を実施し、120万個もの新規遺伝子を発見した、J.C.Venterらによるサルガッソー海における海洋細菌群集をターゲットとしたメタゲノム解析と言えよう³⁾。本研究により、メタゲノム解析が細

菌群集の生態の理解や、細菌が持つ膨大な未知の遺伝子資源の発見に有効であることが広く知れわたり、2004年以降盛んにメタゲノム解析が行われるようになった。これまでに、鉱山排水、海水、土壌、ヒト・動物・昆虫腸内、活性汚泥、大気、極限環境、発酵食品などさまざまな環境において細菌群集のメタゲノム解析が行われ、環境中の遺伝子プールに関する基盤データが地球レベルで明らかになりつつある。

2.2 ヒトメタゲノム解析

現在、メタゲノミクスにおいて、最も研究が進んでいる研究対象がヒト常在細菌群集である。ヒトの体内や表層部位には、多様かつ膨大な数の細菌が存在していることは古くから知られている。例えば、ヒト腸内には数百種類の腸内細菌が生息し、その総数は100兆個、重さは1kgにもおよぶ。これらの細菌群集が複雑な共同体を形成しつつ、宿主であるヒトの生理や健康に対して、ビタミン合成や難分解性多糖類の代謝など、ヒトが生存する上で欠くことができない非常に大きな役割を果たしていることも広く認識されている。さらに、細菌群集の異常や変動がさまざまな疾患と密接に結びついていることを示すデータも集積してきている。したがって、ヒト常在細菌群集の解明が進めば、それら細菌群集のモニタリングによる宿主の健康状態の管理、さらには常在細菌群集のコントロールによる疾患の治療や予防など新たな方策の開発に結びつく可能性が高い。このような背景から、世界的にもヒトメタゲノミクスに対する関心は急激に高まっており、微生物学・免疫学・臨床医学などさまざまな分野での研究が急速に活発化している。

このヒトメタゲノミクスは、2004年から世界の主要研究者らによる議論が開始された。その後度重なる議論を経て、2008年には米国でHMP (Human Microbiome Project)⁴⁾、欧州ではMetaHIT (Metagenomics of the Human Intestinal Tract, ヒト腸内におけるメタゲノミクス)⁵⁾ が立ち上がり、それ

ぞれ約120億円、約30億円（いずれも5年間総額）の巨大な研究プロジェクトが開始された。

日本では、著者らが2004年からヒトメタゲノミクスに関する議論を重ね、世界に先駆けて大規模なヒトメタゲノミクスの研究成果を2007年に発表した⁶⁾。この研究では、ヒト腸内細菌群集に共通して存在する特徴を特定し、性別や年齢など環境要素に依存する特異的な構造を見出すことを目的とし、2つの家族を含む乳児から大人までさまざまな年齢、性別の13人の健常な日本人を対象として大規模な比較メタゲノム解析を行った（図1）。合計で727MbのDNA断片配列から約66万個の遺伝子を高精度に予測し、バイオインフォマティクス解析によりさまざまな新規知見を得た。乳児における腸内細菌群集は大人とは大きく異なり単純で、かつ個人差が極めて大きいが、反対に幼児を含む大人における腸内細菌叢はより複雑であるものの、年齢や性別に関わらず機能的に一樣であった。このことは、大人および幼児の腸内細菌群集は、環境のゆらぎに対して頑健であるが、乳

児の腸内細菌群集は脆弱で、小規模な環境変動によっても劇的に変化するし環境に迅速に適応する可能性のあることを示唆する。さらに、大人では237、乳児では136の特異的に強化された遺伝子ファミリーを特定することができた。これらには、食物中に含まれる多糖類の分解、DNAの修復、殺菌性物質に対抗する防御などに関わる遺伝子が含まれる。一方、離乳前乳児では、多糖類を分解する遺伝子よりも、母乳等に多く含まれる低分子の糖類やビタミン等を取り込む遺伝子群が有意に増えていた。これらの特異的に強化された遺伝子ファミリーは、腸の環境に順応するために、腸内細菌の種類によって利用されるさまざまな戦略を示していると考えられる。

このように、メタゲノム解析を実施することで、これまでほとんど解析不可能であったヒト腸内細菌群集の全体像をおおまかに理解することができ、今後検証すべき多数の新規仮説を見いだすとともに、ヒト腸内細菌における新たな研究の方向性を提示するに至った。

乳児における腸内細菌群集は単純で、かつ個人差が極めて大きいことがわかる。反対に幼児を含む大人における腸内細菌群集は、悪玉菌といわれる *Bacteroides* 属がもっとも多く、乳児より菌の分布が複雑であるものの、年齢や性別に関わらず機能的に一樣である。

図1 ヒト腸内細菌群集のメタゲノム解析（日本人13人の腸内細菌群集の構造）

2.3 環境メタゲノミクス

土壌や海洋など自然環境中の細菌群集は、ヒト常在細菌群集と比較して、細菌数も桁違いに多く、構成する種の多様性も圧倒的に高い。メタゲノミクス黎明期に、酸性鉱山排水、深海のクジラ遺骸、サルガッソー海、ミネソタ土壌などのいくつかのメタゲノム研究が発表された。しかし、当時の技術では環境中の巨大なバイオマスのごく一部しか解析することができない上に、レファレンスとなる個別菌のゲノム情報がほとんど得られていなかったことから、圧倒的な解析能力を有する新型シーケンサーが一般化するまで、ほとんど研究されていなかった。

2.4 新型シーケンサーによるメタゲノミクスの発展

新型シーケンサーは2004年に454シーケンサーが登場して以降、Roche 454 GS FLX, IlluminaGAIIx, HiSeq2000などを中心とした新型シーケンサーが相次いで開発された。シーケンシングコストが圧倒的に低下し、解析スピードが圧倒的に向上したことから、大規模なシーケンシングが要求されるメタゲノム解析においても新型シーケンサーが広く一般的に利用されるようになった。現在では、上述した大規模なヒトメタゲノム解析においても、これら新型シーケンサーが利用されており、興味深い研究が次々と発表されている。2009年HMPでは21～32歳までの一卵性双生児31ペアと二卵性双生児23ペア、さらに彼らの母親46人の合計154人について、肥満の指標であるBMI指数を測定し、遺伝的要因と腸内細菌との関係、さらには肥満との関係を解析した⁷⁾。この解析によって、一卵性双生児と二卵性双生児の間で細菌群集の類似度に違いはなく、どちらも他人と比較すると類似していること、肥満タイプのサンプルの細菌群集は種の多様性が少ないこと、154人すべてに共通した細菌は存在しないことなどが明らかになった。また2010年MetaHITでは、欧州人124人から576.7 Gbのシーケンシングによって330万個の遺伝子を発見

し、ヒト腸内の細菌群集の遺伝子セットおよび組成の大枠を示すとともに、クローン病や潰瘍性大腸炎などの患者の腸内細菌群集は、健常人とは大きく異なることを遺伝子プールレベルで詳細に示した⁸⁾。同じくMetaHITは、4か国にまたがる大規模ヒト腸内メタゲノム解析を実施し、ヒト腸内細菌群集が人種や性別に依存せず、大きく3つのタイプ (Enterotype) に分類できることを2011年に発表した⁹⁾。さらに、現在はHMPで300人、MetaHITでは800人、中国BGIでは400人におよぶ主にヒト腸内細菌群集のメタゲノム解析が進行中であり、II型糖尿病、クローン病、潰瘍性腸疾患、精神疾患などの疾病との関係性、食生活との関係性などを中心に研究が進められている。

一方、環境メタゲノミクスにおいても、新型シーケンサーは威力を発揮しており、世界中から相次いで新たな研究が発表されている。現在進行中の大規模研究としては、深海および浅海を含めた地球規模の海洋水のメタゲノム解析、アラスカの永久凍土のメタゲノム解析、Great Prairie土壌のメタゲノム解析、メキシコ湾の原油流出事故における環境評価メタゲノム解析などがある。さらに大規模なプロジェクトとして進行中なのが、2010年に日本を除く米欧中を中心とした12か国70人の研究者が立ち上げたEMP (Earth Microbiome Project)¹⁰⁾である。現時点では公的な研究予算は獲得できておらず、MO BIO Laboratories, Illumina, Eppendorfなどの民間企業が主要スポンサーとなっている。このEMPでは、地球規模での細菌の多様性、分布の地理特異性の有無およびそれを決定している環境要因などを明らかにするため、非常に多様なサンプルで大規模なメタ16S rRNA解析およびメタゲノム解析を実施する予定である。その解析対象は、海洋、深海、土壌、大気、ヒトおよび動物共生、淡水、極域、温泉など、地球上20万か所に及ぶ。現在すでに5,387サンプルの解析を実施しており、大規模統計解析により、地球環境とそこに存在する細菌群集との相関関係が明らかになりつつある。

3. 知識データベースの拡充

環境の根幹を形成する細菌群集の各個体および総体としての生命システムを明らかにするためには、そこに存在する細菌や遺伝子など個々の静的な事象を理解するのみならず、それらがどのような仕組みで連携しているか、環境とどのように連動しているかなど相互作用や動的な関係性に関する理解が必須となる。そのためには、得られた各種データを統合した新たなデータベース（DB）を開発することこそが、今後の当該分野の発展に欠くことのできない課題となっている。

3.1 メタデータの重要性と標準化

メタゲノム解析では、単一のサンプルの解析だけでは遺伝子プールの記載は可能であるものの、細菌群集の環境との相互作用や、群集総体としての生命システムに関する新規知見を見いだすことは困難である。これを克服するためには、個別菌のゲノム解析の戦略同様、複数のサンプルにおける解析結果を比較することが重要となる。特に、各細菌がその環境において果たす役割や、温度やpHなどの環境要素が細菌群集に与える影響などを明らかにするには、複数のメタゲノムデータを比較解析する、比較メタゲノム解析が必須となる。そこで必須となるのがメタデータである。メタデータとは、「データに関するデータ」であり、メタゲノム解析で得られた配列データに関する情報、すなわち、サンプリング地点の地理情報や温度、pHなどの環境についての情報、さらにはDNA抽出法やシーケンシング法、遺伝子予測法などの実験および解析方法などに関する情報のことである。

これらのメタデータは以下の2つの理由から重要である。

(1) メタゲノム解析は事前に仮説を立てて研究を行う仮説検証型研究というより、興味深い現象を発見し、新たな仮説を立てるデータ駆動型研究の性格が

強く、メタゲノムデータとメタデータ間の相関解析や多変量解析などが、生物学的な発見を行う上で非常に重要であるため。

(2) 自然界から得られたサンプルは動的であるため、厳密な意味での再現実験を行うことは不可能であり、可能な限り条件を合わせた再現実験をするためには、詳細な環境の記述が不可欠であるため。

しかし、一括りにメタデータと言っても、それが表すものは多岐にわたり、複数のメタゲノムプロジェクト間でどのような情報をどのように記述するかが統一されておらず、現在大きな問題となっている。さらに、海水や土壌、鉱山排水、ヒトや動物由来のサンプルなど、メタゲノム解析の対象となるサンプルがまったく異なるさまざまな環境由来であり、それらすべてに適合する統一した基準を作るのが難しいことが、問題解決をより困難なものにしている。最近になって、ゲノム情報の記述を標準化することを目的とした国際コンソーシアムであるGenomic Standards Consortium（GSC）が、メタゲノム解析におけるメタデータの記述の統一、およびDBへのメタデータの登録項目として最低限必要な項目リストの作成などを行っている¹¹⁾。メタゲノムデータおよびメタデータの整備、さらにはそれらを統合した知識DBの開発こそが、膨大なデータから新たな知見を見いだすために極めて重要となる。

3.2 知識DBの開発

細菌はあらゆる環境に生息しており、多種多様な研究分野や産業に関与しているため、微生物学以外のさまざまな研究分野の研究者が微生物学上の知識に容易にアクセスできるような検索システムを持った知識DBを構築することが、微生物学の裾野を広げ、本分野のさらなる発展を促すためには重要になる。そのような検索システムを持った知識DBを構築するためには、さまざまな困難を克服する必要がある。例えば以下のような課題がある。

(1) 専門用語を用いて記述された知識を、専門用語

を知らない研究者がどのように検索するか。

(2) 知りたい内容がそれほど明確でなく、したがって漠然とした検索クエリ（例: 熱いところにいる微生物、病気を起こす微生物など）に対して、どのように検索を実行し結果をどう表示するか。

これらの困難を克服する鍵となるのが、セマンティックWebの技術である。セマンティックWebでは、Web上に存在するさまざまなデータの意味をメタデータとして記述し、さらにそれらのデータ間の関係性を記述し連結することで、さまざまな分野の専門家が持つ概念の明確化・共有やデータを基にしたコンピューターによる推論検索などの、より賢いWebを実現することができると期待されている¹²⁾。HTML文書間のリンクに基づいた従来のWebは、あくまでヒトが読んでリンクをたどることを想定しており、HTML文書間のリンクが何を意味しているのかは、リンク自体には記述されていなかった。また、HTML文書中で使われている各語句の意味についても、特に記述されているわけではないため、2つのHTML文書中での語句が同じ意味であるのかどうかなどは、文脈による推論などの高度な自然言語処理が必要であった。これに対してセマンティックWebでは、RDF (Resource Description Framework) やRDFスキーマ (Resource Description Framework Schema), OWL (Web Ontology Language) 等の表現能力の異なる多数のモデル構築言語が標準で提供されており、これらを用いることで、データの意味であるメタデータの記述やデータ間の関係性の記述と連結などを実現している。具体的には、RDFによってデータ間の関係性を主語、述語、目的語という3つ組み（トリプル）を用いてモデル化するとともにデータ間で記述方法を統一し、RDFスキーマによってデータ間の関係性の基本を記述し、OWLによって各データの意味の記述やデータ間の関係性の詳細などの概念体系であるオントロジーを記述する。これらをあわせることで、専門家が持つ概念を記述し共有可能にするとともにデータの意味やデータ間の関係性を用いたコ

ンピューターによる推論検索を実現する。

3.3 微生物統合DBプロジェクト「MicrobeDB.JP」

著者らは、NBDC微生物統合DBプロジェクト「MicrobeDB.JP」^{13),14)}において、さまざまな微生物学上のデータをRDFを用いて記述し、それらのデータの意味やデータ間の関係性をRDFスキーマやOWLを用いて記述する作業を現在行っている。一例として、微生物の生息環境について以下に述べる。自分が興味を持っている微生物が環境中のどのような場所に生息しているのかを知ることは、その微生物の生態や進化を考える上で非常に重要になる。メタゲノム解析によってその環境に生息する微生物のリストを得ることが比較的容易になり、既に公共の配列DBにはメタゲノム解析由来の多様なデータが75,000サンプル以上登録されており、各サンプルのデータは配列データとそのサンプルについて記述されたメタデータから構成されている。そこには、例えばオランダ人の36歳の男性の腸内や4月のメキシコ湾の海水など、微生物の生息環境についてのメタデータが多数含まれている。したがって、ある微生物の生息環境を知りたい、というニーズに応えることができるデータは既にある程度そろっていると言える。しかしながら、上述した通り、それらのメタデータの記述方法や語彙は統一されておらず、特に語彙については多数の同義語がデータ登録者ごとにバラバラに使われているのが現状である。このため、それらのメタデータをそのままDB化しても、検索クエリに対して、多数の同義語が含まれた生息環境を表す語彙のリストが列挙される、という検索結果となり、語彙間の意味的な関係性をユーザーが推測する必要がある。さらにその中には同じ語彙でもデータ登録者によって意味する生息環境が異なる場合もあるため、そこからその微生物の生息環境を推定することは困難を極める。

MicrobeDB.JPプロジェクトにおいては、この問題を解決するため、微生物の生息環境の記述に用いら

れる語彙の定義と語彙間の意味的な関係性を記述したオントロジーであるMEO (Metagenome/Microbes Environmental Ontology) を新しく構築した^{15),16)} (図2)。

微生物は生息環境に関して以下のような特徴をもつ。

- (1) 地球上のいたる所に生息する
- (2) 他の生物の体内にも生息している
- (3) 同じ種であっても場所によってゲノムや生態が微妙に異なる場合がある

そこでこれらを記述することが可能なオントロジーを構築する必要がある。MEOは、生物の生息環境についてのオントロジーのENVO (Environmental Ontology), 地名オントロジーのGAZETTEER, 生物の系統名についてのオントロジーのNCBI Taxonomy, 動物の解剖学オントロジーのFMA (Foundational Model of Anatomy) の4つのオントロジーと、人体各部位の分類区分を記述したDBであるBodyParts3Dにおける語彙間の関係性のデータを統合し、それらに含まれる語彙間の関係性を図2のような階層構造で表現した巨大なオントロジーである。このMEOを用

いることによって、メタゲノムデータや菌株保存機関が持つ微生物の分離元のデータなどに含まれるさまざまな微生物の生息環境についてのデータを、各語彙の意味や語彙間の関係性を明確にして容易に検索することが可能となった。

4. 今後の展開

最初に述べたように、細菌は地球上のあらゆる環境において、多様な種が群集を構成し生息することで、環境における物質循環の根幹を形成している。したがって、それら細菌群集が、個別の種ではなく群集として担う機能を理解することは、その環境の生態系を理解し、生物間の相互作用や物質循環のプロセスをモデル化する上で必須となる。さまざまな環境におけるメタゲノム解析から得られた基盤データの蓄積により、環境における群集を構成する種の組成や、群集が担う生命システムとしての機能、環境との相互作用などについて理解が進んできている。また、細菌は地球上いたる所に存在することから、細菌が関与する多くの異分野との融合研究が実施さ

Metagenome/Microbes Environmental Ontology (MEO) Ver. 0.1

図2 メタゲノム環境オントロジー MEO (Metagenome/Microbes Environmental Ontology) の階層構造

れつつあり、例えば、ヒトに有害な細菌や菌類の繁殖を抑制する新たな建築技術、地球規模のメタゲノム解析による地球物質循環モデリング・温暖化予測など、多様な研究と融合し発展しつつある。これら融合領域の発展には、微生物学上の知識に容易にアクセスできるような検索システムを持った知識DBの構築が必須である。また微生物と他の生物とのデータの統合化も重要となる。例えば光合成細菌、菌根菌、エンドファイト、植物病原菌などと植物DBとのデータの統合化は次世代農業の基盤となり得るだろうし、ヒトの免疫と細菌との関係性などの観点から、ヒトメタゲノム研究における究極のメタデータとして、ヒトゲノムデータを微生物DBに統合することもパーソナルゲノムとニュートリゲノミクスの融合に大いに貢献するだろう。

今後シーケンシングおよびコンピューター技術のさらなる発展により、より大規模なメタゲノム研究

が期待できる。その結果、集団としての細菌を解析することで、これまでの単一の細菌種にターゲットを絞った細菌学から、ゲノムレベルでの細菌群集生態学が新たに展開すると思われる。環境中の細菌は、環境との平衡を常に維持しており、集団として環境の変化に迅速に対応する能力を有する。したがって、ゲノム科学の発展からゲノムデザインが可能になりつつあるように、環境中の遺伝子プールを網羅的に解析することで、環境を科学的にデザインする「環境デザイン」が可能になってくると考えられる。遺伝子プールにおける環境デザインが可能となれば、口腔内や腸内など、ヒト常在菌群をデザインすることによるヒトの健康維持・増進、より効率的かつ簡便な農業や栽培工場、畜産業、安全安心な食品開発、環境浄化など細菌に関わる多岐にわたる分野で応用が期待でき、より低エネルギーかつ持続的な社会構築に向けて大きく貢献すると考えられる。

参考文献

- 1) Handelsman, J., et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*. 1998, vol. 5, no. 10, p. 245-249.
- 2) Beja, O. et al. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology*. 2000, vol. 2, no. 5, p. 516-529.
- 3) Venter, J. C. et al. Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*. 2004, vol. 304, no. 5667, p. 66-74.
- 4) "Human Microbiome Project (HMP)". NIH Common Fund. <http://commonfund.nih.gov/hmp/>, (accessed 2012-04-25).
- 5) MetaHIT. <http://www.metahit.eu/>, (accessed 2012-04-25).
- 6) Kurokawa, K. et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Research*. 2007, vol. 14, no. 4, p. 169-181.
- 7) Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. *Nature*. 2009, vol. 457, no. 7228, p. 480-484.
- 8) Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010, vol. 464, no. 7285, p. 59-70.
- 9) Arumugam, M. et al. Enterotypes of the human gut microbiome. *Nature*. 2011, vol. 473, no. 7346, p. 174-180.

- 10) Earth Microbiome Project. <http://www.earthmicrobiome.org/>, (accessed 2012-04-26).
- 11) Genomic Standards Consortium (GSC). http://genc.org/gc_wiki/index.php/, (accessed 2012-04-25).
- 12) "W3C Semantic Web Frequently Asked Questions". W3C. <http://www.w3.org/2001/sw/SW-FAQ>, (accessed 2012-04-25).
- 13) JST. バイオサイエンスデータベースセンター (NBDC). <http://biosciencedbc.jp/>, (accessed 2012-04-25).
- 14) 微生物統合DB「Microbe DB」. <http://microbedb.jp/>, (accessed 2012-04-25).
- 15) "meo". Microbe DB.jp. <http://mdb.bio.titech.ac.jp/meo>, (accessed 2012-04-25).
- 16) "Metagenome/Microbe Environmental Ontology". BioPortal. <http://bioportal.bioontology.org/ontologies/3009>, (accessed 2012-04-25).

Author Abstract

Microbes are essential for every part of life on Earth. Numerous microbes inhabit the biosphere, many of which are uncharacterized or uncultivable. They form a complex microbial community that deeply affects against surrounding environments. Metagenome analysis provides a radically new way of examining such complex microbial community without isolation or cultivation of individual bacterial community members. In this article, we present a brief discussion about a metagenomics and the development of knowledge bases, and also discuss about the future trends in metagenomics.

Key words

metagenomics, bacterial community, data intensive science, next generation sequencer, bioinformatics, knowledge base, semantic web, ontology