

独立行政法人科学技術振興機構
バイオインフォマティクス推進センター事業

創造的な生物・情報知識融合型の研究開発
平成18年度・平成21年度採択研究開発課題
(追跡調査報告書)

研究開発課題：
メタゲノムオーソログ遺伝子統合解析システムの開発
(平成18年11月～平成21年9月)
メタゲノム統合解析システムの開発
(平成21年10月～平成23年9月)

代表研究者氏名：黒川 顕
(東京工業大学大学院生命理工学研究科 教授)

目次

本編

1. 研究開発による成果.....	1
1.1. 研究開発課題の目標及び新規性・独創性	1
1.2. 研究開発開始時の課題を取り巻く研究・技術水準及び分野における 課題の位置づけ	2
1.3. 研究開発終了時の成果概要	3
1.4. 研究開発の達成度	5
2. 研究開発による成果の活用状況や発展状況.....	7
2.1. 構築されたデータベース・ソフトウェア等の活用状況.....	7
2.1.1. データベース・ソフトウェア等の継続状況や発展状況	7
2.1.2. 第三者によるデータベース・ソフトウェア等の活用事例	7
2.1.3. データベース・ソフトウェア等へのアクセス数／ダウンロード数、visit 数 .	9
2.2. 課題終了後の研究開発成果の継続状況や発展状況.....	9
2.2.1. BIRD 終了から現在まで、BIRD で取り組んだ課題に関連した研究開発成果の 継続状況や発展状況（国内外の研究者との共同研究へ発展した等）	9
2.2.2. BIRD 終了後に発表された論文リスト.....	10
2.2.3. 研究開発成果の現在の国際的な評価・位置づけ	11
2.3. 現在の科学技術研究における研究開発成果の活用状況、発展状況のまとめ.....	12
3. 研究開発による成果の科学技術的、社会的及び経済的な効果.....	12
3.1. バイオインフォマティクス分野及びライフサイエンス分野の研究への貢献.....	12
3.1.1. バイオインフォマティクス分野の研究への貢献	12
3.1.2. ライフサイエンス分野（バイオインフォマティクス以外）や その他、科学技術分野の研究への貢献	12
3.2. 人材育成の面から参加研究者の活動状況	13
3.2.1. 研究開発に参加した研究者のキャリアアップ	13
3.3. 社会的及び経済的な効果	13
3.3.1. 研究開発成果が大学や公的研究機関、企業等で、応用に向けて 継承または発展した例	13
3.3.2. その他、研究開発成果が社会的、経済的な効果・効用につながる兆し、 可能性	14

資料編

1. 論文リスト	15
2. 主要論文の被引用回数.....	18

3.	学会招待講演・基調講演.....	20
4.	新聞発表等	22
5.	特許出願・成立	22
6.	学会賞等の受賞	22
7.	グラントの獲得実績.....	22
8.	書籍等の執筆実績	23
9.	総説の執筆実績	23
10.	参加研究者の活動状況.....	23

本調査報告書は平成 26 年 4 月に作成
表紙の代表研究者所属は平成 25 年 10 月時点

1. 研究開発による成果

1.1. 研究開発課題の目標及び新規性・独創性

自然環境中の様々な細菌叢を、培養することなくまるごとゲノム解析する「メタゲノム解析」が現実化し、環境の根幹を形成する細菌叢の生命システムを明らかにすることが可能となってきた。メタゲノム解析によりもたらされる大量の情報から有益な知識を効率よく発見するためには、データベースの整備ならびにデータベース整備に伴うバイオインフォマティクス技術の開発が急務であった。さらに、単一のメタゲノムデータだけでなく、複数のメタゲノムデータを比較することではじめて、環境の違いによる生命システムの相違を見出すことが可能になることから、メタゲノムデータの比較を容易におこなうための、メタゲノムデータベースの視覚化技術開発が必須となる。メタゲノム遺伝子を比較し、環境の違いによる生命システムの相違を効率よく発見するため、以下の4つの新規バイオインフォマティクス技術、①メタゲノム配列からの遺伝子予測、②メタゲノム遺伝子クラスタリング、③メタゲノムオーソログ遺伝子データベースの構築、④比較メタゲノムにおけるオーソログクラスターの視覚化、を開発・整備し、それらを導入したメタゲノム統合解析システムの開発を目標とした。

また、細菌群集を構成する種や遺伝子のリストを記述することに主眼をおいたメタゲノム研究は、新型シーケンサーの普及によりゲノム科学分野および環境微生物学分野などにおいて一般化してきており、研究代表者らが開発してきたシステムの集大成とも言えるメタゲノム解析ワークフローの整備が、国内外のメタゲノム研究にとって研究遂行上必須なものとなった。メタゲノム研究を実施する際、細菌叢を取り巻く環境因子の各要素を記述したメタデータは欠く事のできない重要な情報となる。多岐に渡る高精度なメタデータとメタゲノムデータを統合し、環境と遺伝子プールとの関係性を解析することが今後の重要な課題となった。

継続研究開発（平成21年10月～平成23年9月）では、メタゲノム解析のワークフロー整備を主な目的として、上記の4つのバイオインフォマティクス技術の開発に加えて、⑤メタゲノムデータベースシステム整備およびワークフロー作成、⑥免疫研究に資するヒトメタゲノム情報解析技術の開発、⑦メタデータと遺伝子データの統合解析技術の開発、⑧スパコンによるメタゲノム解析、の4項目を追加してシステム開発することを目指した。

本研究開発においては、メタゲノム解析の中でも特にヒトメタゲノム解析に着目し、大量のメタゲノムデータから有用な知識を効率よく発見するためのインフラ構築を目的に、ヒトメタゲノムデータベースHuman Meta BodyMapの整備、メタ16S rRNA解析用ソフトVITCOMIC (Visualization tool for taxonomic Compositions of microbial community) の開発、ヒトメタゲノム検索システムBody-BLASTの開発を行った。具体的な内容は以下のとおりである。

- 1) メタゲノム解析で得られるのは、これまでのゲノム解析とは異なり、シーケンサー

から産出される短い断片配列である。しかし、従来のソフトウェアのほとんどはゲノム配列のような比較的長い配列に対するものであり、700bp程度の断片配列から正確に遺伝子領域を予測することは困難であった。そこで本研究では、短い塩基配列断片から正確に遺伝子を予測可能なソフトウェアを世界に先駆けて開発することとした。

2) 系統マッピング法の開発においては、既存の系統樹上にマッピングするという新たな方法論を提案した結果、比較メタ16S rRNA解析の際、多様性の差異を直感的に見出すことを可能とし、さらに分類群情報を付加したマッピング方法を開発したことにより、環境と群集構成のダイナミクス、さらには新規の代謝パスウェイの示唆など、これまでのゲノム研究にはない新しい可能性の提案を目指した。

3) 今後のメタゲノム研究においては新型シーケンサーによる解析が主流になると予想されるが、新型シーケンサーでは産出されるリード数は膨大であるものの、個々のリード長が短く、メタゲノム解析には不適であると考えられていた。本研究ではサンガー法と同じ情報が得られる解析手法を開発することとした。

4) メタゲノムデータを、メタデータを中心とした表現型データと捉え、配列相同性を介してメタデータ検索するという、これまでに無い新たな概念に基づいて、ヒットした配列のメタデータを結果として提示するシステムの開発を目指した。

1.2. 研究開発開始時の課題を取り巻く研究・技術水準及び分野における課題の位置づけ

本研究開発に着手しはじめた頃、大量の配列を一度にシーケンス可能な新型シーケンサーが登場し、あらゆる環境におけるメタゲノム解析が本格化した。環境中の細菌は、数百から数万種の集団（細菌叢）を形成し、細菌間相互作用だけでなく、環境と密接に関連しながら複雑なシステムを構成している。これら細菌叢をまるごとゲノム解析する「メタゲノム解析」により、環境の根幹を形成する細菌叢の生命システムを明らかにすることが可能となってきた。メタゲノム解析では極めて大量の配列データが公開されるが、これら大量のメタゲノム遺伝子を比較し、環境の違いによる生命システムの相違を見出すためには、まず未知の遺伝子を含むこれら大量の遺伝子を配列相同性にてクラスタリングし、かつその冗長度を見極めることが重要であった。

本研究では、メタゲノム配列から遺伝子配列を予測、クラスタリングし、それらを既存のオーソログ遺伝子データベースと組み合わせ機能別にデータベース化することを目標とした。将来的には、環境中の様々な要素をインデックス化し、遺伝子クラスターと組み合わせ解析することで、生態系の根幹を形成する細菌叢と環境との因果関係に焦点を絞った解析、「エンバイロメントーム解析」が実現できると考えられ、本研究はその解析基盤を形成する重要な研究になると期待された。

またメタゲノムのデータベースならびに解析システムは、主に欧米で開発、運用されていた。しかし、基本的には環境中のメタゲノムデータを対象としており、地理情報システ

ムとの融合などを積極的に推進していた。一方、本研究で開発しているシステムは、ヒトメタゲノムにターゲットを絞り、**Human Meta BodyMap**など、今後爆発的に産出されるヒトメタゲノムデータの解析に特化した。また、配列相同性検索において、検索した配列を結果として提示するのではなく、得られた配列のメタデータを提示する**Body-BLAST**は、メタゲノムデータを遺伝型データとして捉えるのではなく、メタデータを中心とした表現型データと捉えるという、これまでに無い新たな概念に基づいており、今後のメタゲノム研究の方向性にインパクトを与える点で意義深い研究であった。

1.3. 研究開発終了時の成果概要

1) メタゲノム遺伝子統合データベースの開発

メタゲノム解析で得られる塩基配列断片から高精度に遺伝子を予測するために、プロファージ遺伝子群の確率モデルを作成し、RBS (Ribosomal Binding Site) モデルを取り入れることで、外来遺伝子も含め、より短い配列断片から遺伝子領域を正確に予測可能な**MGA** (MetaGeneAnnotator) を開発した。MGAにより予測した遺伝子群から、細菌叢全体としての全代謝経路ならびに細菌種間の代謝相互作用を解明するために、メタゲノム遺伝子の分類群情報を付加した上で代謝マップ上にマッピングするシステムを開発した。メタゲノム遺伝子の代謝マップ上へのマッピングは、データベースに対する配列相同性検索により実施可能であるが、遺伝子由来の分類群を同定するために、(1)相同性検索結果のトップヒット、(2)相同性検索結果パターンのクラスタリング、(3)配列上の特徴から分類予測のため**MGC** (MetaGeneCluster) を開発し、3種の手法を導入することで代謝マップ上に分類群情報を付加した遺伝子マッピングを実現した。

2) ヒトメタゲノムBodyMapシステムの構築

メタゲノム由来の配列断片を既知のゲノム配列上に高速にマッピング可能なソフトウェアを別途開発し、マッピング結果を可視化するソフトウェアを開発した。本機能を用いることで、細菌叢に含まれる当該種のゲノム多様性を論じることが可能となった。メタゲノム解析に必須となる解析法さらには本研究で開発した上述の技術をすべて結集し、それらを統合したメタゲノム解析パイプラインを構築した。さらに、現在入手可能なメタゲノムデータ、特にヒトメタゲノムプロジェクトを中心に、腸内、口腔内、皮膚や膣内などヒトの各部位におけるメタゲノムデータの配列を対象にMGAによる遺伝子予測をおこない、予測遺伝子群データベースを構築するとともに、メタゲノムコンティグビューア、Taxonomy 逆引きビューア、クラスタリング解析結果ビューア、種分類情報可視化ビューアによる大量データの可視化など、すべての解析結果を可視化した上で、部位ごとの比較解析などを可能とするヒトメタゲノムBodyMapシステムを構築しウェブを通して公開した。

3) ヒト腸内細菌叢のメタゲノム解析

ヒトの腸内に常在する細菌叢は、ヒトの健康と病気に大きく関係していることが知られていた。ビフィズス菌や大腸菌などは腸内細菌叢を構成する代表的な細菌であるが、その構成細菌種は1,000菌種にのぼっており、その全体像やこれらが担う生理的役割や働きはほとんど知られていなかった。その理由の一つは大部分の細菌が個別に培養できない菌種であるところによると考えられていた。本研究では、細菌の培養を必要としないメタゲノム解析を駆使することによって、健康な日本人（離乳前乳児4名、離乳後幼児2名、成人7名の年齢が3ヶ月から40歳代の男女計13名、うち3及び4名からなる2家族を含む）の腸内細菌叢のゲノムDNAを抽出し、そのメタゲノム解析を行った。この解析から、腸内細菌叢の菌種組成、ヒト腸内細菌叢全体がもつ特異的な機能、大人と乳児など年齢による違い、家族間、個人間での差などについての新規知見を得ることができた。

4) メタ16SrRNA系統解析技術の開発

細菌叢のメタゲノム解析において、群集にどのような種が存在するかを明らかにすることが重要な課題の一つとなっていた。群集構造を種レベルで明らかにするために、遺伝子配列から種を分類する際に良いマーカーとなる16S rRNA遺伝子の特異的にシーケンスし系統解析を実施した。本研究では、より容易にサンプル間の相違を表現することを可能とする新たな解析手法および可視化技術としてVITCOMICを開発した。Illuminaなどに代表される新型シーケンサーは、得られるデータ量はこれまでと比較にならないほど膨大であるものの、個々の配列は100塩基程度と非常に短く、このような短い配列から生物学的な情報を正確に得るために、新たな解析法を開発した。

また、メタゲノムサンプル間で遺伝子プール内の遺伝子量を比較する際、シーケンスに用いたゲノム量ならびにシーケンス量による偏りを補正する必要があった。Raesら(2007)によって特定された35個のUniversal single copy genesに、系統推定などで広く利用されておりほとんどの細菌のゲノム中に1コピーしか存在しないgyrBを加えた36個のUniversal single copy genesについて、あるサンプルにおけるそれらの遺伝子にアサインされた平均的なリード数を、そのサンプルにおいてシーケンスされたゲノムの数と仮定した。これらの遺伝子のKEGGのアミノ酸配列データベース中における全リストは、各遺伝子のオーソログ情報をKEGG Orthology Database (<http://www.genome.jp/kegg/ko.html>) から取得することによって作成した。各EC番号にアサインされたリード数を、このUniversal single copy genesにアサインされた平均的なリード数で割ることによって、サンプル間のシーケンスされたリード数の違いや平均的なゲノムサイズの違いを補正する手法を開発した。

5) メタゲノムデータベースHuman Meta BodyMapおよびBody-BLASTの開発

新型シーケンサーによる大量16S rRNAシーケンスに対応するために、特にヒトメタゲノム研究に特化する形で、ヒトメタ16S rRNA BodyMapを開発した。本システムは、研

究開発終了時には、当時入手可能なヒト由来バクテリア16S rRNAデータ53,460本を収集し、大腸菌16S rRNA遺伝子に対応する位置、さらに分類情報を相同性にに基づき注意深く再アノテーションし、これらすべてのメタデータを格納したデータベースとなった。また一般的に利用されている比較的精度の高い16S rRNAデータベースSILVA (304,036本) も格納している。本データベースを16S 配列のみならずメタゲノムデータに拡張し、研究開発終了時現在入手可能な69,876サンプルにおよぶヒトメタゲノムデータを収録した、**Human Meta BodyMap**として整備した。

メタゲノム研究において最も価値がある情報は、それら配列が発見された細菌叢を取り巻く環境情報、すなわちメタデータである。したがって、メタゲノム配列データに対する相同性検索は、クエリ配列と相同な配列がメタゲノムデータ中に存在するか？という従来と同様の概念に基づくのではなく、クエリ配列がどのような環境に存在するのか？という配列相同性を介したメタデータ検索こそが重要となる。そこで、上述したメタゲノムデータベースを対象とした相同性検索システム**Human Body-BLAST**を開発した。さらに**Body-BLAST**の結果として得られたヒットのメタ情報を要約・可視化してユーザに提供する機能**BodyParts3D**を開発した。

6) スパコンによるリアルタイムメタゲノム解析システムの開発

本研究では、サンガー型シークエンサーならびに新型シークエンサーによるメタゲノム解析手法の開発ならびに解析パイプラインの整備を実施してきており、主にヒトメタゲノムを対象として、本研究で開発した解析手法を適用して実施した。しかし、シークエンス技術の圧倒的な発展により、メタゲノム解析では膨大な配列データが得られるようになり、通常の解析システムでは、計算コストが高く開発した解析プロトコルが適用困難となってきた。そこで、すでに開発してきた解析パイプラインを、東工大スパコンTSUBAME2.0に実装し、メタゲノム解析における大量相同性検索をリアルタイムで解析可能なシステムを開発した。さらに、GPGPU (General-Purpose computing on Graphics Processing Units) による相同性検索ソフトウェアを併せて開発し、より低電力かつ高速な開発を可能とするシステムを開発し、世界に先駆けて「リアルタイムメタゲノム解析」を実現した。

1.4. 研究開発の達成度

メタゲノム解析のための遺伝子発見ソフトウェアや可視化ツールなど、当初の研究開発計画で目標としていたデータベースや解析システムは全て開発された。また、ヒトゲノム**BodyMap**や代謝マップなど、新規の可視化への試みがなされた。

当初計画では想定していなかった新型シークエンサーの登場により、大量の配列データが産出されるようになり、データ量の問題からヒトメタゲノムに絞った解析となった。このような大量のデータに対処すべく、計算機システムならびにアルゴリズムの改良が必要

となったが、本研究開発はこのような環境変化に対応する方向に進んでおり、我が国を代表するメタゲノム情報解析システムを作りつつあるだけではなく、実験研究へのフィードバックも意識された。

当初の研究開発計画では、①メタゲノム配列からの遺伝子予測、②メタゲノム遺伝子クラスタリング、③メタゲノムオーソログ遺伝子データベースの構築、④比較メタゲノムにおけるオーソログクラスターの視覚化、の4項目を中心にシステム開発することを目標としていた。上記すべては、本研究で開発したデータベースおよび解析システムに搭載されており、当初の目標を達成することができた。しかしながら、ヒトメタゲノムのみならず海洋、土壌などメタゲノム研究が世界的に進展し、扱うデータが爆発的に増大したため、これらすべてのデータを扱うことが困難となった。さらに新型シーケンサーの一般化により今後さらにメタゲノムデータが増大することが容易に想像できるため、本研究では、対象とするメタゲノムデータをヒトメタゲノムに特化させ、高度な解析システムを維持しつつ特異性を確保することができた。

継続研究開発においては、⑤メタゲノムデータベースシステム整備およびワークフロー作成、⑥免疫研究に資するヒトメタゲノム情報解析技術の開発、⑦メタデータと遺伝子データの統合解析技術の開発、⑧スパコンによるメタゲノム解析、の4項目を中心にシステム開発することを目標としており、このうち⑤と⑧については、本研究開発で開発したデータベースおよび解析システムに搭載されており、当初の目標を達成することができた。⑥に関しては、当初メタゲノムからのペプチド予測法を開発し、研究協力者による実験的証明を経て、有用なペプチドの同定&データベース化を試みるというものであったが、研究協力者である徳島大学・桑原知巳准教授（現香川大学教授）および宮崎大学・林哲也教授から、予測したペプチド配列の表現型を実験的に証明するのはコスト面で困難であるとの助言を得て、当初の計画を断念せざるを得なかった。しかしながら、ヒトの免疫システムの発達に密接に関与すると考えられている腸内細菌SFB（Segmented Filamentous Bacteria）のゲノムを解読し、本システムを利用したゲノムとメタゲノムの関連性解析結果を含む本菌のゲノム論文を発表するに至った。⑦に関しては、メタゲノムデータにおけるメタデータのXML化ならびに詳細なメタデータ検索システムを開発したことで、現在入手可能なすべてのメタゲノムデータに関して、メタゲノム配列とメタデータとの相関解析を可能にした。

Illumina に代表される新型シーケンサーが登場し、メタゲノム研究にも新型シーケンサーが急速な勢いで浸透し、その膨大な配列データが公開されている。本研究が採択された当初は、これほどの勢いでデータ爆発が起きるとは想像しておらず、公開されたデータは速やかに解析プロトコルに投入し解析する予定であった。しかし、EU の MetaHIT（Metagenomics of Human Intestinal Tract : <http://www.metahit.eu/>）、米国の HMP（Human Microbiome Project : <http://commonfund.nih.gov/hmp/index>）から膨大なデータが産出されるに至り、解析の高速化および本研究の方向性の変更が避けられない状況と

なった。というのも、メタゲノム配列から予測した遺伝子をクラスタリングしデータベース検索する必要があったが、本研究で導入した 144 コアクラスタにて約 6 ヶ月間連続計算しても解析が終了しないという事態に陥った。この問題は、多数のメタゲノムデータを横断的に解析するグループのみの問題ではなく、今後のメタゲノム研究の発展において憂慮すべき問題に発展するであろうと考えられた。そこで、本継続課題の当初からスパコンおよび GPGPU を活用した超高速解析パイプラインの可能性を模索された。また、2011 年 3 月の東日本大震災による原発事故の影響による電力不足から、メタゲノム解析にも高速かつ低電力な解析が要求されるようになり、スパコンを利用して超並列計算をする際も、可能な限り低電力で実施する必要が生じた。上述の通り、早い段階でメタゲノム解析にスパコンおよび GPGPU を積極的に導入し、より高速にかつより低電力で解析を可能とするプロトコルの開発を実施していたこともあり、世界に先駆けて「リアルタイムメタゲノム解析」を実現することができた。

2. 研究開発による成果の活用状況や発展状況

2.1. 構築されたデータベース・ソフトウェア等の活用状況

2.1.1. データベース・ソフトウェア等の継続状況や発展状況

	ツール名	URL	公開状況	備考
(1)	Human Meta BodyMap (VITCOMIC、Body-BLAST 含む)	http://microbedb.jp/MDB/	無料公開中	JST バイオサイエンスデータベースセンター (NBDC) 統合化推進プログラムにて開発中の「MicrobeDB.jp」に継承され、改良・発展している
(2)	Human Meta BodyMap	http://metagenomics.jp/mg/	無料公開中	

2.1.2. 第三者によるデータベース・ソフトウェア等の活用事例

本研究開発に関連する主要な論文について、それを引用している具体的な活用事例を示す。

- 1) Kurokawa, K. et al., Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res., 14:169-181,(2007).

本論文の結果は、肥満の原因としての腸内細菌に関する文献 (Cox1, L.W. and Blaser,

M.J., *Cell Metabolism* 17, 2013) に引用されており、哺乳動物の消化管の微生物叢の分類が、人それぞれで非常に多様であり、かつ変動するが、一方では高度な機能を共有している理由を説明するデータとして利用されている。

もう一つの論文 (PLoS Comput Biol 5(4): e1000352.) も、腸内細菌叢の変化と肥満の関係を検討しているが、そこでは本論文の結果、13人の腸内細菌叢のオルソログスグループのクラスター (COG) プロファイル、及び本論文で用いられた戦略が利用されている。

2) Noguchi, H., Park, J., and Takagi, T. MetaGene: prokaryotic gene finding from

environmental genome shotgun sequences. *Nucleic Acids Res.*, 34:5623–5630, 2006

本論文で報告されている「MetaGene」はメタゲノム断片配列から遺伝子を予測するソフトウェアであり、論文も102回 (国内:6、海外:96、2012年まで) 引用されている。例えば、Sierra-García, I.N. ら (PLoS ONE 9(2): e90087) は、細菌による炭化水素の生分解経路の検討においてメタゲノム解析をおこない、そこで解析プログラムである「MetaGene」を使用している。その他、Cecchini, D.A.ら (PLoS ONE 8(9): e72766)、Ferrer, M.ら (Environmental Microbiology, 15(1), 211(2013)) は機能性食品などを目的とした腸内細菌のメタゲノム解析を行っており、本解析ソフトウェア「MetaGene」利用した旨が記載されている。これを引用する文献の多くは、「MetaGene」を使って解析したものであり、多くは実験項で引用されている。

3) Arumugam, M. et al., Enterotypes of the human gut microbiome. *Nature*, 473:174-180 (2011)

BIRD にて開発したヒトメタゲノム DB を徹底的に利用することで、ヒト腸内細菌叢は大きく 3 つのタイプ「Enterotype」に分類することができるという新たな概念を、ドイツ EMBL との共同研究で提唱することができた。本論文は 900 報近く (2013 年 12 月現在) 引用されており、極めてハイインパクトな報告となっている。具体的には、Ghosh T.S.ら (PLoS ONE 8(12): e83823) は、8 か国のヒト腸内細菌について抗生物質耐性遺伝子の可能性をプロファイリングにより耐性遺伝子を予測している。この報告では Arumugam によって解析されたデータをダウンロードして利用している。また、Tong M ら (PLoS ONE 8(11): e80702(2013)) は、クローン病、潰瘍性大腸炎、炎症性腸疾患と糞便中細菌の関連性の検討についてメタゲノム解析を行っているが、過去のメタゲノム解析の例として本論文を引用している。同様に Gary, D.ら (Anaerobe 24 (2013) 117e120) は疾患、食事、腸内細菌叢の関連研究としてメタゲノム解析を行っており、その結果について Arumugam らの論文を引用し比較している。また、Fredrik H ら (NATURE communications, 3, 1245(2012)) はアテローム性動脈硬化症に関連する動脈壁における脂質蓄積および炎症とその原因物質としての腸内細菌叢の関連を、メタゲノム解析によって検討しており、その中で本開発研究をそれまでの最大の腸内細菌のエンテロタイプ研究と位置づけ引用し、結果

を比較している。

- 4) Mori, H., Maruyama, F., and Kurokawa, K., VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences., BMC Bioinfo., 11:332 (2010)

上記のメタ 16S 解析ソフトウェア VITCOMIC に関する論文は、出版社ウェブページ上で highly accessed paper に評価されるとともに、研究開発終了時までに国内外からすでに 1,000 件近い解析実績を挙げている。実験科学者がメタゲノム由来の膨大な配列情報を、自らのデスクトップ PC にて解析できる本ソフトウェアは、メタゲノム解析のハードルを下げ、新たな研究分野として確立することに貢献できたと言える。例えば本プログラムの活用として、Sundarakrishnan, B. ら (Indian J Microbiol.52(4):544(2012)) は、細菌集団を Web ツールである VITCOMIC を用いて BLASTN 分析を行い、細菌叢を分類している。

2.1.3. データベース・ソフトウェア等へのアクセス数／ダウンロード数、visit 数

	ツール名	アクセス数、ダウンロード数、visit 数	集計期間	備考
(1)	メタ 16S 解析ソフトウェア VITCOMIC	解析実績：1,000 件近い		事後評価書作成時
(2)	Human Meta BodyMap	アクセス数：15,180	2012 年度 公開日：2012.3.25	NBDC 統合化推進プログラムの H24 年度実施報告書
(3)	MicrobeDB.jp	アクセス数：15,171	2012 年度 公開日：2011.12.12	NBDC 統合化推進プログラムの H24 年度実施報告書

2.2. 課題終了後の研究開発成果の継続状況や発展状況

2.2.1. BIRD 終了から現在まで、BIRD で取り組んだ課題に関連した研究開発成果の継続状況や発展状況（国内外の研究者との共同研究へ発展した等）

BIRD 研究開発期間終了後、H23 年度からは JST NBDC 統合化推進プログラムにおいて、「ゲノム・メタゲノム情報を基盤とした微生物 DB の統合化」として研究開発を実施している。本研究開発では、現在世界中に散在している微生物に関する膨大なデータをセマンティックウェブ技術により統合し、データ駆動型研究に資するデータベースシステム「MicrobeDB.jp」を開発している (<http://microbedb.jp/>)。BIRD にて開発してきたすべての技術、データを本 DB システムに継承するとともに、ヒトのみならず自然環境のメタゲ

ノムデータに対してもこれらの技術を応用することで、微生物データを genome、taxonomy、environment の 3 つの軸に沿って俯瞰し、横断的に知識を取得することが可能な DB システムとなっている。

また、BIRD にて開発したヒトメタゲノム DB を徹底的に利用することで、ヒト腸内細菌叢は大きく 3 つのタイプ「Enterotype」に分類することができるという新たな概念を、ドイツ EMBL との共同研究で提唱することができた。2011 年発表の論文はすでに 900 報近く引用されており（2013 年 12 月現在）、極めてハイインパクトな報告となっている。この概念はヒト環境のみならず、あらゆる環境の微生物群集にも応用可能と考えられており、今後の発展性が期待できる。

2.2.2. BIRD 終了後に発表された論文リスト

Nonaka, L., Maruyama, F., Miyamoto, M., Miyakoshi, M., Kurokawa, K., Masuda, M. (2012) Novel Conjugative Transferable Multiple Drug Resistance Plasmid pAQU1 from *Photobacterium damsela* subsp. *damsela* Isolated from Marine Aquaculture Environment. *Microbes Environ.*, Mar 23

Suzuki, S., Ishida, T., Kurokawa, K., Akiyama, Y. (2012) GHOSTM: A GPU-accelerated homology search tool for metagenomics. *PLoS one*, 7:e36060

Tuama, EÓ., Deck, J., Dröge, G., Döring, M., Field, D., Kottmann, R., Ma, J., Mori, H., Morrison, N., Sterk, P., Sugawara, H., Wieczorek, J., Wu, L., Yilmaz, P. (2012) Meeting Report: Hackathon-Workshop on Darwin Core and MIxS Standards Alignment (February 2012). *Stand Genomic Sci.*, 7:166-70.

Minegishi, K., Aikawa, C., Furukawa, A., Watanabe, T., Nakano, T., Ogura, Y., Ohtsubo, Y., Kurokawa, K., Hayashi, T., Maruyama, F., Nakagawa, I., Eishi, Y. (2013) Complete Genome Sequence of a *Propionibacterium acnes* Isolate from a Sarcoidosis Patient. *Genome Announc.*, 1:e00016-12.

Masuda, S., Hori, K., Maruyama, F., Ren, S., Sugimoto, S., Yamamoto, N., Mori, H., Yamada, T., Sato, S., Tabata, S., Ohta, H., Kurokawa, K. (2013) Whole-genome sequence of the purple photosynthetic bacterium *Rhodovulum sulfidophilum* strain W4. *Genome Announc.*, 1:e00577-13.

Aikawa, C., Furukawa, N., Watanabe, T., Minegishi, K., Furukawa, A., Eishi, Y., Oshima, K., Kurokawa, K., Hattori, M., Nakano, K., Maruyama, F., Nakagawa, I., Ooshima, T. Complete genome sequence of the serotype k *Streptococcus mutans* strain LJ23 (2012) *Journal of Bacteriology*, 194 (10), pp. 2754-2755.

Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, Nagata Y, Fujiyama A, Tsuda M, Kurokawa K (2013) Design and experimental application of a novel

non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. DNA Res., dst052.

黒川顕 (2011) 微生物統合データベース”MicrobeDB.JP”, 細胞工学, vol.30, No.12.

森宙史, 山田拓司, 黒川顕 (2012) メタゲノム解析の現状と将来, 情報管理, 55(3):167-175.

森宙史, 山田拓司, 黒川顕 (2013) ヒトメタゲノムの大規模コホート研究, 生物の科学「遺伝」, 67(3):279-282.

2.2.3. 研究開発成果の現在の国際的な評価・位置づけ

メタゲノムのデータベースならびに解析システムは、主に米国で開発、運用されている。代表的なものとして、UCSD（カリフォルニア大学サンディエゴ校）のCAMERA（Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis）システム（<http://camera.calit2.net/>）とJGI（Joint Genome Institute, 米国エネルギー省）のIMG/M（Integrated Microbial Genomes with Microbiome Samples : <https://img.jgi.doe.gov/cgi-bin/m/main.cgi>）がある。しかし、これらはいずれも環境中のメタゲノムデータを対象としており、地理情報システムとの融合などを積極的に推進している。一方、本研究開発で開発しているシステムは、ヒトメタゲノムにターゲットを絞り、今後爆発的に産出されるヒトメタゲノムデータの解析に適したものである。また、KEGG マッピング法などの技術も他に類を見ないものであることから、今後のヒトメタゲノム計画に大きく貢献できる点で意義深い研究であると考ええる。

メタ16S rRNA解析においては、コロラド大学で開発されたQIIME（Quantitative Insights Into Microbial Ecology : <http://qiime.org/>）がよく用いられている。一方、本研究開発で開発しているシステムHuman Meta BodyMapなどは、今後爆発的に産出されるヒトメタゲノムデータの解析に特化したものになっている。また、配列相同性検索において、検索した配列を結果として提示するのではなく、得られた配列のメタデータを提示するBody-BLASTは、メタゲノムデータを遺伝型データとして捉えるのではなく、メタデータを中心とした表現型データと捉えるという、これまでに無い新たなコンセプトに基づいており、今後のメタゲノム研究の方向性にインパクトを与える点で意義深い研究であると考ええる。

また、BIRD にて開発した DB システムは、MicrobeDB.jp へと発展的に継承されている。この MicrobeDB.jp は、セマンティックウェブ技術に基づく他に類を見ない新たな DB システムであり、多様なデータに基づく推論検索により新たな知識を引き出すことが可能なシステムとなっている。本 DB システムは 2012 年から公開しているものの、論文による公表はしていない。それにも関わらず、国外の生物多様性研究グループからの共同研究の申し出や、国外研究者の論文に引用されるなど、注目されている状況にある。今後開発を加速させ、論文公表することでより高い評価を受けるものと考えられる。

2.3. 現在の科学技術研究における研究開発成果の活用状況、発展状況のまとめ

BIRD研究開発期間終了後、H23年度からはJST バイオサイエンスデータベースセンター (NBDC) 統合化推進プログラムにて開発中の「MicrobeDB.jp」に継承され、改良・発展している (<http://microbedb.jp>)。MicrobeDB.jpでは、BIRDにて開発してきたすべての技術、データを継承し、ヒトのみならず自然環境のメタゲノムデータに対してもこれらの技術を応用する。

ヒトメタゲノムDBを利用し、ドイツEMBLとの共同研究で提唱したヒト腸内細菌叢の分類に関する概念はあらゆる環境の微生物群集にも応用可能と考えられ、今後の発展が期待できる。

3. 研究開発による成果の科学技術的、社会的及び経済的な効果

3.1. バイオインフォマティクス分野及びライフサイエンス分野の研究への貢献

3.1.1. バイオインフォマティクス分野の研究への貢献

新型シーケンサーによるメタゲノムデータの情報量は、ゲノムデータやRNA-seqデータよりも圧倒的に大きい。したがって、メタゲノム解析においては、一般的な解析ツールを応用するだけではスムーズな研究が困難である。これを克服するためには、高度なバイオインフォマティクス技術が要求され、本研究開発課題によって、GPGPU (General-purpose computing on graphics processing units) を利用した超高速相同性検索ツールの開発や、スパコンによるメタゲノム解析手法の開発、SINET (Science Information NETwork : <http://www.sinet.ad.jp/>) を利用した超高速データ転送技術開発など、新たな技術に基づく方向性を見出すことができた。

また、数理生態学分野では、これまで数理モデルに基づくシミュレーションによる研究が一般的であったが、二次解析結果を付加したメタゲノムデータを利用することで、数理モデルの証明が可能となりつつあり、本分野にも貢献できている。

3.1.2. ライフサイエンス分野（バイオインフォマティクス以外）やその他、科学技術分野の研究への貢献

BIRD 研究開発期間中に、世界的にメタゲノム研究が実施されるようになってきた。特にヒト細菌叢に対するメタゲノム研究は、HMPやMetaHITなど欧米の巨大な研究費を背景に、大規模な研究が実施されるようになった。一方、本研究分野における我が国の状況は芳しくなかったため、メタゲノム研究の啓蒙と奨励、若手研究者の育成などが喫緊の課題

となっていた。研究代表者らは、2007 年世界に先駆けて発表したヒト腸内細菌叢の研究成果および BIRD において開発した成果を軸に、Human MetaGenome Consortium JAPAN (HMGJ) を 2008 年に立ち上げ、研究の啓蒙、技術指導、技術移転などを実施してきた。2013 年には Japanese Consortium for Human Microbiome (JCHM : <http://bioinfo.bio.titech.ac.jp/jchm/>) を発足させ、HMGJ の活動を継承するとともに、産業界での応用なども含めより発展させた活動を継続している。さらに、ライフサイエンス分野においては、新学術領域研究「ゲノム科学の総合的推進に向けた大規模ゲノム情報生産・高度情報解析支援（通称：ゲノム支援）」にて申請される多様なメタゲノム研究課題に対して、BIRD にて開発した解析パイプラインおよび可視化技術を徹底活用し、より迅速な解析支援を実現している。

3.2. 人材育成の面から参加研究者の活動状況

3.2.1. 研究開発に参加した研究者のキャリアアップ

本研究開発課題に参加した者で、その後新たなポジションを得た者は 2 名いた。

3.3. 社会的及び経済的な効果

3.3.1. 研究開発成果が大学や公的研究機関、企業等で、応用に向けて継承または発展した例

メタゲノム解析が国内外で一般化したことを受け、BIRD にて開発してきたデータベース、可視化技術および解析パイプラインの活用事例が急速に高まった。特に 2010 年度から開始された新学術領域「ゲノム科学の総合的推進に向けた大規模ゲノム情報生産・高度情報解析支援」事業において、多数のメタゲノム解析支援依頼があり、解析から可視化までシームレスな情報提供・技術支援を行っている。

また、ベンチトップシーケンサーの登場によるデータ生産コストの低下に伴い、民間企業においてもメタゲノム解析に対する障壁が取り払われようとしている。研究代表者の研究室においても、BIRD および MicrobeDB.jp の開発を基盤として、化学、食品、飲料水、製薬、農業、環境、商社、情報、住宅など多業種にわたる企業と、メタゲノムに関して共同研究中である。しかし、研究室レベルでこれらニーズに対応するのは限定的であるため、上述した通り、JCHM を発足させ、オールジャパン体制で次世代産業の育成に努めている。

3.3.2. その他、研究開発成果が社会的、経済的な効果・効用につながる兆し、可能性

次世代シーケンサーの普及に伴い、環境ゲノム研究が進展することが予想され、大量配列データ解析技術の開発などの課題が存在するものの、本課題で開発されたシステムはそのような研究を支えるプラットフォームとなる可能性がある。ヒトメタゲノム研究は、食や健康に関する研究や産業分野において重要性が増すことが予想されることから、本課題での成果の貢献度もまた大きくなっていくと思われる。

本研究では新型シーケンサーを用いたメタゲノム情報解析の基盤技術を確立しており、今後の生命科学研究に大きく貢献することが期待できる。また、微生物工学においてゲノムレベルでの理解が可能となったため産業への応用も期待できる。

新型シーケンサーが一般化したことで、ヒトのみならずあらゆる環境におけるメタゲノム研究が世界的に行われるようになってきた。しかし、メタゲノム研究では複雑なバイオインフォマティクス技術を駆使する必要があるため、本研究の解析パイプラインが必須となる。すでに国内外の研究者からの共同研究の申し込みが多く寄せられている。また、本研究に関して、例えば、健康産業、医薬品業など、ヒトの健康と密接に関わる企業のみならず、食品産業や化学産業など、さらには農業など多岐にわたる産業分野から問合せを受けている。本研究によりメタゲノム研究は、これまで曖昧であった微生物環境をはじめで数値化しただけでなく、可視化することで生物環境の理解が圧倒的に促進されと考えている。

新型シーケンサーが一般化したことで、ヒトのみならずあらゆる環境におけるメタゲノム研究が世界的に行われるようになってきた。しかし、メタゲノム研究では複雑なバイオインフォマティクス技術を駆使し、かつ大規模な計算をする必要があるため、解析パイプラインの構築ならびに大型計算機（スパコン、クラウド）における解析が必須となる。すでに国内外の研究者と共同研究をするとともに、メタゲノム研究を志す海外（台湾、タイ、オランダ）からの学生を当該研究室で受け入れ、メタゲノム解析に関する教育・研究を開始している。また、健康産業、医薬品業など、ヒトの健康と密接に関わる企業のみならず、食品産業、化学産業やIT産業、さらには農業など多岐にわたる産業分野から問合せを受け、IT産業および化学メーカーとの共同研究を開始するに至った。本研究により、これまでその困難さゆえ避けられてきたメタゲノム研究のハードルを下げることができ、微生物が関わるあらゆる産業において、これまでブラックボックスであった微生物集団をデジタル化することで、生産の安定化やQOLの向上など、これまで科学的検証が困難であった広大な未踏領域に新たな展開をたらすものと期待できる。

資料編

1. 論文リスト

1. 黒川顕 メタゲノミクスのバイオインフォマティクス, 化学と生物, 45:495-501, 2007.
2. 服部正平, 林哲也, 黒川顕, 伊藤武彦, 桑原知巳, ヒト腸内細菌叢のゲノムシーケンス, 腸内細菌学雑誌, 21:187-197, 2007
3. 黒川顕, 服部正平 細菌叢のメタゲノム解析, 細胞工学別冊「比較ゲノム学から読み解く 生命システム」, 東京, 161-164, 2007
4. 黒川顕, 服部正平 メタゲノムデータベース, 実験医学 4 月増刊号「バイオインフォマティクスツールの開発と生命研究への応用の最前線」, 37-42, 羊土社, 東京, 2008.
5. #* Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K., Srivastava, T. P., Taylor, T. D., Noguchi, H., Mori, H., Ogura, Y., Ehrich, D. S., 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 (査読あり)
6. #* Hattori, M., and Taylor, T.D. The human intestinal microbiome: a new frontier of human biology. DNA Res., 16:1-12, 2009 (査読あり)
7. # Noguchi, H., Park, J., and Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res., 34:5623-5630, 2006 (査読あり)
8. #* Noguchi, H., Taniguchi, T., and Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res., 15:387-396, 2008(査読あり)
9. #* Goto, N., Kurokawa, K., and Yasunaga, T. Analysis of invariant sequences in 266 complete genomes. Gene, 401:172-180, 2007 (査読あり)
10. 森宙史, 林哲也, 黒川顕 「メタゲノム研究の最前線」, 蛋白質核酸酵素, 54:1264-1270, 2009.
11. # Arumugam, M., Raes, J., Pelletier, E., Paslier, DL., Yamada, T., Mendel, DR., Fernandes, GR., Tap, J., Bruls, T., Batto, JM., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, HB., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, EG., Wang, J., Guarner, F., Pedersen, O., deVos, WM., Brunak, S., Doré, J. MetaHIT Consortium, Weissenbach J, Ehrlich S D, Bork P (2011) Enterotypes of the human gut microbiome. Nature, 473:174-180.

12. Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., Maruyama, F., Nakagawa, I. (2011) CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. PLoS one, 6:e19543.
13. Watanabe, T., Maruyama, F., Nozawa, T., Aoki, A., Okano, S., Shibata, Y., Oshima, K., Kurokawa, K., Hattori, M., Nakagawa, I., Abiko, Y. (2011) Complete genome sequence of the periodontogenic bacterium *Porphyromonas gingivalis* TDC60. J Bacteriol. Epub Jun 24.
14. Tajima, N., Sato, S., Maruyama, F., Kaneko, T., Sasaki, NV., Kurokawa, K., Ohta, H., Kanesaki, Y., Yoshikawa, H., Tabata, S., Ikeuchi, M., Sato, N. (2011) Genomic Structure of the Cyanobacterium *Synechocystis* sp. PCC 6803 Strain GT-S. DNA Res., 18:393-399.
15. Kuwahara, T., Ogura, Y., Oshima, K., Kurokawa, K., Ooka, T., Hirakawa, H., Itoh, T., Nakayama-Imaohji, H., Ichimura, M., Itoh, K., Ishifune, C., Maekawa, Y., Yasutomo, K., Hattori, M., Hayashi, T. (2011) The Lifestyle of the Segmented Filamentous Bacterium: A Non-Culturable Gut-Associated Immunostimulating Microbe Inferred by Whole-Genome Sequencing. DNA Res., 18:291-303
16. Mori, H., Maruyama, F., and Kurokawa, K. VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences., BMC Bioinfo., 11:332, 2010.
17. Nakayama, K., Kurokawa, K., Fukuhara, M., Urakami, H., Yamamoto, S., Yamazaki, K., Ogura, Y., Ooka, T., Hayashi, T. Genome comparison and phylogenetic analysis of *Orientia tsutsugamushi* strains., DNA Res., 17:281-291., 2010.
18. Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., Kodama, T., Abe, H., Nakayama, K., Kurokawa, K., Tobe, T., Hattori, M., and Hayashi, T. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*., Proc. Natl. Acad. Sci. USA, 106:17939-17944, 2009.
19. Masahira, H., Oshima, K., Kim, S., Kurokawa, K., Toh, H., and Taylor, T. Metagenomics and genomics decoding human gut microbiome, Genes Genet. Syst., 84:479-479, 2009.
20. Toh, H., Oshima, K., Toyoda, A., Ogura, Y., Ooka, T., Sasamoto, H., Park, S.H., Iyoda, S., Kurokawa, K., Morita, H., Itoh, K., Taylor, T.D., Hayashi, T., and Hattori, M. Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15 belonging to phylogenetic group B2., J. Bacteriol., 192:1165-1166, 2009.
21. Hirakawa H., Akita H., Fujiwara T., Sugai M., Kuhara S. Structural insight into the binding mode between the targeting domain of ALE-1 (92AA) and pentaglycine of peptidoglycan. Protein Eng. Des. Sel., 22, 385-391, 2009.
22. 森宙史, 丸山史人, 黒川顕, メタゲノムインフォマティクス, 難培養微生物研究の最新

技術 II, pp82-91, 2010, シーエムシー出版, 東京

23. 森宙史, 丸山史人, 黒川顕 (2010) メタゲノムデータベース, メタゲノム解析技術の最前線, シーエムシー出版, pp42-53
24. Nonaka, L., Maruyama, F., Miyamoto, M., Miyakoshi, M., Kurokawa, K., Masuda, M. (2012) Novel Conjugative Transferable Multiple Drug Resistance Plasmid pAQU1 from *Photobacterium damsela* subsp. *damsela* Isolated from Marine Aquaculture Environment. *Microbes Environ.*, Mar 23
25. # Suzuki, S., Ishida, T., Kurokawa, K., Akiyama, Y. (2012) GHOSTM: A GPU-accelerated homology search tool for metagenomics. *PLoS one*, 7:e36060
26. Tuama, EÓ., Deck, J., Dröge, G., Döring, M., Field, D., Kottmann, R., Ma, J., Mori, H., Morrison, N., Sterk, P., Sugawara, H., Wiczorek, J., Wu, L., Yilmaz, P. (2012) Meeting Report: Hackathon-Workshop on Darwin Core and MIxS Standards Alignment (February 2012). *Stand Genomic Sci.*, 7:166-70.
27. Minegishi, K., Aikawa, C., Furukawa, A., Watanabe, T., Nakano, T., Ogura, Y., Ohtsubo, Y., Kurokawa, K., Hayashi, T., Maruyama, F., Nakagawa, I., Eishi, Y. (2013) Complete Genome Sequence of a *Propionibacterium acnes* Isolate from a Sarcoidosis Patient. *Genome Announc.*, 1:e00016-12.
28. Masuda, S., Hori, K., Maruyama, F., Ren, S., Sugimoto, S., Yamamoto, N., Mori, H., Yamada, T., Sato, S., Tabata, S., Ohta, H., Kurokawa, K. (2013) Whole-genome sequence of the purple photosynthetic bacterium *Rhodovulum sulfidophilum* strain W4. *Genome Announc.*, 1:e00577-13.
29. Aikawa, C., Furukawa, N., Watanabe, T., Minegishi, K., Furukawa, A., Eishi, Y., Oshima, K., Kurokawa, K., Hattori, M., Nakano, K., Maruyama, F., Nakagawa, I., Ooshima, T. Complete genome sequence of the serotype k *Streptococcus mutans* strain LJ23 (2012) *Journal of Bacteriology*, 194 (10), pp. 2754-2755.
30. # Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, Nagata Y, Fujiyama A, Tsuda M, Kurokawa K (2013) Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. *DNA Res.*, dst052.
31. 黒川顕 (2011) 微生物統合データベース” MicrobeDB.JP” , 細胞工学, vol.30, No.12.
32. 森宙史, 山田拓司, 黒川顕 (2012) メタゲノム解析の現状と将来, 情報管理, 55(3):167-175.
33. 森宙史, 山田拓司, 黒川顕 (2013) ヒトメタゲノムの大規模コホート研究, 生物の科学「遺伝」, 67(3):279-282.

: 今回の追跡調査において研究代表者が主要な論文として指定したもの (上限 30 報)

* : 研究開発期間終了後の終了報告書において研究代表者が主要な論文として指定したもの

2. 主要論文の被引用回数

論文	国内外別 件数		分野別件数																				出版年別件数									
	国内 件数	海外 件数	農学	生物 学・ 生化学	化学	臨床 医学	コン ピュー ター サイ エンス	経済 学・ 経営 学	工学	環 境・ 生態 学	地球 科学	免疫 学	材料 科学	数学	微生物 学	分子 生物 学・ 遺伝 学	複合 領域	神経 科学・ 行動 科学	薬理 学・ 毒物 学	物理 学	植物 学・ 動物 学	精神 医 学・ 心理 学	社会 科 学・ 一般	宇宙 科学	2006 年	2007 年	2008 年	2009 年	2010 年	2011 年	2012 年	
5. *Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K., Srivastava, T. P., Taylor, T. D., Noguchi, H., Mori, H., Ogura, Y., Ehrich, D. S., 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(. 査読あり)	24	200	11	25	2	57	7	0	0	11	0	2	0	1	59	40	0	3	5	0	0	0	0	0	—	0	19	41	52	57	55	
6. *Hattori, M., and Taylor, T.D. The human intestinal microbiome: a new frontier of human biology. DNA Res., 16:1–12, 2009.(査読あり)	1	56	5	6	1	20	1	0	0	3	0	1	0	0	14	1	0	0	3	0	1	0	1	0	—	—	—	2	17	24	14	
7. Noguchi, H., Park, J., and Takagi, T. MetaGene: prokaryotic gene fi nding from environmental genome shotgun sequences. Nucleic Acids Res., 34:5623–5630, 2006.(査読あり)	6	96	2	21	0	17	15	0	1	4	0	0	0	1	22	16	0	0	0	0	1	0	0	0	0	5	9	23	19	23	23	
8. *Noguchi, H., Taniguchi, T., and Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res., 15:387–396, 2008.(査読あり)	16	33	0	7	0	9	8	0	0	0	0	0	0	0	21	4	0	0	0	0	0	0	0	0	—	—	0	4	5	13	27	
9. *Goto, N., Kurokawa, K., and Yasunaga, T. Analysis of invariant sequences in 266 complete genomes. Gene, 401:172–180, 2007.(査読あり)	0	3	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	0	0	0	0	1	2	0
11. Arumugam, M., Raes, J., Pelletier, E., Paslier, DL., Yamada, T., Mendel, DR., Fernandes, GR., Tap, J., Bruls, T., Batto, JM., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, HB., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz–Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, EG., Wang, J., Guarner, F., Pedersen, O., deVos, WM., Brunak, S., Doré, J. MetaHIT Consortium, Weissenbach J, Ehrlich S D, Bork P (2011) Enterotypes of the human gut microbiome. Nature, 473:174–180.	8	263	19	32	1	86	8	0	1	11	0	21	0	0	52	25	3	1	6	1	3	0	0	0	—	—	—	—	—	55	216	
25. Suzuki, S., Ishida, T., Kurokawa, K., Akiyama, Y. (2012) GHOSTM: A GPU-accelerated homology search tool for metagenomics. PLoS one, 7:e36060	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	—	—	—	—	—	—	0

- ・本追跡調査において研究代表者が主要な論文として指定したもの（上限 30 報）について、トムソン・ロイター社 Web of Science で調査した。調査対象は、2012 年 12 月まで。被引用情報が取得できたものののみ記載した。
- ・論文番号は「資料編 1. 論文リスト」に対応している。
- ・＊：研究開発期間終了後の終了報告書において研究代表者が主要な論文として指定したもの。
- ・[国内外別件数] は、被引用文献の国内外別件数。被引用文献の著者の所属国のうち JAPAN が一つであれば国内としている。国内外の合計が全被引用文献数となる。
- ・[分野別件数] は、被引用文献の分野別件数。被引用文献が論文の場合のみカウントしている。
- ・[出版年別件数] は、被引用文献の出版年別件数。「—」は、データなしを表す。

論文	Web of Science 国内外別件数(2012 年 12 月まで)			Google Scholar	
	国内件数	海外件数	合計	～2014 年 4 月 24 日	～2012 年
5. * Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K., Srivastava, T. P., Taylor, T. D., Noguchi, H., Mori, H., Ogura, Y., Ehrich, D. S., 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(. 査読あり)	24	200	224	459	361
6. * Hattori, M., and Taylor, T.D. The human intestinal microbiome: a new frontier of human biology. DNA Res., 16:1–12, 2009.(査読あり)	1	56	57	149	117
7. Noguchi, H., Park, J., and Takagi, T. MetaGene: prokaryotic gene fi nding from environmental genome shotgun sequences. Nucleic Acids Res., 34:5623–5630, 2006.(査読あり)	6	96	102	209	168

8. *Noguchi, H., Taniguchi, T., and Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res., 15:387–396, 2008.(査読あり)	16	33	49	113	75
9. *Goto, N., Kurokawa, K., and Yasunaga, T. Analysis of invariant sequences in 266 complete genomes. Gene, 401:172–180, 2007.(査読あり)	0	3	3	7	3
11. Arumugam, M., Raes, J., Pelletier, E., Paslier, DL., Yamada, T., Mendel, DR., Fernandes, GR., Tap, J., Bruls, T., Batto, JM., Bertalan, M., Borrue!l, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, HB., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz–Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, EG., Wang, J., Guarner, F., Pedersen, O., deVos, WM., Brunak, S., Doré, J. MetaHIT Consortium, Weissenbach J, Ehrlich S D, Bork P (2011) Enterotypes of the human gut microbiome. Nature, 473:174–180.	8	263	271	1045	505
25. Suzuki, S., Ishida, T., Kurokawa, K., Akiyama, Y. (2012) GHOSTM: A GPU-accelerated homology search tool for metagenomics. PLoS one, 7:e36060	0	0	0	7	5
30. Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, Nagata Y, Fujiyama A, Tsuda M, Kurokawa K (2013) Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. DNA Res., dst052.	—	—	—	(引用元記載なし)	—

- ・ 情報科学・計算機科学分野における研究開発成果では、会議予稿集での発表に引用されることが多いため、Google Scholar での被引用件数を参考情報として添付した。
- ・ Google での調査方法

対象論文のタイトルを検索し、当該論文を「引用元」とする件数を取得した。
 「引用元記載なし」とあるのは、「引用元」の表記がなく引用件数のデータがないもの。
 最新の被引用件数と、それから 2013 年以降の被引用件数を引いたものを 2012 年までとした。ただし、年が明確でないものは引かれていないため、2012 年までの被引用件数が実際よりも多くなっている可能性がある。
- ・ 論文番号は「資料編 1. 論文リスト」に対応している。
- ・ *：研究開発期間終了後の終了報告書において研究代表者が主要な論文として指定したもの。

3. 学会招待講演・基調講演

1. 黒川顕（奈良先端大），ゲノム解析からメタゲノム解析へ，第3回アグリバイオインフォマティクスシンポジウム，東京，2007年3月
2. 黒川顕（奈良先端大），ヒト腸内フローラのメタゲノム解析，日本乳酸菌学会 2007 年度大会，相模原，2007年7月
3. Masahira Hattori（東大），Sequencing of Bacterial Genomes and Metagenomes, Metagenomics 2007, SanDiego, USA, 2007年7月
4. 黒川顕（奈良先端大），ヒト腸内フローラのメタゲノム解析，新化学発展協会，東京，2007年10月
5. 黒川顕（奈良先端大），メタゲノムインフォマティクス，第2回放射線防御研究センターシンポジウム，千葉，2007年12月
6. 黒川顕（東工大），次世代シーケンサーによるメタゲノム解析の可能性，第288回CBI学会研究講演会，東京，2008年8月
7. 黒川顕（東工大），メタゲノムインフォマティクス，日本農芸化学会 2009 年度大会シンポジウム，福岡，2009年3月
8. 黒川顕（東工大），新型シーケンサーとメタゲノム解析，第61回日本生物工学会年会シンポジウム，名古屋，2009年9月
9. 平成21年度採択継続
10. 黒川顕（東工大），ゲノム情報立脚型社会，日本生体医工学会バイオメクフォーラム，大阪，2010年12月11日
11. 黒川顕（東工大），ゲノムからメタゲノムへ，平成22年度千里ライフサイエンスセミナー
12. 黒川顕（東工大），ゲノムからメタゲノムへ，日本進化学会第12回東京大会，東京，2010年8月3日
13. 黒川顕（東工大），ゲノムからメタゲノムへ，第55回日本ブドウ球菌研究会，東京，2010年7月21日
14. 黒川（東工大），次世代シーケンサーと環境メタゲノム解析，JBA バイオエンジニアリング研究講演会「次世代シーケンサーと産業応用」，東京，2009年12月
15. Ken Kurokawa (TITECH), Metagenomics analysis using Next-generation sequencing technology、第32回日本分子生物学会年会シンポジウム，神奈川，2009年12月
16. 黒川顕（東工大），新型シーケンサーによるメタゲノム解析，IPAB（並列生物情報処

- 理 イニシアティブ) 創立 10 周年記念シンポジウム, 東京, 2009 年 11 月
17. 黒川顕 (東工大), 新型シーケンサーによるメタゲノム解析, バイオインフォマティクス研究者と医学研究者の交流会, 千葉, 2009 年 11 月
 18. JBA バイオエンジニアリング研究講演会「次世代シーケンサーと産業応用」, 東京, 2009 年 12 月, 次世代シーケンサーと環境メタゲノム解析, 黒川 顕
 19. 第 32 回日本分子生物学会年会シンポジウム, 神奈川, 2009 年 12 月, Metagenomics analysis using Next-generation sequencing technology, Ken Kurokawa
 20. 日本細菌学会総会, 黒川顕, 新型シーケンサーとメタゲノム解析, 2010 年 3 月 28 日, 神奈川.
 21. 日本ブドウ球菌研究会, 黒川顕, ゲノムからメタゲノムへ, 2010 年 7 月 21 日, 東京.
 22. 日本進化学会, 黒川顕, ゲノムからメタゲノムへ, 2010 年 8 月 3 日, 東京.
 23. 千里ライフサイエンスセミナー, 黒川顕, ゲノムからメタゲノムへ, 2010 年 10 月 22 日, 大阪.
 24. 日本生体医工学会バイオメクフォーラム, 黒川顕, ゲノム情報立脚型社会, 2010 年 12 月 11 日, 大阪
 25. 日本ゲノム微生物学会, 黒川顕, ゲノム・メタゲノム情報を基盤とした微生物 DB の統合, 2011 年 8 月 21 日, 仙台
 26. 第 63 回日本生物工学会大会, 黒川顕, 新型シーケンサーによるメタゲノム解析, 2011 年 9 月 27 日, 東京
 27. CBI/JSBi2011 合同大会, 黒川顕, ゲノムからメタゲノムへ, 2011 年 11 月 9 日, 神戸
 28. 農業環境技術研究所公開セミナー, 生態系の理解に向けたゲノム情報の解析と活用, 2011 年 12 月 6 日, つくば
 29. 第 12 回 iPAB (並列生物情報処理イニシアティブ) シンポジウム, 黒川顕, 新型シーケンサーによるメタゲノム解析, 2011 年 12 月 9 日, 東京
 30. BiWO2011 (Bioinformatics Week in Odaiba 2011), Development of an integrated analysis system for metagenomics and its applications for soil metagenome analysis, Ken Kurokawa, 2012 年 1 月 26 日, 東京
 31. 日本農芸化学会 2012 年度大会シンポジウム, 黒川顕, 環境メタゲノミクス, 2012 年 3 月 25 日, 京都
 32. 第 28 回日本微生物生態学会シンポジウム, Development of an integrated analysis system for metagenomics and a global microbial database MicrobeDB.jp, Ken Kurokawa, 2012 年 9 月 20 日, 豊橋
 33. 情報・システム研究機構シンポジウム 2012 「生命科学のビッグデータ革命」, 黒川顕,

メタゲノミクスの現状と未来, 2012 年 11 月 9 日, 東京

34. 社団法人企業研究会 CAMM フォーラム, 黒川顕, メタゲノム解析の現状と将来, 2013 年 1 月 11 日, 東京
35. The 1st ELSI International Symposium, Deciphering ancient ecosystems with environmental genomics and “Earth-Life DB”, Ken Kurokawa, 2013 年 3 月 29 日, 東京

4. 新聞発表等（著作権の関係により非公開）

5. 特許出願・成立

該当なし

6. 学会賞等の受賞

1. 日本ゲノム微生物学会研究奨励賞（2008 年 3 月）
2. IBM SUR Awards “Stream Computing for Real-Time Traffic Monitoring and Bioinformatics” 徳田雄洋・東工大大学院理工学研究科教授, 鈴木豊太郎・東工大大学院理工学研究科客員准教授らと受賞
3. Linked Open Data Challenge 2011 ライフサイエンス賞（平成 23 年度）

7. グラントの獲得実績

1. 新世代 DNA 塩基配列決定技術を用いた大腸菌 pan-genome の解明
林 哲也 2008-2011 年 基盤研究(B)
2. ゲノム科学の総合的推進に向けた大規模ゲノム情報生産・高度情報解析支援
小原 雄治 2010-2015 年 新学術領域研究
3. 環境汚染物質添加土壌での微生物集団応答様式の研究
津田 雅孝 2010-2013 年 基盤研究(B)
4. ゲノム・メタゲノム情報を基盤とした微生物 DB の統合
黒川 顕 2011-2013 年 JST NBDC 統合化推進プログラム

8. 書籍等の執筆実績

該当なし

9. 総説の執筆実績

該当なし

10. 参加研究者の活動状況（個人情報が含まれるため非公開）