

共生、相互作用など、複雑なゲノム構成系を解析するための情報基盤研究

●久原 哲¹⁾ ◇内山 郁夫²⁾ ◇黒川 顕³⁾ ◇平川 英樹¹⁾

1) 九州大学大学院農学研究院 2) 自然科学研究機構基礎生物学研究所 3) 東京工業大学生命情報専攻

<研究の目的と進め方>

生命・生物の特徴である普遍性と多様性のメカニズムを解明するために、ゲノム配列の比較と遺伝子の使われ方あるいは相互作用・ネットワークの違いを解析する情報学的基盤システムの構築を行う。基盤システムとしては、モデル生物として微生物間のオルソログ、パラログ遺伝子あるいは特異的な遺伝子の自動作成とデータベース化を行い、このデータに種々の生物学情報、遺伝子発現データ等を統合した配列比較情報解析システムに拡張する。また、最近、特に注目されている土壌や海洋あるいは腸内といった環境中における微生物の群集をまるごとゲノム解析するメタゲノム解析にも上記で開発した手法を適用する。このメタゲノム解析の困難さの原因とされている、遺伝子データベースのデータおよび解析ツールの不十分さに注目する。本研究では、これらの原因を解消するため、基盤となるデータベースの整備、ツールの開発も同時に行う。同時に今後の微生物研究の基盤となる難培養性微生物を含む微生物集団（土壌中、腸管内等）でのポピュレーションを計測する新チップの設計・製作も行う。

<2008年度の研究の当初計画>

1) 配列データの整備

これまでに作成したMBGDにユーザ独自のデータを載せて解析する「My MBGD」機能をさらに進め、より大規模な解析を可能とするため、MBGDのサーバ構築機能をローカルインストールして利用できるようにする。その際、メタゲノム解析のような不完全な情報の解析も念頭におき、データ構築方式を検討する。また、MBGDを「リファレンスデータベース」として使うようにすることを目標に、オーソロググループごとのアノテーション付けやIDの付与方式について検討する。あわせて、MBGDではこれまで不完全だった、菌類など真核微生物のデータを充実させる。また、昨年度開発したコア構造構築手法を、より広範なデータに適用することも引き続き検討する。（連携研究者の内山を中心として行う）

2) 生物学データの整備と比較ゲノム解析およびメタゲノム遺伝子配列の種同定を行うシステム開発

細菌のゲノムに存在するすべての遺伝子について、個々に系統関係を分子進化学的解析により明らかにし、オーソログ、パラログ、シングルトンに分類することで、種への分化途上で起きたイベントを詳細に記述する。メタゲノム解析により得られた遺伝子の断片配列を既存のデータベースから、種、属、科、目などの分類群と、遺伝子相同性との関連性を明らかにするツールを開発し、腸内細菌科等に適用する。今後はさらに解析の対象を拡げ、他の科、属に対しても予測精度を推定する。また、相違度データベースからの情報を取り入れることで、より詳細な分類推定を可

能にする。（連携研究者の黒川を中心として行う）

3) 情報学的ツールの整備

細菌のゲノムに存在する遺伝子について、オルソログ遺伝子等に注目し、各オルソログ間の類似度等を計算し、オルソログマトリックスを構築し、このマトリックスに細菌の形質等生物学的情報を付加したゲノムマトリックス等を構築する。このマトリックスを基盤として、数理統計学的手法を用いて、形質等を説明する遺伝子群の同定を行う。（代表者の久原を中心として行う）

4) 新規微生物分野の情報収集のための基盤整備

現在作製している微生物叢解析用マイクロアレイの問題点であるクロスハイブリの原因は、ミスマッチが正確に見積もられておらず、プローブ配列の二次構造を考慮していないためと考えている。これらの問題を解決し、より精度が高いマイクロアレイを作製する。さらに、土壌や海洋などの環境中から16S rDNA配列を収集し、各微生物叢に生息する微生物群についてプローブ設計を行い、マイクロアレイを新たに作製する。また、16S rDNAの他の流域、もしくは、DNAポリメラーゼなどのハウスキーピング遺伝子をプローブに用いることを検討する。（連携研究者の平川を中心として行う）

<2008年度の成果>

1) 配列データの整備

MBGDのサーバ構築機能をローカルインストールして独自データを解析できるようにしたものを作成し、専用の解析インターフェイスを持つ比較ゲノムワークベンチRECOGの一部として公開した。また、MBGDのオーソロググループごとのアノテーションを充実させるとともに、菌類などの真核微生物を拡充したバージョンを作成し、公開した。中程度の類縁ゲノム間でオーソログな遺伝子の並びを比較し、並びが保存された領域を抽出する「ゲノムコア構造アライメント」手法（CoreAligner）を開発した。この手法を用いてバチルス科と腸内細菌科のゲノムデータからコア構造を構築し、機能カテゴリ、必須遺伝子、GC含量、系統樹等の観点から、コア遺伝子の特徴付けを行った。その結果、抽出されたコア遺伝子群は必須遺伝子をはじめとする重要な遺伝子の多くを含んでおり、また非コア遺伝子と比べると垂直的に伝搬してきた割合が大きいことを示唆する結果を得た。開発したCoreAlignerアルゴリズムを比較ゲノムワークベンチRECOGに組み込んで様々なゲノム解析に適用可能にした。

2) 生物学データの整備と比較ゲノム解析およびメタゲノム遺伝子配列の種同定を行うシステム開発
新型シーケンサーを用いた16S rRNAによる群集構造解析に

に向けたソフトウェア開発ならびにデータベース整備を行った。ヒト常在細菌の中で、Gut (大腸)、Skin (皮膚)、Lung (肺)、Oral (口腔)、Vagina (膣) の5部位に生息する細菌の16S rRNA 遺伝子配列、計 53,750 本を GenBank から抽出し、菌種組成や配列相同性により比較し、各フローラにおける細菌群集の特徴を抽出した。ヒト常在細菌のうち未同定の新規の細菌と思われる細菌種については、16S rRNA 遺伝子配列の相同性により他の環境中に生息する細菌と比較し、新規の細菌の由来と環境中における分布についても明らかにした。また、取得したヒト常在細菌の16S rRNA 遺伝子配列情報をデータベース化するとともに、群集構造を分子系統的に容易に理解するための可視化ソフトウェア、ヒト常在細菌の配列&メタ情報を検索できる web アプリケーションも作成した。また、454 シークエンサーによる大規模群集構造解析を効率良く推進するためのユニバーサルプライマー設計手法を開発し、これまで公開されている 724 の細菌ゲノム配列から各分類群をターゲットとした 454 シークエンサーに特化したプライマーセットを設計した。

3) 情報学的ツールの整備

黄色ブドウ球菌の176の臨床分離株における性質(病状、部位、感染場所など)を特徴付ける遺伝子群を明らかにするため、クラス識別解析を行った。各株における遺伝子の有無は、アレイ CGH (Comparative Genomic Hybridization) 実験に基づき判定した。2つの病状間で発現が異なる遺伝子をクラス識別により調べたところ、それぞれの病状に関連が深いと思われる遺伝子群が抽出され、さらに、幾つかの機能未知な遺伝子も抽出された。

4) 新規微生物分野の情報収集のための基盤整備

微生物叢解析用マイクロアレイのクロスハイブリを解消する実験条件の検討を行なった。その結果、ハイブリの際、増幅断片に対するフラグメンテーションをすることが効果的であることが分かった。一方、幾つかの海水土壌に対してマイクロアレイを適用した結果、全体的な発光のパターンは類似していたが、各サンプルにのみ発光が見られるプローブも存在しており、微生物叢の違いを反映しているものと考えられた。また、全ゲノム配列が決定された微生物のオルソログ遺伝子の中から微生物の検出に用いるプローブの検討を行なった。

<国内外での成果の位置づけ>

MBGD データベースを基盤とした、微生物ゲノムを系統的に比較する環境の整備は着々と進んでおり、国内外でユニークな位置づけのシステムとして確立しつつある。オーソログ解析を基にしたゲノムコア構造アライメントは、手法としての斬新さはそれほどないかもしれないが、ゲノム比較へのアプローチとしてはユニークなものであり、今後重要となる近縁微生物ゲノムの大規模比較を行う際などに有用なアプローチとなりうるものと考えている。

クラス識別解析法は、従来、癌患者の予後予測に主に適用されてきた。我々は、この解析手法を、黄色ブドウ球菌の幾つかの異なる病原性に関与する遺伝子群を推定するために適用しており、これまでの結果によりその方法は有効であると考えている。

様々な環境における微生物叢を調べるマイクロアレイのうち代表的なものとして、Affymetrix 社の Phylochip が作製されているが、非常に高価である。また、近年、次世代シークエンサーによ

り様々な環境化における微生物叢が調べられているが、やはり高価である。そのため、マイクロアレイを用いた微生物叢の解析は、非常に有効であると考えられる。我々は、安価で汎用性が高いマイクロアレイの作製と実験方法の確立を目指している。

<達成できなかったこと、予想外の困難、その理由>

MBGD に不完全なデータを取り込むことについては、原理的にはそれほど難しくもないものの、スキーマや更新手続きの変更がかなり加わるために作業が遅れている。

GenBank から 16S rRNA 遺伝子データを抽出しデータベース化を行ったが、新規の大規模 16S rRNA シークエンスが猛烈な勢いで登録されており、既存の計算機設備を用いた配列アライメントおよびクラスタリングなどの計算が困難であったため、GenBank 上のすべての 16S rRNA 遺伝子データを対象とする事ができなかった。しかしながら、メタ情報を伴った配列情報のみを慎重に抽出した上でデータベース化し、大規模シークエンスに関しては既データベースにマッピングすることで、すべてのデータを対象とした解析を実施することが可能となった。

黄色ブドウ球菌の176の臨床分離株における性質(病状、部位、感染場所など)を特徴付け、クラス識別した結果、性質によっては、識別の精度が低いという問題が生じた。その原因としては、性質による特徴付けの問題と、クラス識別の手法の問題が考えられた。

微生物叢解析用マイクロアレイの開発では、クロスハイブリがなるべく生じない実験条件の検討を主に行なった。このため、二次構造を考慮したプローブの再設計をすることができなかった。

<今後の課題>

MBGD のリファレンスデータベースとしての使い勝手の向上を図りつつ、より広範なデータに適用可能とするためのスキーマの変更を含めた改良作業を続ける。また、新規に解読されたゲノムの解析や近縁種間ゲノム比較、およびメタゲノム解析に MBGD/RECOG システムを効果的に活用するための方策について検討する。

構築したヒト常在細菌の16S rRNA 遺伝子データベースを公開する予定である。また、設計した 454 シークエンサーに特化したプライマーセットが、実際の細菌群集構造解析に対して有効であるかを判断するために、細菌群集を対象として PCR や DGGE 等による多様性解析実験を実施する。

クラス識別については、幾つかの手法が提案されているため、今後はそれらの手法を適用する。

微生物叢解析用マイクロアレイの開発では、実験条件を検討したことにより、クロスハイブリが生じたプローブ配列をより高い精度で同定することができるようになったため、今後は二次構造を考慮したプローブ配列の再設計を行なう。また、オルソログ遺伝子を用いた新規なマイクロアレイの作製を行う。

<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. 0901122149

Uchiyama, I.: Multiple genome alignment for identifying the core structure among moderately related microbial genomes. BMC Genomics, 9, 515 (2008) .