

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⁴⁾

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) 奈良先端科学技術大学院大学
4) 九州大学大学院システム生命科学府

<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う。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。最後に相互作用・ネットワークの相違を解析するためのネットワーク比較解析システムを統合した、総合的システムを目指す。また、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。そのための基盤整備を行う。

<2007年度の研究の当初計画>

1) 配列データの整備

内山を中心に、MBGD (Microbial Genome Database) のオルソログ、パラログの自動生成プログラムの改良を行い、微生物の比較ゲノム解析の基本データベースとしての安定な運用を図る。これに加えて生物学研究者を中心とした統一アノテーション情報を付加する。これにより、比較ゲノムにおける基盤データベースの構築を図る。このシステムを基礎として、より深い比較解析を行う方向に進むと考えており、その一環として、特に近縁ゲノム間に共通に存在する「コア構造」を、ゲノム構造比較を通じて抽出する手法を開発する。コア構造は、主に垂直的に伝播した構造として、原核生物ゲノムの進化を考える際の基盤となるものと考えることができ、解析にこのような進化的視点を取り入れていくことが今後の重要になる。このことを踏まえて、開発中のRECOGシステムに、こうした新しい機能を取り込んで完成させる。

2) 生物学データの整備と比較ゲノム解析

黒川を中心として、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的解析により明らかにし、オルソログ、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。特に、すべての遺伝子の進化的履歴を明らかにするにあたり、種の代表的な進化速度の指標となり得る参照遺伝子として、種分化解析で頻りに用いられている16S rRNA 遺伝子を用い、参照遺伝子である16S rRNA 遺伝子と、他の遺伝子との進化速度を比較し、種分化と遺伝子進化との関係を明らかにする。

またインフォマティクス技術を用いたメタゲノム解析を行う。現在進行中の腸内細菌叢のメタゲノム配列解析データ、腸管ショットガンリード合計数が727Mb、予測した遺伝子合計数が462,223遺伝子に対して、既存のデータベースに対して相同性検索を行い、腸内細菌叢の属分類解析を行う。

3) 情報学的ツールの整備について

細菌のゲノムに存在する遺伝子について、オルソログ遺伝子に注目し、各オルソログ間の類似度を計算し、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

付加した細菌ゲノムマトリックスを構築する。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。この問題には、教師付き機械学習法の応用が考えられる。

4) 新規微生物分野の情報収集のための基盤整備について

平成18年度に作成した菌叢解析用チップでは、23merという短いプローブ配列を使っているため、クロスハイブリを少なくするために、ハイブリする際の蛍光試薬やバッファー等の実験条件を検討し、特定の微生物を検出することができるかどうかを調べる。これにより、DNAチップの精度を調べ、さらに、サンプリングされた混合微生物叢の解析を試みる。

<2007年度の成果>

1) 配列データの整備

MBGDに関しては、機能カテゴリ分類の情報を、COG、KEGG、TIGRの各データベースから取得してアサインする仕組みを組み込んだ。これにより、オルソロググループの特徴付けを行う指標が整備された。近縁ゲノムのコア構造アライメントについては、2遺伝子ではなく3遺伝子の並び順による近傍関係グラフを用いることにより、保存セグメントが不自然に融合することを防ぐとともに、複数のインパラログを含むオルソロググループについて、文脈に応じた適切なオルソログ関係を見つけて、必要に応じてグループを分割する処理を組み込んだ。これにより、構築されるコア構造の妥当性が改善された。また、このプログラムを用いてパチルス科と腸内細菌科のコア構造を構築して、機能カテゴリ、必須遺伝子等の、いくつかの観点からコア構造に含まれる遺伝子の特徴付けを行った。さらに、このコア構造アライメントプログラムを、比較ゲノムワークベンチ、RECOGシステムに組み込む作業を現在進めている。RECOGシステムは、内群、外群を指定したオルソログ解析を可能とするDomClustの改良版を中心とした比較ゲノムワークベンチで、特に興味ある菌とその近縁種を中心としつつ、それ以外のゲノムを外群として取り込んだ比較解析を効果的に行えることを目標としている。

2) 生物学データの整備と比較ゲノム解析

ヒト腸内細菌叢はヒトと密接に関わりながら複雑な共同体を形成している。それらヒト腸内細菌叢に共通なゲノムの特徴を特定するために、乳児から大人まで様々な年齢の13人の健康人を対象として大規模な比較メタゲノム解析を実施した。その結果、幼児における腸内細菌叢は大人とは大きく異なり単純で、かつ個人間における変異が極めて大きかった。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。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。それら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様々な戦略を示していると考えられる。また、特異的な conjugative トランスポゾンが、ヒト腸内において爆発的に増幅したことを発見した。このことは、ヒト腸内細菌間における遺伝子水平伝播のホットスポットであることを示唆している。

3) 情報学的ツールの整備について

クラス識別と遺伝子選択の手法については、Fisher の基準をはじめとするクラス分離度を測る複数の基準がカーネル化できることを示した。実験では、Fisher の基準がカーネル部分空間法におけるカーネルパラメータの選択において有効であることを示した。これらの手法を黄色ブドウ球菌のアレイ CGH データに対し適用することで、院内感染株と市中感染株間、とびひを発症する株と発症しない株間、敗血症を発症する株と発症しない株間に特徴がある遺伝子群の抽出を行った。

4) 新規微生物分野の情報収集のための基盤整備について

作製した DNA チップに対して、4 種の *Bacillus* 属それぞれについてのハイブリ実験を行い、対応した微生物のプロープ配列に対する特異性を調べたところ、クロスハイブリのため、特異性があまり高くなかった。その原因の一つとして、プロープ配列の二次構造がステム構造をとっている可能性が高いため、蛍光強度が低くなっていた。従って、プロープの設計では Tm 以外に、二次構造も計算する必要があることが分かった。さらに、クロスハイブリの原因として、配列の類似性の評価が不十分であった。プロープ配列を設計した際に配列の相同性は BLAST を用いて調べていたため、プロープ配列全体での類似性の評価ができていなかった。このため、配列の類似度を正確に見積もるには、FASTA を用いてプロープ配列の全体での類似度を評価する必要があることが分かった。なお、これらの問題点は、ハイブリ温度を上げることで実験的に解決できる可能性がある。

<国内外での成果の位置づけ>

1) 配列データの整備

MBGD の動的なオーソロググループ作成機能はきわめてユークである。また、網羅的なオーソログデータベースはいくつか存在するが、特に多様なデータが急速に蓄積している微生物に特化して、定期的にデータを更新しているという点でも特徴がある。コア構造については、近年近縁ゲノム比較の視点から多くの研究者が注目しつつあるが、中程度の類縁性を持つゲノムにおいて、遺伝子の並び順によるアライメント構築に基づくアプローチは、現在のところユニークである。また、近縁ゲノム比較と遠縁ゲノム比較を重ねるアプローチは、オーソドックスながら、これをシステムティックに進める方策は確立していない。

2) 生物学データの整備と比較ゲノム解析

論文発表と同時に国内の新聞各紙に記事が掲載された。さらに科学的側面だけでなく、ネット上の多くのウェブサイトにて取り上げられたことから、一般に対する興味の啓発にも大きく貢献したと考えられる。また Science 誌の主編集者からも取材を受けた。さらに、日米欧を中心として、国際的なコンソーシアムが発足した。この中でも本研究は大きく取り上げられたことから、本研究が先導的な研究として国際的に認識されていると考えられる。

3) 情報学的ツールの整備について

従来から開発してきた識別器、及び遺伝子選択のシステム開発は、マイクロアレイ等のデータに基づく識別問題において重要な役割を担っており、データ解析における中心的なテーマになっている。

4) 新規微生物分野の情報収集のための基盤整備について

現在のところ、微生物叢解析用の DNA マイクロアレイは、Affymetrix 社製の Phylochip が市販されている。しかし、そのマイクロアレイは非常に高価であり、また、クロスハイブリするプロープもスポットされている。我々はクロスハイブリがなるべく少ない DNA マイクロアレイを開発している。このため、プロープの設計については、Phylochip で行われた BLAST を用いた方法よりも精度が高い方法を検討している。また、ハイブリ実験に

についてもより精度が高い条件を検討している。

<達成できなかったこと、予想外の困難、その理由>

1) MBGD の更新に時間がかかるようになってきた。これまでは年間3回以上は更新してきたが、今年度は、今のところ何とか2回の更新が行える見込み、という状況である。実際にはいろいろなトラブルが重なっていることが遅れの原因ではあるが、その背景には、1回あたりの更新時間がかかるようになってきたことがある。ホモロジーデータは1年で倍増するペースで増えており、さらに加速する可能性もあることから、今後何らかの対策が必要になると考えている。

2) メタゲノム解析では膨大なデータを扱うため、計算機資源の確保が重要な課題である。また、解析が進むにつれ過去の計算結果も併せて保存する必要があるため、ストレージ領域も予想外に大量に確保する必要がある。

3) クラス識別問題では結果の信頼性を向上させることが問題になっている。この件については、十分なデータを得ることが必要であるが、実際の実験結果を得ることが十分にできなかった。

4) DNA マイクロアレイを作製する目標は達成できたが、クロスハイブリのため、その微生物に対する特異性があまり高くないことがわかった。また、作製している DNA マイクロアレイには、プロープの二次構造など幾つかの問題があることがわかったため、今年度に予定していた混合微生物叢への適用を試みることはできなかった。

<今後の課題>

データの増加の多くの部分は、既知ゲノムの近縁種であることから、近縁ゲノム比較と遠縁ゲノム比較を重ねるアプローチを前提とした処理方式が、データの急激な増加への対応策としても有効であると考えられる。そこで今後この方向性を推し進めつつ、データ構築の方策を検討する。また、国際的にも微生物に特化したゲノムデータ整理はあまり進んでいないことから、特に遠縁比較に向けた部分については、「標準データベース」として確立し、情報を充実させていくことも検討していきたい。

細菌叢全体での機能的側面も解析しきれていない。そのためにもデータベース整備や、解析技術の開発を今後は精力的に行う必要がある。

マイクロアレイについては、ハイブリ実験条件の検討、および、プロープの設計方法の検討を行う。また、16S rDNA では分解能が低いということが指摘されている。このため、RNA polymerase や DNA gyrase といった 16S rDNA 以外のハウスキーピング遺伝子について、プロープの設計を検討し、より精度が高い微生物叢解析用の DNA マイクロアレイを完成させる。

<成果公表リスト>

- ・ 0801220119
Kurokawa, K. et al., Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes., *DNA Res.*, 14, 169-181 (2007).
- ・ 0801231328
Ogura, Y. et al., Extensive genomic diversity and selective conservation of virulence-determinants in enterohaemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes., *Genome Biol.*, 8, R138 (2007).
- ・ 0801242034
Izutsu, K. et al., Comparative genomic analysis using microarray demonstrates a strong correlation between the presence of the 80-kilobase pathogenicity island and pathogenicity in Kanagawa phenomenon-positive *Vibrio parahaemolyticus* strains., *Infect. Immun.*, in press.