

研究開発課題別事後評価結果

1. 研究開発課題名

大規模な比較ゲノム研究を展開するためのワークベンチの構築

2. 代表研究者名

内山 郁夫（自然科学研究機構基礎生物学研究所 助教）

3. 代表研究者による成果開発概要

3-1. 研究目的

主要なゲノムプロジェクトの終了後も、ゲノム解読は多様な生物種へと拡大を続けている。プロジェクトの目的は様々であるが、興味のある生物種を中心とした近縁種の解読と、より幅広い系統をカバーする解読とが並行して進められており、いずれも比較解析が重要な役割を担うようになっている。ただし、これまでのゲノムプロジェクトでは、基本的には対象とする生物の理解が目標であり、比較対象もそれほど多くはなかった。しかしながら、蓄積しつつあるゲノムデータを駆使して、より普遍的な視点からゲノムの機能や進化を理解するには、各データを体系づけて横断的な解析を行う必要がある。本研究は、こうした見地から、当面 500 ゲノム程度をめどにした大規模な比較ゲノム研究を展開するためのワークベンチの開発を目指した。ただし、ゲノム進化を考える上で近年蓄積している近縁ゲノム情報は非常に有用であることを考慮し、ゲノム横断的な比較だけでなく、様々な類縁度の近縁ゲノム間比較をサポートすることも目指した。その際、従来は目的やアプローチの違いから別々に扱われていた近縁種比較と遠縁種比較を統合的に扱うことも目指した。

3-2. 研究開発概要

ゲノム横断的な解析を進めるためには、ゲノム間で網羅的なオーソログの対応をつける必要がある。我々はこれまでにそのような機能を持つ微生物比較ゲノムデータベース MBGD を作成しており、これを開発の土台として用いた。自動的なオーソログの対応付けについては、階層的クラスタリングに基づく独自の方法 (DomClust) を開発してきており、この技術の改良と活用とを中心課題のひとつに据えた。近縁種比較と遠縁種比較の統合については、利用者が興味のある系統群に含まれる生物種群を内群、その他を外群として指定して解析を行うことを想定し、DomClust の持つ階層的クラスタリングの枠組みを生かして、内群・外群指定に基づく系統的制約をオーソログ分類プロセスに取り込む手法の組み込みを行うこととした。

一方、オーソログ分類結果を活用した、特に近縁ゲノム比較を対象とした解析法として、遺伝子の並び順（シンテニー）の保存性に基づくゲノムアライメントの問題にも取り組んだ。ここでは、通常の配列アライメントとは違って遺伝子の並び順だけを考慮するものを考えた。これは、大規模解析向けの効率の良さに加え、通常のアライメントが難しいような類縁度の低いゲノム比較まで適用可能と思われるので、本ワークベンチの方向性に合致している。ただし、類縁度の低いゲノム間ではシンテニーが保存されている領域は限られてくるため、共通に保存されている構造を抽出すること自体が生物学的な意味を持ちうる。これは、特に水平移動が多く存在して、近縁種間でも一般に著しい多型性が見られる原核

生物ゲノムの解析においては本質的な問題と考えられるため、これを「コアゲノム構造」と呼んで重点的に取り組んだ。

以上のような検討課題をふまえて、これらの成果を取り込んだ包括的な比較ゲノム解析ワークベンチ RECOG を開発した。RECOG の開発は、データベースの構築、管理、検索等を行うサーバと、サーバと通信して解析結果を表示するクライアントとに分けて行い、このうちサーバは MBGD とデータベースを共有する形をとったため、開発の一環として MBGD の拡張も併せて行った。また、クライアントはオーソログ解析実行とその結果の表示を行うが、その際網羅的なオーソログ分類結果から得られた大規模な系統パターン解析機能の実装とその活用方法についても検討した。

3-3. 研究成果概要

1) オーソログ分類プログラム DomClust の開発と改良

これはすでに MBGD の一部として実装されていたが、プログラムの性能評価と論文発表、単体プログラムの公開を本プロジェクトの一環として行った。また、内群種・外群種を指定した際のオーソログ判定を行うよう改良した。その際、内群・外群間のオーソログに対して、内群系統固有の重複により生じたインパラログは、サブクラスターとして階層的な関係として出力される。

2) 微生物ゲノムデータベース MBGD の拡張

MBGD 上に利用者の持つ独自のゲノムを追加できる MyMBGD 機能を実装し、公開済みのゲノムデータに非公開のデータを合わせて比較解析を行えるようにした。また、COG, KEGG 等で定義された機能カテゴリ情報を取り込んでアノテーションに利用できるようにしたほか、オーソロググループ内の各遺伝子の記述を要約して、グループ全体のアノテーションとする機構を組み込んだ。

3) 近縁ゲノム比較ツール CGAT の開発

RECOG はアミノ酸レベルの比較を対象としているが、ごく近縁のゲノム比較では塩基配列レベルの比較が必要になるので、補完的なツールとして CGAT を開発した。CGAT はドットプロット表示とアライメント表示の組み合わせによって、ゲノム構造変化の詳細な観察を容易にするツールで、各ゲノム上に種々の情報を表示して比較する機能を持っており、ゲノム多型と特定の構造との関連等を調べるのに利用できる。

4) 近縁ゲノムのコア構造抽出手法の開発

中程度の類縁度のゲノム間で保存されたコア構造を抽出するアルゴリズムを開発した。これは、ゲノム上近傍関係が保存されたオーソログ対で構成されたグラフを用いて定義した。この手法をバチルス科、腸内細菌科のゲノムに適用して、抽出されたコア構造についてコア構造らしさを評価した。特に GC 含量や系統樹解析から、コア遺伝子が主に垂直的に伝搬してきたという仮説を裏付ける結果を得た。

5) 比較ゲノムワークベンチ RECOG の開発

以上のような成果を基に比較ゲノムワークベンチ RECOG を開発した。RECOG では、生物種セットを選択してオーソログテーブルを作成し、これを基に系統パターン解析やコア構造アライメントなどの解析を行える。生物種選択の際は、興味のある生物種を中心として内群、外群を設定でき、結果が入れ子型のテーブルとして表示される。作成されたオーソログ解析結果は、系統パターンビューアによって表示され、並べ替えやフィルタリングなどを通じて解析できる。

4. 事後評価結果

4-1. 当初計画の達成度

○当初の研究開発計画から見た進捗状況や達成度等

ゲノム比較を行うためのオーソログ分類の改良や補助ツールの開発が行われ、包括的な微生物ゲノム比較を可能とする微生物ゲノムデータベースが開発された。

○当初計画では想定されていなかった新たな展開

計画後にさらに多くの生物種のゲノムが明らかにされたので、近縁ゲノム比較ツール CGAT が開発され、近縁種の比較も同時に行えるシステムの開発につながった。

4-2. 知的財産権、外部発表（論文等）等研究開発成果の状況

○知的財産権

微生物ゲノム比較に関する数多くのデータベースやプログラムが開発され、着実に公開が行われている。

○外部発表（論文等）

ワークベンチの開発というテーマの関係上、論文発表数は多くないが、適切に論文発表されている。

4-3. 研究開発成果の公開による波及効果

○データベース、プログラムや論文、学会発表など、これまでの成果公開によるバイオインフォマティクス領域およびライフサイエンス分野全体への影響や効果について

データベースは公開されており、すでにユーザの利用に供されている。開発されたワークベンチは公開されているが、まだ間がないこともあり、ツール等が比較解析を行う上で、どのように効果的に機能するか検証することが期待される。

○国内外の類似研究開発の動向と比較して本課題の意義

海外の同種のデータベースの更新が停止されている中で、急激な微生物ゲノムデータベースの増大に対応した比較ゲノムデータベースを維持してきたことに大きな意義がある。

4-4. 成果の実用化の可能性及び成果から予想される波及効果

○今後、成果の生命科学、研究基盤、産業への貢献度

微生物ゲノム研究の基盤データベースとなる可能性があり、微生物の基礎研究や有用遺伝子発見等の用途で利用価値が高いと期待される。

4-5. 総合評価

研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。日本発の大規模微生物ゲノム比較データベースとしての存在価値は大きい。微生物ゲノムについては、次世代シーケンサーによるデータ、メタゲノムデータの大幅な増加が見込まれる。比較ゲノムを維持していくことがきわめて重要である。

5. 主な論文発表等

Uchiyama, I., Hierarchical clustering algorithm for comprehensive orthologous-domain

classification in multiple genomes. Nucleic Acids Res. 34, 647–658, 2006.

Uchiyama, I., Higuchi, T., Kobayashi, I.: CGAT: a comparative genome analysis tool for visualizing alignments in the analysis of complex evolutionary changes between closely related genomes. BMC Bioinformatics, 472, 2006.

Uchiyama, I.: MBGD: a platform for microbial comparative genomics based on the automated construction of orthologous groups. Nucleic Acids Res. 35, D343–D346, 2007.

Uchiyama, I.: Multiple genome alignment for identifying the core structure among moderately related microbial genomes. BMC Genomics, in press.

参考

論文発表 国内 0 件、海外 4 件

口頭発表 国内 2 件、海外 0 件

ポスター発表 国内 3 件、海外 4 件

公開されている成果プログラムおよびデータベース等

- ・ MBGD

<http://mbgd.genome.ad.jp/>

利用者が生物種を選んでオーソログ解析を行える機能を持った微生物比較ゲノムデータベース。

- ・ DomClust

<http://mbgd.genome.ad.jp/domclust/>

ゲノム中の遺伝子間の総当たりのホモロジー検索結果を用いて、独自に開発された階層的クラスタリングアルゴリズムによって、ドメイン単位のオーソロググループを作成するプログラム。

- ・ MyMBGD

<http://mbgd.genome.ad.jp/MyMBGD/>

公開ゲノムデータに利用者のゲノムを加えた上でオーソログテーブルを作成し、MBGD のインターフェイスを通して比較解析を行えるようにするシステム。

- ・ CGAT

<http://mbgd.genome.ad.jp/CGAT/>

近縁ゲノム間のアライメントに基づく比較ゲノム解析ツール。アライメントの計算やデータの管理を行うサーバと、ドットプロットとアライメント表示とを組み合わせたビューアとから構成され、特にゲノム上の繰り返し構造とゲノム多型との関連を詳細に調べるのに適している。スタンドアローンのプログラムとして配布しているほか、ビューアは MBGD の機能の一部としても利用可能。

- ・ CoreAligner

<http://mbgd.genome.ad.jp/CoreAligner/>

類縁ゲノム間の遺伝子の並び順の保存性に基づいてコア構造を構築するプログラム。MBGD のクラスタリング結果と MBGD からダウンロードしたゲノムデータを入力として実行する。

- ・ RECOG

<http://mbgd.genome.ad.jp/RECOG/>

MBGD を発展させて作成された比較ゲノムワークベンチ。クライアント・サーバ型ソフトウェアとして動作する。基本的な使い方は MBGD と同様に、生物種を選択してオーソログクラスタリングを実行して、その結果を基にして様々な比較解析を行う、というものであるが、系統パターンマップの全体像の表示、内群・外群を指定したオーソログクラスタリング (DomClust) の実行と表示、コア構造プログラム (CoreAligner) の実行と表示などの解析機能を含んでいる。