

氏名	藤 潤 航
学位(専攻分野)	博士(理学)
学位記番号	論理博第1340号
学位授与の日付	平成10年5月25日
学位授与の要件	学位規則第4条第2項該当
学位論文題目	ゲノム間比較に基づく遺伝子クラスターの解析

(主査)

論文調査委員 教授 金久 實 教授 岡 穆宏 教授 西田 栄介

論文内容の要旨

分子生物学の実験技術の進歩により、多くの生物種の全ゲノム配列が決定され、大量の配列データがデータベースに蓄積されるようになった。一方、分子生物学関連のデータベースの種類も増加し、大量かつ多様なデータベースから、いかに生物学的に意味のある情報を取り出すかが問題となっている。そこでの重要な概念は、データ間のリンク(関係)である。実際、分子間の反応やシグナル伝達の経路など、殆どどの生物学的現象は、分子間相互作用という関係に基づいている。一方、配列データベースの相同性(ホモロジー)検索とは配列類似性という関係をもつデータを検索していることであり、WWW(World Wide Web)におけるハイパーリンクも意味的につながりのあるデータとの関係を表現するのに用いられている。本論文では、様々なタイプのリンクを用いて分子生物学データベースを統合化し、またゲノム構造の比較解析を行ったものである。

まず分子生物学データベースの統合に関して、本論文は次のような基本構造を考案した。(1)あらゆる分子生物学データベースをエントリーのレベルで統合し、同じ検索方式でどのデータベースからもエントリー検索が可能であること。(2)エントリー間の関連情報であるリンクは全て抽出し、リンクを次々と辿ることで新たなリンクを生み出せるようにすること。これらの構造を基礎において、DBGET/LinkDB統合データベース検索システムを作成した。リンク情報は、もともとデータベースに書かれているエントリー間の相互参照リンク(ファクトリンク)だけでなく、生物学的な現象として関連のあるエントリー間のリンク(生物学的リンク)や相同性検索の結果を結び付けたリンク(相同性リンク)を組み込んで、統合的なリンク検索ができるようになってきている。このうち、生物学的リンクは特に、KEGG(Kyoto Encyclopedia of Genes and Genomes)プロジェクトで進めているデータの蓄積に依るところが大きい。

次に統合リンク検索に基づく生物学的な応用例として、本論文では全塩基配列が決定された10生物種のゲノムで、種内および種間で保存される遺伝子クラスターの解析を行った。ゲノム上で対応する遺伝子を抽出する過程ではDBGET/LinkDBと同様の統合リンク検索の手法が用いられている。この解析の結果では、種間の違いを越えて保存された遺伝子クラスターの存在が確認され、さらに近縁種の*Mycoplasma pneumoniae*と*Mycoplasma genitalium*の比較では、ゲノム量の増幅が遺伝子クラスターの増加に起因していると考えられることが示された。また、各生物種において*Escherichia coli*ゲノムとの比較から抽出したクラスターを原著者らの予測したオペロンに照らし合わせたところ、オペロン内の遺伝子群と完全または部分的に一致した遺伝子をもつクラスターが多く、その数が80%にまで達する生物種もあった。この一致度は相手のゲノムが小さいほど、また遠縁の種になるほど高い値を示す傾向があり、オペロン予測を行う上で重要な手掛かりとなる可能性を示した。さらに、抽出した遺伝子クラスターの機能の特徴を*Methanococcus jannaschii*との比較から抽出したクラスターで行ったところ、予期されたトランスポーター関連クラスター以外にも、エネルギー代謝や転写制御、DNA複製など非常に多様なクラスターが抽出できていることがわかった。

論文審査の結果の要旨

ゲノム解析計画の進展により、ここ数年の間に微生物の全DNA配列が次々と決定されてきている。しかしながら、実験で確かめられる遺伝子の機能のデータには限りがあるので、それらの大量データを計算機科学的にどのように意味付けを行うかがまさに現実的な問題として取り上げられている。この問題には、多様な既存のデータベースをいかに有効にまた統合的に利用するかと、そこからいかに新たな知識を獲得するかが焦点になっている。本論文では、生物学のデータ解析のもっとも基本となることはデータ間のつながり、すなわちリンクの概念であるとし、2つのテーマでその実践を行っている。1つは多様化する生物学の知識を統合的に検索できるようなシステムの開発であり、もう1つは種間比較に基づくゲノム構造、とくにオペロン構造の解析である。一見異なるテーマではあるが、概念的にも方法的にも共通の枠組みで研究が行われており、全体としてリンクの概念の実用性を示した論文となっている。

まず、データベースの統合方法について、既存の異なるデータベースが異なる見方で生物学データのコンピュータ化を行っているため、データの内部構造も含めた統合化は現実問題として不可能である。そこで、本論文ではエントリーとエントリーが関連するかどうかだけに着目し、関連するならリンクを張るといった最も簡略化された手法を用いた統合リンク検索を開発している。リンクされたエントリーに書かれたデータの意味を解釈することは利用者にまかされていることになるが、この簡略化により抽象化レベルの高い生物学の知識でも関連付けを行えることになり、また三段論法のようなリンク計算ができることから、その実用的な価値は高い。本論文中では17種類の既存のデータベースを統合し、さらにDNAやアミノ酸配列間の類似性検索もリンクとして取り込み、代謝系上の酵素やシグナル伝達と言った分子間の生物学的な関連も組み込むことが可能となった。また、データの内部構造の統合までを行う従来の関係データベースシステムなどと比較して、すべてのデータベースの日々更新が現実的な時間でできることが示された。本研究の成果であるDBGET/LinkDBシステムはゲノムネットの基本システムとして、世界中の研究者に広く利用されていることは特筆すべきことである。

次に本論文では、統合リンク検索をゲノムデータベースに適用し、種間比較による遺伝子クラスターの抽出を行っている。遺伝子クラスターとは、機能的に関連した複数遺伝子がゲノム上の連続した領域にコードされているグループである。既知のグループを元にして、新たに配列が決定された遺伝子の機能を、グループ単位でまとめて意味付けをすることにも繋がるものである。本論文では、ゲノム上で隣接する遺伝子間の関係を示すリンク情報と、種間の類似遺伝子の関係を示すリンク情報とを、DBGET/LinkDBと同じような統合リンク検索によってうまく検索することで、自動的に遺伝子クラスターを抽出している。また、この検索により抽出された遺伝子クラスターが生物学的にどのような意味を持つのかの解析も行っている。その結果、種間比較によって抽出される遺伝子クラスターはオペロンの部分構造となっている場合が多く、今後オペロン予測および遺伝子クラスターの機能予測を行う方向性への重要な足掛かりを残した点がとくに評価できる。

以上により申請者の統合データベース検索システムの開発と統合ゲノム解析の研究は、ゲノム塩基配列から遺伝子の機能予測を行うための基礎となるものであり、高く評価できる。よって、本論文は博士（理学）の学位論文として価値あるものとして認める。なお、主論文および参考論文に報告されている研究業績に関連した試問の結果、合格と認めた。