

氏名	上野国光	学籍番号	250010
主テーマ指導教員	佐藤賢二	副テーマ指導教員	ホーツーバオ
主指導教員	小長谷明彦	副指導教員	佐藤賢二

<研究テーマ>

原核生物オペロンのマイクロアレイデータベースシステムの構築とそれによるオペロン同定法の研究

<研究の目的>

原核生物は一つの mRNA により複数の構造遺伝子（シストロン）を同時に発現するオペロン転写機構を持つが、真核生物は持たないとされ、オペロンが生物進化を考える上で重要な手がかりとなることが期待される。近年、多数のゲノム配列が解明され、特徴的な配列情報から盛んにオペロン予測がなされているが、実験的負担から未検証のままとなっている。一方、ゲノムの全遺伝子をブローブとして、発現している遺伝子を網羅的にとらえるマイクロアレイ解析も増加しており、そのデータも着々と公開される状況となってきた。そこで、本研究では、ゲノム配列から候補オペロンを予測し、その構造遺伝子とマイクロアレイの発現データとを一括管理できるデータベースシステムを構築し、アレイ実験データによるオペロン同定法の研究支援環境を開発することを第一の目的とする。次にこれを用いて、大腸菌によるオペロンの同定を試みる。

<研究の背景・特色>

現存生物は、大きく原核生物と真核生物に二分される。原核生物は、一つの mRNA により複数の構造遺伝子（cistron）を同時に発現するオペロン構造をもつが、真核生物はもたない、とされてきた。原核から真核への進化を考える上で、この転写機構の違い（policistronic vs monocistronic）が注目される。

原核生物のオペロン構造は、1961年に Jacob と Monod によって、大腸菌ラクトース代謝系で初めて見出されて以来、長期にわたり仮説としての地位にあった。しかし、現在では、古細菌も含めた原核生物において普遍的な転写機構と考えられるようになった。また、近年、多くのゲノム配列が解読されたことにより、オペロン構造の塩基配列等の特徴を用いてゲノムワイドにオペロン予測も行われるようになった。その結果、大腸菌では予測オペロン数が630～700で全遺伝子の60%以上を占めるという報告もある。しかしながら、実験の負担から実験系で実際に確認されたオペロンは意外に少なく、その意味で、いまだにオペロン構造の存在を疑っている学者もいる。そこで、これら多数の予測オペロンの存在を実証する効率的な方法が求められている。

一方、DNA マイクロアレイは、1995年米国 Stanford 大の Brown らにより開発されて以来、ゲノムの全遺伝子をブローブとして、ある環境下の細胞（組織）で発現している mRNA の発現強度を網羅的に測定できる極めて効率的な実験系である。その測定データも年々増加しており、データ公開も行われる状況になってきた。オペロン構造の存在を検証するためには、原則的には、オペロン構造遺伝子がこのアレイデータ上で共発現していることを示すことが必要であるが、残念ながら、①候補オペロンとアレイ発現データを結合したデータベースがないこと②単独種のアレイデータでは、統計解析に必要なデータ量が確保できないこと（測定データのバラツキ）③各構造遺伝子のサンプルとブローブのハイブリダイゼーション（結合反応）のバラツキ等の問題があり、今のところ成功していない。そこで、まず①②の問題を解決するため、ゲノム配列が解読されアレイデータの豊富な大腸菌をモデルケースに選び、ゲノム配列から候補オペロンを抽出し、その構造遺伝子と公共のアレイデータベースから収集した遺伝子発現データとをリンクしたデータベースシステムを構築し、統計解析機能を加えたオペロン同定法の研究支援システムを開発する。次にこのデータベースシステムを用いて大腸菌のオペロン同定を試みる。

<研究計画・方法>

1、データベースシステムの構築

以下に手順を示す。モデル生物として、大腸菌を選択する。

1) 全塩基配列の取得

公共の配列データベースである GenBank または TIGER Microbial Genomes から、全塩基配列を取得する。

2) 候補オペロンの抽出

オペロン構造の配列特性（隣接、同一転写向き等）を利用して、全塩基配列データから広く候補オペロンの構造遺伝子（ORF）を抽出する。多少の false positive は許容する。

3) マイクロアレイ発現データの収集

Wisconsin 大のデータベース（wisc.u）を中心に、Medline の関連論文を手がかりにして遺伝子発現データを収集する。同時にアノテーション情報も収集する。

4) アレイデータの前処理と構造遺伝子発現データベースの作成

オリジナルデータと前処理データ（発現量の閾値によるデータ削除、missing value の処理等）を保存し構造遺伝子のみの発現データベースを作成する。

5) 統計解析機能とクラスタリング機能の付加

・ データの層別機能

① 機能別

既存のアノテーションを参考にオペロンの生化学的機能と生物学的（生理的）機能による層別（オペロンの機能別分布）

② 検証済みオペロンと未検証オペロンの層別

③ オペロン構造遺伝子数による層別

・ 構造遺伝子間の発現強度の相関解析

・ 統計解析（t-検定、分散等）

・ クラスタリング機能：階層型、非階層型（k-means,SOM）

2、大腸菌のオペロン同定トライアル

構築したデータベースシステムを使用して、候補オペロンの発現解析を実施する。

・ 構造遺伝子間の発現相関

上記層別を組合わせて、構造遺伝子間の発現相関をみる。

・ 補正法の検討

構造遺伝子の発現強度のバラツキ要因（文献情報）

① 構造遺伝子の塩基数と配列特異性によるハイブリダイゼーション効率

② クロスハイブリダイゼーションの影響

③ mRNA の分解速度の DNA 上の位置との関連

等を参考に、補正法を検討し、オペロン同定を試みる。

・ 同定精度の評価

検証済みオペロンのアレイデータを補正し、クラスタリングで各オペロンの構造遺伝子の切り分け精度を確認する。

<現在までに単位取得した専門教科>

概論：3科目、方法論：2科目、専門：3科目

K114 計算の科学B

K418 知識表現論

K217 物理科学概論B

K215 インノベーション概論B

K417 知識創発論

K616 生命知識特論

K216 知能科学概論A

K414 複雑系解析論

015 国際特許法

K213 システム科学方法論

K212 知識ベース方法論B