

「ゲノムの構造と機能」
平成10年度採択研究代表者

森 浩禎

(奈良先端科学技術大学院大学遺伝子教育研究センター 教授)

「大腸菌におけるゲノム機能の体系的解析」

1. 研究実施の概要

研究のねらい：ヒトゲノムの決定終了が間近に迫っている現在においても、機能未知遺伝子群の機能解明、遺伝子ネットワーク解明、病原性および有用微生物の有効利用等において大腸菌研究の重要性は明らかである。このような現状を踏まえ、ゲノム生物学からの大腸菌研究を推進することで一生物としての大腸菌の完全理解を目標とし、今後のゲノム生物学の基礎を築こうというものである。

これまでの研究の概要：そのためには、1)大腸菌ゲノム配列解析より明らかになった情報の解析および整理とデータベース化、2)機能未知遺伝子群の網羅的な機能解析のための研究材料の構築、3)この材料を用いた網羅的機能解析と情報科学的解析との連携、を目標とした研究開発を行った。

成果：上記研究開発の結果、以下の成果を得ている。

- 1)ゲノム配列解析方法の改良と自動化、今後の網羅的機能解析の結果を広く公開するためのホームページの開設とそのためのシステム開発の完了。
- 2)研究材料の作製は a)全予測遺伝子のクローン化の完成とそれを利用したDNAチップの完成、b)全遺伝子の網羅的破壊株作製方法の完成。この方法を用いた破壊株作製は全遺伝子のほぼ1/3を終了。c)これまで必須遺伝子と考えられてきた領域以外の約150箇所の欠失株作製から半数の領域に新規必須遺伝子の存在が示唆され、残る半数の欠失の完了、d)質量分析計を利用したRHFRタンパク質二次元電気泳動法によるgene-protein indexを140スポットについて終了。
- 3)実験系と情報系との連携において、a)DNAマイクロアレーからの情報を元に遺伝子ネットワーク解明のためのソフト開発、b)コドン利用頻度を利用した遺伝子産物の分類、c)質量分析計からのデータ解析システムの開発、を行った。

今後の見通し：全遺伝子のクローン化の完成により、そのクローンを利用した新たな解析システムの構築が可能となった。現在、細胞分裂等に関する変異(広田コレクション)同定の系の確立を進めている。DNAマイクロアレーの完成により、

網羅的な転写解析が可能となり、今後はクラスター解析の結果を利用しながら転写因子等による発現調節の全体像の解明が短期間に可能と考えられる。破壊株、欠失株を利用し、遺伝子ネットワーク解明へ向けた解析を進める予定である。破壊株は2000年中の完成を予定しており、網羅的な機能解析の材料とする。同時に広く公開を行い、機能未知遺伝子の機能解析を加速させていきたい。

2. 研究実施内容

目的：本年度の研究開発は a) DNAマイクロアレー作製、目的遺伝子のタンパク質精製、個々の遺伝子研究を目的に全予測遺伝子のクローン化、b) 各遺伝子の機能解析および必須遺伝子の同定を目的とする網羅的破壊株と欠失変異株の作製、など今後のシステムティック機能解析に向けた研究材料の作製とその完成を第一の目的とする。次いで、大腸菌ゲノムデータベースの構築と公開のためのシステム開発およびホームページ開設を目的とする。

方法と結果：研究開発を大きく4つ、材料、情報解析、データベースそして網羅的機能解析、に分けて進めているが、今年度は材料開発に特に集中して行った。材料開発は a) 全遺伝子のクローン化とDNAマイクロアレーの作製、b) トランスポゾンを用いた網羅的破壊株の作製、c) 非必須領域の欠失株作製、d) 質量分析計を用いた網羅的タンパク質インデックスの作成を進めてきた。以下にそれぞれの研究開発の方法と結果を示す。

1) 材料開発

(ア) 全遺伝子のクローン化およびDNAマイクロアレー作製：クローン化のためのベクターの開発を行い、各遺伝子のクローン化は小原クローンをテンプレートとしてPCRにより目的遺伝子を増幅し、クローン化を行った。図1にベクターの構造とクローン化の方法を示す。PCR増幅断片はアガロースゲル電気泳動による分離・精製を行い、末端の平滑化を行った。1999年4月よりクローン化を始め、同年12月に全遺伝子のクローン化を完成した。2000遺伝子のクローン化が終了した段階で宝酒造株式会社と共同で本クローンを利用した大腸菌DNAマイクロアレーの作

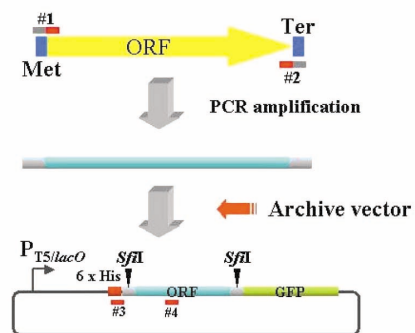


図1 ORFのクローン化とベクター

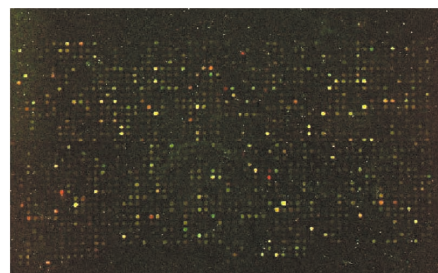


図2 DNAマイクロアレー

製を行い、評価を含めて網羅的転写解析を開始した。現在は大腸菌全遺伝子を打ったDNAマイクロアレーも完成し、それをを用いた転写解析を行っている。DNAマイクロアレーのイメージを図2に示す。

- (イ) トランスポゾン挿入による網羅的破壊株作製：小原クローンに、ランダムにmini-Tn10(Km)を挿入し、これを大腸菌に感染させ、ファージ上のTn10挿入遺伝子を宿主ゲノム上の相当遺伝子と置き換えて、目的の破壊株の作製を行う(図3)。

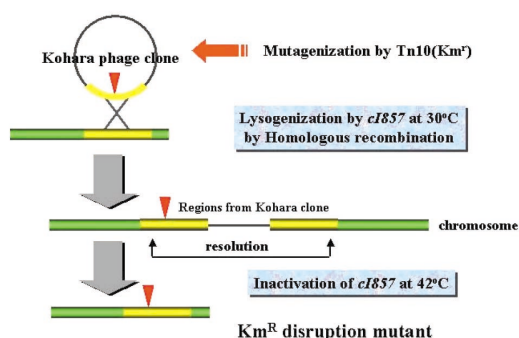


図3 網羅的破壊株作製

具体的には、新たに構築したmini-Tn10(Km)を有するプラスミド、pKP2371をTn10の供給源とする。このプラスミドを保持する大腸菌に小原クローンを感染させて、溶菌液を回収する。この溶菌液には、プラスミドからTn10が転移したファージが含まれる。この溶菌液を大腸菌に感染させ、Km耐性の溶原菌(シス部分2倍体)を分離、そこからKm耐性の非溶原菌(一倍体)を分離すると、これが目的の破壊株である。ここで用いるファージは、cIリプレッサー欠損の為に溶原化できないが、宿主側に高温感受性cI857リプレッサーを持つmini-Rプラスミドを導入しておくことにより低温で溶原化が可能となる。破壊された遺伝子の同定は、Tn10挿入部位を挿入点近傍の配列を決定する事により行う。また、このKm耐性菌の分離不可能な遺伝子は、必須遺伝子の可能性が高く、高温感受性のmini-Fベクターを利用した相補の系を利用して非溶原菌の分離を行い必須性の確認を行う。パイロット実験の結果、大腸菌染色体4-6分の112個のORFの内、104個のORFの変異株分離に成功した。変異株が得られた104個の中で、24個は必須遺伝子、71個は非必須遺伝子で、8個は現時点では必須性の判断は不能であった。必須遺伝子と判定された24個の遺伝子の中、4個は新規の必須遺伝子であり、3個は従来必須ではないと判断されていた遺伝子であったが今回新たに必須であることを明らかにできた。また、変異株が得られなかった8個の遺伝子の中、2個はtRNA遺伝

子、1個はIS配列上の遺伝子であり、蛋白質をコードすると考えられる遺伝子は5個であった。以上の結果より、本システムは系統的な大腸菌全遺伝子の破壊に使用可能なものとして、現在、残る染色体領域における遺伝子破壊株の作製を進めている。今年度において約1/3が終了しており、2000年10月を目処に破壊を進めている。

- (ウ) 非必須領域の欠失株作製:系統的、網羅的に欠失株を作製することにより、細胞の増殖にシスまたはトランスに必須な遺伝子領域の同定及び機能未知非必須遺伝子の機能解析の材料とする。方法は

欠失させたい領域の両側約1.5kbずつ(A, B)を、順番にベクター(664BSCK2)にKm耐性遺伝子をはさむ形でクローニングする。

作製したプラスミドをMG1655 rpsL polA12株に導入する。この株の中では、42において664BSCK2プラスミドは複製できないので、Cm耐性のコロニーを選択することによりプラスミドが染色体に、AまたはB領域の領域で相同的組み換えにより挿入されたものを単離することができる。

さらに35で培養を続け、Km耐性さらにはSm耐性かつCm感受性のコロニーを選択することにより、プラスミドが挿入された状態からもう一度相同的組み換えを起こし、染色体に欠失変異を起こし、プラスミドは染色体外に出て、さらには細胞の中からプラスミドがいなくなったものを得ることができる。

すでに知られている大きな欠失変異(10箇所、計約970 kb)のある領域については欠失株を作らず、既知の必須遺伝子の存在する領域についても欠失株の作製を行わない。欠失できなかった領域については、その領域またはその領域内の遺伝子を持つプラスミドで相補させた状態で欠失株を単離し、必須遺伝子の同定を行う。この方法で(1)156個の欠失株を作製するためのプラスミドの作製、(2)作製したプラスミドによる欠失の結果、72個の欠失株作製、を行ったが、84個の領域に関しては欠失ができず、未知の必須遺伝子の存在が示唆された。

- (エ) 質量分析計によるタンパク質インデックス作製:大腸菌ゲノムの全塩基配列から予想されるORFのうち、機能未知のORFが50%近く存在する。種々の条件のもとで発現されてくる蛋白質を系統的に解析し、機能分類すると同時にRHFR二次元電気泳動に基づくgene-protein indexを作成し、そのデータベース化を目指した。今年度は対数期で比較的多く発現されている蛋白質140個のgene-protein indexを作成することを目標に行った。方法はタンパク質の泳動後、スポットを切り出し、タンパク質分解酵素による処理の後に質量分析計にかけ、その正確な質量を求める。そのデータをもとに、大腸菌ゲノムから

予測されるORFの質量データベースとの比較検索により、遺伝子とタンパク質スポットとの対応を取る。この方法で大腸菌に熱ショックを与えた時に発現が増加又は減少する蛋白質群の同定を行った結果、DNAマイクロアレイによる転写レベルでの遺伝子発現とは異なった蛋白質の増減が見られた。これはDNAマイクロアレイによるtranscriptomeの解析を補完するものである。

2) 情報解析

- (ア) ゲノム配列解析システムと遺伝子ネットワーク解析のためのシステム開発：大腸菌ゲノムデータベース構築を目的とした解析自動システムの開発を行った。このシステムを利用し大腸菌ゲノムデータベースの構築を行った。現在進行しているDNAマイクロアレイを利用した網羅的転写解析のデータを利用し、遺伝子ネットワークの解明を行うシステム開発を行っている。
- (イ) コドン組成を利用した遺伝子の分類：類義語コドンの選択は、合成される蛋白質の構造に影響を及ぼさないにもかかわらず、種固有のコドン利用特性が存在することが知られている。自己組織的にコドン使用により遺伝子进行分类し種固有のコドン使用特性を把握することを試みた。競合学習型ニューラルネットワークとして知られているコホネンの自己組織化マップ法 (Self-Organizing Map : 以下 SOM とする) では、データの入力順序により学習後作成されるマップが異なる。そこで入力順序に依存しないSOM法 (Batch-Learning SOM : 以下BLSOMとする) を開発し、この方法を用いて大腸菌及び全塩基配列が決定されている16種のバクテリアの遺伝子に対し、コドン利用特性に基づいた遺伝子の分類を行い、本方法の評価を行った。その結果、これまでのSOM法と比較しての有効性を確認できた。

3) データベース

- (ア) ホームページの開設とシステム開発：大腸菌ゲノム配列より予測されるORFを中心にゲノムデータベースの構築を行ってきた。データベースシステムとしてはリレーショナルデータベースシステムを採用し、管理を行っている。公開のシステムはUNIXワークステーション上にWWWサーバーの apache を利用したホームページを構築し、データベースとの連携はHTML、Javaを中心としたCGIプログラムにより実現させた。今後は機能解析などの実験結果も含めて公開を行うためのシステム開発を行う。
- (イ) 変異データベースの構築:大腸菌の遺伝学的研究に関する文献から始めて、大腸菌研究の文献を網羅し、検索できるデータベースを作成し公開する。そのために、"Escherichia coli and Salmonella typhimurium"のBerlyn et al.の引用論文(約5000報)から出発して、それぞれの文献をOCRでデジタル化し、タイトル、著者名、要約、引用論文を抽出してまとめる。その一方、公開のため

のweb pagesならびに必要なCGI-programsを作成する。その結果、これまでに、約2,700報の文献をデジタル化し、データベースとして整備し、公開の準備を行っている。

4) システムティック機能解析

- (ア) タンパク質非コード領域に関するシステムティック解析：タンパク質をコードする遺伝子以外の領域のシステムティック機能解析を目的に進める。RNA関連遺伝子を中心に解析方法の確立を行うために、(1)RNA関連遺伝子の「数」と「位置」について、細胞機能全体からの意義を把握、(2) tRNAや機能未知の低分子RNAについて新しい機能の検索を行い、最終的には全細胞機能における役割の把握を目指す。

遺伝子間領域 (inter genic region (IGR)) における偽 tRNA遺伝子や t m RNA様遺伝子の網羅的検索を行った。(i) tRNAの3'末端部分からGTTCアームまでの半分子を一つのtRNAモチーフとして、全IGRについて検索した。その結果、2種類の偽tRNA遺伝子、2種類 (Arg, Thr) の半tRNA分子の配列を見出したが、発現は確認できていない。(ii) 大腸菌ゲノムの全塩基配列からすべてのIGRを取りだし、大きいものから100個のIGRについて、「くりかえし構造」などの詳細な検索を行った。その結果、13個についてユニークなrepeating配列が見出されたが、その機能は不明である。(iii) 大腸菌ゲノム上のtRNA遺伝子の欠失、あるいは破壊を行った。これまでにLeu6、Ser2、Thr2、Leu1クラスターについて成功した。現在、Gly1、Arg4、Gln2、Pro1などのtRNA遺伝子について実験が進行中である。(iv) 大腸菌ゲノム上のある遺伝子やIGRを欠失させる必要のため、全遺伝子のクローンとcre/loxP系を利用した方法の開発を行った。

- (イ) タンパク質コード領域：破壊株を利用した網羅的・システムティックな機能解析の方法論の確立を行っている。現在、大量のサンプルの処理を目的に開発されたロボットシステム、Biomec2000を利用したアッセイ系の確立を簡単なテスト (糖代謝機能等) を利用して行っている。

3. 主な研究成果の発表 (論文発表)

Itoh T., Okayama T., Hashimoto H., Takeda J., Davis RW., Mori H. and Gojobori T. A low rate of nucleotide changes in Escherichia coli K-12 estimated from a comparison of the genome sequences between two different substrains, FEBS Letters 450, 1-2, 72-76, 1999

Itoh T., Takemoto K., Mori H. and Gojobori T. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes, Mol Biol vol.16, No.3, 332-346, 1999

Itoh T., Matsuda H. and Mori H. Phylogenetic analysis of the third hsp70 homolog in *Escherichia coli*; a novel member of the Hsc66 subfamily and its possible co-chaperone. *DNA Research* 6, No.5, 299-305, 1999

Kanjo N. & Inokuchi H., Genes for tRNA Arg located in the upstream region of the Shiga toxin II operon in Enterohemorrhagic *Escherichia coli* O157:H7, *DNA Res.*6, 71-73, 1999

Meider R., Morad I., Amitsur M., Inokuchi H. & Kaufmann G., Detection of anticodon nuclease residues involved in tRNA^{Lys} cleavage specificity, *J. Mol. Biol.*287, 499-510, 1999

Narita S., Taketani S. & Inokuchi H., Oxidation of protoporphyrinogen IX in *Escherichia coli* is mediated by the aerobic coproporphyrinogen oxidase, *Mol. Gen. Genet.*261, 1012-1020, 1999

Guo L., Katayama T., Seyama Y., Sekimizu K. and Miki T., Isolation and characterization of novel cold-sensitive dnaA mutants of *Escherichia coli*. *FEMS Microbiol. Lett.* 176, 357-366, 1999

Takata M., Guo L., Katayama T., Hase M., Seyama Y., Miki T. and Sekimizu K., Mutant DnaA proteins defective in opening of oriC, the origin of chromosomal DNA replication in *Escherichia coli*. *Mol. Microbiol.*35, 454-462, 2000

Tohsato, Y., Matsuda, H. and Hashimoto, A., Application to Gene Cluster Analysis of Inductive Inference of Languages over Patterns with Conceptual Hierarchy, *Genome Informatics* No.10, 324-325, 1999

Stokes, A.J., Matsuda, H. and Hashimoto A., GXML: A Novel Method for Exchanging and Querying Complete Genomes by Representing them as Structured Documents, *IPSJ Transactions on Databases* Vol.40, No.3, 66-78, 1999

Stokes, A.J., Matsuda, H. and Hashimoto A., Making High-level Queries on Diverse Genome Data: A Structured Genome Document Database System based on GXML and GQL, *Genome Informatics* No.10, 76-185, 1999

Azam T. A., Iwata A., Nishimura A., Ueda S. and Ishihama A., Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid, *J. Bacteriol* 181, 6361-6370, 1999

Kanaya S., Yamada Y., Kudo Y. and Ikemura T., Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis, *Gene* 238, 143-155, 1999

Nakayama K., Kanaya S., Ohnishi M., Terawaki Y. and Hayashi T., The complete

nucleotide sequence fCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages, *Mol. Microbiol.* 31, 399-419,1999

Abe T., Kanaya S., Kinouchi M., Kudo Y., Mori H. and Matsuda H., C.D. Carpio, T. Ikemura, Gene classification method based on batch-learning SOM, *Genome Informatics Series No.10*, 314-315,1999