

大量 DNA シーケンシングと生命システム比較解明への応用

●小原 雄治

情報・システム研究機構 国立遺伝学研究所生物遺伝資源情報総合センター

<研究の目的と進め方>

ゲノム解析の進展は遺伝子からゲノムへというパラダイムシフトをもたらした。遺伝子の機能は、相互作用する相手（ゲノム環境）によって変わりうるのであり、その組み合わせの数は莫大になる。これがゲノムが高次かつ多様な生命現象をつくりだすことの基本である。このことは、ゲノムを単位として研究することにより、生命現象の素過程を統合し生物を形作り働かせる仕組み（生命システム）や生物個体、環境との相互作用により進化・多様化を生み出す仕組み（生物システム）の解明にアプローチできることを示している。

本研究では *C.elegans* 線虫という基軸生物のシステムを徹底的に明らかにし、近縁線虫のゲノム機能・発現と比較することにより、それら生命システムの構造、進化・多様化を解明することを目的とする。先端レベルの研究に基づく研究支援という基盤ゲノムの役割を明示するためにも、正面突破の研究を進めたい。

<2006 年度の研究の当初計画>

1) 線虫 *C.elegans* の解析

・線虫 *C.elegans* ゲノムについて発現パターン（mRNA、タンパク質）・機能パターンの徹底的解析を続け、発現パターンのクラスタリング解析などから、発生過程の遺伝子カスケード抽出をおこない、この結果をもとに発現制御領域の体系的同定をおこなう。さらに制御領域への結合タンパク質の同定を進め、関与する遺伝子群の全貌、さらには遺伝子カスケードの全貌解明をめざす。いくつかの現象、例えば初期胚における翻訳制御・局在化のメカニズム全貌の解明や、そのことによる細胞運命決定機構の解明もおこなう。

・これらの解析のために、発生パターン（細胞系譜や細胞配置など）及び発現パターンの自動解析ソフトウェアを開発・改良する。さらに、以上の結果をもとに発生のコンピュータモデル化/シミュレーションをおこなう。

2) 比較システム解析

・近縁線虫のうち、まず *Diploscapter coronatus* (100Mb程度) のドラフトゲノム配列を支援班において決定し、*C.elegans* とのオルソログ遺伝子を同定し、DNAチップや *in situ* ハイブリダイゼーション法、RNAi などにより、それらの発現パターン、機能パターンを比較する。

・D02との連携により完全長cDNAライブラリを作成し、トランスクリプトーム解析をおこない、*C.elegans* と比較する。

・以上により比較システム解析の基礎データを得る。

<2006 年度の成果>

1) NEXTDBの拡張

昨年度に引き続き、EST 配列・分類、ゲノムとのアラインメントによるゲノム構造、whole mount *in situ* ハイブリダイゼーションによる発現パターン解析（詳細なアノテーション付け）、発現パターンに基づく RNAi 解析、などの結果を NEXTDB で統合、公開を続けた。WormBase や WormGenes とともに直接リンクされている。URL は <<http://nematode.lab.nig.ac.jp>> これらをもとに

して、様々な遺伝子の機能研究や共同研究を遂行した。

完全長 cDNA についてできる限り網羅するために昨年度より開始した V-capping 法による胚発生から成虫期及び雄から cDNA ライブラリークローンの EST 情報整理を進めた。さらに代表クローンについて混合ショットガン法により内部配列の解析を進行中である。また、cDNA ライブラリ作成に用いた各期 mRNA をランダムヘキサマで cDNA 化し、Solexa シーケンサーで解析し、各期における全トランスクリプトーム情報を得た。ゲノム、遺伝子モデル、完全長 cDNA 配列との対応を解析中である。

2) 母性 mRNA の局在機構

線虫 *C.elegans* では受精後第 1 卵割において前後の細胞運命が決定され、引き続き卵割と細胞間相互作用によりさらに細かく細胞運命が決定されていく。この過程の主役は母性遺伝子である。*pos-1* 遺伝子は翻訳制御を通じて生殖細胞系譜の確立に働く母性遺伝子であるが *pos-1* mRNA は卵母細胞内には一様に分布するが、受精後第一卵割時に後極へ局在化し、その後の初期発生過程を通じて生殖細胞系譜へと局在化する。*Pos-1* タンパク質は mRNA と同じ分布を示すことも興味深い。そこで、我々はこの *pos-1* mRNA 局在化機構を解析した。

まず、局在化に必要な配列（シス因子）の解析のため内在性 *pos-1* mRNA の局在を再現する *in vivo* assay 系を構築した。この assay 系では VENUS::*pos-1* 3' UTR 融合遺伝子の形質転換体を作成し、VENUS プローブを用いた *in situ* hybridization で融合 mRNA の分布を調べる。線虫で外来遺伝子を母性発現させる事は困難であることから bombardment 法による形質転換条件の最適化により解決した。十分なシグナル強度を得るために、活性の高い *pos-1* プロモーターを単離、採用し、更に GC 含量の高い VENUS 配列をタグ配列とするなどの工夫をおこなった。この assay 系において、VENUS::*pos-1* 3' UTR 融合 mRNA は内在性 *pos-1* mRNA の局在を再現したが、他の母性 mRNA (*pal-1*, *mex-3*) の 3' UTR では、融合 mRNA は各々の mRNA の局在 (*pos-1* と異なる) を再現したことから、*pos-1* 3' UTR は局在化に必要な十分な活性を持つと判断出来た。*pos-1* 3' UTR について更に deletion 解析を行った結果、必須領域を 58 塩基の配列にまで絞りこむことができた。この配列は近縁種間でも保存されていた。以上の結果は、線虫の母性 mRNA の局在化のための領域を明らかにした初めての例である。トランス因子については candidate gene approach による解析を行い、*mex-5* 及び *mex-5*; *mex-6* 変異体では *pos-1* mRNA の局在が異常になる事を見出した。

3) 新しいモチーフ探索アルゴリズムの開発

C. elegans の遺伝子の細胞・時期特異的制御モチーフを見つけることを目的としている。コンパクトな *C.elegans* ゲノムでは遺伝子上流数 Kb の配列が特異的発現の制御に十分であることが多く知られているからである。そこで、私達は、陽性を残しつつ、探索する領域を狭めるアルゴリズム「filtering step」を開発した。人工・実際のいろいろなモチーフを検出可能であるかどうかをテストした結果、モチーフを想定以上の長い領域から見つけ出すこ

とに成功した。それに加え、このアルゴリズムは複数のモチーフを、同時に検出できることが分かった。本研究室の *in situ* hybridization database から時期・部位特異的に発現している遺伝子群に、この新たに開発したモチーフ予測プログラムを適用することで、時期・部位特異的制御に関わるモチーフを予測し、検証実験を行なうことでモチーフの同定に至った。

4) 温度感受神経AFDでの発現制御調節配列の系統的解析

本研究は細胞特異的な遺伝子の発現がどのようにして制御されているのか、その分子機構の解明を目指すものである。その第一歩として我々は、線虫 *C. elegans* の温度感受神経 AFD に特異的な発現を示す 10 種類の遺伝子を選び、これらの上流調節配列の系統的な欠損解析を行った。その結果これまでに、1) グアニリル環状化酵素 *gcy-8* の場合、AFD 特異的な発現を誘導するには 50 bp の調節配列で十分であること、2) 同じ発現パターンを示す遺伝子でも異なる機構によって制御される場合があること、3) 神経細胞特異的な発現制御にはエンハンサーだけでなくサイレンサーもまた重要な役割を果たすこと、4) 離れた位置にある複数のエンハンサーが同時に必要になる場合があることなどを明らかにした。

さらに我々は *gcy-8* を含む複数の遺伝子の AFD 特異的な発現が、ホメオボックス型転写因子をコードする二つの遺伝子 *ttx-1*、*ceh-14* の二重変異によって失われることを見出した。また精製した TTX-1、CEH-14 タンパク質を使ったゲルシフト解析によって、これらが *gcy-8* の 50 bp に結合することを明らかにした。続いて我々は AWB 化学受容神経に *ttx-1*、*ceh-14* を強制発現させることによって、*gcy-8* が本来の発現細胞である AFD に加えて AWB にもエクトピクに発現するようになることを示した。以上の結果は二つの転写調節遺伝子 *ttx-1*、*ceh-14* の両方が、*gcy-8* を含む AFD 特異的な発現に重要な役割を持つことを強く示唆する。

4) 線虫胚発生における細胞形状モデルの半自動生成システム

線虫 *C. elegans* は、透明で細胞数が少ないため、発生の過程を詳細に観察することができる優れた系である。発生においては細胞間相互作用が重要な役割を果たすので細胞の形状・配置・隣接関係の把握が非常に重要である。本研究では、細胞膜を蛍光染色して共焦点顕微鏡で観察した時系列 3 次元画像をもとに、計算機を用いて半自動的に細胞の形状モデルを構築するシステムの開発を進めてきた。

細胞膜に局在する GFP 融合遺伝子を発現する線虫株の 1 ~ 24 細胞期の画像データを多光子共焦点顕微鏡を用いて取得し、このシステムを適用して形状モデルを構築したが、これまでは細胞の形状がボックスで表現されていて画像のノイズによる細かな凹凸があったため、そのままでは表面積が求められなかった。そこで今回、形状を三角メッシュ表現に変換し、動的輪郭法を用いて滑らかな細胞形状を抽出する後処理システムを開発した。その結果、MS 細胞と接触しないために新しい運命に誘導されないと思われていた ABarp 細胞が実際には MS 細胞と接触していること、そしてその接触表面積が誘導が起こる MS 細胞と ABara 細胞の間の接触や MS 細胞と ABalp 細胞の間の接触よりも小さいことが明らかになった。今後 RNAi の表現型解析に応用する予定であるが、手始めに *par* 遺伝子の RNAi 胚との比較を進めた。

5) 線虫胚発生初期の力学モデル化

生物の発生過程では、細胞分裂時に形成される紡錘体の配置が重要な要因のひとつである。線虫 *C. elegans* の初期胚の 1 細胞期では、紡錘体が後極に偏って位置することで不等分裂が起き、正しく発生が進む。加えて、紡錘体は後極に移動しながら大きく振動する。紡錘体の非対称配置は Microtubule(MT) のダイナミクスと、細胞膜上に局在するモータータンパクが MT を引っ張るこ

とによって起こることを示す生物学実験の結果が報告されている。しかし、細胞膜上のどこで・どのような力が働いているかについての詳細はまだわかっていない。本研究では、どのようなメカニズムによって紡錘体が動くのかを調べる方法の一つとして、力の発生メカニズムのモデルと観測データを組み合わせた統計学的解析を行った。

まず共焦点顕微鏡を用いて 1 細胞期の紡錘体の動きを時系列画像として取得し、二つの中心体の軌跡 (二次元) を観測データとして抽出した。次に、細胞膜上のモータータンパクと紡錘体から広がる MT の負荷依存的な相互作用によって力を発生するモデルを構築した。ここで、構築したモデルと中心体の軌跡データによってモデルのパラメータを推定したいのだが、このような「システムの出力からシステムの内部状態を推定する問題」を一般に逆問題と呼び、解析的に解くことは難しい。そこで、統計学的に推定する方法の一つである「データ同化」を用いた。テストケースとして、現在知られているたんばく質の局在パターンから考えられる、細胞膜を二分割するモデルと三分割するモデルのどちらが観測データを説明しやすいのかをデータ同化によって推定した。その結果、三分割するモデルの方が観測データを説明しやすく、かつ推定された細胞膜上のダイナミクスは生物学実験から考えられたワーキングモデルと似た結果となった。その一つとして、*let-99* が胚の真ん中後ろよりに位置し引く力を弱めているという仮説を支持している。

<国内外での成果の位置づけ>

線虫 *C. elegans* トランスクリプトームについて世界のセンターとして、引き続き機能した。

線虫比較ゲノム研究は欧米で *C. briggsae* など 3 種がドラフト配列決定がなされ、計 5 種が決定間近であるなど先行しているが、本研究では発生様式 (システム) の多様性の遺伝基盤の解明を目指してユニークな貢献をしたい。

<達成できなかったこと、予想外の困難、その理由>

4) 近縁線虫の研究

共同研究者の E. Schierenberg も確認しているが、この種を含め *C. elegans* 以外の多くの線虫は 2 本鎖 RNA の導入による RNAi が効かない。このためホモログの機能解析が一直線には進まなかった。

<今後の課題>

1) NEXTDBの拡張

発現パターンのアノテーションに多大の時間が取られているが、継続する。それをもとに発現パターンクラスタリングを推進する。また、完全長 cDNA 配列決定や、Solexa 等による全トランスクリプトーム解析を進め、スプライシング変動の全貌をつかみたい。

<成果公表リスト>

0802072142

Hiroshi Kagoshima, Rachael Nimmo, Nicole Saad, Junko Tanaka, Yoshihiro Miwa, Shohei Mitani, Yuji Kohara and Alison Woollard: The *C. elegans* CBFb homologue BRO-1 interacts with the Runx factor, RNT-1, to promote stem cell proliferation and self-renewal. *Development* 134, 3905-3915 (2007)
0705142123

Kagoshima H, Shigesada K, Kohara Y.: RUNX regulates stem cell proliferation and differentiation: Insights from studies of *C. elegans*. *J Cell Biochem.* 2007 Apr 1;100(5):1119-30

大量 DNA シーケンシング体制の構築と支援

●小原 雄治

情報・システム研究機構 国立遺伝学研究所生物遺伝資源情報総合センター

<目的>

「基盤ゲノム」は、特定領域ゲノム3領域(「生命情報システム」「比較ゲノム」「応用ゲノム」と密接な連携のもとに、これまでの特定領域研究等によって整備された施設・設備を有効活用しスケールメリットがありかつゲノム研究に必須のデータ取得を推進することにより、ゲノム研究のさらなる展開をめざすものである。その中において本支援班は平成16年度で終了した特定領域「統合ゲノム」で設置されたDNAシーケンシング施設を活用し、より高効率のシーケンシング体制を構築し、他3領域特に「比較ゲノム」領域との連携のもとにそれぞれの課題におけるゲノム/ESTシーケンシングを集中的におこなうことを目的とする。コストや効率の改良を図り、進化の観点から重要な生物種のゲノムやESTの集中的な解析に応用し、配列・発現比較などから生命システムの解明をめざすものである。

<2008年度の活動方針>

①大量シーケンシング体制の構築・改良

現有施設を活用してより高効率のシーケンシング体制を構築する・新しいシーケンシング技術の導入も視野に入れ、一層の低コスト化、効率化をはかる。また、大量情報処理のための基盤を整備する。

- ・whole genome shotgun(WGS)法を最適化するために、偏りの少ないショットガンライブラリー構築、長鎖データ産出の効率的な方法を検討する。
- ・cDNAの全長配列の決定を、低コスト、ルーチン処理で行うために昨年度開発したmix cDNA shotgun法を改良し、さらなる効率化、高速化を目指す。
- ・ライブラリー管理からシーケンスアッセンブリ、アノテーション、公開まで一貫した情報処理体制を構築し、処理の高速化と間違いの最小化を図る。この部分では本領域支援班「情報解析と成果公開のための支援活動」との連携をはかる。

②大量シーケンシング支援

- ・他3領域と連携し、それぞれの課題におけるゲノム/ESTシーケンシングを集中的におこなう。
- ・限られた能力を最大限有効に利用するために、総括班DNAシーケンシングセンター委員会において支援内容(配列決定の対象、順番など)を決定する。

③リソース管理

- ・シーケンスをしたクローンは研究者コミュニティで共有する必要がある。これは個々の依頼研究室には大変な手間であるので、集中的にクローン維持配布に応じる体制を構築する。

<2007年度の成果>

①大量シーケンシング体制の構築・改良

- ・コストダウンについては、ゲノム、ESTの両方の配列解析においてTempliPhiを中心とするシステムを導入し、徹底した希釈により目標を達成した。

②大量シーケンシング支援(年度末までの予定も含む)

1. ゲノム

- 1) ヒメギボシムシ 500-600 Mb(「比較ゲノム」佐藤矩行(京大) 藤山秋佐夫(情報研)) 昨年度、WGSシーケンス100万リードを終了し、総延長で550 Mbの配列を決定した。本解析は2008年度も継続し、新しいシーケンシング技術を検討しつつ、全ゲノムのアセンブルを目指す。
- 2) 立襟鞭毛虫 100 Mb(「比較ゲノム」岩部直之(京大)藤山秋佐夫(情報研)) WGSシーケンス30万リードを終了し、これまでに総延長で700 Mbの配列を決定した。RAMENアセンブラにより、約1万個のscaffold(総延長63 Mb、最長1.7 Mb)を得た。本解析は2008年度も継続し、全ゲノムアセンブルを目指す。
- 3) 細胞性粘菌 *Acytostelium subglobosum* 40 Mb(「比較ゲノム」漆原秀子(筑波大)) WGSシーケンス40万リードを終了し、総延長330 Mbの配列を決定した。PCAPアセンブラによって7,314個のscaffold(総延長36 Mb、最長1.7 Mb)を得た。
- 4) ヌクレオモルフ 合計450 kb(直鎖状染色体3本)(「比較ゲノム」石田健一郎(筑波大)) 2万リードを終了し、総延長10 Mbの配列を決定した。PCAPアセンブラによって得られた740個のscaffold(総延長1 Mb、最長88 kb)の提供を行った。
- 5) マウスMSM系統(「比較ゲノム」城石俊彦(遺伝研)) すでに取得した全ゲノムショットガンデータをもとに標準B6系統との間の比較解析をおこなった。

2. EST

- 1) ヒメツリガネゴケ(「比較ゲノム」長谷部光泰(基生研)西山智明(金沢大)) 5'SAGE 5ライブラリー計73,000クローン
- 2) ゼニゴケ(「比較ゲノム」福澤秀哉(京大)) 2ライブラリー、計115,000クローン
- 3) アピコンプレクサ原虫(「比較ゲノム」渡辺(東大)) 3ライブラリー、計35,000クローン
- 4) ギボシムシ(「比較ゲノム」佐藤矩行(京大)) 4ライブラリー、計54,000クローン
- 5) 細胞性粘菌(「比較ゲノム」漆原秀子(筑波大)) 2ライブラリー、計12,000クローン
- 6) コムギ(比較ゲノム・荻原保成(横浜市大)) 8ライブラリー、計61,000クローン
- 7) タマーワラビー(「比較ゲノム」黒木陽子(理研)) 3ライブラリー、計92,000クローン
- 8) ショウジョウバエ(「生命システム情報」相垣敏郎(首都大東京)) 2ライブラリー、計34,000クローン
- 9) 線虫(「基盤ゲノム」小原雄治(遺伝研)) 8ライブラリー、計177,000クローン

③リソース管理

- ・本課題で解析したcDNAクローンは全て2組のレプリカを作成し、依頼研究室とシーケンシングセンターで分けて保管し、生物資源の安全管理を行っている。また、依頼に応じてクローンの配布を行っている。