

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ものであり、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。

<2008 年度の研究の当初計画>

①線虫 *C.elegans* の解析

1) 線虫 *C.elegans* トランスクリプトームの徹底解析。

発生時期毎の mRNA について完全長 cDNA ライブラリー解析、Solexa 等を用いた解析を行い、トランスクリプトームを徹底的に解析する。スプライシング特異的エキソンをプローブにした発現パターン解析や 5' 端情報の分類整理により線虫遺伝子の転写様式の一覧を作成する。

2) 発現制御ネットワークの解析。

発現パターンのクラスタリング解析をおこない、発現制御領域候補の体系的同定さらには結合タンパク質の同定を進め、関与する遺伝子群の全貌解明をめざす。

3) 初期胚における翻訳制御・局在化のメカニズム

母性 mRNA 局在化シス領域の体系的同定を行う。また一群の遺伝子について翻訳制御シス因子を同定し、母性遺伝子の翻訳・局在制御ネットワークの解明をめざす。

4) モデル化に向けた胚の細胞配置データ取得

胚発生の細胞配置（細胞輪郭）を生きたまま時系列で計測する方法をさらに進め、突然変異体の細胞系譜解析をおこない、遺伝子発現変動と形の関係を研究する。

②比較システム解析

1) 近縁線虫 *Diploscapter coronatus* の解析

C.elegans とのシステムの違いの候補を探るために、ゲノムドラフト配列や cDNA 配列を高精度に完成させ、オルソログ、ホモログ遺伝子を同定し、発現比較を行い、違いを探る。また、様々なシステムの違いが見られる他の線虫候補を探る。

<2008 年度の成果>

①線虫 *C.elegans* の解析

1) 線虫 *C.elegans* トランスクリプトームの徹底解析。

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。

これらの拡張として、次世代型シーケンサ SOLEXA による全トランスクリプトーム解析を試みた。昨年度に V-capping による完全長 cDNA ライブラリー解析を行ったものと同じトータル RNA 試料（胚、幼虫、成虫期の雌雄同体と雄リッチな混合時期集団の 4 試料）を用いた。まず、ポリ A-RNA を精製し、ランダムヘキサマーをプライマーとして cDNA を合成し、定法で 2 本鎖 DNA にし、これを SOLEXA シーケンサで解析した。各試料から 800 万リード（有効塩基長 32）を得、ELAND プログラム及び我々がこのために開発したイントロンをまたぐギャップを許すプログラムにより 75% がゲノムにマップできた。5' 端については、線虫特有のスプライスリーダーを検出し、マップした。これらのデータは NEXTDB に統合化し、エキソン構造、発現レベル（リード数）、発生ステージ、雌雄同体と雄などについてのビューアを構築した。全トランスクリプトームのアセンブルは非常に望まれる技術であることから、合計 3200 万リードを用いて、いくつかのソフトウェアをテストした。EBI で開発された VELVET アセンブラを用いた場合、最長 5Kb で N50 が 450bp となる 36000 のコンティグにアセンブルされた。VELVET アセンブラの特質から、枝分かれ選択スプライシング領域は枝と幹が別々のコンティグにアセンブルされるためにコンティグ長はどうしても短くなるが、そのような情報でも WormBase の多数の遺伝子モデルの修正（エキソン構造やスプライシング変動）ができた。今後メイトペア情報付加による改善が見込まれた。

2) 発現制御ネットワークの解析。

・腸系譜の時期特異的なモチーフの予測と同定

生物の発生では、遺伝子の特定の領域にタンパク質が結合することにより転写の制御が行われ、その結合／制御サイトに特徴的は配列がモチーフと呼ばれる。しかし、その重要性にも関わらず、多細胞生物において、細胞・時期特異的に制御するモチーフはあまり多くは知られていない。これらは発生の遺伝子ネットワークシステム解明のために必須であり、したがって、*in silico* によるそのようなモチーフ予測や実験検証は重要かつチャレンジングな問題である。

本研究では、私達は *C. elegans*（線虫）の初期胚で、共通の発現パターンを示す複数の遺伝子群の上流配列比較から細胞・時期特異的制御を行なうモチーフの同定を目的としている。比較的単純な多細胞生物である線虫でさえ、過去 20 年間で 10 例弱しか成体を使つての成功例がない。さらに、あるひとつの予測プログラムで複数のモチーフを予測・同定した例もない。そこで、私達は、複数のモチーフを同時に検出し、また複数の群に対しても適

用できる「FEsystem」を開発してきた。今回はこのシステムを用いて、線虫の初期胚での複数遺伝子を制御するモチーフを同定することにした。本研究の *in situ* hybridization database から原腸陥入初期・中期の腸細胞特異的に発現が開始される遺伝子群に、この開発したプログラムでモチーフを予測し、reporter assay 等の検証実験を行なった。結果として、それぞれの時期から別のモチーフを同定することができた。現在、同定されたモチーフの時期特異的な機能を調べると同時に、原腸陥入後期の腸細胞特異的なモチーフの同定も行っている。

・温度感受神経AFDでの発現を制御する調節配列の系統的解析
我々はさらに *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。さらに我々は AWB 化学受容神経に *ttx-1*、*ceh-14* を強制的に発現させ、この神経への *gcy-8* の異所的発現の誘導に成功した。以上の結果は *ttx-1*、*ceh-14* の両方が、*gcy-8* の 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。複数の近縁線虫のゲノム配列データを使った比較解析を行ったところ、*gcy-8* 相同遺伝子群の上流配列中には線虫種間で高度に保存された配列があり、実験的に求めた 50 bp とその一部が一致した。我々はこの配列が TTX-1、CEH-14 タンパク質の結合配列である可能性が高いと考えている。今後は、近縁線虫のゲノム情報を活用し、塩基置換やさらに詳細な欠損解析を進めることによって AFD 神経に特異的な発現を制御するメカニズムの解明に迫って行きたいと考えている。

・内胚葉誘導におけるSRCシグナル経路の解析
C.elegans の内胚葉は、4細胞期のEMS割球の後ろ側に、P2割球からEMS割球の後端へもたらされるWnt及びSrcの重複するシグナルにより誘導される。これらのシグナルは、内肺葉形成遺伝子の発現を抑制するPOP-1の核局在を阻害することにより誘導を引き起こすが、この阻害がEMSの娘細胞EおよびMSのうち内肺葉を形成するEに於いて起こる。我々はSrcシグナル経路の解析を進める中でHMP-2のSRC-1に依存したチロシンリン酸化によるHMR-1からの解離及び核移行を見出し、これがPOP-1の核局在の阻害を通じて内肺葉形成に寄与する事を示した。

今回我々は、POP-1やHMP-2及びHMP-2のチロシンリン酸化などの定量をおこない、更にHMP-2の機能を解析した。この結果、HMP-2はWntシグナルと協調してEにおけるPOP-1の核局在を阻害している事が示された。また、リン酸化されたHMP-2はEの核にMSの核よりも強く局在し、局在量及び非対称性はSRC-1に依存していた。従ってHMP-2のEにおけるPOP-1の核局在の阻害はSRC-1依存的にリン酸化されたHMP-2がEの核に多く局在する事によると示唆された。

β -カテニンHMP-2はカドヘリンHMR-1にアンカーされているが、SRC-1に加えHMR-1もRNAiで変異させたところSRC-1単独に比べてHMP-2の量はEおよびMS両方の核で増大し、非対称性は増大しなかった。HMR-1の下流ではHMP-2の局在の非対称性が生じないことから、SRC-1に依存したHMR-1からのHMP-2の解離がHMP-2の核への局在に重要であることが示された。しかし *src-1*(RNAi); *hmr-1*(RNAi) はリン酸化チロシンの量は増大させず、HMP-2のリン酸化はHMR-1から解離したHMP-2の核局在に不要である事が示された。*hmr-1*(RNAi) がWntとSrcシグナルの除去に依る内肺葉欠損を抑圧する事を考慮すると、HMP-2のリン酸化はHMP-2の活性にも不必要であり、HMR-1からの解離によりHMP-2の内肺葉誘導活性が制御され

ていると考えられる。

3) 初期胚における翻訳制御・局在化のメカニズム

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、昨年度までに、局在化に必要な配列(シス因子)を *pos-1* 3' UTR の 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おり、トランス因子の解析のために候補遺伝子アプローチとともに、このRNA領域に結合する遺伝子蛋白質を酵母 tri-hybrid や three-hybrid 法などにより探索中である。

また、*pos-1* mRNA とは逆に Anterior 側に局在する *mex-3* mRNA の局在シス因子を 3-UTR 内に見出した。

4) モデル化に向けた胚の細胞配置データ取得

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が、完成に向けて微調整を進めた。

②比較システム解析

1) 近縁線虫 *Diploscapter coronatus* の解析

ゲノム配列についてサンガー法による whole genome ショットガンに 3x 程度に加え、次世代シーケンサ SOLEXA リードを加えてハイブリッド法によるアセンブルを試みた。ハイブリッド法の最適化の検討を進めている。また、フォスミドライブラリーの構築を進めた。

<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。

<達成できなかったこと、予想外の困難、その理由>

・近縁線虫の研究

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が効かないために、siRNA, shRNA など試してみたがこれまでのところ効果が得られていない。このためホモログの機能解析が一直線には進んでいない。

<今後の課題>

- ・NEXTDBの拡充
- ・母性mRNA局在のトランス因子の同定。
- ・近縁線虫ホモログの機能解析系のたちあげ

<成果公表リスト>

1) 論文/プロシーディング (査読付き)

James Douglas McGhee; Tetsunari Fukushima; Michael W Krause; Stephanie E Minnema; Barbara Goszczynski; Jeb Gaudet; Yuji Kohara; Olaf Bossinger; Yongjun Zhao; Jaswinder Khattri; Martin Hirst; Steven J Jones; Marco A Marra; Peter Ruzanov; Adam Warner; Richard Zapf; Donald G Moerman; John M Kalb: ELT-2 Is the Predominant Transcription Factor Controlling Differentiation and Function of the *C. elegans* Intestine, from Embryo to Adult. *Developmental Biology* (2009, in press)

大量 DNA シーケンシング体制構築と支援

●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。その中にあって本支援班は平成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>

- ①一層効果的な大量シーケンシング体制の構築・改良を進める
 - ・新技術を常にサーベイし、全ステップを再度見直し、一層の低コスト化をはかる。
 - ・SOLEXA等新型シーケンサとのハイブリッド法の最適化を図る。
 - ・ドラフト配列の完成に向けたフィニッシングプラットフォームを構築し、研究コミュニティと連携して可能な限りの完成へ努力する。
 - ・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処理に加え、新型シーケンサ対応の一層の大量情報の処理体制を構築運用する。
- ②大量シーケンシング支援
 - ・本年度のターゲットとしては「比較ゲノム」領域からの要望を中心にゲノムのドラフト、各種ゲノムのBACクローン、多様な生物種のESTの解析を行う
 - ・全長cDNA配列決定：これまでにEST整備したもののから領域の戦略に従い順次進める。
- ③リソース管理（解析済みクローン等の維持配布）も並行して続ける。

<2008年度の成果>

- ①大量シーケンシング体制の構築・改良
 - ・「比較ゲノム」支援班の協力により、線虫及び近縁線虫を用いて次世代シーケンサSOLEXAによるショートリードのde novoアッセンブル解析の試行及び、サンガーリードと併用するハイブリッド法の検討を進めた。またショートリードのゲノム貼り付けのためのソフトウェアの比較、これらによる多型検出（SNP, In/Delなど）のためのビューアを構築した。
 - ・de novoアッセンブルのためには間隔の大きいメイトペアリードが必要であることから、SOLiD3の導入に向けて検討をおこなった。
 - ・ドラフト配列の完成に向けたフィニッシングプラットフォーム

を情報解析支援班の支援で構築し、近縁線虫*Diploscapter coronatus*の全ゲノム解読（サンガーリードとSOLEXAリード及びフォスミド配列などのハイブリッド）へ試用し、改良を進めた。

②大量シーケンシング支援（2009年1月初め現在）

1. ゲノム

- ・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)） 昨年度に引き続き、WGSシーケンス280万リードを追加し、サンガーリード約4X相当を得た。次世代シーケンサリードと合わせ、日米共同で全ゲノムのアセンブルを目指す。
- ・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)） 前年度に、極限まで希釈したシーケンス条件のために高GC比率で配列クオリティが低かったものについて条件を見直しサンガーリードを追加した。論文作成に向けた作業中である。
- ・細胞性粘菌 *Acytostelium subglobosum* 40 Mb（「比較ゲノム」漆原秀子(筑波大)） 昨年度までのWGSシーケンスに加え、フォスミド両端配列解析1.2万リードを得た。
- ・クラミドモナスのフォスミド両端配列決定（「比較ゲノム」福澤秀哉(京大)）：GCリッチのためかTempliphiによる鋳型調製に難航し、プラスミド調整で進行中。
- ・近縁メダカ（インドメダカ、ルソンメダカ、ハブシメダカ）（「比較ゲノム」成瀬清（基生研））：計23のフォスミド/BACクローンのドラフト配列決定。

2. EST

- ・カイコ（「比較ゲノム」嶋田透(東大)）：4ライブラリー約40,000クローン
- ・ゼニゴケ（「比較ゲノム」福澤秀哉(京大)）：2ライブラリー、約62,000クローン
- ・細胞性粘菌 *Acytostelium subglobosum*（「比較ゲノム」漆原秀子(筑波大)）：1ライブラリー、約6,000クローン
- ・コムギ(比較ゲノム・荻原保成(横浜市大))：4ライブラリー、約14,000クローン
- ・ヒメミカツキモ（「比較ゲノム」西山智明（金沢大））：1ライブラリー約17,000クローン

3. 全長cDNA配列のde novo決定

- ・EST配列で分類し、代表クローンを混合し、EST配列をアンカーとしてSOLEXA等ショートリードによるアッセンブルを行い内部全長配列決定する条件を設定・検討した。線虫cDNAで試行した。

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た。