

研究種目：特定領域研究

研究期間：2005 ～ 2009

課題番号：17020009

研究課題名（和文） 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

研究課題名（英文） Large-scale DNA sequencing for comparative systems biology

研究代表者

小原 雄治 (KOHARA YUJI)

国立遺伝学研究所・所長

研究者番号：70135292

研究成果の概要（和文）：線虫 *C.elegans* の cDNA 解析を徹底的に進めた。V-capping 法により作成した発生時期ごとの完全長 cDNA ライブラリ約 15 万クローンの EST 解析を行い、新たに 4909 の cDNA 種を得た。whole mount in-situ hybridization 法による発現解析を継続し、11,309 種について発現画像を取得した。これらの結果は NEXTDB に統合し、WormBase ともリンクした。さらに次世代 DNA シーケンサで全トランスクリプトーム解析を行い、NEXTDB に統合した。発現制御領域の体系的同定、発現制御ネットワークの解析では、腸系譜の時期特異的なモチーフの同定とシス因子の同定、温度感受神経 AFD での発現を制御する調節配列と制御遺伝子を同定した。翻訳制御・局在化のメカニズムの解明の研究では、*pos-1* mRNA 局在化のシス配列とトランス因子を同定した。*mex-3* mRNA 局在化でもシス配列を同定した。モデル化に向けた胚の細胞配置データ取得を進めた。近縁線虫 *D. cronotus* のゲノム/cDNA 解析をサンガー法と次世代シーケンサのハイブリッド法で進めた。

研究成果の概要（英文）：We performed cDNA analysis of the nematode *C.elegans* thoroughly. ESTs of about 150,000 full-length cDNA clones that had been made by the V-capping method were obtained to produce 4,909 new cDNA species. Altogether 11,309 cDNA species have been analysed for their expression patterns by the whole mount in situ hybridization. The mRNAs for the cDNA libraries were analysed by the next generation DNA sequencer to obtain the whole transcriptome information. These data were integrated in NEXTDB, which were linked with the *C.elegans* central database WormBase. As to the studies for gene expression regulation network, we identified 1) the stage-specific cis-regulatory sequences and their trans-factors in the early intestine lineage, and 2) the regulatory sequence for the expression in the thermo-sensory neuron AFD and its regulatory genes. As to the studies for mRNA localization mechanisms, we identified 1) cis-element in its 3'-UTR for *pos-1* mRNA localization and its trans-factor, and 2) cis-element for *mex-3* mRNA localization. We developed a semi-automatic system for the creation of cell shape models in *C. elegans* embryogenesis. We carried out whole genome sequencing of a nematode *D. cronotus* by the hybrid method using the Sanger method and the new generation sequencer.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2005 年度	150,200,000	0	150,200,000
2006 年度	150,100,000	0	150,100,000
2007 年度	150,000,000	0	150,000,000
2008 年度	150,100,000	0	150,100,000
2009 年度	151,200,000	0	151,200,000
総計	751,600,000	0	751,600,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・ゲノム生物学

キーワード：(1) 胚発生 (2) 母性遺伝子 (3) mRNA 局在化
 (4) UTR(非翻訳領域) (5) トランスクリプトーム (6) in situ ハイブリダイゼーション
 (7) 転写調節領域 (8) 多光子共焦点顕微鏡

1. 研究開始当初の背景

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おり、このような方向が求められていた。

2. 研究の目的

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した。さらにこれを通して支援班の解析技術向上につなげることも目指した。

3. 研究の方法

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

リーを作成し、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。

4. 研究成果

1) 線虫 *C.elegans* の解析

(A) NEXTDB の拡張と *C.elegans* トランスクリプトーム

2004年度までに136,466のcDNAクローン(λ -ZAPおよびオリゴキャップ)についてEST解析を行い11,904のユニークなcDNA種を得ていた。本研究では、V-capping法により発生時期ごとの胚および雄をサンプルとしてcDNAライブラリを作成し、得られた150,041クローンについてEST解析を行い、新たに4,909のユニークなcDNA種を得た。これらEST配列はゲノム上に割当てることによって、既知遺伝子や既予測遺伝子と対応付けた。

分類同定されたcDNA種について whole mount *in-situ* hybridization 法による発現解析を継続し、最終的に11,309種について、発現ステージおよび位置を同定する画像を取得した。これら画像について、手動によるアノテーション付けを行い、検索可能な形式に整理した。また、発現パターンより選択したcDNA種についてRNAi解析を継続し、920種について表現型画像を取得した。

上記の結果はNEXTDBとして統合され、WormBaseをはじめとする外部リソースとのリンクを加えた形で公開されている。また、本データベースを発現パターンから検索するためのエンジンとしてSearchEXを開発した。

さらに、次世代型シーケンサ SOLEXA を用いた全トランスクリプトーム解析を試みた。具体的には、胚、幼虫、成虫期の雌雄同体と、雄リッチな混合時期の計4種のmRNA試料をランダムヘキサマでcDNA化し、SOLEXAによりシーケンスし、各試料600-1000万、計3500万リードを得た。これらをゲノム配列にマッピングすることで、エキソン構造の推定と同時に、マップ可能リード数を指標に発現レベルのステージ間/性別間比較を行った。これらの結果は、ショートリード/コンティグ専用ビューアを開発しNEXTDBに導入することで、同データベースへ統合した。この解析手法は支援班で活用した。

NEXTDB <http://nematode.lab.nig.ac.jp>

(B) 発現制御領域の体系的同定、発現制御ネットワークの解析

i) 腸系譜の時期特異的なモチーフの予測と同定

先行研究から線虫の腸の発生には、複数の GATA 因子が、時期特異的に複数の遺伝子の活性に働いていることが知られているが、その GATA 因子がどのように適当なターゲット遺伝子を見分けているのかは謎である。今回はこのために開発したモチーフ予測プログラム FESystem を用いて、本研究室の *in situ* hybridization database から 3 つの連続した原腸陥入期 (E2-, E4-E8-時期) で発現が開始する遺伝子群に応用し、GATA サイト以外に、線虫の初期胚での腸細胞で複数遺伝子を時期特異的に制御するモチーフを同定した。

さらに転写因子候補についても、FESystem で線虫の yeast-one-hybrid のデータセット (Deplancke *et al.*, 2006) の中から転写因子結合サイトを予測した結果、それぞれの時期で転写因子候補が見つかった。特に興味深い発現パターンを示した E8 時期のモチーフ (DAF-12 結合サイトに類似) について詳細な解析を行った。その結果は、DAF-12 が E8 時期特異的に転写制御に寄与していることを強く示唆、さらに GATA 因子との関係では ELT-2 が発現に寄与していることが分かった。

ii) 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

AFD 神経で特異的に発現している 11 種類の遺伝子プロモーター:GFP コンストラクトを構築し、系統的な欠損を導入し、AFD 特異的な発現に必要な最小限な調節領域 (50 bp~600 bp) を決定した。この中で、我々は特にグアニル酸環状化酵素 *gcy-8*、*gcy-18* 遺伝子のプロモーターに着目した。*C. elegans* に加え 4 種類の近縁線虫 *C. briggsae*、*C. brenneri*、*C. remanei*、*C. japonica* の *gcy-8* と *gcy-18* の相同遺伝子の上流領域についての発現解析を行ったところ、*C. elegans* の場合と全く同様に AFD 特異的 (排他的) な発現を示すことが判った。また、これらの近縁線虫のゲノム配列データを使った比較解析では、*gcy-8*、*gcy-18* 相同遺伝子群の上流配列中に種間で高度に保存された配列があり、実験的に求めた *gcy-8* の 50 bp、および *gcy-18* の 68 bp のコア配列の一部と一致した

温度感受神経 AFD では 4 種類の転写因子 (*ceh-14*、*ceh-23*、*nhr-38*、*ttx-1*) が発現しているが、このうち 2 つの転写因子、LIM 型、及び OTX 型ホメオボックスである CEH-14、TTX-1 をコードする遺伝子の変異だけが、この神経の機能である温度走性に異常を示す。そこでこれらの変異体での *gcy-8*、*gcy-18* プロモーター:GFP の AFD 神経での発現を観察したところ、*ceh-14* または *ttx-1* どちらか一方の変異体では発現レベルがやや低下に過ぎなかったが、これらを二重変異にすると発現

が完全に消失した。この結果を受けて次に、これらの転写因子を別の臭覚受容神経 AWB に強制発現させたところ、本来 AFD 神経だけでしか発現しない *gcy-8* と *gcy-18*:GFP を、AWB 神経で発現させることに成功した。このことから、*gcy-8*、*gcy-18* の発現を制御する上流転写因子は *ceh-14* および *ttx-1* であり、この両者が共同して転写制御を行っている可能性が示唆された。

(C) 翻訳制御・局在化のメカニズムの解明

i) *pos-1* mRNA 局在化機構の解析

線虫初期胚において生殖細胞系譜へ局在化する母性 *pos-1* mRNA の局在化機構に着目した。POS-1 タンパク質は他の母性遺伝子の翻訳制御を通じて生殖細胞系譜の確立に働くが、*pos-1* mRNA の局在化はこの POS-1 タンパク質の局在制御に関与していると考えられる。本研究で我々は、1. 局在に必要な mRNA 上の配列 (シス因子) 同定のための *in vivo* アッセイ系の構築、2. *in vivo* アッセイ系を用いたシス因子の同定、3. シス因子に結合して局在を制御するタンパク質 (トランス因子) の探索を行った。

シス因子の解析は欠損解析と置換解析により進めた。この結果、*pos-1* mRNA の局在化には 3' 非翻訳領域 (UTR) 上の近接した二つの保存配列 (CCCACA が二回繰り返すタンデムリピート配列、及び 30 残基の配列) が必要と判明した。他生物における研究から、mRNA の局在制御は多くの場合 3' UTR 上のシス因子により制御されている事が示されているが、この結果は線虫においても同様の機構が mRNA の局在制御に働いている事を示唆するものである。

トランス因子については、Yeast three-hybrid 法を利用し、同定したシス因子と結合するタンパク質を線虫 cDNA ライブラリーから探索した。この結果、候補遺伝子の一つ、*mex-1* が *pos-1* mRNA の局在に必要な事が分かった。

ii) *mex-3* mRNA 局在化のメカニズム

線虫初期胚において、前側割球へと局在化する母性 mRNA として有名なものに母性 *mex-3* mRNA があり、本研究では母性 *mex-3* mRNA の局在化機構について解析を行なった。

まず母性 *mex-3* mRNA の局在化に必要なシス因子の解析を行なった。その結果、*mex-3* 遺伝子の 3' 側非翻訳領域上 (3' UTR) の 179nt が母性 mRNA の局在化に十分であり、179nt を欠失した *mex-3* 遺伝子から転写された VENUS 融合 mRNA は局在化しないことが分かった。この 179nt はステムループを構成し、特にループ部付近を構成する 35nt が mRNA の局在化に重要であることが明らかとなった。この 35nt は線虫種間でも高度に保存されて

いた。

(D) モデル化に向けた胚の細胞配置データ取得

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行った。

このシステムは、ある時点における細胞形状を3次元画像上での領域拡張法により抽出するプログラムとそれに必要な種点を対話的操作で修正するためのユーザーインターフェースプログラムで構成されている。細胞膜に局在するGFP融合遺伝子を発現する線虫株の撮影条件を検討した。その結果、多光子共焦点顕微鏡を用いて撮影後も発生が進行し孵化するような条件で、1~28細胞期の胚の画像データが取得できるようになった。

2) 近縁線虫のゲノム/cDNA 解析

発生過程の多様性および可塑性に関してその遺伝的基盤を比較解析するために、原腸陥入までの卵割およびその配置が異なる近縁線虫 *Diploscapter coronatus* (E. Schierenberg, Univ. Koeln より供与) をサンプルとしてゲノムおよび cDNA 解析を行った。

cDNA 解析では通常法、オリゴキャップ法、V-capping 法によりライブラリーを構築し、得られた計 147,256 クローンについて EST 解析を行った結果、13,245 のユニークな cDNA 種に分類され、うち 6,351 については *C. elegans* にホモログが見つかった。一方、非常に高頻度 (3,063/147,256) に出現するが、既知のタンパク質に相同性の見られない cDNA 種も見られた。全長 cDNA クローンの 5' EST 先頭を解析することにより、*D. coronatus* においてもスプライスリーダー配列が見られることが判明した。高頻度であったのは *C. elegans* における SL1、SL2 と同一の配列であったが、そのほかにこれらと僅かに異なる配列パターンも一定の頻度で見られた。

ゲノム解析では、全ゲノムショットガン (WGS) 法によるゲノム配列再構成を試みた。ショットガンクローンの両端をサンガー法によりシーケンスして 120 万リードを得、高クオリティ領域抽出およびベクタ削除を行って総延長 657Mbases を得たが、これは推定ゲノム長より×4 相当と考えられる。本リードセットを PCAP (Huang, Iowa State Univ.) を用いてアセンブルし、36,000 コンティグ (最長 48,519 bases、N50 8,945 bases)、21,873 scaffolds (最長 61,666 bases、N50 13,166 bases)、総延長 175Mbases のアセンブリを得た。本アセンブリ上より *C. elegans* ホモログを探索することで、EST 解析では未発見の

3,066 ホモログが見つかった。また、cDNA 解析で得られていた EST 配列をアセンブリ上に精確にアラインメントすることで、一部の遺伝子についてエキソン構造を得ることができ、さらに、複数のスプライシングパターンの存在を示唆する結果が得られた。

ゲノムアセンブリの質的向上のため、次世代型シーケンサ SOLEXA リードを用いたハイブリッド法を試行した。具体的には、*D. coronatus* ゲノムを SOLEXA によりシーケンスし、得られた 1.7 億本、総延長 650Mbases のリードを VELVET (Zerbino, EBI) によりアセンブルし、得られたコンティグを WGS リードセットに加えて、全体を PCAP でアセンブルした。新アセンブリを、WGS リードのみを素材としたアセンブリと比較した結果、16,684 コンティグは伸長され、4,167 コンティグ組はその間がブリッジされたことが分かった。

上記のゲノム/cDNA 解析結果は、NEXTDB と併せて公開できるよう、WWW 上のデータベースとして整備した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 11 件) すべて査読あり

1) Mangone, M., Manoharan, A.P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S., Mis, E., Zegar, C., Gutwein, M.R., Khivansara, V., Attie, O., Chen, K., Salehi-Ashtiani, K., Vidal, M., Harkins, T.T., Bouffard, P., Suzuki, Y., Sugano, S., Kohara, Y., Rajewsky, N., Piano, F., Gunsalus, K.C., Kim, J.K.: The Landscape of *C. elegans* 3'UTRs. **Science** Published Online June 3, 2010, DOI: 10.1126/science.1191244

2) Ogura, K., Okada, T., Mitani, S., Gengyo-Ando, K., Baillie, D.L., Kohara, Y. and Goshima, Y.: Protein phosphatase 2A cooperates with the autophagy-related kinase UNC-51 to regulate axon guidance in *Caenorhabditis elegans*. **Development** 137, 1657-1667 (2010)

3) Langenhan, T., Prömel, S., Mestek, L., Esmaeili, B., Waller-Evans, H., Hennig, C., Kohara, Y., Avery, L., Vakonakis, I., Schnabel, R., and Russ, A.P.: Latrophilin signalling links anterior-posterior tissue polarity and oriented cell divisions in the *C. elegans* embryo. **Developmental Cell** 17(4), 494 – 504 (2009)

4) X. Wang, Y. Zhao, K. Wong, P. Ehlers, Y. Kohara, S. J. Jones, M. A. Mara, R. A. Holt, D. G. Moerman and D. Hansen: Identification of

genes expressed in the hermaphrodite germ line of *C. elegans* using SAGE. **BMC Genomics** 2009, 10:213

5) Kawano T, Zheng H, Merz DC, Kohara Y, Tamai KK, Nishiwaki K, Culotti JG.: *C. elegans* mig-6 encodes papilin isoforms that affect distinct aspects of DTC migration, and interacts genetically with mig-17 and collagen IV. **Development**. 136, 1433-1442 (2009)

6) J. D. McGhee; T. Fukushige; M. W. Krause; S. E. Minnema; B. Goszczynski; J. Gaudet; Y. Kohara; O. Bossinger; Y. Zhao; J. Khattrra; M. Hirst; S. J. Jones; M. A. Marra; P. Ruzanov; A. Warner; R. Zapf; D. G. Moerman; J. M. Kalb: ELT-2 Is the Predominant Transcription Factor Controlling Differentiation and Function of the *C. elegans* Intestine, from Embryo to Adult. **Developmental Biology** 327, 551-565 (2009)

7) H. Kagoshima, R. Nimmo, N. Saad, J. Tanaka, Y. Miwa, S. Mitani, Y. Kohara and A. 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)

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

9) J.D. McGhee, M. C. Sleumer; M. Bilenky; K. Wong; S. J. McKay; B. Goszczynski; H. Tian; N. D. Krich; J. Khattrra; R. A. Holt; D. L. Baillie; Y. Kohara; M. A. Marra; S. J.M. Jones; D. G. Moerman; A. G. Robertson: The ELT-2 GATA-Factor and the Global Regulation of Transcription in the *C. elegans* Intestine. **Developmental Biology** 302, 627-645 (2007)

10) E. Lazakovitch, J. M Kalb, R. Matsumoto, K. Hirono, Y. Kohara, R. M. Gronostajski: nfi-1 affects behavior and life-span in *C. elegans* but is not essential for DNA replication or survival. **BMC Developmental Biology** 2005, 5:24

11) H. Kagoshima, H. Sawa, S. Mitani, T. R. Burglin, K. Shigesada and Y. Kohara: The *C. elegans* RUNX transcription factor MAB-2/RNT-1 is required for asymmetrical cell division of the T blast cell. **Developmental Biol**. 287, 262-273 (2005)

[学会発表] (計 22 件)

[その他]
ホームページ等
NEXTDB <http://nematode.lab.nig.ac.jp>

6. 研究組織
(1) 研究代表者
小原 雄治 (KOHARA YUJI)
国立遺伝学研究所・所長
研究者番号 : 70135292