

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 4 月 1 日現在

機関番号：12102

研究種目：基盤研究(C)

研究期間：2009～2011

課題番号：21510214

研究課題名（和文）線虫の初期発生段階における遺伝子調節ネットワークの予測とシミュレーション

研究課題名（英文）Simulation and prediction of gene regulatory network of *C. elegans* early embryo

研究代表者

真栄城 哲也 (MAESHIRO TETSUYA)

筑波大学・図書館情報メディア系・准教授

研究者番号：30361356

研究成果の概要（和文）：線虫 *C.elegans* の最初の 40 分までの初期胚を対象に、ゲノムワイドな遺伝子発現の時系列データを 10 分間隔で DNA マイクロアレイを用いて計測した。そのデータを用いて、超高速シミュレータ Starpack を基盤とする予測システムによって、初期胚における遺伝子調節ネットワークを予測した。

研究成果の概要（英文）：Whole genome gene expression data with ten minutes interval of *C.elegans* early embryo (initial 40 min) was measured using DNA microarray. This data was processed by a prediction system based on ultra high speed simulator Starpack, and the gene regulatory network of early embryo was predicted.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	2,100,000	630,000	2,730,000
2010 年度	1,200,000	360,000	1,560,000
2011 年度	500,000	150,000	650,000
年度			
年度			
総計	3,800,000	1,140,000	4,940,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・ゲノム情報科学

キーワード：線虫, シミュレーション, マイクロアレイ

## 1. 研究開始当初の背景

近年、様々なハイスループット実験によって細胞の挙動をゲノムワイドで計測することが可能となる一方、これらの実験データから根底にある生命機構についての情報を得る方法が求められている。本研究で扱う遺伝子調節ネットワークの推定については、個々の遺伝子を生物実験で調べるには膨大な時間とコストがかかるため、DNA マイクロアレイ等の実験データを基にした予測手法がこれまでに提案されている (Basso et.al, Nature Genetics, 2005, Friedman et.al, J. Comp. Biol., 2000, Yu et.al, ICSB, 2002 等)。

これらの手法の中で動的ベイジアンネットワーク (DBN) が最も予測精度が高い。しかし、DBN の欠点は大量のデータ (サンプル数) を必要とすることである。

DBN の予測精度の評価として、評価用の遺伝子数 20 の単純なネットワークを用いた検証結果が発表されている (Hartemink, Nature Biotech., 2005)。各遺伝子の発現量の時系列データを計算機シミュレーションによって生成し、遺伝子発現量の時系列データのサンプル数に対する予測精度の変化を調べた結果、DBN が 100% の予測精度を達成するには約 300 から 600 のサンプル数

が必要であった。この必要なサンプル数は予測対象の遺伝子数の増加とともに増加すると考えられる。多くの場合、一連の実験で得られるマイクロアレイデータはたかだか数十だと考えられ、数百の DNA マイクロアレイデータを実験で得るには膨大な時間とコストが必要であり、現実的ではない。一方、複数の変異体の実験から得られるマイクロアレイデータから総合的な解析を行うこともあるが、合計数百のサンプル数を得られることは、あまりないと思われる。なお、サンプル数が 100 個の場合、DBN の予測精度および予測感度は、20 個の遺伝子を対象とする単純な場合でも、それぞれ 62%と 67%である。

本研究で扱う線虫 *C.elegans* はモデル生物の一種であり、分子レベルにおける様々な機構が明らかにされつつあるが、未解明の機構や現象も多い。本研究は *C.elegans* の初期発生段階に注目し、遺伝子発現機構と遺伝子間の関係の詳細を解析する。発生の初期段階は詳細が未解明であり、初期段階から得られる知見には発生全体の基本的な知見が含まれると考えられる。また、初期段階の遺伝子発現機構の予測ができれば、より複雑な過程の予測に応用できる。さらには、マウスやヒト等のより高等な生物に関連する知見が得られると期待できる。

*C.elegans* は発生の研究に適した生物であり、L1, L2 等の発生段階毎の DNA マイクロアレイ実験データも公開されている (Kim et al, Science, 2001 等)。しかし、本研究のように遺伝子調節ネットワークの予測に必要な短い時間間隔 (数十分以下) で実施した DNA マイクロアレイ実験データは見当たらない。

## 2. 研究の目的

本研究の目的は、線虫 *C.elegans* の初期発生段階における遺伝子発現の時系列データをマイクロアレイ実験によって計測し、さらにはシミュレーションによって発生の初期段階における遺伝子発現の調節機構を、遺伝子調節ネットワークとして予測することである。なお、遺伝子調節ネットワークは遺伝子間の相互作用関係のネットワークを指す。

## 3. 研究の方法

基本的な研究方法は、(A) ゲノムワイドな DNA マイクロアレイ実験によって遺伝子調節ネットワークの予測に必要なデータを生成し、このデータを基に (B) 線虫の初期発生段階の遺伝子調節ネットワークを予測する。

線虫の発生の初期段階の DNA マイクロア

レイデータを実験によって得る。マイクロアレイデータの計測は短い間隔で行い、初期 40 分程度のデータを得る。これによって、発生初期における全遺伝子の発現の時系列データが得られる。発現データの再現性を保証するため、実験を複数回繰り返す。

本研究では遺伝子調節ネットワークのシミュレーションと予測は密接に関連している。遺伝子調節ネットワークの予測は、超高速シミュレータ Starpack を利用して、2 段階で行う。Starpack は、生体内の化学反応ネットワークやシグナル伝達ネットワーク、遺伝子調節ネットワーク等、要素と要素間の相互関係として捉えることができる現象を、直接ハードウェアで高速シミュレーションするシステムである。複数の規模の異なる化学反応ネットワークを対象に実行速度を測定したところ、Starpack の実行速度は、最新の計算機上で実行したソフトウェアによるシミュレーションの 1 万倍以上であった。

第 1 段階は低精度なシミュレーションであり、多数の遺伝子調節ネットワークを高速で探索することを目的とする。これは大局的な探索に相当する。第 2 段階では、第 1 段階で得られる候補を対象に高精度のシミュレーションを行い、予測精度を高める。それぞれの段階の予測は、進化的計算手法に基づく手法で行なう。初期集団は、既知の遺伝子調節関係は固定し、それ以外の関係をランダムに生成したネットワークで構成される。それぞれの個体 (ネットワーク) を Starpack によってシミュレーションし、シミュレーション結果を適応度関数を用いて評価する。全個体の評価後、エリート戦略とルーレット戦略に従って選択し、変異を加え、再度シミュレーションを行なう。ここで、変異は、遺伝子間の接続と接続種類 (活性化か抑制) の変更を意味する。この繰り返しを、生成されるネットワークの評価値が収束するまで繰り返す。

シミュレータ内では、遺伝子調節ネットワークは、専用の表現体系を用いて記述する。これは、シミュレーションの記述と、時間を含むネットワークの論理的矛盾を検出するためである。

シミュレーション対象の遺伝子数を段階的に増やすことで、遺伝子調節ネットワークの予測精度を徐々に向上させることができる。また、シミュレーションのパラメータとして遺伝子間の作用関係を定義する必要があるが、本研究では作用関係の定義は予測によって行われる。

基本的な推測手法は、まず遺伝子調節ネットワークの候補を複数生成し、それぞれの候補を高速シミュレーションによって評価し、評価の高いネットワークに多少の変化を加える、という処理の繰り返しである。

#### 4. 研究成果

##### マイクロアレイデータ

線虫の DNA マイクロアレイデータ実験について、線虫の同調精度を高める実験方法を確立した。この手法を用い、4 回実験を行い、初期胚の 1 細胞期 (0 min.) から 10 分間隔で 40 分までのゲノムワイド (遺伝子数 22,548) な遺伝子発現の時系列データを、4 セット分取得した。これにより、4 セットの 0 分から始まる 10 分間隔の遺伝子発現の時系列変化が判明する。40 分までで、1 細胞期から 12 細胞期までが対象となる。

線虫 *C. elegans* の初期胚発生について、このような短い時間間隔の発現データは見当らず、本研究によって計測された高精度なデータは、より詳細な遺伝子調節ネットワークの推測や、遺伝子間の相互作用関係の理解を探るために重要である。一方、本プロジェクト実施中に、使用している Affymetrix 社のマイクロアレイ実験プロトコルが One cycle 法から Express kit 法に変更され、後者のみが今後利用可能となる。計測した発現データは、3 セットが One cycle 法、1 セットが Express kit 法であるため、同一サンプルを両手法で 3 回ずつ発現データを計測し、両手法によるデータの互換性を解析した。その結果、相関係数は 0.93 であり、Affymetrix が公表している値 0.97 よりも低かった。さらに、発現量の計測値の変動が大きい遺伝子群 (全体の 1.14%) も抽出し、値の変換式を作成した。これらから、今後計測する Express kit 法によるマイクロアレイデータとの比較が可能である。

##### 遺伝子調節ネットワークの記述

このマイクロアレイデータを用いて、遺伝子調節ネットワークのシミュレーションを行うために、線虫の遺伝子群で細胞分裂に関わると考えられる遺伝子群を文献およびデータベースから抽出した。これらの遺伝子群のデータベースを構築するために、遺伝子と調節関係の両側面から捉えると同時に可視化が可能な基本的なデータ構造を設計した。さらに、この構造は、本研究で用いる遺伝子ネットワークシミュレータのアーキテクチャに適しており、表現とシミュレーションが一体化されたモデルであるため、静的および動的な特徴が同時に表現できる利点を持つ。表現モデルは、ハイパーネットワークモデルと呼ぶ我々が考案したモデルである。このモデルは、意味ネットワークや ER モデルのようにグラフに基づく従来の表現モデルとは異なり、相対性や多項関係を記述でき、よりの確に遺伝子調節関係を表現できる。なお、

予備実験としての記述とシミュレーションの結果、モデルの有効性を確認すると同時に、遺伝子間のより詳細な調節関係の表現が必要なことが判明したため、遺伝子調節ネットワークのための表現モデルの拡張を行った。

このモデルを用いて既知の遺伝子調節関係を表現する。遺伝子調節ネットワークのシミュレーションは、この表現を直接読み込んで実行する。さらには、ネットワークの動的な論理解析を実行するため、静的な解析では検出できない遺伝子調節ネットワークの問題を明らかにできる。

また、遺伝子調節ネットワークで実行する進化的計算手法で実行する変異は、この記述モデルに直接作用できるため、表現の変換という余分な処理が不要となり、予測時間の短縮にもつながる。記述が一体化されていなければ、進化的計算による処理の繰り返し毎に、全ネットワーク (個体) に対して、シミュレーション用の記述を進化的計算用に変換し、変異を加え、元のシミュレーション用に再変換するという処理が必要になる。

##### 遺伝子調節ネットワークの評価

本研究では、遺伝子調節ネットワークの予測に、進化的計算手法を用いる。その際に、それぞれの遺伝子調節ネットワークを評価する必要があるが、遺伝子調節ネットワークの予測精度は、そこで用いる評価関数 (適応度関数) の質に大きく依存する。さらには、予測処理の最初に行なう初期集団の生成でも、より尤もらしいネットワークを生成すれば、同様に予測精度を向上できる。このように、遺伝子調節ネットワークの評価関数は重要である。従って、本研究では、対象とする全遺伝子で構成されるネットワークの全体的な構造の尤もらしさを計算する方法 HPA (Harmonic Pulse Analysis) を提案した。この手法は、ネットワークを構成する遺伝子に、順次一定幅の仮想的なパルスを入力し、ネットワーク全体への伝搬をシミュレーションによって計測することで、ネットワーク全体の尤度、または生物らしさを計算する方法である。各遺伝子には、基本パルスの 2 倍、4 倍のように高次周波数の振幅幅を持つパルスを入力する。この尺度によって、必要とする精度に依存するが、遺伝子調節ネットワーク候補数を 10%程度に絞ることが可能である。この尺度は、ネットワークの動的な特性を定量化する。

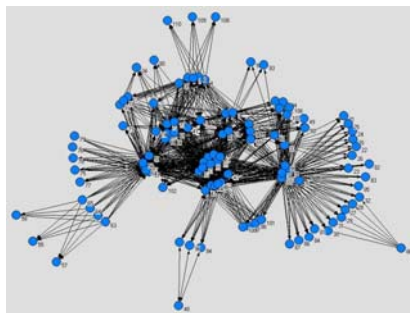
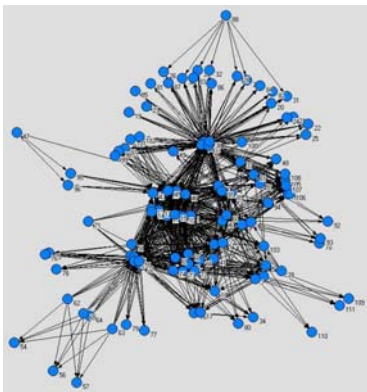
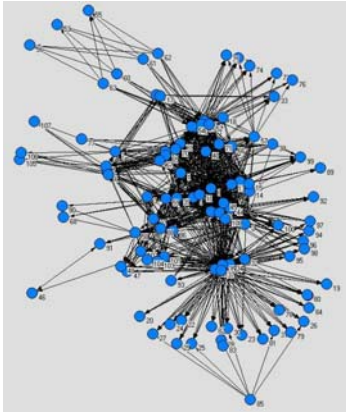
##### 発現遺伝子数

マイクロアレイデータの解析から、発現する遺伝子数は一定ではなく、10,394 個と 10,986 個の間を変動していることが明らかになった。最大差は 5.8%である。

### 遺伝子調節ネットワークの予測

超高速シミュレータ Starpack を利用した予測手法によって、初期胚の 10 分間隔の全遺伝子発現データから、各時点 (0 min, 10 min, 20 min, 30 min, 40 min) の計 5 時点での遺伝子調節ネットワークを予測した。図は 0 min, 20 min と 40 min における予測されたネットワークの一部で、初期胚で発生に中心的な役割を果たしていると考えられる遺伝子を含む。

また、予測された遺伝子調節ネットワークで、多くの遺伝子を調節し、ネットワーク科学の概念でハブとして機能する遺伝子を対象に、既知の機能について調べた。GO を用いて分子機能を調べたところ、41.5% がテロメアに結合する因子、32.3% が転写因子、13.8% が DNA に結合する因子であり、以下は核酸結合、亜鉛イオン結合等と主に細胞複製に関わる機能が続く。



### 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)

[学会発表] (計 12 件)

- ① T. Maeshiro, S. Nakayama, M. Ito, “Information and structural properties of C. elegans gene regulatory networks”, 17th International C. elegans Meeting, CDROM, June 24–28, 2009 UCLA USA.
- ② T. Maeshiro, K. Shimohara, S. Nakayama, “Fitness function for evolutionary system to predict unknown gene regulatory networks”, ICROS-SICE International Joint Conference 2009, 2722–2727, Aug. 18–21, 2009 Fukuoka, Japan.
- ③ T. Maeshiro, S. Nakayama, “Description of Worm Genes for High Speed Gene Regulatory Network Simulator”, 日本分子生物学会第 32 回年会, 2009 年 12 月 9 日～12 日 横浜
- ④ T. Maeshiro, S. Nakayama, “Network Representation Model of Gene Regulatory Network with Multiple Viewpoints”, 日本分子生物学会第 32 回年会, 2009 年 12 月 9 日～12 日 横浜
- ⑤ T. Maeshiro, S. Nakayama, “Harmonic Pulse Analysis to detect biologically plausible gene regulatory networks”, SICE Annual Conference 2010, 3233–3239, Aug. 18–21, 2010 Taipei, Taiwan.
- ⑥ T. Maeshiro, S. Nakayama, “Cronus Two: Computer System for C. elegans Gene Regulatory Networks Research”, Proceedings of the 11th International Conference on Systems Biology, 3233–3239, Oct. 10–15, 2010 Edinburgh, UK.
- ⑦ 岡野文香, 中山伸一, 伊藤将弘, 真栄城哲也, 「線虫の初期胚発生に関連する遺伝子ネットワークの推定」, 情報処理学会第 73 回全国大会, 4-697–698, 2011 年 3 月 2–4 日東京工業大学 大岡山キャンパス.
- ⑧ 岡野文香, 中山伸一, 伊藤将弘, 真栄城哲也, “Prediction of partial gene regulatory network of C. elegans early embryo”, 第 33 回日本分子生物学会年会, 2010 年 12 月 7–10 日 神戸.
- ⑨ 真栄城哲也, “Dynamic Network Analysis and Properties of Gene Regulatory Networks”, 第 34 回日本分子生物学会年会, 2011 年 12 月 13–16 日 横浜.
- ⑩ T. Maeshiro, S. Nakayama, K. Monobe, M.

Ito. “Analysis of temporal expression data of C.elegans early embryo”, 18<sup>th</sup> International C. elegans Meeting, CDROM, Los Angeles, USA June 22-26, 2011

- ⑪ A. Okano, K. Monobe, S. Nakayama, M. Ito, T. Maeshiro, “Estimation of gene regulatory networks of C.elegans early embryo”, 18<sup>th</sup> International C. elegans Meeting, CDROM, Los Angeles, USA June 22-26, 2011
- ⑫ 真栄城哲也, “関係性に基づく表現と知識”, SICE 関係論的システム科学調査研究会, 滋賀, 2012年5月18-19日

## 6. 研究組織

### (1) 研究代表者

真栄城 哲也 (MAESHIRO TETSUYA)  
筑波大学・図書館情報メディア系・准教授  
研究者番号：30361356

### (2) 研究分担者

伊藤 将弘 (ITO MASAHIRO)  
立命館大学・生命科学部・准教授  
研究者番号：50388112