

植物遺伝子の発現ネットワーク推定法

昇 博也† 三井 重之† 長村 吉晃‡ アントニオ バルタザール‡
小松 節子§ 西林 双龍† 渡辺 克昭† 坂田 克己†

†三菱スペース・ソフトウェア (株)

‡ (独) 農業生物資源研究所

§ (独) 農業・食品産業技術総合研究機構 作物研究所

遺伝子の発現時系列データからネットワークを推定する方法について述べる。我々はMINOS (Mathematical Gene Interaction Network Optimization Software) を使い、遺伝子ネットワークを推定する手法を開発してきた。MINOSによれば、線形およびSシステム微分方程式に基づいて、100個以上の遺伝子からなる系の相互作用を容易に推定できる。本報告では、イネゲノム情報を応用した解析についても示す。MINOSで求めたネットワーク解を配列解析から得られる情報と統合することにより、未知遺伝子の機能推定や重要遺伝子の同定が期待できる。

A Method for Estimating Plant Gene-network Based on Expression Profile and Genomic Information

Hiroya Nobori† Shigeyuki Mitsui† Yoshiaki Nagamura‡ Baltazar A. Antonio‡
Setsuko Komatsu§ Soryu Nishibayashi† Katsuaki Watanabe† Katsumi Sakata†

†Mitsubishi space software Co., Ltd.

‡National Institute of Agrobiological Sciences

§National Institute of Crop Science

A method for estimating plant gene-network is reported. We have developed a computer program called MINOS (Mathematical Gene Interaction Network Optimization Software). MINOS allows calculation of the interaction among more than 100 genes from expression time course based on both the linear and S-system model of differential equations. An integrated analysis using rice genome together is also presented. The results suggest that the integration of mathematical solutions and genomic information is effective to capture gene networks involved in complex biological processes.

1. まえがき

遺伝子の発現時系列データから相互作用ネットワークを推定する方法として、シミュレーテッドアニーリングや遺伝的アルゴリズムに基づいたプログラム[1, 2]が研究されてきた。これらの手法では、相互作用が解ける遺伝子数が5個程度に限られていた。これに対し、我々が開発したプログラムMINOS (Mathematical Gene Interaction Network Optimization Software) では、探索範囲を限定して段階的に最小2乗法を適用することにより(段階的的最小2乗法)、効率的なネッ

トワークの推定が可能になった。例えば100個以上の遺伝子による相互作用でもパーソナルコンピュータで容易に推定できる。

MINOSの開発により、遺伝子ネットワークを数学的に解くための現実的な手法が得られたので、植物の環境応答ネットワーク解析などへ応用を進めている。そこでは、ゲノム情報を活用した統合解析による信頼性の向上を目指している。

我々は、イネ科植物のサトウキビに低温ストレスを与えた場合の遺伝子発現プロファイル[3]について、MINOSとイネゲノム情報を用いた統合解析を試みた。その結果、

複数の転写因子が関連した制御ネットワークの存在など、新たな生物学的知見の発見が示唆された。

2. 相互作用解析プログラム-MINOS

MINOS は様々なデータ型に対応した解析ソルバーを備えている。時系列データに対しては、線形微分方程式モデル (式(1)) およびSシステムモデル (式(2)) が用意されている。

$$\dot{x}_i = \mathbf{a}_i \cdot \mathbf{x} + w_i \quad (1)$$

$$\dot{x}_i = \alpha_i \prod_{j=1}^n x_j^{g_{ij}} - \beta_i \prod_{j=1}^n x_j^{h_{ij}} \quad (2)$$

式(1)、(2)においては、 x_i が i 番目の遺伝子の発現データを表している。

Sシステム (Synergism and saturation system) は生化学の領域で度々用いられてきた数学モデルであり、式(2)のように 2 つのべき乗項から構成されている[4]。線形微分方程式モデルよりも系の挙動を詳細に表現できると考えられている。

MINOS の時系列データソルバーは、Fig.1 に示すように発現データの変化率についても良い再現性を示すように設計されている。また効率的な計算を行うため、上流の候補遺伝子を 2 個またはそれ以下の組み合わせに絞り、最適な組み合わせの遺伝子を求めるようにしている(Fig.2)。

MINOS により与えられるネットワーク解には単独評価解 (Stand-alone solution) と連立評価解 (Simultaneous solution) の 2 種類がある。

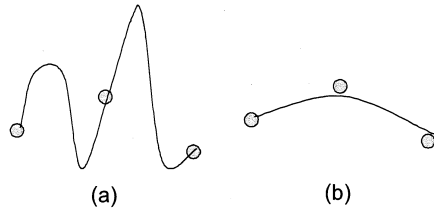


Fig.1 The target of solver is not only (a) to minimize the residuals, (b) but to obtain a good approximation for the rate of change.

単独評価解は以下の手順で計算される。(i) ある遺伝子について上流遺伝子の組み合わせを変えて適用度を計算する。ここで、推定残差が小さいほど適用度が大きくなるような関数を選んでいく。(ii) (i)で計算した適用度を上流の候補となる遺伝子ごとに集計したスコアを計算する。(iii)スコアが大きい候補遺伝子を上流遺伝子として選定する。(iv)他の遺伝子についても(i)-(iii)の処理を行い、ネットワークを得る。

一方、連立評価では全遺伝子の微分方程式を連立し、システム全体の適用度を示す評価関数の上昇が止まるまでネットワークを逐次的に見直していく。その手順は、(i)単独評価手順の(ii)までに得られた上流の候補遺伝子について係数の推定値を記憶しておく。(ii)あるネットワーク解において、一つの遺伝子に着目し、その上流遺伝子を入れ替えたときの評価関数が元の評価関数より大きくなれば、上流の候補遺伝子を入れ替える。(iii)全ての遺伝子について、それ以上評価関数が上昇しなくなるまで手順(ii)を繰り返す。

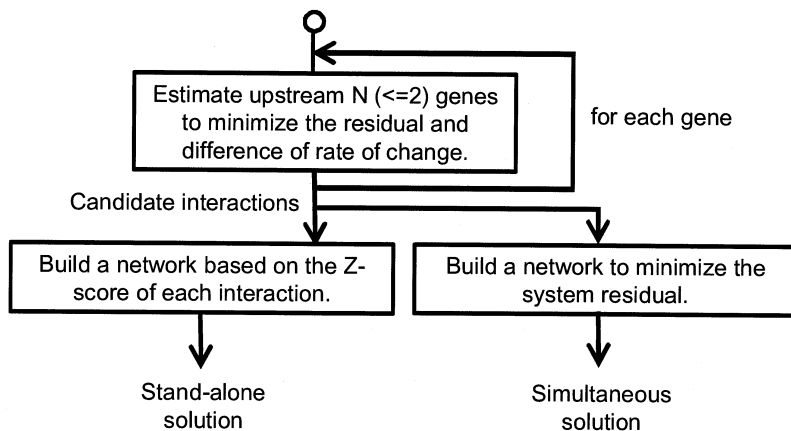


Fig.2 Two type of solutions by MINOS.

Fig.3 は、32 個の遺伝子が示す発現プロファイル[5] から得られた 4 種類の連立評価解を、代表的な 4 つの遺伝子について示した例である。通常、連立評価で得られるネットワーク解は繰り返し計算の初期値により異なる。しかし、図のように係数の推定値を連立微分方程式に代入して数値積分した波形は殆ど一致している。

計算に要する時間を示す。10 個の遺伝子について 6 時点における発現データが与えられている系について、S システムモデルを前提とした連立評価解の計算を行った。初期値を変えて 5 回実行したところ、最長 32sec、最短 12sec、平均 25sec という結果であった (3.4GHz、1GB RAM のパーソナルコンピュータを使用)。

計算で得られたネットワークは、描画モジュールにより 2 次元および 3 次的に表示できる。この時、ネットワークが見易くなるように自動的に最適化された描画が行われる (Fig.4)。

3. ゲノム情報を活用した植物遺伝子ネットワーク解析

MINOS で得られるネットワーク解とゲノム情報を統合した解析の例を示す。

サトウキビ (*Saccharum*) はイネ科の単子葉植物である。低温ストレス下のサトウキビ遺伝子が示す発現プロファイル[3]を、

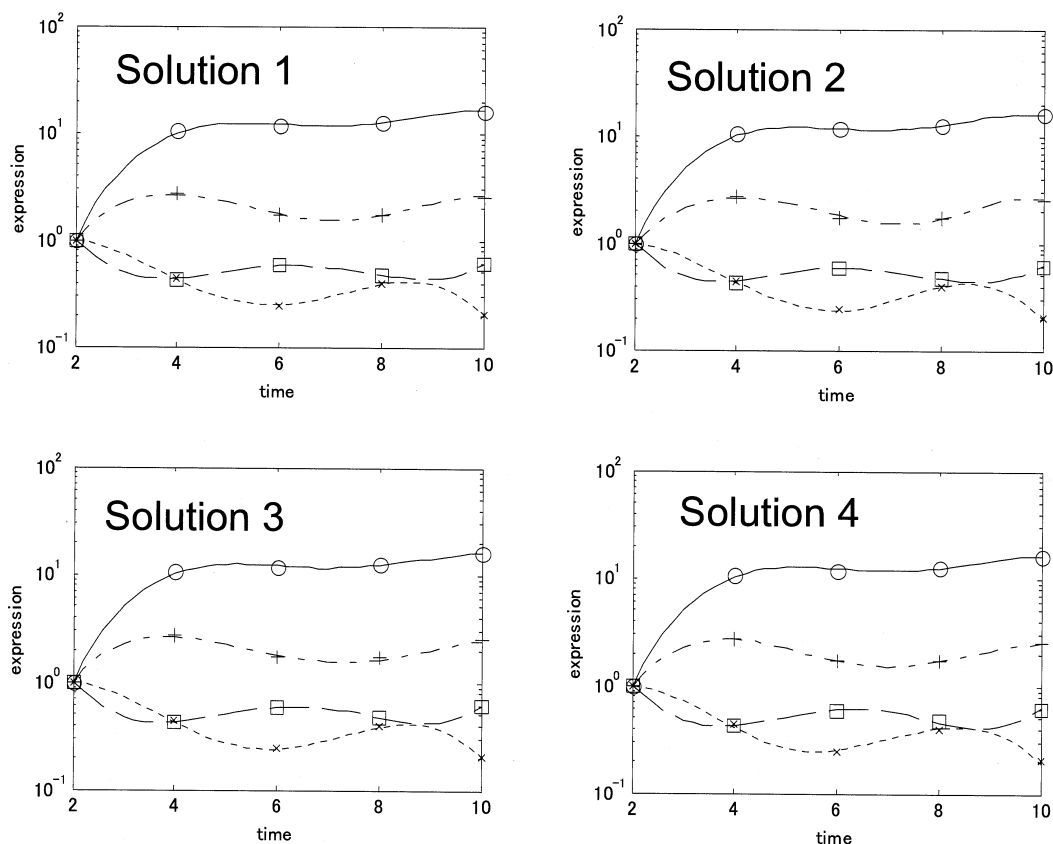


Fig.3 An example of time course by numerical integral of MINOS. A line indicates the numerical integral of simultaneous-solution.

The mark indicates an observation of representative gene. This example shows four different simultaneous-solutions.

MINOS により解析した。得られたネットワーク解をイネの配列データから得られる情報と統合した(Fig.5)。イネは単子葉のモデル植物として、全ゲノム配列や cDNA など、様々な研究リソースが整備されている。

本解析では、低温応答の関与が疑われるサトウキビの EST (25 個) の発現プロファイル[3]を、発現強度およびその変化率で別々にクラスタリングし、得られたクラスタの共通部分 18 クラスタを得た。各クラスタの中央値を MINOS の S システムソルバーで計算した。初期値を変えて 10 回の連立評価計算を行い、5 回以上の計算で共通に検出された相互作用をネットワーク解とした。

一方で、サトウキビ EST 配列と高い相同性を有するイネ完全長 cDNA を、cDNA 解析サイト (cdna01.dna.affrc.go.jp/cDNA/) [6]で検索した。そして、見つけられた cDNA 配列がイネゲノム上にマップされたときゲノム近傍領域に存在する、シス要素 (cis-acting regulatory DNA element) と呼ばれる転写制御に関する塩基配列に注目した。同じイネ科植物であるサトウキビとイネにおいては、遺伝子の近傍に見つかるシス要素も共通の可能性が高いだろうという発想に基づいた解析法である。シス要素の検索は、植物のシ

ス要素を集めた PLACE データベース[7]を参照して行われている。

Fig.6 に統合解析の結果から推定される低温応答モデルを示す。サトウキビ遺伝子の近傍に存在が予想されるシス要素に基づき、18 クラスタは 3 グループに分けられる。グループ 1 は低温誘導に関与するシス要素を近傍に持つ遺伝子群、グループ 2 は NAC と呼ばれる転写因子に関連したシス要素を近傍に持つ遺伝子群、グループ 3 はグループ 1、2 のどちらにも当てはまらない遺伝子群である。配列解析からは、Fig.6 の破線で示すように、3 種類の転写因子が関与した誘導経路が示唆された。MINOS によるネットワーク解を重ねて描画すると実線のようになり、グループ 1 と 2 の間にネガティブフィードバックを含む相互作用が見られた。また、5 番と 8 番の遺伝子は多くの相互作用の起点となっていて、機能的に重要な遺伝子であることが予想される。

4. むすび

MINOS の概要と計算能力について報告した。また、ゲノム情報と統合した植物遺伝子ネットワーク解析法の例を示した。今後

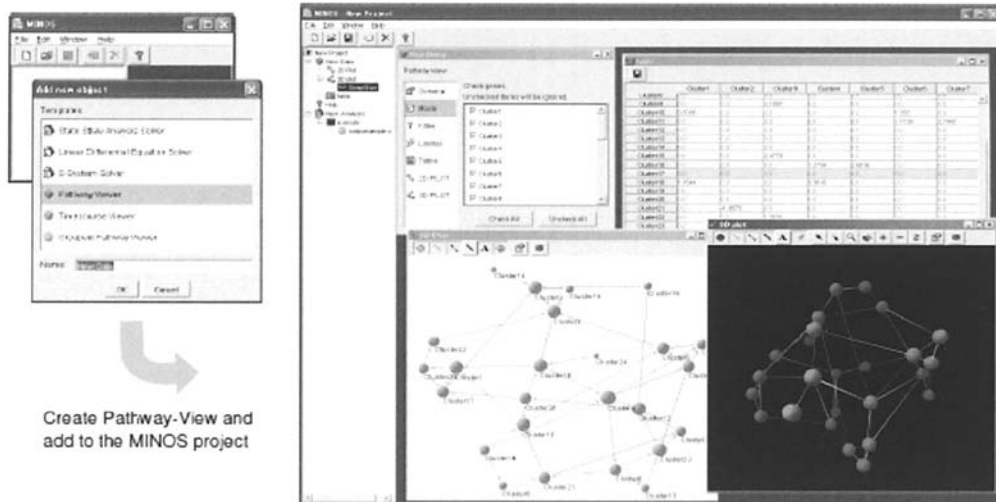


Fig.4 A network view of MINOS. The element with many interactions is automatically positioned in the middle of network diagram. The position (in 2-D view) and viewpoint (in 3-D view) is manually changeable. A filter is available to display the elements of interest. The data are managed in a project style.

は解析法の更なるチューニングを進める計画である。

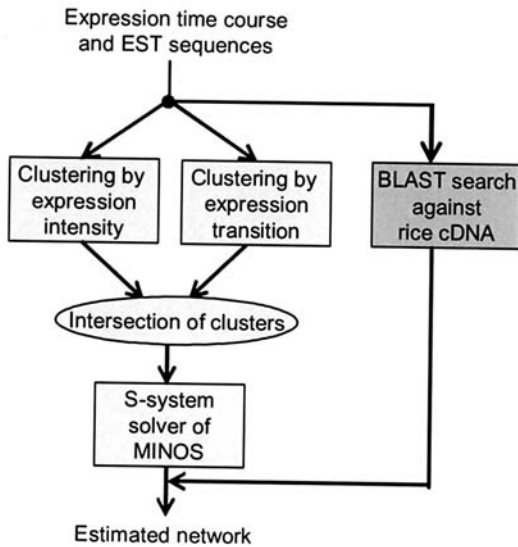


Fig.5 Analysis flowchart for cold inducible sugarcane ESTs. The expression profiles of 25 ESTs are provided by Nogueira et al. 2003.

文献

- [1] Tominaga D. et al., Nonlinear numerical optimization technique based on a genetic algorithm for inverse problems: towards the inference of genetic networks. Proc. German Conference on Bioinformatics, 127-140, 1999
- [2] Voit E. O. and Almeida J., Dynamic profiling and canonical modeling: powerful partners in metabolic pathway identification. Kluwer Academic Publishing, Dordrecht, The Netherlands, 2003
- [3] Nogueira F. T. S. et al., RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiology, 132: 1811-1824, 2003
- [4] Savageau M. A., Biochemical systems analysis: A study of function and design in molecular biology. Addison-Wesley Pub. Co., 1976
- [5] Tanaka N. et al., Expression and function of proteins during development of the basal region in rice seedlings. Mol. Cell. Proteomics, 4: 796 – 808, 2005
- [6] Kikuchi S. et al., Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science, 301, 5631: 376-379, 2003
- [7] Higo K. et al., Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research, 27, 1: 297-300, 1999

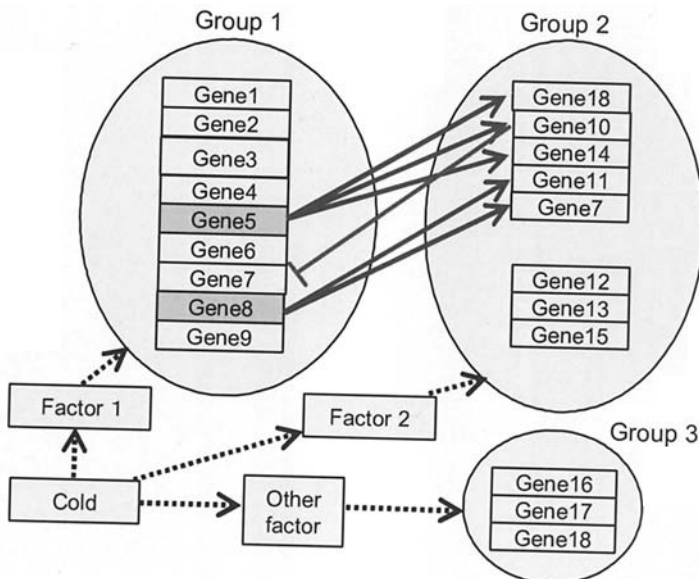


Fig.6 Estimated model from an integrated analysis of mathematical solutions and sequence analysis.