

表現型を利用した遺伝子ネットワーク推定

阿部奈津美 (指導教員:瀬々潤)

1 概要

細胞の形態は観測が容易である上、細胞内の状態を反映するものとして、遺伝学で重要視されてきたが、主に定性的な形態の変化を調べていたのみであった。図 1,2 に遺伝子破壊前の酵母と遺伝子 SHE4 を破壊後の酵母の例を示す。この図のように、特定の遺伝子破壊により、芽の大きさや、芽と母細胞を結合する角度など、様々な表現型が変化している。これに対し、大矢ら [1] は遺伝子を網羅的に破壊した出芽酵母についての形態変化を 512 のパラメータにおいて定量的に観測し、遺伝子と形態の関係を明らかにした。そして、この形態変異データから、各遺伝子は独立して形態を制御しているのではなく、協調して形態を制御したり、複数の形態を同時に制御したりしている事が推察される。そこで本研究では、この形態変異データから、形態を制御するネットワークを記述することで、各遺伝子が制御している形態がどのようなものであるかを明らかにする。

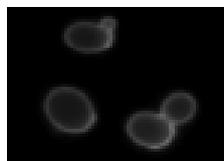


図 1: 遺伝子を破壊する前の酵母

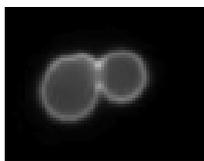


図 2: SHE4 遺伝子を破壊した後の酵母

2 本研究の動機付け例題

本研究で求めるネットワーク例を表 1 および図 3 に示す。表 1 は、破壊した遺伝子とそれに伴って変化した形態パラメータの番号を示している。例えば、遺伝子 G_A を破壊するとパラメータ番号 1,2,3 が変化している。この表から本研究では、図 3 に示すようなネットワークを求めたい。図 3 では、円が遺伝子を、四角が形態を、矢印が制御関係を示している。表 1 では、 G_A が破壊されると形態 1,2,3 が変異していたが、図 3 のネットワークでは G_A が破壊されると、その下流に関連している形態が変異することを示している。同様に、遺伝子 G_B, G_C についても G_B の下流には形態 2,3 が G_C の下流には形態 3 が関連付けられているので、それぞれ形態 2,3 と形態 3 が変化することがわかり、表 1 の情報が矛盾無く図 3 のネットワークに著されている。そして、ここから上流の遺伝子はそれより下流の遺伝子すべての制御に関わっているといえる。

図 3 は最も簡単な例であり、本研究では実際の形態データから細胞内で活動している情報を推定する為、分岐や合流のあるネットワークやノイズを含むデータの扱いも考慮する。

表 1: 形態変異データの例

破壊した遺伝子	変異した形態
G_A	1,2,3
G_B	2,3
G_C	3

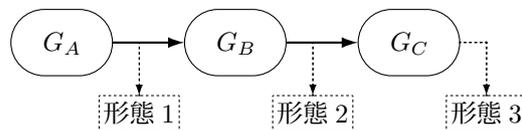


図 3: 表 1 から求めた形態変異のネットワーク

3 関連研究

遺伝子ネットワーク推定の研究は、今までマイクロアレイなどにより採取された発現量情報から行われてきた [2]。しかし今後、画像解析技術などの進歩により、表現型データの採取が進むと思われる。一方で、表現型に関連したネットワーク推定はいまだ行われていない為、本研究を行う事により、今後の表現型情報解析の手法確立を行う。

4 手法

大矢ら [1] は 512 の形態のパラメータに渡り定量的に観測しているが、ここでは簡単のため各パラメータについて、変異があるか、無いかの 2 値の値としてとらえる。遺伝子 G を破壊したときに変異する形態パラメータのセットを $B(G)$ と表す。

$$B(G) = \left\{ x \mid x \in \text{G(遺伝子)を破壊したときに変異を起こした形態パラメータ} \right\}$$

例えば遺伝子 G_a を破壊したとき形態パラメータ 1,2,3,4 に異常が起こったとしたら $B(G_a) = \{1, 2, 3, 4\}$ となる。表現型の変異をこのように表すことで遺伝子は表現型 1,2,3,4 の制御に関わっているといえる。

ここで遺伝子 G_a, G_b に対し $B(G_a) = \{1, 2, 3, 4\}, B(G_b) = \{3, 4\}$ と仮定する。この時、 G_a と G_b のいずれかが遺伝子ネットワーク上、上流にあるかを考えよう。 G_a を破壊したことで変異した形態パラメータ 1,2,3,4 は、 G_b を破壊した事で起こる形態パラメータ 3,4 を包含している ($B(G_a) \supseteq B(G_b)$) から、 G_a の方が G_b より形態への影響が大きく、よって G_a が G_b の上流にあると考えられる。以上は理想的な例であり、生命情報の観測にはノイズがあるため、必ずしも変異パラメータが包含関係にあるとは限らない。そこで、遺伝子 G_m と G_n の間に次のような距離 $D(G_m, G_n)$ を定める。

$$D(G_m, G_n) = \frac{|B(G_n) \setminus B(G_m)|}{(|B(G_m)| |B(G_n)|)^k}$$

ここで $k (> 0)$ は任意に定めるパラメータである。 k は後に構成される遺伝子ネットワークの1ノード内に含まれる遺伝子数を調節し、遺伝子がネットワークの上流や、下流に固まって表示されることを防ぐために設定する。

この式は、遺伝子 G_n が遺伝子 G_m の下流に存在しやすい程度を表現しており、値が小さいほど下流に存在しやすい。例えば、先ほどの $B(G_a), B(G_b)$ の例で $k = \frac{1}{2}$ とし、 $D(G_m, G_n)$ を計算すると、それぞれ $D(G_a, G_b) = 0, D(G_b, G_a) = \frac{1}{\sqrt{2}}$ となる。

$D(G_a, G_b) < D(G_b, G_a)$ より、 G_a は遺伝子ネットワーク上で G_b の上流にあると推定できる。

次に、定義した距離を利用し、遺伝子間の関係を考える (図4)。

まず、(1) のグループ化を行う。もし遺伝子 G_m, G_n において $D(G_m, G_n) = D(G_n, G_m) = 0$ であれば、その2つの遺伝子が誘発する形態異常は同じであり同一の形態異常を誘発するグループと考えられる。このような場合図4(1)のように表記する。この時、多少、誘発する形態変異に違いがあっても観測誤差により距離が生まれてしまう事があつて、閾値を M とし $D(G_m, G_n) < M, D(G_n, G_m) < M$ であれば、同一のグループと見なす。この操作は2つ以上の遺伝子間に対しても適用し、グループ化できるものはすべてグループ化する。

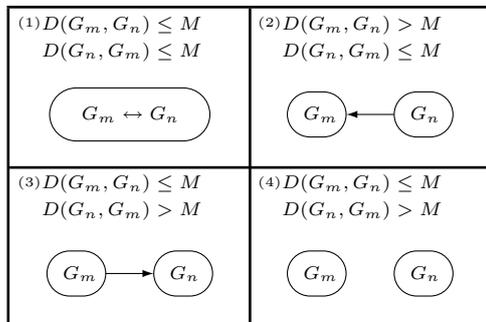


図4: 距離とノード・エッジの関係

その後各グループ間にエッジを引く。この時、距離は下流に存在しやすい程度を表現していたので、例えば図4(2)では、遺伝子 G_n から G_m の距離が小さいので、 G_m は G_n の下流に存在しやすい為、図4(2)のようにネットワークが描ける。また、両遺伝子間の距離が遠い場合、両遺伝子間に関係はないとみて、図4(4)に示すように両ノード間にエッジはひかない。

以上の方法により、遺伝子ネットワークは構築できるが、例えば G_d, G_e, G_f 間に $G_d \rightarrow G_e \rightarrow G_f$ という関係がある時、 G_e を除いた $G_d \rightarrow G_f$ のエッジも引かれてしまう。遺伝子ネットワークでは、上流の遺伝子は下流の遺伝子の制御に関わることが前提なので、このような冗長なエッジを削除する。

5 結果

ここでは $M = 2.0, k = 0.375$ とし、ネットワークを構築する。対象を遺伝子間ネットワークの重要な転写制御関連遺伝子 (241 個) かつ異変が起こった表現型が多い

(12 以上) 遺伝子に絞りに構築した。実行環境は Mac OS X 10.4, 使用言語は Java5. CPU は、Core 2 Duo 2.0GHz であり、実行時間は 1 分 30 秒である。計算の結果、ノード数 66, エッジ数 193, のグラフが得られた。一部を図5(A)に示す。

図5で四角に囲まれた ARG2 と CTK2 に着目し、これらの遺伝子の KEGG [3] パスウェイ上での位置を図5(B)に示す。図5(A)はイノシトールリン酸化代謝経路の一部と一致し、本手法の信頼性が確かめられた。この他、HXK2 \rightarrow THI3 や、CYC8 $\rightarrow \dots \rightarrow$ SWI5 などの関係が既知のネットワークのものと一致しており、本手法で描いたネットワークが高精度であることが示された。

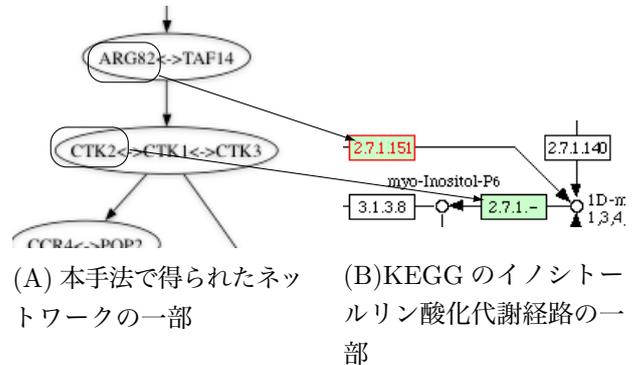


図5: 本手法のネットワークと KEGG の比較。

6 今後の課題

このプログラムとデータでは、非循環グラフしか作成出来ない為、フィードバック制御に対応できない。また、下流の方では、「遺伝子ネットワークでは上流の遺伝子はそれより下流の遺伝子すべての制御に関わる」という遺伝子ネットワーク上での決まりが守られない可能性があるため、検証したい。さらに、上流の遺伝子に比べ下流の遺伝子は変異している形態が多いはずであるため、あるノードからエッジが出る為の変異している細胞形態の違いを調べ、ネットワークと細胞形態についての関係の明確化を行いたい。

参考文献

- [1] Ohya, et.al *High-dimensional and large-scale phenotyping of yeast mutants*. Proc. Natl. Acad. Sci. Vol. 102. No.52. 19015-19020, 2005
- [2] T.Akutsu, S.Miyano, and S.Kuhara. *Inferring qualitative relations in genetic networks and metabolic pathways..* Bioinformatics, 16, 727-734, 2000
- [3] M. Kanehisa, and S. Goto. *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Res, 28: 27-30, 2000. Proc. Natl. Acad. Sci. 98, 4569-4574, 2001.