

ゲノムスキャンニング

7F-4

2次元電気泳動パターンの画像解析

渡辺弥寿夫*、中沢政幸**、堀岡雅清*、林崎良英***

*金沢工業大学、**石川職業能力開発短期大学校、***理化学研究所

1. はじめに

ゲノム解析の研究は、現在世界的な規模で精力的に進められており、中でも、ゲノムマッピングの高速化技術の開発は、重要なテーマの一つである。ゲノムDNAの全塩基配列の決定には、ゲノム上のいかなる座位(landmark)をどのような信号として検出して、いかに多くの座位を検出し高速に解析するかという点が鍵になる。多数の座位について一度にスキャンニングできれば、複数のゲノムDNAに適用して、欠失・増幅・転座などの変化を検出することができる。

近年、林崎が中心となってゲノムスキャンニングの画期的な方法、RLGS(Restriction Landmark Genomic Scanning)法を開発した。これは、まず、ゲノムDNAを制限酵素で切断し、その末端をラジオアイソトープで直接放射標識した後、それを二次元電気泳動法で展開・分類し、オートラジオグラフィを行う方法である。

詳細は、文献¹⁾に譲るが、この方法の特徴として、(1)1枚のゲルで2000~3000のスポット座位が解析できる。(2)ランドマークとなる制限酵素の種類を変えることによりスキャンニングの精度をいくらかでも上げることができる(3)スポットの強度は、ゲノムのランドマークのコピー数を反映する、などがある。

ところが、オートラジオグラフィで得られたX線フィルム上の数千のスポットを読みとる作業は現在、人手で目視に頼っているため、読みとり速度は遅く、人為的なミスも入る。そこで、RLGSパターン的高速自動解析システムの開発を目指して、RLGSパターンの画像処理における基本的課題とその解決法について検討を行ったので、ここに報告する。

2. RLGSパターンの画像解析における基本的課題

図1. は、RLGS法により得られた画像パターン(14"×17")の一例である。このような画像から解析すべきことは、主として(1)各スポットの位置と濃度、(2)微かなパターンの相異がある複数の画像におけるスポット

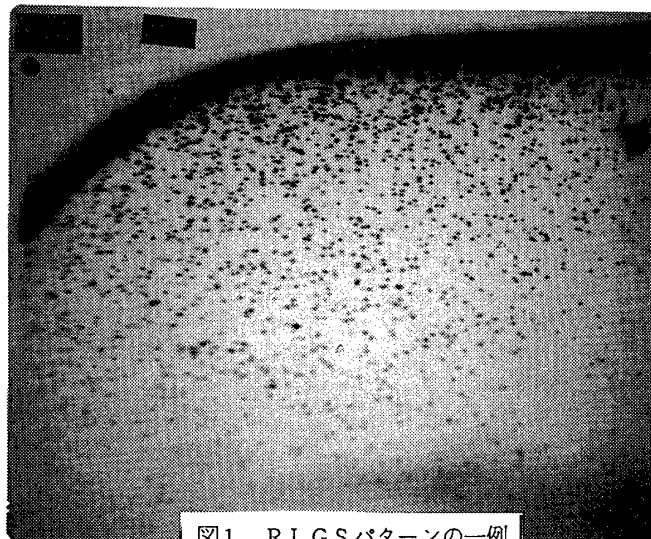


図1. RLGSパターンの一例

の照合・比較である。以下に、処理の概要を述べる。

3. RLGSスポットの位置検出と幾何学的構造の抽出

RLGSパターンは、まず、X線フィルムスキャナで読み込み(1760×2140画素、8ビット/画素)、前処理として平滑化操作を行う。この画像を $\phi(x, y)$ とする。次に、画像 ϕ に対して2値化を行った画像 $f(x, y)$ 、 $|\text{grad}\phi|$ に対して2値化を行った画像 $g(x, y)$ を求める。その後、 $f(x, y) \cdot g(x, y) = 1$ なる範囲($|\text{grad}\phi \neq 0|$)で、リングオペレータによって $\phi(x, y)$ のスポットの位置を検出し、ラベルを付ける。

ここで、画像全体には、ゆるやかなシェーディングがあるので、ある程度、重なりをもたせて、全体を256×256のサイズの部分画像に分割し、その中で、大津の2値化処理²⁾を施す。また、リングオペレータの出力 $h(x, y)$ は、次式のように、濃淡画像の極大点を検出するためのものである。

$$h(x, y) = \int_C \phi(u, v) du dv - \int_R \phi(u, v) du dv$$

ただし、 $C(x, y) = \{(x, y) \mid (u-x)^2 + (v-y)^2 \leq r_m^2\}$

$$R(x, y) = \{(x, y) \mid r_m^2 \leq (u-x)^2 + (v-y)^2 \leq r_M^2\}$$

次に、ドロネー網(Delauney net)の部分グラフである相対隣接グラフRNG(Relative Neighbourhood Graph)⁴⁾を利用して、幾何学的構造を抽出する。ドロネー網とは、平面上の2点が次に示すボロノイ領域(Voronoi region)の意味で隣接するとき、線分で結んだグラフである。³⁾

Image Analysis of RLGS Pattern with Two-Dimensional Electrophoresis for Genome Scanning:
Yasuo Watanabe*, Masayuki Nakazawa**, Masakiyo Horioka*, Yoshihide Hayashizaki***
*Kanazawa Institute of Technology, **Ishikawa Polytechnic College, ***The Institute of Physical and Chemical Research

$$V(p_i) = \{p \in R^2 \mid d(p, p_i) < d(p, p_j), j \neq i\}$$

ただし、 d はユークリッド距離

さらに、RNGは、ドロネー網において枝でつながっている2点 p_i, p_j に対して、次の条件を満たす枝を除いたものである。

$$d(p_i, p_j) \geq \max\{d(p_i, p_k), d(p_j, p_k)\}, k \neq i, k \neq j$$

ここでは、各々のノードに各スポットの位置座標や特徴を属性として持つ属性付きRNGとして表現する。

属性：スポットの位置座標、隣接するスポットの数、スポットの濃度、隣接するスポットへの相対位置ベクトル

このうち、スポットの濃度以降の属性は、特徴ベクトルとして、照合のために利用する。

図2は、部分画像のスポットの位置とRNGを求めた例である。画像の中心付近に見える3角形は、人が目視で照合するため、X線フィルム上に描いたものであり、RLGSパターンとは全く関係ない。

4. RLGSパターンの照合

RLGSパターンは、前記のように構造化すれば、2枚の画像パターンを照合比較することにより、欠失・増幅・転座を検出することが可能となる。

2枚のRLGSパターンは、ほとんど同じスポットを持つが、微妙な非線形の位置ずれが見られる。そのため、前節のように構成された属性付きRNGと原画像を用いて照合する。グラフの照合は、最も計算量の多い問題の一つであるが、ノードの属性と原画像上のパターン照合を利用して、ほとんど後戻りなしにトップダウン的に行うことができる。

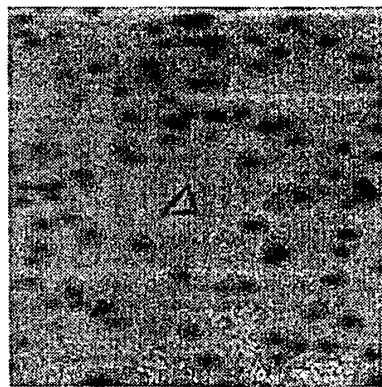
画像全体の非線形な写像を求めることは、能率が悪いので、階層的に画像分割を行い、部分画像で、対応関係を求める。対応は、部分画像間でパターン照合を行い、最も良い位置を、その対応を表すベクトルとする。このベクトルは、RNGのノードの対応すなわち、スポットの対応を求めるとき大まかな位置を予測するのに利用する。このとき、部分画像において、スポットが多い特徴的な領域から対応をとることが望ましいと考えられるため、その領域の濃度分散が大きいところから照合を開始する。また、スポットの対応がつかない場合、それを除去したときのRNGを生成して再び照合する。この後戻りは、高々1つだけ許すこととし、さらに対応がつかないようなら、ここで打ち切り、別の特徴的な領域のノード(スポット)から、追跡する。

図3は、図2の画像に対し、それと同様のパターンをもつ別の画像との照合結果を示している。対応する点が存在する場合■、対応点が無い場合□としている。

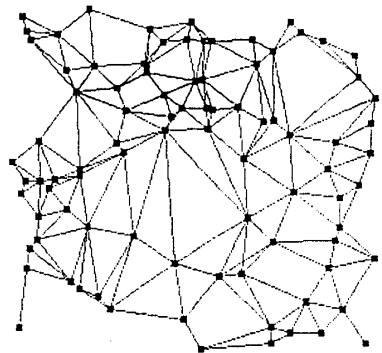
5. おわりに

本稿で示したゲノム解析の画像処理技術は、RLGS法を支援し、高速なゲノムマッピング技術に、貢献すると思われる。今後の課題として、スポットの重なりがある場合や、尾を引くような場合の位置検出、さらに、多くのRLGSパターンに対して実験を重ねて、実用化をはかりたい。

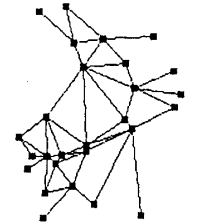
最後に、本研究のためにX線画像を提供していただいた 弘田 憲史氏(NBT(株))および、X線フィルムスキャナ入力にご協力いただいた 真田 茂氏(金沢大学医療技術短期大学部)に感謝いたします。



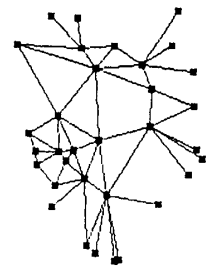
(a) 図1. の部分画像



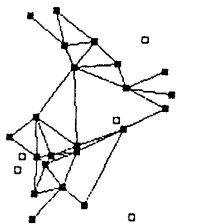
(b) (a)のスポットの位置とRNG



(a)



(b)



(c)

図2.

図3. 照合結果

(処理途中)

[参考文献] 1) Y. Hayashizaki, et al. : "Restriction landmark genomic scanning method and its various application," *Electrophoresis*, 14, pp. 251~258 (1993)

2) 大津 : "判別および最小2乗規準に基づく自動しきい値選定法," *信学論*, J63-D, 4, pp. 349~356 (1980)

3) 伊理(監)、腰塚(編)他 : "計算幾何学と地理情報処理," 共立出版 (1986)

4) G. T. Toussaint : "The relative neighbourhood graph of finite planar set," *Pattern Recognition*, 12, pp. 261~268 (1980)