

DNA 2次元電気泳動像による 遺伝子情報解析システムに関する研究

5 G - 1

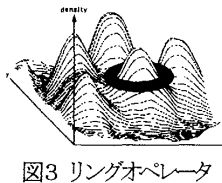
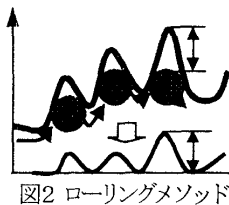
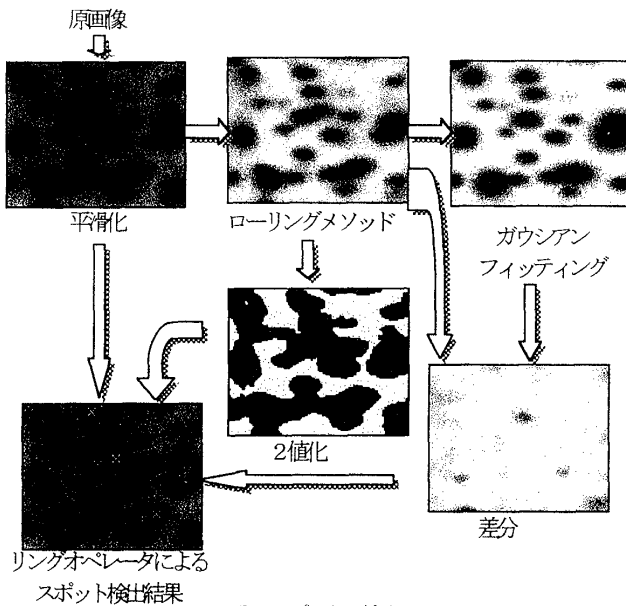
渡辺 弥寿夫 勇川 淳一 戸田 貴之 福澤 勝之
金沢工業大学

1. はじめに

遺伝子情報解析は、ヒトゲノム計画をはじめとする多くの関連研究プロジェクトの進行とともに、目を見張るばかりの進展を遂げてきた。しかし、ヒトのように30億bpという大きなサイズのDNAを一度に解析したり、遺伝子異常の検出や遺伝子発現制御機構の推定を効率良く行うには、従来技術だけでは困難である。そこで、我々は、それが可能なRLGS(Restriction Landmark Genomic Scanning)法²⁾に着目し、自動化を目指して研究を行ってきた³⁾。今回は、その画像解析とデータベース技術について報告する。

2. スポットの検出

スポット検出の概略手順は、次のとおりである。①画像の平滑化、②ローリングメソッド(図2)による背景処理、③背景処理した画像の2値化、④背景処理した画像に対するガウス関数によるフィッティング処理、⑤背景処理とフィッティングによって作成された画像の差分取得、⑥2値化により絞り込んだ領域及び差分画像に対するリングオペレータ(図3)の適用。



3. スポットの照合

RLGSパターンは非線形な歪みと濃度の変動があるため、スポットの照合は単なる画像の重ね合わせでは、うまくいかず。また、従来の方法^{3,4)}でも、歪みが大きいときやある部分領域にしか対応点が存在しないときは照合がうまくいかず。そこでスポットパターン⁵⁾の構造としてドロネー網DN(Delauney Net)とその部分グラフである相対隣接グラフRNG(Relative Neighborhood Graph)を利用して照合を行う方法を開発してきた⁶⁾。DNより枝の数が少ないRNG(レファレンス)をガイドラインとしてDN(オブジェクト)中に対応する節点(スポット)を求めることによって、探索の数を減らし、高速化を行っている。

グラフの探索過程において、レファレンスであるRNG上の節点に連結する枝*i*とオブジェクトであるDN上の節点に連結する枝*j*との整合を関数 $E(i,j)$ 、
$$E(i,j) = \sum_x [(1 - S(a_i^{(k)}, F(a_i^{(k)})))^2 + (1 - S(P_r(a_i^{(k)}), P_o(F(a_i^{(k)}))))^2]$$
 ... (1)

を用いて評価する。但し、 S はベクトル u, v に関して、

$$S(u, v) = \frac{(u, v)}{\|u\| \|v\|}$$
 ... (2)

と定義される。 $a_i^{(k)}$ は、RNGにおける深さ*k*の*i*番目の枝とし、 $F(x)$ はRNGの枝*x*からDNの枝への写像である。また、 $P_r(x), P_o(x)$ は、レファレンス画像、オブジェクト画像の各々において*x*の終端点を中心とする正方形領域(一辺=1)の画像パターン(ベクトル)をあらわすものとする。 $E(i,j)$ を最小にする $F(x)$ が $a_i^{(k)}$ に対応する枝であると考え、それぞれの連結された点を対応づけ、 $a_i^{(k)}$ 方向にグラフ探索を継続する。但し、

$$S(a_i^{(k)}, F(a_i^{(k)})) \leq Th_e, |a_i^{(k)} - F(a_i^{(k)})| \geq Th_r, Th \geq E(i,j)$$
 ... (3)

のいずれかの条件を満たさない場合、枝は対応しているとは考えず、それ以降のグラフ探索を中止する(Th_e, Th_r, Th は閾値)。

本アルゴリズムでは、レファレンス上の点に対応するオブジェクト上の点が存在しない場合、グラフ探索はそれ以上進行しない。従って、その先に対応点が存在していたとしても、この対応点を認識しないままグラフ照合が終了してしまう可能性がある。これを防ぐため、冗長な仮想枝を設定した上でグラフ探索を行う。1枚の画像には、2000個前後のスポットがあり、その照合にPentium 166MHzのDOS/Vコンピュータを用いた場合、約8~12秒の処理時間で、93~97%の精度の結果が得られる。

図4は $Th_e = \cos(\pi/6), Th_r = 10, Th = 0.04, l = 41$ で実験を行ったときの結果であり、■は対応するスポットが存在する所で、□は対応するスポットが存在しない所を示している。

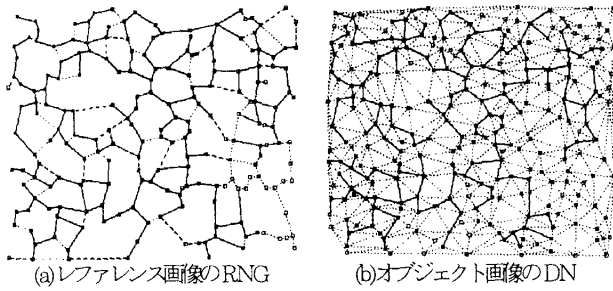


図4 スポットの照合の結果

4. 標準スポットパターンへの作成とスポット出現率

RLGS 画像のスポットパターンは、個体差があり、またガンなどの遺伝子異常によっても異なっている。そこで全 RLGS スポットパターンを統合した、標準スポットパターンの生成を行う必要がある。本研究では、既に作成された複数組のスポットマッチングテーブルを利用して、次のような手順で標準スポットパターンを作成している。

- ① 座標の基準となる RLGS 画像の決定、② 全スポットの基準画像上での座標取得、③ クラスタ分析による同一スポットの検出、④ 各クラスター内のスポットの平均値取得、⑤ 各スポットについての出現率取得、⑥ 標準スポットパターン出力。

この内②の処理は、ドロネー網を利用した三角形の重心座標による線形変換によって行う (図5)。また③の処理では、2つのクラスター間の距離として、クラスター内に含まれる全スポットの座標から求めた分散 σ_x^2 と分散 σ_y^2 を使って得られる式(4)の d を利用し、

$$d = \sqrt{\sigma_x^2 + \sigma_y^2} \quad \dots (4)$$

d の値が最小となるクラスター同士を、新たに1つのクラスターとしてまとめていき、 d がある閾値以上になるまで処理を繰り返す。こうして、クラスターを代表する一点を決定する (図6)。

標準スポットパターンの生成後、各スポットが何枚の画像中何枚の画像に現れているのかを取得し、スポットの出現率を求めるのが⑤の処理である。各スポットの出現率は疾病毎に違ってくる重要な情報となる。

このようにして生成された各スポットの出現率を含む標準スポットパターンは、図7のようなファイルとして生成される。

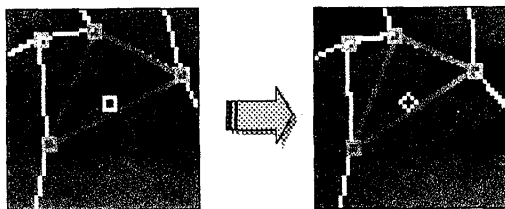


図5 三角形の重心座標によるスポット位置の線形変換

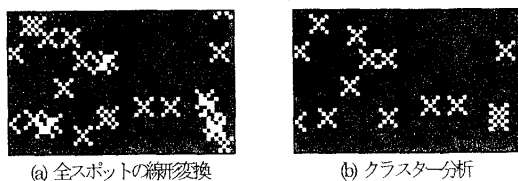


図6 クラスタ分析による標準パターンのスポットの生成

##	xloc	yloc	int	vol	area	state	Nexst/Ns	Textst/Ts
1	194	491	114	3610	95	0	10 15	3 18
2	211	403	108	3869	101	0	12 15	16 18
3	242	404	133	2787	92	0	6 15	7 18
4	360	562	124	539	10	0	14 15	14 18
...								

##: スポット番号
 xloc/yloc: スポット座標
 int: スポット座標の強度値
 vol: スポットの累積強度値
 area: スポットの面積値
 state: スポット検出手法
 Nexst/Ns: 全正常パターン上でのスポット出現率
 Textst/Ts: 全異常パターン上でのスポット出現率

図7 標準スポットパターンテーブル

5. データベース

これまで述べてきた画像データや画像処理から派生する様々な解析結果を遺伝子診断に利用するために、我々はデータベースを構築している。RLGS画像やスポットの情報、マッチング結果などのデータはデータベースに格納され、必要な情報はすべてこのデータベースから参照される。このデータベースを利用することによって、膨大なデータを効率良く管理することが可能となり、各々の処理間のデータの受け渡しを円滑に行うことができる。

例として、スポットテーブルについてのデータ検索を図8に示す。

##	xloc	yloc	int	vol	area	state	条件:
1	454	123	211	1367	35	0	スポットテーブルのデータのうち 'vol>1400' であるもの
2	346	115	212	0	0	0	SQL文:
3	231	109	239	0	0	0	select * from spot where vol > 1400;
4	329	92	197	1589	37	0	
5	369	84	227	1764	36	0	
6	290	118	230	2408	40	0	

検索

図8 SQLによるデータ検索例

6. おわりに

このシステムを利用することによって RLGS パターンの解析に対する大幅な時間と労力の削減が期待できる。また、今後このシステムによって構成されたデータベースを利用し、生成された標準スポットパターンを解析して、ある特定の疾病に対してのみ出現あるいは欠失するスポットなど、疾病毎に特徴のあるスポットのパターンを検出することで、実用的な遺伝子診断システムとして構築していく予定である。

本研究を行うにあたり、ご指導を賜りました北陸先端科学技術大学院大学の 高橋勝利先生に心から感謝の意を表します。

[参考文献]

- 1) 林崎, 畑田, 広常, 向井: "ゲノムスキャニング法—restriction landmark genomic scanning(RLGS)法", 蛋白質・核酸・酵素, Vol.38, No.3, pp.573-580 (1993).
- 2) Y.Watanabe and K.Takahashi: "A Fast Structural Matching and Its Application to Pattern Analysis of 2-D Electrophoresis Images", Proc. Int. Conf. Pattern Recognition (1998).
- 3) S.T.Barnard and W.B.Thompson: IEEE Trans., Pattern Analysis and Machine Intelligence, PAMI-2, pp.333-340 (1980).
- 4) M.A.Fischler and R.A.Eischlager: IEEE Trans. Computer, C-22, 1, pp.66-92 (1973).
- 5) S.Ranade and A.Rosenfeld: Pattern Recognition, 12, pp.269-275 (1980).
- 6) K.Takahashi, M.Nakazawa and Y.Watanabe: "DNAinsight An Image Processing System for 2-D Gel Electrophoresis of Genomic DNA", Genome Informatics 1997, pp.135-146 (1997).