High-Speed Genome Scanning Analysis Based on Automated Detection and Matching Spots in Autoradiogram Images of 2-D Gel Electrophoresis

Tomoyasu Takahashi [1]
takahasi@infor.kanazawa-it.ac.jp
Katsutoshi Takahashi [1]
sltaka@infor.kanazawa-it.ac.jp
Masayuki Nakazawa [2]
nakazawa@nakazawa.ishikawa-pc.ac.jp
Yasuo Watanabe [1]
watanabe@infor.kanazawa-it.ac.jp

[1] Kanazawa Institute of Technology,
7-1 Ohgigaoka, Nonoichi-machi, Ishikawa 921, JAPAN

[2] Ishikawa Polytechnic College,
I 45-1 Yuigaoka, Anamizu-machi, Ishikawa 927, JAPAN

Abstract

In this paper, we present the pattern analysis of autoradiogram images obtained in two-dimensional gel electrophoresis based on the RLGS (restriction landmark genome scanning) method [1]. On a RLGS autoradiogram image, several thousands of spots can be observed; one spot corresponds to one genetic locus and its intensity reflects the copy number of the restriction landmark. In the first step of the computer analysis, geometrical location and density of each spot was detected by applying some kinds of operators, including gradient and ring operators, to the autoradiogram image. As the second step, structured features of the spot pattern was extracted and represented as structured graph. Two or more structured graphs can be matched and compared by computer efficiently, enabling high-speed and fully-automated scanning of genomic DNA.

1 Introduction

RLGS (restriction landmark genome scanning) method developed by Hayashizaki et. al [1] is one of the most powerful experimental procedures which can be applied to high-speed genome mapping. Through this method, restriction enzyme sites on genomic DNA are used as landmarks, being guideposts on the genome. These landmark information can be scanned over whole genome by employing direct end-labeling of genomic DNA fragments and high-resolutional two-dimensional gel electrophoresis. Such landmarks can be detected as several thousands of spots on the autoradiogram images of two-dimensional gel electrophoresis. Since each variant spot segregates as an independent genetic locus, simultaneous mapping of a large number of loci can be achieved by recognizing locations and intensities of the spots. It is impossible, however, either to separate all of the several thousands of spots or to determine their intensities by visual inspection. It is also difficult by hand to compare two or more images, in order to perceive the translocation, amplification and deletion of spots. Such difficulties on RLGS analysis make the high-speed multiplex genome mapping impractical. To overcome the above problems, we have developed automated image processing scheme which extracts locations and intensities of several thousands of spots simultaneously and reliably. We also propose the computer representation of RLGS spot patterns and reliable pattern matching scheme.

2 Detection of RLGS spots

The analysis of RLGS spots begins with digitization of X-ray film with an image scanner. The digitized image is, then, preprocessed for image enhancement and smoothing. Here, let the preprocessed image be $\phi(x,y)$. By applying a thresholding operator [2] to $\phi(x,y)$ and $|grad\phi|$, we obtain the resultant binary images $f(x,y)$ and $g(x,y)$, where

$$f(x,y) = \begin{cases} 1, & \text{if} \phi(x,y) \geq t_1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad g(x,y) = \begin{cases} 1, & \text{if} |grad\phi| > t_2 \\ 0, & \text{otherwise} \end{cases} .$$

As the proceeding step, a ring operator is applied to $\phi(x,y)$ on the domain $\{(x,y) \mid f(x,y) \cdot \overline{g(x,y)} = 1\}$ to detect the maximal points, as follows.

$h(x,y) = \max \phi(x,y) - \max \phi(x,y)$, where $C(x,y) = \{(u,v) \mid (u-x)^2 + (u-y)^2 \leq r_M^2\}$ and $R(x,y) = \{(u,v) \mid r_m^2 \leq (u-x)^2 + (u-y)^2 \leq r_M^2\}$.

Finally, the spots detected in the above steps are labeled for their identification.

3 Extraction and representation of structured features

Patterns of spot locations and intensities found on RLGS autoradiogram image should be extracted and should be represented as computer accessible form for further analysis. Here, we represent the RLGS spot pattern as two kinds of structured graph, that is, the Delaunay net and relative neighborhood graph [3], rather than a simple set of feature points. Two or more RLGS patterns represented as structured graphs can be matched by means of a graph search method [4]. The high-speed and reliable comparison reveals the differences among RLGS patterns, such as translocation, amplification and deletion of the spots, which specify the changes in genomic DNA.

4 Summary

We have investigated an analysis system of autoradiogram images for genome scanning, which contributes to automation of the RLGS method and spurs the development of a gene-based-diagnosis system.

Acknowledgment

We are grateful to Dr. Hirota at NBT Inc. for his technical supports on our research, and to Prof. Ueda at Keio University for his kind donation of RLGS autoradiogram images.

References

[1] Y. Hayashizai, et al. : Electrophoresis, 14, 251-258 (1993).
[2] K. Palaniappan, et al. : Proc. of Int. Workshop on Industrial Application of Machine Intelligence and Vision, 360-367(1989)
[3] G.T. Tousssaint : Pattern Recognition, 12, 261-268 (1980)
[4] F. Preparate and M.I. Shaos : Computational Geometry :An Introduction, Springer-Verlag, New York (1985)