A Computer Algorithm for the Standardization of Marker Size Obtained by Electrophoresis for Microsatellite Marker Analysis

Tsukamoto, Yasuo

Shiozawa, Shunichi

(Citation)
Bulletin of allied medical sciences Kobe : BAMS (Kobe),13:137-141

(Issue Date)
1997-12-26

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(URL)
https://hdl.handle.net/20.500.14094/00188180

A Computer Algorithm for the Standardization of Marker Size Obtained by Electrophoresis for Microsatellite Marker Analysis

Yasuo Tsukamoto and Shunichi Shiozawa

We have applied the microsatellite marker anaysis to identify the gene loci that predispose to rheumatoid arthritis (RA). During execution of an improved computational algorithm after Holmans & Clayton, which is able to estimate the missing marker genotypes of parents by incorporating one unaffected sibling into an affected sib pair, we encountered the problem that the sizes of microsatellite markers obtained by electrophoresis were not always concentrated strictly around the expected, two-base distant, values. To overcome this problem, we have devised a new algorithm to standardize the size of marker alleles for automatic computational processing. Three feasible markers were identified under the criterion of the lod score being more than the prescribed level 3, which is generally used for the linkage analysis.

Key Words

Marker size,
Computational algorithm,
Linkage analysis,
Affected-sib-pair method,
Rheumatoid arthritis.

INTRODUCTION

A recent development in molecular biology has enabled us to investigate the cause of intractable diseases in a precise manner. We have applied the microsatellite marker analysis[1] to the study of genetic trait of rheumatoid arthritis[2]. Since this application was new and without preceding studies, we needed to advent some devices for the calculation of the result of family study using microsatellite marker analysis.

Faculty of Health Science, Kobe University School of Medicine, Kobe

For estimation of linkage, we employed the affected sib-pair method which was commonly used for clinical genetic studies. The affected-sib-pair method proposed by Rish[3,4] is based on the total information of marker genotypes of parents together with their affected sib-pair. This method is advantageous since it can be used for the disease without the knowledge of Mendelian inheritance. While, in the case of RA, it is often impossible to type the markers of parents because of late onset of the disease. We have employed a modified affected-sib-pair method of Holmans & Clayton[5], which enabled us to estimate the missing genotype of parents by incorporating an unaffected sib into the affected sib pair[6].

Holmans & Clayton introduced a likelihood approach by assigning a probability z_j to each possible amount of j (j=0,1,2) marker alleles identity-by-descent (IBD) shared between the affected sib-pair, where the restriction

is that $z_0 + z_1 + z_2 = 1$. The likelihood for all the family was maximized with respect to z_j among all markers in each family. We have developed a practical algorithm[6] for the purpose of applying the classical sib-pair linkage method to the study of families with RA. One of the problems we encountered during the study was a very small but significant gel-to-gel variation in sizing each marker, and therefore we tried to standardize the marker size obtained from different sets of electrophoresis. The purpose of the present study was to develop a practical algorithm for the standardization of the marker size for total family in microsatellite DNA sizing.

METHOD

Study design A panel of 41 Japanese families, each with at least two affected siblings, was typed for genome-wide 358 polymorphic microsatellite marker loci, with a mean heterozygosity of 0.79. A sex-averaged genetic distance between two consecutive marker loci was 10.8cM. Microsatellite markers were amplified by the polymerase chain reaction (PCR) using fluorescence-tagged primers and sized based on the difference of CA repeats on DNA fragments using ABI 377 sequencer operated under the software GENESCAN™. The DNA fragments were properly sized by lane-to-lane gel comparison and automatic allele calling using GENOTYPER™. We have typed one unaffected sibling in addition to the affected sib-pair, since parents were not available for typing in most of the RA families at the time of the diagnosis of RA because of late onset nature of the disease.

Family study Japanese families including affected sib-pairs who met the diagnostic criteria of the American College of Rheumatology (formerly, the American Rheumatism Association[7]) with clear and identifiable joint destruction of more than stage 2 of Steinbroker's X-ray classification[8] were studied. Peripheral blood (10ml) was withdrawn by using EDTA and gently mixed with 20ml of buffer I (0.32M sucrose, 5%v/v Triton X-100, 5mM $MgCl_2$, 12mM Tris HCl (pH 7.6)) to lyse cell membrane. After centrifugation, the precipitate of nuclei was reacted with buffer II (4M guanidine thiocyanate, 12mM EDTA, 375mM NaCl, 0.5% sodium N-lauroyl sarcosinate, 0.1M β-mercaptoethanol, 12mM Tris HCl (pH7.6)) at 55℃ for 10min to lyse nuclear membrane. Genomic DNA was then extracted by ethanol precipitation.

Microsatellite DNA sizing[9] Fluorescence-tagged primers for 358 microsatellite markers that define a 10.8cM resolution human index map were purchased from PE Applied Biosystems-Roche Molecular Systems, Inc. (Branchburg, New Jersey). However, the markers D1s502 and D6s344 were not used because of technical difficulty in amplifying their DNA and the marker D6s276 was substituted to D6s299, D6s265 and D6s273 to specifically examine the HLA-D region. Other microsatellite markers were newly synthesized according to the published primer sequences[10] (Biologica Co., Nagoya, Japan). PCR was performed in 96-well microtiter plates (# 6511, Corning Coster Corp., Cambridge, MA) in

a 15λl volume containing 30ng of DNA, 0.2λM of primer mixture, 0.2mM of each dNTP, 1 unit of AmpliTaq Gold DNA polymerase (PE Applied Biosystems), 2.5mM MgCl2 and 1x PCR buffer II (PE Applied Biosystems), and amplification condition in a MJ Research PTC-100 thermocycler was 94℃ for 10 min followed by 27 cycles of 94℃ for 30 sec, 55℃ for 1 min, 72℃ for 2 min, and 72℃ for 5 min after the last cycle. PCR products labeled with 6-FAM, TET or HEX were mixed together with TAMURA-labeled size standards and applied on a 36cm-well-to-read gel (4% acrylamide/ 6M urea) plate. Electrophoresis was carried out in a ABI 377 sequencer (PE Applied Biosystems) at 3,000V for 2hr. The peak, height and area of DNA fragments were identified using the GENESCAN (Ver.2.0.0) computer analysis software (PE Applied Biosystems)
GENOTYPER™ (Ver.1.1) (PE Applied Biosystems).

Numerical analysis Linkage analysis was made using the computer program SIB-ADONE which was developed by us[6] according to the method of Holmans & Clayton, whose work opened the way for the clinical use of the possible triangle method of Risch. This approach was based on the IBD (identity-by-decent) sharing states of the affected pair only, using the unaffected siblings to infer parental genotypes, and was expected to yield a close approximation to the full likelihood especially when penetrance was low.

RESULTS AND DISCUSSION

From theoretical viewpoints, the size of two different microsatellite markers should be separated at least two-base distance, because of the nature of PCR amplification. The set of marker sizes for the same locus should then be represented by odd number or even one. The markers measured using electrophoresis technique were, however, found to be often distributed in continuous and not discrete fashion around the expected values. Then, the marker size must be standardized before going into the calculation of the frequency of each sized marker in the population.

If the marker distribution coincided with an odd scale rather than an even scale, the integrated power of the histogram of markers for the odd scale would be expected to be much greater than the one for the even scale. According to such an idea, we invented an evaluation function for the odd and even case, respectively, as follows:

$$Q_o = \sum_{i=1}^{n} h_{o,i}^{\,2} \qquad (1)$$

$$Q_e = \sum_{i=1}^{n} h_{e,i}^{\,2} \qquad (2)$$

where Q_* (*=odd or even) is the evaluation function, $h_{*,i}$ is the histogram of the marker i standardized by odd or even scale in the population, and n is the total number of standardized classes of sized markers for a locus of interest. n' in equation (2) is not necessarily the same as n in equation (1). According to the evaluation function, we proposed a criterion that, if $Q_o \geqq Q_e$, then the markers should be classified with respect to the odd scale. If otherwise, they were classi-

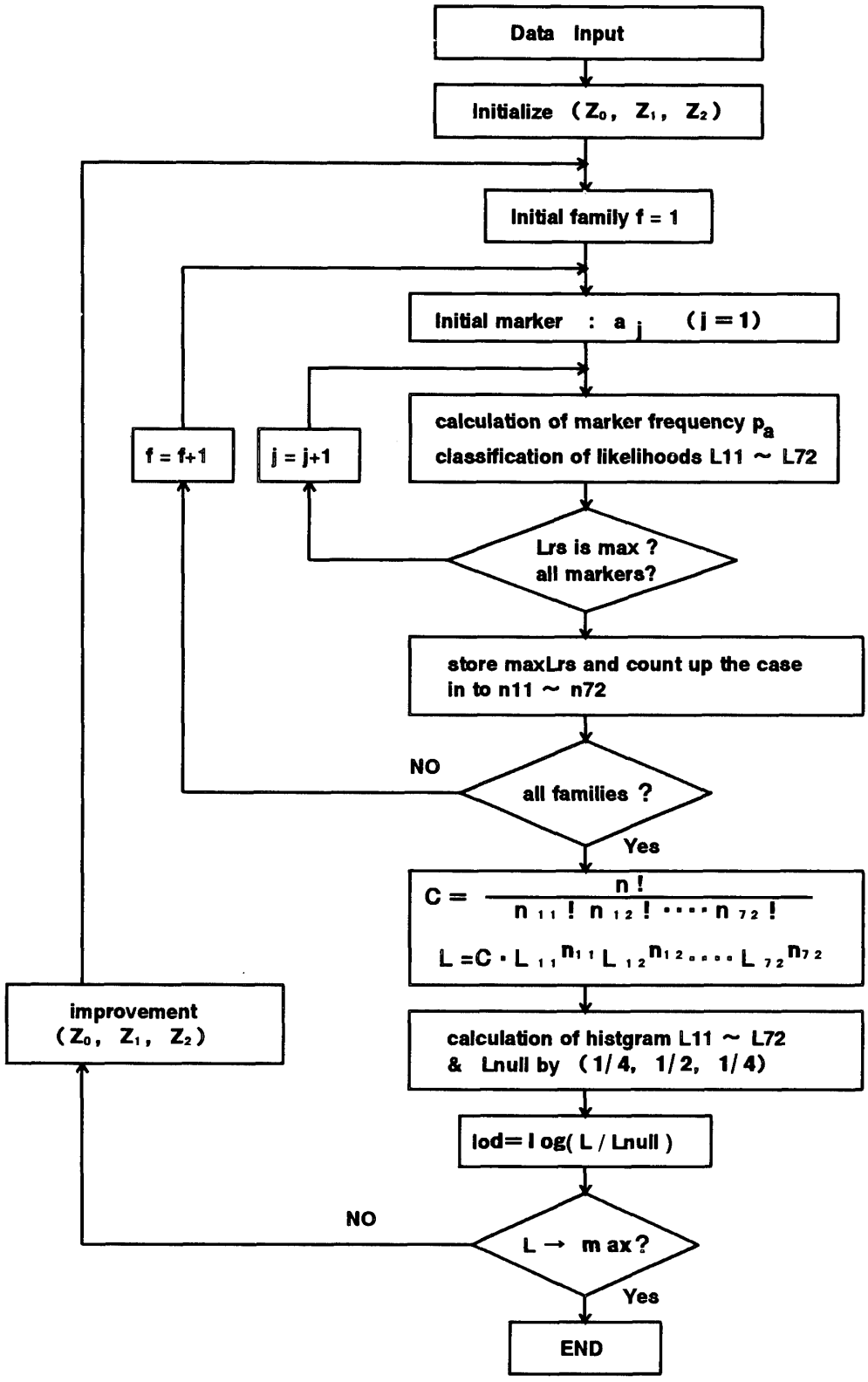

Fig.1. Improved computational algorithm for linkage analysis by an affected sib pair with one unaffectedsibling

fied with respect to the even scale.

Taking this evaluation function into account for preprocessing the marker size, we proposed a computational algorithm shown in Fig. 1 to obtain the condition for maximizing the likelihood. In the figure, p_a is the gene frequency of the marker a_j, and L_{rs} and n_{rs} is the likelihood and the number of the sibship, respectively, typed by the marker a_j and the family f of interest. According to the likelihood L_{rs} and the amount n_{rs} for each category, the total likelihood L for all families can be described by a multinomial distribution. The algorithm in Fig. 1 was composed mainly of three iteration loops; the maximization of likelihood L_{rs} for each marker a_j, each family f and a set of z-values (z_0, z_1, z_2). These procedures were programmed and calculated by using a personal computer.

The algorithm compiled with the proposed method was applied to estimate the gene loci responsible for the development of RA. According to the maximum likelihood obtained by the proposed method, which was integrated into our SIB-ADONE program[6], we evaluated the extent of linkage by the lod score, which was calculated for the log ratio by dividing the maximized likelihood by its value under null hypothesis of $z_{null}=(1/4,1/2,1/4)$. Three principal regions of linkage, D1s214, D8s556 and the locus between DXS1001 and DXS1227, were identified under the criterion of the lod score being more than the prescribed level 3, which is generally used for the linkage analysis[2]. It should therefore be possible to assign the gene loci of RA disease to the region near the markers D1 S 214, D8 S 556 and the locus between DXS1001 and DXS1227[2].

REFERENCES

1. Vyse T, Todd JA. Genetic analysis of autoimmune disease. Cell 85:311-318,1996
2. Shiozawa S, Hayashi S, Tsukamoto Y, et al. Identification of the gene loci that predispose to rheumatoid arthritis. ,submitted.
3. Risch N. Linkage strategies for genetically complex traits. Ⅲ. The effect of marker polymorphism on analysis of affected relative pairs. Am J Hum Genet 46:242-253,1990
4. Risch N. Assessing the role of HLA-linked and unlinked determinations of disease. Am J Hum Genet, 40:1-14,1987
5. Holmans P, Clayton D. Efficiency of typing unaffected relatives in an affected-sib-pair likage study with single-locus and multiple tightly linked markers. Am J Hum Genet 57:1221-1232,1995
6. Tsukamoto Y, Yasuda N, Shiozawa S. Improved computational argorithm for typing unaffected relatives in an affected-sib-pair linkage study. , submitted.
7. Amett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31:315-324,1988
8. Steinbroker O, Traeger GH, Batterman RC: Therapeutic criteria in rheumatoid arthritis. JAMA 140:659-662,1949
9. Reed PW, Davis JL, Copeman JB, et al. Chromosome-specific microsatellite sets for fluorescence-based, semiautomated genome mapping. Nature Genet 7:390-395,1994
10. Gyapay G, Morissette J, Vignal A, et al. The 1993-94 Genethon human genetic linkage map. Nature Genet 7:246-339,1994