Development of a simulation system for SLE disease model mice crossing experiment

Nobutaka KUWAHARA[1)], Yasuhiro IIDA[2)], Hiroyuki NISHIMURA[1)] and Yukiyasu IIDA[1)]

(2011年3月15日　受理)

1. Introduction

With the growth of genetic research, it has been learned that there are two types of disorders. One is single-gene disorder due to serious abnormality of one gene such as hemophilia, Duchenne muscular dystrophy and Huntington's disease, the other is multifactorial genetic disease due to two or more genetic factor and an environmental factor such as diabetes and high-blood pressure.

In the analysis of a genetic disease, it is effective to use the disease model mouse that presents the symptom similar to the disease. In other words, it is very important for a clarification of human affected gene to specify a causative gene of the model mouse and to analyze an influence of the causative gene.

To analyze causative genes using disease model mouse, researchers utilize quantitative statistical genetics tools such as Mapmaker/QTL[1] and MapManager QTX[2] based on gene data of hybrid mice. However, crossing experiments need to prepare hundreds of disease model mice to get enough genetic data of hybrid mice, it takes several years to analyze those data, and it is difficult to modify the experimental conditions.

One of our goals is to design a more accurate crossing experimental plan targetting for a disease model mouse which suffers from SLE (systemic lupus erythematosus)[3]. The other is to develop a simulation program which enables to estimate an influence of an environmental factor and a genetic factor which relates disorders.

2. Systemic lupus erythematosus (SLE) and the disease model mouse

Systemic lupus erythematosus (SLE) is one of diseases which discriminate between self and non-self, generate an antibody to biological macromolecules such as own DNA and the proteins and destroy them. SLE can affect any part of a body. An appearance of a red rash on the skin resembles a wound bitten by a wolf, so this symptom is called Systemic lupus(lupus is Latin for wolf) erythematosus. SLE causes fever, general malaises, various symptoms with the joint, the skin, and internal organs, etc. Though it's not known

[1)] Faculty of Biomedical Engineering, Toin University of Yokohama, 1614 Kurogane-cho, Aoba-ku, Yokohama, Japan 225-8503
[2)] Information sharing platform Laboratories, NTT

exactly why, it is thought that the disease receptivity is formed by a combination of gene polymorphisms. Although several genetic analyses of SLE patients and their families have been done recently, it is not clear the causative gene now.

In the analysis of a genetic disease, it is effective to use the disease model mouse.

NZB (New Zealand Black) is known as one of mice that causes an autoimmune systemic disease.

We study a backcross normal mice and F1 mice which hybridized NZB mice and normal mice, which produce self-tolerance, or F2 inter cross mice (F1×F1). NZW(New Zealand White Mice) or C57BL/6 Mice is used as a normal mice according to study objectives.

3. Simulation system

3-1 QTL Analysis

Phenotype, which varies in degree such as a blood sugar and a weight, is called quantitative character (quantitative trait; QT). This quantitative character is thought to be determined by a combination of alleles of several gene loci. QTL Analysis is to clear a behavior of quantitative trait loci(QTLs) which relates to the quantitative character. Some representative examples of QTL Analysis software are MapMaker/QTL and MapManager QTX.

We can estimate a location of a gene locus that affects a related trait of diseases by analyzing genes data of hybridized mice with these software. Our simulation system makes simulated gene information of hybridized mice. We can obtain a wide knowledge of diseases by analyzing our simulation outcomes with QTL analysis software.

3-2 Mathematical model of our simulation

Our simulation is based on the assumption that a causative gene thought to be participation in the disease exists on two or more chromosomes. The relation among a phenotype value, a heritable factor, and an environmental factor was shown by the expression of (3-1) according to Fisher's polygenic model[4].

$$V_i = \Sigma A_j * X_{ji} + E * N(0,1) + C \quad \cdots\cdots (3\text{-}1)$$

where V_i is an phenotype value of the diseases mouse of i. and X_{ji} is a random variable that takes value of 0 or 1 whether causative gene that takes part in phenotype on target chromosomes of j, A_j is a coefficient of X_{ji}, $N(0,1)$ is a random variable of an environmental factor, can be expressed as a random number of a standard normal distribution (that is, the normal random number of mean value 0 and dispersion 1). E is a coefficient of an environmental factor, C is a constant coefficient.

In our simulation, we determine A_j,E, and C adjusting for the real data of a mice crossing experiment.

We have to consider a crossover for gametes transmitted to children from parents. A crossover is an exchange of genetic material between chromosome of mother and that of father. Though the probability of occurring a crossover is low, there is theoretically no limitation in frequency and occurs at random on the location of chromosome. The probability distributions are known to be a Poisson distribution when the event probability is very low and there is no limitation in frequency.

The probability that an event occurs k times is obtained in expression (3-2).

$$P(k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \cdots\cdots (3\text{-}2)$$

where e is Napier's constant and λ in expression (3-2) is an expected value (that is, average frequency in which the event happens).

A genetic distance is shown by an expected value of a crossover frequency. When the expected value in which the crossover in the targeted chromosome occurs is one time, the genetic distance is 1 M (=Morgan). The crossover occurrence probability is so low that cM (=centi Morgan; 1/100 M) is also used as the unit of the genetic distance. The genetic distances vary depend on the chromosome, and that of mouse is 0.7~1.3 M.

In the first step of our simulation, crossover frequency is determined based on the Poisson distribution. We set the value of the Poisson distribution parameter λ based on the length of the target chromosome. For example, λ is set to 0.9 at a chromosome of 0.9 M.

In the second step, crossover points are randomly chosen on the target chromosome.

The interference is known that a crossover rarely occur when the crossover point is near the former point.

The final step of our simulation, we set a genetic distance threshold θ in which crossover does not occur.

There are two widely known methods for generating normally distributed random numbers N(0,1) which stands for an environmental factor,

[Method A] : One is a method of generating 12 uniform random numbers and subtracting 6.

[Method B] : Another is a Box-Muller transform[5].

With the Box-Muller transform, the outcome of the expression 3-3 is a random number of normal distribution of average 0 and standard deviation 1, where α and β denotes random numbers on the interval (0, 1].

$$\sqrt{-2\log(\alpha)} \cdot \sin(2\pi\beta) \qquad \cdots\cdots (3\text{-}3)$$

Generally speaking, the method of B is more accurate than that of A. So we adopt the method of B.

Figure 1 shows a distribution of A and B respectively generating 10,000 random numbers.

4. Verification of our simulation system

To verify our simulation system, we conducted a verification experiment which evaluates (1) a frequency of genotype, (2) recombination rate. In the verification experiment, F2 inter cross simulation of 10,000 mice for 121 markers on 20

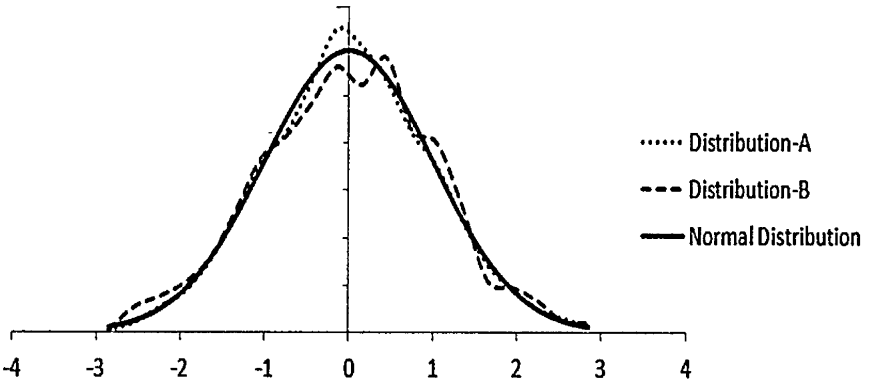

Figure 1 Comparison between distribution-A and distribution-B

chromosome of a mouse.

0 stands for an allele of a normal mouse, 1 stands for an allele of a disease mouse, and (a, b) stands for duplotype of marker xx a stands for a father, b stands for a mother

Combinations of (a, b) are (0,0)、(0,1)、(0,1)、(1,1). If F1 mice are breed randomly, these 4 pairs occurs at an equal probability.

Because genotype distinguishes neither (0,1) nor (1,0), the frequency of 0/0, 0/1, and 1/1 Genotype becomes 0.25 and 0.5 and 0.25 respectively.

Table 1 shows the frequency obtained by the verification experiment, and indicates good agreement with a theory.

Figure 2 is a graph of a genetic distance M (horizontal axis) and a recombination rate (vertical axis) varying the parameter θ expressing an interference of a crossover described in 3-2. When there is no interference (that is, $\theta = 0$), the graph identifies the Holden's map function[6] (that produces the recombination rate from the genetic distance). And then, we confirmed that Kosambi's map function[6] considering an interference is located between two graphs of θ =16cM and $\theta = 32$cM.

Though we do not discuss a genetic distance here, we can speculate that a crossover simulation considering an interference is possible with selecting an appropriate θ.

5. Conclusion

We develop the crossing simulation program of SLE disease model mouse as part of identifying the causative gene of SLE disease studies. Our goal of this study is to clarify the influence of the genetic factor and the environmental factor relating SLE quantitatively. We adopted Fisher's polygenic model in that how the genetic factor and the environmental factor affect the phenotype value.

In the simulation, we decided the number of occurrences of chromosome crossover according Poisson distribution, then decided a crossover point as the number of occurrences, and finally decided the allele of a normal mouse and the disease mouse as marker points respectively. The frequency of genotype(a pair of alleles) obtained by the

Table 1 Frequency of genotypes

Genotype	Frequency
0/0	0.249
0/1	0.501
1/1	0.250

0:allele from a normal mouse
1:allele from a disease model mouse

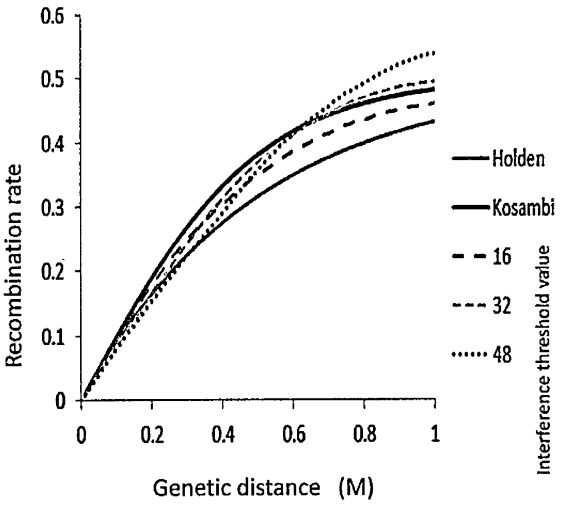

Figure 2 Graph of recombination rate

simulation shows good agreement with the actual data. In our simulation, we modeled the interference in which a crossover rarely occur when the crossover point is near the former point. As a result, we obtained the function while varying genetic distances that is quite close to Kosambi's map function known to be considering the interference, we confirmed that our simulation can demonstrate the actual crossing.

In our simulation analysis, coefficients described by 3-2 must be close to the actual data. Therefore, we have to vary the coefficient value, generate data, and then perform QTL analysis. If we can use QTL analysis software from our simulation system, we can automate this procedure and analyze more efficiently. As MapMaker/QTL, one of the QTL analysis software, is an open source software and available to the public, we are going to adopt it to our simulation system as our future work.

REFERENCE

[1]M.J.Kearsey, and V.Hyne "QTL analysis: a simple 'marker-regression' approach", Theor Appl Genet, Vol.89, No.6, p.698-702, 1994

[2]Kenneth F. Manly, Robert H. Cudmore, Jr. and Jane M. Meer "Map Manager QTX, cross-platform software for genetic mapping", MAMMALIAN GENOME, Vol.12, No.12, 930-932, 2001

[3]Robert George Lahita, Systemic lupus erythematosus, 3rd Revised edition, Elsevier Science & Technology Academic Press, 1998

[4]Fisher RA: The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb, 52:399-433., 1918

[5]S. C. Chay, R. D. Fardo and M. Mazumdar, "On Using the Box-Muller Transformation with Multiplicative Congruential Pseudo-Random Number Generators", Appl. Statistics, Vol.24, No.1, p.132-135, 1975

[6]Jurg Ott, J "Analysis of Human Genetic Linkage", Third Edition, The Johns Hopkins University Press, 1999