Two-Dimensional Interval Mapping toward Construction of Maker Interval Networks

Master's Thesis

Department of Computational Biology
The University of Tokyo

Nakaya Laboratory
46905 Yasumasa Suzushino

Abstract

Identifying a set of loci affecting a trait that shows a continuous distribution of values is attracting great interest. Such loci are called *quantitative trait loci* (QTLs) and a series of analyses including prediction of QTL locations along chromosomes is called QTL analysis. During the primary analysis, each candidate QTL location is individually investigated. The framework called *interval mapping* (IM), which is most frequently used in QTL analysis, follows this single scan strategy. However, since the distribution of quantitative trait values is considered to be affected by an aggregation of multiple QTLs, the single scan strategy introduces approximation by neglecting the correlated relationships among multiple loci, reducing detecting power of the method.

To cope with this problem, we develop a new framework called *two-dimensional interval mapping* (2DIM). This framework can evaluate simultaneous effects of the loci by improving the single scan framework so that it can carry out estimation in a multi-dimensional manner. After introducing the statistical background of two-dimensional interval mapping, usefulness of this method is also demonstrated by using mainly the real datasets of mice, from a practical point of view.

On the other hand, two-dimensional interval mapping can reveal another aspect of QTL analysis. The two-dimensional method can extract pairs of marker intervals that are significant to the quantitative trait of interest instead of significant intervals that are obtained by using the usual interval mapping in a one-dimensional manner. This means that the whole set of the significant interval pairs constitutes a network among marker intervals. We call the network a *marker interval network* to emphasize that it can represent the synergetic architecture behind the trait. Applying the method to schizophrenia model mice, we constructed an interaction map among marker intervals in relation to this disease. Using the interaction map, correlated behavior among genetic factors is discussed.

We also applied two-dimensional interval mapping to more than twenty datasets of mice quantitative traits that are publicly available. The results are stored in a database named *marker interval network database* (MINTDB). This database provides analytical facilities focusing on characteristics of the network architecture in the datasets. Using them, we can compare and integrate marker interval networks each of which is significant to a specific quantitative trait.

Contents

Chapter 1

Introduction

A quantitative trait is a phenotype that is shown by a continuous amount. Body height and body weight are examples. Since the distribution of values of a quantitative trait is considered to be regulated by multiple genetic factors, effects of combination of genetic factors to the quantitative trait are attracting interest. Although such correlated behavior among genetic factors is essential to a quantitative trait, analytical techniques that can cope with interactions among those factors are not necessarily sufficient.

1.1 Motivation

A gene locus that is concerned with a quantitative trait is called a quantitative trait locus (QTL). Series of analyses including prediction of locations of such QTLs on chromosomes is called QTL analysis. The history of QTL analysis can go back to, for example, Sax (1923) who showed that two traits of haricot color (qualitative) and weight (quantitative) were mutually interacting, and the former can be used as a kind of a marker of the latter. Based on this observation, Thoday (1961) showed that locations of quantitative loci can be identified by using qualitative traits. However, since the number of the qualitative traits available was limited, sufficient results were not obtained. Currently, densely located DNA markers such as micro-satellite markers are available, without suffering from such the problems as in the past, we can carry out QTL analysis precisely to our heart's content.

1.2 Two-dimensional interval mapping (2DIM)

Effects of combinations of multiple genetic factors are often concerned with a quantitative trait. Interval mapping method that estimates significance of at possibility of existence a QTL at each single location on chromosomes is insufficient. Some extension of the framework of interval mapping is required. In this research, we developed a new framework called two-dimensional interval mapping that can extract significant pairs of chromosome segments in relation to the behavior of correlated characteristics behind a quantitative trait. Two-dimensional interval mapping can evaluate interactions among genetic factors (i.e., epistatic effects). It can take into account additive and dominant effects of each candidate QTL. We discuss the efficiency of two-dimensional interval mapping by using the dataset of schizophrenia model mice.

1.3 Marker interval network

Two-dimensional interval mapping can reveal another aspect of QTL analysis. The two-dimensional method can extract pairs of marker intervals that are significant to the quantitative trait of interest instead of significant intervals that are obtained by using the usual interval mapping in a one-dimensional manner. This means that the whole set of the significant interval pairs constitutes a network among marker intervals. We call the network a *marker interval network* to emphasize that it can represent the synergetic architecture behind the trait. Applying the method to schizophrenia model mice, we constructed an interaction map among marker intervals in relation to this disease. Using the interaction map, correlated behavior among genetic factors is discussed.

On the other hand, in two-dimensional interval mapping, as well as the traditional one-dimensional interval mapping, estimated QTL locations between markers are indicated in genetic map distance (recombination fraction) along chromosomes. Thus, except the locations just on the marker loci, results of two-dimensional interval mapping cannot be compared directly with data using physical map distance (base pairs). By comparing marker interval networks and gene networks, for example, protein-protein interaction (PPI), we tried to find candidate genes within the significant interval pairs. The results of preliminary analysis are presented using the marker interval network in relation to size of sexual organs of fly.

1.4 Marker interval network database (MINTDB)

We also applied two-dimensional interval mapping to more than twenty datasets of mice quantitative traits that are publicly available. The results are stored in a database named *marker interval network database* (MINTDB). This database provides analytical facilities focusing on characteristics of the network architecture in the datasets. Using them, we can compare and integrate marker interval networks each of which is significant to a specific quantitative trait.

1.5 Outline of this Thesis

Chapter 2 introduces two-dimensional interval mapping. In Chapter 3, using two-dimensional interval mapping, marker interval networks are constructed in relation to mice datasets. In Chapter 4, fly marker interval networks are compared to and integrated with gene networks. In Chapter 5, discussion and future work are described.

Chapter 2

Two-dimensional interval mapping (2DIM)

We develop a new framework called *two-dimensional interval mapping* (2DIM). This framework can evaluate simultaneous effects of the loci by improving the single scan framework so that it can carry out estimation in a multi-dimensional manner. After introducing the statistical background of two-dimensional interval mapping, usefulness of this method is also demonstrated by using mainly the real datasets of mice, from a practical point of view.

2.1 Data for QTL Analysis

The data needed in QTL analysis is a set of values of the target trait and genotypes of marker loci. There are three genotypes at marker loci in the datasets we use in this study: a homozygote of the maternal line, a homozygote of the paternal line, and a heterozygote of both lines. Since the marker loci are discretely located along chromosomes, genotypes at locations (sometimes called pseudomarkers) between flanking markers are estimated by using a map function, for example, Haldane's mapping function in a probabilistic manner.

2.2 Interval mapping

2.2.1 Outline of the Method

As the foundation of the two-dimensional interval mapping, we first formalize the traditional one-dimensional interval mapping. If we assume that quantitative trait values are affected by genetic factors and environmental effects, the genetic model between trait values and genotypes is shown as follows:

$$P = G + E.$$

Here, P represents phenotype values of the quantitative trait, G represents the genotype values, and E is the environmental effects. This model can be rewritten if we want to take additive effects (effects when the genotype is the homozygote) and dominant effects (effects when the genotype is the heterozygote) into account as follows:

$$P = u + a + d + \sigma^2.$$

Here, u is a constant, a represents additive effects, d represents dominant effects, and σ^2 is a residue term that is assumed to be normally distributed. As for additive and dominant effects, it becomes $u + a$ in case of the homozygote of the maternal line, $u - a$ in case of the homozygote of the paternal origin and $u + d$ in case of the heterozygote. σ^2 is assumed to be normally distributed.

Genotype frequency at a pseudomarker is missing. Using an EM algorithm, therefore, the values of the four parameters above are estimated so that the likelihood is maximized and the genetic model can explain the data as much as possible. The EM algorithm is carried out in two steps (the E step and the M step). The expected value of genotype frequency is obtained in the E step and the parameter values are maximized in the M step.

As the initial values, zero is used for a and d, and the average and the variance of the trait values are respectively used for u and σ^2. The E step and the M step are carried out using theses initial values, and both steps are alternately repeated until the values of the four parameters converge. If the estimations of the parameters are obtained, the ratio of the likelihood of the model that assumes a and d are not zero against the likelihood of the null hypothesis that does not assume genetic effects in the trait values. The logarithm of the ratio is the LOD score. This procedure is carried out to calculate the LOD score at each location between the flanking markers.

2.2.2 Details of the Method

To carry out interval mapping, an EM algorithm is designed for each of inbred line. Here, using an F_2 intercross line, details of the EM algorithm are explained. Note that the calculation of the other lines, for example, P_1 backcross, and P_2 backcross and so on can be carried out in the same manner. The details of the E step and the M steps are as follows. Note that those calculations are carried out at each location among flanking markers.

2.2.3 The E Step

At the location where we calculate its LOD score (i.e., at pseudomarker), the genotype is estimated in a probabilistic manner. The genotype at the pseudomarker is estimated using the following nine cases of the flanking marker genotypes.

1: $A_1A_1B_1B_1$ 2: $A_1A_1B_1B_2$ 3: $A_1A_1B_2B_2$ 4: $A_1A_2B_1B_1$ 5: $A_1A_2B_1B_2$
6: $A_1A_2B_2B_2$ 7: $A_2A_2B_1B_1$ 8: $A_2A_2B_1B_2$ 9: $A_2A_2B_2B_2$

Here, A and B represent the flanking markers and the suffixes one and two represent from whether maternal or paternal the marker inherits. On the other hand, the pseudomarker genotype is represented one of Q_1Q_1, Q_1Q_2, and Q_2Q_2. Here, Q represents the pseudomarker between marker A and B. Assume that the probability that a recombination happens between A and Q is r_1, the probability that a recombination happens between Q and B is r_2, the probability that a recombination happens only by one degree between A and B is assumed to be r_{1+2}, and the probability that a recombination happens both between A and Q and between Q and B is r_{12}. When the genotypes of the flanking markers are case *i* as above, we denote the probability that the genotype of the pseudomarker is Q_1Q_1, Q_1Q_2, and Q_2Q_2 by p_{i1}, p_{i2}, and p_{i3}, respectively. Here, *i* ranges from one to nine, and each is corresponds to one of the nine marker genotypes above. The probabilities of p_{i1}, p_{i2}, and p_{i3} are represented as follows.

	Q_1Q_1 (p_{i1})	Q_1Q_2 (p_{i2})	Q_2Q_2 (p_{i3})
1: $A_1A_1B_1B_1$	q_1^2	$2q_1q_2$	q_2^2
2: $A_1A_1B_1B_2$	q_1q_3	$q_1q_4+q_2q_3$	q_2q_4
3: $A_1A_1B_2B_2$	q_3^2	$2q_3q_4$	q_4^2
4: $A_1A_2B_1B_1$	q_{14}	$q_1q_3+q_2q_4$	q_2q_3
5: $A_1A_2B_1B_2$	$z_1q_1q_2+z_2q_3q_4$	$z_1(q_1^2+q_2^2)+z_2(q_3^2+q_4^2)$	$z_1q_1q_2+z_2q_3q_4$
6: $A_1A_2B_2B_2$	q_2q_3	$q_1q_3+q_2q_4$	q_1q_4
7: $A_2A_2B_1B_1$	q_4^2	$2q_3q_4$	q_3^2
8: $A_2A_2B_1B_2$	q_2q_4	$q_1q_4+q_2q_3$	q_1q_3
9: $A_2A_2B_2B_2$	q_2^2	$2q_1q_2$	q_1^2

Here,

$$q_1 = \frac{(1-r_1-r_2+r_{12})}{(1-r_{1+2})} \quad q_2 = \frac{r_{12}}{(1-r_{1+2})} \quad q_3 = \frac{(r_2-r_{12})}{r_{1+2}} \quad q_4 = \frac{(r_1-r_{12})}{r_{1+2}}$$

$$z_1 = \frac{(1-r_{1+2})^2}{\{(1-r_{1+2})^2+r_{1+2}^2\}} \qquad z_2 = 1-z_1 .$$

Using the assumption that the residue terms of Q_1Q_1, Q_1Q_2, and Q_2Q_2 cases are normally distributed, the probability densities Φ_1, Φ_2, and Φ_3 represented as follows.

$$\phi_1 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-a)^2}{2\sigma^2}} \quad \phi_2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d)^2}{2\sigma^2}} \quad \phi_3 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u+a)^2}{2\sigma^2}}$$

Therefore, frequencies of the three genotypes at the pseudomarker described as follows.

Q_1Q_1: $z_1 = \Phi_1 p_{i1} / (\Phi_1 p_{i1} + \Phi_2 p_{i2} + \Phi_3 p_{i3})$
Q_1Q_2: $z_2 = \Phi_2 p_{i2} / (\Phi_1 p_{i1} + \Phi_2 p_{i2} + \Phi_3 p_{i3})$
Q_2Q_2: $z_3 = \Phi_3 p_{i3} / (\Phi_1 p_{i1} + \Phi_2 p_{i2} + \Phi_3 p_{i3})$

2.2.4 The M Step

The M step is carried out by using the result of the E step. The likelihood is represented as follows.

$$L \propto \prod_{i}^{9} \prod_{j}^{n_i} \left(p_{i1}\phi_{ij1}\right)^{z_{ij1}} \left(p_{i2}\phi_{ij2}\right)^{z_{ij2}} \left(p_{i3}\phi_{ij3}\right)^{z_{ij3}}$$

Here, i indicates the genotype of the marker (one of the nine types), and j indicates each individual ($1 \sim n_i$) that has the marker genotype.

Logarithm of the likelihood calculated as follows:

$$\ln(L) = const + \sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}\ln p_{i1} + z_{ij2}\ln p_{i2} + z_{ij3}\ln p_{i3}\right) - 0.5N\ln\left(2\pi\sigma^2\right)$$

$$-\sum_{i}^{9}\sum_{j}^{n_i}\left\{\frac{z_{ij1}\left(y_{ij}-u-a\right)^2}{2\sigma^2} + \frac{z_{ij2}\left(y_{ij}-u-d\right)^2}{2\sigma^2} + \frac{z_{ij3}\left(y_{ij}-u+a\right)^2}{2\sigma^2}\right\}.$$

Here, *const* is a constant, and N is the number of individuals. By differentiating the log-likelihood with respect to u, a, d, and σ^2, and setting the derivatives to zero, we have

$$\frac{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1}y_{ij}}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1}} = \hat{u} + \hat{a}\,,\quad \frac{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3}y_{ij}}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3}} = \hat{u} - \hat{a}\,,\quad \frac{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij2}y_{ij}}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij2}} = \hat{u} + \hat{d}\,,$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i}^{9}\sum_{j}^{n_i}\left\{z_{ij1}\left(y_{ij}-u-a\right)^2 + z_{ij2}\left(y_{ij}-u-d\right)^2 + z_{ij3}\left(y_{ij}-u+a\right)^2\right\}.$$

These equations are sufficient to carry out the M step, but if one wants to have the estimators of each parameter, they are represented as follows:

$$\hat{u} = \frac{\dfrac{\left(\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1} y_{ij}\right)}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1}} + \dfrac{\left(\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3} y_{ij}\right)}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3}}}{2}, \quad \hat{a} = \frac{\dfrac{\left(\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1} y_{ij}\right)}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1}} - \dfrac{\left(\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3} y_{ij}\right)}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3}}}{2},$$

$$\hat{d} = \frac{\left(\sum_{i}^{9}\sum_{j}^{n_i} z_{ij2} y_{ij}\right)}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij2}} - \frac{\dfrac{\left(\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1} y_{ij}\right)}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij1}} + \dfrac{\left(\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3} y_{ij}\right)}{\sum_{i}^{9}\sum_{j}^{n_i} z_{ij3}}}{2}.$$

The details of the derivations are given in Appendix A.1. Now we have updated values of the four parameters. By using these values, the next E step is carried out. Until the values of parameters converge, the rounds of the E step and the M steps are iterated.

2.2.5 LOD Score

A LOD score is calculated using the values of the log-likelihood ln(L) and ln(L_0). ln(L_0) is calculated by the parameters that are obtained under the null hypothesis.

$$LOD = \log_{10} L - \log_{10} L_0 = \frac{1}{\ln(10)}\{\ln(L) - \ln(L_0)\}$$

The constant value (*const*) disappears. P_1 backcross and P_2 backcross can be explained similarly, and refer to appendix A.2 for details.

When the LOD score is obtained, the pseudmarker is moved to the next of 1cM, and the calculation is repeated on all chromosomes. And, the graph where the position on the chromosome (cM) is taken in the horizontal and the LOD score was taken in the ordinate is written. Figure 2 shows the example of applying interval mapping to the dataset of schizophrenia model mice's forced swim test (Yoshikawa et al, 2002, chromosome 3).

2.2.6 An Example

When interval mapping is applied to the datasets of schizophrenia model mice's forced swim test (Yoshikawa et al, 2002), how the LOD score changes by the repetition of the E steps and the M steps in the EM algorithm is shown in Figure 2 (part of chromosome 8). The LOD score rises whenever repeating.

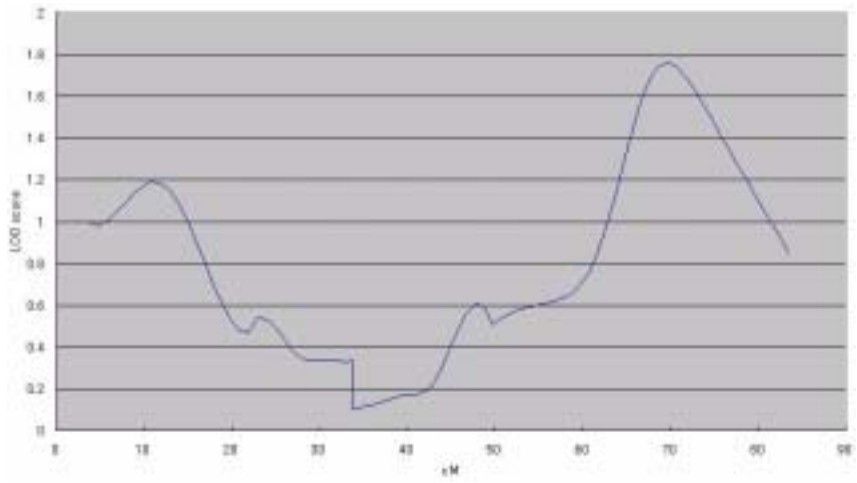


Figure 1 Example of interval mapping

Interval mapping was applied to dataset of schizophrenia model mice's forced swim test (Yoshikawa et al, 2002, chromosome 3).

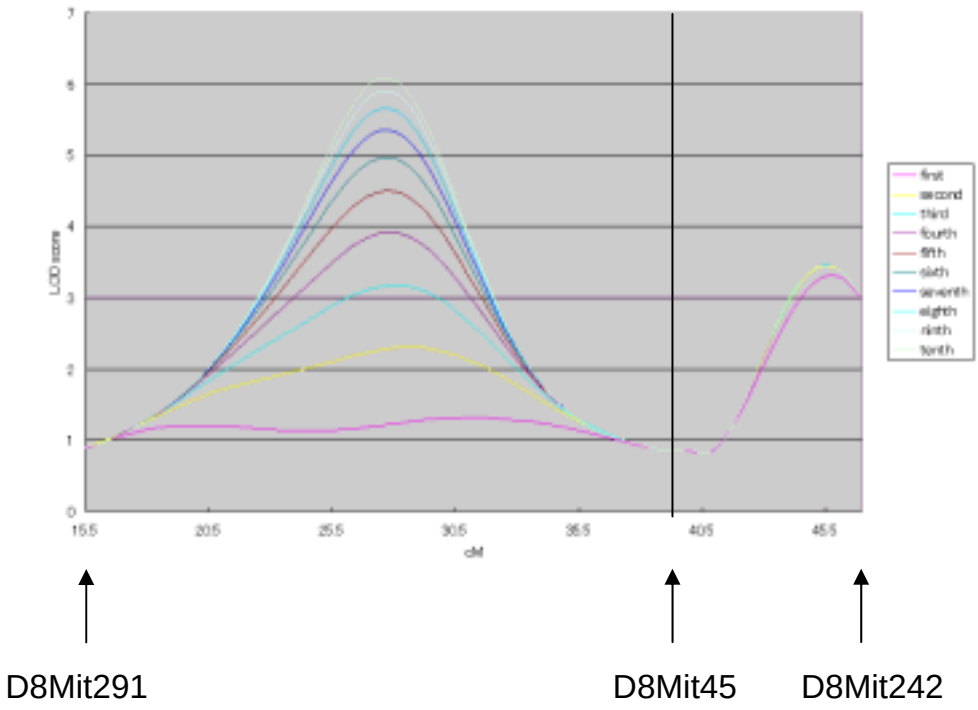


Figure 2 Repetition frequency and LOD score of the E step and the M step in the EM algorithm

Transition of LOD score by repetition frequency of the E step and the M step in EM algorithm when interval mapping is applied to dataset of schizophrenia model mice's forced swim test (Yoshikawa et al, 2002, part of chromosome 8)

2.3 Existing Two-Dimensional QTL Analysis

The QTL analysis method that considers gene loci and their interactions has been

developed before. Insufficient points are of two-dimensional QTL analysis method though TWOSCAN and multiple interval mapping (MIM) are enumerated as an example.

2.3.1 Multiple interval mapping

Multiple interval mapping is QTL analysis method that considered two gene loci (Kao et al, 1999). MIM uses the EM algorithm by advancing interval mapping. The place of QTL is presumed beforehand by the Composite interval mapping (CIM), and it calculates based on the place. The genetic model is shown as follows.

$$y_{ijk} = u_{ij} + \sum_{r=1}^{m} \alpha_{ir} X_{ijkr} + \sum_{r \neq s \subset (1,\ldots,m)}^{t} \beta_{irs} \left(X^*_{ijkr} X^*_{ijks} \right) + e_{ijk}$$

Here, the first term indicates the phenotype value, the second term is the effect of QTL of m piece presumed beforehand, the third term is the effect of the interaction of the pair of QTL (The pair of the same place is excluded) to which is presumed beforehand, and the fourth term is the rest errors.

In MIM, the interaction is considered, but the calculation result is shown in the graph of one dimension of each QTL presumption place, and only the place presumed beforehand is calculated. That is, the place where the LOD score is low is not calculated by the calculation done beforehand, and the place where the LOD score rises for the first time when pairing off cannot be presumed.

2.3.2 TWOSCAN

TWOSCAN is two-dimensional QTL analysis method (Sen et al, 2001). The Monte Carlo algorithm is used instead of the EM algorithm for easing and the flexibility of the calculation. Genetic model's calculation type is shown by the next expression.

$$p_H(y \mid m) = \frac{1}{q} \sum_{s_i=1}^{q} \sum_{u} W_H(r_i(u))$$

Here, y is the phenotype value, m is the marker genotype, u is a present QTL presumption place (1 ~ q piece), r_i is the genotype calculated there, and s is the number of QTL presumption places of p class.

The calculation result of TWOSCAN is shown as two dimensions. The place where the LOD score rises for the first time when pairing off can be presumed, differing from MIM. However, it is an approximation method that uses the Monte Carlo algorithm, and the term of an interaction is only one term.

2.4 Two-dimensional interval mapping (2DIM)

2.4.1 Outline of the method

In this research, it was thought that an existing two-dimensional QTL analysis method was insufficient, and two-dimensional interval mapping that advanced existing interval mapping to two dimensions was developed. Here, it explains two-dimensional interval mapping.

In existing interval mapping, it was a genetic model which consisted of one gene locus and the additive effect and the dominant effect. However, in two-dimensional interval mapping, if we want to superadd two gene loci and epistatic effects, the genetic model can be rewritten as follows:

$$P = u + a_1 + d_1 + a_2 + d_2 + i_{12} + j_{12} + j_{21} + l_{12} + \sigma^2.$$

Here, u represents a genotype value, a_1 represents additive effects in the first pseudmarker, d_1 represents dominant effects in the first pseudmarker, a_2 represents additive effects in the second pseudmarker, d_2 represents dominant effects in the second pseudmarker, i_{12} represents additive × additive epistatic effects, j_{12} represents additive × dominant epistatic effects, j_{21} represents dominant × additive epistatic effects, l_{12} represents dominant × dominant epistatic effects, and σ^2 is a residue term that is assumed to be normally distributed.

The EM algorithm of this genetic model is calculated as well as interval mapping. Here, using an F_2 intercross line, details of the EM algorithm are explained also in two-dimensional interval mapping. Note that the calculation of the other lines, for example, P_1 backcross, and P_2 backcross and so on can be carried out in the same manner.

2.4.2 The E step

At the location where we calculate its LOD score (i.e., at pseudomarker), the genotype is estimated in a probabilistic manner. The genotype at the pseudomarker is estimated using the 81 cases (Multiplication of nine cases ($A_1A_1B_1B_1$ ~ $A_2A_2B_2B_2$) and nine cases ($C_1C_1D_2D_2$ ~ $C_2C_2D_2D_2$)) of the flanking marker genotypes. Here, A and B represent the flanking markers and the suffixes one and two represent from whether maternal or paternal the marker inherits. On the other hand, the pseudomarker genotype is represented one of the following nine cases.

1: $Q_1Q_1Q_3Q_3$ 2: $Q_1Q_1Q_3Q_4$ 3: $Q_1Q_1Q_4Q_4$ 4: $Q_1Q_2Q_3Q_3$ 5: $Q_1Q_2Q_3Q_4$
6: $Q_1Q_2Q_4Q_4$ 7: $Q_2Q_2Q_3Q_3$ 8: $Q_2Q_2Q_3Q_4$ 9: $Q_2Q_2Q_4Q_4$

Here, Q_1 and Q_2 represent the pseudomarker between marker A and B, and Q_3 and Q_4 represent the pseudomarker between marker C and D. The suffixes one and two represent from whether maternal or paternal the pseudmarker inherits, and the suffixes three and four represent from whether maternal or paternal the pseudmarker inherits.

Assume that the probability that a recombination happens between A and Q is r_1, the probability that a recombination happens between Q and B is r_2, the probability that a recombination happens only by one degree between A and B is assumed to be r_{1+2}, and the probability that a recombination happens both between A and Q and between Q and B is r_{12}. Moreover, assume that the probability that a recombination happens between C and Q is r_3, the probability that a recombination happens between Q and D is r_4, the probability that a recombination happens only by one degree between C and D is assumed to be r_{3+4}, and the probability that a recombination happens both between C and Q and between Q and D is r_{34}. When the genotypes of the flanking markers are case *i* as above, we denote the probability that the genotype of the pseudomarker is $Q_1Q_1Q_1Q_1$ ~ $Q_2Q_2Q_2Q_2$ by p_{i1} ~ p_{i9}, respectively. Here, *i* ranges from one to nine, and each is corresponds to one of the nine marker genotypes above. The probabilities of p_{i1} ~ p_{i9} are obtained from the product of the probability in the gene locus of the first pseudmarker and the second pseudmarker.

The probabilities in the gene locus of the first pseudmarker are represented as follows.

	Q_1Q_1 ($p_{i1\sim3}$)	Q_1Q_2 ($p_{i4\sim6}$)	Q_2Q_2 ($p_{i7\sim9}$)
1: $A_1A_1B_1B_1$	q_1^2	$2q_1q_2$	q_2^2
2: $A_1A_1B_1B_2$	q_1q_3	$q_1q_4+q_2q_3$	q_2q_4
3: $A_1A_1B_2B_2$	q_3^2	$2q_3q_4$	q_4^2
4: $A_1A_2B_1B_1$	q_1q_4	$q_1q_3+q_2q_4$	q_2q_3
5: $A_1A_2B_1B_2$	$z_1q_1q_2+z_2q_3q_4$	$z_1(q_1^2+q_2^2)+z_2(q_3^2+q_4^2)$	$z_1q_1q_2+z_2q_3q_4$
6: $A_1A_2B_2B_2$	q_2q_3	$q_1q_3+q_2q_4$	q_1q_4
7: $A_2A_2B_1B_1$	q_4^2	$2q_3q_4$	q_3^2
8: $A_2A_2B_1B_2$	q_2q_4	$q_1q_4+q_2q_3$	q_1q_3
9: $A_2A_2B_2B_2$	q_2^2	$2q_1q_2$	q_1^2

Here,

$$q_1 = \frac{(1-r_1-r_2+r_{12})}{(1-r_{1+2})} \quad q_2 = \frac{r_{12}}{(1-r_{1+2})} \quad q_3 = \frac{(r_2-r_{12})}{r_{1+2}} \quad q_4 = \frac{(r_1-r_{12})}{r_{1+2}}$$

$$z_1 = \frac{(1-r_{1+2})^2}{\{(1-r_{1+2})^2 + r_{1+2}{}^2\}} \quad z_2 = 1 - z_1$$

The probabilities in the gene locus of the second pseudmarker are represented as follows.

	Q_3Q_3 ($p_{i1,4,7}$)	Q_3Q_4 ($p_{i2,5,8}$)	Q_4Q_4 ($p_{i3,6,9}$)
1: $C_1C_1D_1D_1$	q_5^2	$2q_5q_6$	q_6^2
2: $C_1C_1D_1D_2$	q_5q_7	$q_5q_8+q_6q_7$	q_6q_8
3: $C_1C_1D_2D_2$	q_7^2	$2q_7q_8$	q_8^2
4: $C_1C_2D_1D_1$	q_5q_8	$q_5q_7+q_6q_8$	q_6q_7
5: $C_1C_2D_1D_2$	$z_3q_5q_6+z_4q_7q_8$	$z_3(q_5^2+q_6^2)+z_4(q_7^2+q_8^2)$	$z_3q_5q_6+z_4q_7q_8$
6: $C_1C_2D_2D_2$	q_6q_7	$q_5q_7+q_6q_8$	q_5q_8
7: $C_2C_2D_1D_1$	q_8^2	$2q_7q_8$	q_7^2
8: $C_2C_2D_1D_2$	q_6q_8	$q_5q_8+q_6q_7$	q_5q_7
9: $C_2C_2D_2D_2$	q_6^2	$2q_5q_6$	q_5^2

Here,

$$q_5 = \frac{(1-r_3-r_4+r_{34})}{(1-r_{3+4})} \quad q_6 = \frac{r_{34}}{(1-r_{3+4})} \quad q_7 = \frac{(r_4-r_{34})}{r_{3+4}} \quad q_8 = \frac{(r_3-r_{34})}{r_{3+4}}$$

$$z_3 = \frac{(1-r_{3+4})^2}{\{(1-r_{3+4})^2 + r_{3+4}{}^2\}} \quad z_4 = 1 - z_3$$

Using the assumption that the residue terms of $Q_1Q_1Q_3Q_3 \sim Q_2Q_2Q_4Q_4$ cases are normally distributed, the probability densities $\Phi_1 \sim \Phi_9$ represented as follows.

$$\phi_1 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-a_1-a_2-i_{12})^2}{2\sigma^2}} \quad \phi_2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-a_1-d_2-j_{12})^2}{2\sigma^2}}$$

$$\phi_3 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-a_1+a_2+i_{12})^2}{2\sigma^2}} \quad \phi_4 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d_1-a_2-j_{21})^2}{2\sigma^2}}$$

$$\phi_5 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d_1-d_2-l_{12})^2}{2\sigma^2}} \quad \phi_6 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d_1+a_2+j_{21})^2}{2\sigma^2}}$$

$$\phi_7 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u+a_1-a_2+i_{12})^2}{2\sigma^2}} \qquad \phi_8 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u+a_1-d_2+j_{12})^2}{2\sigma^2}}$$

$$\phi_9 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u+a_1+a_2-i_{12})^2}{2\sigma^2}}$$

Therefore, frequencies of the nine genotypes at the pseudomarker described as follows.

$$Q_1Q_1Q_3Q_3 : z_1 = \frac{\phi_1 p_{i1}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_1Q_3Q_4 : z_2 = \frac{\phi_2 p_{i2}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_1Q_4Q_4 : z_3 = \frac{\phi_3 p_{i3}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_2Q_3Q_3 : z_4 = \frac{\phi_4 p_{i4}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_2Q_3Q_4 : z_5 = \frac{\phi_5 p_{i5}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_2Q_4Q_4 : z_6 = \frac{\phi_6 p_{i6}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_2Q_2Q_3Q_3 : z_7 = \frac{\phi_7 p_{i7}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_2Q_2Q_3Q_4 : z_8 = \frac{\phi_8 p_{i8}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_2Q_2Q_4Q_4 : z_9 = \frac{\phi_9 p_{i9}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

2.4.3 The M Step

The M step is carried out by using the result of the E step. The likelihood is represented as follows.

$$L \propto \prod_{i}^{81} \prod_{j}^{n_i} \left(p_{i1}\phi_{ij1}\right)^{\tau_{ij1}} \left(p_{i2}\phi_{ij2}\right)^{\tau_{ij2}} \left(p_{i3}\phi_{ij3}\right)^{\tau_{ij3}} \left(p_{i4}\phi_{ij4}\right)^{\tau_{ij4}} \left(p_{i5}\phi_{ij5}\right)^{\tau_{ij5}} \left(p_{i6}\phi_{ij6}\right)^{\tau_{ij6}} \left(p_{i7}\phi_{ij7}\right)^{\tau_{ij7}} \left(p_{i8}\phi_{ij8}\right)^{\tau_{ij8}} \left(p_{i9}\phi_{ij9}\right)^{\tau_{ij9}}$$

Here, i indicates the genotype of the marker (one of the 81 types), and j indicates each individual ($1 \sim n_i$) that has the marker genotype.

Logarithm of the likelihood calculated as follows:

$$\ln(L) = const + \sum_{i}^{81}\sum_{j}^{n_i}\begin{pmatrix} z_{ij1}\ln p_{i1} + z_{ij2}\ln p_{i2} + z_{ij3}\ln p_{i3} + z_{ij4}\ln p_{i4} + z_{ij5}\ln p_{i5} \\ + z_{ij6}\ln p_{i6} + z_{ij7}\ln p_{i7} + z_{ij8}\ln p_{i8} + z_{ij9}\ln p_{i9} \end{pmatrix} - 0.5N\ln\left(2\pi\sigma^2\right)$$

$$-\sum_{i}^{81}\sum_{j}^{n_i}\left\{\begin{array}{l} \dfrac{z_{ij1}\left(y_{ij}-u-a_1-a_2-i_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij2}\left(y_{ij}-u-a_1-d_2-j_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij3}\left(y_{ij}-u-a_1+a_2+i_{12}\right)^2}{2\sigma^2} \\ + \dfrac{z_{ij4}\left(y_{ij}-u-d_1-a_2-j_{21}\right)^2}{2\sigma^2} + \dfrac{z_{ij5}\left(y_{ij}-u-d_1-d_2-l_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij6}\left(y_{ij}-u-d_1+a_2+j_{21}\right)^2}{2\sigma^2} \\ + \dfrac{z_{ij7}\left(y_{ij}-u+a_1-a_2+i_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij8}\left(y_{ij}-u+a_1-d_2+j_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij9}\left(y_{ij}-u+a_1+a_2-i_{12}\right)^2}{2\sigma^2} \end{array}\right\}.$$

Here, *const* is a constant, and N is the number of individuals. By differentiating the log-likelihood with respect to u, a_1, d_1, a_2, d_2, i_{12}, j_{12}, j_{21}, l_{12}, and σ^2, and setting the derivatives to zero, we have as follows.

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij4}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij4}} = \hat{u} + \hat{d}_1 + \hat{a}_2 + \hat{j}_{21} \qquad \cdots (1)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij6}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij6}} = \hat{u} + \hat{d}_1 - \hat{a}_2 - \hat{j}_{21} \qquad \cdots (2)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij2}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij2}} = \hat{u} + \hat{a}_1 + \hat{d}_2 + \hat{j}_{12} \qquad \cdots (3)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij8}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij8}} = \hat{u} - \hat{a}_1 + \hat{d}_2 - \hat{j}_{12} \qquad \cdots (4)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5}} = \hat{u} + \hat{d}_1 + \hat{d}_2 + \hat{l}_{21} \qquad \cdots (5)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij1} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij1}} = \hat{u} + \hat{a}_1 + \hat{a}_2 + \hat{i}_{12} \qquad \cdots (6)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij3} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij3}} = \hat{u} + \hat{a}_1 - \hat{a}_2 - \hat{i}_{12} \qquad \cdots (7)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij7} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij7}} = \hat{u} - \hat{a}_1 + \hat{a}_2 - \hat{i}_{12} \qquad \cdots (8)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij9} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij9}} = \hat{u} - \hat{a}_1 - \hat{a}_2 + \hat{i}_{12} \qquad \cdots (9)$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i}^{9}\sum_{j}^{n_i}\left\{\begin{array}{l} z_{ij1}\left(y_{ij} - u - a_1 - a_2 - i_{12}\right)^2 + z_{ij2}\left(y_{ij} - u - a_1 - d_2 - j_{12}\right)^2 + z_{ij3}\left(y_{ij} - u - a_1 + a_2 + i_{12}\right)^2 \\ + z_{ij4}\left(y_{ij} - u - d_1 - a_2 - j_{21}\right)^2 + z_{ij5}\left(y_{ij} - u - d_1 - d_2 - l_{12}\right)^2 + z_{ij6}\left(y_{ij} - u - d_1 + a_2 + j_{21}\right)^2 \\ + z_{ij7}\left(y_{ij} - u + a_1 - a_2 + i_{12}\right)^2 + z_{ij8}\left(y_{ij} - u + a_1 - d_2 + j_{21}\right)^2 + z_{ij9}\left(y_{ij} - u + a_1 + a_2 - i_{12}\right)^2 \end{array}\right\}$$

These equations are sufficient to carry out the M step, but if one wants to have the estimators of each parameter, they are represented as follows:

$$\hat{u} = \frac{(6)+(7)+(8)+(9)}{4}, \quad \hat{a}_1 = \frac{(6)+(7)-(8)-(9)}{4}, \quad \hat{a}_2 = \frac{(6)-(7)+(8)-(9)}{4},$$

$$\hat{d}_1 = \frac{(1)+(2)}{2} - \frac{(6)+(7)+(8)+(9)}{4}, \quad \hat{d}_2 = \frac{(3)+(4)}{2} - \frac{(6)+(7)+(8)+(9)}{4},$$

$$\hat{i}_{12} = \frac{(6)-(7)-(8)+(9)}{4}, \quad \hat{j}_{21} = \frac{(1)-(2)}{2} - \frac{(6)-(7)+(8)-(9)}{4},$$

$$\hat{j}_{21} = \frac{(3)-(4)}{2} - \frac{(6)+(7)-(8)-(9)}{4},$$

$$\hat{l}_{12} = (5) - \frac{(1)+(2)+(3)+(4)}{2} + \frac{(6)+(7)+(8)+(9)}{4}.$$

The details of the derivations are given in Appendix A.3. Now we have updated values of the four parameters. By using these values, the next E step is carried out. Until the values of parameters converge, the rounds of the E step and the M steps are iterated.

2.4.4 LOD Score

A LOD score is calculated using the values of the log-likelihood ln(L) and ln(L_0). ln(L_0) is calculated by the parameters that are obtained under the null hypothesis.

$$LODscore = \log_{10} L - \log_{10} L_0 = \frac{1}{\ln(10)}\{\ln(L) - \ln(L_0)\}$$

The constant value (*const*) disappears. P_1 backcross and P_2 backcross can be explained similarly, and refer to appendix A.2 for details.

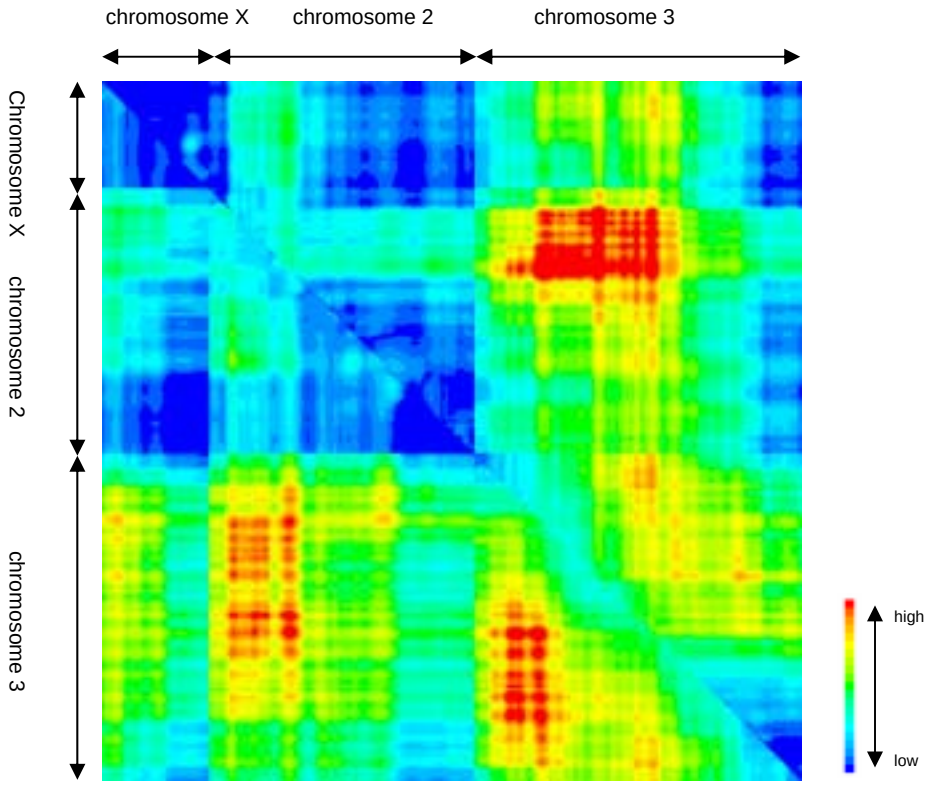


Figure 3 Example of calculating two-dimensional interval mapping

Calculation example of applying two-dimensional interval mapping to datasets of size of sexual organs of fly (Zeng et al,2000). Father is simulans, upper right, and father is maulitiana, under the left. It displayed it in blue in the place where the LOD score was low and in red in the place where the LOD score was high.

When the LOD score is obtained, the second pseudmarker is moved to the next of 1cM, and the calculation is repeated on all chromosomes. Next, the first pseudmarker is moved to the next of 1cM, and the calculation is repeated on all chromosomes. Thus, all chromosome pair is calculated.

2.4.5 An Example

Figure 3 is the example of applying two-dimensional interval mapping to datasets (Zeng et al, 2000) of the size of the sexual organs of fly. The point where the LOD score is low was displayed in blue, and high was displayed in red.

2.5 Evaluation

The result by two-dimensional interval mapping is evaluated by a permutation test and explained variance. In permutation test, the threshold of a significant LOD score is obtained. Explained variance evaluates how much variance is exchanged by the parameters obtained by two-dimensional interval mapping.

2.5.1 Permutation test

The threshold of the significant LOD score for the existence of QTL is obtained. The data of the marker genotype and the data of the phenotype value are permuted at random, and a similar calculation is repeated enough frequency in the data. The highest LOD score is obtained each time. The lowest LOD score of high rank (100-N) % is assumed to be a threshold for the significance level of N%.

2.5.2 Explained variance

Explained variance evaluates how much variance is exchanged by the parameters obtained by two-dimensional interval mapping. Explained variance is calculated as follows:

$$ev = \frac{\left(\sigma_0^2 - \sigma^2\right)}{\sigma_0^2} \times 100 .$$

Here, σ_0^2 is σ^2 when each parameter is the initial value (that is, variance between individuals of the phenotype value), and σ^2 is σ^2 when each parameter has been the calculated value in each pseudmarker.

Chapter 3

Maker interval network

In two-dimensional interval mapping, QTL of each 1cM is presumed as well as existing interval mapping. Therefore, the result is given as LOD score in which significant of the existence of QTL in each 1cM is shown. In a word, the result of two-dimensional interval mapping, which is the data by a genetic map distance in which cM is assumed to be a unit, cannot be compared directly with data by a physical map distance in which the base pair is assumed to be a unit.

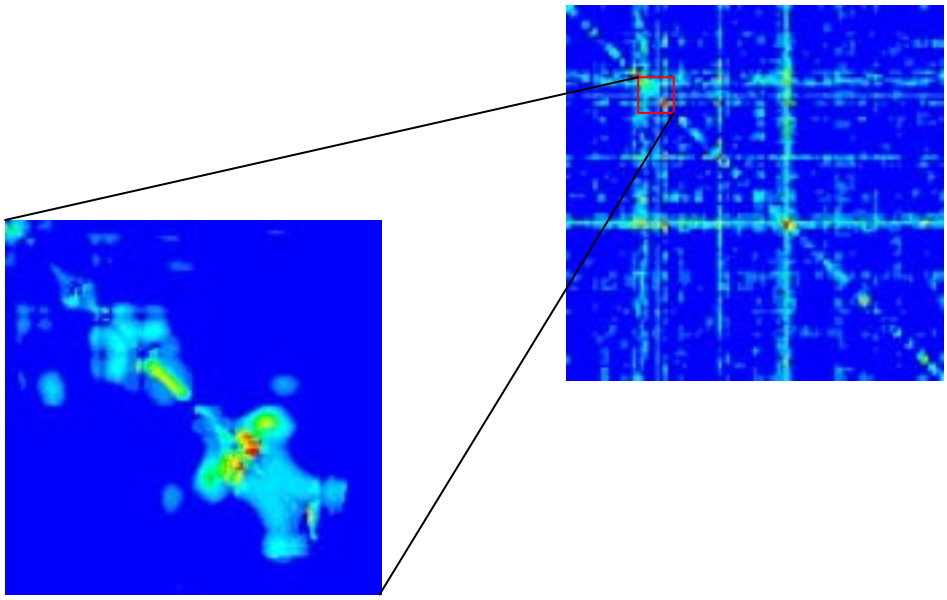


Figure 4 Example of making marker interval network

The left under is a part of the result of applying two-dimensional interval mapping to the dataset of schizophrenia model mice's forced swim test (Yoshikawa et al, 2002, chromosome 4 and 5). When this part is assumed to be a marker interval network, it is shown in the part enclosed by a red square in upper right.

There, the result of two-dimensional interval mapping is treated as the data of which unit was a marker interval. Whereat, the data enabled the comparison with the data of a physical map distance, because a position of a marker by a physical map distance is clear. This data of which the unit was a marker interval can be assumed a network where marker intervals are assumed to be nodes and marker interval pairs are assumed to be edges. This network is named a *marker interval network*. Here, the making method of marker interval network is

explained, and the result of two-dimensional interval mapping is evaluated by using the marker interval network.

3.1 Making method of marker interval network

The result by two-dimensional interval mapping is assumed data of which the unit is the marker interval so that it can be compared with data by a physical map distance. In interval mapping, QTL is presumed according to whether LOD scores in a marker interval exceed the threshold calculated by a permutation test. That is, the maximum value among LOD scores of a marker interval only has to be referred. Therefore, the maximum value among LOD scores of a marker interval is treated as the LOD score in each marker interval.

Figure 4 is an example of making marker interval network from the result of two-dimensional interval mapping. The left under is a part of the result of applying two-dimensional interval mapping to the dataset of schizophrenia model mice's forced swim test (Yoshikawa et al, 2002, chromosome 4 and 5). When this part is assumed to be a marker interval network, it is shown in the part enclosed by a red square in upper right.

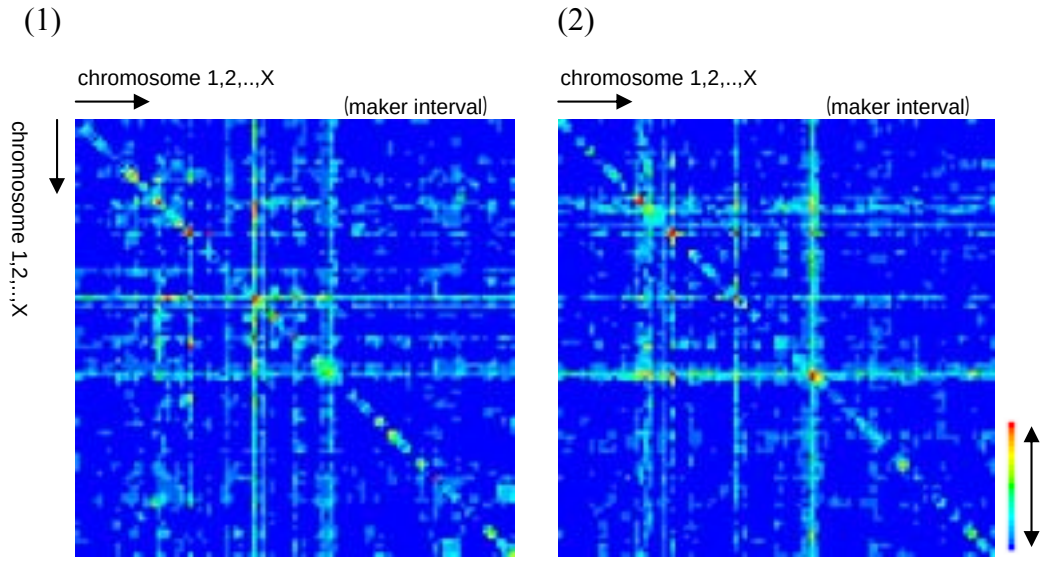


Figure 5 Result of two-dimensional interval mapping of schizophrenia model mice

The unit of each mass is a marker interval. Chromosome 1, 2, .., X queue up from the left sequentially. In the chromosome, it queues up from the left sequentially in order in which cM of the marker is low. ((1): FST, (2): TST)

(1)

ch1	M1	M2	ch2	m1	m2	LOD	u	a_1	a_2	d_1	d_2	i_{12}	j_{12}	j_{21}	l_{12}	σ^2
3	49.7	83.5	3	49.7	83.5	66.87	145.26	5.04	4.95	48.22	20.84	3.69	-47.99	12.36	-75.63	994.2
5	1	18	5	18	54	44.75	185.93	-3.94	-5.49	20.58	-17.45	-35.15	42.71	-5.83	-43.10	1222.8
5	18	54	5	18	54	53.23	164.79	-32.01	21.48	3.65	30.25	-15.41	13.83	17.16	-57.99	1128.7
5	18	54	6	2.5	26.5	42.04	168.94	-8.60	-7.65	-29.15	-35.67	23.46	2.02	9.97	83.43	1149.7
8	15.5	40	5	18	54	41.30	193.42	-20.12	-13.59	-52.96	-56.06	-6.84	22.85	9.69	77.46	1218.3
8	15.5	40	8	15.5	40	59.81	164.79	-2.26	-6.91	-28.13	42.57	-12.64	-13.21	11.57	-38.36	1068.5
10	17	36	8	15.5	40	42.78	145.49	-8.05	-9.38	43.51	24.19	-2.32	11.95	-5.81	-80.41	1241.6
10	17	36	10	17	36	39.52	151.11	0.43	-2.77	-22.32	58.21	8.94	14.25	-4.35	-41.23	1266.4
10	40	44	5	18	54	47.38	153.08	-20.66	-25.44	13.66	15.49	19.18	44.88	33.67	-43.67	1165.2
16	16.9	43	16	16.9	43	40.51	156.25	-10.01	15.06	23.04	10.70	-6.36	-32.54	23.98	-38.13	1258.8

(2)

ch1	M1	M2	ch2	m1	m2	LOD	u	a_1	a_2	d_1	d_2	i_{12}	j_{12}	j_{21}	l_{12}	σ^2
3	33.7	49.7	3	49.7	83.5	60.64	133.47	-5.00	7.31	100.17	162.43	3.07	17.44	-37.40	-251.30	6103.1
3	49.7	83.5	3	33.7	49.7	58.32	127.30	6.89	-4.46	179.44	100.99	9.11	-33.20	9.72	-260.64	6090.3
3	49.7	83.5	3	49.7	83.5	46.16	144.17	5.01	-2.09	134.94	86.99	6.04	-70.61	37.09	-212.97	6820.4
5	1	18	5	18	54	51.94	206.48	-59.30	10.93	4.62	-7.74	-33.63	150.21	44.89	-39.58	6475.4
5	18	54	5	18	54	49.03	152.44	5.71	-19.17	158.64	32.10	1.90	47.82	19.38	-186.40	6603.7
5	18	54	8	15.5	40	47.63	196.37	-86.19	40.18	-19.34	19.42	-26.60	152.06	-75.41	-35.73	6516.6
8	15.5	40	8	15.5	40	48.89	208.63	68.12	-73.46	-42.98	87.77	-42.94	-55.48	26.23	-99.04	6655.8
11	31	47.64	5	18	54	50.69	273.32	17.80	32.71	-132.82	-142.20	-41.65	-7.20	-41.18	213.07	6737.7
11	31	47.64	11	31	47.64	51.38	168.91	-24.67	13.76	22.24	60.04	-1.28	-78.90	74.04	-76.35	6668.7
11	47.64	55.6	11	31	47.64	45.36	126.01	-7.43	15.79	170.33	122.42	12.25	52.75	-31.86	-255.08	6926.6

Table 1 Value of each parameter calculated by two-dimensional interval mapping

Ch1, M1, and M2 are one of chromosomes in the marker interval and the starting and the terminal points of markers. Ch2, m1, and m2 are one of chromosomes in the marker interval and the starting and the terminal points of markers. ((1): forced swim test, (2): tail suspension test)

3.2 Maker interval network

The dataset of the schizophrenia model mice (Yoshikawa et al, 2002) was used. The mice are soaked compulsorily in water, and time until the mice don't struggle is measured in forced

swim test (FST). And, the mice are caught by the tail, and time until the mice don't move is measured in tail suspension test (TST). Two-dimensional interval mapping was applied to these datasets, and the marker interval network was made (Figure 5). Moreover, ten intervals that are evaluated as obviously significant in marker interval pairs of marker interval networks in FST and TST, were listed (Table 1). A marker interval network is evaluated from the LOD score, each parameter, and explained variance by using these datasets.

3.2.1 LOD score

The interval pairs with high LOD scores obtained two-dimensional interval mapping are classified into two types. One is a pair of marker intervals each of which belongs to the same chromosome segment. The other is a pair of marker intervals each of which belongs to the different chromosome segment respectively.

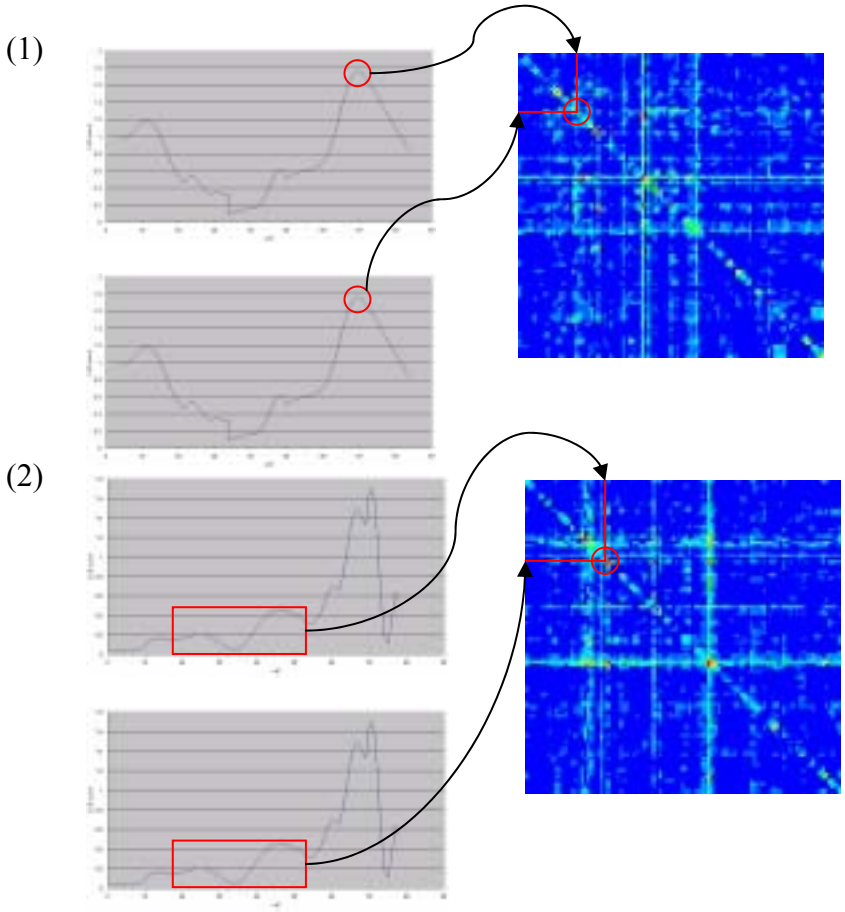


Figure 6 Example of a pair of marker intervals each of which belongs to the same chromosome segment

One is the interval pair with a high LOD score calculated by interval mapping in the example of chromosome 3 of FST (1). One is the interval pair with a low LOD score calculated by interval mapping in the example of chromosome 5 of TST (2).

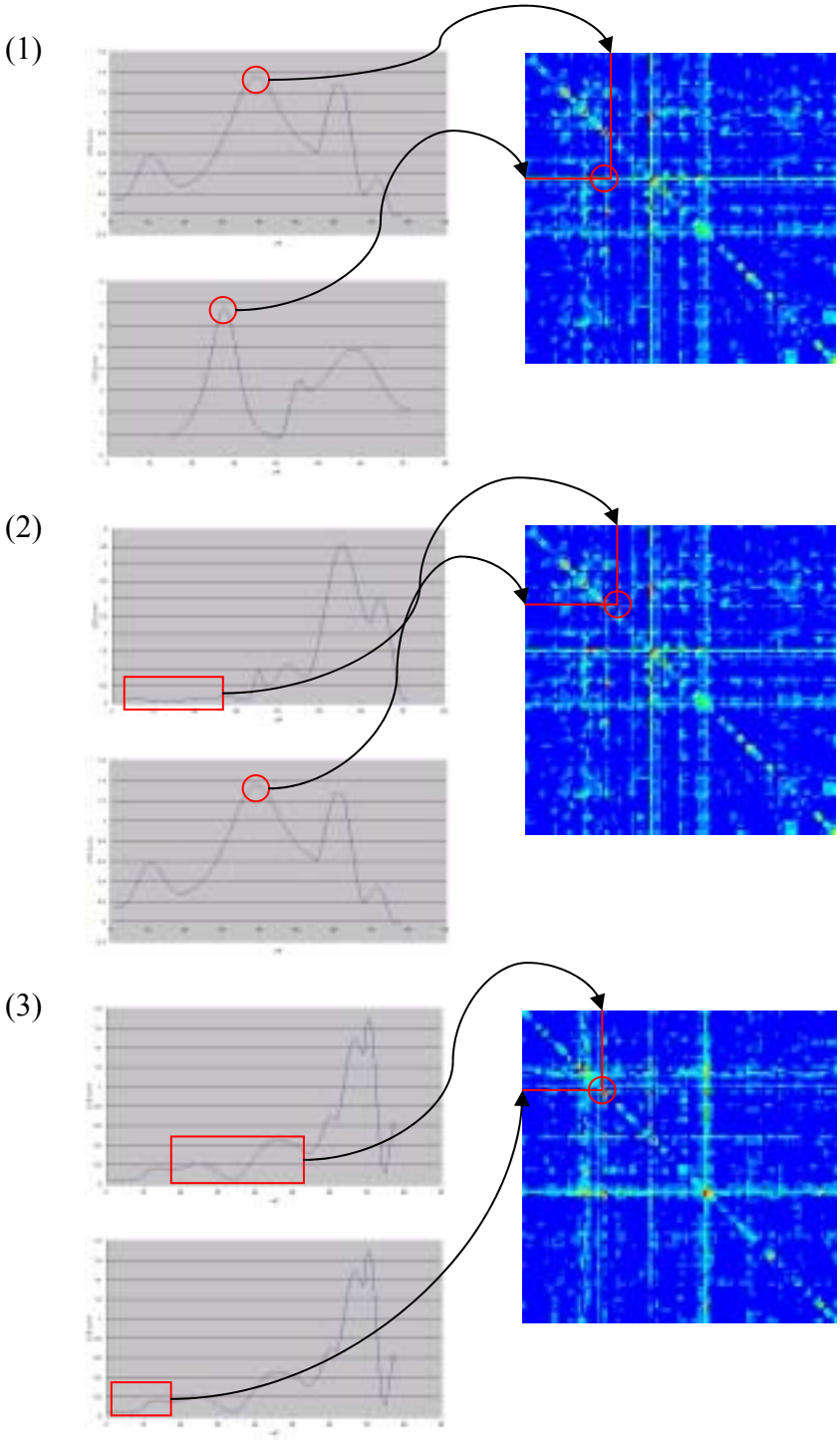


Figure 7 Example of the marker interval where the pair is different

The first case is the interval pair with both high LOD scores calculated by interval mapping in the example of chromosome 8 and 5 of FST (1). The second case is the interval pair with only one high LOD score calculated by interval mapping in the example of chromosome 6 and 5 of FST (2). The third case is the interval pair with both low LOD scores calculated by interval mapping in the example of chromosome 5 of TST (3).

In addition, the significant pairs of the former case are classified into two types (Figure 6). One is the interval pair with a high LOD score calculated by interval mapping. The other is the interval pair with a low LOD score calculated by interval mapping. On the other hand, the significant pair of the latter case are classified into three types (Figure 7). The first case is the interval pair with both high LOD scores calculated by interval mapping. The second case is the interval pair with only one high LOD score calculated by interval mapping. The third case is the interval pair with both low LOD scores calculated by interval mapping. It's notable that the gene in the marker interval affects mutually and is related to the phenotype in the third case.

(1)

ch1	M1	M2	ch2	m1	m2	ev (%)
3	49.7	83.5	3	49.7	83.5	66.87
5	1	18	5	18	54	44.75
5	18	54	5	18	54	53.23
5	18	54	6	2.5	26.5	42.04
8	15.5	40	5	18	54	41.30
8	15.5	40	8	15.5	40	59.81
10	17	36	8	15.5	40	42.78
10	17	36	10	17	36	39.52
10	40	44	5	18	54	47.38
16	16.9	43	16	16.9	43	40.51

(2)

ch1	M1	M2	ch2	m1	m2	ev (%)
3	33.7	49.7	3	49.7	83.5	60.64
3	49.7	83.5	3	33.7	49.7	58.32
3	49.7	83.5	3	49.7	83.5	46.16
5	1	18	5	18	54	51.94
5	18	54	5	18	54	49.03
5	18	54	8	15.5	40	47.63
8	15.5	40	8	15.5	40	48.89
11	31	47.64	5	18	54	50.69
11	31	47.64	11	31	47.64	51.38
11	47.64	55.6	11	31	47.64	45.36

Table 2 Explained variance for two-dimensional interval mapping

Explained variance was calculated from each σ^2 calculated by two-dimensional interval mapping. ev is the value of explained variance. ((1): FST, (2): TST)

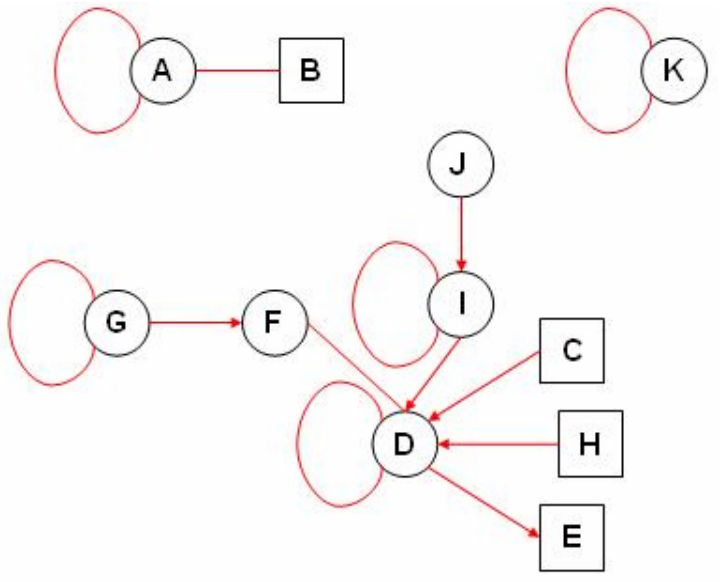
3.2.2 Prameters

Next, the term of epistatic effects is paid to attention among the results of each parameter. In chromosome 5 (1.0-18.0cM and 18.0-54.0cM) in FST, the values of additive effects (a_1, a_2) are low, but the value of additive × additive epistatic effects (i_{12}) is high. In chromosome 3 (49.7-83.5cM) in FST, the value of the first of additive effects (a_1) and the value of the second of dominant effects (d_2) are low, but the value of additive × dominant epistatic effects (j_{12}) is high. In chromosome 11 (31.0-47.64cM) in TST, the value of the first of dominant effects (d_1) and the value of the second of additive effects (a_2) are low, but the value of dominant ×

additive epistatic effects (j_{21}) is high. Each epistatic effect affects the phenotype more effectively in these places.

3.2.3 Explained variance

Explained variance was calculated from each σ^2 calculated by two-dimensional interval mapping. (Table 2). Values are 40 ~ 60 %, and a lot of parts are shown by another parameter by two-dimensional interval mapping.



A: chromosome 3 (33.7-49.7)
B: chromosome 3 (49.7-83.5cM)
C: chromosome 5 (1.0-18.0cM)
D: chromosome 3 (18.0-54.0cM)
E: chromosome 6 (2.5-26.5cM)
F: chromosome 8 (15.5-40.0cM)
G: chromosome 10 (17.0-36.0cM)
H: chromosome 10 (40.0-44.0cM)
I: chromosome 11 (31.0-47.64cM)
J: chromosome 11 (47.64-55.6cM)
K: chromosome 16 (16.9-43.0cM)

Figure 8 The prediction of marker interval network

A marker interval network was predicted from the ten high-ranking marker interval pairs. It enclosed full shows the marker interval where the high score was calculated in interval mapping. It enclosed with the square shows the marker interval where the low score was calculated in interval mapping. The line where each marker interval is connected shows the interaction. The arrow shows the epistatic effects that their starting point is high rank. The line connected with oneself shows that it has the pair in the marker interval.

3.2.4 Prediction of maker interval network

Finally, the network is predictable from the ten high-ranking marker interval pairs of the marker interval network, is described. When both FST and TST results are synthesized, the network is like Figure 8 being composed is predictable in these ten marker intervals. It enclosed full shows the marker interval where the high score was calculated in interval mapping. It enclosed with the square shows the marker interval where the low score was calculated in interval mapping. The line where each marker interval is connected shows the interaction. The arrow shows the epistatic effects that their starting point is high rank. The line connected with oneself shows that it has the pair in the marker interval.

In Figure 8, a big network is composed centering on the marker interval of chromosome 5 (18.0-54.0cM). Although it was considered the ten high-ranking marker interval pairs in this research, a bigger marker interval network can be predicted by paying attention to more marker interval pairs.

Chapter 4

Data of physical map distance

According to the result of two-dimensional interval mapping treated as the data of which unit was a marker interval, the data enabled the comparison with the data by a physical map distance, because a position of a marker by a physical map distance is clear. Here, as one example of the comparison between the data by marker interval network and a physical map distance, it compares with the PPI (protein-protein interaction) database, and the candidate gene of QTL is extracted.

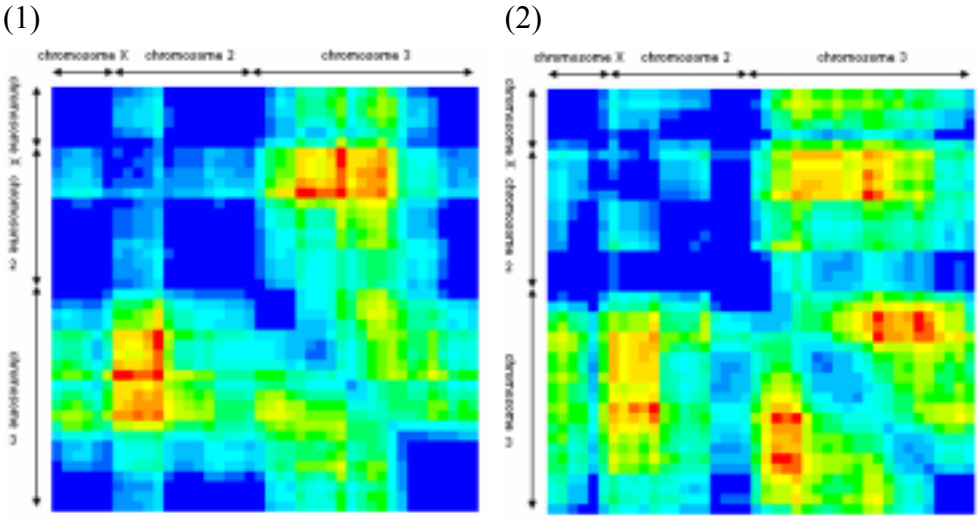


Figure 9 Result of two-dimensional interval mapping of size of sexual organs of fly (Zeng et al, 2000) The unit of each mass is a marker interval. Chromosome 1, 2, .., X queue up from the left sequentially. In the chromosome, it queues up from the left sequentially in order in which cM of the marker is low. ((1): father is maulitiana, (2): father is simulans)

4.1 Marker interval network and PPI database

Two-dimensional interval mapping was applied to the data of the size of the sexual organs of fly (Zeng et al, 2000), and Marker interval network was made (Figure 9). The gene list in each marker interval can be made by listing the gene placed between the franking markers (Figure 10). Thus, the allocated LOD score of each marker interval pair means how strongly at least one pair among the genes in the marker interval pair is related to the phenotype. Whether it is interactive immediately or acts indirectly is not asked, the possibility where at least one or more gene pairs with high possibility of being related mutually to the expression of the phenotype exist, is high in the marker interval pair with high LOD score (For example,

it may be a part of the metabolic pathway for the expression of the phenotype, or be the one like the edge and the edge in the pathway, etc.).

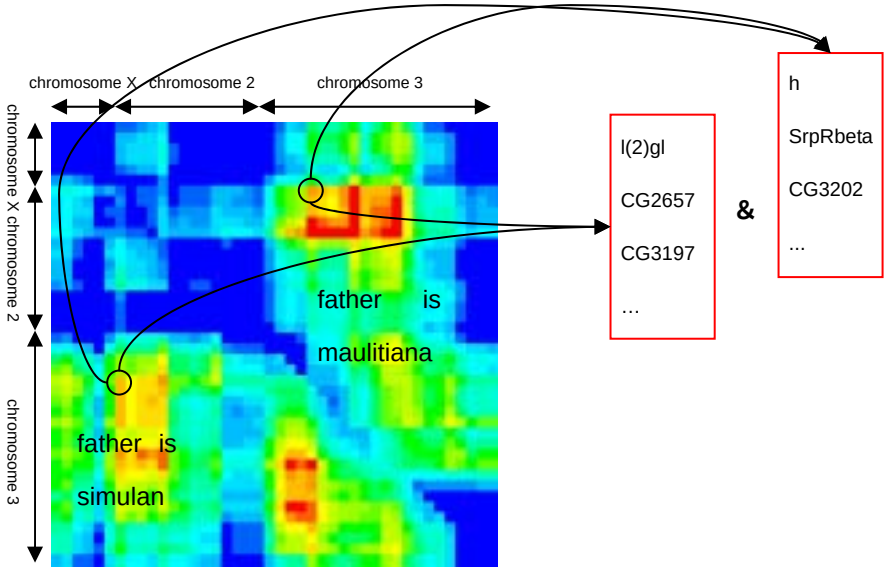


Figure 10 The making of the gene list from the marker interval network

Two data with different father was arranged to be comparable. The unit of each mass is a marker interval. Chromosome 1, 2, X queue up from the left sequentially. In the chromosome, it queues up from the left sequentially in order in which cM of the marker is low. The gene list can be made respectively by listing the gene placed between the franking markers from each marker interval pair.

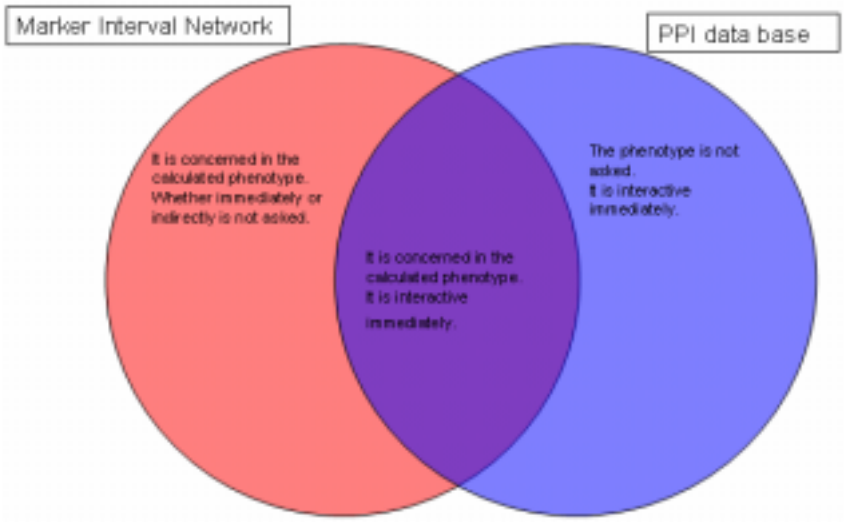


Figure 11 The comparison between marker interval network and PPI database

The comparison between a marker interval network (left) by two-dimensional interval mapping and PPI database (right).

The gene being related to the phenotype is included in one or more gene pair included in

the marker interval pair with high LOD score, whether immediately or indirectly is not asked. On the other hand, the protein that has been interactive immediately is included in PPI database, which phenotype is related is not asked. Here, the gene to which each protein of the pair of the protein in PPI database that has been interactive immediately is coded is retrieved, and the list of the gene pair is made, and each gene pair is allocated in each marker interval pair. Then, it is related to the phenotype, and the gene pair with high possibility of interactive immediately is included in the retrieved gene pair in PPI database included in the marker interval where the LOD score is high (Figure 11).

4.2 Extraction of the candidate genes of QTL

The extracting the candidate genes of QTL become possible by the comparison with data by physical map distance having become possible. Here, as continuation of 4.1 the one example, the candidate genes of QTL of the size of the sexual organs of fly is extracted.

courtship behavior	qtc, Est-6, per, ple, pros, spin, tko, y
male courtship behavior	qtc, fru, tko
mating behavior	dsf, per
mating behavior, sex discrimination	dsf, fru
sex discrimination	dsf, br, BtbVII, CG3056, CG6118, fru, lolal, mod(mdg4), sc, Sox100B, Sox14, Stat92E, ttk, vir
somatic sex discrimination	dsf
copulation	dsf, fru, ken

Table 3 Biological process and gene thought to be related to the size of sexual organs

List of kind of Biological process thought to be related in the size of sexual organs used by this research and the gene that belongs to it.

One of the gene pair at least is related to the size of sexual organs by using the data of Biological process of Gene ontology extracted from among the retrieved gene pair in PPI database included in the marker interval pair with high LOD score. Biological process and the gene belonged to it that had been used at this time were summarized in Table 3. The gene pair for one of the pair to contain these genes at least is sorted in order with high LOD scores (Table 4).

interval1	gene1	interval2	gene2	LOD score
2-28.5-34.7	CG6415	2-147.7-157.7	mod(mdg4)	28.59117255
2-22.0-28.5	CG14534	2-43.2-50.0	Est-6	26.54336379
2-0.0-7.0	CG15631	2-134.6-147.7	fru	26.36016013
2-28.5-34.7	Nup170	2-14.3-21.3	ple	25.21665039
2-34.7-55.2	BG:DS06874.2	2-147.7-157.7	Stat92E	24.94588041
2-34.7-55.2	CG4959	2-147.7-157.7	CG31160	24.94588041
2-34.7-55.2	CG4959	2-134.6-147.7	fru	24.94588041
2-14.3-21.3	Src64B	2-147.7-157.7	mod(mdg4)	23.00990654
2-14.3-21.3	CG12607	2-147.7-157.7	mod(mdg4)	23.00990654
2-14.3-21.3	ple	2-147.7-157.7	RpS30	23.00990654
interval1	gene1	interval2	gene2	LOD score
2-14.3-21.3	Src64B	2-147.7-157.7	mod(mdg4)	27.48448759
2-14.3-21.3	CG12607	2-147.7-157.7	mod(mdg4)	27.48448759
2-14.3-21.3	ple	2-101.3-114.2	hb	27.48448759
2-14.3-21.3	ple	2-147.7-157.7	RpS30	27.48448759
2-28.5-34.7	CG6415	2-147.7-157.7	mod(mdg4)	27.11054392
2-21.3-28.7	Nmt	2-147.7-157.7	CG31160	27.02120768
2-28.7-43.2	SH3PX1	2-123.3-126.6	pros	25.99129802
2-14.3-21.3	ple	2-114.2-123.3	CG14684	24.81583541
2-0.0-7.0	CG15631	2-134.6-147.7	fru	24.28721228
2-22.0-28.5	CG14534	2-43.2-50.0	Est-6	24.03511424

Table 4 List of the gene pair for one of the pair thought to be concerned in the size of sexual organs at least.

Ten high-ranking gene pair of each dataset was listed. interval1 and interval2 show the chromosome number and the starting point and the terminal of the marker of the gene pair. gene1 and gene2 show the gene pair. LOD score indicates the LOD score of the marker interval pair that includes the gene pair. (top10: father is maulitiana, under10: father is simulans)

As for the threshold, when assuming the significance level of 95% by the permutation test of 400 times, the threshold in the dataset that father is mauritiana was 8.17, and the threshold in the dataset that father is simulans was 8.87(Figure 12). When the LOD score pays attention to the gene pair more than the threshold, the gene pair with high possibility of being related to

the expression of the size of sexual organs can be extracted.

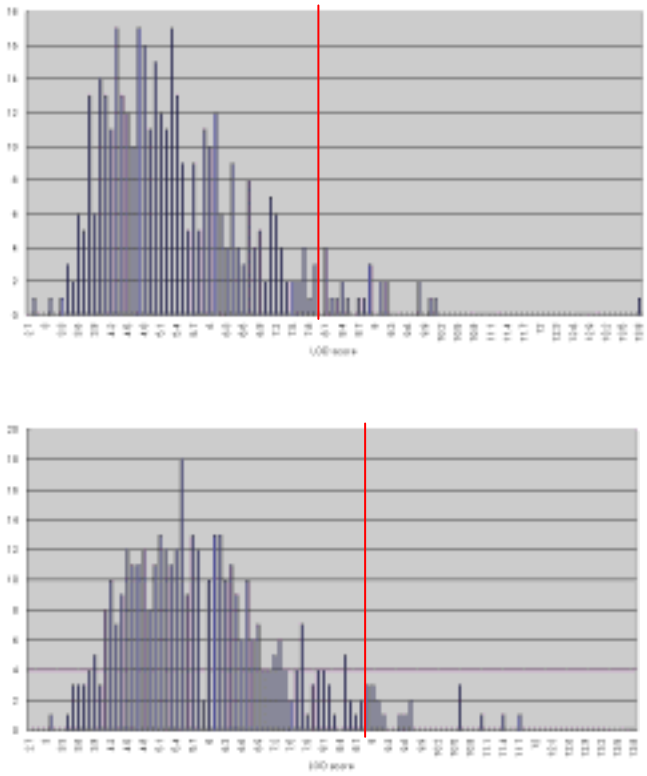


Figure 12 The result of the permutation test of 400 times

When assuming the significance level of 95%, the threshold in the dataset that father is mauritiana is 8.17, and the threshold in the dataset that father is simulans is 8.87.

The gene in no examination beforehand by Gene ontology of the gene pair extracted thus was referred to the thesis. As a result, the genes that seem that it is related to the size of sexual organs as follows:

CG6415, SrC64B, Nmt, Nup170, hb, CG4959, CG4328, krz, AP-50, CG31973, CG7291, CG17666, HLHm7, CG33070, tun, Mst84a, Mst84b, Gld.

The genes related to the size of sexual organs are included in the list of these genes though it has not been included in Gene ontology yet. There are the genes that are related to the size of sexual organs though it has not been known yet in the extracted genes other than these.

Chapter 5

Discussion and future work

In this research, two-dimensional interval mapping that is considered multiple gene loci and their interactions is developed and a marker interval network was made from the result of the two-dimensional interval mapping. Further discussions of these and the views in the future are described.

5.1 Two-dimensional interval mapping (2DIM)

There are two advantages of advancing interval mapping to two dimensions. One is to have come to calculate the terms of the interactions that are not considered in the interval mapping. The other one is to have come to be able to treat as a network.

The terms of interactions make four parameters calculably, not only the additive ×additive epistatic effects but also the additive × dominant epistatic effects, the dominant × additive epistatic effects, and the dominant × dominant epistatic effects is included. Therefore, it is clear which becomes significant by high-ranking time among pairs of the pseudmarker. Moreover, which the interaction affects a lot is calculated for the pair of each QTL presumption place.

The network by two-dimensional interval mapping specializes in whether is related to the phenotype, and it doesn't ask whether the pair has physically interactive or relation by a distance like the edge and the edge in the metabolic pathway. This is a new network with the side in the existing one without.

5.2 Marker interval network

It came to be able to make the network in pseudmarkers by two-dimensional interval mapping. The network by the marker interval was made in this research, though a network by a genetic map distance where cM was assumed to be a unit can be made. As a result, it comes to be able to list the genes that exist in the marker interval assumed to be significant by two-dimensional interval mapping.

In this research, the new candidate genes of QTL that are related to the phenotype are extracted by using the marker interval network, PPI database, and Gene ontology. Thus, it proposes what the new candidate genes of QTL that are related to the phenotype are predictable by the comparison between a marker interval network and a database by a

physical map distance. If the already-known genes increase in PPI database and Gene ontology, a higher prediction of reliability becomes possible.

5.3 Marker interval network database (MINTDB)

A marker interval network is made from the result of two-dimensional interval mapping for a variety of phenotypes, and the database that integrates them is being constructed. This is named *marker interval network database* (MINTDB). Now, the database is composed by the result of each two-dimensional interval mapping and the marker interval network for 24 phenotypes (the dataset of intercross about the blood pressure (Sugiyama et al, 2002, and DiPetrillo et al, 2004), the dataset of intercross about the bone density (Beamer et al, 1999), the dataset of backcross about the cholesterol (Wittenberg et al, 2002), and the dataset of backcross about the blood pressure of salt-induced hypertension mice (Sugiyama et al, 2001)). These datasets can be compared with data with a different marker as long as it is the QTL dataset of mice. Moreover, it is possible to compare it with the dataset that integrates the datasets of 24 phenotypes (Figure 13).

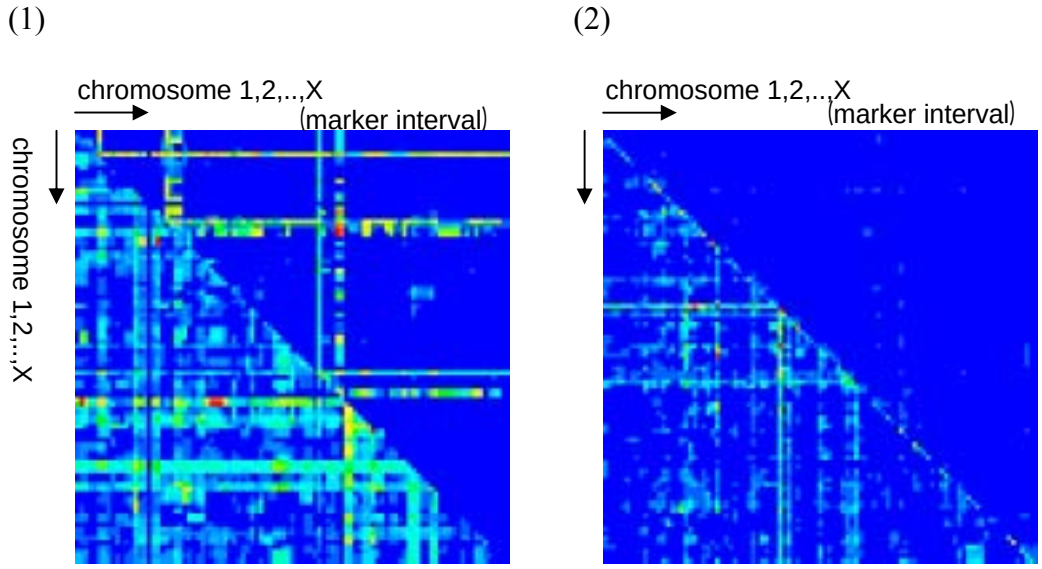


Figure 13 The comparison of marker interval networks.

In the comparison of the marker interval networks (1), the comparison between the dataset of the blood pressure of salt-induced hypertension mice (upper right, Sugiyama et al, 2002) and the dataset of the blood pressure (left under, Sugiyama et al, 2002) was enumerated as an example.

In the comparison with the database of marker interval networks (2), the comparison between the dataset of FST (left under, Yoshikawa et al, 2002) and the dataset of the number of marker interval pairs of the high rank of ten in 24 phenotypes (upper right) was enumerated as an example.

In this research, the comparison of the marker interval networks of a different phenotype

was enabled by the database constructed with the datasets of marker interval networks of 24 phenotypes. There are the datasets that assumes that are related mutually though they are the different phenotypes (for example, blood sugar and weight, etc.). Even if they are the same phenotypes, there is the difference by the lineage. Therefore, it is important to enable these comparisons. Hereafter, two-dimensional interval mapping is applied to more phenotypes, and it aims at the construction of a larger database.

Appendix

A.1 Detail of the derivation in interval mapping

Here, the detail of the derivation interval mapping that is abbreviated in 2.2.4 is explained.

$$\frac{\partial \ln(L)}{\partial u} = \frac{1}{\sigma^2}\sum_{i}^{9}\sum_{j}^{n_i}\left\{z_{ij1}\left(y_{ij}-u-a\right)+z_{ij2}\left(y_{ij}-u+d\right)+z_{ij3}\left(y_{ij}-u+a\right)\right\}$$

$$= \frac{1}{\sigma^2}\sum_{i}^{9}\sum_{j}^{n_i}\left\{\left(z_{ij1}+z_{ij2}+z_{ij3}\right)y_{ij}-\left(z_{ij1}+z_{ij2}+z_{ij3}\right)u-\left(z_{ij1}-z_{ij3}\right)a-z_{ij2}d\right\}$$

$$= \frac{1}{\sigma^2}\sum_{i}^{9}\sum_{j}^{n_i}\left(y_{ij}-u-\left(z_{ij1}-z_{ij3}\right)a-z_{ij2}\right)=0$$

As a result,

$$\sum_{i}^{9}\sum_{j}^{n_i}y_{ij} = \hat{u}\sum_{i}^{9}\sum_{j}^{n_i}(1)+\hat{a}\sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}\right)+\hat{d}\left(\sum_{i}^{9}\sum_{j}^{n_i}(1)-\sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}\right)\right)$$

Here,

$$N = \sum_{i}^{9}\sum_{j}^{n_i}(1)$$

$$A = \sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}\right)$$

$$B = \sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}\right)$$

Then,

$$\sum_{i}^{9}\sum_{j}^{n_i}y_{ij} = N\hat{u}+A\hat{a}+\left\{\sum_{i}^{9}\sum_{j}^{n_i}z_{ij2}y_{ij}-\hat{u}(N-B)\right\}$$

$$\sum_{i}^{9}\sum_{j}^{n_i}\left(1-z_{ij2}\right)y_{ij} = B\hat{u}+A\hat{a} \quad \cdots (1)$$

Moreover,

$$\frac{\partial \ln(L)}{\partial a} = \frac{1}{\sigma^2}\sum_{i}^{9}\sum_{j}^{n_i}\left\{z_{ij1}\left(y_{ij}-u-a\right)-z_{ij3}\left(y_{ij}-u+a\right)\right\}=0$$

$$\hat{a}\sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}\right)=\sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}\right)y_{ij}-\hat{u}\sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}\right)$$

As a result,

$$\sum_{i}^{9}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}\right)y_{ij} = A\hat{u}+B\hat{a} \quad \cdots (2)$$

Then, from (1) + (2),

$$\sum_{i}^{9}\sum_{j}^{n_i}\left(1+z_{ij1}-z_{ij2}-z_{ij3}\right)y_{ij}=\left(A+B\right)\hat{u}+\left(A+B\right)\hat{a}$$

$$\sum_{i}^{9}\sum_{j}^{n_i}\left(2z_{ij1}\right)y_{ij}=\hat{u}\sum_{i}^{9}\sum_{j}^{n_i}\left(2z_{ij1}\right)+\hat{a}\sum_{i}^{9}\sum_{j}^{n_i}\left(2z_{ij1}\right)$$

$$\frac{\sum_{i}^{9}\sum_{j}^{n_i}z_{ij1}y_{ij}}{\sum_{i}^{9}\sum_{j}^{n_i}z_{ij1}}=\hat{u}+\hat{a}\quad\cdot\cdot\cdot(3)$$

Moreover, from (1) – (2),

$$\sum_{i}^{9}\sum_{j}^{n_i}\left(1-z_{ij1}-z_{ij2}+z_{ij3}\right)y_{ij}=\left(B-A\right)\hat{u}-\left(B-A\right)\hat{a}$$

$$\sum_{i}^{9}\sum_{j}^{n_i}\left(2z_{ij3}\right)y_{ij}=\hat{u}\sum_{i}^{9}\sum_{j}^{n_i}\left(2z_{ij3}\right)-\hat{a}\sum_{i}^{9}\sum_{j}^{n_i}\left(2z_{ij3}\right)$$

$$\frac{\sum_{i}^{9}\sum_{j}^{n_i}z_{ij3}y_{ij}}{\sum_{i}^{9}\sum_{j}^{n_i}z_{ij3}}=\hat{u}-\hat{a}\quad\cdot\cdot\cdot(4)$$

Next,

$$\frac{\partial\ln\left(L\right)}{\partial d}=\frac{1}{\sigma^2}\sum_{i}^{9}\sum_{j}^{n_i}\left\{z_{ij2}\left(y_{ij}-u+d\right)\right\}=0$$

As a result,

$$\sum_{i}^{9}\sum_{j}^{n_i}z_{ij2}y_{ij}=\left(\hat{u}+\hat{d}\right)\sum_{i}^{9}\sum_{j}^{n_i}z_{ij2}$$

$$\frac{\sum_{i}^{9}\sum_{j}^{n_i}z_{ij2}y_{ij}}{\sum_{i}^{9}\sum_{j}^{n_i}z_{ij2}}=\hat{u}+\hat{d}\quad\cdot\cdot\cdot(5)$$

Finally,

$$\frac{\partial\ln\left(L\right)}{\partial\sigma^2}=\frac{N}{2}\frac{1}{\sigma^2}-\sum_{i}^{9}\sum_{j}^{n_i}\left\{-\frac{1}{\left(\sigma^2\right)^2}\right\}\left\{\frac{z_{ij1}\left(y_{ij}-u-a\right)^2}{2}+\frac{z_{ij2}\left(y_{ij}-u+d\right)^2}{2}+\frac{z_{ij3}\left(y_{ij}-u+a\right)^2}{2}\right\}=0$$

As a result,

$$\hat{\sigma}^2=\frac{1}{N}\sum_{i}^{9}\sum_{j}^{n_i}\{z_{ij1}\left(y_{ij}-u-a\right)^2+z_{ij2}\left(y_{ij}-u-d\right)^2+z_{ij3}\left(y_{ij}-u+a\right)^2\}\quad\cdot\cdot\cdot(6)$$

A.2 P_1 backcross and P_2 backcross of interval mapping

Here, in P_1 backcross, and P_2 backcross lines, details of the EM algorithm are explained.

A.2.1 The E step of P_1 backcross

At the location where we calculate its LOD score (i.e., at pseudomarker), the genotype is estimated in a probabilistic manner. The genotype at the pseudomarker is estimated using the following nine cases of the flanking marker genotypes.

1: $A_1A_1B_1B_1$ 2: $A_1A_1B_1B_2$ 4: $A_1A_2B_1B_1$ 5: $A_1A_2B_1B_2$

Here, A and B represent the flanking markers and the suffixes one and two represent from whether maternal or paternal the marker inherits. On the other hand, the pseudomarker genotype is represented one of Q_1Q_1 and Q_1Q_2. Here, Q represents the pseudomarker between marker A and B. Assume that the probability that a recombination happens between A and Q is r_1, the probability that a recombination happens between Q and B is r_2, the probability that a recombination happens only by one degree between A and B is assumed to be r_{1+2}, and the probability that a recombination happens both between A and Q and between Q and B is r_{12}. When the genotypes of the flanking markers are case *i* as above, we denote the probability that the genotype of the pseudomarker is Q_1Q_1 and Q_1Q_2 by p_{i1} and p_{i2}, respectively. Here, *i* ranges from one to four, and each is corresponds to one of the fore marker genotypes above. The probabilities of p_{i1} and p_{i2} are represented as follows.

	Q_1Q_1 (p_{i1})	Q_1Q_2 (p_{i2})
1:$A_1A_1B_1B_1$	q_1	q_2
2:$A_1A_1B_1B_2$	q_3	q_4
4:$A_1A_2B_1B_1$	q_4	q_3
5:$A_1A_2B_1B_2$	q_2	q_1

Here,

$$q_1 = \frac{(1-r_1-r_2+r_{12})}{(1-r_{1+2})} \quad q_2 = \frac{r_{12}}{(1-r_{1+2})} \quad q_3 = \frac{(r_2-r_{12})}{r_{1+2}} \quad q_4 = \frac{(r_1-r_{12})}{r_{1+2}}$$

Using the assumption that the residue terms of Q_1Q_1 and Q_1Q_2 cases are normally distributed, the probability densities Φ_1 and Φ_2 represented as follows.

$$\phi_1 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-a)^2}{2\sigma^2}} \qquad \phi_2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d)^2}{2\sigma^2}}$$

Therefore, frequencies of the three genotypes at the pseudomarker described as follows.

$$Q_1Q_1 : z_1 = \phi_1 p_{i1} / (\phi_1 p_{i1} + \phi_2 p_{i2})$$
$$Q_1Q_2 : z_2 = \phi_2 p_{i2} / (\phi_1 p_{i1} + \phi_2 p_{i2})$$

A.2.2 The M Step of P_1 backcross

The M step is carried out by using the result of the E step. The likelihood is represented as follows.

$$L \propto \prod_i \prod_j^{n_i} \left(p_{i1}\phi_{ij1}\right)^{z_{ij1}} \left(p_{i2}\phi_{ij2}\right)^{z_{ij2}}$$

Here, *i* indicates the genotype of the marker (one of the four types), and *j* indicates each individual ($1 \sim n_i$) that has the marker genotype.

Logarithm of the likelihood calculated as follows:

$$\ln(L) = const + \sum_i \sum_j^{n_i} \left(z_{ij1} \ln p_{i1} + z_{ij2} \ln p_{i2}\right) - 0.5N \ln\left(2\pi\sigma^2\right)$$

$$- \sum_i \sum_j^{n_i} \left\{ \frac{z_{ij1}\left(y_{ij} - u - a\right)^2}{2\sigma^2} + \frac{z_{ij2}\left(y_{ij} - u - d\right)^2}{2\sigma^2} \right\}.$$

Here, *const* is a constant, and *N* is the number of individuals. By differentiating the log-likelihood with respect to *u*, *a*, *d*, and σ^2, and setting the derivatives to zero, we have

$$\frac{\sum_i \sum_j^{n_i} z_{ij1} y_{ij}}{\sum_i \sum_j^{n_i} z_{ij1}} = \hat{u} + \hat{a}\,, \quad \frac{\sum_i \sum_j^{n_i} z_{ij2} y_{ij}}{\sum_i \sum_j^{n_i} z_{ij2}} = \hat{u} + \hat{d}\,,$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_i \sum_j^{n_i} \left\{ z_{ij1}\left(y_{ij} - u - a\right)^2 + z_{ij2}\left(y_{ij} - u - d\right)^2 \right\}.$$

These equations are sufficient to carry out the M step. We don't have the estimators of each parameter. Now we have updated values of the four parameters. By using these values, the next E step is carried out. Until the values of parameters converge, the rounds of the E step and the M steps are iterated. The way to obtain the LOD score are the same as intercross.

A.2.3 The E step of P_2 backcross

At the location where we calculate its LOD score (i.e., at pseudomarker), the genotype is estimated in a probabilistic manner. The genotype at the pseudomarker is estimated using the following nine cases of the flanking marker genotypes.

5: $A_1A_2B_1B_2$ 6: $A_1A_2B_2B_2$ 8: $A_2A_2B_1B_2$ 9: $A_2A_2B_2B_2$

Here, A and B represent the flanking markers and the suffixes one and two represent from whether maternal or paternal the marker inherits. On the other hand, the pseudomarker genotype is represented one of Q_1Q_2 and Q_2Q_2. Here, Q represents the pseudomarker between marker A and B. Assume that the probability that a recombination happens between A and Q is r_1, the probability that a recombination happens between Q and B is r_2, the probability that a recombination happens only by one degree between A and B is assumed to be r_{1+2}, and the probability that a recombination happens both between A and Q and between Q and B is r_{12}. When the genotypes of the flanking markers are case *i* as above, we denote the probability that the genotype of the pseudomarker is Q_1Q_2 and Q_2Q_2 by p_{i2} and p_{i3}, respectively. Here, *i* ranges from one to four, and each is corresponds to one of the fore marker genotypes above. The probabilities of p_{i2} and p_{i3} are represented as follows.

	Q_1Q_2 (p_{i2})	Q_2Q_2 (p_{i3})
5:$A_1A_2B_1B_2$	q_1	q_2
6:$A_1A_2B_1B_2$	q_3	q_4
8:$A_2A_2B_1B_2$	q_4	q_3
9:$A_2A_2B_2B_2$	q_2	q_1

Here,

$$q_1 = \frac{(1-r_1-r_2+r_{12})}{(1-r_{1+2})} \quad q_2 = \frac{r_{12}}{(1-r_{1+2})} \quad q_3 = \frac{(r_2-r_{12})}{r_{1+2}} \quad q_4 = \frac{(r_1-r_{12})}{r_{1+2}}$$

Using the assumption that the residue terms of Q_1Q_2 and Q_2Q_2 cases are normally distributed, the probability densities Φ_2 and Φ_3 represented as follows.

$$\phi_2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d)^2}{2\sigma^2}} \qquad \phi_3 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u+a)^2}{2\sigma^2}}$$

Therefore, frequencies of the three genotypes at the pseudomarker described as follows.

$$Q_1Q_2 : z_2 = \phi_2 p_{i2} / (\phi_2 p_{i2} + \phi_3 p_{i3})$$
$$Q_2Q_2 : z_3 = \phi_3 p_{i3} / (\phi_2 p_{i2} + \phi_3 p_{i3})$$

A.2.2 The M Step of P_2 backcross

The M step is carried out by using the result of the E step. The likelihood is represented as follows.

$$L \propto \prod_i \prod_j^{n_i} \left(p_{i2}\phi_{ij2}\right)^{z_{ij2}} \left(p_{i3}\phi_{ij3}\right)^{z_{ij3}}$$

Here, i indicates the genotype of the marker (one of the four types), and j indicates each individual ($1 \sim n_i$) that has the marker genotype.

Logarithm of the likelihood calculated as follows:

$$\ln(L) = const + \sum_i \sum_j^{n_i} \left(z_{ij2} \ln p_{i2} + z_{ij3} \ln p_{i3}\right) - 0.5N \ln\left(2\pi\sigma^2\right)$$

$$- \sum_i \sum_j^{n_i} \left\{ \frac{z_{ij2}\left(y_{ij} - u - d\right)^2}{2\sigma^2} + \frac{z_{ij3}\left(y_{ij} - u + a\right)^2}{2\sigma^2} \right\}.$$

Here, *const* is a constant, and N is the number of individuals. By differentiating the log-likelihood with respect to u, a, d, and σ^2, and setting the derivatives to zero, we have

$$\frac{\sum_i \sum_j^{n_i} z_{ij2} y_{ij}}{\sum_i \sum_j^{n_i} z_{ij2}} = \hat{u} + \hat{d}\,, \quad \frac{\sum_i \sum_j^{n_i} z_{ij3} y_{ij}}{\sum_i \sum_j^{n_i} z_{ij3}} = \hat{u} - \hat{a}\,,$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_i \sum_j^{n_i} \left\{ z_{ij2}\left(y_{ij} - u - d\right)^2 + z_{ij3}\left(y_{ij} - u + a\right)^2 \right\}.$$

These equations are sufficient to carry out the M step. We don't have the estimators of each parameter. Now we have updated values of the four parameters. By using these values, the next E step is carried out. Until the values of parameters converge, the rounds of the E step and the M steps are iterated. The way to obtain the LOD score are the same as intercross.

A.3 Derivation of two-dimensional interval mapping

Here, the detail of the derivation interval mapping that is abbreviated in 2.4.3 is explained.

$$\frac{\partial \ln(L)}{\partial u} = \frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i} \begin{Bmatrix} z_{ij1}\left(y_{ij} - u - a_1 - a_2 - i_{12}\right) + z_{ij2}\left(y_{ij} - u - a_1 - d_2 - j_{12}\right) + z_{ij3}\left(y_{ij} - u - a_1 + a_2 + i_{12}\right) \\ + z_{ij4}\left(y_{ij} - u - d_1 - a_2 - j_{21}\right) + z_{ij5}\left(y_{ij} - u - d_1 - d_2 - l_{12}\right) + z_{ij6}\left(y_{ij} - u - d_1 + a_2 + j_{21}\right) \\ + z_{ij7}\left(y_{ij} - u + a_1 - a_2 + i_{12}\right) + z_{ij8}\left(y_{ij} - u + a_1 - d_2 + j_{12}\right) + z_{ij9}\left(y_{ij} - u + a_1 + a_2 - i_{12}\right) \end{Bmatrix}$$

$$= \frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i} \begin{Bmatrix} \left(z_{ij1} + z_{ij2} + z_{ij3} + z_{ij4} + z_{ij5} + z_{ij6} + z_{ij7} + z_{ij8} + z_{ij9}\right) y_{ij} \\ -\left(z_{ij1} + z_{ij2} + z_{ij3} + z_{ij4} + z_{ij5} + z_{ij6} + z_{ij7} + z_{ij8} + z_{ij9}\right) u \\ -\left(z_{ij1} + z_{ij2} + z_{ij3} - z_{ij7} - z_{ij8} - z_{ij9}\right) a_1 - \left(z_{ij4} + z_{ij5} + z_{ij6}\right) d_1 \\ -\left(z_{ij1} - z_{ij3} + z_{ij4} - z_{ij6} + z_{ij7} - z_{ij9}\right) a_2 - \left(z_{ij2} + z_{ij5} + z_{ij8}\right) d_2 \\ -\left(z_{ij1} - z_{ij3} - z_{ij7} + z_{ij9}\right) i_{12} - \left(z_{ij2} - z_{ij8}\right) j_{12} - \left(z_{ij4} - z_{ij6}\right) j_{21} - z_{ij5} l_{12} \end{Bmatrix}$$

$$= \frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i} \begin{pmatrix} y_{ij} - u - \left(z_{ij1} + z_{ij2} + z_{ij3} - z_{ij7} - z_{ij8} - z_{ij9}\right) a_1 - \left(z_{ij4} + z_{ij5} + z_{ij6}\right) d_1 \\ -\left(z_{ij1} - z_{ij3} + z_{ij4} - z_{ij6} + z_{ij7} - z_{ij9}\right) a_2 - \left(z_{ij2} + z_{ij5} + z_{ij8}\right) d_2 \\ -\left(z_{ij1} - z_{ij3} - z_{ij7} + z_{ij9}\right) i_{12} - \left(z_{ij2} - z_{ij8}\right) j_{12} - \left(z_{ij4} - z_{ij6}\right) j_{21} - z_{ij5} l_{12} \end{pmatrix} = 0$$

As a result,

$$\hat{u}\sum_i^{81}\sum_j^{n_i}(1) = \sum_i^{81}\sum_j^{n_i} y_{ij} - \hat{a}_1\sum_i^{81}\sum_j^{n_i}\left(z_{ij1} + z_{ij2} + z_{ij3} - z_{ij7} - z_{ij8} - z_{ij9}\right) - \hat{d}_1\sum_i^{81}\sum_j^{n_i}\left(z_{ij4} + z_{ij5} + z_{ij6}\right)$$

$$- \hat{a}_2\sum_i^{81}\sum_j^{n_i}\left(z_{ij1} - z_{ij3} + z_{ij4} - z_{ij6} + z_{ij7} - z_{ij9}\right) - \hat{d}_2\sum_i^{81}\sum_j^{n_i}\left(z_{ij2} + z_{ij5} + z_{ij8}\right)$$

$$- \hat{i}_{12}\sum_i^{81}\sum_j^{n_i}\left(z_{ij1} - z_{ij3} - z_{ij7} + z_{ij9}\right) - \hat{j}_{12}\sum_i^{81}\sum_j^{n_i}\left(z_{ij2} - z_{ij8}\right) - \hat{j}_{21}\sum_i^{81}\sum_j^{n_i}\left(z_{ij4} - z_{ij6}\right) - \hat{l}_{12}\sum_i^{81}\sum_j^{n_i}\left(z_{ij5}\right)$$

・・・(1)

Moreover,

$$\frac{\partial \ln(L)}{\partial a_1} = \frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i} \begin{Bmatrix} z_{ij1}\left(y_{ij} - u - a_1 - a_2 - i_{12}\right) + z_{ij2}\left(y_{ij} - u - a_1 - d_2 - j_{12}\right) + z_{ij3}\left(y_{ij} - u - a_1 + a_2 + i_{12}\right) \\ - z_{ij7}\left(y_{ij} - u + a_1 - a_2 + i_{12}\right) - z_{ij8}\left(y_{ij} - u + a_1 - d_2 + j_{12}\right) - z_{ij9}\left(y_{ij} - u + a_1 + a_2 - i_{12}\right) \end{Bmatrix}$$

$$= \frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i} \begin{Bmatrix} \left(z_{ij1} + z_{ij2} + z_{ij3} - z_{ij7} - z_{ij8} - z_{ij9}\right) y_{ij} - \left(z_{ij1} + z_{ij2} + z_{ij3} - z_{ij7} - z_{ij8} - z_{ij9}\right) u \\ -\left(z_{ij1} + z_{ij2} + z_{ij3} + z_{ij7} + z_{ij8} + z_{ij9}\right) a_1 - \left(z_{ij1} - z_{ij3} - z_{ij7} + z_{ij9}\right) a_2 - \left(z_{ij2} - z_{ij8}\right) d_2 \\ -\left(z_{ij1} - z_{ij3} + z_{ij7} - z_{ij9}\right) i_{12} - \left(z_{ij2} + z_{ij8}\right) j_{12} \end{Bmatrix} = 0$$

As a result,

$$\hat{a}_1\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}+z_{ij2}+z_{ij3}+z_{ij7}+z_{ij8}+z_{ij9}\right)$$

$$=\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}+z_{ij2}+z_{ij3}-z_{ij7}-z_{ij8}-z_{ij9}\right)y_{ij}-\hat{u}\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}+z_{ij2}+z_{ij3}-z_{ij7}-z_{ij8}-z_{ij9}\right)$$

$$-\hat{a}_2\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)-\hat{d}_2\sum_i^{81}\sum_j^{n_i}\left(z_{ij2}-z_{ij8}\right)-\hat{i}_{12}\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}-z_{ij3}+z_{ij7}-z_{ij9}\right)-\hat{j}_{12}\sum_i^{81}\sum_j^{n_i}\left(z_{ij2}+z_{ij8}\right)$$

・・・(2)

Moreover,

$$\frac{\partial\ln(L)}{\partial d_1}=\frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i}\left\{z_{ij4}\left(y_{ij}-u-d_1-a_2-j_{21}\right)+z_{ij5}\left(y_{ij}-u-d_1-d_2-l_{12}\right)+z_{ij6}\left(y_{ij}-u-d_1+a_2+j_{21}\right)\right\}$$

$$=\frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i}\left\{\begin{matrix}\left(z_{ij4}+z_{ij5}+z_{ij6}\right)y_{ij}-\left(z_{ij4}+z_{ij5}+z_{ij6}\right)u-\left(z_{ij4}+z_{ij5}+z_{ij6}\right)d_1\\-\left(z_{ij4}-z_{ij6}\right)a_2-\left(z_{ij5}\right)d_2-\left(z_{ij4}-z_{ij6}\right)j_{21}-z_{ij5}l_{12}\end{matrix}\right\}$$

As a result,

$$\hat{d}_1\sum_i^{81}\sum_j^{n_i}\left(z_{ij4}+z_{ij5}+z_{ij6}\right)$$

$$=\sum_i^{81}\sum_j^{n_i}\left(z_{ij4}+z_{ij5}+z_{ij6}\right)y_{ij}-\hat{u}\sum_i^{81}\sum_j^{n_i}\left(z_{ij4}+z_{ij5}+z_{ij6}\right)$$

$$-\hat{a}_2\sum_i^{81}\sum_j^{n_i}\left(z_{ij4}-z_{ij6}\right)-\hat{d}_2\sum_i^{81}\sum_j^{n_i}\left(z_{ij5}\right)-\hat{j}_{21}\sum_i^{81}\sum_j^{n_i}\left(z_{ij4}-z_{ij6}\right)-\hat{l}_{12}\sum_i^{81}\sum_j^{n_i}\left(z_{ij5}\right)$$

・・・(3)

Moreover,

$$\frac{\partial\ln(L)}{\partial a_2}=\frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i}\left\{\begin{matrix}z_{ij1}\left(y_{ij}-u-a_1-a_2-i_{12}\right)-z_{ij3}\left(y_{ij}-u-a_1+a_2+i_{12}\right)+z_{ij4}\left(y_{ij}-u-d_1-a_2-j_{21}\right)\\-z_{ij6}\left(y_{ij}-u-d_1+a_2+j_{21}\right)+z_{ij7}\left(y_{ij}-u+a_1-a_2+i_{12}\right)-z_{ij9}\left(y_{ij}-u+a_1+a_2-i_{12}\right)\end{matrix}\right\}$$

$$=\frac{1}{\sigma^2}\sum_i^{81}\sum_j^{n_i}\left\{\begin{matrix}\left(z_{ij1}-z_{ij3}+z_{ij4}-z_{ij6}+z_{ij7}-z_{ij9}\right)y_{ij}-\left(z_{ij1}-z_{ij3}+z_{ij4}-z_{ij6}+z_{ij7}-z_{ij9}\right)u\\-\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)a_1-\left(z_{ij1}+z_{ij3}+z_{ij4}+z_{ij6}+z_{ij7}+z_{ij9}\right)a_2-\left(z_{ij4}-z_{ij6}\right)d_1\\-\left(z_{ij1}+z_{ij3}-z_{ij7}-z_{ij9}\right)i_{12}-\left(z_{ij4}+z_{ij6}\right)j_{21}\end{matrix}\right\}=0$$

As a result,

$$\hat{a}_2\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}+z_{ij3}+z_{ij4}+z_{ij6}+z_{ij7}+z_{ij9}\right)$$

$$=\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}-z_{ij3}+z_{ij4}-z_{ij6}+z_{ij7}-z_{ij9}\right)y_{ij}-\hat{u}\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}-z_{ij3}+z_{ij4}-z_{ij6}+z_{ij7}-z_{ij9}\right)$$

$$-\hat{a}_1\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)-\hat{d}_1\sum_i^{81}\sum_j^{n_i}\left(z_{ij4}-z_{ij6}\right)-\hat{i}_{12}\sum_i^{81}\sum_j^{n_i}\left(z_{ij1}+z_{ij3}-z_{ij7}-z_{ij9}\right)-\hat{j}_{21}\sum_i^{81}\sum_j^{n_i}\left(z_{ij4}+z_{ij6}\right)$$

・・・(4)

Moreover,

$$\frac{\partial \ln(L)}{\partial d_2} = \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{z_{ij2}\left(y_{ij}-u-a_1-d_2-j_{12}\right)+z_{ij5}\left(y_{ij}-u-d_1-d_2-l_{12}\right)+z_{ij8}\left(y_{ij}-u+a_1-d_2+j_{12}\right)\right\}$$

$$= \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{\begin{array}{l}\left(z_{ij2}+z_{ij5}+z_{ij8}\right)y_{ij}-\left(z_{ij2}+z_{ij5}+z_{ij8}\right)u-\left(z_{ij2}-z_{ij8}\right)a_1\\-\left(z_{ij5}\right)d_1-\left(z_{ij2}+z_{ij5}+z_{ij8}\right)d_2-\left(z_{ij2}-z_{ij8}\right)j_{12}-z_{ij5}l_{12}\end{array}\right\}$$

As a result,

$$\hat{d}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2}+z_{ij5}+z_{ij8}\right)$$

$$=\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2}+z_{ij5}+z_{ij8}\right)y_{ij}-\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2}+z_{ij5}+z_{ij8}\right)$$

$$-\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2}-z_{ij8}\right)-\hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij5}\right)-\hat{j}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2}-z_{ij8}\right)-\hat{l}_{21}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij5}\right)$$

・・・(5)

Moreover,

$$\frac{\partial \ln(L)}{\partial i_{12}} = \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{\begin{array}{l}z_{ij1}\left(y_{ij}-u-a_1-a_2-i_{12}\right)-z_{ij3}\left(y_{ij}-u-a_1+a_2+i_{12}\right)\\-z_{ij7}\left(y_{ij}-u+a_1-a_2+i_{12}\right)+z_{ij9}\left(y_{ij}-u+a_1+a_2-i_{12}\right)\end{array}\right\}$$

$$= \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{\begin{array}{l}\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)y_{ij}-\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)u\\-\left(z_{ij1}-z_{ij3}+z_{ij7}-z_{ij9}\right)a_1-\left(z_{ij1}+z_{ij3}-z_{ij7}-z_{ij9}\right)a_2\\-\left(z_{ij1}+z_{ij3}+z_{ij7}+z_{ij9}\right)i_{12}\end{array}\right\}$$

As a result,

$$\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}+z_{ij7}+z_{ij9}\right)$$

$$=\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)y_{ij}-\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)$$

$$-\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}+z_{ij7}-z_{ij9}\right)-\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}-z_{ij7}-z_{ij9}\right)$$

・・・(6)

Moreover,

$$\frac{\partial \ln(L)}{\partial j_{12}} = \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{z_{ij2}\left(y_{ij}-u-a_1-d_2-j_{12}\right)-z_{ij8}\left(y_{ij}-u+a_1-d_2+j_{12}\right)\right\}$$

$$= \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{\left(z_{ij2}-z_{ij8}\right)y_{ij}-\left(z_{ij2}-z_{ij8}\right)u-\left(z_{ij2}+z_{ij8}\right)a_1-\left(z_{ij2}-z_{ij8}\right)d_2-\left(z_{ij2}+z_{ij8}\right)j_{12}\right\}$$

As a result,

$$\hat{j}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2}+z_{ij8}\right)$$

$$= \sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} - z_{ij8}\right)y_{ij} - \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} - z_{ij8}\right) - \hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} + z_{ij8}\right) - \hat{d}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} + z_{ij8}\right) \quad \cdots (7)$$

Moreover,

$$\frac{\partial \ln(L)}{\partial j_{21}} = \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{z_{ij4}\left(y_{ij} - u - d_1 - a_2 - j_{21}\right) - z_{ij6}\left(y_{ij} - u - d_1 + a_2 + j_{21}\right)\right\}$$

$$= \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{\left(z_{ij4} - z_{ij6}\right)y_{ij} - \left(z_{ij4} - z_{ij6}\right)u - \left(z_{ij4} - z_{ij6}\right)d_1 - \left(z_{ij4} + z_{ij6}\right)a_2 - \left(z_{ij4} + z_{ij6}\right)j_{21}\right\}$$

As a result,

$$\hat{j}_{21}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} + z_{ij6}\right)$$

$$= \sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} - z_{ij6}\right)y_{ij} - \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} - z_{ij6}\right) - \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} - z_{ij6}\right) - \hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} + z_{ij6}\right) \quad \cdots (8)$$

Moreover,

$$\frac{\partial \ln(L)}{\partial l_{12}} = \frac{1}{\sigma^2}\sum_{i}^{81}\sum_{j}^{n_i}\left\{z_{ij5}\left(y_{ij} - u - d_1 - d_2 - l_{12}\right)\right\}$$

As a result,

$$\hat{l}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij5}\right) = \sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij5}\right)y_{ij} - \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij5}\right) - \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij5}\right) - \hat{d}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij5}\right) \quad \cdots (9)$$

From (3) – (9),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} + z_{ij6}\right)y_{ij}$$

$$= \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} + z_{ij6}\right) + \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} + z_{ij6}\right) + \hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} - z_{ij6}\right) + \hat{j}_{21}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij4} - z_{ij6}\right) \quad \cdots (10)$$

From (8) + (10),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij4}\right)y_{ij} = \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij4}\right) + \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij4}\right) + \hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij4}\right) + \hat{j}_{21}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij4}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij4}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij4}} = \hat{u} + \hat{d}_1 + \hat{a}_2 + \hat{j}_{21} \quad \cdots (11)$$

From (8) – (10),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij6}\right)y_{ij} = \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij6}\right) + \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij6}\right) + \hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij6}\right) + \hat{j}_{21}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij6}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij6} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij6}} = \hat{u} + \hat{d}_1 - \hat{a}_2 - \hat{j}_{21} \qquad \cdots (12)$$

From (5) – (9),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} + z_{ij8}\right)y_{ij}$$

$$= \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} + z_{ij8}\right) + \hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} - z_{ij8}\right) + \hat{d}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} + z_{ij8}\right) + \hat{j}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij2} - z_{ij8}\right) \qquad \cdots (13)$$

From (7) + (13),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij2}\right)y_{ij} = \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij2}\right) + \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij2}\right) + \hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij2}\right) + \hat{j}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij2}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij2} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij2}} = \hat{u} + \hat{a}_1 + \hat{d}_2 + \hat{j}_{12} \qquad \cdots (14)$$

From (7) – (13),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij8}\right)y_{ij} = \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij8}\right) + \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij8}\right) + \hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij8}\right) + \hat{j}_{21}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij8}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij8} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij8}} = \hat{u} - \hat{a}_1 + \hat{d}_2 - \hat{j}_{12} \qquad \cdots (15)$$

From (9),

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5}} = \hat{u} + \hat{d}_1 + \hat{d}_2 + \hat{l}_{21} \qquad \cdots (16)$$

From addition of (1) ~ (9),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij1} + 4z_{ij2} + 4z_{ij4} + 4z_{ij5}\right)y_{ij}$$

$$= \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij1} + 4z_{ij2} + 4z_{ij4} + 4z_{ij5}\right) + \hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij1} + 4z_{ij2}\right) + \hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij1} + 4z_{ij4}\right) + \hat{d}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij4} + 4z_{ij5}\right)$$

$$+ \hat{d}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij2} + 4z_{ij5}\right) + \hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij1}\right) + \hat{j}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij2}\right) + \hat{j}_{21}\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij4}\right) + \hat{l}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(4z_{ij5}\right)$$

From (11), (14), (16),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}\right)y_{ij} = \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}\right)+\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}\right)+\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}\right)+\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij1}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij1}} = \hat{u}+\hat{a}_1+\hat{a}_2+\hat{i}_{12} \qquad \cdots (17)$$

From (4) – (8),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}+z_{ij7}-z_{ij9}\right)y_{ij}$$

$$=\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}+z_{ij7}-z_{ij9}\right)+\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)+\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}+z_{ij7}+z_{ij9}\right)$$

$$+\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}-z_{ij7}-z_{ij9}\right) \qquad \cdots (18)$$

From (6) + (18),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}-2z_{ij3}\right)y_{ij}$$

$$=\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}-2z_{ij3}\right)+\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}-2z_{ij3}\right)+\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}+2z_{ij3}\right)+\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}+2z_{ij3}\right)$$

$$\cdots (19)$$

From (17),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij3}\right)y_{ij} = \hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij3}\right)+\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij3}\right)-\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij3}\right)-\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij3}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij3}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij3}} = \hat{u}+\hat{a}_1-\hat{a}_2-\hat{i}_{12} \qquad \cdots (20)$$

From (18) – (6),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij7}-2z_{ij9}\right)y_{ij}$$

$$=\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij7}-2z_{ij9}\right)-\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij7}-2z_{ij9}\right)+\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij7}+2z_{ij9}\right)-\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij7}+2z_{ij9}\right)$$

$$\cdots (21)$$

From (2) – (7),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}-z_{ij7}-z_{ij9}\right)y_{ij}$$

$$=\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}-z_{ij7}-z_{ij9}\right)+\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}+z_{ij3}+z_{ij7}+z_{ij9}\right)+\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}-z_{ij7}+z_{ij9}\right)$$

$$+\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij1}-z_{ij3}+z_{ij7}-z_{ij9}\right)$$

・・・(22)

From (6) + (12),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}-2z_{ij7}\right)y_{ij}$$

$$=\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}-2z_{ij7}\right)+\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}+2z_{ij7}\right)+\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}-2z_{ij7}\right)+\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(2z_{ij1}+2z_{ij7}\right)$$

・・・(23)

From (17),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij7}\right)y_{ij}=\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij7}\right)-\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij7}\right)+\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij7}\right)-\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij7}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i}z_{ij7}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i}z_{ij7}}=\hat{u}-\hat{a}_1+\hat{a}_2-\hat{i}_{12}$$ ・・・(24)

From (21),

$$\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij9}\right)y_{ij}=\hat{u}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij9}\right)-\hat{a}_1\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij9}\right)-\hat{a}_2\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij9}\right)+\hat{i}_{12}\sum_{i}^{81}\sum_{j}^{n_i}\left(z_{ij9}\right)$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i}z_{ij9}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i}z_{ij9}}=\hat{u}-\hat{a}_1-\hat{a}_2+\hat{i}_{12}$$ ・・・(25)

A.4 P_1 backcross and P_2 backcross of two-dimensional interval mapping

Here, in P_1 backcross, and P_2 backcross lines, details of the EM algorithm in two-dimensional interval mapping are explained.

A.4.1 The E step of P_1 backcross

At the location where we calculate its LOD score (i.e., at pseudomarker), the genotype is estimated in a probabilistic manner. The genotype at the pseudomarker is estimated using the

16 cases (Multiplication of four cases ($A_1A_1B_1B_1 \sim A_1A_2B_1B_2$) and four cases ($C_1C_1D_1D_1 \sim C_1C_2D_1D_2$)) of the flanking marker genotypes. Here, A and B represent the flanking markers and the suffixes one and two represent from whether maternal or paternal the marker inherits. On the other hand, the pseudomarker genotype is represented one of the following four cases.

1: $Q_1Q_1Q_3Q_3$ 2: $Q_1Q_1Q_3Q_4$ 4: $Q_1Q_2Q_3Q_3$ 5: $Q_1Q_2Q_3Q_4$

Here, Q_1 and Q_2 represent the pseudomarker between marker A and B, and Q_3 and Q_4 represent the pseudomarker between marker C and D. The suffixes one and two represent from whether maternal or paternal the pseudmarker inherits, and the suffixes three and four represent from whether maternal or paternal the pseudmarker inherits.

Assume that the probability that a recombination happens between A and Q is r_1, the probability that a recombination happens between Q and B is r_2, the probability that a recombination happens only by one degree between A and B is assumed to be r_{1+2}, and the probability that a recombination happens both between A and Q and between Q and B is r_{12}. Moreover, assume that the probability that a recombination happens between C and Q is r_3, the probability that a recombination happens between Q and D is r_4, the probability that a recombination happens only by one degree between C and D is assumed to be r_{3+4}, and the probability that a recombination happens both between C and Q and between Q and D is r_{34}. When the genotypes of the flanking markers are case *i* as above, we denote the probability that the genotype of the pseudomarker is $Q_1Q_1Q_1Q_1 \sim Q_1Q_2Q_1Q_2$ by p_{i1}, p_{i2}, p_{i4}, and p_{i5}, respectively. Here, *i* ranges from one to nine, and each is corresponds to one of the nine marker genotypes above. The probabilities of p_{i1}, p_{i2}, p_{i4}, and p_{i5} are obtained from the product of the probability in the gene locus of the first pseudmarker and the second pseudmarker.

The probabilities in the gene locus of the first pseudmarker are represented as follows.

	Q_1Q_1 ($p_{i1,2}$)	Q_1Q_2 ($p_{i4,5}$)
1：$A_1A_1B_1B_1$	q_1	q_2
2：$A_1A_1B_1B_2$	q_3	q_4
4：$A_1A_2B_1B_1$	q_4	q_3
5：$A_1A_2B_1B_2$	q_2	q_1

Here,

$$q_1 = \frac{(1-r_1-r_2+r_{12})}{(1-r_{1+2})} \quad q_2 = \frac{r_{12}}{(1-r_{1+2})} \quad q_3 = \frac{(r_2-r_{12})}{r_{1+2}} \quad q_4 = \frac{(r_1-r_{12})}{r_{1+2}}$$

The probabilities in the gene locus of the second pseudmarker are represented as follows.

	Q_3Q_3 ($p_{i1,4}$)	Q_3Q_4 ($p_{i2,5}$)
1 : $C_1C_1D_1D_1$	q_1	q_2
2 : $C_1C_1D_1D_2$	q_3	q_4
4 : $C_1C_2D_1D_1$	q_4	q_3
5 : $C_1C_2D_1D_2$	q_2	q_1

Here,

$$q_5 = \frac{(1-r_3-r_4+r_{34})}{(1-r_{3+4})} \quad q_6 = \frac{r_{34}}{(1-r_{3+4})} \quad q_7 = \frac{(r_4-r_{34})}{r_{3+4}} \quad q_8 = \frac{(r_3-r_{34})}{r_{3+4}}$$

Using the assumption that the residue terms of $Q_1Q_1Q_3Q_3 \sim Q_1Q_2Q_3Q_4$ cases are normally distributed, the probability densities Φ_1, Φ_2, Φ_4, and Φ_5 represented as follows.

$$\phi_1 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-a_1-a_2-i_{12})^2}{2\sigma^2}} \quad \phi_2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-a_1-d_2-j_{12})^2}{2\sigma^2}}$$

$$\phi_4 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d_1-a_2-j_{21})^2}{2\sigma^2}} \quad \phi_5 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d_1-d_2-l_{12})^2}{2\sigma^2}}$$

Therefore, frequencies of the nine genotypes at the pseudomarker described as follows.

$$Q_1Q_1Q_3Q_3 : z_1 = \frac{\phi_1 p_{i1}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_1Q_3Q_4 : z_2 = \frac{\phi_2 p_{i2}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_2Q_3Q_3 : z_4 = \frac{\phi_4 p_{i4}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_2Q_3Q_4 : z_5 = \frac{\phi_5 p_{i5}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

A.4.2 The M Step of P_1 backcross

The M step is carried out by using the result of the E step. The likelihood is represented as follows.

$$L \propto \prod_{i}^{16} \prod_{j}^{n_i} \left(p_{i1}\phi_{ij1}\right)^{z_{ij1}} \left(p_{i2}\phi_{ij2}\right)^{z_{ij2}} \left(p_{i4}\phi_{ij4}\right)^{z_{ij4}} \left(p_{i5}\phi_{ij5}\right)^{z_{ij5}}$$

Here, i indicates the genotype of the marker (one of the 16 types), and j indicates each individual ($1 \sim n_i$) that has the marker genotype.

Logarithm of the likelihood calculated as follows:

$$\ln(L) = const + \sum_{i}^{16}\sum_{j}^{n_i}\left(z_{ij1}\ln p_{i1} + z_{ij2}\ln p_{i2} + z_{ij4}\ln p_{i4} + z_{ij5}\ln p_{i5}\right) - 0.5N\ln\left(2\pi\sigma^2\right)$$

$$-\sum_{i}^{16}\sum_{j}^{n_i}\left\{\begin{array}{l}\dfrac{z_{ij1}\left(y_{ij}-u-a_1-a_2-i_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij2}\left(y_{ij}-u-a_1-d_2-j_{12}\right)^2}{2\sigma^2} \\ + \dfrac{z_{ij4}\left(y_{ij}-u-d_1-a_2-j_{21}\right)^2}{2\sigma^2} + \dfrac{z_{ij5}\left(y_{ij}-u-d_1-d_2-l_{12}\right)^2}{2\sigma^2}\end{array}\right\}.$$

Here, *const* is a constant, and N is the number of individuals. By differentiating the log-likelihood with respect to u, a_1, d_1, a_2, d_2, i_{12}, j_{12}, j_{21}, l_{12}, and σ^2, and setting the derivatives to zero, we have as follows.

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij1}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij1}} = \hat{u} + \hat{a}_1 + \hat{a}_2 + \hat{i}_{12} \qquad \frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij2}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij2}} = \hat{u} + \hat{a}_1 + \hat{d}_2 + \hat{j}_{12}$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij4}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij4}} = \hat{u} + \hat{d}_1 + \hat{a}_2 + \hat{j}_{21} \qquad \frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5}y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5}} = \hat{u} + \hat{d}_1 + \hat{d}_2 + \hat{l}_{21}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i}^{9}\sum_{j}^{n_i}\left\{\begin{array}{l} z_{ij1}\left(y_{ij}-u-a_1-a_2-i_{12}\right)^2 + z_{ij2}\left(y_{ij}-u-a_1-d_2-j_{12}\right)^2 \\ + z_{ij4}\left(y_{ij}-u-d_1-a_2-j_{21}\right)^2 + z_{ij5}\left(y_{ij}-u-d_1-d_2-l_{12}\right)^2\end{array}\right\}$$

These equations are sufficient to carry out the M step. We don't have the estimators of each parameter. Now we have updated values of the ten parameters. By using these values, the next E step is carried out. Until the values of parameters converge, the rounds of the E step and the M steps are iterated. The way to obtain the LOD score are the same as intercross.

A.4.3 The E step of P_2 backcross

At the location where we calculate its LOD score (i.e., at pseudomarker), the genotype is estimated in a probabilistic manner. The genotype at the pseudomarker is estimated using the 16 cases (Multiplication of four cases ($A_1A_2B_1B_2 \sim A_2A_2B_2B_2$) and four cases ($C_1C_2D_1D_2 \sim C_2C_2D_2D_2$)) of the flanking marker genotypes. Here, A and B represent the flanking markers and the suffixes one and two represent from whether maternal or paternal the marker inherits. On the other hand, the pseudomarker genotype is represented one of the following four cases.

5: $Q_1Q_2Q_3Q_4$ 6: $Q_1Q_2Q_4Q_4$ 8: $Q_2Q_2Q_3Q_4$ 9: $Q_2Q_2Q_4Q_4$

Here, Q_1 and Q_2 represent the pseudomarker between marker A and B, and Q_3 and Q_4 represent the pseudomarker between marker C and D. The suffixes one and two represent from whether maternal or paternal the pseudmarker inherits, and the suffixes three and four represent from whether maternal or paternal the pseudmarker inherits.

Assume that the probability that a recombination happens between A and Q is r_1, the probability that a recombination happens between Q and B is r_2, the probability that a recombination happens only by one degree between A and B is assumed to be r_{1+2}, and the probability that a recombination happens both between A and Q and between Q and B is r_{12}. Moreover, assume that the probability that a recombination happens between C and Q is r_3, the probability that a recombination happens between Q and D is r_4, the probability that a recombination happens only by one degree between C and D is assumed to be r_{3+4}, and the probability that a recombination happens both between C and Q and between Q and D is r_{34}. When the genotypes of the flanking markers are case *i* as above, we denote the probability that the genotype of the pseudomarker is $Q_1Q_2Q_1Q_2 \sim Q_2Q_2Q_2Q_2$ by p_{i5}, p_{i6}, p_{i8}, and p_{i9}, respectively. Here, *i* ranges from one to nine, and each is corresponds to one of the nine marker genotypes above. The probabilities of p_{i5}, p_{i6}, p_{i8}, and p_{i9} are obtained from the product of the probability in the gene locus of the first pseudmarker and the second pseudmarker.

The probabilities in the gene locus of the first pseudmarker are represented as follows.

	Q_1Q_2 ($p_{i5,6}$)	Q_2Q_2 ($p_{i8,9}$)
1：$A_1A_2B_1B_2$	q_1	q_2
2：$A_1A_2B_2B_2$	q_3	q_4
4：$A_2A_2B_1B_2$	q_4	q_3
5：$A_2A_2B_2B_2$	q_2	q_1

Here,

$$q_1 = \frac{(1 - r_1 - r_2 + r_{12})}{(1 - r_{1+2})} \quad q_2 = \frac{r_{12}}{(1 - r_{1+2})} \quad q_3 = \frac{(r_2 - r_{12})}{r_{1+2}} \quad q_4 = \frac{(r_1 - r_{12})}{r_{1+2}}$$

The probabilities in the gene locus of the second pseudmarker are represented as follows.

	Q_3Q_4 ($p_{i5,8}$)	Q_4Q_4 ($p_{i6,9}$)
1 : $C_1C_2D_1D_2$	q_1	q_2
2 : $C_1C_2D_2D_2$	q_3	q_4
4 : $C_2C_2D_1D_2$	q_4	q_3
5 : $C_2C_2D_2D_2$	q_2	q_1

Here,

$$q_5 = \frac{(1 - r_3 - r_4 + r_{34})}{(1 - r_{3+4})} \quad q_6 = \frac{r_{34}}{(1 - r_{3+4})} \quad q_7 = \frac{(r_4 - r_{34})}{r_{3+4}} \quad q_8 = \frac{(r_3 - r_{34})}{r_{3+4}}$$

Using the assumption that the residue terms of $Q_1Q_2Q_3Q_4 \sim Q_2Q_2Q_4Q_4$ cases are normally distributed, the probability densities Φ_5, Φ_6, Φ_8, and Φ_9 represented as follows.

$$\phi_5 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d_1-d_2-l_{12})^2}{2\sigma^2}} \quad \phi_6 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u-d_1+a_2+j_{21})^2}{2\sigma^2}}$$

$$\phi_8 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u+a_1-d_2+j_{12})^2}{2\sigma^2}} \quad \phi_9 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-u+a_1+a_2-i_{12})^2}{2\sigma^2}}$$

Therefore, frequencies of the nine genotypes at the pseudomarker described as follows.

$$Q_1Q_2Q_3Q_4 : z_5 = \frac{\phi_5 p_{i5}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_1Q_2Q_4Q_4 : z_6 = \frac{\phi_6 p_{i6}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_2Q_2Q_3Q_4 : z_8 = \frac{\phi_8 p_{i8}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

$$Q_2Q_2Q_4Q_4 : z_9 = \frac{\phi_9 p_{i9}}{\phi_1 p_{i1} + \phi_2 p_{i2} + \phi_3 p_{i3} + \phi_4 p_{i4} + \phi_5 p_{i5} + \phi_6 p_{i6} + \phi_7 p_{i7} + \phi_8 p_{i8} + \phi_9 p_{i9}}$$

A.4.4 The M Step of P_2 backcross

The M step is carried out by using the result of the E step. The likelihood is represented as follows.

$$L \propto \prod_{i}^{16} \prod_{j}^{n_i} \left(p_{i5}\phi_{ij5}\right)^{z_{ij5}} \left(p_{i6}\phi_{ij6}\right)^{z_{ij6}} \left(p_{i8}\phi_{ij8}\right)^{z_{ij8}} \left(p_{i9}\phi_{ij9}\right)^{z_{ij9}}$$

Here, i indicates the genotype of the marker (one of the 16 types), and j indicates each individual ($1 \sim n_i$) that has the marker genotype.

Logarithm of the likelihood calculated as follows:

$$\ln(L) = const + \sum_{i}^{81} \sum_{j}^{n_i} \left(z_{ij5} \ln p_{i5} + z_{ij6} \ln p_{i6} + z_{ij8} \ln p_{i8} + z_{ij9} \ln p_{i9}\right) - 0.5N \ln\left(2\pi\sigma^2\right)$$

$$- \sum_{i}^{81} \sum_{j}^{n_i} \left\{ \begin{array}{l} \dfrac{z_{ij5}\left(y_{ij} - u - d_1 - d_2 - l_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij6}\left(y_{ij} - u - d_1 + a_2 + j_{21}\right)^2}{2\sigma^2} \\ + \dfrac{z_{ij8}\left(y_{ij} - u + a_1 - d_2 + j_{12}\right)^2}{2\sigma^2} + \dfrac{z_{ij9}\left(y_{ij} - u + a_1 + a_2 - i_{12}\right)^2}{2\sigma^2} \end{array} \right\}.$$

Here, *const* is a constant, and N is the number of individuals. By differentiating the log-likelihood with respect to u, a_1, d_1, a_2, d_2, i_{12}, j_{12}, j_{21}, l_{12}, and σ^2, and setting the derivatives to zero, we have as follows.

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij5}} = \hat{u} + \hat{d}_1 + \hat{d}_2 + \hat{l}_{21} \qquad \frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij6} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij6}} = \hat{u} + \hat{d}_1 - \hat{a}_2 - \hat{j}_{21}$$

$$\frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij8} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij8}} = \hat{u} - \hat{a}_1 + \hat{d}_2 - \hat{j}_{12} \qquad \frac{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij9} y_{ij}}{\sum_{i}^{81}\sum_{j}^{n_i} z_{ij9}} = \hat{u} - \hat{a}_1 - \hat{a}_2 + \hat{i}_{12}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i}^{9} \sum_{j}^{n_i} \left\{ \begin{array}{l} z_{ij5}\left(y_{ij} - u - d_1 - d_2 - l_{12}\right)^2 + z_{ij6}\left(y_{ij} - u - d_1 + a_2 + j_{21}\right)^2 \\ + z_{ij8}\left(y_{ij} - u + a_1 - d_2 + j_{21}\right)^2 + z_{ij9}\left(y_{ij} - u + a_1 + a_2 - i_{12}\right)^2 \end{array} \right\}$$

These equations are sufficient to carry out the M step. We don't have the estimators of each parameter. Now we have updated values of the ten parameters. By using these values, the next E step is carried out. Until the values of parameters converge, the rounds of the E step and the M steps are iterated. The way to obtain the LOD score are the same as intercross.

Acknowledgement

I thank associate professor Nakaya for his instruction very much.

References

Beamer,W.G., Shultz,K.L., Churchill,G.A., Frankel,W.N., Baylink,D.J., Rosen.C.J., Donahue,L.R. (1999) Quantitative trait loci for bone density in C57BL/6J and CAST/EiJ inbred mice. Mammalian Genome, 10, 1043-1049.

Dempster, A., N. Laird and D. Rubin. (1977) Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B 39, 1–22.

DiPetrillo,K., Tsaih,S-W., Sheehan,S., Johns,C., Kelmenson,P., Gavras,H., Churchill,G.A., and Paigen, B. (2004) Genetic analysis of blood pressure in C3H/HeJ and SWR/J mice. Genomics, 17, 215-220

Haley,C.S. and Knott,S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity, 69, 315–324.

Kao, C.-H., Z-B. Zeng and R. D. Teasdale. (1999) Multiple interval mapping for quantitative trait loci. Genetics 152, 1203–1216.

Lander, E., and D. Botstein. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185–199.

Sax, K. (1923) The association of size difference with seed-coat pattern and pigmentation in phaseolus vulgaris. Genetics 8, 552–560.

Sen,´S. and Churchill,G.A. (2001) A statistical framework for quantitative trait mapping. Genetics, 159, 371–387.

Sugiyama,F., Churchill,G.A., Higgins,D.C., Johns,C., Makaritsis,K.P., Gavras,H. and Paigen,B. (2001) Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. Genomics, 71, 70–77.

Sugiyama,F., Churchill,G.A., Li,R., LAURA J. M. LibbyL.J.M., Carver,T., Yagami,K., John,S.W.M., and Paigen,B. (2002) QTL associated with blood pressure, heart rate, and heart weight in CBA/CaJ and BALB/cJ mice. Genomics, 10, 5-12.

Thoday, J. (1961) Location of polygenes. Nature 191,368–370.

Wittenburg H., Lammert F., Wang D.Q-H., Churchill.G.A., LI R., Bouchard G., CAREY,M.C., and Paigen,B. Interacting QTLs for cholesterol gallstones and gallbladder mucin in AKR and SWR strains of mice. (2002) Genomics, 8, 67-77.

Yoshikawa T., Watanabe A., Ishitsuka Y., Nakaya A., and Nakatani N. Identification of Multiple Genetic Loci Linked to the Propensity for “Behavioral Despair” in Mice. (2002) Genome Research, 12, 357-366.

Zeng Z-B., Liu J., Stam L. F., Kao C-H., Mercer J. M. and Laurie C. C. Genetic Architecture of a Morphological Shape Difference Between Two Drosophila Species. (2000) Genetics, 154, 299-310.