DNA Research **14**, 217–225, (2007) doi:10.1093/dnares/dsm020

Genotype Matrix Mapping: Searching for Quantitative Trait Loci Interactions in Genetic Variation in Complex Traits

Sachiko Isobe[1,3,*], Akihiro Nakaya[2], and Satoshi Tabata[3]

National Agricultural Research Center for Hokkaido Region, Histujigaoka 1, Toyohira, Sapporo 062-8555, Japan[1]; Department of computational Biology, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8561, Japan[2] and Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu, Chiba 292-0818, Japan[3]

(Received 25 July 2007; accepted October 8, 2007; published online 13 November 2007)

Abstract

In order to reveal quantitative trait loci (QTL) interactions and the relationship between various interactions in complex traits, we have developed a new QTL mapping approach, named genotype matrix mapping (GMM), which searches for QTL interactions in genetic variation. The central approach in GMM is the following. (1) Each tested marker is given a virtual matrix, named a genotype matrix (GM), containing intersecting lines and rows equal to the total allele number for that marker in the population analyzed. (2) QTL interactions are then estimated and compared through virtual networks among the GMs. To evaluate the contribution of marker combinations to a quantitative phenotype, the GMM method divides the samples into two non-overlapping subclasses, S_0 and S_1; the former contains the samples that have a specific genotype pattern to be evaluated, and the latter contains samples that do not. Based on this division, the F-measure is calculated as an index of significance. With the GMM method, we extracted significant marker combinations consisting of one to three interacting markers. The results indicated there were multiple QTL interactions affecting the phenotype (flowering date). GMM will be a valuable approach to identify QTL interactions in genetic variation of a complex trait within a variety of organisms.

Key words: genotype matrix mapping; QTL interaction; genetic variation

1. Introduction

Many complex traits of medical and agricultural importance, such as blood pressure, diabetes, crop yield, and plant stress resistance, are controlled by quantitative trait loci (QTLs).[1,2] The phenotypic variation of a complex trait usually results from multiple QTLs, QTL–QTL interactions, and QTL–environmental interactions.[3] These complex interactions make identification of individual QTL and QTL interactions difficult. However, recent advances in molecular biology and genomics will now enable us to identify and dissect QTLs related to complex traits in connection with genomic information by employing a statistical method. Once candidate QTLs have been identified, such genetic regions can be broken down into responsible genes by a map-based cloning approach with the aid of DNA markers and genetic linkage maps.[4] The first successful example of gene identification and cloning from a naturally occurring allelic variation was *Hd1*, a major QTL responsible for photoperiod sensitivity in rice.[5]

Two major approaches have been used to investigate QTLs contributing to complex traits: linkage analysis and association analysis. Linkage analysis exploits the shared inheritance of functional polymorphisms and adjacent markers within families or pedigrees of known ancestry,[1] whereas association analysis is an approach to detect QTL localization based on linkage disequilibrium in unrelated individuals or natural populations.[6] Both these approaches have contributed to our understanding of single QTLs in a great variety of species; however,

Edited by Masahiro Yano

* To whom correspondence should be addressed. Kazusa DNA Research Institute 2-6-7, Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan. Tel. +81 438-52-3928. Fax +81 438-52-3934. E-mail: sisobe@kazusa.or.jp

neither approach can be used to detect interactions among multiple QTLs. Although composite interval mapping (CIM)[7] and multiple QTL mapping (MQM)[8] can identify multiple QTLs by including additional background markers in the model, these approaches can only precisely locate QTLs with significant main effects and do not detect epistatic QTLs.[9]

Recently, several approaches for the detection of epigenetic gene interactions have been developed, primarily in the field of human genetics, by application of regression analysis, neural networks, and non-parametric methods.[10] One of the typical approaches is fitting a multiple regression model and, thus, relating the trait values to marker genotypes.[9,11–13] In this approach, the number of QTLs in the population by first making estimates based on the Bayesian information criterion and then QTL location and interaction fitting in the model are assumed. The weak point of this approach is that the detection power decreases as the number of parameters increases. So far, identification of presumptive epistatic QTLs using the multiple regression model has been demonstrated only with inter-crossed populations. The other commonly used approach is the multifactor dimensionality reduction (MDR) method, which uses non-parametric methods and is most efficiently used with case–control data.[14–16] The MDR method pools multilocus genotypes into a single dimension with two groups classified as either case or control data. Because this approach was developed for genetic epidemiologists, sequential phenotypic data cannot be adapted to this analysis.

Similar to the multiple regression model and MDR method, other current approaches available for detection of epistatic QTLs have their own advantages and disadvantages. For example, some methods are incapable of detecting epistatic QTLs in the absence of main effect QTLs or interactions among more than three loci, and others have not developed software programs for general users.[10] In addition, we recognized a common disadvantage in these current approaches; that is, all of them were developed to generate a best solution (QTL interaction) among numerous candidate interactions, and do not consider relations among subsets of interacting QTLs. Because the final phenotypic value of a complex trait should be determined through multifactor interactions, there may not be just one best interaction in a complex trait, but multiple interactions related to a complex trait. Therefore, we considered that comparison of many QTL interactions at once is vital for a comprehensive understanding of QTL expression affecting a complex trait.

In order to reveal QTL interactions and interaction–interaction relationships in complex traits, we have developed a new QTL mapping approach named genotype matrix mapping (GMM), which searches for QTL interactions not only in family data, but also in various genetic backgrounds. In GMM, each marker is given a matrix in which each of the total number of alleles for the marker in the tested population is represented by intersecting lines and rows. QTL interactions are estimated and compared through virtual networks generated among the locus matrixes. In this report, we described the concept of GMM and investigate the efficiency of the concept using a real genotype–phenotype data set.

2. Materials and methods

2.1. Red clover data set

For development and investigation of the GMM algorithm, a real genotype–phenotype data set derived from red clover (*Trifolium pratense* L.) was used. Red clover is diploid and the genome size is less than 440 Mb.[17] It is an allogamous species and each variety contains a high level of heterozygosity.[18]

A total of 74 red clover individuals originating from ten varieties bred in different countries were used: 'Natsuyu' (Japan), 'Hokuseki' (Japan), 'Sapporo' (Japan), 'Rannij2' (Russia), 'Start' (Czechoslovakia), 'Kurano' (Denmark), 'Renova' (Switzerland), 'Merviot' (Belgium), 'Kenland' (USA) and 'Altaswede' (Canada). Genomic DNA was extracted from five to eight individuals of each variety and subjected to PCR examination using 106 primer pairs for selected microsatellite markers distributed throughout the genome.[17] The presence or absence of amplification and differences in fragment size were scored as different alleles. For phenotype data, the flowering date was estimated for each individual grown under conditions of 20°C and an 18 h light/6 h dark cycle in a greenhouse. The possibility of population structure in the tested 74 individuals was estimated by the 'Structure' (version 2.0) program with the following parameters: length of burning period = 10 000, Number of MCMC population in the burning period = 1000 (http://pritch.bsd.uchicago.edu/software.html).[19]

Phenotype values for red clover flowering date were distributed over 26 days (Fig. 1A). The average number of alleles per marker was 6.5, and the range was between 2 and 10 (Fig. 1B). The number of genotypes (observed two allele combinations per marker) ranged from 2 to 31 (Fig. 1C). The absence of population structure was confirmed by comparison of Ln $P(D) = -22476.9$ ($K = 1$) and Ln $P(D) = -22507.9$ ($K = 2$). A schematic view of the dataset is presented in Fig. 2.

2.2. Genotype matrix mapping

To carry out our analysis, we used a newly developed program written in C++ language. This program uses the obtained marker genotypes and the phenotypes of the individuals as its input data set (Fig. 2), and extracts all significant QTLs and QTL interactions in an exhaustive manner without omission. The results are available in a computer-parsable format and can be converted into graphical presentations, as shown in this report. In

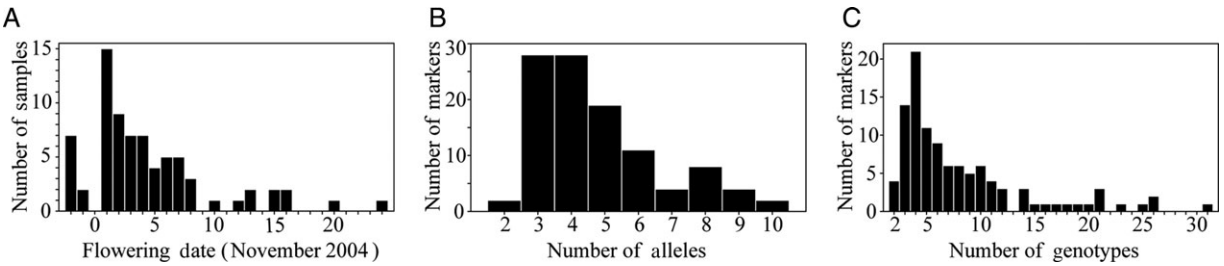

Figure 1. Phenotype and genotype data from the red clover germplasms. Distribution of flowering date (**A**), number of alleles per locus (**B**), and number of genotypes per marker (**C**) are summarized.

contrast to interval mapping, the GMM algorithm does not necessarily require relative genetic distances between markers, since it does not estimate genotypes at putative loci between two flanking marker loci. It uses map information, if available, only for visualizing the results.

3. Results

3.1. Central concept of genotype matrix mapping

Any type of population, including unrelated individuals, family data, and mapping populations for linkage analysis, can be used for GMM, providing that there is no population structure within the tested data set. The number of alleles in the population should be determined for every marker before analysis. Next, each marker is given a virtual matrix, named a genotype matrix (GM), in which line and row numbers are allotted based on the total number of alleles for the marker in the population (Fig. 3A). For example, when the number of alleles of 'Marker A' is five, 'Marker A' is given half a 5×5 matrix covering all possible allele combinations. Individual genotypes are plotting onto the matrix. For example, 'Individual X' with genotype 'ab' for 'Marker A', plots to the cell where line 'a' and row 'b' intersect. The estimated existence of a significant relationship between genotype and phenotype values based on analysis of variance (ANOVA), as in association analysis, is determined when a number of individuals are plotted to the same cell. Allelic interaction between 'a' and 'b' on 'Marker A' is estimated by comparison the phenotype distribution of 'aa', 'ab', and 'bb' cells on the GM of 'Marker A'. Additionally, the general effect of allele 'a' can be estimated by comparison the phenotypes among all the cells in 'line a'.

Finally, virtual networks created by through multiple GMs, named genotype matrix networks (GMNs), connected by each cell are generated. When a particular network indicates a significant relationship to the phenotype, the marker–allele combinations assigned on the GMs are considered a QTL interaction combination. In the example in Fig. 3B, the combination of Marker A(ab)–Marker B(bc)–Marker C(bd) strongly influences the phenotype (red line), the combination of Marker A(ab)–Marker B(ac)–Marker C(bd) has an intermediate influence (orange line), and the combination of Marker A(ab)–Marker B(ac)–Marker C(ae) has a weak influence (green line). The letters in parentheses indicate allele types. In summary, QTL interactions are identified by finding significant interactions between multiple GMs using GMNs. In theory, the potential number of conjunct

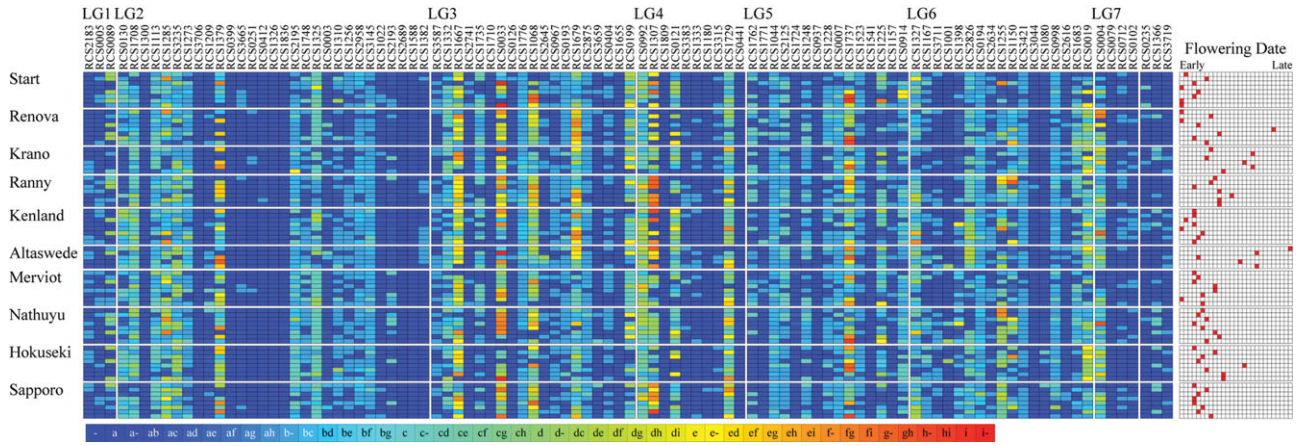

Figure 2. A schematic view of the genotype data set. Each line and row indicates red clover individuals and marker genotypes, respectively. Colors of cells indicate allele combination as listed at the bottom. Allele types are indicated by lower case, and '-' indicates the absence of amplifications or undetermined alleles of dominant markers. The matrix at the right indicates phenotypic values (flowering date) of each individual (red dots) from early to late, corresponding to 29 October 2004 to 24 November 2004, respectively.

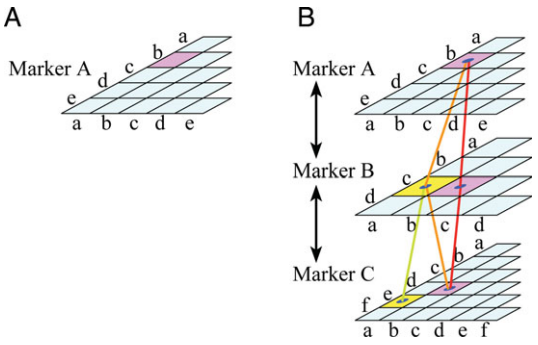

Figure 3. Schematic representation of GM and GMN. (**A**) A virtual matrix given for 'Marker A', which is composed of five alleles. The pink cell indicates the 'ab' genotype. (**B**) Detection of marker locus interactions by virtual networks among multiple matrices. Red, orange, and green lines indicate interactions which influence the phenotype to various extent (see text).

markers per combination ranges from one to the maximum number of tested markers.

3.2. Detection of QTL interactions

The first step of the GMM method is to construct a list of locus combinations. Each locus combination in the list is then depicted on the GM of the marker, as explained above. To evaluate the significance of the locus combinations, we used the *F*-measure.[20] The total set S that consists of N individuals is divided into two non-overlapping subclasses S_0 and S_1 according to their marker genotypes. Here, we have defined S_1 as the samples that have the specific genotype pattern to be evaluated, and S_0 as the samples that do not. When the aggregative effect of markers is evaluated, the samples are subdivided using the multiple markers instead of a single marker. Next, the mean square among classes is calculated as follows:

$$\begin{aligned}\mathrm{MSA} &= \sum_{j=0}^{1} |S_j|(\mu_j - \mu)^2 \\ &= |S_0|(\mu_0 - \mu)^2 + |S_1|(\mu_1 - \mu)^2.\end{aligned}$$

Here, μ, μ_0, and μ_1 are the means of the phenotype values in S, S_0, and S_1, respectively. |X| is the number of individuals in X.

The mean square within each class is defined as follows:

$$\begin{aligned}\mathrm{MSW} &= \frac{\sum_{j=0}^{1} \sum_{s_i \in Sj} (P_i - \mu_j)^2}{N-2} \\ &= \frac{\sum_{s_i \in S_0} (P_i - \mu_0)^2 + \sum_{s_i \in S_1} (P_i - \mu_1)^2}{N-2}.\end{aligned}$$

Here, P_i is the phenotype value of the i-th individual s_i. The *F*-measure is obtained by dividing the MSA by the MSW ($F = \mathrm{MSA}/\mathrm{MSW}$) and indicates the bias of the distribution of phenotype values in the two subclasses. If the distribution of the phenotype in the two subclasses differs, one can conclude that the condition (i.e., the pattern of marker genotypes) used for sample division is associated with the phenotype of interest. In such cases, the *F*-measure value is large.

Significant locus combinations that have large *F*-measure values are searched in an incremental manner. During the searching procedure, the maximum *F*-measure obtained is F_{opt}, and the value of F_{opt} is updated when a better combination whose *F*-measure is higher than the current F_{opt} is identified. Additionally, d is a given margin value and the final result of this method includes all combinations whose *F*-measure is less than the maximum by this margin value d. To fulfill this condition, starting with a single marker, another marker is concatenated to the first marker with each round of the incremental search until the given maximum length L is reached. When each additional marker is concatenated to a combination, the *F*-measure is evaluated to judge whether the combination is significant or not (i.e., higher than F_{opt}-d or not). If the combination is significant, it is included in the result, and the upper boundary of the *F*-measure that the combination can reach by further concatenation is calculated.[21] If this upper boundary is less than F_{opt} - d, significant combinations cannot be obtained by further computation, and the current search is therefore terminated and evaluation of another set of combinations is started. When all searches are terminated, all the optimum and sub-optimum combinations are included in the result.

3.3. GMM search on a red clover data set

We examined the feasibility of the GMM method using the datasets from red clover, which consist of 106 microsatellite markers distributed along the entire genome for the genotype dataset and the flowering date for the phenotype dataset. Using the GMM algorithm, we tentatively extracted the significant locus combinations consisting of at most $L = 3$ markers. Table 1 shows all the combinations of markers whose *F*-measure value was higher than 35.0 (searching was carried out setting the margin value $d =$ 10.0 and only combinations higher than $F = 35.0$ are shown). All the combinations with high *F*-measure values listed in Table 1 are concatenations of three loci. This means that the *F*-measure values obtained using combinations of one or two loci were lower than those obtained using combinations of three loci. The combination of RCS2958(b-), RCS0914(b-), and RCS1300(ab) produced the highest *F*-measure score (43.9), and extracts samples with the three highest phenotype values (late-flowering samples). Fig. 4 illustrates the distribution of phenotype values extracted using each combination consisting of one, two, or three of the three loci. The *F*-measure values obtained using the single loci RCS2958(b-), RCS0914(b-)

Table 1. Significant locus combinations consisting of one to three interacting markers whose *F*-measures were >35.0

F-measure	Relevant samples[a]		Others		Number of loci	Interacted locus (allele)		
	Number of samples	Mean[b]	Number of samples	Mean[b]				
43.9	3	20.0	71	3.8	3	RCS2958(b-)	RCS0914(b-)	RCS1300(ab)
43.9	3	20.0	71	3.8	3	RCS1113(bc)	RCS0914(b-)	RCS0235(-)
43.9	3	20.0	71	3.8	3	RCS1113(bc)	RCS0914(b-)	RCS1300(ab)
41.0	3	19.7	71	3.8	3	RCS1366(a-)	RCS1167(b-)	RCS1022(ab)
41.0	3	19.7	71	3.8	3	RCS1167(b-)	RCS3665(a-)	RCS1022(ab)
41.0	3	19.7	71	3.8	3	RCS1167(b-)	RCS2741(a-)	RCS1022(ab)
41.0	3	19.7	71	3.8	3	RCS1167(b-)	RCS1022(ab)	RCS0937(-)
38.2	5	15.8	69	3.6	3	RCS2958(b-)	RCS0914(b-)	RCS3315(ab)
38.2	5	15.8	69	3.6	3	RCS1366(a-)	RCS0914(b-)	RCS3719(a-)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS2958(b-)	RCS1113(bc)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS1113(bc)	RCS1541(a-)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS1113(bc)	RCS1167(b-)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS1113(bc)	RCS0235(-)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS1113(bc)	RCS3315(ab)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS1113(bc)	RCS2689(-)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS2645(ab)	RCS0235(-)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS1167(b-)	RCS1022(ab)
35.8	3	19.0	71	3.8	3	RCS1325(cd)	RCS0235(-)	RCS1022(ab)
35.8	3	19.0	71	3.8	3	RCS2958(b-)	RCS1167(b-)	RCS1022(ab)
35.8	3	19.0	71	3.8	3	RCS2958(b-)	RCS0235(-)	RCS1022(ab)
35.8	3	19.0	71	3.8	3	RCS1113(bc)	RCS2645(ab)	RCS1167(b-)
35.8	3	19.0	71	3.8	3	RCS1113(bc)	RCS1167(b-)	RCS1022(ab)
35.8	3	19.0	71	3.8	3	RCS2645(ab)	RCS1167(b-)	RCS0235(-)
35.8	3	19.0	71	3.8	3	RCS1167(b-)	RCS0235(-)	RCS1022(ab)

[a] Samples harbouring the locus (allele) combination in the right-most column.
[b] Mean value of phenotypes (flowering date).

and RCS1300(ab) were 7.9×10^{-1}, 3.9×10^{0}, and 3.7×10^{-4}, respectively. These low values indicate that there is almost no, or only a weak, effect of each single locus alone. Similar to the single locus results, combinations of two loci did not strongly affect the phenotype. However, the triple loci combinations are highly associated with the phenotype.

Several loci are repeatedly listed in Table 1, which indicates that they interact with multiple combinations of loci. For example, RCS1022(ab) and RCS1167(b-) were each found in five combinations in Table 1. The redundancy of the list presentation does not facilitate an intuitive understanding of the results, and the obtained information was therefore also visualized in two styles of graphical presentation, as shown in Figs 5 and 6. Fig. 5 illustrates the combinations of interacting triple loci and their positional information on the red clover linkage map.[17] Locations of interacting loci are interlinked by lines on the linkage group maps. By illustrating the positions of interacting loci, the loci which affect the phenotype through interactions with multiple combinations of genotypes are readily identified; in this case, 50–52 cM and 84–89 cM on LG2, 76 cM on LG5, and 27 cM on LG6. In a different type of graphic, Fig. 6 illustrates the relationships between interacting loci and allele type using GMs and a GMN. Line colors (GMNs) and cell colors (GMs) represent the magnitude of *F*-measure values for interacting and single locus/allele effects, respectively. By comparing several GMNs, the existence of 'hub loci/alleles' in these interactions emerges, such as RCS1022(ab) and RCS1167(b-), as well as multiple GMNs. The hub loci/alleles, however, did not show the highest effect on phenotypes when evaluated as a single locus/allele.

4. Discussion

Although there is a growing awareness of the importance of gene interactions in genetic studies of complex traits, classical genetic analysis either ignores gene interactions or defines the effect of gene interactions as a

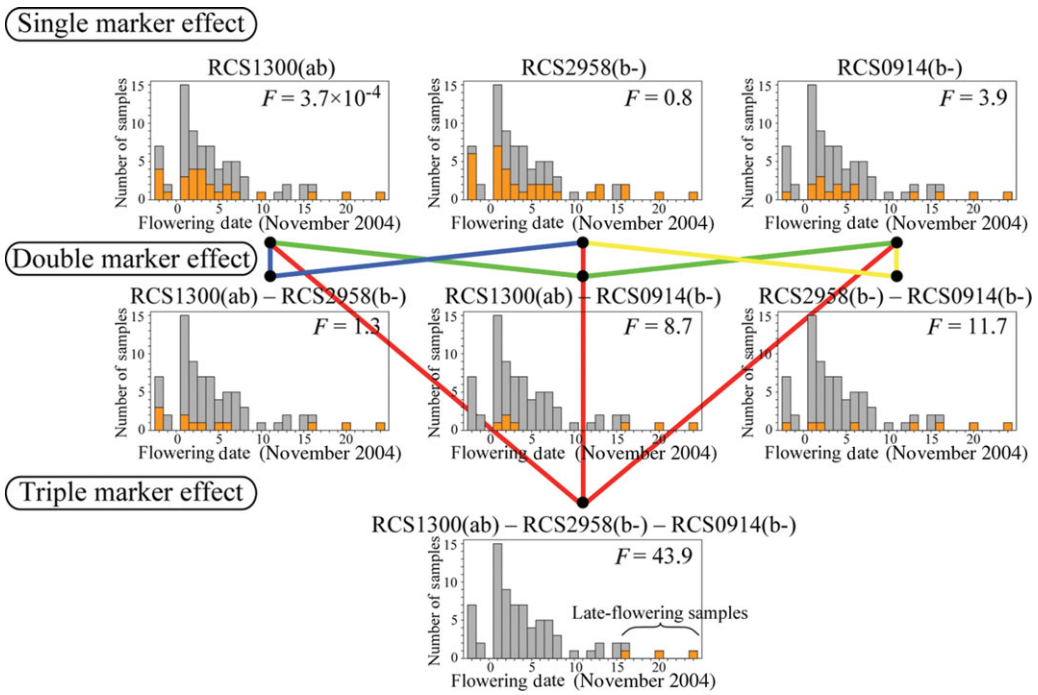

Figure 4. Distribution of phenotype values (flowering date) resulting from multiple locus/allele combinations. Combinations of two out of three locus/allele types are shown by green, blue, and yellow lines, and that of three types by red lines. Orange and gray bars indicate samples possessing the relevant locus/allele combinations and others, respectively (see Table 1).

deviation from genetic additive effects.[22] The demonstration in this study, using a real data set from red clover, indicates that the GMM algorithm efficiently detects multi-QTL interactions in genetic variation in a complex phenotype. The novel aspect of GMM is that the algorithm is capable of comparing multiple QTL interactions at once, which has not previously been possible in QTL detection approaches.

One of the advantages of comparison of multiple QTL interactions is the detection of hub locus/allele in the interactions. In this study, GMM demonstrated the existence of loci which affected the phenotype through multiple combinations of interacting loci. Interestingly, these hub loci/alleles did not necessarily display the highest effect on the phenotype as a single locus. To date, most findings have suggested that quantitative variation is determined by a few QTLs with a relatively large effect and a large number of genes having progressively smaller effects.[2] However, Jannink[23] recently identified QTLs by analyzing genetic background interactions in association studies, and was able to detect loci that have no main effect but which influence a trait only through their interactions with other loci. Our results, together with those of Jannink, suggest that multiple QTL interactions might be buried under the smaller effect of single QTLs. The hub locus/allele may be a key locus for identifying QTLs networks in which each component contributes directly to the final phenotype.

The hub loci/alleles that appeared in the multiple interactions that affected the flowering date of red clover were identified on 50–52 cM and 84–89 cM on LG2, 76 cM on LG5 and 27 cM on LG6. Herrmann et al.[24] have detected seven QTLs for the flowering date of red clover on all of the linkage groups except for on LG1. Although the tested DNA markers and the total lengths of the reference maps were different between Herrmann et al.'s (444 cM) and our study (869 cM), it could be presumed that the hub loci/alleles found on LG2 and LG5 will correspond to the QTLs identified by Herrmann et al., once a comparison of the two maps used in these studies has been performed. In this study, we used a set of unrelated individuals to demonstrate the GMM algorithm; however, any type of population, including family data, which is used for interval mapping analysis, can be analyzed by the GMM algorithm. Therefore, using GMM for reassessment of QTLs that have been identified by other methods (e.g., interval mapping) might provide us with insights into QTL interactions in the analyzed population. It is expected that the comparison of multiple combinations of interacting QTLs will lead to the identification of genes related to complex traits. Recently, expression quantitative trait locus (eQTL) mapping, which is a combination of gene expression profiling and classical genetic mapping, has been adopted to reveal quantitative heritable variation in the transcriptome.[25] In this method, however, thousands of expression profiles are related with sequence polymorphisms across the genome through their correlated variations, which results in a large number of mappings that make it difficult to consider simultaneously the relationships

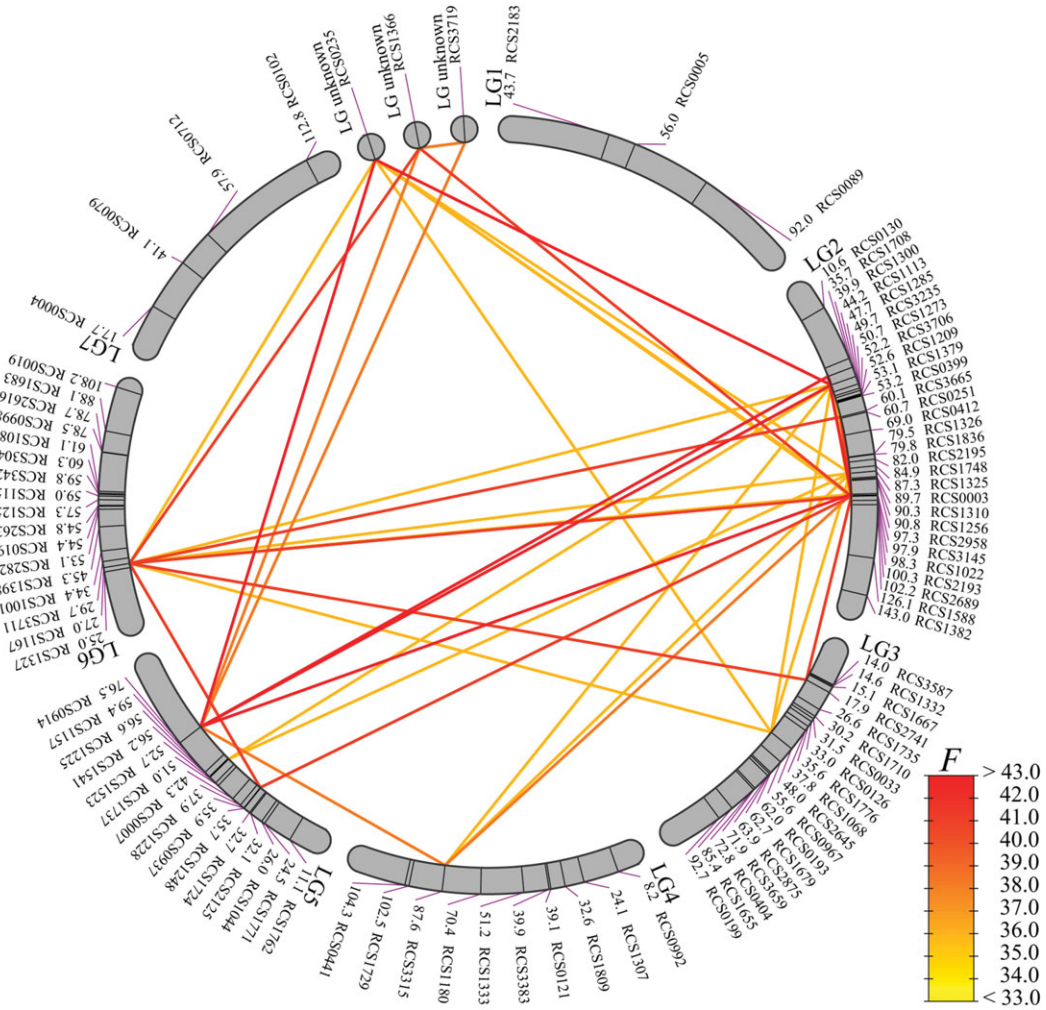

Figure 5. Graphical presentation of interacting triple loci and their positions on the genetic linkage map. Seven linkage groups and unmapped markers are arranged tandemly as a circle. Triangles in the circle indicate GMNs, i.e. interacted triple loci combination. Magnitude of the *F*-measure is show by a color-code.

between multiple genomic regions and multiple expression profiles.[26] Prioritization and re-organization of a large number of original results by the GMM method may facilitate efficient dissection of expression QTLs and subsequent identification of genes.

Though the concept of GMM and its potential feasibility were demonstrated with a real genotype–phenotype data set in this article, there are several issues yet to be evaluated, examined, and solved before GMM can be applied more precisely for the detection of QTL interactions. The most prioritized issue to be discussed is the population structure. We used a data set of unrelated individuals without population structure in the present demonstration. However, QTL detection by GMM is performed on the basis of analysis of variance; thus, it is possible that hidden population structure leads to false positive associations between genotypes and phenotypes, as was observed in LD mapping.[27] Therefore, estimation of the degree of population structure of a data set prior to GMM calculation should be an essential requirement. In order to address the problem of population structure in population-based samples, Yu et al.[28] have developed a new statistical method for LD mapping that simultaneously accounts for both population structure and familial relatedness. A combination of such method with GMM might provide a solution to the problem of the current version of GMM.

Effectual data size and degree of significance are also indispensable issues to be evaluated. One of the most probable reasons for insufficient output from the GMM analysis is the limited size of the data set. Increasing the number of individuals for data acquisition would improve the accuracy of the whole analysis, but this is not always feasible, especially in the early stages of data preparation. On the other hand, there is a strong demand to sort locus combinations according to their significance so that the loci that are expected to interact together can be prioritized. Therefore, searching for probable locus combinations will be useful from a practical point of view. In LD mapping, for example, many

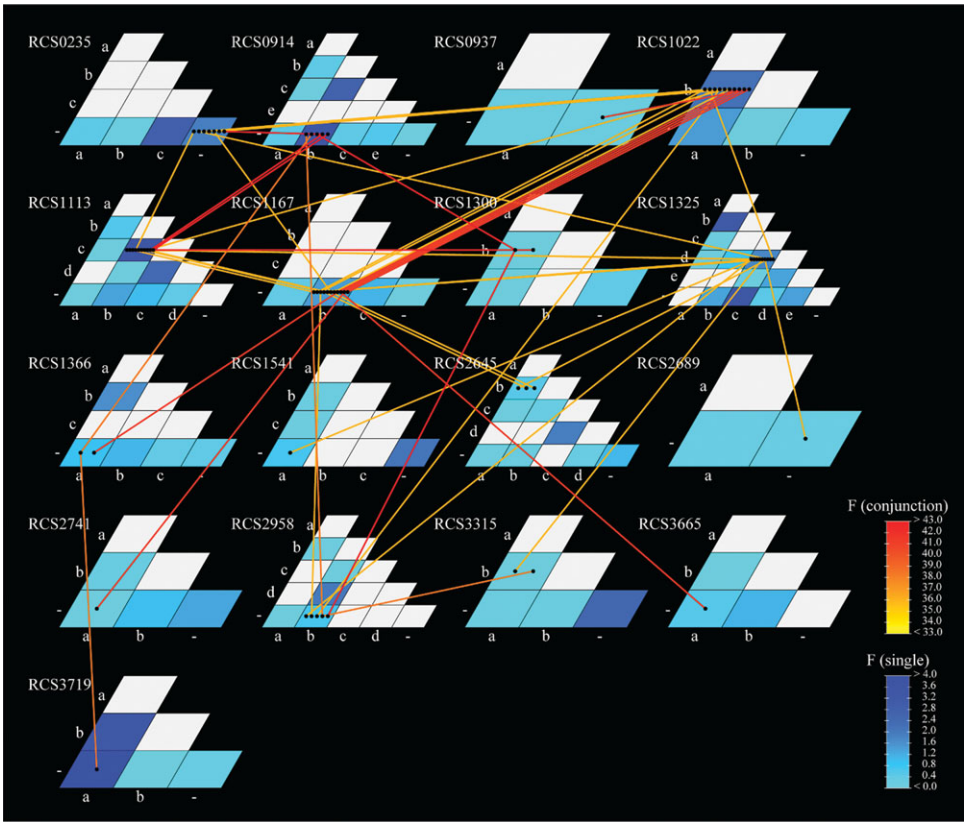

Figure 6. Graphical presentation of interacting loci and allele type by GMs and a GMN. Significant locus/allele combinations of one to three interacting loci whose *F*-measures were >35.0 are shown by GMs and GMN. Matrices and connecting lines indicate GMs and GMNs, respectively. Magnitude of the *F*-measure of combination and single locus/alleles effects, respectively, is shown by a color-code.

simulation studies have been carried out to obtain the optimal experiment design to detect QTLs, by considering a wide range of parameters, such as population size, number of markers, heritability and effects of QTLs, length of LDs, and the number of generations.[29–32] As with LD mapping, it is probable that the appropriate size of a data set for GMM will vary depending on the experimental design. Moreover, if the number of alleles of each locus increases, the risk rate of detecting false-positive interactions could also increase. To solve such problems, we need to evaluate the detection power of GMM in relation to the number of possible combinations of individuals and their genotypes. Both simulation studies and the application of the GMM method to real data sets will be useful to estimate the effectual data size. Additionally, the robustness of the significance test should also be investigated in the near future.

For easy handling and presentation of the results for users, it will be fundamental to design an appropriate interface for the GMM software program. Because of the large number of, and sometimes redundant, candidate QTL combinations detected by the GMM analysis, extraction of meaningful information from a tabulated set of data will often be difficult. In this study, we presented two types of graphic presentations for the results from the GMM analysis: a circular linkage map (Fig. 5) and a combination GM and GMN chart (Fig. 6). Development of a user interface that allows intuitive understanding of the relationships among multiple QTL interactions is under way.

Though GMM currently has several unclear issues which have to be resolved, it should give us an additional dimension of QTL impacts by indication of multiple QTL interactions. We hope the new ideas for dissection of QTL expression will be sparked on genetics and genomics by handling GMM. The GMM service and software package will be provided at http://www.kazusa.or.jp/GMM in December, 2007.

Acknowledgement: We thank S. Sasamoto for excellent technical assistance.

Funding

This work was supported by the "Development of DNA Marker-aided Selection Technology for Plants and Animals (DM-1604)" Project, Ministry of Agriculture, Forestry and Fisheries, Japan, and the Kazusa DNA Research Foundation.

References

1. Yu, J. and Buckler, E. S. 2006, Genetic association mapping and genome organization of maize, *Curr. Opin. Biotech.*, **17**, 155–160.
2. Farrall, M. 2004, Quantitative genetic variation: a post-modern view, *Hum. Mol. Genet.*, **13**, R1–R7.
3. Majumder, P. P. and Ghosh, S. 2005, Mapping quantitative trait loci in humans: achievements and limitations, *J. Clin. Invest.*, **115**, 1419–1424.
4. Flaherty, L., Herron, B. and Symula, D. 2005, Genomics of the future: identification of quantitative trait loci in the mouse. *Genome Res.*, **15**, 1741–1745.
5. Yano, M., Katayose, Y., Ashikari, M. et al. 2000, *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*, *Plant Cell*, **12**, 2473–2484.
6. Morton, N. E. 2005, Linkage disequilibrium maps and association mapping, *J. Clin. Invest.*, **115**, 1425–1430.
7. Zeng, Z. B. 1993, Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci, *Proc. Natl. Acad. Sci. USA*, **90**, 10972–10976.
8. Jansen, R. C. and Stam, P. 1994, High resolution of quantitative trait into multiple loci via interval mapping, *Genetics*, **136**, 1447–1445.
9. Zak, M., Baierl, A., Bogdan, M. and Futschik, A. 2007, Locating multiple interacting quantitative trait loci using rank-based model selection, *Genetics*, **176**, 1845–1854.
10. Heidema, A. G., Boer, J. M. A., Nagelkerke, N. et al. 2006, The Challenge for genetic epidemiologist: how to analyze large number of SNPs in relation to complex diseases, *BMC Genetics*, **7**, 23.
11. Kao, C. H. and Zeng, Z. B. 2002, Modeling epistasis of quantitative trait loci using Cckerham's model, *Genetics*, **160**, 1243–1261.
12. Bogdan, M., Ghosh, J. K. and Doerge, R. W. 2004, Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci, *Genetics*, **167**, 989–999.
13. Narita, A. and Sasaki, Y. 2004, Detection of multiple QTL with epistatic effects under a mixed inheritance model in an outbred population, *Genet. Sel. Evol.*, **36**, 415–433.
14. Ritchie, M. D., Hahn, L. W., Roodi, N. et al. 2001, Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer, *Am. J. Hum. Genet.*, **69**, 138–147.
15. Coffey, C. S., Hebert, P. R., Ritchie, M. D. et al. 2004, An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene–gene interactions on risk of myocardial infarction: the importance of model validation, *BMC Bioinformatics*, **5**, 49.
16. Cho, Y. M., Ritchie, M. D., Moore, J. H. et al. 2004, Multifactor-dimensionality reduction shows a two locus interaction associated with Type 2 diabetes mellitus, *Diabetologia*, **47**, 549–554.
17. Sato, S., Isobe, S., Asamizu, E. et al. 2005, Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.), *DNA Res.*, **12**, 301–364.
18. Kölliker, R., Herrmann, D., Boller, B. and Widmer, F. 2003, Swiss Mattenklee landraces, a distinct and diverse genetic resource of red clover (*Trifolium pratense* L.), *Theor. Appl. Genet.*, **107**, 306–315.
19. Pritchard, J. K., Stephens, M. and Donnelly, P. 2000, Inference of Population Structure using multilocus genotype data, *Genetics*, **155**, 945–959.
20. Nakaya, A., Hishigaki, H. and Morishita, S. 1999, Tracing synergetic behavior of the QTLs affecting oral glucose tolerance in the OLETF rat, *Genome Informatics*, **10**, 155–165.
21. Sese, J., Kurokawa, Y., Monden, M. et al. 2004, Constrained clusters of gene expression profiles with pathological features, *Bioinformatics*, **20**, 3137–3145.
22. Zhao, J., Jin, L. and Xiong, M. 2006, Test for interaction between two unlinked loci, *Am. J. Hum. Genet.*, **79**, 831–845.
23. Jannink, J. 2007, Identifying quantitative trait locus by genetic background interactions in association studies, *Genetics*, **176**, 553–561.
24. Herrmann, D., Boller, B., Studer, B. et al. 2006, QTL analysis of seed yield components in red clover (*Trifolium pratense* L.), *Theor. Appl. Genet.*, **112**, 536–545.
25. Schadt, E. E., Monks, S. A., Drake, T. A. et al. 2003, Genetics of gene expression surveyed in maize, mouse and man, *Nature*, **422**, 297–302.
26. Zou, W., Aylor, D. L. and Zeng, Z. B. 2007, eQTL Viewer: visualizing how sequence variation affects genome-wide transcription, *BMC Bioinformatics*, **8**, 7–11.
27. Lander, E. S. and Schork, N. J. 1994, Genetic dissection of complex traits, *Science*, **265**, 2037–2048.
28. Yu, J., Pressoir, G., Briggs, W. H. et al. 2006, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, *Nat. Genet.*, **38**, 203–208.
29. Terwilliger, J. D., Haghighi, F., Hiekkalinna, T. S. et al. 2002, A bias-ed assessment of the use of SNPs in human complex traits, *Curr. Opin. Genet. Develop.*, **12**, 726–734.
30. Zhang, K. and Sun, F. 2005, Assessing the power of tag SNPs in the mapping of quantitative trait loci (QTL) with extremal and random samples, *BMC Genetics*, **6**, 51.
31. Ball, R. D. 2005, Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies, *Genetics*, **170**, 859–873.
32. Zhao, H. H., Fernando, R. L. and Dekkers, J. C. M. 2007, Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci, *Genetics*, **175**, 1975–1986.