3D graphical visualization of the genetic architectures underlying complex traits in multiple environments[*]

HU Cheng-cheng[†1], YE Xiu-zi[†‡1], ZHANG Yin[1], YU Rong-dong[2], YANG Jian[3], ZHU Jun[3]

(*1College of Computer Science, Zhejiang University, Hangzhou 310027, China*)
(*2James D. Watson Institute of Genomic Science, Zhejiang University, Hangzhou 310029, China*)
(*3Institute of Bioinformatics, Zhejiang University, Hangzhou 310029, China*)
[†]E-mail: cechu@ucdavis.edu; yxz@zju.edu.cn

Abstract: An approach for generating interactive 3D graphical visualization of the genetic architectures of complex traits in multiple environments is described. 3D graphical visualization is utilized for making improvements on traditional plots in quantitative trait locus (QTL) mapping analysis. Interactive 3D graphical visualization for abstract expression of QTL, epistasis and their environmental interactions for experimental populations was developed in framework of user-friendly software QTLNetwork (http://ibi.zju.edu.cn/software/qtlnetwork). Novel definition of graphical meta system and computation of virtual coordinates are used to achieve explicit but meaningful visualization. Interactive 3D graphical visualization for QTL analysis provides geneticists and breeders a powerful and easy-to-use tool to analyze and publish their research results.

Key words: Quantitative trait locus (QTL), Visualization, 3D graphical, Interactive visualization
doi:10.1631/jzus.2007.A0563   **Document code:** A   **CLC number:** TP391.41

INTRODUCTION

Understanding the genetic architecture for a complex trait has great impact on evolution, agriculture and human health. Quantitative trait locus (QTL) is a putative genetic locus identified through the statistical analysis of complex traits, such as plant height or body weight (Doerge, 2002). The complexity of phenotypic traits, particularly of those involved in adaptation, probably arises from segregation of alleles at many interacting loci, the effects of which are sensitive to the environments (Mauricio, 2001). In the past 15 years, many systematic statistical methods have been well established to scan genome for multiple loci (Lander and Botstein, 1989; Haley and Knott, 1992; Zeng, 1994), to detect the intergenic allele interaction (epistasis) (Kao *et al*., 1999; Wang *et al*., 1999) as well as the QTL-by-environment interaction (Wang *et al*., 1999; Piepho and Gauch, 2001). However, the computational results obtained by those QTL mapping methods are incredibly complicated, and researchers have gradually become tired of such large scale output in text format. Graphical visualization is urgently needed to make an intuitive view of these QTL mapping results for further study. Compared with the fast development in statistical models, the pace of graphical visualization for QTL mapping results seems much slower. Expertise and automatic software visualization tools for QTL analysis are highly desired (Chesler *et al*., 2005).

Among several free computer programs to generate graphical visualization for QTL mapping, the WinQTLCart (Wang *et al*., 2005) is one of the most popular ones, particularly for its powerful graphical tool for presenting and summarizing mapping results. However, its capability of allowing user interactions

is rather limited, and it does not support visualizing QTL and epistasis of complex traits in multiple environments. Another influential software tool for QTL visualization is Expressionview (Fischer *et al.*, 2003). The graphics part of this tool mainly focuses on combined visualization of gene-expression data and QTL, as well as the visualization of gene clusters and the selection of functional candidate genes in the context of research on complex traits. However, its graphical expression is monotone, and inadequate to emphasize the QTL. R/qtl (Broman *et al.*, 2003) is also famous for its excellent 2D visualization of LOD scores for 2D genome scan, but it lacks the expressional visualization and the user interface is poor. Yang and Zhu (2005) and Yang *et al.*(2006) made a simple visualization of epistasis and QTL×environment (QE) interaction effects. But the 2D graphics fails to reveal an explicit picture to QTL researchers, because of mass crowded objects in scene.

Up to date, no satisfactory work has been done in QTL visualization to fully meet the demands from the research community. Furthermore, no previous work has been attempted in 3D visualization of QTL and epistasis of complex traits in multiple environments.

In this paper, we present our recent work in 3D visualization of QTL and epistasis of complex traits in multiple environments. Our graphical visualization is based on the data resulted from the mixed linear model based composite interval mapping (MCIM) method (Zhu, 1999; Wang *et al.*, 1999; Yang and Zhu, 2005; Yang *et al.*, 2006), which is currently the unique method for simultaneously mapping the epistasis and QE interaction by data from multiple-environment trials (METs). We will describe our research work in two parts. In the first part, we visualize the test statistics (*F* values) obtained from marker interval analysis and genome scanning for QTL and epistasis by introducing 3D graph in the plots. In the second part, we define, compute and render the abstract graphical expression for QTL and epistasis of complex traits in multiple environments.

PLOTTING *F* TEST STATISTICS OF MARKER INTERVAL AND QTL ANALYSIS

Visualizing the test statistics of marker interval analysis

The marker interval analysis is aimed to select

the candidate intervals and interval/interval interaction which will be used to control the background genetic variation during the procedure of scanning genome for QTL and epistasis. Our approach is to plot a whole view graphics describing the *F* values of the intervals in all chromosomes with each chromosome having a detailed sub-picture. The genetic distance of each marker in chromosomes is shown in non-overlapping horizontal labels. Users can easily obtain explicit numeric values in views (Fig.1). In Fig.2a, the colorful matrix represents the *F* value between each pair of intervals, and the color scalar bar provides continuous mapping from color to *F* value. The black separate lines delimit the ranges of chromosomes. Using color to reflect the *F* value is popular but not intuitive enough, because users have to transform the color to its corresponding numeric value in mind. Thus, we introduce 3D molecular marker interaction plot in our visualization. By taking *F* value as the height value, we can construct a 3D visualization similar to terrain visualization, giving the users more direct information (Fig.2b).
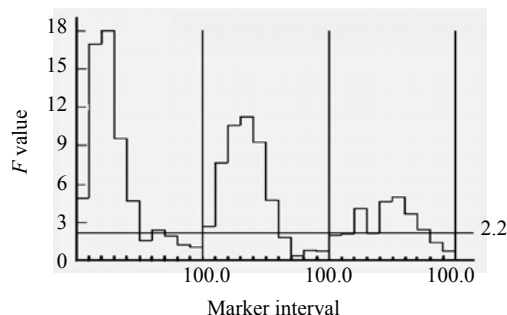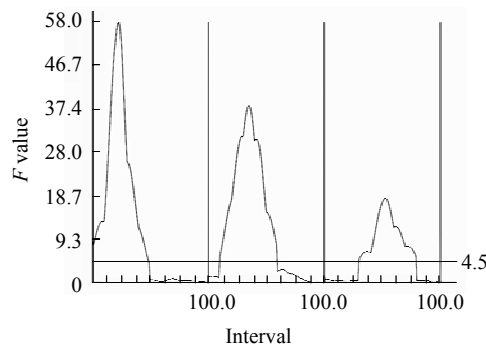

Fig.1 *F* statistical profile of marker interval

Visualization of the test statistics of genome scan for QTL and epistasis

After the candidate marker intervals and interval interactions are selected, the 1D genome scan is performed to map the main-effect QTL (Fig.3), and partial 2D genome scan is performed to detect epistasis (Fig.4a). In the 1D genome scan procedure, *F* test statistics values are calculated step by step separated by 1 cm along the whole genome. The black curve in Fig.3 describes the genome scanning results in chromosomes with each chromosome giving a detailed view. The 2D genome scan procedure is only conducted in the chromosome regions near the in-

tervals involved in interactions with other intervals. Since the testing regions are trivial in the whole genome, other parts are omitted as small blanks (Fig.4a). To calculate the position of a test point on the horizontal axis, let P be the distance of the test point in the nth testing region

$$P = \sum_{i=1}^{n-1} c_i + \varepsilon_0 + \varepsilon(n-1) + d,$$

where c_i is the length of the ith testing region; ε is the length of the omitted blank; $\varepsilon_0=0$ only if the testing region starts at the chromosome origin, otherwise $\varepsilon_0=\varepsilon$, d is the position of the test point in its testing region.

Fig.3 F statistical profile of 1D genome scan for putative QTL

A 2D graph analogue to that of visualizing the F values in interval/interval interaction selection is constructed according to the F statistical profile generated by the partial 2D genome scan (Fig.4a). The F statistical profile of partial 2D genome scan is also displayed in a 3D view in order to give more impressive images (Fig.4b). For the 6 graphs in Figs.1~4 described above, we provide better methods in accommodating scales and detecting the accurate coordinate values of a point in a scene. Users need only to use a mouse to interact with the software tools.

GRAPHICAL PRESENTATION OF THE GENETIC ARCHITECTURE

In order to get terse and explicit presentation of the genetic architecture underlying a complex trait, a graphical meta system is defined based on the complexity of the genetic architecture. We first predigest some chromosome parts in the main meta system containing neither QTL nor epistasis, and restore them in detailed pictures. To achieve quality rendering, virtual spatial coordinates are also computed.

The graphical meta system

A meaningful and concise graphical meta system is the base for excellent visualization of abstract expressions. As human being is very sensitive to shape and color, we choose them as the basis to construct our graphical meta system, as listed in Table 1.

Predigesting chromosomes

When the resultant data contains many chromosomes, in some cases close to 50 chromosomes in one view, depicting all these chromosomes in their full sizes may lead to crowded and overlapped graphics objects. Thus, an appropriate simplification is necessary here. In the following two situations, chromosomes are predigested or omitted. First, if no QTL (including epistatic QTL) is prognosticated in a chromosome, this chromosome will be totally omitted in the view. Second, for chromosomes with QTL, we clip those chromosome regions where no QTL is located and no overlapping with the confidence ranges of adjacent QTLs exists.

Generate virtual spatial coordinates

In order to gain an appropriate compromise between overview and detail, we use a virtual spatial coordinate to replace the original one designated. We compute the virtual position of a point Q in an interval on chromosome a by the pseudo-code listed below:

```
1: a'=Convert(a);
2: Current_Pos=Chromosome_Head(a');
3: do
4:     Interval=Search_Next_QTL();
5:     Molecular_Interval=Calculate_MI_Length(a', Interval);
6:     Confidence_Interval=Calculate_CI_Region(a', Interval);
7:     if Confidence_Interval>=Molecular_Interval
8:         Current_Pos+=Omitted_Line+Confidence_Interval;
9:     else Current_Pos+=Omitted_Line+Molecular_Interval;
10:    end if
11: while !Reach(Q)
12: Final_Pos=Currrent_Pos+Relative_Pos(Q);
```
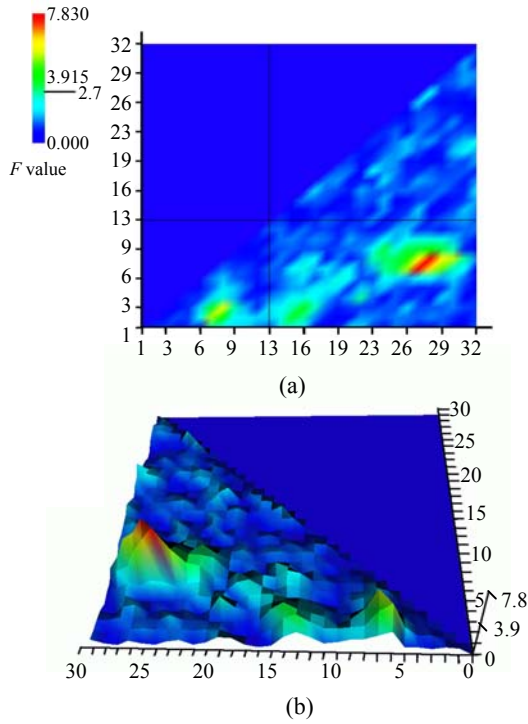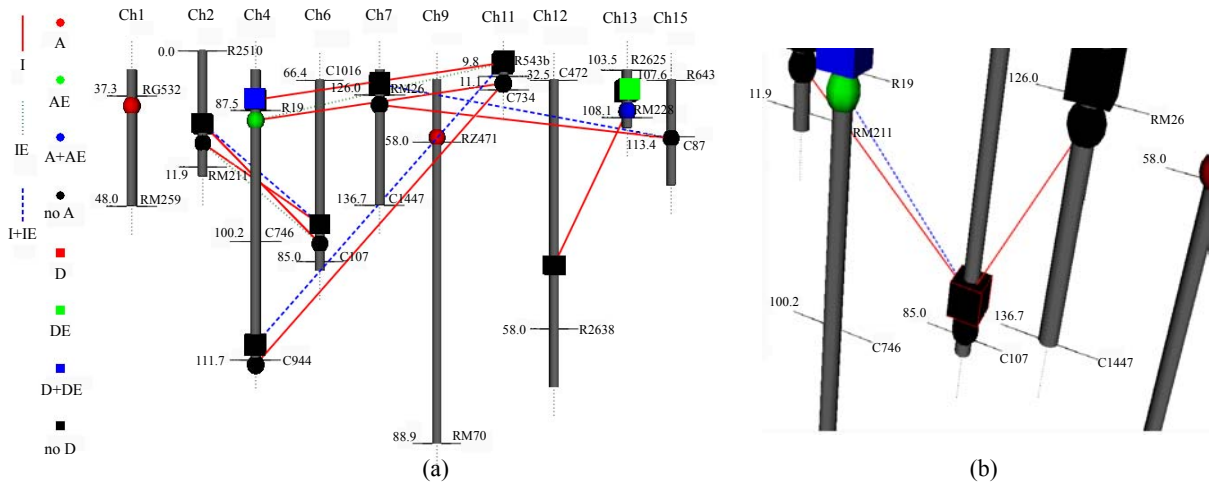
Fig.2 *F* statistical profile of marker interval analysis. The marker interval is given in Fig.1. (a) 2D Marker interval interaction; (b) 3D Marker interval interaction
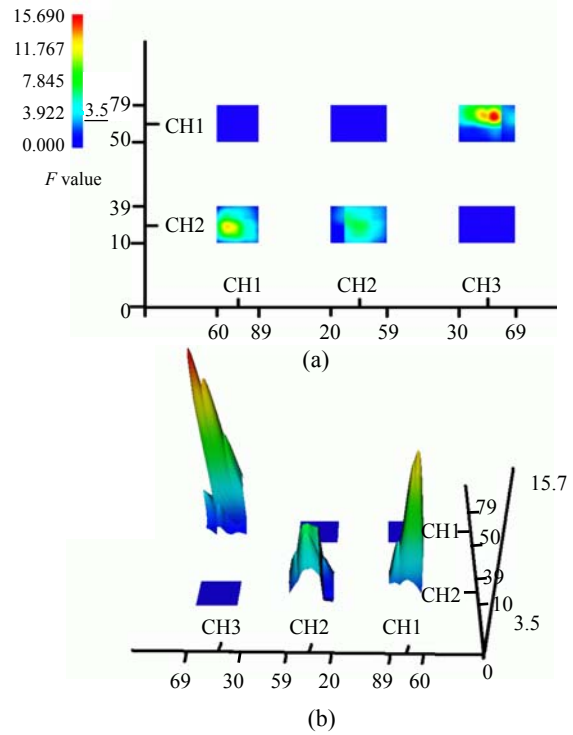
Fig.4 *F* statistical profile of (a) 2D genome scan for epistasis and (b) 3D genome scan for epistasis

Fig.5 Example of visualizing the complex genetic architecture obtained by real experimental data from METs. (a) Full view of the detected genetic architecture; (b) Interactive view showing the interactions of a focal QTL with other QTLs

Table 1 Definition of the graphical meta system for genetic architecture presentation

Graphical meta system	Line (Epistasis)	Shape (QTL)	
		Circle	Square
Red	— With only epistatic main effect (I)	● With only additive effect (A)	■ With only dominance effect (D)
Green	····· With only epistasis×environment interaction effect (IE)	● With only additive×environment interaction effect (AE)	■ With only dominance×environment interaction effect (DE)
Blue	– – With both I and IE	● With both A and AE	■ With both D and DE
Dark	— Not available	● With no additive related effect	■ With no dominance related effect

Convert(*a*) in Step 1 is to change real chromosome into displaying chromosome. *Search_Next_QTL*() in Step 4 is to find the nearest interval containing QTL or epistasis. And *Relative_Pos*(*Q*) in the last step denotes the inside distance of point *Q* in its local interval.

However, the accurate position in the display chromosome is still required for labels in the interactive view. That is to say, by picking QTL in the view, the real location of this QTL should be popped up in a balloon. This can be achieved by keeping the mapping between the virtual coordinate and real coordinate designated.

CONCLUSION

In our visualization tool, we have improved the existing graphical visualization of QTL mapping results and provided a better user interaction method. We define and render the abstract graphics expression for QTL and epistasis of complex traits in multiple environments. With our graphical meta system and virtual spatial coordinate, we present the users with an explicit and meaningful view with all necessary information. Users can simply cut the pictures and paste them on their papers to illustrate their research results. This 3D view also offers the researchers an easy and comfortable interactive approach to focus on every detail in the scene, i.e. users can conveniently use view operations such as rotation, scale, translation to locate an accurate point in view.

Based on the method proposed in the paper, we developed a user-friendly software tool using MFC (Microsoft Foundation Class) and VTK (Visualization Toolkit), an object-oriented open source graphics library. Our software tool runs on Microsoft Windows 9.x/NT/2000/XP/2003 server, and is freely available for downloading from http://ibi.zju.edu.cn/software/qtlnetwork. We have applied our software to analyze some real experimental datasets. Fig.5 in the preceding page illustrates one such example.

References

Broman, K.W., Wu, H., Sen, S., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**(7):889-890. [doi:10.1093/bioinformatics/btg112]

Chesler, E.J., Lu, L., Shou, S., Qu, Y.H., Gu, J., Wang, J.T., Hsu, H.C., Mountz, J.D., Baldwin, N.E., Langston, M.A., *et al*., 2005. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.*, **37**(3):233-242. [doi:10.1038/ng1518]

Doerge, R.W., 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.*, **3**(1): 43-52. [doi:10.1038/nrg703]

Fischer, G., Ibrahim, S.M., Brockmann, G.A., Pahnke, J., Bartocci, E., Thiesen, H.J., Serrano-Fernández, P., Möller, S., 2003. Expressionview: visualization of quantitative trait loci and gene-expression data in ensembl. *Genome Biology*, **4**(11):R77. [doi:10.1186/gb-2003-4-11-r77]

Haley, C.S., 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., Zeng, Z.B., Teasdale, R.D., 1999. Multiple interval mapping for quantitative trait loci. *Genetics*, **152**:1203-1216.

Lander, E.S., Botstein, D., 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**:185-199.

Mauricio, R., 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nat. Rev. Genet.*, **2**(5):370-381. [doi:10.1038/35072085]

Piepho, H.P., Gauch, H.G., 2001. Marker pair selection for mapping quantitative trait loci. *Genetics*, **157**:433-444.

Wang, D.L., Zhu, J., Li, Z.K., Paterson, A.H., 1999. Mapping QTLs with epistatic effects and QTL environment interactions by mixed linear model approaches. *Theor. Appl. Genet.*, **99**:1255-1264. [doi:10.1007/s001220051331]

Wang, S.C., Basten, J., Zeng, Z.B., 2005. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. Http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

Yang, J., Zhu, J., 2005. Methods for predicting superior genotypes under multiple environments based on QTL effects. *Theor. Appl. Genet.*, **110**(7):1268-1274. [doi:10.1007/s00122-005-1963-2]

Yang, J., Zhu, J., Williams, R., 2006. Mapping genetic architecture of complex trait in experimental populations. *Bioinformatics*, under revision.

Zeng, Z.B., 1994. Precision mapping of quantitative trait loci. *Genetics*, **136**:1457-1468.

Zhu, J., 1999. Mixed model approaches of mapping genes for complex quantitative traits. *J. Zhejiang Univ. (Natural Science)*, **33**:327-335 (in Chinese).