

HOSTED BY



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect



QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations

Lei Meng¹, Huihui Li¹, Luyan Zhang¹, Jiankang Wang^{*,1}

The National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, and CIMMYT China Office, Chinese Academy of Agricultural Sciences, Beijing 100081, China

ARTICLE INFO

Article history:

Received 20 December 2014

Received in revised form

20 January 2015

Accepted 16 February 2015

Available online 23 February 2015

Keywords:

Biparental populations

Map construction

QTL mapping

Software

ABSTRACT

QTL IciMapping is freely available public software capable of building high-density linkage maps and mapping quantitative trait loci (QTL) in biparental populations. Eight functionalities are integrated in this software package: (1) BIN: binning of redundant markers; (2) MAP: construction of linkage maps in biparental populations; (3) CMP: consensus map construction from multiple linkage maps sharing common markers; (4) SDL: mapping of segregation distortion loci; (5) BIP: mapping of additive, dominant, and digenic epistasis genes; (6) MET: QTL-by-environment interaction analysis; (7) CSL: mapping of additive and digenic epistasis genes with chromosome segment substitution lines; and (8) NAM: QTL mapping in NAM populations. Input files can be arranged in plain text, MS Excel 2003, or MS Excel 2007 formats. Output files have the same prefix name as the input but with different extensions. As examples, there are two output files in BIN, one for summarizing the identified bin groups and deleted markers in each bin, and the other for using the MAP functionality. Eight output files are generated by MAP, including summary of the completed linkage maps, Mendelian ratio test of individual markers, estimates of recombination frequencies, LOD scores, and genetic distances, and the input files for using the BIP, SDL, and MET functionalities. More than 30 output files are generated by BIP, including results at all scanning positions, identified QTL, permutation tests, and detection powers for up to six mapping methods. Three supplementary tools have also been developed to display completed genetic linkage maps, to estimate recombination frequency between two loci, and to perform analysis of variance for multi-environmental trials.

© 2015 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Genetic studies of QTL (mapping quantitative trait loci) mapping based on fine-scale linkage maps have greatly increased our

understanding of the inheritance of quantitative traits in the last 20 years [1–4]. Information identified by QTL mapping is important for fine gene mapping, map-based cloning, and efficient use of gene information in molecular breeding [5–7]. Linkage

* Corresponding author.

E-mail addresses: wangjiankang@caas.cn, jkwang@cgiar.org (J. Wang).

Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

¹ The authors made equal contributions to this work.

analysis and map construction in plants were first performed in genetically segregating populations derived from two inbred parental lines, such as F_2 , backcross, doubled haploid, and recombinant inbred line populations. Key to linkage map construction is the accurate estimation of recombination frequency, which has been long studied in genetics for various populations [4,8]. Sun et al. [9] investigated the estimation efficiency of recombination frequency in 12 biparental populations and concluded that larger population size and smaller recombination frequency resulted in higher LOD score and smaller deviation. Advanced backcrossing and selfing populations yielded lower precision in estimating recombination frequency.

Based on constructed linkage maps, several statistical methods have been developed for QTL detection and effect estimation. Composite interval mapping (CIM) [10] represents one of the most commonly used methods, but the algorithm used in CIM cannot ensure complete background control [3,4,11]. Inclusive composite interval mapping (ICIM) was accordingly developed and proved to be more efficient for background control via a two-step mapping strategy [3,11,12]. In the first step of ICIM, stepwise regression is applied to identify most-significant regression variables. In the second step, interval mapping is performed using phenotypes adjusted by the markers identified in the first step. ICIM retains all advantages of CIM over the simple interval mapping and avoids the possible increase of sampling variance and the complicated background marker selection process in CIM [13,14]. The method has been extended to mapping additive and dominant QTL [12,15], epistatic QTL [16,17], and QTL-by-environment interactions [18].

More recently, populations consisting of introgression lines have been used for fine gene mapping and map-based cloning [6,19]. Owing to high selection intensity during the process of population development, gene and marker frequencies do not follow ratios expected in standard biparental populations. A likelihood ratio test based on stepwise regression has been proposed for these special populations [5,19]. A NAM population is derived from a multiple-cross mating design sharing one common parent. Li et al. [20] extended ICIM to this population, calling it joint ICIM (JICIM).

The speed of generation of genetic data and acquisition of gene information is increasing, owing to the development of user-friendly genetic software (for examples, see Tables S1–S3). Linkage analysis and QTL mapping are two closely related aspects of genetic studies but have hitherto been handled by separate software packages (Tables S2 and S3). For example, to use MapQTL for QTL mapping, one needs first to use JoinMap to build linkage maps. Over the last 10 years, we have developed integrated software called QTL IciMapping. It is freely available from <http://www.isbreeding.net/> and can be used for linkage analysis, map construction, and QTL mapping in most biparental populations. Our objective in this article is to introduce the functionalities, interfaces, inputs, and outputs of the software.

2. Materials and methods

2.1. Genetic mapping populations

Any genetic study requires one or several genetically segregating populations. Among populations used in plant genetic

studies, such as F_2 , backcross (BC), doubled haploids (DH), and recombinant inbred lines (RIL), two categories can be defined: temporary and permanent populations [3]. In a temporary population such as F_2 or BC, individuals in the population can segregate after self-pollination. In contrast, in a permanent population such as DH or RIL, each individual in the population is genetically homozygous, and the genetic structure will not change with self-pollination. With permanent populations, random environmental errors in phenotyping can be better controlled by replication, and accordingly the accuracy of QTL mapping can be improved. By using the QTL IciMapping software, genetic studies can be conducted in 20 biparental populations, populations of chromosome segment substitution (CSS) lines, and NAM populations (Fig. 1). Linkage map construction is limited to the 20 biparental populations, among which 10 are permanent and the other 10 are temporary.

Allelic and genotypic frequencies define the structure of a genetic population. We may denote by A and B the two alleles at one locus, and the genotypes of the two parental lines P_1 and P_2 as AA and BB, respectively. Table 1 gives the frequencies of the three genotypes AA, AB and BB and of the two alleles A and B in the 20 biparental populations. According to the frequency of allele A, these populations can be roughly classified into five categories: (1) F_1 -derived, where the allele A frequency is 0.5 (i.e., F_2 , F_3 , F_1 DH, and F_1 RIL); (2) P_1 BC $_1$ F $_1$ and P_1 BC $_1$ F $_1$ -derived, where the allele A frequency is 0.75 (i.e., P_1 BC $_1$ F $_1$, P_1 BC $_1$ F $_2$, P_1 BC $_1$ DH, and P_1 BC $_1$ RIL); (3) P_2 BC $_1$ F $_1$ and P_2 BC $_1$ F $_1$ -derived, where the allele A frequency is 0.25 (i.e., P_2 BC $_1$ F $_1$, P_2 BC $_1$ F $_2$, P_2 BC $_2$ DH, and P_2 BC $_1$ RIL); (4) P_1 BC $_2$ F $_1$ and P_1 BC $_2$ F $_1$ -derived, where the allele A frequency is 0.875 (i.e., P_1 BC $_2$ F $_1$, P_1 BC $_2$ F $_2$, P_1 BC $_2$ DH, and P_1 BC $_2$ RIL); and (5) P_2 BC $_2$ F $_1$ and P_2 BC $_2$ F $_1$ -derived, where the allele A frequency is 0.125 (i.e., P_2 BC $_2$ F $_1$, P_2 BC $_2$ F $_2$, P_2 BC $_2$ DH, and P_2 BC $_2$ RIL).

CSS (chromosome segment substitution) lines (also called introgression lines) are normally generated by repeated backcrossing, assisted by use of markers for donor segment selection and background uniformity (Fig. 1). In the ideal case in which each CSS line has a single segment from the donor parent, a standard analysis of variance, followed by multiple comparison between each line and the background parent, can readily be used to test whether a segment in one CSS line carries a QTL controlling the trait of interest. Unfortunately, it takes much labor and time to develop ideal CSS lines. Usually in a preliminary CSS population, each line carries a few segments from the donor parent. Owing to high intensity of selection in the process of population development, the gene and marker frequencies do not follow Mendelian ratios as in standard mapping populations such as F_2 , BC, DH, and RIL. The method for QTL mapping with CSS lines is called a stepwise regression-based likelihood ratio test (RSTEP-LRT) [5,19]. A NAM population is derived from a multiple-cross mating design with one common parent (Fig. 1) and affords high power and resolution via joint linkage and association analysis, and a broader genetic resource for quantitative trait analysis than biparental populations. The method for QTL mapping in NAM populations is called joint inclusive composite interval mapping (JICIM) [20].

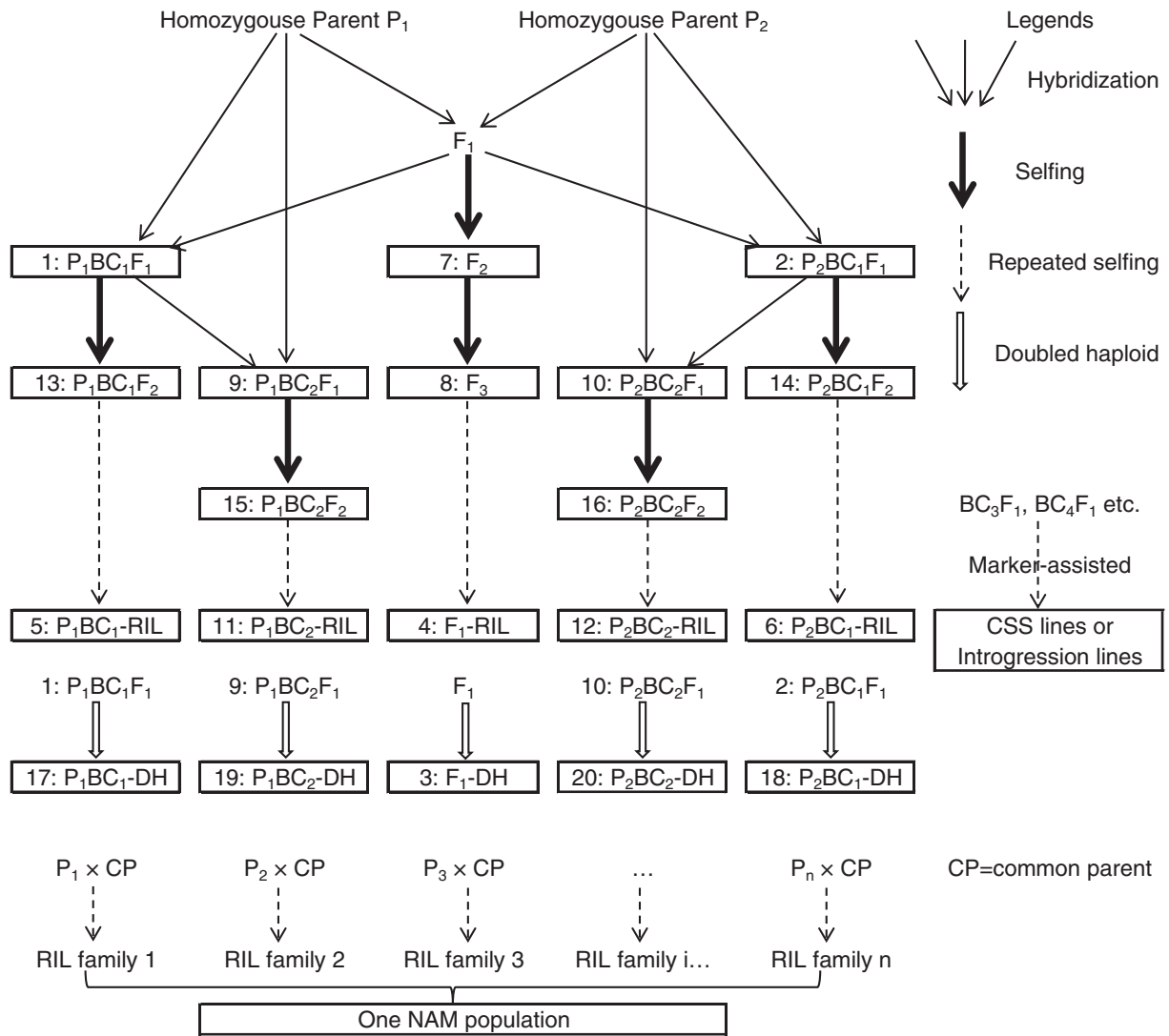


Fig. 1 – Genetic populations handled in QTL IciMapping. Two homozygous parental lines are represented by P_1 and P_2 , respectively. Twenty biparental populations (numbered 1–20) can be used for both map construction and QTL mapping. Where more than two rounds of backcrossing have been applied, the derived populations are treated as chromosome segment substitution lines with respect to the recurrent parent. A nested association mapping (NAM) population is derived from a multiple-cross mating design, which can also be handled in QTL IciMapping by a joint linkage mapping approach.

2.2. Encoding criteria for marker types

We denote AA as the genotype of P_1 (or parent A), BB as the genotype of P_2 (or parent B), and AB as the genotype of their F_1 hybrid. When one marker is dominant, AB + AA represents the two non-separated genotypes AB and AA. When one marker is recessive, AB + BB represents the two non-separated genotypes AB and BB. Some marker types may be missing. In total, we may have six possible types in the 20 biparental populations, i.e., AA, AB, BB, missing, AB + AA, and AB + BB. In fact, dominant or recessive markers can be treated as partially missing. It should be noted that not all the six types will be present at one marker. One marker can be only codominant, dominant, or recessive.

Accepted encoding options for the six possible types are summarized in Table 2. Each marker type can be encoded by a given number or by single or double capital letters. When numbers are used, the six types are encoded as 2, 1, 0, -1, 12, and 10. In this case, the numbers given for marker types AA, AB, and BB can be viewed as the numbers of A alleles in the marker type. When single letters are used, the six types are encoded as A, H, B, X (or *), D, and R. When double letters are used, the six types are encoded as AA, AB (or BA), BB, XX (or **), AX (or AH, or HA, or XA, or A*, or *A, A_, or _A), and BX (or BH, or HB, or XB, or B*, or *B, B_, or _B). Mixed encoding with numbers and capital letters is allowed in the software but is not recommended. The three recommended encoding criteria are given in the second part of Table 2. Assuming single letters

Table 1 – Theoretical frequencies of genotypes AA, AB, and BB and of alleles A and B at one locus in the 20 biparental populations.

Population no.	Full name	Genotype			Allele	
		AA	AB	BB	A	B
1	P ₁ BC ₁ F ₁	0.5	0.5	0	0.75	0.25
2	P ₂ BC ₁ F ₁	0	0.5	0.5	0.25	0.75
3	F ₁ DH	0.5	0	0.5	0.5	0.5
4	F ₁ RIL	0.5	0	0.5	0.5	0.5
5	P ₁ BC ₁ RIL	0.75	0	0.25	0.75	0.25
6	P ₂ BC ₁ RIL	0.25	0	0.75	0.25	0.75
7	F ₂	0.25	0.5	0.25	0.5	0.5
8	F ₃	0.375	0.25	0.375	0.5	0.5
9	P ₁ BC ₂ F ₁	0.75	0.25	0	0.875	0.125
10	P ₂ BC ₂ F ₁	0	0.25	0.75	0.125	0.875
11	P ₁ BC ₂ RIL	0.875	0	0.125	0.875	0.125
12	P ₂ BC ₂ RIL	0.125	0	0.875	0.125	0.875
13	P ₁ BC ₁ F ₂	0.625	0.25	0.125	0.75	0.25
14	P ₂ BC ₁ F ₂	0.125	0.25	0.625	0.25	0.75
15	P ₁ BC ₂ F ₂	0.8125	0.125	0.0625	0.875	0.125
16	P ₂ BC ₂ F ₂	0.0625	0.125	0.8125	0.125	0.875
17	P ₁ BC ₁ DH	0.75	0	0.25	0.75	0.25
18	P ₂ BC ₁ DH	0.25	0	0.75	0.25	0.75
19	P ₁ BC ₂ DH	0.875	0	0.125	0.875	0.125
20	P ₂ BC ₂ DH	0.125	0	0.875	0.125	0.875

AA and BB denote the genotypes of the two parental lines.

are used, A, H, B, and X are the possible allowed values for a codominant marker in an F₂ population. When HA or D is present at this marker locus, the software will report an error and ask the user to correct it.

2.3. Development of the QTL IciMapping software

In QTL IciMapping, the core modules for recombination frequency estimation and QTL mapping were written in Fortran 90/95, those for building linkage maps were written in C#, and all user interfaces were written in C#. The software runs on Windows XP/Vista/7/8, with Microsoft.NET Framework 2.0 (x86)/3.0/3.5. QTL IciMapping is project-based software. Once a user starts the software, the first step is to build a new project or open an existing project. The use of a project

Table 2 – Potential marker types and their encoding criteria in the QTL IciMapping software.

Genotype	AA	AB	BB	Missing	AB + AA	AB + BB
Encoding options	2 A AA	1 H AB BA	0 B BB	-1 X * **	12, D AH, HA AX, XA A*, *A A_, _A	10, R BH, HB BX, XB B*, *B B_, _B
Encoding by number	2	1	0	-1	12	10
Encoding by single letter	A	H	B	X	D	R
Encoding by double letters	AA	AB	BB	XX	AX	XB

The three encoding recommendations are given at the bottom of the table.

will assure that all operations and results will be properly saved when the software is closed. When the project is next opened, all operations and results previously performed can be properly retrieved.

Eight functionalities have been integrated in the software. The first, BIN, is designed to remove redundant markers. Redundant markers have identical segregation in the genetic population and can make no additional contribution to a genetic study. The second, MAP, can be used for the construction of genetic linkage maps in biparental populations. The well-designed and user-friendly interface takes the user easily through the grouping, ordering, and rippling procedures in map construction. The third, CMP, is designed for building a consensus map from multiple genetic linkage maps sharing common markers. The common markers can be used as anchors in CMP. The fourth, SDL, was developed for mapping segregation distortion loci in biparental populations.

The four functionalities described earlier need only genotypic data. The next four perform QTL mapping in different genetic populations and require both genotypic and phenotypic data. The first, BIP, performs mapping of additive, dominant, and digenic epistasis genes in biparental populations. Several traits are allowed in this functionality. The second, MET, performs QTL-by-environment interaction analysis in biparental populations. Only one trait is allowed in this functionality but may have been phenotyped in a number of environments. When multiple traits are phenotyped in multiple environments, the user must build one input file for each trait to use MET. The third, CSL, performs mapping of additive and digenic epistasis genes with chromosome segment substitution lines. The fourth, NAM, performs QTL mapping in NAM populations.

In addition to these eight functionalities, there are three supplementary tools in the software. The first, MapShow, displays completed genetic linkage maps. The second, 2pointREC, estimates recombination frequencies between two loci in 20 biparental populations. The third, ANOVA, performs analysis of variance and heritability estimation for multi-environment trials. The input file of each functionality or tool can be arranged in three formats: plain text, MS Excel 2003, and MS Excel 2007. In Excel format, each value takes one cell. Each marker is arranged in one row, with the marker name followed by marker types of all individuals in the population. In TEXT, values can be separated by a space, comma, or tab character. Below we will use BIN, MAP, and BIP functionalities as examples to describe the input files and interfaces of the software.

2.4. BIN functionality

By redundant markers, we mean markers that are completely correlated or identical in a population. They cannot provide additional information if more than one of them is used. In linkage analysis, the recombination frequency between them will be estimated as 0. Redundancy is commonly seen when one genetic population has a limited size, but thousands of molecular markers, such as by genotyping by sequencing, are screened. The BIN functionality was accordingly developed to remove redundant markers and generate the input file for map construction. Markers with correlation coefficients of 1

will be deleted either at random or by missing proportion, as chosen by the user.

For convenience, the input files for BIN and MAP have the same format, with three parts: (1) general information, (2) marker types, and (3) anchor information. Fig. 2 shows part of the data in the MS Excel format using a DH population as an example. Five parameters are included in the general information of the population (Fig. 2A). (1) Population type: 1 to 20 for the 20 biparental populations in Fig. 1; (2) mapping function: 1 for Kosambi function, 2 for Haldane function, and 3 for Morgan function; (3) marker space type: 1 for marker positions, and 2 for marker intervals; (4) number of markers; and (5) population size. The example given in Fig. 2A represents an F_1 -derived DH population. The Haldane mapping function is used to convert the recombination frequency to marker distance in cM. The 2378 markers on their chromosomes will be defined by positions and the population size is 225. It should be noted that general information on population type, mapping function, and marker space type will not be used in the BIN functionality. These parameters are used solely to generate an output file that can be directly used by the MAP functionality.

Genotypic data at all marker loci for all individuals in the population are defined in the second part of the input file. Fig. 2B shows data for 26 individuals for the first seven markers in the population. Column A in Fig. 2B gives marker names. Column B gives the marker genotypes of the first individual, column C those of the second individual, and so on. The number of columns following the marker name must equal the population size. For example, the first marker has the name “M1” and must have 225 data points of marker types at the first row. Anchor information is given in the order of markers defined in the third part of the input file. Fig. 2C shows anchor information for the first seven markers. In this example, no marker is anchored, so the anchor ID is set at 0 for each marker.

Fig. 3 shows the interface of the BIN functionality. The menu and tool bars are located at the top of the interface. The project window is located at left, displaying the loaded input files and their output files under each functionality node. At middle right is the marker summary display window, which shows marker ID, anchor information, marker names, missing rates, BIN ID, and deleted marker ID after the input file has been properly loaded. At bottom right is the parameter setting

A. General information

	A	B	C	D	E	F	G	H	I	J	K
1	3	Mapping Population Type (see remarks above)									
2	1	Mapping Function (1 for Kosambi; 2 for Haldane; 3 for Morgan)									
3	2	Marker Space Type (1 for intervals; 2 for positions)									
4	2378	Number of Markers									
5	225	Population size of the mapping population									
6											
7											

B. Marker types or genotypic data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
1	M1	2	0	0	0	2	2	2	2	2	2	2	2	2	0	2	2	2	0	0	0	2	2	2	2	0	0	
2	M2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	2	0	2	0	2	2	2	0	0	0	
3	M3	0	0	0	-1	2	0	0	2	2	0	0	2	0	0	2	0	2	0	0	0	0	0	2	0	0	2	0
4	M4	0	0	0	0	2	0	0	2	2	0	0	2	0	0	2	0	2	0	0	0	0	2	0	0	2	0	
5	M5	0	0	0	2	0	2	2	2	0	2	0	0	0	2	2	2	0	0	2	2	2	0	0	2	0	2	
6	M6	2	2	2	2	0	2	2	2	2	2	2	2	0	0	2	0	2	2	0	0	0	0	0	0	2	2	
7	M7	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	2	0	2	0	2	2	2	0	0	0	

C. Anchor information

	A	B	C	D	E	F	G	H	I	J	K	L
1	M1	0										
2	M2	0										
3	M3	0										
4	M4	0										
5	M5	0										
6	M6	0										
7	M7	0										

Fig. 2 – Input data for the BIN and MAP functionalities. (A) General information of the population, including population type, mapping function, marker space type, number of markers, and population size. (B) Genotypic data at all markers for all individuals in the population. (C) Anchor group information for all markers.

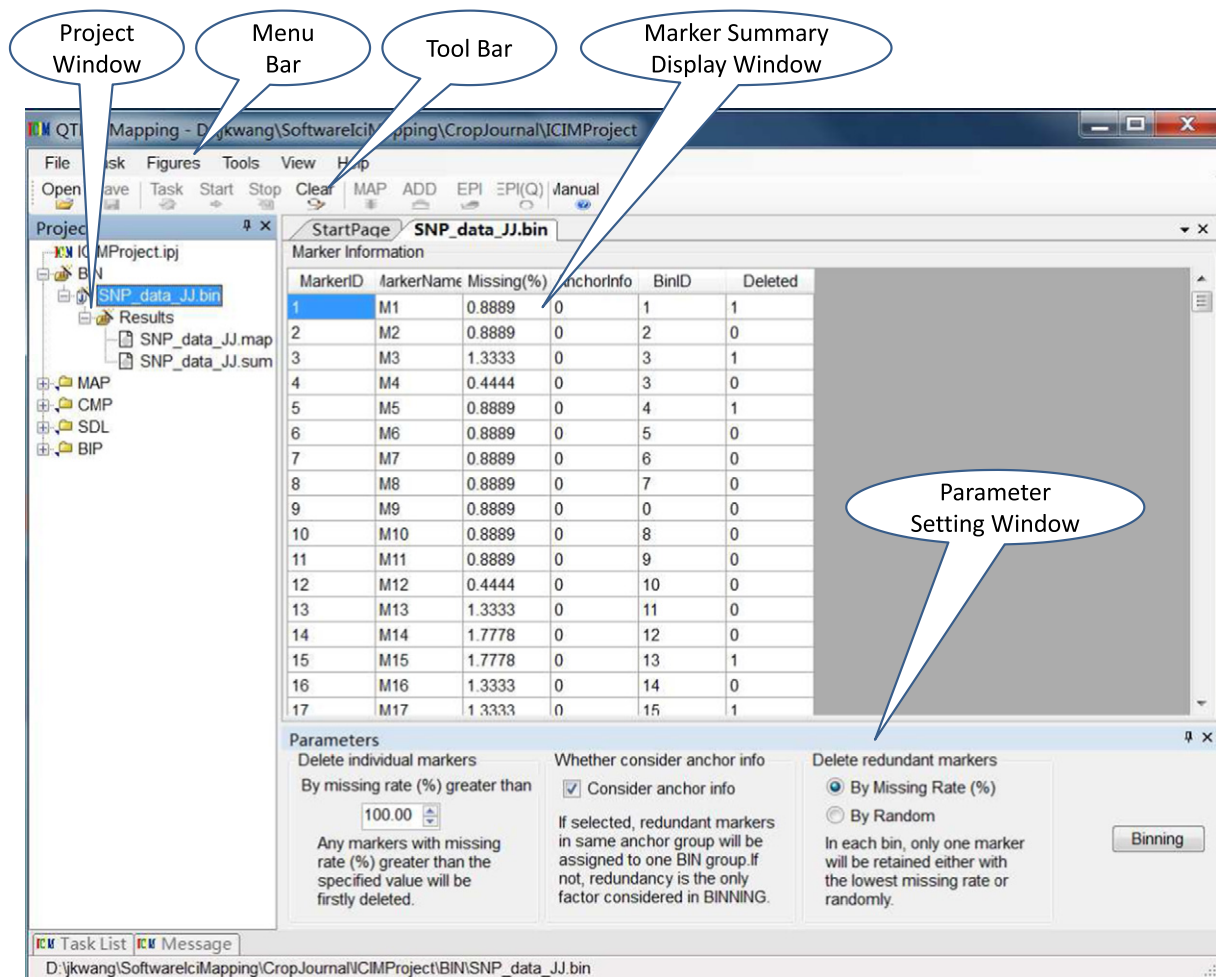


Fig. 3 – Interface of the BIN functionality. At the top of the interface are the menu and tool bars. At the left is the project window displaying the loaded input files and their output files under each functionality node. At the middle right is the marker summary display window. At the bottom right is the parameter setting window.

window. In this window users can set the threshold of the missing-data proportion for deleting markers, specify whether to consider anchor information in removing redundancy, and select the method for deleting redundant markers in each bin. Any markers with missing proportion greater than the specified value will be deleted first. In Fig. 3, 100% was used, indicating that no markers will be deleted according by missing proportion. For deleting by redundancy, only one marker will be retained in each bin, either with the lowest missing proportion or at random.

Clicking the “Binning” button in the parameter setting window will activate the BIN functionality to identify redundant markers in different bin groups and remove the redundancy based on the selected deletion criteria. Bin IDs and deleted marker IDs will be updated in the marker summary display window. Output files of this procedure will be listed in the project window. Anchor information may also be used in the BIN functionality (Fig. 3). In this case, redundancy was first determined for markers in each anchor group. In this manner, at least one marker will be retained in each anchor group when the anchor information is defined.

2.5. MAP functionality

Input files of the MAP functionality have exactly the same format as that of BIN (see Fig. 2) and will not be repeated here. Fig. 4 shows the interface of the MAP functionality. At the top are the menu and tool bars. At left is the project window. In the middle is the marker summary display window, which shows marker ID, marker names, anchor ID, group or chromosome ID, sample size in each marker class, results of χ^2 test for marker segregation distortion, and marker categories. At middle right is the linkage map display window. At bottom right is the parameter setting window.

Three steps (i.e., Grouping, Ordering, and Rippling) must be followed in order, after which a user can click Outputting to see the completed maps (lower right, Fig. 4). Once the MAP functionality is initiated, anchoring information is first displayed in a window at top right. Before Grouping is applied, the user can manage the anchor information (Fig. 5A). Note that moving markers in the display window can also be done by dragging the mouse.

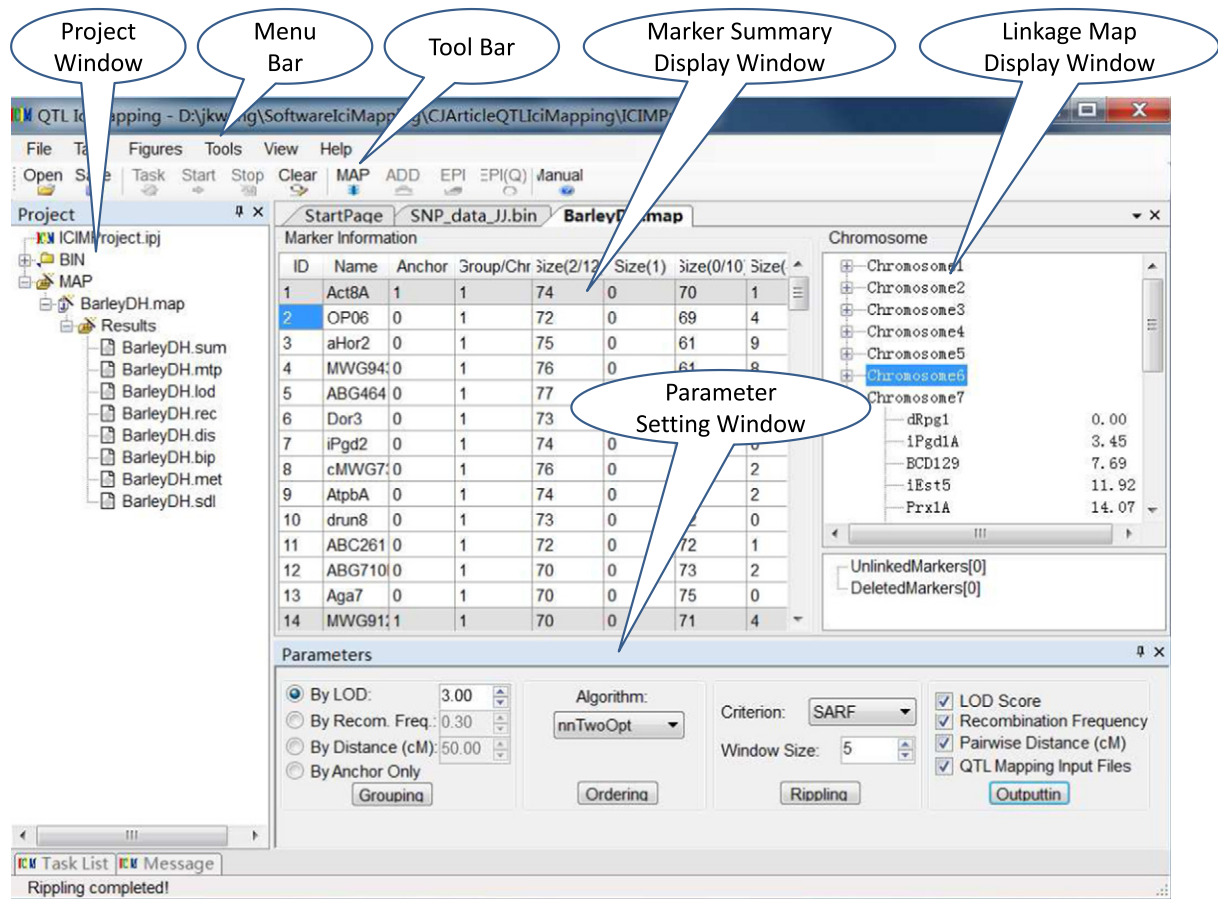


Fig. 4 – Interface of the MAP functionality. At the top of the interface are the menu and tool bars. At the left is the project window displaying the loaded input files and their output files under each functionality node. In the middle is the marker summary display window, and at the middle right is the linkage map display window. At the bottom right is the parameter setting window.

2.5.1. Grouping of markers

After the anchor information is correctly managed, a user can click the Grouping button so that the unanchored markers can be properly grouped (Fig. 4). Grouping can be based on (1) anchored marker information and a threshold of LOD score for unanchored markers, (2) anchored marker information and a threshold of marker recombination frequency for unanchored markers, (3) anchored marker information and a threshold of marker distance for unanchored markers, and (4) anchored marker information only.

For grouping, the ungrouped markers have priority to be added to the existing anchor groups. If one marker cannot be fitted into any anchor group, a new group will be generated. Thus, additional groups other than the anchor groups may be generated if some unanchored markers cannot be grouped with any anchor group. By right-clicking a grouped chromosome, the user can adjust the chromosome order, build a new chromosome group, delete the current chromosome, or perform ordering for markers within the chromosome (Fig. 5B). By right-clicking a marker, the user can move one group up or down, move one marker to another chromosome group, or delete the current marker (Fig. 5B). When any chromosomes or markers are deleted, those markers will be

shown in a list called “DeletedMarkers” below the Chromosome Display Window (Figs. 4 and 5). By right-clicking a deleted marker, one can link it to any existing chromosome group.

2.5.2. Ordering of markers

After all markers are correctly grouped, the user can click the Ordering button to make the genetic linkage maps (Fig. 4). Three ordering algorithms are available: (1) SER, a seriation algorithm [21]; (2) RECORD, recombination counting and ordering algorithm [22]; and (3) nnTwoOpt, an efficient approximate algorithm for solving traveling salesman problems in which nearest neighbors are used for tour construction and two-opt is used for tour improvement. Two-opt tour improvement is an efficient approximate algorithm in solving the traveling salesman problems [23,24].

We also implemented “By Input” and “By Anchor Order” in the software. When “By Input” is used, the marker order will be exactly the same as the order in the input file. Only the recombination frequency and genetic distance are re-estimated. This option is useful when the marker order is known, such as from a physical map. The interest is not in ordering the markers but in estimating genetic distances.

A. Handling anchor information

ID	Name	Anchor	Group/Chr	size(2/12)
1	Act8A	1	1	74
2	OP06	0	0	72
3	aHor2	0	0	75
4	MWG94	0	0	76
5	ABG464	0	0	77
6	Dor3	0	0	78
7	iPgd2	0	0	79
8	cMWG7	0	0	80
9	AtpbA	0	0	81
10	drun8	0	0	82
11	ABC261	0	0	83
12	ABG710	0	0	84
13	Aga7	0	0	85
14	MWG911	1	1	86
15	MWG84	2	2	87

B. Handling group information

C. Handling chromosome information

ID	Name	Anchor	Group/Chr	size(2/12)
1	Act8A	1	1	74
2	OP06	0	1	72
3	aHor2	0	1	75
4	MWG94	0	1	76
5	ABG464	0	1	77
6	Dor3	0	1	78
7	iPgd2	0	1	79
8	cMWG7	0	1	80
9	AtpbA	0	1	81
10	drun8	0	1	82
11	ABC261	0	1	83
12	ABG710	0	1	84
13	Aga7	0	1	85
14	MWG911	1	1	86
15	MWG84	2	2	87
16	ABG058	0	2	88
17	MWG516	0	2	89

D. Handling individual markers

Fig. 5 – Handling of groups and markers information at each step in map construction. A, Handling of anchor groups; B, handling of marker groups; C, handling of chromosomes; and D, handling of individual markers.

When “By Anchor Order” is used, the anchored marker order will be fixed. Unanchored markers are ordered without change in the order of anchor markers. Markers with no anchor information will be tried in the best positions. This option is useful when the order of some markers is known, such as from a physical map. The interest is not in ordering the anchored markers but in ordering unanchored markers.

After the Ordering button is clicked, a preliminary map indicating all marker positions will be shown in the Display Window (Fig. 5C). By right-clicking an ordered chromosome, one can adjust the chromosome order by selecting up or down, rename or reverse the current chromosome, delete the current chromosome, re-perform ordering or rippling for markers within the current chromosome, split the current chromosome into two sub-chromosomes from the longest marker interval, build a new chromosome group, or draw the current linkage map. One can also adjust the marker order, move the marker to other chromosome, or delete the current marker (Fig. 5D). By right-clicking a deleted marker, one can link it to any ordered chromosome.

2.5.3. Rippling of markers

SER, RECORD, and nnTwoOpt are heuristic algorithms. When there are too many markers, there is no guarantee that the final identified orders from these algorithms represent the global optimum solution. For this reason, after ordering, each marker sequence needs to be rippled for fine tuning. Rippling is performed by permutation of a window of m markers ($m = 2$ to 8 in QTL IciMapping) and comparison among all $m!$ possible orders. Initially, positions 1, ..., m are permuted, then position 2, ..., $m + 1$, and so on until the whole map is covered. Four rippling criteria available in the software are (i) SARF: sum of adjacent recombination frequencies; (ii) SAD: sum of adjacent distances; (iii) SALOD: sum of adjacent LOD scores; and (iv) COUNT: number of recombination events.

The intuitive criterion for choosing the ordering algorithm and rippling criteria is the length of the linkage map. The best method should yield the shortest linkage map. Rippling is used for fine tuning the chromosome orders. Generally, the shorter the chromosome length, the better are the rippling results. Operations similar to those shown in Fig. 5C and D are

available after Rippling. As shown in Fig. 5C, each group can be rippled separately by right-clicking pop-up menus. As with Ordering, by right-clicking an ordered chromosome, one can adjust the chromosome order by selecting up or down, build a new chromosome group, delete the current chromosome, perform ordering or rippling for markers within the chromosome, rename or reverse the current chromosome, or draw the linkage map. One can also adjust the marker order, move the current marker to another chromosome group, or delete the marker. By right-clicking a deleted marker, one can link it to any ordered chromosome.

2.6. BIP functionality

The input file for the QTL functionality consists of five parts: (1) general information, (2) marker number information, (3) linkage map information, (4) marker types, and (5) phenotypes. Fig. 6 shows part of the data in MS Excel format using a DH population as an example. Eight parameters are included in the general information of the population (Fig. 6A): (1) indicator—1 for QTL mapping in actual populations, and 2 for power simulation in simulated populations; (2) population type—1 to 20 for the 20 biparental populations in Fig. 1;

A. General information

	A	B	C	D	E	F	G	H	I	J	K	
1		1	Indicator: 1 for mapping; 2 for simulation									
2		3	Mapping Population Type									
3		1	Mapping Function (1 for Kosambi; 2 for Haldane; 3 for Morgan)									
4		1	Marker Space Type (1 for intervals; 2 for positions)									
5		1	Marker Space Unit (1 for centiMorgan; 2 for Morgan)									
6		7	Number of Chromosomes (or Linkage Groups)									
7		145	Size of the mapping population									
8		3	Number of traits									

B. Linkage group or chromosome information

	A	B	C	D	E	F	G	H	I	J	K
1	-Ch1H	14									
2	-Ch2H	18									
3	-Ch3H	15									
4	-Ch4H	13									

C. Definition of each chromosome or linkage group

	A	B	C	D	E	F	G	H	I	J	K
1	Act8A	1	0								
2	OP06	1	10.9								
3	aHor2	1	7.6								
4	MWC943	1	59.7								

D. Marker type or genotypic data

	A	B	C	D	E	F	G	H	I	J	K
1	Act8A	0	2	-1	2	0	2	2	0	0	2
2	OP06	0	2	2	2	0	2	2	0	0	2
3	aHor2	0	2	2	2	0	2	2	0	0	2
4	MWC943	2	2	2	2	2	2	0	0	0	0

E. Phenotype or phenotypic data

	A	B	C	D	E	F	G	H	I	J	K
1	KWT1	41.0129	40.4104	40.107	40.2816	41.4958	45.7471	40.1984	44.0952	42.0887	45.655
2	KWT2	45.75	44.35	44.6	45.9	47.05	51.4	45.95	49.15	49.9	50.7
3	KWT3	48.85	47.75	46.4	46.6	46.85	50.2	46.9	51.75	51	52.05

Fig. 6 – Input data for the QTL mapping functionality in actual populations. (A) General information of the population, including population type, mapping function, marker space type, marker space unit, number of chromosomes, population size, and number of phenotypic traits. (B) Marker number information, including name of the linkage group or chromosome, and number of markers on the group or chromosome. (C) Linkage map information, including marker name, group or chromosome ID, and position or interval of the marker. (D) Genotypic data at all markers for all individuals in the population. (E) Phenotypic data of all traits.

(3) mapping function—1 for Kosambi function, 2 for Haldane function, and 3 for Morgan function; (4) marker space type—1 for marker positions and 2 for marker intervals; (5) marker space unit—1 for centimorgan and 2 for morgan; (6) number of chromosomes or linkage groups; (7) population size; and (8) number of traits. The example given in Fig. 6A shows an actual DH population. The Haldane mapping function was used to convert marker distance to recombination frequency. Markers on the seven chromosomes are defined by positions. The unit of marker space is cM and the population size is 145. The number of phenotypic traits of interest is 3.

Marker number and linkage map information are defined as in Fig. 6B and C, respectively. In the second part of the input file are marker numbers on chromosomes. Fig. 6B shows the information of the first three chromosomes. There are 14, 18, and 15 markers, respectively. In the third part of the input file are marker names, group or chromosome ID, and position or interval of the marker. Fig. 6C shows the positions of the first three markers on chromosome 1. Genotypic data at all markers for all individuals in the population are defined

in the fourth part of the input file. Fig. 6D shows data for nine individuals at the first three markers in the population. Phenotypic data for all traits are presented in the fifth part of the input file. Any missing phenotypes are represented by -100.0. Fig. 6E shows the phenotypes of the first ten individuals for the trait.

QTL IciMapping can also simulate the 20 biparental populations, given the QTL positions and effects. In this case, the indicator in the general information is equal to 2 (Fig. 7A). Population type, mapping function, marker space type, marker space unit, number of chromosomes, and population size are defined as in the case of an actual mapping population. Marker number and linkage map information are defined as in Fig. 6B and C and were not given in Fig. 7. Instead of marker type and phenotype in an actual QTL mapping population, a number of QTL for one or multiple traits must be defined before power simulation studies are performed. Fig. 7B shows QTL information for the first trait. For this trait, there is one QTL each on the first three chromosomes and no QTL on the last two chromosomes. The QTL number on each

A. General information

	A	B	C	D	E	F	G	H	I
1	2	Indicator (1 for mapping, 2 for simulation)							
2	3	Mapping Population Type							
3	2	Mapping Function (1 for Kosambi; 2 for Haldane; 3 for Morgan)							
4	2	Marker Space Type (1 for intervals; 2 for positions)							
5	1	Marker Space Unit(1 for centiMorgan; 2 for Morgan)							
6	5	Number of Chromosomes (or Linkage Groups)							
7	200	Size of the mapping population							
8	4	Number of traits							

B. QTL information for each trait

	A	B	C	D	E	F	G	H	I
1	1	1	35						
2	1	1	53						
3	1	1	22						
4	1	0							
5	1	0							
6	1	1							
7	1	0	1.25						
8	1	0	0	1.5					
9	1	1							
10	1	0.6							
11	1	10							
12	2	2	35	70					

Fig. 7 – Input data for the QTL simulation functionality. (A) General information of the population, including population type, mapping function, marker space type, marker space unit, number of chromosomes, population size, and number of phenotypic traits. (B) QTL information, including number of QTL on each chromosome and their positions, additive effects, additive-by-additive epistatic effects, heritability or error variance, and population mean. Chromosome and linkage map information are the same as in actual mapping populations (Figs. 5B and 5C). Marker type and phenotype will be simulated using the specified QTL information and accordingly are not needed.

chromosome is followed by the QTL position. For example, the first QTL is located at 35 cM on the first chromosome.

A lower triangular matrix is used to define the additive and additive-by-additive epistatic effects of all QTL for the trait of interest. For example, in Fig. 7B, a 3×3 lower triangular matrix is defined for trait 1. The heritability in the broad sense or error variance must be given, in order to determine the phenotypic value of each individual in the mapping population. For trait 1, 0.6 is defined as the heritability, as the indicator 1 has been specified. Then the mean of the simulated phenotype corresponding to each trait needs to be specified. The software will use linkage map and QTL information to simulate marker type and phenotype, which are then used for QTL mapping.

Fig. 8 shows the interface of the BIP functionality. At top are the menu and tool bars. At left is the project window. At middle right is the input file display window. The file to be displayed in this window can be selected from the project window. At bottom right is the parameter setting window. Six methods are available in BIP: (1) SMA: single marker analysis [4]; (2) IM-ADD: the conventional interval mapping of additive and dominant QTL [1]; (3) ICIM-ADD: inclusive composite

interval mapping of additive and dominant QTL [3,11,12]; (4) SGM: selective genotyping mapping [25]; (5) IM-EPI: conventional interval mapping of epistatic QTL; (6) ICIM-EPI: inclusive composite interval mapping of epistatic QTL [4,11,16,17]. After mapping method selection and parameter setting, when the “Start” item in the menu bar is clicked, the BIP functionality will be activated and the mapping results will be shown in the project window.

3. Results

3.1. Output files of the BIN functionality

For the eight functionalities in QTL IciMapping, the output files have the same prefix name as the input file but with different extensions. There are two output files after BIN is run (see Project Window in Fig. 3). One has the extension *.sum and contains summary information of the identified bin groups and deleted markers in each bin (Fig. S1). The bin group with number 0 contains markers that were not redundant and not deleted. Number 1 in the last column

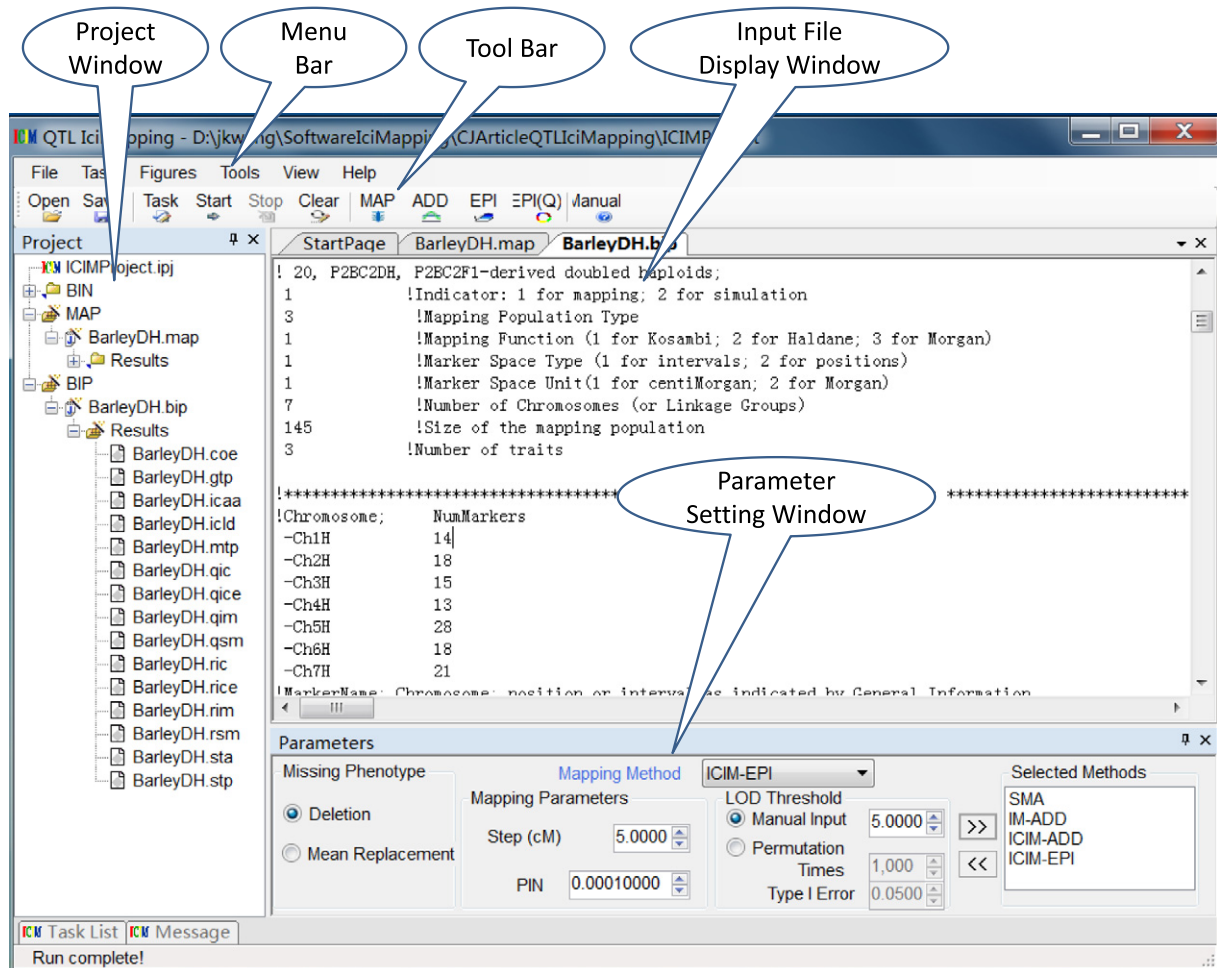


Fig. 8 – Interface of the QTL functionality. At the top of the interface are the menu and tool bars. At left is the project window displaying the loaded input files and their output files under each functionality node. At middle right is the input and output file display window. At bottom right is the parameter setting window.

indicates that this marker was deleted, and number 0 that this marker was retained for further genetic analysis. The other output file has the extension *.map and contains the nonredundant markers and can be directly loaded for linkage map construction.

3.2. Output files of the MAP functionality

Eight output files are generated by the MAP functionality (Table 3; also see Project Window in Fig. 4). A file with extension *.mtp contains a marker summary, which is similar to the information shown in the marker summary display window in Fig. 4. Files with extension *.sum contain summary information of the completed linkage maps. Figure S2 shows the information of seven completed linkage maps from one barley DH population. The chromosomal position of each marker is given first. The upper part of Fig. S2 shows the information for all markers on the first chromosome. The first column is chromosome ID, followed by chromosome name, marker name, interval length in cM from the previous marker, position in cM, recombination frequency with one previous marker, and coefficient of interference between two marker intervals. After the chromosomal positions of all markers, the number of markers and chromosomal length are given for each chromosome. It can be seen that the total length of the seven chromosomes is 1274.61 cM when the Haldane mapping function was used. Given at the end is the deleted marker information. No markers were deleted or unlinked in this population.

Files with extension *.rec contain the estimates of pairwise recombination frequencies between all markers, arranged in a lower triangular matrix (Table 3). Figure S3 shows the estimates of recombination frequencies between the first ten markers in the population. For example, the estimated recombination frequencies of marker 1 with markers 2, 3, and 8 were 0.1071, 0.1111, and 0.4507, respectively. Estimates above 0.5 are allowed and may arise from sampling error or indicate wrong linkage phases in the two parents. Wrong linkage phases provide evidence that some markers were misclassified between the two parents. Files with extension *.lod contain the pairwise LOD scores between all markers for testing linkages (Table 3). Files with extension *.dis contain the pairwise genetic distances (cM) between all markers (Table 3). For using genetic distance in ordering, a distance of 1000.00 is used to indicate no linkage between two markers. The formats of *.lod and *.dis output files are similar to that of the *.rec output file (Fig. S3).

Files with extension *.sdl have the same format as the input file for the SDL functionality. Files with extension *.bip have the same format as the input file for the BIP functionality, except that the number of traits is set temporarily at 0. Changing the number of traits and adding phenotypic data renders this output file ready for QTL mapping in BIP. Files with extension *.met have exactly the same format as the input file for the MET functionality, except that number of environments is set at 0 temporarily. Changing the number of environments and adding phenotypic data renders this output file ready for QTL-by-environment analysis in MET.

The software also provides a function for drawing and editing linkage maps after the MAP functionality has been run (Fig. S4). The linkage maps can be drawn for individual chromosomes, all chromosomes, or a set of selected chromosomes. The user can add or delete marker positions, marker names, and chromosome names in the display window. Advanced editing functions are also available in this window; for example, modifying the color of chromosomes, adding centromeres, and adding or deleting chromosomes.

3.3. Files of the BIP mapping functionality

More than 30 output files can be generated by the BIP mapping functionality in actual populations (Table 4). Four files recording general information of the mapping population have extensions *.coe, *.mtp, *.sta, and *.stp. The file *.coe contains pairwise correlation coefficients between markers. This output may be used to check the quality of the linkage map. The first part of the *.mtp file is similar to the *.mtp file from the MAP functionality and the second part gives the marker types after imputation of incomplete and missing markers. Files with extension *.sta contain basic statistical analyses of phenotypic data in the population, and file *.stp contains results from stepwise regression.

Each QTL mapping method (SMA, IM-ADD, ICIM-ADD, SGM, IM-EPI, or ICIM-EPI) has three kinds of output, denoted as Q for significant QTL, R for results at all scan positions, and T for permutation tests (Table 4). As an example, Fig. S5 shows the content in file *.qic from a barley DH population with three traits and Fig. S6 shows partial contents of file *.rice from the same population. From the one-dimensional scan by ICIM, six QTL were identified for trait KWT1, five for trait KWT2, and four for trait KWT3 (Fig. S5). For each identified QTL, the chromosomal position, nearest left marker, nearest right marker, LOD, PVE (%), and additive effect are reported (Fig. S5). For populations such as F₂ or F₃ with three genotypes

Table 3 – Description of output files from the MAP functionality.

Output file	Description
MTP	Frequency of different marker types and chi-square test for segregation distortion in the genetic population for each marker
SUM	Summary of the constructed combined linkage maps
LOD	Lower triangular matrix of marker pairwise LOD score testing the combined linkage relationship (against independent inheritance)
REC	Lower triangular matrix of marker pairwise combined recombination frequency estimates
DIS	Lower triangular matrix of marker pairwise genetic distance (cM) calculated from combined recombination frequency
SDL	Ready-to-use input file for the SDL functionality. The file can be loaded into the software for segregation distortion locus mapping.
BIP	Ready-to-use input file for the BIL functionality. The file can be loaded into the software for both additive and epistatic mapping after addition of phenotypic data at the end.
MET	Ready-to-use input file for the MET functionality. The file can be loaded into the software for QTL-by-environment analysis after addition of multi-environmental phenotypic data.

Table 4 – Description of output files from the BIP functionality of actual mapping populations.

Output file	Description
COE	Lower triangular matrix of pairwise correlation coefficient between markers in the mapping population
MTP	Frequency of different marker types and the chi-square test for segregation distortion in the genetic population for each marker
STA	Basic statistics of phenotypes
STP	Stepwise regression in the first step of Inclusive composite interval mapping for both additive and epistatic QTL
QSM, QIM, QIC, QSG	Additive (and dominant) QTL identified from single marker analysis (SM), interval mapping of additive QTL (IM), inclusive composite interval mapping (IC) of additive QTL, and selective genotyping (SG)
RSM, RIM, RIC, RSG	Results of every one-dimensional scanning position from single marker analysis (SM), interval mapping of additive QTL (IM), inclusive composite interval mapping of additive QTL (IC), and selective genotyping (SG)
TSM, TIM, TIC, TSG	Permutation tests of additive (and dominant) mapping from single marker analysis (SM), interval mapping of additive QTL (IM), inclusive composite interval mapping of additive QTL (IC), and selective genotyping (SG)
QIME, QICE	Digenic epistatic QTL identified from interval mapping of epistatic QTL (IME) and inclusive composite interval mapping of epistatic QTL (ICE)
RIME, RICE	Results of every two-dimensional scanning from Interval Mapping of epistatic QTL (IME) and inclusive composite interval mapping of epistatic QTL (ICE)
TIME, TICE	Permutation tests of digenic epistatic mapping from Interval Mapping of epistatic QTL (IME) and inclusive composite interval mapping of epistatic QTL (ICE)
IMLD, ICLD	Lower triangular matrix of LOD score at every two-dimensional scanning positions from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMAA, ICAA	Lower triangular matrix of additive-by-additive effect at every two-dimensional scanning from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMAD, ICAD	Lower triangular matrix of additive-by-dominance effect at every two-dimensional scanning from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMDA, ICDA	Lower triangular matrix of dominance by additive effect at every two-dimensional scanning from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMDD, ICDD	Lower triangular matrix of dominance by dominance effect at every two-dimensional scanning from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
GTP	Bayesian classification of genotypes of additive and dominance QTL identified from inclusive composite interval mapping of additive and dominant QTL

at each locus, a dominance effect will be reported as well. For two-dimensional scanning of epistatic QTL, the two chromosomal positions, two nearest left markers, two nearest right markers, LOD, PVE (%), two additive effects, and additive-by-additive epistasis are reported in populations with two genotypes at each locus (Fig. S6). For populations with three genotypes, two dominance effects and dominance-associated epistasis will be reported as well.

For IM-EPI and ICIM-EPI, LOD (IMLD and ICLD) and epistasis (IMAA, ICAA, etc.) are also output as a lower triangular matrix (Table 4). Users may use these outputs to draw contour or 3-D graphs in other graphics software. File *.gtp shows the posterior probability and predicted genotype of each identified QTL in each individual, and the predicted genotypic value for each individual (Table 4). This file is only for ICIM-ADD.

The software also provides plots of LOD scores and genetic effects on one, all, or selected chromosomes. Fig. S7 shows the one-dimensional LOD and additive profiles for three traits in the barley DH population. The user can select chromosomes, traits, LOD scores, and genetic effects for displaying in the Figure window. The software also provides figures that combine LOD scores or identified QTL with the linkage map, for IM and ICIM. Once the plot is satisfactory, the user can copy/paste, save, or print the figure via a pop-up menu activated by right-clicking the mouse.

3.4. Files of the BIP simulation functionality

More than 20 output files are generated by the BIP simulation functionality (Table 5). Files with extension *.sta contain basic genetic parameters of the specified QTL information for each

trait to be simulated, some of which are additive variance, epistatic variance, ratios of additive and epistatic variance to total genetic variance, and error variance. It can be seen in Fig. S8 that the fourth simulated trait has both additive and epistatic variances, which are equal to 2.25, and 4.0, respectively. Total genetic variance is 6.25, and error variance is 2.68, resulting in a heritability of about 0.8. File *.sta also gives the theoretical coefficients of the flanking markers of the defined QTL (second part of Fig. S8). File *.stp contains results from stepwise regression in each simulated population.

Each QTL mapping method (SMA, IM-ADD, ICIM-ADD, SGM, IM-EPI, or ICIM-EPI) has three kinds of output: Q for significant QTL in each simulated population, R for results at scan positions, and P for QTL detection power (Table 5). Q and R outputs have the same formats as those in actual mapping populations. By using the P outputs, the user can compare QTL detection powers of different mapping methods. For IM-EPI and ICIM-EPI, average LOD (i.e., IMLD and ICLD) and epistatic effects (i.e., IMAA, ICAA etc.) across all simulated populations are output as a lower triangular matrix as well (Table 5). As with actual mapping populations, users may use these outputs to draw contour or 3-D graphs in other graphics software.

The software provides graphs of LOD scores from one- and two-dimensional scanning. Using one genetic model including linkage, Fig. S9A and B show LOD profiles for IM and ICIM in additive QTL mapping across 10 simulated populations. ICIM distinguishes the two linked QTL on the first chromosome, but IM does not. Using one genetic model including epistasis, Fig. S9C and D show LOD profiles for IM and ICIM in epistatic QTL mapping across 10 simulated populations. ICIM

Table 5 – Description of output files from the BIP functionality of simulated populations.

Output file	Description
STA	Basic statistics of the QTL information
STP	Stepwise regression in the first step of inclusive composite interval mapping for both additive and epistatic QTL
QSM, QIM, QIC, QSG	Additive (and dominant) QTL identified from single marker analysis (SM), interval mapping of additive QTL (IM), inclusive composite interval mapping (IC) of additive QTL, and selective genotyping (SG) in each simulated population
RSM, RIM, RIC, RSG	Average results of every one-dimensional scanning position from single marker analysis (SM), interval mapping of additive QTL (IM), inclusive composite interval mapping of additive QTL (IC), and selective genotyping (SG) across all simulated populations
PSM, PIM, PIC, PSG	Power analysis of additive (and dominant) mapping from single marker analysis (sm), interval mapping of additive QTL (IM), inclusive composite interval mapping of additive QTL (IC), and selective genotyping (SG)
QIME, QICE	Digenic epistatic QTL identified from interval mapping of epistatic QTL (IME) and inclusive composite interval mapping of epistatic QTL (ICE) in each simulated population
RIME, RICE	Average results of every two-dimensional scanning from interval mapping (IME) and inclusive composite interval mapping (ICE) across all simulated populations
PIME, PICE	Power analysis of digenic epistatic mapping from interval mapping of epistatic QTL (IME) and inclusive composite interval mapping of epistatic QTL (ICE)
IMLD, ICLD	Lower triangular matrix of LOD score at every two-dimensional scan position from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMAA, ICAA	Lower triangular matrix of additive-by-additive effect at every two-dimensional scan position from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMAD, ICAD	Lower triangular matrix of additive-by-dominance effect at every two-dimensional scan position from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMDA, ICDA	Lower triangular matrix of dominance by additive effect at every two-dimensional scan position from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)
IMDD, ICDD	Lower triangular matrix of dominance by dominance effect at every two-dimensional scan position from interval mapping of epistatic QTL (IM) and inclusive composite interval mapping of epistatic QTL (IC)

has a much sharper peak at the intersection of the two linked QTL, indicating its high detection power and accuracy for epistatic QTL as well.

4. Discussion

Efficient genetic studies depend on large populations with high-quality genotypic and phenotypic data. Genetic analysis based on large amounts of data requires extensive computing capabilities. The current version of the QTL IciMapping software has eight integrated functionalities, greatly reducing the workload in data format transformation from one computer software program to others. For example, output from BIN can be directly used as input for MAP and output from MAP can be directly used as input for SDL. Some output files from MAP can be readily used as input files for BIP and MET, after phenotypic data are added. The integration of the software provides a seamless pipeline from the treatment of redundant markers, to the estimation of recombination frequency, to the construction of linkage maps, and finally to the mapping of QTL.

Many QTL mapping methods have been developed, but only the most efficient and powerful method should be applied. Statistically, the best QTL mapping method should have high detection power and low false discovery rate. However, the asymptotic properties of most test statistics used in QTL mapping are barely known theoretically. The simulation functionality in QTL IciMapping provides a useful tool to compare different mapping methods under a wide range of genetic models [3,4,13–15]. Statistical power is the probability that the null hypothesis is rejected when it is indeed false. In QTL mapping, power represents the probability that a true QTL can be detected. Extensive previous simulation studies have shown the great advantage of ICIM in

improving QTL detection power [3,12], separating linked QTL [13], mapping interacting QTL [15,16], and assessing multiparental populations [20] and QTL-by-environment interaction [18]. In addition, many frequently asked questions in QTL mapping can be properly investigated [26], such as the effect of missing and distorted markers [15], detection power of epistasis [17], detection power of selective genotyping [25], use of mathematically-derived traits [27], and choice of LOD threshold [28].

CSS lines are ideal genetic materials for gene fine mapping and map-based cloning, but severe selection occurs during the development of CSS lines. For this reason, genetic analysis methods for typical biparental populations become inapplicable. A likelihood ratio test based on stepwise regression has been implemented in QTL IciMapping as the CSL functionality. NAM is an extreme case of multiparental design, where multiple crosses share a common parent. The population derived from one cross is a typical biparental population. QTL mapping for the NAM design has been implemented in QTL IciMapping as the NAM functionality.

Clonal species are common among plants. Clonal F_1 progenies are derived from hybridization between two heterozygous clonal lines. In self- and cross-pollinated species, double crosses (or four-way crosses) can be made from four inbred lines in order to extend the diversity in genetic studies and plant breeding. Clonal F_1 and double-cross populations have more alleles at each locus than the biparental populations [29]. We have developed linkage analysis and QTL mapping in populations of clonal F_1 and double crosses as another integrated software called GACD, which can be used for linkage analysis, map construction, and QTL mapping in clonal F_1 and double-cross populations. Methods and software for genetic analysis in populations derived from eight parental lines are under development.

Availability

QTL IciMapping is freely available from <http://www.isbreeding.net/>. The download package also contains a manual and sample datasets.

Acknowledgments

This work was supported by the Natural Science Foundation of China (31271798), the Generation Challenge Program (GCP), and HarvestPlus Challenge Program of CGIAR.

Supplementary material

Supplementary material to this article can be found online at <http://dx.doi.org/10.1016/j.cj.2015.01.001>.

REFERENCES

- [1] E.S. Lander, D. Botstein, Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics* 121 (1989) 185–199.
- [2] D.S. Falconer, T.F.C. Mackay, *Introduction to Quantitative Genetics*, Fourth Edition Longman Group Ltd., 1996.
- [3] H. Li, G. Ye, J. Wang, A modified algorithm for the improvement of composite interval mapping, *Genetics* 175 (2007) 361–374.
- [4] J. Wang, H. Li, L. Zhang, *Genetic Mapping and Breeding Design* (in Chinese), The Science Press of China, Beijing, 2014.
- [5] J. Wang, H. Li, X. Wan, W. Pfeiffer, J. Crouch, J. Wan, Application of identified QTL-marker associations in rice quality improvement through a design breeding approach, *Theor. Appl. Genet.* 115 (2007) 87–100.
- [6] X. Wan, J. Weng, H. Zhai, J. Wang, C. Lei, X. Liu, T. Guo, L. Jiang, N. Su, J. Wan, QTL analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5, *Genetics* 179 (2008) 2239–2252.
- [7] J. Wang, H. Li, X. Zhang, C. Yin, Y. Li, Y. Ma, X. Li, L. Qiu, J. Wan, Molecular design breeding in crops in China, *Acta Agron. Sin.* 37 (2011) 191–201.
- [8] N.T.J. Bailey, *Introduction to the Mathematical Theory of Genetic Linkage*, Oxford University Press, Oxford, UK, 1961.
- [9] Z. Sun, H. Li, L. Zhang, J. Wang, Estimation of recombination frequency in biparental genetic populations, *Genet. Res.* 94 (2012) 163–177.
- [10] Z. Zeng, Precision mapping of quantitative trait loci, *Genetics* 136 (1994) 1457–1468.
- [11] J. Wang, Inclusive composite interval mapping of quantitative trait genes, *Acta Agron. Sin.* 35 (2009) 239–245.
- [12] L. Zhang, H. Li, Z. Li, J. Wang, Interactions between markers can be caused by the dominance effect of quantitative trait loci, *Genetics* 180 (2008) 1177–1190.
- [13] H. Li, S. Hearne, M. Bänziger, Z. Li, J. Wang, Statistical properties of QTL linkage mapping in biparental genetic populations, *Heredity* 105 (2010) 257–267.
- [14] H. Li, L. Zhang, J. Wang, Estimation of statistical power and false discovery rate of QTL mapping methods through computer simulation, *Chin. Sci. Bull.* 57 (2012) 2701–2710.
- [15] L. Zhang, S. Wang, H. Li, Q. Deng, A. Zheng, S. Li, P. Li, Z. Li, J. Wang, Effects of missing marker and segregation distortion on QTL mapping in F_2 populations, *Theor. Appl. Genet.* 121 (2010) 1071–1082.
- [16] H. Li, J.M. Ribaut, Z. Li, J. Wang, Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations, *Theor. Appl. Genet.* 116 (2008) 243–260.
- [17] L. Zhang, H. Li, J. Wang, Statistical power of inclusive composite interval mapping in detecting digenic epistasis showing common F_2 segregation ratios, *J. Integr. Plant Biol.* 54 (2012) 270–279.
- [18] S. Li, J. Wang, L. Zhang, Inclusive composite interval mapping (ICIM) of QTL-by-environment interactions in biparental populations, *PLoS One* (2015) (under review).
- [19] J. Wang, X. Wan, J. Crossa, J. Crouch, J. Weng, H. Zhai, J. Wan, QTL mapping of grain length in rice (*Oryza sativa* L.) using chromosome segment substitution lines, *Genet. Res.* 88 (2006) 93–104.
- [20] H. Li, P. Bradbury, E. Ersoz, E. Buckler, J. Wang, Joint QTL linkage mapping for multiple-cross mating design sharing one common parent, *PLoS ONE* 6 (2011) e17573. <http://dx.doi.org/10.1371/journal.pone.0017573>.
- [21] K.H. Buetow, A. Chakravarti, Multipoint gene mapping using seriation: I. General methods, *Am. J. Hum. Genet.* 41 (1987) 180–188.
- [22] H. van Os, P. Stam, R.G. Visser, H.J. Van Eck, RECORD: a novel method for ordering loci on a genetic linkage map, *Theor. Appl. Genet.* 112 (2005) 30–40.
- [23] S. Lin, B.W. Kernighan, An effective heuristic algorithm for the traveling-salesman problem, *Oper. Res.* 21 (1973) 498–516.
- [24] G. Laporte, The traveling salesman problem: an overview of exact and approximate algorithm, *Eur. J. Oper. Res.* 59 (1992) 231–247.
- [25] Y. Sun, J. Wang, J.H. Crouch, Y. Xu, Efficiency of selective genotyping for genetic analysis and crop improvement of complex traits, *Mol. Breed.* 26 (2010) 493–511.
- [26] H. Li, L. Zhang, J. Wang, Analysis and answers to frequently asked questions in quantitative trait locus mapping, *Acta Agron. Sin.* 36 (2010) 918–931.
- [27] Y. Wang, H. Li, L. Zhang, W. Lu, J. Wang, On the use of mathematically-derived traits in QTL mapping, *Mol. Breed.* 29 (2012) 661–673.
- [28] Z. Sun, H. Li, L. Zhang, J. Wang, Properties of the test statistic under null hypothesis and the calculation of LOD threshold in quantitative trait loci (QTL) mapping, *Acta Agron. Sin.* 39 (2013) 1–11.
- [29] L. Zhang, H. Li, J. Wang, Linkage analysis and map construction in genetic populations of clonal F_1 and double cross, *Genes Genomes Genet.* 5 (2015) 427–439.