THE CROP JOURNAL 3 (2015) 269–283

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 Meng[1], *Huihui Li*[1], *Luyan Zhang*[1], *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.

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.

http://dx.doi.org/10.1016/j.cj.2015.01.001

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_1DH, and F_1RIL); (2) $P_1BC_1F_1$ and $P_1BC_1F_1$-derived, where the allele A frequency is 0.75 (i.e., $P_1BC_1F_1$, $P_1BC_1F_2$, P_1BC_1DH, and P_1BC_1RIL); (3) $P_2BC_1F_1$ and $P_2BC_1F_1$-derived, where the allele A frequency is 0.25 (i.e., $P_2BC_1F_1$, $P_2BC_1F_2$, P_2BC_2DH, and P_2BC_1RIL); (4) $P_1BC_2F_1$ and $P_1BC_2F_1$-derived, where the allele A frequency is 0.875 (i.e., $P_1BC_2F_1$, $P_1BC_2F_2$, P_1BC_2DH, and P_1BC_2RIL); and (5) $P_2BC_2F_1$ and $P_2BC_2F_1$-derived, where the allele A frequency is 0.125 (i.e., $P_2BC_2F_1$, $P_2BC_2F_2$, P_2BC_2DH, and P_2BC_2RIL).

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].