

Comparison of *in silico* Quantitative Trait Loci (QTL) Analysis Tools

Hemani Sharma^{1*}, Sushma Jha², Sunita Jain¹ and Sudhir Kumar¹

¹Bioinformatics Section, CCSHAU, Hisar

²Department of Biotechnology and Molecular Biology, CCSHAU, HISAR

Abstract A Quantitative trait loci (QTL) is the location of a gene that affects a trait that is measured on a quantitative scale. QTL analysis has become an important tool to dissect the genetics of complex characters especially in molecular breeding and human disease diagnosis. A large numbers of software are available for analysis of QTL. Decision about which tools to use is one of the important problems, especially for the non-specialist users. Therefore, a comparison of different QTL analysis software is necessary to interpret the biological data correctly. In current study, three free software were chosen namely WIN QTL Cartographer, QTL Mapper and QTL Network to detect QTLs for Plant height, Panicle no./effective tiller and yield components in a RIL population of rice. Study revealed marked difference in the type and number of analysis, data type they support, methodology on which software are based and their graphical interface. WIN QTL Cartographer is based on single marker analysis, interval mapping, composite interval mapping, Bayesian interval mapping, multiple interval mapping, multiple trait analysis and categorical trait analysis. It represents the result graphically and provide information regarding additive, dominant, R^2 traits. QTL Mapper is based on mixed linear model approach whereas QTL Network is based on mixed model composite interval mapping. Our results shows WIN QTL Cartographer to be the best for general QTL finding as it uses both CIM and MQM approaches. QTL Network, however, can be used for specific epistatic estimation and QTL environmental interaction effects.

Index Terms— Single marker analysis, Composite interval mapping, Interval mapping, Mixed model, WIN QTL Cartographer, QTL Mapper, QTL Network and Quantitative Trait Loci.

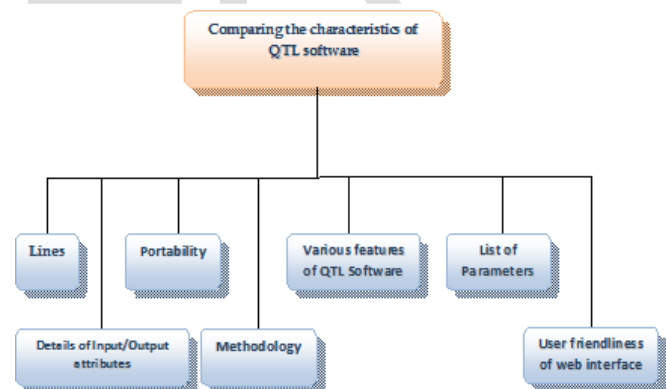
1 INTRODUCTION

Traditional methods of plant breeding have made a significant contribution to crop improvement but they have been slow in targeting complex traits like grain yield, grain quality etc. One method receiving growing attention is the mapping of chromosomal regions affecting qualitative or quantitative traits.

Quantitative trait locus (QTL) analysis is a statistical method that links two type of information - phenotypic data (trait measurements) and genotypic data (usually molecular markers) in an attempt to explain the genetic basis of variation in complex traits. QTL analysis depends on the fact that where such type of linkage occurs, the marker locus and the QTL will not segregate independently and so differences in those marker genotypes will be associated with different trait phenotypes. The basic problem while studying a quantitative trait has always been that the phenotype of a given genotype tells us little about the genotype itself. The discovery of extensive and easily recognizable molecular variation has opened up the possibility of studying individual QTL (Apotikar et al, 2011). Molecular markers are preferred for genotyping because these markers are unlikely to affect the trait of interest.

Then to carry out the QTL analysis, the parental strains are

crossed, resulting in heterozygous (F_1) individuals and these individuals are then crossed using one of a number of different schemes. Finally, the phenotypes and genotypes of the



derived various populations are scored. Both DH and RIL's are better than F_2 and backcross generations and the power of detection of QTL depends upon the heritability of a trait. Over the past decades more than fifty QTL packages were developed. This reflects the importance of QTL tools. This study addresses this critical issue in relation to MSA algorithms by systematically comparing and evaluating three famous QTL analysis software (Figure1).

Fig 1: The characteristics evaluated in QTL Software

• Co-Author name Hemani Sharma
• E-mail: himani6606@gmail.com

2 METHODOLOGY

Data used for analysis:

The data used for QTL analysis consisted molecular marker data of 50 SSR markers used on a subset of 198 Recombinant inbred lines (RIL) of parent variety of HBC19 x NPT2 of rice. The phenotypic data consisted of data for plant height, panicle no./effective tiller and yield of F₇ population of 198 lines derived from HBC19 x NPT2 of rice (Jha ,2007).

From the large number of software available freely, three namely WIN QTL Cartographer, QTL Mapper and QTL Network were chosen for the present study Table 1. These were chosen because each can accommodate a variety of crosses and perform many different types of analyses. WIN QTL Cartographer analyze data with single-marker analysis (SIM), interval mapping (IM), composite interval mapping (CIM), Bayesian interval mapping, multiple interval mapping (MIM), multiple trait analysis and categorical trait analysis. QTL network and QTL Mapper analyze data on mixed linear model and mixed composite model approach.

3 RESULT

The three QTL finding software were compared for their functionality and usability (Table 2). The WIN QTL Cartographer and QTL network were user-friendly for all the three operating system like Windows, Mac and Linux whereas QTL Mapper used only windows operating system. There were also more options for various type of map functions in WIN QTL cartographer. WIN QTL Cartographer also is the only software that can analyze data with several methods (SIM, IM, CIM, Bayesian interval mapping, MIM, Multiple trait analysis, Categorical trait analysis). The output file of QTL cartographer contains estimate of QTL position, measure of statistical significance, %R² (Variance explained), epistatic effects and estimates of additive and dominance effects. This software is able to handle mixture of dominant and codominant markers, discrete traits (e.g threshold traits and disease resistant traits with arbitrary scale) and multivariate analysis. QTL mapper is a command line software for mapping Quantitative Trait Loci (QTLs) with main effects, epistatic effects and QTL X environment interactions. It is based on mixed linear approach. The mixed model based methods are suitable for efficiently searching the QTLs along the whole genome. QTL Mapper 1.0 can not analyze F₂ data sets with design for more than one environment factor and less user friendly. Also Mapping epistatic QTLs with F₂ data is not yet available in QTL Mapper even when the data set is collected from multiple environments. QTL Network is also based on mixed model linear approach but more user friendly than QTL Mapper. It can simultaneously map QTLs with individual effects, epistasis and QTL environment interaction. All the detected QTLs and epistasis are fitted by a full QTL model to estimate the main effect of QTLs and epistasis and their interactions effects with environment by MCMC algorithm. A significant disadvantage is computational burden. High order epistatic is only available in case of F₂ and IF₂ population.

3 DISCUSSION

We compared the three software using the data that consisted of 50 SSR marker study on a subset of 198 recombinant inbred lines (RIL) of Basmati HBC19 x high yielding japonica NPT2 of rice. The phenotypic data consisted of data for plant height, panicle no. /effective tiller and yield of F₇ population of 198 lines derived from HBC19 x NPT2 of rice (Jha ,2007)

A total of 14 QTLs were detected for plant height by QTL Cartographer while QTL Mapper and QTL Network were unable to detect any QTL for plant height. Similarly one QTL was detected for yield and 8 QTLs were detected for panicle no. /effective tiller by QTL Cartographer whereas only 2 QTLs were detected for yield and 3 QTLs were detected for panicle no. / effective tiller by QTL Mapper. QTL Network, however, was unable to detect any QTL for yield and only 2 QTLs were detected for panicle no./effective tiller by QTL Network. The comparison of the QTLs identified for plant height, yield and panicle no. /effective tiller in three software showed that some QTL were common. The SSR markers associated with them were validated with Gramene database. The QTLs present on chromosome no. 2, 3 and 11 for panicle no. /effective tiller was detected by all the three software with LOD ≥ 2.5 and can be considered to be stable while the additional QTLs were considered being suggestive.

Every method of QTL mapping has certain advantages accompanied with some drawbacks. It is, therefore, important that the analysis is performed following different approaches in order not to miss the QTLs, particularly those which explain relatively less phenotypic variation.

(Apotikar et al., 2011) also reported that it is important to use more than one mapping population as well as more than one approaches of QTL mapping to find out as many QTLs as possible for their effective use in the marker-assisted breeding programmes. (Balyan *et al.*, 2005) analyzed the QTLs by QTL Mapper and QTL Cartographer for three grain quality traits in bread wheat using inter-varietal mapping populations. Some common QTLs were detected by both approaches. The common QTLs were considered to be stable and for use in Marker Assisted Selection (MAS).

(Ravi et al, 2011) demonstrated that for complex traits like drought resistance, in addition to simple QTLs, there is a need to identify epistatic QTLs (E-QTLs). For this QTL Network had been used to identify main effect QTLs (M-QTL), epistatic QTLs (E-QTL) and QTL- environment (QE) effects in several crop species. For complex traits that involve QE interaction/epistatic interaction QTL Mapper and QTL Network was found to be the best. However, in case of epistatic effects, best results were obtained only when a larger sample size with data from different locations and within-location replications were entered in the analyses (MCIM). QTLs would only be detected because of the increased statistical power due to a larger sample size.

(Marsan et al, 2001) made a comparison of SIM and CIM models while characterizing QTLs for grain yield and grain related traits in a segregating population of maize. The results indicated that the CIM method employed had greater power in the detection of QTLs, and provided more precise and accurate estimates of QTL positions and effects than SIM. (Nagab-

hushana et al, 2006) did comparative studies on QTL mapping by SIM and CIM for selected growth and yield traits in rice. The QTL Mapmaker is based on SIM and QTL Cartographer was based on CIM approach. But the CIM approach was found to be more accurate and precise. However, some common QTLs were detected by both approaches. QTL Cartographer is based on various methods that increased its accuracy. Table 3 explains some consideration about choosing the proper QTL analysis software.

- chis hypogaea L.)" (2011) *Theor Appl Genet*, Vol. 122, pp.1119-1132.
- [8] WINQTLCartographer2.5
<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>
- [9] QTLNetwork 2.0 <http://ibi.zju.edu.cn/software/qtlnetwork/>
- [10] QTLMapper1.0
<http://ibi.zju.edu.cn/software/qtlmapper/index.htm>

4 CONCLUSION

The aim of this study was to evaluate well-known QTL software used by biologists in order to select proper software which corresponds best to their specific needs. If only one QTL exists, the result from simple single locus association and simple interval mapping analyses are likely to be very similar, with interval mapping having the advantage of giving a convenient graphical output for localizing the QTL. If more than one QTL exists, it makes sense to take account of this using a CIM/MQM approach. Because in practice we may not know how many QTLs there are so use of a statistical software QTL Cartographer that allow implementation of both CIM/MQM and simpler methods is recommended especially based on its user friendly and multiple choice approach. In addition one can use either QTL Mapper or QTL Network in order not to miss the QTLs, particularly those which explain relatively less phenotypic variation and for epistatic estimation and QTL environment interaction effect.

REFERENCES

- [1] D. B. Apotikar, D. Venkateswarlu, R. B. Ghorade, , R. M. Wadaskar, J.V. Patil and P. L. Kulwal "Mapping of shoot fly tolerance loci in sorghum using SSR markers" (2011) *Journal of Genetics*, Vol. 90, pp.59-66.
- [2] H. S. Balyan, P K Gupta, P L Kulwal and N kumar"QTL analysis for three grain quality traits in bread wheat using intervarietal mapping populations" (2005), *Genet. Plant Breed*, Vol. 41, pp.281-283.
- [3] S Jha "Phenotyping and Molecular Marker Analysis of New Plant Type II (*jaпонica*) x Taraori Basmati Derived Recombinant Inbred Lines" (2007), Ph D thesis, CCSHAU, Hisar, India.
- [4] E S Lander and D Botstein "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps"(1989), *Genetics*, Vol. 121, pp.185-199.
- [5] A Marsan, C Gorni, A Chitto, R Redaelli, R Vijk, P Stam and M Motto "Identification of QTLs for grain yield and grain related traits of maize using AFLP map, different testers, and cofactor analysis"(2001) *Theor. Appl. Genet*, Vol. 102, pp.230-243.
- [6] K Nagabhushana, S P Mane and H Shailaja "Comparative studies on QTL mapping by simple interval mapping and composite interval-mapping models for selected growth and yield traits in rice (*Oryza sativa* L.)" (2006) *Indian J. Crop Science*, Vol. 1, pp.97-101.
- [7] K Ravi, V Vadez, S Isobe, R R Mir, Y Guo, S N Nigam, M V C Gowda, T Radhakrishnan, D J Bertioli, S J Knapp and R K Varshney "Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Ara-*

TABLE 1: SOFTWARE SELECTED FOR PRESENT STUDY

S. No.	SOFTWARE	OPERATING SYSTEM	AUTHOR	URL
1	WIN QTL CARTOGRAPHER 2.5	WINDOWS	SHENGCHU WANG, CHRISTOPHER J. BASTEN AND ZHAO-BANG ZENG (2005)	HTTP://STATGEN.NCSU.EDU/QTLCART/WQTLCART.HTM
2	QTLNETWORK 2.0	WINDOWS + WEBSERVICE	JIAN YANG, CHENGCHENG HU, XIUZI YE AND JUN ZHU (2005-2006)	HTTP://IBI.ZJU.EDU.CN/SOFTWARE/QTLNETWORK/
3	QTLMAPPER 1.0	WINDOWS	DAOLONG WANG, JUN ZHU, ZHI-KANG LI, ANDREW H. PATERSON (2003)	HTTP://IBI.ZJU.EDU.CN/SOFTWARE/QTLMAPPER/INDEX.HTM

Table2: Summary of the Comparison of the functionality and usability of the QTL Software

Characters	WINQTL Cartographer	QTLMapper	QTL Network
Input format/Import	Mapmaker/QTL Microsoft excel CSV format QTL Cartographer mcd format	Mapmaker/QTL LRmap.exe and Rcross.exe of QTL Cartographer QTL mapper	Mapmaker/QTL QTL cartographer QTL mapper
Output format	QTL Cartographer .mcd format Microsoft excel text	.one, .two, .qtl, eps, flq, .fle	QTLNetwork QTL Cartographer .qnk
Portability	Windows Macos linux	Windows (command line)	Windows Linux Unix macos
Portability	Windows Macos linux	Windows (command line)	Windows Linux Unix macos
WEB	No	No	Yes
Stand-alone	Yes	Yes	Yes
Result/E*	No	No	yes
Lines	Backcross RIL Selfed Intercross line Randomly mated intercross T(B)SF	Backcross RIL F2 Double Haploid	Backcross RIL F2 IF2 BxFy Double Haploid
Map function	Haldane Kosambi Fixed	Haldane Kosambi	Haldane Kosambi

*The Result field indicates whether the results of a query are obtained instantly

(I) through the web interface, or are sent via e-mail to the user (E).

Table 3: Some consideration in choosing the proper QTL software

Program	Some major Advantages
WINQTL Cartographer	<ul style="list-style-type: none"> ➤ This software is one of the oldest and famous and creditable QTL software. ➤ There are a series of remarkable parameters that are accessible for user to select. ➤ Based on various statistical approaches ➤ Produce a simulation file ➤ View, copy and print chromosome information graphically ➤ We can append, delete and edit source data file (trait, chromosome) ➤ Highly user friendly ➤ Simulation study demonstrated the power of this approach across levels of trait heritability and marker data were sparse. ➤ Can detect eQTL ➤ Can handle variety of input /output format ➤ Easily interpretable output file. ➤ Able to handle dominant and co-dominant markers
QTLMapper	<ul style="list-style-type: none"> ➤ Capable of controlling BGV that increase its precision. ➤ Able to map quantitative trait loci (QTL) with main effects, epistatic effects and QTL x environment interaction
QTL Network	<ul style="list-style-type: none"> ➤ There are a series of remarkable parameters that are accessible for user to select. ➤ Based on linear, mixed linear models having both fixed and random effects. ➤ Fixed Q effects and Random QE effects estimated/predicted with no bias. ➤ Can handle complex effects. ➤ User friendly ➤ Able to predict superior genotype
Program	Some major Disadvantages
WINQTL Cartographer	<ul style="list-style-type: none"> ➤ Did not support NCD3 data type. ➤ Currently cannot perform a complete pairwise Analysis of map segments without regard to already identified additive loci. .
QTLMapper	<ul style="list-style-type: none"> ➤ Not user friendly ➤ Not able to handle so many data types ➤ So many and Highly complicated output file ➤ Cannot analyse F2 data set for more than one environment ➤ Mapping epistatic QTLs with F2 data is not available
QTL Network	<ul style="list-style-type: none"> ➤ cannot append, delete and edit source data file (trait, chromosome) ➤ High order epistasis is available only for F2 and IF2 population ➤ Not able to export in variety of format ➤ Command line version of QTLNetwork CL is only able to analyse data but not able to present the results graphically