BioMercator: integrating genetic maps and QTL towards discovery of candidate genes

Anne Arcade¹, Aymeric Labourdette², Matthieu Falque¹, Brigitte Mangin², Fabien Chardon¹, Alain Charcosset¹, Johann Joets¹

¹UMR INRA UPS-XI INAPG CNRS Génétique Végétale, Ferme du Moulon, 91190 Gif-sur-Yvette - France
²INRA, Biométrie et Intelligence Artificielle, Toulouse, BP27, 31326 Castanet-Tolosan - France

ABSTRACT

Summary: Breeding programs face the challenge of integrating information from genomics and from quantitative trait loci (QTL) analysis, in order to identify genomic sequences controlling the variation of important traits. Despite the development of integrative databases, building a consensus map of genes, QTL and other loci gathered from multiple maps remains a manual and tedious task. Nevertheless, this is a critical step to reveal co-locations between genes and QTL. Another important matter is to determine whether QTL linked to a same trait or related ones, detected in independent experiments and located in a same region, represent a single locus or not. Statistical tools such as meta-analysis can be used to answer this question. BioMercator has been developed to automate map compilation and QTL meta-analysis and to visualize co-locations between genes and QTL through a graphical interface.

Availability: Available upon request (http://moulon/~bioinfo/BioMercator/). Free of charge for academic use.

Contact: joets@moulon.inra.fr

INTRODUCTION

Discovering genes involved in the variation of quantitative traits represents a first step in understanding biological processes underlying these traits. A promising approach consists in
exploiting consensus genetic maps to reveal co-locations between regions of the genome statistically involved in trait variation (i.e. Quantitative Trait Loci or QTL) and candidate genes (Flint and Mott, 2001). The construction of consensus maps must be based on informative loci from many individual maps. Valuable softwares such as CarthaGene (http://www.inra.fr/bia/T/CarthaGene/) or JoinMap (Stam, 1993) can be used to construct a synthetic map from the raw segregation data of several populations. Unfortunately, raw data are seldom available through literature or databases. On the contrary, loci position values are always described, and can be used to merge genetic maps by a homothetic projection process. Indeed, if two genetic maps share a sufficient number of common loci, those loci can be considered as bridges between maps. Thus, projection of the remaining loci, including QTL, from the first map to the other one is possible. Finally, a compiled map can be built from multiple ones by iterative projection.

A QTL, when detected, is described as a most-likely position on a linkage group and a confidence interval around this position. Unfortunately, confidence interval length may represent dozens of centimorgans, while gene density in plant genomes usually vary from 30 to 50 genes per centimorgan (Sasaki et al., 2002; Feng et al., 2002). A single QTL may therefore correspond to many candidate genes. On the other hand, QTL detected in independent experiments and located in a given region of a chromosome may represent several estimations of the position of one single QTL. This hypothesis can be tested by appropriate statistical tools such as meta-analysis, which consists in combining data from diverse sources into a single study. Goffinet and Gerber (2000) have developed a meta-analysis approach suitable for QTL data. This method indicates the most likely number of “real” QTL underlying a pool of QTL of independent experiments. It also provides with consensus positions for these QTL. In most cases, the confidence interval length of the consensus position is reduced when compared to the smallest confidence interval of the initial
QTL. This decreases the number of matching genes and therefore facilitates the identification of relevant positional candidate genes.

We have developed BioMercator (1) to provide users with a tool for map display (2) to automate map compilation and (3) to compute QTL meta-analysis. BioMercator is intended for scientists involved in QTL mapping projects in any kind of organism.

PROGRAM OVERVIEW

Map projection algorithm. The purpose is to project QTL, genes and other loci from a genetic map onto another one, in order to pool all the information on an unique map.

This computation is only based on loci position data. Each genetic map to be loaded in BioMercator is described in a text file created by the user. Data describing QTL (trait names, LOD scores, R²) are loaded from a second file for the purposes of the graphical display. Maps to be projected and projection parameters are chosen in a dialog window. For each pair of homologous chromosomes, common loci (sharing the same name) are listed. A specific distance ratio is then computed for each interval between two common loci. Inverted pairs of loci can be automatically discarded from the list of common loci. Finally, QTL and/or remaining loci positions on the target map are computed by application of the appropriate distance ratio through an homothetic projection process. Users are encouraged to review the resulting map as its relevance is not assessed by the software. Caution in designing the order of maps in case of iterative projection is particularly required. The projection process should start with maps showing the highest quality in loci order, in order to limit error propagation. The software was successfully used to project over 5,000 loci from a public map onto our current working map of maize.
**Meta-analysis algorithm.** The algorithm devised by Goffinet and Gerber (2000) can help to determine if $N$ QTL detected from independent experiments in a same region of a chromosome are consistent with 1-, 2-, 3-, 4- or N-QTL models (the N-QTL model being the case where there are as many "real" QTL as input QTL). For each of these five models, the most likely QTL distribution is determined by means of the maximum likelihood method. Then, an Akaike-type statistical criterion indicates the best model among the five. Meta-analysis computing is based on the position of the QTL and on the variance of this position, which is assessed through confidence interval values (deduced from LOD curves, or from population size and proportion of variance explained by the QTL (Darvasi and Soller, 1997)). The QTL dataset should include 10 to 40 QTL lying within a genome region no longer than 200 cM.

The graphical user interface is used to build the set of QTL to be included in the meta-analysis experiment, and displays the positions of consensus QTL as well as statistical results for each meta-analysis model.

**Graphical user interface.** In addition to functionalities described above, BioMercator interface provides many display options: zoom on each chromosome of the map, adaptable scale for the genetic unit, export of map display to JPEG image format. Several map files can be loaded, and selected for graphical display (Figure 1). BioMercator can be run on Windows, Unix, MacOS X and Linux platforms supporting Java 1.4.

**ACKNOWLEDGEMENTS**

This work was supported by the french genomic initiative Génoplante through its projects ‘Maize Bioinformatics’ and ‘Integrative Tools for Genetic Mapping’.
REFERENCES


Figure 1. General interface of BioMercator. Upper screenshot: loaded maps are added to an arborescence of projects (upper left corner) and displayed in the central panel. The lower screenshot presents the results of a meta-analysis experiment.