MapDisto: fast and efficient computation of genetic linkage maps

Mathias Lorieux

UMR DIADE, Institut de Recherche pour le Développement (IRD), 34394 Montpellier Cedex 5, France.

Address: Rice Genetics and Genomics Laboratory, International Center for Tropical Agriculture (CIAT), AA6713, Cali, Colombia

E-mail: mathias.lorieux@ird.fr — Tel: +57 2 445 00 00 - Fax: +572 445 00 94

ABSTRACT

Motivation: Several options are available to the scientific community for genetic map construction but few are simple to install and use. Either available programs lack intuitive interface, or they are commercial, expensive for many laboratories. We present MapDisto, a free, user-friendly and powerful program to construct genetic maps from experimental segregating populations.

Availability: MapDisto is freely available at http://mapdisto.free.fr/Download_Soft/. Current version: 1.7.5.

Keywords: Genetic mapping – Segregation distortion – Locus ordering algorithms – Molecular markers – Maximum likelihood – Genotyping errors

INTRODUCTION

Several efficient and powerful programs have been developed to construct genetic linkage maps from molecular marker data in experimental segregating populations. One might cite Mapmaker/EXP (Lander et al. 1987), MapManager (Manly et al. 2001), Gmendel (Holloway and Knapp 1994), JoinMap (Stam 1993), MultiPoint (Mester et al. 2003), RECORD (Van Os et al. 2005), CarthaGene (Schiex et al. 1997), MSTmap (Wu et al. 2008). Though, these programs usually don't offer a used-friendly interface. Others are commercial and may not be affordable for many laboratories. Also, multipoint maximum likelihood (MML) locus ordering (eg, Mapmaker/EXP or CarthaGene) are CPU intensive, making the genetic map construction a long and tedious process. With the recent advances in single nucleotide polymorphism (SNP) marker technologies, large mapping datasets can be generated. In this context, availability of software that implement fast ordering algorithms combined with user-friendly interface is crucial. We present MapDisto, a free, intuitive software that provides a series of tools for fast computing (Table 1), verifying and comparing genetic linkage maps. MapDisto can also be used as a teaching tool in classes of theoretical and applied genetics.

FEATURES

1.1 Data handling

Multiple datasets at a time can be handled. MapDisto can import .raw Mapmaker/EXP files and .qdf QGene 4 files (Nelson 1997). It can also export computed maps and data for advanced QTL analysis programs including QGene 4 or Win QTL Cartographer.

Table 1. Comparison of performance between MapDisto 1.7.5 and Mapmaker/EXP 3.0. Platform: Intel i5 @ 3.6 GHz– 4 GB 1333 MHz RAM running Windows XP. MapDisto version: 1.7.5.1 for Excel 2010. Data: simulated BC1 population of 100 individuals, chromosome of 50-200 cM with markers evenly dispersed every cM. Time is expressed in seconds. *m* indicates the number of markers simulated.

	Ordering a sequence	Ripple	Bootstrap
m=50			
MapDisto	< 0.5	0.8	9
Mapmaker	14	16	N/A**
m=100			
MapDisto	0.4	1.5	40
Mapmaker	150	47	N/A
m=200			
MapDisto	4	9	312
Mapmaker	E*	E	N/A

E*: Unreachable on our platform (Mapmaker crashed)
N/A**: Mapmaker doesn't perform boostrap resampling.

1.2 Population types supported; maximum dataset size

The present version of MapDisto handles F2 backcross (BC1), F2 intercross (F2), doubled haploid (DH), single-seed descent (SSD) and Highly Recombinant Lines (HRIL) populations. MapDisto can handle large datasets, typically 1,000s of loci (*n*) x 100s of individuals (*m*). Maximum dataset size (MDS) actually depends on the computer's available memory. For example, we could analyze a simulated dataset made of 500 BC1 individuals scored by 15,000 markers on a 4 GB MacOS X computer.

1.3 Map construction

1.3.1 Building linkage groups

The search for linkage groups (LGs) is done in a similar way to Mapmaker, in specifying a minimum LOD score and a maximum recombination frequency between linked pairs of markers.

1.3.2 Determining marker orders

MapDisto can automatically order loci on LGs. The process is very fast – typically a few seconds per LG – compared to MML algorithms; several ordering algorithms (Branch and bound II, Seriation II, Unidirectional growth) and criteria (SARF, SAD,

COUNT, Salod) can be chosen (See Wu et al. 2011 for a review of ordering methods). We could verify that MapDisto and Mapmaker/EXP give very similar results, using true and simulated datasets and different combinations of algorithms and criteria. Automated commands like "Ripple" and "Check inversions" are implemented. These two commands check local stability of the map using two different algorithms. "Ripple" tests all possible orders of a sliding window of 5 loci along the LG, while "Check inversions" tests nested inversions of chromosome blocks in order to detect possible errors due to large gaps in the map that can lead to macro-inversions that are not detectable by local rippling. "Detailed map" allows users to compute several map parameters for a specified LG and order: classical or corrected recombination fractions, map distances and their associated standard deviations, linkage and independence chi^{2}s, LOD scores and population size for each interval. Finally, one can compute tables of recombination fractions (various estimates), map distances, linkage and independence chi^{2}s, and two-point LOD scores for all pairs of loci in a linkage group.

1.3.3 Graphical maps

Once computed, the map can be graphically displayed, either drawing the LGs one by one or all LGs at once.

One can also draw genetic maps imported from other mapping programs.

1.4 Mapmaker interfacing

From the beta version (1.7.5 b) of MapDisto, it is possible to automatically send data and commands to Mapmaker/EXP, and to retrieve its outputs. A previous installation of Mapmaker/EXP is required. This function is useful to compare MapDisto results to MML computations (Fig. 1).

1.5 Testing map robustness

A nice way of evaluating the stability or robustness of a given order is to use resampling methods. The user can choose an LG and perform a bootstrapping of orders. This allows the identification of problematic loci and regions that are difficult to order.

1.6 Detecting genotyping errors

A way to identify potential genotyping errors is to look at double recombinants with low probability of occurrence. This is the purpose of the "Color genotypes" window that displays error candidates with a chosen threshold. The user can then choose to replace candidate errors by missing data or by inferred data from flanking genotypes. The new dataset can then be used to recompute the map. An automated algorithm to iterate these operations with a varying threshold value until no more error is detected is available.

1.7 Handling of segregation distortion

The program allows computing segregation chi^{2}s that measure the deviation of allelic or genotypic frequencies from Mendelian expectations, along with their associated probabilities for the loci of an LG or all loci at once. Various recombination fractions estimates are computed: the classical one (Allard 1956), Bailey's estimate (Bailey 1949) and a corrected estimate that handles for selection against any genotypic class of the progeny (Lorieux et al. 1995a, Lorieux et al. 1995b).

1.8 Comparing maps

It is often useful to compare the results of different ordering algorithms, or to compare one's map with results from an external program. MapDisto displays graphical comparisons of two genetic maps, allowing quick identification of inconsistencies. MapDisto can also extract the positions of markers on a physical map

and compare the computed genetic map to the true order in each LG. This is particularly helpful to increase the power of searching for genotyping errors.

1.9 Simulating populations

A simulation module allows users to simulate BC1 or F2 populations and to include markers exhibiting segregation distortion (SD). This module is useful for teaching purposes, e.g. to evidence the differences of mapping resolution between different population types. It can also be used to test the effect of SD on the map (Garavito et al. 2010).

1.10 QTL search

A simple one-way ANOVA to search for quantitative trait locus (QTL) to marker associations is implemented. With saturated maps, ANOVA-1 is almost as powerful as Interval mapping (Rebai et al. 1995). A graphical output of QTLs on the genetic map can be browsed trait by trait.

1.11 Compatibility

MapDisto is written in Microsoft Visual Basic for Applications and runs within Microsoft Excel spreadsheet (versions 2003 to 2011). It is therefore compatible with the two main platforms generally used by geneticists, Microsoft Windows and Mac OS X.

1.12 Tutorial

A tutorial is available for download at the MapDisto Website. Although most of the commands are straightforward to use, users are encouraged to read it in order to fully take advantage of the program's features.

REFERENCES

Allard RW (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235-278

Bailey NTJ (1949) The estimation of linkage with differential viability, II and III. Heredity 3:220-228

Garavito A, Guyot R, Lozano J, Gavory F, Samain S, Panaud O, Tohme J, Ghesquiere A, Lorieux M (2010) A genetic model for the female sterility barrier between Asian and African cultivated rice species. Genetics 185 (4):1425-1440.

Holloway J and Knapp SJ (1994) Gmendel 3.0 Users Guide. (http://www.css.orst.edu/G-Mendel/Default.htm)

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Lorieux M, Goffinet B, Perrier X, González de León D, Lanaud C (1995a) Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. Theor Appl Genet 90:73-80

Lorieux M, Perrier X, Goffinet B, Lanaud C, González de León D (1995b) Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F_2 populations. Theor Appl Genet 90:81-89

Manly KF, Cudmore RH, Meer JM (2001) Map manager QTX, cross-platform software for genetic mapping. Mamm Genome 12 (12):930-932

Mester DI, Ronin YI, Minkov D, Nevo E, Korol AB (2003) Constructing large scale genetic maps using an evolutionary strategy algorithm. Genetics 165:2269–2282

Nelson JC (1997) QGene: Software for marker-based genomic analysis and breeding. Mol Breed 3 (3):239-245

Rebai A, Goffinet B, Mangin B (1995) Comparing Power of Different Methods for Qtl Detection. Biometrics 51 (1):87-99

Schiex T and Gaspin C (1997) CARTHAGENE: Constructing and joining maximum likelihood genetic maps. In: Proceeding of ISMB (1997), pp. 258–267

Stam P (1993) Construction of Integrated Genetic-Linkage Maps by Means of a New Computer Package - Joinmap. Plant Journal 3 (5):739-744

Van Os H, Stam P, Visser RGF, Van Eck HJ (2005) RECORD: a novel method for ordering loci on a genetic linkage map. Theor Appl Genet 112 (1):30-40. doi:Doi 10.1007/S00122-005-0097-X

Wu JX, Jenkins JN, McCarty JC, Lou XY (2011) Comparisons of four approximation algorithms for large-scale linkage map construction. Theor Appl Genet 123 (4):649-655. doi:Doi 10.1007/S00122-011-1614-8

Wu YH, Bhat P, Close TJ, Lonardi S (2008) Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph. PLoS Genetics vol. 4 (10) pp. e1000212

Figure 1. Comparison between MapDisto (left) and Mapmaker/EXP (right) maps obtained from a simulated data set — an F2 population of 100 individuals, three chromosomes and 200 markers—, showing the similarity of locus orders and map distances.