MAP-O-MAT: internet-based linkage mapping

X. Kong and T. C. Matise*

Department of Genetics, Rutgers University Piscataway, NJ 08854, USA

Received on August 3, 2004; accepted on September 7, 2004
Advance Access publication September 16, 2004

ABSTRACT

Summary: MAP-O-MAT is a web-based server for automated linkage mapping of human polymorphic DNA markers. MAP-O-MAT facilitates the verification of order and map distances for custom mapping sets using genotype data from the CEPH database, and from the Marshfield, SNP Consortium and Rutgers linkage maps (exclusive to the deCODE genotyping data). The CRI-MAP program is used for likelihood calculations and some mapping algorithms, and physical map positions are provided from the human genome assembly.

Availability: MAP-O-MAT is located at http://compgen.rutgers.edu/mapomat/

Contact: matise@biology.rutgers.edu

INTRODUCTION

Genetic linkage maps are the foundation of both linkage and linkage disequilibrium studies for mapping disease genes. Over 16,000 polymorphic markers have been genotyped in reference pedigrees suitable for map construction, and numerous genome-wide linkage maps have been published over the past two decades (NIH/CEPH Collaborative Mapping Group, 1992; Matise et al., 1994; Murray et al., 1994; Dib et al., 1996; Broman et al., 1998; Kong et al., 2002). The availability of genome-wide physical mapping and sequencing data from the Human Genome Project has increased the accuracy of the more recently published maps. However, any single linkage map contains only a subset of the available polymorphic markers. In addition, often only a portion of the mapped markers are localized to single map positions with high statistical support while others may only be localized to larger intervals. Therefore, it can be a difficult and tedious exercise to determine the order and map distances for any specific set of polymorphic markers. Frequently, users performing linkage/association studies may be interested in the answers to questions such as: What are the map distances for a specific ordered map of markers? What is the statistical support for a specific ordered map of markers? What is the position of a given marker on a specific ordered map of markers? Where does a marker of interest lie with respect to the Rutgers or deCODE linkage maps?

*To whom correspondence should be addressed.

SYSTEMS AND METHODS

We have developed a web-based server (MAP-O-MAT) for automated linkage mapping of polymorphic DNA markers. We imported a comprehensive set of marker genotype data from three sources: the CEPH database version 9.0 (Dausset et al., 1990), the Center for Medical Genetics in Marshfield, WI (Broman et al., 1998) and The SNP Consortium (Matise et al., 2003). The genotypes from these sources were determined using the CEPH standard reference pedigrees (Dausset et al., 1990). To our knowledge, our combined collection of genotype data is the largest of its type.

We initially identified genotype data for over 16,000 markers. However, after careful screening of marker names and aliases, primer sequences, two-point linkage mapping results, and physical positions to determine markers not previously identified as replicated copies, our final working set contains 15,624 unique markers. This set consists mainly of short tandem repeat markers (61%) but also contains SNPs (26%) and RFLP/VNTR and other hybridization-based markers (9%), plus a small set of markers whose type could not be determined (4%). These markers have an average of 219 informative meioses.

The CRI-MAP program (Lander and Green, 1987) was integrated into our system to perform likelihood calculations and most of the mapping algorithms. MAP-O-MAT can be set to interface with either all of the markers or with only the SNP markers. Note: the genotype data themselves are not distributed by MAP-O-MAT, they are only available for computation by the MAP-O-MAT server. Users wishing to access the genotype data must obtain them directly from their sources.

MAP-O-MAT provides the following functions:

FIXED: Given an ordered map of markers, the FIXED function (implemented in CRI-MAP) estimates the inter-marker recombination fractions and map distances. Users choose whether to compute a sex-averaged or sex-specific map, and whether to apply the Haldane or Kosambi map function. There is also an option to perform a simple FLIPS analysis (see below) and include the FLIPS results in the table of the FIXED results. For each marker on the map, the results also include the physical location from the NCBI
FLIPS: Given a map of markers in putative order, the FLIPS function (implemented in CRI-MAP) computes the statistical support for local map order. This is done by computing the difference in log₁₀ likelihood (equivalent to a lod score) between the original marker order and numerous alternate marker orders. The alternate marker orders are determined by specifying the tuple size, from two to a maximum of six, and the marker order is then altered by permuting all possible orders of the markers in each successive tuple while retaining the original order of the rest of the map. For example, if there are four markers on a map (A-B-C-D) and the tuple size is three, then the nine alternate marker orders for which likelihoods will be calculated are: A-C-B-D, B-A-C-D, B-C-A-D, C-A-B-D, C-B-A-D, A-B-D-C, A-C-D-B, A-D-B-C, A-D-C-B. For larger tuple sizes, the number of alternate orders greatly increases.

ALL: The ALL function is used to locate markers on a specified map. Given an ordered set of markers, an unmapped marker is localized on the map using the ALL function (implemented by CRI-MAP) by placing the marker into each map interval and computing the multipoint likelihood. The interval with the highest likelihood is the most likely location of the unmapped marker, with the difference between the likelihood of the best position and the second-best position providing a lod score that can be used to assess the statistical confidence of the localization. Typically, a lod score of at least three is required to be confident of a marker placement, indicating that the marker position is at least 1000 times more likely than any other.

View/query Rutgers maps: This function allows a user to view or query the combined linkage-physical maps we have recently constructed [http://compgen.rutgers.edu/maps, X.Kong, K.Murphy, T.Raj, C.He, P.S.White and T.C.Matise (submitted for publication)]. These maps were constructed using both meiotic recombination data as well as physical mapping data to place almost 15,000 polymorphic markers onto a single map. The user can view our map of each chromosome in pre-formatted HTML tables, and the user can query the map to check the position of any of the marker.

decODE: The decODE function is used to position markers on the decODE linkage map (Kong et al., 2002). It is identical to the ALL function, except that the ordered set of markers onto which the unmapped marker is placed is the decODE linkage map rather than a user-specified map. In this way, the position of markers not on the decODE map can be readily determined. Since the decODE linkage map is a fixed ordered set of markers, we have pre-computed the position of all markers not already on the decODE map. MAP-O-MAT accesses this database of pre-computed positions for a very quick display of results.

Multiple map comparison: This function allows the user to directly compare the likelihood and statistical support of several different orders of the same set of markers.

REMAP: This function provides another approach to evaluating the local support for order of a map. Given an ordered set of markers, each marker is successively removed from the map and then run through the ALL function to evaluate its position on the remaining map. Ideally, on a well-ordered and statistically well-supported map, each marker would map back to its putative map position with high statistical support.

Marker information: The marker information tables provide marker-specific information not easily found elsewhere, which include observed heterozygosity, the number of families genotyped and the number of meioses available for each marker.

IMPLEMENTATION

MAP-O-MAT is accessed remotely via the web. It uses Java servlet technology to allow platform-independent computation, and is running on a Linux 8.0 server. For each analysis, the user fills out a simple query form, the servlet prepares data for input to the CRI-MAP program, CRI-MAP is invoked, and the results from CRI-MAP are then parsed into easy-to-read results presented back to the user as web-pages. Most computations are very quick and take only seconds to run.

ACKNOWLEDGEMENTS

We gratefully acknowledge the input of Peter Reiner and Joshua Gitlin in the development of MAP-O-MAT, Suzanne Leal for helpful discussions, use of the CRI-MAP program by Phil Green, and generation and/or curation of the genotype data by the CEPH database, the Marshfield Center for Medical Genetics and the TSC SNP database. Development and maintenance of MAP-O-MAT was supported in part by NIH grants R29HG01691 and HG00008.

REFERENCES


