BIOINFORMATICS APPLICATIONS NOTE Vol. 21 no. 2 2005, pages 263–265
doi:10.1093/bioinformatics/bth457

Haploview: analysis and visualization of LD and haplotype maps

J. C. Barrett[1,*], B. Fry[2], J. Maller[1] and M. J. Daly[1,3]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA, [2]MIT Media Lab, Cambridge, MA 02139, USA and [3]Broad Institute of Harvard and MIT, Cambridge, MA, USA

Received on June 23, 2004; revised and accepted on July 23, 2004
Advance Access publication August 5, 2004

ABSTRACT

Summary: Research over the last few years has revealed significant haplotype structure in the human genome. The characterization of these patterns, particularly in the context of medical genetic association studies, is becoming a routine research activity. Haploview is a software package that provides computation of linkage disequilibrium statistics and population haplotype patterns from primary genotype data in a visually appealing and interactive interface.

Availability: http://www.broad.mit.edu/mpg/haploview/

Contact: jcbarret@broad.mit.edu

INTRODUCTION

Knowledge of local linkage disequilibrium (LD) and common haplotype patterns in disease association and positional cloning studies is becoming increasingly widespread since it has become clear (Van Eerdewegh *et al*., 2002; Rioux *et al*., 2001; Geesaman *et al*., 2003; Stoll *et al*., 2004) that intelligent use of this information has the potential to make them much more comprehensive and efficient. Early studies identifying unexpected extent of correlation and structure in haplotype patterns (Reich *et al*., 2001; Daly *et al*., 2001; Gabriel *et al*., 2002) have led to the initiation of the Human Haplotype Map project (HapMap) to make this information available to all medical genetics researchers (International HapMap Consortium, 2003). Given the dramatic increase in the size and number of disease association studies worldwide and the enormous amount of public genotype data from HapMap, tools for analyzing, interpreting and visualizing these data are of critical importance to researchers everywhere.

Haploview is designed to provide a comprehensive suite of tools for haplotype analysis for a wide variety of dataset sizes. Haploview generates marker quality statistics, LD information, haplotype blocks, population haplotype frequencies and single marker association statistics in a user-friendly format. All the features are customizable and all computations performed in real time, even for datasets with hundreds of individuals and hundreds of markers.

FEATURES

Haploview accepts input in a variety of formats. Pedigree data can be loaded as either partially or fully phased chromosomes or as unphased diplotypes in the standard Linkage format. The latter format also allows the user to specify family structure information as well as disease affection or case/control status. Marker information, including name and location is loaded separately. Haploview also directly accepts genotype data dumped from the Human HapMap website (http://www.hapmap.org). A graphical genome browser maintained at that site allows researchers to navigate to a particular region of the genome and dump HapMap genotype data for all genotyped markers in the selected region in a format accepted by Haploview.

Upon loading a dataset, the software presents to the user a series of marker genotyping quality metrics. These include a check for conformance with Hardy–Weinberg equilibrium, a tally of Mendelian inheritance errors and the percentage of individuals successfully genotyped for that marker. The program filters out markers which fall below a preset threshold for these tests. The user can adjust these thresholds as well as handpick markers to add or remove from the subsequent steps. At any time later in the process, the user may return to this quality control panel, add or remove additional markers, and have the changes immediately reflected in the ongoing analyses.

Haploview calculates several pairwise measures of LD, which it uses to create a graphical representation (Fig. 1). The user has the option to select one of several commonly used block definitions (Gabriel *et al*., 2002; Wang *et al*., 2002) to partition the region into segments of strong LD. Alternatively, the user may manually select groups of markers for subsequent haplotype analysis. This view also allows a number of different color schemes to represent the LD relationships.

*To whom correspondence should be addressed.

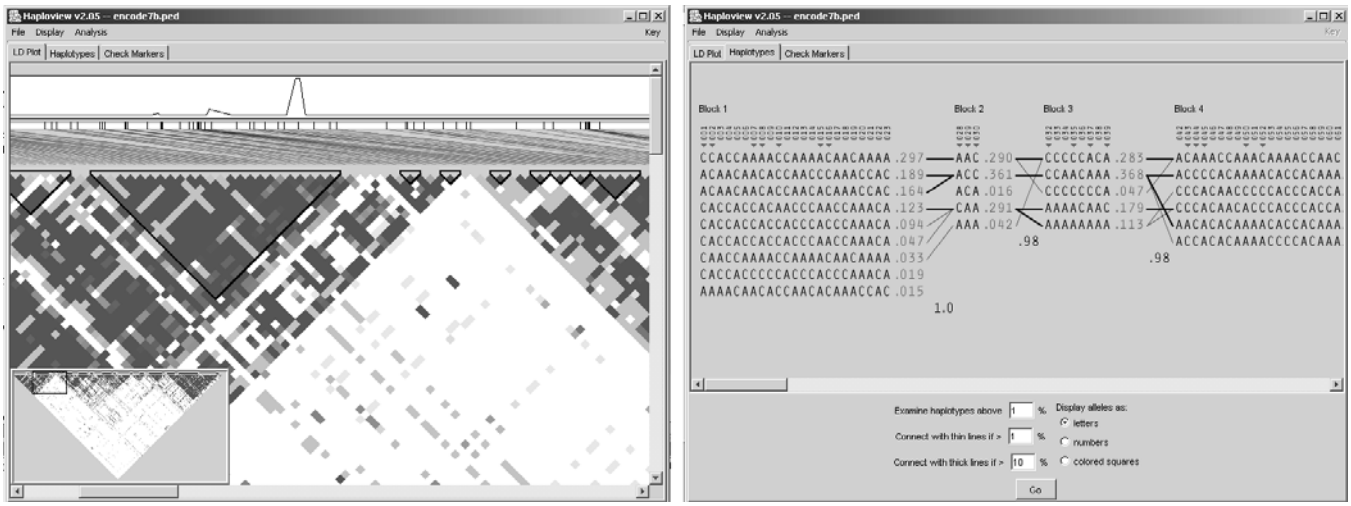

Fig. 1. Haploview LD display with recombination rate plotted above (left) and haplotypes display (right). Interface developed at MIT Media Lab by B.Fry (http://acg.media.mit.edu/people/fry/).

Further, the program allows the display of an 'analysis track' above the LD plot, to display continuous variables such as recombination rate estimates (McVean *et al*., 2004) (Fig. 1).

Once groups of markers are selected (either automatically or manually), the program generates haplotypes and their population frequencies (Fig. 1). This display shows lines to indicate transitions from one block to the next with frequency corresponding to the thickness of the line and also presents Hedrick's multiallelic D′, which represents the degree of LD between two blocks, treating each haplotype within a block as an 'allele' of that region. Again, customization is available for nearly all aspects of the display, including displaying alleles as letters, numbers or colored boxes and displaying only those haplotypes above an adjustable threshold in the population.

If affection status is included in the input file, Haploview also calculates the standard TDT statistic (for trio data) or simple χ^2 (for case/control data) for each marker that can be used for association studies. Future versions will include several haplotype-tag SNP selection methods as well as haplotype-based association testing and evaluation of significance using permutation testing. These final features allow the user to go from raw genotype data through exploring genetic associations in one easy to use software package. Haploview is maintained as an open source project (http://sourceforge.net/projects/haploview/), which allows external parties to add their own methods in addition to the continuing development by the authors.

Each of these views of the data is shown on a separate tab (Fig. 1), allowing the user to move from one to the next, with interactive modifications made by the user in any panel reflected in all the others. For example, one can return at any time to the review of marker quality and manually include or exclude individual markers—these changes are instantly reflected in the LD and haplotype panels. This provides the ability to analyze the data in real-time. The information on each panel is also able to be exported to a PNG for use in presentations or publications or dumped to a text file. Additionally, the program has a fully functional command-line mode, which allows users to run all the analyses without opening the GUI on one or more files at once.

IMPLEMENTATION

Haploview is written entirely in Java, which means it is usable on any platform with Java 1.3 or later installed. Running on a 1.8 GHz Pentium 4 with 1 GB of RAM, Haploview can display a dataset with 200 markers genotyped in 400 individuals and adjust parameters with no noticeable delay. The program is also able to be used (from the command line) to do the LD calculations on very large datasets in comparatively small amounts of time. Haploview was able to compute 3.3 million pairwise LD values (comparisons of all markers closer than 500 KB in a 45 500 marker dataset) in 30 min.

Haploview uses a two marker EM (ignoring missing data) to estimate the maximum-likelihood values of the four gamete frequencies, from which the D′, LOD and r^2 calculations derive. Haplotype phase and population frequency are inferred using a standard EM algorithm with a partition–ligation approach for blocks with greater than 10 markers. Conformance with Hardy–Weinberg equilibrium is computed using an exact test (G.Abecasis and J.Wigginton, personal communication).

REFERENCES

Daly,M.J. Rioux,J.D., Schaffner,S.F., Hudson,T.J. and Lander,E.S. (2001) High-resolution haplotype structure in the human genome. *Nat. Genet.*, **29**, 229–232.

Gabriel,S.B. Schaffner,S.F., Nguyen,H., Moore,J.M., Roy,J., Blumenstiel,B., Higgins,J., Defelice,M., Lochner,A., Faggart,M. *et al*. (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.

Geesaman,B.J., Benson,E., Brewster,S.J., Kunkel,L.M., Blanche,H., Thomas,G., Perls,T.T., Daly,M.J. and Puca,A.A. (2003) Haplotype-based identification of a microsomal transfer protein marker associated with the human lifespan. *Proc. Natl Acad. Sci., USA*, **100**, 14115–20.

The International HapMap Consortium (2003) The International HapMap Project. *Nature*, **18**, 789–796.

McVean,G.A., Myers,S.R., Hunt,S., Deloukas,P., Bentley,D.R. and Donnelly,P. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581–584.

Reich,D.E., Cargill,M., Bolk,S., Ireland,J., Sabeti,P.C., Richter,D.J., Lavery,T., Kouyoumjian,R., Farhadian,S.F., Ward,R. and Lander,E.S. (2001) Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.

Rioux,J.D., Daly,M.J., Silverberg,M.S., Lindblad,K., Steinhart,H., Cohen,Z., Delmonte,T., Kocher,K., Miller,K., Guschwan,S. *et al*. (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.*, **29**, 223–228.

Stoll,M., Corneliussen,B., Costello,C.M., Waetzig,G.H., Mellgard,B., Kroch,W.A., Rosenstiel,P., Albrecht,M., Croucher,P.J., Seegert,D. *et al*. (2004) Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat. Genet.*, **36**, 476–480.

Van Eerdewegh,P., Little,R.D., Dupuis,J., Del Mastro,R.G., Falls,K., Simon,J., Jorrey,D., Pandit,S., McKenny,J., Braunschweiger,K. *et al*. (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature*, **418**, 426–430.

Wang,N., Akey,J.M., Zhang,K., Chakraborty,R. and Jin,L. (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.*, **71**, 1227–1234.