BIOINFORMATICS APPLICATIONS NOTE Vol. 21 no. 8 2005, pages 1730–1732
doi:10.1093/bioinformatics/bth488

Genetics and population analysis

HaploPainter: a tool for drawing pedigrees with complex haplotypes

Holger Thiele[1,3,*] and Peter Nürnberg[1,2,3]

[1]Gene Mapping Center, Max Delbrueck Center (MDC) for Molecular Medicine, Berlin-Buch, Germany, [2]Institute of Medical Genetics, Charité—University Medicine of Berlin, Berlin, Germany and [3] Cologne Center for Genomics (CCG), University of Cologne, Germany

Received on July 6, 2004; revised and accepted on August 16, 2004
Advance Access publication September 17, 2004

ABSTRACT

Summary: HaploPainter is a user-friendly pedigree-drawing application with special features for easy visualization of complex haplotype information. It has been developed to facilitate gene mapping in Mendelian diseases in terms of fast and reliable definition of the smallest critical interval harbouring the underlying gene defect. HaploPainter is written in Perl and may be used for visualization of haplotypes calculated by any of the common linkage programs. With special features like haplotype compression or the ability of marker section cut-out it particularly addresses the requirements for viewing large haplotypes as obtained by using for genome scans high-density marker panels of many thousands of single nucleotide polymorphisms (SNPs).

Availability: http://haplopainter.sourceforge.net/

Contact: holger.thiele@uni-koeln.de

Construction and visualization of haplotypes is an essential step in gene mapping projects. It is often instrumental in choosing the right region for further follow-up. Moreover, it defines precisely the size of the critical interval by pinpointing the location of the recombination breakpoints in the families under investigation. While a graphical presentation of haplotypes is very useful for interpretation and publication of genotyping data, there is only inadequate software support available. Existing solutions as implemented in Cyrillic (Chapman, 1990), Pedigree/Draw (Mamelka *et al*., 1990), CoPE (Brun-Samarcq *et al*., 1999) or Pelican (Dudbridge *et al*., 2004) either suffer from lack of user-friendliness, limited data compatibility, or insufficient drawing alternatives for haplotypes.

Recently, we performed several genome scans with the Affymetrix GeneChip® Human Mapping 10K SNP (single nucleotide polymorphism) array (Janecke *et al*., 2004; Kaindl *et al*., 2004; Uhlenberg *et al*., 2004). The new approach, using >10 000 SNPs instead of about 400 microsatellites as DNA markers in genome-wide linkage studies, clearly revealed the shortcomings of the existing solutions for haplotyping. Dealing with very long haplotypes composed of hundreds of SNPs was impossible with any existing software and turned out to be a time consuming bottleneck in our data analysis. This prompted us to develop HaploPainter as a user-friendly tool for the handling of haplotype information in extended pedigrees. In particular, we aimed at a clearly arranged presentation of large marker blocks and have therefore implemented features like selective definition and narrowing of marker regions or haplotype compression at a chosen length.

The program is written in Perl/Tk, running under Windows and Linux. The algorithms are all oriented at practical considerations. The majority of families should be drawable without extensive computational time. Every family is represented in a single multidimensional array data structure which is built up following a top-to-bottom strategy. In case of loops the order of sibs is randomly chosen, keeping loop-starting and connecting family members close to each other. Further optimization is performed by minimization of line crossings. From this point different pedigree drawing solutions may be found. Although the majority of simple and moderately complex pedigrees were drawn correctly, there are limitations, for example, when a person occurs in different generations, that is, in the typical 'backcross' situation of animal breeds. This can be avoided by allowing for person duplications. We will implement this feature into future versions of HaploPainter.

HaploPainter accepts haplotype outputs from Simwalk (Weeks *et al*., 1995), Allegro (Gudbjartsson *et al*., 2000), Genehunter (Kruglyak *et al*., 1996) and Merlin (Abecasis *et al*., 2002). Any supplementary information provided by programs like Simwalk is discarded at this stage. Points of recombination are recognized using HaploPainter's own algorithm. Starting at the p-telomer, the program identifies the first informative marker revealing that linkage phase has changed. Pedigree data are imported either in a pre- or post-makeped format as provided by standard linkage files. Map information can be easily added in a file format used by Mega2 [Mukhopadhyay *et al*. (2001) in http://www.rfcgr.mrc.ac.uk/Menu/Help/mega2/]. Up to three lines of case information may be attached to each person using an excel sheet-like table format. After drawing of the pedigree, the following modifications are possible. Symbols can be moved per drag-and-drop and haplotype phases can be switched by double clicking on uninformative marker alleles. Many different drawing styles are selectable from the configuration menu. The output from HaploPainter can be similar to what is shown in Figure 1. The graphic is directly printable or can be exported as a postscript file.

In conclusion, there is a growing need for the support of haplotype drawing in pedigrees. Many programs suitable for haplotype

*To whom correspondence should be addressed.

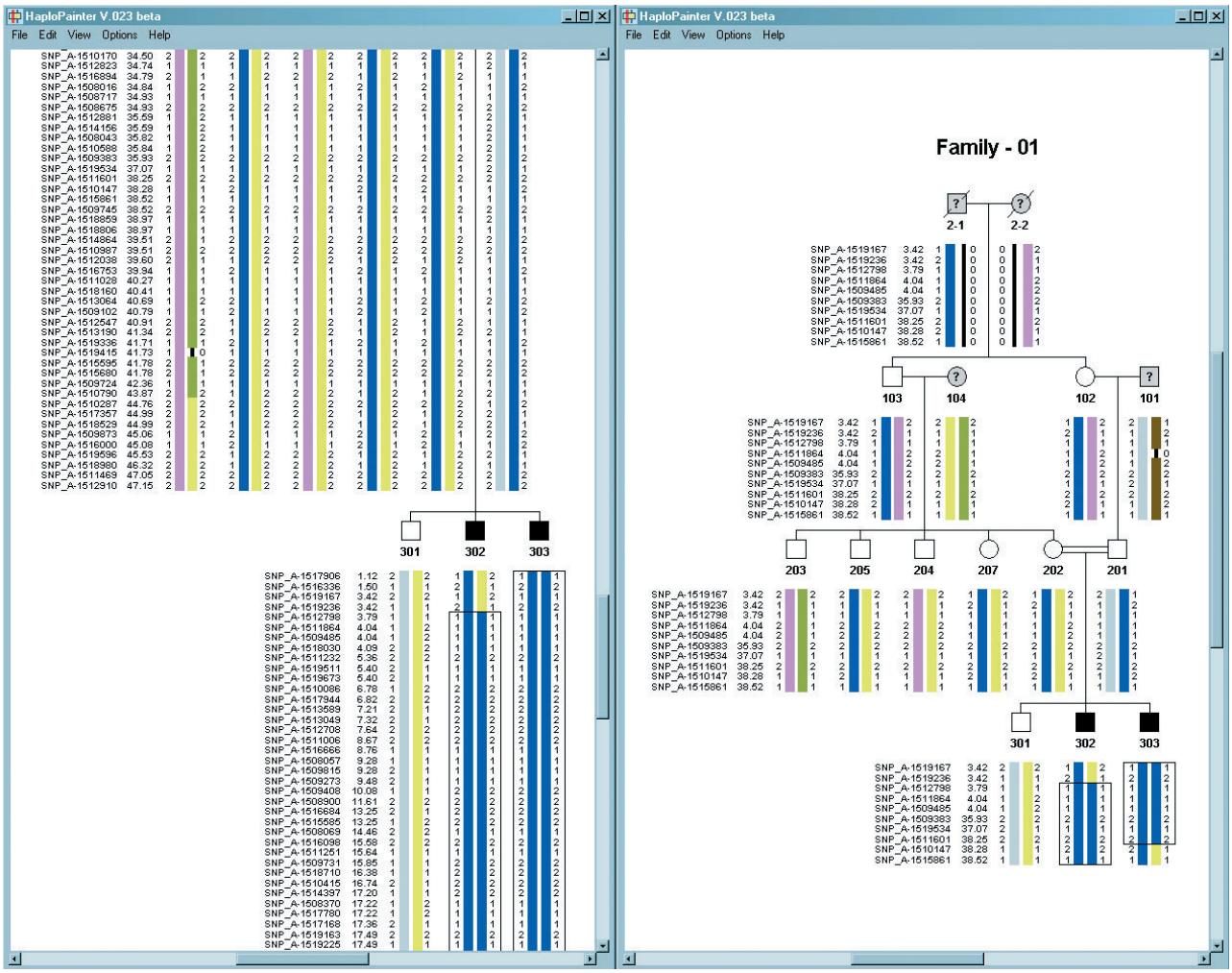

Fig. 1. Using the Affymetrix GeneChip® Human Mapping 10K SNP array more than 10 000 SNPs were genotyped in 13 family members of a consanguineous pedigree. Data conversion, handling and haplotype calculation were performed with the programs ALOHOMORA (Ruschendorf and Nürnberg, 2005) and Genehunter-Imprinting (Strauch *et al*., 2000). The left window demonstrates HaploPainter's ability to draw even very long haplotypes—composed of 100 SNPs in this case. Cutting out most of the homozygous region between the critical recombination events results in a clearly arranged graphic with all necessary information as shown in the right window. Filled symbols represent affected individuals and open symbols non-affected ones. Persons with an unknown affected status are shown in grey with a question mark inside. A diagonal slash indicates that the person is deceased. Marker names and positions are displayed on the left side of each generation. The crucial recombinations occurred in individuals 302 and 303. Their disease haplotype fragments are boxed.

calculation like Genehunter, Merlin, Simwalk or Allegro are insufficient in proper visual presentation. We offer a platform-independent, Perl-based solution, HaploPainter, as an open source software for the scientific community to fill this gap. Features like haplotype compression and the ability of marker section cut-out are particularly helpful for viewing large SNP-derived haplotypes. The software is equipped with an intuitive graphical interface and is powerful enough to draw even complex consanguineous pedigrees. It is addressed to geneticists and physicians handling human pedigrees with the necessity of graphical haplotype representation.

ACKNOWLEDGEMENTS

We would like to thank our colleagues and guests of the Gene Mapping Center at the MDC for advice and software testing. This work was funded by the Federal Ministry of Science and Education of Germany through the National Genome Research Network.

REFERENCES

Abecasis,G.R., Cherny,S.S., Cookson,W.O. and Cardon,L.R. (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.*, **30**, 97–101.

Brun-Samarcq,L., Gallina,S., Philippi,A., Demenais,F., Vaysseix,G. and Barillot,E. (1999) CoPE: a collaborative pedigree drawing environment. *Bioinformatics*, **15**, 345–346.

Chapman,C.J. (1990) A visual interface to computer programs for linkage analysis. *Am. J. Med. Genet.*, **36**, 155–160.

Dudbridge,F., Carver,T. and Williams,G.W. (2004) Pelican: pedigree editor for linkage computer analysis, *Bioinformatics*, **20**, 2327–2328.

Gudbjartsson,D.F., Jonasson,K., Frigge,M.L. and Kong,A. (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.*, **25**, 12–13.

Janecke,A.R., Thompson,D.A., Utermann,G., Becker,C., Hübner,C.A., Schmid,E., McHenry,C.L., Nair,A.R., Rüschendorf,F., Heckenlively,J. *et al*. (2004) Mutations in

RDH12 encoding a photoreceptor cell retinol dehydrogenase cause childhood-onset severe retinal dystrophy. *Nat. Genet.*, **36**, 850–854.

Kaindl,A.M., Rüschendorf,F., Krause,S., Goebel,H.H., Koehler,K., Becker,C., Pongratz,D., Müller-Höcker,J., Nürnberg,P., Stoltenburg-Didinger,G., Huebner,A. (2004) Missense mutations of *ACTA1* cause dominant congenital myopathy with cores. *J. Med. Genet.*, **41**, 842–848.

Kruglyak,L., Daly,M.J., Reeve-Daly,M.P. and Lander,E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347–1363.

Mamelka,P.M., Dyke,B. and MacCluer,J.W. (1990) Pedigree/Draw for the Apple Macintosh, Version 4.0. PGL Technical Report No. 1, Pop. Genetic Laboratory, Department of Genetics, South-west Foundation for Biomedical Research, San Antonio, Texas 78284, USA.

Ruschendorf,F. and Nürnberg,P. (2005) ALOHOMORA: A tool for linkage analysis using 10K SNP array data. *Bioinformatics*, Jan 12. [Epub ahead of print] doi:10.1093/bioinformatics/bti264

Strauch,K., Fimmers,R., Kurz,T., Deichmann,K.A., Wienker,T.F. and Baur,M.P. (2000) Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. *Am. J. Hum. Genet.*, **66**, 1945–1957.

Uhlenberg,B., Schuelke,M., Rüschendorf,F., Ruf,N., Kaindl,A.M., Henneke,M., Thiele,H., Stoltenburg-Didinger,G., Aksu,F., Topaloğlu,H. *et al*. (2004) Mutations in the gene encoding gap junction protein α12 (connexin 46.6) cause Pelizaeus-Merzbacher-like disease. *Am. J. Hum. Genet.*, **75**, 251–260.

Weeks,D.E., Sobel,E., O'Connell,J.R. and Lange,K. (1995) Computer programs for multilocus haplotyping of general pedigrees. *Am. J. Hum. Genet.*, **56**, 1506–1507.