Pelican: pedigree editor for linkage computer analysis

Frank Dudbridge*, Tim Carver and Gary W. Williams

MRC Rosalind Franklin Centre for Genomics Research, Hinxton, Cambridge CB10 1SB, UK

ABSTRACT

Summary: Linkage analysis software requires an input text file that describes the structure of the pedigrees to be analysed. Manual creation of these files is tedious and error-prone, and a graphical input tool is desirable. This is currently only available in commercial packages that include much greater functionality. We have therefore developed Pelican, a lightweight graphical pedigree editor for rapid construction of linkage pedigree files and diagrams.

Availability: The software runs on any Java-enabled machine (version 1.2 or higher). A Java Web Start launch, class files, a demonstration applet, source code and documentation are freely available at http://www.rfcgr.mrc.ac.uk/Software/ PELICAN/

Contact: support@rfcgr.mrc.ac.uk

Linkage analysis is used to map genes according to recombination events within pedigrees. A wide range of software exists for this purpose, with the common requirement of an input file that describes the structure of the pedigrees. Although several file formats are in use, the most common is that used by the LINKAGE package (Terwilliger and Ott, 1994); software is available to convert between different formats (Mukhopadyay *et al.*, 1999). A linkage pedigree file is a plain text file with one line for each subject, each line consisting of space-delimited columns that identify the subject, its parents in the pedigree, its gender, and its phenotype and genotype information (we shall only be concerned with the 'pre-makeped' version of the format here). Although the format is easily understood, manual creation of pedigree files is tedious and error-prone, and a graphical input tool would considerably simplify the process.

Several programs are freely available for graphical display of pedigrees, but they either do not allow graphical input (Curtis, 1990; Stajich *et al.*, 1998; Brun-Samarcq *et al.*, 1999; Trager, 2001, http://eyegene.ophthy.med. umich.edu/madeline/index.html; Loredo-Osti and Morgan, 2003, http://www.stat.washington.edu/Thompson/Genepi/

Pedfiddler.shtml) or do not create linkage files (Quiaoit, 1996, http://qge.fhcrc.org/pdpsys/; GenoPro, 2003, http:// www.genopro.com/). Graphical construction of linkage files is currently only available in some commercial packages as one component of an integrated project environment with much greater functionality (Cyrillic Software, 2003, http://www.cyrillicsoftware.com/; Epicenter Software, 2003, http://icarus2.hsc.usc.edu/epicentre/; Progeny Software, 2003, http://www.progeny2000.com/).

Often one would like to quickly create a pedigree file without investing significant effort in assimilating new software. A situation we have encountered is when some families have been ascertained for a clinical phenotype and an investigator wishes to estimate the power to detect linkage to a causal gene. Simulation software exists for this purpose, e.g. SLINK (Weeks *et al.*, 1990), which requires linkage files as input, but there is little motivation to create these from commercial software at the exploratory stage. Another situation where rapid pedigree creation is desirable is in the preparation of presentation graphics.

We have therefore developed Pelican, with the aim of providing a lightweight, intuitive graphical editor for quickly creating pedigree files and images. The Java language allows wide portability, and with the Web Start facility, users can launch Pelican directly from a web browser. A pedigree can be created from scratch, or an existing pedigree file can be displayed and edited, allowing any errors to be quickly highlighted and corrected. Changes to the pedigree are made by clicking the mouse on a subject and selecting actions from a popup menu. The pedigree can be saved as a linkage file (with optional extensions for the SLINK program), as an image file in PNG or JPEG format (note that this option requires Java 1.4) or as a PostScript file that can be sent directly to a printer.

A graphical editor should provide an automated layout facility, but this is a difficult problem for general pedigrees (Tores and Barillot, 2001). It can be further complicated when the user imposes constraints on the layout, e.g. placing siblings in their birth order. We have therefore implemented a layout algorithm that is sensitive to the input order of the pedigree,

---

*To whom correspondence should be addressed.

and allows the user to supervise the layout by graphically changing this input order.

We define a root subject to be an orphan subject with no non-orphan mate in the same generation (an orphan is a subject with no parents in the pedigree). Thus, a pedigree descending from a single founder mating has two root subjects. For each root subject that has not yet been laid out, the sub-tree layout is arranged in depth-first order, propagating the horizontal space used by each sibship back up the tree. Root subjects, siblings and mates are laid out from left to right following the input order of the pedigree, which can be changed by the user using mouse drag-and-drop. Inbreeding loops, inter-generational matings and experimental crosses can be constructed and displayed. The automatic layout can be disabled so that the user can arrange the pedigree in any desired fashion.

In LINKAGE format, a mating is only defined by its offspring, and sibships are defined only by their parents. Thus, unlike some other pedigree drawing programs, matings cannot be created without children, and if the only child of a mating is deleted, and one parent is a founder, then that parent is redundant for linkage analysis and is also deleted. Similarly, if one parent of a sibship is deleted, then all the offspring are also deleted, since most linkage programs require both parents to be defined even if no data is available for them.

Although not part of the LINKAGE format, Pelican can display standard symbols for deceased subjects, obligate carriers and proband status, as these are common features in pedigree diagrams (Bennett *et al.*, 1995). It can also indicate which subjects have DNA available for genotyping, as this information is used by SLINK. However, more extensive annotation features are not currently included, because this is not our primary goal and there is already software available for this purpose. This includes genotype data, which is not managed by Pelican although it is part of a linkage pedigree file. We feel that a sufficiently flexible database for the genotype data must largely duplicate the functionality of existing software, whereas our intention is to provide a convenient tool for creating the pedigree structure in situations where the genotype data is archived separately. However, Pelican does allow the display spacing between pedigree members to be varied, so that custom annotations could be subsequently added to the output image using another application.

REFERENCES

Bennett,R.L., Steinhaus,K.A., Uhrich,S.B., O'Sullivan,C.K., Resta,R.G., Lochner-Doyle,D., Markel,D.S., Vincent,V. and Hamanishi,J. (1995) Recommendations for standardized human pedigree nomenclature. *Am. J. Hum. Genet.*, **56**, 745–752.

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.

Cyrillic Software (2003) Cyrillic software. *Technical Report*, http://www.cyrillicsoftware.com/

Curtis,D. (1990) A program to draw pedigrees using LINKAGE or LINKSYS data files. *Ann. Hum. Genet.*, **54**, 365–367.

Epicenter Software (2003) GAP software. *Technical Report*, http://icarus2.hsc.usc.edu/epicentre/

GenoPro, Inc. (2003) GenoPro software. *Technical Report*, http://www.genopro.com/

Loredo-Osti,J.C. and Morgan,K. (2003) Pedfiddler: a suite of pedigree drawing programs. *Technical Report*, http://www.stat.washington.edu/Thompson/Genepi/Pedfiddler.shtml

Mukhopadyay,N., Almasy,L., Schroeder,M., Mulvihill,W.P. and Weeks,D.E. (1999) Mega2, a data-handling program for facilitating genetic linkage and association analyses. *Am. J. Hum. Genet.*, **65**, (suppl) A436.

Progeny Software, LLC (2003) Progeny software. *Technical Report*, http://www.progeny2000.com/

Quiaoit,F. (1996) PDPSys software. *Technical Report*, http://qge.fhcrc.org/pdpsys/

Stajich,J.E., Haynes,C. and Pericak-Vance,M.A. (1998) PEDPLOT: pedigree plotting program for the unix operating system. *Am. J. Hum. Genet.*, **63** (suppl), A242.

Terwilliger,J.D. and Ott J. (1994) *Handbook of Human Genetic Linkage*. Johns Hopkins University Press, Baltimore, MD.

Tores,F. and Barillot,E. (2001) The art of pedigree drawing: algorithmic aspects. *Bioinformatics*, **17**, 174–179.

Trager,E. (2001) Madeleine software. *Technical Report*, http://eyegene.ophthy.med.umich.edu/madeline/index.html

Weeks,D.E., Ott,J. and Lathrop,G.M. (1990) SLINK: a general simulation program for linkage analysis. *Am. J. Hum. Genet.*, **47**, (suppl), A204.