Journal of Heredity 2012:103(6):903–907
doi:10.1093/jhered/ess060
Advance Access publication October 19, 2012

Computer Note

Pedimap: Software for the Visualization of Genetic and Phenotypic Data in Pedigrees

ROELAND E. VOORRIPS, MARCO C. A. M. BINK, AND W. ERIC VAN DE WEG

From the Wageningen University & Research Centre, Plant Breeding, PO Box 16, 6700 AA Wageningen, The Netherlands (Voorrips and van de Weg); and the Wageningen University & Research Centre, Biometris, PO Box 100, 6700 AC Wageningen, The Netherlands (Bink).

Address correspondence to Roeland E. Voorrips at the address above, or e-mail: roeland.voorrips@wur.nl.

Pedimap is a user-friendly software tool for visualizing phenotypic and genotypic data for related individuals linked in pedigrees. Genetic data can include marker scores, Identity-by-Descent probabilities, and marker linkage map positions, allowing the visualization of haplotypes through lineages. The pedigrees can accommodate all types of inheritance, including selfing, cloning, and repeated backcrossing, and all ploidy levels are supported. Visual association of the genetic data with phenotypic data simplifies the exploration of large data sets, thereby improving breeding decision making. Data are imported from text files; in addition data exchange with other software packages (FlexQTL™ and GenomeStudio™) is possible. Instructions for use and an executable version compatible with the Windows platform are available for free from http://www.plantbreeding.wur.nl/UK/software_pedimap.html.

Key words: *genetic data, pedigree software, phenotypic data, plant breeding*

The availability of extensive genetic information is transforming plant breeding. Chromosome regions that contain desirable alleles can theoretically be systematically combined and selected by breeders to increase the probability of obtaining elite individuals. However, increasing quantities of phenotypic and genetic data in breeding programs are rapidly becoming challenging to manage and mine for useful information.

A considerable number of software programs are available for diagramming human and animal pedigrees and combining this information with trait and/or marker data: Cyrillic (Chapman 1990), PEDRAW (Curtis 1990), PEDHUNTER (Agarwala et al. 1998), CoPE (Brun-Samarcq et al. 1999), Pelican (Dudbridge et al. 2004), HaploPainter (Thiele and Nürnberg 2005), Madeline (Trager et al. 2007), PediGrach (Garbe and Da 2008), Progeny (http://www.progenygenetics.com/clinical/pedigree.html), and R package Kinship (http://cran.r-project.org/web/ packages/kinship/kinship.pdf). For plant breeding and genetics some specialized packages are also available, including PediTree (Van Berloo and Hutten 2005) and E-Brida (http://www.e-brida.nl). Many of these programs go beyond the visualization of pedigrees; they carry out calculations such as inbreeding coefficients and/or they include a database for storing breeding information. However, none of these programs offer the combination of user-definable phenotypic traits, linkage map–based genetic marker data, and the flexibility to handle plant pedigrees, in which selfing, clonal propagation, doubled haploid production, and repeated backcrossing with the same parent may occur. Other features frequently encountered in plant (breeding) pedigrees that are not (or to a smaller extent) found in human and animal pedigrees are polyploidy, hermaphroditic individuals (an individual can be used both as male and as female parent in different crosses), and mutants. A comprehensive pedigree visualization package for use in plant breeding and genetics must accommodate these specific plant needs.

Pedimap was developed as a tool to visualize and assist in the exploration of the flow of phenotypes and marker alleles of individuals through pedigrees to inform plant breeding decisions, and to understand and correct possible errors in genotyping and in assumed pedigree relations. It does not perform statistical or quantitative genetic calculations; rather it presents available genotypic and phenotypic data in a way that enables the understanding of the data. Pedimap simplifies the analysis of large and complex pedigrees by allowing the selection of groups of individuals (subpopulations) based on their pedigree relations.

The primary input files include data for the 3 core components of breeding records: parentage, phenotypic, and genotypic data. Pedimap can represent all possible forms of parentage in plant pedigrees including biparental crosses and selfings, sib-mating, backcross relationships, vegetative propagation, mutants, and doubled haploids. Population sizes are not restricted and any level of polyploidy can be accommodated. All categories of phenotypic data are supported, including continuous quantitative, discrete quantitative, and qualitative data. Also, all marker types are supported including representation of putative and confirmed null alleles, and calculated Identity-by-Descent (IBD) probabilities can be supplied as well. These phenotypic and genotypic data are not required nor do they have to be present for all individuals, so Pedimap can also be used to study just the structure of a pedigree itself.

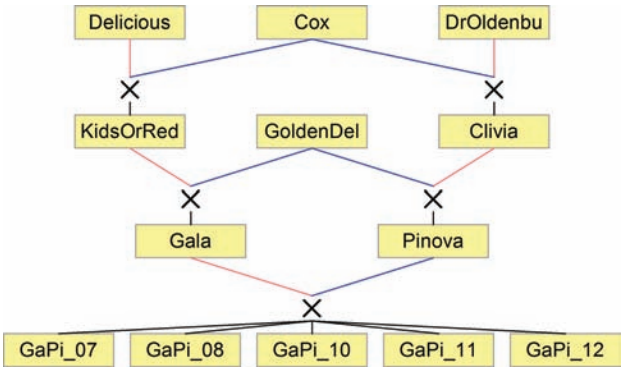

Figure 1. Basic presentation of part of an apple pedigree. Female parentage is indicated by pink (light grey) lines and male parentage by blue (dark grey) lines. This subpopulation was obtained by automatically selecting all ancestors and sibs of individual GaPi_07.

Features and Functionality

Pedigree Structure

The interface for the first set of data, pedigree relationships, is based on the concept of subpopulations. From the total population, various subpopulations can be selected and each can be viewed in different graphical presentations. These subpopulations can be assembled manually, but tools are available to automatically include or exclude further individuals based on their pedigree relations with another individual.

In the simplest case, the pedigree of a selected individual can be visualized. In Figure 1, the individual (apple seedling) "GaPi_07" was selected, and all its ancestors and full sibs were added automatically using the tool for selection based on pedigree relationships.

The pedigree visualization forms the scaffold on which the next set of data, phenotypic data, is superimposed. Qualitative phenotypic data can be color coded to reflect the different trait classes, whereas a color gradation can be used to represent variation for a quantitative trait (Figures 2 and 3).

Genetic Marker Data

Next to the pedigree and phenotypic data, genetic data (observed marker genotypes) can be added, organized by linkage group. The required information for each linkage group includes a linkage map: a list of locus names and map positions, and for each locus, the possible allele identifiers. The locus genotype of every individual is specified by 2 (or more, in the case of polyploids) allele identifiers. Instead of an allele,

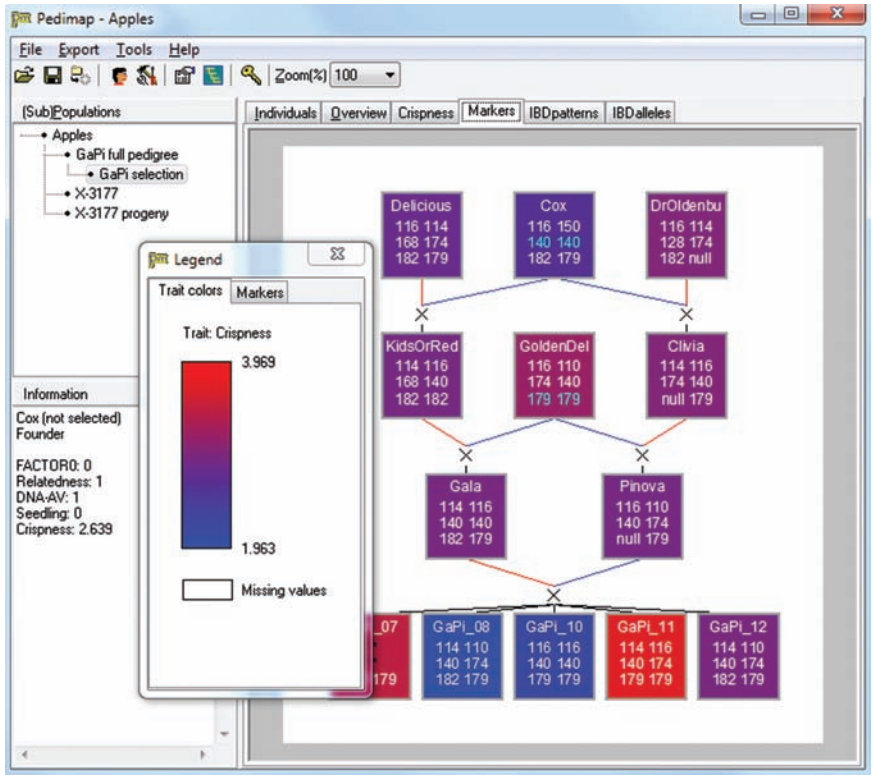

Figure 2. Screenshot of the user interface of Pedimap with the subpopulations panel (top left, with the "GaPi selection" subpopulation selected), information panel (bottom left, with information on the individual under the cursor), and the pedigree view (right). The pedigree view shows the same subpopulation as Figure 1 with apple crispness values indicated by color and with observed marker genotypes for 3 Simple Sequence Repeat loci. In several cases the alleles are highlighted in different colors.

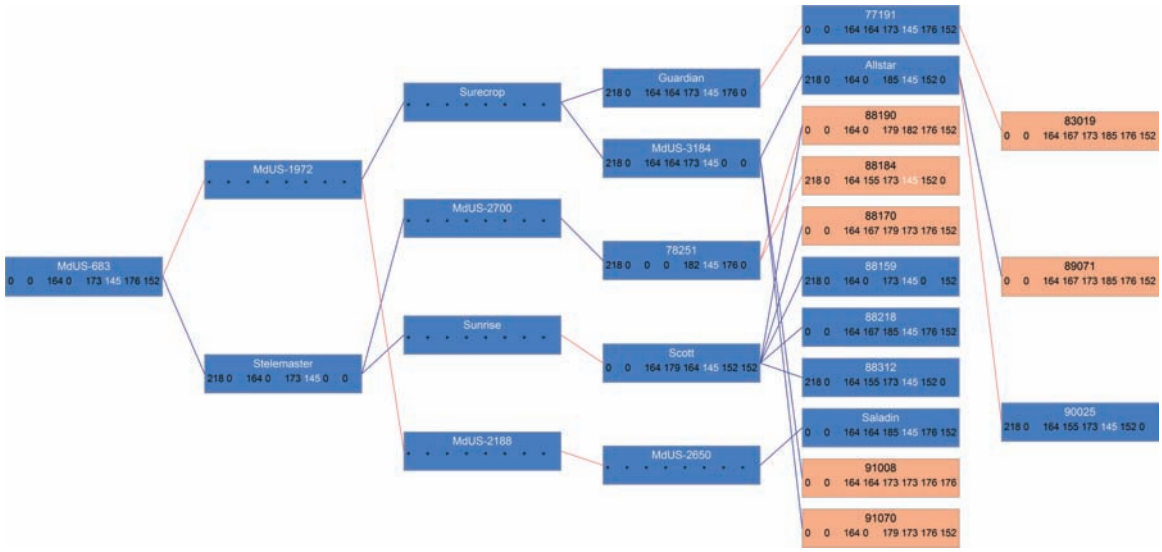

Figure 3. Part of a pedigree of octoploid strawberry. Individuals colored blue (dark grey) or pink (light grey) when resistant or susceptible to red root rot (*Phytophthora fragariae*), respectively. The octoploid genotypes at an SSR marker are shown; the allele 145 (in white) is closely linked to the resistance gene (data from A. Pikunova et al., Wageningen UR - Plant Breeding).

a missing value or a symbol indicating a possible or a confirmed null allele can be given. It is possible to assign an allele a specific color, which can be useful to investigate the inheritance of specific alleles or to easily identify situations (e.g., contradictory genotypes) that need further attention. The genetic data for selected marker loci can then be visualized along with the pedigree and phenotypic data (Figures 2–4).

IBD

Pedimap can also visualize IBD probabilities of founder alleles. These must be given in the input file as Pedimap does not calculate these values itself. A package that can perform these calculations and produces output files compatible with Pedimap is FlexQTL (Bink et al. 2008; see also Interfacing with Other Software). This visualization can help to follow the inheritance of a specific haplotype through the pedigree, even if several founders have the same marker allele at that position (Figure 5A). It can also illustrate how the founder chromosomes are progressively recombined through the generations (Figure 5B; Van Dijk et al. 2012).

Interfacing with Other Software

Pedimap is being used in several projects in close relation with FlexQTL, software for QTL analysis and IBD calculations in multiple pedigreed populations (Bink et al. 2008). In order to allow for easy exchange of data, both Pedimap and FlexQTL export data in the format required for the other program. In the case of Pedimap, this involves the export of data from the complete pedigree or from selected subpopulations, and the filling-in of missing parents of "semi-founders": individuals of which only 1 parent is known, which are accepted by Pedimap but not by FlexQTL.

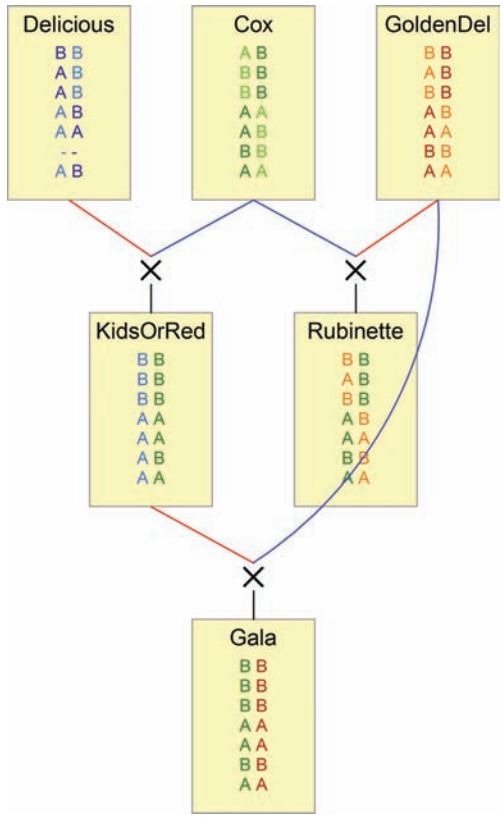

Figure 4. A different part of the apple pedigree is shown with genotypes for 7 closely linked Single Nucleotide Polymorphism (SNP) markers; the alleles per locus are listed in alphabetical order. SNP haplotypes are reconstructed starting at the completely homozygous individual Gala and shown in different colors (SNP data from Dr Michela Troggio, Istituto Agrario di San Michele all'Adige).

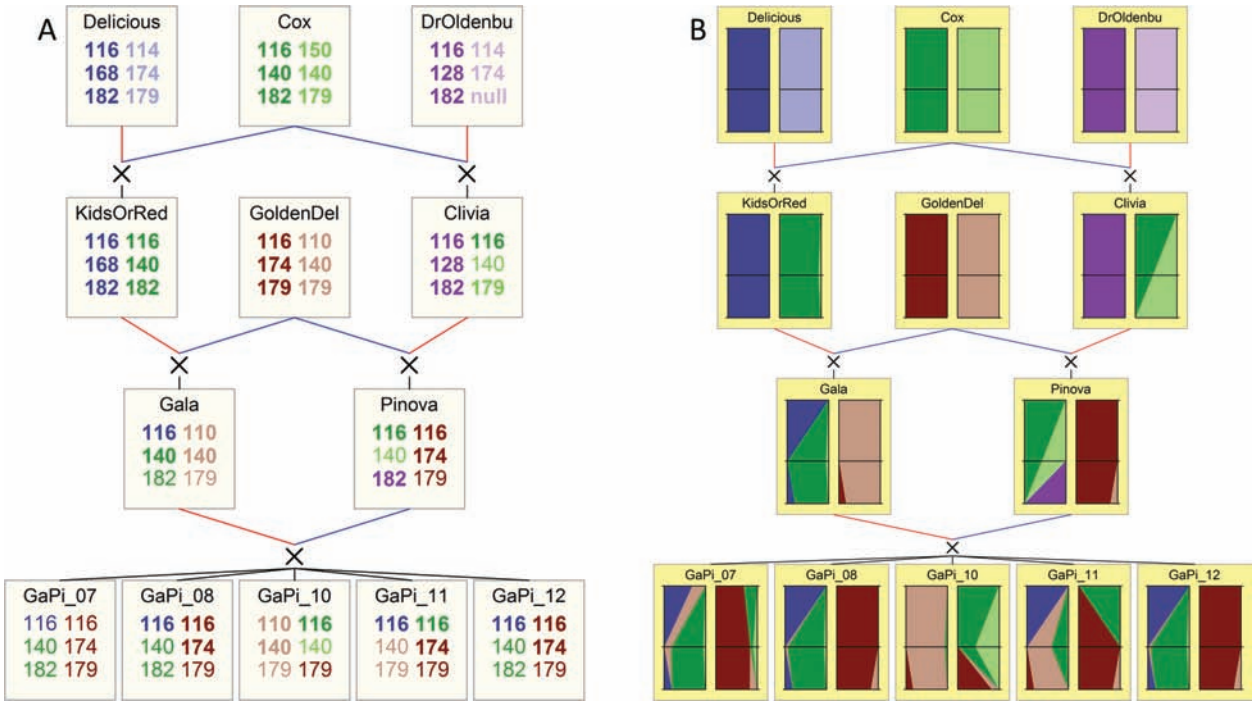

Figure 5. The same apple pedigree as in the first 3 figures is shown with precalculated IBD data for part of chromosome 1. Panel A shows the most probable founder alleles instead of the observed alleles. The color of the allele indicates the founder allele. A bold typeface indicates a high probability, normal and italic typefaces indicate lower probabilities according to user-specified thresholds. In panel B, again each color represents a different founder haplotype. Each rectangle in an individual represents 1 of the 2 copies of the selected chromosome. The vertical dimension represents the positions along the chromosome; the width of a color at a certain position indicates the IBD probability that the corresponding founder allele is present.

In order to import large amounts of SNP data generated by Illumina's genotyping platforms, a conversion script is available from the Pedimap website that builds a data file in Pedimap format from a text file exported by the GenomeStudio software (http://www.illumina.com/software/genomestudio_software.ilmn).

Availability

Pedimap is available at no cost for the Windows platform (all 32- or 64-bit versions). Pedimap can be downloaded from http://www.plantbreeding.wur.nl/UK/software_pedimap.html. The download package contains also a manual and a small sample dataset. A license can be obtained free of charge from the same website.

Conclusion

Pedimap allows the visualization of phenotypic and genotypic information in pedigreed populations. It does not do any calculations such as inbreeding coefficients or IBD probabilities, and it does not offer database capabilities; it only visualizes the data imported from an input file. However, it does so in a very flexible way and it is unique in its ability to handle a combination of items including selfing, hermaphroditism, backcrossing, polyploidy, and others that are especially important in plant pedigrees. As such it is uniquely suited for use in a plant (breeding) context, whereas it can also be applied to other organisms.

Funding

European projects HiDRAS (QLK5-CT-2002-01492; http://www.hidras.unimi.it) and FruitBreedomics (265582-EC-GA; http://fruitbreedomics.com/); USDA-NIFA-SCRI project RosBREED: Enabling marker-assisted breeding in Rosaceae (2009-51181-05808; http://www.rosbreed.org); Wageningen UR Department of Plant Breeding.

Acknowledgements

We thank Dr Amy Iezzoni and Dr Umesh Rosyara (Michigan State University) for their stimulating involvement and for critically reading the manuscript.

References

Agarwala R, Biesecker LG, Hopkins KA, Francomano CA, Schäffer AA. 1998. Software for constructing and verifying pedigrees within large genealogies and an application to the old order Amish of Lancaster County. Genome Res. 8:211–221.

Bink MCMA, Boer MP, Ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE. 2008. Bayesian analysis of complex traits in pedigreed plant populations. Euphytica. 161:85–96.

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

Chapman CJ. 1990. A visual interface to computer programs for linkage analysis. Am J Med Genet. 36:155–160.

Curtis D. 1990. A program to draw pedigrees using LINKAGE or LINKSYS data files. Ann Hum Genet. 52:259.

Dudbridge F, Carver T, Williams GW. 2004. Pelican: pedigree editor for linkage computer analysis. Bioinformatics. 20:2327–2328.

Garbe JR, Da Y. 2008. Pedigraph user manual version 2.4. Department of Animal Science, University of Minnesota.

Thiele H, Nürnberg P. 2005. HaploPainter: a tool for drawing pedigrees with complex haplotypes. Bioinformatics. 21:1730–1732.

Trager EH, Khanna R, Marrs A, Siden L, Branham KE, Swaroop A, Richards JE. 2007. Madeline 2.0 PDE: a new program for local and web-based pedigree drawing. Bioinformatics. 23:1854–1856.

van Berloo R, Hutten RC. 2005. Peditree: pedigree database analysis and visualization for breeding and science. J Hered. 96:465–468.

van Dijk T, Noordijk Y, Dubos T, Bink MC, Meulenbroek BJ, Visser RG, van de Weg WE. 2012. Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids. BMC Plant Biol. 12:25.

Received January 9, 2012; Revised June 14, 2012; Accepted July 8, 2012

Corresponding Editor: John Stommel