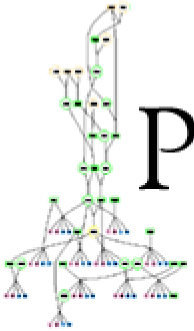


Pedigraph

A Software Tool for the Graphing and Analysis of Large Complex Pedigree

User Manual Version 2.4

(INTERNET VERSION WITH LOWER GRAPHIC QUALITY)

JOHN R. GARBE AND YANG DA

DEPARTMENT OF ANIMAL SCIENCE
UNIVERSITY OF MINNESOTA

AUGUST 19, 2008

1 INTRODUCTION

Graphical pedigree visualization is helpful for studying the relationships among individuals, gene flows from generation to generation, and the population structure. However, graphical visualization of large complex pedigrees is often a humanly impossible task. Pedigraph provides rapid graphical visualization of large, complex pedigrees, with options for controlling colors, drawing size, page size and margins, drawing styles, extraction and highlighting of partial pedigrees involving selected individuals, and calculation of inbreeding and coancestry coefficients. Pedigraph can display all individuals in the data set, or display the number of offspring in each family by gender and trait status such as disease or normal phenotypes.

2 GETTING STARTED

The current version of Pedigraph can be run under two operation systems: Windows and Linux.

2.1 On Microsoft Windows

PEDIGRAPHTo run Pedigraph on Windows, open an MS/DOS window and move to the directory where the Pedigraph executable is located. The type the following command:

```
pedigraph pedigree_file_name
```

where ‘pedigraph’ stands for ‘pedigraph.exe’, and ‘pedigree_file_name’ is the file name of the input pedigree file.

Alternatively, simply double-click the pedigraph.exe file, and type in the name of the input file at the prompt. A PostScript file named output.ps will be produced which contains the pedigree drawing. If Adobe Acrobat Distiller is installed on the computer, double-click on the PostScript file will generate a PDF file of the pedigree drawing.

2.2 On Linux

To run Pedigraph on Linux, open a Linux terminal and move to the directory where the Pedigraph executable is located. Type pedigraph followed by the name of the input file to run the program. A PostScript file named output.ps will be produced which contains the pedigree drawing. Use gv to view the Postscript file. (NOTE: before running Pedigraph for the first time on Linux you may need to change the permissions on the executable with this command:

```
$chmod u+x pedigraph.
```

3 INPUT FILES

Pedigraph requires one input file, the pedigree file containing pedigree information, and an optional parameter file containing program options. The parameter file must be named as ‘pedigraph.dat’. In the event pedigraph.dat is missing or empty, the pedigree is drawn using default options as described in **4.1**.

3.1 Pedigree file

The data file is a simple text file that contains the pedigree to be drawn. Each line gives the ID of an individual (column 1) as well as the IDs of the individual’s parents (columns 2 and 3, the gender of the individual (column 4), and, optionally, a trait status (column 5).

Individual ID’s in column 1 can be characters or numbers other than 0 (zero). Parental IDs in columns 2 and 3 can be characters or numbers. A parental ID of 0 indicates an unknown parent. The default gender code is M for males and F for females. The user may use other gender codes using the ‘male’ and ‘female’ options in the parameter files. Gender codes other than the default codes or the user specified codes are considered as missing gender. The Pedigraph grogram has mechanism to deduce the gender of an individual that does not have a gender code whenever possible. A fifth column must be present in the pedigree file if a trait status is to be displayed, and the value indicating an affected individual must be specified in the parameter file. If the pedigree file does not have a fifth column to define the trait status, the parameter file may not activate the options ‘affected’ and ‘affectedcolor’.

An example of the input pedigree file is shown below.

```
A23     A13     A14     F       1
A21     A13     A9      M       0
A22     A13     A10     F       1
A20     A7      A8      F       1
A17     A7      A5      F       0
A15     A13     A5      F       0
A19     A7      A8      M       0
A25     A7      A23     M       0
A26     A7      A23     M       1
A29     A7      A23     F       0
A27     A7      A23     M       0
A28     A7      A23     F       1
A30     A7      A23     F       0
A16     A7      A5      F       1
A18     A7      A5      M       1
A24     A7      A20     F       0
A6      A1      0       M       1
A10     A1      A2      F       1
A11     A1      A2      M       1
A7      A1      A2      M       1
A9      A1      0       F       0
A5      A1      A3      F       0
A12     A1      A4      F       0
```

3.2 Parameter file

The parameter file named “pedigraph.dat” contains Pedigraph options to allow custom pedigree drawings and analysis. Each line in “pedigraph.dat” should contain only one option. Lines beginning with # are ignored, as are blank lines. **Table 1** summarizes the currently available program options. The usages of these options are illustrated in **PEDIGREE DRAWING**.

Table 1. Program Options

Option Name	Default Value	Description
male	M	A string that indicates a male individual
female	F	A string that indicates a female individual
affected		A string that indicates an affected individual
summary		Activate the ‘summary’ option. Specify 1 or 2.
individual		Draw the pedigree for the specified individual
descendants		Draw the pedigree of the individual’s offspring (following ‘individual’ statement).
sibs		Draw the pedigree of the individual’s siblings (following ‘individual’ statement).
ancestors		Draw the pedigree of the individual’s ancestors (following ‘individual’ statement).
extractedoutputfile		Print the pedigree data of the selected individual to a file
affectedcolor	lightgrey	The color used to indicate affected individuals
nocolor		The pedigree is drawn in black and white
nodecolor	black	The color used to outline individual nodes
Fillcolor	white	The color used to indicate unaffected individuals
familysize	0	Minimum number of offspring for a mating to be drawn
title		A title to be printed on the drawing
unknown		Display “?” nodes for unknown individuals
gengap	2.2	Distance (in inches) between generations (rows of nodes)
verbose		Print progress information to the screen while the program is running
width		Width (in inches) of the drawing
height		Height (in inches) of the drawing
pagewidth		Width (in inches) of a page (e.g. 8.5)
pageheight		Height (in inches) of a page (e.g. 11)
nomargin		Prevents margins on multi-page drawings

outputfile jpg	output.ps	Name of the output file Create the output file in jpg format. Supported on Windows only.
png		Create the output file in the png format instead of PostScript. Supported on Windows only
inbreeding		Turns on calculation of inbreeding coefficients
inbreedfile		Name of the file in which inbreeding coefficients will be printed. If a file is not specified the inbreeding coefficients will be printed to the screen.
coancestryfile	No coancestry calculation	Name of the file containing a list of pairs of individual IDs. Pedigraph will compute the coancestry coefficient between each pair of individuals in the file.
coancestryoutputfile	No output	Name of the file in which coanscestry coefficients will be printed. If a file is not specified the coancestry coefficients will be printed to the screen.
coancestry	No coancestry calculation	Calculates coancestry coefficients between the selected individual and all relatives of the individual. This must be used with 'individual' command jointly.
coancestryped output_file_name		This option print coancestry coefficients from the statements 'individual' and 'coancestry' to the file named 'output_file_name'
arrows	No arrows	Draws pedigrees using arrowed lines.
title	No title	A title to be printed on the drawing

3.3 Color definitions

Codes for defining a color in the color control options in **Table 1** (affectedcolor, fillcolor, and nodecolor) are given below.

blue	blue1	blue2	blue3	blue4
brown	brown1	brown2	brown3	brown4
cyan	cyan1	cyan2	cyan3	cyan4
gold	gold1	gold2	gold3	gold4
green	green1	green2	green3	green4
grey	grey1	grey2	grey3	grey4
orange	orange1	orange2	orange3	orange4
pink	pink1	pink2	pink3	pink4
purple	purple1	purple2	purple3	purple4
red	red1	red2	red3	red4
tan	tan1	tan2	tan3	tan4
yellow	yellow1	yellow2	yellow3	yellow4
white	lightgrey	black		

Each of three color control statements, affectedcolor, fillcolor, and nodecolor, may use any one of the above 63 colors.

4 PEDIGREE DRAWING

In the pedigree drawings produced by Pedigraph, each individual listed in the input file is represented by a node. The shape and color of the node indicate information about the individual. The label inside the node is the ID of the individual. A circle indicates a female, a square a male, and a diamond an individual of unknown gender. A dotted orange line indicates that the individual has no ancestral and gender information in the input file. Such individuals often are ‘founders’ of the pedigree. A filled node indicates an individual with a special status, such as a diseased or an excellent individual, or an individual with (or without) DNA sample. All individuals in the same generation are placed in the same row, and individuals in a row are ordered to reduce line crossings. In some cases the generation of an individual cannot be determined and may be placed between generations. Colored lines are used to connect offspring with their parents.

4.1 Full pedigree drawing with default settings

To illustrate the pedigree drawings using Pedigraph, the pedigree file in **3.1** is used. The pedigraph.dat parameter file that comes with the Pedigraph software has a # sign at the beginning of each line, meaning that none of the options is activated. The use of this parameter file without any modification will produce a full pedigree drawing using default options. **Figure 1** shows the full pedigree drawing using the above data and the default settings.

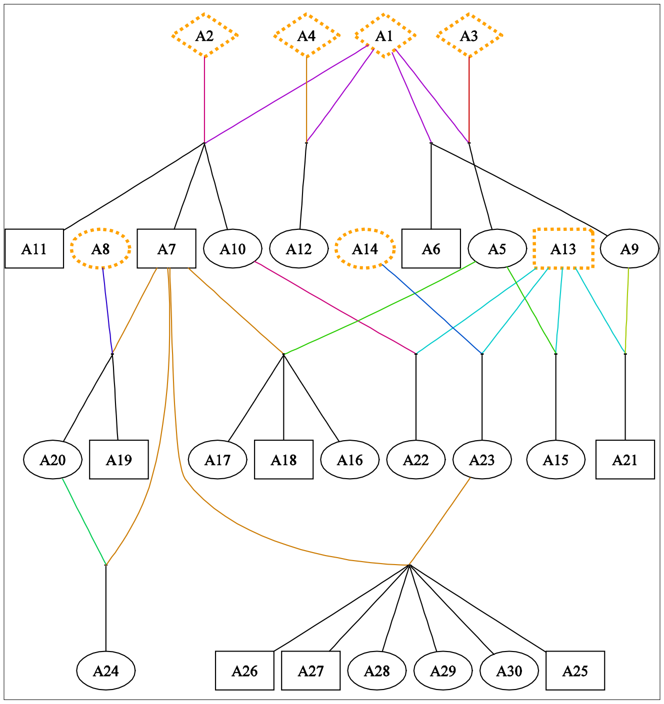


Figure 1. Example of full pedigree drawing with default settings.

In **Figure 1**, individuals that with dotted orange lines do not have ancestral and gender information in the testped.txt file. The top four individuals with diamond nodes, individuals A1, A2, A3, and A4, have unknown genders and their genders cannot be inferred from the data. The other individuals with dotted orange lines also have unknown genders in the data but their genders are determined in the pedigree drawing based on the known genders of their mates. The color lines help identify the offspring from a particular parent, e.g., the three purple lines connecting to individual A1 show that A1 has three offspring.

4.2 Display of trait status

The testped.txt file has a fifth column defining the trait status but that column is not used in **Figure 1** using the default settings. The trait status can be defined as 'affected' versus 'unaffected', 'diseased' versus 'normal', or 'good' versus 'bad', or the availability of DNA sample of the individual. To display the trait status in column 5, the 'affected' option needs to be activated in pedigraph.dat. If a custom color is to be used to highlight the 'affected' nodes, the 'affectedcolor' option also needs to be activated in the pedigraph.dat file.

Using 'affected 1' and 'affectedcolor gold' statements in pedigraph.dat and the same data of **Figure 1** and, the full pedigree with 'affected' individuals highlighted in gold color is produced, as shown in **Figure 2**. As examples of color options for highlighting trait status, **Figure 3** shows six optional colors of 'affected' individuals.

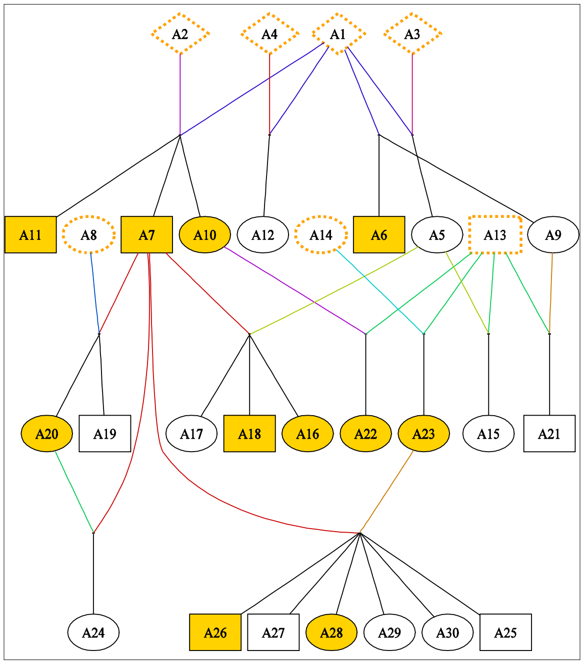


Figure 2. Full pedigree drawing with 'affected' individuals highlighted in gold color.

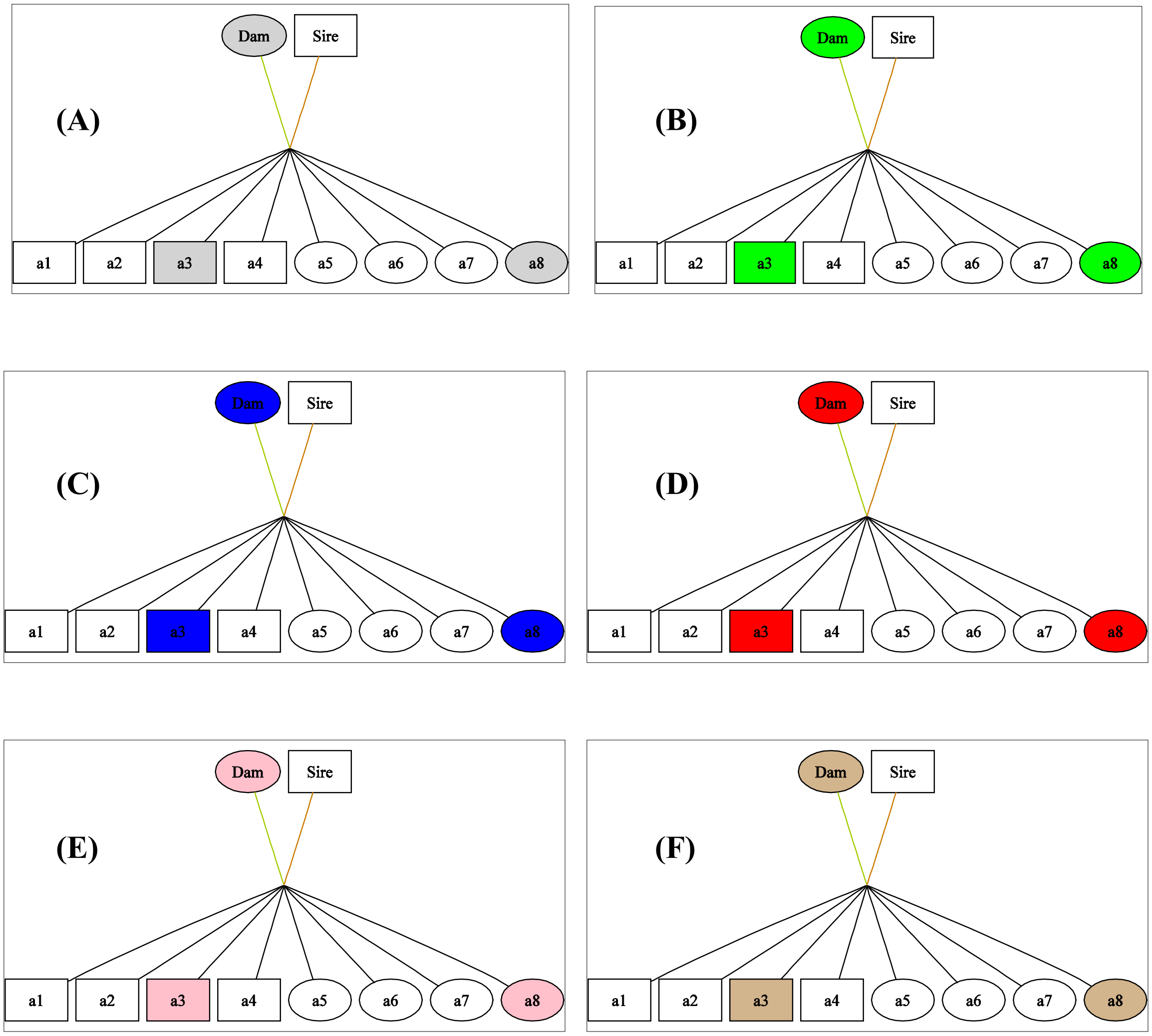
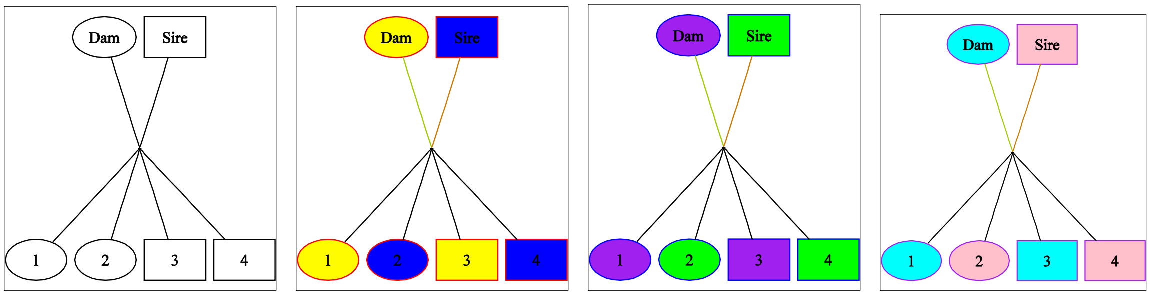


Figure 3. Selected color options to highlight 'affected' individuals in the pedigree drawing. **(A)** 'affected 1' statement is used in pedigraph.dat. **(B)** 'affected 1' and 'affectedcolor green' statements are used in pedigraph.dat. **(C)** 'affected 1' and 'affectedcolor blue' statements are used in pedigraph.dat. **(D)** 'affected 1' and 'affectedcolor red' statements are used in pedigraph.dat. **(E)** 'affected 1' and 'affectedcolor pink' statements are used in pedigraph.dat. **(F)** 'affected 1' and 'affectedcolor tan' statements are used in pedigraph.dat.

4.3 Color controls

Colors of pedigree nodes are controlled by four options: nocolor, affectedcolor, nodecolor, and fillcolor. The ‘nocolor’ option draws a black-white pedigree, while the other three color options can be used individually or jointly to draw color pedigrees. Each of the last three color options can use any of the 63 colors defined in Section **3.3**. Therefore, the number of possible color combinations for pedigree drawing is 63^3 = 250,047. In **Figure 4**, **(A)** is a pedigree drawing with ‘nocolor’ option, and **(B-D)** are 3 of the 250,047 possible color combinations for drawing color pedigrees.

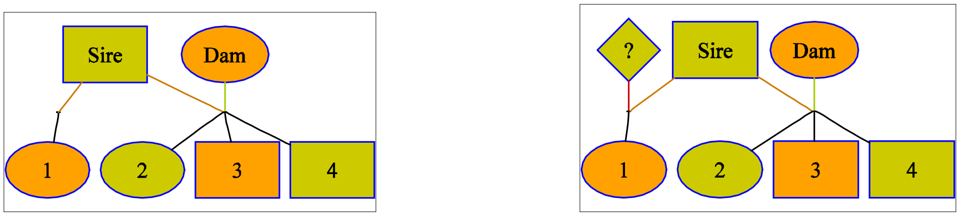


(A) **(B)** **(C)** **(D)**

Figure 4. Examples of color controls. **(A)** nocolor. **(B)** affectedcolor yellow, nodecolor red, fillcolor blue. **(C)** affectedcolor purple, nodecolor blue, fillcolor green. **(D)** affectedcolor cyan, nodecolor purple, fillcolor pink.

4.4 Unknown parents

Pedigraph does not include unknown parents in the pedigree drawing by default. The ‘unknown’ option allows each unknown individual be drawn using a question mark.

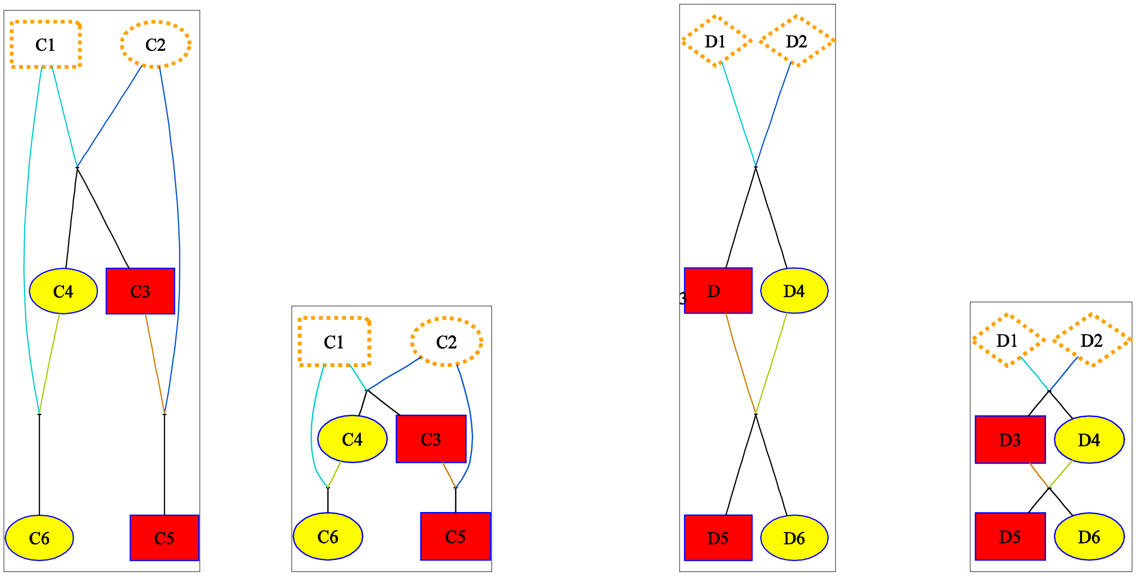


(A) **(B)**

Figure 5. Pedigree drawing for unknown parents. **(A)** Unknown dam of individual 1 is not drawn (default). **(B)** Unknown dam of individual 1 is drawn as a question mark using ‘unknown’ statement in pedigraph.dat.

4.5 Generation gap

Pedigraph has an option to control the distance between nodes of two adjacent generations, referred to as ‘generation gap’. For certain pedigrees, such as multigenerational pedigrees with relatively simple mating structures, the default value for generation gap may be unnecessarily large. The ‘gengap’ statement in pedigraph.dat allows custom values of generation gap. **Figure 6** shows two comparisons between pedigree drawings using the default and custom ‘gengap’ values.



(A) **(B)** **(C)** **(D)**

Figure 6. Control of generation gap in pedigree drawing. **(A)** Default value of 2.2 for generation gap. **(B)** The statement ‘gengap 0.5’ is used in pedigraph.dat. **(C)** Default value of 2.2 for generation gap. **(D)** The statement ‘gengap 0.5’ is used in pedigraph.dat.

4.6 Tests for drawing full pedigrees of large complex populations

The Pedigraph program has been tested for its capability of drawing full pedigrees of large complex populations. Thus far the Pedigraph program has not failed due to population size and complexity such as complicated inbreeding structure. Sample full pedigree drawings for such large complex populations are available from http://animalgene.umn.edu .

The largest sample tested for full pedigree drawing was the European Royal Genealogical Data (www.dcs.hull.ac.uk/public/genealogy/royal/). The database is unofficial and is known to have errors, but is an ideal set of data for testing Pedigraph because the population is large and complex. The full pedigree contains 48,605 individuals spread out over 100 generations. The sample size and complexity of this data set did not cause any problems for Pedigraph. The full pedigree drawing required 40 minutes of CPU time to print on a 2 GHz Pentium 4 desktop computer. The size of the pedigree printout appears to be the only major limitation for pedigree drawing using Pedigraph. A readable printout of the main body of the full pedigree could be as large as 100 feet wide and 80 feet tall (**Figure 7A**). The drawings gives a good sense of the overall structure of the full pedigree and reveals highly complex relationships, a visual effect that is impossible from reading the genealogy data. On the computer screen, details of a small area of the full pedigree could be viewed clearly using 'zoom in' (**Figure 7B**).

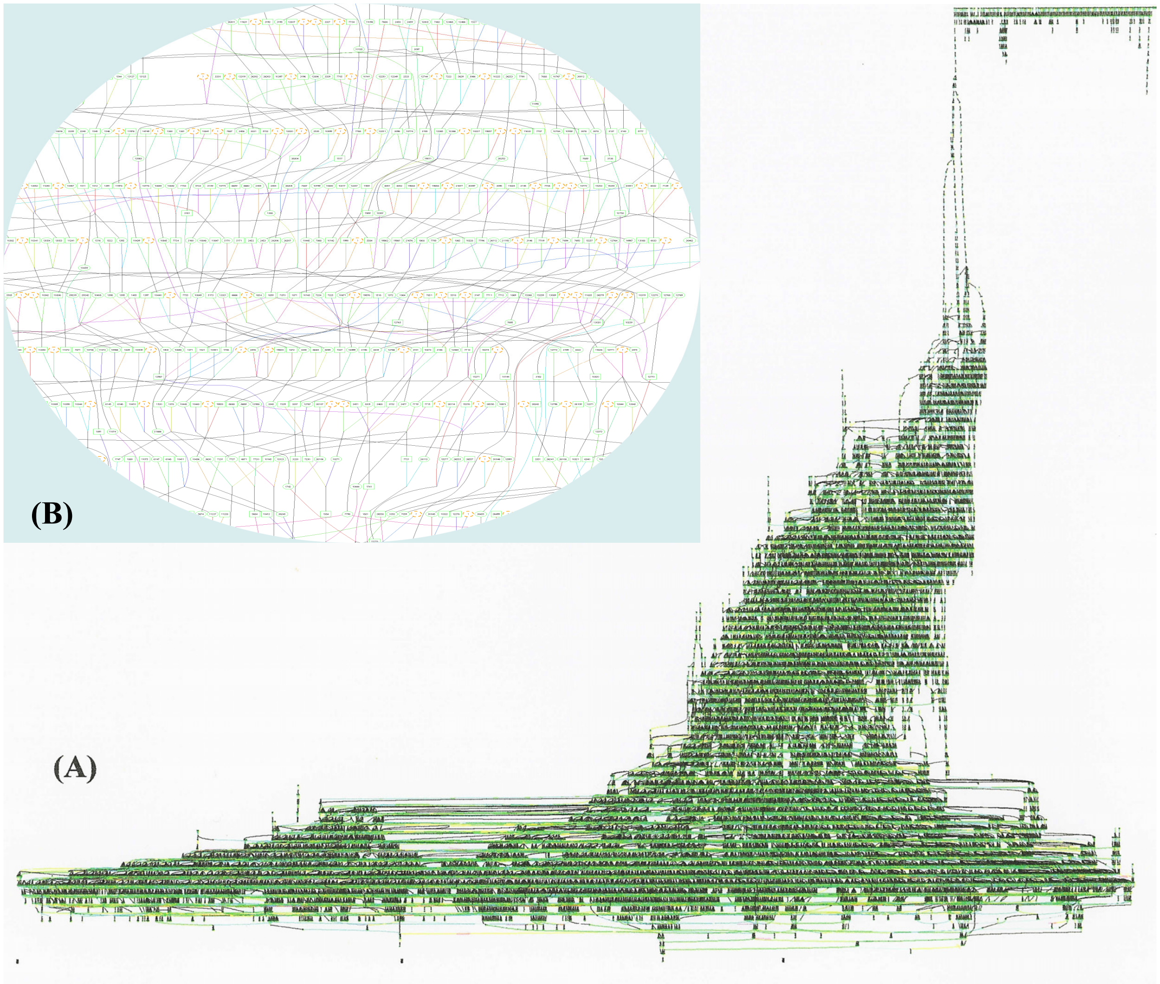


Figure 7. Full pedigree drawing of the European Royal Genealogical Data. (A) Main body of the full pedigree drawing (omitting a long horizontal string at the top-right). **(B)** 'Zoom-in' of a small area of the full pedigree.

4.7 Compact pedigree drawing using ‘summary’ option

Pedigraph has two ‘summary’ options, ‘summary 1’ and ‘summary 2’, for compact pedigree drawing. These two ‘summary’ options display parents and the number of offspring in each family. The ‘summary 1’ option displays each offspring group in a separate node while the ‘summary 2’ option displays all offspring groups in each family in different cells of one node, which yields more compact but less artistic pedigree drawing than ‘summary 1’ option. A compact pedigree drawing can be much smaller than the full pedigree drawing. The drawing size difference between a compact pedigree and the full pedigree increases as family size increases. For populations with large families, the ‘summary’ options can be an effective tool for the graphical visualization of the gene or disease flow in the population.

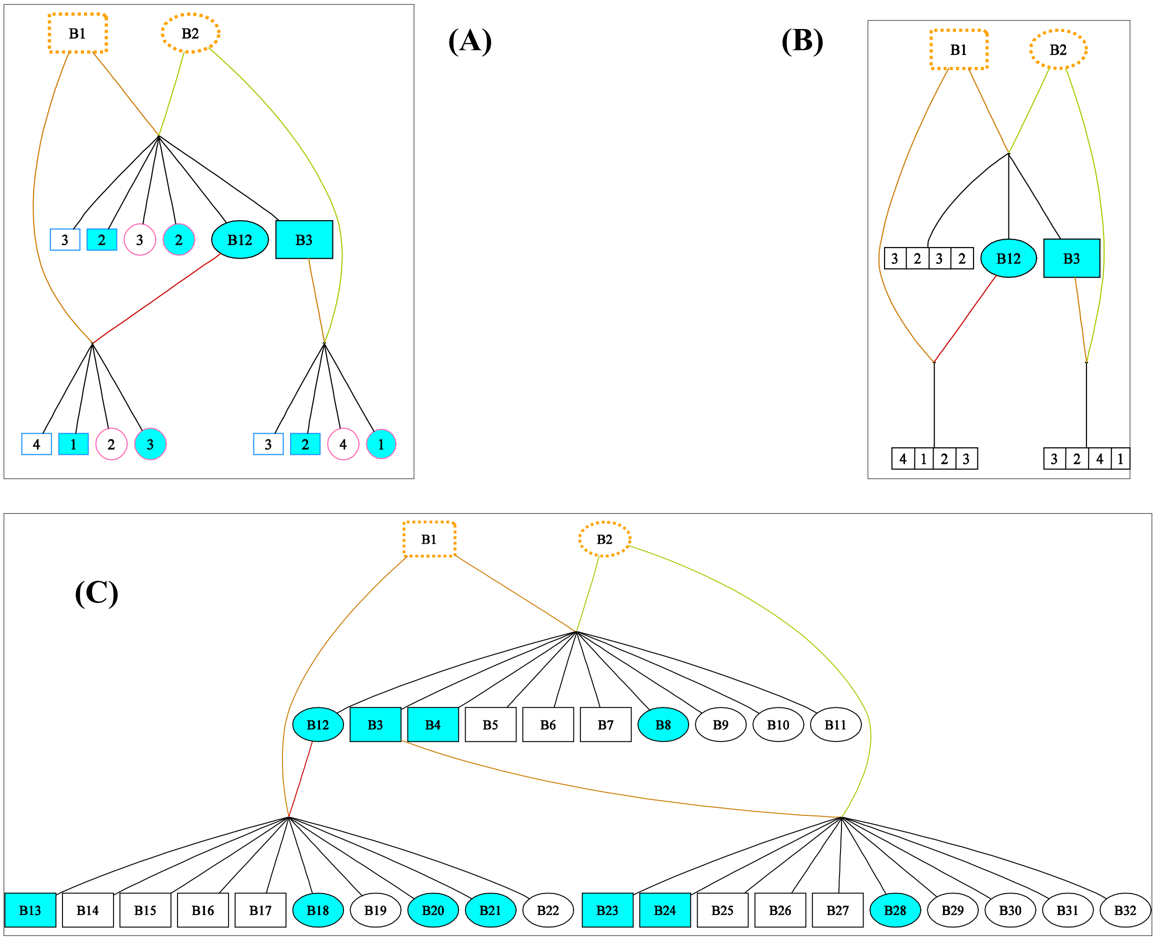
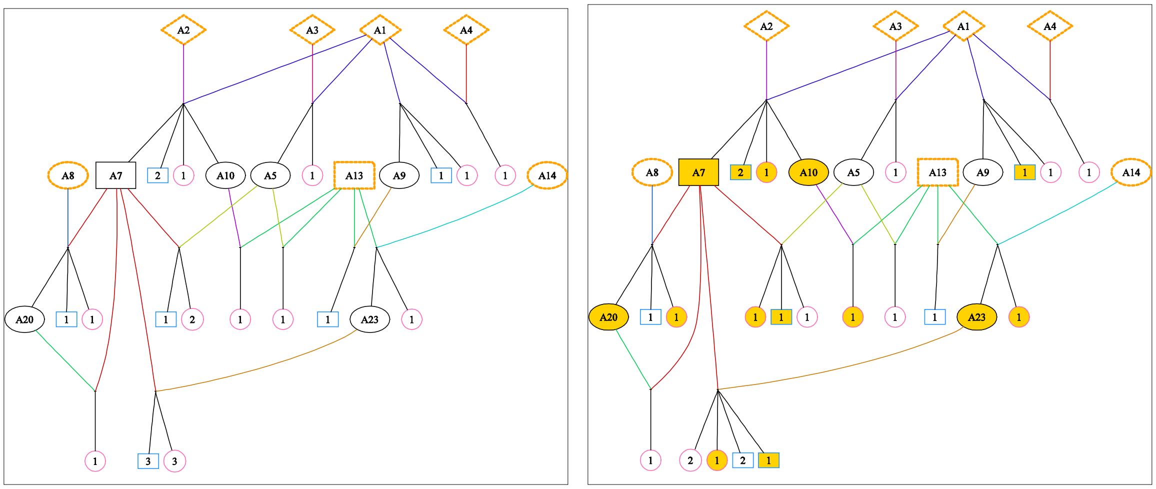


Figure 8. Compact and full pedigree drawings. (A) Compact drawing using ‘summary 1’ statement in pedigraph.dat**. (B)** Compact drawing using ‘summary 2’ statement in pedigraph.dat. **(C)** Full pedigree.

Figure 8 shows the compact pedigree drawing under each ‘summary’ option and the full pedigree drawing. In a compact pedigree, the maximum number of nodes is four for all offspring in a family. Therefore, as the family size increases, the reduction in drawing size increases. In **(A)** of **Figure 8**, the family with B1 and B2 as parents has 10 offspring, 3 'unaffected' male offspring, 2 'affected' male offspring, 3 'unaffected' female offspring and 2 'affected' female offspring. Offspring B12 and B3 are also parents and they are drawn in the pedigree. Other offspring in the family are not parents and are not drawn in the pedigree. In **(B)** of **Figure 8**, the same summary principle is implemented but the offspring counts are represented in one node with four cells. In comparison, **(B)** is more compact than **(A)** but **(A)** has a regular pedigree appearance and is more artistic than **(B)**.



(A) **(B)**

Figure 9. Compact pedigree drawings drawing with two offspring groupings. (A) Offspring are grouped by gender. **(B)** Offspring are grouped by gender and trait status.

Figure 9 shows compact pedigree drawing with two offspring groupings using the same pedigree data for drawing **Figures 1-2. (A)** 'summary 1' statement is added to the pedigraph.dat used by **Figure 1**. This results in offspring grouping by gender only. **(B)** 'summary 1' and 'affected 1' statements are added to the pedigraph.dat used by **Figures 1**. This results in offspring grouping by gender and trait status.

4.8 Comparison between compact and full pedigrees

Comparisons between compact and full pedigrees of large complex populations are available from http://animalgene.umn.edu . For the unofficial European Royalty Genealogy Database with over 48,000 individuals spanning over 100 generations, 'summary 1' reduced the pedigree size by about 2/3. For the swine melanoma population with about 2069 individuals in large families (**Figure 10**), 'summary 1' reduced the pedigree size by about 95%, i.e., the compact pedigree drawing is about only 5% as large as the full pedigree drawing. The pedigree size reduction is even more dramatic for a dairy population with large halfsib families (unpublished data).

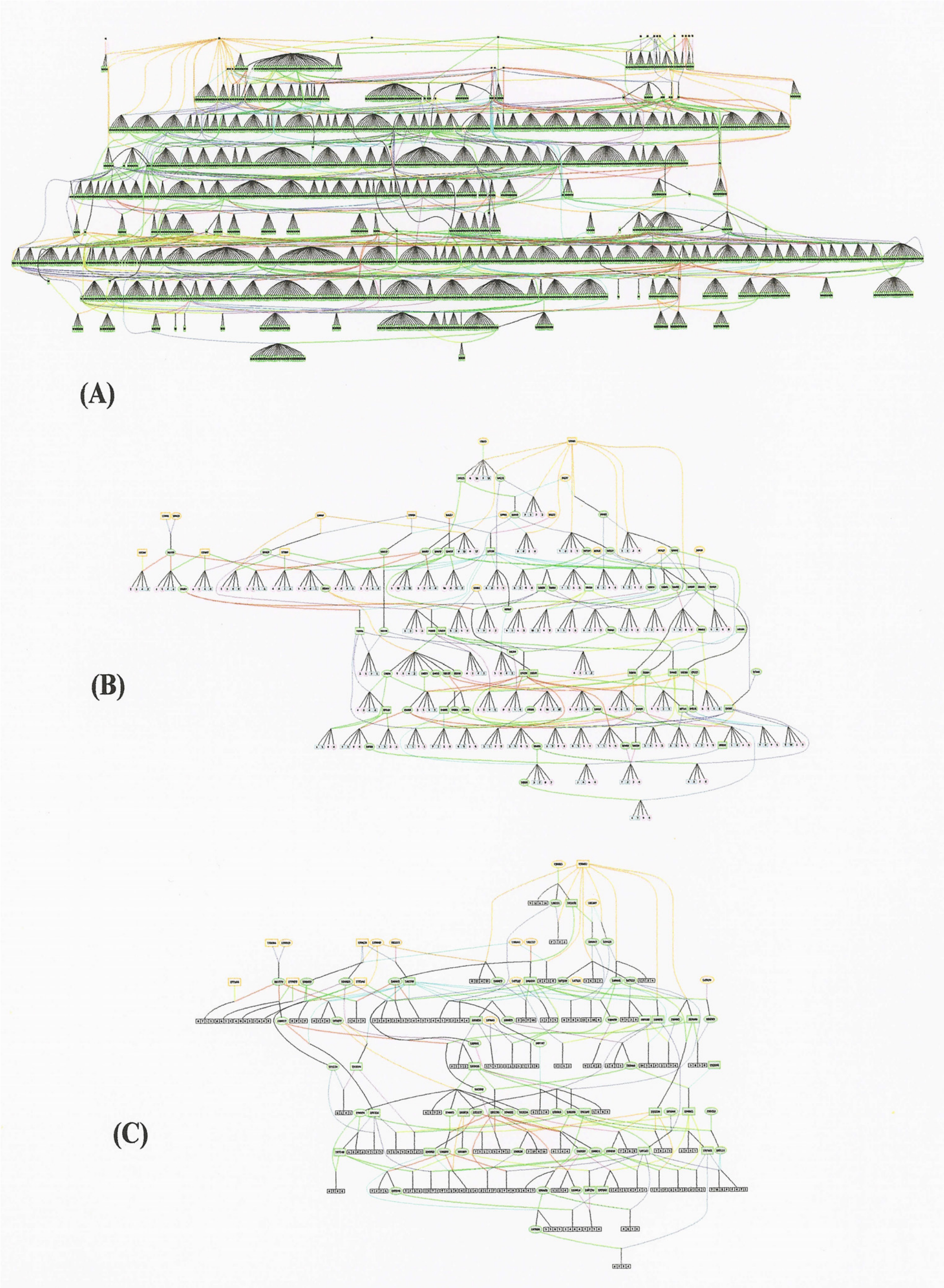


Figure 10. Comparison of compact and full pedigree drawings. (A) Full pedigree of 2069 individuals with complex inbreeding structure. **(B)** Compact pedigree drawing using ‘summary 1’. **(C)** Compact pedigree drawing using ‘summary 2’.

4.9 Pedigree extraction

Pedigraph offers pedigree extraction for a specified individual. The user may select whether the siblings, ancestors, and/or descendents of the individual for pedigree drawing. A special node is used to identify the selected individual in the pedigree drawing. Results of the extraction options are shown in **Figure 11**.

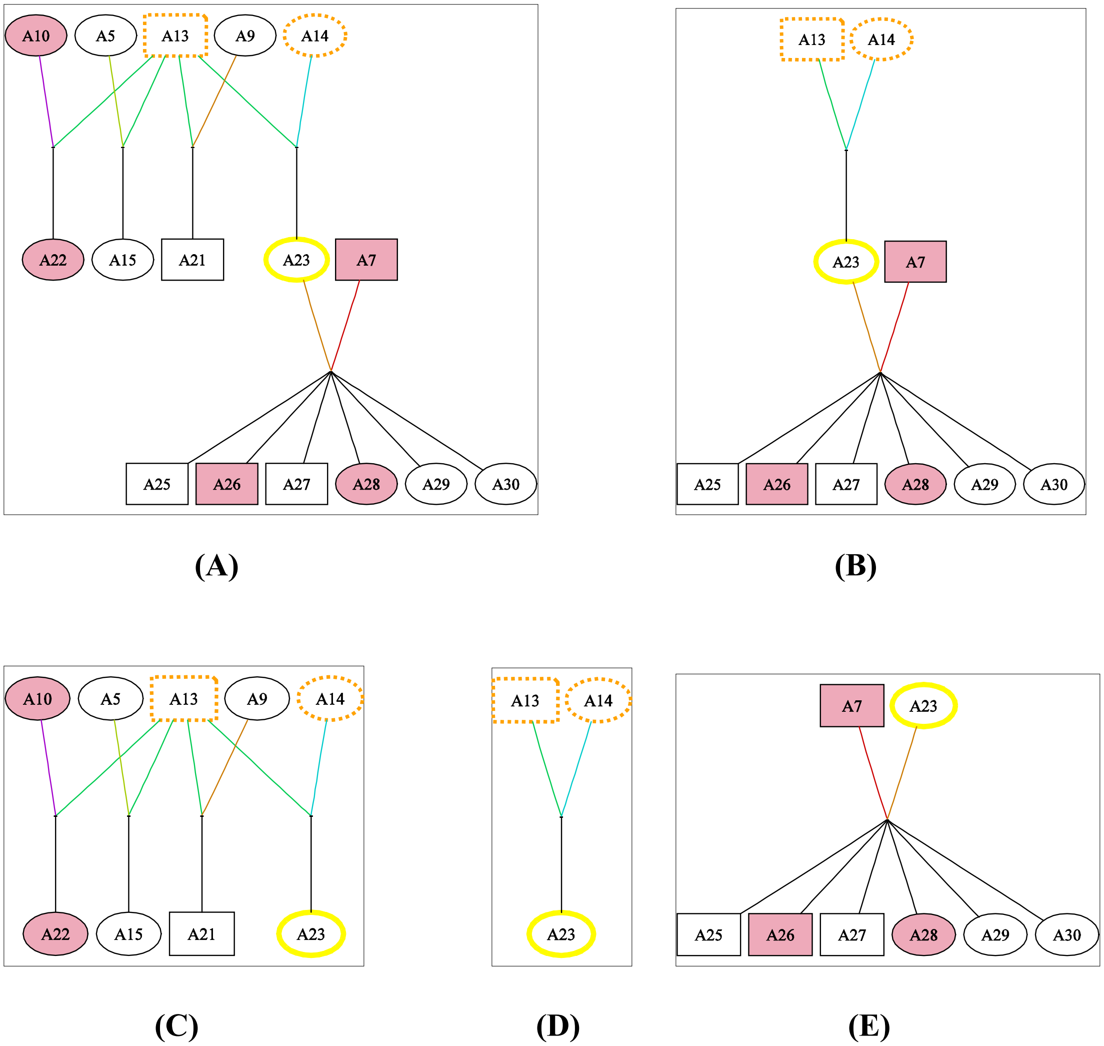


Figure 11. Pedigree extraction for individual A23 in Figure 2. (A) All relatives of individual A23 are drawn using 'individual A23' statement in pedigraph.dat**. (B)** Ancestors and descendents of individual A23 are drawn using 'individual A23', 'ancestors' and 'descendents' statements in pedigraph.dat. **(C)** Sibs of individual A23 are drawn using 'individual A23' and 'sibs' statements in pedigraph.dat. **(D)** Ancestors of individual A23 are drawn using 'individual A23' and 'ancestors' statements in pedigraph.dat. **(E)** Descendents of individual A23 are drawn using 'individual A23' and 'descendents' statements in pedigraph.dat.

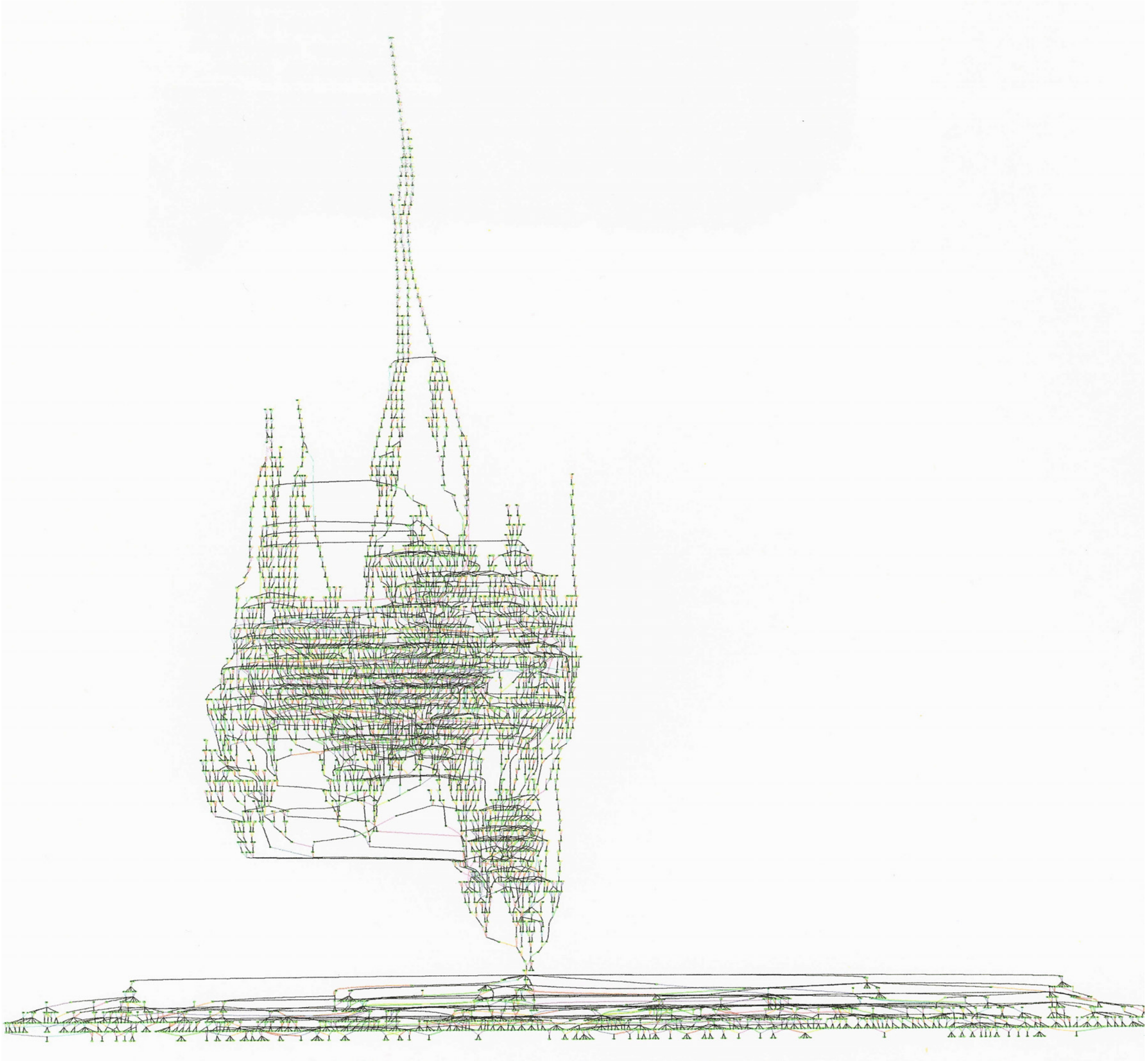


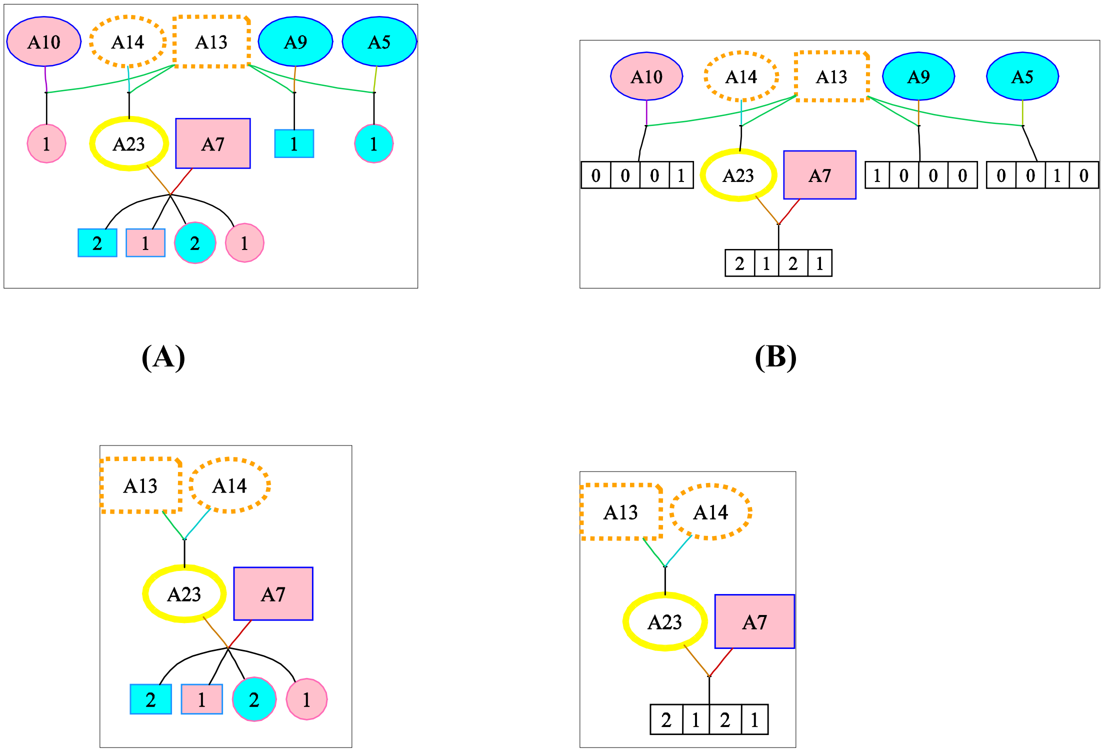
Figure 12. Pedigree of all relatives of Victoria Hanover, Queen of Britain.

Figure 12 is extracted from the full pedigree of the European Royal Genealogical Data shown in Figure 2 using the ‘individual ID’ statement. The resulting pedigree, when printed out in full size, is about 30 feet wide and 24 feet tall, which is about 11% of the main body of the full pedigree of **Figure 7**. The pedigree drawing of the Queen contains 3975 ancestors and 1148 descendents of the Queen.

The ‘extractedoutputfile’ statement in pedigraph.dat allows the pedigree data of the extracted individual to be printed in file. For example, the command ‘extractedoutputfile extract.txt’ will print the pedigree data of all relatives including the extracted individual to a file named ‘extract.txt’. The ‘extractedoutputfile’ statement can be used in combination with ‘descendants’, ‘sibs’ and ‘ancestors’.

4.10 'Summary' and 'extraction'

The 'summary' and 'extraction' options can be used jointly to further reduce the pedigree size by using the two options individually. For example, the 'summary' option reduces the full pedigree of **Figure 7** by 2/3 and the pedigree of Queen Victoria is about 11% of the full pedigree. Therefore, using the two options jointly, the size of Queen Victoria's pedigree could be reduced to 8% of the full pedigree. **Figure 13** shows examples of using these two options jointly for the same data as used by **Figure 11**.



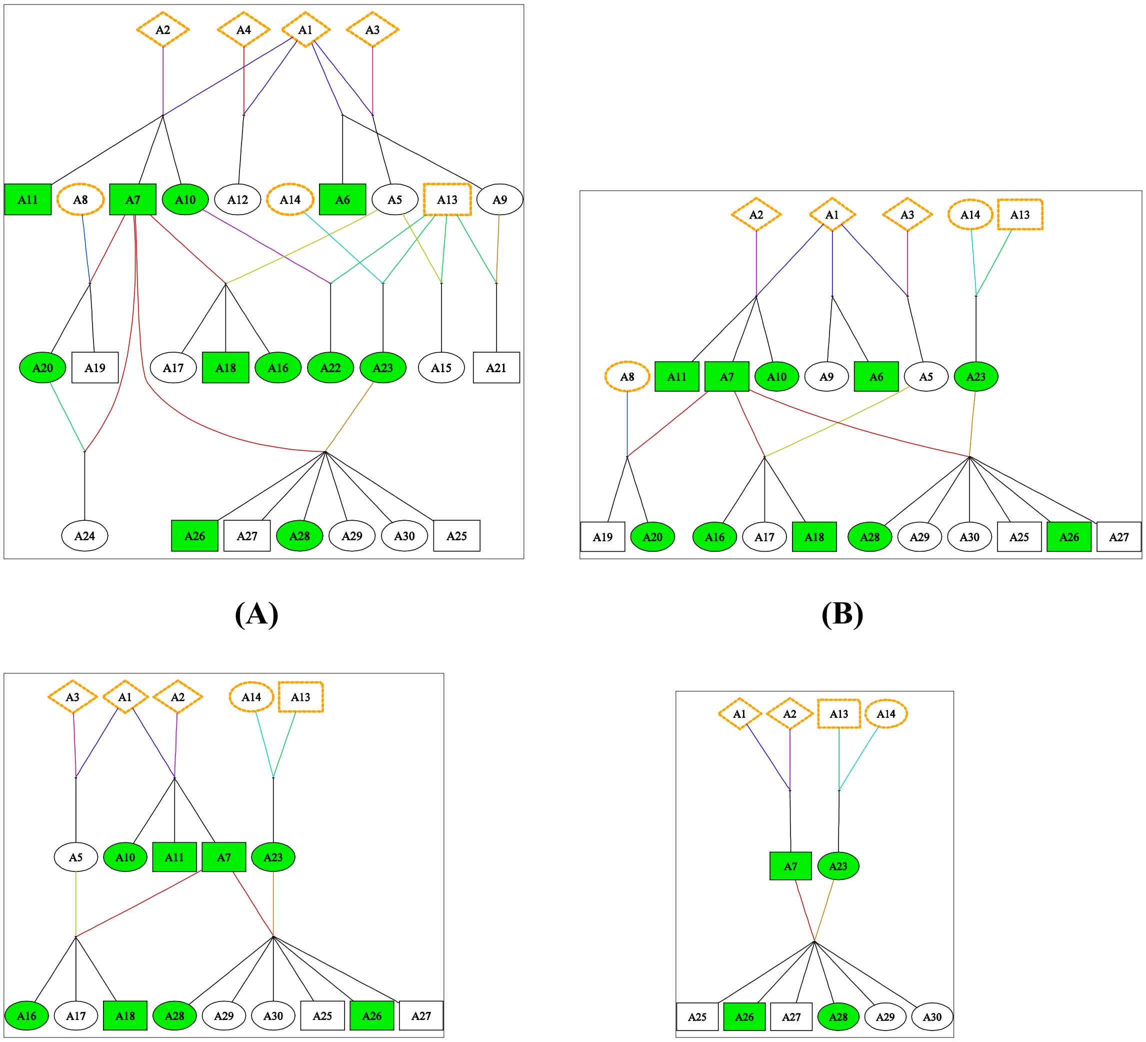
(A) **(B)**

(C) **(D)**

Figure 13. Pedigree drawing using 'summary' and 'extraction' options jointly. (A) 'individual A23' and 'summary 1' are used in pedigraph.dat. **(B)** 'individual A23' and 'summary 2' are used in pedigraph.dat. **(C)** 'individual A23', 'ancestors', 'descendents' and 'summary 1' are used in pedigraph.dat. **(D)** 'individual A23', 'ancestors', 'descendents' and 'summary 2' are used in pedigraph.dat.

4.11 Minimum family size

The user may specify the minimum number of offspring that two parents must have for the family to be included in the pedigree drawing. This is achieved by the ‘familysize number’ statement in pedigraph.dat, where ‘number’ is any integer number. For example, the statement ‘familysize 2’ will all families with at least two offspring in the pedigree drawing. **Figure 11** shows examples of pedigree drawing with minimal family size requirements.



(A)

(B)

(C)

(D)

Figure 14. Pedigree drawings with minimal family size requirement. **(A)** Full pedigree**.** **(B)** Minimal family size of 2. **(C)** Minimal family size of 3. **(D)** Minimal family size of 4.

4.12 Page Controls

Pedigraph offers a number of options to control page layout, including pagewidth, pageheight, nomargin, width, height, and title statements (**Table 1**). For a large pedigree, the pedigree drawing may be displayed on several pages under the default page size. Increasing the page size using 'pagewidth' and 'pageheight' statements may achieve the goal to print the entire pedigree drawing on one page.

5 INBREEDING COEFFICIENTS

Pedigraph has a utility to calculate inbreeding coefficients. The inbreeding coefficient of each individual can be displayed along with the individual's ID or be stored in a file.

5.1 Display of inbreeding coefficients in pedigree drawing

The 'inbreeding' statement in pedigraph.dat executes the calculation and display of inbreeding coefficient for each individual.

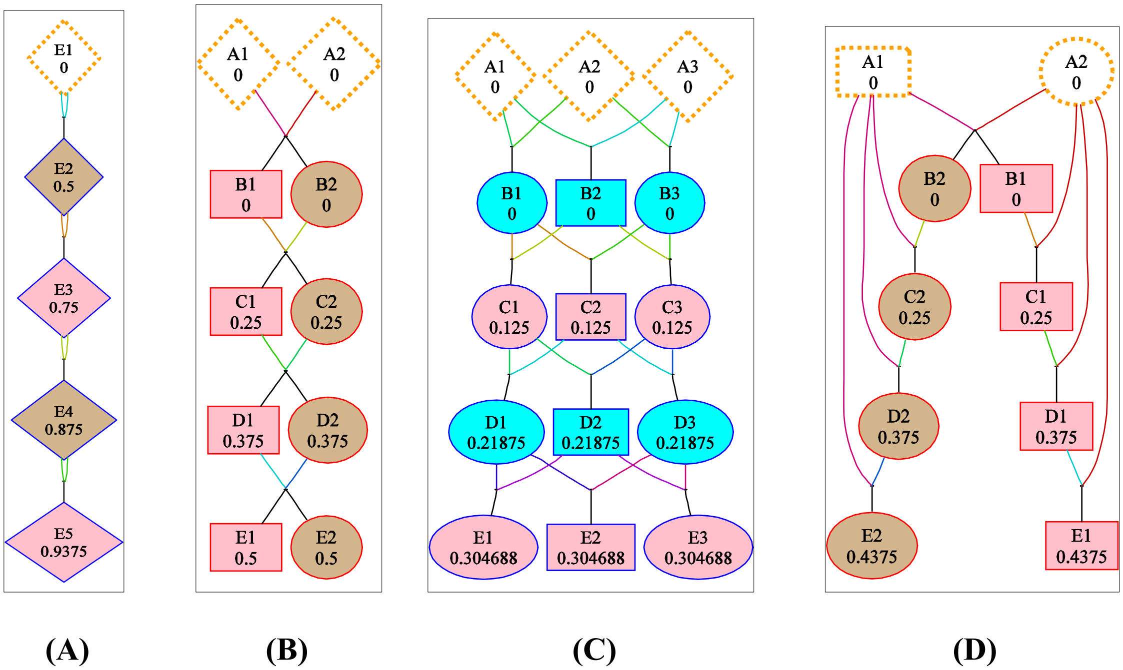


Figure 14. Display of inbreeding coefficient for regular mating systems. (A) Inbreeding coefficients from selfing. **(B)** Inbreeding coefficients from a fullsib mating system. **(C)** Inbreeding coefficients from a halfsib mating system. **(D)** Inbreeding coefficients from a backcross mating system.

In **Figure 14**, **(A-D)** are four examples of regular mating systems where the same mating scheme is applied every generation (D. S. Falconer and T. F. C. Mackay, 1996. Introduction to Quantitative Genetics, page 84). The display of inbreeding coefficients in **(A-D)** shows the rate of increase in inbreeding coefficients as the number of generations increase under the four regular mating systems, with the selfing system having the most rapid increase. **Figure 15** shows two examples of calculating and displaying inbreeding coefficient for irregular mating systems.

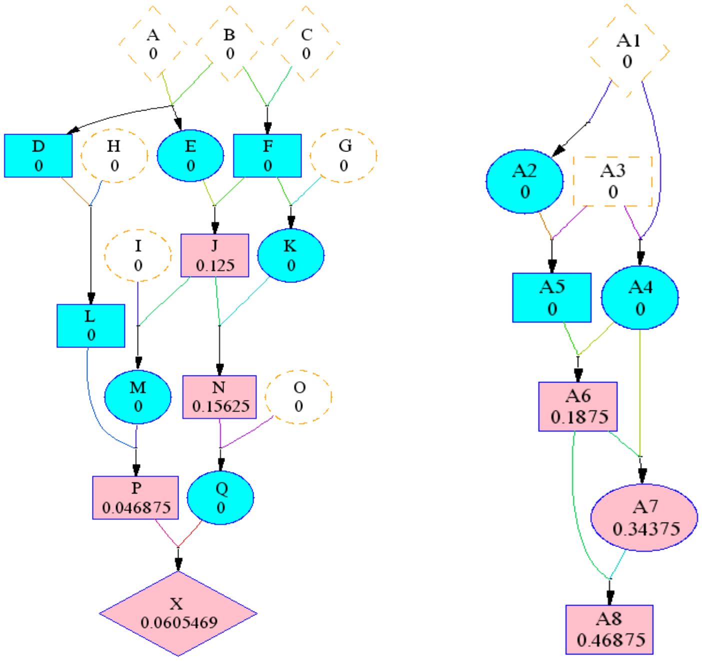


Figure 15. Display of inbreeding coefficients for two examples of irregular mating systems.

5.2 Print inbreeding coefficients to a file

The inbreeding coefficients of all individuals can be printed to a file using the statement 'inbreedingfile filename', where 'filename' is the name of the output file to store the increeding coefficients. The resulting output file has two columns, the ID and the inbreeding coefficient of each individual. On top of the file are some summary statistics. The following is output file for **Figure 14C**.

```
Inbreeding Coefficients
Max: 0.304688
Average: 0.129688
Smallest non-zero: 0.125
A1      0
A2      0
A3      0
B1      0
B2      0
B3      0
C1      0.125
C2      0.125
C3      0.125
D1      0.21875
D2      0.21875
D3      0.21875
E1      0.304688
E2      0.304688
E3      0.304688
```

6 COANCESTRY COEFFICIENTS

Pedigraph has two options to calculate and print coancestry coefficients between two individuals. Coancestry coefficient is a measure of relatedness of two individuals. A list of individuals is required as the input file. Coancestry coefficient for each pair of individuals is calculated and printed on computer screen or to a file.

6.1 Print coancestry coefficients on computer screen

In the pedigraph.dat file, the "coancestryfile" statement activates the calculation of coancestry coefficient between each pair of individuals in the file. For example, the statement ‘coancestryfile test.ids’ will calculate coancestry coefficient for each pair of individuals in test.ids file and print the coancestry coefficients on the computer screen.

6.2 Print coancestry coefficients to a file

In the pedigraph.dat file, the "coancestryoutputfile" statement following the "coancestryfile" statement will print the coancestry coefficients to an output file. For example, the statements

```
coancestryfile test.ids
coancestryoutputfile test.out
```

will calculate coancestry coefficient for each pair of individuals in test.ids file and print the coancestry coefficients in the output file named test.out.

6.3 Coancestry coefficients between a selected individual and relatives of this individual

In the pedigraph.dat file, the "coancestry" statement and the "individual" statement will print the coancestry coefficients to an output file. For example, the statements

```
individual 1000
coancestry
coancestryped test.out
```

will calculate coancestry coefficient between individual 1000 and each relative, and print the coancestry coefficients in the output file named test.out.