

DmuTrace

*A program to trace the pedigree for a sub set
of animals from a larger pedigree file.*

Per Madsen

*Dept. of Genetics and Biotechnology
Faculty of Agricultural Sciences, Aarhus University
Research Centre Foulum
DK-8830 Tjele, Denmark
Ph.: (+45) 89 99 12 16
Fax: (+45) 89 99 13 00
E-mail: per.madsen@agrsci.dk*

Program description

Trace is a program to extract a subset from a large pedigree file. The traced pedigree file can be used as input for DMU.

Initially Trace check the pedigree file for consistency, i.e. it should be possible to trace all individuals back to “base populations” (phantom parent groups) indicated by negative ID's. If day of birth is given in the format “yyyymmdd”, it is checked that there is at least a minimum specified number of days between between an animals day of birth and day of birth for its parents. Different minimum days can be specified for sire and dam.

A file with ID's of the individuals to be traced is read. First it is checked that all individuals specified for tracing are included in the total pedigree file. A pedigree file with all the specified individuals specified for tracing and all their ancestors. Additionally a pruned pedigree file is build. The pruning can be more or less aggressive. The mildest pruning removes non-informative individuals, which is individuals not specified for tracing and with only one link to individuals specified for tracing is extracted with all the specified animals and all their ancestors. More aggressive pruning can be performed by removing individuals more than a specified number of generations before individuals specified for tracing.

The program can also calculate contributions from each of the base populations to each individual and the degree of heterozygosity between all possible combination's of the base populations.

If the base populations specified in the pedigree file are phantom parent groups (PHG's), and the PHG's have contributions from different breeds, the breed contribution to the PHG's can be specified in a file. Calculation of breed contributions and heterozygosity will then be based

on breed contributions to the PHG's.

The trace programs are distributed free of charge, but the Faculty of Agricultural Sciences (DJF) owns the source code of the trace program and the subroutines in the distributed.

Great care has been put into the development in order to insure correct computations. If, however, errors are found they will be attended as quickly as possible. Correct use of the programs are entirely the responsibility of the end user and neither DJF nor the author will in any way be financially responsible for the results of use or misuse of results from the programs.

Basic program parameters

A number of basic program parameters can be set/adjusted via a configuration file named "DmuTrace.conf".

DmuTrace looks for the configuration file in the directory where it is executed. If it exist. It is read and the parameters specified overwrite the default values set in DmuTrace. The file is read using "NAMELIST-directed" IO. The format of the file is:

```
&PARM
MAX_NIV =      25,
MAX_A   = 10000000,
MAX_PHG =      500,
MIN_ACC_BD   =      0,
MAX_ACC_BD   = 20100204,
MIN_D_SBD    =      1,
MIN_D_DBD    =      1
/
```

The parameter values above are the default values set by DmuTrace and have the following meaning:

MAX_NIV	Maximum number of generations to go back when looking for base population (PHG's). If the recursive procedure for looking up ancestors exceeds this value without reaching the base population (PHG) , the animal is marked as an animal that can not be traced back to base population (PHG) . This value should be set to a value that is reasonable for the maximum number of generations that is possible for the spices in question and the time frame in the pedigree file.
MAX_A	Maximum number of animals that can be in the pedigree. This must be at least the number of animals in the total pedigree.

MAX_PHG	Maximum number of PHG's. This value must be at least the number of different PHG's in the total pedigree file
MIN_ACC_BD	Smallest value accepted as a legal day of birth
MAX_ACC_BD	Largest value accepted as a legal day of birth. If this is set to 0 (zero) the date of execution is used, which is also the default value if no configuration file is specified
MIN_D_SBD	If number of days between an animals day of birth and the day of birth of its sire is below this value, information is written to a error file.
MIN_D_DBD	If number of days between an animals day of birth and the day of birth of its dam is below this value, information is written to a error file.

Description of input read from standard input

The program reads from standard input (one line per item):

1. Name of the total pedigree file (the "Total_Ped" file).
2. Name of the file with ID's of individuals to trace (the "PROB" file)

AND OPTIONAL

3. Opt1 Opt2 Gprune

where: Opt1 = 1 -> Tracing

2 = Tracing + calculation of breed proportions and Heterozygosity

Opt2 = 1 -> Using ID's from the total pedigree file as ID's in the output files

2 -> Using renumbered ID's in the output files

(This can be used to create ID's for use in DMU and MiX99 for pedigree files using ID'S > 2**31 -1. In this case the original ID is included in the output files)

Gprune = n -> pruning to max n generations back form individuals specified for pruning

If no input line 3 is specified, the default values of 1 1 and 0 is used for Opt1, Opt2 and Gprune.

4. Name of file with proportions from base populations in PHG's

5. N_BPOP = number of base populations (breeds)

If input line 4 and 5 are specified, input line 3 must also be specified.

Description of input data sets

The pedigree file (Total_Ped), in ASCII format, must have the following information on each line:

1. ID
2. Sire ID
3. Dam ID
4. Birth date for individual ID (sort var.)

The three ID's are read and treated as long integers (64 bit). This means that the largest ID that can be handled is $2^{63} - 1 = 9\ 223\ 372\ 036\ 854\ 775\ 807$. Base populations or PHG's must have negative numbers. The Birth date (sort var.) must be in a format so animals can be accessed in birth date order (yyyymmdd, day number from some fixed date etc.). The value in the birth date column is included in the files with the traced pedigree. Thereby the files can be used directly as pedigree files for the DMU-packaged.

The "PROB" file (in ASCII format) must have the following information on each line:

1. ID (correspond to ID in the pedigree file)

If specified, the file with proportions from base populations in PHG's must have the following information on each line (in ASCII format):

1. PHG ID
2. N_BPOP+1. Proportions of genes from each of the N_BPOP in this PHG (real).

Description of output data sets

If errors are found in the input pedigree file (Total_Ped), a number of error files are produced. The error conditions, name of the error files and their contents are explained in the listing from the program.

The program produces a number of output files, which will be described in the following:

The data sets are named as the name of the "PROB" data set with an extension.

A pedigree file with all individuals specified in the "PROB" file and their ancestors. The extension for this data set is "PED".

The format of this file is ASCII and the variables depends on Opt2 of directive line 3.

<i>Var. no.</i>	<i>Opt2 = 1 Meaning</i>	<i>Opt2 = 2 Meaning</i>
1	Original ID (long integer)	Renumbered ID (integer)
2	Original sire ID (long integer)	Renumbered sire ID (integer)
3	Original dam ID (long integer)	Renumbered dam ID (integer)
4	Birth date (sort var.) (integer)	Birth date (sort var.) (integer)
5	Completeness of the pedigree expressed as “known” generations (real)	Completeness of the pedigree expressed as “known” generations (real)
6		Original ID (long integer)

A pruned pedigree file with all individuals specified in the “PROB” file and all their informative ancestors. The extension for this data set is “PRUNE”. In this file all pruned parents are allocated to the base population (PHG) -999999997.

The format of this file is ASCII and the variables depends on Opt2 of input line 3.

<i>Var. no.</i>	<i>Opt2 = 1 Meaning</i>	<i>Opt2 = 2 Meaning</i>
1	Original ID (long integer)	Renumbered ID (integer)
2	Original sire ID (long integer)	Renumbered sire ID (integer)
3	Original dam ID (long integer)	Renumbered dam ID (integer)
4	Birth date (sort var.) (integer)	Birth date (sort var.) (integer)
5		Original ID (long integer)

If no files with breed proportions in the PHG's is specified, a file with renumbered phantom parent group ID's is created. The extension for the data set named is PHG_RENUM. The format of this file is ASCII and it has 2 variables.

<i>Var. no.</i>	<i>Meaning</i>
1	Original phantom parent (base population) ID (long integer)
2	Renumbered phantom parent ID (-1* assigned base population number (integer))

If calculation of base population (PHG) contributions and heterozygosity are requested (Opt1 = 2 on the third line read from standard input), two additional files are created. One with proportions from base populations in each animal in the traced pedigree, and one with

heterozygosity. The extension for the file with breed contributions is “BCONTRIB”.

The format of this file is ASCII and the variables depends on Opt2 of input line 3.

<i>Var. no.</i>	<i>Opt2 = 1 Meaning</i>	<i>Opt2 = 2 Meaning</i>
1	Original ID (long integer)	Renumbered ID (integer)
2	Renumbered base population ID (integer)	Renumbered base population ID (integer)
3	Contribution from base population (real)	Contribution from base population (real)
4		Original ID (long integer)

The extension for the file with heterozygosity is “HET”, the format of this file is ASCII and the variables depends on Opt2 of input line 3.

<i>Var. no.</i>	<i>Opt2 = 1 Meaning</i>	<i>Opt2 = 2 Meaning</i>
1	Original ID (long integer)	Renumbered ID (integer)
2	Renumbered ID for base population 1 (integer)	Renumbered ID for base population 1 (integer)
3	Renumbered ID for base population 2 (integer)	Renumbered ID for base population 2 (integer)
4	Heterozygosity between the two base populations (real)	Heterozygosity between the two base populations (real)
5		Original ID (long integer)

Example on running the trace program

Example 1: Tracing and standard pruning

The total pedigree file (TOTAL_PEDIGREE):

```
1 -1 -2 20000101
2 -1 -2 20010201
3 -1 -2 20000301
4 -3 -4 20010401
5 -3 -4 20000501
6 -3 -4 20010601
101 1 2 20030101
102 3 4 20030201
103 5 4 20030301
104 5 4 20030401
105 5 6 20030501
201 101 102 20060101
202 103 102 20060201
203 103 104 20060301
204 105 104 20060401
206 105 102 20060501
301 201 202 20080101
302 203 202 20100601
303 203 204 20000301
```

The file with animals to trace (PROB):

```
301
302
303
```

The directive to be read from standard input:

```
TOTAL_PEDIGREE
PROB
```

The listing and produced files from the trace program is on the following pages.

```

****      *      * *      * ***** ****      **      ***      *****
* * * * *      * * *      * *      * *      * *      * *      * *      *
* * * * *      * * *      * * *      * *      * *      * *      * *      *
* * * * *      * * *      * * *      * *      * *      * *      * *      *
* * * * *      * * *      * * *      * *      * *      * *      * *      *
****      *      * *      * *      * *      * *      * *      * *      *

```

A program to extract pedigree information for a subset of individuals from a larger pedigree file

Per Madsen
Faculty of Agricultural Science, University of Aarhus
Dept. of Genetics and Biotechnology
Research Centre Foulum
DK-8830 Tjele, Denmark
Ph.: (+45) 89 99 12 16 Fax: (+45) 89 99 13 00
E-mail: per.madsen@agrsci.dk

Name of pedigree file TOTAL_PEDIGREE
Name of file with individuals to trace . PROB
Option 1 not specified, only tracing
Option 2 not specified, use original ID's

Check of day of birth assumes that day of birth is yyyymmdd format

Smallest accepted day of birth: 19000101
(Values smaller is assumed unknown and not used for)
checking age differences between progeny and parent)

Largest accepted day of birth: 20100205
(Animals with a day of birth larger is written to a)
error file)

Minimum number of days between day of birth and:
Sire day of birth: 600
Dam day of birth: 300

ERROR/WARNINGS

Individuals with possible errors in birth date

Errors can be:

Born before one or both its parents
Age of its sire < 600
Age of its dam < 300
Born before 19000101
Born after 20100205

Obs. no.	ID	Birth date	Parent ID	Parent Birth date
18	302	20100601	0	0
19	303	20000301	203	20060301
19	303	20000301	204	20060401

There are 0 errors of this type
They are in the file PROB.err8

The file has 6 variables:

1 : Obs. number
2 : ID
3 : Birth date (sort var.)
4 : Parent no. (1 = sire, 2 = dam)
5 : Parent ID
6 : Parent birth date (sort var)

For animals with birth date in yyyymmdd format, but born before or after the specified dates, variables 4, 5 and 6 is 0 (zero)

Statistics for the pedigree file

```

-----
No. of individuals ..... 19
No. of individuals inserted ..... 0
No. of phantom parent groups ..... 4
No. of individuals that should be traced ..... 3
No. of phantom parent groups with repeated information ..... 0
No. of individuals with repeated pedigree information ..... 0
No. of individuals with inconsistent pedigree information ..... 0
No. of individuals that should be traced not in pedigree file . 0
No. of individuals that can not be traced back to PHG's ..... 0
No. of these that should be traced ..... 0

```

PHANTOM PARENT GROUPS = BASE POPULATIONS(B POP)

Direct links and contribution expressed as no. of genomes and gene proportions

PHG	B POP	Sire	Dam	All in traced ped.		Marked for traced	
				Genomes	Gene prop	Genomes	Gene prop
-1	1	3	0	3.06	0.17013889	0.31	0.10416667
-2	2	0	3	3.06	0.17013889	0.31	0.10416667
-3	3	3	0	5.94	0.32986111	1.19	0.39583333
-4	4	0	3	5.94	0.32986111	1.19	0.39583333
SUM				18.00	1.00000000	3.00	1.00000000

Description of output files

A pedigree file has been created.
It contains 18 records
File name: PROB.PED

The format is ASCII and it has 5 variables

The variables are:
1. Original ID (long integer)
2. Original sire ID (long integer)
3. Original dam ID (long integer)
4. Birth date (sort var.) (integer)
5. Known generations (real)

A pruned pedigree file has been created.
It contains 13 records
File name: PROB.PRUNE

The format is ASCII and it has 4 variables

The variables are:
1. Original ID (long integer)
2. Original sire ID (long integer)
3. Original dam ID (long integer)
4. Birth date (sort var.) (integer)

A file with renumbered phantom parent groups ID's has been created.
-1 times the assigned base population number is used as the renumbered PHG ID
It contains 4 records
File name: PROB.PHG_RENUM

The format is ASCII and it has 2 variables
(2 integer).

The variables are:
1. The original PHG ID (long integer)
2. Renumbered PHG ID (integer)

EOJ

The complete traced pedigree file (PROB.PED):

1	-1	-2	20000101	0.0000000E+00
2	-1	-2	20010201	0.0000000E+00
3	-1	-2	20000301	0.0000000E+00
4	-3	-4	20010401	0.0000000E+00
5	-3	-4	20000501	0.0000000E+00
6	-3	-4	20010601	0.0000000E+00
101	1	2	20030101	1.000000
102	3	4	20030201	1.000000
103	5	4	20030301	1.000000
104	5	4	20030401	1.000000
105	5	6	20030501	1.000000
201	101	102	20060101	2.000000
202	103	102	20060201	2.000000
203	103	104	20060301	2.000000
204	105	104	20060401	2.000000
301	201	202	20080101	3.000000
302	203	202	20100601	3.000000
303	203	204	20000301	3.000000

The pruned pedigree file (PROB.PRUNE):

4	-3	-4	20010401	
5	-3	-4	20000501	
102	-999999997	4	20030201	
103	5	4	20030301	
104	5	4	20030401	
105	5	-999999997	20030501	
201	-999999997	102	20060101	
202	103	102	20060201	
203	103	104	20060301	
204	105	104	20060401	
301	201	202	20080101	
302	203	202	20100601	
303	203	204	20000301	

The file with recoded base population ID's (PROB.PHG_RENUM):

-1	-1
-2	-2
-3	-3
-4	-4

The file with errors related day of birth (PROB.err8):

18	302	20100601	0	0	0
19	303	20000301	1		203 20060301
19	303	20000301	2		204 20060401

Example 2: Tracing and aggressive pruning

The full pedigree file (TOTAL_PEDIGREE) as well as the file with individuals to trace (PROB) are the same as in example 1.

The directive to be read from standard input:

```
TOTAL_PEDIGREE
PROB
1 1 1
```

The listing and produced files from the trace program is on the following pages.

```
*****
* * * * *
* * * * *
* * * * *
* * * * *
*****
* * * * *
* * * * *
* * * * *
* * * * *
*****

A program to extract pedigree information for a subset of
individuals from a larger pedigree file

Per Madsen
Faculty of Agricultural Science, University of Aarhus
Dept. of Genetics and Biotechnology
Research Centre Foulum
DK-8830 Tjele, Denmark
Ph.: (+45) 89 99 12 16      Fax: (+45) 89 99 13 00
E-mail: per.madsen@agrsci.dk
```

```
Name of pedigree file ..... TOTAL_PEDIGREE
Name of file with individuals to trace . PROB
```

```
Option 1 specified..... 1
(Trace only)
```

```
Option 2 specified..... 1
(Using original ID's)
```

```
Pruning to max. generation back from data      1
```

Check of day of birth assumes that day of birth is yyyyymmdd format

```
Smallest accepted day of birth:      19000101
(Values smaller is assumed unknown and not used for)
checking age differences between progeny and parent)
```

```
Largest accepted day of birth:      20100205
(Animals with a day of birth larger is written to a)
error file)
```

```
Minimum number of days between day of birth and:
Sire day of birth:      600
Dam day of birth:      300
```

ERROR/WARNINGS

 Individuals with possible errors in birth date

Errors can be:

Born before one or both its parents
 Age of its sire < 600
 Age of its dam < 300
 Born before 19000101
 Born after 20100205

Obs. no.	ID	Birth date	Parent ID	Parent Birth date
18	302	20100601	0	0
19	303	20000301	203	20060301
19	303	20000301	204	20060401

There are 0 errors of this type
 They are in the file PROB.err8

The file has 6 variables:

- 1 : Obs. number
- 2 : ID
- 3 : Birth date (sort var.)
- 4 : Parent no. (1 = sire, 2 = dam)
- 5 : Parent ID
- 6 : Parent birth date (sort var)

For animals with birth date in yyyymmdd format, but born before or after the specified dates, variables 4, 5 and 6 is 0 (zero)

Statistics for the pedigree file

No. of individuals	19
No. of individuals inserted	0
No. of phantom parent groups	4
No. of individuals that should be traced	3
No. of phantom parent groups with repeated information	0
No. of individuals with repeated pedigree information	0
No. of individuals with inconsistent pedigree information	0
No. of individuals that should be traced not in pedigree file ..	0
No. of individuals that can not be traced back to PHG's	0
No. of these that should be traced	0

PHANTOM PARENT GROUPS = BASE POPULATIONS(B POP)

Direct links and contribution expressed as no. of genomes and gene proportions

PHG	B POP	Sire	Dam	All in traced ped.		Marked for traced	
				Genomes	Gene prop	Genomes	Gene prop
-1	1	3	0	3.06	0.17013889	0.31	0.10416667
-2	2	0	3	3.06	0.17013889	0.31	0.10416667
-3	3	3	0	5.94	0.32986111	1.19	0.39583333
-4	4	0	3	5.94	0.32986111	1.19	0.39583333
SUM				18.00	1.00000000	3.00	1.00000000

Description of output files

 A pedigree file has been created.
 It contains 18 records
 File name: PROB.PED

The format is ASCII and it has 5 variables

The variables are:

- 1. Original ID (long integer)
- 2. Original sire ID (long integer)
- 3. Original dam ID (long integer)
- 4. Birth date (sort var.) (integer)
- 5. Known generations (real)

A pruned pedigree file has been created.
It contains 7 records
File name: PROB.PRUNE

The format is ASCII and it has 4 variables

The variables are:
1. Original ID (long integer)
2. Original sire ID (long integer)
3. Original dam ID (long integer)
4. Birth date (sort var.) (integer)

A file with renumbered phantom parent groups ID's has been created.
-1 times the assigned base population number is used as the renumbered PHG ID
It contains 4 records
File name: PROB.PHG_RENUM

The format is ASCII and it has 2 variables
(2 integer).

The variables are:
1. The original PHG ID (long integer)
2. Renumbered PHG ID (integer)

EOJ

The complete traced pedigree file (PROB.PED) and the file with renumbered PHG ID's (PROB.PHG_RENUM) are as for example 1.

The pruned pedigree file (PROB.PRUNE) contains now only 7 individuals:

201	-999999997	-999999997	20060101
202	-999999997	-999999997	20060201
203	-999999997	-999999997	20060301
204	-999999997	-999999997	20060401
301	201	202	20080101
302	203	202	20100601
303	203	204	20000301

Example 3: Tracing and computation of breed composition and breed heterozygosity

The full pedigree file (TOTAL_PEDIGREE) as well as the file with individuals to trace (PROB) are the same as in example 1.

The directive to be read from standard input:

```
TOTAL_PEDIGREE
PROB
2 1 0
```

The listing and produced files from the trace program is on the following pages.

```
**** * * * * ***** **** ** *** *****
* * * * * * * * * * * * * * * * * * * *
* * * * * * * * * * * * * * * * * * * *
* * * * * * * * * * * * * * * * * * * *
* * * * * * * * * * * * * * * * * * * *
**** * * * * * * * * * * * * * * * * * * * *

A program to extract pedigree information for a subset of
individuals from a larger pedigree file

Per Madsen
Faculty of Agricultural Science, University of Aarhus
Dept. of Genetics and Biotechnology
Research Centre Foulum
DK-8830 Tjele, Denmark
Ph.: (+45) 89 99 12 16 Fax: (+45) 89 99 13 00
E-mail: per.madsen@agrsci.dk
```

```
Name of pedigree file ..... TOTAL_PEDIGREE
Name of file with individuals to trace . PROB

Option 1 specified..... 2
(Trace + breed proportions + heterozygosity)

Option 2 specified..... 1
(Using original ID's)

Check of day of birth assumes that day of birth is yyyymmdd format

Smallest accepted day of birth: 19000101
(Values smaller is assumed unknown and not used for)
checking age differences between progeny and parent)

Largest accepted day of birth: 20100205
(Animals with a day of birth larger is written to a)
error file)

Minimum number of days between day of birth and:
Sire day of birth: 600
Dam day of birth: 300
```

Statistics for the pedigree file

```

-----
No. of individuals ..... 19
No. of individuals inserted ..... 0
No. of phantom parent groups ..... 4
No. of individuals that should be traced ..... 3
No. of phantom parent groups with repeated information ..... 0
No. of individuals with repeated pedigree information ..... 0
No. of individuals with inconsistent pedigree information ..... 0
No. of individuals that should be traced not in pedigree file . 0
No. of individuals that can not be traced back to PHG's ..... 0
No. of these that should be traced ..... 0
  
```

PHANTOM PARENT GROUPS = BASE POPULATIONS (B POP)

Direct links and contribution expressed as no. of genomes and gene proportions

```

-----
                                All in traced ped.      Marked for traced
                                -----
                                Genomes  Gene prop      Genomes  Gene prop
-----
      PHG  B POP      Sire      Dam      Genomes  Gene prop      Genomes  Gene prop
-----
      -1    1         3         0       3.06  0.17013889      0.31  0.10416667
      -2    2         0         3       3.06  0.17013889      0.31  0.10416667
      -3    3         3         0       5.94  0.32986111      1.19  0.39583333
      -4    4         0         3       5.94  0.32986111      1.19  0.39583333
-----
SUM                                18.00  1.00000000      3.00  1.00000000
-----
  
```

Simple statistics for the traced pedigree

Contribution from base populations to all individuals

```

-----
Base population      N-obs      Average      S.D.      Min value      Max value
-----
      1              18     0.17013890     0.21200803     0.00000000     0.50000000
      2              18     0.17013890     0.21200803     0.00000000     0.50000000
      3              18     0.32986110     0.21200803     0.00000000     0.50000000
      4              18     0.32986110     0.21200803     0.00000000     0.50000000
-----
  
```

Contribution from base populations to individuals with contribution > 0

```

-----
Base population      N-obs      Average      S.D.      Min value      Max value
-----
      1              9      0.34027779     0.17430417     0.62500000E-01     0.50000000
      2              9      0.34027779     0.17430417     0.62500000E-01     0.50000000
      3             14      0.42410713     0.12542853     0.12500000         0.50000000
      4             14      0.42410713     0.12542853     0.12500000         0.50000000
-----
  
```

Heterozygosity all individuals

Base populations

```

-----
      1      2      N-obs      Average      S.D.      Min value      Max value
-----
      1      2         18     0.21354167     0.38351995     0.00000000     1.00000000
      1      3         18     0.39930556E-01     0.74110501E-01     0.00000000     0.25000000
      1      4         18     0.39930556E-01     0.74110501E-01     0.00000000     0.25000000
      2      3         18     0.39930556E-01     0.74110501E-01     0.00000000     0.25000000
      2      4         18     0.39930556E-01     0.74110501E-01     0.00000000     0.25000000
      3      4         18     0.37326390     0.36006922     0.00000000     1.00000000
-----
  
```

Heterozygosity for individuals with specific heterozygosity > 0

Base populations

```

-----
      1      2      N-obs      Average      S.D.      Min value      Max value
-----
      1      2         6      0.64062500     0.41446030     0.93750000E-01     1.00000000
      1      3         5      0.14375000     0.68465322E-01     0.62500000E-01     0.25000000
      1      4         5      0.14375000     0.68465322E-01     0.62500000E-01     0.25000000
      2      3         5      0.14375000     0.68465322E-01     0.62500000E-01     0.25000000
      2      4         5      0.14375000     0.68465322E-01     0.62500000E-01     0.25000000
      3      4        12      0.55989581     0.29389453     0.93750000E-01     1.00000000
-----
  
```

Description of output files

A pedigree file has been created.
It contains 18 records
File name: PROB.PED

The format is ASCII and it has 5 variables

The variables are:
1. Original ID (long integer)
2. Original sire ID (long integer)
3. Original dam ID (long integer)
4. Birth date (sort var.) (integer)
5. Known generations (real)

A pruned pedigree file has been created.
It contains 13 records
File name: PROB.PRUNE

The format is ASCII and it has 4 variables

The variables are:
1. Original ID (long integer)
2. Original sire ID (long integer)
3. Original dam ID (long integer)
4. Birth date (sort var.) (integer)

A file with renumbered phantom parent groups ID's has been created.
-1 times the assigned base population number is used as the renumbered PHG ID
It contains 4 records
File name: PROB.PHG_RENUM

The format is ASCII and it has 2 variables
(2 integer).

The variables are:
1. The original PHG ID (long integer)
2. Renumbered PHG ID (integer)

A file with breed (PHG) contributions has been created.
It contains 46 records
File name: PROB.BCONTRIB

The format is ASCII and it has 3 variables

The variables are:
1. Original ID (long integer)
2. Renumbered breed (PHG) ID (integer)
3. Contribution from this breed (PHG) (real)

A file with breed (PHG) heterozygosity has been created.
It contains 38 records
File name: PROB.HET

The format is ASCII and it has 4 variables

The variables are:
1. Original ID (long integer)
2. Renumbered breed (PHG) ID for breed (PHG) 1 (integer)
3. Renumbered breed (PHG) ID for breed (PHG) 2 (integer)
4. Heterozygosity with respect to these two breeds (PHG's) (real)

EOJ

The complete traced pedigree file (PROB.PED), the pruned pedigree file (PROB.PRUNE) and the file with renumbered PHG ID's (PROB.PHG_RENUM) are as for example 1.

The file with breed (PHG) contributions (PROB.BCONTRIB):

1	1	0.5000000000000000
1	2	0.5000000000000000
2	1	0.5000000000000000
2	2	0.5000000000000000
3	1	0.5000000000000000
3	2	0.5000000000000000
4	3	0.5000000000000000
4	4	0.5000000000000000
5	3	0.5000000000000000
5	4	0.5000000000000000
6	3	0.5000000000000000
6	4	0.5000000000000000
101	1	0.5000000000000000
101	2	0.5000000000000000
102	1	0.2500000000000000
102	2	0.2500000000000000
102	3	0.2500000000000000
102	4	0.2500000000000000
103	3	0.5000000000000000
103	4	0.5000000000000000
104	3	0.5000000000000000
104	4	0.5000000000000000
105	3	0.5000000000000000
105	4	0.5000000000000000
201	1	0.3750000000000000
201	2	0.3750000000000000
201	3	0.1250000000000000
201	4	0.1250000000000000
202	1	0.1250000000000000
202	2	0.1250000000000000
202	3	0.3750000000000000
202	4	0.3750000000000000
203	3	0.5000000000000000
203	4	0.5000000000000000
204	3	0.5000000000000000
204	4	0.5000000000000000
301	1	0.2500000000000000
301	2	0.2500000000000000
301	3	0.2500000000000000
301	4	0.2500000000000000
302	1	6.250000000000000E-002
302	2	6.250000000000000E-002
302	3	0.4375000000000000
302	4	0.4375000000000000
303	3	0.5000000000000000
303	4	0.5000000000000000

The file with breed (PHG) heterozygosity ((PROB.HET):

1	1	2	1.000000000000000
2	1	2	1.000000000000000
3	1	2	1.000000000000000
4	3	4	1.000000000000000
5	3	4	1.000000000000000
6	3	4	1.000000000000000
101	1	2	0.500000000000000
102	1	3	0.250000000000000
102	1	4	0.250000000000000
102	2	3	0.250000000000000
102	2	4	0.250000000000000
103	3	4	0.500000000000000
104	3	4	0.500000000000000
105	3	4	0.500000000000000
201	1	2	0.250000000000000
201	1	3	0.125000000000000
201	1	4	0.125000000000000
201	2	3	0.125000000000000
201	2	4	0.125000000000000
202	1	3	0.125000000000000
202	1	4	0.125000000000000
202	2	3	0.125000000000000
202	2	4	0.125000000000000
202	3	4	0.250000000000000
203	3	4	0.500000000000000
204	3	4	0.500000000000000
301	1	2	9.37500000000000E-002
301	1	3	0.156250000000000
301	1	4	0.156250000000000
301	2	3	0.156250000000000
301	2	4	0.156250000000000
301	3	4	9.37500000000000E-002
302	1	3	6.25000000000000E-002
302	1	4	6.25000000000000E-002
302	2	3	6.25000000000000E-002
302	2	4	6.25000000000000E-002
302	3	4	0.375000000000000
303	3	4	0.500000000000000

Example 4: Tracing and computation of breed composition and breed heterozygosity based on specified breed proportions in phantom parent groups.

The full pedigree file (TOTAL_PEDIGREE) as well as the file with individuals to trace (PROB) are the same as in example 1. In addition a file (B-PROB) containing breed proportions in each Phantom Parent Group (PHG) is needed.

The directive to be read from standard input:

```
TOTAL_PEDIGREE
PROB
2 1 0
B_PROP
3
```

The last 3 lines read from standard input contains the name of the file with breed proportions (here B_PROP) and the number of breeds (here 3)

The file with breed proportions (B_PROB) must contains one line PHG defined in the total pedigree. The first column contains the PHG ID (negative number), followed by as many columns, as there are breeds contributing to the population. For this example, there are defined 4 PHG's and there are 3 breeds contributing to the population. PHG -1, -2 and -3 are pure breed of breed 1, 2 and 3 respectively while PHG -4 is a cross with contribution .25 from breed 1 and 2 and .5 from breed 3. The file (B_PROB) content is:

```
-1 1 0 0
-2 0 1 0
-3 0 0 1
-4 .25 .25 .5
```

The listing and produced files from the trace program is on the following pages.

```

**** * * * * ***** **** ** *** *****
* * * * **** * * * * * * * * * *
* * * * * * * * * * * * * * * *
* * * * * * * * * * * * * * * *
* * * * * * * * * * * * * * * *
**** * * * * * * * * * * * * * *

```

A program to extract pedigree information for a subset of individuals from a larger pedigree file

Per Madsen
Faculty of Agricultural Science, University of Aarhus
Dept. of Genetics and Biotechnology
Research Centre Foulum
DK-8830 Tjele, Denmark
Ph.: (+45) 89 99 12 16 Fax: (+45) 89 99 13 00
E-mail: per.madsen@agrsci.dk

Name of pedigree file TOTAL_PEDIGREE
Name of file with individuals to trace . PROB

Option 1 specified..... 2
(Trace + breed proportions + heterozygosity)

Option 2 specified..... 1
(Using original ID's)

Name of file with proportions in PHG's ..B_PROP
No. of base populations 3

Check of day of birth assumes that day of birth is yyyymmdd format

Smallest accepted day of birth: 19000101
(Values smaller is assumed unknown and not used for)
checking age differences between progeny and parent)

Largest accepted day of birth: 20100205
(Animals with a day of birth larger is written to a)
error file)

Minimum number of days between day of birth and:
Sire day of birth: 600
Dam day of birth: 300

Statistics for the pedigree file

```

No. of individuals ..... 19
No. of individuals inserted ..... 0
No. of phantom parent groups ..... 4
No. of individuals that should be traced ..... 3
No. of phantom parent groups with repeated information ..... 0
No. of individuals with repeated pedigree information ..... 0
No. of individuals with inconsistent pedigree information ..... 0
No. of individuals that should be traced not in pedigree file . 0
No. of individuals that can not be traced back to PHG's ..... 0
No. of these that should be traced ..... 0

```

PHANTOM PARENT GROUPS

Direct links and contribution expressed as no. of genomes and gene proportions

PHG	As Sire	As Dam	All in traced ped.		Marked for traced	
			Genomes	Gene prop	Genomes	Gene prop
-1	3	0	3.06	0.17013889	0.31	0.10416667
-2	0	3	3.06	0.17013889	0.31	0.10416667
-3	3	0	5.94	0.32986111	1.19	0.39583333
-4	0	3	5.94	0.32986111	1.19	0.39583333
SUM			18.00	1.00000000	3.00	1.00000000

BASE POPULATIONS

Contribution expressed as genomes and gene proportions

Base population	All		Traced	
	Genomes	Gene prop	Genomes	Gene prop
1	4.55	0.25260417	0.61	0.20312500
2	4.55	0.25260417	0.61	0.20312500
3	8.91	0.49479167	1.78	0.59375000
SUM	18.00	1.00000000	3.00	1.00000000

Simple statistics for the traced pedigree

Contribution from base populations to all individuals

Base population	N-obs	Average	S.D.	Min value	Max value
1	18	0.25260416	0.15900603	0.12500000	0.50000000
2	18	0.25260416	0.15900603	0.12500000	0.50000000
3	18	0.49479166	0.31801206	0.00000000	0.75000000

Contribution from base populations to individuals with contribution > 0

Base population	N-obs	Average	S.D.	Min value	Max value
1	18	0.25260416	0.15900603	0.12500000	0.50000000
2	18	0.25260416	0.15900603	0.12500000	0.50000000
3	14	0.63616073	0.18814279	0.18750000	0.75000000

Heterozygosity all individuals

Base populations

1	2	N-obs	Average	S.D.	Min value	Max value
1	2	18	0.24641927	0.36916602	0.00000000	1.00000000
1	3	18	0.17903645	0.10952467	0.00000000	0.37500000
2	3	18	0.17903645	0.10952467	0.00000000	0.37500000

Heterozygosity for individuals with specific heterozygosity > 0

Base populations

1	2	N-obs	Average	S.D.	Min value	Max value
1	2	15	0.29570311	0.38713410	0.31250000E-01	1.00000000
1	3	14	0.23018973	0.54826066E-01	0.18750000	0.37500000
2	3	14	0.23018973	0.54826066E-01	0.18750000	0.37500000

Description of output files

A pedigree file has been created.
It contains 18 records
File name: PROB.PED

The format is ASCII and it has 5 variables

The variables are:

1. Original ID (long integer)
2. Original sire ID (long integer)
3. Original dam ID (long integer)
4. Birth date (sort var.) (integer)
5. Known generations (real)

A pruned pedigree file has been created.
It contains 13 records
File name: PROB.PRUNE

The format is ASCII and it has 4 variables

The variables are:
1. Original ID (long integer)
2. Original sire ID (long integer)
3. Original dam ID (long integer)
4. Birth date (sort var.) (integer)

A file with breed (PHG) contributions has been created.
It contains 50 records
File name: PROB.BCONTRIB

The format is ASCII and it has 3 variables

The variables are:
1. Original ID (long integer)
2. Renumbered breed (PHG) ID (integer)
3. Contribution from this breed (PHG) (real)

A file with breed (PHG) heterozygosity has been created.
It contains 43 records
File name: PROB.HET

The format is ASCII and it has 4 variables

The variables are:
1. Original ID (long integer)
2. Renumbered breed (PHG) ID for breed (PHG) 1 (integer)
3. Renumbered breed (PHG) ID for breed (PHG) 2 (integer)
4. Heterozygosity with respect to these two breeds (PHG's) (real)

EOJ

The complete traced pedigree file (PROB.PED), the pruned pedigree file (PROB.PRUNE) and the file with renumbered PHG ID's (PROB.PHG_RENUM) are as for example 1.

The file with breed (PHG) contributions (PROB.BCONTRIB):

```
1 1 0.5000000000000000
1 2 0.5000000000000000
2 1 0.5000000000000000
2 2 0.5000000000000000
3 1 0.5000000000000000
3 2 0.5000000000000000
4 1 0.1250000000000000
4 2 0.1250000000000000
4 3 0.7500000000000000
5 1 0.1250000000000000
5 2 0.1250000000000000
5 3 0.7500000000000000
6 1 0.1250000000000000
6 2 0.1250000000000000
6 3 0.7500000000000000
101 1 0.5000000000000000
101 2 0.5000000000000000
102 1 0.3125000000000000
102 2 0.3125000000000000
102 3 0.3750000000000000
103 1 0.1250000000000000
103 2 0.1250000000000000
103 3 0.7500000000000000
104 1 0.1250000000000000
104 2 0.1250000000000000
104 3 0.7500000000000000
```

105	1	0.125000000000000
105	2	0.125000000000000
105	3	0.750000000000000
201	1	0.406250000000000
201	2	0.406250000000000
201	3	0.187500000000000
202	1	0.218750000000000
202	2	0.218750000000000
202	3	0.562500000000000
203	1	0.125000000000000
203	2	0.125000000000000
203	3	0.750000000000000
204	1	0.125000000000000
204	2	0.125000000000000
204	3	0.750000000000000
301	1	0.312500000000000
301	2	0.312500000000000
301	3	0.375000000000000
302	1	0.171875000000000
302	2	0.171875000000000
302	3	0.656250000000000
303	1	0.125000000000000
303	2	0.125000000000000
303	3	0.750000000000000

The file with breed (PHG) heterozygosity ((PROB.HET):

1	1	2	1.0000000000000000
2	1	2	1.0000000000000000
3	1	2	1.0000000000000000
4	1	3	0.2500000000000000
4	2	3	0.2500000000000000
5	1	3	0.2500000000000000
5	2	3	0.2500000000000000
6	1	3	0.2500000000000000
6	2	3	0.2500000000000000
101	1	2	0.5000000000000000
102	1	2	0.1250000000000000
102	1	3	0.3750000000000000
102	2	3	0.3750000000000000
103	1	2	3.125000000000000E-002
103	1	3	0.1875000000000000
103	2	3	0.1875000000000000
104	1	2	3.125000000000000E-002
104	1	3	0.1875000000000000
104	2	3	0.1875000000000000
105	1	2	3.125000000000000E-002
105	1	3	0.1875000000000000
105	2	3	0.1875000000000000
201	1	2	0.3125000000000000
201	1	3	0.1875000000000000
201	2	3	0.1875000000000000
202	1	2	7.812500000000000E-002
202	1	3	0.2812500000000000
202	2	3	0.2812500000000000
203	1	2	3.125000000000000E-002
203	1	3	0.1875000000000000
203	2	3	0.1875000000000000
204	1	2	3.125000000000000E-002
204	1	3	0.1875000000000000
204	2	3	0.1875000000000000
301	1	2	0.1777343750000000
301	1	3	0.2695312500000000
301	2	3	0.2695312500000000
302	1	2	5.468750000000000E-002
302	1	3	0.2343750000000000
302	2	3	0.2343750000000000
303	1	2	3.125000000000000E-002
303	1	3	0.1875000000000000
303	2	3	0.1875000000000000