Blood group determination of Danish Jersey cattle in the F-blood group system

Lene Kolind Rasmussen
Dina Foulum
P.O.box 23, 8830 Tjele, Denmark
E-mail: kolind@dina.foulum.min.dk

Introduction

Bovine blood typing is carried out in many laboratories throughout the world as a regular part of modern cattle breeding programs. The purpose is primarily parentage verification for pedigree registration. The introduction of embryo transplantation technology and increasing international trade with semen and embryos have stressed the importance of sophisticated methods for individual identification and parentage control in cattle.

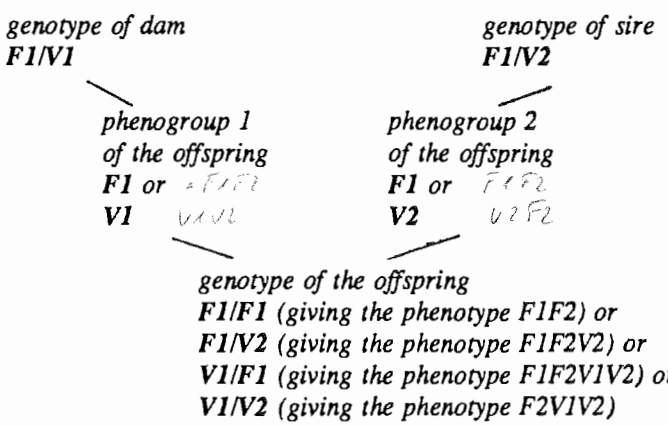
In Denmark 52 different blood group factors are used for blood typing of cattle, controlled by 10 different blood group systems (named A, B, C, F, L, M, S, Z, R' and T'). In eight of these systems the blood group determination is relatively simple (controlling from one and up to 4 blood group factors only). However, two of the systems (B- and C-) are rather complicated, controlling 26 and 10 of the above mentioned 52 blood group factors, respectively.

This note supplements an earlier paper (Rasmussen, 1992), which contains a description of using causal probabilistic network (CPN) to determine the genotypes and verify the parentage of Danish Jersey cattle. In Rasmussen (1992) it was difficult to describe the model in detail because the technique was applied to one of the most complicated blood group systems (the B-blood group system). This paper is meant as a tutorial, providing a more comprehensive description of the CPN applied to the more manageable F-blood group system.

Blood group determination

To determine blood groups of cattle each blood group factor is observed by measuring the reaction between an antibody to the blood group factor and the blood group factor itself. The reactions are measured by a photometer (lysis values) and expressed by integer values from 0 through 7, where 7 is a very high reaction (the blood group factor is present) and 0 is no reaction (the blood group factor is absent).

In the F-blood group system, 4 different blood group factors (F1, F2, V1 and V2) are used. The phenotype is a listing of the observed blood group factors (e.g., F1F2V2). The phenotype is an expression of the genotype which is composed of two alleles (the genotype is denoted **x/x**). Each allele contains a subset of the blood group factors contained in the phenotype. The blood group factors occur, however, only in a small number of combinations, the so-called phenogroups, which represent from none to several blood group factors inherited as a unit (see Figure 1). In the F-blood group system, 3 phenogroups are known, **F1F2**, **V1V2** and **V2F2**, (in practice, the phenogroups in the F-blood group system are usually denoted by **F1**, **V1** and **V2**, where the occurrence of F2 is implied in **F1** and **V2**, and the occurrence of V2 is implied in **V1**).



As an example: If the phenotype of the offspring, given by the observed blood group factors, is F1F2V2, then is the genotype given by F1/V2.

Figure 1: An example of the connection between phenotypes, phenogroups and genotypes.

Prototype

The overall structure of a causal probabilistic network to determine genotypes and verify assumed parentage from one blood group system is shown in Figure 2 (the structure of the system is independent of the blood group systems and the breed of cattle). The biological background is the inheritance of blood groups (one phenogroup is inherited from each parent), and assumptions concerning the risk of misstated parents. The input to the model is the set of lysis values obtained by the photometer from bloodtyping the offspring, and the phenogroups of the assumed parents of the offspring (if these are known).

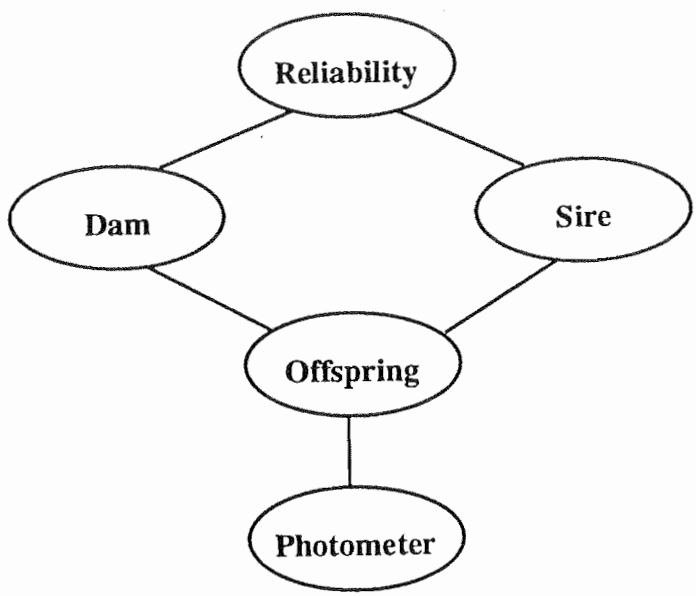


Figure 2: The overall structure of a causal model for determining the genotypes of cattle.

In the overall structure the concept "reliability" represents the nodes related to parental error and the implied assumption about the parents. "Dam" contains nodes representing the phenogroups of the real dam and the phenogroups of the stated dam. Similarly for "Sire". "Offspring" represents the phenogroups of the offspring, the genotype of the

offspring and the blood group factors of the offspring. "Photometer" contains nodes representing the lysis values as measured by the photometer.

Figure 3a. shows the causal probabilistic network for blood group determination and testing the parentage in the F-blood group system. The inputs to the system are marked as node which are hatched.

To each node in the CPN, a set of mutually exclusive states is associated. The states are shown in Figure 3b. Starting from the top of the model:

"Parental error" has 4 states corresponding to the possibilities: Both parents incorrect, sire incorrect, dam incorrect, and no error.

"Dam correct" and "Sire correct" each has two states (yes and no).

All nodes containing phenogroups have 3 different states corresponding to the number of known phenogroups in the F-blood group system of Danish Jersey Cattle (**F1**, **V1** and **V2**). As earlier mentioned: The phenogroups **F1** and **V2** correspond to the occurrence of both the blood group factors itself and the blood group factor F2, as the phenogroup **V1** corresponds to the occurrence of both the blood group factors V1 and V2.

"Genotype" has 6 states corresponding to the number of possible genotypes in the F-blood group system (**F1/F1**, **F1/V1**, **F1/V2**, **V1/V1**, **V1/V2** and **V2/V2**).

The four nodes representing the blood group factors in the F-blood group system each has two states (present and absent).

The four nodes representing the lysis factor for the blood group factor, each has 8 states corresponding to the clutting levels 0 to 7 measured by the photometer.

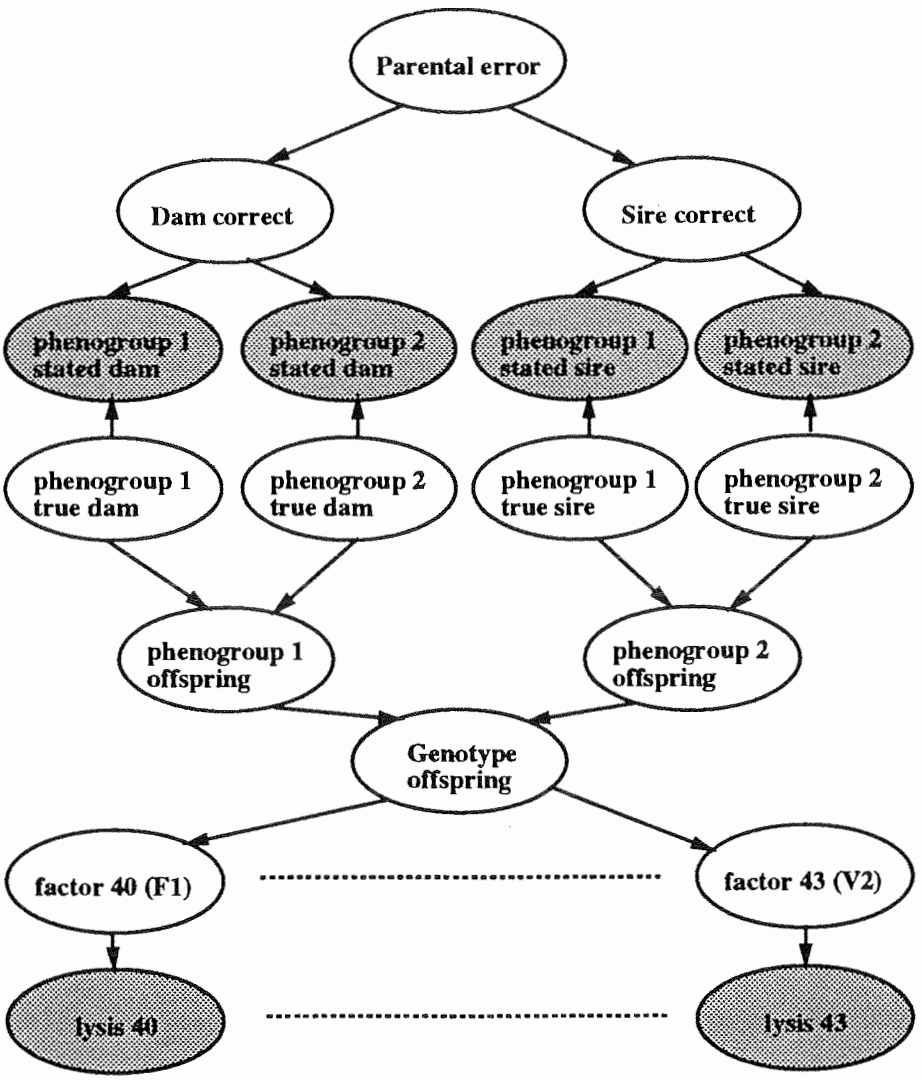


Figure 3a: Causal probabilistic network for determining the genotype and verifying parentage in the F-blood group system (input to the system is marked as hatched nodes).

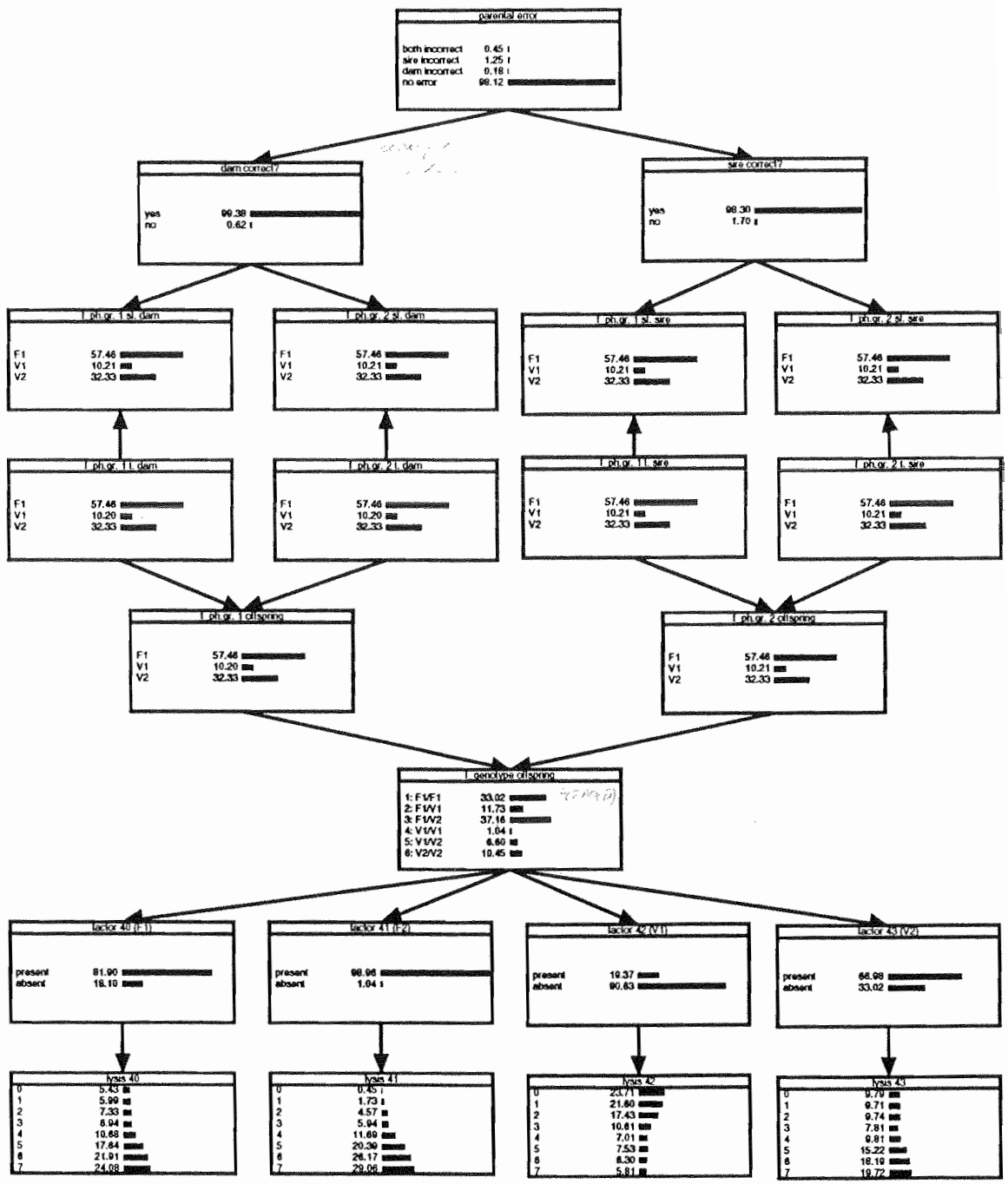


Figure 3b: Causal probabilistic network for determining the genotype and verify the parentage in the F-blood group system (with the states of each node and the initial probabilities).

In the following part of the paper it will be referred to tables placed in Appendix A.

The orphan nodes in the network are assigned unconditional probabilities. In the node "parental error", the distribution of the parental errors is determined from previously cases (Table 1). The nodes concerning the true parents ("phenogroup 1 of true dam", "phenogroup 2 of true dam", "phenogroup 1 of true sire", and "phenogroup 2 of true sire") have the distribution of the phenogroups of the F-blood group system in the population of Danish Jersey cattle, Table 2, (these distributions were determined at the blood group laboratory, Research Centre Foulum).

For all other nodes in the network conditional probabilities are specified. Table 3 and Table 4 display the probabilities of dam and sire being correct, logically determined from the parental error. The conditional distribution of the phenogroup of the stated sire, given the sire is correct and the phenogroup of the true sire, is logically determined, e.g., if phenogroup of true sire has the state F1, then the phenogroup of stated sire also has the state F1. Given the sire is incorrect, it is the distribution of the phenogroups in the population of Danish Jersey cattle, as above (Table 5).

The conditional probabilities of each phenogroup of offspring, given the phenogroups of one of the parents, are given by the random genetic transmissions as shown in Table 6.

The conditional probability of the genotype (Table 7), given the phenogroups of the offspring, is logical, from the states of the phenogroups (because the state **F1/F1** of the genotype, is determined when both phenogroups has the state **F1**, the state **F1/V1** of the genotype, is determined when one of the phenogroups has the state **F1** and the other has the state **V1**, corresponding for the other states of the genotype).

In Table 8 to 11 the conditional probabilities for each blood group factor, given the genotype, are described. The probabilities are logically determined from the genotype.

The relationship between "factor n" and "lysis n" (Table 12) are given by the following model: If a blood group factor is present there is a low probability of the lysis factor being zero. This probability increases exponentially to lysis factor 7. This is reversed if the blood group factor is absent. It has to be mentioned, that the tables are constructed from rules of thumb. In a final system, it would be desirable to determine the tables after a statistical analysis of data from the photometer and the manually determined phenotype.

Examples

The causal probabilistic network is implemented with the UNIX-based software package, Hugin (© Hugin Expert Ltd., Andersen et al., 1989). Appendix C shows the specification file used to the blood group determination of the F-blood group system.

Figure B1 and B2 (Appendix B) display the results, given by Hugin, of a single test case corresponding to two different sort of tests (one where the input to the system was the figures obtained by the photometer (lysis factors) only, and one where also the phenogroups of the stated parents were used as input), Rasmussen, 1992. As a remark to the figures: Input to the CPN can be recognized with blank bars and and the filled bars represent the probabilities calculated by Hugin. As it can be seen in Figure B1 the following input is used: lysis 40 has the state 4, lysis 41 has the state 5, lysis 42 has the state 0 and lysis 43 has the state 3. This input gives high probabilities to two states of the genotype, **F1/F1** and **F1/V2**, and it is undecided which of the states is the correct one. When we add the information about the phenogroups of the stated parents (Figure B2), only one state with high probability remains. Thus we conclude that the genotype probably is **F1/V2** and that there is no reason to question the parentage.

References

Andersen, S.K., Olesen, K.G., Jensen, F.V. & Jensen, F., 1989. HUGIN - a shell for Building Bayesian Universes for Expert Systems. Proceedings 11th International Joint Conference on Artificial Intelligence, Detroit.

Rasmussen, L.K., 1992. A causal probabilistic network for blood group determination of Danish Jersey cattle. Dina Research Report, No. 3.

Appendix A:

Tables with the probabilities used
in the causal probabilistic network
to determine blood groups
and verify the parentage in Danish Jersey Cattle.

In this appendix the following abbreviations are used:

ph.gr.: phenogroup
t.: true
st.: stated

Table 1: The probability distribution of parental error in the population of Danish Jersey cattle. **P**(*parental error*).

both incorrect	sire incorrect	dam incorrect	no error
0.45	1.25	0.18	98.12

Table 2: The probability distribution of phenogroups in the population of Danish Jersey cattle. **P**(*ph.gr. 1 of t. dam*)

F1	V1	V2
58	10	32

Corresponding for **P**(*ph.gr. 2 of t. dam*), **P**(*ph.gr. 1 of t. sire*) and **P**(*ph.gr. 2 of t. sire*).

Table 3: Conditional probability distribution of dam correct, given parental error. **P**(*dam correct/parental error*)

dam correct		
parental error	yes	no
both incorrect	0	100
sire incorrect	100	0
dam incorrect	0	100
no error	100	0

Table 4: Conditional probability distribution of sire correct, given parental error. **P**(*sire correct/parental error*)

sire correct		
parental error	yes	no
both incorrect	0	100
sire incorrect	0	100
dam incorrect	100	0
no error	100	0

Table 5: The conditional probability distribution of phenogroup 1 of the stated dam given dam correct and phenogroup 1 of the true dam.
P(*ph.gr. 1 st. dam|dam correct, ph.gr. 1 t. dam*)

		ph.gr. 1 st. dam		
dam correct	*ph.gr. 1 t. dam*	F1	V1	V2
yes	F1	100	0	0
yes	V1	0	100	0
yes	V2	0	0	100
no	F1	58	10	32
no	V1	58	10	32
no	V2	58	10	32

Corresponding tables for
P(*ph.gr. 2 st. dam/correct dam, ph.gr. 2 t. dam*)
P(*ph.gr. 1 st. sire/correct dam, ph.gr. 2 t. sire*)
P(*ph.gr. 2 st. sire/correct dam, ph.gr. 2 t. sire*)

Table 6: The conditional probability distribution of phenogroup 1 of offspring given the phenogroup 1 and phenogroup 2 of the true dam.
P(*ph.gr. 1 offspring|ph.gr. 1 t. dam, ph.gr. 2 t. dam*)

		ph.gr. 1 offspring		
ph.gr. 1 t. dam	*ph.gr. 2 t. dam*	F1	V1	V2
F1	F1	100	0	0
F1	V1	50	50	0
F1	V2	50	0	50
V1	F1	50	50	0
V1	V1	0	100	0
V1	V2	0	50	50
V2	F1	50	0	50
V2	V1	0	50	50
V2	V2	0	0	100

A corresponding table for
P(*ph.gr. 2 offspring|ph.gr. 1 t. sire, ph.gr. 2 t. sire*)

Table 7: The conditional probability distribution of the genotype given phenogroup 1 and phenogroup 2 of offspring. **P**(*F genotype|ph.gr. 1 offspring, ph.gr. 2 offspring*)

F genotype							
ph. gr. 1	*ph. gr. 2*	F1/F1	F1/V1	V1/V1	F1/V2	V1/V2	V2/V2
F1	F1	100	0	0	0	0	0
F1	V1	0	100	0	0	0	0
F1	V2	0	0	0	100	0	0
V1	F1	0	100	0	0	0	0
V1	V1	0	0	100	0	0	0
V1	V2	0	0	0	0	100	0
V2	F1	0	0	0	100	0	0
V2	V1	0	0	0	0	100	0
V2	V2	0	0	0	0	0	100

Table 8: The conditional probability distribution of the blood group factor F1 given the genotype of the F-blood group system. **P**(***F1|F genotype***)

blood group factor F1		
F genotype	present	absent
F1/F1	100	0
F1/V1	100	0
V1/V1	0	100
F1/V2	100	0
V1/V2	0	100
V2/V2	0	100

Table 9: The conditional probability distribution of the blood group factor F2 given the genotype of the F-blood group system. **P(*F2|F genotype*)**

	blood group factor F2	
F genotype	present	absent
F1/F1	100	0
F1/V1	100	0
V1/V1	0	100
F1/V2	100	0
V1/V2	100	0
V2/V2	100	0

Table 10: The conditional probability ditribution of the blood group factor V1 given the genotype of the F-blood group system. **P(*V1|F genotype*)**

	blood group factor V1	
F genotype	present	absent
F1/F1	0	100
F1/V1	100	0
V1/V1	100	0
F1/V2	0	100
V1/V2	100	0
V2/V2	0	100

Table 11: The conditional probability distribution of the blood group factor V2 given the genotype of the F-blood group system. **P(*V2|F genotype*)**

	blood group factor V2	
F genotype	present	absent
F1/F1	0	100
F1/V1	100	0
V1/V1	100	0
F1/V2	100	0
V1/V2	100	0
V2/V2	100	0

Table 12: The conditional probability distribution of lysis F1, given the blood group factor F1. **P**(*lysis F1/F1*)

	lysis F1							
F1	0	1	2	3	4	5	6	7
present	0.15	1.47	4.41	5.87	11.75	20.56	26.43	29.37
absent	29.37	26.43	20.56	11.75	5.87	4.41	1.47	0.15

Corresponding for the other 3 lysis.

Appendix B

Illustrations of the examples used in this paper

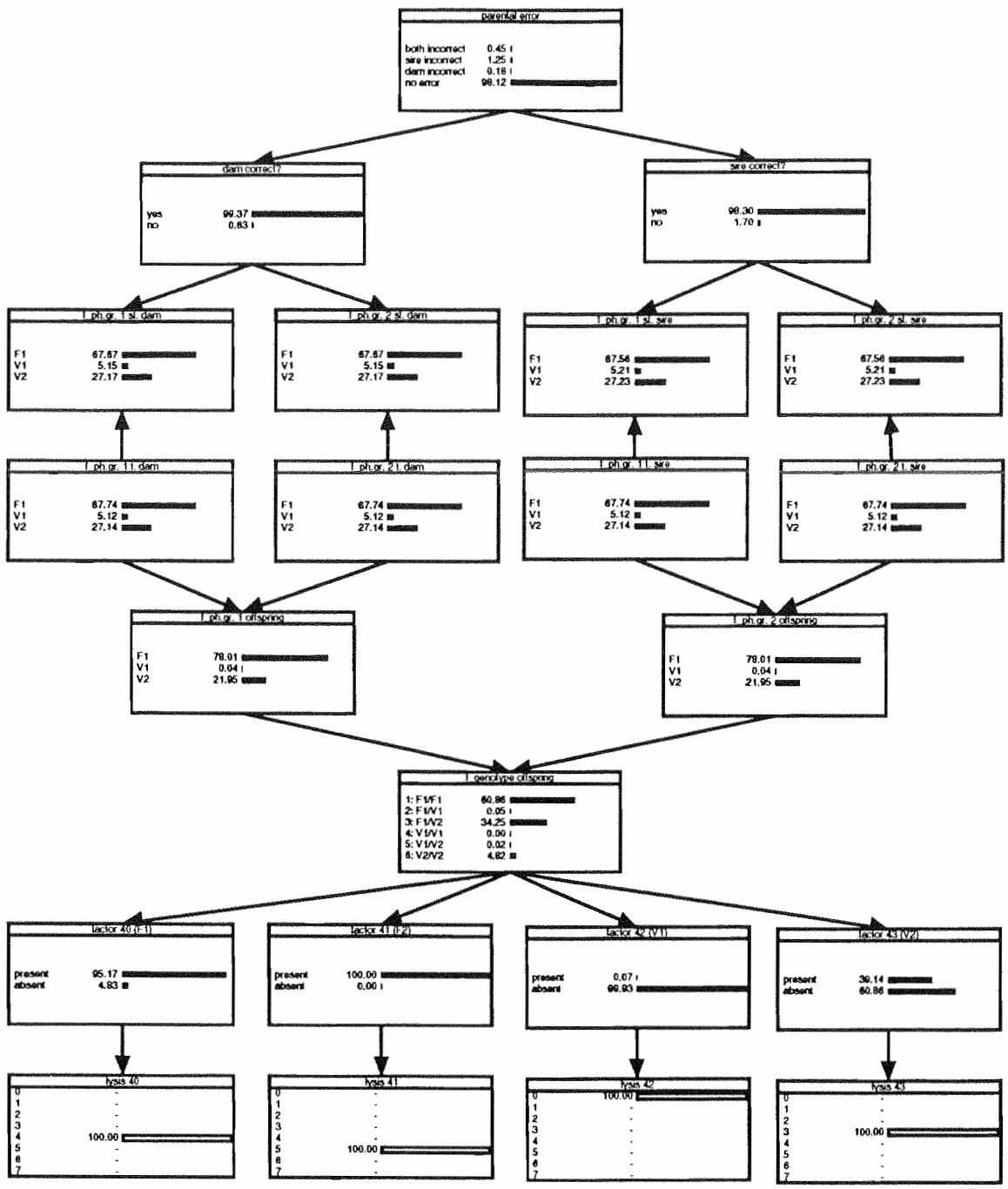


Figure B1: The probabilities of the nodes, where input to the system were the figures obtained by the photometer (lysis factors) only. The genotypes **F1/F1** and **F1/V2** are both having high probability.

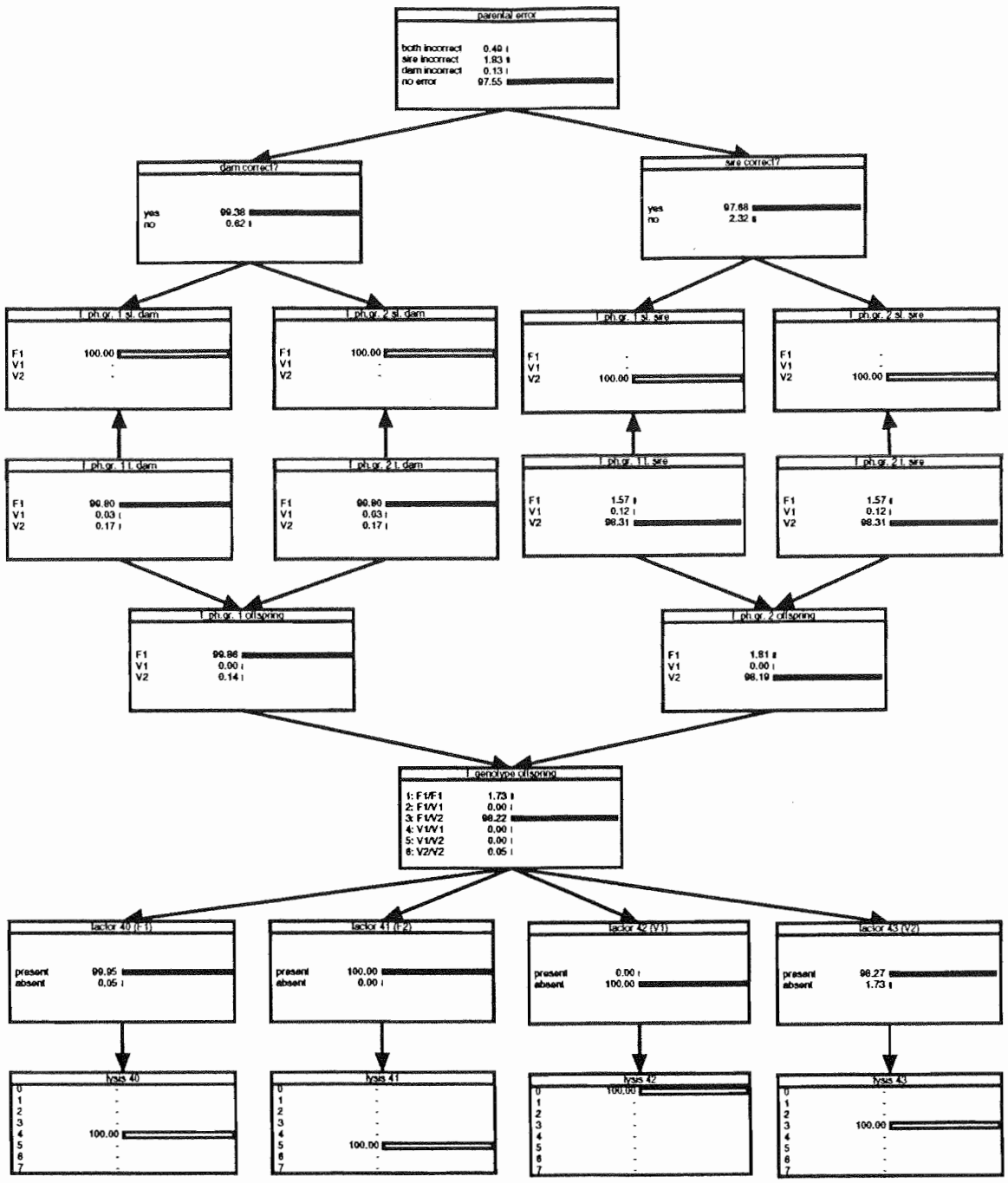


Figure B2: The probabilities of the nodes, where both the measurements of the photometer (lysis factors), and the phenogroups of the stated parents were used as input to the system. The genotype is essentially identified as **F1/V2** and there seeems to be no reason to question the parentage.

Appendix C

Specification file for the F blood group system

(used by Hugin)

```
f_blodtype "f_blodtype"          %% CPN name and label
  90.000000 40.000000            %% Node width and height
  1.000000       %% CPN scale factor

pe "parental error"              %% Node name and label
187.788177 487.490540            %% Node position
( "both incorrect" "sire incorrect" "dam incorrect" "no error" )
%% State vector
                 %% No parents
( 0.45 1.25 0.18 97.94 )


dc "dam correct?"                %% Node name and label
84.409836 426.791321             %% Node position
( "yes" "no" )             %% State vector
 pe                        %% Parents
(( 0 1 )          %  both incorrect
 ( 1 0 )          %  sire incorrect
 ( 0 1 )          %  dam incorrect
 ( 1 0 ))         %  no error


sc "sire correct?"               %% Node name and label
285.475983 427.739746            %% Node position
( "yes" "no" )             %% State vector
 pe                        %% Parents
(( 0 1 )          %  both incorrect
 ( 0 1 )          %  sire incorrect
 ( 1 0 )          %  dam incorrect
 ( 1 0 ))         %  no error


ftd1 "f_ph.gr. 1 t. dam"                 %% Node name and label
31.298031 310.135010             %% Node position
( "F1" "V1" "V2" )               %% State vector
                 %% No parents
( 57.4 10.2 32.3 )


ftd2 "f_ph.gr. 2 t. dam"                 %% Node name and label
137.521652 310.135010            %% Node position
( "F1" "V1" "V2" )               %% State vector
                 %% No parents
( 57.4 10.2 32.3 )


fts1 "f_ph.gr. 1 t. sire"                %% Node name and label
237.106293 311.083435            %% Node position
( "F1" "V1" "V2" )               %% State vector
                 %% No parents
( 57.4 10.2 32.3 )


fts2 "f_ph.gr. 2 t. sire"                %% Node name and label
340.484619 310.135010            %% Node position
( "F1" "V1" "V2" )               %% State vector
                 %% No parents
( 57.4 10.2 32.3 )
```

```
fstd1 "f_ph.gr. 1 st. dam"               %% Node name and label
31.298031 369.885803             %% Node position
( "F1" "V1" "V2" )               %% State vector
 dc ftd1                         %% Parents
((( 1 0 0 )      %  yes  F1
  ( 0 1 0 )      %  yes  V1
  ( 0 0 1 ))     %  yes  V2
 (( 57.4 10.2 32.3 )     %  no  F1
  ( 57.4 10.2 32.3 )     %  no  V1
  ( 57.4 10.2 32.3 )))   %  no  V2


fstd2 "f_ph.gr. 2 st. dam"               %% Node name and label
137.521652 369.885803            %% Node position
( "F1" "V1" "V2" )               %% State vector
 dc ftd2                         %% Parents
((( 1 0 0 )      %  yes  F1
  ( 0 1 0 )      %  yes  V1
  ( 0 0 1 ))     %  yes  V2
 (( 57.4 10.2 32.3 )     %  no  F1
  ( 57.4 10.2 32.3 )     %  no  V1
  ( 57.4 10.2 32.3 )))   %  no  V2


fsts1 "f_ph.gr. 1 st. sire"              %% Node name and label
237.106293 367.988953            %% Node position
( "F1" "V1" "V2" )               %% State vector
 sc fts1                         %% Parents
((( 1 0 0 )      %  yes  F1
  ( 0 1 0 )      %  yes  V1
  ( 0 0 1 ))     %  yes  V2
 (( 57.4 10.2 32.3 )     %  no  F1
  ( 57.4 10.2 32.3 )     %  no  V1
  ( 57.4 10.2 32.3 )))   %  no  V2


fsts2 "f_ph.gr. 2 st. sire"              %% Node name and label
339.536194 367.988953            %% Node position
( "F1" "V1" "V2" )               %% State vector
 sc fts2                         %% Parents
((( 1 0 0 )      %  yes  F1
  ( 0 1 0 )      %  yes  V1
  ( 0 0 1 ))     %  yes  V2
 (( 57.4 10.2 32.3 )     %  no  F1
  ( 57.4 10.2 32.3 )     %  no  V1
  ( 57.4 10.2 32.3 )))   %  no  V2


fph1 "f_ph.gr. 1 offspring"              %% Node name and label
80.616135 249.435822             %% Node position
( "F1" "V1" "V2" )               %% State vector
 ftd1 ftd2                       %% Parents
((( 1 0 0 )      %  F1  F1
  ( 1 1 0 )      %  F1  V1
  ( 1 0 1 ))     %  F1  V2
 (( 1 1 0 )      %  V1  F1
  ( 0 1 0 )      %  V1  V1
  ( 0 1 1 ))     %  V1  V2
 (( 1 0 1 )      %  V2  F1
  ( 0 1 1 )      %  V2  V1
  ( 0 0 1 )))    %  V2  V2
```

```
fph2 "f_ph.gr. 2 offspring"              %% Node name and label
293.063385 248.487396                    %% Node position
( "F1" "V1" "V2" )                       %% State vector
 fts1 fts2                               %% Parents
((( 1 0 0 )      %  F1  F1
  ( 1 1 0 )      %  F1  V1
  ( 1 0 1 ))     %  F1  V2
 (( 1 1 0 )      %  V1  F1
  ( 0 1 0 )      %  V1  V1
  ( 0 1 1 ))     %  V1  V2
 (( 1 0 1 )      %  V2  F1
  ( 0 1 1 )      %  V2  V1
  ( 0 0 1 )))    %  V2  V2


fgeno "f_genotype offspring"             %% Node name and label
187.788177 187.788177                    %% Node position
( "1: F1/F1" "2: F1/V1" "3: F1/V2" "4: V1/V1" "5: V1/V2" "6: V2/V2" )
%% State vector
 fph1 fph2                               %% Parents
((( 1 0 0 0 0 0 )        %  F1  F1
  ( 0 1 0 0 0 0 )        %  F1  V1
  ( 0 0 1 0 0 0 ))       %  F1  V2
 (( 0 1 0 0 0 0 )        %  V1  F1
  ( 0 0 0 1 0 0 )        %  V1  V1
  ( 0 0 0 0 1 0 ))       %  V1  V2
 (( 0 0 1 0 0 0 )        %  V2  F1
  ( 0 0 0 0 1 0 )        %  V2  V1
  ( 0 0 0 0 0 1 )))      %  V2  V2


f40 "factor 40 (F1)"                     %% Node name and label
31.298031 127.088966                     %% Node position
( "present" "absent" )                   %% State vector
 fgeno                                   %% Parents
(( 1 0 )         %  1: F1/F1
 ( 1 0 )         %  2: F1/V1
 ( 1 0 )         %  3: F1/V2
 ( 0 1 )         %  4: V1/V1
 ( 0 1 )         %  5: V1/V2
 ( 0 1 ))        %  6: V2/V2


f41 "factor 41 (F2)"                     %% Node name and label
134.676376 127.088966                    %% Node position
( "present" "absent" )                   %% State vector
 fgeno                                   %% Parents
(( 1 0 )         %  1: F1/F1
 ( 1 0 )         %  2: F1/V1
 ( 1 0 )         %  3: F1/V2
 ( 0 1 )         %  4: V1/V1
 ( 1 0 )         %  5: V1/V2
 ( 1 0 ))        %  6: V2/V2


f42 "factor 42 (V1)"                     %% Node name and label
238.054718 125.192123                    %% Node position
( "present" "absent" )                   %% State vector
 fgeno                                   %% Parents
(( 0 1 )         %  1: F1/F1
 ( 1 0 )         %  2: F1/V1
 ( 0 1 )         %  3: F1/V2
 ( 1 0 )         %  4: V1/V1
 ( 1 0 )         %  5: V1/V2
 ( 0 1 ))        %  6: V2/V2


f43 "factor 43 (V2)"                     %% Node name and label
339.536194 124.243698                    %% Node position
( "present" "absent" )                   %% State vector
 fgeno                                   %% Parents
(( 0 1 )         %  1: F1/F1
 ( 1 0 )         %  2: F1/V1
 ( 1 0 )         %  3: F1/V2
 ( 1 0 )         %  4: V1/V1
 ( 1 0 )         %  5: V1/V2
 ( 1 0 ))        %  6: V2/V2


140 "lysis 40"                           %% Node name and label
31.298031 66.389763                      %% Node position
( "0" "1" "2" "3" "4" "5" "6" "7" )                    %% State vector
 f40                                     %% Parents
(( 0.1 1 3 4 8 14 18 20 )                %  present
 ( 20 18 14 8 4 3 1 0.1 ))               %  absent


141 "lysis 41"                           %% Node name and label
134.676376 66.389763                     %% Node position
( "0" "1" "2" "3" "4" "5" "6" "7" )                    %% State vector
 f41                                     %% Parents
(( 0.1 1 3 4 8 14 18 20 )                %  present
 ( 20 18 14 8 4 3 1 0.1 ))               %  absent


142 "lysis 42"                           %% Node name and label
238.054718 65.441338                     %% Node position
( "0" "1" "2" "3" "4" "5" "6" "7" )                    %% State vector
 f42                                     %% Parents
(( 0.1 1 3 4 8 14 18 20 )                %  present
 ( 20 18 14 8 4 3 1 0.1 ))               %  absent


143 "lysis 43"                           %% Node name and label
339.536194 64.492912                     %% Node position
( "0" "1" "2" "3" "4" "5" "6" "7" )                    %% State vector
 f43                                     %% Parents
(( 0.1 1 3 4 8 14 18 20 )                %  present
 ( 20 18 14 8 4 3 1 0.1 ))               %  absent
```