Pedigree verification with multiple candidate parents using Pedverif software

J. B. C. H. M. van Kaam, R. Finocchiaro, M. Vitale, S. Caracappa

1 Istituto Zooprofilattico Sperimentale della Sicilia “A. Mirri”. Palermo, Italy
2 Dipartimento di Scienze Entomologiche, Fitopatologiche, Microbiologiche Agrarie e Zootecniche. Università di Palermo, Italy

Corresponding author: Jan-Thijs van Kaam. Istituto Zooprofilattico Sperimentale della Sicilia “A. Mirri”. Via Gino Marinuzzi 3, 90129 Palermo, Italy - Tel. +39 091 6565372 - Fax: +39 091 6565314 - Email: jtkaan@unipa.it

ABSTRACT – Species like sheep and beef cattle are commonly raised in large herds and often on pasture with multiple sires joining the females for unrecorded natural insemination. This leads to offspring with multiple candidate parents and therefore uncertain parentage. Twins or triplets can be from multiple sires as well. Pedigrees from such populations are often problematic and need proper verification. The Pedverif computer program is able to verify normal pedigrees as well as pedigrees with multiple candidate parents.

Key words: Pedigree, Verification, Uncertain parentage, Candidate parents.

INTRODUCTION – Commonly species like sheep and beef cattle are raised in large herds and often on pasture with multiple sires joining the females for natural insemination. Consequently inseminations are unrecorded. This leads to offspring with multiple candidate parents and therefore uncertain parentage. Twins or triplets can be from different sires as well due to multiple matings within one ovulatory cycle. In such populations pedigree recording is often problematic and needs proper verification. Therefore a new computer program named Pedverif has been developed. Pedverif is able to verify normal pedigrees as well as pedigrees with multiple candidate parents. Molecular marker information can be used if available. Output of the program can be used as input data for breeding value estimation with multiple candidate parents using, for example, the methodology described in Cardoso and Tempelman (2003), which was adapted to account also for twins and triplets (Van Kaam et al., 2003).

MATERIAL AND METHODS – Pedverif is developed in Fortran 95. Pedverif can use pedigree files with the different formats. Special file format are available for handling multiple candidate parents. In these file formats each record contains for a single individual: Identification of individual, Sex of individual (optional), Number of candidate sires, Number of candidate dams, Identification of every candidate sire and Identification of every candidate dam. Identifications and sex can be alphanumerical.

Pedverif is able to verify pedigrees and report problems. The following actions are undertaken:
• If needed set the sex of specified animals.
• Check if sex codes need to be converted to another format.
• Check if animals do not have identification equal to the code for missing identifications.
• Check for absent candidate sires and candidate dams within a record. Add absent candidate parents to the record.
• Check if animals do not have a candidate parent occurring multiple times. Remove duplicate occurrences.
• Sort candidate sires and candidate dams per record.
• Check if each animal has just one pedigree record. Identical duplicate records are removed. Non-identical duplicate records are reported.
• Check if all parents have a pedigree record. Add parents with a missing record as base parents to the pedigree.
• Check if all sires and dams have known sex in the pedigree file. Add derived sex to the record.
• Verification of a pedigree for loops: animals who are mistakenly their own ancestor.
• Verification for non-Mendelian inheritance of alleles on multiple loci.
• Add a missing parent if an animal has no possible combination of the given candidate parents.
• Produce an output file with all possible combinations of given (and if necessary missing) candidate parents for each animal.

When marker information is available:
• Combinations of candidate parents need to be reported and not single candidate parents, otherwise there will be impossible combinations between possible parents. For example an animal with genotype AB can have possible candidate sires and candidate dams with genotype AA, however it is not possible that both the sire and dam have genotype AA.
• A combination of candidate parents for an individual (i.e., a specific trio) may be possible alone but maybe not in combination with other possible trios.

Pedverif can be used for checking a pedigree file. If problems are found, the user can correct the pedigree or genotype files and rerun Pedverif until all problems are resolved.

RESULTS AND CONCLUSIONS – To demonstrate and test the Pedverif program a dataset of Valle del Belice dairy sheep was analyzed. The dataset contained 99,792 pedigree records. Two animals with birthdates before 1975 were removed. Spaces within the alphabetical part of the IDs were replaced by underscores. Six times the numerical part of the IDs needed to be right-adjusted and leading spaces were filled with zeros. Genotypes on 5 loci of 170 animals were available of which 162 were present in the pedigree file.

The first run of the Pedverif program revealed the following results:

| First run                                      | Number |
|-----------------------------------------------|--------|
| Animals with missing ID                       | 0      |
| Animals with absent candidate parents         | 0      |
| Animals with multiple occurring candidate parents | 0      |
| Animals with multiple identical records       | 0      |
| Animals with multiple non-identical records   | 0      |
| Parents with missing records                  | 1348   |
| Animals with two sexes                        | 45     |

After the first run a file was created in which the sex of the animals with two sexes was set to missing. The second run gave the following results:

| Second run                                      | Number |
|-----------------------------------------------|--------|
| Animals registered as their own ancestor first order | 3      |
| Animals registered as their own ancestor second to fourth order | 0      |

After the second run in the pedigree record of the three animals which were recorded as their own progeny the faulty parent was set to missing.

A third run was carried out using the genotype data. Only 37 offspring could be linked to 23 dams. No offspring could be linked to any sire. This was likely due to the cross sectional nature of the data collection. Of these 37 offspring only 8 had all genotypes matching with their dam. Another 8 offspring had only one genotype mismatch with their supposed dams. This might be a pedigree or genotype recording error. The remaining 21 offspring had multiple genotype mismatches. These animals therefore are probably from another dam than recorded.

In conclusion software has been created which is capable of rapid and automatic verification of pedigrees with and without uncertain parentage. Verification can be done with and without genotype information.
The program distinguishes itself from similar programs because:

- Pedverif will report several problems simultaneously, after which the user can correct the pedigree or genotype file and rerun Pedverif. This results in fewer runs.
- Pedverif can use marker information.
- Pedverif is suitable for pedigrees with multiple candidate parents.

Main differences of pedigree verification with multiple candidate parents:

- More pathways need to be verified to avoid pedigree loops.
- More possible trios of individuals with candidate parents can occur per individual.

The first author was supported by a Marie Curie European Reintegration Grant of the European Community Programme ‘Quality of Life’, contract number MERG-CT-2004-516458.

REFERENCES – Cardoso, F.F., Tempelman, R.J. (2003) Bayesian inference on genetic merit under uncertain paternity. Genet. Sel. Evol. 35 : 469-487. Van Kaam, J.B.C.H.M., Cardoso, F.F., Tempelman, R.J., Portolano B. (2003) Comparison of methods for breeding value estimation with uncertain paternity and twins. Book of Abstracts of the 54th Annual Meeting of the European Association for Animal Production 9 : 78.