**Parental reconstruction in rural goat population with microsatellite markers**

**Maria Siwek,1,2 Egbert Frank Knol3**

1 Dipartimento di Scienze Entomologiche, Fitopatologiche, Microbiologiche Agrarie e Zootecniche, Università di Palermo, Italy
2 Animal Biotechnology Department, University of Technology and Life Sciences, Bydgoszcz, Poland
3 IPG-Institute for Pig Genetics, Beuningen, The Netherlands

**Abstract**

Proper knowledge about individual origin and pedigree is a major point for animal breeding and livestock genetic improvement. Mainly for goat, but also for sheep breeding in Sicily information on these points is limited, and sometimes even misleading. Goat milk production is, in many regions and especially in Sicily, a rural enterprise, where males and females graze together and pedigree of offspring can be largely unknown. Genetic improvement in this situation is challenging. Modern DNA techniques can help to identify the parentage (PI, parental identification). This principle has been shown by Isberg et al. (2004), Lee and Cho (2006) and Rohrer et al. (2007). Especially the work of Isberg et al. (2004) on saltwater crocodiles is a good example on how complicated the parental identification and/or verification can be in non-commercial conditions. If the problem of erratically assigned parents can be solved for animals kept under natural conditions, e.g. saltwater crocodile, it can also be implemented in organic and/or rural goat farming. The central animal identification system for goats in Sicily provides incomplete and possibly imprecise pedigree. The most common mating approach is the use of multiple sires. The combination of these males makes the paternal origin of the new born individual a challenging question. On the maternal side there might be a natural cross fostering. We challenged ourselves with the task of verifying and parental identification based on the molecular marker information. Can the paper pedigree be verified based on the molecular markers? Is it possible to identify sire and dam from a list of candidate fathers and mothers? Which conditions should be fulfilled to make this work? The most commonly used markers at the moment are single nucleotide polymorphism (SNP) and microsatellites. Rohrer et al. (2007) performed a study on SNP for pig identification and parental exclusion, and concluded that a sufficient number of markers for this type of analysis is 60 SNP. The number of SNP available for goat is still limited; recently Pariset et al. (2006), reported 27 SNP characterized on 8 Italian goat breeds. Given the demand for heterozygosity, microsatellite markers with a high number of alleles are more suitable and easier to apply for pedigree reconstruction in goat breeds than biallelic SNP markers. The goat linkage map is still in its infancy, as Vaiman concluded in 1996, compared to other livestock species, e.g. dairy cattle or pig (Vaiman et al., 1996). An update of the goat linkage map has been performed by Maddox and Cockett (2007). Still the only available goat linkage map is a male map, which spans 2737 cM and is comprised of 307 loci. A good number of microsatellite markers are known and available (http://dga.jouy.inra.fr/cgi-bin/Agbo/main.pl?BASE=goat); some of them are goat specific (11 markers), others successfully applied from the cattle (492 markers) or from the sheep genome (109 markers) (Vaiman et al., 1996).

In the GoMilkSicily project, we set up: i) a trajectory to develop a multiplex set of microsatellites; ii) to show validity of the principle in Girgentana, an endangered goat breed in Sicily. Validation of the pedigree results would be challenged by the large amount of uncertainties; dams were not fully certain, since natural cross fostering might have occurred or a dam or an offspring could be misidentified; identity of animals on the farm was sometimes unknown, or animals were cross identified with an already existing tattoo; tattoo logic was and is not completely unambiguous. In this paper we aim to: i) describe our set of microsatellites in relation to the ECONOGENE set; ii) estimate accuracy of microsatellite typing for a goat breed; iii) test the principle of verification and identification on one goat breeding farm; iv) discuss the direct use of the technique in a running goat breeding program.
Materials and methods

Sampling and laboratory techniques

Blood samples were taken from the Girgentana breed located on Pietranera farm. Blood samples were collected twice (in winter 2004 and in summer 2006), each time all the animals present at the farm were sampled (190 individuals; 19 males). Genomic DNA was extracted using GFX Genomic Blood DNA Purification Kit (Nycod-Amersham PLC, Little Chalfont, UK), according to the provided protocol. Twenty-three microsatellite markers were chosen either from the goat linkage map (Vaiman et al., 1996), selected from the ISAG panels recommended for parental verification, or provided by ECONOGENE project (http://www.econogene.org) and assigned in four multiplex sets. The idea behind these sets was to obtain molecular information from markers equally distributed over the whole goat genome. Only unlinked markers with high heterozygosity were taken into consideration. One of the primers from each marker primer pair was labeled with one of the fluorescent labels: VIC, FAM, PET, NED (Applied Biosystems, Carlsbad, CA, USA). Marker scores were obtained by analysis with Gene Mapper software (Applied Biosystems, Carlsbad, CA, USA). Microsatellite markers information and basic characteristics are presented in Table 1.

Statistical analysis

Web available software Cervus 3.0 (Kalinowski et al., 2007) was used. This has the following features.

Allele frequency analysis

Allele frequencies and Hardy-Weinberg equilibrium analysis was performed on a set of 21 markers. Two markers, out of a starting set of 23 markers, with high genotyping error were excluded after a priori from this and further analysis (INRA003, OarFCB11). One marker (INRA0063) was excluded since it deviated significantly from Hardy-Weinberg equilibrium (HW). Our working set consisted of 20 microsatellite markers.

Power estimation using simulation

For parental analysis two levels of confidence were used: strict (95%) and relaxed (80%). Simulation was run to estimate the resolving power of a series of loci given their allele frequencies, and to estimate critical values of the log-likelihood statistics LOD or Delta, so that the confidence of parentage assignments made using the parental analysis module could be evaluated statistically. In the simulation analysis we considered: 1,000 offspring, 120 mothers and 15 fathers, 95% of sampled typed, 92% of loci typed, and 1% of genotyping error. Simulation was performed for paternal side only, maternal side only, and for both parents with known sexes, each time for two levels of confidence.

Parental verification

The assumption was made that one of the parents was known. At first we assigned the dam as known, in a second run the sire was assigned as known. Error rate of paper pedigree was thus determined. Erratic dams and sires were put to missing.

Parental identification

Parental analysis was carried out in three steps: first assuming dams known and sires unknown; second assuming sires known and dams unknown; and thirdly assuming both parents unknown. During this analysis for each offspring tested, parentage is either assigned to the most-likely candidate parent (or parent pair) with a pre-determined level of confidence or is left unassigned.

Results and discussion

Allele frequencies, probability exclusion analysis

Analysis performed by Cervus resulted in a basic marker descriptive (Table 2). The highest number of alleles (10) and PIC (0.74) was observed for locus HSC; the lowest number of alleles (4) and PIC (0.44) for locus INRA0063, which also deviated from HW, therefore we excluded it from further analysis. The average number of alleles in the set of 20 markers used in the analysis was 6.50, and the average PIC 0.64.

For an accuracy control of our results and to test for genotyping error for each locus, we repeated the analysis (PCR reaction and ABI product visualization) of 20 random samples. The average estimated genotyping error was 1%.

Power estimates

For the paternal analysis for both confidence levels, for father alone and for father given...
The aim of parental identification was to ‘fill the holes’ in the paper pedigree, and to identify the parents assigned as erratic in the verification step. When parental identification was performed for both parents at the same time, assuming all parents missing, 49 parent pairs were successfully assigned. For paternal identification alone, 44 fathers could be assigned, while for maternal identification alone, 68 mothers were assigned.

**Choice of markers**

The GoMilk Sicily project aimed to apply new technology in a Sicilian environment from the beginning. At the start of the project we selected markers from all available sources: ISAG panels recommended for parental verification, goat linkage map, ECONOGENE project, and multiplexed them to make the lab work fast and cost efficient. During the project, the goat seventeen-plex/sheep nineteen-plex (Glowatzki-Mullis et al., 2007) became available. It was not, and is still not, our aim to propagate our multiplexes. State of the art nowadays seems to be SNPs. Although, for verification and identification, a good set of microsatellite markers also seems to work quite well. The number of available SNP markers for goat, as stated already, is still a limiting factor (Pariset et al., 2006).

**Reasons for making the choice of markers:**
- part of ECONOGENE (markers selected and advised by the United Nations Food and Agriculture Organization)

Known mother, the assignment rate, given current marker set, was 95%. For the maternal analysis, for mother alone, the assignment rate was 96% for strict confidence level and 97% for relaxed confidence level. For maternal analysis, when father was assumed to be known, the assignment level was 95% for both strict and relaxed confidence levels. For the analysis of both parents with known genders the assignment rate was 95% for strict and 94% for relaxed confidence interval.

### Verification

Our data set consisted of 184 records present in the paper pedigree data set. In this data set 31% mothers and 39% fathers were missing. Parental verification was performed for paternal or maternal side alone, and for both parents together, assuming 1% of loci mistyped, we allowed for one mistyping. When we performed one-sided (paternal or maternal) verification, 26% mothers and 40% fathers were considered erratic, since they had two or more mismatches between parents and offspring. For verification of both parents at the same time, 17% of parent pairs were considered wrong when compared with paper pedigree information.

### Identification of missing parents

The aim of parental identification was to ‘fill the holes’ in the paper pedigree, and to identify the parents assigned as erratic in the verification step. When parental identification was performed for both parents at the same time, assuming all parents missing, 49 parent pairs were successfully assigned. For paternal identification alone, 44 fathers could be assigned, while for maternal identification alone, 68 mothers were assigned.

### Table 2. Microsatellite markers characteristics.

| Locus          | K | N   | H obs | H exp | PIC | NE-1P | NE-2P | NE-PP | NE-1 | NE-SI | HW  |
|----------------|---|-----|-------|-------|-----|-------|-------|-------|------|-------|-----|
| BM0321         | 6 | 189 | 0.651 | 0.681 | 0.624 | 0.738 | 0.574 | 0.395 | 0.159 | 0.450 | ns  |
| BM1329         | 7 | 171 | 0.743 | 0.724 | 0.688 | 0.670 | 0.489 | 0.292 | 0.112 | 0.417 | ns  |
| HSC            | 10| 154 | 0.657 | 0.773 | 0.744 | 0.599 | 0.419 | 0.224 | 0.080 | 0.385 | ns  |
| MAF064         | 6 | 183 | 0.667 | 0.756 | 0.713 | 0.651 | 0.474 | 0.292 | 0.101 | 0.398 | ns  |
| SRCRSP0005     | 7 | 174 | 0.695 | 0.697 | 0.665 | 0.695 | 0.510 | 0.094 | 0.123 | 0.433 | ns  |
| SRCRSP008      | 4 | 170 | 0.588 | 0.587 | 0.499 | 0.823 | 0.700 | 0.554 | 0.258 | 0.522 | ns  |
| ILST0087       | 6 | 171 | 0.690 | 0.683 | 0.646 | 0.720 | 0.538 | 0.340 | 0.137 | 0.444 | ns  |
| MAF70          | 6 | 184 | 0.723 | 0.703 | 0.652 | 0.715 | 0.544 | 0.363 | 0.138 | 0.434 | ns  |
| McM073         | 4 | 187 | 0.567 | 0.615 | 0.545 | 0.803 | 0.658 | 0.498 | 0.218 | 0.498 | ns  |
| McM64          | 8 | 187 | 0.759 | 0.733 | 0.695 | 0.665 | 0.485 | 0.292 | 0.108 | 0.412 | ns  |
| SRCRSP024      | 6 | 185 | 0.551 | 0.643 | 0.604 | 0.758 | 0.580 | 0.384 | 0.166 | 0.471 | ns  |
| ILST011        | 6 | 177 | 0.633 | 0.698 | 0.648 | 0.717 | 0.545 | 0.361 | 0.141 | 0.437 | ns  |
| INRABern172    | 7 | 184 | 0.734 | 0.731 | 0.687 | 0.672 | 0.497 | 0.308 | 0.116 | 0.414 | ns  |
| OarA54         | 7 | 185 | 0.514 | 0.564 | 0.535 | 0.817 | 0.640 | 0.444 | 0.219 | 0.523 | ns  |
| OarFBC48       | 7 | 188 | 0.635 | 0.810 | 0.783 | 0.549 | 0.372 | 0.189 | 0.061 | 0.362 | ns  |
| SRCRSP009      | 5 | 182 | 0.742 | 0.758 | 0.716 | 0.647 | 0.469 | 0.209 | 0.099 | 0.397 | ns  |
| SPS113         | 8 | 178 | 0.685 | 0.701 | 0.653 | 0.705 | 0.533 | 0.343 | 0.137 | 0.455 | ns  |
| INRA0063       | 4 | 140 | 0.257 | 0.549 | 0.447 | 0.848 | 0.748 | 0.621 | 0.305 | 0.553 | ns  |
| MAF065         | 9 | 158 | 0.707 | 0.766 | 0.731 | 0.618 | 0.448 | 0.249 | 0.089 | 0.390 | ns  |
| McM0227        | 5 | 172 | 0.578 | 0.579 | 0.360 | 0.524 | 0.703 | 0.632 | 0.405 | 0.662 | ns  |
| SRCRSP0023     | 6 | 187 | 0.701 | 0.668 | 0.631 | 0.733 | 0.552 | 0.354 | 0.147 | 0.453 | ns  |

K, number of alleles; N, number of samples analyzed; H obs, heterozygosity observed; H exp, heterozygosity expected; PIC, polymorphic information content; NE-1P, average non-exclusion probability for one candidate parent; NE-2P, average non-exclusion probability for one candidate parent given the genotype of a known parent of the opposite sex; NE-PP, average non-exclusion probability for a candidate parent pair; NE-1, average non-exclusion probability for identity of 2 unrelated individuals; NE-SI, average non-exclusion probability for identity of 2 siblings; HW, significance of deviation from Hardy-Weinberg equilibrium; ns, not significant; *significant at the 0.1% level.

### Table 3. Comparison of power of analysis and rate of assignment for paternal or maternal alone, and for both parents (Cervus analysis).

| Year of birth | No. | Correct assignment, % |
|---------------|-----|-----------------------|
|               | Dam alone | Sire alone | Both parents |
| 1998 and older| 30     | 6       | -       | -       |
| 2000          | 3      | 33      | -       | -       |
| 2001          | 41     | 66      | -       | -       |
| 2002          | 27     | 48      | 63      | 63      |
| 2003          | 39     | 56      | 46      | 61      |
| 2004          | 27     | 59      | 44      | 55      |
| 2005          | 14     | 78      | 50      | 57      |
| 2006          | 12     | 83      | 66      | 83      |
We compared three existing microsatellite sets, ECONOGENE, Glowatzki-Mullis and our own (see Appendix). On the basis of these arguments we conclude somewhat arbitrarily that the ‘Glowatzki-Mullis’ is the best. The major advantage of this set is that it is cost efficient. It allows amplification of 17 markers in a single PCR reaction, thus speeding up the process of data generation and significantly reducing analysis costs, which are the two most important features from a practical point of view. In this set, 7 loci, out of 30 listed in the ECONOGENE project, are used. It would be quite helpful from a practical point of view if the same markers could be used for goat and sheep.

Our results seem to suggest that this might be feasible, but extra attention should be given to the choice of markers.

Parental verification

Our results showed that 18% of the dam-offspring relation could not be real. The offspring was cross-fostered or erratic identification of markers or paper pedigree occurred. For father-offspring relation the ratio of erratically assigned father (given correct dam) was 23%.

Mating of sheep and goat in Sicily follows this procedure: first one ram joins the flock and mates 80% of ewes. Later the second ram is used and on occasion even the third one. Some of the pedigree can be derived relatively easily on the basis of the date of birth; in this current analysis we, arbitrarily, ignored this fact.

Paternal (partially maternal) identification

Parental identification was performed in three ways: paternal alone, maternal analysis alone, and for both parents at the same time. Results of this analysis are presented in Table 3, where different types of the analysis (paternal or maternal alone, or both parents in the same time) are presented in relation to the number of the individuals available in the date set and their date of birth. In other words, how likely are we able to perform correct assignment of each or both parents given the availability of candidate parents in the dataset.

A limited number of sampled parents born before 2001 made the proper analysis impossible, since there was no comprehensive list of candidate sires and dams.

The key to successful parental analysis is routine sampling of relevant animals in overlapping generations.

Implementation

Assuming a good sized 200 goat farm and assuming 5 effective parities, some 50 replacement females are needed annually. We propose to use milk sampling of first parity females for DNA extraction. This DNA will then be used for: i) maternal verification, to check for potential maternal cross-fostering; ii) for paternal identification.

Similarly, assuming the need for 4 replacement males annually, 10 males could be used for rearing. We recommend taking tissue samples of around 15 males for: i) maternal verification; ii) paternal identification. For long-term genetic selection the males are more relevant than the females. We recommend to genotype sires twice to reduce errors and to have as many markers typed as possible. This should help in increasing the percentage of offspring assigned and therefore should efficiently increase the power of the analysis considerably.

Direct costs for this procedure are: i) once a year tissue collection for the young males (we recommend blood samples). These could be taken during milk sampling (costs pm in that situation); ii) microsatellite analysis for two multiplexes, amounting to €10 per sample, or €650 for 50 females and 15 males, for the total farm per year. Bear in mind that these are mainly direct costs, consumables and collection and lab labor.

The advantages are:

i) strongly improved pedigree (see above), therefore better EBVs;
ii) better gene pool conservation through thorough choice of new males and females;
iii) opportunity to extend the procedure to candidate gene and QTL selection.

Conclusions

In this paper we have presented a case study of microsatellite markers usage for parental identification and parental verification in a rural goat population. Field data might include missing sire and/or dams; or sires/dams which do not produce any progeny. The key for successfully solving problems with offspring origin and paternity testing is routine sampling of key animals in overlapping generations. In general a set of 20 polymorphic microsatellite markers might be used in routine verification of paper parental information and parental identification.

References

Glowatzki-Mullis, M.L., Muntwyler, J., Gaillard, C., 2007. Cost-effective parentage verification with 17-plex PCR for goats and 19-plex PCR for sheep. Anim. Genet. 38:86-88.

Isberg, S.R., Chen, Y., Barker, S.G., Moran, C., 2004. Analysis of microsatellites and parentage testing in saltwater crocodiles. J. Hered. 95:445-449.

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16:1099-2006.

Lee, S.Y., Cho, G.J., 2006. Parentage testing of Thoroughbred horse in Korea using microsatellite DNA typing. J. Vet. Sci. 7:63-67.

Maddock, J.F., Cockell.N.E., 2007. An update on sheep and goat linkage maps and other genomics resources. Small Ruminant Res. 70:4-20.

Pariset, L., Cappuccio, I., Ajmone Marsan, P., Dunner, S., Luikart, G., England, P.R., Obexer-Ruff, G., Peter, C., Marletta, D., Pilla, F., Valenti, A., 2006. Assessment of population structure by single nucleotide polymorphisms (SNPs) in goat breeds. J. Chromatogr. B. 833:117-120.

Rohrer, G.A., Freking, B.A., Nonneman, D., 2007. Single nucleotide polymorphisms for pig identification and parentage exclusion. Anim. Genet. 38:253-258.

Vaiman, D., Schiller, L., Bourgeois, F., Oustry, A., Amigues, Y., Criqui, E.P., 1996. A genetic linkage map of the male goat genome. Genetics 144:279-305.
### APPENDIX

**Appendix. Comparison of three microsatellite sets: ECONOGENE, Glowatzki-Mullis, GoMilkSicily, loci used in set, number of alleles, mean number of alleles calculated. ECONOGENE markers used as a reference set.**

| Locus       | Number of alleles |
|-------------|-------------------|
|             | ECONOGENE (54 breeds) | Glowatzki-Mullis (11 breeds) | GoMilkSicily (4 breeds)* |
| BM6444      | 43                 | not used                       | not used                       |
| CSR2247     | 15                 | 8                              | not used                       |
| DRBP1       | 22                 | not used                       | not used                       |
| ETH10       | 5                  | 4                              | not used                       |
| ILSTS0805   | 9                  | 5                              | not used                       |
| ILSTS0911   | 12                 | not used                       | 8                             |
| ILSTS0929   | 13                 | not used                       | not used                       |
| ILSTS0937   | 11                 | not used                       | 8                             |
| INRA023     | 15                 | not used                       | not used                       |
| INRA063     | 9                  | 6                              | 5                             |
| INRABERN172 | 13                 | not used                       | 9                             |
| INRABERN185 | 14                 | not used                       | not used                       |
| MAF209      | 5                  | not used                       | not used                       |
| MAF65       | 24                 | not used                       | 16                            |
| MAF70       | 17                 | not used                       | 9                             |
| McM527      | 12                 | not used                       | 7                             |
| OarAES4     | 17                 | not used                       | 10                            |
| OarFCB20    | 11                 | 10                             | not used                       |
| OarFCB48    | 14                 | not used                       | 8                             |
| P19         | 17                 | not used                       | not used                       |
| PPS113      | 17                 | 13                             | 11                            |
| SRCRS9      | 17                 | not used                       | 9                             |
| SRCRS15     | 11                 | not used                       | not used                       |
| SRCRS23     | 21                 | not used                       | 13                            |
| SRCRS3      | 11                 | not used                       | not used                       |
| SRCRS5      | 13                 | not used                       | 11                            |
| SRCRS7      | 7                  | not used                       | not used                       |
| SRCRS8      | 17                 | 15                             | 8                             |
| TCRVB6      | 19                 | not used                       | not used                       |
| TGLA53      | 18                 | not used                       | not used                       |
| BM1258      | not used           | 11                             | not used                       |
| BM1329      | not used           | 9                              | 10                            |
| BOBT24A     | not used           | 11                             | not used                       |
| HSC         | not used           | 24                             | not used                       |
| INRA132     | not used           | 4                              | not used                       |
| INRA400     | not used           | 13                             | not used                       |
| INRA005     | not used           | 5                              | not used                       |
| OarFCB128   | not used           | 4                              | not used                       |
| SPS115      | not used           | 6                              | not used                       |
| SRCRS1201   | not used           | 17                             | not used                       |
| BM0321      | not used           | 11                             | not used                       |
| INRA003     | not used           | not used                       | 3                             |
| MAF64       | not used           | not used                       | 7                             |
| McM3        | not used           | not used                       | 5                             |
| McM6        | not used           | not used                       | 13                            |
| SRCRS24     | not used           | not used                       | 9                             |
| OarFCB111   | not used           | not used                       | 7                             |
| Mean        | 10.96              | 9.7                            | 9.3                           |

*Analysis was carried out with Genetix on four goat breeds used in the GoMilkSicily project (Girgentana, Maltese, Messinese, Derrara di Siria).*