Inference of genotype probabilities and derived statistics for PrP locus in sheep[1]

Gregor Gorjanc [a,*] Drago Kompan [a]

[a] *University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Groblje 3, 1230 Domžale, Slovenia*

Abstract

The aim of our work was to infer PrP genotype probabilities in sheep to provide additional genotype identifications. The pedigree data consisted of 10,429 animals of Jezersko-Solcava sheep breed with 3,669 animals having PrP genotype data. There were 2,673 live non-genotyped animals. Five PrP alleles were present with the following frequencies: ARR 0.174, AHQ 0.074, ARH 0.083, ARQ 0.632, and VRQ 0.037. Iterative allelic peeling with incomplete penetrance model as implemented in the GenoProb program was used for inferring the genotype probabilities. There were only some additional identification of PrP genotype and NSP (national scrapie plan) type with high probability. We maintain that the main reasons for a low number of additional identifications can be attributed to the large number of alleles with moderate frequencies, incomplete penetrance model, uniform prior, and inherent pedigree and genotype data structure. In order to overcome the limits of additional genotype identifications we derived novel statistics (maximal NSP type, average NSP value and its variance and accuracy) to facilitate practical implementation of selection for scrapie resistance based on NSP types. Maximal NSP type can be used to infer maximal potential scrapie susceptibility of individual animals as well as for the entire flocks. The average NSP value encompasses all information contained in PrP genotype probabilities and is the most useful statistic for the selection on NSP type and therefore PrP genotype. These novel statistics can be used as a criterion for the selection against scrapie susceptibility for the whole population taking into the account the possible errors in the genotype and/or pedigree data.

Key words: genotype probabilities, PrP locus, NSP type, scrapie, sheep

1. Introduction

Scrapie is a type of transmissible spongiform encephalopathy disease in sheep. Recent review on scrapie in sheep has been given by Ulvund (2008). The susceptibility to scrapie is strongly associated with polymorphisms on the 136^{th}, 154^{th}, and 171^{st} codon of the PrP gene (Hunter, 1997). The following five PrP alleles (haplotypes) are the most frequent in many populations (e.g. Lühken et. al, 2008): $A_{136}R_{154}R_{171}$ (ARR), $A_{136}H_{154}Q_{171}$ (AHQ), $A_{136}R_{154}H_{171}$ (ARH), $A_{136}R_{154}Q_{171}$ (ARQ), and $V_{136}R_{154}Q_{171}$ (VRQ).

Due to the large number of possible genotypes and to some extent similar effect, PrP genotypes are usually classified into risk groups based on scrapie susceptibility (e.g. Dawson et al. , 2008). Risk groups are sometimes denoted as NSP types or groups after the National Scrapie Plan in the UK. Commonly, the most resistant genotype ARR/ARR is of NSP type 1, while the most susceptible genotypes are of NSP type 5. This grouping often represents the organizational implementation of selection programs for scrapie resistance in sheep in many countries. In these programs the aim is to remove the VRQ allele and to increase the frequency of ARR allele, especially in the ARR/ARR form.

Since the infecting agent causing scrapie is still not known, the information about the PrP genotype is the most important factor that can be used for the regulation of scrapie susceptibility/resistance in sheep. The number of sheep involved in genotyping is massive - Dawson et al. (2008) reported that 700,000 rams have been genotyped in Great Britain since the start of National Scrapie Plan in year 2001. The costs of genotyping have decreased in recent years. However, the collection of tissue samples and genotyping of a large number of individuals are still of

* Corresponding author.
Email address: gregor.gorjanc@bfro.uni-lj.si (Gregor Gorjanc).

[1] *This is a shortened version of the paper submitted to Livestock Science*

considerable cost in sheep breeding programmes. Therefore, methods for the calculation of genotype probabilities for non-genotyped animals could be exploited to reduce the costs and to increase selection intensity. The aim of this paper is to present the results of the inference of PrP genotype probabilities and derived statistics that can be used for selection on scrapie resistance.

2. Material

For the purpose of this study PrP genotype and pedigree data of the Slovenian autochthonous Jezersko-Solcava sheep were used. This is a meat-type breed of considerable value for Slovenian sheep production. All animals from the Jezersko-Solcava herdbook were taken into consideration, except those that were not informative for the calculation of PrP genotype probabilities and were therefore pruned from the pedigree. Pruning was applied in the direction from ascendants to descendants. The following criteria had to be met for the removal of an animal: known date of death or culling, only one descendant, and no PrP genotype data. Live animals were retained since the aim was to infer genotype probabilities for live animals, the potential selection candidates.

Altogether 10,429 animals were used in the analysis (Table 1). Among all animals in the study, there were 397 rams and 3,530 ewes (Table 1). The percentage of unknown sires or dams was 23.0 % for all animals, 19.1 % for sires and 33.5 % for dams. PrP genotype data was available for 3,669 animals of which 114 were sires and 1,443 were dams. PrP genotype data was first available in year 2005 for performance tested rams at the national central test station. Later, PrP genotype data were available also for rams and ewes of all ages in the flocks all around the country. The number of rams and ewes with PrP genotype data increased by the year of birth, but there was a large variability between herds (data not shown). There were 2,673 live animals, 28 live sires and 508 live dams (Table 1) that were not (yet) genotyped. Allele and genotype frequencies for the used dataset are presented in the results.

Table 1
Data structure

	Animals	Sires	Dams
All	10,429	397	3,530
Unknown sire or dam	2,403	76	1,184
With PrP genotype data	3,669	114	1,443
Live and without PrP genotype data	2,673	28	508

3. Methods

PrP genotype probabilities were calculated with iterative allelic peeling (Thallman et. al, 2001a,b) using incomplete penetrance model as implemented in the GenoProb program (Thallman, 2002). Input data were PrP genotypes and pedigree. In GenoProb program prior probability for alleles in non-genotyped founders is assumed to be uniform i.e.

$$\Pr(ARR) = \Pr(AHQ) = \ldots = \Pr(VRQ) = 1/5. \quad (1)$$

Incomplete penetrance model allowed the use of animals with potential errors in either genotype or pedigree data. Error rate was conservatively set to 0.1. For each animal a vector $\mathbf{g}$ (2) with 15 genotype probabilities was obtained. Only probabilities bigger than 0.0001 were retained. From the obtained genotype probabilities a vector $\mathbf{n}$ (3) with 5 NSP type probabilities was calculated according to PrP genotype grouping in NSP types (e.g. Dawson et al. , 2008).

$$\mathbf{g}' = \left(\Pr\left(ARR/ARR\right), \ldots, \Pr\left(VRQ/VRQ\right)\right), \quad (2)$$

$$\mathbf{n}' = \left(\Pr\left(NSP_1\right), \ldots, \Pr\left(NSP_5\right)\right). \quad (3)$$

In order to assess potential scrapie susceptibility, maximal NSP type (4) for each animal was defined as:

$$\max\left(NSP, l\right) = \max\left(I\left(\mathbf{n} > l\right) \bullet \mathbf{a}\right), \quad (4)$$

where $I()$ is an indicator function returning value 0 when the condition is false, or 1 when the condition is true, l is an arbitrary NSP type probability threshold value such as 0.05, $\mathbf{a}' = (1, 2, 3, 4, 5)$ is a vector of NSP type values, and $\bullet$ is an elementwise multiplication operator.

Additionally, an average NSP value (5) was calculated for each animal as a weighted average of arbitrary values $\mathbf{a}$ weighted with NSP type probabilities ($\mathbf{n}$):

$$\overline{NSP} = \mathbf{a}' * \mathbf{n}. \quad (5)$$

The variance of individual average NSP value (6) was derived as:

$$\sigma^2_{\overline{NSP}} = \left(\mathbf{a}' - \overline{NSP}\right)^2 * \mathbf{n}. \quad (6)$$

If all 15 PrP genotypes are equally likely such us under uniform prior, all values of $\mathbf{g}$ are equal to 1/15, $\mathbf{n}' = (1/15, 3/15, 6/15, 1/15, 4/15)$, $\overline{NSP} = 3.27$, and $\sigma^2_{\overline{NSP}} = 1.53$. These results were used to derive the accuracy of average NSP value:

$$r_{\overline{NSP}} = 1 - \sigma^2_{\overline{NSP}}/k, \quad (7)$$

where $k = \sigma^2_{\overline{NSP}}$ under the uniform distribution of PrP genotypes. In case of uniform probabilities for NSP types, $\overline{NSP} = 3$, and $\sigma^2_{\overline{NSP}} = 2$. If the prior is different as in our setting, then it should be used in the above derivation.

4. Results with Discussion

4.1. *Frequencies*

Frequency of the favourable ARR allele was 0.174, while the frequency of VRQ allele was 0.037 (Table 2). The most frequent allele was ARQ with frequency 0.632. All 15 possible PrP genotypes were found in this breed, with the highest frequencies for ARQ/ARQ (0.4) and ARR/ARQ (0.212). Other genotypes had frequencies lower than 0.12. Heterozygosity was equal to 0.56, while polymorphism information content was equal to 0.52. In comparison with PrP allele and genotype frequency estimates of 56 sheep breeds compiled by Lühken et. al (2008), Jezersko-Solcava breed had low frequency of ARR allele, high frequency of ARQ, AHQ, and ARH alleles, and high frequencies of heterozygotes. Frequencies of NSP type 1, 2, 3, 4, and 5 for Jezersko-Solcava breed were 0.035, 0.267, 0.628, 0.011, and 0.061, respectively. These frequencies correspond to an average NSP value (5) of 2.8.

Table 2
PrP allele, PrP genotype and NSP type frequencies (n=3,669)

PrP allele				
ARR	AHQ	ARH	ARQ	VRQ
0.174	0.074	0.083	0.632	0.037
PrP genotype				NSP type
ARR/ARR				1
0.035				0.035
ARR/AHQ	ARR/ARH	ARR/ARQ		2
0.026	0.029	0.212		0.267
AHQ/AHQ	AHQ/ARH	AHQ/ARQ		
0.008	0.010	0.092		3
ARH/ARH	ARH/ARQ	ARQ/ARQ		0.628
0.007	0.111	0.400		
ARR/VRQ				4
0.011				0.011
AHQ/VRQ	ARH/VRQ	ARQ/VRQ		
0.004	0.005	0.050		5
VRQ/VRQ				0.061
0.002				

4.2. *Additional identifications*

Ideally, the calculation of genotype probabilities is performed in order to acquire additional identification or exclusion of genotype(s) for some animals and partial information on genotype(s) for other animals. Unfortunately, there were no additional identifications of PrP genotype with certainty via the calculated genotype probabilities (Table 3), not even for non-genotyped sires or dams. When probability of identification was lowered to 0.99, 66 animals had additionally identified PrP genotype. However, this was only 2.5 % of all live non-genotyped animals. There were more additional identifications (as measured by the percentage of animals in a group) for sires, but not for dams. This is clearly due to the higher number of progeny per sire than per dam. Lowering the probability of genotype identification down to 0.80 increased the number of additional identifications, but not substantially.

There were also none additional identifications of NSP type with certainty (Table 3). However, 209 (7.8 %) animals had additionally identified NSP type with the probability of identification equal to 0.99. This number more than doubled when the probability of identification was lowered to 0.8. The higher number of additional identifications for NSP type in comparison to PrP genotype is due to the fact that 15 genotypes are grouped into 5 NSP types. There were also more additional NSP type identifications than genotype identifications for non-genotyped sires and dams, although the increase was proportionally higher in dams. Additional exclusion of PrP genotype and NSP type was also assessed but is not presented. Those results were not very informative, since there are 15 possible PrP genotypes and 5 NSP types.

Table 3
Additional identifications of PrP genotype and NSP type for live animals (n=2,673), sires (n=28), and dams (n=508)

PrP genotype						
	Animals		Sires		Dams	
Probability	No.	%	No.	%	No.	%
1.00	0	0.0	0	0.0	0	0.0
0.99	66	2.5	6	21.4	9	1.8
0.95	101	3.8	8	28.6	15	3.0
0.90	116	4.3	9	32.1	19	3.7
0.85	134	5.0	9	32.1	22	4.3
0.80	144	5.4	10	35.7	27	5.3
NSP type						
	Animals		Sires		Dams	
Probability	No.	%	No.	%	No.	%
1.00	0	0.0	0	0.0	0	0.0
0.99	209	7.8	7	25.0	28	5.5
0.95	283	10.6	10	35.7	47	9.2
0.90	329	12.3	12	42.8	54	10.6
0.85	382	14.3	12	42.8	65	12.8
0.80	440	16.5	15	53.6	81	15.9

Tier and Henshall (2005) have evaluated the limits of additional genotype identification for single loci in com-

mon livestock pedigrees. They concluded that the increased number of alleles, intermediate frequencies of alleles, and smaller families decreased success as measured with additionally inferred genotypes with certainty. Other reasons for low number of additional genotype identifications in our example are the use of a uniform prior and the incomplete penetrance model. The assumption about a uniform prior distribution for PrP alleles in founders is surely debatable given the non-uniform distribution of published allele frequencies (e.g. Lühken et. al, 2008). However, this is the most non-informative prior information and Kerr and Kinghorn (1996) have shown that such a prior reduces the number of erroneous genotype exclusions. This is very important for the PrP genotype case, where erroneous statements about $VRQ/*$ genotypes are not wanted. Therefore, a uniform prior for allele frequencies in founders is warranted as a conservative choice. Incomplete penetrance model introduces additional uncertainty via the penetrance (error) function, but enables the usage of all the data that might contain genotype and/or pedigree errors. With the complete penetrance model, the erroneous data must be either corrected or excluded. This is very important issue for the applied work. It is far too often assumed that genotype data is accurate. Sobel et al. (2002) have stated that results from incomplete penetrance model are not so informative, but more secure.

The PrP locus in sheep is an example with considerable limits for additional genotype identifications via inferred genotype probabilities. That is the reason why we derived other statistics that could be used as selection criteria for scrapie resistance in non-genotyped animals. Since the selection on PrP genotype is mainly driven by grouping of genotypes in NSP types, we focused on the derivation of statistics for NSP type, though the same approach could also be used for the PrP genotypes.

4.3. *Maximal NSP type*

In comparison to rare additional identification of PrP genotype and NSP type based on genotype probabilities, maximal NSP type (4) can be always inferred. It should be stressed that the NSP type of a genotyped animal can differ from inferred maximal NSP type due to information coming from relatives. Maximal NSP type can be used to infer maximal potential scrapie susceptibility of individual animals as well as for the entire flocks. However, with the incomplete penetrance model there is a practical problem of setting the NSP type probability threshold (i.e. value of l in (4) in order to skip NSP types with negligible probability. We have used 0.05, but other values could also be used.

Distribution of NSP type probabilities of genotyped animals and maximal NSP type of non-genotyped animals for Jezersko-Solcava dataset is shown in Fig. 1. The majority of non-genotyped animals had a maximal NSP type of 5, but there was a considerable amount of animals with lower (better) NSP type.

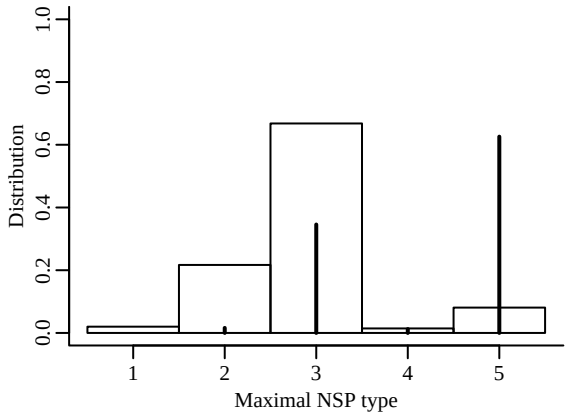

Fig. 1. Distribution of live animals by maximal NSP type: bars - animals with PrP genotype data, vertical lines - animals without PrP genotype data

4.4. *Average NSP value*

The maximal NSP type does not encompass all information contained in inferred PrP genotype probabilities. Therefore, the average NSP value (5) was derived as a weighted average of NSP type values, where weights are NSP type probabilities. The average NSP value of a genotyped animal is the same as NSP type value, but only if the information from the relatives confirms this. Otherwise there might be some change. This statistic uses all information contained in inferred PrP genotype probabilities and is therefore the most useful for the selection on PrP genotype. The distribution of average NSP values for Jezersko-Solcava breed is shown in Fig. 2. The full use of genotype information is manifested as a good accordance in distributions of average NSP values for the genotyped and non-genotyped animals. Although this statistic is continuous, the peaks were observed at values around 1.5, 2, 2.5, 3, 3.5, 4, and 5.

The average NSP value is in essence similar to a breeding value, but with emphasis on practical use of PrP genotype in the selection process. NSP type values ($\mathbf{a}$) could reflect the scrapie susceptibility in a more precise way, but in that case the calculation of breeding value would be preferred. Breeding value would be calculated as a weighted average of PrP breeding values i.e. a sum of PrP allele average effects, weighted with PrP genotype probabilities. Unfortunately, average effects of PrP alleles are not known. The NSP type values could also be defined in such a way that they would more closely reflect the scrapie susceptibility of NSP types. As it is usually the case with breeding values, the average NSP value can be represented as a deviation from the population NSP average.

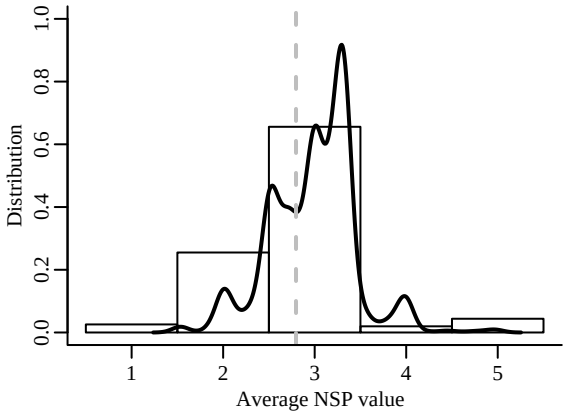

Fig. 2. Distribution of live animals by average NSP value $\left(\overline{NSP}\right)$: bars - animals with PrP genotype data, superimposed curve - animals without PrP genotype data; vertical dashed line - population $\overline{NSP}$

The average NSP value can be used as a selection criterion for the non-genotyped animals, but this statistic should not be used blindly. When there is no genotype information from relatives and a uniform prior is used, all inferred PrP genotypes are equally likely and $\overline{NSP}$ is equal to 3.27. In this case the average NSP value is indistinguishable from the same value for an individual with genotype information from relatives. Variance of the average NSP value (6) or accuracy (7) can be used to distinguish such cases. Therefore, the average NSP value and its accuracy could be used in the selection for scrapie resistance in order to include all animals in the selection progress and to account for the possible errors in the genotype and/or pedigree data.

5. Conclusion

PrP genotype probabilities for non-genotyped animals of Jezersko-Solcava breed were calculated with iterative allelic peeling using incomplete penetrance model. There were only some additional PrP genotype and NSP type identifications with high probability. This can be attributed to the large number of alleles with moderate frequencies, incomplete penetrance model, uniform prior, and inherent pedigree and genotype data structure. Nevertheless, PrP genotype probabilities can be used to calculate maximal NSP type and more importantly average NSP value and its accuracy. These novel statistics can be used as a criterion for the selection against scrapie susceptibility for the whole population taking into the account the possible errors in the genotype and/or pedigree data.

References

Dawson, M., Moore, R.C., Bishop, S.C. 2008. Progress and limits of PrP gene selection policy. Vet. Res. (39) 25–37. http://dx.doi.org/10.1051/vetres:2007064 (31. mar. 2008)

Hunter, N. 1997. PrP genetics in sheep and the applications for scrapie and BSE. Trends Microbiol. (5) 331-334. http://dx.doi.org/10.1016/S0966-842X(97)01081-0 (31. mar. 2008)

Kerr, R.J., Kinghorn, B.P. 1996. An efficient algorithm for segregation analysis in large populations. J. Anim. Breed. Genet. (113) 457–469.

Lühken, G., Lipsky, S. Peter, C. Erhardt, G. 2008. Prion protein polymorphisms in autochtonous European sheep breeds in respect to scrapie eradication in affected flocks. Small. Rumin. Res. (75) 43–47. http://dx.doi.org/10.1016/j.smallrumres.2007.07.010 (10. mar. 2008)

Sobel, E., Papp, J.C., Lange, K. 2002. Detection and integration of genotyping errors in statistical genetics. Am. J. Hum. Genet. (70) 496–508. http://dx.doi.org/10.1086/338920 (31. mar. 2008)

Thallman, R.M., Bennet, G.L., Keele, J.W., Kappes, S.M. 2001a. Efficient computation of genotype probabilities for loci with many alleles: I. Allelic peeling. J. Anim. Sci. (79) 26–33. http://jas.fass.org/cgi/content/abstract/79/1/26 (31. mar. 2008)

Thallman, R.M., Bennet, G.L., Keele, J.W., Kappes, S.M. 2001b. Efficient computation of genotype probabilities for loci with many alleles: II. Iterative method for large complex pedigrees. J. Anim. Sci. (79) 34–44. http://jas.fass.org/cgi/content/abstract/79/1/34 (31. mar. 2008)

Thallman, R.M. 2002. User's Manual for GenoProb Version 2.0. Computation of genotype and phase probabilities in complex pedigrees by iterative allelic peeling. 18 pp.

Tier, B., Henshall, J.M. 2005. Limits to genotypic probabilities for single loci. In: Application of new genetic technologies to animal breeding: Proceedings of the 16th Conference of the Association for the Advancement of Animal Breeding and Genetics 2005-09-25/2, CSIRO, Noosa Lakes, Queensland, Australia 366-369.

Ulvund, M.J. 2008. Ovine scrapie disease: Do we have to live with it? Small Rumin. Res. (in press) http://dx.doi.org/10.1016/j.smallrumres.2007.12.004 (31. mar. 2008)