
Article:
Slate, J. (2005) Molecular evolution of the sheep prion protein gene. Proceedings of the Royal Society B: Biological Sciences, 272 (1579). pp. 2371-2377. ISSN 1471-2954

https://doi.org/10.1098/rspb.2005.3259

ELECTRONIC APPENDIX

This is the Electronic Appendix to the article

Molecular evolution of the sheep prion protein gene

by

Jon Slate

Proc. R. Soc. B (doi:10.1098/rspb.2005.3259)

Electronic Appendix, part A

Table 2: GenBank Identifiers of sequences used in this study. Consensus sequences are shown in bold.

Species	GenBank Identifier (GI)
Sheep (*Ovis aries*)	13784993, **16215707**, 16215709, 6110614, 5714675, 2398736, 2330625, 2330621, **1778172**, 6010644, 2398744, 2398742, 2330623, 2407959, **2769529**, **2398746**, 2398740, 2809230, 38145690, 38145692, 38145694, **38145696**, 38145698, 38145702, 38145704, 42557972
Goat (*Capra hircus*)	1848295, 19698515, 400442, 1149616, 19698509, 19698513, 19698507, **19698511**, 4406468, 4406470
Cattle (*Bos taurus*)	217593, 188342, **X55882.1**, **217595**, **13810180**, **266111**, **21617492**, 29838421, 21666989, **34334035**, 34334037, 34334039, **34334041**, 34334043, 34334045, 41386734
Auroch (*Bos primigenius*)	**217597**
Red deer (*Cervus elaphus*)	**1711299**
Elk (*Cervus elaphus canadensis*)	8885889, 5069432, **8885891**, 5069439, 29650244, 2734628
Whitetail deer (*Odocoileus virginianus*)	8885893, 30984059, 30692036, 30692032, **8885895**
Mule deer (*Odocoileus hemionus*)	**6166131**, 881349, 29824919

Electronic Appendix, part B

McDonald-Kreitman tests on sheep and goat PRNP variation.

It has been demonstrated that tests of neutral evolution based on the relative rates of d_N and d_S in a phylogeny are sensitive to the effects of undetected recombination (Anisimova et al. 2003; Shriner et al. 2003). Here, only one sheep and one goat sequence can possibly be recombinant because most branches represent a single, unique mutation. When these possible recombinant sequences were excluded from the codeml analysis (sheep GI13784993 and goat GI19698513) the results were qualitatively identical.

However, additional evidence for the relaxation of purifying selection in sheep and goats is provided by McDonald-Kreitman tests (McDonald & Kreitman 1991). Here the ratio of nonsynonymous to synonymous polymorphism within species is compared to the ratio of nonsynonymous to synonymous substitutions between species in a 2x2 contingency table. Fisher's Exact Test is used to determine departures from neutrality. Because this analysis is based simply on the presence of substitutions and polymorphisms (rather than their frequency or where they arise on a phylogeny) it is not sensitive to recombination. Both sheep and goats were compared to the consensus sequence of each of the other 6 ruminant species (Electronic Appendix Table 2). Only polymorphism within the focal species (i.e. sheep or goat) was included in these analyses to avoid the potentially confounding effect of alternative selection regimes between the focal species and comparison species. In 5 of the 6 comparisons involving sheep there was a significant excess of within species nonsynonymous polymorphism. Only 2 of the 6 comparisons involving goats were statistically significant, although in every case the nonsynonymous:synonymous ratio was greater within species than between species. These findings lend support to the codeml analysis and are consistent with the relaxation of purifying selection (i.e. positive selection, balancing selection or segregating slightly deleterious alleles) in sheep and goats, in marked contrast to elsewhere in ruminant evolution.

Table 3: McDonald Kreitman tests comparing sheep and goats to 6 other ruminant species.

	Sheep			Goat		
Comparison	N/S within	N/S between	*P*	N/S within	N/S between	*P*
cattle	14/5	6/10	0.044	6/2	6/10	0.193
auroch	14/5	3/6	0.095	6/2	3/7	0.153
red deer	14/5	5/11	0.018	6/2	5/11	0.082
elk	14/5	5/12	0.018	6/2	5/12	0.081
whitetail deer	14/5	4/14	0.003	6/2	4/14	0.026
mule deer	14/5	4/11	0.014	6/2	4/12	0.032

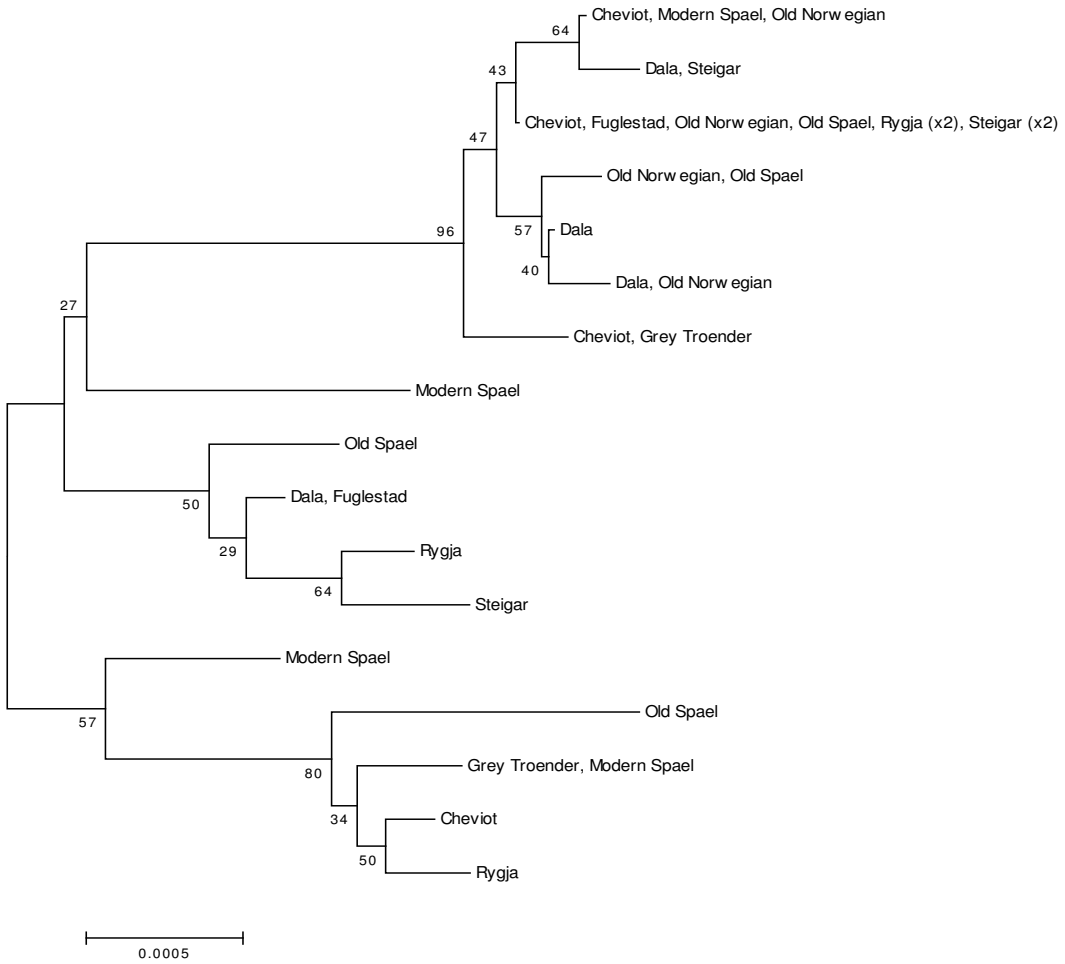
N/S is the number of nonsynonymous and synonymous substitutions within species and between species. The statistical significance (*P*) of the difference between the two ratios was tested using Fisher's Exact Test (2-tailed).

Electronic Appendix, part C

Can positive values of allele frequency tests be caused by population structure?

Positive values of D_T and other allele frequency tests can be caused not only by balancing selection but also by population structure. If population structure was present in this sample then haplotypes from specific breeds should cluster on neighbouring branches. The Juke-Cantor estimate of the number of nucleotide substitutions per site was low (d <0.002), so a neighbour-joining tree was constructed using the p-distance estimate of sequence divergence (Kumar et al. 2001) (see Figure 3)

Figure 3. An unrooted neighbour-joining tree of all 32 sheep exon 3 haplotypes. Bootstrap values (1000 replications) are shown.



The 32 *PRNP* haplotypes appear to form 3 distinct groups and haplotypes from the same breed are distributed apparently randomly across these groups. Thus, there is no evidence of population structure caused by between-breed differences at *PRNP*. However, the presence of 3 distinct haplogroups is consistent with long-term balancing selection.

Assuming a molecular clock and a divergence date between sheep and cattle of 19.6 million years ago (MYA) (Kumar & Hedges 1998) the time to the most recent common ancestor (TMRCA) of the 32 sheep haplotypes is approximately 1.3 MYA. If population structure is present in this sample then it can only be attributed to sheep of different geographical origins (i.e. different domestication events) being represented in the ancestors of modern Norwegian sheep breeds. Sheep probably have been domesticated more than once, but European breeds all appear to be descended from a single mouflon population in the Near East (Hiendleder et al. 2002). Thus, population structure as a cause of positive values of D_T cannot be unequivocally rejected, but it is a less parsimonious explanation of the data than balancing selection.

Can positive values of allele frequency tests be caused by population bottlenecks?

It is likely that sheep populations underwent a bottleneck during domestication. Depending on their duration and intensity, population bottlenecks can cause D_T and related statistics to take positive or negative values. Rare polymorphisms tend to be lost faster than intermediate frequency polymorphism, thereby causing allele frequency statistics such as D_T to take positive values, at least in the short term (Fay & Wu 1999; Wooding et al. 2004). There are several reasons to suspect that population bottlenecks are not the cause of the apparent departure from neutrality of sheep *PRNP*. First, the genealogy of the haplotypes does not include a star-like pattern of recent evolution that might be expected following population growth after a bottleneck caused by domestication (Electronic Appendix Fig 1). Second, cattle which are likely to have undergone similar population demographies to sheep following domestication, do not provide positive values of D_T at *PRNP* (Seabury et al. 2004). Third, ovine *PRNP* has a high nucleotide diversity (relative to non-domesticated ruminants) which is consistent with balancing selection but not with population

bottlenecks. Fourth, it is approximately 5000 generations since sheep were domesticated (assuming domestication occurred 10,000 years ago and a mean generation time of 2 years). At 5000 generations post-bottleneck D_T had returned to ~ zero for neutrally evolving genes under all scenarios investigated by Fay & Wu (1999).

The definitive test of whether *PRNP* nucleotide diversity shows patterns consistent with balancing selection is to compare allele frequency test statistics at *PRNP* with the distribution of test statistics obtained from other genes. Furthermore the same sample of sheep must be examined at each gene. If *PRNP* has evolved by balancing selection it will be an outlier. Population structure and/or population bottlenecks should affect all genes. Unfortunately the data are not available to perform this test. D_T and related statistics were estimated for 19 sheep genes that were represented in GenBank 5 or more times. *PRNP* provided greater values of D_T, D* and F* than any other gene. However, GenBank submissions will not be representative of the sample studied here, so too much weight should not be attached to this additional evidence of balancing selection.

References

Anisimova, M., Nielsen, R. & Yang, Z. 2003 Effect of Recombination on the Accuracy of the Likelihood Method for Detecting Positive Selection at Amino Acid Sites. *Genetics* **164**, 1229-1236.

Fay, J. C. & Wu, C. I. 1999 A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution* **16**, 1003-1005.

Hiendleder, S., Kaupe, B., Wassmuth, R. & Janke, A. 2002 Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 893-904.

Kumar, S. & Hedges, S. B. 1998 A molecular timescale for vertebrate evolution. *Nature* **392**, 917-920.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. 2001 MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244-1245.

McDonald, J. H. & Kreitman, M. 1991 Adaptive Protein Evolution at the Adh Locus in Drosophila. *Nature* **351**, 652-654.

Seabury, C. M., Honeycutt, R. L., Rooney, A. P., Halbert, N. D. & Derr, J. N. 2004 Prion protein gene (PRNP) variants and evidence for strong purifying selection in functionally important regions of bovine exon 3. *PNAS* **101**, 15142-15147.

Shriner, D., Nickle, D., Jensen, M. A. & Mullins, J. I. 2003 Potential impact of recombination on sitewise approaches for detecting positive natural selection. *Genetical Research* **81**, 115-121.

Wooding, S., Kim, U. K., Bamshad, M. J., Larsen, J., Jorde, L. B. & Drayna, D. 2004 Natural selection and molecular evolution in PTC, a bitter- taste receptor gene. *American Journal of Human Genetics* **74**, 637-646.