High Inbreeding in Sheep or Erroneous Estimation?

PHILIP W. HEDRICK

From the School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501.

Address correspondence to Philip W. Hedrick at the address above, or e-mail: philip.hedrick@asu.edu.

Paiva et al. (2011) examined how microsatellite loci data from sheep surveys in the United States and Brazil could be merged when they are analyzed in different laboratories. In addition, they calculated the inbreeding coefficient using $F_{IS}$ from the software FSTAT (Goudet 2001) using the formula $F_{IS} = 1 - H_O / H_E$, where $H_O$ and $H_E$ are the observed and expected Hardy–Weinberg heterozygosities. For 11 microsatellite loci, they found that for the 12 sheep breeds from the United States, the average $F_{IS}$ value was 0.205, and for the 9 sheep breeds from Brazil, the $F_{IS}$ was –0.005. They suggested that “The higher levels of inbreeding ($F_{IS}$) observed in some US breeds could possibly be due to smaller population sizes or a higher level of selection pressure” (Paiva et al. 2011).

It is unclear how either smaller population sizes or more selection pressure could increase $F_{IS}$ to this high level found in the US breeds. On the other hand, the high $F_{IS}$ values found in the US breeds could have been generated not by inbreeding within each breed but by heterogeneity among the flocks sampled within each breed, caused by the Wahlund effect (Hedrick 2011), and the sampling of only a few animals per flock. Specifically, the formula used above for $F_{IS}$ assumes that there is no population structure within each breed among the flocks sampled, that is, no Wahlund effect, which occurs when heterogeneous samples (different flocks here) with different allele frequencies are lumped together. Even with Hardy–Weinberg proportions within each flock (no inbreeding), combining heterogeneous samples results in an apparent deficiency of observed heterozygotes, a finding that Paiva et al. (2011) attributed to inbreeding.

In another article (Blackburn et al. 2011), some of the same authors also found, using some of the same data, similarly high $F_{IS}$ values in the US breeds when 31 microsatellite loci were used. Hedrick (2012), in a theoretical effort to replicate these data for the Navajo-Churro sheep breed, showed that the $F_{IS}$ values in the study by Blackburn et al. (2011) were probably generated by a combination of population structure and small sample size from different flocks within a breed and not inbreeding within a breed. More specifically, Hedrick (2012) assumed that individual flocks within a breed were founded by samples of a given size, assumed Hardy–Weinberg proportions within each flock, and then drew a random sample of a given size from each flock for genetic analysis. For example, if the flocks were initiated with approximately 3 effective founders and 2 sheep were sampled from each flock as in the study by Blackburn et al. (2011), there were different enough allele frequencies by chance in different flocks so that even with every flock in Hardy–Weinberg proportions, $H_O$ was similar to that found by Blackburn et al. (2011), and $H_E$ was substantially less than $H_O$. Therefore, using the formula above as Blackburn et al. (2011) did to calculate $F_{IS}$, high values are found, not due to inbreeding, but due to the heterogeneity in allele frequencies among flocks, caused by a combination of chance founding events and sampling effects.

For the different breeds, Paiva et al. (2011) gave the number of flocks (breeders) sampled. The mean number of breeders per breed in the United States was 13.4 and the mean number of breeders per breed in Brazil was 1.8. After analysis of the 28 breeds examined by Blackburn et al. (2011), Hedrick (2012) found a similar positive association of high $F_{IS}$ and high number of breeders (low sample size per flock). The positive association of the $F_{IS}$ estimate and the number of breeders sampled is an unexpected result if the association is the result of inbreeding; but this is not unexpected if the association is the result of population structure and the sampling of small numbers of sheep per flock within a breed.

Maiwashe and Blackburn (2004) used pedigree data to examine inbreeding in Navajo-Churro sheep and estimated a fairly low average inbreeding level of 0.012. Surprisingly, given this low estimate from pedigree data, Blackburn et al. (2011) and Paiva et al. (2011) were not skeptical of the very high inbreeding estimates for this breed from the microsatellite data (0.152 and 0.265, respectively). The large discrepancy between these estimates should have led them to reexamine their approach and alerted them to their erroneous estimation of inbreeding using microsatellite data.

As an illustration of the comparison of pedigree inbreeding and inbreeding estimation from molecular data, Li et al. (2011) recently examined individual inbreeding coefficients in 99 Finnsheep using both a large pedigree composed of 319 000 sheep and a molecular data set of 48 000 single nucleotide polymorphisms. The inbreeding levels estimated using these 2 different data sets and approaches were generally consistent, particularly for sheep with inbreeding coefficients $>0.0625$. Li et al. (2011)
also identified 3 genetic subpopulations in Finnsheep, which generally corresponded to different coat colors \((F_{ST} = 0.054)\), indicating substantial genetic structure. If this within-breed population structure was ignored, then estimates of inbreeding using the molecular data were significantly inflated, consistent with the suggestions above for the analysis of Paiva et al. (2011).

Sampling a few sheep from many flocks within a breed is a good strategy to capture a substantial amount of variation in each breed but not a good approach to estimate inbreeding, as demonstrated in the preceding paragraphs (see also discussion by Keller and Waller 2002). On the other hand, estimation of inbreeding from molecular data is possible if the sample comprises a large number of sheep from a given flock (as for the Brazilian breed samples that had near-zero \(F_{IS}\) estimates). Fortunately, now that sheep (and other species) have extensive genomic data, many independent loci can also be used in each individual to estimate individual inbreeding coefficients within a breed (Keller et al. 2011; Li et al. 2011).

Acknowledgments
I wish to thank Jérôme Goudet and Marty Kardos for discussions.

References
Goudet J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Lausanne (Switzerland): University of Lausanne.

Received October 8, 2012; Revised December 2, 2012; Accepted December 6, 2012

Corresponding Editor: Ernest Bailey