A Note on Interannual Variation of F_{is} for Bowhead Whales Sampled at Barrow

GEOF H. GIVENS, Dept. of Statistics, Colorado State University

Jorde and Schweder (2007) present a boxplot of locus-specific estimates of F_{is} for a sample of Bering-Chukchi-Beaufort Seas (BCB) bowhead whales split by sample year, for 206 whales and 33 microsatellite loci. In their figure, the 8 samples from 1992 exhibit F_{is} estimates that vary more widely and average higher than do samples from other years (Jorde and Schweder, 2007, Figure 1). They cite this result as evidence that BCB bowhead samples are "not representing a single biological population."

There are other more likely explanations. I re-analyzed thebowhead data to investigate the finding of Jorde and Schweder. The data I used comprised microsatellite scores for 213 whales caught at Barrow, for 33 loci. These are the same loci and nearly the same whales used by Jorde and Schweder. Indeed, using my data I can replicate their boxplot; see Figure 1.

There are several aspects of the 1992 samples that are special. First, 1992 is among the years for which the fewest samples (8) are available, so 1992 estimates of allele frequencies and F_{is} are highly variable compared to other years.

Second, three 1992 individuals (92B3, 92B5, and 92B6) have unusual homozygosity patterns: all are homozygous for rare alleles. Table 1 summarizes the microsatellite genotypes for these individuals. Furthermore, the extent of homozygosity in these whales (19/31, 20/33, and 17/31, loci respectively) is about triple the median homozygosity in the dataset; see Figure 2. Such results evoke concerns over the possibility of allelic dropout, unequal allele amplification, null alleles and related scoring errors. Morin and LeDuc (2004) examined such issues for 11 of the bowhead loci used here. They reported (among other findings) that homozygosity and PCR success rate were negatively correlated, suggesting that low DNA quality or quantity can affect genotype quality. Morin et al. (2007) report on microsatellite scoring errors in the present data and their potential influence on detection of deviations from Hardy-Weinberg equilibrium. Three whales from 1992 are identified in Table 2 of Morin et al. (2007), including two of the three whales listed in Table 1 here. The other outlier in my Figure 2 is also identified by Morin et al.

Third, two of the 1992 whales listed in Table 1 were caught on the very same day at Barrow, raising the possibility that they might be relatives. (Incidentally, two whales from 1992, including 92B3, are detected by Skaug and Givens (2007) as related to other whales in the dataset, but not to other 1992 whales).

Fourth, these three whales (92B3, 92B5, and 92B6) are all rather large old whales. They were caught on August 31, September 2, and September 2, respectively, at the very start of the fall whaling season at Barrow during the period renowned for the presence of the largest, oldest whales. Figure 3 shows a boxplot of birth year index split by sampling year. This birth year index is an estimate based on the fit of a von Bertalanffy growth curve (Rosa et al., 2004). Age estimates were truncated at 100 years before translation to birth year. This index is not meant to be taken as a direct estimate of birth year because the growth curve model is very approximate. Instead this index is meant as a rough approximation that adjusts for the important size difference between genders and which provides a consistent relative index across sampling years. Figure 3 shows that the whales sampled in 1992 have birth year index values that vary more widely and average much earlier than do whales caught in other years.

Fifth and finally, these three whales are highly influential on the finding of Jorde and Schweder. Figure 4 shows the results of the same analysis after omitting these three whales.

In conclusion, when data are repeatedly stratified into increasingly smaller groups the impact of a few outliers can be magnified. In the present case, the most likely explanation is that these three whales cause the 1992 anomaly because of some erroneously scored alleles and/or because of genetic drift effects (for which birth year is a signal carrier) as simulated by Archer et al. (2007) and detected by Givens et al. (2007). The alternative explanation that these three whales were members of some other stock which happened to be present or detected in 1992 but not since then seems comparatively unlikely.

Acknowledgements

This work was supported by the North Slope Borough (Alaska) and the National Oceanic and Atmospheric Administration (through the Alaska Eskimo Whaling Commission).

References

Archer, E., Martien, K., Taylor, B.L., LeDuc, R.G., Givens, G.H., and George, J.C. (2007) Use of an individual-based simulation of BCB bowhead whale population dynamics to examine empirical genetic data. Paper SC/59/BRG17.

Givens, G.H., Huebinger, R.M., Bickham, J.W., George, J.C., and Suydam, R. (2007) Patterns of genetic differentiation in bowhead whales (*Balaena mysticetus*) from the western Arctic. Paper SC/59/BRG14.

Jorde, P.E. and Schweder, T. (2007) Further analysis of stock structure for BCB bowhead whales using genetic data. Paper SC/59/BRG27.

Morin, P.A. and LeDuc, R.G. (2004) Analysis of bowhead DNA quantity and microsatellite characteristics: implications for potential biases in population structure analysis. Paper SC/56/BRG34.

Morin, P.A., LeDuc, R.G., Archer, E., Martien, K.K., Taylor, B.L., Huebinger, R., and Bickham, J.W. (2007) Estimated genotype error rates from bowhead whale microsatellite data. Paper SC/59/BRG15.

Rosa, C., George, J.C., Zeh, J., O'Hara, T.M., Botta, O., and Bada, J. (2004) Update on age estimation of bowhead whales using aspartic acid racemization. Paper SC/56/BRG6.

Skaug, H.J. and Givens, G.H. (2007) Relatedness among individuals in BCB bowhead microsatellite samples. Paper SC/59/BRG20.

All these papers are available from the IWC Secretariat.

Locus	92B3	92B5	92B6	Locus	92B3	92B5	92B6	Locus	92B3	92B5	92B6
Bmy1	H	H	H	**Bmy26**	9	H		**Ev1**		-	-
Bmy2	-	H	-	**Bmy33**	-	H	H	**Ev104**		-	-
Bmy7	-	H	11	**Bmy36**	-	H	-	**Gata28**	-	-	-
Bmy8	-	H	H	**Bmy41**	10	-	H	**Tv7**	H	-	-
Bmy10	9	H	H	**Bmy42**	H	H	-	**Tv11**	H	H	H
Bmy11	H	H		**Bmy49**	-	H	-	**Tv13**	-	-	H
Bmy12	-	-	-	**Bmy53**	H	-	13	**Tv14**	-	H	-
Bmy14	15	H	H	**Bmy54**	H	-	H	**Tv16**	-	H	H
Bmy16	H	H	H	**Bmy55**	H	H	H	**Tv17**	H	-	-
Bmy18	H	H	H	**Bmy57**	2	-	H	**Tv19**	H	-	-
Bmy19	-	H	H	**Bmy58**	15	13	-	**Tv20**	H	-	-

Table 1: Summary of homozygosity for three 1992 samples. 'H' represents homozygosity, '-' represents heterozygosity, and blank indicates missing data. The numbers represent homozygosity for rare alleles, where rareness is taken to be allele frequencies less than about 5% (20 or fewer alleles in the sample of 426 alleles). In these cases, the number indicates the total frequency for the allele for which the individual is homozygous.

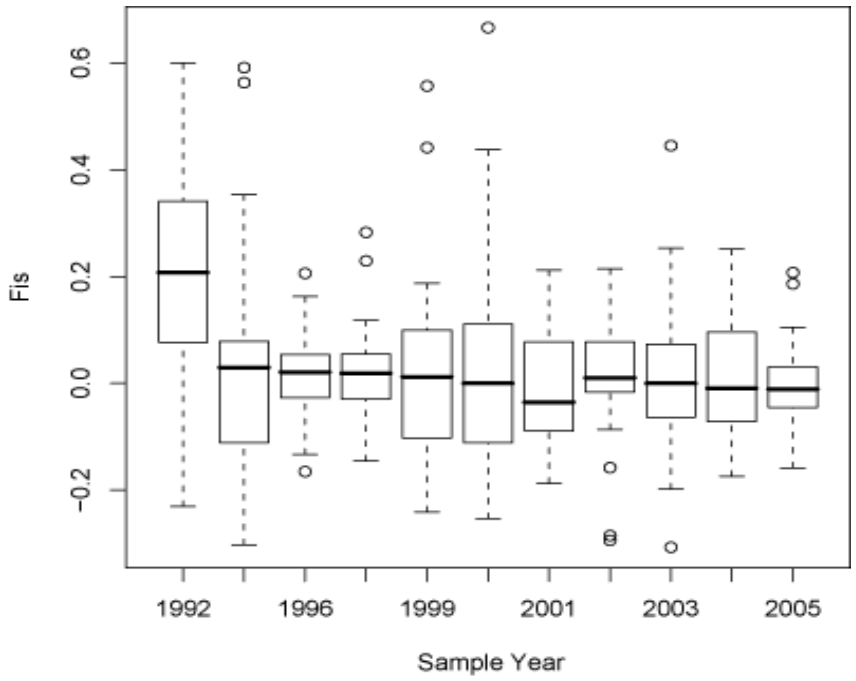

Figure 1: Replication of the result from Jorde and Schweder (2007). Although not all years are labeled, they are the same years used in the Jorde and Schweder analysis, specifically 1992, 1993, 1996, 1997, and 1999-2005.

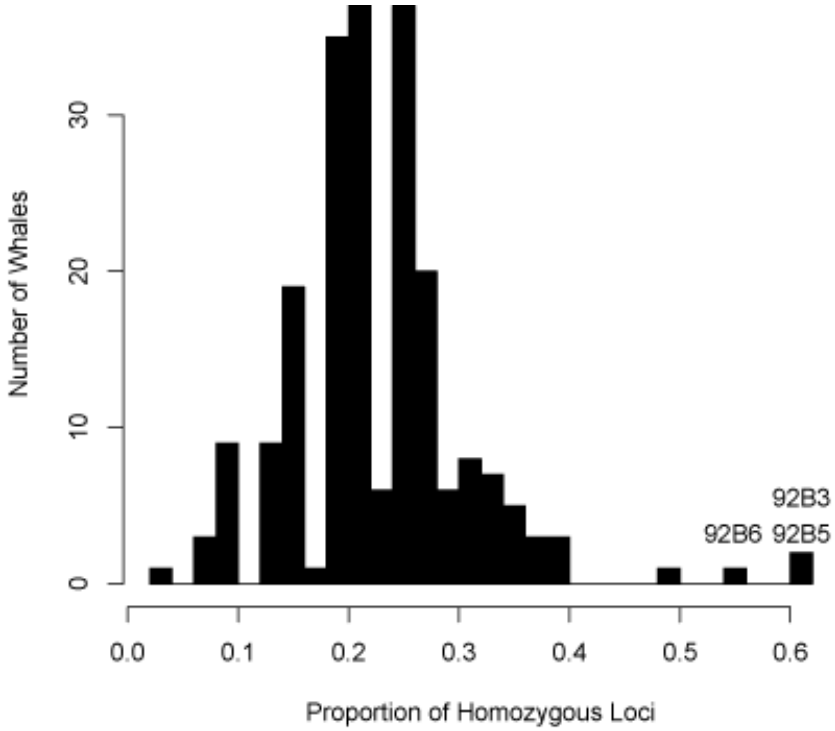

Figure 2: Proportion of scored loci (for each whale) that were homozygous. Three outliers are labeled.

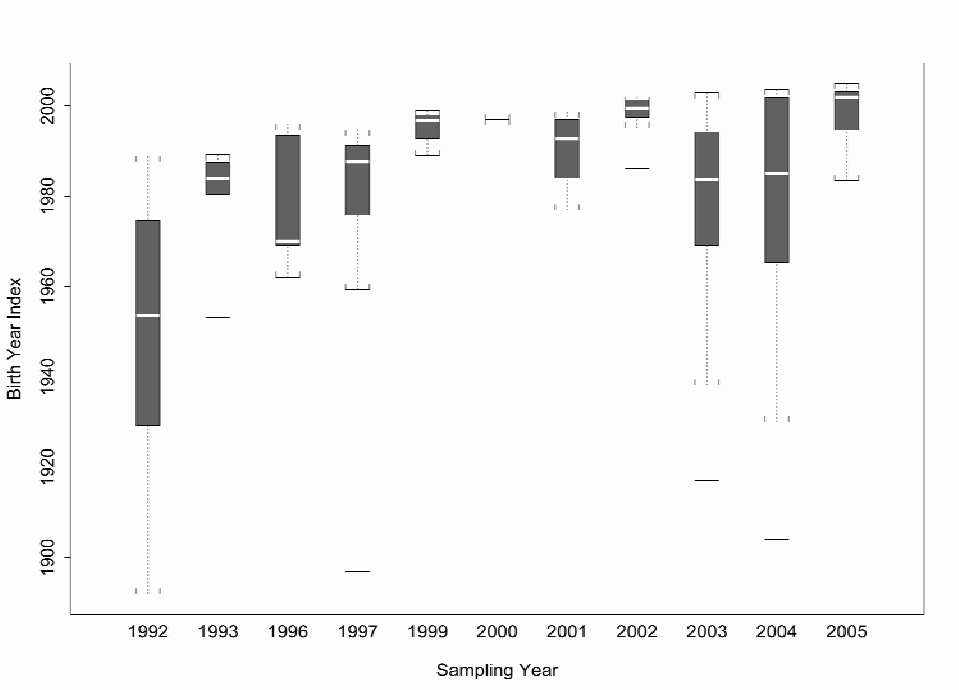

Figure 3: Birth year index for whales caught in each year.

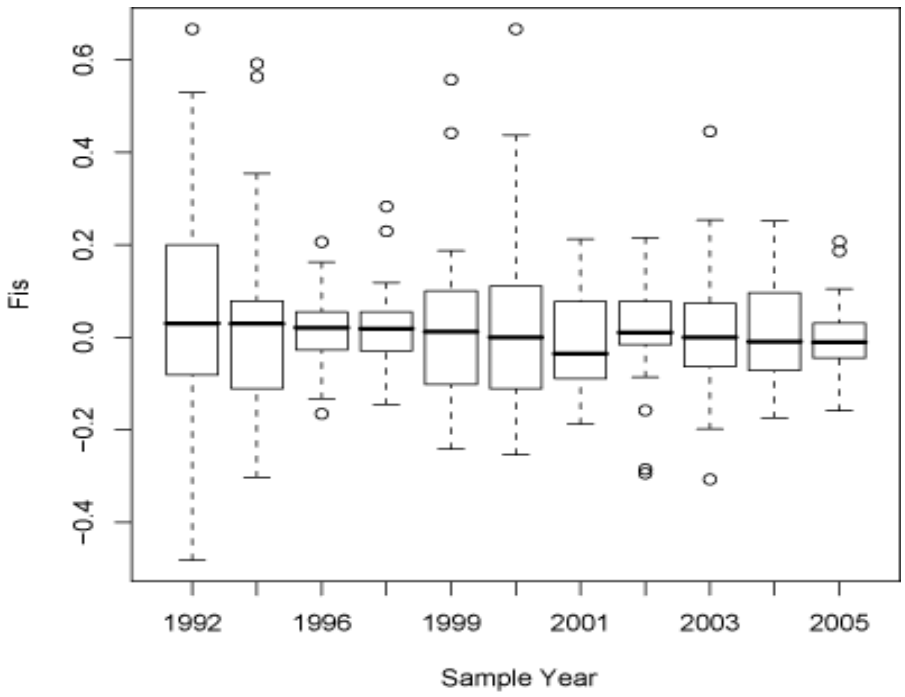

Figure 4: Result after omitting three whales. Although not all years are labeled, they are the same years used in the Jorde and Schweder analysis.