Distribution of correlation coefficients between θ and F_{IS} in mixed samples of two distinct stocks: Comments to Waples (2011)

Naohisa Kanda and Hiroshi Hatanaka

The Institute of Cetacean Research, 4-5, Toyomi-cho, Chuo-ku, Tokyo 104-0055, Japan

Contact email:kanda@cetacean.jp

ABSTRACT

Waples (2011) presented very elegant analyses based on standard population genetics theory to test the two-stock hypothesis (i.e., O and J stocks in Hypotheses A and B) proposed for the North Pacific minke whales around Japan. Logics of the analyses were that, when only two distinct stocks exist in a given area, the largest departures from equilibrium (described as F_{IS}) or higher linkage disequilibrium should be seen at the loci or pairs of loci that show the largest or strong allele frequency differences between the two distinct stocks (described as θ). Because the observed relationships of these genetic indices in the samples of the North Pacific minke whales from SA7 bycatch, SA7-Kushiro, SA7-Sanriku, SA2, and SA11 were weaker than the expected relationship estimated from artificial mixtures of only putative J and O individuals, it was suggested the samples might have contained individuals from more than two stocks. However, it was also indicated that the robustness of the analyses should be evaluated because this approach was novel and the behaviors of these indices were unclear under various situations such as different samples sizes. In this document, using computer simulation, we examined the effect of different sample sizes to the distributions of the correlations of these genetic indices.

Among the results presented in Waples (2011), we only examined the correlation between θ and F_{IS} in 1:1 mixture of individuals from the two distinct stocks that reflected the situation of the bycatch sample from SA7 (J=93, O=90, unclassified=29) because we thought the argument to this single case should cover the other cases. We generated 100 genetic data that contained genotypic data at 16 loci through computer simulations. We examined distribution of the correlation coefficients under the three different cases in regard to the sample size: 5000 (same as simulated Ne, total sample size=10000), 400 (SA6 and SA9, 800), and 100 (SA7 bycatch, 200) individuals from each of the two putatively distinct stocks that were differentiated each other with F_{ST} values around 0.05. This level of genetic differentiation was comparable to that previously estimated between the putative J and O stocks (Kanda *et al.*, 2009). After the data generation, we look at the distribution of correlation coefficients between θ and F_{IS} of the 16 loci among the 100 data in each of the three cases. The correlation coefficients distributed as expected from Waples (2011) in the cases of 5000 and 400, but fluctuated quite widely and some of the values were quite low in the case of 100. The results of our simulation exercises indicated that lower level of the relationships observed in the SA2, SA7, and SA11 samples in Waples (2011) could be due to the small sample size rather than due to mixture of the individuals from more than two stocks. The conclusion of Waples (2011) therefore needs further evaluation to be used for the evidence against Hypotheses A and B.

KEY WORDS: NORTH PACIFIC MINKE WHALE, SIMULATION, O STOCK, J STOCK, *IMPLEMENTATION* ASSESSMENT

INTRODUCTION

We have presented the evidence of statistically significant departures from the expected Hardy Weinberg genotypic proportions with homozygous excess in the samples from SA7 (including bycatches and scientific catches) as one of the supportive results for that minke whales in the area are the mixture of the two stocks, J and O, rather than an additional intermediate stock (Kanda et al., 2010). Waples (2011) indicated based on standard population genetics theory that if there are only two stocks, loci with the largest allele frequency differences between them (described as θ) should show the largest departures from equilibrium (described as F_{IS}), and linkage disequilibrium should be higher for pairs of loci that show strong allele frequency differences between the two distinct stocks. In order to test the two-stock hypothesis, then, Waples (2011) used the samples from SA6 and SA9 to characterize putative pure J and O stocks, artificially mixed the two samples to look at the expected relationships of the above genetic indices, and compared these genetic indices to the ones observed from the natural samples obtained from SA2, SA7 and SA11 where it was believed that whales from only the

two stocks existed according to Hypotheses A and B. Because the observed values were lower than the expected ones, Waples (2011) concluded that the samples from SA2, SA7 and SA11, respectively, might have contained individuals from more than two stocks. However, it was also raised that the robustness of the analyses should be evaluated because this approach was novel and the behaviors of these values were unclear.

In this document, in order to verify Waples (2011), we describe the expected distribution of the correlations between θ and F_{IS} using computer simulation and examine the effect of sample size on the distributions. Among the results presented in Waples (2011), we only examined the case of 1:1 mixture of individuals from the two distinct stocks that reflected the situation of the bycatch sample from SA7 (J=93, O=90, unclassified=29).

METHODS

The software EASYPOP (Balloux, 2001) was used to generate 100 sets of genotypic data for each of the scenarios and the software FSTAT 2.9.3 (Goudet, 1995) was to calculate θ (Weir and Cockerham, 1984) per locus between the stocks and F_{IS} per locus for the data two stocks combined as one.

For the generation of the genotypic data, two stocks were assumed, each of which consists of diploid individuals with a constant size and equal sex ratio with random mating. Effective population size (Ne) of each stock was set to be 5000. Census population size (N) of Ne=5000 can be approximately 20000 when ratio of Ne to N to be 1/3 to 1/4 (Roman and Palumbi, 2003), which was comparable to the IWC's accepted population abundance of the minke whales in the North Pacific. The simulation produces genotype data set for 16 independent nuclear gene loci for each individual. The number of the loci simulated and maximum number of the allelic states (13) was set based on the observed minke whale data. Bidirectional migration was assumed with an equal migration rate of 0.0002. Mutation rate of $5x10^{-4}$ was chosen to represent microsatellite loci. For each simulation parameter set, we made 100 replicates. Number of generation was 5000 for each replicate before collecting data. In order to look at the effect of sample size, we conducted three different cases of the simulations: at the final generation of each replicate, all of the 5000 individuals each was sampled, 2) 400 individuals each were sampled, 3) 100 individuals each were sampled from two simulated stocks for genetic analysis. The sample size of 400 (total sample size=800) approximately reflects the case of artificial mix of the samples from SA6 (411) and SA9 (466) and the sample size of 100 (total sample size=200) does the case of the SA7 bycatch sample (J=93, O=90, unclassified=29) in Waples (2011).

RESULTS AND DISCUSSION

Values of θ between the two generated stocks ranged from 0.038 to 0.065 (avg=0.051, sd=0.006) among the 100 data set in the case of sample size 5000, 0.037 to 0.066 (0.050, 0.006) among the 100 data set in the case of sample size 400, and 0.033 to 0.070 (0.050, 0.007) among the 100 data set in the case of sample size 100. Because the F_{ST} value between the J and O stocks we observed was 0.049 (Kanda et al., 2009), these generated data should be reasonable to examine the distribution of correlations between the F_{IS} and θ for our purpose.

Fig.1 shows the distributions of the correlation coefficient values among the 100 data in each of the three different sample size scenarios. When all the individuals in the simulated stocks were used, very high correlations were observed in all of the 100 data set. This result demonstrated that, in theory, the proposed hypothesis in Waples (2011) based on population genetics was correct. Although the observed correlation coefficients from the case of the sample size 400 fluctuated, most of the values were high and the results supported the high correlation (0.83) observed from the artificial mixture of the individuals from SA6 and SA9 in Waples (2011). Contrary to these two cases, the correlation coefficients among the 100 data set in the case of sample size 100 fluctuated widely with some negative values, and about 40% of the values were around 0.3 or smaller. The reliability of the results can be affected strongly by the sample sizes and thus the low correlation coefficient observed from the SA7 bycatch sample (0.3) in Waples (2011) could be due to the small sample size.

In this paper, we only examined the case of SA7 bycatch sample. Because the sample sizes for the other cases (SA7-Kushiro, SA7-Sanriku, SA2, and SA11) were similar to or lower than the SA7 bycatch sample, we think that the same pattern of fluctuated distributions will likely be observed when we simulate the other samples. Similarly, because both F_{IS} and linkage disequilibrium describe the departure from equilibrium, the sample size effect will likely be appeared for the relationship between θ and linkage disequilibrium, too.

In regard to the stock structure issue for North Pacific minke whales, the conclusion of Waples (2011) therefore needs further evaluation to be used for the evidence against Hypotheses A and B that proposed only two stocks in the western North Pacific.

ACKNOWLEDGEMENTS

We thank Toshihiro Mogoe for technical assistance and Luis A. Pastene for valuable comments on the paper.

REFERENCES

Balloux, F. 2004. EASYPOP (version 1.7): a computer program for the simulation of population genetics. J. Hered. 92:301-302.

Goudet, J. 1995. FSTAT, version 1.2: a computer program to calculate F-statistics. J. Hered. 86:485-486.
Kanda, N., Goto, M., Kishiro, T., Yoshida, H., Kato, H., and Pastene, L.A. 2009. Microsatellite analysis of minke whales in the western North Pacific. Paper SC/J09/JR30 presented to the JARPN II Review Workshop, Tokyo, January 2009.
Kanda, N., Goto, M., Nagatsuka, S., Kato, H., Pastene, L.A., and Hatanaka, H. 2010. Analyses of genetic and non-genetic data do not support the hypothesis of an intermediate stock in sub-area 7. Paper SC/S10/NPM9 presented to the First Intersessional Workshop for western North Pacific common minke whales, Tokyo, September 2010.
Roman, J, and Palumbi, S.R. 2003. Whales before whaling in the North Atlantic. Science 301:508-510.
Waples, R.S. 2011. Can evidence for spatial and/or temporal genetic heterogeneity of North Pacific minke whales be explained by different mixture fractions of the same two core stocks, or is it necessaru to postulate an additional stock (s)? Paper SC/63/RMP7 presented to the IWC Scientific Committee, May 2011, Tromso, Norway
Weir, B. S. and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

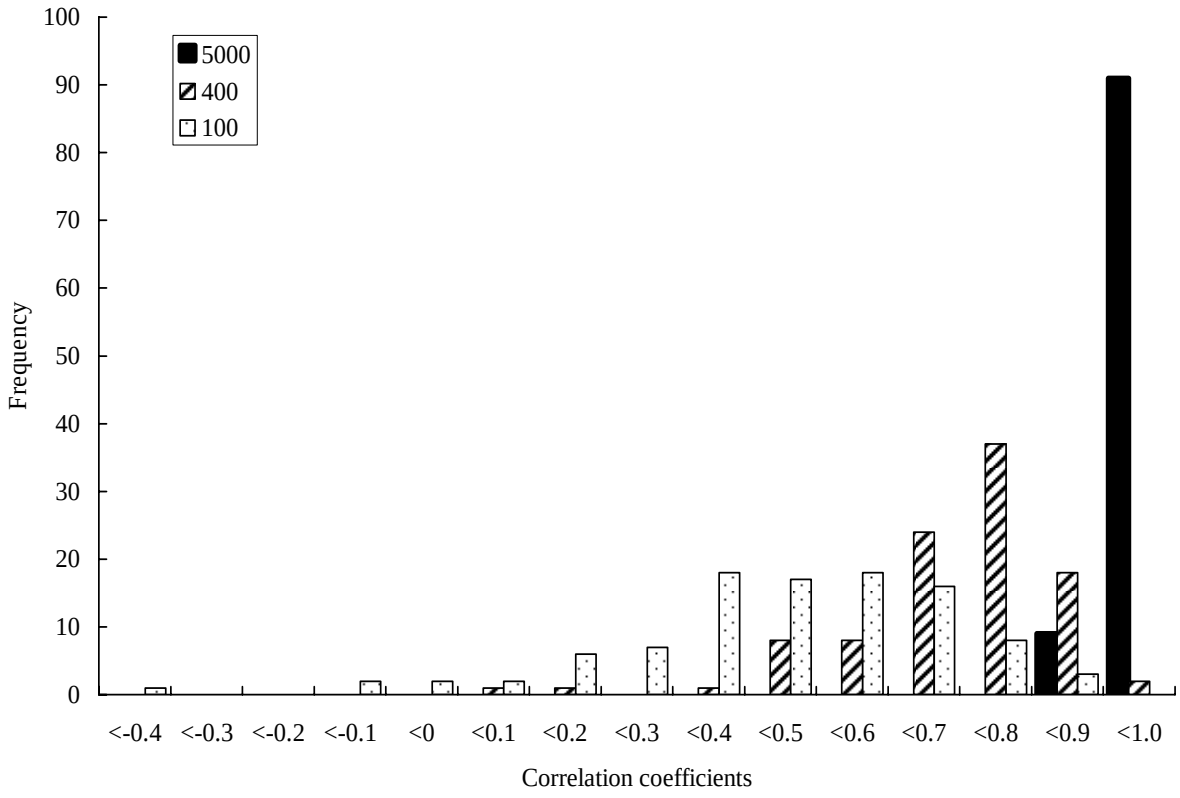

Fig.1. Distribution of correlation coefficients between θ and F_{IS} among the 100 computer generated genotypic data set for different sample sizes (5000, 400, and 100) from each of the two putatively distinct stocks.