Supplementary File 1) Justification for the use of ANOVAs

White-adjusted ANOVAs were used to test for differences in δ15N and δ13C among the broad time intervals (high ice, low ice, historic, and modern) because the White adjustment makes these parametric tests more robust to heteroscedasticity (White 1980; Long and Ervin 2000); however, it remains possible that the ANOVAs were affected by the non-normality of the data (though ANOVAs are typically robust to non-normality; Blanca et al. 2017). To assess this possibility, the analyses were re-run using Kruskal-Wallis Tests, followed by Dunn’s Multiple Comparison post hoc tests. The Kruskal-Wallis Test is a non-parametric test that is commonly used in place of a one-way ANOVA when data violate the assumption of normality. For δ15N, the results of the White-adjusted ANOVA and Kruskal-Wallis Test were identical. For δ13C, the same tests give slightly different results. When the non-parametric test is used, the difference between the modern samples and those from previous intervals of low ice is significant (p < 0.001), whereas this result was not significant when the White Adjusted ANOVA was used with the Dunnett’s Modified Tukey-Kramer Pairwise Multiple Comparison (DTK) Test (p = 0.072).

Though the Kruskal-Wallis Test does not assume normality, it is still sensitive to heteroscedasticity (McDonald 2014). The ANOVA, in contrast, is relatively robust to deviations from normality and the White Adjustment makes it robust to unequal variances. Given the structure of the data (non-normal and heteroscedastic), the results of the ANOVA are at least as likely to be correct as those of the Kruskal-Wallis Test. This is particularly true, given that the DTK post hoc test used in conjunction with the ANOVA was adjusted for unequal sample sizes and unequal variances(Dunnett 1980), whereas no such adjustment exists for the post hoc tests available for use with the Kruskal-Wallis Test. The single difference in the results of these two approaches does not substantially impact the conclusions of this paper. If anything, it serves to highlight the fact that the difference in mean δ13C between modern walruses and animals from previous periods of low sea ice was not large (magnitude of difference = 0.2 ‰), whether or not it was determined to be statistically significant.

Blanca, M. J., R. Alarcón, J. Arnau, R. Bono, and R. Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? Psicothema **29**: 552–557. doi:10.7334/psicothema2016.383

Dunnett, C. 1980. Pairwise multiple comparisons in the unequal variance case. Journal of the American Statistical Association **75**: 796–800.

Long, J. S., and L. H. Ervin. 2000. Using heteroscedasticity consistent standard errors in the linear regression model. The American Statistician **54**: 217–24. doi:10.2307/2685594

McDonald, J. H. 2014. Handbook of Biological Statistics, Sparky House Publishing, Baltimore, Maryland. Available at: http://www.biostathandbook.com/wilcoxonsignedrank.html.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica **48**: 817–838.