**Appendix B.** Questions asked in the Literature Review

**A. Descriptive questions about samples**

1. Samples analyzed were archaeological?

2. Samples analyzed were modern?

3. Number of samples that were analyzed?

4. Samples analyzed were bone or tooth collagen?

5. Samples analyzed were hair?

6. Samples analyzed were plants (modern)?

7. Samples analyzed were plants (archaeological)?

8. Samples analyzed were other materials not listed?

**B. Descriptive questions about archaeological collagen**

Answers to these questions were only recorded for archaeological bone collagen. These questions applied only to studies presenting isotopic data from archaeological bone collagen. Note that for the three collagen quality criteria (C:N, %C and %N, yield), these were only counted as being provided if individual values were listed. If it was stated that all of the samples had C:N ratios within the range of 2.9-3.6, this was counted as C:N ratio not presented. For the ultrafiltration, this step was recorded as being performed if it was explicitly stated to have been performed, if the collagen yields were listed as >30 kDa collagen yield, or if Brown et al. (1988) was cited in the methods section.

9. C:N ratio provided?

10. %C and %N provided?

11. Collagen yield provided?

12. Collagen was ultrafiltered?

**C. IRMS**

For the following questions, answers were only recorded if the details were explicitly stated. References made to other papers were not considered. The laboratory was not counted as being specifically named if it was only listed in the acknowledgements section and not in the materials and methods section.

13. Isotopic compositions were determined on-line or off-line?

14. Lab was specifically named?

15. Instrumentation specifically named?

**D. Data Presentation**

Data could be presented either in tables in the manuscript or in an online supplement. Summary statistics (means and standard deviations for particular contexts or sites) were not counted.

16. All values are presented and available online?

**E. Standards and Calibration**

17. Calibration standards are mentioned?

\*Note that this question does not refer to VPDB and AIR, but to the materials that were analyzed to generate the calibration curve.

18. Calibration standards are specifically named?

19. Calibration standards are international standard reference materials?

\*International standard reference materials are those produced by the IAEA, NBS, and USGS with known *δ­*-values that can be easily obtained.

20. Was a two-point calibration used?

21. Was the reference gas calibrated once and subsequently monitored but not calibrated during each analytical session?

22. If in-house standard reference materials were used to calibrate the data, were the known *δ*-values given?

23. Were the number of calibration standards that were analyzed tallied?

24. Check standards are mentioned?

25. Were the known *δ*-values for check standards provided?

26. Check standards have similar elemental compositions to the materials analyzed?

\*For example, most organic standards were considered to have similar elemental compositions to collagen or keratin, while inorganic standards such as ammonium sulfate or NBS oil were considered to be dissimilar to collagen or keratin.

27. Were the number of check standards that were analyzed tallied?

**F. Analytical Accuracy and Precision**

28. Analytical accuracy and precision (or error) is mentioned?

29. Analytical accuracy and precision are differentiated?

30. Vague reference to long-term or laboratory analytical error?

\*The statement was judged to be vague if analytical error is reported but it is unclear how this value was determined.

31. Specific reference to long-term or laboratory analytical error?

\*The reported error is linked to long-term analyses of calibration/check standards.

32. Analytical error reported is specific to the data produced in the paper?

\*The reported error is not long-term and was determined for the analytical sessions that produced the results in the paper.

33. Accuracy was specifically mentioned?

\*Accuracy had to have been mentioned separately from precision.

34. Accuracy was quantified via check standards?

35. Analytical precision was linked to check standards, sample duplicates, uncertainty of calibration standard measurements, or some combination of these?

**Duplication**

36. Were samples duplicated?

\*Sample duplication had to be explicitly stated but did not need to be specifically quantified.

37. What was the rate of sample duplication?

\*If all samples were analyzed in duplicate or triplicate, these were both scored as 100%.

38. Was the method of duplication defined?

39. Were the actual values of the duplicate pairs reported?

40. Was the average of the difference of the duplicates reported?

41. Was the standard deviation of each sample duplicate reported?

42. Was the average of the standard deviation of each duplicate pair reported?

**Scoring**

43. Calibration (Scale of 0-6)

44. Precision (Scale of 0-7)

45. Accuracy (Scale of 0-4)