

This is a repository copy of *Calling all archaeologists : guidelines for terminology, methodology, data handling, and reporting when undertaking and reviewing stable isotope applications in archaeology*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/125898/>

Version: Accepted Version

Article:

Roberts, Patrick, Fernandes, Ricardo, Craig, Oliver E orcid.org/0000-0002-4296-8402 et al. (4 more authors) (2017) *Calling all archaeologists : guidelines for terminology, methodology, data handling, and reporting when undertaking and reviewing stable isotope applications in archaeology*. *Rapid communications in mass spectrometry*. ISSN 1097-0231

<https://doi.org/10.1002/rcm.8044>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Calling all archaeologists: guidelines for terminology, methodology, data handling, and reporting when undertaking and reviewing stable isotope applications in archaeology

Patrick Roberts^{1*}, Ricardo Fernandes^{1,2}, Oliver E Craig³, Thomas Larsen⁴, Alexandre Lucquin³, Jillian Swift¹, Jana Zech¹

¹ Max Planck Institute for the Science of Human History, Kahlaische Str. 10, Jena D-07745, Germany

² McDonald Institute for Archaeological Research, Downing St, Cambridge CB2 3ER, UK

³ Archaeology, University of York, York YO10 5DD, UK

⁴ Leibniz-Laboratory for Isotope Research, Christian-Albrechts-Universität, Kiel D-24118, Germany

*correspondence to: Patrick Robert; email: roberts@shh.mpg.de

ABSTRACT

Stable isotope analysis has been utilized in archaeology since the 1970s, yet standardized protocols for terminology, sampling, pretreatment evaluation, calibration, quality assurance and control, data presentation, and graphical or statistical treatment still remain lacking in archaeological applications. Here, we present recommendations and requirements for each of these in the archaeological context of: bulk stable carbon and nitrogen isotope analysis of organics; bulk stable carbon and oxygen isotope analysis of carbonates; single compound stable carbon and nitrogen isotope analysis on amino acids in collagen and keratin; and single compound stable carbon and hydrogen isotope analysis on fatty acids. The protocols are based on recommendations from the Commission on Isotopic Abundances and Atomic Weights of the International Union of Pure and Applied Chemistry (IUPAC) [1] as well as an expanding geochemical and archaeological science experimental literature. We hope that this will provide a useful future reference for authors and reviewers engaging with the growing number of stable isotope applications and datasets in archaeology.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/rcm.8044

INTRODUCTION

The archaeological literature has seen an exponential increase in references to stable isotopes over the last half century due to reductions in equipment and sample processing costs, an increasing number of stable isotope laboratories and archaeological science units, and proliferating knowledge of archaeological science applications (Figure 1). Since some of the earliest archaeological applications in the late 1970s (e.g. [2]), stable isotope analysis of human tissues, and of faunal and plant remains, have been used to study past diets, ecologies, and environments. More recently, stable isotope analysis of specific compounds has emerged as a more refined tool for studying ancient diet. For example, stable isotope analysis of individual amino acids isolated from bone collagen can be used to determine the proportion of marine versus terrestrial protein in the diet [3]. The application of these methods to specific sites, periods, and regions of archaeological interest is increasingly commonplace and often led by archaeologists. While this democratization is certainly of benefit to the discipline, it comes with an enhanced responsibility on the part of archaeological users and reviewers. Commonly accepted guidelines in archaeology are essential to stimulate so-called 'big data' approaches that allow data scientists and modelers to readily compile and access published data.

On the one hand, studies of pretreatment effects on stable isotope data [4-5], the role of diagenesis in changing isotope ratios [6-7], and understandings of ecological variability [8-9] are emerging from archaeologically-focused laboratories in greater number. On the other hand, many papers appear without appropriate mention of basic measurement and calibration criteria (see [10] for an estimate of the scale of this problem), quality control, justification of pretreatment selection or sampling, full presentation of methods and datasets, and with inappropriate use of graphs and statistics. Here we seek to disseminate, in Open Access form, best practices in this regard relating to the bulk stable carbon and nitrogen isotope analysis of organics; bulk stable carbon and oxygen isotope analysis of carbonates; single compound stable carbon and nitrogen isotope analysis on amino acids (AA) isolated from collagen and hair keratin, and stable carbon and hydrogen isotope analysis of fatty acids from artefacts, bone, and sediments. While some reviews have touched on a few of these themes in the context of geochemistry as a whole [11-12], and more recently in forensics [13], we have written this article to directly increase information flow to archaeological science practitioners, students, and reviewers less familiar with these techniques. We hope this will ensure that archaeologists continue to make substantial contributions to cross-disciplinary advancements of mass spectrometry methods and applications.

A NOTE ON TERMINOLOGY

Tyler Coplen [11] has previously provided a thorough discussion of stable isotope terminology in *Rapid Communications in Mass Spectrometry*. However, many archaeological isotope publications use incorrect terminology, some of which has also been highlighted by Zachary Sharp for geochemistry in general [14]. First, the isotope ratio for samples and standards obtained from Isotope Ratio Mass Spectrometers is in the form of $^{13}\text{C}/^{12}\text{C}$, $^{15}\text{N}/^{14}\text{N}$, and $^{18}\text{O}/^{16}\text{O}$. The corresponding $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, and $\delta^{18}\text{O}$ values are not ratios *per se* but rather values produced from an equation that relates the measured isotope ratio of a (given) sample to the isotope ratio of a standard reference material (SRM):

$$\delta^i E_{sample} = \frac{\left(\frac{i_E}{j_E}\right)_{sample} - \left(\frac{i_E}{j_E}\right)_{reference}}{\left(\frac{i_E}{j_E}\right)_{reference}} \quad (1)$$

The delta notation ($\delta^i E_{sample}$), defined through Equation 1, is commonly employed in reporting stable isotope results. For a certain chemical element (E) the ratios of the abundances of the heavier (i) versus lighter (j) isotope are measured in a sample ($\left(\frac{i_E}{j_E}\right)_{sample}$)

and in a SRM with an internationally accepted δ value ($\left(\frac{i_E}{j_E}\right)_{reference}$) [15]. Equation 1 expresses the relative difference in absolute isotopic abundance ratios of a sample against the chosen SRM. Per mil notation (‰) is used as a convenient means of reporting small numerical values [16]. This is not a SI unit of measurement; it is simply a unit of comparison of the sample with a standard with an internationally recognized isotopic abundance [11,15].

It is correct to state that an archaeological material such as bone collagen has a “stable carbon isotope composition”. However, it is not possible for bone collagen to have a $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{13}\text{C}_{\text{AA}}$, $\delta^{15}\text{N}_{\text{AA}}$, or $\delta^{18}\text{O}$ ‘composition’ or ‘signature’. As Sharp [14] points out, δ values are numbers and “a composition of numbers has no meaning”. Similarly, a signature is something that is individualized and constant.

Example:

Not:

“The $\delta^{13}\text{C}$ composition/signature of the bone collagen was = -20.3 ‰”

But:

“The $\delta^{13}\text{C}$ value of the bone collagen was -20.3 ‰”

There are often also terminological issues when comparing different samples or measurements during the reporting of results or interpretation. The term ‘isotopically depleted’ is inappropriate when referring to δ values of archaeological materials. This is not only due to the vagueness of this term (as is also the case with discussing an ‘isotopic composition’ or ‘isotopic value’ above), but also because a sample of collagen, plant, or carbonate is not depleted or enriched in isotopes generally. Instead the terms depleted or enriched should be reserved for changes in the proportion of the heavy isotope (e.g. ^{13}C) of the element in a given substance or fractionation process.

Example:

Not:

“The $\delta^{18}\text{O}$ value of the cattle teeth is enriched compared with that of the pig teeth”

But:

“The cattle tooth enamel was more ^{18}O -enriched than the pig teeth”

Similarly, it is not possible to have heavier or lighter $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{13}\text{C}_{\text{AA}}$, $\delta^{15}\text{N}_{\text{AA}}$, or $\delta^{18}\text{O}$ values. This refers to the isotopes within the ratio that was measured. $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{13}\text{C}_{\text{AA}}$, $\delta^{15}\text{N}_{\text{AA}}$, and $\delta^{18}\text{O}$ can either be more negative or positive or lower or higher. Something is also not ‘isotopically negative’. The terms ‘more negative’ and ‘more positive’ in reference to delta-notated values are passable, but the use of high/higher and low/lower is more appropriate terminology for the comparison of numbers.

Example:

Not:

“Humans from later periods had heavier $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values than those from earlier periods”

But:

“Humans from later periods at the site have higher $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values than those from earlier periods”

The use of a standard appropriate terminology aids clarity in publications that can often use a multitude of different terms when referencing δ values.

ACKNOWLEDGING DIAGENESIS AND SELECTING A PRETREATMENT

The potential for variability in burial environment (e.g., humidity, pH, microbial attack, temperature and time) to alter the *in vivo* isotope values of bone collagen, bone bioapatite, dentine collagen, and tooth enamel bioapatite has been documented in archaeological applications since the 1980s [17-20]. Furthermore, the mechanisms behind these changes, especially for tooth enamel bioapatite and bone collagen, are relatively well-known [6,18,21]. While this has led to basic diagenetic checks that are applied in many archaeological science publications (Table 1), this information often goes unpublished or is not utilized even where burial environments are problematic.

Restrictions of cost, equipment availability, training and sample size availability, and the need for continued development mean that we do not recommend routine application of complex methods, particularly for bone collagen and tooth enamel bioapatite in relatively well-known environments during the Holocene and even the Late Pleistocene. However, there are materials for which regular diagenetic assessment is essential. For example, Loftus *et al.* [22] recently demonstrated a rapid way of using Fourier-Transform Infrared Spectroscopy (FTIR) to determine whether marine shell calcite had diagenetically converted to aragonite. They documented that this can occur locally, altering *in vivo* δ values, making the testing of sub-samples essential in the sequential analyses of this material.

Similarly, charred crop remains are increasingly isotopically analyzed in order to provide insights into growing conditions, climate change, and farming strategies. It has been documented that in some cases burial environments can modify pre-burial isotopic ratios of charred grains [23]. Thus, it is recommended that a preliminary study is carried out relying on

a sub-sample to compare the isotopic ratios of untreated and pre-treated charred grains [24]. In many instances pre-treatment may not be necessary, allowing for a greater sample throughput and reduction in operating costs, but this should be checked when the method is being applied to new burial environments [24].

Elemental compositions and collagen yields are regularly reported in the context of collagen preservation, and the percentage carbonate reported for tooth enamel. We would recommend more in-depth diagenetic studies over longer time periods, such as the Miocene, Pliocene, and Early Pleistocene, and in challenging taphonomic environments where these materials have been less well studied, such as those with high water throughput or potentially high trace element abundance. Fourier-Transform Infrared Spectroscopy (FTIR), Fluorescence Microscopy, and X-ray Fluorescence (XRF) or Inductively-Coupled Plasma Mass Spectrometry (ICP-MS) measurements of trace elements have all proved promising in this regard [6-7].

Diagenesis is an important problem to address. However, it is also important to note that although studies might document diagenetic alteration of isotopic ratios in certain archaeological materials and contexts (e.g. [6]), this does not mean that the same materials should not be analyzed in other settings [25-27]. This has been a particular problem in stable carbon and oxygen isotope studies of tooth enamel, which have only recently begun to expand in archaeology following a period of distrust. Opinions of the utility of different materials for stable isotope analysis need therefore to be formed on the basis of the wider, current literature rather than a few papers and preconceptions.

It is also important for researchers and reviewers to probe the impacts that different pretreatment protocols might have on a sample's material structure and δ value [4,23,28]. A plethora of different pretreatment techniques are still being used for the isotope analysis of bone collagen, bone bioapatite, tooth enamel bioapatite, tooth dentine, and plant remains, impacting the reliability of cross-comparison [e.g. 28]. In many cases, such as for tooth enamel bioapatite and bone collagen, the effects of different techniques are minimal [4,29]. However, choices of pretreatment for bone bioapatite [4,28], shell and soil carbonate [30], and crop remains [23] can be more significant in attempts to replicate the *in vivo* δ value. In these latter cases, it is not unreasonable for a reviewer to request information regarding sample pretreatment and sample preparation if this has not already been provided.

Diagenetic and pretreatment biases for compound-specific approaches remain relatively under-explored and un-reported in the context of archaeological science. For the stable isotope analysis of proteinogenic amino acids, similar measures to those employed in bulk isotope analyses of bone collagen are used to assess diagenesis (Table 1). When evaluating amino acid $\delta^{15}\text{N}$ data, there should be only minor differences between proline (Pro) and hydroxyproline (Hyp) $\delta^{15}\text{N}$ values in collagen since the nitrogen in Hyp derives from Pro [31]. Similarly, the expected slope for the $\delta^{13}\text{C}_{\text{Hyp}}$ value as a function of the $\delta^{13}\text{C}_{\text{Pro}}$ value is 1 because, after formation of immature collagen, Hyp is synthesized exclusively from Pro [32-33]. Empirical evidence from five archaeological studies of mammal bone collagen samples shows that Hyp (y) and Pro (x) $\delta^{13}\text{C}$ values usually follow a 1:1 line ($y = 0.9239x - 1.4745$, $R^2 = 93.6\%$, $n = 171$) thus lending credence to using the relationship between Hyp and Pro $\delta^{13}\text{C}$ values for data quality control (Figure 2).

With regard to pretreatment biases, it is important to homogenize samples before analysis by milling or grinding, particularly for composite materials. After the subsequent acid hydrolysis

we recommend filtering out particles from the digest using glass fiber or inert membranes. For complex or composite samples, amino acids should be separated from other compounds in the digests via ion exchange chromatography. This procedure may induce isotope effects on certain amino acids as reported previously in the literature [34]. For this reason, it is recommended that each lab investigates whether their cleaning protocols have isotope effects in order to correct for these offsets. For lipids from artefacts, bone and sediments, the normal practice is to screen the samples with gas chromatography (GC) and gas-chromatography mass spectrometry (GC/MS) to verify the compounds of interest and check for co-elution and contamination (Table 1). Contaminants, such as from plastic sample bags, should be reported and accounted for. Negative controls involving blank extractions and extractions of associated sediments and pottery should be routinely performed. Similarly, amino acid chromatograms should be checked for co-elution from non-proteinogenic compounds.

SAMPLING

Problems of diagenesis and taphonomy (e.g. funeral practices) are also the primary cause of an inevitable issue in archaeological science stable isotope applications: low or variable sample size. The nature of archaeological preservation makes it very difficult to be prescriptive in sample size necessity in different isotopic applications. However, it is important that both the variability within a given sample (e.g. a single bone) and the variability within the population under study (e.g. a sample of the same skeletal element from a series of different individuals) are taken into account when interpretations relating to δ differences between groups are being made. Furthermore, if a confident difference is to be asserted, the sample size must be large enough for the chosen statistical test to operate effectively.

In order to determine variability within a given type of sample for a particular study we recommend a pilot study measuring at least three repeat aliquots that are extracted and pretreated separately from a single sample (e.g. a bone). The measurement standard deviation for these extracts will provide a useful evaluation of δ uncertainty resulting from burial environment, pretreatment, and natural heterogeneity in a sample. It is especially important to do this when an archaeological material is being analyzed for the first time in a particular burial environment. Suitable statistical testing, that meets sample size requirements, will address the problem of whether variability within a given population is greater than the variability between populations. Where only small sample sizes are available for a context, care should be taken when comparing the data with that from other contexts.

CALIBRATING ISOTOPE DATA

Perhaps one of the most significant omissions in stable isotope applications in archaeology is the proper reporting of the stable isotope abundance measurements of standards and reference materials used in the analysis, together with their adopted values and uncertainties [15]. Szpak *et al* [10] recently found that a large majority of archaeological isotope studies do not provide adequate information regarding isotope measurement calibration and analytical uncertainty. Isotope Ratio Mass Spectrometers generate raw δ values using a single-point calibration relative to a laboratory working gas, the composition of which is arbitrary and can change over time [35]. The working and sample gases are introduced into the isotope ratio mass spectrometer using two different systems. This means that the working gas does not undergo the same physical and chemical processes as those affecting the sample gas, thus

voiding the important principle of ‘Identical Treatment’ that must hold for calibration materials [36].

To make the raw δ values meaningful for cross-study comparison and interpretation, it is essential that the measurements be calibrated to internationally accepted δ scales, using standard reference materials (SRMs) with known isotopic ratios interspersed among samples in each analytical sessions [36-37]. Following Coleman and Meier-Augenstein [12] and Coplen [11], and established IUPAC requirements for data calibration in isotope geochemistry more generally, authors should:

- i) Express δ values relative to current international measurement standards. References to these standards should also avoid using redundant standards such as SMOW (it is now VSMOW) or PDB (it is now VPDB) [11,15];
- ii) Where a second international measurement standard is used to define the range of the δ scale, for example SLAP (SLAP2) water for $\delta^2\text{H}$ and $\delta^{18}\text{O}$ measurements, normalize δ values using both standards and state this clearly;
- iii) Report δ values of SRMs with known, internationally-accepted isotopic ratios (previously calibrated to VPDB, AIR, or VSMOW); at least two calibration SRMs, such as USGS40 or USGS41a [38-39] in the case of $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ values, should be used to anchor the raw sample isotopic ratios at the high and low ends of the isotopic range, enabling shifting and stretching onto an international scale via a ‘two-point’ calibration [36,40-41] (**Figure 2**).

Most archaeological science applications of isotope analysis meet requirements i) and ii) (where required). However, the names and δ values of calibration SRMs, and the process of normalization used for iii), are often lacking (see [10]). For example, sometimes only one calibration SRM is used, making the appropriate stretching of the data impossible [10,36]. Inadequate reporting may be the result of archaeologists sending their samples away to a commercial chemistry lab that provides ‘final’ values and no insight into the process of measurement, normalization, or the standards that were run alongside the samples. However, the necessary information must be requested and verified from commercial labs prior to the sending of samples. Notably, laboratories with ISO/IEC 17025 [42] “General Requirements for the Competence of Testing and Calibration Laboratories” (see clauses 4.13.2.1 and 5.10) accreditation must provided the client with all records and documentation pertaining to the analysis of the client’s samples including raw data, data calibration, and data evaluation.

Archaeological science authors and reviewers should pay close attention to the selection and reporting of the utilized calibration SRMs chosen to ‘bracket’ the expected δ range of the samples. If internal SRMs are used, it is necessary to specify their accepted values, and how these values were obtained relative to internationally-accepted standards [12,36-37]. Calibration SRMs should also be ‘matrix-matched’, as far as possible, to the samples under study. In other words, they should have a similar elemental composition and structure to the samples. The number of calibration SRMs used in a study should also be stated. Szpak *et al* [10] recommend that at least 10% of the total analyses should be calibration standards.

For compound-specific isotope analysis, check standards should be composed of a mixture of compounds of interest (e.g. palmitic and stearic acid methyl esters) and of other stable compounds eluting closely (i.e. eicosane), corresponding to the expected range of isotopic value (see [15]). SRMs with known isotope ratios that bracket the expected isotope range

should also be run for calibration purposes (e.g. [43]). Ideally, multiple compounds found within the sample should be used as SRMs. Finally, internal reference compounds can also be included in each analytical run to help with calibration or precision estimation.

QUALITY CONTROL AND QUALITY ASSURANCE

It is also important to document systematic errors, such as instrumental drift, related to the analytical accuracy of a system, as well as the relationship between the δ value measured and the true δ value of the sample. This should be assessed using check SRMs that are not used for calibration. These are treated as having ‘unknown’ values just like the samples (Figure 3). As with the calibration SRMs these should be matrix-matched to the samples. While Szpak *et al* [10] recommend that 10% of the total analyses should be ‘check’ standards and placed at regular intervals within a session, the exact proportion will depend on the accuracy required for a given application. The δ values of these check SRMs (mean $\pm 1\sigma$) for each measurement session should be reported, even if only in supplementary information [10]. We recommend that these records of accuracy be kept up-to-date and used as an ongoing measure of laboratory performance that can even be publicly displayed.

We would recommend using more than one check SRM to test whether this magnitude is similar across slightly different materials. Furthermore, tests for thresholds of linearity effects (whereby IRMS-based techniques return different δ values depending on the mass of sample) should be made for each new material or SRM run on an IRMS system [13]. This all helps to assess the discrepancy between measured δ values and ‘true’ sample isotope composition within a study, and ensure that interpretations of the data are balanced accordingly. If this systematic error is related to instrumental drift, for example in the form of temperature or humidity changes affecting the IRMS system, and is shown to be continuous across the entire analytical session, drift corrections can be made using check SRMs placed at regular intervals throughout the session [10,36]. If these corrections are made details should be provided in the text or supplementary information.

Precision refers to the repeatability or reproducibility of measurements and random errors rather than systematic errors. This can be assessed for a given study using either check or calibration SRMs, as well as repeated measurement of samples. We recommend reporting duplicate measurements of the same aliquots of archaeological samples where possible. The precision of repeated SRMs or samples should be reported as the mean $\pm 1\sigma$. As Szpak *et al* [10] note, it is essential that matrix-matched SRMs are used when precision estimates are used to validate interpretation. For example, a precision of $\delta^{13}\text{C}$ measurements of $\pm 0.01\text{‰}$ may be attainable with a pure carbonate reference material. However, this is almost meaningless when provided for a study focused on tooth enamel (c. 3-10 % carbonate) sample $\delta^{13}\text{C}$ values, on which a precision of $\pm 0.01\text{‰}$ is impossible on even the most refined systems as a result of natural inhomogeneity. It is often best to use in-house reference materials, or repeat sample aliquots, for a more realistic assessment of precision in archaeological science applications.

Precision measured through repeat analyses is not necessarily the best overall measure of uncertainty in sample δ values, however [44]. A better measure is obtained by propagating errors across the whole process of sample selection, preparation, measurement, and normalization. This can be done using a ‘bottom up’ approach (e.g. [44-45]), or via a ‘top down’ approach [13]. Together, these criteria enable archaeological science-focused

laboratories to validate their pretreatment and instrumental methods more widely, the importance of which has recently been emphasized in forensic studies [13]. Given the growing importance of data compatibility between laboratories, and over long time scales for archaeological interpretation between populations and sites, it is essential that each study and laboratory demonstrate that its methods of sample selection, pretreatment/extraction, measurement, and calibration meet accepted criteria.

In the case of a given study, we recommend the analysis of an SRM with a known isotope δ value and treating it alongside samples to establish the degree to which treatment and measurement causes sample δ values to deviate from their 'true' value. This is particularly important in single-compound approaches. For example, in Figure 4 we show the results of $\delta^{13}\text{C}$ analyses of fatty acids repeatedly extracted from homogenized pulverized pottery sherds to assess within-lab reproducibility. The uncertainty achieved through this exercise ($\pm 0.6\text{‰}$ 1σ) was greater than obtained from repeated measurements of a single extract (typically $\pm 0.3\text{‰}$ 1σ), emphasizing the importance of propagating errors beyond measurement precision and the importance of extracting a SRM 'sample' during extraction. It is also particularly important when reporting this procedure to account for the addition of C and H during the derivatization process. As a result, the isotope ratios of reference compounds should be reported before and after derivatization with measurement uncertainties.

In the case of inter-laboratory validation, inter-laboratory comparisons should be designed to take into account the points raised above relating to uncertainty arising from pretreatment, calibration, and standard use, in order to enable laboratories to critically identify the largest sources of errors during sample cleaning, extraction, and instrument calibration. Until now this has been rarely undertaken, or has been published without suitable determinations of analytical uncertainty [e.g. 46], particularly in the case of single compound analysis. Swapping and repeated extraction of samples between laboratories is therefore recommended in order to document the precision of a given archaeological measurement. In addition, for amino acid and fatty acid stable isotope analysis we propose an anonymous, blind inter-laboratory study of the δ values of two calibrated mixtures known by the coordinator. For AA analysis it would also be useful to include a matrix sample of bone collagen already measured for bulk $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values.

LEAVING A FULL PICTURE: METHOD AND DATA AVAILABILITY

Another potential problem in archaeological science applications of isotope methods is the omission of the full methodological details and datasets in publications. While most papers focused on bulk isotope measurements also provide method information that could be practically followed by any researcher seeking to emulate the study, these data are often missing when isotope data is provided as supplementary information to radiocarbon dating. $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values are often presented for bone collagen samples that have been radiocarbon dated, but information relating to their preparation, normalization, and comparison with reference materials should also be provided so that it can be determined whether these results are useful for subsequent stable isotope comparisons.

Full data reporting is also essential in publications. This is, in part, because it is useful to have datasets available for comparison with existing literature, particularly given the rise of 'big data' approaches also within archaeology [47]. One of the most common oversights in this regard is the production of a mean and standard deviation plot for a given human group or faunal group as a useful, simplified graphical representation of one's data 'average' and

'spread'. Nevertheless, in some cases the corresponding data table will also only report the mean and standard deviation for that group without an additional table that includes the individual measurements or samples that comprise that group. It should be said that this is relatively rare, although often things such as associated fauna can be treated as "background data" and not reported.

A paper relying on data that is not fully provided should not make it past review. Similarly, where modern samples have been incorporated as a useful analogy, collection locations, growing conditions (for plants), and local climate or environmental data should be reported. The date of sample collection is particularly important so that the appropriate atmospheric or oceanic *Suess effect* can be applied [48]. If *Suess effect* corrections are made these need to be explicit and include appropriate uncertainties. In general, all calculation and correction stages applied to raw data should be documented in a publication and its associated tables for transparent evaluation. Where groups are compared, full archaeological context information, and the logic behind such groupings, should be provided.

Transparent methodology is also often missing in many compound-specific isotope papers, hindering replication and application beyond select groups. Given the rapid development of these specialist approaches and the ample space available in supplementary information sections this should be remedied. There is also a growing problem of data reporting in more recently-applied compound-specific approaches. Here studies must provide the final δ value for each sample (mean and analytical standard deviation) across replicate injections or include chromatograms of representative samples.

One potential solution to these problems is a central repository for archaeological isotopic data, and calls for this have been put forward [49-51]. Central data repositories, such as GenBank, exist for the field of archaeogenetics and proteomics [52-53]. However, it is also important to notice that applications of stable isotope data are extremely diverse and different research fields have specific data requirements. Partnership-based initiatives, such as IsoMemo, attempt to bring together multiple repositories of stable isotope data from archaeology, ecology, and environmental sciences [54]. The goals of the initiative are to coordinate data collection efforts, sharing and centralization of data, creating tools (e.g. user friendly graphical interfaces) for facilitating data access, building interdisciplinary projects, and establishing common data standards.

The latter includes, for instance, adoption of common terminologies and the assignment of unique codes to stable isotope labs and for reported measurements. Among IsoMemo partners are stable and radiogenic databases devoted to the storing of archaeological data from varied regions and time periods (e.g. [55-57]) although there is still a general lack of awareness of their availability. While data collection requires the overcoming of political and data-retention concerns, it seems reasonable that for archaeological science isotope papers to be published all reported data should either be placed in a similar repository or made fully available in a table within or attached to the publication. Journals should also consider changing their requirements for table formatting in some cases, as .xls or .csv format makes for greater ease of re-use of the data than PDF forms.

GRAPHS AND SCALING

The development of open access graphical and statistical analysis programs makes sophisticated plotting tools widely available. However, there are several caveats to keep in

mind when selecting appropriate methods for data display. First, a graph should be chosen that fits the data at hand: for normally distributed data a mean and standard deviation plot are appropriate when summarizing the dataset. Here the sigma of the standard deviation should be specified, as well as the range of the sample distribution that it describes (i.e. 68 % or 95 %) when interpretations are made. For non-normally distributed data, box plots would better summarize the data. Similarly, choices of data groupings (e.g. species, period, or stratigraphic layer) should be fully justified and considered in the text. For example, plotting humans from one period/site and fauna from another period/site may be the only option left to the researchers, but this should be explained clearly.

There can be interpretive issues with the use of axes scales (in terms of both scale length and scale divisions) in bi-plots of two isotope parameters, the most common being $\delta^{13}\text{C}$ values plotted against $\delta^{15}\text{N}$ values for bone collagen, dentine collagen, or crop remains. Variation in ‰ in a given human or ecological system may be greater for $\delta^{15}\text{N}$ than $\delta^{13}\text{C}$ values, and *vice versa*. $\delta^{15}\text{N}$, $\delta^{13}\text{C}$, and $\delta^{18}\text{O}$ values will also vary significantly between sites, periods, and ecological systems. These factors may lead to decisions to adopt different scale lengths and divisions for different isotopic parameters or archaeological periods and sites (Figure 5A). However, if this is done, researchers must be wary of statements such as:

“**Figure 5A** shows that the variation in $\delta^{15}\text{N}$ values is greater or the same as for $\delta^{13}\text{C}$ values”

Or

“**Figure 5A** shows large variation in $\delta^{13}\text{C}$ values”

Similarly, in Figure 5B, two sequential plots of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values from tooth enamel can look very different depending on the scale, leading to comments such as:

“Individual 1 shows larger variation in $\delta^{18}\text{O}$ than $\delta^{13}\text{C}$ values throughout the period of enamel formation”

In both cases, these comments are either inaccurate or require legitimate comparison with another sample group or δ value on the same scale (of the same total length in ‰ and with the same ‰ divisions). In practice, data comparisons and evaluations are always best made using statistical tests. One must also ask whether measurement uncertainty, isotopic heterogeneity of the sample matrix, and physiological effects outweigh the archaeological significance of the variation demonstrated [58-59]. This is particularly true when scales drop close to or below the measured ‰ uncertainty for $\delta^{15}\text{N}$, $\delta^{13}\text{C}$, and $\delta^{18}\text{O}$ values within a given study or laboratory (see above).

It is important that all data points are displayed in graphs where possible, even if only in the supplementary information, and not just a summary representation (e.g. a boxplot or mean and standard deviation plot). This is a particularly significant problem with the growing use of ellipses to summarize archaeological δ parameters for a particular data group. These can be drawn using different methods and principles that should be clearly outlined. Furthermore, the confidence level of the ellipse should be stated. In single compound analysis of fatty acids in archaeological pottery, for example, the archaeological data are commonly plotted against 68 % confidence ellipses of modern foods corrected for the *Suess effect* (e.g. [60]). As can be

seen in Figure 6, a confidence ellipse of 50 % and one of 95 % can have very different relationships to the real spread of the data. This is particularly sensitive for the relatively small sample sizes that are frequent in archaeological research.

STATISTICS: THE FINAL FRONTIER

In terms of basic reporting, summary statistics such as mean or median (or both), standard deviation or interquartile range, and range should be provided where appropriate for any group of interest. Standard deviation and standard error should be used depending on the research question or point being made. Standard deviation will demonstrate how much variation there is among individual observations in a given sample, while the standard error shows how good the estimation of the mean is [61]. As noted above, if there is concern about the homogeneity of a given sample it is also often useful to provide the δ values from multiple measurements on separate sample aliquots and provide the summary statistics for that comparison.

When comparing the δ values of different groups it should always be remembered that the phrase:

“There is a significant difference between the $\delta^{15}\text{N}$ values of period x and period y”

Is only valid if followed up by a relevant statistical test, its parameters, and full results, and would be even more appropriately phrased as:

“The null hypothesis that there are no differences in $\delta^{15}\text{N}$ values of period x and period y can be rejected”

An obvious difference in periods x and y might be observed from a graph of box plots or means and standard deviations where there is no overlap and can be stated as such. However, a “significant” difference is a term limited to statistical analysis and must be used accordingly.

It is also important that the appropriateness of the chosen statistical test is discussed [62]. It should be clear that the data has been evaluated, either through a histogram or normality test. From this point the appropriate choice of parametric (normally distributed data) and non-parametric test (non-normally distributed data) should be made (see Figure 7 for a comparison of basic tests used to compare stable isotope data from groups of archaeological samples). Many archaeological scientists now utilize complex statistical models and tests. In these cases, every effort should be made to walk a non-specialist reader (and reviewer) through the building of the model, including the reasoning as to why certain parameters have been included or not. This not only facilitates future replication but also enables more adequate evaluation of the methodology and the validity of the assumptions used.

When reporting the results of the statistical test, full information for the test chosen should be provided (Figure 7) plus the chosen significance level (typically set at or below 5 %). This then defines the boundary of a significant or non-significant result. It is therefore not worthwhile to compare data as being more or less close to being significantly different, as that would ignore this basic statistical concept. Statistical test result tables should be fully provided where appropriate, in supplementary information if necessary. It should also be

clear how the data presented was “plugged into” a given method so that the test can be replicated easily.

As discussed in the ‘Sampling’ section above, a statistical method is only as good as the data that it is used to analyze. This is obvious from the development, and occasional misuse, of Bayesian mixing models that have been applied to model food source contributions to human and animal diets based on stable isotope ratios [63]. These models quantify, as probability distributions, the contributions from food sources by considering multiple parameters (e.g. consumer and food isotopic signals, diet-to-consumer isotopic offsets, food nutrient concentrations, complex dietary routing mechanisms) and respective uncertainties. Here, it is essential to adequately characterize the ‘true’ uncertainty of ‘consumer’ and ‘food’ isotope ratios and also utilize meaningful groups or sources.

For example, a limited number of isotopic proxies, isotopic similarities among foodstuffs, and the impossibility of identifying all potential source contributions mean that identifying inputs from a particular taxon is almost impossible. Generic groupings (e.g. terrestrial plants, terrestrial animals, freshwater fish, marine fish) with associated conservative uncertainties, often around one order of magnitude larger than measurement uncertainties, are more appropriate [64]. Diet-to-consumer isotopic offsets and routing mechanisms are quantified through controlled feeding experiments on humans or omnivorous mammals. The complexity of dietary routing mechanisms means that mixing models that account for separate signal contributions from multiple food nutrients towards consumer tissues are best employed in certain contexts [65-66].

Evaluation of compound-specific data rests increasingly on classification methods and multivariate statistical approaches that are not yet widely applied in archaeology. Ecologists typically use these methods to emphasize variation and highlight strong patterns in multivariate isotopic datasets by either applying principal component analysis (PCA) or linear discriminant function analysis (LDA). While both methods are linear transformation techniques, PCA treats all data unsupervised to maximize variance among individual samples. In contrast, LDA is a supervised approach that maximizes variance among predefined groups but minimizes the variance within these groups. Hence, LDA should only be applied when additional data or observations (e.g. marine vs terrestrial) support an *a-priori* classification. With PCA, one has to assess whether isotope baseline information should be factored out. In that case, center the isotope value of each compound to the mean isotope values across all compounds for each sample.

Regardless of the chosen statistical method, a central goal is to reduce the number of dimensions and identify the most informative variables. For linear transformation techniques such as PCA and LDA, we recommend reporting the significant differences between classes by employing MANOVA analyses or similar. In LDA, we suggest using leave-one-out cross-validation to predict the probability of class membership of training data samples. To enhance transparency, it is good practice to report the statistical results either by describing the main outcome in the text or by including the statistical output in the supplementary information. Care should be taken in applying multivariate statistics when the number of variables considered is large relative to the number of samples. For discriminant analysis we recommend, as per Carter, [67], that the total number of observations must be significantly greater than the number of independent variables (>5:1), and that the number of observations in the smallest group must be greater than the number of independent variables.

CONCLUSION

This review is not meant as criticism directed towards any particular group of researchers, and indeed the authors have certainly been guilty of some of the above-listed failings at some point. In the interest of best scientific practice, we have attempted to summarize the basic requirements of terminology, sampling, measurement, reporting, display, and analysis in the presentation and publication of isotope data. Much of this is dictated by IUPAC and influential members within geochemistry, mass spectrometry, and archaeological science itself, have leveled similar criticisms. However, as an increasing number of students and researchers from different academic backgrounds enter archaeological science it is beneficial to circulate these widely within an archaeology-focused format that is as accessible as possible to this readership (including by making this article Open Access).

We have focused on bulk stable carbon and nitrogen isotope analysis of organics; bulk stable carbon and oxygen isotope analysis of carbonates; and single compound stable carbon and nitrogen isotope analysis on amino acids in collagen and keratin; and single compound stable carbon and hydrogen isotope analysis on fatty acids. However, with developments in archaeological science applications, the same principles of terminology, sampling, calibration, quality assurance and control, graphical representation, and statistical analysis will apply to the currently rarer, or more experimental, applications of bulk organic sulfur, hydrogen, and oxygen stable isotope analyses and stable hydrogen analysis of amino acids.

We request that reviewers agreeing to evaluate publications involving isotope analysis familiarize themselves with these requirements, making the appropriate critique and suggestions where necessary, so as to raise the standard of science, data production, and data availability in this ever-expanding and advancing field.

Acknowledgments

We would like to thank the Max Planck Society for the funding to write this paper and make it available to the archaeological community as Open Access. OEC and AL would like to thank the Arts and Humanities Research Council (AH/L0069X/1) for their support. We would also like to thank four anonymous Reviewers for their thorough and constructive comments on an earlier draft of this manuscript.

REFERENCES

- [1] IUPAC. *Nomenclature of Inorganic Chemistry. IUPAC Recommendations 2005*. Prepared for publication by Connelley NG, Damhus T, Hartshorn RM, Hutton AT. Cambridge, UK: RSC Publishing; 2005.
- [2] Van der Merwe NJ, Vogel JC. ^{13}C content of human collagen as a measure of prehistoric diet in woodland North America. *Nature*. 1978;276:815-816.
- [3] Corr LT, Sealy JC, Horton MC, Evershed RP. A novel marine dietary indicator utilizing compound-specific bone collagen amino acid values of ancient humans. *J Arch Sci*. 2005;32(3):321-330.
- [4] Pellegrini M, Snoeck C. Comparing bioapatite carbonate pre-treatments for isotopic measurements: Part 2 – Impact on carbon and oxygen isotope compositions. *Chem Geol*. 2016;420:88-96.
- [5] Guiry EJ, Szpak P, Richards MP. Effects of lipid extraction and ultrafiltration on stable carbon and nitrogen isotopic compositions of fish bone collagen. *Rapid Commun Mass Spectrom*. 2016;30:1591-1600.
- [6] Schoeninger MJ, Hallin K, Reeser H, Valley JW, Fournelle J. Isotopic alteration of mammalian tooth enamel. *Int J Osteoarchaeol*. 2003;13:11-19.
- [7] Fernandes R, Hüls H, Nadeau MJ, et al. Assessing screening criteria for the radiocarbon dating of bone mineral. *Nuc Instrum Methods Phys Res B*. 2013;294:226-232.
- [8] O'Connell TC, Hedges REM, Healey MA, Simpson AHR. Isotopic composition of hair, nail and bone: Modern analyses. *J Arch Sci*. 2001;28:1247-1255.
- [9] Szpak P, White CD, Longstaffe FJ, Millaire J-F, Vásquez Sánchez VF. Carbon and nitrogen isotopic survey of northern Peruvian plants: baselines for paleodietary and paleoecological studies. *PLoS ONE*, 2013;8:e53763.
- [10] Szpak P, Metcalfe JZ, Macdonald RA. Best practices for calibrating and reporting stable isotope measurements in archaeology. *J Archaeol Sci Reports*. 2017;13:609-616.
- [11] Coplen TB. Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results. *Rapid Commun Mass Spectrom*. 2011;25(17):2538-2560.
- [12] Coleman M, Meier-Augenstein W. Letter to the Editor “Ignoring IUPAC guidelines for measurement and reporting of stable isotope abundance values affects us all”. *Rapid Commun Mass Spectrom*. 2014;28(17):1953-1955.
- [13] Dunn PJH, Salouros H, Carter JF, Doyle SP. Letter to the Editor “Forensic application of stable isotope delta values: Proposed minimum requirements for method validation”. *Rapid Commun Mass Spectrom*. 2017;31:1476-1480.
- [14] Sharp Z. *Principles of stable isotope geochemistry*. London, UK: Prentice Hall; 2016. Open access at: http://digitalrepository.unm.edu/unm_oer/1/
- [15] Brand W, Coplen TB, Vogl J, Rosner M, Prohaska T. Assessment of international reference materials for isotope-ratio analysis (IUPAC Technical Report). *Pure Appl Chem*. 2014;86(3):425-467.
- [16] Urey HC. Oxygen isotopes in Nature and in the Laboratory. *Science*. 1948;108:489-496.

- [17] Hare PE. Organic geochemistry of bones, and its relation to the survival of bone in the natural environment. In: Behrensmeyer AK, Hill AP. Eds; *Fossils in the making*. Chicago, US: The University of Chicago Press, Chicago, 1980:208-219.
- [18] Collins MJ, Nielsen-Marsh CM, Hiller J, et al. The survival of organic matter in bone: a review. *Archaeometry*. 2002;44(3):383-394.
- [19] Lee-Thorp J.A. Two decades of progress towards understanding fossilization processes and isotopic signals in calcified tissue minerals. *Archaeometry*. 2002;44(3):435-446.
- [20] Berna F, Matthews A, Weiner S. Solubilities of bone mineral from archaeological sites: the recrystallization window. *J Arch Sci*. 2004;31(7):867-882.
- [21] Lee-Thorp JA, van der Merwe NJ. Carbon isotope analysis of fossil bone apatite. *S Afr J Sci*. 1987;83(11):712-715.
- [22] Loftus E, Rogers K, Lee-Thorp J. A simple method to establish calcite:aragonite ratios in archaeological mollusk shells. *J Quat Sci* 2015; 30(8): 731-735.
- [23] Vaiglova P, Snoeck C, Nitsch E, Bogaard A, Lee-Thorp J. Impact of contamination and pre-treatment on stable carbon and nitrogen isotopic composition of charred plant remains. *Rapid Commun Mass Spectrom* 2014; 28(23): 2497-2510.
- [24] Brinkkemper O, Braadbaart F, van Os B, van Hoesel A, van Brussel AAT, Fernandes R. Effectiveness of different pre-treatments in recovering pre-burial isotopic ratios of charred plants. *Rapid Commun Mass Spectrom*. In press.
- [25] Stuart-Williams HLQ, Schwarcz HP, White CD, Spence MW. The isotopic composition and diagenesis of human bone from Teotihuacan and Oaxaca, Mexico. *Palaeogeog Palaeoclim Palaeoecol*. 1996;126:1-14.
- [26] Lee-Thorp JA, Sponheimer M. Three case studies used to reassess the reliability of fossil bone and enamel isotope signals for palaeodietary studies. *J Anth Archaeol*. 2003;22(3):208-216.
- [27] Roberts P, Perera N, Wedage O, et al. Fruits of the forest: Human stable isotope ecology and rainforest adaptations in Late Pleistocene and Holocene (~36 to 3 ka) Sri Lanka. *J Hum Evol*. 2017;106:102-118.
- [28] Crowley BE, Wheatley PV. To bleach or not to bleach? Comparing treatment methods for isolating biogenic carbonate. *Chem Geol*. 2014 381:234-242.
- [29] Sealy J, Johnson M, Richards M, Nehlich O. Comparison of two methods of extracting bone collagen for stable and nitrogen isotope analysis: comparing whole bone demineralization with gelatinization and ultrafiltration. *J Arch Sci*. 2014;47:64-69.
- [30] Schöne BR, Schmitt K, Maus M. Effects of sample pretreatment and external contamination on bivalve shell and Carrara marble $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ signatures. *Palaeogeog Palaeoclim Palaeoecol*. 2016;484:22-32.
- [31] O'Connell TC, Collins MJ. Comment on "Ecological niche of Neanderthals from Spy Cave revealed by nitrogen isotopes of individual amino acids in collagen". *J Hum Evol*. 2017;93:82-90.
- [32] Rennie MJ, Meier-Augenstein W, Watt PW, Patel A, Begley IS, Scrimgeour CM. Use of continuous-flow combustion MS in studies of human metabolism. *Biochem Soc Trans*. 1996;24(3):927-932.

- [33] Babraj J, Cuthbertson DJ, Rickhuss P, et al. Sequential extracts of human bone show differing collagen synthetic rates. *Biochem Soc Trans.* 2002;30:61-65.
- [34] Larsen T, Bach LT, Salvatelli R, et al. Assessing the potential of amino acid ^{13}C patterns as a carbon source tracer in marine sediments: effects of algal growth conditions and sedimentary diagenesis. *Biogeosciences.* 2015;12:4979-4992.
- [35] Paul D, Skrzypek G, Forizas I. Normalization of measured stable isotopic compositions to isotope reference scales – a review. *Rapid Commun Mass Spectrom.* 2007;21(18):3006-3014.
- [36] JCarter JF, Fry B. Ensuring the reliability of stable isotope ratio data – beyond the principle of identical treatment. *Anal Bioanal Chem.* 2013;405:2799-2814.
- [37] Werner RA, Brand WA. Referencing strategies and techniques in stable isotope ratio analysis. *Rapid Commun Mass Spectrom.* 2001;15(7):501-519.
- [38] Qi H, Coplen TB, Geilman H, Brand WA, Böhlke JK. Two new organic reference materials for $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ measurements and a new value for the $\delta^{13}\text{C}$ of NBS 22 oil. *Rapid Commun Mass Spectrom.* 2003;17(22):2483-2487.
- [39] Qi H, Coplen TB, Mroczkowski SJ, et al. A new organic reference material, L-glutamic acid, USGS41a for $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ measurements – a replacement for USGS41. *Rapid Commun Mass Spectrom.* 2016;30(7):859-866.
- [40] Coplen TB. Reporting of stable hydrogen, carbon, and oxygen isotopic abundances. *Pure Appl Chem.* 1994;66:273-276.
- [41] Coplen TB. New guidelines for reporting stable hydrogen, carbon, and oxygen isotope ratio data. *Geochim Cosmochim Acta.* 1996;60(17):3359-3360.
- [42] ISO/IEC 17025 (E). General requirements for the competence of testing and calibration laboratories (second edition 2005-05-15). <https://www.iso.org/standard/39883.html>
- [43] Schimmelmann A, Qi H, Coplen TB, et al. New organic reference materials for hydrogen, carbon, and nitrogen stable isotope-ratio measurements: caffeine, n-alkanes, fatty acid methyl esters, glycines, L-valines, polyethylenes, and oils. *Anal Chem.* 2016;88(8):4294-4302.
- [44] Dunn PJH, Hair L, Malinovsky D, Goenaga-Infante H. Simple spreadsheet templates for the determination of the measurement uncertainty of stable isotope ratio delta values. *Rapid Commun Mass Spectrom.* 2015;29(22):2184-2186.
- [45] Kragten J. Tutorial review. Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique. *Analyst.* 1994;119:2161-2165.
- [46] Pestle WJ, Crowley BE, Weirauch MT. Quantifying inter-laboratory variability in stable isotope analysis of ancient skeletal remains. *PLoS ONE.* 2014;9:e102844.
- [47] Killick D. The awkward adolescence of archaeological science. *J Arch Sci.* 2015;56:242-247.
- [48] Keeling CD. The Suess effect: ^{13}C - ^{14}C interrelations. *Environ Int.* 1979;2(4-6):229-300.
- [49] Stansbie D, Mallet S. “Big, bad (?) data. New approaches to the study of food, identity and landscape in early medieval England.” *Medieval Settlement Research.* 2015;30:16-24.

- [50] Pauli JN, Newsome SD, Cook JA, et al. Opinion: Why we need a centralized repository for isotopic data. *Proc Natl Acad Sci USA*. 2017;114(12):2997-3001.
- [51] Grupe G, Hölzl S, Mayr C, Söllner F. The Concept of Isotopic Landscapes: Modern Ecogeochemistry versus Bioarchaeology. In: Grupe G, Grigat A, McGlynn GC, eds. *Across the Alps in Prehistory*. The Netherlands: Springer International Publishing, 2017: 27-48.
- [52] Benson DA, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res*, 2013,41,D36-D42.
- [53] Genbank. How to submit data to GenBank 2017. <https://www.ncbi.nlm.nih.gov/genbank/submit/> Accessed October 10, 2017.
- [54] Fernandes R, Larsen T, Knipper C, Feng F, Wang Y. IsoMemo.com: a database of isotopic data for ecology, archaeology, and environmental science 2017. <http://www.isomemo.com> Accessed October 10, 2017.
- [55] Van Strydonck M, De Roock E. Royal Institute for Cultural Heritage Web-Based Radiocarbon Database. *Radiocarbon*. 2011;53(2):367-370.
- [56] Etu-Sihvola H, Arppe L, Junno A, Uusitalo J, Oinonen M. Dietary isotopic baseline for the ancient north 2017. www.oasisnorth.org/diana, 2017. Accessed October 10, 2017.
- [57] Salesse K, Fernandes R, de Rochefort X, Bruzek J, Castex D, Dufour E. IsoArch.eu: An open-access and collaborative isotope database for bioarchaeological samples from the Graeco-Roman world and its margins. *J Arch Sci Reports*. 2017; In Press: <https://doi.org/10.1016/j.jasrep.2017.07.030>
- [58] DeNiro MJ, Schoeniger MJ. Stable carbon and nitrogen isotope ratios of bone collagen: variations within individuals, between sexes, and within populations raised on monotonous diets. *J Arch Sci*. 1983;10(3):199-203.
- [59] Waters- Rist AL, Bazaliiskii VI, Weber AW, Katzenberg MA. Infant and child diet in Neolithic hunter-fisher-gatherers from cis-baikal, Siberia: Intra-long bone stable nitrogen and carbon isotope ratios. *Am J Phys Anth*. 2011;146(2):225-241.
- [60] Copley MS, Berstan R, Dudd SN, et al. Direct chemical evidence for widespread dairying in prehistoric Britain, *Proc Natl Acad Sci USA*. 2003;100:1524-1529.
- [61] Baxter M. *Statistics in Archaeology*. London, UK: John Wiley and Sons; 2003.
- [62] Biochemia-medica. Comparing groups statistical differences, how to choose the right statistical test. <http://www.biochemia-medica.com/content/comparing-groups-statistical-differences-how-choose-right-statistical-test> Accessed October 10, 2017.
- [63] Fernandes R, Millard AR, Brabec M, Nadeau MJ, Grootes P. Food reconstruction using isotopic transferred signals (FRUITS): a Bayesian model for diet reconstruction. *PLoS One*. 2014;9:e87436.
- [64] Fernandes R, Grootes R, Nadeau MJ, Nehlich O. Quantitative diet reconstruction of a Neolithic population using a Bayesian mixing model (FRUITS): the case study of Ostorf (Germany). *Am J Phys Anth*. 2015;158:325-340.
- [65] Newsome SD, Wolf N, Peters J, Fogel ML. Amino acid $\delta^{13}\text{C}$ analysis shows flexibility in the routing of dietary protein and lipids to the tissue of an omnivore. *Integr Comp Biol*. 2014;54(5):890-902.

- [66] Fernandes R. A simple (r) model to predict the source of dietary carbon in individual consumers. *Archaeometry* 2016; 58(3): 500-512.
- [67] Carter JF, Yates HS, Tinggi U. Isotopic and elemental composition of roasted coffee as a guide to authenticity and origins. *J Agric Food Chem.* 2015;63(24):5771-5779.
- [68] Colonese AC, Collins M, Lucquin A, et al. Long-term resilience of Late Holocene coastal subsistence system in southeastern South America. *PLoS ONE.* 2014;9:e93854.
- [69] Honch NV, McCullagh JS, Hedges RE. Variation of bone collagen amino acid $\delta^{13}\text{C}$ values in archaeological humans and fauna with different dietary regimes: Developing frameworks of dietary discrimination. *Am J Phys Anth.* 2012;148(4):495-511.
- [70] Raghavan M, McCullagh JSO, Lynnerup N, Hedges REM. Amino acid $\delta^{13}\text{C}$ analysis of hair proteins and bone collagen using liquid chromatography/isotope ratio mass spectrometry: paleodietary implications from intra-individual comparisons. *Rapid Commun Mass Spectrom.* 2010;24(5):541-548.
- [71] Webb EC, Honch NV, Dunn PJH, Eriksson G, Lidén K, Evershed RP. Compound-specific amino acid isotopic proxies for detecting freshwater resource consumption. *J Arch Sci.* 2015;63:104-114.
- [72] Webb EC, Honch NV, Dunn PJH, et al. Compound-specific amino acid isotopic proxies for distinguishing between terrestrial and aquatic resource consumption. *Archaeol Anthropol Sci.* 2016; 1-18. <https://doi.org/10.1007/s12520-015-0309-5>
- [73] Craig O, Saul H, Lucquin A, et al. Earliest evidence for the use of pottery. *Nature.* 2013;496:351-354.
- [74] R Core Team. R: A language and environment for statistical computing 2013. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: <http://www.R-project.org/> Accessed October 10, 2017.
- [75] McCrum-Gardner E. Which is the correct statistical test to use? *Br J Oral Maxillofac Surg.* 2008;46:38-41.
- [76] Marusteri M, Bacarea V. Comparing groups for statistical differences: how to choose the right statistical test? *Biochem Med.* 2010;20(1):15-32.

Table 1. Summary of diagenetic checks and available methods for different archaeological materials.

Sample type	Common checks	Other methods
Bone collagen	%C, %N, C:N atomic ratio, % collagen yield	FTIR – Bone collagen content Raman spectroscopy – Bone collagen content GC/MS – Amino acid profiling
Dentine collagen	%C, %N, C:N atomic ratio, % collagen yield	
Collagen amino acids	GC-FID or GC/MS to assess impurities and compare amino acid profiles to modern reference samples of the same taxa	
Fatty acids	GC and GC/MS to assess sample quality and lipid yield	
Crop remains	%C, %N, C:N atomic ratio	FTIR – check for the presence of carbonates and humic acids
Tooth enamel bioapatite	%CO ₃ , expected $\delta^{13}\text{C}$ range according to species and region (e.g. grazers vs. non-grazers)	FTIR – check for calcite, changes in crystallinity parameters (API, BPI, IRSF, BAI) Microscopic luminescence – transferal of metallic elements across material boundaries Measurement of trace elements – bulk values or section profiles
Bone bioapatite	%CO ₃ , $\delta^{13}\text{C}$ pattern between grazers and non-grazers	FTIR – check for calcite, changes in crystallinity parameters (API, BPI, IRSF, BAI) Microscopic luminescence – transferal of metallic elements across material boundaries Measurement of trace elements – bulk values or section profiles
Terrestrial snail shell	%CO ₃	Light microscopy, scanning electron microscopy, X-ray diffraction, and FTIR – check for conversion from aragonite to calcite and presence of secondary calcite.
Marine shell	%CO ₃	Light microscopy, scanning electron microscopy, X-ray diffraction, and FTIR – check for conversion from aragonite to calcite and presence of secondary calcite.

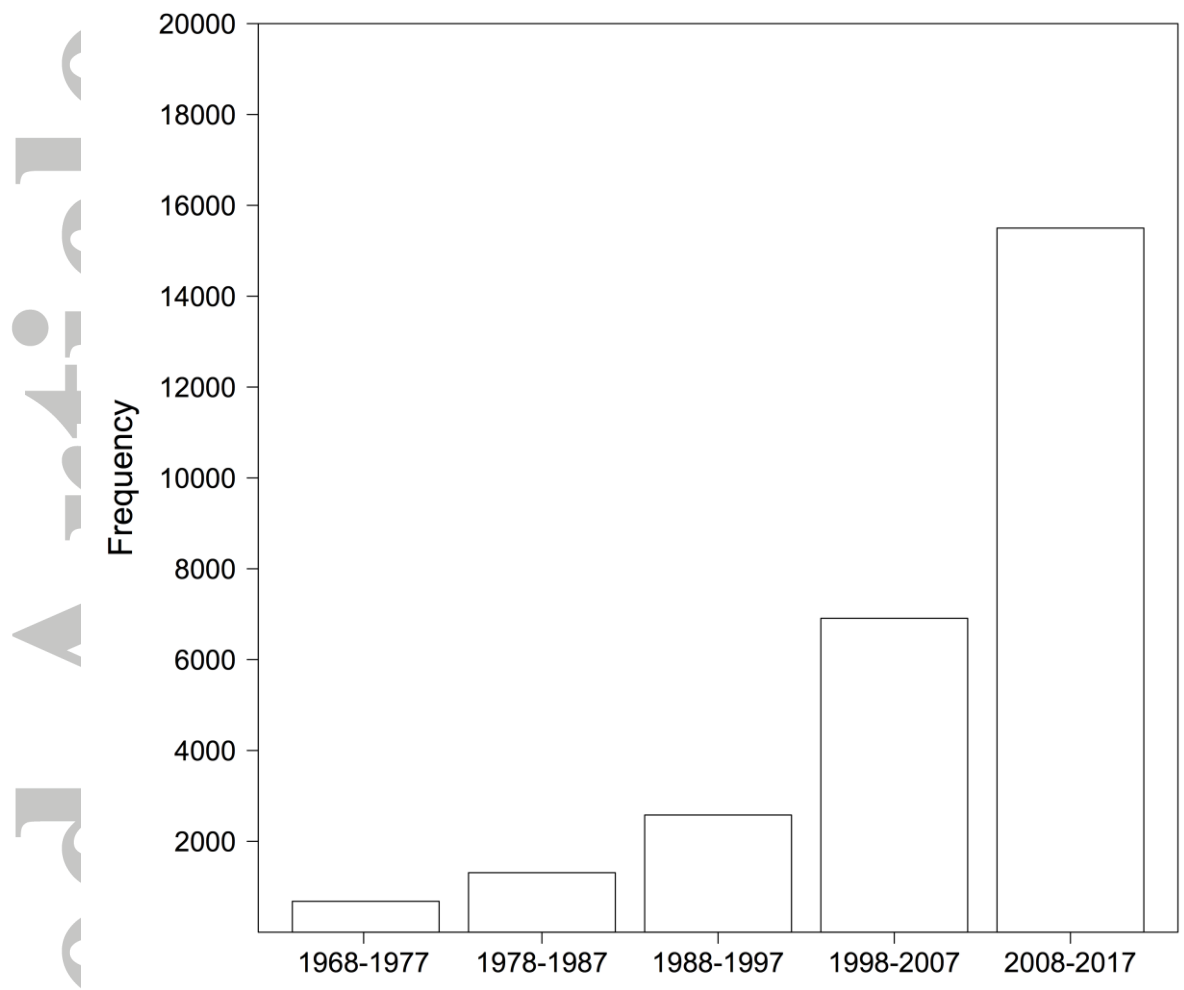


Figure 1. Bar-plot showing the number of mentions of ‘archaeology’ and ‘isotopes’ in archaeological publications according to Google Scholar for the past five decades.

Accepted Article

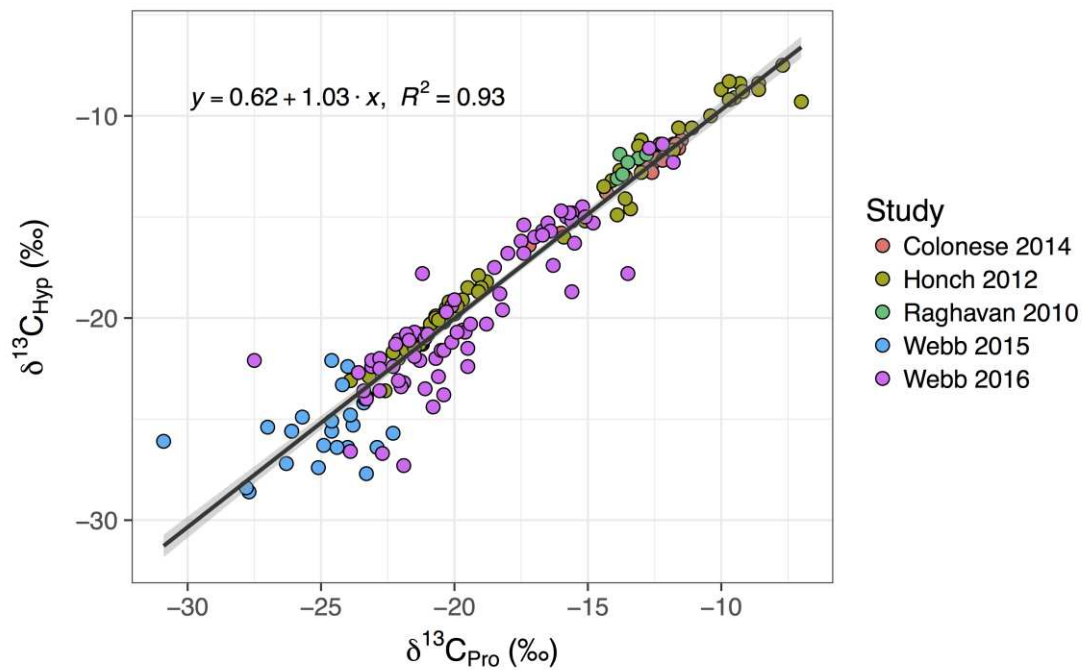


Figure 2. Comparison of published collagen and dentin proline and hydroxyproline $\delta^{13}C$ values from archaeological materials shows that there are only minor offsets between these amino acids ($y = 0.62 + 1.03x$, adjusted $R^2=0.9252$, $F(1,184)=2290$, $P<0.001$). The data were obtained from five studies that used LC/IRMS [68-72]. The solid lines depict the linear correlations, and the shaded areas depict 95% confidence intervals.

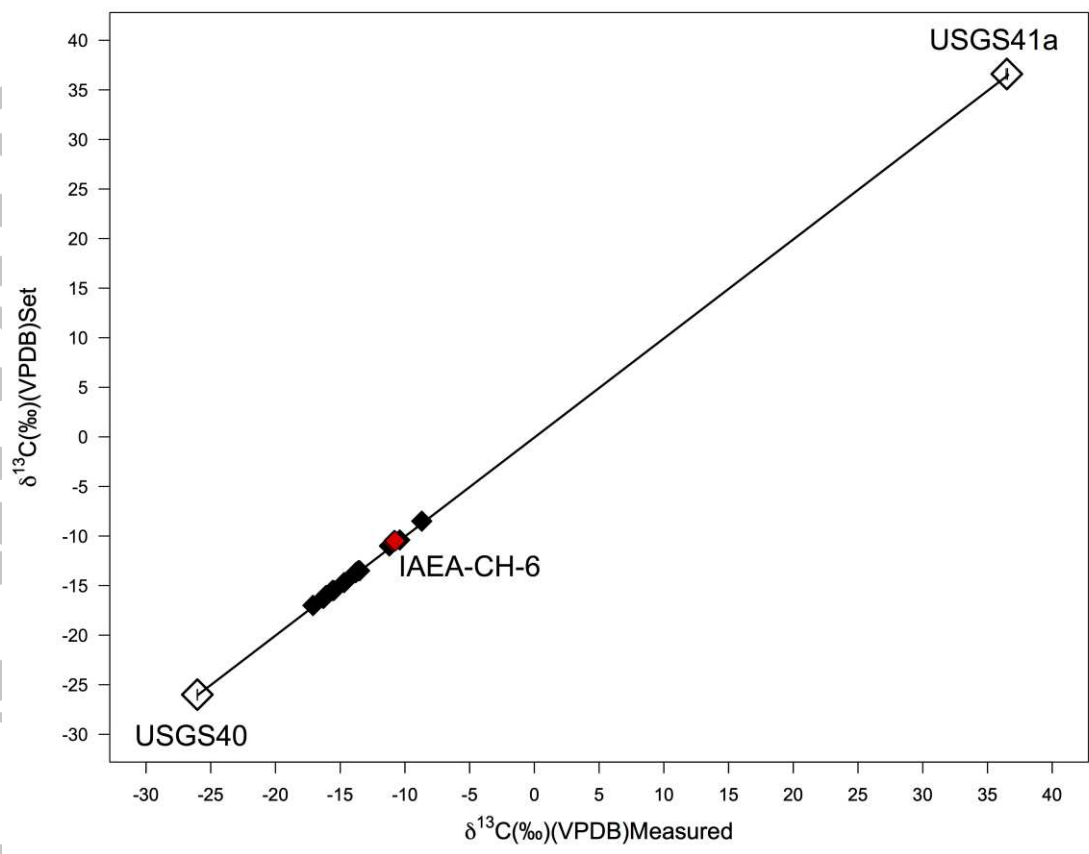


Figure 3. Plot demonstrating data normalization using secondary measurement standards.

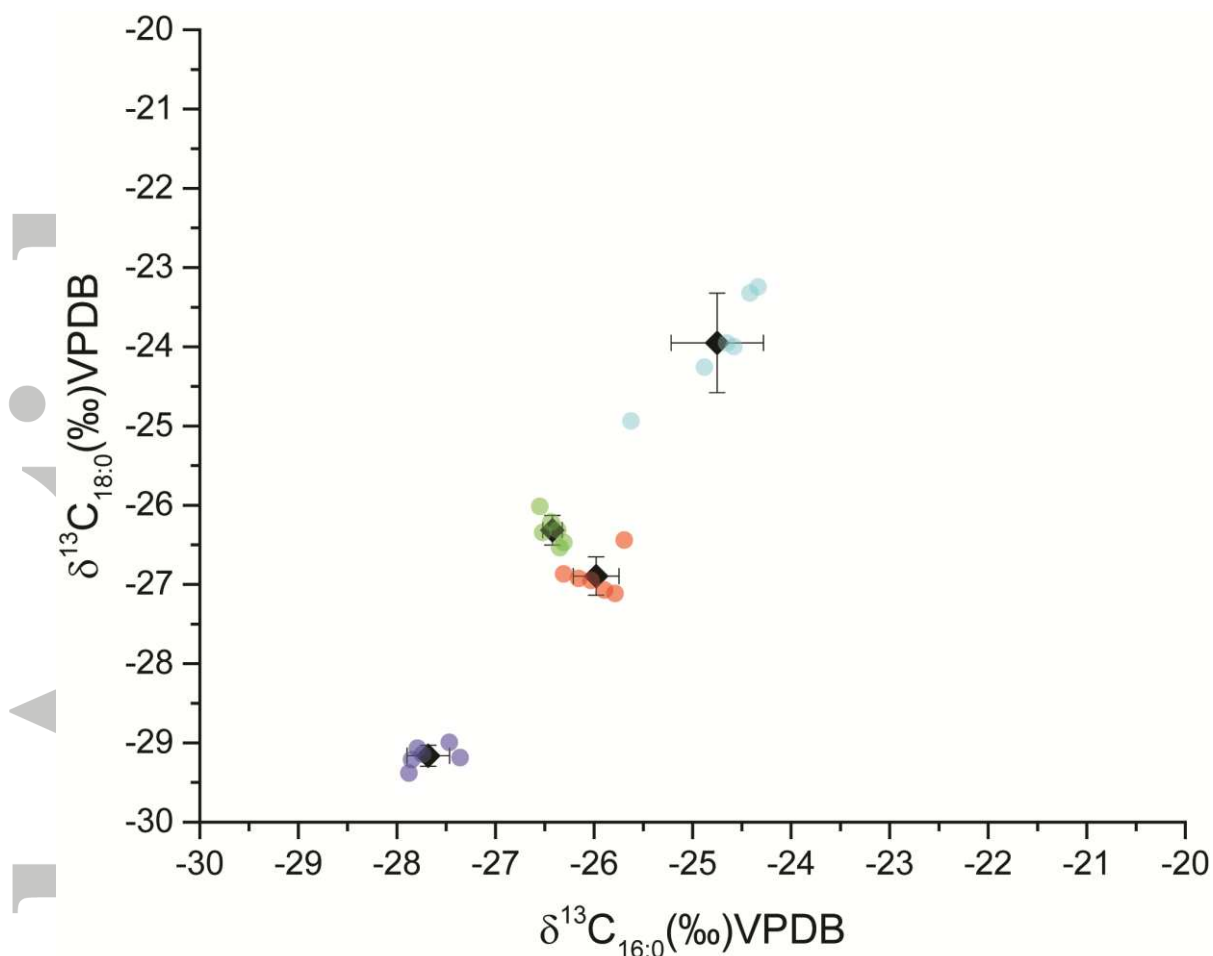


Figure 4. An example of intra-laboratory variability in compound-specific stable isotope measurements. Four pottery sherds were sampled by drilling and the resulting powder homogenized. Aliquoted subsamples were distributed to three different analysts at the University of York and extracted in duplicate according to established protocols [73]. Plot shows the $\delta^{13}\text{C}$ values of palmitic ($\text{C}_{16:0}$) and stearic acid ($\text{C}_{18:0}$), following calibration and correction for derivatization, for each separate extract and the mean and standard deviation of each sherd. Each sherd is represented by an individual color.

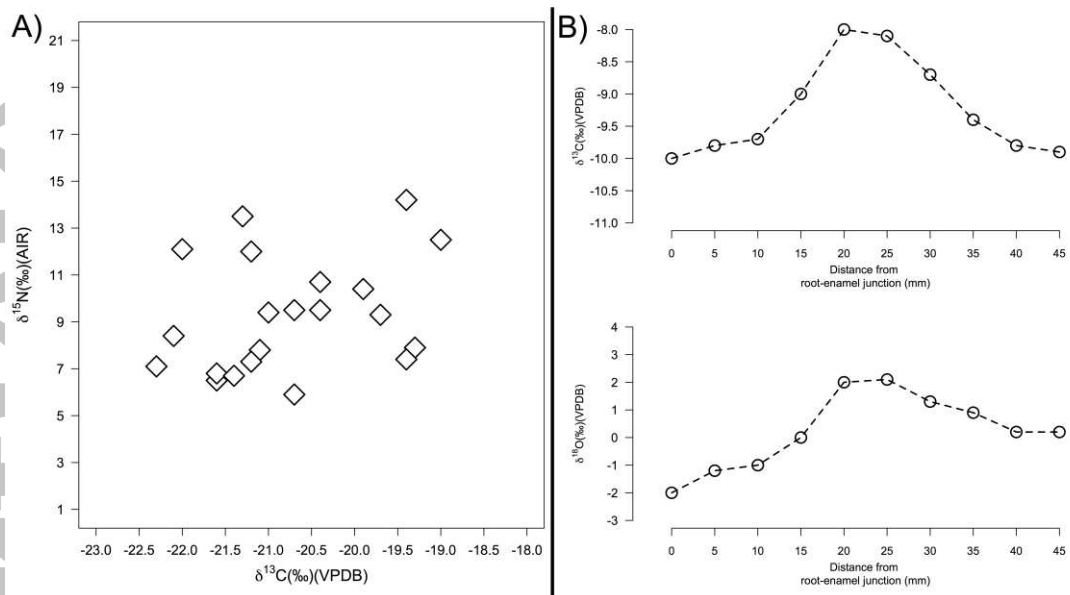


Figure 5. A) Example of differential scaling of $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ values in a scatterplot. **B)** Example of differential scaling in a sequential plot of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values of tooth enamel.

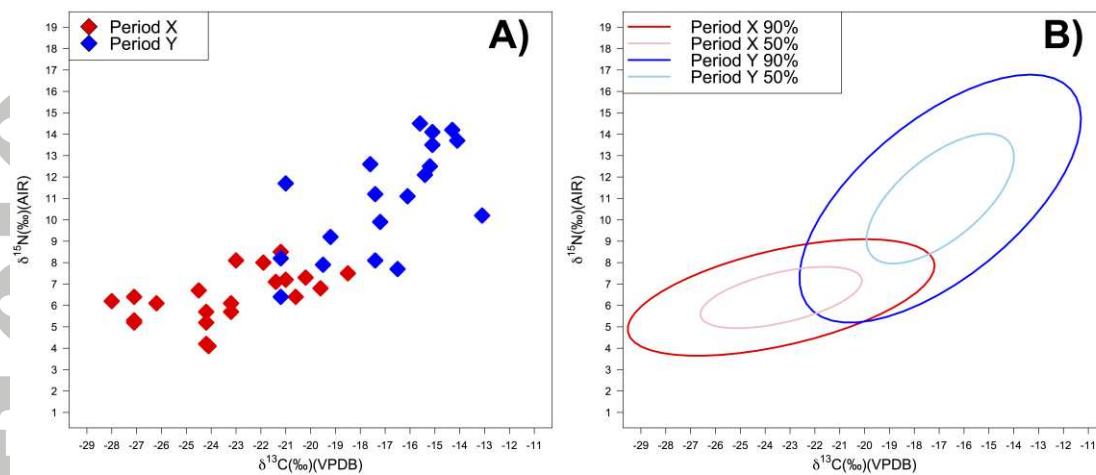


Figure 6. Scatterplot of $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ data for two groups overlain by Ellipses calculated at the 50% and 95% confidence intervals using the R [74] function ‘Ellipse’.

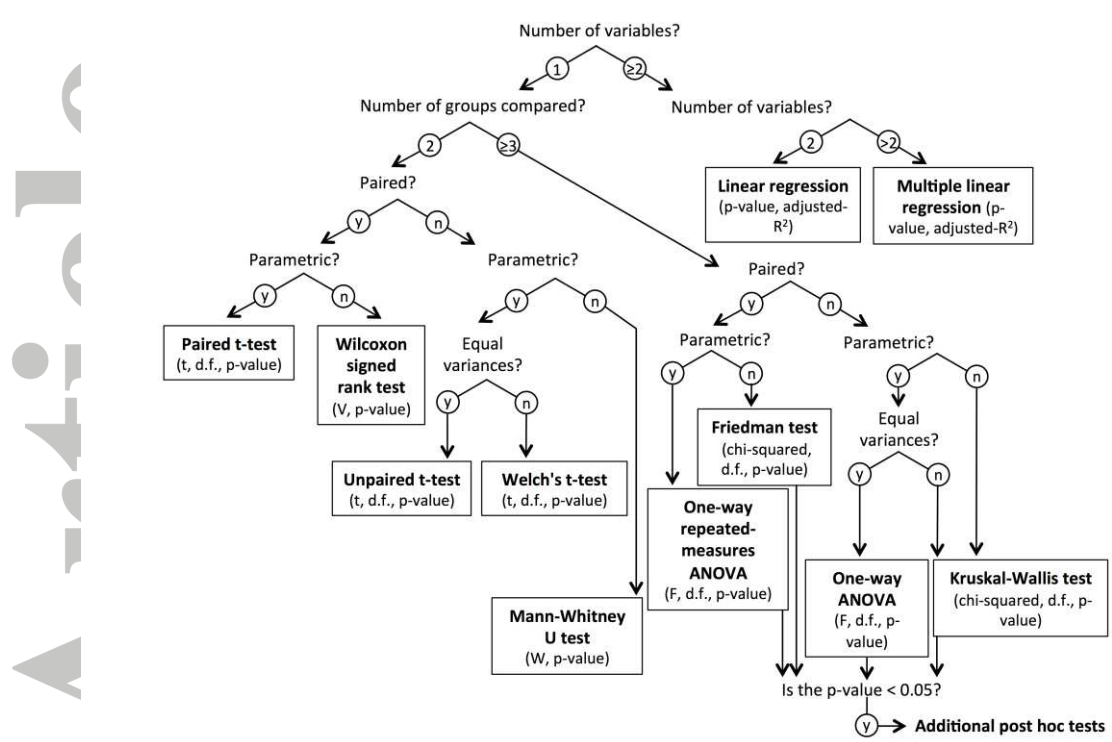


Figure 7. Comparison of different, commonly-used statistical analyses potentially useful for evaluating δ value differences between groups in archaeological datasets (based on McCrum-Gardner [75] and Marusteri and Bacarea [76]).

Accepted