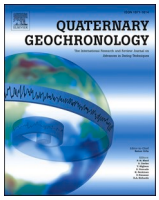
Quaternary Geochronology 85 (2024) 101631

Contents lists available at ScienceDirect

Quaternary Geochronology

journal homepage: www.elsevier.com/locate/quageo



Rapid, effective screening of tar seep fossils for radiocarbon and stable isotope analysis

Robin B. Trayler [a,*], Lauren E. Lopes [a], Patricia A. Holroyd [b], Sora L. Kim [a], John R. Southon [c]

[a] Department of Life and Environmental Sciences, University of California, Merced, CA, USA
[b] Museum of Paleontology, University of California, Berkeley, CA, USA
[c] Department of Earth System Science, University of California, Irvine, Irvine, CA, USA

A B S T R A C T

Tar seeps trap and preserve diverse fossil assemblages that reflect unique environmental histories. While the macro preservation of the fossils is usually good, preservation of organic bone collagen is often variable. Radiocarbon dating and stable isotope analysis of tar seep taxa can reveal distinct paleoecological insights but are complicated by the contamination from tar infiltration. Additionally, the removal of tar is complex and time-consuming. Therefore identifying fossils with preserved collagen prior to further investigation minimizes unnecessary damage to fossil collections and improves success rates for analyses that require organic preservation.

We used tar pit fossils where the collagen preservation state was independently known to test non-destructive (visual inspection) and minimally-destructive (infrared spectroscopy; FTIR) methods to determine the most reliable methods to identify bones with well preserved collagen. We found that while collagen is less often preserved in heavily weathered bones, visual cues alone are not a reliable indicator. Instead, the Water-Amide-on-Phosphate FTIR index is highly sensitive and specific at identifying fossils suitable for radiocarbon dating and stable isotope analysis. While our protocol and analysis code were developed using fossils from two California tar seep localities (McKittrick and Rancho La Brea), they are likely to be broadly applicable to other tar seep localities that have yielded fewer fossils, thus requiring even greater care when selecting specimens for further analysis.

1. Introduction

Asphalt seep *lagerstätten*, colloquially known as “tar pits”, have produced remarkable fossil assemblages, giving an unparalleled look into the ecology of floras and faunas during the late Pleistocene and Holocene (see reviews in Stock and Harris (1992) and McDonald et al. (2015)). Tar pits form when natural tar seeps through subsurface fractures to form a viscous sticky layer on the surface. This tar tends to trap animal and plant remains, leading to large accumulations of fossils. Importantly the rapid entrapment and later impregnation of tar into the remains can preserve fossils in environments where local conditions are otherwise not amenable to long term preservation (e.g., the neotropics; Lindsey and Seymour (2015)), and therefore act as an important source of information about these ecosystems. The notable morphological preservation of tar impregnated fossils comes at a cost however, for researchers interested in applying geochronological ($^{14}C_{collagen}$; Berger and Libby, 1968; Fox-Dobbs et al., 2014; Fuller et al., 2015; Fuller et al., 2014; Ho et al., 1969; Marcus and Berger, 1984; O’Keefe et al., 2023; Stock and Harris, 1992) or stable isotope techniques ($\delta^{13}C_{collagen}$; Coltrain et al., 2004; Fox-Dobbs et al., 2007, 2006; Fuller et al., 2020). Natural tar is 80–90% carbon by weight and is, by consequence of its age, radiocarbon-free. Consequently, even a small amount of tar is a potent contaminant that can bias $^{14}C_{collagen}$ and $\delta^{13}C_{collagen}$ measurements, and it must be removed prior to geochemical analysis (Berger and Libby, 1968; Coltrain et al., 2004; Fox-Dobbs et al., 2006; Fuller et al., 2014; Ho et al., 1969). Briefly, the current most common practice for extracting collagen from tar impregnated fossils involves repeatedly washing an aliquot of bone powder with organic solvents (e.g., methanol, toluene, acetone, petroleum ether) followed by acid digestion (HCl) of the bone mineral, gelatinization using low-molarity HCl and finally collagen purification via ultrafiltration (Fuller et al., 2014). Taken together, these methods require multiple days of work, large sample sizes (~150 mg) and due to variations in collagen preservation, may not yield viable collagen viable for further analysis. Therefore, identifying fossils with well-preserved collagen *prior* to tar removal and collagen extraction can reduce unnecessary damage to fossil collections, lower analytical costs, and improve research outcomes for geochemical studies of these fossils.

In this study we investigate potential screening methods to identify tar impregnated fossils with a high likelihood of collagen preservation. We developed a training data set of fossils from two geographically distinct seep areas in California, Rancho La Brea in Los Angeles County

* Corresponding author.
E-mail address: rtrayler@ucmerced.edu (R.B. Trayler).

https://doi.org/10.1016/j.quageo.2024.101631
Received 26 July 2024; Received in revised form 19 September 2024; Accepted 24 September 2024
Available online 26 September 2024

and McKittrick in Kern County, where the collagen preservation state was known from previous studies. These data were used to test two potential screening methods: visual, non-destructive, taphonomic scoring (Behrensmeyer, 1978) and minimally-destructive Fourier Transform Infrared (FTIR) spectroscopy. We calculated several commonly used FTIR indices to assess diagenesis in bone as well as a new index to assess tar impregnation. We used these initial training data to identify the most useful predictors of collagen preservation and content in other fossils (Fig. 1). We used these predictions of collagen content to assess the limitations of our initial training data, thereby resulting in a robust, freely available reference data set for researchers studying tar pit fossils for further analysis.

2. Background

2.1. Depositional and curatorial context

2.1.1. McKittrick

The McKittrick tar seeps are located along the eastern edge of the Temblor Range in the Southern San Joaquin Valley of California, just outside the town of McKittrick. Here, faulting allows oil to migrate from the underlying shale through porous sandstones of the Pliocene Etchegoin and Quaternary Tulare Formations to form layers of asphaltum in a NW-SE trending brea belt. The faunas appear to span from the late Pleistocene into the Holocene and the small number of radiocarbon dates for the locality support a late Pleistocene age for the majority of the fossils (Fox-Dobbs et al., 2014; France, 2008).

Specimens used in the current study are from collections made by University of California personnel and by Charles Sternberg for both the University of California and California Institute of Technology which primarily took place from 1921 to 1927 in several different pits (Merriam and Stock, 1921; Schultz, 1938; Sternberg, 1932; Stock, 1928). As reported by Sternberg (1932), fossils from McKittrick were cleaned using kerosene on site, which removed most surface tar. Based on direct examination of the collections, plaster was also used in joining broken bones, and there is no indication that animal-based glues were used in preparation.

2.1.2. Rancho La brea

The Rancho La Brea tar seeps occur in the northern part of the Los Angeles Basin, south of the Santa Monica Mountains, in urban Los Angeles. The seeps have produced a diverse fauna spanning the late Pleistocene to Holocene (see Stock and Harris (1992)). The tar seeps form pools of liquid asphaltum, which act as natural traps. The fossils included in this study are from University of California excavations conducted from 1908 to 1912 (Merriam, 1911; Stoner, 1913) and include those with radiocarbon dates reported in O'Keefe et al. (2009) and stable isotope analysis from Fox-Dobbs et al. (2006).

2.2. Bone collagen preservation and isolation

Bone is a composite material consisting of approximately 20% organic and 80% mineral components, both of which are common tissues for geochemical analysis. The mineral fraction is bioapatite [$Ca_5(PO_4)_3OH$] with carbonate (CO_3^{-2}) substitutions in the hydroxyl and phosphate sites (Driessens and Verbeeck, 1990; Elliott, 2002). The organic phase is primarily collagen (Collins et al., 2002), which is relatively robust and can persist in bone for tens of thousands of years (Clementz, 2012; Collins et al., 2002), although it can be altered or removed by a variety of mechanisms (see review of Collins et al. (2002)). The high collagen content of bone relative to tooth enamel and dentine has made it a tissue of choice for many paleontologists and archaeologists interested in stable ($\delta^{13}C$, $\delta^{15}N$) and radioactive (^{14}C) isotope analysis of fossil specimens. However, collagen preservation can vary considerably between localities depending on a variety of taphonomic factors.

While the most precise assessment of bone collagen quality is via quantitative analysis of amino acids (Guiry and Szpak, 2020; Wyckoff et al., 1963), most archaeological and paleontological studies rely on measurements of bulk bone collagen to assess preservation. Bone collagen preservation in modern and fossil specimens is usually assessed using a combination of collagen yield, carbon and nitrogen content, and the atomic carbon-to-nitrogen ratio (C/N_{atomic}). The amino acid compositions of modern bone collagen are conservative across a variety of vertebrate groups (Szpak, 2011) allowing broad comparisons of these parameters. Collagen content in fresh bone ranges from 12 to 33% with an average of about 22% (Ambrose, 1990; Collins et al., 2002; Stafford Jr et al., 1988; Van Klinken, 1999). Low bone collagen yield in fossil specimens can indicate degradation and loss of individual amino acids or as peptides (Bada et al., 1989; Van Klinken, 1999). Since the stable isotope composition of different amino acids within a tissue can vary substantially (Hare et al., 1991) preferential loss of some amino acids will change in stable and or radioactive isotope composition. However, amino acid profiles in fossil collagen are often remain unchanged until substantial collagen loss has occurred (<1% collagen remaining; Van Klinken (1999)). Likewise, low carbon and nitrogen contents can indicate alteration (Guiry and Szpak, 2020). Several C/N_{atomic} ranges have been proposed as indicative of well-preserved collagen including 2.9–3.6 (Ambrose, 1990), 3.1–3.5 (Van Klinken, 1999), and a variety of taxon specific ranges (Guiry and Szpak, 2021). Schwarcz and Nahal (2021) suggested a C/N_{atomic} midpoint of 3.243 based on a theoretical mass-balance of carbon and nitrogen contents of individual amino acids. Weathering is another potential pathway of collagen loss (Koch et al., 1999, 1999, 1999; Trueman et al., 2004), although it may be less pronounced in temperate climates (Fernández-Jalvo et al., 2010). In this study we use two criteria to assess collagen preservation. For specimens analyzed as part of this study we use a C/N_{atomic} of 2.9–3.6 to indicate well preserved collagen. For values taken from the literature, we rely on the original study authors assessment.

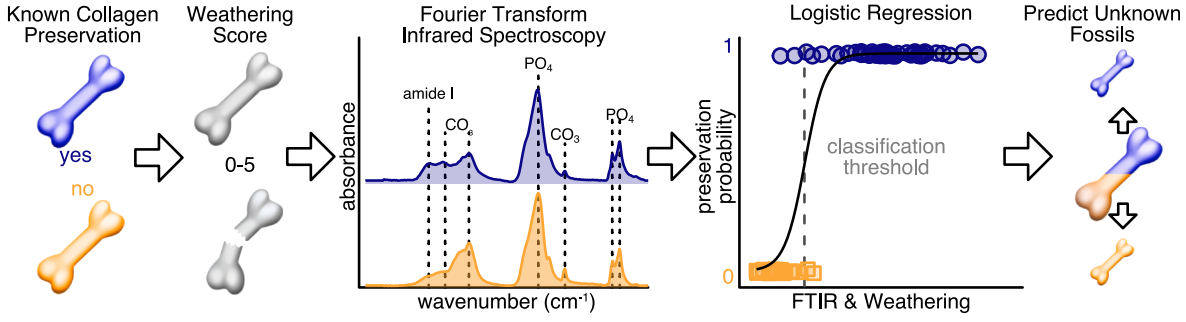


Fig. 1. Workflow of model development.

2.3. The problem of tar

Collagen is also susceptible to contamination from a variety of other organic carbon sources, including endogenous lipids, humic acids, and non-collagenous proteins (Guiry and Szpak, 2021). In particular, lipids and humics have high carbon, but low nitrogen, contents and lead to higher C/N_{atomic} than the accepted ranges (e.g., >3.6, Ambrose (1990); >3.5, Van Klinken (1999), >3.5, Guiry and Szpak (2021)). Tar is a particularly potent contaminant; it has a very high carbon content, readily impregnates bone pore space, and is pernicious and difficult to completely remove. Several methods have been developed to purify and extract collagen from tar-impregnated fossils (Coltrain et al., 2004; Fox-Dobbs et al., 2006; Fuller et al., 2014, 2015; Ho et al., 1969) that follow a similar set of steps. First, bone (powder or chunks) is repeatedly washed with solvents (e.g., toluene, methanol, acetone) to extract the tar. Second, bone bioapatite is dissolved using either acid (HCl) or a chelating agent (EDTA), leaving isolated collagen. While these methods differ in their specific details, in all cases the tar removal is a complex, time consuming process (~3 days, Coltrain et al. (2004); 5–6 days, Fox-Dobbs et al. (2006); 2–3 days, Fuller et al. (2014)). Finally, even after these steps, not all specimens will yield viable collagen that meets the criteria for good preservation. Collagen preservation can also vary substantially between tar-pit deposits. For example, Coltrain et al. (2004) reported that 13 of 143 bones from Rancho La Brea failed to yield collagen (8% failure), whereas France (2008) (following the methods of Coltrain et al. (2004)) reported a 78% failure rate (22 of 28) for fossils from McKittrick. Similarly, preliminary attempts at collagen extraction for this study from randomly selected McKittrick fossils yielded a success rate of 28%.

2.4. Infrared spectroscopy

Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR), is a minimally destructive vibrational spectroscopic technique that characterizes the molecular functional groups of a material by irradiating it with infrared light. Since molecular structure determines which infrared wavelengths are transmitted or absorbed, ATR-FTIR can be used to attribute absorbance bands to different functional groups, and semi-quantitatively determine the chemical composition of a material (Stuart, 1991) .

Within the context of paleontological and archaeological specimens, ATR-FTIR is often used to investigate the crystal-chemical properties of bone bioapatite and collagen, with a particular focus on diagenesis and alteration during fossilization (Chadefaux et al., 2009; Hassan et al., 1977; Roche et al., 2010; Sponheimer and Lee-Thorp, 1999). The absorbance band-positions of the major components of bone are known; including inorganic phosphate (PO_4) and carbonate (CO_3) (Fleet, 2009; Sponheimer and Lee-Thorp, 1999), and organic amides and lipids (Chadefaux et al., 2009; Lebon et al., 2016; Liden et al., 1995) (Table 1; Fig. 1and 2). Infrared spectra of tar contains multiple absorbance bands that partially, or completely overlap with amide, carbonate, and lipid band positions in collagen (Fig. 2), and the degree to which this overprinting may interfere with accurate identification of collagen is unclear. Previous work has shown that FTIR is an effective tool at identifying organic preservation in archaeological contexts (Lebon et al., 2016), but to our knowledge, there have been no prior attempts to apply the methodology to tar impregnated fossils.

Table 1
Nominal FTIR band positions of several relevant chemical groups. Actual band positions may be shifted by several cm^{-1}.

Band position (cm^{-1})	Functional Group
2925	Lipid
2854	Lipid
1650	Amide I
1551	Amide II
1545	A-Type Carbonate
1415	B-Type Carbonate
1231	Amide III
1020	Phosphate
880	Carbonate
605	Phosphate
565	Phosphate

3. Materials and methods

3.1. Sample selection

We selected specimens from the University of California, Museum of Paleontology (UCMP) collection where radiocarbon dating and/or isotopic analysis ($\delta^{13}C$, $\delta^{15}N$, ^{14}C) of the organic collagen fraction has previously been attempted (Table 2) as part of other research projects (Fox-Dobbs et al., 2006, 2014; O'Keefe et al., 2009) as well as unpublished data for Rancho La Brea fossils (Southon & Holroyd, unpublished). These specimens fall into two groups: 1) fossils that produced viable, well-preserved collagen (n = 50) and 2) fossils that failed to yield well-preserved collagen (n = 20). A complete list of samples including UCMP catalog number, locality, taxon, genus and species is available in the linked data supplement. Additional details including photographs of some specimens, skeletal element and previous publication information are available in the online UCMP database at https://ucmpdb.berkeley.edu. We considered collagen to be well-preserved and viable for isotopic analysis if it passed the standard metrics for collagen preservation (see reviews of Guiry and Szpak (2020); Guiry and Szpak (2021)) as reported by the original study authors. We collected 5–10 mg of cortical bone powder using a handheld rotary tool and a dental drill bit. These powders were stored in 1.5 mL micro-centrifuge vials prior to analysis. Additionally, the bone powders were not chemically treated to remove tar (Fuller et al., 2014), or otherwise treated to remove carbonates or organics (Koch et al., 1997), prior to FTIR analysis. All specimens are from either Rancho La Brea (n = 48) or McKittrick (n = 22) and cover a wide range of carnivore and herbivore taxa, and include mammals, birds, and reptiles.

3.2. FTIR indices

ATR-FTIR spectra were collected for all fossil specimens using a *Bruker Vertex 70 Far-Infrared Fourier Transform Infrared Spectrometer* housed in the Nuclear Magnetic Resonance Facility at the University of California, Merced. The spectra were collected from 400 to 4000 cm-1 over 32 scans at a spectral resolution of 4 cm^{-1}. Each spectrum was background-corrected using several baseline points and slightly smoothed prior to index calculation using custom R scripts modified from those of Trayler et al. (2023), available in the supplementary material.

The resulting 70 spectra were used to calculate two FTIR indices commonly used to investigate organic content and diagenesis in fossil bone. The Water-Amide-on-Phosphate-Index (WAMPI) is the ratio of the Amide-I and ν_2PO_4 phosphate absorbance bands ($\frac{B_{1650}}{B_{605}}$). The WAMPI tracks bone collagen content where higher WAMPI values should indicate more prominent amide-I bands and by proxy better collagen preservation (Lebon et al., 2016; Roche et al., 2010; Trayler et al., 2023). The Phosphate-Crystallinity-Index (PCI) is the sum of the ν_2PO_4 and ν_4PO_4 phosphate absorbance band maxima, normalized by the depth of valley between these two peaks ($\frac{B_{605}+B_{565}}{V_{590}}$ (Sponheimer and Lee-Thorp, 1999);). Since increases in bone crystallinity sharpen the two phosphate peaks and deepen the valley, higher PCI values reflect greater diagenetic alteration of the bone mineral (Sponheimer and Lee-Thorp, 1999). Furthermore, higher PCI values can also reflect heat-induced changes to crystal order and structure, resulting from deliberate (cultural) or natural burning (wildfires), which is also expected to remove organic

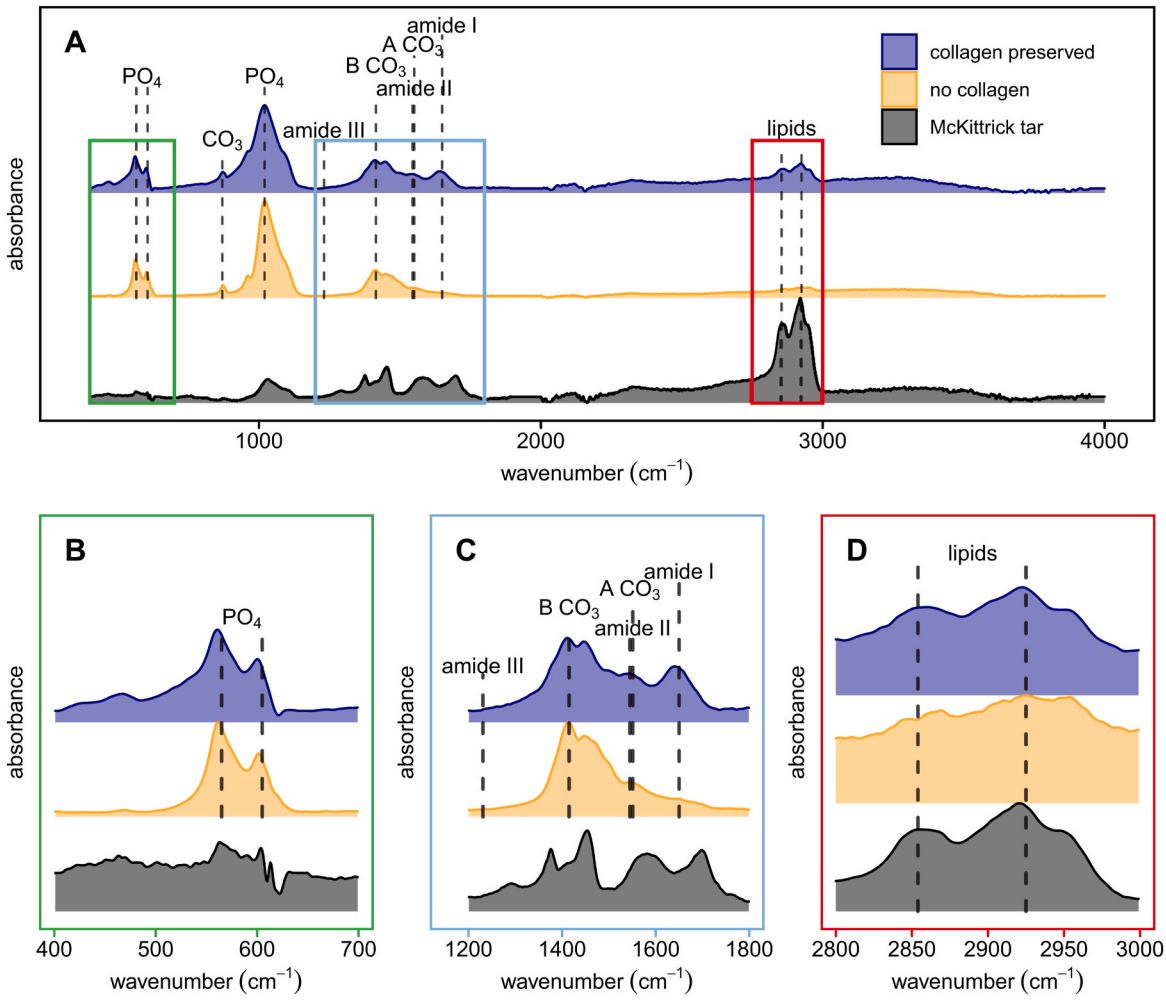


Fig. 2. Representative example ATR-FTIR spectra for McKittrick tar seep bones with/without preserved collagen and McKittrick tar. Vertical dashed lines indicate band positions from Table 1. Carbonate, phosphate, and amide bands are only applicable to bone spectra while the lipid bands occur in in both materials. A) full spectrum for all materials. Bottom Panel: zoomed view of the B) phosphate, C) amide/carbonate, and D) lipid regions highlighted by colored boxes in panel A.

Table 2
Number of specimens in the well preserved and poorly preserved groups for Rancho La Brea and McKittrick asphalt seeps.

Locality	Collagen	No Collagen	Collagen Preservation %
Rancho La Brea	43	5	90%
McKittrick	7	15	31%

material (Thompson et al., 2009, 2013).

We also calculated the ratio of two lipid absorbance bands (Liden et al., 1995) normalized to the ν_2PO_4 phosphate absorbance band ($\frac{B_{2925}+B_{2854}}{B_{605}}$). Endogenous lipids exhibit prominent absorbance bands at about 2925 cm^{-1} and 2854 cm^{-1}, as does McKittrick and Rancho La Brea asphaltum (Fig. 2). Thus, this Lipid-on-Phosphate-Index (LPI) should reflect excess tar or lipid content in bone. While the loss of endogenous lipids is variable in the fossil record (Collins et al., 2002; Koch et al., 1999), tar seep fossils are impregnated with geologic oils, which have been proposed as a "preservative" (Stock and Harris, 1992), suggesting that higher LPI values could correspond to better collagen preservation.

3.3. Taphonomy and weathering stages

To assess if visual assessment of weathering was a reliable indicator of collagen preservation, we scored each specimen according to the set of weathering stages established by Behrensmeyer (1978) for large mammals and using images in Behrensmeyer and Miller (2012) and Fernández-Jalvo and Andrews (2016) as visual referents. These stages are numbered 0 to 5, with stage 0 showing no modification and stage 5 showing significant splintering, flaking, as well as a greater tendency to break. If weathering is a significant factor, we would expect specimens with less collagen to have higher weathering scores.

3.4. Collagen isolation and analysis

We attempted to extract collagen from a subset of fossils to validate model performance (discussed below in Section 4.2). Collagen was isolated at the UC Irvine Keck Carbon Cycle AMS Laboratory following a procedure based on that of Fuller et al. (2014) for tar impregnated fossils. Briefly, crushed bone samples of ~200 mg were sonicated at ~40 °C to remove bulk asphalt, in 2:1 toluene:methanol with solution changes every hour until the solution was visually clear, then in methanol for 1 h, and deionized water for 1 h. The samples were then demineralized in 1N HCl overnight at 25 °C, using a measured amount of acid calculated as just sufficient to dissolve the bone if the composition was 100% bioapatite, ie, if no collagen was present. The samples were then gelatinized over several hours in 0.01 N HCl at 80 °C, with individual samples removed from heat as soon they were fully dissolved. The gelatin was ultra-filtered at 100 kDa to discriminate against refractory

high molecular weight hydrocarbon aggregates (asphaltenes) that had survived the solvent treatments, and at 5 kDa to remove small stray peptides and other low molecular weight debris, and the intermediate 5–100 kDa fraction was then lyophilized. This extracted collagen was analyzed for its carbon and nitrogen contents using a Fisons NA1500NC elemental analyzer coupled with a Finnigan Delta Plus Isotope Ratio Mass Spectrometer.

3.5. Statistical methods

3.5.1. Statistical analyses

To determine the best predictors of collagen preservation, we used the weathering scores and FTIR indices from the fossils of known collagen preservation to develop a training dataset to assess which predictors are most strongly associated with collagen presence or absence. We fit a logistic regression model in the form of:

$$ln(odds) = ln\left(\frac{p}{1-p}\right) = \beta \times x + \alpha \quad (1)$$

to each predictor (WAMPI, PCI, LPI, weathering score) separately, where p is the probability of collagen preservation (between 0 and 1), x is the value of an individual predictor with probability p, and β and α are regression coefficients. These logistic regression models were used to predict collagen presence for each fossil using a log-odds >1 as a classification threshold, such that specimens with $p \geq 0.5$ were predicted to contain collagen and specimens where p was < 0.5 were predicted to lack collagen.

Model performance was assessed by calculating the sensitivity and specificity of each logistic regression model to determine the most useful predictor(s) for collagen presence. Sensitivity is defined as:

$$sensitivity = \frac{true\ positives}{true\ positives + false\ negatives}$$

and specificity as:

$$specificity = \frac{true\ negatives}{true\ negatives + false\ positives}$$

Both sensitivity and specificity vary between 0 and 1, and in general, higher values for both indicate better model performance. A perfectly performing model that predicts zero false negatives and zero false positives would therefore have both a sensitivity and specificity of 1. However, a model that *always* predicts the presence of collagen, regardless of preservation state, would have a high sensitivity (~1), and a low specificity (~0), with the opposite being true for a model that always predicts the absence of collagen. Therefore, an ideal model is both sensitive and specific, with both values close to one.

4. Results & discussion

4.1. Training data

The WAMPI, PCI, and weathering score logistic regression model-fits were statistically significant, and the LPI was not (Fig. 3). The Water-Amide-on-Phosphate Index (WAMPI) was the best performing predictor and is both highly sensitive and specific when predicting collagen presence and absence (Table 3). The WAMPI is calculated using the Amide-I band height, which directly correlates to bone nitrogen and organic content (Lebon et al., 2016; Roche et al., 2010). The Phosphate-Crystallinity Index and weathering scores are highly sensitive, but are only weakly specific. Both models have negative β coefficient, indicating increasing predictor values correspond to lower collagen preservation (Fig. 4). Weathering scores and PCI track taphonomic processes; macro scale weathering and microscopic apatite recrystallization, respectively. In other words, high degrees of weathering (high PCI, high weathering score), are somewhat predictive of collagen absence, but less weathering is not necessarily predictive of collagen presence. The proposed Lipid-on-Phosphate Index performed poorly and was not able to distinguish collagen presence or absence.

4.2. Validation

To validate model performance, we collected FTIR spectra on 235 other UCMP McKittrick fossils for which collagen preservation was unknown. Using these data we calculated the best performing FTIR index (WAMPI) to predict collagen presence. A WAMPI threshold of 0.43 (Table 3) predicts that 75 of 235 fossils (31%) have preserved collagen, similar to the rate we observed from random selection of McKittrick fossils (28%). From these 235 fossils we selected 67 specimens for attempted collagen extraction based on their WAMPI values; 18 were predicted to contain no collagen (WAMPI <0.43) and 49 were predicted to have well preserved collagen (WAMPI >0.43).

Of the 67 fossils in the validation set, 34 produced well preserved collagen (C/N_{atomic} = 3.30 ± 0.08, mean±1S.D); 32 were from the group predicted to contain collagen, and 2 from the group were predicted to contain no collagen (Fig. 4). Given that the underlying collagen preservation rate in McKittrick fossils appears to about 20–30%, a 65% chance of successfully extracting collagen is a substantial improvement. However, the validation dataset had a sensitivity of 0.94 and a specificity of 0.5. While the sensitivity of the validation data is similar to that of the training data (Table 3), the specificity is notably lower. This discrepancy suggests that when applied to only McKittrick fossils, the training classification threshold (WAMPI >0.43) has a higher false positive rate. This is likely because of imbalances in the training data set. About 80% of training samples with well preserved collagen are from Rancho La Brea, whereas 76% of the samples without preserved collagen are from McKittrick which reflects the underlying preservation rate of the two localities.

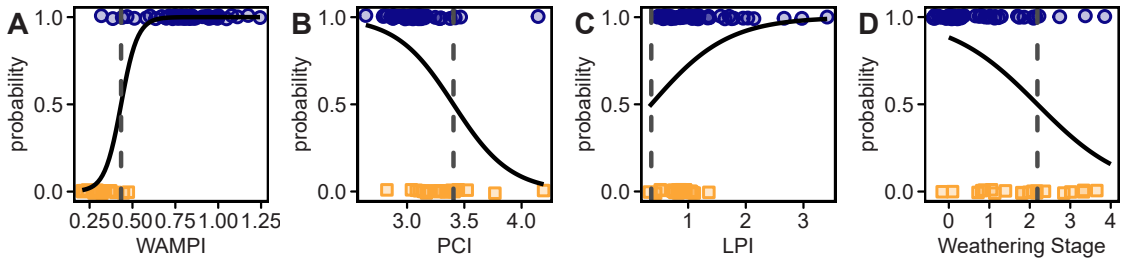


Fig. 3. Training data logistic regression results for each of the four possible predictors of collagen presence [A) WAMPI, B) PCI, C) LPI, D) Weathering Stage]. The vertical dashed line is the predictor value with an odds-ratio of 1 which we used as the prediction threshold (see Table 3 for details). Blue dots indicate samples with known collagen preservation and yellow dots indicate samples with known collagen absence. Points have been vertically-jittered slightly for visual clarity.

Table 3
Summary of training data logistic regression results. Model coefficients (α, β) correspond to eq. (1). The *Threshold* column indicates the predictor value with an odds-ratio of 1. If β is positive then *Predictor* values higher than the *Threshold* predict collagen presence, whereas if β is negative, *Predictor* values lower than the *Threshold* predict collagen presence.

Predictor	α	β	p value	Threshold	Sensitivity	Specificity
WAMPI	−9.13	21.06	0.003	0.43	0.96	0.90
PCI	13.80	−4.05	0.004	3.40	0.92	0.30
Weathering Score	2.06	−0.93	0.001	2	0.94	0.25
LPI	−0.49	1.47	0.057	0.34	1.00	0.05

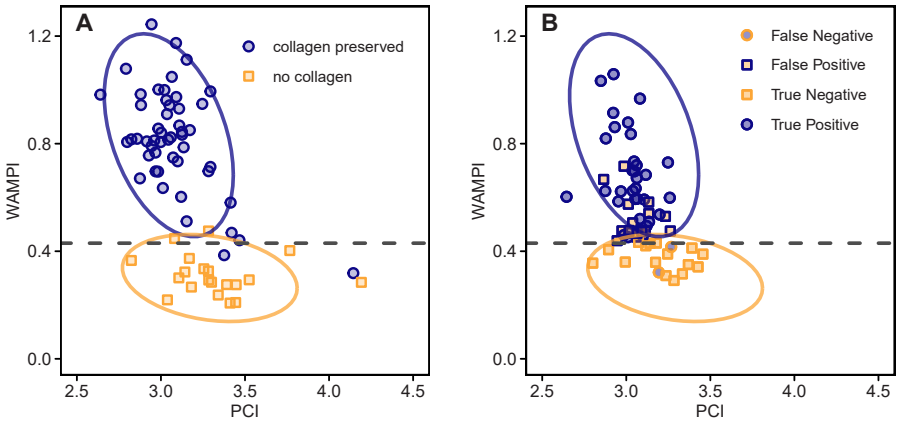


Fig. 4. Plot of the PCI and WAMPI FTIR indices for the A) training and B) validation data sets. These two indices have the highest combined sensitivity and specificity. The colored ellipses on both panels contain 95% of the *training* data in each group. The horizontal dashed line indicate the initial WAMPI classification threshold of 0.43 (Table 3). A) Initial training data colored by collagen presence or absence. B) Model validation results using 67 fossils with unknown collagen preservation. The perimeter color of each point in B) indicates the WAMPI based prediction of presence/absence and the fill color indicates actual collagen presence/absence. The colored ellipses and horizontal dashed line are the same as shown in the left panel.

4.3. Combined data and use

To address the mismatch in the training data set, we combined the training and validation data sets and recalculated the WAMPI logistic regression model (Table 4). The sensitivity and specificity for the combined data are both 0.87, indicating a low false positive and low false negative rate. The threshold with a greater than 50% chance of collagen preservation is slightly higher (WAMPI >0.5) than the training data threshold. Importantly, it should be noted however that our choice of an odds-ratio of 1 (>50% preservation probability) is somewhat arbitrary, and that other thresholds could be calculated and used depending on the needs of a particular study (Fig. 5). Rearranging eq. (1) to solve for the threshold and using the regressions coefficient in Table 4 gives:

$$WAMPI\ threshold = \frac{ln(odds) - \alpha}{\beta} = \frac{ln(odds) + 6.94}{13.9}$$

This equation can be used to calculate classification thresholds at arbitrary probabilities. For example, a higher odds-ratio of 3 (>75% preservation probability) gives a WAMPI threshold of 0.58 and an odds-ratio of 19 (>95% preservation probability) gives a WAMPI threshold of 0.78 (Fig. 5). These higher odds ratios come with a corresponding higher false-negative rate however, if fossils are abundant, then this trade off may be worthwhile to ensure only the best-preserved specimens are chosen for further geochemical work.

Table 4
Logistic regression result for the Water-Amide-on-Phosphate Index using the combined, training and validation datasets.

Predictor	α	β	p value	Threshold	Sensitivity	Specificity
WAMPI	−6.94	13.9	1×10^{-8}	0.50	0.87	0.87

5. Conclusions

Preparing tar seep fossils for isotope analysis ($\delta^{13}C$, $\delta^{15}N$, ^{14}C) is costly, time consuming, and can be hampered by poor collagen preservation. Here, we have presented methods and a reference data set for quickly identifying tar seep fossils with a high likelihood of collagen preservation. Entirely non-destructive visual taphonomic indicators and the amount of impregnating tar are moderate to poor predictors of collagen preservation. In contrast, minimally destructive FTIR analysis is highly sensitive and specific, sampling and data collection are rapid, and Fourier Transform Infrared spectrometers are available at most major research universities. The Water-Amide-on-Phosphate index is the best predictor, and predictive threshold of >0.5 is effective at identifying tar impregnated fossils with preserved collagen. This threshold performs well for both Rancho La Brea and McKittrick tar pit fossils and is likely broadly applicable to other tar pit fossils. The underlying data and code are freely available, such that it can be used as a screening tool for fossils from other tar pit deposits and strengthened by the inclusion of additional data. Given the increasing interest in tar pit faunas (Lindsey and Seymour, 2015; Seymour, 2015; Solorzano et al., 2015), their importance for understanding paleoclimate and extinction (O'Keefe et al., 2023), and the associated difficulties in excavating and curating these specimens, minimizing large-scale destructive sampling is crucial to continued research on these faunas.

Data availability

All data (infrared spectra, collagen preservation state, taxonomic details) and analysis code are available in a Zenodo Repository (https://doi.org/10.5281/zenodo.13798959) and at github.com/robintrayler/mckittrick_FTIR.

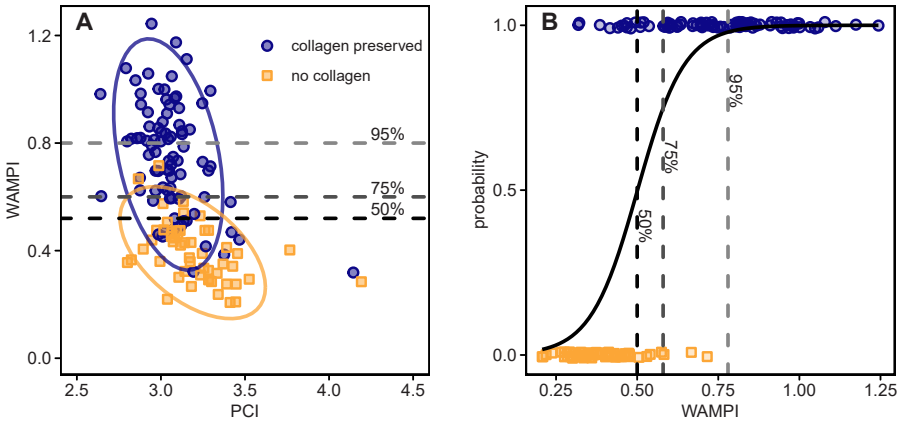


Fig. 5. Plot of final reference data set (training and validation data combined). A) Plot of PCI and WAMPI FTIR indices for the combined data set. The colored ellipses contain 95% of the data in each group. The horizontal dashed lines indicate the 50%, 75%, and 95% WAMPI thresholds. B) Final logistic regression model using the Water-Amide-on-Phosphate index for the combined data set. The vertical dashed lines indicate the 50%, 75%, and 95% preservation probability WAMPI thresholds.

CRediT authorship contribution statement

Robin B. Trayler: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Lauren E. Lopes:** Writing – review & editing, Writing – original draft, Software, Formal analysis, Data curation, Conceptualization. **Patricia A. Holroyd:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sora L. Kim:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **John R. Southon:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Robin Trayler reports financial support was provided by National Science Foundation. Sora L Kim reports financial support was provided by National Science Foundation. Patrica A Holroyd reports financial support was provided by National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is available in an online repo.

Acknowledgments

We would like to thank HyeJoo Ro and Dr. Gina Palefsky for their help with sample collection, and Dr. David Rice for his assistance with FTIR data collection, and Dr. Elizabeth Black for her discussion of statistical analyses. This work was supported by NSF-EAR-2138163 (RBT and SLK) and NSF-EAR-2138164 (PAH).

References

Ambrose, S.H., 1990. Preparation and characterization of bone and tooth collagen for isotopic analysis. J. Archaeol. Sci. 17, 145.

Bada, J.L., Schoeninger, M.J., Schimmelmann, A., 1989. Isotopic fractionation during peptide bond hydrolysis. Geochem. Cosmochim. Acta 53, 3337–3341.

Behrensmeyer, A.K., 1978. Taphonomic and ecologic information from bone weathering. Paleobiology 4, 150–162.

Behrensmeyer, A.K., Miller, J.H., 2012. Building links between ecology and paleontology using taphonomic studies of recent vertebrate communities. Paleontology in ecology and conservation 69–91.

Berger, R., Libby, W.F., 1968. UCLA radiocarbon dates VIII. radiocarbon. an international journal of cosmogenic isotope research 10, 402–416. https://doi.org/10.1017/S0033822200010997.

Chadefaux, C., Le Hô, A.-S., Bellot-Gurlet, L., Reiche, I., 2009. Curve-fitting Micro-ATR-FTIR studies of the amide I and II bands of type I collagen. In: Archaeological Bone Materials.

Clementz, M.T., 2012. New insight from old bones: stable isotope analysis of fossil mammals. J. Mammal. 93, 368–380.

Collins, M.J., Nielsen–Marsh, C.M., Hiller, J., Smith, C., Roberts, J., Prigodich, R., Wess, T.J., Csapo, J., Millard, A.R., Turner–Walker, G., 2002. The survival of organic matter in bone: a review. Archaeometry 44, 383–394.

Coltrain, J.B., Harris, J.M., Cerling, T.E., Ehleringer, J.R., Dearing, M.D., Ward, J., Allen, J., 2004. Rancho La Brea stable isotope biogeochemistry and its implications for the palaeoecology of late Pleistocene, coastal southern California. Palaeogeogr. Palaeoclimatol. Palaeoecol. 205, 199–219.

Driessens, F.C., Verbeeck, R.K., 1990. Biominerals. CRC Press, Boca Raton, FL.

Elliott, J.C., 2002. Calcium phosphate biominerals. Rev. Mineral. Geochem. 48, 427–453.

Fernández-Jalvo, Y., Andrews, P., 2016. Atlas of Taphonomic Identifications: 1001+ Images of Fossil and Recent Mammal Bone Modification. Springer.

Fernández-Jalvo, Y., Andrews, P., Pesquero, D., Smith, C., Marín-Monfort, D., Sánchez, B., Geigl, E.-M., Alonso, A., 2010. Early bone diagenesis in temperate environments: Part I: surface features and histology. Palaeogeogr. Palaeoclimatol. Palaeoecol. 288, 62–81.

Fleet, M.E., 2009. Infrared spectra of carbonate apatites: N2 -Region bands. Biomaterials 30, 1473–1481.

Fox-Dobbs, K., Bump, J.K., Peterson, R.O., Fox, D.L., Koch, P.L., 2007. Carnivore-specific stable isotope variables and variation in the foraging ecology of modern and ancient wolf populations: case studies from Isle Royale, Minnesota, and La Brea. Can. J. Zool. 85, 458–471.

Fox-Dobbs, K., Dundas, R.G., Trayler, R.B., Holroyd, P.A., 2014. Paleoecological implications of new megafaunal ^{14}C ages from the McKittrick tar seeps, California. J. Vertebr. Paleontol. 34, 220–223.

Fox-Dobbs, K., Stidham, T.A., Bowen, G.J., Emslie, S.D., Koch, P.L., 2006. Dietary controls on extinction versus survival among avian megafauna in the late Pleistocene. Geology 34, 685–688.

France, C., 2008. A Carbon and Nitrogen Isotopic Analysis of Pleistocene Food Webs in North America: Implications for Paleoecology and Extinction.

Fuller, B.T., Fahrni, S.M., Harris, J.M., Farrell, A.B., Coltrain, J.B., Gerhart, L.M., Ward, J. K., Taylor, R., Southon, J.R., 2014. Ultrafiltration for asphalt removal from bone collagen for radiocarbon dating and isotopic analysis of Pleistocene fauna at the tar pits of Rancho La Brea, Los Angeles, California. Quat. Geochronol. 22, 85–98.

Fuller, B.T., Harris, J.M., Farrell, A.B., Takeuchi, G., Southon, J.R., 2015. Sample preparation for radiocarbon dating and isotopic analysis of bone from Rancho La Brea. La Brea and beyond: the paleontology of asphalt-preserved biotas. In: Harris, J. M. (Ed.), Natural History Museum of Los Angeles County, Science Series, pp. 151–167.

Fuller, B.T., Southon, J.R., Fahrni, S.M., Farrell, A.B., Takeuchi, G.T., Nehlich, O., Guiry, E.J., Richards, M.P., Lindsey, E.L., Harris, J.M., 2020. Pleistocene paleoecology and feeding behavior of terrestrial vertebrates recorded in a pre-LGM asphaltic deposit at Rancho La Brea, California. Palaeogeogr. Palaeoclimatol. Palaeoecol. 537, 109383.

Guiry, E.J., Szpak, P., 2021. Improved quality control criteria for stable carbon and nitrogen isotope measurements of ancient bone collagen. J. Archaeol. Sci. 132, 105416.
Guiry, E.J., Szpak, P., 2020. Quality control for modern bone collagen stable carbon and nitrogen isotope measurements. Methods Ecol. Evol. 11, 1049–1060.
Hare, P.E., Fogel, M.L., Stafford Jr, T.W., Mitchell, A.D., Hoering, T.C., 1991. The isotopic composition of carbon and nitrogen in individual amino acids isolated from modern and fossil proteins. J. Archaeol. Sci. 18, 277–292.
Hassan, A.A., Termine, J.D., Haynes, C.V., 1977. Mineralogical studies on bone apatite and their implications for radiocarbon dating. Radiocarbon 19, 364–374.
Ho, T.Y., Marcus, L.F., Berger, R., 1969. Radiocarbon dating of petroleum-impregnated bone from tar pits at Rancho La Brea, California. Science 164, 1051–1052.
Koch, P., Behrensmeyer, A.K., Stott, A.W., Tuross, N., Evershed, R.P., Fogel, M.L., 1999. The effects of weathering on the stable isotope composition of bones. Anc. Biomol. 3, 117–134.
Koch, P.L., Tuross, N., Fogel, M.L., 1997. The effects of sample treatment and diagenesis on the isotopic integrity of carbonate in biogenic hydroxylapatite. J. Archaeol. Sci. 24, 417–429.
Lebon, M., Reiche, I., Gallet, X., Bellot-Gurlet, L., Zazzo, A., 2016. Rapid quantification of bone collagen content by ATR-FTIR spectroscopy. Radiocarbon 58, 131–145.
Liden, K., Takahashi, C., Nelson, D.E., 1995. The effects of lipids in stable carbon isotope analysis and the effects of NaOH treatment on the composition of extracted bone collagen. J. Archaeol. Sci. 22, 321–326.
Lindsey, E.L., Seymour, K.L., 2015. Tar Pits" of the western Neotropics: paleoecology, taphonomy, and mammalian biogeography. Nat. Hist. Mus. Los Angel. Cty. Sci. Ser. 42, 111–123.
Marcus, L.F., Berger, R., 1984. The significance of radiocarbon dates for Rancho La Brea. Quaternary Extinctions: a Prehistoric Revolution, pp. 159–183.
McDonald, H.G., Harris, J.M., Lindsey, E.L., 2015. Introduction. La Brea and beyond: the paleontology of asphalt-preserved biotas. In: Harris, J.M. (Ed.), Natural History Museum of Los Angeles County, Science Series, vol. 42, pp. 1–4.
Merriam, J.C., 1911. The Fauna of Rancho La Brea: Part 1: Occurrence, vol. 1. Memoirs of the University of California, pp. 197–213.
Merriam, J.C., Stock, C., 1921. Occurrence of Pleistocene vertebrates in an asphalt deposit near McKittrick, California. Science 54, 566–567.
O'Keefe, F.R., Dunn, R.E., Weitzel, E.M., Waters, M.R., Martinez, L.N., Binder, W.J., Southon, J.R., Cohen, J.E., Meachen, J.A., DeSantis, L.R.G., Kirby, M.E., Ghezzo, E., Coltrain, J.B., Fuller, B.T., Farrell, A.B., Takeuchi, G.T., MacDonald, G., Davis, E.B., Lindsey, Emily L., 2023. Pre–Younger Dryas megafaunal extirpation at Rancho La Brea linked to fire-driven state shift. Science 381, eabo3594. https://doi.org/10.1126/science.abo3594. New York, N.Y.
O'Keefe, F.R., Fet, E.V., Harris, J.M., 2009. Compilation, Calibration, and Synthesis of Faunal and Floral Radiocarbon Dates, Rancho La Brea, California.
Roche, D., Ségalen, L., Balan, E., Delattre, S., 2010. Preservation assessment of miocene–pliocene tooth enamel from tugen hills (Kenyan rift valley) through FTIR, chemical and stable-isotope analyses. J. Archaeol. Sci. 37, 1690–1699. https://doi.org/10.1016/j.jas.2010.01.029.
Schultz, J.R., 1938. A Late Quaternary Mammal Fauna from the Tar Seeps of McKittrick. Carnegie Instutution of Washington Publication, California, pp. 111–215.
Schwarcz, H.P., Nahal, H., 2021. Theoretical and observed C/N ratios in human bone collagen. J. Archaeol. Sci. 131, 105396.
Seymour, K.L., 2015. Perusing talara: overview of the late Pleistocene fossils from the tar seeps of Peru. Natural history Museum of Los Angeles county. Science 42, 97–109.
Solorzano, A., Rincon, A.D., McDonald, H.G., 2015. A new mammal assemblage from the Late Pleistocene El Breal de Orocual, northeast of Venezuela. La Brea and Beyond: The Paleontology of Asphalt-Preserved Biotas, Science Series. Natural History Museum of Los Angeles County, Los Angeles 125–150.
Sponheimer, M., Lee-Thorp, J.A., 1999. Alteration of enamel carbonate environments during fossilization. J. Archaeol. Sci. 26, 143–150.
Stafford Jr, T.W., Brendel, K., Duhamel, R.C., 1988. Radiocarbon, 13C and 15N analysis of fossil bone: removal of humates with XAD-2 resin. Geochem. Cosmochim. Acta 52, 2257–2267.
Sternberg, C.H., 1932. The Pleistocene fossil bed at McKittrick, California. In: Hunting Dinosaurs in the Bad Lands of the Red Deer River. NeWest Press, Alberta, Canada, pp. 214–221. Edmonton, Alberta, Canada.
Stock, C., 1928. A Peccary from the McKittrick Pleistocene, California.
Stock, C., Harris, J.M., 1992. Rancho La brea: a record of Pleistocene life in California. Natural History Museum of Los Angeles, 37, pp. 14–15.
Stoner, R.C., 1913. Recent Observations on the Mode of Accumulation of the Pleistocene Bone Deposits of Rancho La Brea. University of California Press.
Stuart, A.J., 1991. Mammalian extinctions in the late Pleistocene of northern eurasia and north America. Biol. Rev. 66, 453–562.
Szpak, P., 2011. Fish bone chemistry and ultrastructure: implications for taphonomy and stable isotope analysis. J. Archaeol. Sci. 38, 3358–3372.
Thompson, T., Gauthier, M., Islam, M., 2009. The application of a new method of Fourier Transform Infrared Spectroscopy to the analysis of burned bone. J. Archaeol. Sci. 36, 910–914.
Thompson, T., Islam, M., Bonniere, M., 2013. A new statistical approach for determining the crystallinity of heat-altered bone mineral from FTIR spectra. J. Archaeol. Sci. 40, 416–422.
Trayler, R.B., Landa, P.V., Kim, S.L., 2023. Evaluating the efficacy of collagen isolation using stable isotope analysis and infrared spectroscopy. J. Archaeol. Sci. 151, 105727.
Trueman, C.N., Behrensmeyer, A.K., Tuross, N., Weiner, S., 2004. Mineralogical and compositional changes in bones exposed on soil surfaces in Amboseli National Park, Kenya: diagenetic mechanisms and the role of sediment pore fluids. J. Archaeol. Sci. 31, 721–739.
Van Klinken, G.J., 1999. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. J. Archaeol. Sci. 26, 687–695.
Wyckoff, R.W., Wagner, E., Matter III, P., Doberenz, A.R., 1963. Collagen in fossil bone. Proc. Natl. Acad. Sci. USA 50, 215–218.