

Supplementary information to:

## Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne.

### Authors:

F. Welker<sup>a,b,1</sup>, M. Hajdinjak<sup>c</sup>, S. Talamo<sup>a</sup>, K. Jaouen<sup>a</sup>, M. Dannemann<sup>c,d</sup>, F. David<sup>e</sup>, M. Julien<sup>e</sup>, M. Meyer<sup>c</sup>, J. Kelso<sup>c</sup>, I. Barnes<sup>f</sup>, S. Brace<sup>f</sup>, P. Kamminga<sup>g</sup>, R. Fischer<sup>h</sup>, B. Kessler<sup>h</sup>, J.R. Stewart<sup>i</sup>, S. Pääbo<sup>c</sup>, M.J. Collins<sup>b</sup>, J.-J. Hublin<sup>a</sup>

<sup>a</sup>Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany.

<sup>b</sup>BioArCh, University of York, York, YO10 5DD, United Kingdom.

<sup>c</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany.

<sup>d</sup>Medical Faculty, University of Leipzig, 04103 Leipzig, Germany.

<sup>e</sup>Paris Unité Mixte de Recherche 7041 du Centre National de la Recherche Scientifique, Archéologies et Sciences de l'Antiquité, 92023 Nanterre, France.

<sup>f</sup>Department of Earth Sciences, Natural History Museum, London SW7 5BD, United Kingdom.

<sup>g</sup>Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA Leiden, The Netherlands.

<sup>h</sup>Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, United Kingdom.

<sup>i</sup>Faculty of Science and Technology, Bournemouth University, Talbot Campus, Fern Barrow, Poole, Dorset, BH12 5BB, United Kingdom.

<sup>1</sup>To whom correspondence should be addressed. Email:

[frido\\_welker@eva.mpg.de](mailto:frido_welker@eva.mpg.de)/[frido.welker@palaeo.eu](mailto:frido.welker@palaeo.eu)

Materials and Methods	2
Section 1: ZooMS screening	7
Section 2: Morphological description	9
Section 3: Palaeoproteomics	13
Section 4: Archaic ancestry of rs142463796 in <i>COL10A1</i>	20
Section 5: Ancient DNA	22
Section 6: Radiocarbon dating and Bayesian modelling of AR-14	24
Section 7: Stable isotope analysis and associated fauna	26
Figures, Tables and Bayesian CQL Code	29
References	69

## **Materials and Methods:**

### ***The Grotte du Renne: Research history and chronostratigraphy***

Excavated between, primarily, 1948 and 1966 (1, 2), the Grotte du Renne (Arcy-sur-Cure) has gained prominence for a suggested direct attribution of “advanced” cultural artefacts found within the Châtelperronian layers, artefacts of a type often associated solely with AMHs, and hominins morphologically identified as Neandertals in the same layers (3, 4). As discussed in the main text, various hypotheses have been proposed to explain this association. Here we specifically test the hypothesis that the hominin remains in the Châtelperronian layers at the Grotte du Renne are Neandertals but derive from the underlying Mousterian (hypothesis **d** in the main text; (5, 6)).

The stratigraphy of the Grotte du Renne as described originally (1, 2) contains 3 Châtelperronian layers (VIII, IX and X), overlying various Mousterian layers (of which layer XI directly underlies the Châtelperronian layer X). The uppermost Châtelperronian layer is capped by a Proto-Aurignacian layer (VII) and various additional Aurignacian and Gravettian layers. The general stratigraphy has been verified and confirmed more recently by a small-scale excavation (7). Complex taphonomic processes are common to cave deposits, and the Grotte du Renne is no exception. The front rows (1-6) display extensive sloping (6, 8), while digging and/or levelling activities have disturbed part of the Mousterian and Châtelperronian deposits in squares X14 to B11 (6, 9, 10). It is therefore not inconceivable that some mixing of material has occurred, as demonstrated by lithic refits between different Châtelperronian layers and the presence of outlying radiocarbon dates in constructed Bayesian models (5, 8, 11). In relation to hypothesis **d**, the question is to what extent this has affected the hominin remains at the site. Direct radiocarbon dating of hominin specimens is capable to test such a hypothesis, and has been applied elsewhere to other hominin specimens or specific artefacts to answer similar questions (12–15).

Lithic refits are present between layers IX and X (8, 16), and therefore these archaeological layers are currently combined into a single phase for Bayesian modelling (8, 11). From the available Bayesian models, it seems there is additional evidence for the movement of specimens between the Châtelperronian layer VIII and the combined Châtelperronian layers IX+X (for example dates EVA-29, EVA-56 and possibly dates OxA-21577, OxA-21591 and OxA-21593). For the Bayesian modelling conducted here, we keep the phasing separate for layer VIII and layers IX+X.

### ***The Grotte du Renne sample selection and ZooMS COL1 screening***

Bone specimen selection was based on two criteria. First, we selected 69 bone specimens that were put aside during the Leroi-Gourhan excavations as possible hominin fragments (1, 2). All these specimens come from squares C7 and C8, Châtelperronian Layers Xb2 and X(c?). These bone specimens remained unanalyzed after their discovery and preserve few to no diagnostic criteria to substantiate a direct identification as hominin. Their post-excavation history seems to have been separate from other bone specimens from the site. Second, 127 bone specimens were selected based on the spatial association with Neandertal teeth and the temporal bone (17), hominin-modified artefacts (18) or a combination thereof. These latter specimens come from squares A7, A11, B7, B10, B11, C8, C9, C10, C11, C12, D8, D9, D10, Y11, Z13, in Châtelperronian Layers IX, Xb and Xc.

ZooMS screening (Fig. S1) was performed on bone samples weighing between 0.9 and 64.9 mg using a non-destructive ammonium-bicarbonate buffer protocol (19) to screen for the presence of diagnostic peptide markers B and D, the mass values of which are guiding in the identification of a potential hominin (20–23). In the case of a positive or unidentifiable result, the remaining bone sample was demineralised in 200 µl 0.5 M HCl. After demineralisation, samples were centrifuged (1 minute at 13k rpm) and the acid supernatant, containing the acid-soluble protein fraction, was stored separately before resuspending the demineralised sample in 100 µl ammonium-bicarbonate buffer. Further extraction steps and analysis followed protocols outlined elsewhere (24). A blank was included during all stages of analysis to monitor the introduction of exogenous, contaminating, proteins.

For positive hominin identifications, Qubit fluorimetry (Life Sciences) was performed on selected samples to assess protein concentration in the analysed extracts. Deamidation calculations were performed on ZooMS spectra obtained through ammonium-bicarbonate buffer extraction to assess the exogenous/endogenous origin of the extracted collagen using the code described by (25) and compared to values obtained in a similar manner for ammonium-bicarbonate COL1 database extracts (see Collagen Type I database section).

### ***Palaeoproteomic extraction***

An additional sample of 83.1 mg of AR-30 was powdered using a sterile micropestle. The sample was incubated at room temperature for one day in 0.5M EDTA to demineralise, centrifuged and resuspended in ammonium-bicarbonate buffer (50 mM; Sigma). The solution was gelatinized at 65°C for 1 hour, centrifuged, and the pellet resuspended in GuHCl/Tris-HCl (6.0 M; Sigma and Life Technologies). The ammonium-bicarbonate fraction is not analyzed here. After homogenization, the solution was stored at 4°C for one day to further gelatinize. After protein extraction, cysteines were reduced (5 mM DTT at 60°C for 1 hour; Sigma) and alkylated (15 mM Chloroacetamide for 45 minutes in the dark; Sigma). Proteins were digested using trypsin (Promega) overnight at 37°C. Digestion was terminated by the addition of trifluoroacetic acid (TFA; Sigma). Peptide elution and purification took place on C18 Stage Tips (Thermo Scientific), with elution conditions similar to those reported elsewhere (26). The pH was checked throughout and adjusted where needed. Blanks were included throughout all stages to monitor potential contamination, and included in subsequent LC-MS/MS analysis. Peptide concentrations were determined using Qubit (Life Technologies) following manufacturers protocol. The palaeoproteomic extract from AR-30 was analyzed independently twice in different LC-MS/MS runs (hereafter named AR-30A and AR-30B).

### ***LC-MS/MS analysis***

LC-MS/MS analysis was performed on three ZooMS extracts (samples AR-7, AR-16 and AR-30) and two additional extracts from AR-30 (AR-30A and AR-30B). These extracts were selected based on their ZooMS protein concentration as determined by Qubit analysis (AR-7 and AR-16), or morphologically non-informative status that would also allow ancient DNA analysis (AR-30). Peptide extracts were dried down, resuspended and analyzed following Quadrupole-Orbitrap details presented in (27). All LC-MS/MS runs were preceded by blanks to avoid the possibility of protein carryover between runs.

.raw files were converted to mgf files using ProteoWizard 3.0 (28) and merged for bone specimen AR-7 using MergeMGF (ProteinMetrics). Next, the obtained data was searched in a four-step process. First, .mgf files were searched against a custom collagen type I database (see

below) built *de novo* for this project and containing COL1 sequences and common lab contaminants for all relevant genera to verify the ZooMS MALDI-TOF-MS taxonomic identification. Second, .mgf files were searched against the reviewed human Swiss-Prot database (20,206 entries) and common lab contaminants (115 entries). Third, .mgf files were searched against the reviewed human Swiss-Prot database with the addition of archaic-specific substitutions present for human proteins identified in the second search. These three searches were all performed in PEAKS (29). Fourth, .mgf files were searched on Mascot against the complete UniProt database to estimate the presence of non-vertebrate proteins in the analyzed datasets. Significant spectral hits from this search were then removed from analyses of the other three searches.

In the first three cases, files were searched using PEAKS 7 (*de novo* to SPIDER) using a precursor mass tolerance of 10.0 ppm and fragment ion tolerance of 0.07 Da. Precursor masses were automatically corrected by PEAKS to enable accurate detection of possibly deamidated glutamine and asparagine residues. For the first search, variable modifications were set for hydroxylation (P), oxidation (M) and deamidation (NQ). For samples AR-30A and AR-30B, carbamidomethylation (C) was set as an additional fixed modification. In each search, the maximum number of PTMs was set at 5, with a maximum number of 2 missed cleavages allowed at both sides of the peptide. Spectral matches were accepted with a False Discovery Rate (FDR) equal to 1.0%, protein matches were accepted with a cumulative  $-10\lg P \geq 20.0$  and 1 unique peptide in two different datasets or 2 unique peptides in one dataset. *De novo* only hits were inspected manually when  $ALC\% \geq 50.0$ . For the fourth search, Mascot parameters were similar to those in PEAKS. Spectral matches were accepted with a Mascot ion score  $\geq 30$  and a significance threshold of  $p < 0.05$ . The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD003208.

### ***Collagen type I database***

To facilitate reliable taxonomic identification of Pleistocene bone and dental specimens we constructed a reference collagen type I database composed of species present in the Late Pleistocene of western Eurasia, including members of all present Mammalian genera that are medium- or larger-sized (30–35). Some of the used specimens have been analyzed genetically before (36, 37). The only genera absent from this database are *Homotherium* and *Stephanorhinus*, both of which make sporadic appearances in the faunal record of the region during MIS4/3 (38, 39). Instead, we included samples from four extant Rhinocerotidae species (*Diceros bicornis*, *Rhinoceros sondaicus*, *Rhinoceros unicornis*, *Dicerorhinus sumatrensis*), while the COL1 sequence of the fifth extant rhino species (*Ceratotherium simum*) is available through genomic sources (Dataset S2). We included several species of Cervidae not present in our study region or chronological period as previous ZooMS studies reported the absence of discriminating peptide marker masses for some clades within this family (24, 40). In addition and to be complete, we include additional species for which ZooMS peptide markers have been published or for which COL1 sequences are available through genetic sources.

Skin, bone, dental and antler specimens of 41 species (Dataset S2) were treated and analysed similarly to the Grotte du Renne bone specimens to extract COL1. Obtained peptide mass fingerprints were compared to published peptide marker libraries (20, 41–43).

Next, COL1 sequences were obtained using *de novo* peptide sequencing for a subset of the analyzed species (n=27). Analytical procedures followed those detailed in (27). In short,

extracts previously analysed by MALDI-TOF-MS were analyzed using LC-MS/MS. Obtained .raw data were converted to .mgf files using ProteoWizard 3.0 (28). Data was searched against online available and previously published COL1 sequences (27) using PEAKS v.7 (29) with a FDR of 0.5% (other settings similar to those reported above for LC-MS/MS analysis). Leucines (L) were converted into isoleucines (I) as these two amino acids are isobaric and cannot be separated based on mass alone.

On average, we obtained 91.8% coverage for the COL1 sequences for the analyzed species. These include the first available COL1 sequences for the Cervidae, a group which is difficult to separate based on ZooMS peptide marker series (24, 40), as well as a large number of carnivores. COL1 sequences include examples where increased taxonomic specificity can be achieved using LC-MS/MS analysis (amongst cervids, caballine versus non-caballine horses and differentiation among Felidae), as well as examples where COL1 sequences are identical between closely related species (for example within the genus *Ursus*).

The obtained COL1 sequences, in addition to those obtained from online genetic sources (44, 45), as well as those published before (27, 46), comprise a total of 127 species in 105 genera. For those species for which ZooMS peptide marker masses are present as well, both COL1 amino acid sequences and observed peptide masses agree with each other. Peptide marker masses for species with only COL1 sequences available were predicted using the amino acid sequence and inferred (MALDI-TOF-MS) or directly observed (LC-MS/MS) hydroxylation frequency for closely related species. We did not predict peptide markers C, P2 and E for Leporidae and Rodentia due to several inconsistencies between amino acid sequences available online through genetic sources and MALDI-TOF-MS peptide marker masses published previously. Taxonomic identification of these taxa can be achieved through observation of additional peptide markers for which such conflicts were absent. No peptide marker masses were predicted for Chiroptera as no COL1 MALDI-TOF-MS spectra or LC-MS/MS datasets are available for this order.

This database facilitates confident ZooMS identification of MALDI-TOF-MS COL1 spectra, as well as subsequent verification of such identifications using LC-MS/MS. The database can easily be expanded to cover contemporaneous faunal communities in adjoining regions or chronologically preceding Early and Middle Pleistocene faunal communities in western Eurasia (23). The COL1 mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD003190. Generated COL1 consensus sequences are attached in Dataset S1, while accession data, coverage and peptide marker masses used in this study can be found in Dataset S2.

### ***Ancient DNA extraction and library preparation***

Samples of 8.2 mg and 22 mg were removed from AR-14 and AR-30, respectively, using a sterile dentistry drill. Prior to DNA extraction, the bone powder was treated with phosphate buffer (47) to remove some of the microbial and modern human DNA typically present in ancient specimens. Two DNA extracts (E3293 and E3295) were prepared using a silica-based method (48) in the implementation of (47). Ten  $\mu\text{L}$  (20% of the total volume) of each extract were converted into DNA libraries (A9328 and A9330) using a single-stranded DNA library preparation method (49) as modified elsewhere (47). Extraction and library negative controls were included in the experiments. The number of DNA molecules in each library was determined by digital droplet PCR (QX200 system, Bio-Rad), using 1  $\mu\text{L}$  of a 5,000-fold library dilution in EB buffer (10 mM Tris-HCl pH 8.0, 0.05% Tween 20) as a template in an Eva Green

assay (Bio-Rad) with primers IS7 and IS8 (50). Each library was amplified with AccuPrime Pfx DNA polymerase (Life Technologies) as described (51) but using indexed primers for both adapters (52) at a final concentration of 1  $\mu$ M in a reaction volume of 100  $\mu$ l. Fifty  $\mu$ L of each amplified library were purified using the MinElute PCR purification kit (Qiagen) and eluted in 30  $\mu$ L TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA concentrations of the purified libraries were determined using a NanoDrop 1000 Spectrophotometer.

### ***Mitochondrial DNA (mtDNA) captures and sequencing***

Amplified libraries were enriched for human mitochondrial DNA (mtDNA) using a bead-based hybridization method (53). Capture baits covering the mtDNA genome in 1bp tiling were designed using the revised Cambridge Reference Sequence (rCRS, NC\_012920) and cleaved from a microarray as described in (53). Enriched libraries were pooled and sequenced on Illumina's MiSeq platform in a double index configuration (2 x 76 cycles)(52).

Overlapping paired-end reads were merged into single sequences (54) and mapped to the rCRS using the Burrows-Wheeler Aligner (BWA; (55)). BWA parameters were adjusted for ancient DNA sequences (“-n 0.01 -o 2 -l 16500”) by allowing more mismatches and indels and turning off the seeding(56). Only sequences with index reads that perfectly matched one of the expected index combinations were retained for further analysis. PCR duplicates were removed by calling a consensus from sequences with identical alignment start and end positions using bam-rmdup (<https://github.com/udo-stenzel/biohazard>). We restricted further analyses to sequences mapped to the rCRS that were 35bp or longer and had a mapping quality of at least 30.

### ***Collagen extraction for isotopic analysis***

Bone collagen was extracted from 21 faunal bone specimens and 1 hominin fragment (AR-14) at the Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Leipzig, Germany, using the method described in (57). All faunal specimens come from Layers IX and X at the Grotte du Renne. In short, the outer surface of the bone is cleaned by a shot blaster and  $\approx$ 500mg of the bone is taken and subsequently decalcified in 0.5M HCl at room temperature until no CO<sub>2</sub> effervescence is observed. 0.1M NaOH is added for 30 minutes to remove humics and followed by a final 0.5M HCl step for 15 minutes. The resulting solid is gelatinized following Longin (1971) at pH3 in a heater block at 75°C for 20h. The gelatin is then filtered in an Eeze-Filter™ (Elkay Laboratory Products (UK) Ltd.), to remove small (<80  $\mu$ m) particles, and ultrafiltered with Sartorius “Vivaspin Turbo” 30 KDa ultrafilters. Finally, the samples are lyophilized for 48 hours.

To identify the preservation of the collagen, C:N ratios, %C, %N, collagen yield and  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values must be evaluated. The C:N ratio should be between 2.9 and 3.6 and the collagen yield not less than 1% of weight (58, 59). Stable isotope analysis was conducted using a ThermoFinnigan Flash EA coupled to a Delta V isotope ratio mass spectrometer.  $\delta^{13}\text{C}$  values are measured relative to the V-PDB standard and  $\delta^{15}\text{N}$  values are measured relative to the AIR standard.

For AR-14, 6.8 mg of the collagen was weighed into a pre-cleaned tin capsule and sent to the Mannheim AMS laboratory (Lab Code MAMS), where it was graphitized and dated (60). The date was corrected for a residual preparation background estimated from pretreated <sup>14</sup>C free bone samples, kindly provided by the MAMS and pretreated in the same way as the archaeological samples.

## **1. ZooMS screening**

*Frido Welker\**, Matthew J. Collins, Ian Barnes, Selina Brace, John R. Stewart, Pepijn Kamminga, Francine David, Michèle Julien, Jean-Jacques Hublin

\*to whom correspondence should be addressed:

[frido\\_welker@eva.mpg.de](mailto:frido_welker@eva.mpg.de)/[frido.welker@palaeo.eu](mailto:frido.welker@palaeo.eu)

We screened 196 bone specimens using a stepwise workflow for the presence of possible hominin bone specimens (Fig. S1). Ammonium-bicarbonate spectra indicative of hominine identifications were verified with subsequent demineralization of the same bone sample. In all cases such follow-up analysis confirmed initial ZooMS identifications (Fig. S2). As the workflow aimed to identify hominins, we did not follow-up on incomplete ZooMS peptide marker series obtained through ammonium-bicarbonate screening when these spectra allowed excluding an identification as hominin. We note that ammonium-bicarbonate spectra of some bone specimens, but not all, contain repetitive peptides in the mass range up to  $\approx 1200$  m/z (Fig. S2a). They likely result from the fractionation of a non-organic consolidate or preservative. Mass values associated to this series did not interfere with peptide marker identifications and were not present in follow-up acid demineralised spectra. In addition, no identifications were made solely based on peptide markers P1 and A, for which some peptide marker masses fall within the range where these repetitive peptides were present.

Following this approach, Hominoidea marker B (1477.8 m/z) and the presence of Homininae-informative peptide markers D (2115.1 m/z) and F (2832.4 m/z; Tab.S1), 28 bone specimens were identified as hominines (Fig. S10). All these specimens come from squares C7 and C8, the Grotte du Renne Layers Xb and X(c?). Based on ZooMS peptide markers, the correct taxonomic assignment of these specimens is as Homininae (extant genera *Pan* + *Homo*; (61)). This would not have been possible without an extensive peptide marker database including COL1 sequences from most medium and large-sized mammals present in western Eurasia during the Late Pleistocene, as well as the incorporation of COL1 sequences available through genetic sources of closely related species (eg. Hominidae). Furthermore, it facilitates verification of ZooMS identifications by subsequent LC-MS/MS analysis (see below).

We assessed the presence of possible COL1 contamination from modern sources by obtaining deamidation values for two peptides (P1105 and P1706) (62). Deamidation values were calculated for all COL1 database bone specimens extracted using ammonium-bicarbonate buffer, all faunal specimen studied for Arcy, as well as all specimens identified as hominins, following (25). Focusing on peptides P1105 and P1706 (Fig. 1A; (25)), we identify that all modern and permafrost samples are not to barely deamidated (values of 0.0 or close to 0.0%), indicating that our ammonium-bicarbonate buffer extraction did not induce glutamine deamidation. This is in agreement with observations made elsewhere using the same extraction protocol (62, 63). It also corresponds with previous observations that deamidation is dependent on local preservation conditions rather than age, given the chronological difference between our modern bone specimens and the permafrost samples (which are around 10 ka old, Fig. 1A; (64)). All studied reference specimens of Pleistocene age display values around or slightly above 50.0%. Deamidation values for the Arcy specimens are comparable between faunal and hominin specimens, with no significant differences between the two (t-tests; P1105:  $p=0.85$ ; P1706:  $p=0.55$ ). The values are slightly below those for the included Pleistocene reference samples, but comparable to the values reported for Les Cottés previously (62). Conversely, deamidation data

for the Arcy hominins is significantly different from that obtained for modern bone specimens (t-test;  $p=2.31E-11$ ).

Database searching against the COL1 database using LC-MS/MS ZooMS extract data for AR-7, AR-16 and AR-30 verified the identification of the analyzed extracts as hominines, with hits to *Homo sapiens* having the largest number of spectrum-peptide hits. The number of unique spectrum-peptide hits could not be used a criterion as there are no *Homo*-specific substitution in either COL1 chain (Tab. S1). In addition to spectral matches covering positions *COL1 $\alpha$ 1* 784S and *COL1 $\alpha$ 1* 1194A, positive spectral matches to *COL1 $\alpha$ 2* 420P firmly identify these bone specimens as hominins (e.g., on the Homo side of the Homo-Pan split; Tab. S1). ZooMS extracts analyzed by LC-MS/MS were selected based on their ZooMS protein concentration as determined by Qubit analysis (AR-7 and AR-16), or morphologically non-informative status that would also allow aDNA analysis (AR-30).

## **2. Morphological description**

*Jean-Jacques Hublin\**, *Frido Welker*

\*to whom correspondence should be addressed: [hublin@eva.mpg.de](mailto:hublin@eva.mpg.de)

A total of 28 bone specimens were identified as homininae (see SI Appendix Section 1). These specimens are generally very small with an average size of 17.7 mm ( $\pm 6.0$ ) by 13.4 mm ( $\pm 4.6$ ). All these specimens are from squares C7 and C8, outside the sloping section of the cave stratigraphy and outside the area supposedly affected by digging/levelling activities (6). They derive from the Châtelperronian layers Xb2 and X(c?), and have associated stratigraphic and contextual information written on them with ink (similar to other bone specimens excavated from the Grotte du Renne).

Below, a short morphological description of these 28 bone specimens follows. Due to the fragmentary nature of the specimens studied here, only in a small number of instances is a specimen attributed to an element or successfully sided (Fig. S10).

### *AR-3*

Square C8; Layer Xb2.

AR-3 appears to be a portion of the right zygomatic arch. The specimen preserves parts of the frontal, maxillary and temporal processes. The zygomaticofacial foramen is present, while the masseteric origin is not preserved. The initial suggestion that this specimen represents a hominin was confirmed by ZooMS. The specimen measures 26.0 x 14.2 mm.

### *AR-5*

Square C8; Layer Xb2.

AR-5 is a cervical vertebral element composed of an inferior articular surface and a portion of the pedicle and circumference of the foramen. Maximum dimensions are 13.2 x 8.8 mm.

### *AR-6*

Square C8; Layer Xb2.

AR-6 is a square fragment measuring 11.7 x 11.4 mm with thin cortical bone on both surfaces.

### *AR-7*

Square C8; Layer Xb2.

AR-7 comprises a left hemi-arch of a cervical vertebra. Anterior, both the lower and upper articular facets are partly preserved with trabecular exposure of both facets laterally. Part of the foramen transversarium is present although closure cannot be assessed as both the posterior and anterior bars are not preserved. Further, the neurocentral junction is not preserved either, preventing assessment of the closure with the centrum. Posterior, there appear two unfused surfaces, one directed medially and the other laterally. The absence of fusion of the neural arch suggests it represents an individual younger than 2 years (65). The relative size and shape exclude attribution to an axis or atlas (66). The trabecular exposure on the lateral side of both articular facets and the absence of a preserved transverse process inhibits the assessment of this specimen to a more specific typical cervical vertebra (65, 67). Width and thickness of the lamina were measured following (68) and compared to the Neandertal La Ferrassie 8 C3/C4 left hemi-arch, C3/C4/C5 right hemi-arch and C6 left hemi-arch and found to be consistent with similar dimensions (Tab. S13; (67)).

A sample of 10.1 mg was taken for initial ZooMS screening and subsequently analyzed by LC-MS/MS three times.

*AR-8*

Square C8; Layer Xb2.

AR-8 is a rectangular fragment measuring 9.8 x 9.0 mm with thin cortical bone on both surfaces.

*AR-9*

Square C8; Layer Xb2.

AR-9 is an irregular specimen with a U-shaped cross-section and a large open space in its center. The specimen retains cortical bone on the outer surface of the cross-section but is broken on both ends. An element could not be assigned to this specimen.

*AR-11*

Square C8; Layer Xb2.

AR-11 has an unclear shape but preserves portions of cortical bone. The maximum dimensions are 13.5 x 9.7 mm.

*AR-12*

Square C8; Layer Xb2.

AR-12 is a rectangular fragment measuring 13.8 x 12.8 mm with thin cortical bone on both surfaces.

*AR-13*

Square C7; Layer X(c?).

AR-13 is a rectangular fragment measuring 13.9 x 11.4 mm with thin cortical bone on both surfaces. The specimen preserves a suture on 1 side. The closed or unclosed status of this suture is unclear.

*AR-14*

Square C7; Layer X(c?).

AR-14 is a relatively flat fragment measuring 21.1 x 20.2 mm with cortical bone on both surfaces and a possible suture visible on the interior surface. This specimen was subsequently used for direct radiocarbon dating, stable isotope analysis and ancient DNA analysis.

*AR-16*

Square C7; Layer X(c?).

AR-16 is a partial right maxilla preserving part of the infraorbital sulcus, the roof of the maxillary sinus and the infraorbital foramen. The suture to the zygomatic process is partly present as well. The maximum dimensions are 36.0 x 16.6 mm. The initial suggestion that this specimen represents a hominin was confirmed by ZooMS.

A sample of 31.2 mg was taken for initial ZooMS screening and subsequently analyzed by LC-MS/MS once.

*AR-29*

Square C7; Layer X(c?).

AR-29 is a slightly curving square piece with thin cortical bones on both sides. The maximum dimensions are 30.2 x 30 mm.

*AR-30*

Square C7; Layer X(c?).

AR-30 is an irregularly shaped, slightly curved fragment with thin cortical bone on both sides measuring 23.6 x 18.0 mm. The overall shape suggests a cranial origin. This specimen was selected for ancient DNA and palaeoproteomic analysis.

A sample of 6.1 mg was taken for initial ZooMS screening and subsequently analyzed by LC-MS/MS once. Following, an additional extraction was performed on this specimen and analyzed twice (AR-30A and AR-30B).

*AR-31*

Square C7; Layer X(c?).

This specimen resembles AR-29, the maximum dimensions are 24.5 x 16.4 mm.

*AR-33*

Square C7; Layer X(c?).

AR-33 has thin cortical bone on both surfaces and displays a slight curve. The maximum dimensions for this specimen are 21.0 x 13.3 mm.

*AR-34*

Square C7; Layer X(c?).

AR-34 is a slightly elongated fragment with a curved cross-section and cortical bone on both sides. The specimen is broken on both ends and might represent the neck portion of an otherwise unspecified rib. Maximum dimensions are 12.9 x 12.4 mm.

*AR-35*

Square C7; Layer X(c?).

The overall shape, including the presence of cortical bone on the majority of the surfaces, suggests this specimen preserves part of the onset of the spinal process of an unidentified vertebra. The specimen measures 12.4 x 9.4 mm.

*AR-163*

Square C7; Layer X(c?).

This small (16.1 x 13.8 mm) irregular specimen preserved a possible unfused suture on one side.

*AR-164*

Square C7; Layer X(c?).

This specimen is slightly curved and irregular. Maximum dimensions are 15.1 x 14.4 mm.

*AR-165*

Square C7; Layer X(c?).

Specimen AR-165 preserved an unfused suture. The overall shape of the fragment is irregular, and measures 20.9 x 16.3 mm.

*AR-167*

Square C7; Layer X(c?).

Specimen AR-167 preserves thin cortical bone on both surfaces. Maximum dimensions are 14.7 x 13.0 mm.

*AR-169*

Square C7; Layer X(c?).

Specimen AR-169 preserves thin cortical bone on both surfaces. Maximum dimensions are 14.6 x 9.7 mm.

*AR-170*

Square C7; Layer X(c?).

Specimen AR-170 preserves thin cortical bone on both surfaces. Maximum dimensions are 18.7 x 17.6 mm.

*AR-171*

Square C7; Layer X(c?).

Specimen AR-171 preserves thin cortical bone on both surfaces. Maximum dimensions are 16.1 x 11.5 mm.

*AR-172*

Square C7; Layer X(c?).

Specimen AR-172 preserves thin cortical bone on both surfaces. Maximum dimensions are 17.7 x 9.8 mm.

*AR-175*

Square C7; Layer X(c?).

AR-175 is an irregular bone specimen with cortical bone preserved on one side and trabecular bone exposed on the opposing side. The specimen measures 15.2 x 9.0 mm.

*AR-176*

Square C7; Layer X(c?).

Specimen AR-176 preserves thin cortical bone on both surfaces. Maximum dimensions are 13.6 x 9.1 mm.

*AR-177*

Square C7; Layer X(c?).

Specimen AR-177 preserves thin cortical bone on both surfaces. Maximum dimensions are 16.2 x 9.9 mm.

### **3. Palaeoproteomics**

*Frido Welker\**, R. Fischer, B. Kessler, Matthew J. Collins

\*to whom correspondence should be addressed:

[frido\\_welker@eva.mpg.de](mailto:frido_welker@eva.mpg.de)/[frido.welker@palaeo.eu](mailto:frido.welker@palaeo.eu)

LC-MS/MS analysis of ZooMS extracts of AR-7, AR-16 and AR-30, as well as additional palaeoproteomic data generated for AR-30 (A and B) resulted in the identification of 73 proteins (Tabs. 1, S2). All three ZooMS extracts contain proteins other than COL1, including various non-collagenous proteins (NCPs; Tab. 1). The main difference between the ZooMS extracts and the two palaeoproteomic extracts (AR-30A and AR-30B) is in the number of protein groups identified, especially for NCPs (Tab. 1).

Of the identified proteins, 10 are known protein contaminants (human skin keratins, trypsin, bovine albumin and beta-casein). Some of these proteins were also present in the extraction blank (bovine trypsin, human skin keratins, trypsin), which also contained several proteins not present in other extracts (*Histone H4*, *Histone H2A.V*). Searching against the entire UniProt database using Mascot resulted in the detection of a limited number of significant spectral matches against non-vertebrate proteins. These consist of bacterial, fungi and other Eukaryota proteins (Fig. S4), but other than the blank, always comprise <1% of the total number of spectral matches. These spectra were removed from subsequent analysis. For the extraction blank, a total of 11 spectra were matched by Mascot, with three of those to Fungi and eight to vertebrate proteins.

Deamidation spectral count analysis of LC-MS/MS spectra indicate elevated frequencies for 35 proteins, while for 24 proteins glutamine/asparagine spectral data was absent (Figs. 1B, S5; Tab. S2). Some of the proteins for which deamidation frequency could not be obtained have been reported previously in palaeoproteomic analysis of non-hominin bone specimens (*POSTN*, *THBS1*, *ACTB*, *C3*, *IGHG1*, *NUCB2*; (26, 69, 70)) or are supposed to have a role in the bone ossification process (*DLL3*; (71)).

Next to the cellular-specific or development-specific proteins mentioned in the main text (*COL10a1*, *COL27a1*, *DLL3*, *COL24a1*), there are numerous spectral matches to collagenous propeptides. Propeptides for *COL1a1* and *COL1a2*, for example, are cleaved off before triple-helix formation and therefore before bone mineralization. Together with an overall enrichment for proteins involved in cartilage and bone developmental processes (Tab. S5), there are numerous indications of an active tissue in the process of formation and mineralization. Further, the presence of *COL10a1* indicates that at least part of the sampled bone tissue has not yet undergone remodelling, consistent with osteological (SI Appendix Section 2) and isotopic (SI Appendix Section 7) evidence.

Below, specific aspects such occurrence and frequency of post-translational modifications (PTMs), GO annotation analysis, and amino acid sequence analysis are presented.

#### ***Post-translational modifications***

Post-translational modifications (PTMs) in ancient proteins have received increased attention in recent years. The majority focuses on glutamine (and asparagine) deamidation (62, 72, 73), while others focus on the detection of hitherto unseen PTMs in ancient proteins (74). The study of such modifications attempts to understand protein diagenesis and degradation mechanisms, while a recent study also highlights the potential of identifying *in vivo* PTMs preserved in ancient bone specimens (75).

PEAKS SPIDER analysis allows for the detection of additional PTMs that were not specified manually. Below, an overview of common classes of PTMs is given, excluding hydroxylation because of its ubiquitous presence in collagens.

### Deamidation and the identification of endogenous proteins

Collagen deamidation has received considerable attention in palaeoproteomic studies as a suggested marker of non-enzymatic, diagenetic, alteration (62, 64, 72, 76). Recently, it was reported that non-collagenous bone proteins (NCPs) in a Pleistocene bone specimen had complete glutamine deamidation (74) in contrast to collagen glutamine deamidation from the same specimen.

Here we take advantage of this observation and the presence of known contaminants in palaeoproteomic analysis (e.g. Trypsin, human skin keratins) to demonstrate that the spectral counts obtained for the identified proteins and their associated (non)deamidated status cluster in 3 major groups (Figs. 1B, S5). These clusters are based on unsupervised clustering analysis through the R package Mclust (77) for protein groups containing 3 or more asparagine/glutamine positions. Analysis of glutamine and asparagine positions separately resulted in similar membership assignment (Fig. S5). The presence of known contaminants within our extracts allows assessment of deamidation due to extraction and analysis procedures, as these contaminants are either present before protein extraction (keratins) or introduced during protein extraction (trypsin). In addition, it implies that no a priori assumptions are made on expected deamidation frequencies for endogenous or exogenous proteins. We explicitly assume that all spectral matches to a protein come from a single source, which can either be endogenous or exogenous to the analyzed bone specimens. We realize that this is a limitation of our approach that needs to be addressed in future work.

- Cluster 1 (Figs. 1B, S5; triangles): Includes (all identified) known contaminants containing no or few deamidated glutamines and asparagines. Among others, these include keratins, *DCD* and several non-hominin proteins (trypsin, bovine albumin and bovine *CSN2*). All proteins identified in our extraction blank with sufficient glutamine and/or asparagine positions fall within this cluster. The presence of *COL4a6* as the only collagen member of this cluster can be explained as the *COL4* heterotrimer it takes part in, is present in skin basement membranes (78, 79). This is in contrast to the other *COL4* members, some of which are more ubiquitously distributed throughout basement membranes (78), including basement membranes located in vascularized areas of mineralized tissues. The low frequency of deamidation for known contaminants suggests protein deamidation during extraction has not significantly influenced overall protein deamidation. We acknowledge that exceptionally well-preserved proteins, or proteins that undergo different diagenetic pathways, might erroneously be included in this cluster (64). We see no reason to suspect the presence of such “exceptional” proteins in our analysis for AR-7, AR-16 and AR-30, however, and cautiously interpret all proteins of this cluster to represent contaminants.
- Cluster 2 (Figs. 1B, S5; squares): This cluster is subdivided in two subclusters in all three analyses. The two subclusters both include fibrillar collagens, network forming collagens, and FACITs. Several collagen gene products form triple-helices together but are separated over both sub-clusters, for example *COL4a3*, *COL4a4* and *COL4a5* (Figs. 1B, S5; (78)). Components include most collagens (except *COL4a6*, see above, and

*COL26a1* and *COL8a1*, see below) and two NCPs (lumican and human albumin).

Deamidation frequencies for proteins included in this cluster range between 30-75%, well differentiated from both cluster 1 and cluster 3. There are no known contaminants present in this cluster. The two NCPs present in this cluster have the highest deamidation frequencies of the included proteins.

-Cluster 3 (Fig. 1B, S5; filled circles): Includes NCPs and collagens (*COL26a1*, *COL8a1*) of which all determined glutamine/asparagine positions are deamidated. There are no known contaminants present in this cluster. Proteins included in this cluster have also been identified in previous palaeoproteomic studies on non-hominin bone specimens (*CLEC3B*, *OMD*, *CHAD*, *AHSG*, *BGN*). To this group can potentially be added several proteins with one or two determined and deamidated glutamine/asparagine positions (*SERPINF1*, *OGN*, *CI*), which have also been identified in previous palaeoproteomic studies. An increase in deamidation spectral frequency for NCPs compared to most collagens has been reported previously (74), and might be explained by differences in diagenetic pathways or protein structure (64, 74).

It is well known that deamidation rates of glutamines and asparagines are dependent on the acidity/basicity of the following amino acid ( $n+1$ ), as well as protein structure (80, 81). We did not correct for or separate by this as the number of glutamine and asparagine residues not followed by a glycine is relatively low. Averaging frequencies across a single protein, assuming all spectral matches originate from a single source, should partly mitigate such issues.

Glutamines and asparagines followed by a glycine are known to have deamidation half-times one to two orders of a magnitude faster than when glutamines and asparagines are followed by a different amino acid (72, 80). We note that when protein members of cluster 1 do contain a deamidated spectrum (*KRT1*, *KRT9* and trypsin) such deamidated positions are followed by glycine, serine and threonine. Glutamines and asparagines followed by these amino acids are relatively fast-deamidating (80). Non-deamidated positions in the two NCPs that are not fully deamidated, lumican and human albumin, concern glutamines followed by a valine and a tyrosine, respectively. Glutamines followed by these amino acids are relatively slow-deamidating (80). Finally, the two proteins with the lowest overall deamidation frequency in cluster two, *COL10a1* and *COL7a1*, have asparagine positions followed by valine (*COL10a1*) and proline and glycine (*COL7a1*). Asparagines followed by a valine are relatively slow-deamidating, explaining the presence of non-deamidated asparagines at position 128 for *COL10a1*. The picture for *COL7a1* is slightly more complex, with Q+P being slow-deamidating and Q+G being fast-deamidating (80), while both are present in deamidated and non-deamidated states.

It has been observed elsewhere that deamidated peptides and  $^{13}\text{C}$  peaks of non-deamidated peptides can be misassigned during MS/MS selection, resulting in the misidentification of non-deamidated spectra as deamidated spectra (82, 83). This is visible as an overrepresentation of positive ppm mass errors for spectra identified as containing a deamidated residue (but actually referring to the  $^{13}\text{C}$  peak of a non-deamidated peptide; (82)). First, PEAKS was enabled to correct for the misassignment of  $^{13}\text{C}$  peaks. Second, we observe no difference in mass accuracy (ppm) distribution between deamidated and non-deamidated spectral matches for glutamines (Kolmogorov-Smirnov test;  $D=0.0216$ ,  $p=0.7583$ ) and asparagines (Kolmogorov-Smirnov test;  $D=0.0202$ ,  $p=0.9277$ ). Third, overall ppm distributions show a single continuous distribution primarily between -4 and 4 ppm (Fig. S7d). Together, this suggests incorrect  $^{13}\text{C}$

peak assignment during data acquisition has not influenced the subsequent identification of deamidated positions.

The observations made above allow us to conclude that the proteins in clusters 2 and 3 (n=35) are likely endogenous to the analysed bone specimens, with a deamidation frequency significantly different from that observed for known contaminants. Cluster 1 (deamidation average  $0.60 \pm 2.07\%$ ) contains known contaminants, some of which were also present in our extraction blanks, and some additional proteins not previously mentioned as potential contaminants (*DCD*, *COL4A6*). Differences in deamidation frequency between the proteins included in clusters 2 and 3 are most likely explained by differences in primary sequence and protein structure. Experimental studies on protein preservation and the occurrence of diagenetic PTMs, including deamidation, methionine oxidation, methionine di-oxymethionine and others (75), have the potential to provide key insights into protein preservation in a variety of environmental conditions. For 24 proteins, no or only few spectral ( $\leq 2$ ) observations of glutamine/asparagine positions were available. Their endogenous or exogenous status must be determined through other means. Future work incorporating total ion current (TIC) or other quantification approaches to spectral counts has the potential to provide more accurate quantification of deamidation. This has not been performed here due to the absence of TIC data when searching .mgf formatted data files.

### **Oxidation and sulfones**

No reports of oxidation levels of methionines in ancient proteins exist, despite methionine oxidation being included as a variable PTM in almost all palaeoproteomic analyses reported previously. For COL1, we find that methionines are oxidized in a high frequency of matched spectra (79.4%), while only a small portion of spectra is not oxidized (11.6%). The remainder (9.0%) contains sulfones (di-oxymethionine). Sulfones are detected on two positions in *COL1 $\alpha$ 1* (181, 580) and two in *COL1 $\alpha$ 2* (417, 785), despite large numbers of identified spectra for other methionine positions (Tab. S3). Two of these positions, *COL1 $\alpha$ 1* 181 and *COL1 $\alpha$ 2* 417, contained sulfones in all three studied bone specimens. The presence of sulfones in bone specimens of this age is not surprising, and has been shown before in comparative studies on collagen diagenesis (84). In addition, oxidation levels generally are very comparable between the three specimens. Although assumed as originating from bone diagenesis and/or protein extraction procedures, we find it interesting to note that methionine oxidation levels are not uniformly distributed over the COL1 triple helix (Fig. S6). Although a majority of positions is oxidized in  $\geq 75\%$  of observations, *COL1 $\alpha$ 1* 181 (position 3 in the triple helix) is oxidized between in  $\approx 40$ -60% of observations, *COL1 $\alpha$ 1* 217 (position 39 in the triple helix) is consistently oxidized in  $\approx 40$ /50% of observations and *COL1 $\alpha$ 1* 1000 (position 822 in the triple helix) in  $\approx 50$ -60% of observations. These three positions are all preceded by a proline and followed by a glycine. The only other instance where a methionine is preceded by a proline and followed by a glycine is *COL1 $\alpha$ 2* 93 (position 4 in the triple helix). This position is oxidized at a high frequency (86.3%) and preceded within the triple helix by *COL1 $\alpha$ 1* 181, which is oxidized at a lower frequency (67.9%).

In addition to the presence of oxidation on COL1 and other collagens, we note the presence of oxidized methionines in *ANT3* and *ALB*. For albumin, oxidation and sulfone frequency is very similar in all three extracts where this protein is identified (Tab. S3). Future proteomic investigation of oxidation location and frequency has the potential to elucidate the

biological or diagenetic origin of methionine sulfoxide and methionine sulfone in collagens and NCPs.

### **Acetylation and formylation**

Protein acetylation is a reversible enzymatic modification of lysine residues and N-terminal residues (85). Widely studied on histones and other nuclear proteins, lysine acetylation has also been investigated for numerous non-histone proteins (85, 86). The presence of other types of acyl-modifications of lysine residues such as formylation (87) are reported as well (reviewed in (88)). The presence of lysine formylation in the analytes might be due to the formic acid during LC-MS/MS analysis. Lysine acetylation has been implied in protein-protein interaction (89), and suggested to increase protein stability (90).

We find a large number of acetyl- and formyl-modifications of lysines in the datasets. Focusing on positions for which total ion count (TIC) exceeds 2% and that are not N-terminal, we identify positions that are variably acetylated/formylated, positions that are consistently acetylated/formylated, and positions that contain both PTM modifications. For acetylated positions, the PTM modification was manually evaluated using the occurrence of reporter ions at 143.1 and 126.1 m/z (Tab. S4; (91)). These reporter ions were present for all suggested acetylated positions with the exception of *COL1a1* 520 and *COL5a2* 968. For positions *COL1a1* 505, *COL1a1* 594, *COL1a1* 1096, *COL1a2* 654 and *COL1a2* 1064, both formylation and acetylation are present in the same extract.

### **Other PTMs**

Due to the nature of the PEAKS SPIDER search, multiple additional PTMs were identified. A complete analysis of these is beyond the scope of the current paper. These include several of the PTMs listed by (75). The presence of galactosyl and glucosyl-galactosyl, present in numerous previous palaeoproteomic studies (74) can be noted for *COL5a1* (position 795) in extracts AR-7 and AR-30. For additional lysine glycation sites mentioned by (74), we found the presence of formylation or acetylation modifications in *COL1a1* and *COL1a2* (see above). As well, we note the presence of a small number of hexose modifications. The diagenetic or biological origin of such PTMs requires further study.

### **GO annotation analysis**

After exclusion of protein contaminants based on deamidation frequency, we performed GO annotation analysis on the remaining proteins in order to infer enrichment for specific molecular functions, biological processes and (extra)cellular components. The analysis indicates enrichment for extracellular, fibrillary, triple-helical domain containing proteins as cellular components and related biological processes (Tab. S5). This is not surprising given the dominance of collagens in the protein list. A large number of biological processes enriched within the network relate to neuron development ( $p=4.84E-06$ , GO:0031175;  $p=4.93E-05$ , GO:0048666) and projection ( $6.73E-06$ , GO:0048812) or blood vessel development ( $p=0.012$ , GO:0001568). Finally, several enriched GO biological processes are related to cartilage development ( $p=4.99E-05$ , GO:0051216;  $p=7.08E-03$ , GO:0060351). Proteins associated with such GO terms are *COL1a1*, *COL2a1*, *COL10a1*, *COL11a1*, *COL11a2*, *COL27a1*, *CHAD*, *LUM* and *THBS1*.

There is a significant group of blood microparticles (GO:00725262; *ALB*, *SERPINA3*, *AHSG*, *APOA4*, *SERPINC1* and *C3*) which is enriched in biological processes involved in the

regulation of endopeptidase activity ( $p=6.32E-3$ , GO:0010951), peptidase activity ( $7.62E-3$ , GO:0010466), proteolysis ( $p=2.72E-2$ , GO:0045861) and protein metabolic processes ( $p=4.32E-2$ , GO:0051248). These proteins have been identified in non-hominin palaeoproteomes as well (26, 69) and are consistently incorporated into the mineralized bone matrix. Furthermore, none of these proteins cluster with the known contaminants (Figs. 1B, S5) when spectral data for glutamine and asparagine positions is available and are therefore interpreted as endogenous to the analyzed bone tissues.

### ***Amino acid variants***

Matching peptides and suggested amino acid substitutions were compared to Neandertal protein sequences provided by (92), Denisovan protein sequences provided by (56), and the 1000 genomes (93) non-synonymous SNPs (phase 3). This resulted in the identification of several variable amino acid positions among the 1000G data at relatively high frequency (Tab. S6). Here, such positions are included when both amino acid allele states have a global frequency of  $\geq 1.0\%$ . We identified 6 such positions covered by spectral matches in the dataset (Tab. S6).

We note that in each instance the identified amino acid state for the studied bone specimens is the ancestral, Hominidae amino acid state. In one case, *COL28a1* 472, this concerns the ancestral state (amino acid P) for which the dominant variant among present-day humans (amino acid A at a global frequency of 93.5%) is a human-specific substitution (rs17167927; (92)).

For *COL10a1*, we identified spectral matches to a peptide spanning positions 125-138 (Fig. S7). At amino acid position 128, this protein is translated as a D in all three Neandertal exomes available (92), a D in 0.9% and an N in 99.1% of modern humans (dbSNP rs142463796). The high-coverage Denisovan genome from Denisova Cave (56) has a heterozygous distribution with alleles translating for both N and D. Comparison with available Hominidae amino acid sequences indicates that the D represents the ancestral amino acid state for *COL10a1*, position 128 (Tab. S7).

*COL10a1* is a network-forming collagen protein exclusively expressed by hypertrophic chondrocytes when considering bone formation (94). Hypertrophic chondrocytes and the expression and translation of *COL10a1* RNA have been reported for endochondral ossification (95–97), intramembranous ossification, and unfused cranial sutures (98, 99) in Tetrapods. The deposition of *COL10a1* networks precedes mineralization of the deposited bone matrix, into which they are incorporated. This indicates *COL10a1* is not excreted during bone remodeling by osteoblasts. Furthermore, *COL10a1* has been detected in dental tissues as well, again suggesting a role in dental tissue mineralization (100, 101).

We tested whether the peptide-spectrum matches to *COL10a1* belonged to the same protein population as the contaminating proteins (Cluster 1). Comparing the deamidation frequency for all spectral matches to this protein to the frequencies observed for the population of contaminant proteins (Cluster 1; Figs. 1B, S5), we obtained a Z-score of 16.98. We conclude therefore that *COL10a1* is not part of the contaminant population ( $p<0.00001$ ). The same comparison to proteins assigned to cluster 2 results in a Z-score of -1.62, which corresponds to a non-significant  $p$ -value of 0.053. We can therefore not reject that *COL10a1* is part of the same population as proteins assigned to cluster 2. We interpret this cluster as representing endogenous ancient bone proteins.

The restricted cellular origin of the protein, its specific occurrence during initial bone development and deamidation values obtained for this protein (Fig. 1B) and the absence of this

protein in our extraction blanks all support the notion that the identification of *COL10a1* in AR-7, AR-16 and AR-30 is endogenous.

For Hominidae, there are several amino acid positions of phylogenetic interest present in *COL10a1* (Tab. S7). We note the presence of one amino acid position, *COL10a1* 27 T>M, that could discriminate between Neandertals and Denisovans/AMHs but for which we have no spectral matches; one position that has a unique substitution on the Homo-lineage (*COL10a1* 61 I>T) and for which we find spectral matches in AR-30 supporting the Homo-lineage state; as well as a unique non-synonymous substitution on position 128, D>N, which is derived in archaic hominins (Neandertals + Denisovan, see main text).

The hominin specific peptide spanning positions 121-138 was identified in bone specimen AR-30 for four spectra in two independent extracts (AR-30A and AR-30B). Differentiation between both possible amino acid states for position 128, N and D, is complicated as the asparagine (N, 132.12 g/mol) can be deamidated (+0.985), and therefore appear in mass equal to aspartic acid (D, 133.10 g/mol), a problem further enhanced by the possibility of incorrect MS1 <sup>13</sup>C peak selection during MS/MS acquisition. Manual validation confirmed that the monoisotopic peak was selected for fragmentation and that the correct monoisotopic fragment ions were assigned for three out of four spectra. The fourth spectrum was triggered on the first isotopic peak and the spectrum is therefore discarded from further consideration here. The parts per million (ppm) accuracy on all three remaining spectra matching to *COL10a1* are within 3 $\sigma$  of the mean error range and within the main distribution for extracts AR-30A and AR-30B (Fig. S7D+E). This further suggests that these spectra do not deviate from the other mass error assignments in our experiments.

The identification of well-fragmented non-deamidated asparagines in two biological replicates (AR-30A and C; Fig. S7) indicates that these represent the unmodified amino acid sequence. The absence of y-ions in these spectra on the N-terminal end of the first proline is a common feature of peptides rich in prolines, as is the presence of (intense) non-assigned fragment ions (102, 103). All three spectra contain non-enzymatic cleavages at their N-terminal end. Such cleavages are a common feature of palaeoproteomic analysis, likely caused by protein diagenesis during burial (26, 70, 104). Backbone cleavage products are a common feature of glutamine and asparagine-containing peptides, and should therefore be expected in palaeoproteomic studies (76, 80). Furthermore, our data indicates identical placement of a hydroxylated proline towards the C-terminal end for all three spectra (position 135; Fig. 7A-C), again supporting consistency between our replicate analyses.

We BLAST-searched the identified peptide sequence against the complete protein and nucleotide sequence databases of Genbank (protein\_blast) and Uniprot to test whether there were significant matches for the identified amino acid sequence outside those present in the databases searched using PEAKS. In addition, this allowed testing of the presence of the identified amino acid sequence in Genbank, which contains a more comprehensive set of Bacteria, Fungi and Archaea protein sequences. In each case, sequence hits to the corresponding *COL10a1* human sequence had the highest score/lowest E-value. Similarly, limiting the search to Bacteria, Fungi or Archaea did not result in significant sequence matches (Tab. S8).

Taken together, we see the above as conclusive evidence of the presence of an aspartic acid (N) at amino acid position 128 for protein *COL10a1* in bone specimen AR-30.

#### **4. Archaic ancestry of rs142463796 in COL10a1**

*Michael Dannemann\*, Janet Kelso*

\*to whom correspondence should be addressed: [michael\\_dannemann@eva.mpg.de](mailto:michael_dannemann@eva.mpg.de)

We identified an adenine variant in rs142463796 (hg19; chr6:116442897) that is at low frequency in some non-African populations, absent in African populations, and which matches the Neandertal and Denisovan genome sequences (56, 105). To test whether this allele may have been introduced into modern humans by admixture with Neandertals or Denisovans we tested whether the archaic-like variant of rs142463796 is located on an extended archaic-like haplotype.

We used the 1000 Genomes individuals (phase III; (93)) to define archaic-like SNPs (aSNPs) which are likely to have been introduced by admixture between modern humans and Neandertals/Denisovans. We first defined a set of potentially aSNPs by identifying genomic sites where at least one allele in the Neandertal and/or Denisovan differs from the alleles carried by 109 Yoruba individuals (an African group with no evidence for admixture from Neandertals or Denisovans). At a subset of these sites some of the non-African genomes carry the Neandertal or Denisovan-like allele. In each non-African individual we then scanned for consecutive aSNPs at which individuals carry the archaic-like allele. Among the 2028 non-African genomes we found 43 individuals carrying the archaic allele at rs142463796 on 46 haplotypes. We found consecutive stretches of aSNPs ranging from 222kb to 1689kb in 46 haplotypes. We define a region of 146kb (chr6: 116436420-116582525) shared by all 46 haplotypes as the core archaic-like haplotype (Fig. S8).

#### ***Additional evidence for introgression***

To verify that these archaic-like haplotypes are the result of admixture between modern humans and Neandertals and/or Denisovans we carried out three additional analyses.

- i. We examined the genomes of the Altai Neandertal (105) and the Denisovan (56) to determine which is most similar to the 46 archaic-like haplotypes identified in non-Africans. At all 72 of the aSNPs in the 146kb region all 46 archaic-like haplotypes are identical to the Altai Neandertal sequence while the Denisovan is homozygous for the Yoruba allele at 19 positions and heterozygous at 53 positions. Using the exome data of two additional Neandertals we found that all three Neandertals (92) are homozygous for the archaic-specific (not seen in Yoruba) allele of rs142463796 while the Denisovan is heterozygous (Fig. S8). The higher sequence similarity to the Neandertal rather than to the Denisovan genome suggests that the introgressed haplotype may be of Neandertal rather than Denisovan origin.
- ii. The presence of ancestral alleles in present-day human genomes is consistent with both introgression and with lineage sorting (ILS). Under ILS genomic regions can carry haplotypes that do not match the expected phylogeny (106), but such regions are expected to be short due to the extended time since the species split. By contrast, because introgression is more recent than the species split the genomic regions are expected to be longer. Using the approach presented by (107) and the parameters for the age of the Neandertal and Denisovan bones used by (108) we computed that a haplotype of at least 146kb in a region with a recombination rate of 0.4 cM/Mb (109) is unlikely to arise due to ILS ( $p=2.2E4$  and  $p=8.0E-9$  for mutation rates of  $1 \times 10^{-9}$  and  $0.5 \times 10^{-9}$  per base pair per year respectively).

iii. We examined the geographic distribution of the archaic-like allele of rs142463796 in the 1000 Genome data and the Simons Genome Diversity Panel (SGDP; [www.simonsfoundation.org/life-sciences/simons-genome-diversity-project-dataset/](http://www.simonsfoundation.org/life-sciences/simons-genome-diversity-project-dataset/)). The archaic allele is almost exclusively found in South and East Asia and Oceania (additionally one archaic-like haplotype was identified in a Toscani individual from Italy; Tab. S9). The archaic-like allele of rs142463796 is at highest frequency in Oceania – present in 33% of the Australians (two haplotypes in three Australians) and 47% of the Papuans (15 of 32 haplotypes) (Fig. 2). In mainland Asia the archaic-allele of rs142463796 is at lower frequency with the maximum observed in the Bengali from Bangladesh (6.3%).

We conclude that an archaic haplotype carrying rs142463796 was introduced into modern humans by admixture with archaic humans. Although the geographic distribution is consistent with admixture from either Neandertals or Denisovans, the higher sequence similarity of the haplotypes in modern humans to the Altai Neandertal genome suggests that the origin was more likely Neandertal.

## **5. Ancient DNA**

*Mateja Hajdinjak, Matthias Meyer, Svante Pääbo*

\*to whom correspondence should be addressed: [mateja\\_hajdinjak@eva.mpg.de](mailto:mateja_hajdinjak@eva.mpg.de)

Ancient DNA analysis was performed on two hominin specimens. AR-14 was also directly dated (Fig. 3), and included in stable isotope analysis (Fig. S10). AR-30 was used for palaeoproteomic analysis, which included the archaic amino acid sequence identified for *COL10a1* (Fig. 1; Tab. S7).

### ***Yield of DNA molecules and sequences***

The numbers of DNA molecules in the libraries from AR-14 and AR-30, as determined by digital droplet PCR, were  $2.82 \times 10^9$  and  $2.57 \times 10^9$ , and thus higher than in the extraction and library negative controls (Tab. S10). After enrichment, mapping and duplicate removal, 48,469 and 7,015 unique mtDNA sequences were obtained for both specimens (Tab. S10). Each unique sequence was seen 8.71 and 14.86 times on average, indicating that deeper sequencing of the libraries would not substantially increase the number of unique mtDNA sequences.

### ***Nucleotide substitutions and authentication of ancient DNA sequences***

Deamination of cytosine (C) to uracil (U) residues, which occurs primarily at single-stranded DNA overhangs, leaves characteristic C to T substitutions in ancient DNA sequence alignments, which are particularly frequent close to alignment ends (110). Elevated C to T substitution frequencies can thus provide evidence for the presence of authentic ancient DNA (111, 112). Of the sequences from specimen AR-14 that start or end at a C in the reference genome, 12.1% show a T at the 5'-end and 11.5% at the 3'-end, respectively (Tab. S11). These numbers are 20.5% and 28.1% for AR-30, suggesting that at least some of the sequences recovered from the two specimens are of ancient origin.

### ***Phylogenetic inferences based on „diagnostic” positions***

To assess which hominin group the mitochondrial DNA of the two specimens belongs to, we studied the state of sequences overlapping four sets of positions that are „diagnostic“ for the modern human, the Neandertal, the shared modern human/Neandertal and the Denisovan/Sima de los Huesos branch in the mtDNA tree. These positions were identified using the mtDNA genome sequences of 311 present-day humans (113), ten Neandertals (105, 110, 113–115), three Denisovans (116–118), one Sima de los Huesos individual (119) and the chimpanzee (120), and requiring that the sequences of the individuals on the respective branch differ from those of all other individuals. We identified 14, 19, 26 and 34 such positions, respectively. We then determined the percentage of derived variants supporting the state diagnostic for each branch, using all unique mtDNA sequences and only those sequences with a C to T difference to the reference genome within the first and/or last three alignment positions, thereby using the deamination signal to enrich for putative endogenous ancient sequences. This analysis is similar to that introduced in the paper describing nuclear DNA analysis of the Sima de los Huesos individuals (121).

When using all sequences, both specimens support the modern human state in more than 70% of their sequences (Fig. S9). However, we also observe considerable sharing of derived variants with the Neandertal branch (17.7% for AR-14 and 31.6% for AR-30, respectively) but not with the Denisovan/Sima de los Huesos branch, indicating that DNA fragments from both

modern humans and Neandertals are present. When restricting the analysis to putatively deaminated DNA fragments, the support of the Neandertal branch increases to above 70%, while that of the human branch drops to below 30% in both specimens. This would be compatible with the human DNA fragments in both specimens occurring due to recent contamination, which is much less deaminated than the Neandertal DNA.

To demonstrate this more directly, we separated putative Neandertal and present-day human sequences *in silico* using a fifth set of diagnostic positions where only those two hominin groups differ and studied the C to T substitution frequencies in both population of sequences. As expected, we observe low frequencies of damage-induced substitutions among the present-day human sequences ( $\leq 4\%$ ) and high frequencies ( $\geq 25\%$ ) among the sequences identified as Neandertal (Tab. S16). Even though there is too much residual contamination in the fraction of putatively deaminated sequences and their number is too small to enable reconstruction of useful mtDNA consensus sequences from the two specimens, the above analyses allow us to conclude that both specimens carry mtDNA of the type seen in Late Pleistocene Neandertals.

## **6. Radiocarbon dating and Bayesian modelling of AR-14**

*Sahra Talamo\**, *Jean-Jacques Hublin*

\*to whom correspondence should be addressed: [sahra.talamo@eva.mpg.de](mailto:sahra.talamo@eva.mpg.de)

In order to test the hypothesis of movement of the hominin specimens from the underlying Mousterian into the Chatelperronian layers directly (hypothesis **d** in the main text), one hominin sample (AR-14) was selected for direct AMS radiocarbon dating. Unfortunately, numerous other specimens are too small and could not be directly dated. Bone specimen AR-14 originated from square C7, archaeological Layer X. Specimen AR-14 was also analyzed genetically. After bone collagen extraction, isotopic measurements and the C:N ratios of the hominin bone at the Grotte du Renne fall in the acceptable range (Tab. S14). The collagen yield results in 3.9% which is above the accepted threshold of 1%. Consequently, it was concluded that the AR-14 specimen contained well-preserved protein suitable for radiocarbon dating. Together with the results obtained for additional bone specimens used for stable isotope analysis (Tab. S14), this generally confirms the good state of bone preservation at the site observed earlier (8).

The uncalibrated radiocarbon date of AR-14 results in an age of  $36,840 \pm 660$   $^{14}\text{C}$  BP (Tab. S16). Stable isotopic values associated with this date are discussed in the next section in a dietary/ecological context (Fig. S11).

We constructed four Bayesian models to test the compatibility of this date with previously obtained radiocarbon dates for the Grotte du Renne (5, 8), each addressing different criticisms on the Grotte du Renne chronostratigraphy. Both chronologies presented previously have been criticized. The first one (5) for the poor quality of extracted collagen (8, 18), the second one (8) for bone specimen selection criteria (11), and both for including specimens from possibly disturbed squares (9). The construction of four different models was used to estimate whether individual models provided different outlier probabilities for AR-14, and whether the estimated age boundaries for the Châtelperronian at the Grotte du Renne were significantly different between the models.

In each model, the prior outlier probability of EVA-29 and EVA-56 was set to 100%, following suggestions by (8), but we also ran models with the prior outlier probability set to 5% for these two specimens. Phases were defined based on archaeological layers, with layers IX and X combined based on the presence of lithic refits between these two layers (8, 16).

The first model, which we call the Hublin et al. model, included all radiocarbon dates reported in (8) (included are 26 samples in Layers IX+X and five samples in Layer VIII). The second model, which we call the hominin-modified model, included only radiocarbon dates from the (8) dataset from specimens containing hominin modifications (included are eight samples in Layers IX+X and two samples in Layer VIII). The third model, which we call the Discamps et al. model, excluded all radiocarbon dates from the Hublin et al. (8) dataset for specimens obtained from possibly disturbed squares, as discussed recently by (9) (included are 10 samples for Layer X). The fourth model, which we call the Higham et al. model, included all radiocarbon dates used by (11) (also containing a subset of (8) and only incorporating hominin-modified bone specimens). Bayesian modelling was performed using OxCal 4.2 (122) and IntCal13 (123). Bayesian CQL codes for the individual models are included at the end of this SI.

The date obtained fits within the chronological boundaries for Layers IX/X at the Grotte du Renne for each of the four models (Tab. S16), with posterior outlier probabilities of 4, 8, 5 and 4%, respectively (Fig. 3). Such probabilities would be sufficient to argue that the radiocarbon date from AR-14 does not deviate statistically from age estimate boundaries for the

Châtelperronian at the Grotte du Renne for each of the four Bayesian models. The radiocarbon date is also statistically similar to the date obtained for the Châtelperronian-associated Saint-Césaire Neandertal (Fig. 3; (8)), see (124) for a discussion of the stratigraphic integrity in Saint-Césaire).

We initially noticed a low agreement index for the three Bayesian models based on the Hublin et al. (8) dataset due to two outliers (EVA-56 and EVA-29). These two outliers are located in archaeological Layers VIII and X and squares C8 and Z11 respectively, and are both attributed to the Châtelperronian. Their ages fit well within the phase boundaries for the other Châtelperronian layers at the site, and are suggestive of the presence of some mixing between the Châtelperronian layers (IX+X and VIII). Following Hublin et al. (8), the outlier probability for EVA-56 and EVA-29 was set at 100% and afterwards the agreement index for the first three models is above the threshold of 60% (Tab. S16). The main difference between the first tree models and the Higham et al. model is the presence of additional outliers between Châtelperronian and non-Châtelperronian layers in the Higham et al. model. The agreement index for this model remains low ( $A_{\text{overall}} = 7.5\%$ ).

We stress that in all four models, with and without the outlier probability for EVA-56 and EVA-29 at 100%, the radiocarbon date for AR-14 does not deviate from the modelled boundaries for the Châtelperronian at the site, and neither does its outlier probability.

According to the Hublin et al. model, the Châtelperronian at the Grotte du Renne starts at 44,730 and finishes at 39,940 cal BP (68.2% probability). In the four models, the start boundaries for the Châtelperronian (Layers IX+X and VIII) agree within a range of 250 years (68.2% probability). A similar agreement is observed for the end of the Châtelperronian, with the exception of the hominin-modified model where the end Châtelperronian boundary VIII/VII is  $\approx 1000$  years older (68.2% probability) compared to the other three models, most probably due to a reduced number of dates.

The radiometric results therefore reject the hypothesis of movement of the hominin remains from the underlying Mousterian into the Châtelperronian (hypothesis **d** in the main text). The notion that the stratigraphic position of specimen AR-14 could result from major stratigraphic admixture can be rejected, as its date is incompatible with radiocarbon dates presented in (8) or (11) for both the Middle Palaeolithic Mousterian and Upper Palaeolithic Aurignacian assemblages at the site.

## **7. Stable isotope analysis of AR-14 and associated fauna**

*Klervia Jaouen\**

\*to whom correspondence should be addressed: [klervia\\_jaouen@eva.mpg.de](mailto:klervia_jaouen@eva.mpg.de)

Carbon (C) and nitrogen (N) stable isotope analyses are commonly used in archaeology for palaeodietary reconstructions. This approach relies on the fact that isotopic compositions of body tissues record the signature of the types of food eaten by an individual, following a predictable isotopic fractionation (125). The isotopic composition of the chemical elements are expressed with the delta notation ( $\delta$ ), which corresponds to the normalized difference between an isotopic ratio measured in a sample to the same ratio measured in a standard. Carbon isotopic compositions ( $\delta^{13}\text{C}$ ) are very low in terrestrial environments ( $\delta^{13}\text{C} \gg -20 \text{‰}$ ), except if the plants at the base of the food web have a C4 metabolism, such as corn or sorghum. However, such C4 plants were not present in Europe in the Late Pleistocene (126). In this context, carbon isotope ratios can therefore be used to detect the consumption of marine foods, characterized by high isotope ratios ( $\delta^{13}\text{C} \gg -12 \text{‰}$ ). Nitrogen isotopic compositions ( $\delta^{15}\text{N}$ ) of animal tissues are correlated to the trophic level of the animal: that is, there is an increase of 3-5‰ between the tissues of an individual and the protein in its diet. The  $\delta^{15}\text{N}$  values will be also higher in aquatic environments than in terrestrial ones at identical trophic levels. Therefore, these N isotope ratios can be used to assess whether a diet of an individual relied on fish, terrestrial meat or plants. Another parameter is, however, likely to significantly increase  $\delta^{15}\text{N}$  bone values: when infants are breastfeeding, they are in effect consuming their mother's tissues. Consequently, in isotopic terms, breastfeeding infants are at a higher trophic level. The elevation of the collagen  $\delta^{15}\text{N}$  value between the infant and its mother is equivalent to the magnitude of a trophic level effect. During the weaning process, this elevation progressively disappears (127, 128).

Over the past twenty years, C and N isotope analyses have been applied to Neandertals and anatomically modern humans (AMHs). These analyses documented a homogeneous diet of Neandertals, in time and space, with collagen isotopic compositions indicating that they were top-level carnivores but did not exploit significant amounts of aquatic resources (129). On the other hand, early AMH show a wider range of  $\delta^{15}\text{N}$  values, which can be explained by the consumption of variable amounts of freshwater or marine fish (129).

Bone collagen is the material of choice for C and N isotope analyses. Because bone is slowly renewed throughout the life of the individual, collagen isotopic values record the average isotope composition of food consumed over several years before death (130, 131). In a protein, like collagen, these values actually reflect more the dietary protein sources rather than the diet as a whole (132). Variations of C and N isotopic baseline can exist between different regions due to environmental factors (133). It is therefore preferable to assess the bone isotopic compositions of the local fauna to avoid any bias in the interpretation of the diet of their potential consumers (129). For this reason, in addition to the AR-14 hominin bone, we conducted C and N isotope analyses on the associated fauna.

### ***Results and discussion***

The stable isotope results for the AR-14 specimen and associated fauna are presented in Tab. S14 and Fig. S11. The amount of collagen extracted (>30 kDa fraction) and its C:N ratio meet published quality controls in all the analyzed samples, except one sample that failed collagen extraction (S-EVA 32918). The data previously reported in (8) for which taxonomic information was available are plotted in Fig. S11 and are taken into account in the average values for each

animal group reported in Tab. S15. Only samples from Layers IX and X were included, as these comprise one chronological and stratigraphic layer (Fig. 3).

### ***Isotopic composition and diet of the Grotte du Renne fauna***

The data obtained for this study are consistent with previous ones published in (8). The herbivores show a signature of a diet relying on C3 terrestrial plants. The reindeer exhibit higher  $\delta^{13}\text{C}$  values than other herbivores, which is commonly observed and interpreted as the result of lichen consumption (134, 135).  $\delta^{13}\text{C}$  values for one horse and the single Rhinocerotidae overlap with that of reindeer. This suggests that horse and Rhinocerotidae diet may have included a significant amount of lichen for some individuals. As previously observed for cave bears (136, 137), the Ursidae specimen shows a low  $\delta^{15}\text{N}$  value indicating a plant-based diet. Two out of three Elephantidae exhibit relatively high  $\delta^{15}\text{N}$  isotope ratios (Fig. S11). This is in agreement with what is generally reported for this family (137–139). There is one unusual result: a single reindeer and a bovid with a high  $\delta^{15}\text{N}$  value (S-EVA 32904). As has been previously suggested for mammoths, this individual may be from an arid region characterized by higher baseline  $\delta^{15}\text{N}$  values (139–141). However, this could also be explained by suckling, since the analyses were conducted on small bone fragments and it was, therefore, not possible to assess if the animals were juveniles or not. Such an explanation would also account for its  $\delta^{13}\text{C}$  value, which does not differ from the other reindeer.

On average, the carnivore  $\delta^{15}\text{N}$  values show an enrichment of 4 ‰ compared to that of herbivores (Tab. S15). The Pantherinae and Canidae have overlapping  $\delta^{13}\text{C}$  values with reindeer, pointing in the direction of a diet mostly relying on these animals. Considering this hypothesis, the trophic level effect would correspond to an enrichment of almost 4.6‰ (Tab. S15). The hyena's prey choice looks less selective, given the fact that it exhibits intermediate  $\delta^{13}\text{C}$  value compared to the range observed for herbivores.

### ***Isotopic values and diet of the AR-14 hominin***

The  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values of AR-14 are -19.5‰ and 14.3‰ respectively. Similar to the values observed for the associated fauna, the C isotopic composition of AR-14 indicates a diet based on terrestrial C3 plants and/or animals that consumed terrestrial C3 plants (Fig. S11a). There is no evidence of marine food consumption for this individual. The nitrogen isotope value is very high: the difference between AR-14 and the local herbivores is 9.3‰ on average ( $\Delta^{15}\text{N}_{\text{AR-14-herbivores}}$ ) and 5.4‰ between AR-14 and the carnivores on average ( $\Delta^{15}\text{N}_{\text{AR-14-carnivores}}$ ; Tab. S15).

Two factors could account for such a high value: 1) consumption of freshwater fish, or 2) breastfeeding. When compared to Neandertals and contemporary AMH (>30 000 ky BP), the AR-14  $\delta^{15}\text{N}$  value is higher than all the values reported for Neandertals and falls in the range of those observed for AMH (Fig. S11b; (129, 133, 136, 142–147)). Among these early AMH, isotopic data for the associated fauna are only available for three individuals. These three individuals show significantly higher  $\delta^{15}\text{N}$  values than the local carnivores (Fig. S11b), a pattern generally interpreted as a signature of fish consumption (129, 142, 143). Higher  $\delta^{15}\text{N}$  values of hominins relative to carnivores were observed elsewhere for Neandertals before (148), but are potentially explained by hominin predation of megafauna such as mammoths. Similar to what was observed for the Neandertals from Goyet and AMH, AR-14 is above all the local carnivore  $\delta^{15}\text{N}$  values, but the magnitude of this  $\Delta^{15}\text{N}_{\text{AR-14-carnivores}}$  value is, in this case, twice as high than the  $\Delta^{15}\text{N}_{\text{AMH-carnivores}}$  (Fig. S11b).

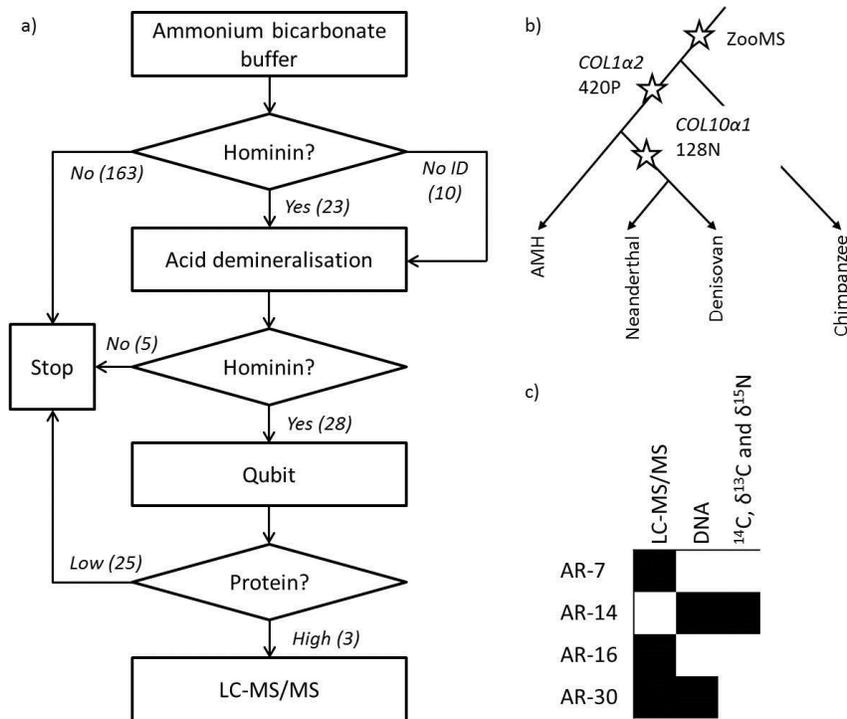
Another possible explanation that could account for such a high  $\delta^{15}\text{N}$  value is breastfeeding. As mentioned before, nitrogen isotope ratios of bone collagen are between 3 and 5‰ higher than dietary protein (149). Interestingly, the difference between  $\delta^{15}\text{N}$  values of AR-14 and the associated herbivores is 9.3‰, which is consistent with the range expected for a secondary carnivore ( $\Delta^{15}\text{N}$ : +6 to +10 ‰). If AR-14 was breastfed, the  $\delta^{15}\text{N}$  collagen value of the mother should be one trophic level below the AR-14 value, around 9.65 ‰. This expected value falls in the range observed for the local carnivores (Fig. S11). This hypothesis is consistent with the fact that *COL10a1* was identified in an associated bone specimen, indicating that the bone still contained (pre)hypertrophic chondrocytes and that 1-year old infant remains have been found nearby the AR-14 specimen. The fact that AR-14 was breastfed would also imply that the AR-14 mother was a top-level carnivore and had similar prey choices as hyena's, considering the AR-14  $\delta^{13}\text{C}$  bone value (Fig. S11).

### **Conclusion**

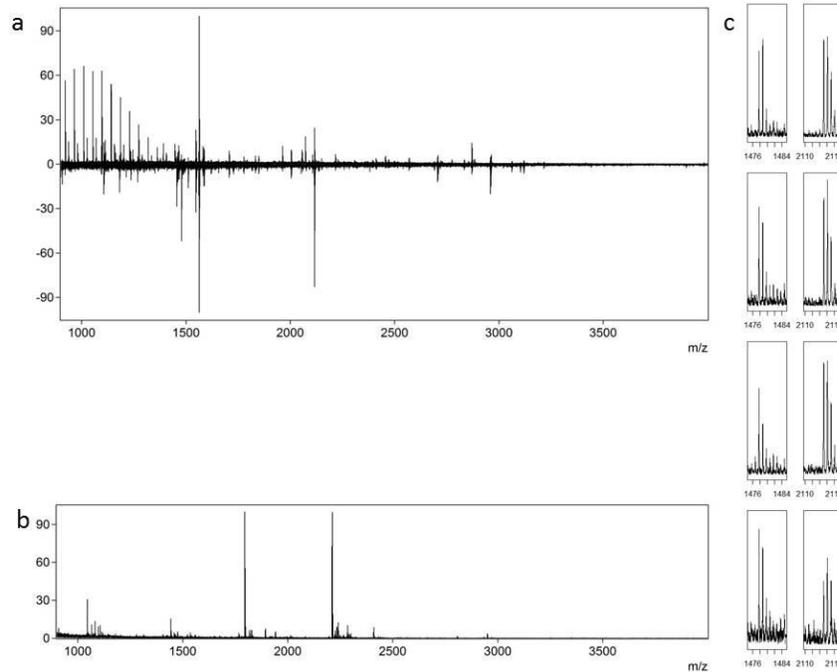
The AR-14 hominin is characterized by the highest  $\delta^{15}\text{N}$  value compared to all European Late Pleistocene hominins for which we have data on associated fauna. This high value is most likely explained by breastfeeding, a conclusion supported by osteological analysis and protein identifications obtained for related bone specimens identified here through ZooMS. This would suggest its mother was a top-level carnivore, primarily consuming large herbivore.

**Dataset S1: FASTA file containing *de novo* sequenced COL1 sequences.** For 27 species the complete triple helical regions of *COL1 $\alpha$ 1* and *COL1 $\alpha$ 2* (length 1014 amino acids for both) is included. Both sequences are concatenated with a K and leucines have been converted into isoleucines.

**Dataset S2. COL1 database generated for this study.** The file includes sample details, COL1 sequence accession numbers and ZooMS peptide marker masses observed in MALDI-TOF-MS experiments or predicted from available COL1 sequences. Accession numbers refer to Genbank, UniProt, ENSEMBL or specific papers through which data is made available. For peptide markers, displayed masses are computed based on amino acid sequence and hydroxylation occupation when "Data origin = Genetic", and observed masses when "Data origin = Proteomic". For the latter, these masses have been verified using subsequent LC-MS/MS analysis of the same or an extract from a closely related species.

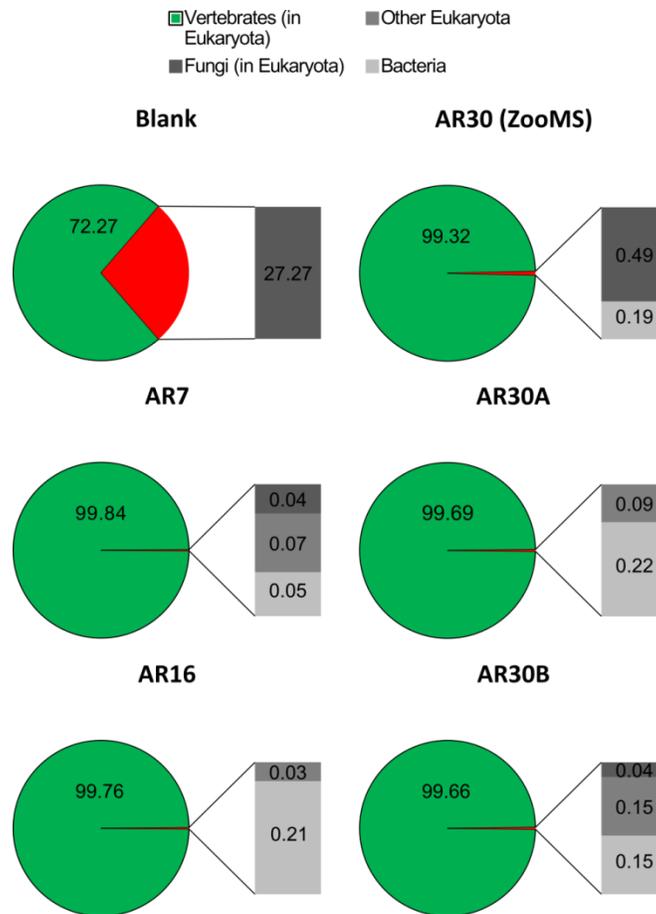


**Figure S1. Flowchart depicting the palaeoproteomic workflow. a)** ZooMS screening protocol, including the selection step for subsequent LC-MS/MS analysis. Numbers in brackets refer to number of samples for which a particular decision was made. **b)** Schematic nuclear hominin phylogeny and key protein amino acid positions. **c)** Hominin specimens subsequently used for multi-disciplinary analysis. Stable isotope analysis ( $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ ) was performed on an additional 21 faunal bone specimens (see table S14).

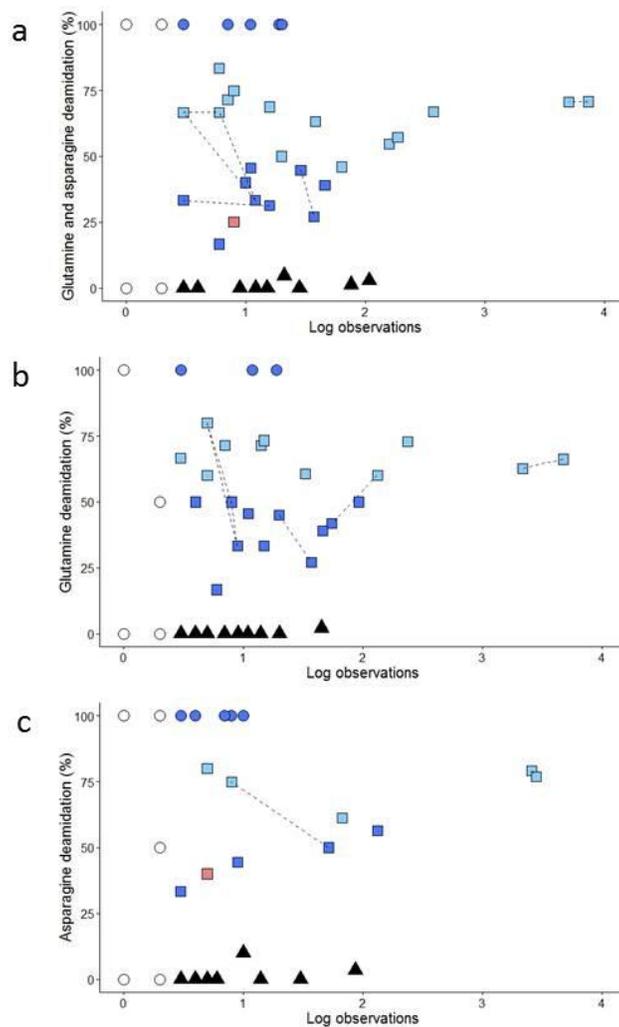


**Figure S2. ZooMS MALDI-TOF-MS screening. a)** Non-destructive ammonium-bicarbonate and demineralised COL1 ZooMS spectra of hominin AR-6. **b)** extraction blank spectrum containing keratin and trypsin peaks. **c)** Additional close-ups of hominin ZooMS peptide markers B (left) and D (right) for AR-7, 14, 16 and 30 (top to bottom). These are the bone specimens used for subsequent biomolecular and chronometric analyses.

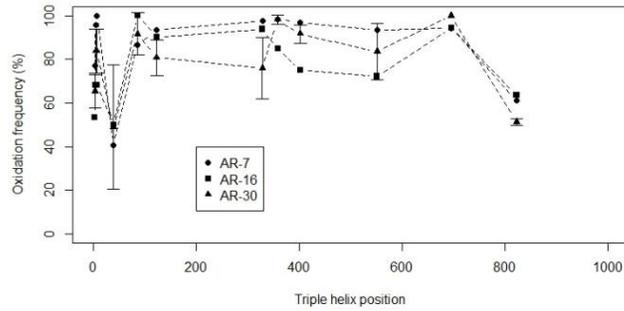




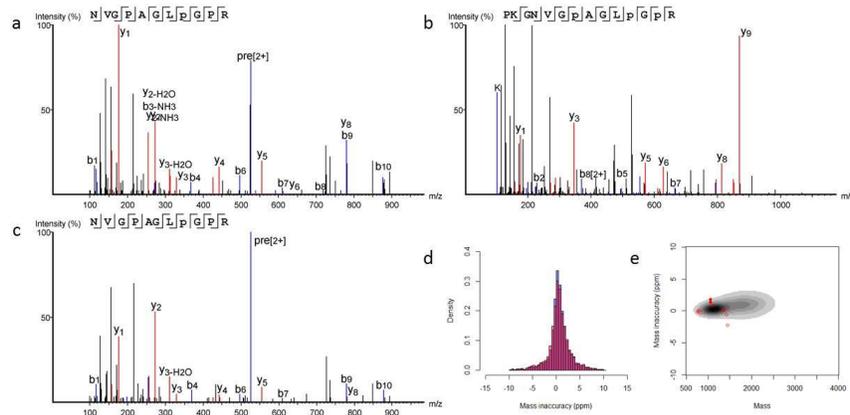
**Figure S4. Protein composition from the Mascot search to the complete UniProt database (in percentages).** Spectral matches (in %) are dominated to those from vertebrate origin, with only minimal presence of non-vertebrate proteins.



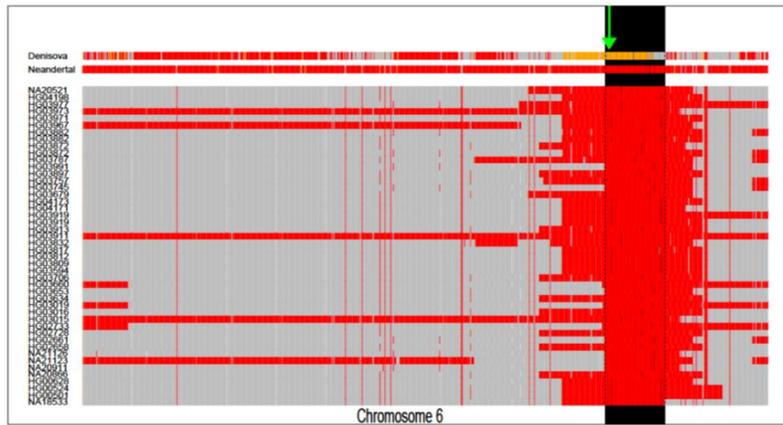
**Figure S5. Deamidation frequency cluster analysis.** a) Deamidation frequencies for all positions. b) Deamidation frequencies for glutamine positions. c) Deamidation frequencies for asparagine positions. *COL10a1* is indicated in pink, and dotted lines indicate helical association between collagen alpha chains. Cluster analysis provides 3 clusters (filled triangles, squares and circles). Open circles represent proteins that have 2 or less spectral matches, and were not included in cluster analysis. *COL10a1* is indicated in pink. 0% indicates no deamidation while 100% indicates complete deamidation.



**Figure S6. Methionine oxidation frequency for the COL1 helical region.** Methionine oxidation frequencies across the COL1 triple helix for all three studied bone specimens. Values for AR-30 are averaged across the three biological replicates, and error bars represent the standard deviation. Sulfones are included. See Tab. S3 for individual position frequencies for oxidation and sulfones.

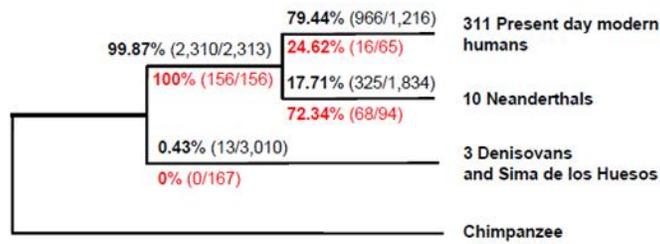


**Figure S7. LC-MS/MS spectra covering position *COL10a1* 128N (a-c) and their associated mass error (d-e).** Biological replicates AR-30B (a-b) and AR-30A (c). For prolines, “p” represents hydroxylated prolines and “P” non-hydroxylated prolines. **d**) PPM error distribution for all peptides (blue, n=7,498) and those containing glutamines and asparagines (red, n=3,867). **e**) PPM error distribution by peptide mass with spectra matching to *COL10a1* in open red circles and those matching to *COL10a1* 128N in filled red circles. The shaded error distribution area includes over two-thirds of the total number of assigned spectra.

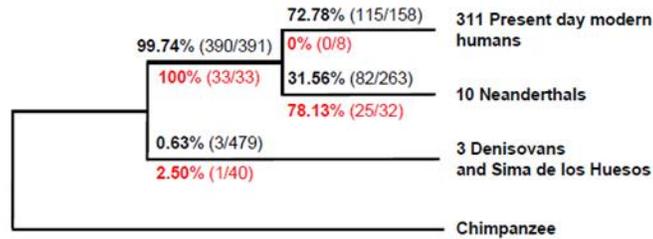


**Figure S8. Archaic-like haplotypes in individuals with archaic-like allele at rs142463796.** Sharing of archaic-like alleles encompassing 817 aSNPs on chromosome 6 (115355562-117076597) representing the longest archaic-like haplotypes for 46 haplotypes in the 1000 Genomes data that carry the archaic-like allele at rs142463796 as well as the genome sequences across these positions of the Altai Neandertal and Denisovan. The individual identifiers for the individuals carrying these 46 haplotypes are shown on the left hand site. Genomic aSNP positions for the 46 modern human haplotypes carrying the Yoruba-like allele are shown in gray while archaic-like alleles are shown in red. The archaic human genome sequences are color coded accordingly with the addition of orange highlighting of heterozygous positions. A region of consecutive aSNPs with shared archaic alleles in all 46 modern human haplotypes is highlighted in black (chr6:116436420-116582525). The position of rs142463796 is highlighted by a green arrow.

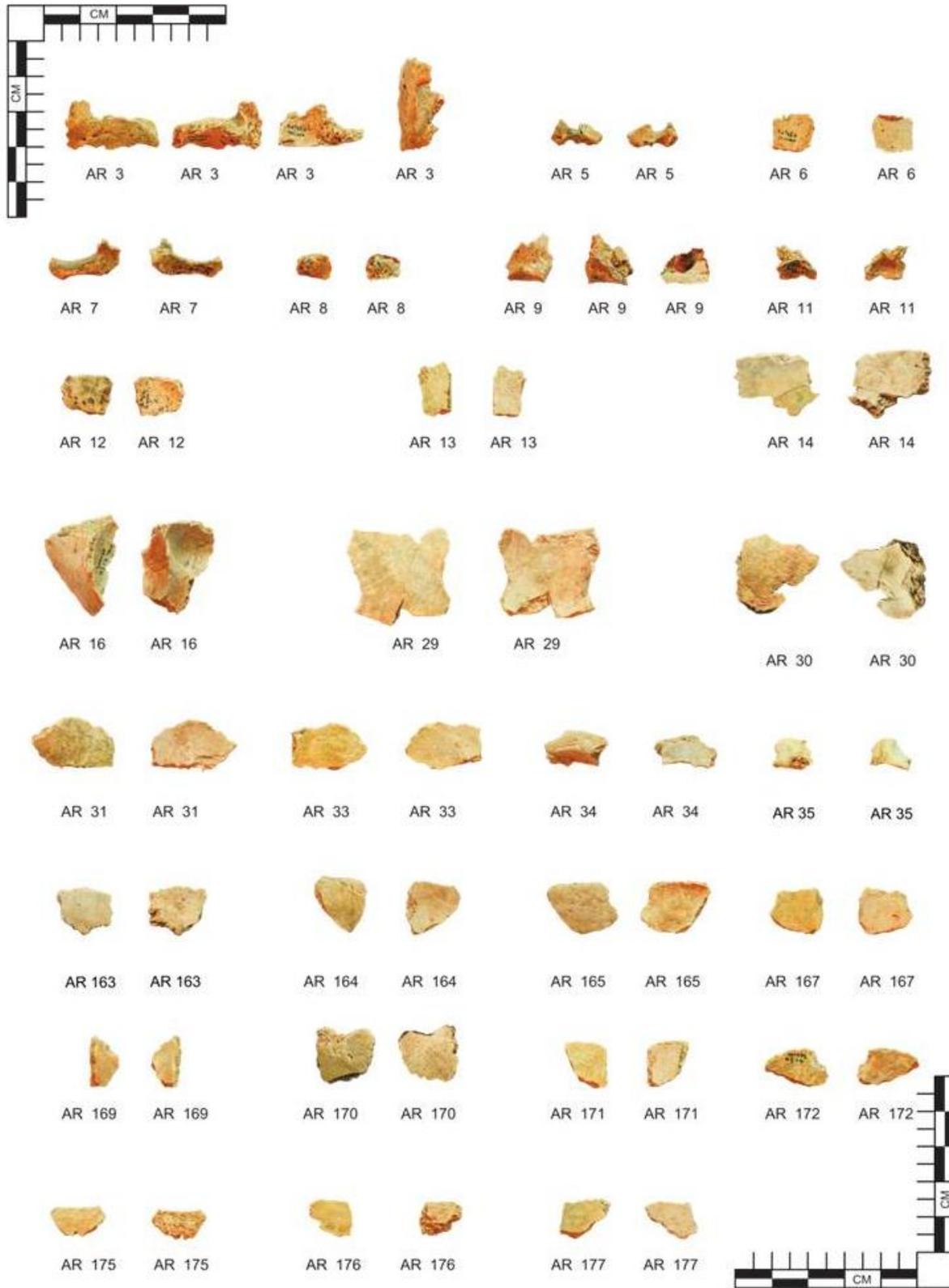
### AR-14



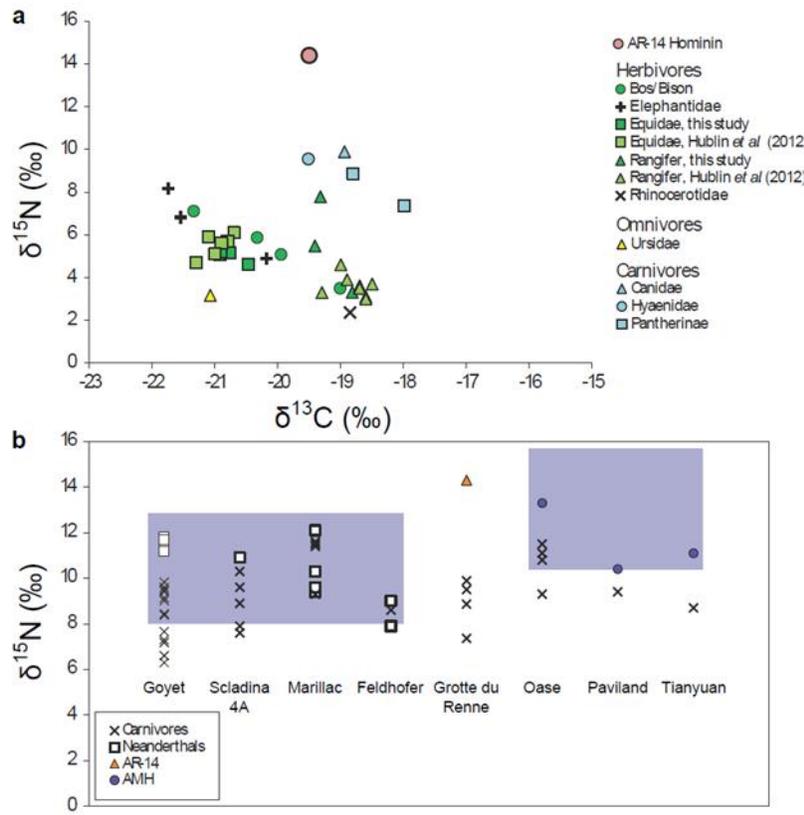
### AR-30



**Figure S9. Percentage of variants supporting the derived state at positions diagnostic for each branch in the hominin mtDNA tree.** Provided in brackets are the number of derived variants shared by sequences overlapping diagnostic positions as well as the total number of observations. Grey numbers above the branch were determined using all mtDNA sequences and red numbers below the branch are from sequences with terminal C to T substitutions only.



**Figure S10. Overview of all 28 bone specimens identified as hominines using ZooMS.**



**Figure S11. Isotopic analysis of AR-14 and associated fauna. a)** Dietary stable isotopes for AR-14, herbivores and carnivores from the Grotte du Renne Layers IX and X. **b)**  $\delta^{15}\text{N}$  values of carnivores and hominins for different Late Pleistocene sites. Blue areas represent the range of values reported for all Neanderthals (on the left) and AMHs (on the right), including individuals with no data available for associated fauna remains. Bone collagen values for Neanderthals, AMHs and associated fauna come from the literature (see SI Appendix Section 7). Foxes are only represented for the sites where their diet has been interpreted as carnivorous.

**Table S1. COL1 amino acid substitutions within the Hominidae family.** Positions are according to UniProt accession number P02452 for *COL1a1* and P08123 for *COL1a2*. The ancestral state is suggested from comparison to available *Nomascus leucogenys* and Cercopithecidae COL1 sequences. For *COL1a1* 1032 no amino acid is given in the *Nomascus leucogenys* reference sequence, while all present Cercopithecidae contain a proline (P). ZooMS peptide markers D and F are marked by \* and \*\*, respectively.

Protein	Position	Ancestral AA	<i>Pongo abelii</i>	<i>Pongo pygmaeus</i>	<i>Gorilla gorilla</i>	<i>Pan troglodytes</i>	<i>Pan paniscus</i>	<i>Homo sp.</i>	AR-7	AR-16	AR-30
<i>COL1a1</i>	172	T	T	T	A	T	T	T	T	T	T
<i>COL1a1</i>	784**	T	T	T	S	S	S	S	S	S	S
<i>COL1a1</i>	1032	P/A?	A	A	A	A	A	A	A	A	A
<i>COL1a1</i>	1083	A	A	T	A	A	A	A	A	A	A
<i>COL1a1</i>	1194	G	G	G	G	A	A	A	A	A	A
<i>COL1a2</i>	96	M	N	M	M	M	M	M	M	M	M
<i>COL1a2</i>	279	A	A	A	S	A	A	A	A	A	A
<i>COL1a2</i>	420	P	P	P	P	A	A	P	P	P	P
<i>COL1a2</i>	435	N	S	S	N	N	N	N	N	N	N
<i>COL1a2</i>	564	A	A	A	A	T	A	A	A	A	A
<i>COL1a2</i>	743	A	T	A	A	A	A	A	-	A	-
<i>COL1a2</i>	894*	P	P	P	S	P	P	P	P	P	P
<i>COL1a2</i>	990	A	V	V	A	A	A	A	A	A	A
<i>COL1a2</i>	1013	D	E	E	E	E	E	E	E	E	E
<i>COL1a2</i>	1044	P	P	P	S	P	P	P	P	P	P

**Table S2. Identified proteins per analyzed bone specimen.** Endogenous and exogenous state determination is based on deamidation spectral counts.

Gene	UniProt Accession	AR-7	AR-16	AR-30 (ZooMS)	AR-30A	AR-30B	Blank	Endogenous/Exogenous state
<i>COL1a1</i>	P02452	•	•	•	•	•		endogenous
<i>COL1a2</i>	P08123	•	•	•	•	•		endogenous
<i>COL2a1</i>	P02458	•	•	•	•	•		endogenous
<i>COL3a1</i>	P02461	•	•	•	•	•		endogenous
<i>COL5a1</i>	P20908	•	•	•	•	•		endogenous
<i>COL5a2</i>	P05997	•	•	•	•	•		endogenous
<i>COL5a3</i>	P25940	•	•	•	•	•		endogenous
<i>COL7a1</i>	Q02388	•	•	•	•	•		endogenous
<i>COL10a1</i>	Q03692	•	•		•	•		endogenous
<i>COL11a1</i>	P12107	•	•		•	•		endogenous
<i>COL11a2</i>	P13942	•	•		•	•		endogenous
<i>COL22a1</i>	Q8NFW1	•	•		•	•		endogenous
<i>BGN</i>	P21810	•	•		•	•		endogenous
<i>COL9a1</i>	P20849	•	•	•				endogenous
<i>COL4a1</i>	P02462		•		•	•		endogenous
<i>COL4a3</i>	Q01955		•		•	•		endogenous
<i>ALB (human)</i>	P02768			•	•	•		endogenous
<i>CLEC3B</i>	P05452	•	•					endogenous
<i>COL16a1</i>	Q07092	•				•		endogenous
<i>COL24a1</i>	Q17RW2	•				•		endogenous
<i>COL26a1</i>	Q96A83	•				•		endogenous
<i>COL27a1</i>	Q8IZC6	•				•		endogenous
<i>COL4a5</i>	P29400		•	•				endogenous
<i>COL28a1</i>	Q2UY09		•		•			endogenous
<i>SERPINF1</i>	P36955		•			•		endogenous
<i>COL25a1</i>	Q9BXS0				•	•		endogenous
<i>AHSG</i>	P02765				•	•		endogenous
<i>CHAD</i>	O15335				•	•		endogenous
<i>COL9a3</i>	Q14050	•						endogenous
<i>COL8a1</i>	P27658		•					endogenous
<i>LUM</i>	P51884				•			endogenous
<i>OMD</i>	Q99983				•			endogenous
<i>SERPINC1</i>	P01008				•			endogenous
<i>COL4a4</i>	P53420					•		endogenous
<i>OGN</i>	P20774					•		endogenous
<i>COL21a1</i>	Q96P44	•	•	•	•	•		unknown
<i>KCNQ4</i>	P56696	•	•	•	•	•		unknown
<i>IGHG1</i>	P01857	•	•	•	•	•		unknown
<i>DLL3</i>	Q9NYJ7		•	•	•	•		unknown
<i>CDC42EP1</i>	Q00587	•			•	•		unknown
<i>COL18a1</i>	P39060	•		•				unknown
<i>KATNA1</i>	O75449		•	•				unknown
<i>COL6a2</i>	P12110				•	•		unknown
<i>COL12a1</i>	Q99715				•	•		unknown
<i>NUCB2</i>	P80303				•	•		unknown
<i>THBS1</i>	P07996				•	•		unknown
<i>ACTB</i>	P60709	•						unknown
<i>COL4a2</i>	P08572	•						unknown
<i>MLLT4-AS1</i>	Q9Y6Z5	•						unknown
<i>ERICH3</i>	Q5RHP9	•						unknown
<i>USP15</i>	Q9Y4E8	•						unknown
<i>KIF26A</i>	Q9ULI4		•					unknown
<i>COL8a2</i>	P25067			•				unknown
<i>APOA4</i>	P06727			•				unknown
<i>C3</i>	P01024				•			unknown

<i>P4HB</i>	P07237				•		unknown
<i>POSTN</i>	Q15063				•		unknown
<i>SERPINA3</i>	P01011				•		unknown
HB* protein group	P68871 / P69905 / P02042					•	unknown
<i>HIST1H4A</i>	P62805					•	exogenous
<i>KRT25</i>	Q7Z3Z0					•	exogenous
<i>H2AFV</i>	Q71UI9					•	exogenous
<i>TRYPsin</i>	P00761	•	•	•	•	•	exogenous
<i>KRT1</i>	P04264	•	•	•	•	•	exogenous
<i>COL4a6</i>	Q14031	•	•		•	•	exogenous
<i>ALB</i> (bovine)	P02769			•	•	•	exogenous
<i>DCD</i>	P81605	•		•		•	exogenous
Ubiquitin-like protein group	P0CG47 / P0CG48 / P62987 / P62979			•		•	exogenous
<i>CSN2</i>	P02666	•					exogenous
<i>KRT5</i>	P13647	•					exogenous
<i>KRT6A</i>	P02538	•					exogenous
<i>KRT6B</i>	P04259	•					exogenous
<i>KRT9</i>	P35527	•					exogenous
<i>KRT10</i>	P13645	•					exogenous
<i>KRT16</i>	P08779	•					exogenous
LATEX Gloves	P42212	•					exogenous

**Table S3. Methionine (M) oxidation sites and frequency detected in this study.** Numbers in brackets refer to the total number of spectral observations for a given position. O=oxidation, S=sulfone.

Protein	Position	AR-7		AR-16		AR-30 ZoomMS		AR-30A		AR-30B	
		O (%)	S (%)	O (%)	S (%)	O (%)	S (%)	O (%)	S (%)	O (%)	S (%)
<i>COL1a1</i>	181	59.6	17.5	43.3	10.0	57.1	21.4	53.8	11.5	51.7	13.8
<i>COL1a1</i>	217	40.6		50.0		0.0		52.5		45.5	
<i>COL1a1</i>	264	86.7		100.0		100.0		100.0		83.3	
<i>COL1a1</i>	301	93.6		90.3		93.3		76.9		84.6	
<i>COL1a1</i>	580	95.7	1.1	75.0		98.0		93.8		86.2	3.4
<i>COL1a1</i>	729	93.3		72.2		100.0		92.3		75.0	
<i>COL1a1</i>	1000	61.3		63.6		50.0		52.7		50.0	
<i>COL1a2</i>	93	95.8		68.2		100.0		80.0		87.5	
<i>COL1a2</i>	96	100.0		68.2		100.0		80.0		87.5	
<i>COL1a2</i>	417	87.5	10.0	87.5	6.3	80.0	20.0	75.0		69.2	7.7
<i>COL1a2</i>	447	98.5		84.8		100.0		96.4		100.0	
<i>COL1a2</i>	785	94.3		94.4		100.0		100.0		94.1	5.9
<i>COL2a1</i>	239	100.0		75.0				0.0		0.0	
<i>COL2a1</i>	602	90.9				100.0		100.0		100.0	
<i>COL2a1</i>	751	100.0									
<i>COL3a1</i>	299	100.0				0.0					
<i>COL3a1</i>	398	100.0									
<i>COL3a1</i>	410	33.3									
<i>COL3a1</i>	535	50.0		0.0		0.0					
<i>COL3a1</i>	578	90.9		75.0		100.0		100.0		100.0	
<i>COL3a1</i>	727	100.0									
<i>COL3a1</i>	964	100.0		100.0		100.0		100.0		100.0	
<i>COL4a1</i>	235			100.0							
<i>COL5a1</i>	519							100.0		100.0	
<i>COL5a1</i>	521							100.0		100.0	
<i>COL5a1</i>	564	100.0						100.0		100.0	
<i>COL5a1</i>	701			50.0							
<i>COL5a1</i>	1011	50.0		0.0				0.0			
<i>COL5a2</i>	248	100.0									
<i>COL5a2</i>	763	100.0									
<i>COL5a2</i>	997	100.0		100.0				100.0			
<i>COL5a2</i>	1006			100.0						100.0	
<i>COL7a1</i>	1262							100.0			
<i>COL11a1</i>	492	0.0						100.0		100.0	
<i>COL11a1</i>	494	100.0						100.0		100.0	
<i>COL11a1</i>	668	0.0		50.0				0.0			
<i>COL11a1</i>	534	100.0		66.7							
<i>COL11a2</i>	451			100.0							
<i>COL16a1</i>	1569	16.7								66.7	
<i>ALB</i>	147					75.0	25.0	66.7	33.4	62.5	37.5
<i>ALB</i>	572					100.0		100.0		100.0	
<i>ANT3</i>	455							100.0			

**Table S4. Acetylation sites detected in this study.** N-terminal acetylation is excluded from analysis. For samples, numbers refer to observed acetylated spectra and total observed spectra for a given position (respectively). Only positions with TIC >2% in one of the samples are included.

Protein	Position	AR-7	AR-16	AR-30 (ZoomMS)	AR-30A	AR-30B	Ion 126.1 m/z	Ion 143.1 m/z
<i>COL1a1</i>	265					1/6		yes
<i>COL1a1</i>	505	3/83	5/57		2/24	1/24	yes	yes
<i>COL1a1</i>	520				1/9			
<i>COL1a1</i>	594	5/39	2/34				yes	yes
<i>COL1a1</i>	742				1/9		yes	yes
<i>COL1a1</i>	751		1/40					yes
<i>COL1a1</i>	781				1/65			yes
<i>COL1a1</i>	907	1/3	1/2				yes	yes
<i>COL1a1</i>	1096	1/127	1/66					yes
<i>COL1a2</i>	354			1/25				yes
<i>COL1a2</i>	360			1/21				yes
<i>COL1a2</i>	621	2/64						yes
<i>COL1a2</i>	654				2/9	2/7	yes	yes
<i>COL1a2</i>	1064	8/104	1/32	2/25	3/51	2/76	yes	yes
<i>COL2a1</i>	452			1/1				yes
<i>COL2a1</i>	803			1/3	1/7			yes
<i>COL2a1</i>	929		1/1					yes
<i>COL5a1</i>	1326				1/1			yes
<i>COL5a1</i>	1515					1/1		yes
<i>COL5a2</i>	539	1/1						yes
<i>COL5a2</i>	968			1/2				
<i>COL11a1</i>	933				1/2			yes
<i>COL22a1</i>	954				1/1			yes
<i>COL22a1</i>	1410		1/1					yes

**Table S5. GO enrichment analysis at  $p < 0.05$  (Bonferroni corrected  $p$ -values).**

<b>GO Molecular Functions</b>	<b>Term</b>	<b><math>p</math>-value</b>
GO:0005201	extracellular matrix structural constituent	0.000
GO:0005198	structural molecule activity	0.000
GO:0048407	platelet-derived growth factor binding	0.000
GO:0030020	extracellular matrix structural constituent conferring tensile strength	0.000
GO:0004866	endopeptidase inhibitor activity	0.000
GO:0061135	endopeptidase regulator activity	0.000
GO:0030414	peptidase inhibitor activity	0.000
GO:0061134	peptidase regulator activity	0.000
GO:0019838	growth factor binding	0.001
GO:0005539	glycosaminoglycan binding	0.002
GO:0008201	heparin binding	0.006
GO:0004867	serine-type endopeptidase inhibitor activity	0.011
GO:0050839	cell adhesion molecule binding	0.016
GO:0005178	integrin binding	0.029
GO:0004857	enzyme inhibitor activity	0.044

<b>GO Biological Processes</b>	<b>Term</b>	<b><math>p</math>-value</b>
GO:0044243	multicellular organismal catabolic process	0.000
GO:0030198	extracellular matrix organization	0.000
GO:0043062	extracellular structure organization	0.000
GO:0030574	collagen catabolic process	0.000
GO:0044236	multicellular organismal metabolic process	0.000
GO:0032963	collagen metabolic process	0.000
GO:0044259	multicellular organismal macromolecule metabolic process	0.000
GO:0022617	extracellular matrix disassembly	0.000
GO:0044712	single-organism catabolic process	0.000
GO:0022411	cellular component disassembly	0.000
GO:0009056	catabolic process	0.000
GO:0030199	collagen fibril organization	0.000
GO:0001501	skeletal system development	0.000
GO:0016043	cellular component organization	0.000
GO:0071840	cellular component organization or biogenesis	0.000
GO:0044710	single-organism metabolic process	0.000
GO:0007155	cell adhesion	0.000
GO:0022610	biological adhesion	0.000
GO:0007411	axon guidance	0.000
GO:0007409	axonogenesis	0.000
GO:0061564	axon development	0.000
GO:0031175	neuron projection development	0.000
GO:0048667	cell morphogenesis involved in neuron differentiation	0.000
GO:0000904	cell morphogenesis involved in differentiation	0.000
GO:0048812	neuron projection morphogenesis	0.000
GO:0048513	organ development	0.000
GO:0040011	locomotion	0.000
GO:0009888	tissue development	0.000
GO:0061448	connective tissue development	0.000
GO:0010951	negative regulation of endopeptidase activity	0.000
GO:0071230	cellular response to amino acid stimulus	0.000
GO:0048731	system development	0.000
GO:0010466	negative regulation of peptidase activity	0.000
GO:0006935	chemotaxis	0.000
GO:0042330	taxis	0.000
GO:0048666	neuron development	0.000
GO:0051216	cartilage development	0.000
GO:0006928	movement of cell or subcellular component	0.000
GO:0048858	cell projection morphogenesis	0.000
GO:0032990	cell part morphogenesis	0.000

GO:0009653	anatomical structure morphogenesis	0.000
GO:0000902	cell morphogenesis	0.000
GO:0048699	generation of neurons	0.000
GO:0030030	cell projection organization	0.000
GO:0045861	negative regulation of proteolysis	0.000
GO:0030182	neuron differentiation	0.000
GO:0022008	neurogenesis	0.001
GO:0032989	cellular component morphogenesis	0.001
GO:0035987	endodermal cell differentiation	0.001
GO:0048468	cell development	0.001
GO:0043200	response to amino acid	0.001
GO:0009605	response to external stimulus	0.001
GO:0009887	organ morphogenesis	0.001
GO:0052548	regulation of endopeptidase activity	0.001
GO:0030154	cell differentiation	0.001
GO:0071229	cellular response to acid chemical	0.002
GO:0051346	negative regulation of hydrolase activity	0.002
GO:0052547	regulation of peptidase activity	0.002
GO:0048869	cellular developmental process	0.003
GO:0048856	anatomical structure development	0.003
GO:0006897	endocytosis	0.003
GO:0001706	endoderm formation	0.003
GO:0042221	response to chemical	0.004
GO:0060350	endochondral bone morphogenesis	0.005
GO:0001101	response to acid chemical	0.005
GO:0060351	cartilage development involved in endochondral bone morphogenesis	0.007
GO:0006898	receptor-mediated endocytosis	0.007
GO:0001568	blood vessel development	0.012
GO:0048646	anatomical structure formation involved in morphogenesis	0.012
GO:0007275	multicellular organismal development	0.013
GO:0048705	skeletal system morphogenesis	0.014
GO:0001944	vasculature development	0.015
GO:0072358	cardiovascular system development	0.018
GO:0072359	circulatory system development	0.018
GO:0007492	endoderm development	0.023
GO:0043933	macromolecular complex subunit organization	0.027
GO:0070208	protein heterotrimerization	0.028
GO:0030168	platelet activation	0.028
GO:0001503	ossification	0.034
GO:0030162	regulation of proteolysis	0.041
GO:0090596	sensory organ morphogenesis	0.042
GO:0071822	protein complex subunit organization	0.044
GO:0071560	cellular response to transforming growth factor beta stimulus	0.048
GO:0071559	response to transforming growth factor beta	0.048

GO Cellular components	Term	<i>p</i> -value
GO:0005788	endoplasmic reticulum lumen	0.000
GO:0005578	proteinaceous extracellular matrix	0.000
GO:0031012	extracellular matrix	0.000
GO:0005581	collagen trimer	0.000
GO:0044420	extracellular matrix component	0.000
GO:0005583	fibrillar collagen trimer	0.000
GO:0044432	endoplasmic reticulum part	0.000
GO:0005783	endoplasmic reticulum	0.000
GO:0012505	endomembrane system	0.000
GO:0043233	organelle lumen	0.000
GO:0070013	intracellular organelle lumen	0.000
GO:0044421	extracellular region part	0.000
GO:0005604	basement membrane	0.000
GO:0005576	extracellular region	0.000

GO:0044446	intracellular organelle part	0.000
GO:0044422	organelle part	0.000
GO:0044444	cytoplasmic part	0.000
GO:0098644	complex of collagen trimers	0.000
GO:0043227	membrane-bounded organelle	0.000
GO:0043231	intracellular membrane-bounded organelle	0.000
GO:0005615	extracellular space	0.000
GO:0098642	network-forming collagen trimer	0.000
GO:0098645	collagen network	0.000
GO:0098651	basement membrane collagen trimer	0.000
GO:0072562	blood microparticle	0.000
GO:0043226	organelle	0.000
GO:0043229	intracellular organelle	0.000
GO:0070062	extracellular exosome	0.000
GO:0043234	protein complex	0.000
GO:0005737	cytoplasm	0.000
GO:0005588	collagen type V trimer	0.000
GO:0032991	macromolecular complex	0.000
GO:0005587	collagen type IV trimer	0.000
GO:0005593	FACIT collagen trimer	0.001
GO:0031988	membrane-bounded vesicle	0.001
GO:0031982	vesicle	0.001
GO:0005592	collagen type XI trimer	0.011
GO:0005584	collagen type I trimer	0.011
GO:0044424	intracellular part	0.017
GO:0005622	intracellular	0.017
GO:0030934	anchoring collagen complex	0.032

---

**Table S6. Amino acid variant analysis.** Amino acid positions identified in bone specimens analyzed here and of which the most frequent modern allele present in the 1000G database has a frequency of  $\leq 99.0\%$ . Column “Hominidae” presents the amino acid state for respectively *Pongo abelii*, *Gorilla gorilla* and *Pan troglodytes*.

Protein	Position	dbSNP	Hominidae	Neandertal	Denisovan	Human amino acids	Frequencies (1000G global)	Identified amino acid
<i>COL1a1</i>	1075	rs1800215	A/A/A	A	A	A, T	97.9, 2.1	A
<i>COL1a2</i>	549	rs42524	A/A/A	A	A	P, A	82.2, 17.8	A
<i>COL9a1</i>	621	rs1135056	Q/Q/Q	(?)	(?)	Q, R	67.3, 32.7	Q
<i>COL10a1</i>	128	rs142463796	D/D/D	N	N	D, N	99.1, 0.9	N
<i>COL28a1</i>	327	rs10486176	S/S/S	S	S	T, S	85.7, 14.3	S
<i>COL28a1</i>	472	rs17177927	P/P/P	P	P	A, P	93.5, 6.5	P
<i>COL28a1</i>	677	rs115211979	P/P/P	P	P	P, T	96.9, 3.1	P

**Table S7. *COL10a1* amino acid substitutions present within the Homininae.** Positions for which archaic data is lacking are omitted. States included in Hominidae are *Pongo abelii*/*Gorilla gorilla*/*Pan troglodytes*, respectively. Amino acid states for *Pongo* and *Gorilla* are given as indicators of the ancestral state, and unique substitutions on their phylogenetic lineages are excluded.

Position	Hominidae	Neandertals	Denisovan	Human	AR-30
27	T/T/T	T	M	M	-
61	I/I/I	T	T	T	T
128	D/D/D	N	N	D	N
155	S/S/S	P	P	P	-
214	P/P/P	S	S	S	-
243	M/M/I	M	M	M	-
287	P/P/A	P	P	P	-
354	N/N/N	S	S	S	-
361	K/K/E	K	K	K	-
393	E/E/E	K	K	K	-
592	R/R/R	X	Q	Q	-

**Table S8. BLAST results of amino acid sequence PKGNVGPAGLPGR to GenBank and UniProt.**

Database	Search	Top hit	Score	E-value	Identity (%)	Query coverage (%)
GenBank	pblast	<i>COL10a1 – Homo sapiens</i>	252	3.0E-07	94	100
UniProt	BLAST	<i>COL10a1 – Homo sapiens</i>	123	1.5E-11	94.4	100
GenBank – Archaea	pblast	Cobalamin biosynthesis protein CbiB - <i>Thermoproteus</i> sp. JCHS 4	39.9	19	80	71
GenBank – Bacteria	pblast	hypothetical protein - <i>Rhodopseudomonas palustris</i>	122	16	71	100
GenBank - Fungi	pblast	hypothetical protein - <i>Hypholoma sublateritium</i>	27.8	18	82	78

**Table S9. Frequency of the archaic-like allele of rs142463796 in modern human populations.** The frequency of the archaic-like allele in modern human populations that include individuals carrying the archaic-like allele (1000 Genomes and SGDP data set).

<b>Population</b>	<b>Dataset</b>	<b>Frequency of archaic allele in % (number of haplotypes with archaic allele/total number of haplotypes in population)</b>
Han Chinese in Beijing	1000 Genomes	0.5% (1/206)
Southern Han Chinese	1000 Genomes	1.4% (3/216)
Gujarati Indian	1000 Genomes	1.9% (4/212)
Punjabi	1000 Genomes	5.7% (11/192)
Bengali	1000 Genomes	6.4% (11/172)
Sri Lankan Tamil	1000 Genomes	2.4% (5/206)
Indian Telugu	1000 Genomes	4.9% (10/206)
Toscani	1000 Genomes	0.5% (1/216)
Tajik	SGDP	25% (1/4)
Australian	SGDP	33% (2/6)
Papuan	SGDP	46.9% (15/32)

**Table S10. General characteristics of the libraries enriched for mitochondrial DNA.** ENC – extraction negative control; LNC – library negative control; ddPCR – digital droplet PCR; C – cytosine; T - thymine.

Library ID	Specimen ID	Powder used for extraction (mg)	molecules in library (ddPCR)	sequences generated	Sequences mapped (%)	unique mtDNA sequences	Average number of duplicates	# of sequences with C→T substitution
A9340	AR-14	8.2	2.82E+09	2,624,306	3.52	48,469	8.71	4,083
A9342	AR-30	22	2.57E+09	2,520,378	0.42	7,015	14.86	850
A9344	ENC	-	2.29E+07	343,405	0.17	208	272.11	3
A9345	LNC	-	1.28E+07	273,670	0.03	15	890.27	1

**Table S11. Frequencies of C to T substitutions at the terminal positions of sequence alignments.**

Specimen ID	5' C→T (%) (95% CI) [number of observations]	3' C → T (%) (95% CI) [number of observations]
AR-14	12.1 (11.5-12.7) [1,449/11,978]	11.5 (10.85-12.19) [1,007/8,756]
AR-30	20.5 (18.6-22.6) [325/1,583]	28.1 (24.9-31.5) [202/720]

**Table S12. Frequencies of C to T substitutions at the terminal positions of sequenced DNA fragments identified as present-day human and Neandertal.**

Specimen ID	Human sequences		Neandertal sequences	
	5' C→T (%) (# of observations)	3' C → T (%) (# of observations)	5' C→T (%) (# of observations)	3' C → T (%) (# of observations)
AR-14	1.8 (39/2192)	2.6 (42/1,624)	29.7 (167/562)	24.7 (97/393)
AR-30	3.4 (9/267)	4.4 (6/137)	33.3 (43/129)	71.4 (40/56)

**Table S13. Linear measurements (in mm) of AR-7, a partial cervical hemi-arch.** Measured variables are abbreviated following (68).

<b>Variable abbreviation</b>	<b>AR-7</b>	<b>LF-8, C3/C4 (MAN-1970)</b>	<b>LF-8, C3-C5 (MAN-1970)</b>	<b>LF-8, C6 (MAN-1970)</b>
LamCrCdDi	4.7	5.3	4.3	-
LamTh	2.6	2.1	2.1	2.3

**Table S14. Isotopic results for AR-14 and 35 additional faunal specimens from the Grotte du Renne.**  $\delta^{13}\text{C}$  values are measured relative to the V-PDB standard and  $\delta^{15}\text{N}$  values are measured relative to the AIR standard.

S-EVA	Taxonomy	Layer	% Coll.	$\delta^{13}\text{C}$	$\delta^{15}\text{N}$	%C	%N	C:N	Reference
S-EVA 32886	Hominin (AR-14)	X(c?)	3.9	-19.5	14.3	40.1	14.5	3.2	this study
S-EVA 32867	Bos/Bison	Xb	3.7	-20.3	5.9	41.8	15.3	3.2	this study
S-EVA 32870	Bos/Bison	Xb	3.3	-21.3	7.1	41.6	15.2	3.2	this study
S-EVA 32873	Bos/Bison	Xb	3.1	-20.0	5.1	42.7	15.7	3.2	this study
S-EVA 32874	Bos/Bison	Xb	6.1	-19.0	3.5	43.3	15.8	3.2	this study
S-EVA 32868	Equidae	Xb	8.2	-20.5	4.6	41.0	15.1	3.2	this study
S-EVA 32871	Equidae	Xb	1.1	-20.9	5.1	39.7	14.4	3.2	this study
S-EVA 32915	Equidae	Xb	7.2	-20.8	5.2	44.9	16.4	3.2	this study
S-EVA 32907	Equidae	Xb2	6.1	-20.9	5.2	45.2	16.4	3.2	this study
EVA-44	Equidae	IXa	1.6	-20.7	6.1	35.2	13.0	3.2	(8)
EVA-47	Equidae	IXa	4.0	-21.3	4.7	39.4	14.6	3.2	(8)
EVA-38	Equidae	Xa	1.5	-20.8	5.7	38.6	14.3	3.2	(8)
EVA-23	Equidae	Xb1	2.6	-21.1	5.9	42.5	15.8	3.1	(8)
EVA-29	Equidae	Xb2	1.7	-20.9	5.6	40.4	14.8	3.2	(8)
EVA-32	Equidae	Xb2	5.1	-21.0	5.1	40.6	15.0	3.2	(8)
S-EVA 32910	Rangifer	Xb2	5.8	-18.8	3.3	47.3	17.2	3.2	this study
S-EVA 32904	Rangifer	Xb2	4.8	-19.3	7.8	39.6	14.4	3.2	this study
S-EVA 32906	Rangifer	Xb2	3.1	-19.4	5.5	19.2	6.7	3.3	this study
EVA-33	Rangifer	IXb	4.4	-18.7	3.6	39.1	14.5	3.1	(8)
EVA-34	Rangifer	IXb	4.5	-19.0	4.6	39.3	14.5	3.2	(8)
EVA-35	Rangifer	IXb	3.4	-18.6	3.1	39.5	14.5	3.2	(8)
EVA-36	Rangifer	IXb	3.5	-19.3	3.3	45.6	16.7	3.2	(8)
EVA-37	Rangifer	IXb	4.3	-18.5	3.7	42.0	15.4	3.2	(8)
EVA-27	Rangifer	Xb1	4.4	-18.7	3.5	41.9	15.4	3.2	(8)
EVA-48	Rangifer	Xb2	3.5	-18.6	3.0	43.2	15.4	3.3	(8)
EVA-49	Rangifer	Xb2	4.4	-18.9	3.9	41.9	15.5	3.1	(8)
S-EVA 32916	Rhinocerotidae	Xc	8.0	-18.9	2.4	46.7	17.0	3.2	this study
S-EVA 32918	Rhinocerotidae	Xc	0.0	failed	failed	failed	failed	failed	this study
S-EVA 32913	Elephantidae	Xb2	1.8	-20.2	4.9	44.2	16.0	3.2	this study
S-EVA 32914	Elephantidae	Xb	5.6	-21.8	8.2	47.7	16.5	3.4	this study
S-EVA 32911	Elephantidae	Xb2	1.9	-21.5	6.8	20.1	7.0	3.3	this study
S-EVA 32919	Hyaenidae	Xc	10.3	-19.5	9.5	46.4	16.9	3.2	this study
S-EVA 32908	Pantherinae	Xb2	11.0	-18.8	8.9	48.0	17.6	3.2	this study
S-EVA 32909	Pantherinae	Xb2	8.6	-18.0	7.4	42.8	15.7	3.2	this study
S-EVA 32912	Canidae	Xb2	5.6	-18.9	9.9	46.5	16.8	3.2	this study
S-EVA 32917	Ursidae	Xc	3.9	-21.1	3.2	29.7	10.6	3.3	this study

**Table S15. Average isotopic values and standard deviations for different animal categories at the Grotte du Renne (Layers IX and X).**

Animal category	#	$\delta^{13}\text{C}$		$\delta^{15}\text{N}$		Reference
		Average	SD	Average	SD	
AR-14 Hominin	1	-19.5	-	14.3	-	this study
Herbivores	29	-20.0	1.1	5.0	2.3	(8); this study
Rangifer	11	-18.9	0.3	4.1	1.4	(8); this study
Equidae & Bos	14	-20.7	0.6	5.3	0.6	(8); this study
Omnivore	1	-21.1	-	3.2	-	this study
Carnivores	4	-18.8	0.6	8.9	1.1	this study

**Table S16. Bayesian age models for the Grotte du Renne.** Calibrated boundaries age range provided by OxCal 4.2 (122) using the International Calibration Curve IntCal13 (123) for the four different models (see text for details). Bayesian CQL Code is given below for individual models.

<b>Hublin <i>et al.</i> 2012 model (8)</b>		<b>68.2%</b>		<b>95.4%</b>	
Indices: A_overall 105.1					
End Protoaurignacian VII		<i>from</i>	<i>to</i>	<i>from</i>	<i>to</i>
		34,230	32,760	35,400	30,400
Transition Châtelperronian VIII/Protoaurignacian VII		40,810	39,940	41,050	39,380
Transition Châtelperronian Layer IX+X/VIII		41,280	40,760	41,470	40,500
AR-14 (MAMS-25149; 36840,660)		41,990	41,170	42,510	40,820
Transition Mousterian XI/Châtelperronian Layer IX+X		44,730	44,150	45,050	43,900
Start Mousterian XI		46,750	45,160	48,290	44,670
Saint-Césaire OxA-18099 (36200,750)		41,540	40,110	42,140	39,340
<b>Only hominin-modified bones (n=10 samples in Châtelperronian layers, dates from (8))</b>					
Indices: A_overall 92.5					
End Protoaurignacian VII		<i>from</i>	<i>to</i>	<i>from</i>	<i>to</i>
		38,730	36,270	41,500	32,720
Transition Châtelperronian VIII/Protoaurignacian VII		41,790	40,090	41,890	38,490
Transition Châtelperronian Layer IX+X/VIII		42,140	41,610	42,370	41,250
AR-14 (MAMS-25149; 36840,660)		42,450	41,770	43,160	41,320
Transition Mousterian XI/Châtelperronian Layer IX+X		44,890	44,130	45,280	43,720
Start Mousterian XI		46,100	44,770	47,590	44,360
Saint-Césaire OxA-18099 (36200,750)		41,560	40,120	42,150	39,350
<b>Discamps <i>et al.</i> model (n=24 samples in the Châtelperronian layers, dates from (8))</b>					
Indices: A_overall 90.2					
End Protoaurignacian VII		<i>from</i>	<i>to</i>	<i>from</i>	<i>to</i>
		34,210	32,760	34,540	30,080
Transition Châtelperronian VIII/Protoaurignacian VII		40,860	39,900	41,100	39,340
Transition Châtelperronian Layer IX+X/VIII		41,390	40,960	41,590	40,720
AR-14 (MAMS-25149; 36840,660)		42,000	41,270	42,480	40,980
Transition Mousterian XI/Châtelperronian Layer IX+X		44,820	44,220	45,170	43,940
Start Mousterian XI		47,430	44,950	50,690	44,460
Saint-Césaire OxA-18099 (36200,750)		41,550	40,130	42,150	39,330
<b>Higham <i>et al.</i> 2014 model (11, dates from (5) and (8))</b>					
Indices: A_overall 7.5					
End Protoaurignacian VII		<i>from</i>	<i>to</i>	<i>from</i>	<i>to</i>
		39,690	38,360	40,220	37,330
Transition Châtelperronian VIII/Protoaurignacian VII		40,610	39,690	41,230	39,290
Transition Châtelperronian Layer IX+X/VIII		40,980	40,010	41,800	39,790
AR-14 (MAMS-25149; 36840,660)		42,080	41,050	42,560	40,460
Transition Mousterian XI/Châtelperronian Layer IX+X		44,840	44,180	45,160	43,810
Start Mousterian XI		45,260	44,450	45,810	43,980

## **Bayesian CQL Code**

### **1) Hublin *et al.* Model:**

```
Options()
{
  Outlier_Model("General",T(5),U(0,4),"t");
  Resolution=20;
  Plot("Saint Cesaire")
  {
    R_Date("OxA-18099", 36200, 750)
    {
      color="red";
      fill="Red";
    };
  };
  Plot()
  {
    Sequence(Arcy-sur-Cure)
    {
      Boundary("Start Mousterian XI");
      Phase("Mousterian XI")
      {
        R_Date("EVA-84", 43266, 929)
        {
          Outlier(0.05);
        };
        R_Date("EVA-77", 42122, 805)
        {
          Outlier(0.05);
        };
        R_Date("EVA-83", 41979, 821)
        {
          Outlier(0.05);
        };
        R_Date("EVA-85", 40898, 719)
        {
          Outlier(0.05);
        };
      };
      Boundary("Transition Mousterian XI/CP Layer IX+X");
      Phase("Chatelperronian Layer")
      {
        R_Date("EVA-33", 40968, 424)
        {
          Outlier(0.05);
        };
        R_Date("EVA-28", 40925, 393)
        {
          Outlier(0.05);
        };
        R_Date("EVA-49", 40834, 778)
        {
          Outlier(0.05);
        };
        R_Date("EVA-34", 40519, 389)
        {
          Outlier(0.05);
        };
        R_Date("EVA-27", 40231, 395)
        {
          Outlier(0.05);
        };
        R_Date("EVA-51", 39958, 702)
        {
          Outlier(0.05);
        };
        R_Date("EVA-46", 39932, 361)
        {
          Outlier(0.05);
        };
      };
    };
  };
}
```

```

};
R_Date("EVA-47", 39754, 360)
{
  Outlier(0.05);
};
R_Date("EVA-37", 39448, 340)
{
  Outlier(0.05);
};
R_Date("EVA-26", 39393, 334)
{
  Outlier(0.05);
};
R_Date("EVA-31", 39290, 334)
{
  Outlier(0.05);
};
R_Date("EVA-44", 39277, 351)
{
  Outlier(0.05);
};
R_Date("EVA-35", 39243, 341)
{
  Outlier(0.05);
};
R_Date("EVA-48", 39071, 332)
{
  Outlier(0.05);
};
R_Date("EVA-43", 39015, 352)
{
  Outlier(0.05);
};
R_Date("EVA-41", 38733, 333)
{
  Outlier(0.05);
};
R_Date("EVA-24", 38395, 317)
{
  Outlier(0.05);
};
R_Date("EVA-42", 38065, 311)
{
  Outlier(0.05);
};
R_Date("EVA-30", 37984, 284)
{
  Outlier(0.05);
};
R_Date("EVA-36", 37742, 307)
{
  Outlier(0.05);
};
R_Date("EVA-40", 37512, 275)
{
  Outlier(0.05);
};
R_Date("MAMS-25149", 36840, 660)
{
  color="red";
  fill="red";
  Outlier(0.05);
};
R_Date("EVA-23", 36837, 335)
{
  Outlier(0.05);
};
R_Date("EVA-32", 36815, 257)
{
  Outlier(0.05);
};

```

```

};
R_Date("EVA-38", 36536, 248)
{
  Outlier(0.05);
};
R_Date("EVA-25", 36207, 250)
{
  Outlier(0.05);
};
R_Date("EVA-29", 35498, 216)
{
  Outlier(1.0);
};
};
Boundary("Transition CP Layer IX+X/VIII");
Phase("Chatelperronian Layer VIII")
{
  R_Date("EVA-56", 37712, 533)
  {
    Outlier(1.0);
  };
  R_Date("EVA-55", 36626, 452)
  {
    Outlier(0.05);
  };
  R_Date("EVA-53", 36232, 435)
  {
    Outlier(0.05);
  };
  R_Date("EVA-52", 35984, 432)
  {
    Outlier(0.05);
  };
  R_Date("EVA-54", 35379, 390)
  {
    Outlier(0.05);
  };
};
Boundary("Transition CP VIII/Proto VII");
Phase("Protoaurignacian VII")
{
  R_Date("EVA-95", 34807, 210)
  {
    Outlier(0.05);
  };
  R_Date("EVA-81", 33849, 311)
  {
    Outlier(0.05);
  };
  R_Date("EVA-93", 33007, 182)
  {
    Outlier(0.05);
  };
  R_Date("EVA-92", 31610, 185)
  {
    Outlier(0.05);
  };
  R_Date("EVA-79", 29934, 208)
  {
    Outlier(0.05);
  };
};
Boundary("End Protoaurignacian VII");
};
};
};

```

**2) Cut-Marks bones Model: A\_overall 92.5%**  
Options()

```

{
Outlier_Model("General",T(5),U(0,4),"t");
Resolution=20;
Plot("Saint Cesaire")
{
R_Date("OxA-18099", 36200, 750)
{
color="red";
fill="Red";
};
};
Plot()
{
Sequence(Arcy-sur-Cure)
{
Boundary("Start Mousterian XI");
Phase("Mousterian XI")
{
R_Date("EVA-77", 42122, 805)
{
Outlier(0.05);
};
};
R_Date("EVA-83", 41979, 821)
{
Outlier(0.05);
};
};
R_Date("EVA-85", 40898, 719)
{
Outlier(0.05);
};
};
Boundary("Transition Mousterian XI/CP Layer IX+X");
Phase("Chatelperronian Layer")
{
R_Date("EVA-33", 40968, 424)
{
Outlier(0.05);
};
};
R_Date("EVA-34", 40519, 389)
{
Outlier(0.05);
};
};
R_Date("EVA-26", 39393, 334)
{
Outlier(0.05);
};
};
R_Date("EVA-41", 38733, 333)
{
Outlier(0.05);
};
};
R_Date("EVA-42", 38065, 311)
{
Outlier(0.05);
};
};
R_Date("EVA-30", 37984, 284)
{
Outlier(0.05);
};
};
R_Date("MAMS-25149", 36840, 660)
{
color="red";
fill="red";
Outlier(0.05);
};
};
R_Date("EVA-29", 35498, 216)
{
Outlier(1.0);
};
};
};
Boundary("Transition CP Layer IX+X/VIII");

```

```

Phase("Chatelperronian Layer VIII")
{
  R_Date("EVA-56", 37712, 533)
  {
    Outlier(0.05);
  };
  R_Date("EVA-55", 36626, 452)
  {
    Outlier(0.05);
  };
};
Boundary("Transition CP VIII/Proto VII");
Phase("Protoaurignacian VII")
{
  R_Date("EVA-81", 33849, 311)
  {
    Outlier(0.05);
  };
};
Boundary("End Protoaurignacian VII");
};
};
};

```

### 3) Discamp *et al.* Model: A\_overall 90.2%

```

Options()
{
  Outlier_Model("General",T(5),U(0,4),"t");
  Resolution=20;
  Plot("Saint Cesaire")
  {
    R_Date("OxA-18099", 36200, 750)
    {
      color="red";
      fill="Red";
    };
  };
  Plot()
  {
    Sequence(Arcy-sur-Cure)
    {
      Boundary("Start Mousterian XI");
      Phase("Mousterian XI")
      {
        R_Date("EVA-84", 43266, 929)
        {
          Outlier(0.05);
        };
        R_Date("EVA-85", 40898, 719)
        {
          Outlier(0.05);
        };
      };
      Boundary("Transition Mousterian XI/CP Layer IX+X");
      Phase("Chatelperronian Layer")
      {
        R_Date("EVA-33", 40968, 424)
        {
          Outlier(0.05);
        };
        R_Date("EVA-28", 40925, 393)
        {
          Outlier(0.05);
        };
        R_Date("EVA-49", 40834, 778)
        {
          Outlier(0.05);
        };
        R_Date("EVA-34", 40519, 389)

```

```

{
  Outlier(0.05);
};
R_Date("EVA-27", 40231, 395)
{
  Outlier(0.05);
};
R_Date("EVA-46", 39932, 361)
{
  Outlier(0.05);
};
R_Date("EVA-47", 39754, 360)
{
  Outlier(0.05);
};
R_Date("EVA-37", 39448, 340)
{
  Outlier(0.05);
};
R_Date("EVA-44", 39277, 351)
{
  Outlier(0.05);
};
R_Date("EVA-35", 39243, 341)
{
  Outlier(0.05);
};
R_Date("EVA-48", 39071, 332)
{
  Outlier(0.05);
};
R_Date("EVA-43", 39015, 352)
{
  Outlier(0.05);
};
R_Date("EVA-24", 38395, 317)
{
  Outlier(0.05);
};
R_Date("EVA-30", 37984, 284)
{
  Outlier(0.05);
};
R_Date("EVA-36", 37742, 307)
{
  Outlier(0.05);
};
R_Date("MAMS-25149", 36840, 660)
{
  color="red";
  fill="red";
  Outlier(0.05);
};
R_Date("EVA-23", 36837, 335)
{
  Outlier(0.05);
};
R_Date("EVA-32", 36815, 257)
{
  Outlier(0.05);
};
R_Date("EVA-25", 36207, 250)
{
  Outlier(0.05);
};
};
Boundary("Transition CP Layer IX+X/VIII");
Phase("Chatelperronian Layer VIII")
{
  R_Date("EVA-56", 37712, 533)
}

```

```

{
  Outlier(1.0);
};
R_Date("EVA-55", 36626, 452)
{
  Outlier(0.05);
};
R_Date("EVA-53", 36232, 435)
{
  Outlier(0.05);
};
R_Date("EVA-52", 35984, 432)
{
  Outlier(0.05);
};
R_Date("EVA-54", 35379, 390)
{
  Outlier(0.05);
};
};
Boundary("Transition CP VIII/Proto VII");
Phase("Protoaurignacian VII")
{
  R_Date("EVA-95", 34807, 210)
  {
    Outlier(0.05);
  };
  R_Date("EVA-81", 33849, 311)
  {
    Outlier(0.05);
  };
  R_Date("EVA-93", 33007, 182)
  {
    Outlier(0.05);
  };
  R_Date("EVA-92", 31610, 185)
  {
    Outlier(0.05);
  };
  R_Date("EVA-79", 29934, 208)
  {
    Outlier(0.05);
  };
};
};
Boundary("End Protoaurignacian VII");
};
};
};

```

#### 4) Higham *et al.* Model: A\_overall 7.5%

```

Options()
{
  Resolution=20;
  Curve("IntCal13", "IntCal13.14c");
  Plot()
  {
    Outlier_Model("General", T(5), U(0,4), "t");
  };
  Sequence()
  {
    Boundary("XII");
    Phase("XII Mousterian")
    {
      R_F14C("OxA-21594", 0.00996, 0.00126)
      {
        Outlier("General", 0.05);
      };
      R_F14C("OxA-21595", 0.00862, 0.00127)
      {
        Outlier("General", 0.05);
      };
    };
  };
};

```

```

};
};
Boundary("Start XI");
Phase("XI Mousterian")
{
R_Date("EVA-77*", 42120, 805)
{
Outlier("General", 0.05);
};
R_Date("EVA-83*", 41980, 821)
{
Outlier("General", 0.05);
};
R_Date("EVA-85*", 40900, 719)
{
Outlier("General", 0.05);
};
};
Boundary("end XI/Start X");
Phase("X + IX Chatelperronian")
{
R_Date("EVA-30*", 37980, 284)
{
Outlier("General", 0.05);
};
R_Date("MAMS-25149", 36840, 660)
{
color="red";
fill="red";
Outlier(0.05);
};
R_Date("EVA-29*", 35500, 216)
{
Outlier("General", 1.0);
};
R_Date("EVA-26*", 39390, 334)
{
Outlier("General", 0.05);
};
R_Date("EVA-42*", 38070, 311)
{
Outlier("General", 0.05);
};
R_Date("EVA-41*", 38730, 333)
{
Outlier("General", 0.05);
};
R_F14C("OxA-21576",0.00621,0.00131)
{
Outlier("General", 0.05);
};
R_F14C("OxA-21577",0.01338,0.00132)
{
Outlier("General", 0.05);
};
R_Date("OxA-21590",21150,160)
{
Outlier("General", 1.00);
};
R_F14C("OxA-21591",0.0132,0.00126)
{
Outlier("General", 0.05);
};
R_F14C("OxA-21565",0.00895,0.00099)
{
Outlier("General", 0.05);
};
R_F14C("OxA-21593",0.01231,0.00132)
{
Outlier("General", 0.05);
};

```

```

};
R_F14C("OxA-X-2279-18",0.00639,0.001)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-X-2279-45",0.00618,0.00102)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-X-2279-46",0.00807,0.001)
{
  Outlier("General", 0.05);
};
R_Date("OxA-X-2222-21", 23120, 190)
{
  Outlier("General", 1.00);
};
R_F14C("OxA-X-2226-7",0.0083,0.00134)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-21577",0.01338,0.00132)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-21592",0.01102,0.00149)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-X-2226-12",0.0057,0.00137)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-X-2226-13",0.00782,0.00136)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-X-2279-44",0.00233,0.00104)
{
  Outlier("General", 0.05);
};
R_Date("EVA-34*", 40520, 389)
{
  Outlier("General", 0.05);
};
R_Date("EVA-33*", 40970, 424)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-21574",0.00794,0.00127)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-21575",0.01837,0.00129)
{
  Outlier("General", 0.05);
};
Boundary("end X/Start IX");
Phase("VIII")
{
  R_Date("EVA-56*", 37710, 533)
  {
    Outlier("General", 1.0);
  };
  R_Date("EVA-55*", 36630, 452)
  {
    Outlier("General", 0.05);
  };
};
R_F14C("OxA-21573",0.01018,0.0013)

```

```

{
  Outlier("General", 0.05);
};
R_F14C("OxA-X-2279-14",0.01215,0.00112)
{
  Outlier("General", 0.05);
};
R_F14C("OxA-21683",0.00684,0.00099)
{
  Outlier("General", 0.05);
};
Boundary("VIII/VII");
Phase("VII Aurignacian")
{
  R_F14C("OxA-21569",0.01062,0.00166)
  {
    Outlier("General", 0.05);
};
  R_F14C("OxA-21570",0.01347,0.00137)
  {
    Outlier("General", 0.05);
};
  R_F14C("OxA-21571",0.01444,0.00132)
  {
    Outlier("General", 0.05);
};
  R_F14C("OxA-21572",0.01343,0.00127)
  {
    Outlier("General", 0.05);
};
  R_F14C("OxA-21682",0.0128,0.001)
  {
    Outlier("General", 0.05);
};
};
Boundary("VII/VI");
};
Sequence()
{
  Boundary("=end XI/Start X");
  Date("Arcy Chatelperronian");
  Boundary("=VIII/VII");
};
Sequence()
{
  Boundary("=VIII/VII");
  Date("Arcy Protoaurignacian");
  Boundary("=VII/VI");
};
Sequence()
{
  Boundary("=XII");
  Date("Arcy Mousterian");
  Boundary("=end XI/Start X");
};
};
};
};

```

## References:

1. Leroi-Gourhan A (1961) Les fouilles d'Arcy-sur-Cure (Yonne). *Gallia préhistoire* 4(1):3–16.
2. Leroi-Gourhan A, Leroi-Gourhan A (1964) Chronologie des grottes d'Arcy-sur-Cure (Yonne). *Gallia Préhistoire* 7(1):1–64.
3. Hublin J-J, Spoor F, Braun M, Zonneveld F, Condemi S (1996) A late Neanderthal associated with Upper Palaeolithic artefacts. *Nature* 381(6579):224–226.
4. d'Errico F, Julien M, Liolios D, Vanhaeren M, Baffier D (2003) Many awls in our argument. Bone tool manufacture and use in the Châtelperronian and Aurignacian levels of the Grotte du Renne at Arcy-sur-Cure. *The Chronology of the Aurignacian and of the Transitional Technocomplexes. Dating, Stratigraphies, Cultural Implications, Trabalhos de arqueologia.*, eds Zilhão J, d'Errico F (Instituto Português de Arqueologia, Lisboa), pp 247–270.
5. Higham T, et al. (2010) Chronology of the Grotte du Renne (France) and implications for the context of ornaments and human remains within the Châtelperronian. *Proc Natl Acad Sci USA* 107(47):20234–20239.
6. Bar-Yosef O, Bordes J-G (2010) Who were the makers of the Châtelperronian culture? *J Hum Evol* 59(5):586–593.
7. David F, et al. (2001) Le Châtelperronien de la grotte du Renne à Arcy-sur-Cure (Yonne). Données sédimentologiques et chronostratigraphiques. *Bulletin de la Société préhistorique française* 98(2):207–230.
8. Hublin J-J, et al. (2012) Radiocarbon dates from the Grotte du Renne and Saint-Césaire support a Neandertal origin for the Châtelperronian. *Proc Natl Acad Sci USA* 109(46):18743–18748.
9. Discamps E, Gravina B, Teyssandier N (2015) In the eye of the beholder: contextual issues for Bayesian modelling at the Middle-to-Upper Palaeolithic transition. *World Archaeol* 47(4):601–621.
10. Movius HL (1969) The Châtelperronian in French Archaeology: the Evidence of Arcy-sur-Cure. *Antiquity* 43(170):111–123.
11. Higham T, et al. (2014) The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* 512(7514):306–309.
12. Conard NJ, Grootes PM, Smith FH (2004) Unexpectedly recent dates for human remains from Vogelherd. *Nature* 430(6996):198–201.
13. Semal P, et al. (2009) New data on the late Neandertals: direct dating of the Belgian Spy fossils. *Am J Phys Anthropol* 138(4):421–428.
14. Pinhasi R, Higham TFG, Golovanova LV, Doronichev VB (2011) Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *Proc Natl Acad Sci USA* 108(21):8611–8616.
15. Moreau L, et al. (2015) Reassessing the Aurignacian of Slovenia: techno-economic behaviour and direct dating of osseous projectile points. *J Hum Evol* 78:158–180.
16. Bodu P (1990) L'application de la méthode des remontages à l'étude du matériel lithique des premiers niveaux châtelperroniens d'Arcy-sur-Cure. *Paléolithique Moyen Récent et Paléolithique Supérieur Ancien En Europe. Ruptures et Transitions : Examen Critique Des Documents Archéologiques.*, ed Farzy C (Mémoires du Musée de Préhistoire d'Île de France), pp 309–312.
17. Bailey SE, Hublin J-J (2006) Dental remains from the Grotte du Renne at Arcy-sur-Cure (Yonne). *J Hum Evol* 50(5):485–508.
18. Caron F, d'Errico F, Del Moral P, Santos F, Zilhão J (2011) The reality of Neanderthal symbolic behavior at the Grotte du Renne, Arcy-sur-Cure, France. *PLoS One* 6(6):e21545.
19. van Doorn NL, Hollund H, Collins MJ (2011) A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. *Archaeol Anthropol Sci* 3(3):281–289.
20. Buckley M, Collins M, Thomas-Oates J, Wilson JC (2009) Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 23(23):3843–3854.
21. Buckley M, Kansa SW (2011) Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. *Archaeol Anthropol Sci* 3(3):271–280.

22. Evans S, et al. (2015) Using combined biomolecular methods to explore whale exploitation and social aggregation in hunter–gatherer–fisher society in Tierra del Fuego. *J Archaeol Sci: Reports* 6:757–767.
23. Brown S, et al. (2016) Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci Rep* 6:23559.
24. Welker F, Soressi M, Rendu W, Hublin J-J, Collins MJ (2015) Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. *J Archaeol Sci* 54:279–286.
25. Wilson J, van Doorn NL, Collins MJ (2012) Assessing the extent of bone degradation using glutamine deamidation in collagen. *Anal Chem* 84(21):9041–9048.
26. Cappellini E, et al. (2012) Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J Proteome Res* 11(2):917–926.
27. Welker F, et al. (2015) Ancient proteins resolve the evolutionary history of Darwin’s South American ungulates. *Nature* 522(7554):81–84.
28. Chambers MC, et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 30(10):918–920.
29. Ma B, et al. (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 17(20):2337–2342.
30. von Koenigswald W (2006) Climatic changes, faunal diversity, and environment of the Neanderthals in Central and Western Europe during the middle and upper Pleistocene. *Terra Nostra* 2:35–40.
31. Kahlke R-D (2014) The origin of Eurasian Mammoth Faunas (Mammuthus–Coelodonta Faunal Complex). *Quat Sci Rev* 96:32–49.
32. Curren A, Jacobi R (2001) A formal mammalian biostratigraphy for the Late Pleistocene of Britain. *Quat Sci Rev* 20(16–17):1707–1716.
33. Pettitt P, White M (2012) *The British Palaeolithic: human societies at the edge of the Pleistocene world* (Routledge).
34. Stewart JR (2007) Neanderthal extinction as part of the faunal change in Europe during Oxygen Isotope Stage 3. *Acta Zoologica Cracoviensia - Series A: Vertebrata* 50(1-1):93–124.
35. Stewart JR, Van Kolfschoten M, Markova A, Musil R (2003) The mammalian faunas of Europe during oxygen isotope stage three. *Neanderthals and Modern Humans in the European Landscape during the Last Glaciation*, ed T.H. van Andel WD (Cambridge: McDonald Institute for Archaeological Research), pp 103–130.
36. Lister AM, et al. (2005) The phylogenetic position of the “giant deer” *Megaloceros giganteus*. *Nature* 438(7069):850–853.
37. Teacher A, Thomas J, Barnes I (2011) Modern and ancient red fox (*Vulpes vulpes*) in Europe show an unusual lack of geographical and temporal structuring, and differing responses within the carnivores to historical climatic change. *BMC Evol Biol* 11(1):214.
38. Reumer JWF, et al. (2003) Late Pleistocene survival of the saber-toothed cat *Homotherium* in northwestern Europe. *J Vert Paleontol* 23(1):260–262.
39. Stuart AJ, Lister AM (2007) Patterns of Late Quaternary megafaunal extinctions in Europe and northern Asia. *Courier-Forschungsinstitut Senckenberg* 259:287–297.
40. Ashby SP, Coutu AN, Sindbæk SM (2015) Urban Networks and Arctic Outlands: Craft Specialists and Reindeer Antler in Viking Towns. *European Journal of Archaeology* 18(4):679–704.
41. Kirby DP, Buckley M, Promise E, Trauger SA, Holdcraft TR (2013) Identification of collagen-based materials in cultural heritage. *Analyst* 138(17):4849–4858.
42. Campana MG, Robinson T, Campos PF, Tuross N (2013) Independent confirmation of a diagnostic sheep/goat peptide sequence through DNA analysis and further exploration of its taxonomic utility within the Bovidae. *J Archaeol Sci* 40(2):1421–1424.
43. Buckley M, et al. (2014) Species identification of archaeological marine mammals using collagen fingerprinting. *J Archaeol Sci* 41:631–641.

44. Perry GH, et al. (2013) Aye-aye population genomic analyses highlight an important center of endemism in northern Madagascar. *Proc Natl Acad Sci USA* 110(15):5823–5828.
45. Prado-Martinez J, et al. (2013) Great ape genetic diversity and population history. *Nature* 499(7459):471–475.
46. Buckley M, Larkin N, Collins M (2011) Mammoth and Mastodon collagen sequences; survival and utility. *Geochim Cosmochim Acta* 75(7):2007–2016.
47. Korlević P, et al. (2015) Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* 59(2):87–93.
48. Dabney J, et al. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA* 110(39):15758–15763.
49. Gansauge M-T, Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* 8(4):737–748.
50. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):db.prot5448.
51. Dabney J, Meyer M (2012) Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52(2):87–94.
52. Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40(1):1–8.
53. Fu Q, et al. (2013) DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci USA* 110(6):2223–2227.
54. Renaud G, Stenzel U, Kelso J (2014) leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res* 42(18):e141.
55. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26(5):589–595.
56. Meyer M, et al. (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222–226.
57. Talamo S, Richards M (2011) A comparison of bone pretreatment methods for AMS dating of samples > 30,000 BP. *Radiocarbon* 53(3):443–449.
58. Ambrose SH (1990) Preparation and characterization of bone and tooth collagen for isotopic analysis. *J Archaeol Sci* 17(4):431–451.
59. van Klinken G (1999) Bone Collagen Quality Indicators for Palaeodietary and Radiocarbon Measurements. *J Archaeol Sci* 26(6):687–695.
60. Kromer B, Lindauer S, Synal H-A, Wacker L (2013) MAMS-a new AMS facility at the Curt-Engelhorn-Centre for Archaeometry, Mannheim, Germany. *Nucl Instrum Methods Phys Res B* 294:11–13.
61. Klein RG (2009) *The human career: human biological and cultural origins* (University of Chicago Press).
62. van Doorn NL, Wilson J, Hollund H, Soressi M, Collins MJ (2012) Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun Mass Spectrom* 26(19):2319–2327.
63. Simpson JP, et al. (2016) The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. *J Archaeol Sci* 69:29–38.
64. Schroeter ER, Cleland TP (2016) Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Commun Mass Spectrom* 30(2):251–255.
65. Schaefer M, Black SM, Scheuer L (2009) *Juvenile osteology: a laboratory and field manual* (Academic).
66. Castellana C, Kósa F (1999) Morphology of the cervical vertebrae in the fetal-neonatal human skeleton. *J Anat* 194(1):147–152.

67. Gómez-Olivencia A, Crevecoeur I, Balzeau A (2015) La Ferrassie 8 Neandertal child reloaded: New remains and re-assessment of the original collection. *J Hum Evol* 82:107–126.
68. Gómez-Olivencia A, Been E, Arsuaga JL, Stock JT (2013) The Neandertal vertebral column 1: the cervical spine. *J Hum Evol* 64(6):608–630.
69. Wadsworth C, Buckley M (2014) Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun Mass Spectrom* 28(6):605–615.
70. Kendall R, Hendy J, Collins MJ, Millard AR, Gowland RL (2016) Poor preservation of antibodies in archaeological human bone and dentine. *STAR: Science & Technology of Archaeological Research* 2(1):15–24.
71. Long F (2012) Building strong bones: molecular regulation of the osteoblast lineage. *Nat Rev Mol Cell Biol* 13(1):27–38.
72. Orlando L, et al. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499(7456):74–78.
73. Mikšik I, et al. (2014) Prince Cangrande's Collagen: Study of Protein Modification on the Mummy of the Lord of Verona, Italy (1291–1329 AD). *Chromatographia* 77(21-22):1503–1510.
74. Hill RC, et al. (2015) Preserved Proteins from Extinct Bison latifrons Identified by Tandem Mass Spectrometry; Hydroxylysine Glycosides are a Common Feature of Ancient Collagen. *Mol Cell Proteomics* 14(7):1946–1958.
75. Cleland TP, Schroeter ER, Schweitzer MH (2015) Biologically and diagenetically derived peptide modifications in moa collagens. *Proc Biol Sci* 282(1808):20150015.
76. Robinson NE, Robinson A (2004) *Molecular clocks: deamidation of asparaginyl and glutaminyl residues in peptides and proteins* (Althouse Press).
77. Fraley C, Raftery AE, Murphy TB, Scrucca L (2012) *mclust Version 4 for R: normal mixture modeling for model-based clustering, classification and density estimation* ( Technical Report No 597).
78. Khoshnoodi J, Pedchenko V, Hudson BG (2008) Mammalian collagen IV. *Microsc Res Tech* 71(5):357–370.
79. Ninomiya, Y., Kagawa, M., Iyama, K., Naito, I., Kishiro, Y., Seyer, J.M., Sugimoto, M., Oohashi, T., Sado, Y. (1995) Differential expression of two basement membrane collagen genes, COL4A6 and COL4A5, demonstrated by immunofluorescence staining using peptide-specific monoclonal antibodies. *J Cell Biol* 130(5):1219–1229.
80. Robinson NE, et al. (2004) Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J Pept Res* 63(5):426–436.
81. Catak S, Monard G, Aviyente V, Ruiz-López MF (2009) Deamidation of asparagine residues: direct hydrolysis versus succinimide-mediated deamidation mechanisms. *J Phys Chem A* 113(6):1111–1120.
82. Nepomuceno AI, Gibson RJ, Randall SM, Muddiman DC (2014) Accurate identification of deamidated peptides in global proteomics using a quadrupole orbitrap mass spectrometer. *J Proteome Res* 13(2):777–785.
83. Hao P, Ren Y, Alpert AJ, Sze SK (2011) Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol Cell Proteomics* 10(10):O111.009381.
84. Dobberstein RC, et al. (2009) Archaeological collagen: Why worry about collagen diagenesis? *Archaeol Anthropol Sci* 1(1):31–42.
85. Glozak MA, Sengupta N, Zhang X, Seto E (2005) Acetylation and deacetylation of non-histone proteins. *Gene* 363:15–23.
86. Choudhary C, et al. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325(5942):834–840.
87. Jiang T, Zhou X, Taghizadeh K, Dong M, Dedon PC (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc Natl Acad Sci USA* 104(1):60–65.

88. Choudhary C, Weinert BT, Nishida Y, Verdin E, Mann M (2014) The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* 15(8):536–550.
89. Dhalluin C, et al. (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399(6735):491–496.
90. Takemura R, et al. (1992) Increased microtubule stability and alpha tubulin acetylation in cells transfected with microtubule-associated proteins MAP1B, MAP2 or tau. *J Cell Sci* 103(4):953–964.
91. Kim JY, Kim KW, Kwon HJ, Lee DW, Yoo JS (2002) Probing lysine acetylation with a modification-specific marker ion using high-performance liquid chromatography/electrospray-mass spectrometry with collision-induced dissociation. *Anal Chem* 74(21):5443–5449.
92. Castellano S, et al. (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proc Natl Acad Sci USA* 111(18):6666–6671.
93. 1000 Genomes Project Consortium, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.
94. Linsenmayer TF, Eavey RD, Schmid TM (1988) Type X collagen: a hypertrophic cartilage-specific molecule. *Pathol Immunopathol Res* 7(1-2):14–19.
95. Arnold MA, et al. (2007) MEF2C transcription factor controls chondrocyte hypertrophy and bone development. *Dev Cell* 12(3):377–389.
96. Surmann-Schmitt C, et al. (2009) Wif-1 is expressed at cartilage-mesenchyme interfaces and impedes Wnt3a-mediated inhibition of chondrogenesis. *J Cell Sci* 122(20):3627–3637.
97. Dy P, et al. (2012) Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. *Dev Cell* 22(3):597–609.
98. Pritchard JJ, Scott JH, Girgis FG (1956) The structure and development of cranial and facial sutures. *J Anat* 90(1):73–86.
99. Coussens AK, et al. (2007) Unravelling the molecular control of calvarial suture fusion in children with craniosynostosis. *BMC Genomics* 8:458.
100. Felszeghy S, Holló K, Módis L, Lammi MJ (2000) Type X collagen in human enamel development: a possible role in mineralization. *Acta Odontol Scand* 58(4):171–176.
101. Liu J, et al. (2007) Matrix and TGF- $\beta$ -related gene expression during human dental pulp stem cell (DPSC) mineralization. *In Vitro Cell Dev Biol* 43(3-4):120–128.
102. Dong N-P, Zhang L-X, Liang Y-Z (2011) A comprehensive investigation of proline fragmentation behavior in low-energy collision-induced dissociation peptide mass spectra. *Int J Mass Spectrom* 308(1):89–97.
103. Brechi LA, Tabb DL, Yates JR 3rd, Wysocki VH (2003) Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal Chem* 75(9):1963–1971.
104. Warinner C, et al. (2014) Pathogens and host immunity in the ancient human oral cavity. *Nat Genet* 46(4):336–344.
105. Prüfer K, et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49.
106. Rogers J, Gibbs RA (2014) Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat Rev Genet* 15(5):347–359.
107. Huerta-Sánchez E, et al. (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512(7513):194–197.
108. Dannemann M, Andrés AM, Kelso J (2015) Introgression of Neanderthal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am J Hum Genet* 98(1):22–33.
109. Kong A, et al. (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31(3):241–247.
110. Briggs AW, et al. (2009) Targeted retrieval and analysis of five Neanderthal mtDNA genomes. *Science* 325(5938):318–321.
111. Krause J, et al. (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20(3):231–236.

112. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* 7(3). doi:10.1371/journal.pone.0034131.
113. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328(5979):710–722.
114. Gansauge M-T, Meyer M (2014) Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res* 24(9):1543–1549.
115. Skoglund P, et al. (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci USA* 111(6):2229–2234.
116. Sawyer S, et al. (2015) Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc Natl Acad Sci USA* 112(51):15696–15700.
117. Krause J, et al. (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464(7290):894–897.
118. Reich D, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327):1053–1060.
119. Meyer M, et al. (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505(7483):403–406.
120. Horai S, et al. (1993) Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J Mol Evol* 35(1):32–43.
121. Meyer M, et al. (2016) Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531(7595):504–507.
122. Bronk-Ramsey C, Lee S (2013) Recent and Planned Developments of the Program OxCal. *Radiocarbon* 55(2–3):720–730.
123. Reimer PJ, et al. (2013) IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon* 55(4):1869–1887.
124. Soressi M (2011) Révision taphonomique et techno-typologique des deux ensembles attribués au Châtelperronien de la Roche-à-Pierrot à Saint-Césaire. *Anthropologie* 115(5):569–584.
125. Lee-Thorp JA (2008) On isotopes and old bones. *Archaeometry* 50(6):925–950.
126. Bocherens H (2009) Neanderthal Dietary Habits: Review of the Isotopic Evidence. *The Evolution of Hominin Diets, Vertebrate Paleobiology and Paleoanthropology.*, eds Hublin JJ, Richards M (Springer Netherlands), pp 241–250.
127. Fuller BT, Fuller JL, Harris DA, Hedges REM (2006) Detection of breastfeeding and weaning in modern human infants with carbon and nitrogen stable isotope ratios. *Am J Phys Anthropol* 129(2):279–293.
128. Williams JS, White CD, Longstaffe FJ (2005) Trophic level and macronutrient shift effects associated with the weaning process in the Postclassic Maya. *Am J Phys Anthropol* 128(4):781–790.
129. Richards MP, Trinkaus E (2009) Isotopic evidence for the diets of European Neanderthals and early modern humans. *Proc Natl Acad Sci USA* 106(38):16034–16039.
130. Sealy J, Armstrong R, Schrire C (1995) Beyond lifetime averages: tracing life histories through isotopic analysis of different calcified tissues from archaeological human skeletons. *Antiquity* 69:290–290.
131. Hedges REM, Clement JG, Thomas CDL, O'Connell TC (2007) Collagen turnover in the adult femoral mid-shaft: Modeled from anthropogenic radiocarbon tracer measurements. *Am J Phys Anthropol* 133(2):808–816.
132. Ambrose SH (1993) Isotopic analysis of paleodiets: methodological and interpretive considerations. *Food and nutrition in history and anthropology (USA)* 10:59–130.
133. Bocherens H, Drucker DG, Madelaine S (2014) Evidence for a (15)N positive excursion in terrestrial foodwebs at the Middle to Upper Palaeolithic transition in south-western France: Implications for early modern human palaeodiet and palaeoenvironment. *J Hum Evol* 69:31–43.
134. Beazley MJ, Rickman RD, Ingram DK, Boutton TW, Russ J (2002) Natural Abundances of Carbon Isotopes (14C,13C) in Lichens and Calcium Oxalate Pruina: Implications for Archaeological and Palaeoenvironmental Studies. *Radiocarbon* 44(3):675–684.

135. Batts JE, Calder LJ, Batts BD (2004) Utilizing stable isotope abundances of lichens to monitor environmental change. *Chem Geol* 204(3–4):345–368.
136. Bocherens H, et al. (1999) Palaeoenvironmental and Palaeodietary Implications of Isotopic Biogeochemistry of Last Interglacial Neanderthal and Mammal Bones in Scladina Cave (Belgium). *J Archaeol Sci* 26(6):599–607.
137. Bocherens H, et al. (2011) Isotopic evidence for dietary ecology of cave lion (*Panthera spelaea*) in North-Western Europe: Prey choice, competition and implications for extinction. *Quat Int* 245(2):249–261.
138. Richards MP, Pettitt PB, Stiner MC, Trinkaus E (2001) Stable isotope evidence for increasing dietary breadth in the European mid-Upper Paleolithic. *Proc Natl Acad Sci USA* 98(11):6528–6532.
139. Richards MP, et al. (2008) Isotopic dietary analysis of a Neanderthal and associated fauna from the site of Jonzac (Charente-Maritime), France. *J Hum Evol* 55(1):179–185.
140. Bocherens H (2003) Isotopic biogeochemistry and the paleoecology of the mammoth steppe fauna. *Deinsea* 9:57–76.
141. Szpak P, et al. (2010) Regional differences in bone collagen  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  of Pleistocene mammoths: Implications for paleoecology of the mammoth steppe. *Palaeogeogr Palaeoclimatol Palaeoecol* 286(1–2):88–96.
142. Trinkaus E, et al. (2003) An early modern human from the Peștera cu Oase, Romania. *Proc Natl Acad Sci USA* 100(20):11231–11236.
143. Jacobi RM, Higham TFG (2008) The “Red Lady” ages gracefully: new ultrafiltration AMS determinations from Paviland. *J Hum Evol* 55(5):898–907.
144. Hu Y, et al. (2009) Stable isotope dietary analysis of the Tianyuan 1 early modern human. *Proc Natl Acad Sci USA* 106(27):10971–10974.
145. Fizet M, et al. (1995) Effect of diet, physiology and climate on carbon and nitrogen stable isotopes of collagen in a late pleistocene anthropic palaeoecosystem: Marillac, Charente, France. *J Archaeol Sci* 22(1):67–79.
146. Bocherens H, Drucker DG, Billiou D, Patou-Mathis M, Vandermeersch B (2005) Isotopic evidence for diet and subsistence pattern of the Saint-Césaire I Neanderthal: review and use of a multi-source mixing model. *J Hum Evol* 49(1):71–87.
147. Richards MP, Schmitz RW (2008) Isotope evidence for the diet of the Neanderthal type specimen. *Antiquity* 82(317):553–559.
148. Wißing C, et al. Isotopic evidence for dietary ecology of late Neandertals in North-Western Europe. *Quat Int*. doi:10.1016/j.quaint.2015.09.091.
149. Bocherens H, Drucker D (2003) Trophic level isotopic enrichment of carbon and nitrogen in bone collagen: case studies from recent and ancient terrestrial ecosystems. *Int J Osteoarchaeol* 13(1–2):46–53.