COLLAGENS, KERATINS ET AL.

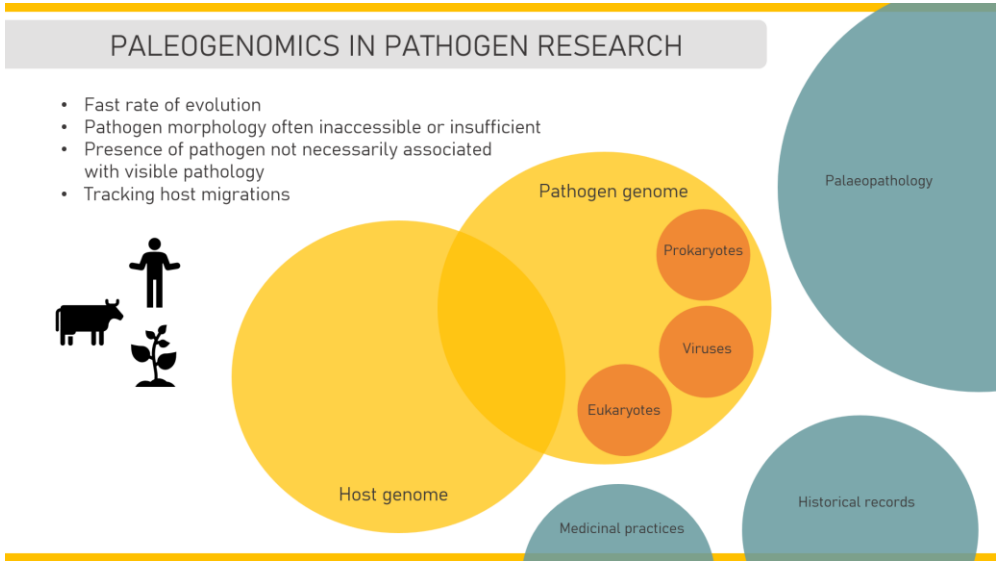
EVA CHOCHOLOVÁ
LABORATORY OF BIOLOGICAL AND MOLECULAR ANTHROPOLOGY
DEPARTMENT OF EXPERIMENTAL BIOLOGY

QUIZ

1. Write at least 3 examples of viruses studied in palaeogenomics:

HBV, variola, influenza, bacteriphages…

2. How can we study past diseases in general?



3. What are the three types of plague?

Bubonic, septicemic, pneumonic

4. Which types of viruses are hardest to study in palaeogenomics?

RNA, ssDNA

5. What is the difference between anthroponosis, zoonosis and sapronosis?

Human-human, human-animal, abiotic-living host

6. What are some examples of sedaDNA sources?

- Lakes and other freshwater sources
- Marine
- Cave
- Burials
- Settlements
- Permafrost
- (Ice, glacier)
- (Latrines)
- (Extraterrestrial sources)

7. What can we study through sedaDNA?

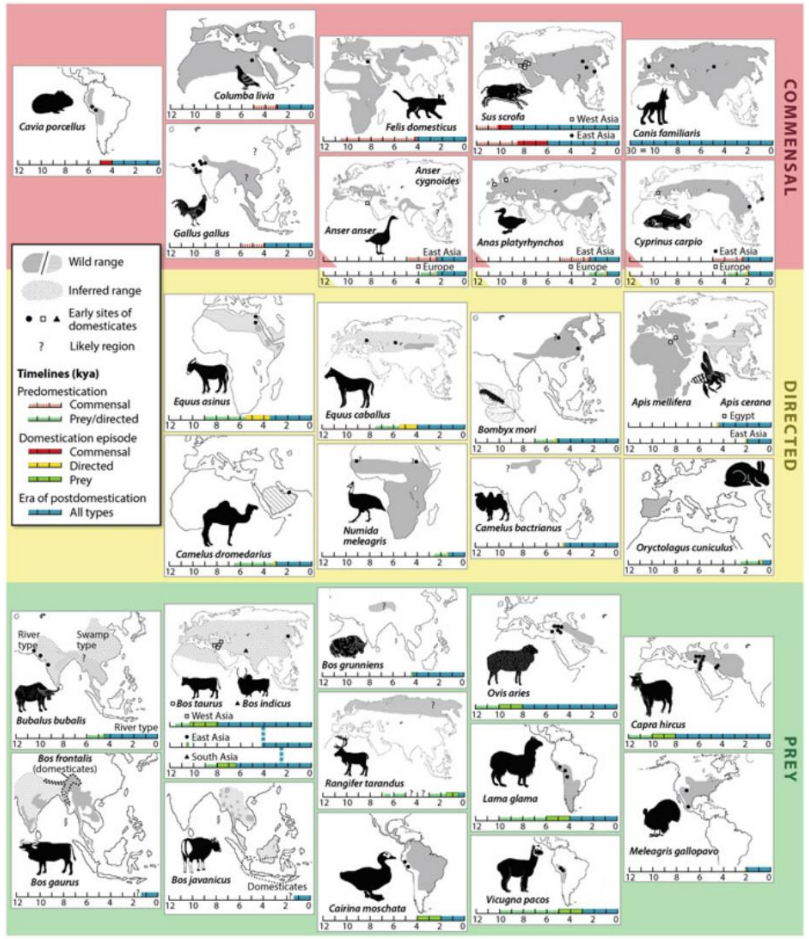
Palaeoenvironment, climate change, biodiversity, domestication, ancient hominins, parasites, …

QUIZ

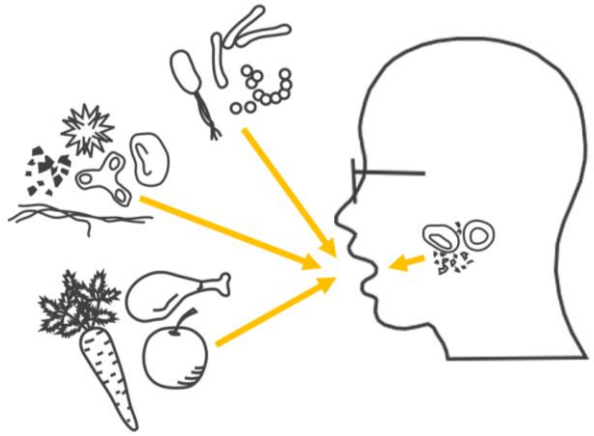
8. What is DNA leaching and where was it observed so far?

Movement of DNA vertically across sediment layers, observed in caves and non-frozen soil

9. Domestication types and example:



10. What can we study from dental calculus?



11. How can we deal with the challenge of limited amount of dental calculus?

UNIFIED PROTOCOLS

MORE SENSITIVE METHODS

OMICS

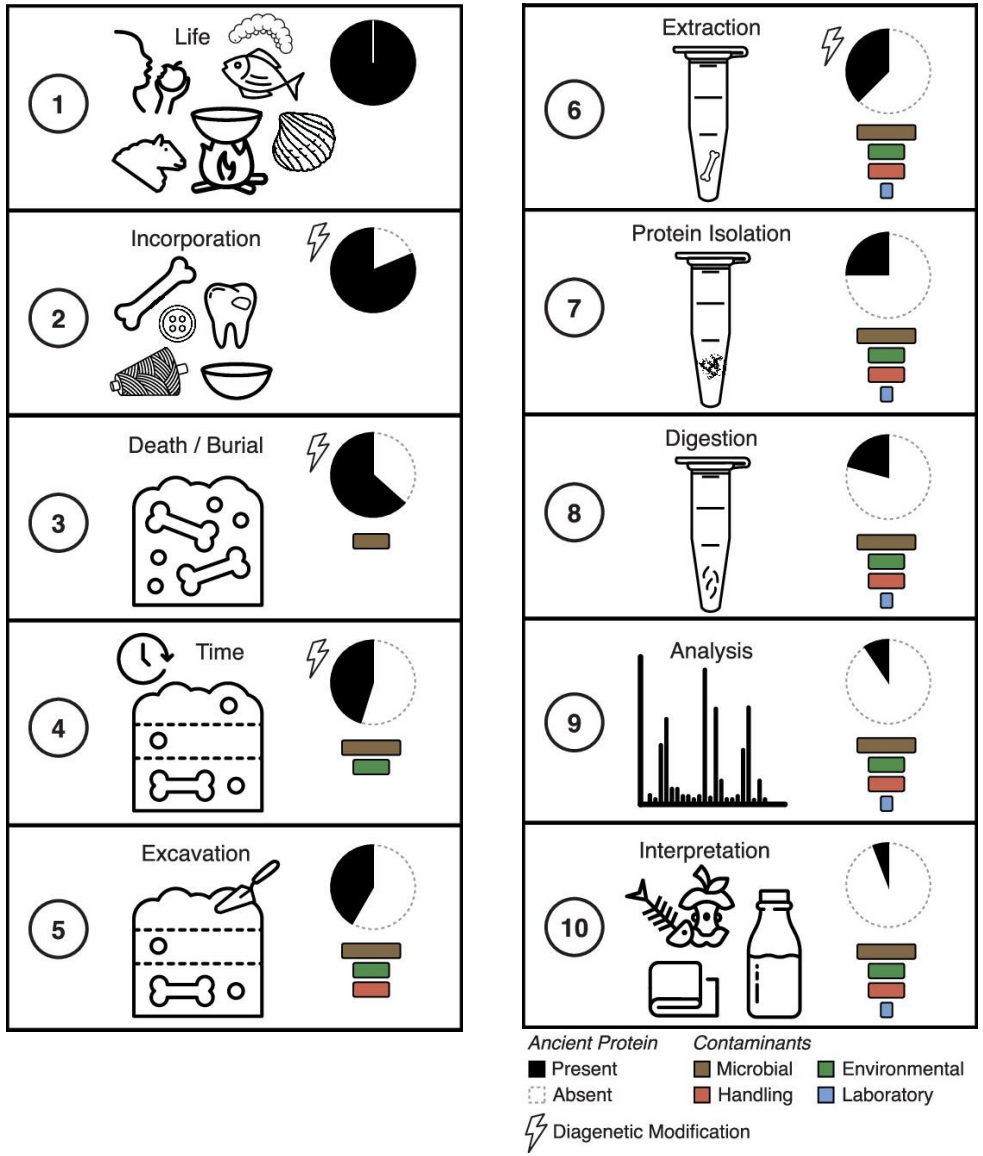
...

COLLAGENS, KERATINS ET AL.

EVA CHOCHOLOVÁ
LABORATORY OF BIOLOGICAL AND MOLECULAR ANTHROPOLOGY
DEPARTMENT OF EXPERIMENTAL BIOLOGY

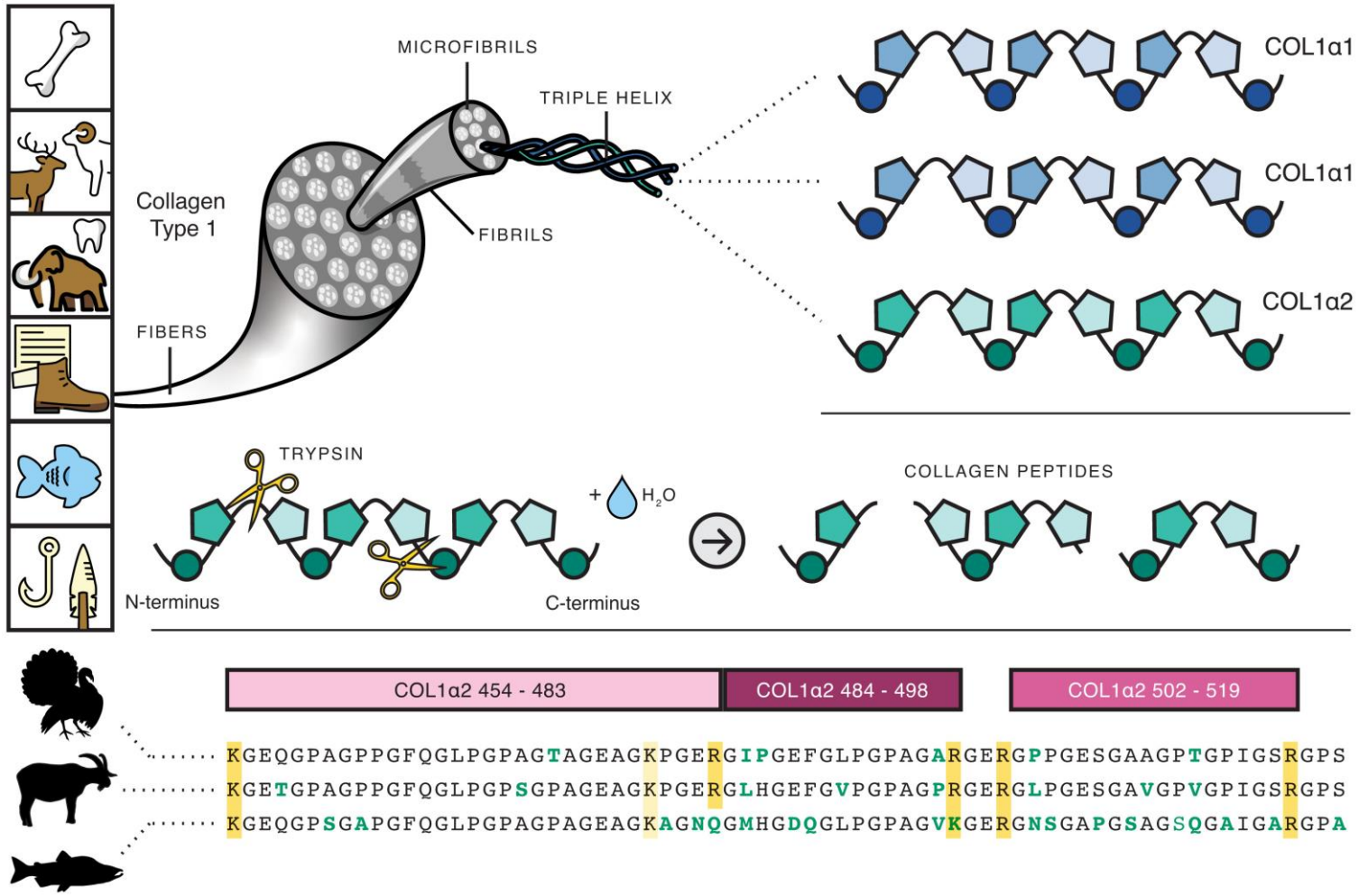
ANCIENT PROTEINS

- Most proteins preserved in mineralised matrix
- Degradation mainly through microbial activity (slowed by cold and dry environment)
- Bottom-up approach
- Palaeoproteomics most often for taxonomy – collagen and keratin – easier authentication through modifications (glutamine deamidation)



Warinner et al., 2022 DOI: 10.1021/acs.chemrev.1c00703

COLLAGENS



Overview of collagen structure and archaeological sources. Collagen can be retrieved from a wide range of animal tissues. In most animals, the COL1 triple helix is composed of two α1-chains and one α2-chain. Five triple helices are bundled into a microfibril. Bundles of microfibrils form a fibril, and bundles of fibrils form fibers. During the initial stages of ZooMS, this structure is denatured, allowing the enzyme trypsin to cut the protein into peptides. Peptides differ in sequence and mass across taxa, as shown for turkey (*M. gallopavo*), goat (C*. hircus*), and coho salmon (*O. kisutch*).

- Aside from palaeoproteomics, collagens are used in stable isotope analysis and radiocarbon dating
- Collagen, type 1 (COL1)
  - Most abundant protein in human body (80 % of bone proteins)
  - Stable – triple helix, hydrogen bonds of hydroxyproline
- Tetrapods 2x COL1α1 and 1x COL1α2
- Teleost fish COL1α1, COL1α2, COL1α3
- Chains consist of G-X-Y motif (G - glycine, X and Y - various AA, often proline and hydroxyproline)
- Taxonomically informative collagen peptides - marker peptides.

Richter et al., 2022 DOI: 10.1073/pnas.2109323119

α-KERATINS AND CORNEOUS β-PROTEINS

- Less used since they‘re often contaminants (human and sheep keratins)

- Most important structural proteins after collagens
- Skin, hair, claws, hooves, horns, feathers, beaks, turtle shells, baleen

- Furs and wool

OTHER TARGETS FOR TAXONOMY

- Databases incomplete
- Best for single origin samples, complex samples studied by LC-MS/MS

- Fibroin in silk (not only from domesticated *Bombyx mori*, also wild silkworms)

- Eggshells
- Mollusk shells
- Corals
- Insect exoskeleton

Characterizing historical textiles and clothing with proteomics

Caroline Solazzo

Proteomics and Biomolecular Mass Spectrometry Laboratory, Museum Conservation Institute (MCI), Smithsonian Institution, 4210 Silver Hill Road, Suitland, MD 20746, USA
solazzoc@si.edu

Table 4
Known sequences of silk proteins in domesticated and wild silks (✓ indicates at least one record exists and × that no sequence is available)

Sequences in NCBI	*Bombyx mori* (Domestic silk)	*Bombyx mandarina* (Wild silkmoth)	*Samia cynthia ricini* (Eri silk)	*Antheraea pernyi* (Tussah silk)	*Antheraea assamensis* (Muga silk)	*Antheraea yamamai* (Tensan silk)
Fibroin heavy chain Fib-H	✓	×	✓	✓	✓	✓
Fibroin light chain Fib-L	✓	✓	×	×	×	×
Sericin	✓	✓	×	×	×	×
P25	✓	✓	×	×	×	×

Richter et al., 2022 DOI: 10.1073/pnas.2109323119

ZooMS

- Zooarchaelogy by mass spectrometry – collagen peptide mass fingerprinting (PMF)
- Routinely MALDI-TOF (matrix-assisted laser desorption/ionisation-time-of-flight)
- Cost-effective, fast, minimally invasive (low input)
- Applied to bone, ivory, antler, parchment and vellum, leather, fish scales, dentine and cementum, horn core, animal glues, works of art and various artefacts…

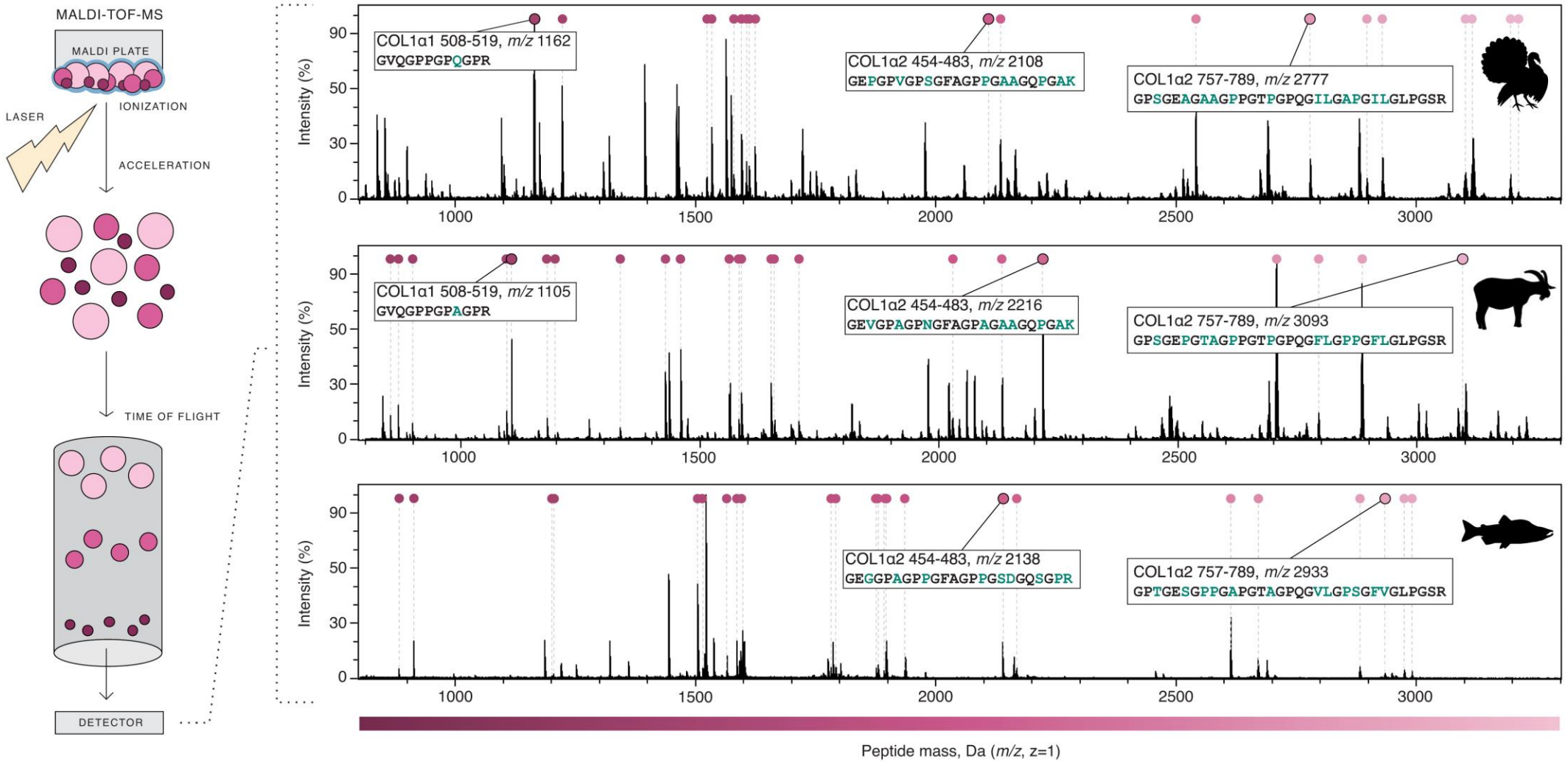
A primer for ZooMS applications in archaeology

Kristine Korzow Richter, Maria C. Codlin, Melina Seabrook, and Christina Warinner Authors Info & Affiliations

Edited by Suzanne Pilaar Birch, The University of Georgia, Athens, GA; received June 18, 2021; accepted January 14, 2022 by Editorial Board Member Dolores R. Piperno

May 10, 2022 | 119 (20) e2109323119 | https://doi.org/10.1073/pnas.2109323119

ZooMS



Steps of MALDI-TOF and representative collagen spectra. Digested collagen peptides (pink) are embedded in the matrix (blue) and ionized with a laser. Charged peptides (+1) are then accelerated through a TOF tube, where they separate by mass. The output of the detector is visualized as spectra, where time is converted to m/z based on calibration standards. Collagen spectra are shown for turkey (*M. gallopavo*), goat (*C. hircus*), and coho salmon (*O. kisutch*). Authenticated collagen peptide peaks are indicated by pink circles (Dataset S1 A). Three taxonomically informative marker peptides are annotated, with Insets indicating the collagen chain, position, m/z, and sequence; amino acids that differ across taxa are highlighted in green. Note that although isoleucine (I) and leucine (L) differences are highlighted, they are not distinguishable by MALDI-TOF. Baseline correction, smoothing, and intensity normalization were performed in mMass (134).

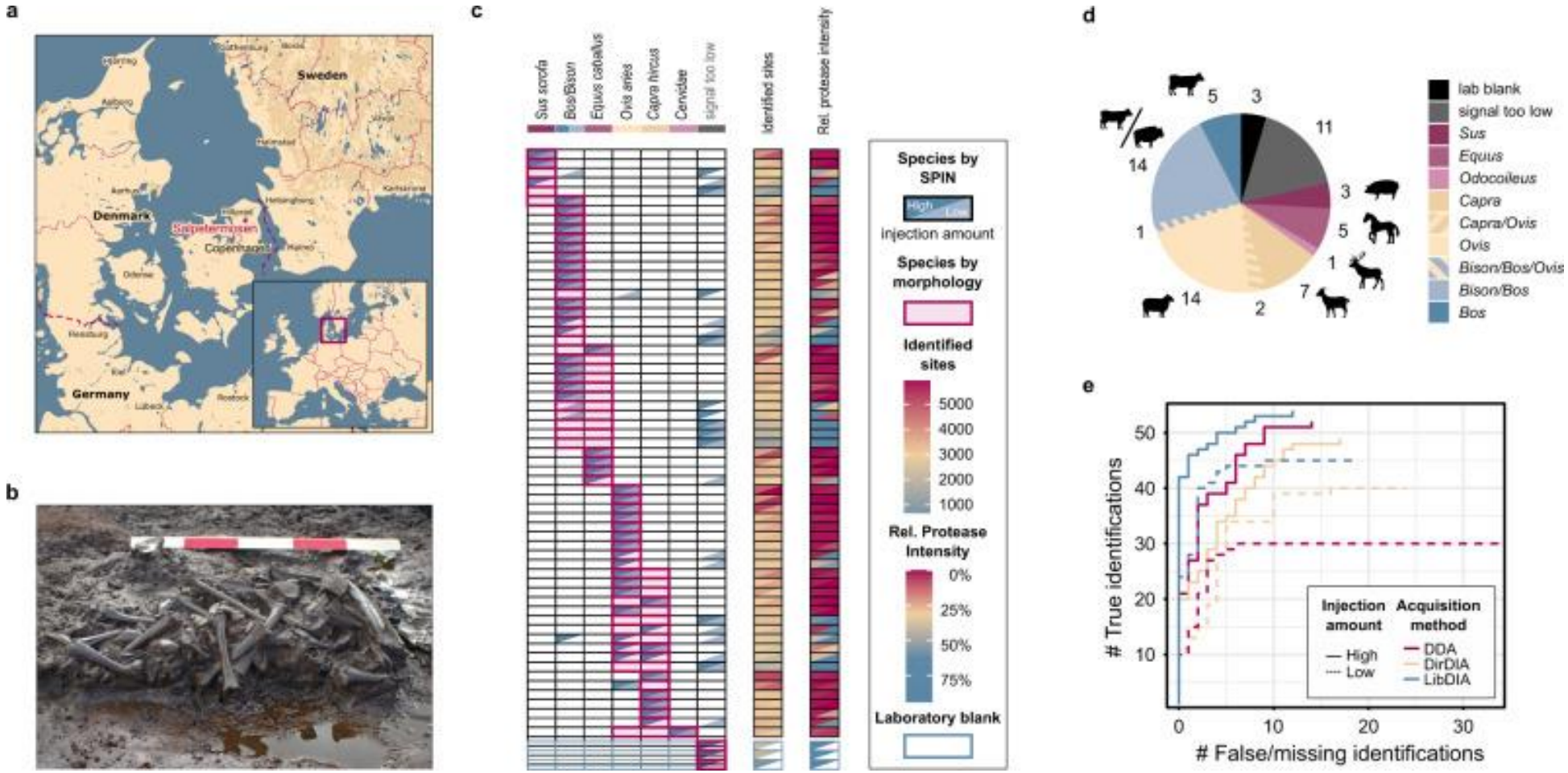
Article | Open access | Published: 05 May 2022

SPIN enables high throughput species identification of archaeological bone by proteomics

Patrick Leopold Rüther, Immanuel Mirnes Husic, Pernille Bangsgaard, Kristian Murphy Gregersen, Pernille Pantmann, Milena Carvalho, Ricardo Miguel Godinho, Lukas Friedl, João Cascalheira, Alberto John Taurozzi, Marie Louise Schjellerup Jørkov, Michael M. Benedetti, Jonathan Haws, Nuno Bicho, Frido Welker, Enrico Cappellini & Jesper Velgaard Olsen

Nature Communications **13**, Article number: 2458 (2022) | Cite this article

a Location of the archaeological site "Salpetermosen Syd 10" on Zealand in Denmark in the Hillerød municipality 30 km north of Copenhagen. Map drawn in Mapbox Studio using a custom style. **b** Cross section of an in situ wetland bone deposit. Scale bar is 50 cm. Four bones were radiocarbon dated between 1720 and 1570 BP. Picture provided by the Museum of North Zealand. **c** Species identification results by SPIN (5 min method, library-based DIA) and by morphological assessment for 63 samples from the Salpetermosen site measured in technical duplicates and 3 blanks. Rows represent individual samples and have been ordered first by morphological species assignment and then by decreasing mean site coverage. The upper left and lower right wedge of each cell represent results measured in two separate experiments, one with higher (upper left, dark blue) and the other with lower (lower right, light blue) MS signal intensity. The first seven columns indicate SPIN species by blue wedges and morphological species possibilities by pink boxes. Bovine species assignments are combined in column two. The eighth and ninth columns are heatmaps showing the absolute number of covered amino acids and relative protease intensity, respectively. **d** Summary of SPIN species identifications from panel c in the replicate with high MS intensity. Bovine identifications are separated into cow (*Bos*) and broader bovine identifications (*Bos/Bison*). Striped colors indicate samples with insufficient sequence coverage to distinguish closely related taxa. Samples with insufficient sequence coverage for confident species identification are marked as "signal too low" and correctly excluded blanks are marked in black. **e** Pseudo receiver operating characteristic (ROC) curves for comparing the sensitivity and success rate of three different data acquisition and analysis strategies. Results of each dataset were sorted by decreasing number of identified sites. The *y*-axis shows the cumulative number of correct species identifications in agreement with the morphology. The *x*-axis shows the cumulative number of false or missing identifications below the relative protease intensity threshold. Color indicates data acquisition and analysis mode with pink for DDA, dark blue for library-based DIA, and sand color for library-free DirectDIA. Experiments with lower MS intensity are shown by dashed and high intensity by solid lines.

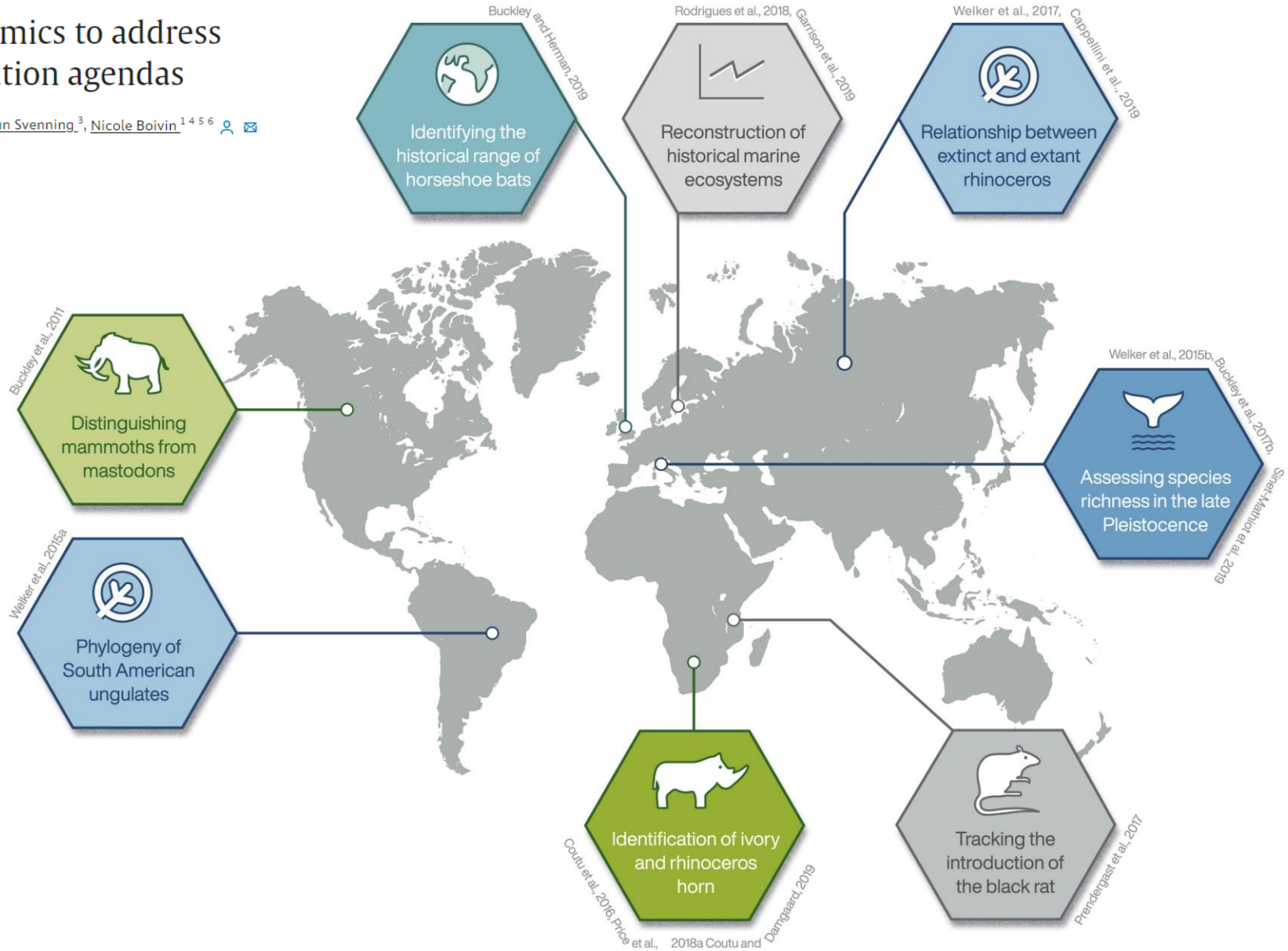


Volume 25, Issue 5, 20 May 2022, 104195

Review

Leveraging palaeoproteomics to address conservation and restoration agendas

Carli Peters [1], Kristine K. Richter [2], Jens-Christian Svenning [3], Nicole Boivin [1 4 5 6]



Peters et al., 2022 DOI: 10.1016/j.isci.2022.104195

- ZooMS
- Clothing, silkworms
- Top-down, bottom-up
- Dust contamination
- Taxonomical resolution
- Whales

Article | Open access | Published: 02 September 2020

Palaeoproteomics gives new insight into early southern African pastoralism

Louise Le Meillour, Séverine Zirah, Antoine Zazzo, Sophie Cersoy, Florent Détroit, Emma Imalwa, Matthieu Lebon, Alma Nankela, Olivier Tombret, David Pleurdeau & Joséphine Lesur

Scientific Reports **10**, Article number: 14427 (2020) | Cite this article

Student answers

What is the main research question addressed in the paper?

What techniques and methods were used in the study to analyze ancient protein samples?

- MS, radiocarbon dating – peptide markers in collagen for taxonomic identification
- ATR FT-IR spectroscopy (23 bones, 3 teeth) – decalcification Tris-EDTA buffer
- UHPLC-MS/MS – reconstruction PEAKS Studio Software
- COL1A1, COLA2

Were there any limitations or challenges associated with the methods used?

What were the main findings of the study?

Were there any unexpected or surprising results?

What are the implications of the results for our understanding of ancient human populations and their lifestyles?

- Populations exploited these domesticated animals mainly for their secondary products (milk, wool, carrying) rather than meat
- Diffusion by gradual infiltration

Are there alternative interpretations of the data presented in the paper?

What additional evidence or analyses could strengthen the conclusions drawn from the data?

How does this paper contribute to your understanding of palaeoproteomics and its applications in archaeology and anthropology?