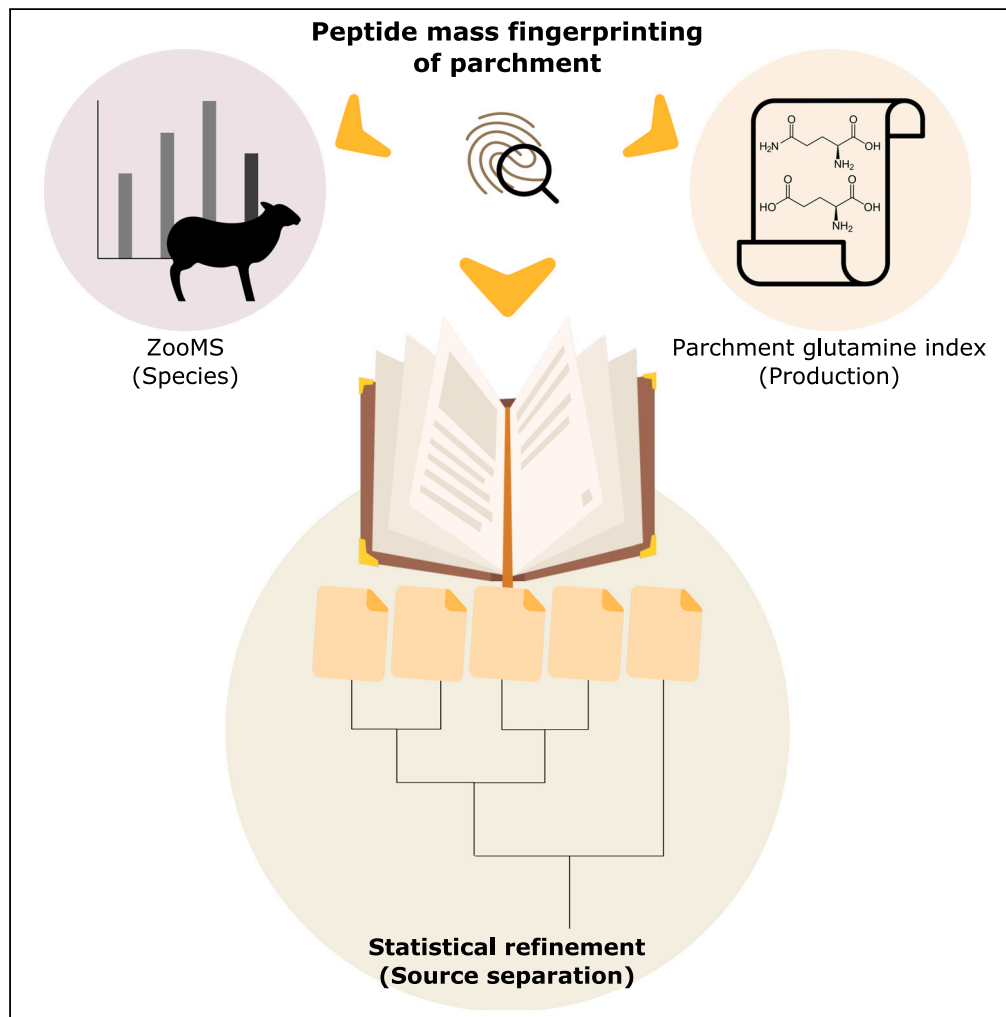


Article

A biological reading of a palimpsest



Laura C. Viñas-Caron, Ismael Rodríguez Palomo, Natasha Fazlic, Jiří Vnouček, Matthew Driscoll, Sarah Fiddymont, Matthew J. Collins

laura@palaeome.org (L.C.V.-C.)
ismael@palaeome.org (I.R.P.)

Highlights

Peptide mass fingerprints (PMFs) of parchment reveal production history, beyond species

PMF clusters are consistent with the source materials used to construct the manuscript

PMF groups the parchment of a 12th-century palimpsest into four original sources

This approach enhances our ability to source parchment based upon production methods

Viñas-Caron et al., iScience 26, 106786
June 16, 2023 © 2023 The Author(s).
<https://doi.org/10.1016/j.isci.2023.106786>

Article

A biological reading of a palimpsest

Laura C. Viñas-Caron,^{1,5,*} Ismael Rodríguez Palomo,^{1,*} Natasha Fazlic,² Jiří Vnouček,³ Matthew Driscoll,² Sarah Fiddymment,⁴ and Matthew J. Collins^{1,4}

SUMMARY

In the Middle Ages, texts were recorded and preserved on parchment, an animal-derived material. When this resource was scarce, older manuscripts were sometimes recycled to write new manuscripts. In the process, the ancient text was erased, creating what is known as a palimpsest. Here, we explore the potential of peptide mass fingerprinting (PMF), widely applied to identify species, to help reconnect the dispersed leaves of a manuscript and reveal differences in parchment manufacturing. In combination with visual methods, we analyzed a whole palimpsest, the codex AM 795 4to from the Arnamagnæan Collection (Copenhagen, Denmark). We find that both sheep and goat skins were used in this manuscript, and that parchment differed in quality. Notably, the PMF analysis distinguished five groups of folios which match the visual groupings. We conclude that this detailed interrogation of a single mass spectrum can be a promising tool to understand how palimpsest manuscripts were constructed.

INTRODUCTION

Palimpsests, recycled writing surfaces, have long fascinated scholars for their ability to preserve evidence of the original text or undertext.¹ The uncovering of the Archimedes Palimpsest, which harbored several unique works of the Greek mathematician, is perhaps one of the best known interdisciplinary collaborations in codicology, bringing together experts in image analysis, philology, conservation, mathematics, and physics.² The reason for recycling such old documents was that parchment, the principal writing support in the European Middle Ages, was a limited and costly resource. In periods of shortage, scribes would reuse manuscripts whose texts were considered less valuable or did not conform to cultural or religious dogma.

Toward the beginning of the 12th century in Barcelona, four manuscripts from the 9th to 11th centuries met this fate. Their texts were erased, and the loose leaves were reshaped to create one of the most remarkable codices in the Arnamagnæan Collection, the manuscript AM 795 4to (Copenhagen, Arnamagnæan Institute, AM 795 4to). Because of this treatment, many parts of the original text remain unreadable, hampering the reassembly of the different leaves composing the original manuscripts.

Biocodicology, in particular peptide mass fingerprinting (PMF), is a new and promising field to examine the biological information preserved in parchment,³ identifying the nature of the animals used⁴ and sometimes aspects of the manufacturing processes involved in the parchment's production.^{5,6} Here, we explore another role for PMF as a means of uncovering the process of constructing this palimpsest. To do so, we attempt a more in-depth analysis of the peptide mass fingerprints of collagen-based materials, demonstrating that these characteristic fingerprints not only capture differences between species but also between the leaves themselves.

Codicological and paleographic research on AM 795 4to

AM 795 4to is a small quarto manuscript (20 × 13 cm, 97 folios) on parchment that contains two distinct texts: the Book of Revelation, or Apocalypse of St John (1r-24r), and then the only surviving version of a commentary on it (24v-97v) by Bishop Apringius of Pace, who lived in the mid-6th century in Pace (now Beja) in Portugal. Apringius' reputation is attested by Isidore of Seville in his *De Viris Illustribus*, and he was influential in later commentaries, such as the one written by the 8th-century Asturian monk Beatus of Liébana.

Paleographic evidence, notably the late Carolingian minuscule script, suggests that AM 795 was written in the late 11th or early 12th century. It is the work of a single hand, possibly the grammarian Renall,⁷ who wrote several

¹Globe Institute, University of Copenhagen, 1350 Copenhagen, Denmark

²The Arnamagnæan Institute, University of Copenhagen, 2300 Copenhagen, Denmark

³Department of Preservation, The Royal Danish Library, 1016 Copenhagen, Denmark

⁴McDonald Institute for Archaeological Research, University of Cambridge, CB2 3ER Cambridge, UK

⁵Lead contact

*Correspondence: laura@palaeome.org (L.C.V.-C.), ismael@palaeome.org (I.R.P.) <https://doi.org/10.1016/j.isci.2023.106786>



documents in Barcelona Cathedral in the 12th century.⁸ A note added at the end of the manuscript in the 17th century states that the text was copied in Barcelona in 1042 (“MXXXXIIJ”) from an older manuscript, but it seems that the date was altered,^{9,10} the original and more likely date being 1132 (“MCXXXIIJ”).

Traces of the undertext are visible on some leaves. Although very little can be read with any certainty, early descriptions suggested that at least four original manuscripts were used.¹¹ Two texts could be responsories or liturgical chants with neumes corresponding to the musical notation system present in Catalonia during the 10th and 11th centuries. Another could contain a *Forum Iudicum*, a set of Visigothic laws, written in 9th-century Catalonia in a Visigothic minuscule with Carolingian influence.

Nothing is known about the manuscript’s whereabouts during the Late Middle Ages, but at the beginning of the 16th century, it appears to have been part of the library of Hernando Colón (1488–1539), son of the famous navigator Christopher Columbus, in Seville.^{12,13} Colón’s library, thought to have been the largest in Europe at the time, comprised some 15,000 volumes. In one of the many biographical tools Colón developed to keep track of his holdings, the so-called “Abecedarium B”, a list in alphabetical order of authors and titles, one entry clearly refers to this same manuscript. We do not know exactly what happened to the manuscript after Colón’s death, but it seems that, through various Danish book collectors and aristocrats, it made its way to Copenhagen during the 17th century and ended up in Árni Magnússon’s hands in 1696. After his death in 1730, his collection of manuscripts was preserved at the University of Copenhagen in the Arnamagnæan Collection.

RESULTS AND DISCUSSION

Visual analysis

To construct the palimpsest, the leaves of the original manuscripts were disbound and formatted into new leaves. The undertext was either chemically or physically removed, a process which had different effects on the parchment surface appearance. In some cases, traces of the text or lines that the scribe ruled as a guide can still be made out on some leaves, but, in others, words and letters remain unreadable or have been completely erased. The parchment was then divided into two or more separate leaves and the edges were trimmed. Interestingly, it is possible to observe that the undertext or ruled lines are orientated both horizontally and vertically on different leaves, indicating that in some cases the leaves were rotated 90°, a common practice in palimpsests to improve the readability of the new text (overtext).

The resulting pieces of parchment were folded in half constituting two connected leaves (bifolios) and positioned inside one another to form quires or gatherings. The quires were stacked side by side and then sewn together into a binding. In all, the final manuscript is composed of 13 regular quires, each comprising four bifolios (or eight folios), with the exception of quire IV, which has only seven leaves and combines different bifolios with single folios (singletons), and the last quire XIII, which has just two singletons. The gathering structure can be seen in [Figure 1C](#). The skins were arranged according to the so-called rule of Gregory,¹⁴ with the hair side of the skin always facing the hair side, and the flesh side facing the flesh side. This folding pattern is consistent throughout the manuscript, with all quires starting with the flesh side.

A closer visual analysis was conducted in an attempt to separate the different original manuscripts. Based on the remnants of the undertext, parchment surface characteristics, traces of production and preservation conditions, the codex was divided into four groups, A–D ([Table 1](#)); there was, however, one folio where it was difficult to discriminate between A and C, whereas another fell between C and D. These groups are likely to represent four source manuscripts which supplied parchment for reuse to make the present manuscript.

eZooMS

Species composition

The entire manuscript was sampled using the non-invasive eraser method, and a total of 54 bifolios, including three duplicates (ff. 1–8, 9–16, 89r–89v), were analyzed using eZooMS (PMF of this eraser waste) to assess the species composition; identifications are included in [Figure 1C](#) and [Data S1](#). Two-thirds of the folios (65%) were made of sheepskin, with the rest comprising goatskin. The presence of sheep and goat is consistent with archaeological evidence, which shows sheep and goats are both well represented in the faunal assemblages of north-east Iberia during the medieval period (c. 30–70% of main domesticates), indicating their wide availability for parchment production.^{15–18} This manuscript is the third to be analyzed in its entirety using eZooMS,^{19,20} but the first from this region, and its composition seems consistent with an

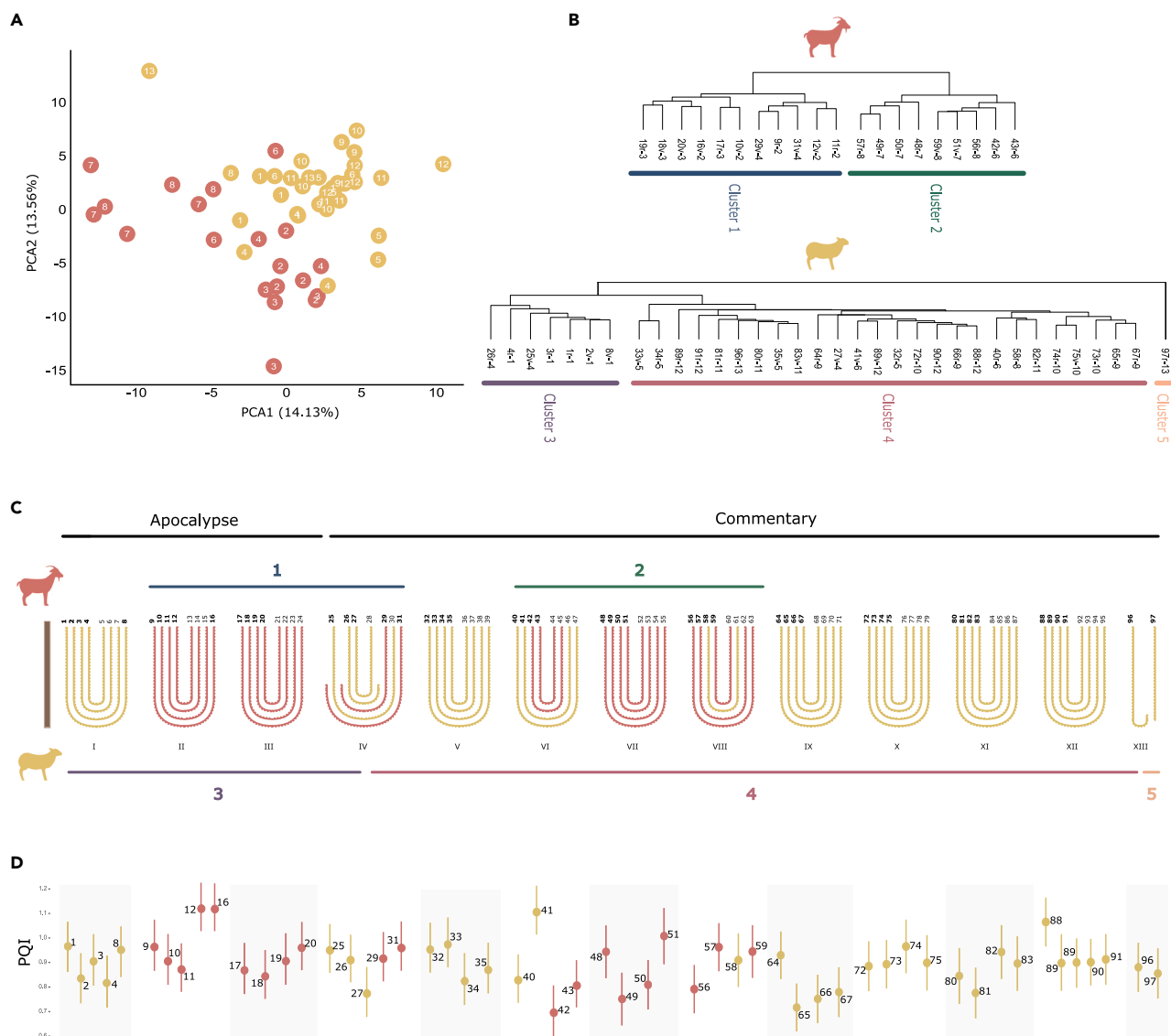


Figure 1. Overview of the proteomic results

(A) PCA with all the leaves analyzed, showing the number of the quire inside the circle and the species identification for each sample.
 (B) Clustering analysis of the leaves separated by sheep and goat. The numbers indicate the folio and the quire.
 (C) Visualization of gathering structure, organization of folios, hair and flesh sides and species identification. The numbers of the sampled folios are highlighted in bold and the proteomic clusters are indicated with the lines. This visualization was created using VisColl v. 2.0 (<https://github.com/leoba/VisColl>).
 (D) PQI estimates from multiple peptides and bars showing the standard error of the mean. PQI should range between 1 (non-deamidated) to 0 (fully deamidated) however values >1 result from baseline variation, whereas values lower than 0.5 are very unlikely as with such high levels of deamidation the collagen would be gelatinized.

earlier study⁴ that suggested geographic preferences in animal use for parchment, with sheep and goat generally preferred in countries from southern Europe. However, other factors such as cost and textual content could also have influenced the selection of animal species in parchment.²¹

Regarding the placement of the folios, the first quire is made from sheepskin, whereas quires II and III are goat. The second text – the Commentary – begins on the last folio of quire III (f. 24v) and continues in quire IV, which has an unusual structure with only seven leaves and alternates sheep and goat. Quire V is made of sheep but quires VI-VIII are largely of goat. The final quires (IX-XIII) are made entirely of sheep. This small sample is no different from the Cistercian library at Orval, where, in almost all cases (88%), individual

Table 1. Summary of the visual and proteomic groupings within the palimpsest

Palimpsest group and number of folios	Text orientation	Column width (cm)	Line spacing (mm)	Proteomic cluster	Species
A (n = 8)	Horizontal	Likely 8 cm	10 - 11 mm	3, 4	Sheep
B (n = 17)	Vertical	16 - 18 cm	9 - 10 mm	4, 5	Sheep
C (n = 17)	Mixed	Likely 2 columns × 10 cm	9 - 11 mm	1, 2, 3, 4	Sheep/goat
D (n = 9)	No ink remains	–	–	1	Goat

For further details see [Data S1](#).

codicological units were made from a single species.²¹ However, as can be seen, there is no particular pattern or intentional order in the placement of the folios, as is the case for the Gospel of Luke produced in Canterbury around 1120.²⁰ Because the manuscript was written by a single scribe, it is unlikely that changes in the choice of skins would be attributed to breaks in production. Instead, the organization of the codex and the reuse of parchment suggest a shortage of resources.

Deamidation and parchment production

We next examined protein damage patterns that can be induced by parchment production processes. Deamidation of glutamine (Q) to glutamic acid (E), in particular, can be caused by the use of lime during the dehairing step and can help to understand the quality of parchment and its manufacture; the greater the total alkalinity, the greater the deamidation rates (Figure 1D). This is captured in a Parchment Glutamine Index (PQI), which is based on deamidation estimates from multiple collagen peptides.⁶ PQI values for all the folios are reported in [Data S1](#).

The parchment used in this manuscript displays variable quality in PQI (0.70–1.13), with no statistical differences between species ($\chi^2 = 1.31$, $df = 1$, $p = 0.25$) or quires ($\chi^2 = 14.25$, $df = 12$, $p = 0.29$). No correlation was found with parchment thickness ($r = -0.1$, $df = 49$, $p = 0.47$), as opposed to what was previously observed in the Orval dataset.⁶ Despite this diversity, it can be observed that some leaves share similar values. If more than one bifolio was taken from the same skin, then all would have been exposed to the same extent to lime and therefore would present the same level of glutamine hydrolysis. This is observed in duplicate samples (ff. 1–8, 9–16, 89r–89v). The similarity in PQI, thus, could indicate that the parchment was obtained from the same production batch, using similar craft techniques or from the same time of year (temperature accelerates deamidation) and can guide the reassembly of original skins when combined with the visual analysis (see discussion below).

Peptide profiling of the leaves

After exploring the animal source and damage patterns, we further examined the spectra to assess whether the similarities and differences in peptide composition across the parchment leaves could enable us to discriminate patterns in production and help identify the original underlying documents.

Principal Component Analysis (PCA) of the entire dataset, consisting of 442 detected peaks and their intensities, perhaps unsurprisingly showed that folios belonging to parchment from the same species predictably group together (Figure 1A). Interestingly, however, there is also variability within species. The distribution showed some patterning, which could be associated with the organization of the quires, explaining c. 28% of the total variance.

Having separated the sheep and goat spectra, we attempted to gain even more resolution of the organization of the folios by hierarchical cluster analysis. This identified five major groups (Figure 1B), as opposed to the four elements previously observed codicologically. Within the goatskin samples, two distinct groups were identified, one encompassing quires II–IV (Cluster 1) and another quires VI–VIII (Cluster 2). One sheepskin cluster (Cluster 3) comprised all the folios of quire I and some of quire IV. The rest of the sheepskin in quires V–XIII is clustered together with one folio from quire IV (Cluster 4). Cluster 5 contains a single sheep folio, the last one in the volume. Duplicate samples indicate some variation within bifolios depending on sampling location but do not alter the global interpretation.

To identify the peaks driving the separation of folios, we employed a binary discriminant analysis, using the clusters as the class. The top 20 discriminatory peaks in sheep and goat are reported in [Table S1](#). We then

attempted to explore the origin of these peptides by matching the masses against type I collagen (COL1) sequences for sheep and goat, the most abundant protein on parchment. As expected, most peptides (approx. 52.5%) were confidently assigned to COL1 (Table S1). The non-identified peptides, approximately half, could correspond to modified type I collagen peptides or other types of collagen and proteins. Interestingly, one of the identified peptide masses (1384.68) has been reported in a previous study²² as a cow-specific milk peptide marker (α s1-casein; mass: 1384.68; position: 38–49; sequence: FFVAPFPEVFGK). Old recipes report the use of milk to erase the writing from the parchment.²³ Thus, parchment production, palimpsest treatment, or specific handling histories are likely driving some of the observed differences between leaves.

Combining visual and biomolecular analyses

When combining the results from the visual and proteomic analysis, which were performed in a blind manner to avoid bias in the interpretation, it can be observed that the proteomic clusters (1–5) mirror the visual groups (A–D) and show essentially the same structure.

This good correlation is particularly true for goatskins, which are separated into two clear proteomic clusters, cluster 1 (quires II, III and IV) and 2 (quires VI, VII and VIII), and these correspond to the two visual groups identified above (D and C, respectively). Interestingly, folio 31 (quire IV) was visually classified to group C but clustered together with other members of group D. In some cases, we were able to establish some visual matches of original skins in quires II and III from Group D (ff. 9/16 and 10/15; 11/14 and 19/22; 17/24 and 18/23). As anticipated, the PQI values (i.e., the extent of glutamine deamidation) from the same skins were similar in most cases.

It was more challenging to detect patterns within the sheepskins because of significant overlap and similarities in their proteomic profiles. Despite this, three different clusters were identified. Folio 97 falls in cluster 5 and is part of the visual group B. This single folio appears as an outlier both in the PCA and hierarchical clustering and may have come from another source or have been extensively handled because it is the last one of the codex and has several notes added in later centuries. The beginning of the manuscript presents another characteristic peptide profile (Cluster 3). This group includes folios from quire I (visual group A) and some folios from quire IV, which were assigned visually to group C, suggesting a similar provenance not seen in the visual analysis. Toward the end of the manuscript, folios from quires V–XIII show high proteomic similarity (Cluster 4), whereas the visual analysis assigned the leaves to two different groups, B and C. In the case of quire V, the leaves were ascribed initially to either group A or C and, based on the proteomic results, would seem to belong to group C.

In this part of the manuscript, it was problematic to establish any direct visual match between individual bifolios, despite similar appearance or very close PQI values. We suspect that this is because of the trimming and formatting undergone during the parchment production and subsequent palimpsest creation. For example, after the leaves of the original manuscript C were disbound, some were rotated 90°, a common practice in many palimpsests, whereas others were kept vertical, as shown by the orientation of the text remains or ruled lines. In addition, it is possible to observe that, in some cases, the edges were trimmed and the undertext is abruptly cut in the margin of the folios. This leads us to believe that the original manuscripts forming Cluster 4 were once quite large and that it was possible to obtain multiple parchment leaves by cutting in different orientations. Previous scholars identified faded traces of neumes or musical notes on f. 93v (quire XII) and suggested it was a responsory *lustus germinabit*, that is, a liturgical chant.^{7,24} These manuscripts usually have larger formats. Thus, it is possible that the two manuscripts identified visually, B and C, could belong to the two responsories from the 10th and 11th centuries and have a similar provenance, as detected by the proteomic analysis. However, as mentioned, the missing connection makes it difficult to match the leaves of the same original skin.

In conclusion, besides the above-mentioned differences in quire IV and V (which is, for several reasons, quite problematic), the visual and proteomic analyses are in remarkable agreement given their very different nature. When the results are in disagreement it is mainly because certain folios did not contain ink residues of the undertext and thus, their visual classification was more arduous.

Scribal agency

The identity of the scribes and the choices they made concerning both the writing of the text and the production of the manuscripts is usually unknown. Our results provide the opportunity to draw some conclusions on the production history of a palimpsest in connection to both its content and materiality, an aspect that cannot be gained solely by looking at the textual evidence.

We hypothesize that the palimpsest was produced as follows: The scribe began copying the Book of Revelation with leaves recycled from two manuscripts (which we call A and D). Manuscript A could be a *Forum Ludicum* from the 9th century, whereas in the case of manuscript D the original text was completely erased. The two manuscripts that were recycled for the copying of the Revelation required only three quires made up of 24 folios. We ignore how large the original manuscripts were, but the first only produced sufficient material for the first quire and the second for quires II and III. We propose that the origin of group D was a manuscript made solely of goatskin, whereas all of the visually discrete members of group A were made originally of sheepskin. At this point, the scribe finished copying the Book of Revelation and the resource was almost spent. We do not know if he always intended to add the Commentary, but our reading of the manuscript casts some doubt on this interpretation.

The only surviving copy of the Commentary of Bishop Apringius of Pace is a much longer text requiring 73 folios and starts on the verso of f. 24, the last one of quire III. The beginning of the next quire (IV) is oddly structured. One bifolio, f. 25 conjoined with f. 30, and one singleton, f. 26, are from the sheepskin MS used for the first quires and begin this one. The second full bifolio (ff. 27/28) uses another much larger sheepskin MS. The last two folios, 29 and 31, are both singletons from the goatskin MS used previously. The use of singletons highlights, like the palimpsest itself, the scarcity of suitable parchment. This inelegant unbalanced quire IV structure sees the scribe using up all of his available resources, which is too small in all but one case to make bifolios, but at the same time introducing the first example of a much more substantial source (C), which is so large it is cut differently and comprises 50% of the rest of the Commentary. This new source contained both sheepskin and goatskin, the last used for quire VI, all of quire VII, and all but one folio of quire VIII. Thus, the group C manuscript was a combination of sheep and goat parchment even before it was recycled into a new palimpsest. The rest of the folios are drawn from a sheepskin manuscript (B) that had sufficient parchment to complete the rest of the text. Manuscripts B and C may have been responsories, copied in the 10th or 11th centuries. The sole exception is the very last leaf, which seems to be a singleton from another source or has been extensively handled.

This supports the idea that the two texts were written one after the other, but a lack of planning suggests that the Commentary may have been an afterthought. It also suggests that the scribes in the scriptorium may have been copying different manuscripts at the same time and reusing parchment from these multiple sources. Sourcing parchment from the same species or manuscript may have not been a priority in this case.

Conclusions

By combining different methods, we shed new light on the history of the palimpsest AM 795 4to and its physical medium, the parchment on which it was inscribed. PMF showed that the manuscript was composed of both sheep and goatskins, whereas the extent of deamidation helped to determine their varying qualities and guide the reconstruction of original skins. It was previously revealed that there were four original documents used as the basis of the final text, our ZooMS analysis revealed four clusters plus a final singleton end leaf.

The visual analysis is broadly in line with the ‘proteomic’ clusters, and where they differ, this is because of limitations in visual evidence provided by specific membranes. Thus, the pattern of collagen fingerprint is not simply an indication of the difference in the species, but clusters by source. The presence of peaks associated with other proteins than collagen type I and their varying intensities may result in a characteristic fingerprint that reflects the overall history of the parchment, such as a particular style of parchment production, palimpsest treatment or handling. This structure would not necessarily be captured by PQI, because we are looking at a few selected collagen peptides and deamidation levels, but it seems to be reflected when looking at the entire set of peptides.

This approach has the added value of providing a myriad of information within a single sample, obtained in a non-invasive way and without additional costs. We are not aware that this combined approach, including within species clustering, has been applied to ZooMS data before, but we believe it holds potential for the analysis of other parchment manuscripts and for the reanalysis of collagen spectra that have previously been obtained through ZooMS. Palimpsests are not as rare as is generally believed with more than 700 new ones recently identified²⁵ and it is, therefore, helpful that codicologists now have another low-cost tool in their armory.

Limitations of the study

This method has the advantage of being a commonly used non-invasive technique which is now revealed to hold additional information and can be applied to an entire manuscript easily. However, one limitation is that PMF does not provide amino acid sequence-level information for precise peptide identification, but it could be circumvented in the future with the use of TOF/TOF mass spectrometers. This study has also highlighted possible issues arising from batch effects and should be accounted for in future studies and further investigated to develop better methods to correct them. Finally, where the undertext is revealed, so is the original source, therefore if more palimpsests are tested with imaging techniques, this method can be directly ground-truthed.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Visual analysis
 - Protein analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Species identification
 - Calculation of glutamine deamidation and PQI
 - MS data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106786>.

ACKNOWLEDGMENTS

We would like to thank Alberto Campagnolo (Université catholique de Louvain) for helping develop the VisColl model, Bharath Nair (University of Copenhagen) for assistance with PQI and Simon Hickinbotham (University of York) for technical support with Bacollite. We also thank Katarzyna A. Kapitan and Beatriz Fonseca, both from the University of Copenhagen, for helpful discussions on image analysis. Thanks also go to Frido Welker (University of Copenhagen) for his constructive feedback on the initial draft of this article and Liam Lanigan (University of Copenhagen) for the title suggestion. The graphical abstract has been designed using images from Flaticon.com. L.C.V.-C. is supported by the European Union's Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie grant agreement No 801199 and Danish National Research Foundation DNRF128. M.C. is supported by Danish National Research Foundation DNRF128 and EU Framework Program for Research and Innovation Horizon 2020 under Grant Agreement No. 787282. S.F. and J.V. are supported by EU Framework Program for Research and Innovation Horizon 2020 under Grant Agreement No. 787282.

AUTHOR CONTRIBUTIONS

N.F., S.F., and M.C. designed the research. N.F. collected samples and M.D. provided information about the manuscript. N.F. and J.V. conducted the visual analysis and interpretation. S.F. performed the ZooMS laboratory experiments. L.C.V.-C., I.R., and M.C. analyzed and interpreted the proteomic data. L.C.V.-C. and M.C. wrote the initial manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 22, 2022

Revised: March 7, 2023

Accepted: April 26, 2023

Published: April 29, 2023

REFERENCES

1. McKitterick, R. (2007). Palimpsests: concluding remarks. In *Early Medieval Palimpsests Bibliologia*, G. Declercq, ed. (Brepols Publishers), pp. 145–151. <https://doi.org/10.1484/M.BIB-EB.3.1428>.
2. R. Netz, W. Noel, N. Tchernetska, and N. Wilson, eds. (2011). *The Archimedes Palimpsest, 2* (Cambridge University Press).
3. Fiddyment, S., Teasdale, M.D., Vnouček, J., Lévêque, E., Binois, A., and Collins, M.J. (2019). So you want to do biocodicology? A field guide to the biological analysis of parchment. *Herit. Sci.* 7, 35. <https://doi.org/10.1186/s40494-019-0278-6>.
4. Fiddyment, S., Holsinger, B., Ruzzier, C., Devine, A., Binois, A., Albarella, U., Fischer, R., Nichols, E., Curtis, A., Cheese, E., et al. (2015). Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. *Proc. Natl. Acad. Sci. USA* 112, 15066–15071. <https://doi.org/10.1073/pnas.1512264112>.
5. Kirby, D.P., Buckley, M., Promise, E., Trauger, S.A., and Holdcraft, T.R. (2013). Identification of collagen-based materials in cultural heritage. *Analyst* 138, 4849–4858. <https://doi.org/10.1039/c3an00925d>.
6. Nair, B., Palomo, I.R., Markussen, B., Wiuf, C., and Collins, M.J. (2022). Parchment Glutamine Index (PQI): a novel method to estimate glutamine deamidation levels in parchment collagen obtained from low-quality MALDI-TOF data. Preprint at bioRxiv. ver. 6, peer-reviewed and recommended by Peer community in Archaeology. <https://doi.org/10.1101/2022.03.13.483627>.
7. Moll, J. (1986). Para una tipología de la notación catalana. *Rev. Musicol.* 9, 399–409. <https://doi.org/10.2307/20795067>.
8. Grifoll, I. (2017). The culture (Ninth–Twelfth centuries): clerics and troubadours. In *The Crown of Aragon - A Singular Mediterranean Empire*, F. Sabaté, ed. (Brill), pp. 125–149. https://doi.org/10.1163/9789004349612_006.
9. Bousset, W. (1895). *Nachrichten über eine Kopenhagener Handschrift (Arnemagnaeanske Legat 1927 A M. 795 4to) des Kommentars des Aringius zur Apocalypse. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. Philologisch-Historische Klasse, 187–209.*
10. Gryson, R. (2003). *Commentaria Minora in Apocalypsin Iohannis* (Brepols).
11. Fischer, B. Letter to Prof Dr Loth (9, August, 1957).
12. Driscoll, M. (2021). Spænsk handrit í safni Árna Magnússonar. In *Á fjarlægum ströndum: tengsl Spánar og Íslands í tímans rás*, E. Erlendsdóttir and K.G. Jónsdóttir, eds. (Háskólaútgáfan), pp. 323–340.
13. Driscoll, M. (2022). A Note on the Provenance of AM 795 4to (Opuscula XX), pp. 326–334.
14. Gregory, C.R. (1886). The quires in Greek manuscripts. *Am. J. Philol.* 7, 27–32. <https://doi.org/10.2307/287262>.
15. Padrós, N., and Valenzuela, S. (2010). La Llosa i els Antigons, una aproximació a la producció ramadera de les villae de l'ager Tarraconensis. Segles III-VI dC. In *Ager tarraconensis*, M. Prevosti Monclús, J. Guitart Duran, ed. (Institut d'Estudis Catalans), pp. 200–206.
16. Novella Dalmau, V. (2014). *Estudi de les pautes d'accés i consum dels recursos animals a partir de l'arqueozoologia l'exemple del Castell de Montsoriu*. Thesis (Autonomous University of Barcelona).
17. Colominas, L., Antolín, F., Ferrer, M., Castanyer, P., and Tremoleda, J. (2019). From vilalba to vila alba: changes and continuities in animal and crop husbandry practices from the early roman to the beginning of the Middle Ages in the north-east of the iberian peninsula. *Quat. Int.* 499, 67–79. <https://doi.org/10.1016/j.quaint.2017.12.034>.
18. Nieto Espinet, A., Huet, T., Trentacoste, A., Guimarães, S., Orenge, H., and Valenzuela-Lamas, S. (2021). Resilience and livestock adaptations to demographic growth and technological change: a diachronic perspective from the Late Bronze Age to Late Antiquity in NE Iberia. *PLoS One* 16, e0246201. <https://doi.org/10.1371/journal.pone.0246201>.
19. Teasdale, M.D., Fiddyment, S., Vnouček, J., Mattiangeli, V., Speller, C., Binois, A., Carver, M., Dand, C., Newfield, T.P., Webb, C.C., et al. (2017). The York Gospels: a 1000-year biological palimpsest. *R. Soc. Open Sci.* 4, 170988. <https://doi.org/10.1098/rsos.170988>.
20. Hickinbotham, S., Fiddyment, S., Stinson, T.L., and Collins, M.J. (2020). How to get your goat: automated identification of species from MALDI-ToF spectra. *Bioinformatics* 36, 3719–3725. <https://doi.org/10.1093/bioinformatics/btaa181>.
21. Ruffini-Ronzani, N., Nieuw, J.-F., Soncin, S., Hickinbotham, S., Dieu, M., Bouhy, J., Charles, C., Ruzzier, C., Falmagne, T., Hermand, X., et al. (2021). A biocodicological analysis of the medieval library and archive from Orval Abbey, Belgium. *R. Soc. Open Sci.* 8, 210210. <https://doi.org/10.1098/rsos.210210>.
22. Calvano, C.D., De Ceglie, C., Monopoli, A., and Zambonin, C.G. (2012). Detection of sheep and goat milk adulterations by direct MALDI-TOF MS analysis of milk tryptic digests. *J. Mass Spectrom.* 47, 1141–1149. <https://doi.org/10.1002/jms.2995>.
23. Reed, R. (1972). *Ancient Skins, Parchments and Leathers* (Seminar Press).
24. Mundó, A.M. (1967). Un fragment molt antic de litúrgia romana a Catalunya. II Congrés litúrgic de Montserrat, vol. III (Abadia de Montserrat), pp. 173–191.
25. Starynska, A., Messinger, D., and Kong, Y. (2021). Revealing a history: palimpsest text separation with generative networks. *Int. J. Doc. Anal. Recogn.* 24, 181–195. <https://doi.org/10.1007/s10032-021-00379-z>.
26. European Commission (2007). *Directorate-general for research and innovation. In Improved Damage Assessment of Parchment, IDAP: Assessment, Data Collection and Sharing of Knowledge*, R. Larsen, ed. (Publications Office).
27. Strohalm, M., Kavan, D., Novák, P., Volný, M., and Havlíček, V. (2010). mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Anal. Chem.* 82, 4648–4651. <https://doi.org/10.1021/ac100818g>.
28. Buckley, M., Whitcher Kansa, S., Howard, S., Campbell, S., Thomas-Oates, J., and Collins, M. (2010). Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *J. Archaeol. Sci.* 37, 13–20. <https://doi.org/10.1016/j.jas.2009.08.020>.
29. Buckley, M., Fraser, S., Herman, J., Melton, N.D., Mulville, J., and Pálsdóttir, A. (2014). Species identification of archaeological marine mammals using collagen fingerprinting. *J. Archaeol. Sci.* 41, 631–641. <https://doi.org/10.1016/j.jas.2013.08.021>.
30. Palarea-Albaladejo, J., Mclean, K., Wright, F., and Smith, D.G.E. (2018). MALDIrppa: quality control and robust analysis for mass spectrometry data. *Bioinformatics* 34, 522–523. <https://doi.org/10.1093/bioinformatics/btx628>.
31. Gibb, S., and Strimmer, K. (2012). MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* 28, 2270–2271. <https://doi.org/10.1093/bioinformatics/bts447>.
32. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. <https://doi.org/10.1093/bioinformatics/bts034>.
33. Gibb, S., and Strimmer, K. (2015). Differential protein expression and peak selection in mass spectrometry data by binary discriminant analysis. *Bioinformatics* 31, 3156–3162. <https://doi.org/10.1093/bioinformatics/btv334>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Parchment manuscript	Arnemagmæan Institute, University of Copenhagen	AM795 4to
Chemicals, peptides, and recombinant proteins		
Ammonium bicarbonate	Sigma-Aldrich	Cat#A6141
Trypsin	Promega	Cat#V5111
Trifluoroacetic acid	Sigma-Aldrich	Cat#302031
Acetonitrile	Sigma-Aldrich	Cat#271004
α -Cyano-4-hydroxycinnamic acid	Sigma-Aldrich	Cat#70990
Calibration standards [des-Arg-Bradykinin, Angiotensin I, Glu-Fibrinopeptide B, ACTH (1 - 17 clip), ACTH (18 - 39 clip), ACTH (7 - 38 clip)]	Sigma-Aldrich	Cat#B1901; Cat#A9650; Cat#F3261; Cat#A2407; Cat#A0673; Cat#A1527
Deposited data		
Spectra in mzML format	This paper	Zenodo: https://doi.org/10.5281/zenodo.7469803
Metadata used in data analysis	This paper	Zenodo: https://doi.org/10.5281/zenodo.7469803
Metadata including visual analysis results	This paper	Zenodo: https://doi.org/10.5281/zenodo.7469803
Software and algorithms		
mMass	(Strohalm et al., 2008)	http://www.mmass.org/
bacollite	GitHub	https://github.com/bioarch-sjh/bacollite
MALDIpqj	GitHub	https://doi.org/10.5281/zenodo.7105461
MALDIquant	CRAN	https://CRAN.R-project.org/package=MALDIquant
MALDIrppa	CRAN	https://CRAN.R-project.org/package=MALDIrppa
sva	Bioconductor	https://doi.org/10.18129/B9.bioc.sva
BinDA	CRAN	https://CRAN.R-project.org/package=binda
Scripts using original code and packages above	This paper	GitHub & Zenodo: https://doi.org/10.5281/zenodo.7406297
Other		
ultrafleX III MALDI-TOF mass spectrometer	Bruker Daltonics	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Laura C. Viñas Caron (laura@palaeome.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw MALDI-TOF-MS spectra and associated metadata have been deposited at Zenodo and are publicly available as of the date of publication. The DOI is listed in the [key resources table](#). Visual and additional data have been deposited at Zenodo and are also available in this paper's supplementary dataset as of the date of publication. The DOI is listed in the [key resources table](#).

- All original code has been deposited on GitHub (https://github.com/ismaRP/a_spanish_book) and is publicly available as of the date of publication. The DOI is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

METHOD DETAILS

Visual analysis

Improved Damage Assessment of Parchment (IDAP) guidelines were used for the visual analysis.²⁶ This procedure included the observation of characteristic hair follicle patterns; surface appearance; ink residues, including the number of columns, line spacing, and margin size; traces of tools left during the parchment-making process or palimpsest preparation such as ruling patterns, holes or scrapings; surface contamination; calcite deposits; gelatinisation; and signs of biodegradation (mould) or water damage. The thickness of the parchment was measured at three different locations along the edge and averaged. In addition, we investigated how the parchment was organised in the manuscript, including the determination of the hair and flesh sides and quire structure.

Finally, the examination also allowed us to reassemble different manuscript leaves of the same skin and estimate the original size of the skin. Visual matches were made based on the observation of traces of tools coming from the manufacture of parchment; the position of the anatomic features of the animal visible on parchment; the size and direction of the hair follicles; the traces left from the production of the original manuscripts which were used for the palimpsest (e.g. trimming, holes, ruling patterns); and finally, comparison with the results of the proteomic analysis, which was done independently.

Preliminary analysis was done by studying the digitised manuscript available online, and virtual reassembling of the whole parchment/animal skin was attempted from pictures of the individual folio (see [Figure S10](#)). Subsequent consultation of the physical manuscript provided further details. Observation and photography using transmitted and raking light was conducted.

Protein analysis

One sample per bifolio was taken using the non-invasive eraser-based sampling technique developed by Fiddyment et al.⁴ Duplicate samples were taken in the case of bifolios 1-8, 9-16 and 89r-89v, and carried out through all steps in an identical manner to assess precision and variance. Overall, 54 samples were obtained for protein analysis. eZooMS was conducted at BioArCh, University of York, following the same protocol described in Fiddyment et al.⁴ This involved the addition of 75 μ L of 50 mM ammonium bicarbonate solution (AmBic, NH_4HCO_3 , pH 8.0) and 1 μ L of trypsin (0.4 $\mu\text{g}/\mu\text{L}$ solution) to the eraser rubbings and then samples were incubated at 37 °C for 4 h. Enzymatic digestion was stopped by adding 1 μ L of 5% (vol/vol) trifluoroacetic acid (TFA). Peptides were extracted and purified using C18 resin ZipTips (Millipore) with a conditioning solution made up of 50% acetonitrile (ACN)/0.1% (vol/vol) TFA and a washing solution of 0.1% (vol/vol) TFA. Then, samples were eluted in a final volume of 50 μ L of conditioning solution. One microliter was spotted on a MALDI plate and 1 μ L of α -cyano-4-hydroxycinnamic acid matrix solution was added on top. Each sample was analysed in triplicate along with calibration standards. Samples were distributed in three different MALDI plates, air-dried and analysed at the Centre of Excellence in Mass Spectrometry, University of York, on a Bruker Ultraflex III matrix-assisted laser desorption/ionisation-time of flight mass spectrometer (MALDI-TOF-MS) in reflector mode over an m/z range of 800-3200.

QUANTIFICATION AND STATISTICAL ANALYSIS

Species identification

Raw MALDI-TOF-MS spectra converted to txt. files were manually analysed using mMass.²⁷ The replicates were averaged for each sample and peaks were detected with a signal-to-noise threshold of 3. Species identification was made after comparison with a reference database containing relevant peptide markers.^{5,28,29}

We also used a new automated software, Bacollite v.1.0.0, that provides a classification with an associated confidence score, to assess how robust the identifications are.²⁰ This automated approach derives a set of peptide masses from the sequence data of possible species (sheep, calf, goat and deer) for comparison

with the sample data, which is analysed by cross-correlation, counting the number of matches above a range of cross-correlation thresholds. It then arrives at a classification with an associated confidence measure. We compared the classifications and efficacy of the manual and automated methods (Data S1). Overall, the classifications were identical for 51 of the 54 samples, but Bacollite allowed us to resolve two unidentified cases. In some cases, the classification was ambiguous (with two scores being non-zero), as for folio 91 (Figure S1). This is because there are more matches for deer at lower correlation thresholds than sheep, but as the threshold increases, the matches for sheep are higher. This situation can happen when some of the key marker peptides for discrimination are poorly preserved. The score could thus be interpreted in different ways, but because more of the (highly-correlated) markers match the theoretical peptides we assign the sample as sheep. In addition, when the score was lower than 10, species identification also relied on the manual examination of the spectra, as suggested in.²¹

Calculation of glutamine deamidation and PQI

The Parchment Glutamine Index (PQI) was estimated following⁶ and using the R package MALDIpqi. First, spectra are preprocessed using default parameters for baseline removal and smoothing and isotopic distributions are extracted for a list of selected peptides of parchment-derived collagen type I. Then, the extent of glutamine deamidation (q) was calculated using a linear regression model for each peptide. In this step, weighted least squares is used to deconvolute the isotopic distribution of peptides, estimating an abundance coefficient of the overlapping deamidated and non-deamidated versions (since they are only 0.98 Da apart). From there, deamidation is calculated as a proportion of the estimates. Finally, an overall index termed Parchment Glutamine Index or PQI is predicted using a linear mixed-effect model at the sample, replicate and peptide level. This value ranges from 0 to 1, with higher values indicating that glutamines are not deamidated at all and, thus, present higher quality. Values above 1 can still occur if during the deconvolution the non-deamidated version has a coefficient lower than 0, due to low signal peaks and noisy background. These values, as long as they are close to 1, can still be considered as a PQI equal to 1. For example, two parts of one goatskin bifolio (9+16) show marked differences in PQI, whilst other paired analyses (sheepskin) show the expected consistency in values. Those differences could be explained by the mentioned reasons of signal quality, as shown in Figure S2.

MS data analysis

Quality control and data preprocessing

Text files were converted into .MzML format and imported into R. An initial screening was conducted to detect potential low-quality mass spectra with the R package MALDIrppa.³⁰ It identifies potential mass spectra outliers using the atypicality score (A score) (Hedges, 2008). Scores above or below the thresholds are possible outliers. We identified some mass spectra with extreme A scores due to higher levels of noise (Figure S3). However, after a visual inspection of the spectra (Figure S4), we did not find a significant difference compared with other samples, so at this stage, we decided to keep them for subsequent analysis. Furthermore, the presence of two clusters with different distributions was observed (samples UoC1 to 24 and UoC24 to 54), which correspond to the different batches.

Subsequently, we used the MALDIquant and MALDIrppa R packages to preprocess all the MS data.^{30,31} We used default parameters unless specified. The preprocessing involved common procedures: square root intensity transformation for variance stabilisation; smoothing using undecimated discrete wavelet transform (UDWT) with a threshold scale of 2.5 to reduce noise; baseline correction using the statistics-sensitive non-linear iterative peak-clipping (SNIP) algorithm with 100 iterations to remove background noise; and total ion current (TIC) normalisation for comparison of intensity values across different spectra. We then performed peak detection using the SuperSmoother noise estimation algorithm with a signal-to-noise ratio (SNR) of 5 and a half-window size set to 20. We previously tried several SNRs and selected a threshold of 5 after visual inspection. This step was followed by peak alignment by calculating a warping “lowess” (Locally Weighted Scatterplot Smoothing) function for each spectrum against a reference spectrum of common peaks. This step helps bring peaks that have the same mass but a certain measurement error closer. However, the mass is still not numerically exact and binning with 0.002 Da tolerance was performed to equalise masses that fall closer. An additional filtering step was done in which only peaks that occurred at least in two replicates were kept. Those peaks not passing this criterion were still kept if they passed it in other groups of replicates. Technical replicates were averaged for each sample. Finally, a feature matrix was generated with the peak lists ($n=442$), where the rows represent the samples, and cells are filled with processed intensity values or 0 for missing peaks.

Additional quality controls were effectuated after the preprocessing, in which the number of peaks (Figure S5) and peak intensity patterns (Figure S6) were explored. We observe that in samples processed in the second batch (UoC 24-54), the number of peaks is more reduced, affecting mostly high molecular weight peptides.

Subsequently, data was adjusted for batch effects to prevent possible bias in posterior analyses and ensure results are biologically meaningful. We used the *sva* function within the *sva* package which estimates surrogate variables for unknown sources of variation.³² We supplied a model with the parchment species and the quires as variables of interest or known sources of variability. Then *sva* estimates unwanted sources of variation, among which we have the batch, and regresses out the surrogate variables.

Downstream analysis

To identify similarities or differences among the samples, we conducted a Principal Component Analysis (PCA) and hierarchical clustering analysis of all the samples (Figure S7) employing the euclidean distance function and “ward.D2” agglomerative method using the R package *pheatmap*. Then, we repeated the PCA (Figure S8) and hierarchical clustering but separating for species, to observe within-species variability. This allows us to better observe the impact of other factors, including damage patterns, without the influence of certain collagen peptides that are known to be species-specific.

Following this, a binary discriminant analysis was performed using the *BinDA* R package³³ on the sheep and goat samples separately and with the previously identified clusters as a class. This produces a score that can be used to rank peak masses by discriminating importance. Then, we examined if the top 20 peaks were collagenous or non-collagenous peptides. We generated a list of theoretical masses after *in silico* trypsin digestion of sheep and goat collagen type I sequences, allowing 1 miscleavage and adding glutamine and asparagine deamidation and hydroxyproline as possible peptide modifications. Then, the theoretical masses of the peptides were matched to the observed peak masses, with a tolerance of 0.6 Da. After this, we performed an additional verification step to the masses that matched the theoretical list, in which *Bacollite* was used to align theoretical spectra (both deamidated and non-deamidated versions) against the raw sheep or goat spectra, using cross-correlation as it is done for species identification.²⁰ Peaks that were not initially found on the theoretical list or gave a low cross-correlation value after the alignment, were considered as potentially derived from non-collagenous proteins.