



**HAL**  
open science

## Paléoprotéomique appliquée au site de Waziers, le “ Bas-Terroir ” (Nord)

Fabrice Bray, Stéphanie Flament, Isabelle Fabrizi, Christian Rolando, Patrick  
Auguste

► **To cite this version:**

Fabrice Bray, Stéphanie Flament, Isabelle Fabrizi, Christian Rolando, Patrick Auguste. Paléoprotéomique appliquée au site de Waziers, le “ Bas-Terroir ” (Nord). Quaternaire, 2023, vol. 34/1, pp.9-22. 10.4000/quaternaire.17615 . hal-04467645

**HAL Id: hal-04467645**

**<https://hal.science/hal-04467645v1>**

Submitted on 20 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PALÉOPROTÉOMIQUE APPLIQUÉE AU SITE DE WAZIERS, LE «BAS-TERROIR» (NORD)



Fabrice BRAY<sup>1</sup>, Stéphanie FLAMENT<sup>1</sup>, Isabelle FABRIZI<sup>1</sup>, Christian ROLANDO<sup>1</sup>, Patrick AUGUSTE<sup>2</sup>

## RÉSUMÉ

Les anciennes molécules conservées dans les ossements fossiles offrent l'opportunité d'acquérir une meilleure connaissance du passé biologique. Ces dernières années, la protéomique appliquée aux échantillons osseux et dentaires est devenue une méthode intéressante pour étudier les espèces éteintes et l'évolution phylogénétique comme alternative à l'analyse de l'ADN qui est limitée par l'amplification de l'ADN présent dans les échantillons anciens et sa contamination. Pour endommager le moins possible les échantillons, il est primordial de prélever une faible quantité de matière. Pour certains échantillons fossiles qui peuvent être très précieux car uniques (restes humains, taxons très rares), ces prélèvements sont problématiques. Une autre difficulté est l'absence de données génomiques pour la plupart des espèces éteintes. Dans cette étude, une méthodologie de protéomique a été appliquée à des os de mammifères âgés d'environ 130 000 à 120 000 ans provenant du site pléistocène de Waziers (Nord, France). La présence d'artefacts lithiques et les traces de boucherie observées sur des ossements indiquent la présence de l'Homme de Neandertal sur ce site. À partir d'échantillons de cinq milligrammes, nos résultats montrent une grande majorité de peptides détectés correspondant aux protéines de collagène I alpha 1 et alpha 2 avec une couverture de séquence jusqu'à 60 %. En utilisant l'homologie avec des séquences modernes, une classification biologique a été réalisée avec succès et les rangs taxonomiques associés à chaque os ont été identifiés de manière cohérente avec les informations obtenues à partir d'études ostéomorphologiques et de données paléoenvironnementales et paléodietétiques

**Mots-clés :** paléoprotéomique, Pléistocène, Waziers, ossements de mammifères, taxonomie, modifications post traductionnelles.

## ABSTRACT

### PALEOPROTEOMICS APPLIED TO THE WAZIERS SITE, «BAS-TERROIR» (NORTH OF FRANCE)

The ancient preserved molecules in bones offer the opportunity to gain a better knowledge on the biological past. In recent years, bones proteomics has become an attractive method to study extinct species and phylogenetic evolution as an alternative to DNA analysis which is limited by DNA amplification present in ancient samples and its contamination. The analysis of fossils must consume a low quantity of material to avoid damaging the samples. For some fossil samples that may be very valuable because they are rare (human remains, very scarce taxa). Another difficulty is the absence of genomic data for most of the extinct species. This study applied a proteomic methodology to mammalian bones of 130,000 to 120,000 years old from the last Pleistocene site of Waziers (France). The presence of lithic artefacts and the traces of butchery observed on the bones indicate the presence of Neanderthal Man on this site. Starting from 5 milligram samples, our results show that most detected peptides match collagen I alpha 1 and alpha 2 proteins with a sequence coverage up to 60%. Using sequence homology with modern sequences, a biological classification was successfully achieved. Each bone taxonomic rank by proteomics was consistent with the existing osteomorphological studies and palaeoenvironmental and palaeodietary data.

**Keywords:** Palaeoproteomics, Pleistocene, Waziers, mammalian bones, taxonomy, post translational modifications.

## 1 - INTRODUCTION

Au cours des dernières décennies, les études sur les biomolécules anciennes ont changé notre compréhension de l'évolution de la vie sur Terre. Traditionnellement, les approches paléontologiques identifient les ossements (humains et animaux) provenant de dépôts naturels ou anthropiques par la comparaison anatomique, puis permettent une étude biostratigraphique et la reconstitution des écosystèmes environnants. Ces analyses ne fournissent qu'une preuve indirecte des facteurs et des mécanismes qui ont créé la biodiversité

actuelle. Les anciennes biomolécules offrent une fenêtre sur le passé biologique et nous permettent de suivre les processus évolutifs (Schroeter *et al.*, 2022). Aujourd'hui, pour analyser ces anciennes biomolécules, un ensemble de techniques est utilisé telles que la paléogénétique (ADN ancien), les études isotopiques, la racémisation des acides aminés et, plus récemment, les techniques de protéomique. Il a été montré que les plus anciennes traces d'ADN ont été détectées à partir d'un mammoth conservé dans le pergélisol (environ 1,6 millions d'années) (Van der Valk *et al.*, 2021) et même de l'ADN environnemental au Groenland vers deux millions

<sup>1</sup> Univ. Lille, CNRS, UAR 3290 - MSAP - Miniaturisation pour la Synthèse, l'Analyse et la Protéomique, FR-59000 LILLE. *Courriels :* fabrice.bray@univ-lille.fr; stephanie.flament@univ-lille.fr; isabelle.fabrizi@univ-lille.fr; christian.rolando@univ-lille.fr

<sup>2</sup> Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paléo - Evolution, Ecologie et Paléontologie, FR-59000 LILLE. *Courriel :* patrick.auguste@univ-lille.fr

d'années (Kjaer *et al.*, 2022). Toutefois, cet ADN se dégrade très vite et, en dehors de ce type de contexte exceptionnel, la limite temporelle de son utilisation est de quelques dizaines de milliers d'années. Les protéines offrent une alternative biomoléculaire dans les zones, les périodes et les tissus où l'ADN ancien ne survit généralement pas. Les protéines conservées dans les tissus minéralisés comme les os ou les dents fournissent des séquences d'acides aminés phylogénétiquement informatives, comme le montrent les analyses réalisées sur le collagène d'ossement de camélidé daté de 3,4 millions d'années découvert en Arctique (Buckley *et al.*, 2019a) ou les protéines de dent de *Gigantopithecus blacki* âgé de 1,9 millions (Welker *et al.*, 2019) ou encore des Rhinocerotidae eurasiens âgés de 1,77 millions (Cappellini *et al.*, 2019). La protéomique est donc une méthode pertinente pour l'analyse des fossiles et devient une alternative à l'analyse de l'ADN qui est limitée par son amplification et la contamination.

Actuellement, la paléoprotéomique utilise deux méthodes : l'approche ZooMS (Zooarcheology by Mass Spectrometry) (Buckley *et al.*, 2009), qui permet l'identification des espèces par empreintes digitales peptidiques et une seconde méthode d'analyse plus élaborée qui est fondée sur la chromatographie liquide couplée à la spectrométrie de masse (LC-MS/MS).

La LC-MS/MS sépare d'abord les peptides d'un extrait à l'aide de la chromatographie liquide, puis applique ces peptides séparés à un système de spectrométrie de masse couplé pour des analyses robustes et rigoureuses à l'aide de la spectrométrie de masse en tandem. Les peptides en solution sont séparés en fonction de leur hydrophobicité sur une colonne en phase inverse, puis sont introduits séquentiellement dans le spectromètre de masse par ionisation par électronebulisation (ESI). Cette méthode va permettre d'obtenir la séquence en acides aminés des peptides, d'identifier et de localiser les modifications post-traductionnelles. La première analyse ESI-LC-MS/MS a été appliquée aux restes de mammoth âgé de 43 000 ans (Cappellini *et al.*, 2011). Depuis, la méthode a été largement appliquée à l'étude d'autres espèces éteintes telles que *Castoroides ohioensis* (Cleland *et al.*, 2016), *Bison latifrons* (Hill *et al.*, 2015), des oiseaux anciens (Horn *et al.*, 2019), des reptiles (Harvey *et al.*, 2019a,b) et des dinosaures (Schroeter *et al.*, 2017). Le séquençage des peptides utilisé en LC MS/MS donne accès à la phylogénie et à l'histoire évolutive des espèces anciennes. Cette approche a été appliquée à l'étude des espèces endémiques d'Amérique du Sud (Welker *et al.*, 2015), d'anciens spécimens d'Hominidés (Brown *et al.*, 2016; Welker *et al.*, 2016) et de paresseux arboricoles (Presslee *et al.*, 2019). Actuellement, la méthode de préparation a été automatisée, la quantité d'échantillons utilisée a été diminuée jusque 5 mg et l'utilisation de banque de librairie peptidique a été employée permettant de réduire le temps des analyses (Ruethe *et al.*, 2021). Les études sur les ossements plus anciens et en particulier sur les fossiles de dinosaures sont très difficiles car elles sont entravées par la présence de contaminants. Ces études montrent l'importance de la préparation des échantillons afin d'éviter toute contamination, la nécessité du contrôle

qualité de l'analyse par spectrométrie de masse et de la validation du traitement des données (Buckley *et al.*, 2008, 2017a).

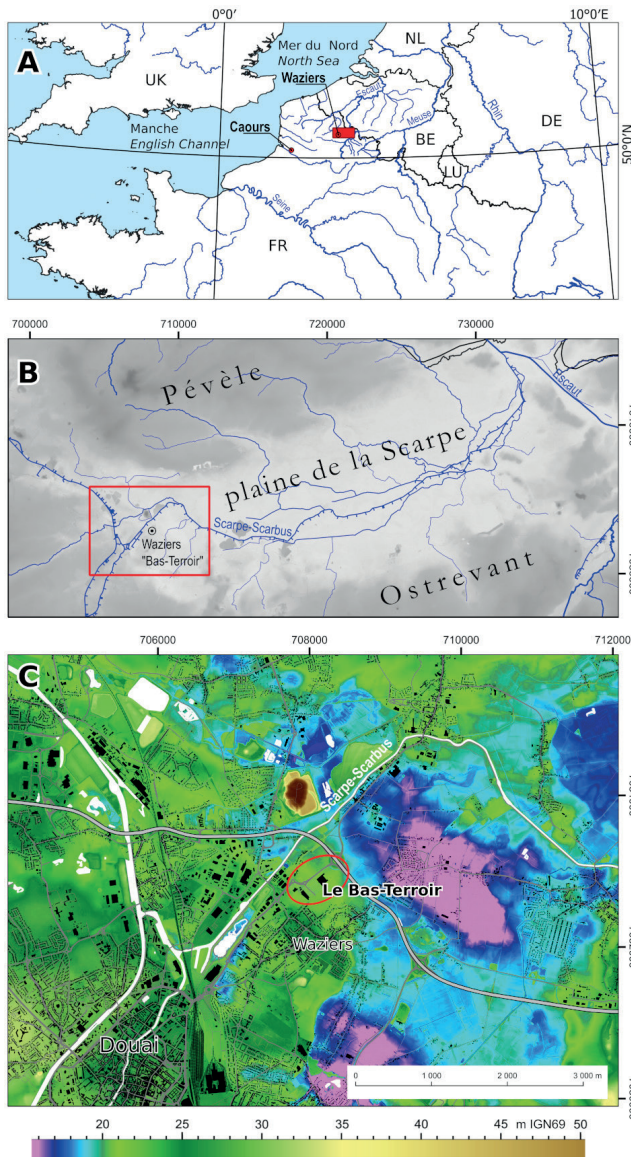
Nous présentons ici une analyse protéomique robuste pour étudier les espèces éteintes du site Le Bas-terroir à Waziers (Nord, France) datées de la fin du Chibanién (nouvelle dénomination du Pléistocène moyen) et du début du Pléistocène supérieur (130 000 / 120 000 ans BP). La méthode a été appliquée à 15 échantillons en utilisant cinq milligrammes de matière. Dix ossements identifiés anatomiquement et cinq incertains ont été analysés. Nos analyses ont abouti à l'identification des rangs taxonomiques de chaque ossement. La conservation de la matière organique dans les ossements a permis d'identifier des modifications post traductionnelles comme des méthylations, di-méthylations, alkylations et hydroxylations sur la séquence de collagène et de localiser leur position. Cette étude montre la puissance de cette technique pour l'analyse d'ossements anciens.

## 2 - MATÉRIEL ET MÉTHODES

### 2.1 - ÉCHANTILLONS

Les fouilles menées à Waziers, le Bas-Terroir en 2014 et 2015 puis en 2017 ont permis la découverte de plusieurs dizaines de vestiges dentaires et osseux de vertébrés. Ce site se localise dans le département du Nord (Hérisson *et al.*, 2022), en périphérie de l'agglomération de Douai (fig. 1). Les vestiges se répartissent dans 10 unités stratigraphiques de synthèses, US2, US3, US4, US5a à US5e, US6 et US7 (fig. 2) et couvrent trois membres litho-stratigraphiques (Deschodt *et al.*, 2022). Le membre 1 est composé de sables et graviers (US3) surmontés par un limon gris (US4). Il est interprété comme alluvions de la fin du Saalien. Le membre 2 est composé de dépôts limono-organique (US5a) et d'une tourbe (US5b, US5e) où s'insèrent des dépôts alluviaux limoneux clastiques (US5c, US5d). Il est interprété comme dépôt de zone humide ou de chenal dans un contexte de passage aux conditions interglaciaires eemiennes. Le membre 3 scelle et débordé largement les membres 1 et 2 qui constituent la Formation de Waziers. Il débute par des dépôts grossiers en ravines (US6) associés à une érosion et à des marqueurs de froids suivi par un dépôt fluvial extensif de graviers de craie (US7a). US6 et US7a remanient des éléments présents dans les unités sous-jacentes eemiennes et sont attribués au glaciaire Weichselien. Le membre 3 est ensuite composé d'un dépôt fluvio-éolien lité (US7b) passant à un loess en partie supérieure (US7c) attribué à la fin du Pléniglaciaire Weichselien.

La méso et macrofaune découvertes au sein de la séquence de Waziers ont fait l'objet d'une analyse exhaustive et ont permis d'établir la liste des taxons présents (Auguste *et al.*, 2022). C'est l'aurochs (*Bos primigenius*) qui est l'espèce la mieux représentée, suivi du chevreuil (*Capreolus capreolus*), du castor (*Castor fiber*), du cheval d'Achenheim (*Equus achenheimensis*), du rhinocéros de prairie (*Stephanorhinus hemitoechus*).



**Fig. 1: Localisation du gisement de Waziers «Bas-Terroir» (France).**

(A) Localisation de la plaine de la Scarpe en Europe du nord-ouest. Coordonnées UTM. (B) Localisation du gisement de Waziers dans la plaine de la Scarpe. Coordonnées en mètres Lambert 93. (C) Localisation du gisement de Waziers «Bas-Terroir» sur un MNT issu du Lidar Scarpe (Source DREAL; D.A.O.: L. Deschodt).

*Fig. 1: Location of the Waziers «Bas-Terroir» deposit (France). (A) Location of the Scarpe Plain in northwestern Europe. UTM coordinates. (B) Location of the Waziers deposit in the Scarpe plain. Coordinates in metres Lambert 93. (C) Location of the Waziers «Bas-Terroir» deposit on a DTM from the Scarpe Lidar (Source DREAL; C.A.D.: L. Deschodt).*

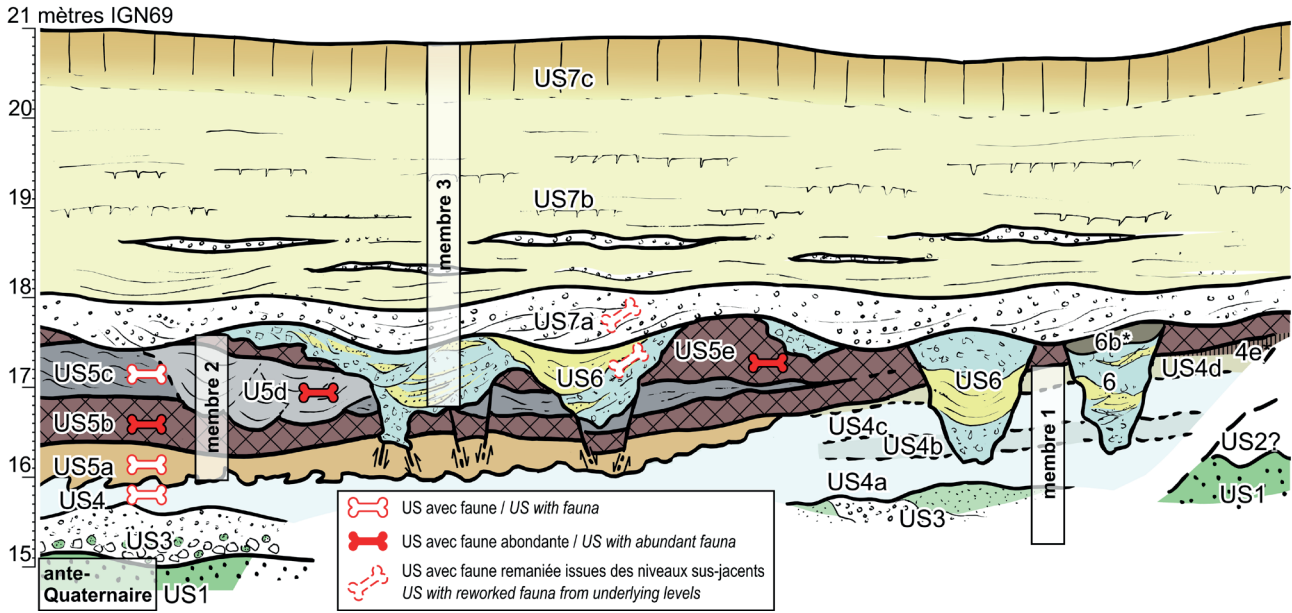
Le cerf élaphe (*Cervus elaphus*), l'ours brun (*Ursus arctos*), le mégacéros (*Megaloceros giganteus*), la loutre (cf. *Aonyx antiqua*) et peut être le daim (cf. *Dama dama*) sont plus anecdotiques.

Quinze ossements provenant du site «Le Bas-terroir» de Waziers ont été échantillonnés. Ils proviennent des fouilles 2014, 2015 et 2017 (tab. 1). La problématique de cet échantillonnage a été de tester la détermination anatomique *versus* l'identification protéomique, avec des pièces attribuées sans ambiguïté à un taxon (Wa-1, 2, 3, 5, 7, 10, 11, 12, 13, 15), ainsi que des pièces à détermination

incertaine et nécessitant la caractérisation de l'espèce (Wa-4, 6, 9, 10, 14). Par ailleurs, deux os ont été prélevés deux fois en utilisant un protocole sensiblement différent, afin de mettre en évidence une éventuelle ou non modification des résultats suivant ce critère (Wa-3 et 13, Wa-5 et 9).

## 2.2 - DIGESTION EFASP

La méthode de digestion des échantillons est basée sur la publication de Bray *et al* (2022). Les échantillons ont été digérés avec la méthode eFASP adapté de Erde *et al.* (2014). Avant utilisation, des filtres ultracentrifuges Amicon® de 0,5mL avec un seuil de coupure de 10 kDa (EMD Millipore, Darmstadt, Allemagne) ont été fraîchement incubés pendant une nuit avec la solution de passivation contenant 5 % (v/v) de Tween®-20 dans l'eau. Les filtres Amicon® ultracentrifuges ont été lavés quatre fois au bain-marie pendant 20 minutes. La suspension de poudre d'os et la fraction déminéralisée ont été transférées sur des filtres Amicon® séparés, 100 µL de tampon d'échange ont été ajoutés (urée 8 M, 0,2 % DCA (acide désoxycholique) p/p, bicarbonate d'ammonium 100 mM pH 8,8) puis les filtres Amicon® ont été centrifugés pendant 30 min à 10 000 g à température ambiante. Les filtrats ont été jetés puis 200 µL de tampon d'échange ont été rajoutés dans les filtres Amicon® qui ont été centrifugés. Cette étape a été répétée deux fois. Les protéines ont été alkylées pendant 1 h à température ambiante dans l'obscurité en utilisant 100 µL de tampon d'alkylation (urée 8 M, iodoacétamide 50 mM et bicarbonate d'ammonium 100 mM, pH 8,8). Les filtres Amicon® ont été centrifugés pendant 30 min à 10 000 g et le filtrat a été jeté. Après l'étape d'alkylation, 200 µL de tampon d'échange ont été ajoutés au filtre Amicon® qui a été centrifugé pendant 30 min à 10 000 g et le filtrat a été jeté. 200 µL de tampon de digestion (0,2 % DCA (acide désoxycholique) p/v, bicarbonate d'ammonium 50 mM pH 8,8) ont été ajoutés au filtre Amicon® puis centrifugés. Cette étape a été répétée deux fois, en jetant le filtrat à chaque étape. Les filtres Amicon® ont été transférés dans de nouveaux tubes de 2 ml. 100 µL de tampon de digestion et 40 µL de trypsine/LysC à 0,05 µg/µL dans un tampon bicarbonate d'ammonium 50 mM pH 8,8 (Promega, Madison, USA) ont été ajoutés et incubés sous agitation dans un tube à bloc chauffant (MHR23, Hettich, Pays-Bas) pendant une nuit à 37°C. Après cette étape, les peptides présents dans le filtre Amicon® ont été récupérés dans le tube inférieur par centrifugation pendant 15 min à 10 000 g. Afin d'obtenir un maximum de peptides, les filtres Amicon® ont été lavés deux fois avec 50 µL de solution de bicarbonate d'ammonium (50 mM pH 8,8). Les filtrats contenant tous les peptides ont été transférés dans des tubes Eppendorf® de 1,5 mL. 200 µL d'acétate d'éthyle et 2,5 µL de TFA ont été ajoutés induisant l'apparition d'un précipité blanc. Aussitôt, 800 µL d'acétate d'éthyle sont à nouveau rajoutés et les tubes Eppendorf® sont centrifugés 10 min à 10 000 g. La phase organique a été éliminée. Cette étape a été répétée deux fois. Le tube Eppendorf® a été placé pendant 5 min



**Fig. 2 : Synthèse lithostratigraphique des campagnes de fouille 2014-2015 et position des vestiges de vertébrés découverts à Waziers, le Bas-Terroir.**

US1/ Substrat antequaternaire (sable éocène). US2/ Formation de versant. US3/ Dépôts grossiers fluviaux (craie, sable glauconieux, grès glauconieux) US4/ Dépôts fluviaux fins à dominante clastique (4a/ limon gris inférieur. 4b/ limon légèrement organique. 4c/ limon gris supérieur. 4d/ limon brunâtre. 4e/ Horizons pédologiques). US5/ Dépôts fluviaux fins à dominante organique (5a/ Limon brunâtre inférieur. 5b/ Tourbe laminée inférieure, riche en avifaune. 5c/ Limon gris laminé, remplissage d'un large chenal. 5d/ limon gris, remplissage de chenaux étroits. 5e/ Limon brunâtre supérieur et tourbe, riche en bois végétal et en os de castors. US6/ Dépôts fluviaux remaniés, graviers d'argile et figures thermokarstiques. US7/ Dépôts de couverture (7a/ graviers de craie. 7b/ sable et limon lité. 7c/loess et pédogenèse). N0 à N3/ niveaux archéologiques. \*/unité exclusivement présente sur la fouille 2015.

*Fig. 2: Lithostratigraphic synthesis of 2014-2015 excavation and location of vertebrate remains at Waziers, Le Bas-Terroir. 1/ Antequaternary substratum (Eocene sand). 2/ Slope formation. 3/ Fluvial coarse deposits (chalk, glauconitic sand and glauconitic sandstones). 4/ Thin fluvial mainly clastic deposits (4a/ lower grey silt. 4b/ slighty organic silt. 4c/ Upper grey silt. 4d/ Upper brownish loam. 4e/ Soil horizons). 5/Thin fluvial mainly organic deposits (5a/ Lower brownish silt. 5b/ Lower laminated peat, with rich avifauna. 5c/ Laminated grey silt, large channel infilling. 5d/ Grey silt, narrow channels infilling. 5e/ Upper brownish silt and peat; rich in wood pieces and beavers bones). 6/ Reworked fluvial deposit, clayey gravels and thermokarstic figures. 7/ Covering deposits (7a/ chalk gavels sheet. 7b/ Bedded sand and silt. 7c/loess and weathering). N0 to N3/ archaeological levels. (\*)/unit exclusively present on the 2015 excavation.*

	identité archéo	espèce probable	os
Wa 1	US5b, 7 et 17	rhinocéros	maxillaire
Wa 2	US5e, 6	rhinocéros	tibia
Wa 3	US5b, 154	castor	coxal
Wa 4	US5b, 173	cygne ?	carpo-métacarpe
Wa 5	US6, 615(*)	aurochs	humérus fracturé
Wa 6	US6, 79	aurochs ?	tibia fracturé
Wa 7	US4, 529 et 530(*)	cheval	humérus
Wa 8	US5, 500(*)	chevreuil	mandibule
Wa 9	US6, 615(*)	aurochs ?	humérus fracturé
Wa 10	US6, 621(*)	mammouth ?	dent
Wa 11	US6, 80-81	chevreuil	radius
Wa 12	US5b, 702	loutre	tibia avec stries
Wa 13	US5b, 154	castor	coxal
Wa 14	US5d, 27	oiseau ?	tibio-tarse
Wa 15	US5e, 87	castor	os long

**Tab. 1: Inventaire des échantillons des faunes mammaliennes de Waziers utilisées par l'analyse protéomique.**

Wa-1 à Wa-15: identification des échantillons. (\*) Fouille 2017, stratigraphie donnée en équivalent US 2014-2015.

*Tab. 1: Inventory of the samples of the Waziers mammalian faunas used in the proteomic analysis. Wa-1 to Wa-15: identification of samples. (\*) Excavation 2017, stratigraphy given in US equivalent 2014-2015.*

à 60°C dans des tubes à bloc chauffant (SBH130, Stuart, Staffordshire, Royaume-Uni) pour éliminer l'acétate d'éthyle restant. Les échantillons ont été évaporés à sec à température ambiante avec un concentrateur SpeedVac™. La paroi des tubes Eppendorf® a été rincée avec 100 µl d'un mélange méthanol/eau (50/50 v/v) et de nouveau évaporée à sec. Pour l'analyse par spectrométrie de masse, l'échantillon a été dissous à nouveau dans 10 µl d'une solution aqueuse contenant 0,1 % d'acide formique. La concentration a ensuite été estimée en mesurant la DO à 215 nm en utilisant 1 µl de la solution à l'aide d'un spectromètre UV gouttelettes (DS-11+, Denovix, Wilmington, USA). Les quantités de peptides obtenues après digestion sont indiquées dans le tableau 2. Les échantillons ont été dilués à une concentration de 1 µg/µL avant analyse dans le solvant A (5 % acétonitrile et 0,1 % acide formique v/v).

### 2.3 - CHROMATOGRAPHIE LIQUIDE-SPECTROMÉTRIE DE MASSE EN TANDEM

Les analyses LC-MS/MS ont été effectuées sur un spectromètre de masse Orbitrap Q Exactive plus couplé

Nom de l'échantillon	Identification LC-MS/MS			Peptides au-dessus du seuil d'identité (p<0,01)	Peptides au-dessus du seuil d'homologie (p<0,01)	Nombre de protéines identifiées	Numéro d'accèsion NCBI	Espèces identifiées par LC-MS/MS	Collagène de type I	Score Mascot	Nombre de peptide	Nombre de séquences distinctes significatives
	Quantité en µg de peptide pour 5 mg de poudre	Famille	Genre									
Wa-1	Rhinocéros	Rhinocerotidae	Ceratotherium	382	1153	113	XP_004434736.1 XP_004431414.1	<i>Ceratotherium simum simum</i> <i>Ceratotherium simum simum</i>	COL1A1 COL1A2	6352 5473	415 326	66 57
Wa-2	Rhinocéros	Rhinocerotidae	Ceratotherium	300	912	202	XP_004434736.1 XP_004431414.1	<i>Ceratotherium simum simum</i> <i>Ceratotherium simum simum</i>	COL1A1 COL1A2	5380 373	371 225	68 53
Wa-3	Castor	Castoridae	Castor	337	1085	109	JAV39636.1 JAV39576.1	<i>Castor canadensis</i> <i>Castor canadensis</i>	COL1A1 COL1A2	5856 6282	427 336	68 66
Wa-4	Cygne*	Anatidae	Cygnus	340	1267	61	XP_035424403.1 XP_032847830.1	<i>Cygnus olor</i> <i>Cygnus olor</i>	COL1A1 COL1A2	11982 7134	722 395	97 85
Wa-5	Aurochs	Bovidae	Bos	177	616	100	XP_006054012.1 XP_006054012.1	<i>Bubalus bubalis</i> <i>Bubalus bubalis</i>	COL1A1 COL1A2	3047 2902	225 174	52 46
Wa-6	Aurochs*	Cervidae	Cervus/Muntiacus	302	986	111	KAB036169.1 XP_043318553.1	<i>Muntiacus muntjak</i> <i>Cervus canadensis</i>	COL1A1 COL1A2	6315 4851	401 311	61 65
Wa-7	Cheval	Equidae	Equus	435	1405	86	ACM24774.1 XP_001492989.1	<i>Equus asinus</i> <i>Equus asinus</i>	COL1A1 COL1A2	9448 6268	623 463	67 63
Wa-8	Chevreuil	Cervidae	Cervus	350	1296	97	XP_023508478.1 XP_001492989.1	<i>Cervus canadensis</i> <i>Cervus canadensis</i>	COL1A1 COL1A2	7793 5506	547 367	73 55
Wa-9	Aurochs*	Bovidae	Bos	475	1578	127	AAI05185.1 AAI49096.1	<i>Bos taurus</i> <i>Bos taurus</i>	COL1A1 COL1A2	7065 7014	494 451	62 62
Wa-10	Mammouth*	Elephantidae	Loxodonta	394	1423	93	XP_010592644.1 XP_010585560.1	<i>Loxodonta africana</i> <i>Loxodonta africana</i>	COL1A1 COL1A2	8625 8979	605 501	65 61
Wa-11	Chevreuil	Cervidae	Cervus	430	1487	96	XP_043327093.1 XP_043318553.1	<i>Cervus canadensis</i> <i>Cervus canadensis</i>	COL1A1 COL1A2	9565 6799	660 466	70 62
Wa-12	Loutre	Mustelidae	Enhydra	410	1439	97	XP_022364932.1 XP_022358799.1	<i>Enhydra lutris kenyoni</i> <i>Enhydra lutris kenyoni</i>	COL1A1 COL1A2	9157 5220	706 372	72 56
Wa-13	Castor	Castoridae	Castor	384	1240	83	JAV39636.1 JAV39576.1	<i>Castor canadensis</i> <i>Castor canadensis</i>	COL1A1 COL1A2	7305 6890	553 406	61 69
Wa-14	Oiseau*	Cervidae	Cervus	371	1441	103	XP_043327093.1 XP_043327121.1	<i>Cervus canadensis</i> <i>Cervus canadensis</i>	COL1A1 COL1A2	7944 7219	591 555	67 63
Wa-15	Castor	Castoridae	Castor	405	1190	72	JAV39636.1 JAV39576.1	<i>Castor canadensis</i> <i>Castor canadensis</i>	COL1A1 COL1A2	8017 6275	552 376	67 59

**Tab. 2 : Informations concernant les identifications par LC-MS/MS des échantillons.**

Le tableau reprend le meilleur résultat d'identification pour le COL1A1 et COL1A2 avec son numéro d'accèsion NCBI et l'espèce identifiée. La quantité de peptide après digestion en µg, le nombre de peptides au-dessus du seuil d'identité (p<0,01), le nombre de peptides au-dessus du seuil d'homologie (p<0,01), le nombre de spectres interrogés, le nombre de protéines identifiées, le score Mascot, le nombre de peptide identifiées, le nombre de séquences distinctes significatives, le nombre de séquences uniques significatives, le pourcentage de protéines de couverture sont indiqués pour chaque collagène. Le tableau comprend l'identification anatomique du paléontologue et l'attribution de la famille et du genre identifiées par spectrométrie de masse. \* correspondants aux échantillons ayant une identification anatomique incertaine.

Tab. 2. Information concerning the LC-MS/MS identifications of the samples. The table shows the best identification result for COL1A1 and COL1A2 with its NCBI accession number and the species identified. The amount of peptide after digestion in µg, the number of peptide matches above identity threshold (p<0,01), the number of peptide matches above homology threshold (p<0,01), the number of queried spectra, the number of identified proteins, the Mascot score, number of peptides identified, the number of significant unique sequences, the percentage of protein coverage are shown for each collagen. The table includes the paleontologist's anatomical identification and assignment of family and genus identified by mass spectrometry. \* corresponds to samples with uncertain anatomical identification.

à un système HPLC microfluidique U3000 RSLC (ThermoFisher Scientific, Waltham, Massachusetts, États-Unis) (Helle *et al.*, 2019). 1 µl du mélange de peptides à une concentration de 1 µg/µL a été injecté avec le solvant A (5 % acétonitrile et 0,1 % acide formique v/v) pendant 3 min à un débit de 10 µl.min<sup>-1</sup> sur un Acclaim PepMap100 Pré-colonne C18 (5 µm, 300 µm id × 5 mm) de ThermoFisher Scientific. Les peptides ont ensuite été séparés sur une colonne en phase inverse C18 Acclaim PepMap100 C18 (3 µm, 75 mm id × 500 mm), en utilisant un gradient linéaire (5-40 %) de la solution B (75 % acétonitrile et 0,1 % acide formique) à un débit de 250 nL.min<sup>-1</sup> en 160 minutes puis 100 % de solution B en 5 minutes. La colonne a été lavée pendant 5 minutes avec le tampon B puis rééquilibrée avec le tampon A. La colonne et la pré-colonne ont été placées dans une étuve à une température de 45°C. La durée totale de l'analyse était de 180 min. Les analyses LC ont été acquises en mode ions positifs avec des balayages MS de  $m/z$  350 à 1 500 dans l'analyseur de masse Orbitrap à une résolution de 70 000 à  $m/z$  400. Le contrôle automatique du gain a été réglé sur 1 10<sup>e</sup>. Séquentiellement, des balayages MS/MS ont été acquis dans la cellule de collision à haute énergie pour les 15 ions les plus intenses détectés dans les spectres MS avec une résolution de 17 500 à  $m/z$  400 et une isolation de 1,6  $m/z$ .

## 2.4 - ANALYSE DES DONNÉES

Les données protéomiques ont été traitées avec Mascot (version 2.5.1, Matrix Science, Londres, Royaume-Uni) par rapport à la base de données NCBI (NCBIProt\_20210815) limitée aux mammifères (12 254 278 séquences) et aux oiseaux (4 179 719 séquences). Les modifications ont été cherchées avec le mode « error tolerants » permettant la recherche de toutes les modifications du logiciel Mascot (1530 modifications), les coupures non spécifiques et les mutations des acides aminés. Le nombre de clivage manqués par l'enzyme est de trois, l'erreur de masse de 10 ppm pour la MS et 15 mmu pour la MS/MS. La carbamidométhylation de la cystéine a été définie comme modification fixe. L'oxydation de la méthionine et l'asparagine, la déamidation de la glutamine ont été sélectionnées comme modifications variables. Le seuil de significativité est à  $p < 0,01$ . La deuxième analyse bioinformatique a été réalisée avec PEAKS X (Bioinformatic solution INC., Waterloo, Canada) par rapport à une base de données maison contenant 1 765 séquences de collagène extraites de la base de données NCBI (All\_Collagen, téléchargée en 12-2021) restreinte aux mammifères et oiseaux. La tolérance de masse du précurseur a été fixée à 10 ppm et la tolérance de masse des ions fragments à 0,05 Da. Trois clivages manqués ont été autorisés. Les mêmes modifications post-traductionnelles (PTMs) ci-dessus ont été autorisées plus l'hydroxylation des acides aminés (RYFPNKD) en tant que modifications variables. Cinq PTM variables ont été autorisés par peptides. PEAKS PTM et SPIDER fonctionnaient avec les mêmes paramètres. Les résultats ont été filtrés en utilisant les critères suivants: score protéique  $-10\log P \geq 20$ , taux

de fausses découvertes de peptides (FDR) de 1 %, PTM avec Ascore = 20, intensité des ions de mutation = 5 % et Denovo ALC  $\geq 50$  %. Les peptides avec des substitutions d'acides aminés ont été filtrés avec une intensité minimale fixée à 1E7. Les peptides identifiés sur le collagène I alpha 1 et I alpha 2 ont été alignés sur la séquence protéique non redondante NCBI (toutes les traductions GenBank CDS non redondantes + PDB + SwissProt + PIR + PRF à l'exclusion des échantillons environnementaux des projets WGS contenant 308 570 119 séquences) pour trouver une similitude avec BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). L'alignement des paramètres de notation a utilisé la matrice BLOSUM62. Les peptides spécifiques ont été validés avec un score de 100 % d'identité et une couverture complète des requêtes. Les spectres de fragmentations pour des peptides spécifiques ont été examinés manuellement pour la qualité. Les peptides identifiés ont été soumis à Unipept (version 4.0, <https://unipept.ugent.be/>) pour une analyse taxonomique (Mesuere *et al.*, 2015). L'appariement peptide-protéine est basé sur les données de la base de données UniProtKB, tandis que les taxons proviennent de la taxonomie NCBI. Les données de protéomique de spectrométrie de masse sont déposées sur le site ProteomeXchange (<http://proteomecentral.proteomexchange.org>) avec l'identifiant du jeu de données PXD039657 (Vizcaino *et al.*, 2012; Deutsch *et al.*, 2023) (supplément d'information 4).

## 3 - RÉSULTATS

### 3.1 - IDENTIFICATION DES PROTÉINES

Les analyses protéomiques par LC-MS/MS ont permis d'identifier le COL1A1 (collagène 1 alpha 1) et COL1A2 (collagène 1 alpha 2) dans tous les ossements. Les résultats permettent d'identifier en moyenne 54 % +/- 3 % de la séquence du COL1A1 et 57% +/- 5% du COL1A2. Le haut pourcentage de couverture montre que le collagène présent dans les ossements est bien conservé. Le nombre de peptides identifiés est en moyenne de 525 +/- 137 pour le COL1A1 et 382 +/- 101 pour le COL1A2. Le nombre de peptides du COL1A1 et COL1A2 varie entre les échantillons car la conservation de la matière organique varie en fonction des conditions environnementales. La prévalence du collagène de type I dans ces échantillons est cohérente avec le tissu osseux, où le collagène I représente 90 % de la protéine osseuse totale (Shoulders & Raines, 2009). Cette information permet d'indiquer que l'analyse protéomique a identifié les protéines des ossements paléontologiques.

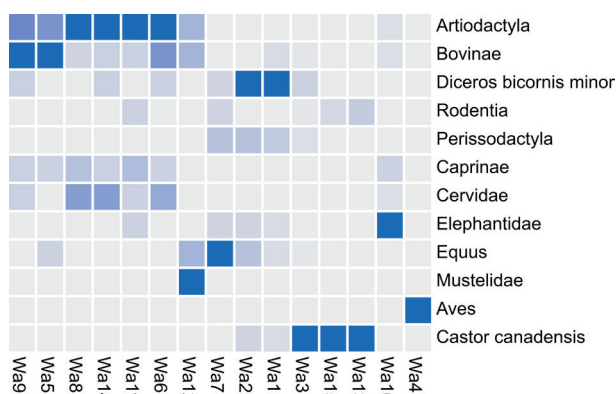
Le tableau 2 montre les résultats des identifications par LC-MS/MS sur les échantillons. Le tableau contient le meilleur résultat d'identification pour le COL1A1 et COL1A2 avec son numéro d'accèsion NCBI et l'espèce identifiée. Dans le tableau 2 est indiqué, la quantité de peptide après digestion en µg, le nombre de peptides au-dessus du seuil d'identité ( $p < 0,01$ ), le nombre de peptides au-dessus du seuil d'homologie ( $p < 0,01$ ),

le nombre de spectres interrogés, le score Mascot, le nombre de peptides identifiées, le nombre de séquences distinctes significatives, le nombre de séquences uniques significatives, le pourcentage de protéines de couverture sont indiqués pour chaque collagène.

Les résultats indiquent une très bonne conservation de la matière organique au vue du nombre de peptides identifiés et de la couverture de séquence du collagène. En plus du collagène de type 1, des protéines comme la kératine, dermicine, honerine sont présentes en quantité plus faible, ce qui indique une contamination via les manipulations des ossements lors de l'excavation et de leur observation. D'autres protéines collagéniques sont présentes comme le collagène 2 alpha 1, collagène 3 alpha 1, collagène 5 alpha 2 et des protéines non collagéniques comme biglycan, prothrombin, desmoplakin.

### 3.2 - IDENTIFICATION DU GENRE TAXONOMIQUE DES OSSEMENTS

Pour permettre d'attribuer des taxonomies à l'aide des peptides identifiés en LC-MS/MS, Le logiciel Unipept a été utilisé. L'ensemble des peptides identifiés par analyse protéomique a été analysé avec ce logiciel pour obtenir une classification taxonomique. Les résultats (figure 3) montrent la spécificité des peptides permettant une classification en fonction de la classe, l'ordre, la famille, le genre ou l'espèce. Les taxons ont été choisis grâce aux données d'identification de Mascot et anatomologique. L'échantillon Wa-4 a des peptides spécifiques des oiseaux, W-12 des Mustelidae, Wa-7 de *Equus*, Wa-10 de *Elephantidae*. W-3, 15, et 13 ont des peptides spécifiques pour le castor et Wa-1 et Wa-2 pour *Diceros*. Les échantillons Wa-5 et Wa-9 ont des peptides spécifiques pour le *Bovinae* et les échantillons Wa-6, 8, 14 et 11 ont des peptides spécifiques pour les *Cervidae*. Cette première analyse permet montrer l'attribution des peptides à des séquences protéiques avec une taxonomie associée. Les peptides permettant d'attribuer les taxonomies pour chaque échantillon sont dans le supplément d'information 1 (Hérisson *et al.*, 2023b).



**Fig 3: Carte de chaleur des peptides des échantillons de Waziers classés par taxonomie dans Unipept.**

Normalisation des données par colonnes. L'intensité de la couleur bleue indique le nombre de peptides associés à une catégorie donnée pour un groupe particulier.

Fig 3: Heat map of peptides from Waziers samples classified by taxonomy in Unipept. Normalization of data by columns. The intensity of the blue color indicates the number of peptides associated with a given category for a particular group.

La protéine ayant le meilleur score est le collagène de type I identifié lors de l'analyse bioinformatique avec le logiciel Mascot et celui-ci permet d'identifier la famille et le genre. L'identification des genres taxonomiques est réalisée sur la protéine avec le meilleur score et le plus grand nombre de peptides identifiés sur la séquence. Les peptides des COL1A1, COL1A2 et COL1A3 des familles et genres identifiés sont récupérés puis seuls les peptides dit uniques par le moteur de recherche Mascot sont sélectionnés. Une analyse avec le logiciel BlastP des peptides permet de comparer les peptides identifiés dans les ossements aux séquences présentes dans la base de données NCBI (National Center for Biotechnology Information) contenant 462 136 597 séquences. Cette analyse permet d'identifier les peptides spécifiques de la famille, genre et espèce. Les familles, genres et espèces identifiés par protéomique ont été mis en corrélation avec les données d'identification du paléontologue et les espèces présentes sur le site de Waziers. Ceci a permis d'affiner de ne garder que les peptides spécifiques des espèces présentes sur le site. Les résultats des analyses Blastp sont présentés dans le tableau 3. Les spectres MS/MS des peptides spécifiques sont dans le supplément d'information 3, figures S1 à S13 (Hérisson *et al.*, 2023b).

Les résultats de protéomique et les identifications anatomiques réalisées par le paléontologue corrélent dans 13 cas sur 15. L'analyse protéomique permet de lever l'incertitude sur l'identification paléontologique pour les ossements Wa-4, 6, 9, 10 et 14. L'identification LC-MS/MS des échantillons Wa-6 et Wa-14 ne correspondent pas à celles proposées avec doute par le paléontologue.

Des peptides spécifiques ont été trouvés pour les protéines COL1A1, COL1A2 et COL1A3. Les peptides ont permis de valider la famille ou le genre. Concernant les échantillons Wa-6, Wa-8, Wa-11 et Wa-14, un peptide spécifique de la famille des *Cervidae* a été identifié sur le COL3A1 sans pouvoir discriminer le genre *Muntiacus* ou *Cervus*.

Un peptide spécifique de la famille des *Rhinocerotidae* a été identifiés échantillons Wa-1 et Wa-2 sur le COL1A1 sans pouvoir discriminer le genre *Ceratotherium* ou *Diceros*.

Deux peptides spécifiques ont été identifiés sur le COL1A2 et un seul sur le COL1A1 pour l'espèce *Castor canadensis* pour les échantillons Wa-3, Wa-13 et Wa-15.

Pour le genre *Bos*, 2 peptides spécifiques ont été identifiés sur le COL1A2 pour les échantillons Wa-5, Wa-9 sans pouvoir préciser l'espèce.

Pour l'échantillon Wa-7, le genre *Equus* a été validé à l'aide d'un peptide spécifique sur le COL1A1 et 2 peptides sur le COL1A2 mais l'espèce n'a pas pu être déterminée.

La famille *Elephantidae* a été validée pour l'échantillon Wa-10 avec 1 peptide spécifique sur le COL1A1 et COL1A2.

L'échantillon Wa-12 a été identifié comme étant du genre *Enhydra* avec 1 peptide spécifique sur le COL1A2.

L'échantillon Wa-4 a été identifié comme étant de la famille des *Anatidae* avec 1 peptide spécifique sur le COL1A2.



* Identification incertaine	Identification LC-MS/MS		Résultats pBlast	COL1A1	COL1A2	Famille / Genre / Espèce	Résultats pBlast	COL3A1	Famille / Genre / Espèce	Résultats pBlast
	Identification paléontologique	Nom de l'échantillon								
Wa-1	Rhinocéros	Rhinocerotidae	Ceratotherium	GETGPAGRPGEAGPPGPPGSGEK		Rhinocerotidae	<i>Diceros bicornis minor, Ceratotherium simum simum</i>			
Wa-2	Rhinocéros	Rhinocerotidae	Ceratotherium	GETGPAGRPGEAGPPGPPGSGEK		Rhinocerotidae	<i>Diceros bicornis minor, Ceratotherium simum simum</i>			
Wa-3	Castor	Castoridae	Castor	GPAGPQGRGDKGETGEQGERGMK	GHNGPDLKGGPAGVYKGGGAGENGTPGOSGAR	Castor	<i>Castor canadensis</i>			<i>Castor canadensis</i>
Wa-4	Cygne*	Aves	Cygnus		AGNRGDPGVGHVGPAGAFQPR	Anatidae	<i>Anser cygnoides, Cygnus olor, Cygnus arcanus, Oxyura jamaicensis</i>			
Wa-5	Aurochs	Bovidae	Bos		GETGLRGDGSFGR	Bos	<i>Bos mutans, Bos indicus, Bubalus bubalis, Odocoileus virginianus texanus</i>			
Wa-6	Aurochs*	Cervidae	Cervus		GYPGNAGPVGAAGAIPGQGPVGPVYK		<i>Bos mutans, Bos laurus, Bubalus bubalis</i>	GAPGAQPPGAPGPIGIAGVTGAR	Cervidae	<i>Muntiacus muntjak, Cervus elaphus, Cervus canadensis</i>
Wa-7	Cheval	Equidae	Equus	GDAGPPGAPGAPGPPGSGVGAIPGKGAR	GPSGEPKRGDKGHLAGLAGAR HGHREGPVPVGSVPVGAAYGPR	Equus	<i>Equus caballus</i>			
Wa-8	Chevreuil	Cervidae	Cervus				<i>Equus caballus</i>			
Wa-9	Aurochs*	Bovidae	Bos		GETGLRGDGSFGR GYPGNAGPVGAAGAIPGQGPVGPVYK	Bos	<i>Bos mutans, Bos indicus, Bubalus bubalis, Odocoileus virginianus texanus</i>			
Wa-10	Mammouth*	Elephantidae	Loxodonta	GIDAGAKGEPGVPVGIQGPAGGEGEK	GERGVVYAGTGTGFSFGPLPGER	Elephantidae	<i>Loxodonta africana, Elephas maximus indicus</i>	GAPGAQPPGAPGPIGIAGVTGAR	Cervidae	<i>Muntiacus muntjak, Cervus elaphus, Cervus canadensis</i>
Wa-11	Chevreuil	Cervidae	Cervus							
Wa-12	Loutre	Mustelidae	Enhydra	GPAGPQGRGDKGETGEQGERGMK	TGETGAHPPGPAAGEKGPSGEPGTGAPPTSGPQGLLGPGLGLPGRS GHRHNGPDLKGGPAGVYKGGGAGENGTPGOSGAR	Lutinae	<i>Enhydra lutris kenyoni, Lutra lutra</i>	GAPGAQPPGAPGPIGIAGVTGAR	Cervidae	<i>Muntiacus muntjak, Cervus elaphus, Cervus canadensis</i>
Wa-13	Castor	Castoridae	Castor	GPAGPQGRGDKGETGEQGERGMK	GHNGPDLKGGPAGVYKGGGAGENGTPGOSGAR	Castor	<i>Castor canadensis</i>			
Wa-14	Oiseau*	Cervidae	Cervus					GAPGAQPPGAPGPIGIAGVTGAR	Cervidae	<i>Muntiacus muntjak, Cervus elaphus, Cervus canadensis</i>
Wa-15	Castor	Castoridae	Castor	GPAGPQGRGDKGETGEQGERGMK	GHNGPDLKGGPAGVYKGGGAGENGTPGOSGAR	Castor	<i>Castor canadensis</i>			

Tab. 3 : Identification des genres déterminés par l'analyse protéomique.

Les peptides spécifiques pour chaque genre taxonomique sont présentés pour les trois groupes de collagène (COL1A1, COL1A2 et COL3A1).

Tab. 3: Identification of genera determined by proteomic analysis. Specific peptides for the three collagen groups (COL1A1, COL1A2, and COL3A1).

Dans le cadre de l'analyse de l'échantillon Wa-10, le logiciel PEAKS a été utilisé pour affiner l'identification en utilisant une base de données contenant les séquences du COL1A1 et COL1A2 de la base de données NCBI et d'espèces éteintes provenant de publications. Par exemple, la séquence du collagène de *Mammuthus sp.* et celle de *Elephas* est présente dans des publications (Buckley *et al.*, 2011, 2019b; Buckley, 2015; Welker *et al.*, 2015). Le logiciel PEAKS identifie la séquence du genre *Elephas* avec 94% de couverture de séquence et 241 peptides pour l'échantillon Wa-10. *Elephas maximus* est le taxon le plus proche phylogénétiquement de *Mammuthus sp.* (Rogaev *et al.*, 2006). Les séquences de COL1A1 et COL1A2 de *Mammuthus sp.* provenant de la publication de Welker *et al.*, 2015 sont identifiées (séquences des COL1A1 et COL1A2 concaténées 224 peptides et 84 % de couverture de séquence) et provenant de la publication de Buckley *et al.* (2011) (COL1A1 57 peptides et 94 % de couverture de séquence, COL1A2 43 peptides et 95 % de couverture de séquence) mais avec de moins bons résultats car les séquences sont partielles.

Pour les échantillons Wa-1 et Wa-2 identifiés en tant que *Rhinocerotidae* en protéomique et rhinocéros de prairie (*Stephanorhinus hemitoechus*) par les paléontologues, une comparaison avec les séquences de COL1A1 et COL1A2 du rhinocéros laineux a été effectuée (Buckley, 2015; Welker *et al.*, 2017). Ces séquences ont été ajoutées au logiciel PEAKS et le logiciel identifie les séquences de collagène de type de *Dicerorhinus sumatrensis* avec le plus grand nombre de peptides et la meilleure couverture de séquence (COL1A1 138 peptides et 69 % de couverture de séquence, COL1A2 133 peptides et 71 % de couverture de séquence). *Dicerorhinus sumatrensis* est le taxon le plus proche phylogénétiquement du genre *Stephanorhinus* (Welker *et al.*, 2017). Les séquences de COL1A1 et COL1A2 de *Stephanorhinus sp.* (Welker *et al.*, 2017) sont identifiées mais avec de moins bons résultats car les séquences sont partielles comme pour l'échantillon Wa-10 (COL1A1 116 peptides et 96 % de couverture de séquence, COL1A2 102 peptides et 90 % de couverture de séquence).

### 3.3 - IDENTIFICATION DES MODIFICATIONS POST TRANSCRIPTIONNELLES

Nous avons détecté une variété de PTMs *in vivo* et diagenétiques dans ces échantillons. L'analyse des modifications post-traductionnelles telles que la déamidation de la glutamine renseigne sur l'état de conservation des os. Le peptide du COL1A1 avec la séquence GVQPPGPAGPR a été utilisé avec et sans déamidation sur l'arginine (spectres MS/MS supplément d'information 2, figures S1 à 2, Hérissou *et al.*, 2023b). Les échantillons plus anciens ou plus endommagés ont un niveau de déamidation plus élevé (Van Doorn *et al.*, 2012; Hill *et al.*, 2015; Schroeter & Cleland, 2016; Ramsøe *et al.*, 2020). Les résultats indiqués dans le tableau 4 montrent une moyenne de 64,3 % de déamidation sur la glutamine et un écart type de 12,2 %, ce qui indiquerait que les échantillons sont anciens.

Nous avons détecté différents PTMs d'origine biologique: méthylation, di-méthylation, alkylation, la carboxylation de la lysine et hydroxylation (spectres MS/MS supplément d'information 4, figure S3 à S9, Hérissou *et al.*, 2023b). La majorité des PTMs *in vivo* identifiées est l'hydroxylation de la proline. L'hydroxylation et la méthylation ont déjà été identifiées dans des études paléoprotéomiques du collagène osseux (Cappellini *et al.*, 2012; Cleland *et al.*, 2015; Welker *et al.*, 2015). Ces échantillons confortent donc la présence de ces modifications dans des échantillons fossiles.

## 4 - DISCUSSION

### 4.1 - MÉTHODOLOGIE ET IDENTIFICATION DES PROTÉINES

Il est primordial de développer des méthodes qui permettent d'obtenir autant d'informations que possible sur les protéines présentes dans les échantillons osseux à partir d'une quantité minimale d'os (Schroeter *et al.*, 2022). Dans nos analyses, nous avons utilisé 5 mg d'ossements pour réaliser les analyses. Pour la digestion, nous utilisons la méthode eFASP (Enhanced Filter Aided Sample Preparation) qui surpasse la digestion en solution ou FASP. Les analyses LC-MS/MS d'os paléontologiques digérés par eFASP optimisé offrent un nombre plus élevé de peptides identifiés, un pourcentage de couverture plus élevé et l'identification fiable des PTM car très peu de fausses modifications sont induites par la digestion eFASP (Erde *et al.*, 2014). L'utilisation de filtre Amicon® avec un seuil de coupure de 10kDa permet d'éliminer les sels, détergents utilisés lors de la préparation des échantillons. Lors du vieillissement des ossements, les protéines sont dégradées et des peptides sont générés permettant d'obtenir des informations sur la conservation (Ntasi *et al.*, 2022). Les peptides générés lors la dégradation des protéines dans les ossements sont éliminés lors des lavages dans notre méthode. Les interrogations avec le moteur de recherche Mascot sont uniquement réalisées avec la trypsine comme enzyme et le mode semi-tryptique n'est pas utilisé.

La préservation osseuse et la présence de contaminants modernes sont deux difficultés majeures dans les études de paléoprotéomiques osseuses. Dans notre étude, les kératines, l'honérine humaine et la dermicine, très souvent décrites comme des contaminants dans les analyses protéomiques des os, ont été détectées à l'état de traces (Wadsworth & Buckley, 2014; Sawafuji *et al.*, 2017). Afin de confirmer la présence de protéines anciennes, plusieurs contrôles ont été réalisés au niveau paléontologique et protéomique. Les ossements ont été prélevés dans un niveau géologique bien caractérisé correspondant à la fin du Saalien et à l'Eemien (entre 130 000 et 110 000 ans) et ces restes osseux ont été déterminés de manière traditionnelle par anatomie comparée avec les espèces actuelles analogues ou proches. Le prélèvement pour les analyses protéomique a été réalisé sur ces

Nom de l'échantillon	Identification paléontologue	Identification LC-MS/MS		Pourcentage de déamidation	Intensité peptide non déamidé	Intensité peptide déamidé
		Famille	Genre			
Wa-1	Rhinocéros	Rhinocerotidae	Ceratotherium	61,2	2,00 10 <sup>9</sup>	3,15 10 <sup>9</sup>
Wa-2	Rhinocéros	Rhinocerotidae	Ceratotherium	48,5	2,82 10 <sup>9</sup>	2,66 10 <sup>9</sup>
Wa-3	Castor	Castoridae	Castor	50,3	2,33 10 <sup>9</sup>	2,36 10 <sup>9</sup>
Wa-4	Cygne*	Anatidae	Cygnus	83,6	4,83 10 <sup>8</sup>	2,47 10 <sup>9</sup>
Wa-5	Aurochs	Bovidae	Bos	47,1	6,36 10 <sup>8</sup>	5,66 10 <sup>8</sup>
Wa-6	Aurochs*	Cervidae	Cervus	64,1	2,80 10 <sup>9</sup>	5,00 10 <sup>9</sup>
Wa-7	Cheval	Equidae	Equus	63,0	8,22 10 <sup>8</sup>	1,40 10 <sup>9</sup>
Wa-8	Chevreuil	Cervidae	Cervus	84,8	1,20 10 <sup>8</sup>	6,67 10 <sup>8</sup>
Wa-9	Aurochs*	Bovidae	Bos	67,6	1,97 10 <sup>9</sup>	4,11 10 <sup>9</sup>
Wa-10	Mammouth*	Elephantidae	Loxodonta	73,9	3,75 10 <sup>8</sup>	1,06 10 <sup>9</sup>
Wa-11	Chevreuil	Cervidae	Cervus	63,1	2,55 10 <sup>9</sup>	4,36 10 <sup>9</sup>
Wa-12	Loutre	Mustelidae	Enhydra	78,8	2,90 10 <sup>8</sup>	1,08 10 <sup>9</sup>
Wa-13	Castor	Castoridae	Castor	69,0	1,74 10 <sup>9</sup>	3,88 10 <sup>9</sup>
Wa-14	Oiseau*	Cervidae	Cervus	56,4	5,35 10 <sup>8</sup>	6,92 10 <sup>8</sup>
Wa-15	Castor	Castoridae	Castor	52,8	2,55 10 <sup>9</sup>	2,85 10 <sup>9</sup>

Tab. 4: Pourcentage de déamidation calculé pour chaque échantillon.

Tab. 4: Percentage of deamidation calculated for each sample.

ossements préalablement nettoyés, avec des gants, sous une sorbonne. L'article de Hendy *et al.* (2018) explique les différentes étapes et vérifications lors des analyses en paléoprotéomique. Pour vérifier la présence de protéines anciennes, la valeur de déamidation sur le COL1A1 a été mesurée. Celle-ci varie entre 50 et 74 % ce qui permet d'indiquer qu'il s'agit de protéine ancienne (Schroeter *et al.*, 2022 ; Warinner *et al.*, 2022).

Les deux chaînes de collagène de type I, représentant 90 % de la matière organique osseuse, ont été identifiées avec jusqu'à 60 % de couverture sur les deux chaînes. Ce résultat montre une préservation osseuse exceptionnelle. L'analyse protéomique basée sur le protocole eFASP sans extraction de protéines après déminéralisation à partir de seulement 5 mg de matière première permet d'identifier un grand nombre de peptide et d'obtenir une grande couverture de séquence. On peut souligner que des matières premières allant de 30 mg à plusieurs centaines de mg sont couramment décrites dans les études paléoprotéomiques (Cappellini *et al.*, 2011 ; Buckley *et al.*, 2017b, c ; Welker *et al.*, 2017). Les progrès de l'instrumentation en spectrométrie de masse permettent de diminuer les quantités d'ossements jusque 5 mg comme le montrent les publications actuelles (Ruether *et al.*, 2021 ; Bray *et al.*, 2022). L'analyse en mode «Data Dependent Acquisition» est utilisée dans notre article. Cette méthode consiste à acquérir un spectre MS et les ions majoritaires sont successivement sélectionnés et fragmentés en MS/MS. Cette méthode peut être remplacée par l'analyse «Data Independent Acquisition». Cette méthode d'acquisition consiste à sélectionner les ions d'une plage de masse et fragmenter tous les ions. Cette méthode permet de réduire le temps des analyses et d'utiliser des bibliothèques de spectre comme dans l'article de Ruether *et al.* (2021). Il faut noter que les nouveaux spectromètres de masses pourraient permettre de diminuer les quantités, d'identifier davantage de peptides et de diminuer les temps d'analyses.

D'autres collagènes ont été identifiés comme COL2A1, un collagène fibrillaire trouvé dans le cartilage, COL3A1,

un composant structurel majeur dans les organes creux, et COL5A2, un composant mineur du tissu conjonctif (Shoulders & Raines, 2009). Welker et ses collègues ont montré la présence d'autres types de collagène, comme COL10A1 et COL27A1, dans des spécimens osseux de la Grotte du Renne (Welker *et al.*, 2016). Des protéines non collagéniques (NCP) ont été détectées comme biglycan, prothrombin, desmoplakin et ont été décrites dans plusieurs études sur des os anciens (Cappellini *et al.*, 2011 ; Wadsworth & Buckley, 2014 ; Sawafuji *et al.*, 2017).

#### 4.2 - IDENTIFICATION DU GENRE

Les identifications ont été réalisées en utilisant la base de données NCBI. Cette base de données contient les séquences de collagène des espèces modernes dont le génome a été annoté. Il existe très peu de génomes annotés d'espèce éteinte comme celui de *Mammuth americanum* (Rohland *et al.*, 2010). Les identifications des espèces données correspondent aux homologues actuels les plus proches. L'identification des espèces a été réalisée en identifiant des peptides spécifiques ce qui signifie que la séquence du collagène est unique parmi toutes les séquences de protéines de NCBI (224 211 842 séquences de protéine provenant de 117 030 organismes).

L'analyse avec MASCOT permet de faire une identification de la famille, du genre en utilisant le collagène de type I avec une limitation qui est les séquences protéiques présentes dans la base de données. Dans notre article, l'identification des ossements est corrélée avec les informations morphologiques proposées par le paléontologue. Cette corrélation permet d'affiner les identifications en protéomique et de valider les espèces identifiées. Lorsque les analyses protéomiques sont réalisées sur des fragments d'ossements, les informations morphologiques ne sont pas disponibles. Dans ce cas, une analyse avec la méthode ZooMS dans un premier temps permet d'identifier rapidement la famille

à l'aide de peptides marqueurs. Une analyse LC-MS/MS peut ensuite être réalisée pour obtenir des informations plus précises comme les protéines, les séquences des protéines, les PTMs. Dans les échantillons d'ossements d'autres protéines sont présentes et pourraient permettre l'identification de la famille, du genre, de l'espèce. (Schroeter *et al.*, 2022; Warinner *et al.*, 2022)

L'utilisation de PEAKS et une base de données contenant des séquences de publications permet d'affiner les identifications. Il est important de noter que les séquences des publications ne sont pas complètes et des trous dans les séquences protéiques sont présents. Lors des analyses protéomiques, il est primordial de travailler avec des bases de données contenant le maximum de séquences de collagène provenant des bases de données NCBI, Swissprot et des publications. Par ailleurs, cette étude montre également l'importance des données protéomiques avec les études paléontologiques et de travailler sur les séquences protéiques des espèces les plus proches.

L'échantillon Wa-3 (et Wa-13) a été identifié sur la base de données NCBI en tant que *Castor canadensis* avec une couverture de séquence de plus de 50 % ce qui est plus élevé que l'étude réalisée par Timothy P. Cleland sur *Castoroides ohioensis* âgé de  $10\,150 \pm 50$  ans BP (Cleland *et al.*, 2016). Ceci indique que la méthode utilisée est plus sensible et que les ossements sont très bien conservés.

L'analyse LC-MS/MS de l'échantillon Wa-6 ne donne pas une identification en corrélation avec celle réalisée par le paléontologue. Cet os a été identifié en tant que *Cervus* en protéomique et en tant que bovidé, probablement l'aurochs, par le paléontologue. Cette différence est due à une identification incertaine liée à l'absence d'éléments diagnostiques permettant une détermination sur des critères ostéomorphologiques. La protéomique permet ainsi d'aider les paléontologues dans leur démarche d'identification des taxons.

### 4.3 - IDENTIFICATION DES PTMS

Dans cette étude, le pourcentage de déamidation varie entre 50 et 74 % sur le COL1A1 dans tous les échantillons, ce qui correspond au pourcentage de déamidation décrit dans la littérature (Welker *et al.*, 2017). Il a été montré que la déamidation de glutamine est plus stable que la déamidation de l'asparagine et que la demi-vie de la déamidation de l'asparagine sur l'échelle de temps physiologique est plus courte que la déamidation de la glutamine (Robinson *et al.*, 2004; Li *et al.*, 2010). De nombreuses études sur les os anciens ont étudié le taux de déamidation ce qui permet une comparaison. Des pourcentages de déamidation de la glutamine et de l'asparagine d'environ 50 % et 60 % respectivement ont été trouvés pour des os de 40 000 et 50 000 ans BP de Kleine Feldhofer Grotte (Allemagne) (Lanigan *et al.*, 2020). L'analyse de la déamidation de l'assemblage de 475 ossements de Quinçay, datant du Châtelperronien (Pléistocène supérieur, 44 500-36 000 ans BP) et de l'Holocène ancien a montré un taux moyen

de déamidation de 60 %. Sur des spécimens osseux du Chibanien du rhinocéros du genre *Stephanorhinus* (337 000-300 000 ans BP), les déamidations de la glutamine et de l'asparagine atteignent près de 100 % (Welker *et al.*, 2017). Des recherches récentes montrent un taux de déamidation variable au sein d'un même assemblage osseux sur 45 000-50 000 ans BP jusqu'à 305 000-284 000 ans BP (Brown *et al.*, 2021). Ces études mettent en évidence la difficulté d'utiliser le taux de déamidation pour dater relativement les échantillons. Les différences de taux de déamidation entre les publications pour les os de la même période peuvent s'expliquer par les différentes procédures d'extraction des protéines pour la déminéralisation et/ou des protéines uniques qui induisent une déamidation variable (Ren *et al.*, 2009; Hao *et al.*, 2011; Van Doorn *et al.*, 2012; Wilson *et al.*, 2012; Simpson *et al.*, 2016). Le contexte sédimentaire et les conditions environnementales peuvent également influencer le pourcentage de déamidation (Schroeter & Cleland, 2016). La préservation des protéines osseuses peut être déduite à la fois du pourcentage en poids de collagène extrait des os et du taux de déamidation qui a été rapporté comme marqueur de dégradation dans les os anciens (Van Doorn *et al.*, 2012).

La détermination de l'endogénéité des protéines fossiles et des PTM biologiques est essentielle pour évaluer leur fonction biologique. Les PTMs identifiées dans cet article, ont pu être observées dans l'article de Timothy Cleland *et al.* (2015).

De nombreuses publications identifient des prolines hydroxylées sur les peptides dans des échantillons anciens. Cette modification est impliquée dans la stabilisation de la triple hélice du collagène (Kotch *et al.*, 2008). Nos analyses ne détectent pas de modification comme l'oxydation des tryptophanes ou la formation d'acide aminoadipique à partir de lysine qui ont été décrits dans des ossements anciens (Cappellini *et al.*, 2011; Cleland *et al.*, 2015). Des oxydations des lysines ont été observées. Cette modification est observée lors du vieillissement ou lors de dégradation (Reeg & Grune, 2015). Des méthylations des lysines ont été détectées et sont identifiées chez les eucaryotes ce qui indique une endogénéité de ces modifications (Sprung *et al.*, 2008). De nombreuses modifications présentes sur les ossements sont dues à l'effet des conditions environnementales mais certaines comme l'hydroxyproline ou la méthylation sont des modifications *in vivo*. Ceci montre que la protéomique peut permettre d'identifier des modifications spécifiques des organismes et des modifications dues aux conditions de conservation.

## 5 - CONCLUSION

Notre étude a démontré la possibilité d'identifier des taxons d'os d'au moins 130 000 ans en utilisant une méthode de protéomique optimisée à partir d'échantillons de 5 mg. Un grand nombre de peptides a été identifié avec

des couvertures de séquences moyennes élevées allant jusqu'à 60 % pour le collagène I alpha 1. Malgré les bases de données restreintes pour plusieurs espèces, plusieurs taxons ont été identifiés. La protéomique a permis d'identifier sans équivoque le rang générique de tous les ossements et d'arriver à l'espèce pour la majorité des ossements. C'est une contribution majeure en particulier pour les études paléontologiques et archéozoologiques souvent aux prises avec l'identification impossible des espèces, de par l'absence de caractères diagnostiques sur les ossements, l'altération des os, la taille des ossements ainsi que les modifications de forme liées aux activités anthropiques (os brûlés, outils en os, supports artistiques, etc.).

## REMERCIEMENTS

Le spectromètre de masse a été financé par l'Université de Lille, le CNRS, la Région Hauts-de-France et le Fonds européen de développement régional. Les auteurs remercient la Mission pour l'Interdisciplinarité du CNRS pour le financement du projet ProtéoMamQuat (2017-2018) et la Région Hauts-de-France pour le financement du projet ProtéOsHdF (2021-2022). Le réseau IBISA est remercié pour le financement de la plateforme. Les auteurs remercient les responsables de la fouille du site de Waziers pour avoir permis d'utiliser les échantillons osseux de cette étude.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- AUGUSTE P., SÈVÈQUE N. & LOUGUET S., 2022 - La mésofaune et macrofaune mammalienne de Waziers (Nord), le Bas-Terroir (campagnes 2014-2015). Données paléontologiques. *Quaternaire*, **33** (4), 311-322. doi: 10.4000/quaternaire.17326.
- BRAY F., FLAMENT S., ABRAMS G., BONJEAN D., ROLANDO C., TOKARSKI C. & AUGUSTE P., 2022 - Extinct species identification from late middle Pleistocene and earlier Upper Pleistocene bone fragments and tools not recognizable from their osteomorphological study by an enhanced proteomics protocol. *Archaeometry*, doi: 10.1111/arc.12800.
- BROWN S., HIGHAM T., SLON V., PÄÄBO S., MEYER M., DOUKA K., BROCK F., COMESKEY D., PROCOPIO N., SHUNKOV M., DEREVIANKO A. & BUCKLEY M., 2016 - Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Scientific reports*, **6**, 23559. doi: 10.1038/srep23559.
- BROWN S., KOZLIKIN M., SHUNKOV M., DEREVIANKO A., HIGHAM T., DOUKA K. & RICHTER K.K., 2021 - Examining collagen preservation through glutamine deamidation at Denisova Cave. *Journal of Archaeological Science*, **133**, 105454. doi: 10.1016/j.jas.2021.105454.
- BUCKLEY M., 2015 - Ancient collagen reveals evolutionary history of the endemic South American 'ungulates'. *Proceedings of the Royal Society B: Biological Sciences*, **282** (1806), 20142671.
- BUCKLEY M., WALKER A., HO S.Y., YANG Y., SMITH C., ASHTON P., OATES J.T., CAPPELLINI E., KOON H. & PENKMAN K., 2008 - Weighing the mass spectrometric evidence for authentic *Tyrannosaurus rex* collagen. *Science (New York, NY)*, **319** (5859), 33. doi: 10.1126/science.1147046.
- BUCKLEY M., COLLINS M., THOMAS-OATES J. & WILSON J.C.W., 2009 - Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, **23** (23), 3843-3854. doi: 10.1002/rcm.4316.
- BUCKLEY M., LARKIN N. & COLLINS M., 2011 - Mammoth and Mastodon collagen sequences; survival and utility. *Geochimica et Cosmochimica Acta*, **75** (7), 2007-2016.
- BUCKLEY M., WARWOOD S., VAN DONGEN B., KITCHENER A.C. & MANNING P.L., 2017a - A fossil protein chimera; difficulties in discriminating dinosaur peptide sequences from modern cross-contamination. *Proceedings of the Royal Society B: Biological Sciences*, **284** (1855), 20170544. doi: 10.1098/rspb.2017.0544.
- BUCKLEY M., HARVEY V.L. & CHAMBERLAIN A.T., 2017b - Species identification and decay assessment of Late Pleistocene fragmentary vertebrate remains from Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting. *Boreas*, **46** (3), 402-411. doi: 10.1111/bor.12225.
- BUCKLEY M., COSGROVE R., GARVEY J. & PRIDEAUX G.J., 2017c - Identifying remains of extinct kangaroos in Late Pleistocene deposits using collagen fingerprinting. *Journal of Quaternary Science*, **32** (5), 653-660. doi: 10.1002/jqs.2964.
- BUCKLEY M., LAWLESS C. & RYBCZYNSKI N., 2019a - Collagen sequence analysis of fossil camels, *Camelops* and c.f. *Paracamelus*, from the Arctic and sub-Arctic of Plio-Pleistocene North America. *Journal of Proteomics*, **194**, 218-225. doi: 10.1016/j.jpro.2018.11.014.
- BUCKLEY M., RECABARREN O.P., LAWLESS C., GARCÍA N. & PINO M., 2019b - A molecular phylogeny of the extinct South American gomphothere through collagen sequence analysis. *Quaternary Science Reviews*, **224**, 105882. doi: 10.1016/j.quascirev.2019.105882.
- CAPPELLINI E., JENSEN L.J., SZKLARCZYK D., GINOLHAC A., DA FONSECA R.A., STAFFORD JR T.W., HOLEN S.R., COLLINS M.J., ORLANDO L. & WILLERSLEV E., 2011 - Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *Journal of Proteome Research*, **11** (2), 917-926. doi: 10.1021/pr200721u.
- CAPPELLINI E., JENSEN L.J., SZKLARCZYK D., GINOLHAC A., DA FONSECA R.A., STAFFORD JR T.W., HOLEN S.R., COLLINS M.J., ORLANDO L. & WILLERSLEV E., 2012 - Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *Journal of Proteome Research*, **11** (2), 917-926.
- CAPPELLINI E., WELKER F., PANDOLFI L., RAMOS-MADRIGAL J., SAMODOVA D., RÜTHER P.L., FOTAKIS A.K., LYON D., MORENO-MAYAR J.V., BUKHSIANIDZE M., JERSIE-CHRISTENSEN R.R., MACKIE M., GINOLHAC A., FERRING R., TAPPEN M., PALKOPOULOU E., DICKINSO M.R., STAFFORD JR T.W., CHAN Y.L., GÖTHERSTRÖM A., NATHAN S.K.S.S., HEINTZMAN P.D., KAPP J.D., KIRILLOVA I.V., MOODLEY Y., AGUSTI J., KAHLKE R.-D., KILADZE G., MARTÍNEZ-NAVARRO B., LIU S., VELASCO M.S., SINDING M.-H.S., KELSTRUP C.D., ALLENTOF M.E., ORLANDO L., PENKMAN K., SHAPIRO B., ROOK L., DALÉN L., GILBERT M.T.P., OLSEN J.V., LORDKIPANIDZE D. & WILLERSLEV E., 2019 - Early Pleistocene enamel proteome from *Dmanisi* resolves *Stephanorhinus* phylogeny. *Nature*, **574** (7776), 103-107. doi: 10.1038/s41586-019-1555-y.
- CLELAND T.P., SCHROETER E.R. & SCHWEITZER M.H., 2015 - Biologically and diagenetically derived peptide modifications in moa collagens. *Proceedings of the Royal Society B: Biological Sciences*, **282** (1808), 20150015. doi: 10.1098/rspb.2015.0015.
- CLELAND T.P., SCHROETER E.R., FERANEC R.S. & VASH-ISHTH D., 2016 - Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proceedings Biological sciences*, **283** (1832), 20160593. doi: 10.1098/rspb.2016.0593.
- DESCHODT L., ANTOINE P., JAMET G., PETITE Y., HÉRISSON D., COUTARD S. & LIMONDIN-LOZOUET N., 2022 - Waziers «Bas-Terroir» (plaine de la Scarpe, nord de la France): synthèse lithostratigraphique des observations effectuées de 2011 à 2015. *Quaternaire*, **33** (4), 247-272.
- DEUTSCH E.W., BANDEIRA N., PEREZ-RIVEROL Y., SHARMA V., CARVER J.J., MENDOZA L., KUNDU D.J., WANG S., BANDLA C. & KAMATCHINATHAN S., 2023 - The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Research*, **51** (D1), D1539-D1548. doi: 10.1093/nar/gkac1040.
- ERDE J., OGORZALEK LOO R.R. & LOO J.A., 2014 - Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *Journal of Proteome Research*, **13** (4), 1885-1895. doi: 10.1021/pr4010019.
- HAO P., REN Y., ALPERT A.J. & SZE S.K., 2011 - Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Molecular & Cellular Proteomics*, **10** (10), doi: 10.1074/mcp.O111.009381.
- HARVEY V.L., EGERTON V.M., CHAMBERLAIN A.T., MANNING P.L., SELLERS W.I. & BUCKLEY M., 2019a - Interpreting the historical terrestrial vertebrate biodiversity of Cayman Brac (Greater Antilles, Caribbean) through collagen fingerprinting. *The Holocene*, **29** (4), 531-542. doi: 10.1177/0959683618824793.
- HARVEY V.L., LEFEBVRE M.J., DEFRANCE S.D., TOFTGAARD C., DROSOU K., KITCHENER A.C. & BUCKLEY M., 2019b - Preserved collagen reveals species identity in archaeological

- marine turtle bones from Caribbean and Florida sites. *Royal Society open science*, **6** (10), 191137, doi: 10.1098/rsos.191137.
- HELLE S., BRAY F., PUTAUX J.-L., VERBEKE J., FLAMENT S., ROLANDO C., D'HULST C. & SZYDLOWSKI N., 2019 - Intra-sample heterogeneity of potato starch reveals fluctuation of starch-binding proteins according to granule morphology. *Plants*, **8** (9), 324, doi: 10.3390/plants8090324.
- HENDY J., WELKER F., DEMARCHI B., SPELLER C., WARINNER C. & COLLINS M.J., 2018 - A guide to ancient protein studies. *Nature ecology & evolution*, **2** (5), 791-799, doi: 10.1038/s41559-018-0510-x.
- HÉRISSON D., DESCHODT L., ANTOINE P., LOCHT J.-L., LACROIX S., SERGENT A., PETITE Y., VALLIN L., RORIVE S., SIMON F. & NOÛS C., 2022 - Waziers, le Bas-Terroir: historique de dix années de recherches archéologiques et géomorphologiques dans un marais pléistocène de la plaine de la Scarpe (2011-2021). *Quaternaire*, **33** (4), 225-246.
- HERISSON D., BRAY F., AUGUSTE P. MARECHAL C., 2023 - Jeux de données issus des recherches sur le gisement de Waziers la Bas Terroir (opérations 2013-2015) en lien avec les articles des volumes 33-4 (2022) et 34-1 (2023) de la revue *Quaternaire*. *Quaternaire*, **34** (1), 65-69.
- HILL R.C., WITHER M.J., NEMKOV T., BARRETT A., D'ALESSANDRO A., DZIECIATKOWSKA M. & HANSEN K.C., 2015 - Preserved Proteins from Extinct Bison latifrons Identified by Tandem Mass Spectrometry; Hydroxylysine Glycosides are a Common Feature of Ancient Collagen. *Molecular & Cellular Proteomics*, **14** (7), 1946-1958, doi: 10.1074/mcp.M114.047787.
- HORN I.R., KENENSY, PALMBLAD N.M., VANDERPLAS-DUI-VESTEIJN S.J., LANGEVELD B.W., MEIJER H.J.M., DALEBOUT H., MARISSSEN R.J., FISCHER A., FLORENS V.F.B., NIEMANN J., RIJSDIJK K.F., SCHULP A.S., LAROS J.F.J. & GRAVENDEEL B., 2019 - Palaeoproteomics of bird bones for taxonomic classification. *Zoological Journal of the Linnean Society*, **186** (3), 650-665. doi: 10.1093/zoolinnean/zlz012.
- KJØER K.H., WINTHER PEDERSEN M., DE SANCTIS B., DE CAHSAN B., KORNELIUSSEN T.S., MICHELSEN C.S., SAND K.K., JELAVIĆ S., RUTER A.H., SCHMIDT A.M.A., KJELDSEN K.K., TESAKOV A.S., SNOWBALL I., GOSSE J.C., ALSOS I.G., WANG Y., DOCKTER C., RASMUSSEN M., JØRGENSEN M.E., SKADHAUGE B., PROHASKA A., KRISTENSEN J.Å., BJERAGER M., ALLENTOF T.M.E., COISSAC E., ALSOS I.G., COISSAC E., ROUILLARD A., SIMAKOVA A., FERNANDEZ-GUERRA A., BOWLER C., MACIAS-FAURIA M., VINNER L., WELCH J.J., HIDY A.J., SIKORA M., COLLINS M.J., DURBIN R., LARSEN N.K., WILLER-SLEV E. & PHYLONORWAY C., 2022 - A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA, *Nature*, **612**. doi: 10.1038/s41586-022-05453-y
- KOTCH F.W., GUZEI I.A. & RAINES R.T., 2008 - Stabilization of the collagen triple helix by O-methylation of hydroxyproline residues. *Journal of the American Chemical Society*, **130** (10), 2952-2953, doi: 10.1021/ja800225k.
- LANIGAN L.T., MACKIE M., FEINE S., HUBLIN J.-J., SCHMITZ R.W., WILCKE A., COLLINS M.J., CAPPELLINI E., OLSEN J.V., TAUROZZI A.J. & WELKER F., 2020 - Multi-proteome analysis of Pleistocene bone proteomes. *Journal of Proteomics*, **228**, 103889, doi: 10.1016/j.jprot.2020.103889.
- LI X., LIN C. & O'CONNOR P.B., 2010 - Glutamine deamidation: differentiation of glutamic acid and  $\gamma$ -Glutamic Acid in peptides by electron capture dissociation. *Analytical Chemistry*, **82** (9), 3606-3615.
- MESUERE B., DEBYSER G., AERTS M., DEVREESE B., VANDAMME P. & DAWYNDT P., 2015 - The Unipept metaproteomics analysis pipeline. *Proteomics*, **15** (8), 1437-1442, doi: 10.1002/pmic.201400361.
- NTASI G., PALOMO I.R., MARINO G., PIAZ F.D., CAPPELLINI E., BIROLO L. & PETRONE P., 2022 - Molecular signatures written in bone proteins of 79 AD victims from Herculaneum and Pompeii. *Scientific reports*, **12** (1), 1-15, doi: 10.1038/s41598-022-12042-6.
- PRESSLEE S., SLATER G.J., PUJOS F., FORASIEPI A.M., FISCHER R., MOLLOY K., MACKIE M., OLSEN J.V., KRAMARZ A. & TAGLIORRETTI M., 2019 - Palaeoproteomics resolves sloth relationships. *Nature ecology & evolution*, **3** (7), 1121-1130, doi: doi.org/10.1038/s41559-019-0909-z.
- RAMSØE A., VAN HEEKEREN V., PONCE P., FISCHER R., BARNES I., SPELLER C. & COLLINS M.J., 2020 - DeamidATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *Journal of Archaeological Science*, **115**, 105080, doi: 10.1016/j.jas.2020.105080.
- REEG S. & GRUNE T., 2015 - Protein oxidation in aging: does it play a role in aging progression? *Antioxidants & redox signaling*, **23** (3), 239-255.
- REN D., PIPES G.D., LIU D., SHIH L.-Y., NICHOLS A.C., TREUHEIT M.J., BREMS D.N. & BONDARENKO P.V., 2009 - An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Analytical Biochemistry*, **392** (1), 12-21, doi: 10.1016/j.ab.2009.05.018.
- ROBINSON N., ROBINSON Z., ROBINSON B., ROBINSON A., ROBINSON J., ROBINSON M. & ROBINSON A.B., 2004 - Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *The Journal of Peptide Research*, **63** (5), 426-436.
- ROGAEV E.I., MOLIKA Y.K., MALYARCHUK B.A., KONDRASHOV F.A., DERENKO M.V., CHUMAKOV I. & GRIGORENKO A.P., 2006 - Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. *PLOS Biology*, **4** (3), e73.
- ROHLAND N., REICH D., MALLICK S., MEYER M., GREEN R.E., GEORGIADIS N.J., ROCA A.L. & HOFREITER M., 2010 - Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLOS Biology*, **8** (12), e1000564.
- RUETHER P.L., HUSIC I.M., BANGSGAARD P., GREGERSEN K.M., PANTMANN P., CARVALHO M., GODINHO R.M., FRIEDL L., CASCALHEIRA J. & JOERKOV M.L.S., 2021 - SPIN-Species by Proteome INvestigation. *bioRxiv*, doi: 10.1101/2021.02.23.432520.
- SAWAFUJI R., CAPPELLINI E., NAGAOKA T., FOTAKIS A.K., JERSIE-CHRISTENSEN R.R., OLSEN J.V., HIRATA K. & UEDA S., 2017 - Proteomic profiling of archaeological human bone. *Royal Society open science*, **4** (6), 161004, doi: 10.1098/rsos.161004.
- SCHROETER E.R. & CLELAND T.P., 2016 - Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Communications in Mass Spectrometry*, **30** (2), 251-255, doi: 10.1002/rcm.7445.
- SCHROETER E.R., DEHART C.J., CLELAND T.P., ZHENG W., THOMAS P.M., KELLEHER N.L., BERN M. & SCHWEITZER M.H., 2017 - Expansion for the *Brachylophosaurus canadensis* Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein. *Journal of Proteome Research*, **16** (2), 920-932, doi: 10.1021/acs.jproteome.6b00873.
- SCHROETER E.R., CLELAND T.P. & SCHWEITZER M.H., 2022 - Deep Time Paleoproteomics: Looking Forward. *Journal of Proteome Research*, **21** (1), 9-19, doi: 10.1021/acs.jproteome.1c00755.
- SHOULDERS M.D. & RAINES R.T., 2009 - Collagen structure and stability. *Annual Review of Biochemistry*, **78**, 929-958, doi: 10.1146/annurev.biochem.77.032207.120833.
- SIMPSON J., PENKMAN K.E.H., DEMARCHI B., KOON H., COLLINS M.J., THOMAS-OATES J.E., SHAPIRO B., STARK M. & WILSON J., 2016 - The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. *Journal of Archaeological Science*, **69**, 29-38, doi: 10.1016/j.jas.2016.02.002.
- SPRUNG R., CHEN Y., ZHANG K., CHENG D., ZHANG T., PENG J. & ZHAO Y., 2008 - Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. *Journal of Proteome Research*, **7** (3), 1001-1006.
- VAN DER VALK T., PEČNEROVÁ P., DÍEZ-DEL-MOLINO D., BERGSTRÖM A., OPPENHEIMER J., HARTMANN S., XENIKOUDAKIS G., THOMAS J.A., DEHASQUE M. & SAĞLIÇAN E., 2021 - Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*, **591** (7849), 265-269, doi: 10.1038/s41586-021-03224-9.
- VAN DOORN N.L., WILSON J., HOLLUND H., SORESSI M. & COLLINS M.J., 2012 - Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Communications in Mass Spectrometry*, **26** (19), 2319-2327.
- VIZCAÍNO J.A., CÔTÉ R.G., CSORDAS A., DIANES J.A., FABREGAT A., FOSTER J.M., GRISS J., ALPI E., BIRIM M. & CONTELL J., 2012 - The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Research*, **41** (D1), D1063-D1069, doi: 10.1093/nar/gks1262.
- WADSWORTH C. & BUCKLEY M., 2014 - Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Communications in Mass Spectrometry*, **28** (6), 605-615, doi: 10.1002/rcm.6821.
- WARINNER C., KORZOW RICHTER K. & COLLINS M.J., 2022 - Paleoproteomics. *Chemical Reviews*, **122** (16), 13401-13446, doi: 10.1021/acs.chemrev.1c00703.

- WELKER F., COLLINS M.J., THOMAS J.A., WADSLEY M., BRACE S., CAPPELLINI E., TURVEY S.T., REGUERO M., GELFO J.N. & KRAMARZ A., 2015** - Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature*, **522** (7554), 81, doi: 10.1038/nature14249.
- WELKER F., HAJDINJAK M., TALAMO S., JAOUEN K., DANNEMANN M., DAVID F., JULIEN M., MEYER M., KELSO J. & BARNES I., 2016** - Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings of the National Academy of Sciences*, **113** (40), 11162-11167, doi: 10.1073/pnas.1605834113.
- WELKER F., SMITH G.M., HUTSON J.M., KINDLER L., GARCIA-MORENO A., VILLALUENGA A., TURNER E. & GAUDZINSKI-WINDHEUSER S., 2017** - Middle Pleistocene protein sequences from the rhinoceros genus *Stephanorhinus* and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ*, **5**, 3033, doi: 10.7717/peerj.3033.
- WELKER F., RAMOS-MADRIGAL J., KUHILWILM M., LIAO W., GUTENBRUNNER P., DE MANUEL M., SAMODOVA D., MACKIE M., ALLENTOFT M.E. & BACON A.-M., 2019** - Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. *Nature*, **576** (7786), 262-265, doi: 10.1038/s41586-019-1728-8.
- WILSON J., VAN DOORN N.L. & COLLINS M.J., 2012** - Assessing the extent of bone degradation using glutamine deamidation in collagen. *Analytical Chemistry*, **84** (21), 9041-9048, doi: 10.1021/ac301333t.