Palaeoproteomics of bird bones for taxonomic classification

Horn, I.R.; Kenens, Y.; van der Plas-Duivesteijn, S.J.; Langeveld, B.W.; Meijer, H.J.M.; Dalebout, H.; Marissen, R.J.; Fischer, A.; Florens, F.B.V.; Niemann, J.; Rijsdijk, K.F.; Schulp, A.S.; Laros, J.F.J.; Gravendeel, B.

Palaeoproteomics of bird bones for taxonomic classification

IVO R. HORN[1,2*], YVO KENENS[1], N. MAGNUS PALMBLAD[3],
SUZANNE J. VAN DER PLAS-DUIVESTEIJN[3], BRAM W. LANGEVELD[4],
HANNEKE J. M. MEIJER[2,5], HANS DALEBOUT[3], ROB J. MARISSEN[3,●], ANJA FISCHER[6],
F. B. VINCENT FLORENS[7], JONAS NIEMANN[8], KENNETH F. RIJSDIJK[9],
ANNE S. SCHULP[2], JEROEN F. J. LAROS[10] and BARBARA GRAVENDEEL[1,2,11]

[1]*University of Applied Sciences Leiden, Faculty of Science and Technology, Zernikedreef 11, 2333 CK, Leiden, The Netherlands*
[2]*Naturalis Biodiversity Center, Endless Forms Group, Darwinweg 2, 2333 CR Leiden, The Netherlands*
[3]*Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands*
[4]*Natural History Museum Rotterdam, Museumpark, Rotterdam, The Netherlands*
[5]*University Museum, Department of Natural History, University of Bergen, Bergen, Norway*
[6]*University of Amsterdam, Faculty of Humanities, Amsterdam, The Netherlands*
[7]*Tropical Island Biodiversity, Ecology and Conservation Pole of Research, University of Mauritius, Réduit, Mauritius*
[8]*Natural History Museum of Denmark, Copenhagen, Denmark*
[9]*BIOMAC group, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Faculty of Natural Sciences, Science Park 904, Amsterdam, The Netherlands*
[10]*Leiden Genome Technology Center, Leiden, The Netherlands*
[11]*Institute of Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands*

We used proteomic profiling to taxonomically classify extinct, alongside extant bird species using mass spectrometry on ancient bone-derived collagen chains COL1A1 and COL1A2. Proteins of Holocene and Late Pleistocene-aged bones from dodo (*Raphus cucullatus)* and great auk (*Pinguinus impennis*), as well as bones from chicken (*Gallus gallus*), rock dove (*Columba livia*), zebra finch (*Taeniopygia guttata*) and peregrine falcon (*Falco peregrinus*), of various ages ranging from the present to 1455 years old were analysed. HCl and guandine-HCL-based protein extractions from fresh bone materials yielded up to 60% coverage of collagens COL1A1 and COL1A2, and extractions from ancient materials yielded up to 46% coverage of collagens COL1A1 and COL1A2. Data were retrieved from multiple peptide sequences obtained from different specimens and multiple extractions. Upon alignment, and in line with the latest evolutionary insights, protein data obtained from great auk grouped with data from a recently sequenced razorbill (*Alca torda*) genome. Similarly, protein data obtained from bones of dodo and modern rock dove grouped in a single clade. Lastly, protein data obtained from chicken bones, both from ancient and fresh materials, grouped as a separate, basal clade. Our proteomic analyses enabled taxonomic classification of all ancient bones, thereby complementing phylogenetics based on DNA.

ADDITIONAL KEYWORDS: ancient proteins – bird taxonomy – dodo – extant birds – great auk – palaeoproteomics – phylogeny.

INTRODUCTION

Biomolecules like DNA and proteins can be retrieved from fossils and complete genomes have been sequenced from fossil samples up to approximately 735 000 years

*Corresponding author. E-mail: ivo.horn@hsleiden.nl

old (Orlando *et al.*, 2013). However, DNA is not always found in ancient samples, as it is liable to decay over time with a half-life in bones of 512 years (Lindahl, 1993; Allentoft *et al.*, 2012). In situations where DNA cannot be retrieved, proteomic profiling using proteins from ancient bones may offer a rapid, efficient and complementary tool for identification. Although the taxonomic resolution offered by protein sequences might be lower because of degeneration of the genetic code, identification of ancient peptide sequences and classification of organisms based on proteomics is possible, and research has predominantly focused on collagens as being the most abundant proteins in ancient bones (Semal & Orban, 1995). Collagen proteins, COL1A1 and COL1A2, are very stable and, therefore, relatively resistant to decay due to their molecular organization, in which three proline-rich polypeptide chains assemble and the overall thermal stability is increased by hydroxylation of the prolines. For an extensive review on collagen biochemistry, see Shoulders & Raines (2009). Because collagens abound in many ancient skeletal remains, studies have focused on optimizing retrieval of specific collagen peptides that are stably preserved in bones (Buckley *et al.*, 2010). In addition, it has been demonstrated that asparagine and glutamine residues are liable to deamidation due to ageing. This process might be of potential use for molecular clock estimates of proteins. Studies on bone degradation and deamidation rates found that deamidation occurs more frequently under alkaline than acid conditions (Wilson *et al.*, 2012). The potential use of deamidation as a marker for antiquity was investigated by Schroeter & Cleland (2016). In the latter study, it was concluded that deamidation might be a good indicator for preservation conditions, but that it is possibly less suited for dating because of the influence of environmental factors such as pH, ionic strength and temperature (Robinson & Robinson, 2001; Hurtado & O'Connor, 2012).

Identification of collagen peptides has successfully been reported from very old and more recent bone material; see, for instance: Asara (2007) and Buckley *et al.* (2010). Cappellini *et al.* (2013) demonstrated the value of a combined genomics and proteomics approach by elucidating the correct identity of a nearly 300-year-old ethanol-preserved elephant foetus. The same group was able to retrieve peptides from 126 unique proteins extracted from a 43 290-year-old femur of a woolly mammoth [*Mammuthus primigenius* (Blumenbach, 1977); Cappellini *et al.*, 2012] and very recently used proteome sequences from 1.77 million years old enamel to investigate the phylogenetic relationships of the Eurasian Pleistocene Rhinocerotidae (Cappellini *et al.*, 2018). It was concluded that palaeoproteomics on enamel can be done on material dating back to the

Early Pleistocene. For a recent review on comparison of proteomic methods, guidelines and the growing number of studies in the field of palaeoproteomics, see Hendy *et al.* (2018).

Recently, taxonomy based on proteomics has been performed on bone specimens of extinct and extant species (Cleland *et al.*, 2015, 2016; Welker *et al.*, 2015; Welker *et al.*, 2016, 2017). These studies included investigations on taxa from very different classes, amongst others from birds. Cleland *et al.* (2015) have investigated the ancient bones from moa remains (Dinornitidae, species undetermined), with special interest in post-translational modifications for taxonomic purposes. It was shown that several post-translational modifications were biologically derived, whereas others were diagenetically derived, a finding that can be used for further studies on physiology, phylogeny and mechanisms leading to preservation or decay of proteins. In an extensive study by Welker *et al.* (2015), South American native ungulates were classified solely on protein sequences using proteomic analysis. The authors concluded that the resulting phylogenetic trees correlated well with mammalian phylogenies obtained using genomic methods. Consequently, and with the ever-ongoing refinements in instrumentation, the investigators foresee an important future role for proteomic methods in palaeontology. In a study by Welker *et al.* (2017), ZooArchaeology by Mass Spectrometry (ZooMS) analysis was performed on Middle and Late Pleistocene peptide sequences from rhinoceros (*Stephanorhinus* sp.) and peptide sequences from various extant species. The investigators were able to group the investigated species of rhinoceros in the same clades, as found in previous morphological studies. The protein degradation and proteome complexity were consistent with an endogenous origin of the identified proteins. A study performed on Late Pleistocene archaic hominins further strengthened the importance of palaeoproteomics studies (Welker *et al.*, 2016). Proteome data supported by mitochondrial DNA data identified hominin material found in France belonging to clades in the genus *Homo*. Proteomic profiling has also been used to identify closely related species in archaeological materials, like those found in Danish peat bogs (Brandt *et al.*, 2014). Based on peptide identification, it could be shown that animal material derived from sheep or goat were used for skin garments and costumes that were 2000 to 3000 years old. The oldest authenticated protein sequences from birds were retrieved from ostrich egg shells in a study by Demarchi *et al.* (2016). The protein sequences retrieved were 3.8 million years old, which is much older than the oldest retrieved DNA sequences.

In the current study, we focus on taxonomic classification of two extinct birds, dodo (*Raphus*

cucullatus Linnaeus, 1758) and great auk (*Pinguinus impennis* Linnaeus, 1758) using proteomics technology. During an interdisciplinary analysis in 2005 of the Mare aux Songes Lagerstätte on Mauritius, we discovered a 4000-year-old bone depository (Rijsdijk *et al.*, 2009; Meijer *et al.*, 2012; Rijsdijk *et al.*, 2016). Skeletal fragments from at least 17 individual dodos (*R. cucullatus*) were retrieved. Most likely, these birds died from miring events (Meijer *et al.*, 2012). Mare aux Songes is situated in a collapsed lava tunnel filled in by sands and clays, and on top of that a matrix of bones, wood and seeds, all submerged in brackish groundwater. The climate conditions are tropical and humid.

We used bone material from the dodo and from the great auk. The latter species, reviewed by Eckert (1963) and Fuller (1999), has been classified in an extensive morphological and molecular study as a member of the Alcidae (Alcids; Smith, 2011), whereas in earlier times it was misclassified as a species belonging to the Spheniscidae (penguins; Linnaeus, 1758). It was common in the Northern Hemisphere, especially at the coasts of the northern Atlantic Ocean and coasts of adjacent seas, like the North Sea and it became extinct in the late 19th century, mainly through overexploitation as food and fuel supplies for sailors (Fuller, 1999). In The Netherlands, one partial skeleton and several individual bones have been found near the coastal regions, either in beach or in inland regions, resulting in differential preservation conditions (Groot, 2005; Langeveld, 2015).

Using bottom-up proteomics on bone-derived collagen chains COL1A1 and COL1A2, we aimed to further investigate the potential of palaeoproteomics to taxonomically classify bones of extinct and extant bird species that lack sufficient diagnostic characters to place them in the correct family solely on morphology or where DNA is insufficiently preserved. We performed these studies partially in accordance with more recently published guidelines to ensure endogeneity (Hendy *et al.*, 2018). Extinct species included dodo and great auk, and extant species in the study were from various taxonomic clades of which the full genome has been sequenced and made accessible in NCBI GenBank and SwissProt databases. These included zebra finch (*Taeniopygia guttata* Vieillot, 1817), chicken (*Gallus gallus* Linnaeus, 1758), peregrine falcon (*Falco peregrinus* Tunstall, 1771), rock dove (*Columba livia* Gmelin, 1789) and brown anole lizard (*Anolis sagrei* Duméril & Bibron, 1837), which served as an outgroup for the phylogenetics analyses. The full genomic sequence of the closely related Caroline anole (*Anolis carolinensis* Voigt, 1832) was used as a reference, since these two *Anolis* species are genetically so close that they can interbreed (Stuart *et al.*, 2014).

MATERIAL AND METHODS

STUDY SITES AND SPECIMENS

Proteomics was performed on a set of bird bones derived from extinct and extant species. We included fresh samples of extant chicken (Phasianidae) alongside museum-preserved bones collected in 1950 from birds reared in The Netherlands and three ancient bones collected by Eugène Dubois between 1887 and 1895 in the Goea Djimbe cave on Java, Indonesia. Two of the ancient bones were radio-carbon dated at the Groningen AMS Facility (University of Groningen) to 1415–1366 calibrated years before present (BP) and 1455–1385 calibrated years BP, respectively (Meijer, unpublished). Two humeri and one femur of great auk (*Pinguinus impennis*) were sampled. The first was collected in the former Roman castellum at the port of Velsen, The Netherlands, and was dated to the first century AD based on its context (Van Wijngaarden-Bakker, 1978; Groot, 2005). The second one, radio-carbon dated to 7000–6890 calibrated years BP, was collected from the beach of the Zandmotor near The Hague (Langeveld, 2015). The third specimen, radio-carbon dated to 46 460–45 690 calibrated years BP, was collected from the beach of Hoek van Holland. Both beach specimens originate from dredged sediments from the North Sea floor and have been preserved under saline anoxic conditions. The Hoek van Holland specimen's provenance is unconfirmed, but likely originates from the Eurogeul area (Langeveld, 2013) where extensive sand-dredging has removed the Holocene (marine) overburden and exposed Late Pleistocene and Early Holocene fluvial sediments, making this area a well-known rich fossil locality for Late Pleistocene terrestrial and marine mammals (Rijsdijk *et al.*, 2013; Mol, 2016). The Zandmotor specimen originates from sand source areas about 10 km north-east of the Eurogeul, from which the same vertebrate fauna was obtained (Van der Valk *et al.*, 2011). For both beaches, material was dredged below at least 20 m of water and within 6 m under the seafloor (Langeveld, 2013).

Two lab samples of a femur and a tibiotarsus (GrA-31362 and GrA-31364) of the extinct dodo (*Raphus cucullatus*) were collected in 2005 at the brackish marsh Mare aux Songes in Mauritius (Meijer *et al.*, 2012) and radio-carbon dated to 4340–4100 and 4285–4095 calibrated years BP, respectively (Rijsdijk *et al.*, 2009). Table 1 summarizes the extinct and extant bird specimens analysed. Museum and ancient specimens were kept as dry bones at room temperature at the Naturalis Biodiversity Center and fresh specimens were frozen prior to analysis.

Bone fragments ranging from 2 to 10 mm in length were sampled in an ancient biomolecules lab using sterilized powder-free nitrile gloves, scalpels

Table 1. Details of bone specimens analysed

Species	Bone type	Specimen identification number	Age (years)	Collecting location
Anolis carolinensis	femur	RMNH.RENA.48333	0	Reared, The Netherlands
Columbia livia	femur	NMR998900003531	0	The Netherlands
Columbia livia	femur	RMNH.AVES.156616	24	The Netherlands
Falco peregrinus	femur	NMR998900003673	0	The Netherlands
Falco peregrinus	femur	RMNH.AVES.5628	89	The Netherlands
Gallus gallus	femur	RMNH.AVES.258108	0	Reared, The Netherlands
Gallus gallus	femur	RMNH.AVES.74692	65	Reared, The Netherlands
Gallus gallus	ulna distal	Coll. Dubois no. 708-3	1415–1360 cal BP	Java, Goea Djimbe, Indonesia
Gallus gallus	ulna	Coll. Dubois no. 806-22	1455–1385 cal BP	Java, Goea Djimbe, Indonesia
Pinguinus impennis	femur	V.53	2000	Velsen, The Netherlands
Pinguinus impennis	humerus	RMNH.5070466	7000–6890 cal BP	North Sea, Zandmotor, The Netherlands
Pinguinus impennis	humerus	RMNH.5070467	46 460–45 690 cal BP	North Sea, Hoek van Holland, The Netherlands
Raphus cucullatus	femur	GrA-31362	4340–4100 cal BP	Mare aux Songes, Mauritius
Raphus cucullatus	tibiotarsus	GrA-31364	4285–4095 cal BP	Mare aux Songes, Mauritius
Taeniopygia guttata	femur	Horn s.n.	0	Reared, The Netherlands
Taeniopygia guttata	femur	RMNH.AVES85442	124	Reared, The Netherlands

and drills. We performed three samplings from all materials separately following established protocols and on different days (Cooper & Poinar, 2000; Hendy *et al.*, 2018). Bone fragments were initially cleaned overnight using phosphate-buffered saline in the presence of protease inhibitors (cOmplete ULTRA tablets, EASYpack, Roche Diagnostics) and distilled water, and subsequently stored at 4 °C for further extraction procedures.

BONE EXTRACTION PROCEDURES AND MASS SPECTROMETRY SAMPLE PREPARATION

Protein extractions were performed on separate days for the various samples. Bone fragments were treated with 1.2 M HCl, 6 M Guanidine-HCl and 100 mM Tris/6 M Guanidine-HCl according to the protocol described by Van der Plas-Duivesteijn *et al.* (2016). A single control sample without bone material was analysed next to ancient samples. Fragments were treated for 24 h with 1.2 M HCl followed by washing steps with sterile water. Subsequently, fragments were treated with 6 M guandine-HCL in 100 mM Tris at pH 7.4 for 72 h. The residues were then incubated with 6 M guandine-HCL in 100 mM Tris and 0.5 M tetrasodium EDTA at pH 7.4 for 72 h. All steps were done in the presence of protease inhibitors (cOmplete ULTRA tablets, EASYpack, Roche Diagnostics) and after each treatment, samples containing extracted proteins were centrifuged at 16 000 × *g* at 4 °C. Supernatants were stored at –80 °C for subsequent mass spectrometry analysis. After acetone precipitation and freeze-drying,

proteins were dissolved in 50 mM ammonium bicarbonate. Concentrations (ranging between 0.5 and 200 µg total protein yield) were determined using the bicinchoninic acid assay (Bio-Rad) treated with 10 mM dithiothreitol (Sigma Aldrich) and alkylated using 25 mM iodoacetamide (Sigma Aldrich). The procedure was essentially as described by Jiang *et al.* (2007) and Van der Plas-Duivesteijn *et al.* (2016). In addition to this extraction procedure, tubes were changed daily for fresh ones, since we noticed that this resulted in a lower level of background noise in the bioinformatic analyses. After alkylation, peptides were digested overnight at 37 °C with 0.25 mg/mL trypsin (Sequencing grade, Promega). Digestion was quenched with 10% trifluoroacetic acid. Peptides were stored at –80 °C until further use. During mass spectrometry, *Escherichia coli* Migula, 1895 negative controls were included for global quality control.

MASS SPECTROMETRY PROCEDURES

Liquid chromatography separation of peptides and MS/MS measurements followed the standard protocol, as described by Van der Plas-Duivesteijn *et al.* (2016). In brief, 2 µL (1.5 µg protein) of bone digests were loaded and desalted on a C18 PepMap precolumn (300 µm, 5 mm i.d., 300 Å; Thermo Scientific) and separated by reversed-phase liquid chromatography using two identical ChromXP C18CL columns (150 mm, 0.3 mm i.d., 120 Å; Eksigent) coupled parallel and connected to a split-less NanoLC-Ultra 2D plus system (Eksigent) with a linear 45-min gradient from 4% to

35% acetonitrile in 0.05% formic acid and a constant (4 µL/min) flow rate. The LC system was coupled to an amaZon speed ETD ion trap (Bruker Daltonics) equipped with an Apollo II electrospray ionization (ESI) source. In general, eight samples were analysed in a row, after which a water sample was analysed. After each MS scan, up to ten abundant multiply charged species in the range m/z 300–1300 were selected for MS/MS using collision-induced dissociation (CID) and actively excluded for 1 min after having been selected twice. The LC system was controlled by HyStar 3.2 and the ion trap by trapControl 7.1. Two technical replicates were acquired.

MASS SPECTROMETRY DATA ANALYSES

After MS/MS, peptides were identified using Mascot server v.2.6.1 (available at http://www.matrixscience.com). Individual spectra were searched against all bony vertebrate sequences in SwissProt and NCBI Protein using the public version of Mascot (Perkins *et al.*, 1999). A search against the Mascot contaminant database was simultaneously performed. Tryptic cleavage was assumed, with two missed cleavages allowed. A semitryptic search was performed on the old samples as well. Carbamidomethylation of cysteines was considered a fixed modification and oxidation of methionine and proline as variable modifications. A peptide tolerance of 0.6 Da with 2 ^{13}C isotope error allowed was used, with an MS/MS tolerance of 0.6 Da. Peptide charges 2+, 3+ and 4+ were considered, as well as all fragment ions corresponding to the ESI-TRAP. To compare the peptides retrieved we used Mascot and the compareMS^2 tool (Palmblad & Deelder, 2012).

We analysed possible deamidation and oxidation of collagens retrieved from the bone samples using the Mascot server as described in the previous paragraph. However, in our tests for endogeneity, we considered oxidation of lysine, methionine and proline as a variable modification, and we looked for deamidation patterns on asparagine and glutamine residues using NQ as a variable modification, which might be an indicator for the environmental conditions in which the bones were preserved (Schroeter & Cleland, 2016). In all cases, we compared at least two biological replicates.

ALIGNMENTS AND PHYLOGENETIC ANALYSES

Peptide sequences retrieved were aligned in GENEIOUS v.10.2.3 (Biomatters, New Zealand) using MUSCLE applying the UPGMB clustering method, the pseudo-rooting method and the CLUSTALW sequence weighting scheme. For comparison purposes, alignments were prepared using a mix of COL1A1 and COL1A2 protein sequences available in NCBI GenBank and concatemers of peptide sequences newly retrieved in this study. Peptide sequences from great auk underwent local BLAST analysis in GENEIOUS using scaffolds from the recently sequenced genome of the razorbill, *Alca torda* Linnaeus, 1758 (Gilbert *et al.*, unpublished). Phylogenetic analyses were performed using PAUP* v.4.1 with the options Maximum Parsimony, heuristic search, random addition with ten replicates and TBR swapping. *Anolis* was used as an outgroup in all analyses. Bootstrap analyses were performed at 1000 iterations using simple stepwise additions, SPR swapping, MULTREES on and holding ten trees per replicate.

DATA DEPOSITION

Protein sequences recovered were deposited under number PDX009204 in the PRIDE Archive using the ProteomeXchange tool (Vizcaino *et al.*, 2016).

RESULTS

COVERAGE OF COL1A1 AND COL1A2 PEPTIDES RETRIEVED

To maximize the number of peptides covering the alpha 1 and 2 chains of collagen I (COL1A1 and COL1A2), two replicates of individual bone fragments were sampled and peptide fractions collected at four different stages during the extraction procedure were subjected to mass spectrometry on different days. Interestingly, in samples collected after a first hydrochloric acid step, we regularly find a high coverage of collagen proteins, whereas in some extracts collected at a later stage in the procedure we find a lower coverage. The reverse is also observed indicating that, in order to obtain a maximum number of peptides of unknown samples, an extensive extraction should be performed, as described previously (Jiang *et al.*, 2007; Van der Plas-Duivesteijn *et al.*, 2016; Cleland & Schroeter, 2018). These findings are also in line with the findings of Schroeter *et al.* (2016), identifying collagens in different extractions in their analyses, but other non-collagen proteins (NCPs) in discrete fractions. For an overview of coverage percentages from the various samples, see Table 2.

In all analyses taken together, COL1A1 and COL1A2 dominate the population of proteins obtained. We identify peptides covering COL1A1 and COL1A2 proteins up to a maximum of 60% and 50%, respectively, from freshly isolated bone materials (Table 2). Coverage percentages are maximally 46% and 32% for COL1A1 and COL1A2 for museum-preserved or ancient bones. Peptide sequence coverage percentages for freshly isolated bones for COL1A1 range from 39% to 60% for the various bird species. The coverage percentages for COL1A2 range from 21% to 50% for these materials.

Table 2. Coverage percentages, MASCOT values and deamidation data for COL1A1 and COL1A2 peptides retrieved from bones of the indicated species analysed; all analyses were at least performed in triplicate

Species	Age (years)	COL1A1			COL1A2		
		% Coverage	Mascot Score	Deamidation	% Coverage	Mascot Score	Deamidation
Fresh material							
Anolis sagrei	0	50	1047	no	21	349	no
Columba livia	0	56	1280	no	47	1051	no
Falco peregrinus	0	39	888	no	28	526	no
Gallus gallus	0	60	1186	no	50	986	no
Taeniopygia guttata	0	39	81	no	35	59	no
Museum / ancient specimens							
Columba livia	24	10	308	yes	10	233	yes
Falco peregrinus	89	45	1289	no	30	640	no
Gallus gallus	65	24	169	no	12	65	no
Gallus gallus	1415–1360 calibrated before present (cal BP)	24	1455	yes	18	191	yes
Pinguinus impennis	2000	25	285	no	22	245	yes
Pinguinus impennis	7000–6890 cal BP	26	150	yes	23	127	yes
Pinguinus impennis	46 460–45 690 cal BP	26	168	yes	23	80	yes
Raphus cucullatus	4340–4100 cal BP	46	308	yes	32	224	yes
Taeniopygia guttata	124	34	215	no	17	183	no

Table 3. Non-collagen proteins retrieved in mass spectrometry analyses on fresh and ancient bones. Cingulin and Calpain15-like proteins were identified based on two peptides each. Keratins I and II should be considered as contaminants

Non-collagen protein	Presence	
	Fresh bone	Museum/ ancient bones
Identified in at least 2 samples:		
α2-HS-Glycoprotein	yes	yes
Decorin	yes	yes
Bone Sialoprotein 2	yes	not found
Tubulin	yes	not found
Kininogen	yes	not found
Phosphoprotein	yes	not found
Keratins I and II	yes	yes
Calpain15-like protein	not found	yes
Cingulin	not found	yes
Ovocleidin-116	yes	not found

The museum-preserved bones produce coverage percentages for COL1A1 ranging from 10% to 46%. The coverage percentages for COL1A2 range between 10% and 32%. Interestingly, the *c.* 4200-year-old dodo bones demonstrate high coverage percentages of 46% and 32% for COL1A1 and COL1A2, respectively. For these studies, three biological samples and multiple extractions were analysed. The percentages are in the same range as those for fresh material or museum-preserved material of peregrine falcon and rock dove of only 24 to 89 years old. Bones of great auk have maximum coverage percentages of 26% and 23% for COL1A1 and COL1A2, respectively. All great auk bones, found at different locations, yield collagen peptide sequences. Other proteins are also retrieved (see below), but not from all samples, impeding a proper comparison of the results. We, therefore, focus on COL1A1 and COL1A2 sequences. Table 2 also shows the Mascot scores, which are up to 1280 for fresh *Columba livia* bone, but, remarkably, also high for museum-preserved *Falco peregrinus* bones having a Mascot value of 1289.

NCPs are retrieved in several analysed fractions (see Table 3). We find α2-HS glycoprotein (fetuin-A), decorin and keratins I and II in both fresh and ancient samples, the latter proteins representing common contaminants. Bone sialoprotein 2, tubulin, kininogen, phosphoprotein and ovocleidin-116 are also found, but only in fresh bone samples. Mascot values are higher than 50 and multiple peptide sequences are minimally found in two analyses. In addition, we find calpain15-like protein and cingulin in ancient samples, but in a limited number of analyses, based on two peptide sequences maximally and at Mascot values lower than 50. Next to NCPs, we frequently retrieve sequences from collagens that are other types than COL1A1 or COL1A2. Only COL22A1 and COL23A1 are solely

retrieved in ancient or museum samples, albeit in a limited number of samples. Sequences infrequently found are subjected to basic local alignment search tool (BLAST) analysis. In this way, we identified cingulin and calpain15-like protein, but two other peptides are rejected as unspecific (e.g. ABCA13).

COMPARISON OF COL1 PEPTIDES RETRIEVED

To compare the peptides retrieved from the ancient bones of chicken, dodo and great auk with modern rock dove spectra, butterfly plots are created using the compareMS[2] tool (Palmblad & Deelder, 2012; see also Fig. 1). In general, the data are of higher quality for rock dove, for which fresh material was analysed. This is reflected in a higher peptide signal to contaminant background ratio and higher confidence in peptide identifications. Ancient chicken bone yields a butterfly plot with a very low background signal (despite the amplification of the signal), which may reflect a good preservation state. Though the spectra compared were matched to the same peptides by Mascot, as indicated
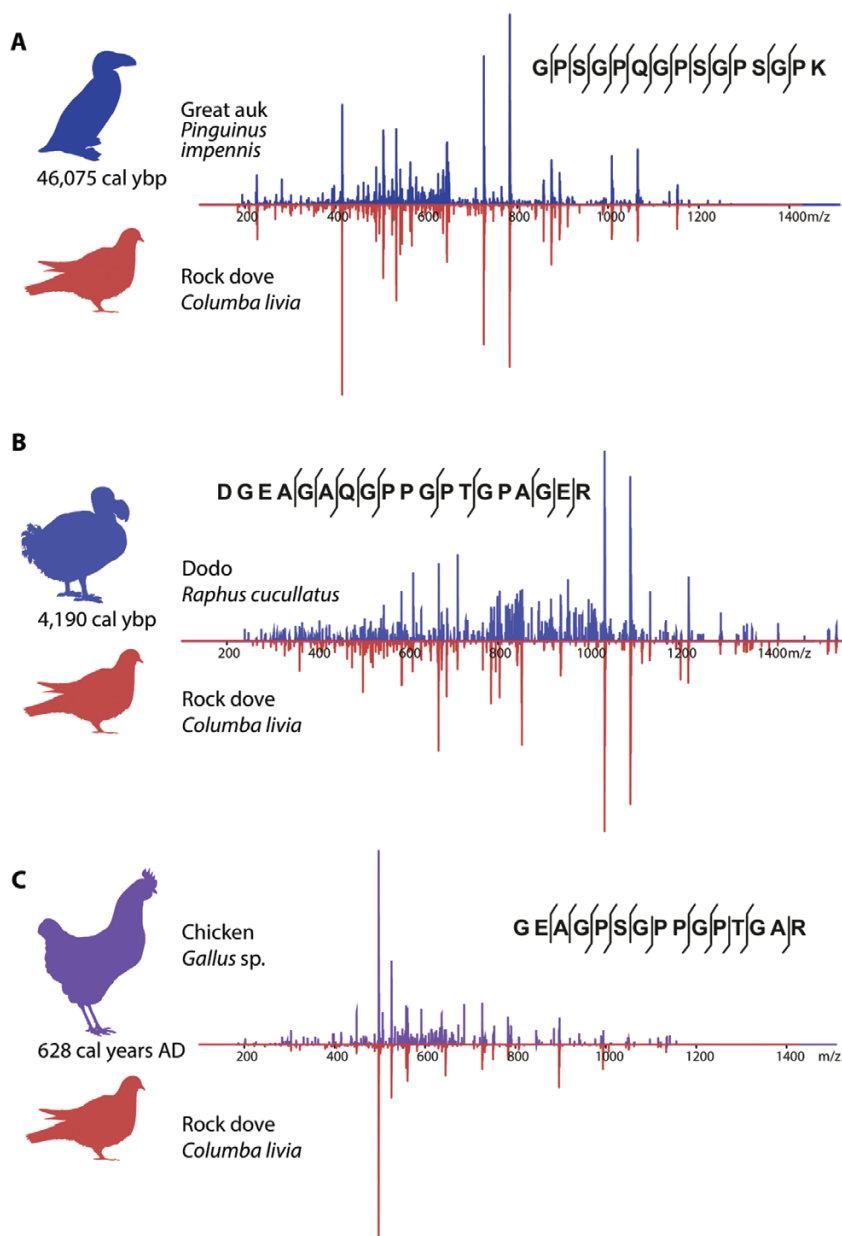

Figure 1. Comparison of tandem mass spectra derived from ancient bones of great auk (A), dodo (B) and chicken (C) with fresh bones of rock dove. Bird illustrations from phylopic.org.

in the spectra, the ions scores are often higher for rock dove (here 57 and 58, respectively) than for the corresponding spectra from dodo and great auk (17 and 44, respectively). The analyses of chicken bones yields similar results: Mascot values of fresh bone material of chicken as compared with bone material of 65-year-old and ancient bone, are 1186, 169 and 270 for COL1A1, respectively. The values for COL1A2 are 986, 65 and 191, respectively. These results show a high confidence in peptide identifications for fresh bones and lower, but reliable values for museum-preserved and ancient bones, where Mascot probability score values higher than 60 are defined as highly reliable. A limited consensus set of peptide sequences could be retrieved for the seven bird species analysed. A partial alignment is shown in Figure 3, the complete alignment is given in Supporting Information, Figure S1. Representative mass spectra for COL1A1 and COL1A2 peptides are shown in Fig. 2.

PHYLOGENETIC ANALYSES

Alignments constructed from of a mix of full-length published proteins of COL1A1 and COL1A2 and smaller peptide sequences retrieved from bones analysed in this study are phylogenetically analysed using Maximum Parsimony and this results in the bootstrap consensus tree depicted in Figure 4. Dodo (*Raphus cucullatus*) and rock dove (*Columba livia*) are part of a single clade, just like great auk (*Pinguinus impennis*) and razorbill (*Alca torda*), and peregrine falcon (*Falco peregrinus*) with zebra finch (*Taeniopygia guttata*). Chicken (*Gallus gallus*) ends up in a separate lineage. The former three lineages are part of an unresolved clade.

DEAMIDATION AND OXIDATION

As previously described for collagens (Hurtado & O'Connor, 2012; Schroeter & Cleland, 2016), we find deamidation of glutamine or asparagine up to 50% in all dodo bones analysed. As expected, we also find deamidation in the ancient bones of great auk and chicken, although not in all samples. Mascot analyses of fresh extant bird bones do not show any deamidation of asparagine and glutamine residues, as expected. Deamidation as a possible parameter for diagenesis is also presented in Table 2 for all analyses. Figure 5 shows the mass spectra of a COL1A2 peptide, which is deamidated in both old specimens but not in fresh rock dove material. Oxidation analyses were carried out by looking at lysine, methionine and proline residues. As expected, we predominantly find hydroxylated proline residues in the collagen peptides retrieved for both COL1A1 and COL1A2. Hydroxylation of methionine and lysine are not, or infrequently, found.

DISCUSSION

COVERAGE OF COLLAGEN PEPTIDES RETRIEVED

We have analysed protein sequences derived from ancient bones obtained from very different localities of two extinct bird species, dodo (*Raphus cucullatus*) and great auk (*Pinguinus impennis*), and one extant bird taxon, chicken (*Gallus*), using liquid chromatography mass spectrometry technology. The high yield obtained may be explained by the fact that we used an extensive extraction method (Jiang *et al.*, 2007) that has been validated (Cleland *et al.*, 2012; Schroeter *et al.*, 2016). We noticed previously that regular tube changing during the procedure may be beneficial for bioinformatic analyses of the data, notwithstanding a probable loss of proteins due to adherence to the plastic consumables used.

Not surprisingly, we find that the Mascot score values tend to be much higher in fresh bones, although we have also obtained a high score for a museum specimen (*Falco peregrinus* sample). In addition, we notice especially that the relatively thicker bones yield higher yields compared to the thinner bones. The high similarity between spectra suggests that spectral library searching using a library generated from a closely related species may infer additional peptide identification through direct spectral matching. Alternatively, tandem mass spectra can be compared directly and the number of shared spectra tallied, as in compareMS[2] (Palmblad & Deelder, 2012). Older material analysed, such as the chicken bones from a cave on Java, yielded a coverage of 24% in mass spectrometry. Although not that high, this is promising, since the coverage of collagens of the older Javanese bones, preserved under very different conditions than the more recent museum-preserved bones, were equally high. The result is strengthened by the mass spectrum shown in Figure 1, being very clean and with a very low background. The great auk bones studied here, and preserved under substantially different conditions, also yielded high quality protein sequences. Coverage percentages were up to 26%, which is comparable to the ancient chicken bones. The dodo bones obtained from the Mare aux Songes location in Mauritius yielded higher coverage percentages, comparable to coverage percentages of freshly isolated materials, despite being preserved anoxically under brackish conditions. The younger age of the dodo bones may explain the higher protein yield for the Mare aux Songes specimens compared with the older great auk bones (the *c*. 6900 and *c*. 46 000 years old specimens) that were preserved under saline marine conditions.

We obtained coverage percentages ranging between 21% and 60 % for fresh materials and coverage percentages ranging between 10% and 46% for museum-preserved or ancient bones. These
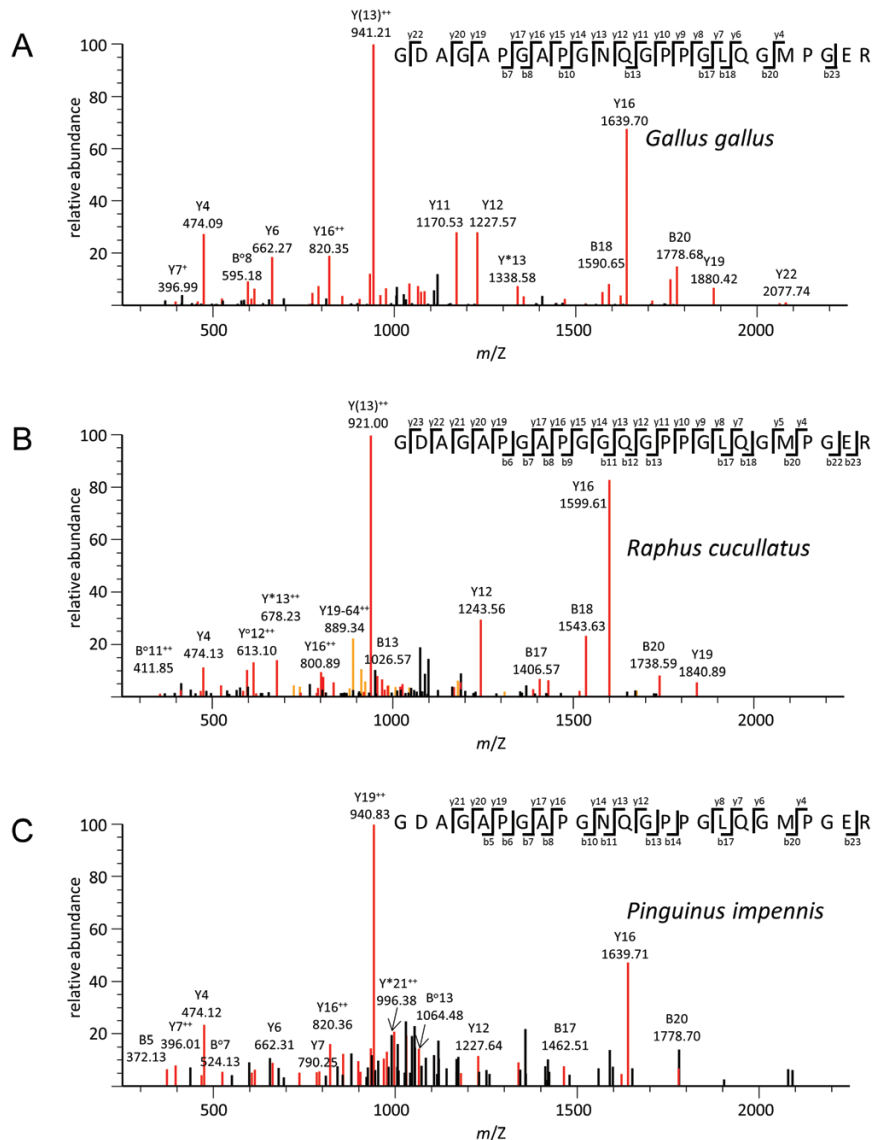
Figure 2. Mass spectra of similar COL1A1 peptide sequences retrieved in: A, chicken (recent); B, dodo (*c*. 4200-year-old specimen); and C, great auk (*c*. 46 000-year-old specimen). Mass spectra of similar COL1A2 peptide sequences retrieved in: D, chicken (recent); E, dodo (*c*. 4200-year-old specimen); and F, great auk (*c*. 46 000-year-old specimen). Black arrows were added to link spectra more clearly to their respective masses.

percentages are lower than reported by others (see, for instance: Cleland *et al.*, 2015), with recorded coverage values up to 84%. We used a capillary flow–ion trap system as this tends to be robust and able to generate comparable data, also from older or less well-preserved samples. Potential improvements could include a more sensitive nano-electrospray source or a recent Orbitrap system. These solutions may have more resolving power aiding phylogenetic studies.

We found that coverage of proteins can be equal in different fractions during the extraction procedure. However, we noticed that in fresh samples, in general,

the coverage was higher in the later extractions. For instance, for a fresh chicken sample we noticed an increase from four matching sequences up to 27 matching sequences retrieved in four consecutive extractions (data not shown). The reverse was true for the older materials (for one sample decreasing from five matching sequences in the first extraction to no matching peptide sequences in the later extractions; data not shown). The results may indicate that the proteins and peptides in the older samples were easier to extract, possibly due to the more degraded state of the bones.
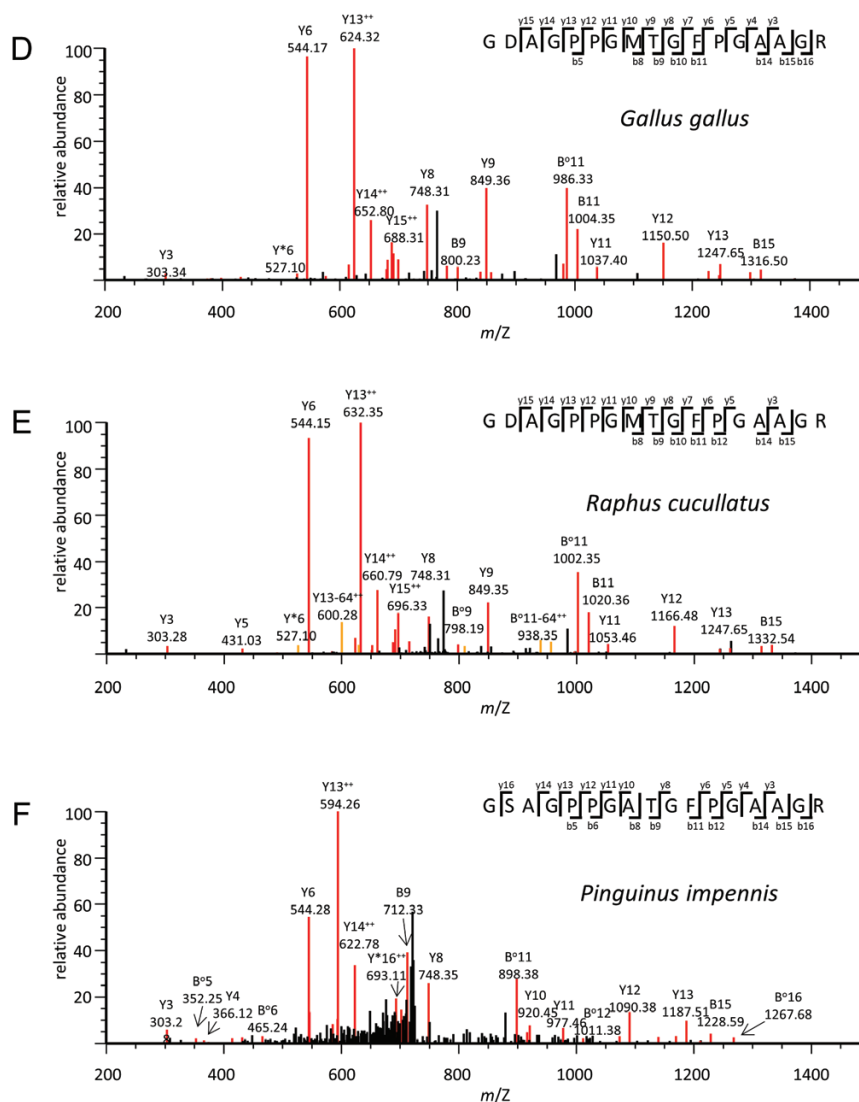
Figure 2. Continued

In a study focused on bone extractomics, it was concluded that coverage percentages of collagens and non-collagenous proteins in different extractions could be rather variable (Schroeter *et al.*, 2016). The authors found that the highest diversity of proteins was generally obtained from fractions that were not yielding the highest protein mass and the number of NCPs was higher in demineralized fractions. In our study, we noticed this as well, although our dataset was limited and we did not specifically investigate this. Bone samples that yielded higher amounts of proteins, in general, over approximately 100 µg per sample, were less diverse than samples that yielded lower amounts of proteins. The highest diversity of proteins was obtained from HCl extracted samples that yielded less than 10 µg of protein. In the bioinformatic analyses, we found NCPs amongst the large proportion of collagen proteins. NCPs that

were regularly encountered were α2-HS-glycoprotein and decorin, which have been reported earlier as proteins that can be retrieved from fresh or ancient bone materials (Cappellini *et al.*, 2012; Wadsworth & Buckley, 2014). A rationale for finding decorins and matrix metalloproteinase (MMP1) peptides, which are associated with the collagen molecules, has been reported (San Antonio *et al.*, 2011). Interestingly, we found some peptides derived from proteins exclusively found in the older materials (calpain15-like protein and cingulin), albeit with lower Mascot values next to the COL1A1 and COL1A2 proteins.

DEAMIDATION AND OXIDATION

Deamidation has been reported as a marker for ancient bone deterioration (Van Doorn *et al.*, 2012). Since

```
Anolis carolinensis      GEPGAP-----GENGTPGQSGAR------VGAPGPVGAR-----------------------------------
Gallus gallus            GEPGAP-----GENGTPGQPGAR------IGAPGPAGARGSDGSAGPTGPAGPIGAAGPPGFPGAPGAKGEIGPAGNVGP
Columba livia            ----------GFPGTPGLPGFK---------------GSDGSAGPTGPAGPIGAAGPPGFPGAPGAKGEIGPAGHVGP
Raphus cucullatus        ----------GFPGTPGLPGFK---------------GSDGSGGPVGPAGPIGAAGPPGFPGAPGAK-----------
Alca torda               TGPHGPR--------------------VGAPGPAGARGSDGSTGPTGPAGPIGAAGPPGFPGAPGAKGEIGPAGNVGP
Pinguinus impennis       -------------------------VGAPGPAGARGSDGSAGPVGPAGPIGAAGPPGFPGAPGAKGEIGPAGNVGP
Falco peregrinus         GEPGAP-----GENGTPGQPGAR--------------------------------------GEIGPAGNVGP
Taeniopygia guttata      GEPGAP-----GENGTPGQPGAR------VGAPGPAGAR-------------------------GEIGPAGNVGP


Anolis carolinensis      -------------------------------GAAGLPGVAGAPGLPGPRGIPGPSGPAGAAGTR------GLTGEPG
Gallus gallus            TGPAGPRGEIGLPGSSGPVGPPGNPGANGLPGAKGAAGLPGVAGAPGLPGPRGIPGPPGPAGPSGAR------GLVGEPG
Columba livia            AGPAGPRGEIGLPGSSGPVGPPGNPGANGLPGAKGAAGLPGVAGAPGLPGPRGIPGPPGPAGPSGAR------GAKGESG
Raphus cucullatus        -------------------------------GAAGLPGVAGAPGLPGPRGIPGPPGPAGPSGAR------GLVGEPG
Alca torda               SGPAGPR------------------------GAAGLPGVAGAPGLPGPRGIPGPPGPAGPGGAR------GLVGEPG
Pinguinus impennis       TGPAGPR------------------------GAAGLPGVAGAPGLPGPRGIPGPPGPAGPPGAR------GLVGEPG
Falco peregrinus         SGPAGPR-------------------------------------GLPGPRGIPGPPGPAGPSGAR------------
Taeniopygia guttata      SGPAGPR------------------------------------GILGPPGPAGPSGAR------GLVGEPG
```

Figure 3. Partial alignment of COL1A2 peptides retrieved (coordinates 198–348 based on CO1A2_CHICK, P02467, Uniprot.org).
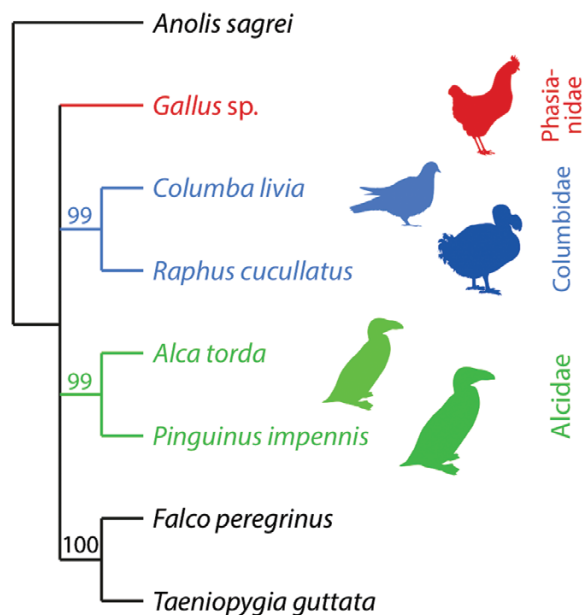

Figure 4. Bootstrap consensus of Maximum Parsimony analyses of a concatenated COL1A1 and COL1A2 alignment. Only bootstrap supports >50% are shown. Bird illustrations from phylopic.org.

deamidation has also been shown in later studies as a marker for preservation state and environmental influence, rather than as an endogeneity test (Schroeter & Cleland, 2016), we performed searches aiming for

peptides with asparagine to aspartic acid or glutamine to glutamic acid conversion. Converted glutamine and asparagine residues were not encountered in the fresh materials, but they were found in the dodo extracts and, to a lesser extent, in the other ancient samples. In the great auk samples of *c*. 46 000 years old and chicken samples dated 1455 years BP, we occasionally found conversion of asparagine or glutamine. Environmental factors might play a role in the possible conversion of these residues, since the great auk samples were preserved in marine conditions under temperate mean temperatures, whereas the dodo samples were retrieved from freshwater or brackish conditions, with the dodo from an environment with tropical mean temperatures (Rijsdijk *et al.*, 2016). This is in line with current ideas attributing a greater effect of environmental conditions than age on deamidation rates (Schroeter & Cleland, 2016). Studies performed on the influence of sampling, location and geological age on specimens indicated that HCl-based extraction plays an important role as an inducer of deamidation (Simpson *et al.*, 2016). In addition, these authors concluded that Pleistocene material, in contrast to recent material, are especially liable to undergo deamidation conversion. In our studies, the oldest great auk bone, which dates back to the Late Pleistocene, demonstrated deamidation in some of the analyses, possibly induced by sampling. We think that the marine conditions might have been beneficial for the preservation of the bone proteins, since the younger great auk bone preserved under
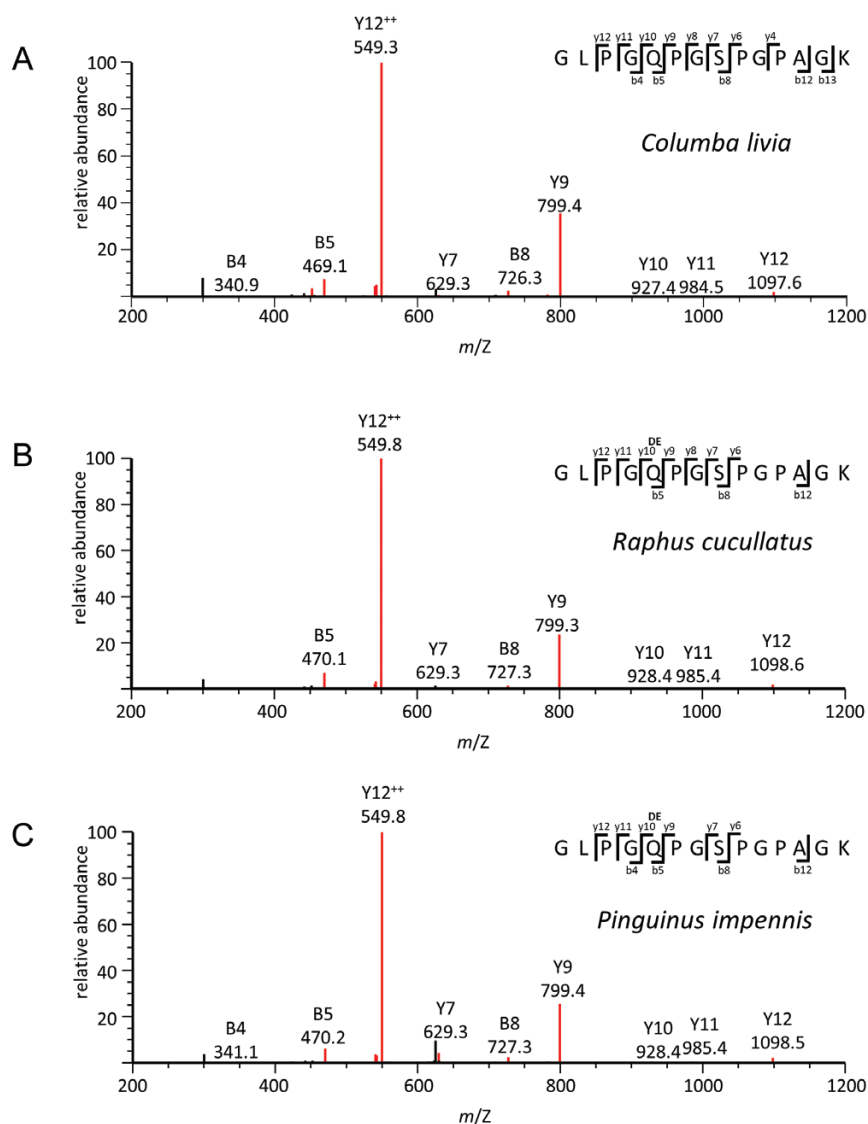
Figure 5. Mass spectra showing deamidation (DE) in two ancient samples: A, recent rock dove sample without deamidation; B, *c.* 4200 years old dodo sample with deamidation on the glutamine (Q) residue; and C, *c.* 46 000 years old great auk sample with deamidation on the glutamine (Q) residue.

saline conditions demonstrated lower deamidation levels, whereas in dodo and ancient chicken the level of deamidation was more prominent. However, the limited set of data obtained in this study does not allow us to draw firm conclusions regarding deamidation of collagen proteins under conditions of high ionic strength or under conditions of high temperatures.

COMPARISON OF COLLAGEN PEPTIDES RETRIEVED

The results of our phylogenetic analyses support current consensus in relatedness of dodo to pigeon (Shapiro *et al.*, 2002; Heupink *et al.*, 2014), great auk to razorbill (Moum *et al.*, 2002) and a more

basal position in the phylogeny for chicken (Jarvis *et al.*, 2014). Resolution of the Maximum Parsimony bootstrap consensus tree was insufficient to resolve the relationship between all lineages in more detail. The topology of the consensus tree produced in this study was congruent with the latest evolutionary insights on bird phylogenetics based on full genome DNA sequences (Zhang *et al.*, 2014; Prum *et al.*, 2015). A valuable addition to the molecular classification of archaeological bird samples was recently presented by Presslee *et al.* (2018). In this study, the authors created a library of mass spectrometry data as a reference for bird materials and successfully compared these to eggshell samples from an archaeological site.

We conclude that peptide sequences from collagens can be retrieved by mass spectrometry analyses to a high coverage from ancient bones of chicken, dodo and great auk up to *c*. 46 000 years old. With these sequences, we were able to correctly classify ancient bones to the family level. Our study provides further support for the general conclusion that classification based on palaeoproteomics can complement traditional classification methods, such as morphology and DNA, and provides another case study of extinct bird species to the set of taxa for which this method proves applicable.

REFERENCES

Allentoft ME, **Collins M**, **Harker D**, **Haile J**, **Oskam CL**, **Hale ML**, **Campos PF**, **Samaniego JA**, **Gilbert MT**, **Willerslev E**, **Zhang G**, **Scofield RP**, **Holdaway RN**, **Bunce M. 2012.** The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B* **279:** 4724–4733.

Asara JM, **Schweitzer MH**, **Freimark LM**, **Phillips M**, **Cantley LC. 2007.** Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* **316:** 280–285.

Brandt LØ, **Schmidt AL**, **Mannering U**, **Sarret M**, **Kelstrup CD**, **Olsen JV**, **Cappellini E. 2014.** Species identification of archaeological skin objects from Danish bogs: comparison between mass spectrometry-based peptide sequencing and microscopy-based methods. *PLoS ONE* **9:** e106875.

Buckley M, **Whitcher Kansa S**, **Howard S**, **Campbell S**, **Thomas-Oates J**, **Collins M. 2010.** Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *Journal of Archaeological Science* **37:** 13–20.

Cappellini E, **Gentry A**, **Palkopoulou E**, **Ishida Y**, **Cram D**, **Roos AM**, **Watson M**, **Johansson US**, **Fernholm B**, **Agnelli P**, **Barbagli F**, **Littlewood DTJ**, **Kelstrup CD**, **Olsen JV**, **Lister AM**, **Roca AL**, **Dalén L**, **Gilbert MTP. 2013.** Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae). *Zoological Journal of the Linnean Society* **170:** 222–232.

Cappellini E, **Jensen LJ**, **Szklarczyk D**, **Ginolhac A**, **da Fonseca RA**, **Stafford TW**, **Holen SR**, **Collins MJ**, **Orlando L**, **Willerslev E**, **Gilbert MT**, **Olsen JV. 2012.** Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *Journal of Proteome Research* **11:** 917–926.

Cappellini E, **Welker F**, **Pandolfi L**, **Ramos Madrigal J**, **Fotakis A**, **Lyon D**, **Moreno Mayar VL**, **Bukhsianidze M**, **Rakownikow Jersie-Christensen R**, **Mackie M**, **Ginolhac A**, **Ferring R**, **Tappen M**, **Palkopoulou E**, **Samodova D**, **Ruther PL**, **Dickinson MR**, **Stafford T**, **Chan YL**, **Gotherstrom A**, **Nathan SKSS**, **Heintzman PD**, **Kapp JD**, **Kirillova I**, **Moodley Y**, **Agusti J**, **Kahlke R-D**, **Kiladze G**, **Martinez-Navarro B**, **Liu S**, **Sandoval Velasco M**, **Sinding M-HS**, **Kelstrup CD**, **Allentoft ME**, **Krogh A**, **Orlando L**, **Penkman K**, **Shapiro B**, **Rook L**, **Dalen L**, **Gilbert MTP**, **Olsen JV**, **Lordkipanidze D**, **Willerslev E. 2018.** Early Pleistocene enamel proteome sequences from Dmanisi resolve Stephanorhinus phylogeny. *preprint*.

Cleland TP, **Schroeter ER. 2018.** A comparison of common mass spectrometry approaches for Paleoproteomics. *Journal of Proteome Research* **17:** 936–945.

Cleland TP, **Voegele K**, **Schweitzer MH. 2012.** Empirical evaluation of bone extraction protocols. *PLoS ONE* **7:** e31443.

Cleland TP, **Schroeter ER**, **Schweitzer MH. 2015.** Biologically and diagenetically derived peptide modifications in moa collagens. *Proceedings of the Royal Society B* **282:** 20150015.

Cleland TP, **Schroeter ER**, **Feranec RS**, **Vashishth D. 2016.** Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proceedings of the Royal Society B* **283:** 20160593.

Cooper A, **Poinar HN. 2000.** Ancient DNA: do it right or not at all. *Science* **289:** 1139.

Demarchi B, **Hall S**, **Roncal-Herrero T**, **Freeman CL**, **Woolley J**, **Crisp MK**, **Wilson J**, **Fotakis A**, **Fischer R**, **Kessler BM**, **Rakownikow Jersie-Christensen R**, **Olsen JV**, **Haile J**, **Thomas J**, **Marean CW**, **Parkington J**, **Presslee S**, **Lee-Thorp J**, **Ditchfield P**, **Hamilton JF**, **Ward MW**, **Wang CM**, **Shaw MD**, **Harrison T**, **Dominguez-Rodrigo M**, **MacPhee R**, **Kwekason A**, **Ecker M**, **Kolska Horwitz L**, **Chazan M**,

Kroger R, **Thomas-Oates J**, **Harding JH**, **Cappellini E**, **Penkman K**, **Collins MJ. 2016.** Protein sequences bound to mineral surfaces persist into deep time. *Elife* **5:** e17092.

Eckert AW. 1963. *The great auk: a novel*. Santa Barbara: Capra Press.

Fuller E. 1999. *The great auk*. New York: Harry N. Abrams Inc. Publishers.

Groot M. 2005. The great auk (*Pinguinus impennis*) in the Netherlands during the Roman period. *International Journal of Osteoarchaeology* **15:** 15–22.

Hendy J, **Welker F**, **Demarchi B**, **Speller C**, **Warinner C**, **Collins MJ. 2018.** A guide to ancient protein studies. *Nature Ecology & Evolution* **2:** 791–799.

Heupink TH, **Van Grouw H**, **Lambert DM. 2014.** The mysterious Spotted Green Pigeon and its relation to the Dodo and its kindred. *BMC Evolutionary Biology* **14:** 136.

Hurtado PP, **O'Connor PB. 2012.** Deamidation of collagen. *Analytical Chemistry* **84:** 3017–3025.

Jarvis ED, **Mirarab S**, **Aberer AJ**, **Li B**, **Houde P**, **Li C**, **Ho SY**, **Faircloth BC**, **Nabholz B**, **Howard JT**, **Suh A**, **Weber CC**, **da Fonseca RR**, **Li J**, **Zhang F**, **Li H**, **Zhou L**, **Narula N**, **Liu L**, **Ganapathy G**, **Boussau B**, **Bayzid MS**, **Zavidovych V**, **Subramanian S**, **Gabaldon T**, **Capella-Gutierrez S**, **Huerta-Cepas J**, **Rekepalli B**, **Munch K**, **Schierup M**, **Lindow B**, **Warren WC**, **Ray D**, **Green RE**, **Bruford MW**, **Zhan X**, **Dixon A**, **Li S**, **Li N**, **Huang Y**, **Derryberry EP**, **Bertelsen MF**, **Sheldon FH**, **Brumfield RT**, **Mello CV**, **Lovell PV**, **Wirthlin M**, **Schneider MP**, **Prosdocimi F**, **Samaniego JA**, **Vargas Velazquez AM**, **Alfaro-Nunez A**, **Campos PF**, **Petersen B**, **Sicheritz-Ponten T**, **Pas A**, **Bailey T**, **Scofield P**, **Bunce M**, **Lambert DM**, **Zhou Q**, **Perelman P**, **Driskell AC**, **Shapiro B**, **Xiong Z**, **Zeng Y**, **Liu S**, **Li Z**, **Liu B**, **Wu K**, **Xiao J**, **Yinqi X**, **Zheng Q**, **Zhang Y**, **Yang H**, **Wang J**, **Smeds L**, **Rheindt FE**, **Braun M**, **Fjeldsa J**, **Orlando L**, **Barker FK**, **Jonsson KA**, **Johnson W**, **Koepfli KP**, **O'Brien S**, **Haussler D**, **Ryder OA**, **Rahbek C**, **Willerslev E**, **Graves GR**, **Glenn TC**, **McCormack J**, **Burt D**, **Ellegren H**, **Alstrom P**, **Edwards SV**, **Stamatakis A**, **Mindell DP**, **Cracraft J**, **Braun EL**, **Warnow T**, **Jun W**, **Gilbert MT**, **Zhang G. 2014.** Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346:** 1320–1331.

Jiang X, **Ye M**, **Jiang X**, **Liu G**, **Feng S**, **Cui L**, **Zou H. 2007.** Method development of efficient protein extraction in bone tissue for proteome analysis. *Journal of Proteome Research* **6:** 2287–2294.

Langeveld B. 2013. De Zandmotor versus het strand van Hoek van Holland: opvallende verschillen in de vondstfrequentie van fossiele kleppen van vier soorten bivalven. *Afzettingen WTKG* **34:** 177–181.

Langeveld B. 2015. Vondsten van de reuzenalk *Pinguinus impennis* (Linnaeus, 1758) (Aves) uit het Eurogeulgebied. *Cranium* **32:** 19–27.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* **362:** 709–715.

Linnaeus C. 1758. *Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis 1–4*. Uppsala: L. Salvius, 1–824.

Meijer HJ, **Gill A**, **De Louw PG**, **Van Den Hoek Ostende LW**, **Hume JP**, **Rijsdijk KF. 2012.** Dodo remains from an in situ context from Mare aux Songes, Mauritius. *Die Naturwissenschaften* **99:** 177–184.

Mol D. 2016. Mammoth fossils recovered from the seabed between the British Isles and the European continent. *Bulletin du Musée d'Anthropologie Préhistorique de Monaco* **6**(Suppl.): 129–142.

Moum T, **Arnason U**, **Arnason E. 2002.** Mitochondrial DNA sequence evolution and phylogeny of the Atlantic Alcidae, including the extinct great auk (*Pinguinus impennis*). *Molecular Biology and Evolution* **19:** 1434–1439.

Orlando L, **Ginolhac A**, **Zhang G**, **Froese D**, **Albrechtsen A**, **Stiller M**, **Schubert M**, **Cappellini E**, **Petersen B**, **Moltke I**, **Johnson PL**, **Fumagalli M**, **Vilstrup JT**, **Raghavan M**, **Korneliussen T**, **Malaspinas AS**, **Vogt J**, **Szklarczyk D**, **Kelstrup CD**, **Vinther J**, **Dolocan A**, **Stenderup J**, **Velazquez AM**, **Cahill J**, **Rasmussen M**, **Wang X**, **Min J**, **Zazula GD**, **Seguin-Orlando A**, **Mortensen C**, **Magnussen K**, **Thompson JF**, **Weinstock J**, **Gregersen K**, **Røed KH**, **Eisenmann V**, **Rubin CJ**, **Miller DC**, **Antczak DF**, **Bertelsen MF**, **Brunak S**, **Al-Rasheid KA**, **Ryder O**, **Andersson L**, **Mundy J**, **Krogh A**, **Gilbert MT**, **Kjær K**, **Sicheritz-Ponten T**, **Jensen LJ**, **Olsen JV**, **Hofreiter M**, **Nielsen R**, **Shapiro B**, **Wang J**, **Willerslev E. 2013.** Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499:** 74–78.

Palmblad M, **Deelder AM. 2012.** Molecular phylogenetics by direct comparison of tandem mass spectra. *Rapid Communications in Mass Spectrometry: RCM* **26:** 728–732.

Perkins DN, **Pappin DJ**, **Creasy DM**, **Cottrell JS. 1999.** Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20:** 3551–3567.

Presslee S, **Wilson J**, **Woolley J**, **Best J**, **Russell D**, **Radini A**, **Fischer R**, **Kessler B**, **Boano R**, **Collins M**, **Demarchi B. 2018.** The identification of archaeological eggshell using peptide markers. *STAR: Science & Technology of Archaeological Research* **3:** 89–99.

Prum RO, **Berv JS**, **Dornburg A**, **Field DJ**, **Townsend JP**, **Lemmon EM**, **Lemmon AR. 2015.** A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526:** 569–573.

Rijsdijk KF, **Hume JP**, **Bunnik F**, **Florens FBV**, **Baider C**, **Shapiro B**, **Van der Plicht J**, **Janoo A**, **Griffiths O**, **Van den Hoek Ostende LW**, **Cremer H**, **Vernimmen T**, **De Louw PGB**, **Bholah A**, **Saumtally S**, **Porch N**, **Haile J**, **Buckley M**, **Collins M**, **Gittenberger E. 2009.** Mid-Holocene vertebrate bone Concentration-Lagerstätte on oceanic island Mauritius provides a window into the ecosystem of the dodo (Raphus cucullatus). *Quaternary Science Reviews* **28:** 14–24.

Rijsdijk KF, **Hume JP**, **Louw PGBD**, **Meijer HJM**, **Janoo A**, **De Boer EJ**, **Steel L**, **De Vos J**, **Van der Sluis LG**, **Hooghiemstra H**, **Florens FBV**, **Baider C**, **Vernimmen TJJ**, **Baas P**, **Van Heteren AH**, **Rupear V**, **Beebeejaun G**, **Grihault A**, **Van der Plicht J**,

Besselink M, **Lubeek JK**, **Jansen M**, **Kluiving SJ**, **Hollund H**, **Shapiro B**, **Collins M**, **Buckley M**, **Jayasena RM**, **Porch N**, **Floore R**, **Bunnik F**, **Biedlingmaier A**, **Leavitt J**, **Monfette G**, **Kimelblatt A**, **Randall A**, **Floore P**, **Claessens LPAM. 2016.** A review of the dodo and its ecosystem: insights from a vertebrate concentration Lagerstatte in Mauritius. *Journal of Vertebrate Paleontology* **35:** 3–20.

Rijsdijk KF, **Kroon IC**, **Meijer T**, **Passchier S**, **Van Dijk TAGP**, **Bunnik FPM**, **Janse AC. 2013.** Reconstructing Quaternary Rhine-Meuse dynamics in the southern North Sea: architecture, seismo-lithofacies associations and malacological biozonation. *Journal of Quaternary Science* **28:** 453–466.

Robinson NE, **Robinson AB. 2001.** Molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 944–949.

San Antonio JD, **Schweitzer MH**, **Jensen ST**, **Kalluri R**, **Buckley M**, **Orgel JP. 2011.** Dinosaur peptides suggest mechanisms of protein survival. *PLoS ONE* **6:** e20381.

Schroeter ER, **Cleland TP. 2016.** Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Communications in Mass Spectrometry* **30:** 251–255.

Schroeter ER, **Dehart CJ**, **Schweitzer MH**, **Thomas PM**, **Kelleher NL. 2016.** Bone protein 'extractomics': comparing the efficiency of bone protein extractions of *Gallus gallus* in tandem mass spectrometry, with an eye towards paleoproteomics. *Peer J* **32:** e2603.

Semal P, **Orban R. 1995.** Collagen extraction from recent and fossil bones: quantitative and qualitative aspects. *Journal of Archaeological Science* **22:** 463–467.

Shapiro B, **Sibthorpe D**, **Rambaut A**, **Austin J**, **Wragg GM**, **Bininda-Emonds OR**, **Lee PL**, **Cooper A. 2002.** Flight of the dodo. *Science* **295:** 1683.

Shoulders MD, **Raines RT. 2009.** Collagen structure and stability. *Annual Review of Biochemistry* **78:** 929–958.

Simpson JP, **Penkman KEH**, **Demarchi B**, **Koon H**, **Collins MJ**, **Thomas-Oates JE**, **Shapiro B**, **Stark M**, **Wilson JC. 2016.** The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. *Journal of Archeological Science* **69:** 29–38.

Smith NA. 2011. Taxonomic revision and phylogenetic analysis of the flightless Mancallinae (Aves, Pan-Alcidae). *ZooKeys* **91:** 1–116.

Stuart YE, **Campbell TS**, **Hohenlohe PA**, **Reynolds RG**, **Revell LJ**, **Losos JB. 2014.** Rapid evolution of a native species following invasion by a congener. *Science* **346:** 463–466.

Van der Plas-Duivesteijn SJ, **Wulff T**, **Klychnikov O**, **Ohana D**, **Dalebout H**, **Van Veelen PA**, **De Keijzer J**, **Nessen MA**, **Van der Burgt YE**, **Deelder AM**, **Palmblad M. 2016.** Differentiating samples and experimental protocols by direct comparison of tandem mass spectra. *Rapid Communications in Mass Spectrometry* **30:** 731–738.

Van der Valk B, **Mol D**, **Mulder H. 2011.** Mammoetbotten en schelpen voor het oprapen: verslag van een onderzoeksexcursie naar fossielen op 'De Zandmotor' voor de kust tussen Ter Heijde en Kijkduin (Zuid-Holland). *Afzettingen WTKG* **32:** 51–53.

Van Doorn NL, **Wilson J**, **Hollund H**, **Soressi M**, **Collins MJ. 2012.** Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Communications in Mass Spectrometry* **26:** 2319–2327.

Van Wijngaarden-Bakker LH. 1978. A subfossil great auk – *Pinguinis impennis* (L.) from the Netherlands. *Ardea* **66:** 57–61.

Vizcaíno JA, **Csordas A**, **del-Toro N**, **Dianes JA**, **Griss J**, **Lavidas I**, **Mayer G**, **Perez-Riverol Y**, **Reisinger F**, **Ternent T**, **Xu QW**, **Wang R**, **Hermjakob H. 2016.** 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research* **44:** D447–D456.

Wadsworth C, **Buckley M. 2014.** Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Communications in Mass Spectrometry* **28:** 605–615.

Welker F, **Collins MJ**, **Thomas JA**, **Wadsley M**, **Brace S**, **Cappellini E**, **Turvey ST**, **Reguero M**, **Gelfo JN**, **Kramarz A**, **Burger J**, **Thomas-Oates J**, **Ashford DA**, **Ashton PD**, **Rowsell K**, **Porter DM**, **Kessler B**, **Fischer R**, **Baessmann C**, **Kaspar S**, **Olsen JV**, **Kiley P**, **Elliott JA**, **Kelstrup CD**, **Mullin V**, **Hofreiter M**, **Willerslev E**, **Hublin JJ**, **Orlando L**, **Barnes I**, **MacPhee RD. 2015.** Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature* **522:** 81–84.

Welker F, **Hajdinjak M**, **Talamo S**, **Jaouen K**, **Dannemann M**, **David F**, **Julien M**, **Meyer M**, **Kelso J**, **Barnes I**, **Brace S**, **Kamminga P**, **Fischer R**, **Kessler BM**, **Stewart JR**, **Pääbo S**, **Collins MJ**, **Hublin JJ. 2016.** Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings of the National Academy of Sciences of the United States of America* **113:** 11162–11167.

Welker F, **Smith GM**, **Hutson JM**, **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:** e3033.

Wilson J, **Van Doorn NL**, **Collins M. 2012.** Assessing the extent of bone degradation using glutamine deamidation in collagen. *Analytical Chemistry* **84:** 8.

Zhang G, **Li C**, **Li Q**, **Li B**, **Larkin DM**, **Lee C**, **Storz JF**, **Antunes A**, **Greenwold MJ**, **Meredith RW**, **Odeen A**, **Cui J**, **Zhou Q**, **Xu L**, **Pan H**, **Wang Z**, **Jin L**, **Zhang P**, **Hu H**, **Yang W**, **Hu J**, **Xiao J**, **Yang Z**, **Liu Y**, **Xie Q**, **Yu H**, **Lian J**, **Wen P**, **Zhang F**, **Li H**, **Zeng Y**, **Xiong Z**, **Liu S**, **Zhou L**, **Huang Z**, **An N**, **Wang J**, **Zheng Q**, **Xiong Y**, **Wang G**, **Wang B**, **Wang J**, **Fan Y**, **Da Fonseca RR**, **Alfaro-Nunez A**, **Schubert M**, **Orlando L**, **Mourier T**, **Howard JT**, **Ganapathy G**, **Pfenning A**, **Whitney O**, **Rivas MV**, **Hara E**, **Smith J**, **Farre M**, **Narayan J**, **Slavov G**, **Romanov MN**, **Borges R**, **Machado JP**, **Khan I**, **Springer MS**, **Gatesy J**, **Hoffmann FG**, **Opazo JC**, **Hastad O**, **Sawyer RH**, **Kim H**, **Kim KW**, **Kim HJ**,

Cho S, **Li N**, **Huang Y**, **Bruford MW**, **Zhan X**, **Dixon A**, **Bertelsen MF**, **Derryberry E**, **Warren W**, **Wilson RK**, **Li S**, **Ray DA**, **Green RE**, **O'Brien SJ**, **Griffin D**, **Johnson WE**, **Haussler D**, **Ryder OA**, **Willerslev E**, **Graves GR**, **Alstrom P**, **Fjeldsa J**, **Mindell DP**, **Edwards SV**, **Braun EL**, **Rahbek C**, **Burt DW**, **Houde P**, **Zhang Y**, **Yang H**, **Wang J**, **Avian Genome C**, **Jarvis ED**, **Gilbert MT**, **Wang J. 2014.** Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346:** 1311–1320.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Figure S1. Full alignments of COL1A1 and COL1A2 peptides generated in this study.