UNIVERSITY of York

This is a repository copy of *The identification of archaeological eggshell using peptide markers*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/126895/

Version: Accepted Version

Article:
Presslee, Samantha Louise, Wilson, Julie Carol orcid.org/0000-0002-5171-8480, Collins, Matthew James orcid.org/0000-0003-4226-5501 et al. (1 more author) (2018) The identification of archaeological eggshell using peptide markers. STAR: Science Technology of Archaeological Research. 4:1. pp. 13-23. ISSN 2054-8923

https://doi.org/10.1080/20548923.2018.1424300

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The identification of archaeological eggshell using peptide markers

Samantha Presslee[1]*, Julie Wilson[2], Jos Woolley[3], Julia Best[4], Douglas Russell[5], Roman Fischer[6], Benedict Kessler[6], Rosa Boano[7], Matthew Collins[1,8], Beatrice Demarchi[7,1]*

[1] *BioArCh, Department of Archaeology, University of York, York, YO10 5DD, UK.*

[2] *Department of Chemistry, University of York, York, YO10 5DD, UK.*

[3] *Independent researcher*

[4] *Department of Archaeology, Anthropology and Forensic Science, Bournemouth University, Poole, BH12 5BB, UK*

[5] *Bird Group, Department of Life Sciences, Natural History Museum, Akeman Street, Tring, Hertfordshire, United Kingdom, HP23 6AP*

[6] *Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Roosevelt Drive, Oxford, OX3 7FZ, UK*

[7] *Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy*

[8] *University of Copenhagen,*

Corresponding authors: slp534@york.ac.uk, beatrice@palaeo.eu

Avian eggshell survives well in alkaline and neutral soils, but its potential as an archaeological resource remains largely unexplored, largely due to difficulties in its identification. Here we exploit the release of novel bird genomes and, for the first time on eggshell, use MALDI-ToF mass spectrometry in combination with peptide sequencing by LC-MS/MS. The eggshell proteome is revealed as unexpectedly complex, with 5755 proteins identified for a reference collection comprising 23 bird species. We determined 782 *m/z* markers useful for eggshell identification, 583 of which could be assigned to known eggshell peptide sequences. These were used to identify eggshell fragments recovered from a medieval site at Freeschool Lane, Leicester. We discuss the specificity of the peptide markers and highlight the importance of assessing the level of taxonomic identification achievable for archaeological interpretation.

Keywords: eggshell; birds; zooarchaeology; mass spectrometry (ZooMS); proteomics.

1. Introduction

Bird eggshell is one of the most underused resources in archaeology and paleontology: it can provide precious information on past environments and the way humans behaved and interacted with animals within a landscape (Saywell, 1934; Thomson, 1981; Miller and Willoughby, 2014; Best et al. 2015, Best & Mulville, 2016; Jonuks et al. 2017). However, the potential of eggshell has been held back by the difficulty in identifying the bird species to which the shells belong. Traditionally, identification is based upon size (curvature and thickness) and organization (shell and pore structure and organization), the latter requiring detailed microscopic examination in addition to a good reference collection (Murphy, 1978; Keepax, 1981; Sidell, 1993; Eastham and Gwynn, 1997; Apolinaire and Turnes, 2010; Medina et al 2011). Morphological examination is time- and cost-intensive and affected by taphonomy, which degrades and alters the shells' structure (Kuhn et al 2015).

MALDI-ToF (Matrix Assisted Laser Desorption Ionisation-Time of Flight) mass spectrometry has been used to identify proteinaceous materials by exploiting the taxonomic specificity of protein sequences (as these are translated from the DNA; Hollemeyer et al. 2002). Identification is achieved by cleaving protein sequences enzymatically and matching the resulting peptide masses to the experimental MALDI-ToF data (including *de novo* peptides obtained by MS/MS). When this is carried out on proteins from organisms of known species, it allows identification of a set of taxon-specific peptide masses (*m/z*), i.e. markers, which are used for identification of archaeological samples. This method was first applied to the analysis of hair keratins from an archaeological sample (Ötzi the Iceman's clothing) in 2008 (Hollemeyer et al. 2008). Keratins have also been studied from archaeological horn, baleen, hair and textiles (Solazzo et al. 2011; 2014; 2017) but the method has seen its widest adoption in archaeology using the more robust and widely occurring protein collagen, found in bone, teeth, antler, ivory, leather and parchment (Buckley et al 2009; 2014; Korsow-Richter et al 2011; von Holstein et al 2014; Fiddyment et al 2015; Buckley 2018).

For eggshell identification, Stewart et al. (2013) adapted this method and proposed a "profiling" approach in which the full mass spectra (i.e. the whole *m/z* list or "peak" list) were matched (using an in-house Virtual Basic application, "ChickenHawk") to a reference database of "peaks" detected in each of the 56 bird taxa they considered. They also identified "marker peaks" for some bird species. However, at the time of their study few bird genomes were available, and therefore taxon-specific peptide markers of known sequence could not be identified, , unless by direct sequencing. In contrast, the present study uses an extensive dataset of peptide sequences, obtained *de novo* by LC-MS/MS and takes advantage of recently released bird genomic data for their identification (Jarvis et al 2014).

The peptide markers determined in this study were used to identify archaeological eggshell from a recent excavation conducted by the University of Leicester Archaeological Services (ULAS) at Freeschool Lane, Leicester (UK). Archaeological and historical records provide evidence for the keeping and consumption of poultry and game in Medieval Leicester (Monkton 2015), as well as evidence of industrial activities such as leather working, with accounts of eggs being used in the leather working process (Thomson 1981; Hurcombe 2014). Thus, identification of eggshell may provide a possible interpretation of past egg use at the site, whether as food stuffs or in industrial activity.

2. Materials and Methods

2.1. Materials

62 samples were used to build the "molecular" reference collection (23 known species belonging to 15

families and 13 orders; Figure 1); we sampled 2 or more individual egg specimens per taxon wherever possible (the number of specimens per taxon is indicated in Figure 1). The reference collection includes birds that are common in the archaeological record, have archaeological significance for interpreting resource use and husbandry, or for which genomes were available (ostrich, chicken, duck, goose, pigeon, falcon, fulmar, budgerigar, barn owl). *Gallus gallus*, *Anas platyrhynchos*, *Coturnix coturnix* and *Anser anser* eggs were purchased at a local supermarket, while all other specimens were obtained from the Natural History Museum (NHM-UK), which currently holds more than 1,000,000 eggs from the period preceding the Protection of Birds Act (1954). A selection of these is set aside as a specific resource of well-catalogued and securely-identified (morphologically), but otherwise data-poor, eggshells, suitable for invasive and potentially destructive research; this includes more than 450 species (containing over 90% of breeding birds of the British Isles) (Russell et al 2010).

Figure 1. Phylogenetic tree (data from Jetz et al. 2012; Jetz et al. 2014) of the species used to build the molecular reference collection for this study. Stars indicate the species for which genomic data were available. Family and Order are in capitals and bold capitals, respectively. The numbers in the circles represent the number of eggshell fragments from different eggs that were used for the MS/MS analyses.

With regard to the archaeological samples, 30 eggshell fragments from the site of Freeschool Lane, Leicester, were randomly selected for analysis. The excavation was conducted by ULAS during the western extension of Shires Shopping Centre, and was located in what was the North-East corner of the medieval city (Buckley, 2015). During medieval times, this street would have been near the town's main trading street and a focus of occupation and industry (Morris[1] et al 2011[P2]). The medieval period of the site is defined by three narrow plots with the remains of timber-framed buildings and associated back yards containing a range of cess pits, wells and outhouses (ULAS 2006[3]). The eggshells under analysis were found at the very bottom of one of these back yard 'industrial' pits and date to 1400-1500 AD (ULAS[4] 2006; Radini 2009).

2.2. Sample Preparation

Eggshell fragments were cleaned using 0.5 M EDTA, rinsed in ultra-pure water and left to air-dry. Once dry, the eggshells were crushed to a fine powder (ca. <500 µm in size) and ~30-35 mg subsampled for analysis. The crushed eggshell was then exposed to bleach, 12% w/v sodium hypochlorite (50 µl of bleach per mg of sample), for 100 hours to isolate the intra-crystalline fraction of proteins (Penkman et al 2008; Crisp et al 2013). After 100 hours the samples were rinsed five times with ultrapure water, briefly suspended in HPLC-grade methanol, in order to remove any residual bleach, and air-dried.

The air-dried reference samples were split into two subsamples (~ 17 mg each), for digestion with trypsin ("T" subsamples) and elastase ("E" subsamples). Two different enzymes were used to increase the protein sequence coverage for LC-MS/MS as seen in other archaeological proteomic studies (Ostrom et al 2006; Buckley et al 2008; Wadsworth & Buckley 2014, Welker et al 2015; Demarchi et

al 2016). The archaeological eggshells were digested with trypsin only. Each subsample was fully demineralised in 0.6 M hydrochloric acid at room temperature, and the solutions neutralised. The neutralised solutions were freeze-dried overnight and the lyophilisate re-suspended in 200 µl 50 mM ammonium bicarbonate ("T" subsamples) or 200 µl 10 mM Tris-HCl ("E" subsamples). Samples were then reduced, alkylated, digested and purified as described elsewhere (Demarchi et al 2016).

2.3. Analytical Methods

2 separate fragments from different reference eggshells were analysed in triplicate, for a total of 132 eggshell MALDI-ToF analyses. All archaeological samples were analysed in triplicate.

1 µl of sample was spotted onto an MTP384 Bruker ground steel MALDI target plate. 1 µl of α-cyano-4-hydroxycinnamic acid matrix solution (1% in 50% Acetonitrile / 0.1% Trifluoroacetic acid (v/v/v)) was added to each sample spot and mixed with the sample. All samples were analysed on a Bruker Ultraflex III MALDI-ToF mass spectrometer. Samples were analysed using the following parameter settings: ion source, 25 kV; ion source, 21.4 kV; lens voltage, 9 kV, laser intensity 40-55%, and mass range 800-4000 Da. Peptide masses below 650 Da were suppressed. Each sample was externally calibrated against an adjacent spot containing a mixture of six peptides (des-Arg1 Bradykinn m/z=904.681, Angiotensin I m/z=1295.685, Glu1-Fibrinopeptide B m/z=1750.677, ACTH (1-17 clip) m/z=2093.086, ACTH (18-39 clip) m/z=2465.198 and ACTH (7-38 clip) m/z=3657.929).

The 62 reference samples (see Figure 1) were analysed by LC-MS/MS as described in Fischer and Kessler (2015). Briefly, peptides were separated on a PepMAP C18 column (75 µm x 500 mm, 2 µm particle size, Thermo) using a Dionex Ultimate 3000 UPLC at 250 nL/min and acetonitrile gradient from 2-35% in 5% dimethyl sulfoxide/0.1% formic acid. Peptides were detected with a Q-Exactive mass spectrometer (Thermo) at a resolution of 70000 @ 200 m/z. Up to 15 precursors were selected for HCD fragmentation.

2.4. Identification of taxonomic markers in reference eggshells

2.4.1 MALDI-ToF dataset

FlexAnalysis software version 3.3 (Bruker Daltonics) was used to perform post-analysis recalibration on the MALDI-ToF spectra using the calibration spot closest to the sample being analysed. The spectra were then normalised and exported as text files using mMass (Strohalm et al 2010), an Open Source mass spectrometry interpretation tool.

The three analytical replicates were averaged and all *m/z* values corresponding to common contaminants (keratin, trypsin, α-cyano MALDI matrix) were excluded from further analyses. The remaining *m/z* values were examined in order to identify the monoisotopic peak for each distribution and then used to build an adjacency matrix. This matrix (Table 1) identifies *m/z* values that are shared among taxa and those that are unique to a certain species.

[P5]

Table 1: Adjacency matrix quantifying the presence/absence of *m/z* values across the species analysed

in this study. Thick cell borders identify closely-related species.

2.4.2 LC-MS/MS dataset

LC-MS/MS data obtained for the 62 reference samples were analysed using PEAKS 7.5 (Bioinformatics Solutions; Zhang et al 2012). Mascot generic format (mgf) files were searched against a public reference database of bird proteomes downloaded from NCBI (https://www.ncbi.nlm.nih.gov/; accessed January 2015). In order to accelerate the bioinformatic analysis the search was limited to sub-databases containing: 1) the proteomes of all taxa for the order of the bird being considered (e.g. all Anseriformes for *Anas platyrhynchos*); 2) the proteomes of all Struthioniformes and of *Gallus gallus* and 3) all common contaminants (cRAP; common Repository of Adventitious Proteins: http://www.thegpm.org/crap/). The search assumed no digestion enzyme and had a fragment ion mass tolerance of 0.06 Da and a parent ion tolerance of 10 ppm. Results obtained by SPIDER searches (i.e. including all modifications) were used for protein identification using the following threshold values for acceptance of high-quality peptides: false discovery rate (FDR) threshold 0.5%, protein scores -10lgP ≥ 40, *de novo* sequences scores (ALC %) ≥ 80. This resulted in the identification of 5755 unique protein identifiers.

2.4.3. Combining the MALDI-ToF and LC-MS/MS datasets

The *m/z* values (unique and shared between samples in our dataset) identified from the MALDI-ToF data were matched directly to the list of 209,117 possible charged peptide masses obtained by LC-MS/MS and identified by bioinformatic analysis (corresponding to 53612 unique peptide sequences). The search was restricted to sequences found in that taxon (e.g. *m/z* values found in *Anas* were only matched to peptides identified in *Anas*), using a wide tolerance of ±0.1 Da. In order to select the most probable match in the cases where many peptide sequences could be assigned to the same *m/z* value within the 0.1 Da interval, we evaluated each of the possible combinations taking into account: the number of experimental replicates in which a certain sequence was identified, the score of the identification for the peptide, the number of spectra, the presence of "unusual" modifications. We expect that proteins will undergo post translational modifications (PTMs) as part of forming the mature protein product, but also due to diagenesis (oxidation, dehydration, deamidation) or sample preparation (carbamidomethylation). Any rare modifications are unlikely to be a frequent feature of the peptide sequences found in the archaeological record.

Studies have shown that some eggshell proteins are very similar, or highly conserved, among a wide range of avian taxa (Nys et al 2004; Marie et al 2015). For example, the protein 'BPI fold containing family B member 4' is recognised in 55 avian taxa and can be very highly conserved (Figure 2). Therefore, all the peptide markers identified were searched against the whole NCBI database (accessed 15/12/2017) and their occurrence in other organisms assessed (birds, but also e.g. mammals, bacteria). This analysis allowed the taxonomic level of identification achievable by each marker to be assessed.

Figure 2. Alignment of a region of the protein "BPI fold containing family B member 4", showing variable and conserved regions.

2.5 Microscopy

Twenty of the Freeschool Lane eggshell fragments were analysed with a Keyence Digital VHX 5000 series microscope.

3. Results and discussion

3.1. Eggshell peptide markers

The identification of taxon-specific peptide sequences in eggshell is more laborious than for bone, or other substrates where collagen dominates the proteome, because eggshell is formed rapidly in the womb of the bird (a matter of a few hours) and, as a result, the whole uterine proteome, as well as a range of biomineralization-specific proteins, can become occluded in the calcite crystals that constitute the mineral matrix. Even the isolation of the intracrystalline fraction by bleaching (Sykes et al 1995; Penkman et al 2008; Crisp et al 2013) does not drastically reduce the number nor the variability of the proteins trapped in the shell (e.g. 273 were identified in bleached ostrich eggshell; Demarchi et al 2016).

In total 782 *m/z* values were determined that, based on the MALDI-TOF analysis, appeared to be specific to particular reference taxa. Of these, 411 were unique (i.e. found only in one taxon), as summarised in Table 1. More than 70% of the *m/z* values for *Struthio*, *Anas*, *Tetrao*, *Coturnix*, *Melopsittacus* were only found within these taxa; between 50 and 70% for *Anser*, *Phasianus*, *Gallus*, *Caprimulgus*, *Phalacrocorax*; between 30 and 50% for *Puffinus*, *Fulmarus*, *Columba livia*, *Larus*, *Fratercula*, *Muscicapa*, *Carduelis*, *Passer*, *Corvus* and *Tyto*; below 30% for *C. palumbus*, *Falco tinnunculus* and *F. peregrinus*. Figure 3 shows the undirected network resulting from the representation of the adjacency matrix (Table 1): while many *m/z* values are shared among members of the same family/order, as would be expected, it is remarkable how many values are shared between different families across the avian clade.

Figure 3: Undirected network diagram showing that *m/z* values tend to be shared among members of the same family, although a high degree of non-specificity can be seen by the large number of lines criss-crossing the network. Numbers in circles are the unique *m/z* values obtained by MALDI-TOF analysis for that taxon.

Of the 782 *m/z* values identified by analysis of the MALDI spectra 583 were successfully assigned to a peptide sequence (see Supplementary Information for the full list of potential markers): 58 yielded a

match to a single peptide and 525 yielded a match to more than one peptide sequence within the 0.1 Da interval. 199 *m/z* values could not be assigned to a peptide sequence; these unmatched values were present in all bird taxa except for *Gallus* (all *m/z* for *Gallus* values produced a match to a characteristic peptide sequence), albeit in different proportions. Interestingly, the highest number of *m/z* values which did not produce a match belonged to *Larus,* a bird for which the genome is not available: of the 26 *m/z* values determined (11 found only in *Larus* and 15 shared with *Fratercula*, *Columba* and *Tyto*), only 4 (3 of which had appeared to be *Larus*-specific) produced a match to a peptide sequence. It is therefore possible that these unassigned markers could be useful for characterisation of *Larus* eggshell. However, 17 of the 46 *m/z* values determined for characterisation of *Anser* (for which genomic data is available) also did not yield any peptide match.

The peptide sequences matched to *m/z* values mostly belonged to eggshell-specific proteins: ovocleidin-116-like (OC-116 is the eggshell ortholog of mammalian MEPE; Bardet et al. 2010; Mann & Mann 2013; Hincke et al. 1999), ovocalyxins (Cordeiro et al. 2013; Mann & Mann 2013; Mann 2015), ovocleidin-17 (Hincke et al. 1995), struthiocalcin and rheacalcin (Mann 2004; Mann & Siedler 2004) and ansocalcin, the struthiocalcin-like protein sequences from goose eggshell (Lakshminarayanan et al. 2003), which has now been retracted from NCBI and yet was recovered with 100% coverage in geese and duck eggshells in the present study. Egg tissues proteins (ovomucoid, ovotransferrin, ovostatin, mucin, proteins from the vitelline membrane, albumin) were also present, as well as a number of other components of the uterine proteome (Mann 2007; Du et al. 2015). This increases the confidence in the reliability of the markers for eggshell identification as it shows the presence of identified peptides should be quite consistent between samples, despite the randomness due to the fast incorporation of the proteins during egg formation. Furthermore, this analysis provided important information on the evolution of the mechanisms of eggshell biomineralisation, which will be explored in a forthcoming study (Demarchi et al in preparation).

Taxonomic identification of eggshell is obviously easier when a certain species, analysed by MALDI, yields spectra characterised by peptide markers that are mostly unique, i.e. not shared with other taxa. This was the case for Struthioniformes, Anseriformes and Galliformes, with the number of unique markers being 36 for *Struthio*, 22 for *Anas*, 22 for *Tetrao*, 20 for *Gallus*, 18 for *Anser*, 18 for *Phasianus* and 16 for *Coturnix*. A high number of unique peptide markers was found for *Melopsittacus* (18) and *Phalacrocorax* (16), while the number of markers found for *Carduelis* (11), *Muscicapa* (10) and *Passer* (9) was the highest among the Passeriformes (*Corvus* = 5). For the other taxa, *Caprimulgus* counted 4 markers, *Corvus* 5, *Columba livia* 6 and *C. palumbus* 4. *Falco peregrinus* (3) and *F. tinnunculus* (8) also displayed a relatively low number of unique peptide markers, similar to *Fratercula* (8) and *Fumarus* (10), *Larus* (3), *Puffinus* (10) and *Tyto* (7). However, each of the peptide markers that are determined here as potentially unique may in fact be present in other bird taxa, which were either not analysed in this study, or present in low concentrations that may not be detectable in our reference spectra, although we try to account for biological variability with the use of replicate samples.

In order to verify the occurrence of the peptide markers in other organisms which were not included in our reference collection, we performed peptide-protein BLAST searches of the whole NCBI database for each of the 583 peptide sequences identified. In each case the organisms (or group of organisms) that yielded a sequence hit with 100% identity and 100% coverage are reported in SI. This was a useful exercise, as it revealed that many of the sequences are in fact very conserved. For example, 129 were found in multiple bird species across the avian clade, therefore we can assume that these will be useful to identify eggshell at the level of Class (Aves). The highest number of these generic "bird" peptides were found in the seabirds (*Fulmarus*, *Fratercula* and *Puffinus*). Furthermore, 71 peptides were found in a range of organisms, and they can sometimes be used only as a marker of "vertebrates" or even higher. Peptides were found in various families and genera belonging to the same order or superorder (89) or to different orders (39), including 45 which we have identified to the level of

protein sequence (struthiocalcin and ansocalcin), as these molecules regulate mineralisation and so their homologues are widespread in avian eggshells.

The most informative peptide sequences were those found in birds belonging to the same family (81, among which 45 Phasianidae) and those found in species belonging to the same genus (117, including 5 peptide sequences for *Anas*, 2 for *Anser*, 7 for *Melopsittacus*, 8 for *Phalacrocorax*, 3 for *Tyto*, 19 for *Struthio* and 28 for *Falco*). However, the latter cannot be used confidently as genus-level markers because they represent the only genus for which a genome is available within their family; we therefore suggest their use as family markers (Anatidae, Psittaculidae, Phalacrocoracidae, Tytonidae, Struthionidae and Falconidae). On the contrary, family Phasianidae is fairly well represented at genomic level, including turkey, quail, black grouse and greater prairie chicken as well as domestic chicken. Therefore, *Gallus* markers can be used to identify chicken to the level of genus (if not species). Similarly, genomes of family Columbidae include both *Columba livia* (the rock pigeon) and *Patagioenas fasciata monilis* (the band-tailed pigeon), therefore sequences found in *Columba* but not in *Patagioenas* can be used fairly confidently as genus-level markers. One sequence (VQPYQGLWLFR, from the vitelline membrane outer layer protein 1-like) might even represent a species-specific marker for the rock pigeon, but only the sequencing of more genomes will be able to assess the validity of this assumption.

3.1. Identification of the Freeschool Lane eggshells

The Freeschool Lane eggshell samples were analysed and the *m/z* list compared (using a tolerance of 0.1 Da) to the peptide markers. All samples which had yielded good-quality spectra (27 out of 30) were identified as *Gallus*. A typical spectrum is shown in Figure 4: we identified 16 unique *Gallus* peptide markers (highlighted by the black chicken symbol in Figure 4). All of these, except for *m/z* 803.4, 878.4 and 1042.6 were only found in *Gallus gallus* sequences by BLAST searches. We also identified the presence of further 7 peptides (orange chicken symbol and *m/z* values) that are shared by *Gallus* and other taxa in our reference dataset (both Phasianidae and sea birds).

Three samples produced poor-quality spectra, and we hypothesise that they had potentially been exposed to high temperatures (boiled or roasted, which cannot necessarily be identified from macroscopic or microscopic alterations of the shell); however further analysis would be required to confirm this (Crisp 2013). The microscopic examination revealed no mobilisation of calcium from the internal surface of the eggshell, which occurs with chick development (Figure 5).

Figure 4: MALDI-MS of sample L-13795 (average of three replicate spectra) from the Freeschool Lane site, showing the identification of the sample as *Gallus* eggshell. This is based on a set of 16 peptide markers, unique to *Gallus*, and further supported by 7 markers which are non-specific but are found in Galliformes and, in some cases, other birds (*Fulmarus*, *Phalacrocorax*).

Figure 5: Digital microscopy image of sample XXX [7] showing the internal eggshell surface with no reabsorption of calcium from the mammillae structures.

Therefore, these samples do not represent post-hatching discard in the pit, but rather eggs which were infertile or relatively freshly laid (within the first 10-12 days of incubation) (Karlsson and Lilja 2008; Best et al in prep). The identification of chicken eggshell ties in well with the other archaeological evidence (e.g. Connor and Buckley 1999; Browning 2009). Overall the results suggest that the chicken eggs at this site were being used as "eggs", rather than for chick production. They may have been eaten, or possibly used in some of the industrial activities present at the site, such as leather working.

5. Conclusions

In this study we identify a set of 782 *m/z* values potentially useful for eggshell identification in the archaeological record. These include 583 *m/z* values which could be matched to peptide sequences, obtained from the in-depth proteomic analysis of 62 eggshell samples belonging to 23 different avian species. The increase in the number of genomes publicly available on the NCBI repository has also allowed us to evaluate the occurrence of each peptide sequence in different organisms, by carrying out BLAST searches against the whole protein sequence database. This analysis has shown that the specificity of most peptides is in fact lower than expected, which implies that the taxonomic identification can be achieved only to family level or higher. It also points towards the curious fact that bits of "molecular machinery" somehow involved in biomineralization probably have very deep origins in time. Notable exceptions were found to be chicken and pigeon, which can be identified with more confidence to the level of genus or even species. In this study, we could successfully identify the eggshell samples from the Freeschool Lane site, Leicester (UK) as chicken, and the observation of their internal microstructure suggested that these eggs were probably exploited for eating or for some other human-related activity, rather than the breeding of chicks.

Overall, we highlight that while using *m/z* values as markers can be extremely useful for taxonomic identification of some archaeological substrates, the complexity and variability of the eggshell proteome is such that the use of markers of known peptide sequence is preferable. Nonetheless, the limitations of this approach are evident: hundreds of avian species, each with a different ecological requirements and behaviours, may be present at any one archaeological site. In order to address meaningful questions related to human-bird interactions in the past, identification at the level of genus or species is needed. The reference collection will need to be expanded in breadth and depth and will need to keep up with advances in avian genomics, which will be crucial for assessing the ability of the markers to distinguish between closely related species.

Acknowledgements

The authors are grateful to one anonymous reviewer and the Editor for their comments on this paper. Kirsty Penkman and Ross D Macphee are thanked for support and useful discussion. [8]

Funding details

This work was supported by the EPSRC under Grant EP/I001514/1; the NSF under Grant [number xxxx][9] ; the Italian Ministry of Education, Universities and Research (Programme "Young Researchers - Rita Levi Montalcini")

Disclosure statement

The authors declare no conflict of interests.

ORCID identifiers

S.P.

J.Wi 0000-0002-5171-8480

R.F. 0000-0002-9715-5951

...

B.D. 0000-0002-8398-4409

Biographical note

S.P is a PhD student at the University of York, working on the application of proteomics in palaeontology. J.Wi. is a reader in the Department of Mathematics at the University of York. J.Wo is an independent specialist in data analysis. D.G.D.R. is a curator for the department of Zoology, NHM, specialising in the national egg and nest collections. R.F is the head of the Discovery Proteomics Facility, University of Oxford. B.M.K is a professor of biochemistry and life science mass spectrometry at the University of Oxford. J.B is a Postdoctoral Research Associate at the University of Bournemouth. A.R is a Postdoctoral Research Associates at the University of York, where M.J.C. is Professor of Biomolecular Archaeology. R.B. and B.D. are Researchers in Anthropology and Archaeology at the University of Turin, with a focus on palaeoproteomics and protein diagenesis.

References

Apolinaire, E., Turnes, L., 2010. Diferenciación específica de reidos a partir de fragmentos de cáscaras de huevo. Su aplicación en sitios arqueológicos del Holoceno Tardío, in: Beron, Monica Luna Leandro Bonomo, Mariano Montalvo, Claudia Aranda, Claudia Carrera Aizpitarte, Manuel Zarate, Marcelo A (ayacucho, Pcia (Ed.), MAMUL Mapu: Past and Present from Archeology

Pampeana. Books Espinillo, Ayacucho, Buenos Aires, pp. 370–389.

Best, J., Demarchi, B., Presslee S., 2015. Avian eggshell. In: Haddow, S.D., (Ed.) Çatalhöyük 2015 Archive Report pp.111-115.

Best, J., Mulville, J., 2016. Birds from the water: Reconstructing avian resource use and contribution to diet in prehistoric Scottish Island environments. Journal of Archaeological Science: Reports, 6, pp. 654–664.

Best, J., Maltby, M., Demarchi, B. In prep. Counting chickens before they hatch: developing identification of embryonic stage in eggshell for archaeological application.

Browning, J., 2009. The Animal Bones from Freeschool Lane, Leciester. In: Coward, J., and Speed, G., (Eds.) Excavations at Freeschool Lane, Highcross Leicester. ULAS Archive Report

Bardet, C. et al 2010. OC-116, the chicken ortholog of mammalian MEPE found in eggshell, is also expressed in bone cells. Journal of experimental zoology. Part B, Molecular and developmental evolution, 314(8), pp.653–662.

Buckley, M., Collins, M. & Thomas-Oates, J., 2008. A method of isolating the collagen (I) alpha2 chain carboxytelopeptide for species identification in bone fragments. Analytical biochemistry, 374(2), pp.325–334.

Buckley, M., Collins, M., Thomas-Oates, J., Wilson, J.C., 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 23, pp. 3843–3854.

Buckley M., 2018. Zooarchaeology by Mass Spectrometry (ZooMS) Collagen Fingerprinting for the Species Identification of Archaeological Bone Fragments. In: Giovas C., LeFebvre M. (eds) Zooarchaeology in Practice. Springer, Cham

Buckley, R., 2015. Archaeology and Medieval Leicester. In: Wood, M., 2015 Medieval Leicestershire. Leicestershire Field Worker, Leicester

Connor, A., and Buckley, R., 1999. Roman and Medieval Occupation in Causeway Lane, Leicester. University of Leicester, Archaeological Services.

Cordeiro, C.M.M. et al 2013. Ovocalyxin-36 is a pattern recognition protein in chicken eggshell membranes. *PloS one*, 8(12), p. e84112.

Crisp, M., Demarchi, B., Collins, M.J., Morgan-Williams, M., Pilgrim, E., Penkman, K.E.H, 2013. Isolation of the intra-crystalline proteins and kinetic studies in Struthio camelus (ostrich) eggshell for amino acid geochronology. Quaternary geochronology, 16(0), pp.110–128.

Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C.L., Woolley, J., Crisp, M.K., Wilson, J., Fotakis, A., Fischer, R., Kessler, B.M., Rakownikow Jersie-Christensen, R., Olsen, J.V., Haile, J., Thomas, J., Marean, C.W., Parkington, J., Presslee, S., Lee-Thorp, J., Ditchfield, P., Hamilton, J.F., Ward, M.W., Wang, C.M., Shaw, M.D., Harrison, T., Domínguez-Rodrigo, M., MacPhee, R.D.E., Kwekason, A., Ecker, M., Kolska Horwitz, L., Chazan, M., Kröger, R., Thomas-Oates, J., Harding, J.H., Cappellini, E., Penkman, K., and Collins, M.J. (2016) Protein sequences bound to mineral surfaces persist into deep time. eLife [online] 5, e17092.

Du, J. Hincke, M.T., Rose-Martel, M., Hennequet-Antier, C., Brionne, A., Cogburn, L.A., Nys, Y., Gautron, J., 2015. Identifying specific proteins involved in eggshell membrane formation using gene expression analysis and bioinformatics. BMC genomics, 16(1), p.792.

Eastham, A., Gwynn, I.A., 1997. Archaeology and the electron microscope. Eggshell and neural network analysis of images in the Neolithic: Methods. Anthropozoologica, pp. 85–94.

Fiddyment, S., Holsinger, B., Ruzzier, C., Devine, A., Binois, A., Albarella, U., Fischer, R., Nichols, E., Curtis, A., Cheese, E., Teasdale, M.D., Checkley-Scott, C., Milner, S.J., Rudy, K.M., Johnson, E.J., Vnouček, J., Garrison, M., McGrory, S., Bradley, D.G., Collins, M.J., 2015. Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. Proceedings of the National Academy of Sciences of the United States of America 112, pp. 15066–15071.

Fischer, R., Kessler, B.M., 2015. Gel-aided sample preparation (GASP)--a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. Proteomics 15, pp. 1224–1229.

Hincke, M.T. et al 1999. Molecular cloning and ultrastructural localization of the core protein of an eggshell matrix proteoglycan, ovocleidin-116. The Journal of biological chemistry, 274(46), pp. 32915–32923.

Hincke, M.T., Tsang, C.P., Courtney, M., Hill, V., Narbaitz, R., 1995. Purification and immunochemistry of a soluble matrix protein of the chicken eggshell (ovocleidin 17). Calcified tissue international, 56(6), pp.578–583.

Hollemeyer, K., Altmeyer, W., Heinzle, E., Pitra, C., 2008. Species identification of Oetzi's clothing

with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry based on peptide pattern similarities of hair digests. Rapid communications in mass spectrometry: RCM, 22(18), pp.2751–2767.

Hollemeyer, K., Altmeyer, W. & Heinzle, E., 2002. Identification and quantification of feathers, down, and hair of avian and mammalian origin using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Analytical chemistry, 74(23), pp.5960–5968.

Hurcombe, L.M. 2014 Perishable Material Culture in Prehistory: Investigating the Missing Majority Abingdon: Routledge.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford, M.W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F., Sheldon, F.H., Brumfield, R.T., Mello, C.V., Lovell, P.V., Wirthlin, M., Schneider, M.P.C., Prosdocimi, F., Samaniego, J.A., Vargas Velazquez, A.M., Alfaro-Núñez, A., Campos, P.F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D.M., Zhou, Q., Perelman, P., Driskell, A.C., 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, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, F.K., Jønsson, K.A., Johnson, W., Koepfli, K.-P., O'Brien, S., Haussler, D., Ryder, O.A., Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren, H., Alström, P., Edwards, S.V., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320–1331.

Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., and Mooers, A.O. (2012) 'The Global Diversity of Birds in Space and Time'. *Nature* 491 (7424), pp. 444–448

Jetz, W., Thomas, G.H., Joy, J.B., Redding, D.W., Hartmann, K., and Mooers, A.O. (2014) 'Global Distribution and Conservation of Evolutionary Distinctness in Birds'. Current Biology: CB 24 (9), pp. 919–930

Jonuks T., Oras E., Best J., Demarchi B., Mänd R., Presslee S., Vahur S., 2017. Multi-method Analysis of Avian Eggs as Grave Goods: Revealing Symbolism in Conversion Period Burials at Kukruse, NE Estonia, Environmental Archaeology, DOI: 10.1080/14614103.2016.1263374.

Karlsson, O., Lilja, C. 2008. Eggshell structure, mode of development and growth rate in birds. Zoology 111, pp. 494-502.

Keepax, C.A., 1981. Avian egg-shell from archaeological sites. Journal of Archaeological Science 8, pp. 315–335.

Korsow-Richter, K.K., Wilson, J., Jones, A.K.G., Buckley, M., van Doorn, N., Collins, M.J., 2011. Fish 'n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone

fragments. Journal of Archaeological Science 38, 1502–1510.

Kuhn, B.F., Carlson, K.J., Hopley, P.J., Zipfel, B., Berger, L.R., 2015. Identification of fossilized eggshells from the Taung hominin locality, Taung, Northwest Province, South Africa. Palaeontol. Electronica 18, pp. 1–16.

Lakshminarayanan, R., Valiyaveettil, S., Rao, V.S., Kini, R.M., 2003. Purification, Characterization, and in Vitro Mineralization Studies of a Novel Goose Eggshell Matrix Protein, Ansocalcin. The Journal of biological chemistry, 278(5), pp.2928–2936.

Mann, K., 2004. Identification of the major proteins of the organic matrix of emu (Dromaius novaehollandiae) and rhea (Rhea americana) eggshell calcified layer. British poultry science, 45(4), pp.483–490.

Mann, K., 2007. The chicken egg white proteome. Proteomics, 7(19), pp.3558–3568.

Mann, K., 2015. The calcified eggshell matrix proteome of a songbird, the zebra finch (Taeniopygia guttata). Proteome science, 13, p.29.

Mann, K. & Mann, M., 2013. The proteome of the calcified layer organic matrix of turkey (Meleagris gallopavo) eggshell. Proteome science, 11(1), p.40.

Mann, K. & Siedler, F., 2004. Ostrich (*Struthio camelus*) eggshell matrix contains two different C-type lectin-like proteins. Isolation, amino acid sequence, and posttranslational modifications. *Biochimica et biophysica acta*, 1696(1), pp.41–50.

Marie, P., Labas, V., Brionne, A., Harichaux, G., Hennequet-Antier, C., Nys, Y., Gautron, J., 2015. Quantitative proteomics and bioinformatic analysis provide new insight into protein function during avian eggshell biomineralization. Journal of Proteomics, 113, pp. 178–193.

Medina, M.E., Hospitaleche, C.A., Turnes, L., 2011. Huevos de Rhea pennata en el Holoceno de la Provincia de Córdoba (Argentina): Implicaciones Ambientales, Zoogeográficas y Arqueológicas. Archaeofauna 20, pp. 157–169.

Miller, J.M., Willoughby, P.R., 2014. Radiometrically dated ostrich eggshell beads from the Middle and Later Stone Age of Magubike Rockshelter, southern Tanzania. Journal of Human Evolution, 74, pp. 118–122.

Monkton, A., 2015. Food for the people of Medieval Leicester: the evidence from Environmental Archaeology. In: Wood, M., 2015 Medieval Leicestershire. Leicestershire Field Worker, Leicester.

Murphy, P., 1978. Norwich Avian Eggshell (No. Report No. 2673). Ancient Monument Laboratory, London.

Nys, Y., Gautron, J., Garcia-Ruiz, J.M., Hincke, M.T., 2004. Avian eggshell mineralization: biochemical and functional characterization of matrix proteins. Comptes Rendus Palevol 3, pp. 549–562.

Ostrom, P.H., Gandhi, H., Strahler, J.R., Walker, A.K., Andrews, P.C., Leykam, J., Stafford, T.W.,

Kelly, R.L., Walker, D.N., Buckley, M., Humpula, J. (2006) Unravelling the Sequence and Structure of the Protein Osteocalcin from a 42ka Fossil Horse. Geochimica et Cosmochimica Acta 70 (8), pp. 2034–2044.

Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. Quaternary Geochronology 3, pp. 2–25.

Radini, A. 2009. The plant remains. In: Coward, J. & Speed, J., (Eds). Excavation at Freeschool Lane. Leicester Highcross Project. ULAS Report 2009, p. 140.

Russell, D., White, J., Maurer, G., Cassey, P., 2010. Data poor egg collections: cracking an important research resource. Journal of Afrotropical Zoology, 6, pp. 77–82.

Saywell, L.G., 1934. Clarification of Wine. Industrial & Engineering Chemistry Research, 26, pp. 981–982.

Sidell, E.J., 1993. A Methodology for the Identification of Archaeological Eggshells. U Penn Museum of Archaeology.

Solazzo, C., Heald, S., Ballard, M.W., Ashford, D.A., DePriest, P.T., Koestler, R.J., Collins, M.J., 2011). Proteomics and Coast Salish Blankets: A Tale of Shaggy Dogs? Antiquity, 85 (330), pp. 1418-1432.

Solazzo, C., Wilson, J., Dyer, J.M., Clerens, S., Plowman, J.E., von Holstein, I., Walton Rogers, P., Peacock, E.E., Collins, M.J., 2014. Modeling Deamidation in Sheep α-Keratin Peptides and Application to Archeological Wool Textiles'. Analytical Chemistry, 86 (1), pp. 567–575

Solazzo, C., Fitzhugh, W., Kaplan, S., Potter, C., Dyer, J.M., 2017. Molecular Markers in Keratins from Mysticeti Whales for Species Identification of Baleen in Museum and Archaeological Collections'. PloS One 12 (8), pp. e0183053

Stewart, J.R.M., Allen, R.B., Jones, A.K.G., Kendall, T., Penkman, K.E.H., Demarchi, B., O'Connor, T., Collins, M.J., 2014. Walking on Eggshells: A Study of Egg Use in Anglo-Scandinavian York Based on Eggshell Identification Using ZooMS. International Journal of Osteoarchaeology, 24, pp. 247–255.

Stewart, J.R.M., Allen, R.B., Jones, A.K.G., Penkman, K.E.H., Collins, M.J., 2013. ZooMS: making eggshell visible in the archaeological record. Journal of Archaeological Science, 40, pp. 1797–1804.

Strohalm, M., Kavan, D., Novák, P., Volný, M., Havlícek, V., 2010. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Analytical Chemistry 82, 4648–4651.

Sykes, G.A., Collins, M.J., Walton, D.I., 1995. The significance of a geochemically isolated intracrystalline organic fraction within biominerals. Organic geochemistry, 23(11–12), pp.1059–1065.

Thomson, R., 1981. Leather Manufacture in the Post-medieval Period with Special Reference to

Northamptonshire. Post-Medieval Archaeology 15, pp. 161–175.

Welker, F., Collins, M.J., Thomas, J.A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S.T., Reguero, M., Gelfo, J.N., Kramarz, A., Burger, J., Thomas-Oates, J., Ashford, D.A., Ashton, P.D., Rowsell, K., Porter, D.M., Kessler, B., Fischer, R., Baessmann, C., Kaspar, S., Olsen, J.V., Kiley, P., Elliott, J.A., Kelstrup, C.D., Mullin, V., Hofreiter, M., Willerslev, E., Hublin, J.-J., Orlando, L., Barnes, I., MacPhee, R.D.E., 2015. Ancient Proteins Resolve the Evolutionary History of Darwin's South American Ungulates. Nature, 522 (7554), pp. 81–84

von Holstein, I.C.C., Ashby, S.P., van Doorn, N.L., Sachs, S.M., Buckley, M., Meiri, M., Barnes, I.,Brundle, A., Collins, M.J., 2014. Searching for Scandinavians in pre-Viking Scotland: molecular fingerprinting of Early Medieval combs. Journal of Archaeological Science, 41, pp. 1–6.

Wadsworth, C. & Buckley, M., 2014. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid communications in mass spectrometry: RCM*, 28(6), pp.605–615.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B., 2012 PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. Molecular & cellular proteomics 11 (4), pp. M111.010587

---

missing in reference list

[P2]reference missing!

is this the Browning paper?

Browning?

[P5]BD to put Word doc in

[P6]SP to put high res version in folder

+slp534@york.ac.uk which one?

anyone else?

+slp534@york.ac.uk +matthew@palaeome.org