Performance of Proteomic Sex Estimation by Amelogenin in the Arid Environments of Mesopotamia and the Nile Valley

By

TAM THANH TANG

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Forensic Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Jelmer W. Eerkens, Chair

Christyann M. Darwent

Glendon J. Parker

Committee in Charge

2024

Abstract

Biological sex is an important criterion in establishing an individual's identity in forensic casework and archaeological studies. In skeletonized cases, biological sex can be estimated by three different approaches: osteology, genomics, and proteomics. This study focuses on the proteomic approach to sex estimation using sexually dimorphic amelogenin proteins in human enamel, the most archaeologically and forensically robust human tissue. This approach is especially useful for working with teeth from juvenile skeletons and in arid environments where aDNA is poorly preserved and more limited. This study uses 41 archaeological samples from Iraq, Egypt, and Sudan to evaluate the proteomic approach under challenging preservation conditions for genomic and osteological sex estimation. The study included both adult and non-adults, with the youngest individual being a neonate and the sample a dental germ. Peptides specific to the X-chromosome isoform of amelogenin were detected in all samples. Peptides specific to the sexually dimorphic Y-chromosome isoform were detected in 14 (n = 14) samples. Comparisons between the intensity of recovered peptides of amelogenin, enamelin, ameloblastin and serum albumin suggest that serum albumin is most subject to degradation compared to the other enamel matrix proteins. Comparison of samples from arid archeological sites in Iraq, Egypt, and Sudan, to those from Central California (site CA-ALA-554) show that enamel proteins are more poorly preserved in arid environments.

Acknowledgments

Thank you to my committee members Drs. Jelmer Eerkens, Glendon Parker, and Christyann Darwent. My thanks to Diana Malarchik for initial training on proteomic analysis in the Parker Lab. Thank you to Dr. Tammy Buonasera and Michelle Salemi who assisted with data analysis for this research. Funding was provided by the Forensic Science Graduate Group student research stipend.

1. Introduction

1.1 Forensic Proteomics and Enamel Proteins

Proteomics is the study of all proteins expressed in a cell, tissue, or organism, plays a crucial role in forensic science. The term also refers to the techniques used to identify and quantitatively analyze a large number of proteins [1]. Proteins are significant body components comprising polypeptide chains of amino acids, are among the most abundant molecules in the body, and are found in all body cells. Proteins are relatively resistant to degradation because they are able to link with other molecules. For instance, in mineral compounds bones or teeth, such as calcium, inhibit the degradation of protein structures and contribute to their preservation [2]. Their abundance and properties make proteins a significant tool in the forensic field. For instance, blood typing, or ABO typing, was one of the first forensic applications in forensics by recovering proteins on the surfaces of red blood cells as markers of blood type to narrow down the pool of possible contributors to the forensic sample [3]..

Protein analysis can also provide crucial forensic investigative leads such as biological sex estimation [4], tissue and bodily fluid identification [5], as well as information in other areas such as archaeology, cultural heritage, paleontology, anti-doping, and detection of bacterial antibiotic resistance phenotypes in clinical medicine [6]. Solid tissues such as bone, tooth, and hair have also been processed and analyzed for forensic, archaeological, and paleontological studies [1]. Proteomes in tooth enamel merit special consideration in degraded human remains, as enamel is the most archaeologically and forensically robust human tissue.

The anatomy of a human tooth consists of the crown, the visible part in the oral cavity, and the root embedded inside. The core of the tooth, the dentin, is a supportive structure that surrounds the pulp. Enamel is the protective layer covering the crown, with the dentin-enamel junction as the interface between the enamel and dentin. During initial formation, tooth enamel consists of 96% minerals [7]. The mineral content is made up of carbonated hydroxyapatite (C-Hap), which forms long, thin crystallites called ribbons. Between these ribbons is the organic matrix. Through mineralization, adult teeth acquire a large amount of calcium phosphate [8]. Thus, the adult tooth protein-rich dentin-enamel junction and inner enamel junction exhibit high fracture toughness and crack growth resistance [9]. During amelogenesis, the formation of enamel, the enamel protein matrix interacts with calcium phosphate minerals to form bundled enamel crystals that are account for the observed physical and biological properties. Amelogenesis occurs in two phases of tooth development: secretory and maturation. Enamel proteins degrade during enamel maturation, resulting in a protein content of only 1% in mature enamel [7]. The composition of enamel makes it possible for teeth to withstand chewing forces and extreme chemical and temperature variations while protecting the tooth structure from external damage [8]. The enamel protein matrix involved in amelogenesis consists of unique enamel proteins secreted by highly specialized cells called ameloblasts [10]. During the secretory phase, ameloblasts move away from dentin as the enamel thickens and proceed to secrete three structural proteins: amelogenin (AMELX), ameloblastin (AMBN), and enamelin (ENAM), and the proteinase MMP-20, an enzyme that degrades the enamel protein matrix, allowing crystallization to occur in a spatially and temporally controlled manner [11], [12]. Amelogenin, a key marker in sex estimation, is the most

abundant protein found in the original matrix of enamel. The AMELX gene on the X chromosome primarily encodes amelogenin proteins, while the AMELY gene on the Y chromosome in males is estimated to be only about 10% as active as AMELX in producing amelogenin proteins [4], [13], [14]. As observed with many of the genes on the sex chromosomes, the X and Y isoforms of amelogenin do not undergo homologous recombination, making amelogenin a reliable genetic marker for biological sex estimation.

1.2 Enamel Proteins for Sex Estimation

Sex cannot be accurately estimated using osteological markers for non-adult remains, and thus forensic science and archaeology studies will turn to DNA for sex estimation [15]. In addition, non-adult bones have lower bone density relative to adult bones (i.e., more porous), there are more skeletal elements because their bones have not yet fused, and their skeletal remains are often more fragmented and more poorly preserved. Lacking DNA, there is no reliable method to estimate biological sex for sub-adult remains. In this respect, the proteomic approach provides a valuable alternative to sex estimation when osteological indicators and DNA recovery is not feasible.

In the past, the whole-crown micro-etch technique successfully recovered enamel proteins from the mature enamel of two recently extracted and two ancient (800 to 1100 AD) third molars [16]. Despite the small sample size, the whole-crown micro-etch technique identified and recovered a few amelogenin peptides. One of the peptides obtained from the ancient tooth using this technique was specific to AMELX, the X-isoform of amelogenin. At the time, it was suggested that amelogenin peptides may have been preserved inside tooth enamel.

A later extraction method using a single tooth instead of the whole crown was found to be more effective, employing nanoflow liquid chromatography coupled with tandem mass spectrometry (MS/MS) to analyze extracted proteins [12][13]. NanoLC-MS/MS offers higher sensitivity and reliability for protein identification. Thus, the analysis method successfully identified dimorphic sequences of AMELY, peptides specific to the Y isoform of amelogenin, allowing the possibility of estimating sex from tooth enamel samples. The same protein extraction procedure effectively recovered amelogenin from thirteen ancient teeth [18]. The age and sex of the thirteen individuals had also been determined by osteological, epigraphic, and grave goods. Sex estimation was performed using the oxidized version of the AMELY peptide as this version was believed to predominate in ancient samples. The biological sex of these samples was estimated based on the presence or absence of AMELY peptides. The results of the thirteen ancient teeth agreed with the predetermination of sex by standard osteological methods. This opened a new door to successful sex estimation using dental enamel from a single tooth. Quantitation of peptides may advance the proteomic approach to sex estimation in rare cases of chromosomal abnormalities and polymorphisms.

Considering the potential of protein quantification, a statistical framework for sex estimation was established to quantify amelogenin proteins and subsequently to estimate sex assignment [19]. Reliable detection of any AMELY peptide indicates the presence of a male sex chromosome. The absence of a male-specific signal, however, can suggest either a low signal sample that is male (since males also secrete AMELX), or female sex. AMELY expression is only 10% of that observed for AMELX [13]. Hence, there is a possibility of undetectable AMELY signals in lower quality samples. Quantifying the

peptide signal helps to address the possibility of false negative samples—male assigned as female—since high quality samples are less likely to have ‘drop-out’ of the AMELY peptides [19]. The statistical framework involved expressing the biological sex of samples as a Pr(F) for samples lacking AMELY due to the potential undetectable AMELY signals. The estimation of sex using this framework agreed with the sex estimates using DNA and osteology. However, it is crucial to note that further research is needed to extend peptide survivability and demonstrate the composition of the peptide signal over archaeological time.

The proteomic approach to sex estimation using enamel proteins has been optimized primarily by analyzing enamel from teeth from non-arid environments and modern samples. Limited studies evaluate the applicability and reliability of sex estimation of nonadult skeletal remains in arid environmental contexts where aDNA preservation is notoriously poor [20], [21]. Due to the lack of sexually dimorphic features, sex estimation of nonadult skeletal remains often relies on genomic DNA analysis. However, this approach frequently fails in arid environments, such as those characterizing Iraq, Egypt, and the Sudan. The ability to implement a proteomic approach to sex estimation of nonadult remains would be highly significant in forensic and archeological casework. The first applications of sex estimation utilizing enamel peptides from nonadult samples of deciduous and permanent teeth were developed in 2018 and published in 2021 [22]. Similar methodology, proposed by Stewart et al. [16], was successfully used for sex estimation [17]. Peptide material from the crowns of developing perinatal teeth that were not fully mineralized was successfully recovered, and there was no evidence suggesting nonadult teeth expressed different amounts of the X and Y isoforms of the protein [22].

Optimizing the approach for children's remains is necessary because an osteological or morphological approach is unreliable for assessing biological sex. Likewise, optimizing the proteomic technique for samples with poor aDNA preservation will widen the applicability of sex estimation [23]. The proteomic method applied under these conditions in forensic casework is also given consideration. The goal is to optimize the amelogenin protein extraction protocol for non-adult samples from arid environments. The result could immediately benefit archaeology and forensic casework, specifically when dealing with infant or subadult remains.

1.3 Environmental Effect on Enamel Proteins

A major challenge when working with ancient samples is degradation. The same environmental factors that work to degrade anatomy and ancient DNA in human tissues, particularly in hot and humid environments, may also serve to degrade proteins. As a result, peptide density in enamel is predicted to be lower in samples collected from poor preservation environments compared to samples run under the same methods but from non-arid environmental contexts. Studies indicate that the amount of proteins contained within the enamel matrix increases by volume across the secretory stage, reaching its peak early during the maturation stage, and decline rapidly thereafter [24]. Amelogenin proteins are distributed across the thickness of developing enamel. Regions closest to the surface contain mostly intact amelogenin, whereas deeper regions of the enamel matrix contain progressively smaller proteins, presumably older amelogenin that has been partially degraded over time [25] [26]. Amelogenin peptides are hypothesized to be preserved in ancient enamel due to their active involvement in the enamel formation process and their binding to hydroxyapatite mineral surfaces. While amelogenin is the

most abundant in the enamel organic matrix, there are other proteins such as ameloblastin (AMBN) and enamelin (ENAM) that are also involved in the amelogenesis [12]. These proteins are essential for proper enamel formation, and their relative expression levels are tightly regulated to ensure the correct structure and maturation of enamel. Ameloblastin is less abundant than amelogenin, but it is essential for the attachment and structural support of ameloblasts to the enamel surface [27]. Ameloblastin expression in ameloblasts peaks during the secretory stage and diminishes during the maturation stage of enamel formation, mostly disappearing from the enamel matrix [28], [29]. Its removal facilitates further mineralization and enamel crystal growth. Enamelin, a phosphorylated glycoprotein, functions mainly in the early stage of enamel development and may be involved in the initiation of hydroxyapatite crystallization [30]. Both amelogenin and enamelin appear to be present in the inner layer of developing enamel and become co-localized within the matrix layer [31], [32]. However, given the lesser abundance of ameloblastin and enamelin in the matrix compared to amelogenin, they are expected to be in lower abundance in archaeological, degraded enamel samples, especially from arid environments.

Unlike amelogenin, ameloblastin or enamelin being enamel structural proteins and are directly involved in the formation of enamel, serum albumin (ALBU) is often found in small amounts in the dentin or pulp chamber of teeth. Recent research proposes that serum albumin infiltrates immature enamel, survives the proteolytic conditions therein, and then binds to enamel crystals to stall their growth [33]. Serum albumin is a major component of dentin, but is also present in the demineralized fraction of porcine and bovine enamel matrix [34]. In developing rat enamel, albumin was detected throughout

the secretion and transitional stages but was lost in early maturation [35]. It was also suggested that albumin might be blocked from binding to crystallites *in vivo* due to the presence of intact amelogenin. Access to the crystal surfaces might only occur when the amelogenin-rich matrix is degraded and withdrawn in the transition and maturation stages [35]. Immunohistochemistry of rabbit teeth suggests that serum albumin once localized in the surface layer of enamel, is not incorporated into the enamel matrix [36]. Hence, in arid environments and extreme climate conditions, amelogenin is expected to be preserved better than serum albumin. Being only on the surface layer of enamel, serum albumin is more susceptible to degradation. Albumin is a globular protein, and so albumin present in enamel at the time of demise would degrade at similar rates to other serum globular proteins such as haptoglobin and immunoglobulins. Albumin is therefore an ideal normalizing molecule to control for degradation of serum proteins trapped in enamel, including immune proteins [37].

2. Material and Methods

2.1 Sample Origin

Archeological tooth samples were collected from burial sites in Iraq, Egypt, and Sudan (**Table 1**). Ur was a prominent Sumerian city-state in ancient Mesopotamia, marked today by Tell el-Muqayyar in Dhi Qar Province, southern Iraq. Enamel samples from Ur, Iraq, were collected during excavation in two different campaigns—American and German. The excavation areas of the American campaign were located close to the city center, while the German campaign excavation areas were closer to the city wall. The samples mainly date back to the Old Babylonian period (c. 2004–1595 B.C.E.), but some are Middle Babylonian/Kassite (c. 1595–1155) to Neo- Babylonian (c. 626–539) in age.

The Ur samples were provided by Dr. Andrea Göhring, an archaeologist at Christian-Albrechts University in Kiel, Germany at the time, to Dr. Jelmer Eerkens at UC Davis for sex estimation.

El-Hesa is situated in the Nile River valley, south of the Egyptian city of Aswan. Enamel samples from El-Hesa were provided to Dr. Eerkens' Archaeometry Lab by Dr. Alex De Voogt, an Assistant Curator at the American Museum of Natural History at the time. These samples are estimated to date between 400 BCE and 1000 CE. A second archaeological site in the region is Tombos, located at the Third Cataract of the east bank of the Nile River in Sudan. The samples from Tombos were provided to the lab by Michele Buzon, a bioarcheologist in the Department of Anthropology at Purdue University. The Tombos samples are estimated to date to c. 1450–300 BCE. The third archaeological site in this region is Sai Island, which is situated between the second and third cataracts of the Nile River in Upper Nubia, Sudan. The Sai Island samples date between c. 0–100 CE and were provided by Dr. Tosha Dupras from Texas Tech University. These tooth samples were originally for a study exploring early childhood dietary patterns in a Late Meroitic population from the island [38], [39]. These sites are all located in a what is classified as a "hot desert climate" zone (BWh) according to the Köppen-Geiger climate classification system [40]. The mean annual temperature for these site localities ranges from 29 to 31 °C (84–86 °F), and the mean annual precipitation ranges from 0.03 to 1.05 cm (0.01–0.41 inches).

Table 1. Archaeological tooth samples analyzed for this study. Average annual temperature and average annual precipitation values obtained from the Red Cross Climate Center, last updated in June 2024 (https://www.climatecentre.org)

	Ur, Iraq	El Hesa, Egypt	Tombos, Sudan	Sai Island, Sudan	Total samples
Adult teeth*	5	6	–	1	12
Non-adult permanent teeth	10	–	3	10	23
Non-adult deciduous teeth	6	–	–	–	6
Average annual temperature	29.85°C	30.83 °C	28.94°C	28.94°C	
Average annual precipitation	1.05 cm	0.03 cm	0.06 cm	0.06 cm	
Total samples					**41**

Adult = individuals with estimated age at death of 18 years old and over.

2.2 Sample Preparation

Samples were provided to UC Davis as either whole enamel fragments or as powdered enamel, and they were stored at room temperature (20°C) until processed. All samples were analyzed in Glendon Parker's laboratory at the University of California, Davis. Samples were extracted over several different batches using the same protocol. A sample blank was analyzed along with each batch of samples following protocols in Parker et al. [4] with modifications described in Buonasera et al. [33]. Each tooth sample was first cut, crushed, or powdered using a dental drill with a diamond burr (Brasseler Inc. Cat # 005098U0, 918B.11.180 HP Medium Flexible Coated Double Sided Diamond Disc) to separate the enamel from the dentine. Samples submitted as enamel fragments or

powdered tooth enamel were processed without any additional preparation. Each sample was weighed (18.9–22.0 mg) and transferred to a pre-weighed milling vial (Omni-International Inc., 2 mL reinforced tubes with screw cap and silicone O-ring) that contains seven ceramic milling beads (Omni-International Inc., Cat# 19-646, 2.8 mm ceramic beads). A blank was run along with each set of samples. The procedure for the blank sample followed the exact step-by-step procedure as the enamel samples.

2.3 Protein Extraction and Digestion

Samples were demineralized by adding 200 µL of 1.2M hydrochloric acid to each vial. Samples were milled using a bead mill homogenizer (Omni-International Inc., Bead-Ruptor 24 Elite) for 3 minutes at 6.00 m/s with a 1-minute rest after 1 minute of milling to reduce heat. Samples were centrifuged for 1 minute at max speed using the Eppendorf non-refrigerated centrifuge. After incubation at 56°C, 1200 rpm for 60 minutes on the Eppendorf Thermomixer, samples were spun down using the centrifuge for 30 seconds at max speed. After consolidating the liquid, samples were neutralized with 2.0 M ammonium bicarbonate and continuously tested for pH using a pH strip. Once the pH has been adjusted to 7.5-8, vials are vortexed and centrifuged. 12 µL of 0.25% (w/v) ProteaseMax solution (Promega Inc.) was added to each sample vial along with 10 µL of trypsin (0.5 µg/µL, Thermo Scientific). Samples were incubated at 37°C, 600 rpm, in the dark for ~20–22 hours for protein digestion.

2.4 Sample Clean-Up

Three 1.5 ml Safe-Lock Eppendorf tubes, each containing 1000 µL of 50% Acetonitrile (ACN), were prepared for the Wetting step. Three 1.5 ml Safe-Lock Eppendorf tubes, each containing 1000 µL of 0.1% Trifluoroacetic acid (TFA), were prepared for the

Equilibration step. One 1.5 ml Eppendorf tube and one 1.5 ml Eppendorf LoBind tube were prepared per sample for washing and elution, respectively. Each washing tube is filled with 500 µL of 0.1% TFA. Each LoBind elution tube is filled with 50 µL of an elution buffer containing 60% ACN and 0.1%TFA.

Prior to the clean-up procedure, samples were centrifuged for 15 minutes. The supernatant, containing peptides, was transferred to 0.22 µm PTFE centrifugal filter tubes and was further centrifuged for another 15 minutes. Filter from the filtration tube is removed, and the tube containing filtrate becomes the loading tube. Samples were then acidified by adding 10 µL of 10% TFA and tested with pH strips until the pH was between 1 and 2. Tubes were pulse centrifuged to move the volume to the bottom of the vial. The clean-up procedure required a fresh Pierce 100 µL C18 pipette tip (Thermo Scientific, Cat# 87784) for each sample. I followed these steps in this order: 1) Wetting, 2) Equilibration, 3) Sample Loading, 4) Washing, and 5) Elution. First, the resin in the C18 tip is wetted by pipetting and discarding 50 µL of 50% ACN three times. The resin is equilibrated by pipetting and discarding 50 µL of 0.1% TFA three times. The C18 tip is transferred from the first pipette to the second one and set to 200 µL. The sample is loaded by drawing and dispensing the sample back in the loading tube twenty times. The C18 tip is then removed from the pipette and transferred back to the first pipette for washing. The resin and loaded sample are washed by pipetting and discarding 50 µL of 0.1% TFA until there is no more wash buffer. The peptides in the sample are then eluted to the elution tube by pipetting the elution buffer up and down the elution tube fifteen times. These steps were repeated for all samples and blanks. Samples and blanks were then labeled and stored in a freezer at -20°C until submitted for analysis.

2.5 Data Processing and Analysis

The samples were submitted to the Proteomics Core Facility for peptide analysis. The peptide concentration was measured using the Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Pierce™) and 0.8 microgram of peptide, or 40 percent of the total sample if the sample was too diluted, was applied to mass spectrometry. LCMS-Peptides were resolved on a Thermo Scientific Dionex UltiMate 3000 RSLC system using a PepMap 75umx25cm C18 column with 2 µm particle size (100 Å pores), heated to 40 °C. A 0.8 µg of total peptide amount was injected for each sample. Peptides were directly eluted into an Orbitrap Exploris 480 instrument (Thermo Fisher Scientific, Bremen, Germany). Spray voltage was set to 1.8 kV, funnel RF level at 45, and heated capillary temperature at 275 °C. The full MS resolution was set to 60,000 at m/z 200 and full MS AGC target was 300% with an IT set to Auto. Mass range was set to 350–1500. AGC target value for fragment spectra was set at 200% with a resolution of 15,000 and injection time was set to Standard and Top40. Intensity threshold was kept at 5E3. Isolation width was set at 1.6 m/z, normalized collision energy was set at 30%. The statistical analysis and quantification of peptides specific to AMELY and AMELX proteins followed the same framework as Parker et al. (2019). This bioinformatic analysis of different isoforms of amelogenin revealed multiple sexually dimorphic amino acid substitutions, identified in blue squares in Figure 1.

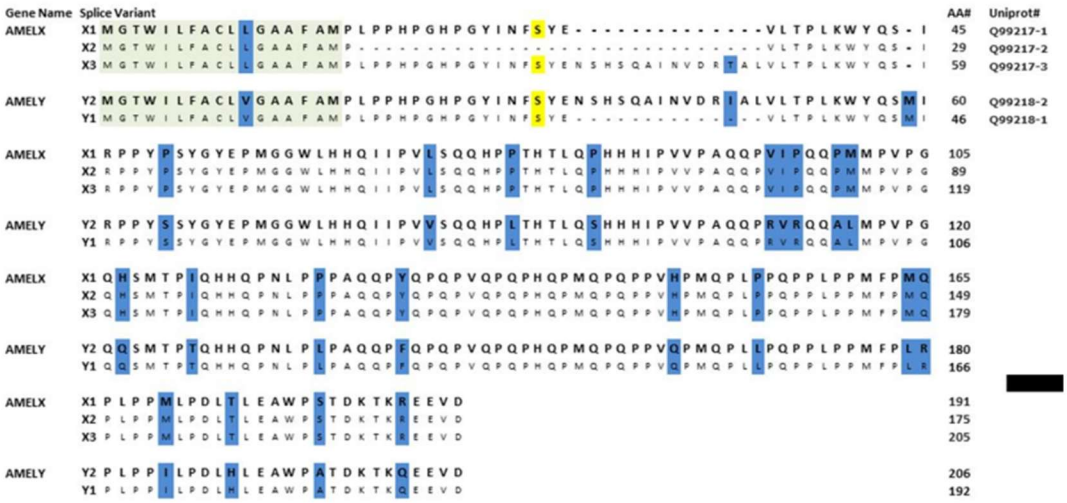

Figure 1. Amelogenin proteins sequence. Isoforms 1,2,3 of X-chromosome amelogenin proteins (Q99217-1,2,3) and isoforms 1,2 of Y-chromosome amelogenin proteins (Q99218-1,2) are shown. Amino acid variations between the X-chromosome amelogenin and the Y-chromosome amelogenin are indicated by blue squares [4].

RAW formats of mass spectrometry datasets were processed using PEAKS™ Software Studio 10.6 and peptide matching software (Bioinformatics Solutions Inc., Waterloo, ON), searched against the default “Human_AMEL_isoforms” database. The FASTA formatted UNIPROT Homo sapiens reference protein database (http://www.uniprot.org/proteomes/UP000005640) was modified to include additional FASTA protein entries of peptide sequences from all splice variants associated with AMELX_HUMAN (Q99217-1, 2,3) and AMELY_HUMAN (Q99218-1,2) proteins gene products [13], [41]. Peptide matching spectral assignment was conducted using default conditions with these exceptions: parent mass error tolerance at 10 ppm, fragment mass error tolerance at 0.04 Da, monoisotopic precursor search type, maximum 2 missed cleavages with trypsin. The algorithm searched for peptides partially modified by oxidation (MHW), pyroglutamate conversion from glutamate and glutamine, and deamidation (NQ).

All peptide sequences assignments were filtered by a 0.9% false discovery rate. Each peptide was quantified by summing the intensity of each signal for the peptide-specific primary precursor mass over charge ratio (m/z). Positive protein identifications required a minimum of two unique peptides. Proteins with only one unique peptide were excluded from analysis.

3. Results and Discussion

3.1 Biological Sex Estimation

The 41 samples (Supplemental Table 1) revealed an average of 1055 ± 421 peptide spectrum matches (PSM), 445 ±156 unique peptides (peptide sequences), and 137 ± 73 proteins. The detection of AMELY peptides in 14 samples is consistent with male sex (Table 2). The ratio of AMELY to AMELX intensity at 8.04 ± 5.73% in these samples is consistent with the relative expression of 10% reported in the literature [13].

Table 2. Biological sex estimation based on proteomic analysis. Male = M (contains AMELY), female = F (probability female ≥ 0.8), sex indeterminate = Ind. (probability female < 0.8).

	Ur			**El Hesa**			**Tombos**			**Sai Island**		
	M	F	Ind.	M	F	Ind.	M	F	Ind.	M	F	Ind.
Adult tooth	1	4	–	3	2	1	–	–	–	–	1	–
Non-adult permanent tooth	5	5	–	–	–	–	1	2	–	2	7	1
Non-adult deciduous	1	4	1	–	–	–	–	–	–	–	–	–
Total samples	**7**	**13**	**1**	**3**	**2**	**2**	**1**	**2**	**–**	**2**	**8**	**1**

Female sex was previously suggested to be expressed as a probability due to high variation in combined peak ion intensity is that male false negative [4]. Enamel samples lacking AMELY_HUMAN peptides were assigned a probability of female sex based on

the normalized cumulative AMELX_HUMAN signal following methods previously described in Parker et al. [4] and slightly modfied in Buonasera et al. [33]. Of the 27 samples lacking AMELY, 24 (89%) have pr(F) ≥ 0.8. This includes the youngest individual (TT004), a female newborn from Sai Island, which produced a probability of female of pr(F) = 0.87 on a dental germ.

Figure 2. examines the combined intensity (CI) of expressed AMELX peptides, which is crucial in understanding the differences in the preservation of amelogenin proteins, for adults vs. non-adults for all the samples combined. Overall, the results suggest that AMELX expression remains relatively consistent regardless of the estimated age of death of the remains. This significant finding supports the use of proteomics for sex estimation when DNA analysis and osteology are not feasible.

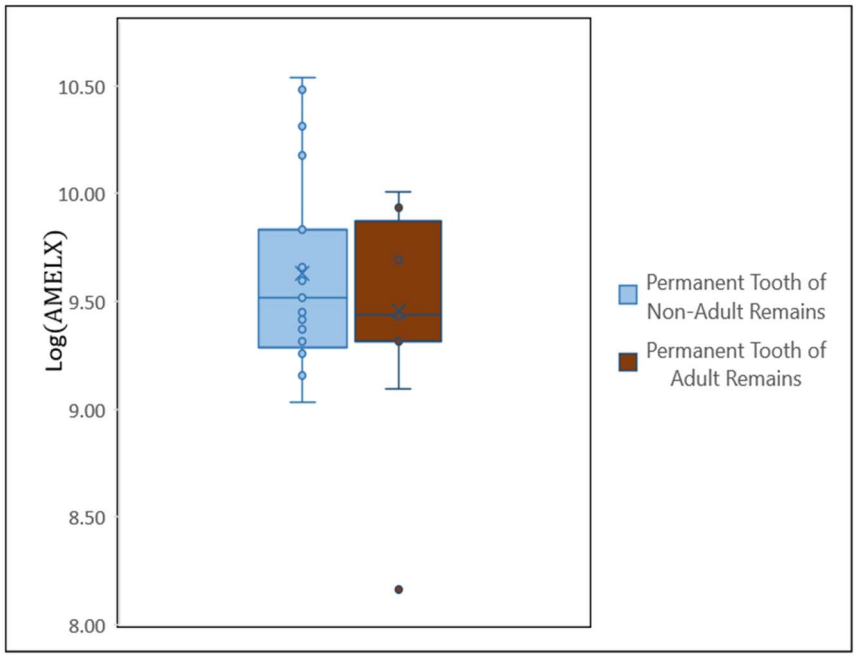

Figure 2. AMELX Expression in Permanent Tooth of Non-adult versus Adult Remains. The combined intensity (CI) of the X-isoform of amelogenin proteins is shown for enamel from non-adult remains compared to adult remains.

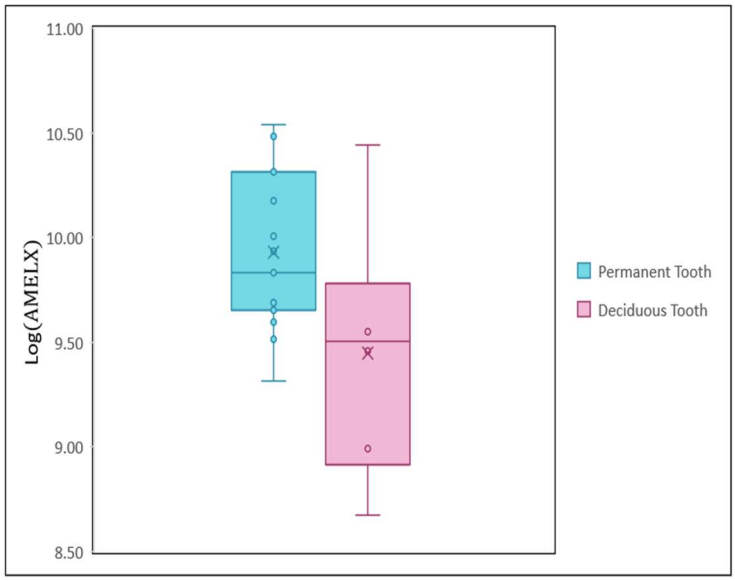

Figure 3. Expression of Amelogenin X-isoform in Permanent versus Deciduous Tooth of Remains from Ur, Iraq. The log transformed total X-isoform of amelogenin peptide signal intensity is shown for enamel from permanent tooth (n=15) compared to deciduous tooth (n=6) from individuals from Ur, Iraq.

The comparison of the AMELX peptides signal intensity between permanent teeth and deciduous teeth from Iraq individuals (**Figure 3.**) indicates that the amount of the protein recovered is generally more consistent and higher in permanent teeth than in deciduous teeth. This may reflect differences in the preservation of enamel proteins between the two types of teeth, possibly due to their differing structures and composition. The higher and more consistent levels of AMELX recovered from permanent teeth suggest that permanent teeth may preserve enamel proteins better over time. This could be attributed to the thicker and denser enamel in permanent teeth, which provides greater protection against environmental degradation factors like temperature, moisture, and soil chemistry. The wider range of AMELX recovered from deciduous teeth indicates less predictable preservation. Previous studies showed that enamel tends to be thinner in deciduous teeth compared to permanent teeth, which could potentially correlate to fewer enamel structural proteins such as amelogenin [42], [43]. Deciduous teeth are also associated with shorter functional lifespan, which could

make them more susceptible to degradation and loss of protein content in archaeological contexts.

3.2 Preservation of Enamel Proteins

Due to the remarkably similar climates of the collection sites of these enamel samples, we compare the results to enamel samples from a precontact period site near Pleasanton, California (CA-ALA-554), for our study on the preservation of enamel proteins. This site dates to between c. 100 BCE to 1500 CE, and thus somewhat contemporaneous with the sites in Iraq, Egypt, and the Sudan. The Pleasanton samples were previously analyzed to examine skeletal pathology phenotypes consistent with biological stress. The San Francisco Bay area is a relatively warm, temperate climate, classified as Csa by the Köppen-Geiger system [40], with mean annual temperature of 16.3 °C and a mean annual precipitation of 59.7 cm. Amelogenin (primarily AMELX due to its consistent presence in all samples) was compared among samples and sites in three ways: 3) normalized by sample weight in mg (**Figure 4.**) to observe the preservation of amelogenin in different environment conditions, 2) relative to other enamel structural proteins: ameloblastin and enamelin (**Figure 6., Figure 7.**) to examine differences in preservation of enamel structural proteins and 3) relative to serum albumin (**Figure 8.**) to examine differences in preservation between an enamel structural protein and a non-structural protein. The IQR (Interquartile Range) method of outlier detection was applied to all data sets prior to computing boxplots to identify and remove potential outliers. Uniprot (https://www.uniprot.org/) sequences and accession numbers used for identifying and comparing intensities of serum proteins in enamel samples were limited to the following: Q99217 AMELX_HUMAN for homo sapiens X isoform of amelogenin protein,

Q99218 AMELY_HUMAN for homo sapiens Y isoform of amelogenin protein, Q9NP70 AMBN_HUMAN for homo sapiens ameloblastin protein, Q9NRM1 ENAM_HUMAN for homo sapiens enamelin protein, and P02768 ALBU_HUMAN for homo sapiens albumin protein. Other sequences with different accession numbers for these same proteins were present in the database and were associated with fragmented or poorly annotated sequences. These were excluded from analysis.

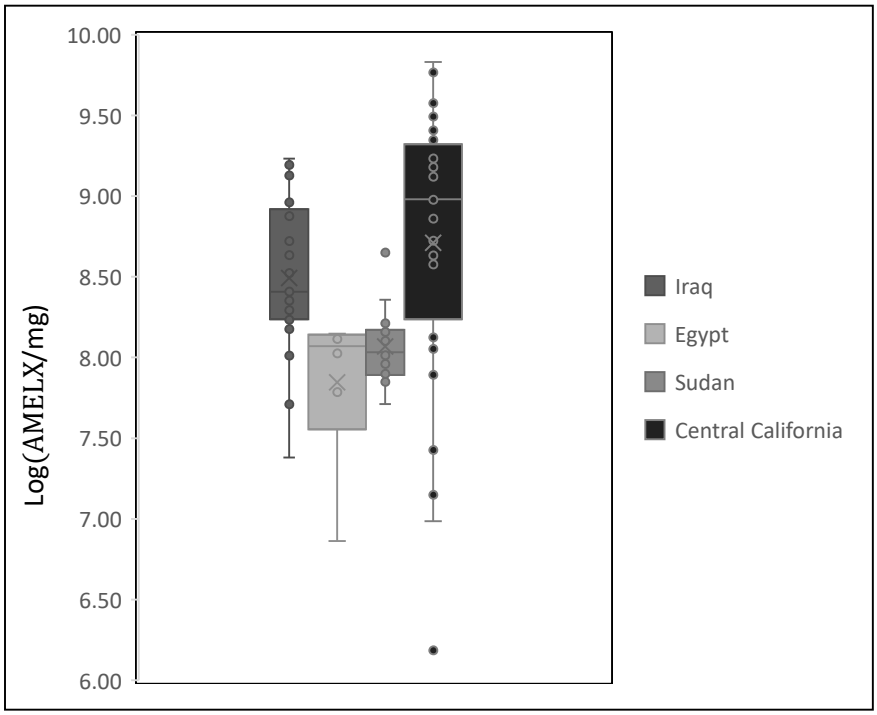

Figure 4. Expression of Amelogenin X-isoform Normalized by Enamel Mass. Box plot of log transformed total AMELX peptide signal intensity normalized by enamel mass in enamel samples from Ur, Iraq (n=21), Egypt (n=6), Sudan (n=14) and from CA-ALA-554 individuals (n=28)

Figure 4 reveals that the intensity of the X-isoform of amelogenin proteins, when normalized by enamel mass, is higher in samples from California than in those from more arid Iraq, Egypt, and Sudan. This finding introduces a new perspective on protein preservation in different environmental conditions. When normalized by enamel mass, the intensity of the X-isoform of amelogenin proteins for enamel from Egypt and Sudan individuals are relatively close; however, both are significantly lower than that of samples from Iraq. This unexpected result challenges our assumptions, as the remains from Iraq

were estimated to be older than the remains from Egypt and Sudan (Supplemental Table 1.). Since the climatic conditions of these three sites are relatively similar, this result might demonstrate the preservation power of amelogenin peptides in an archeological context, marking a significant advancement in our understanding of protein preservation.

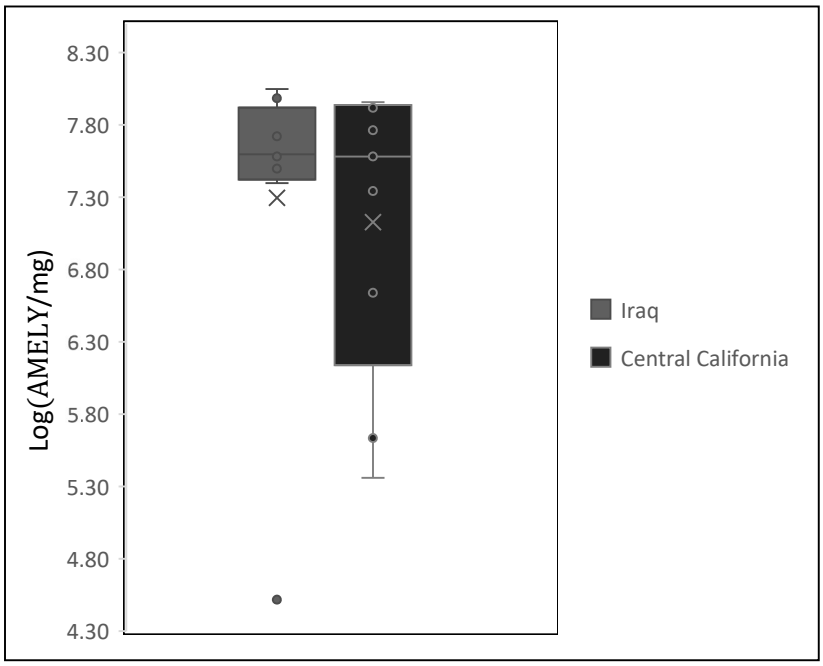

Figure 5. Expression of Amelogenin Y-isoform Normalized by Enamel Mass. Box plot of log transformed total AMELY peptide signal intensity normalized by enamel mass in enamel samples from Ur, Iraq (n=8) and from CA-ALA-554 individuals (n=9)

The Y-isoform of amelogenin proteins was only detected in some samples, and where they were detected, they were normalized by enamel mass. The purpose is to examine the preservation of the Y-isoform of amelogenin proteins. The result revealed that the intensity of the Y-isoform of amelogenin proteins, when normalized by enamel mass (**Figure 5.**), for individuals from Iraq is relatively comparable to that of individuals from the CA-ALA-554 burial site. It is observed that samples from Iraq seem to have lost a significant amount of serum albumin compared to samples from California. While more AMELX peptides were recovered, normalized by enamel, from California samples (**Figure 4.**), AMELY peptides did not show such a distinct difference between the individuals from

Iraq and California. This suggests that the X-isoform and Y-isoform of amelogenin proteins preserve or degrade differently. The differences in their amino acid sequences might lead to variations in their susceptibility to degradation under environmental stresses such as pH, temperature, and enzymatic activity. For instance, small sequence changes can influence their structure, i.e., how tightly these proteins are bound to the enamel matrix [44], potentially making one more susceptible to degradation than the other.

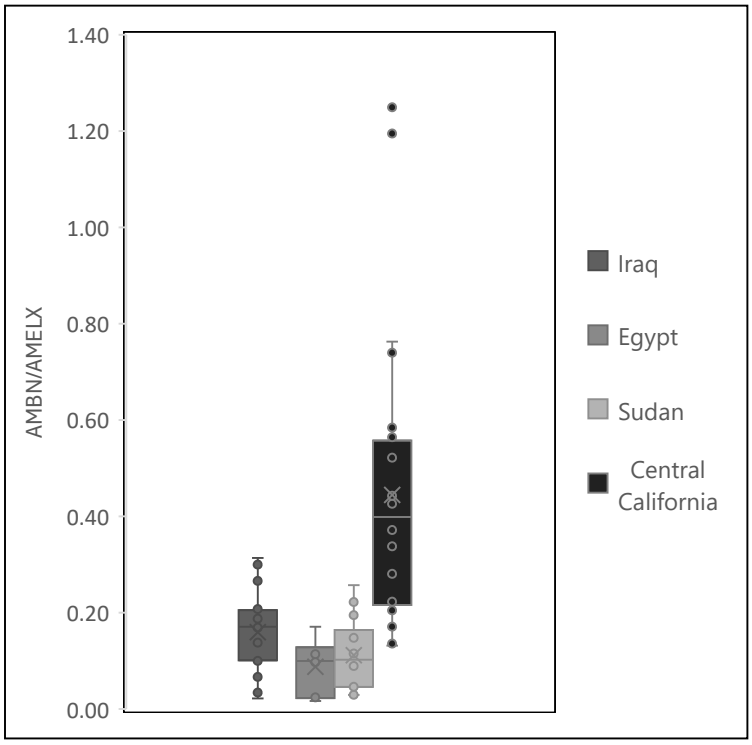

Figure 6. Expression of Ameloblastin Relative to Expression of Amelogenin X-isoform. Box plot of total ameloblastin (AMBN) signal intensity relative to amelogenin X-isoform (AMELX) peptide signal intensity enamel samples from Iraq (n=20), Egypt (n=6), Sudan (n=14) and from California (n=24)

Ameloblastin was detected in all enamel samples analyzed as well as the samples from Central California. The peptide signal of AMBN relative to AMELX is significantly lower in the Iraq, Egypt and Sudan enamel samples than those from the California site. The ratio of AMBN recovered to AMELX recovered are generally less than 0.80 for all samples from arid environment. Overall, the result suggests that ameloblastin degrades at a higher degree compared to amelogenin X- isoform and the

ameloblastin peptides were recovered at higher levels in tooth enamel from the California site compared to arid sites.

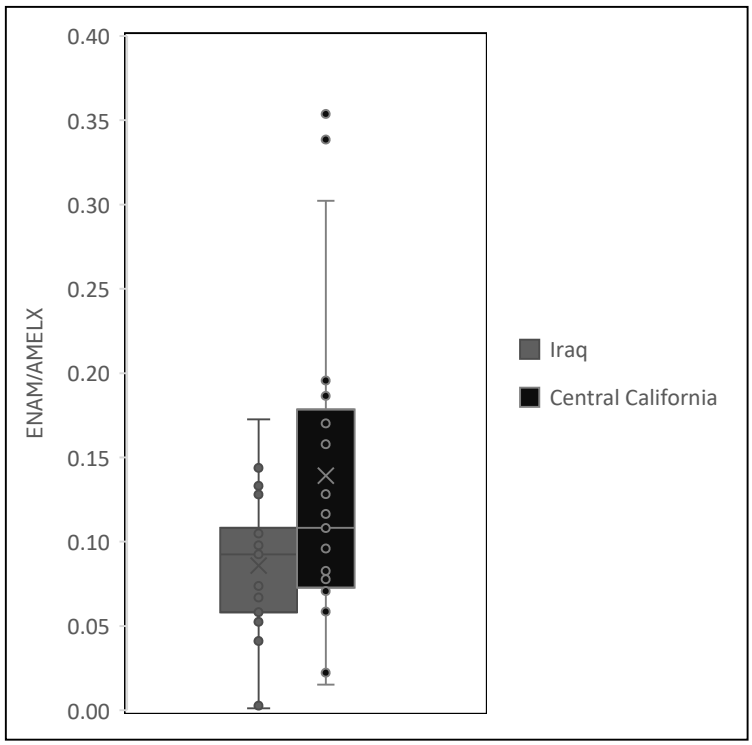

Figure 7. Expression of Enamelin Relative to Expression of Amelogenin X-isoform. Box plot of total enamelin (ENAM) signal intensity relative to amelogenin X-isoform (AMELX) peptide signal intensity enamel samples from Iraq (n=19), and from California (n=25)

Enamelin was not detected in all enamel samples. Out of 21 samples from Iraq, enamelin was only detected in 20 samples. Enamelin peptides were only detected in 1 out of 6 enamel samples from Egypt and in 3 out of 14 samples from Sudan. Due to the lack of data points for a box plot analysis, the comparison of enamelin peptides recovered relative to amelogenin peptides recovered was only performed on two populations: Iraq and Central California (**Figure 7.**). The peptide signal of ENAM relative to AMELX is significantly lower in the Iraq enamel samples than enamel samples from the California site. Hence, enamelin peptides are present at higher levels in tooth enamel from the California site. It is also observed that ENAM peptide intensity is much lower than the AMELX peptide intensity. The ratio ranged from 0.05 – 0.20 for Iraq samples and from 0.06 – 0.35 for Central California samples.

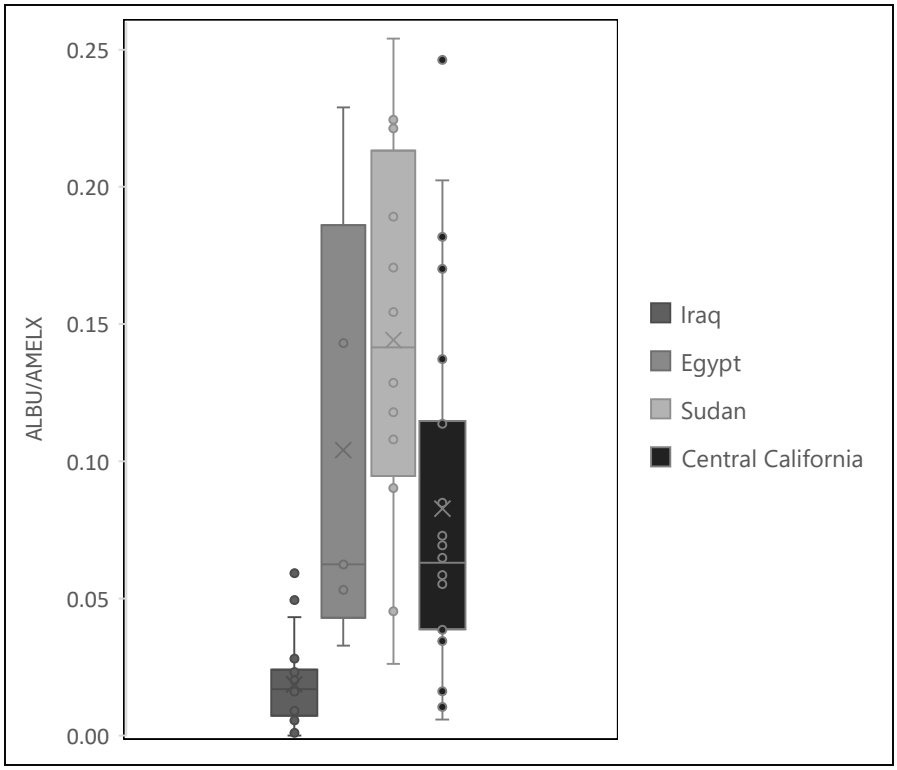

Figure 8. Expression of Serum Albumin Relative to Amelogenin X-isoform. Box plot of total serum albumin (ALBU) peptide signal intensity relative to amelogenin X-isoform (AMELX) peptide signal intensity for enamel samples from Iraq (n=20), Egypt (n=5), Sudan (n=12) and from California (n=24)

Serum albumin was detected in the majority of the enamel samples. Serum albumin peptides were detected in enamel samples at a lower abundance compared to the abundance of amelogenin, ameloblastin, and enamelin (**Figure 8.**). Serum albumin is notably lower among Iraq individuals. These individuals were also estimated to be slightly older in archeological age (2500-4500 BP) compared to individuals from Egypt, Sudan, and California. Compared to ameloblastin and enamelin, serum albumin peptides signal intensity is higher in enamel samples from Egypt and Sudan than in Iraq and Central California. This suggests that temperature and humidity are not the only factors affecting the preservation of serum albumin, as Egypt and Sudan have climate conditions similar to those of Iraq. Microenvironmental information regarding these burial sites must be considered for further analysis.

The AMELX intensity relative to AMBN, ENAM and ALBU suggested that 1) Amelogenin preserves better than ameloblastin, enamelin and serum albumin in ancient enamel samples, and 2) Serum albumin is most susceptible to degradation compared to enamel structural proteins. These observations align with the locations, roles of these proteins in enamel formation, and their relative abundances. Enamelin was found to be more fragile and less abundant than amelogenin, thus making it reasonably more challenging to detect in ancient teeth due to degradation. The significant difference in intensity between amelogenin and enamelin peptides can be due to the relative concentrations of these proteins in the enamel matrix, or there is a potential difference in how these proteins are preserved.

Serum albumin is more susceptible to degradation due to its lower chemical stability. In contrast, amelogenin and enamelin, essential for enamel development, are more resistant to degradation as they are more tightly integrated within the enamel structure. As a soluble protein, albumin is highly susceptible to proteolytic enzymes, which break down proteins [45]. Proteolytic activity from the body's enzymes and microbes targets soluble proteins like albumin after death. In living tissues like the pulp chamber of teeth, this degradation occurs soon after postmortem, limiting the chances of albumin preservation. Even in well-preserved teeth, serum albumin is not abundant in the tooth structure and is mainly found in the living tissues. Thus, after the loss of the soft tissue postmortem, serum albumin, if it remains, would be in minimal quantities.

Proteins are present in all body tissues, but their degradation happens at different rates throughout the body, mainly due to moisture, temperature and bacterial activity [2]. Based on the comparison between different archeological contexts with different climates,

it is observed that amelogenin, ameloblastin, enamelin, and serum albumin were better preserved in enamel samples from California. Several factors affect the preservation of archaeological dental materials: aridity, burial depth, and soil type [46]. The mineral fraction of bone, enamel, and dentine is composed of calcium phosphate and was generally found to be well preserved at arid sites [47] due to extremely low humidity. At the same time, in temperate and cold regions, proteins degrade generally more slowly, as degradation increases with temperature [2]. Thus, there is more likelihood of obtaining well-preserved enamel proteins. The California site, with its cooler climate (annual average temperature of 16.3 °C) than Iraq, Egypt, and Sudan (annual average temperature of 29–30 °C), while being more humid than the arid locations, provides a critical insight. Serum albumin, being part of soft, living tissues, degrades rapidly after death. Serum albumin is mostly found in blood plasma, which circulates throughout the body, making it exposed to external environmental factors (such as heat, moisture, and microbial activity) much more directly after death. These conditions accelerate its breakdown, making it challenging to detect preserved albumin in ancient remains unless the preservation conditions are ideal (e.g., cold and dry environment). In contrast, amelogenin, ameloblastin and enamelin proteins are protected within the enamel matrix, a highly mineralized tissue that is one of the hardest substances in the human body. This protective role of the enamel matrix is crucial in preserving these proteins, making them less susceptible to environmental degradation from moisture and microbial activity, and allowing fragments of these proteins to persist in ancient teeth. The comparison also suggests that the temperature of the burial sites might have a greater effect on protein preservation than precipitation.

4. Limitations and Future Work

Ancient DNA analysis was previously performed on some samples found in Sai Island and showed low potential for DNA recovery [46]. However, with advances in DNA extraction and quantification, ancient DNA analysis can be performed for a comparison of DNA quality and proteomic sex estimation using the same set of samples. The sample preparation was performed using enamel chunks and enamel powder. While they were intentionally sampled to exclude dentine, some samples were provided in powder form. Thus, it is possible that small amounts of dentine were present in one or more samples. It is important to note that any small amount of dentine contamination would be equivalent to more serum albumin present and less enamel mass in the data analysis. The teeth submitted for analysis were classified as permanent teeth. It was previously determined that permanent and deciduous teeth exhibit distinct similarities, including the existence of contrasting but distinct gradual changes in crystal disorder, phase abundance, crystallite size, and hardness [48]. Thus, despite limited studies showing that enamel proteins express differently in deciduous teeth, this study demonstrates that enamel proteins can be successfully extracted and analyzed using the same methods. The comparison between permanent and deciduous teeth would potentially help expand the application of proteomics for sex determination. This study is limited to four archeological sites in Iraq, Egypt, Sudan, and California. The Central California site, while offering a cooler yet more humid climate compared to the others, presents a challenge to protein preservation studies, as proteins typically degrade faster in hot and humid conditions. The microenvironment plays a crucial role in the preservation of proteins, as it directly influences the chemical and physical stability of these molecules. Factors such as

temperature, humidity, pH, and the presence of reactive substances can significantly affect the rate at which proteins degrade. For instance, higher temperatures and humid conditions accelerate hydrolysis and oxidation processes, leading to faster protein breakdown. Similarly, pH variations can destabilize protein structures by altering ionic interactions and hydrogen bonding. Rapid burial and a lower degree of exposure to temperature changes and other atmospheric processes have been shown to improve the preservation of remains [49]. Understanding the specific microenvironmental conditions surrounding a sample allows researchers to predict degradation pathways, develop preservation strategies, and interpret data accurately. Without considering these localized factors, studies on protein preservation risk overlooking critical variables that can compromise results. It is important to consider the type and method of burial for these remains, as they play a significant role in determining the conditions to which teeth samples were directly exposed. Other factors, like soil pH, soil porosity, soil moisture, vegetation, or sedimentation, can start changing archaeological remains even before weathering becomes severe [50] [51]. Archaeological samples from sites with more distinct climates (i.e., hot and humid climate versus cold and dry climate) can potentially confirm the effect of temperature and humidity on the preservation of enamel and enamel proteins. The key enamel protein in this study was amelogenins, hydrophobic proteins rich in proline, histidine, and glutamines. Deamidation of glutamine was previously proposed as an indicator of preservation quality and environmental conditions [52]. Thus, deamidation of glutamine within the amelogenin peptide sequence can be analyzed to compare the preservation quality between samples from different archeological contexts. However, further data analysis is needed to accurately determine deamidation due to

multiple different isoforms of amelogenin (X and Y-isoforms, AMELX-1,2,3 and AMELY-1,2) and the same deamidated peptides may not be present in all samples.

5. Conclusion

Proteomic analysis of human tooth enamel was performed on the skeletal remains of 41 individuals from different archeological contexts in Iraq, Egypt, and Sudan, with the youngest individual being a neonate (dental germ). Results revealed several expected enamel proteins, including amelogenin, ameloblastin, and enamelin, and other proteins like serum albumin, immunoglobulins, and collagen. This study focused on the presence of amelogenin and its X and Y isoforms. The presence of isoform Y of amelogenin helped estimate the biological sex of these remains. The study demonstrated the viability of using a proteomic approach under challenging conditions where genomic or osteological methods might fail. The detection of peptides specific to both X and Y chromosome isoforms of amelogenin reinforces the potential of this method. By focusing on sexually dimorphic amelogenin proteins in enamel, the research highlighted enamel's robustness as a tissue for proteomic analysis, even in environments unsuitable for ancient DNA preservation. The intensity of amelogenin X and Y, compared to the intensity of ameloblastin, enamelin and serum albumin provided information on the preservation quality of these proteins under different environmental factors. The study also compared the intensity of the X isoform of amelogenin and the Y isoform of amelogenin, normalized by enamel mass. The result suggests that the Y isoform of amelogenin degrades at a different rate than the X isoform of amelogenin. This could be due to the variations of amino acids within their peptide sequences. The study notes that serum albumin degrades more rapidly than enamel structural proteins like amelogenin, ameloblastin and

enamelin. The comparison of archaeological samples from hot, arid climate regions (Iraq, Egypt, and Sudan) to those from more temperate contexts (California) suggests the influence of environmental conditions on protein preservation, with enamel proteins showing higher degradation in hotter environments. In conclusion, this research expands the application of proteomics in sex estimation, proving it to be a robust tool in cases where traditional DNA or osteology may be limited due to environmental factors while also emphasizing the need for further studies on protein preservation across varying climatic conditions.

References

[1] G. J. Parker, H. E. McKiernan, K. M. Legg, and Z. C. Goecker, “Forensic proteomics,” *Forensic Sci. Int. Genet.*, vol. 54, p. 102529, Sep. 2021, doi: 10.1016/j.fsigen.2021.102529.

[2] C. Martin and F. Verheggen, “Odour profile of human corpses: A review,” *Forensic Chem.*, vol. 10, pp. 27–36, Aug. 2018, doi: 10.1016/j.forc.2018.07.002.

[3] Corey Harbison, “ABO Blood Type Identification and Forensic Science (1900-1960),” *Ariz. State Univ. Sch. Life Sci. Cent. Biol. Soc. Embryo Proj. Encycl.*, Jun. 2016.

[4] G. J. Parker *et al.*, “Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel,” *J. Archaeol. Sci.*, vol. 101, pp. 169–180, Jan. 2019, doi: 10.1016/j.jas.2018.08.011.

[5] R. P. De Beijer, C. De Graaf, A. Van Weert, T. G. Van Leeuwen, M. C. G. Aalders, and A. Van Dam, “Identification and detection of protein markers to differentiate between forensically relevant body fluids,” *Forensic Sci. Int.*, vol. 290, pp. 196–206, Sep. 2018, doi: 10.1016/j.forsciint.2018.07.013.

[6] E. D. Merkley, “Introduction to Forensic Proteomics,” in *ACS Symposium Series*, vol. 1339, E. D. Merkley, Ed., Washington, DC: American Chemical Society, 2019, pp. 1–8. doi: 10.1021/bk-2019-1339.ch001.

[7] T. Baumann, T. S. Carvalho, and A. Lussi, “The effect of enamel proteins on erosion,” *Sci. Rep.*, vol. 5, no. 1, p. 15194, Oct. 2015, doi: 10.1038/srep15194.

[8] A. Gil-Bona and F. B. Bidlack, “Tooth Enamel and Its Dynamic Protein Matrix,” *Int. J. Mol. Sci.*, vol. 21, no. 12, p. 4458, Jun. 2020, doi: 10.3390/ijms21124458.

[9] M. Jágr, P. Ergang, S. Pataridis, M. Kolrosová, M. Bartoš, and I. Mikšík, “Proteomic analysis of dentin–enamel junction and adjacent protein-containing enamel matrix layer of healthy human molar teeth,” *Eur. J. Oral Sci.*, vol. 127, no. 2, pp. 112–121, 2019, doi: 10.1111/eos.12594.

[10] M. Pandya and T. G. H. Diekwisch, “Amelogenesis: Transformation of a protein-mineral matrix into tooth enamel,” *J. Struct. Biol.*, vol. 213, no. 4, p. 107809, Dec. 2021, doi: 10.1016/j.jsb.2021.107809.

[11]James P. Simmer and Jan C.-C. Hu, “Expression, Structure, and Function of Enamel Proteinases: Connective Tissue Research: Vol 43, No 2-3.” Accessed: Oct. 22, 2024. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03008200290001159

[12] J. D. Bartlett, “Dental Enamel Development: Proteinases and Their Enamel Matrix Substrates,” *Int. Sch. Res. Not.*, vol. 2013, no. 1, p. 684607, 2013, doi: 10.1155/2013/684607.

[13] E. C. Salido, P. H. Yen, K. Koprivnikar, L. C. Yu, and L. J. Shapiro, “The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes,” *Am. J. Hum. Genet.*, vol. 50, no. 2, pp. 303–316, Feb. 1992.

[14] C. W. Gibson, “The Amelogenin Proteins and Enamel Development in Humans and Mice,” *J. Oral Biosci. JAOB Jpn. Assoc. Oral Biol.*, vol. 53, no. 3, pp. 248–256, 2011, doi: 10.2330/joralbiosci.53.248.

[15] E. K. Simpson, “Anthropology: Morphological Age Estimation,” in *Encyclopedia of Forensic and Legal Medicine*, Second Edition., Elsevier Ltd, 2015, pp. 189–195. doi: 10.1016/B978-0-12-800034-2.00025-2.

[16] I. M. Porto, H. J. Laure, R. H. Tykot, F. B. de Sousa, J. C. Rosa, and R. F. Gerlach, “Recovery and identification of mature enamel proteins in ancient teeth,” *Eur. J. Oral Sci.*, vol. 119, no. s1, pp. 83–87, Jan. 2012, doi: 10.1111/j.1600-0722.2011.00885.x.

[17] N. A. Stewart *et al.*, “The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction,” *RSC Adv.*, vol. 6, no. 66, pp. 61673–61679, Jun. 2016, doi: 10.1039/C6RA05120K.

[18] N. A. Stewart, R. F. Gerlach, R. L. Gowland, K. J. Gron, and J. Montgomery, “Sex determination of human remains from peptides in tooth enamel,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 52, pp. 13649–13654, Dec. 2017, doi: 10.1073/pnas.1714926115.

[19] G. J. Parker *et al.*, “Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel,” Aug. 2018, doi: 10.1016/j.jas.2018.08.011.

[20] S. Pääbo, “Preservation of DNA in ancient Egyptian mummies,” *J. Archaeol. Sci.*, vol. 12, no. 6, pp. 411–417, 1985, doi: 10.1016/0305-4403(85)90002-0.

[21] I. Marota, C. Basile, M. Ubaldi, and F. Rollo, “DNA decay rate in papyri and human remains from Egyptian archaeological sites,” *Am. J. Phys. Anthropol.*, vol. 117, no. 4, pp. 310–318, 2002, doi: 10.1002/ajpa.10045.

[22] R. Gowland *et al.*, “Sex estimation of teeth at different developmental stages using dimorphic enamel peptide analysis,” *Am. J. Phys. Anthropol.*, vol. 174, no. 4, pp. 859–869, Dec. 2020, doi: 10.1002/ajpa.24231.

[23] T. Buonasera *et al.*, “A comparison of proteomic, genomic, and osteological methods of archaeological sex estimation,” *Sci. Rep.*, vol. 10, no. 1, p. 11897, Jul. 2020, doi: 10.1038/s41598-020-68550-w.

[24] C. E. Smith, J. R. Pompura, S. Borenstein, A. Fazel, and A. Nanci, “Degradation and loss of matrix proteins from developing enamel,” *Anat. Rec.*, vol. 224, no. 2, pp. 292–316, Jun. 1989, doi: 10.1002/ar.1092240219.

[25] C. Robinson, H. D. Briggs, J. Kirkham, and P. J. Atkinson, “Changes in the protein components of rat incisor enamel during tooth development,” *Arch. Oral Biol.*, vol. 28, no. 11, pp. 993–1000, 1983, doi: 10.1016/0003-9969(83)90053-5.

[26] T. Aoba, T. Tanabe, and E. C. Moreno, “Function of amelogenins in porcine enamel mineralization during the secretory stage of amelogenesis,” *Adv. Dent. Res.*, vol. 1, no. 2, pp. 252–260, Dec. 1987, doi: 10.1177/08959374870010021401.

[27] N. C. Kegulian, G. Visakan, R. A. Bapat, and J. Moradian-Oldak, “Ameloblastin and its multifunctionality in amelogenesis: A review,” *Matrix Biol.*, vol. 131, pp. 62–76, Aug. 2024, doi: 10.1016/j.matbio.2024.05.007.

[28] J. Hatakeyama *et al.*, “Synergistic Roles of Amelogenin and Ameloblastin,” *J. Dent. Res.*, vol. 88, no. 4, pp. 318–322, Apr. 2009, doi: 10.1177/0022034509334749.

[29] S. Fukumoto *et al.*, “Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts,” *J. Cell Biol.*, vol. 167, no. 5, pp. 973–983, Dec. 2004, doi: 10.1083/jcb.200409077.

[30] H. Masuya *et al.*, “Enamelin (Enam) is essential for amelogenesis: ENU-induced mouse mutants as models for different clinical subtypes of human

amelogenesis imperfecta (AI)," *Hum. Mol. Genet.*, vol. 14, no. 5, pp. 575–583, Mar. 2005, doi: 10.1093/hmg/ddi054.

[31] N. Dohi *et al.*, "Immunocytochemical and immunochemical study of enamelins, using antibodies against porcine 89-kDa enamelin and its N-terminal synthetic peptide, in porcine tooth germs," *Cell Tissue Res.*, vol. 293, no. 2, pp. 313–325, Jul. 1998, doi: 10.1007/s004410051123.

[32] D. Fan, C. Du, Z. Sun, R. Lakshminarayanan, and J. Moradian-Oldak, "In vitro study on the interaction between the 32 kDa enamelin and amelogenin," *J. Struct. Biol.*, vol. 166, no. 1, p. 88, Apr. 2009, doi: 10.1016/j.jsb.2009.01.003.

[33] V. A. Perez, J. E. Mangum, and M. J. Hubbard, "Pathogenesis of Molar Hypomineralisation: Aged Albumin Demarcates Chalky Regions of Hypomineralised Enamel," *Front. Physiol.*, vol. 11, p. 579015, Sep. 2020, doi: 10.3389/fphys.2020.579015.

[34] A. G. Fincham, J. Moradian-Oldak, and J. P. Simmer, "The Structural Biology of the Developing Dental Enamel Matrix," *J. Struct. Biol.*, vol. 126, no. 3, pp. 270–299, Jun. 1999, doi: 10.1006/jsbi.1999.4130.

[35] C. Robinson, S. J. Brookes, J. Kirkham, W. A. Bonass, and R. C. Shore, "Crystal growth in dental enamel: the role of amelogenins and albumin," *Adv. Dent. Res.*, vol. 10, no. 2, pp. 173–179; discussion 179-180, Nov. 1996, doi: 10.1177/08959374960100020901.

[36] K. Okamura, "Localization of Serum Albumin in Dentin and Enamel," *J. Dent. Res.*, vol. 62, no. 2, pp. 100–104, Feb. 1983, doi: 10.1177/00220345830620020101.

[37] T. Buonasera *et al.*, “Immune proteins recovered in tooth enamel as a biochemical record of health in past populations: Paleoproteomic analysis of Mission Period Native Californians,” *J. Archaeol. Sci.*, p. 106069, Sep. 2024, doi: 10.1016/j.jas.2024.106069.

[38] J. W. Eerkens, A. de Voogt, T. L. Dupras, S. C. Rose, E. J. Bartelink, and V. Francigny, “Intra- and inter-individual variation in δ13C and δ15N in human dental calculus and comparison to bone collagen and apatite isotopes,” *J. Archaeol. Sci.*, vol. 52, pp. 64–71, Dec. 2014, doi: 10.1016/j.jas.2014.08.020.

[39] J. W. Eerkens, A. de Voogt, T. L. Dupras, V. Francigny, and A. M. Greenwald, “Early childhood diets on the Nile: δ13C and δ15N in serial samples of permanent first molars in an elite Meroitic population from Sai Island, Sudan,” *Int. J. Osteoarchaeol.*, vol. 28, no. 5, pp. 552–562, 2018, doi: 10.1002/oa.2679.

[40] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, “Present and future Köppen-Geiger climate classification maps at 1-km resolution,” *Sci. Data*, vol. 5, no. 1, p. 180214, Oct. 2018, doi: 10.1038/sdata.2018.214.

[41] J. P. Simmer, “Alternative Splicing of Amelogenins,” *Connect. Tissue Res.*, vol. 32, no. 1–4, pp. 131–136, Jan. 1995, doi: 10.3109/03008209509013715.

[42] F. e. Grine, “Enamel thickness of deciduous and permanent molars in modern Homo sapiens,” *Am. J. Phys. Anthropol.*, vol. 126, no. 1, pp. 14–31, 2005, doi: 10.1002/ajpa.10277.

[43] W. K. Seow, W. G. Young, A. K. L. Tsang, and T. Daley, “A Study of Primary Dental Enamel From Preterm and Full-term Children Using Light and Scanning Electron Microscopy,” *Pediatr. Dent.*, vol. 27, no. 5, pp. 374–379, Sep. 2005.

[44] C. Sotomayor-Vivas, E. Hernández-Lemus, and R. Dorantes-Gilardi, “Linking protein structural and functional change to mutation using amino acid networks,” *PLOS ONE*, vol. 17, no. 1, p. e0261829, Jan. 2022, doi: 10.1371/journal.pone.0261829.

[45] A. Fernández-San Millán, A. Mingo-Castel, M. Miller, and H. Daniell, “A chloroplast transgenic approach to hyper-express and purify Human Serum Albumin, a protein highly susceptible to proteolytic degradation,” *Plant Biotechnol. J.*, vol. 1, no. 2, pp. 71–79, 2003, doi: 10.1046/j.1467-7652.2003.00008.x.

[46] V. Francigny, H. Hollund, A. de Voogt, E. Altena, P. de Knijff, and C. Fallet, “Limits of Ancient DNA Extraction from Teeth: The Case of Sudanese Nubia,” *Nyame Akuma*, vol. 79, pp. 13–29, Jun. 2013.

[47] A. Zazzo and J.-F. Saliège, “Radiocarbon dating of biological apatites: A review,” *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, vol. 310, no. 1, pp. 52–61, Sep. 2011, doi: 10.1016/j.palaeo.2010.12.004.

[48] I. M. Low, N. Duraman, and U. Mahmood, “Mapping the structure, composition and mechanical properties of human teeth,” *Mater. Sci. Eng. C*, vol. 28, no. 2, pp. 243–247, Mar. 2008, doi: 10.1016/j.msec.2006.12.013.

[49] D. Todisco and H. Monchot, “Bone Weathering in a Periglacial Environment: The Tayara Site (KbFk-7), Qikirtaq Island, Nunavik (Canada),” *Arctic*, vol. 61, no. 1, pp. 87–101, 2008.

[50] A. B. Marín-Arroyo, B. González-Rabanal, J. L. Arteche, P. Fatás, and M. R. González Morales, “Bone weathering in an Atlantic environment: preliminary results of the Global Weathering Project in Spain,” *Quat. Sci. Adv.*, vol. 12, p. 100112, Oct. 2023, doi: 10.1016/j.qsa.2023.100112.

[51] H. Matthiesen, A. M. Høier Eriksen, J. Hollesen, and M. Collins, “Bone degradation at five Arctic archaeological sites: Quantifying the importance of burial environment and bone characteristics,” *J. Archaeol. Sci.*, vol. 125, p. 105296, Jan. 2021, doi: 10.1016/j.jas.2020.105296.

[52] E. R. Schroeter and T. P. Cleland, “Glutamine deamidation: an indicator of antiquity, or preservational quality?,” *Rapid Commun. Mass Spectrom.*, vol. 30, no. 2, pp. 251–255, 2016, doi: 10.1002/rcm.7445.

Supplemental Data

Supplemental Table 1. Enamel mass, archeological age, estimated age at death, tooth classification, peptides signal intensity of AMELX, peptides signal intensity of AMELY, probability of female Pr(F) and the estimated proteomic sex for enamel from individuals interred at Ur, Iraq, El-Hesa, Egypt, Tombos and Sai Island, Sudan.

Sample ID	Weight (mg)	Site	Burial	Archeological age	Age at Death	Tooth	AMELX	AMELY	Pr(F)	Proteomic Sex
TT001	22.5	Ur, Iraq	Lot 1168	630–540 BC	6-9 years	LRM1	2.06E+10	2.18E+09	0	Male
TT002	20.4	Ur, Iraq	Lot 3079	630–540 BC	6-8 years	ULM1	6.80E+09		0.92	Female
TT003	19.4	Ur, Iraq	241-2219	1845 to 1835 BC	early adult	ULM1	1.02E+10		0.94	Female
TT004	19.2	Ur, Iraq	Lot 5105	630–540 BC	neonate	Dental Germ	2.88E+09		0.87	Female
TT005	20.2	Ur, Iraq	324-3050	1845 to 1835 BC	ı-46 years (historical a	ULP2	4.95E+09		0.91	Female
TT006	19.5	Ur, Iraq	Lot 3980	630–540 BC	3-5 years	LRM1	3.04E+10		0.97	Female
TT007	20.1	Ur, Iraq	Lot 2791	2025–1760 BC	5-8 years	URM1	4.53E+09	8.22E+08	0	Male
TT008	19.9	Ur, Iraq	629-6115	1845 to 1835 BC	⁻72 years (historical a	LRM1	8.59E+09		0.94	Female
TT009	20.2	Ur, Iraq	Lot 2858	630–540 BC	5-9 years	Premolar	3.96E+09		0.89	Female
TT010	19.1	Ur, Iraq	WallQ_child2	2025–1760 BC	4-8 months	Unknown	9.83E+08	6.30E+05	0	Male
TT011	21.4	Ur, Iraq	Lot 2842	2025–1760 BC	4-6 months	ULdp1	3.66E+09		0.88	Female
TT013	20.3	Ur, Iraq	Lot 2941	630–540 BC	4-8 years	LRM1	3.46E+10		0.98	Female
TT014	19.9	Ur, Iraq	Lot 2634	630–540 BC	3-5 years	URM1	1.50E+10	1.05E+09	0	Male
TT015	20.7	Ur, Iraq	Lot 3314	2025–1760 BC	3-5 years	ULdp2	2.78E+10	2.31E+09	0	Male
TT016	19.2	Ur, Iraq	468-4164	1845 to 1835 BC	20-30 years	ULM1	4.91E+09	6.04E+08	0	Male
TT017	18.2	Ur, Iraq	385-3183	1845 to 1835 BC	3-5 years	LRM1	4.87E+09	6.96E+08	0	Male
TT018	19.8	Ur, Iraq	Lot 2861	630–540 BC	2-6 years	Canine	4.75E+08		0.64	Female
TT019	19.7	Ur, Iraq	Lot 2904	630–540 BC	10-12 years	URM1	3.11E+10		0.97	Female
TT020	20	Ur, Iraq	382-3227	1845 to 1835 BC	20-39 years	LRM1	2.06E+09		0.84	Female
TT021	18.9	Ur, Iraq	Lot 2812	2025–1760 BC	8-10 years	URM1	3.28E+09	4.73E+08	0.00	Male
TT022	19.6	Ur, Iraq	Lot 3008	2025–1760 BC	4-8 months	LRdi1	3.54E+09		0.89	Female
TT026	19.6	El Hesa, Egypt	VC 3189B	400 BC - AD 1000	Adult	LLM3	2.08E+09	9.23E+07	0	Male
TT027	19.4	El Hesa, Egypt	VL 3140B	400 BC - AD 1000	Adult	LRM3	2.72E+09	3.27E+07	0	Male
TT028	20	El Hesa, Egypt	VC 3102B	400 BC - AD 1000	Adult	LRM3	2.76E+09		0.87	Female
TT029	20.5	El Hesa, Egypt	VL 3122C	400 BC - AD 1000	Adult	LRM1	1.25E+09		0.78	Female
TT030	20	El Hesa, Egypt	VL 3184A	400 BC - AD 1000	Adult	ULMM2	1.46E+08	2.60E+06	0	Male
TT031	20.8	El Hesa, Egypt	VL 3151A	400 BC - AD 1000	Adult	LRM2	2.71E+09		0.86	Female
TT032	20.2	Tombos, Sudan	1U20Sh7	1450 - 300 BC	5 years	Unknown	1.43E+09		0.80	Female
TT033	20	Tombos, Sudan	4U38Sh2	1450 - 300 BC	12-15 years	Unknown	2.66E+09		0.86	Female
TT034	20.5	Tombos, Sudan	5U36Sh2	1450 - 300 BC	2-3 years	Unknown	2.60E+09	3.24E+08	0	Male
TT036	20.8	Sai Island, Sudan	T-02	AD 0 - 400	5 years	ULM1	1.07E+09		0.76	Female
TT037	20.9	Sai Island, Sudan	T-05	AD 0 - 400	12-15 years	URM1	2.35E+09		0.85	Female
TT038	19.8	Sai Island, Sudan	T-10	AD 0 - 400	6-7 years	ULM1	1.81E+09		0.83	Female
TT039	20.2	Sai Island, Sudan	T-11	AD 0 - 400	9-11 years	LLM1	3.29E+09	1.17E+08	0	Male
TT040	19.4	Sai Island, Sudan	T-17 (individual 1)	AD 0 - 400	4 years	URM1	1.54E+09		0.81	Female
TT041	20.1	Sai Island, Sudan	T-17 (individual 2)	AD 0 - 400	3 years	ULM1	4.59E+09	1.64E+08	0	Male
TT042	20	Sai Island, Sudan	T-20	AD 0 - 400	4 years	ULM1	1.47E+09		0.80	Female
TT043	20.2	Sai Island, Sudan	T-34	AD 0 - 400	young adult	LLM1	9.02E+09		0.94	Female
TT044	19.6	Sai Island, Sudan	T-35	AD 0 - 400	12-13 years	LLM1	2.82E+09		0.87	Female
TT045	19.9	Sai Island, Sudan	Te-01	AD 0 - 400	2-3 years	ULM1	2.05E+09		0.84	Female
TT046	19.8	Sai Island, Sudan	Te-12	AD 0 - 400	3-4 years	ULM1	1.94E+09		0.83	Female

Supplemental Table 2. Amelogenin X-isoform, ameloblastin, enamelin and serum albumin peptide intensity and their relative ratios for enamels from individuals interred in Iraq, Egypt, Sudan.

Sample ID	Site	AMELX	Ameloblastin	Enamelin	Serum Albumin	AMBN/AMELX	ENAM/AMELX	ALBU/AMELX
TT001	Ur, Iraq	2.06E+10	2.06E+09	1.60E+09	4.77E+08	0.10	0.08	0.02
TT002	Ur, Iraq	6.80E+09	1.31E+09	6.39E+08	1.56E+08	0.19	0.09	0.02
TT003	Ur, Iraq	1.02E+10	2.04E+09	9.44E+08	1.11E+08	0.20	0.09	0.01
TT004	Ur, Iraq	2.88E+09	9.80E+07	3.50E+06		0.03	0.00	
TT005	Ur, Iraq	4.95E+09	9.29E+08	6.33E+08	2.74E+07	0.19	0.13	0.01
TT006	Ur, Iraq	3.04E+10	1.40E+09	1.25E+09	2.30E+08	0.05	0.04	0.01
TT007	Ur, Iraq	4.53E+09	1.21E+09	6.03E+08	7.28E+07	0.27	0.13	0.02
TT008	Ur, Iraq	8.59E+09	1.56E+09	9.13E+08	6.11E+07	0.18	0.11	0.01
TT009	Ur, Iraq	3.96E+09	5.48E+08	3.88E+08	3.59E+07	0.14	0.10	0.01
TT010	Ur, Iraq	9.83E+08	1.70E+08	2.67E+06	1.99E+07	0.17	0.00	0.02
TT011	Ur, Iraq	3.66E+09	1.15E+09	3.96E+08	3.67E+06	0.31	0.11	0.00
TT013	Ur, Iraq	3.46E+10	3.66E+09	2.55E+09	9.72E+08	0.11	0.07	0.03
TT014	Ur, Iraq	1.50E+10	1.63E+09	1.12E+09	1.15E+09	0.11	0.07	0.08
TT015	Ur, Iraq	2.78E+10	1.87E+09	1.62E+09	1.20E+09	0.07	0.06	0.04
TT016	Ur, Iraq	4.91E+09	1.33E+09	7.06E+08	2.91E+08	0.27	0.14	0.06
TT017	Ur, Iraq	4.87E+09	1.01E+09	8.40E+08	3.93E+07	0.21	0.17	0.01
TT018	Ur, Iraq	4.75E+08	1.04E+07		2.35E+07	0.02		0.05
TT019	Ur, Iraq	3.11E+10	3.25E+09	2.08E+09	5.51E+08	0.10	0.07	0.02
TT020	Ur, Iraq	2.06E+09	6.19E+08	2.16E+08	4.95E+06	0.30	0.10	0.00
TT021	Ur, Iraq	3.28E+09	5.54E+08	6.34E+08	7.99E+07	0.17	0.19	0.02
TT022	Ur, Iraq	3.54E+09	1.38E+09	1.86E+08	6.29E+07	0.39	0.05	0.02
TT026	El Hesa, Egypt	2.08E+09	3.55E+08		2.98E+08	0.17		0.14
TT027	El Hesa, Egypt	2.72E+09	2.77E+08		8.91E+07	0.10		0.03
TT028	El Hesa, Egypt	2.76E+09	4.68E+07		1.72E+08	0.02		0.06
TT029	El Hesa, Egypt	1.25E+09	3.05E+07		6.64E+07	0.02		0.05
TT030	El Hesa, Egypt	1.46E+08	1.67E+07		2.21E+08	0.11		1.52
TT031	El Hesa, Egypt	2.71E+09	2.63E+08	3.44E+05	6.21E+08	0.10	0.00	0.23
TT032	Tombos, Sudan	1.43E+09	1.69E+08		2.71E+08	0.12		0.19
TT033	Tombos, Sudan	2.66E+09	3.94E+08		4.11E+08	0.15		0.15
TT034	Tombos, Sudan	2.60E+09	5.08E+08		6.81E+07	0.20		0.03
TT036	Sai Island, Sudan	1.07E+09	3.17E+07		2.37E+08	0.03		0.22
TT037	Sai Island, Sudan	2.35E+09	6.93E+07		4.01E+08	0.03		0.17
TT038	Sai Island, Sudan	1.81E+09	8.16E+07		2.33E+08	0.05		0.13
TT039	Sai Island, Sudan	3.29E+09	5.05E+08	1.71E+07	3.88E+08	0.15	0.01	0.12
TT040	Sai Island, Sudan	1.54E+09	3.96E+08		1.66E+08	0.26		0.11
TT041	Sai Island, Sudan	4.16E+09	2.19E+08	5.49E+05	1.06E+09	0.05	0.00	0.25
TT042	Sai Island, Sudan	1.47E+09	1.31E+08		9.12E+08	0.09		0.62
TT043	Sai Island, Sudan	9.02E+09	5.25E+08		8.14E+08	0.06		0.09
TT044	Sai Island, Sudan	2.82E+09	1.31E+08		6.33E+08	0.05		0.22
TT045	Sai Island, Sudan	2.05E+09	4.56E+08		9.28E+07	0.22		0.05
TT046	Sai Island, Sudan	1.94E+09	2.24E+08	2.46E+05	9.57E+08	0.12	0.00	0.49

Supplemental Table 3. Amelogenin X-isoform, ameloblastin, enamelin and serum albumin peptide intensity for enamels from individuals interred in the San Francico Bay, CA (CA-ALA-554) and their relative ratios.

Sample ID	Site	AMELX	Ameloblastin	Enamelin	Serum Albumin	AMBN/AMELX	ENAM/AMELX	ALBU/AMELX
JAS1	CA-ALA-554	5.57E+08	2.80E+09	1.68E+08	2.80E+08	5.03	0.30	0.50
JAS2	CA-ALA-554	2.89E+07	6.33E+08	3.86E+07	2.15E+08	21.90	1.34	7.45
JAS3	CA-ALA-554	2.94E+08	3.02E+09	1.10E+08	1.24E+08	10.29	0.37	0.42
JAS4	CA-ALA-554	1.57E+10	1.16E+10	1.83E+09	2.67E+09	0.74	0.12	0.17
JAS5	CA-ALA-554	8.41E+09	3.73E+09	1.28E+08	1.53E+09	0.44	0.02	0.18
JAS6	CA-ALA-554	3.38E+10	1.26E+10	4.46E+09	1.98E+09	0.37	0.13	0.06
JAS7	CA-ALA-554	1.94E+08	2.42E+08		7.48E+06	1.25		0.04
JAS8	CA-ALA-554	1.50E+09	8.44E+08	3.35E+07	1.06E+08	0.56	0.02	0.07
JAS9	CA-ALA-554	1.03E+10	1.23E+10	3.50E+09	1.19E+09	1.20	0.34	0.11
JAS10	CA-ALA-554	8.34E+09	6.36E+09	1.56E+09	9.48E+08	0.76	0.19	0.11
JAS11	CA-ALA-554	2.10E+09	8.94E+08	2.10E+08	4.24E+08	0.43	0.10	0.20
DM_70	CA-ALA-554	6.91E+09	3.71E+09	2.34E+09	4.12E+08	0.54	0.34	0.06
DM_71	CA-ALA-554	1.91E+10	1.12E+10	6.76E+09	6.58E+08	0.58	0.35	0.03
DM_77	CA-ALA-554	2.77E+09	8.45E+09	3.08E+08	4.80E+09	3.05	0.11	1.73
DM_78	CA-ALA-554	1.61E+10	4.76E+09	2.75E+09	2.22E+09	0.30	0.17	0.14
DM_79	CA-ALA-554	2.22E+10	4.75E+09	3.51E+09	1.36E+09	0.21	0.16	0.06
DM 080	CA-ALA-554	4.52E+10	1.01E+10	4.89E+09	3.84E+09	0.22	0.11	0.08
DM 081	CA-ALA-554	2.04E+10	3.64E+09	1.19E+09	8.13E+08	0.18	0.06	0.04
DM 082	CA-ALA-554	3.09E+10	7.05E+09	2.25E+09	1.71E+09	0.23	0.07	0.06
DM 083	CA-ALA-554	4.81E+10	9.87E+09	3.74E+09	2.80E+08	0.21	0.08	0.01
DM 084	CA-ALA-554	2.89E+10	3.93E+09	3.71E+09	1.87E+09	0.14	0.13	0.06
DM 085	CA-ALA-554	1.13E+11	4.95E+10	1.08E+10	7.83E+09	0.44	0.10	0.07
DM 086	CA-ALA-554	2.33E+10	1.22E+10	4.56E+09	1.70E+09	0.52	0.20	0.07
DM 087	CA-ALA-554	1.37E+11	4.64E+10	9.71E+09	3.38E+10	0.34	0.07	0.25
DM 091	CA-ALA-554	6.35E+10	8.38E+09	5.24E+09	6.66E+08	0.13	0.08	0.01
DM 092	CA-ALA-554	3.43E+10	5.87E+09	2.48E+09	1.28E+09	0.17	0.07	0.04
DM 093	CA-ALA-554	7.15E+10	2.01E+10	5.21E+09	2.80E+09	0.28	0.07	0.04
DM 094	CA-ALA-554	5.04E+10	2.21E+10	4.84E+09	8.17E+08	0.44	0.10	0.02