

Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel



Glendon J. Parker^{a,*}, Julia M. Yip^a, Jelmer W. Eerkens^b, Michelle Salemi^c,
Blythe Durbin-Johnson^d, Caleb Kiesow^e, Randall Haas^b, Jane E. Buikstra^f, Haagen Klaus^g,
Laura A. Regan^h, David M. Rocke^d, Brett S. Phinney^c

^a Department of Environmental Toxicology, University of California, Davis, Davis, CA, 95616, USA

^b Department of Anthropology, University of California, Davis, Davis, CA, 95616, USA

^c Proteomic Core Facility, University of California, Davis, Davis, CA, 95616, USA

^d Department of Public Health Sciences, University of California, Davis, CA, 95616, USA

^e Department of Biology, United States Air Force Academy, Colorado Springs, CO, 80840, USA

^f School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, 85287, USA

^g Department of Sociology and Anthropology, George Mason University, Fairfax, VA, 22030, USA

^h Office of Net Assessment, Department of Defense, Washington, DC, 20301, USA

ARTICLE INFO

Keywords:

Sex estimation

Teeth

Enamel

Amelogenin

Proteomics

Mass spectrometry

Paleoproteomics

ABSTRACT

Amelogenin genes are located on both X and Y sex chromosomes in humans and are a major focus of DNA-based sex estimation methods. Amelogenin proteins, AMELX_HUMAN and AMELY_HUMAN, are expressed in the tooth organ and play a major role in mineralization of enamel, the most taphonomically resistant, archaeologically persistent human tissue. We describe shotgun liquid chromatography mass spectrometry analysis of 40 enamel samples representing 25 individuals, including modern third molars and archaeological teeth from open-air contexts including permanent adult (400 to 7300 BP) and deciduous teeth (100 to 1000 BP). Peptides specific to the X-chromosome isoform of amelogenin were detected in all samples. Peptides specific to the sexually dimorphic Y-chromosome isoform were also detected in 26 samples from 13 individuals, across all time periods, including previously unsexed deciduous teeth from archaeological contexts. While the signal of each gene product can vary by more than an order of magnitude, we show close agreement between osteological and amelogenin-based sex estimation and thus demonstrate that the protein-based signal can be obtained reliably from open-air archaeological contexts dating to at least 7300 years ago. While samples with AMELY_HUMAN peptides are unambiguously male, samples with no AMELY_HUMAN signal may either be low signal male false negative samples or female samples. In order to estimate sex in these samples we developed a probability curve of female sex as a function of the logarithm of AMELX_HUMAN signal ($p < 0.0001$) using logistic regression. This is also the first demonstration using proteomics to estimate sex in deciduous teeth and pushes back the application of the method to teeth that are at least 7300 years old.

1. Introduction

Sex estimation represents one of the essential osteological variables documented in virtually any study of human remains. Sex estimation is fundamental in the bioarchaeological assessment of ancient demographic variation, paleoepidemiological reconstructions of ancient disease patterns, and differential diagnosis in paleopathology (Buikstra and Ubelaker, 1994; White et al., 2012). Currently, sex estimation relies

on two methods: analysis of sexually dimorphic osteological features and detection of DNA markers that are specific to the X- and Y-chromosome (Krishan et al., 2016; McFadden and Oxenham, 2016; Mittnik et al., 2016; Phenice, 1969; Skoglund et al., 2013; Spradley and Jantz, 2011). Though a range of skeletal elements have sexually dimorphic variation across ecogeographic populations (see Krishan et al., 2016 for a recent review), gross analysis of the os coxae are the most reliable. (Buikstra and Ubelaker, 1994; Spradley and Jantz, 2011; White et al.,

* Corresponding author. Department of Environmental Toxicology, University of California, Davis, Rm 4251B Meyer Hall, One Shields Ave., Davis, CA, 95616, USA.

E-mail addresses: gjparker@ucdavis.edu (G.J. Parker), jmyip@ucdavis.edu (J.M. Yip), jweerkens@ucdavis.edu (J.W. Eerkens), msaleti@ucdavis.edu (M. Salemi), bpdurbin@ucdavis.edu (B. Durbin-Johnson), caleb.kiesow@usuhs.edu (C. Kiesow), wrhaas@ucdavis.edu (R. Haas), buikstra@asu.edu (J.E. Buikstra), hklauas@gmu.edu (H. Klaus), laura.a.regan2.mil@mail.mil (L.A. Regan), dmrocke@ucdavis.edu (D.M. Rocke), bsp hinney@ucdavis.edu (B.S. Phinney).

2012). Osteological approaches are non-destructive and time efficient, although they are limited to late adolescents and adults who exhibit well-preserved secondary sexual characteristics of the skeleton. As well, there is a subset of skeletons with ambiguous or contrary traits and a resulting baseline of false positive or indeterminate sex assignment. Sexual dimorphism in osteological features increases as a result of hormonal and behavioral changes that occur during and after puberty (Juliano-Burns et al., 2009). For more than a century, accurate and reproducible estimation of sex from the subadult skeleton has been something of a ‘holy grail’ in physical anthropology. Additionally, sex can be impossible to assess from remains that are degraded by environmental and taphonomic processes (Waldron, 1987). Combined, subadult remains and skeletons lacking clear sexually dimorphic features represent a significant fraction of the osteological record (Gordon and Buikstra, 1981; Waldron, 1987).

An alternative method of sex estimation involves the detection of DNA markers from the X-and Y-chromosome (Loreille et al., 2018; Madel et al., 2016; Salido et al., 1992; Skoglund et al., 2013). Historically the most prominent methods rely on the detection of the amelogenin gene family, which has isoforms on both the X- and Y-chromosome (AMELX and AMELY) (Ballantyne, 2013; Garvin et al., 2012; Madel et al., 2016). Other methods measure the relative number of X- and Y-chromosome reads from massively parallel sequencing runs of DNA retained in samples (Loreille et al., 2018; Mitnik et al., 2016; Skoglund et al., 2013). Detection of these markers therefore can accurately infer the karyotype of sex chromosomes and can be applied in even highly compromised samples (Loreille et al., 2018; Mitnik et al., 2016). They are predicated however on the presence of at least small amounts of endogenous DNA template, which may be missing due to degradation of the relatively fragile DNA molecule (Allentoft et al., 2012; Lindahl, 1996; Ottoni et al., 2017).

Amelogenin genes are expressed as protein in the tooth organ (Fincham et al., 1991; Salido et al., 1992) and play a major role in the biosynthesis of enamel (AMELX_HUMAN and AMELY_HUMAN) (Kwak et al., 2016; Mazumder et al., 2014; Prajapati et al., 2016). The most characterized sex-specific genes are therefore expressed as protein in the most taphonomically resistant human tissue. Further, protein is considerably more robust than DNA molecules, particularly when it is physically associated with mineral surfaces such as in bone, enamel and ostrich egg shell (Buckley, 2016; Demarchi et al., 2016; Schweitzer et al., 2009; Wadsworth and Buckley, 2014). Amelogenin protein therefore has the potential to persist in the archaeological record and provide a reliable complementary, alternative, or superior method for sex estimation in skeletal assemblages (Porto et al., 2011a; Porto et al., 2011b; Stewart et al., 2016).

The last two decades have seen a revolution in the use of mass spectrometry to detect and identify proteins in biological samples (Doerr, 2013; Mallick and Kuster, 2010). The proteins in tissues or cells are broken down in a specific fashion using enzyme proteases; the resulting complex peptide mixtures are applied to liquid chromatography and eluted off the column using precise gradients of organic solvents. The peptides are converted from being in solution to gas phase using electrospray ionization, which gently removes the surrounding liquid and adds a positive charge (Aebersold and Mann, 2016). The mass over charge ratio of peptides can be measured with high accuracy in a mass spectrometer. Peptide ions can be further manipulated to fragment the peptide and measure the resulting masses. The fragmentation typically occurs at the slightly more brittle peptide bonds (Steen and Mann, 2004). The major result of this is that the spectrum of fragmentation masses has the potential to identify a given peptide sequence, and in turn provide evidence of a protein being present in a given sample (Aebersold and Mann, 2016; Nesvizhskii, 2010; Steen and Mann, 2004; Zhang et al., 2012). Current instruments are highly sensitive, accurate, and provide considerable information from even poorly preserved samples (Cappellini et al., 2012; Demarchi et al., 2016; Welker et al., 2015; Welker et al., 2016). For additional information the authors

recommend some informative and accessible reviews (Aebersold and Mann, 2016; Mallick and Kuster, 2010; Woods et al., 2014).

Recently, a method to extract amelogenin peptides from anatomically intact crowns was reported using a brief acid-etching protocol with minimal destructiveness (Stewart et al., 2016). In this breakthrough study, a marker peptide specific to each amelogenin protein was detected in enamel as a single mass in mass spectrometry, providing the first published examples of protein mass spectrometry as a means to sex individuals (Stewart et al., 2017; Stewart et al., 2016). Application of this method to archaeological materials obtained amelogenin peptides, and assigned a sex (rather than probability of sex) in 11 of 13 archaeological samples, ranging up to 5700 BP (Stewart et al., 2017). These findings were similar to a study that focused on demineralization and precipitation of enamel peptides using trichloroacetic acid (Castiblanco et al., 2015).

In this study, we contribute a new extraction technique and statistical framework for estimation of both male and female sex. We maximize the extraction of sexually dimorphic peptides through the use of destructive testing and the summation of all peptide signals that can be unambiguously attributed to either AMELY_HUMAN or AMELX_HUMAN. By comparing proteomic- and osteological-based determinations in 17 archaeological samples spanning over 7000 years, all from open-air depositional contexts, we show that characterization of amelogenin can yield reliable sex estimations even under poor preservation conditions. Confident detection of any AMELY_HUMAN specific peptide can be taken as an indicator of a male sex chromosome. Absence of a male specific signal however can be due to either a low signal sample, that is male, or female sex. We propose that the probability of female sex will increase with AMELX_HUMAN signal. A probability curve of confidently assigned sex, such as with self-identified modern teeth or archaeological material with DNA-based sex estimation, can be used to calculate probability of female sex ($Pr(F)$) as a function of AMELX_HUMAN signal.

2. Methods

2.1. Sample procurement and processing

Three enamel sample types were used in this study. First, modern third molars from individuals self-identified as male or female were collected as part of a study that sought, to document carbon, oxygen, strontium and lead isotopes across the continental United States in order to provenance human remains (Regan, 2006). The subjects were all born between 1964 and 1987 and were procured between 2005 and 2006. This initial study was determined to be exempt by the overseeing institutional review board (HQ USAFA IRB FAC2005026H) and found to be research, not human subject research, by the University of California - Davis institutional review board (IRB 97955-2). Second, archaeological teeth, primarily second and third molars, were collected from individuals with osteological sex estimations. These samples were from open-air contexts in central California and highland Peru, and range in age from 400 to at least 7300 BP (Eerkens et al., 2014; Eerkens et al., 2016a; Eerkens et al., 2016b; Eerkens et al., 2017; Greenwald et al., 2016; Haas et al., 2017; Haas and Viviano Llave, 2015). Third, deciduous teeth completely lacking in contextual sex information were also analyzed. These include three teeth from Mississippian sites (~1000 BP; provided by the Center of American Archeology), and one deciduous tooth from an historic Euroamerican burial ca. AD 1850) (Table 1). Enamel was dissected in 20–35 mg blocks using a diamond coated dental wheel (Brasseler USA, 918B.11.180 DS DIAM DISC) with an NSK Ultimate XL Micromotor. Where possible the dissected dental blocks included the dentin enamel junction (Mitsiadis et al., 2014).

2.2. Enamel protein extraction and digestion

To demineralize the samples (20–35 mg, Table 1), 200 µL of 1.2M

Table 1

Samples Utilized in Proteomic Sex Estimation Project. Samples processed for this project originated from Central Californian (CA-XX-XX), and Andean (Soro Mik'aya Patjxa) contexts, the United States Air Force Academy (AFA-XXX), and the Center of American Archeology (CAA-1). Burials, tooth position and type, sex estimation (M: male, F: female, U: unsexed), sex estimation method, estimated or dated sample age (BP), and number of analyses from each sample (n) are indicated.

Location	Burial	Tooth Position	Type	Sex	Method	Age (BP)	n
AFA-104		UR M3	Permanent	M	Self ID	modern	1
AFA-130		LL M3	Permanent	F	Self ID	modern	1
AFA-022		M3	Permanent	F	Self ID	modern	1
AFA-110		M3	Permanent	M	Self ID	modern	1
AFA-027		M3	Permanent	M	Self ID	modern	1
AFA-011		LL M3	Permanent	M	Self ID	modern	1
AFA-090		LL M3	Permanent	F	Self ID	modern	1
AFA-087		LR M3	Permanent	F	Self ID	modern	1
CA-ALA-554	85D	UR M3	Permanent	M	Osteology	1200	7
CA-Yolo-117	B2	UL M1	Permanent	M	DNA	400	1
CA-Yolo-117	B4	UR M1	Permanent	M	DNA	400	2
CA-ALA-554	B42	LR M3	Permanent	F	Osteology	1000	1
CA-SCL-928	B3	LR M3	Permanent	M	Osteology	6000	1
CA-SCL-919	B16	UR M3	Permanent	F	Osteology	400	1
CA-SJO-112	B35	Incisor	Permanent	F	Osteology	3180	1
CA-SJO-112	B13	LL M3	Permanent	M	Osteology	3160	2
CA-SJO-112	B36	ULM3	Permanent	M	Osteology	3180	1
Soro Mik'aya Patjxa	6	URM3	Permanent	F	Cranium	8000 - 6500	1
Soro Mik'aya Patjxa	11	ULM3	Permanent	F	Cranium	8000 - 6500	2
Soro Mik'aya Patjxa	9	LLM3	Permanent	M	Cranium	7465 - 7317	2
Soro Mik'aya Patjxa	12	UM3	Permanent	M	Cranium	8000 - 6500	1
CAA-1	JEB-4	LL M2	Deciduous	U		1000	2
CAA-1	JEB-	UR M2	Deciduous	U		1000	3
CAA-1	JEB-15	LR M1	Deciduous	U		1000	2
Carrolton, CA	B1	molar fragment	Deciduous	U		100	2

hydrochloric acid was added to each 2 mL sample vial with seven 2.8 mM ceramic beads (Omni-International Inc.), milled for 3 min at 7000 rpm in a MagnaLyzer (Roche Inc.), then centrifuged for 5 min at 16000 g. To reduce soluble proteins, 6 µL of 0.5M dithioerythritol (DTE) was added to each sample vial and incubated at 56 °C for 60 min. After incubation, 2M ammonium bicarbonate was added to each sample vial until the pH of the supernatant was 7.5 - 8.0. Alkylation was performed by adding 12 µL of 0.5M iodoacetamide to the sample vials and incubated in the dark at 25 °C for 60 min. The carbamidomethylation reaction was quenched by adding 12 µL of 0.5M DTE to the sample vials and incubated in room temperature for 5 min. After incubation, 0.01% ProteaseMAX (6 µL of 0.5% w/v, Promega Inc.) was added to the sample vials along with mass spectrometry grade trypsin (1 µL of 0.5 µg/µL, Thermo Pierce Inc.). Each sample was incubated at room temperature for 20 h at 300 rpm. After incubation, the sample vials were centrifuged for 5 min and 200 µL of the supernatant transferred to 0.22 µm centrifugal filters and centrifuged for 30 min. The filtrate was then transferred to clean Eppendorf® Protein LoBind tubes for sample clean up. Organic contaminants in aqueous stocks and solutions were removed by prior passage over solid phase extraction (SepPak, tC18, Waters Inc.).

2.3. Data acquisition

Digested peptides were desalted and concentrated using ZipTip C18 pipette tips (Millipore Inc.) with the eluted material lyophilized and resuspended in 2% (v/v) acetonitrile and 0.1% (v/v) TFA. The peptide concentration was measured using the Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Pierce™) and 1 µg of peptide, or 40% of the total sample if the sample was too diluted, was applied to mass spectrometry. The sample was applied to LC-MS/MS on a Thermo Scientific Q Exactive Plus Orbitrap mass spectrometer in conjunction with a Proxeon Easy-nLC II HPLC (Thermo Scientific) and Proxeon nanospray source. The digested peptides were loaded on a 100 µm x 25 mm Magic C18 100 Å 5U reverse phase trap where they were desalted online before being separated using a 75 µm x 150 mm Magic C18 200 Å 3U reverse phase column. Peptides were eluted using a

65 min gradient with a flow rate of 300 nL/min. An MS survey scan was obtained for the m/z range 300 - 1600, MS/MS spectra were acquired using a method that incorporated an inclusion list of 28 ions (Table A.1) that were subjected to HCD (High Energy Collisional Dissociation). When inclusion list ions were not found, MS/MS was done on other ions in the MS survey scan. An isolation mass window of 1.6 m/z was used for precursor ion selection, and normalized collision energy of 27% used for fragmentation. A 5 s duration was used for the dynamic exclusion. Washes were applied between each sample. After 10 samples a blank run of bovine serum albumin standards was applied to test for sample-to-sample contamination. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD009781 (<http://www.proteomexchange.org>) (Deutsch et al., 2017).

2.4. Bioinformatic processing

Mass spectrometry datasets (.RAW format) were processed using PEAKS™ (version 8.5) peptide matching software (Bioinformatics Solutions Inc., Waterloo, ON) (Zhang et al., 2012). The FASTA formatted UNIPROT *Homo sapiens* reference protein database (<http://www.uniprot.org/proteomes/UP000005640>) was modified to include additional FASTA protein entries of splice variants associated with AMELX_HUMAN (Q99217-1, -2, -3) and AMELY_HUMAN (Q99218-1, -2) gene products (Salido et al., 1992; Simmer, 1995). The reference database was further modified to incorporate a decoy database and was validated in PEAKS™ Software (Zhang et al., 2012). Peptide matching spectral assignment was conducted using default conditions with the following exceptions: error tolerance, precursor mass = 15 ppm, fragment ion = 0.5 Da, cleavage with trypsin with up to 2 missed cleavages, and up to two non-specific cleavages. The algorithm assumed all cysteines were carbamidomethylated, and the peptide was partially modified by deamidation (NQ), oxidation (MHW), pyroglutamate conversion from glutamate and glutamine, and methionine dioxygenation. All peptide assignments were filtered by a 1% false discovery rate. Each peptide was quantified by summing the ion peak intensity of each primary precursor mass (MS1).

2.5. Statistical analysis

AMELX_HUMAN and AMELY_HUMAN protein were quantified using label free quantitation by combining the peak ion intensity measurements (PEAKs™ software, version 8.5) of all peptides ($> 1\%$ FDR) specific to either protein and normalized for enamel mass (CI/mg) (Zhang et al., 2012). A value of 1 was added to all combined measurements prior to normalization to avoid a null value. All values were log transformed prior to analysis. Sex was modeled as a function of log AMELX_HUMAN and log AMELY_HUMAN using Firth's bias-reduced logistic regression (Agresti, 2013; Firth, 1993). Receiver operating characteristic (ROC) curves from models using log AMELY_HUMAN only, log AMELX_HUMAN only, and both log AMELY_HUMAN and log AMELX_HUMAN were compared using a resampling-based cross validation procedure in which, for each of 1000 resampling steps, 5 observations were randomly selected as a test set on which to evaluate the model as fitted on the remaining observations. Analyses were conducted using R, version 3.4.4 (R Core Team, 2018). Firth's bias-reduced logistic regression was implemented using the R package brglm, version 0.6.1 (Kosmidis, 2017). ROC curves were plotted using the R package ROCR, version 1.0–7, and confidence intervals for the cross-validated area under the ROC curve (cvAUC) were calculated using the R package cvAUC, version 1.1.0 (LeDell et al., 2014; Sing et al., 2005). The resulting logistic regression was used to determine the probability of female sex ($\text{Pr}(F)$) as a function of the logarithmic values of AMELX_HUMAN and AMELY_HUMAN signals. Probability values above 0.5 were estimated as female sex, and values below 0.5 estimated as male sex.

3. Results

3.1. The Y-chromosome isoform of amelogenin protein contains sexually dimorphic peptides

A bioinformatic analysis of the X- and Y-chromosome forms of amelogenin identifies numerous amino acid substitutions in the primary protein sequences (Fig. 1, blue squares). These differences increase in frequency after the inserted methionine at M59 in the Y-chromosomal form (AMELY_HUMAN, Q99217-3). The X- and Y-chromosomal protein forms also contain two and one additional sequences respectively as the result of splice variation (Fig. 1, Fig. A.1) (Salido et al., 1992; Simmer, 1995). An enamel sample from a skeleton with male osteological markers (1200 BP, CA-ALA-554, Burial 85D; Eerkens et al., 2016a) was chemically processed as described in the methods and analyzed by application to PEAKs™ peptide spectra matching software (Fig. 2). All unique peptide sequences that matched amino acid sequences in the sexually dimorphic Y-chromosome isoform of amelogenin (AMELY_HUMAN, Q99218-2) are indicated as light blue bars. All

peptides assignments were filtered by a 1% false discovery rate (Zhang et al., 2012). Peptides that contain amino acids specific to the AMELY_HUMAN sequence (dark blue rectangles) are an unambiguous indicator of a Y-chromosome. Many AMELY_HUMAN specific amino acids are represented by multiple peptide spectra. This is consistent with the high level of endogenous tissue protease activity from Kallikrein-4 (KLK4) and Matrix metalloproteinase-20 (MMP20) (Zhu et al., 2014). The ion peak intensity values of all precursor masses specific to the AMELY_HUMAN gene product were combined and normalized for enamel mass (CI/mg) (Table 2).

3.2. Proteomic analysis of enamel

Overall, 2337 ± 785 peptide spectrum matches (PSM), 178 ± 56 proteins, and 929 ± 256 unique peptides (peptide sequences) were observed in modern teeth; 2005 ± 1367 PSM, 102 ± 25 proteins, and 575 ± 329 unique peptides in archaeological adult teeth; and 788 ± 882 PSM, 100 ± 32 proteins, and 281 ± 274 unique peptides in archaeological deciduous teeth. The combined ion peak intensity (CI) attributable to amelogenin proteins was $20 \pm 14\%$ of the combined ion peak intensity values of all peptides in the dataset. When AMELY_HUMAN was detected, the ratio of AMELY_HUMAN to AMELX_HUMAN was $8.54 \pm 6.35\%$, which is consistent with the level of relative expression of 10% reported in the literature and 13% coefficient of linear least-squares regression reported below (Fincham et al., 1991).

3.3. Construction of amelogenin calibration curve

Using PEAKS™ software, the ion peak intensity values from precursor peptides specific for both AMELY_HUMAN and AMELX_HUMAN proteins were combined and normalized for enamel mass (CI/mg), and plotted in Cartesian space (Fig. 3). Samples identified as male (based on DNA, osteologically, or self-identification) are shown as blue squares and cluster away from female samples (plotted as red circles). The exceptions are two low-signal male enamel samples with no detectable AMELY_HUMAN signal and one sample with a high AMELX_HUMAN signal but no detectable AMELY_HUMAN signal (Soro Mik'aya Patjxa, Burial 12). Eight of nine of the unsexed deciduous molars (black diamonds) plot with the main group of male samples. The four originating juvenile skeletons can therefore be estimated as male sex.

A linear line of best fit was calculated using all samples with a detectable AMELY_HUMAN signal ($y = 0.130x - 4.4 \times 10^6$, $r = 0.939$, $DF = 24$, $p < 0.0001$). The slope of the line ($a = 0.13$) indicates that the Y-chromosome form was expressed at around one eighth of the level of the AMELX gene in male enamel, consistent with measured transcription levels (Fincham et al., 1991). The range of AMELY_HUMAN and AMELX_HUMAN signal varies over 3.2 and 3.7 orders of magnitude respectively.

	AA#
X MG TW I L F A C L L G A A F A M P L P P H P G H P G Y I N F S Y E N S H S Q A I N V D R T A L V L T P L K W Y Q S - I	59
Y MG TW I L F A C L V G A A F A M P L P P H P G H P G Y I N F S Y E N S H S Q A I N V D R I A L V L T P L K W Y Q S M I	60
X R P P Y P S Y G Y E P M G G W L H H Q I I P V L S Q Q H P P T H T L Q P H H H I P V V P A Q Q P V I P Q Q P M M P V P G	119
Y R P P Y S S Y G Y E P M G G W L H H Q I I P V V S Q Q H P L T H T L Q S H H H I P V V P A Q Q P R V R Q Q A L M P V P G	120
X Q H S M T P I Q H H Q P N L P P P A Q Q P Y Q P Q P V Q P Q P H Q P M Q P Q P P V H P M Q P L P P Q P P L P P M F P M Q	179
Y Q Q S M T P T Q H H Q P N L P L P A Q Q P F Q P Q P V Q P Q P H Q P M Q P Q P P V Q P M Q P L L P Q P P L P P M F P L R	180
X P L P P M L P D L T L E A W P S T D K T K R E E V D	205
Y P L P P I L P D L H L E A W P A T D K T K Q E E V D	206

Fig. 1. Amino Acid Sequence of Amelogenin Proteins. The amino acid sequence of the X-chromosome specific amelogenin protein (isoform 3; AMELX_HUMAN, Q99217-3) and corresponding Y-chromosome specific protein (isoform 2; AMELY_HUMAN, Q99218-2) are shown. Amino acid variations between the two paralogous sequences are indicated in blue with trypsin protease cleavage sites indicated with thin vertical lines. The excised signal peptide is indicated by green (amino acids 1 to 17). Regions excised through the action of splice variation are indicated in bold rectangles (Q99217-1, Q99217-2, Q99218-1). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

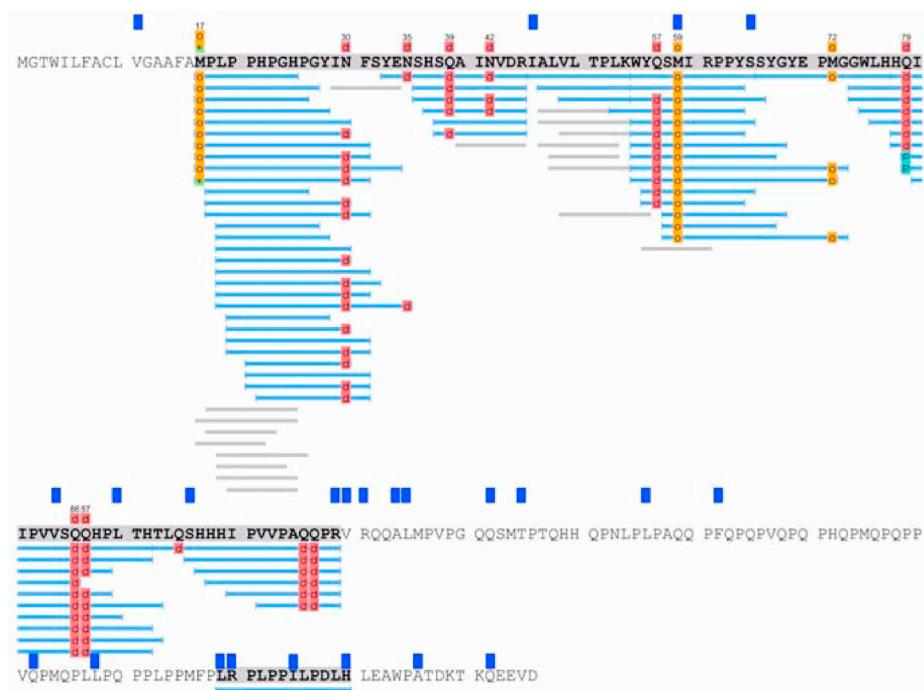


Fig. 2. Multiple Sexually Dimorphic Peptides Can be Detected in Archaeological Male Enamel. An enamel sample from a skeleton with male osteological markers (~1000 BP, CA-ALA-554 B85D) was processed by acid demineralization, trypsinization and tandem mass spectrometry as described in the Methods. Data was processed by application to PEAKs™ peptide spectra matching software. All unique peptide sequences that matched amino acid sequences in the sexually dimorphic Y-chromosome isoform of amelogenin (AMELY_HUMAN, Q99218-2) are indicated as light blue bars. Amino acids specific to the AMELY_HUMAN sequence are indicated by dark blue rectangles. Chemical modifications such as methionine oxidation (yellow squares), deamidation (red squares), and pyro-Glu modification (teal squares) are included. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.4. Use of a logistic regression model to estimate female sex

Detection of peptides from the AMELY_HUMAN gene product is an unambiguous indicator of male sex. The lack of AMELY_HUMAN signal however may be due to either female sex or low signal male samples with no detectable AMELY_HUMAN peptides, either the result of endogenous biological variation or exogenous taphonomic or environmental processes. The probability of a male false negative sex estimate will decrease, and female true positive estimate increase, as the corresponding AMELX_HUMAN signal increases.

To establish a statistical framework for female sex estimation, a probability curve was constructed using data from modern samples and samples with sex estimates based on DNA methodology ($n = 11$, Fig. 4A). Logarithmic values were calculated for both the AMELY_HUMAN and AMELX_HUMAN combined ion peak intensity signal (CI/mg). The logarithmic values of the AMELX_HUMAN cumulative ion peak intensity signal (CI/mg) from samples with no AMELY_HUMAN signal from juvenile skeletons or samples with sex estimated from osteological markers were used to calculate the probability of female sex ($\text{Pr}(F)$) based on the logistic regression (Fig. 4B). The logistic regression of the sex estimation curve using logarithmic values, resulted in the following formula for $\text{Pr}(F)$ given AMELY_HUMAN = 0:

$$\ln(\text{Pr}(F)/(1 - \text{Pr}(F))) = -12.8622 + 0.7496 * \ln(\text{AMELX_HUMAN} + 1)$$

Samples with no detectable AMELY_HUMAN peptides, and an AMELX_HUMAN cumulative peak ion intensity above 5.31×10^8 CI/mg have a greater than 90% probability of having female sex ($\text{Pr}(F)$) (Fig. 4B).

In order to evaluate the predictive capabilities of the above model, a receiver operator curve was constructed after randomly selecting 5 values as a test set, and then comparing the predicted values with the actual values, and repeating this procedure 1000 times. Resulting receiver operating (ROC) curves are included in the appendix (Figs. A.2 and A.3). The cross-validated area under the ROC curve (cvAUC) when using signal from both proteins was calculated as 0.9380 (Appendix Material: A.1 Results).

Estimation of sex using the above protein-based analytical framework was directly compared to sex estimation using osteological

methodology (Table 2). Male samples, defined as having detectable levels of AMELY_HUMAN specific peptides had a null probability of female sex ($\text{Pr}(F) = 0$). Unsexed samples and samples with osteological sex estimation with no detectable AMELY_HUMAN peptides were applied to the logistic regression curve and probability of female sex calculated ($\text{Pr}(F)$). Samples with a $\text{Pr}(F)$ value greater than 0.5 were assigned female sex, samples with a probability of less than 0.5 were assigned male sex. There were three examples of the latter in this sample cohort, two samples (JY2 and JY15) had replicate samples with AMELY_HUMAN peptides and unambiguous male assignment. The remaining sample JY5 from Burial 3 at site SCL-928 (California) had a $\text{Pr}(F)$ value of 0.16 and corresponding probability of male sex ($1 - \text{Pr}(F)$) of 0.84. There was one point of difference with osteological sex estimation, sample JY63 from Burial 12 at Soro Mik'aia Patjxa (Peru, 8000–6500 BP) (Table 2) (Haas et al., 2017). The osteological estimation of male sex in this case is based on secondary sex characteristics of the robust mastoid process and supraorbital margin that by itself has a higher false positive rate (Haas et al., 2017). The probability of female sex in this sample was calculated at $\text{Pr}(F) = 0.96$. Each unsexed deciduous tooth from a juvenile skeleton (JY12 - 19, 65, 66) was unambiguously assigned as male sex ($\text{Pr}(F) = 0.00$, Table 2).

3.5. Persistence of amelogenin peptides over an archaeological time frame

Quantitative data for peptides specific to AMELY_HUMAN (Y) and AMELX_HUMAN protein (X) were plotted for each sample as a function of archaeological time (Fig. 5). Male (blue square) samples contained peptides specific to the sexually dimorphic AMELY_HUMAN gene product (Y) in 18 of 21 measurements from 11 individuals. Samples without detectable AMELY_HUMAN peptides included samples with estimated female sex (red circles) and low signal male sex or unsexed deciduous samples (black diamond). These samples are included in the analysis due to the addition of a single unit to prevent a null result. All samples measured contained at least some peptides specific to the AMELX_HUMAN gene product. While the signal of each gene product can vary by an order of magnitude or greater at each time point, the degree signal is persistent over archaeological time. Indeed, the oldest samples (ca. 7300 BP) are among those with the greatest signal.

Table 2

Comparison of Other Methods with Protein-Based Sex Estimation. Enamel samples (location, burial) from modern teeth, and archaeological adult and deciduous teeth were processed and the level of AMELX_HUMAN (X) and AMELY_HUMAN (Y) protein quantified by combining peak intensities from mass spectrometry and normalizing against enamel mass (CI/mg). Male sex (M) was estimated based on the detection of AMELY_HUMAN protein (Male, M). In order to estimate probability of female sex (Pr(F)), samples with known sex (DNA-based estimation and modern samples) were used to calculate a logistical regression curve. In samples lacking AMELY_HUMAN signal (signal < 1), unsexed samples and samples with osteological sex estimation were applied to the logistical regression curve and probability of female sex (Pr(F)) calculated. These are compared to sex estimates (male, M; female, F; unsexed, U) obtained using osteological methods.

Description	Amelogenin Signal				Sex Estimation		DNA/Osteology		
	Sample	Location	Burial	Signal (CI/mg)		Protein-based		Sex	Method
				X	Y	Sex	Pr(F)		
JY24	AFA-104			3.13×10^8	2.58×10^7	M	0.00	M	Self ID
JY25	AFA-130			3.81×10^8	1.27×10^{-1}	F	0.87	F	Self ID
JY26	AFA-022			3.12×10^9	1.24×10^{-1}	F	0.97	F	Self ID
JY27	AFA-110			4.41×10^8	1.19×10^8	M	0.00	M	Self ID
JY33	AFA-027			1.57×10^8	2.51×10^6	M	0.00	M	Self ID
JY34	AFA-011			2.18×10^8	2.07×10^7	M	0.00	M	Self ID
JY35	AFA-090			1.90×10^8	1.58×10^{-1}	F	0.80	F	Self ID
JY36	AFA-087			5.84×10^8	1.28×10^{-1}	F	0.90	F	Self ID
85D_1	CA-ALA-554	B85D		2.22×10^9	3.08×10^8	M	0.00	M	Osteology
85D_2	CA-ALA-554	B85D		1.55×10^9	2.85×10^8	M	0.00	M	Osteology
85D_3	CA-ALA-554	B85D		3.12×10^9	3.45×10^8	M	0.00	M	Osteology
85D_4	CA-ALA-554	B85D		2.93×10^9	4.88×10^8	M	0.00	M	Osteology
85D_5	CA-ALA-554	B85D		4.12×10^9	5.64×10^8	M	0.00	M	Osteology
85D_6	CA-ALA-554	B85D		2.54×10^9	3.52×10^8	M	0.00	M	Osteology
85D_7	CA-ALA-554	B85D		3.12×10^9	1.90×10^8	M	0.00	M	Osteology
JY3	CA-ALA-554	B42		1.36×10^8	6.17×10^{-2}	F	0.76	F	Osteology
JY1	Yolo-117	B2		1.50×10^8	5.60×10^6	M	0.00	M	DNA
JY2	Yolo-117	B4		7.21×10^6	6.92×10^{-2}	M	0.26	M	DNA
JY30	Yolo-117	B4		5.89×10^7	2.24×10^6	M	0.00	M	DNA
JY5	SCL-928	B3		3.03×10^6	8.76×10^{-2}	M	0.16	M	Osteology
JY6	SCL-919	B16		3.99×10^8	8.00×10^{-2}	F	0.88	F	Osteology
JY7	SJO-112	B35		1.04×10^8	7.79×10^{-2}	F	0.72	F	Osteology
JY8	SJO-112	B13		4.02×10^7	1.19×10^6	M	0.00	M	Osteology
JY31	SJO-112	B13		6.18×10^7	3.01×10^6	M	0.00	M	Osteology
JY32	SJO-112	B36		2.52×10^8	2.14×10^7	M	0.00	M	Osteology
JY57	Soro Mik'aya Patjxa	B6		1.59×10^{10}	1.15×10^{-1}	F	0.99	F	Cranial
JY59	Soro Mik'aya Patjxa	B11		5.58×10^9	1.01×10^{-1}	F	0.98	F	Cranial
JY60	Soro Mik'aya Patjxa	B11		5.47×10^8	1.14×10^{-1}	F	0.90	F	Cranial
JY61	Soro Mik'aya Patjxa	B9		2.49×10^8	1.84×10^7	M	0.00	M	Cranial
JY62	Soro Mik'aya Patjxa	B9		3.64×10^8	2.88×10^7	M	0.00	M	Cranial
JY63	Soro Mik'aya Patjxa	B12		2.34×10^9	4.80×10^{-2}	F	0.96	M	Cranial
JY12	CAA-1	JEB-4		2.08×10^8	1.57×10^7	M	0.00	U	N/A
JY13	CAA-1	JEB-4		1.04×10^8	1.46×10^7	M	0.00	U	N/A
JY15	CAA-1	JEB-		2.49×10^7	7.69×10^{-2}	M	0.47	U	N/A
JY16	CAA-1	JEB-		1.05×10^8	5.13×10^6	M	0.00	U	N/A
JY17	CAA-1	JEB-		1.20×10^8	3.49×10^6	M	0.00	U	N/A
JY18	CAA-1	JEB-15		8.18×10^7	1.58×10^6	M	0.00	U	N/A
JY19	CAA-1	JEB-15		1.55×10^8	5.50×10^6	M	0.00	U	N/A
JY65	Carrolton, CA	B1		2.76×10^7	3.41×10^5	M	0.00	U	N/A
JY66	Carrolton, CA	B1		5.81×10^8	2.10×10^7	M	0.00	U	N/A

3.6. Protein deamidation over archaeological time

The side chains of the amino acids glutamine and asparagine can change over time to form glutamate and aspartate (Clarke, 1987). This process, called deamidation, is accelerated in environments of extreme pH, higher temperatures, and the level of movement permitted at each amino acid residue (Clarke, 1987; Geiger and Clarke, 1987; Robinson, 2002; Tyler-Cross and Schirch, 1991; Wilson et al., 2012). Protein movement is highly restricted when proteins bind to mineralized surfaces (Demarchi et al., 2016; van Doorn et al., 2012; Wilson et al., 2012). Analysis of AMELX_HUMAN peptides in older enamel samples showed almost complete deamidation at glutamine 57 (Q57) compared to modern samples (Fig. 6A). The rate of deamidation at Q57 was quantified by combining the peak intensity of ion currents (CI) for each peptide containing either glutamine 57 (Q57) or glutamate 57 (E57), calculating the proportion of glutamate 57 bearing peptides, and then plotting as a function of time (Fig. 6B). Based on a linear least squares curve fit, deamidation occurred at rate of 1% every 305 years ($R = 0.53$, $p < 0.0001$).

4. Discussion

This study demonstrates the use of amelogenin peptide measurement from enamel to estimate sex. The presence of peptides specific to the Y-chromosome form of amelogenin protein (AMELY_HUMAN) is an unambiguous indicator of male sex. Since the absence of AMELY_HUMAN may be due to female sex or a low signal male sample, female sex should be expressed as a probability (Pr(F)), which will increase with the quantity of peptides specific to the X-chromosome form of amelogenin protein (AMELX_HUMAN). We demonstrate that this framework can be used to estimate sex of subadult teeth. The detection of amelogenin peptides is robust with signal being detected over a 7300 year time frame. This is in spite of a wide range of amelogenin signal. Based on this approach we identified one sample with potential false sex estimation based on osteological markers. Proteomic sex estimation concludes that an individual is likely a female instead of a male ($Pr(F) = 0.96$, Table 2).

The most characterized sex-specific genes, the amelogenin gene family, are expressed as protein in the most robust and taphonomically

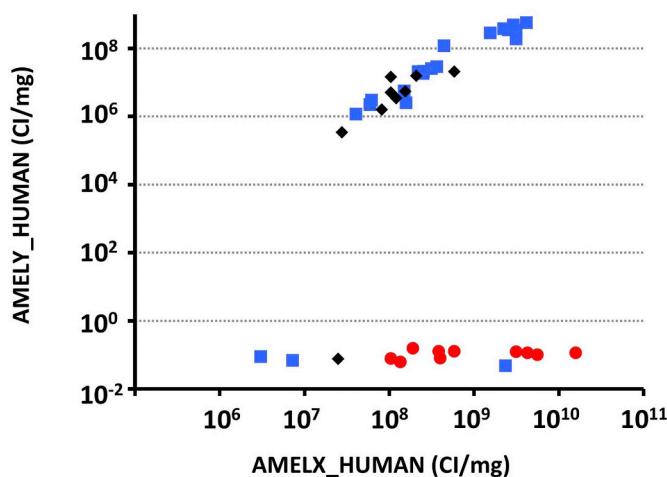


Fig. 3. Specific Amelogenin Signals in Enamel Samples. Precursor peptides with assigned sequences specific to either AMELX_HUMAN and AMELY_HUMAN were measured for the combined peak intensities for each protein and normalized for enamel mass (CI/mg). A single unit was added to each combined measurement to prevent null values. Samples from skeletons with a prior male sex estimation, female sex estimation and unsexed samples are indicated as blues squares, red circles and black diamonds respectively. In samples containing a detectable AMELY_HUMAN signal a least-squares line of linear best-fit was calculated at $y = 0.130x - 4.4 \times 10^6$, $r = 0.939$, $p < 0.0001$. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

persistent tissue in the human body. It is logical that recent studies have raised and then demonstrated the potential for using the Y-chromosome form of amelogenin as a means of sexing enamel samples (Porto et al., 2011a; Porto et al., 2011b; Stewart et al., 2016). Three recent studies have successfully identified peptides from AMELX_HUMAN in archaeological material in samples up to 5700 BP (Stewart et al., 2016, 2017). These innovative studies focus on detection of single masses consistent with characterized peptides from each amelogenin isoform (Stewart et al., 2017; Stewart et al., 2016). Detection of both peptides is taken as an indicator of male sex, the detection of AMELX_HUMAN peptide alone is used to determine female sex. Another innovative aspect of these studies is the minimal impact to the enamel by use of surface demineralization, with brief acid-etching, without engaging in more extensive destruction of the tooth or loss of gross anatomical features.

This project differs from the recently published work by Stewart et al., in four aspects. Firstly, the extraction of peptides from each amelogenin protein is maximized through destructive analysis and more extensive chemical extraction through milling and longer treatment with acid demineralization. Secondly, the resulting detected

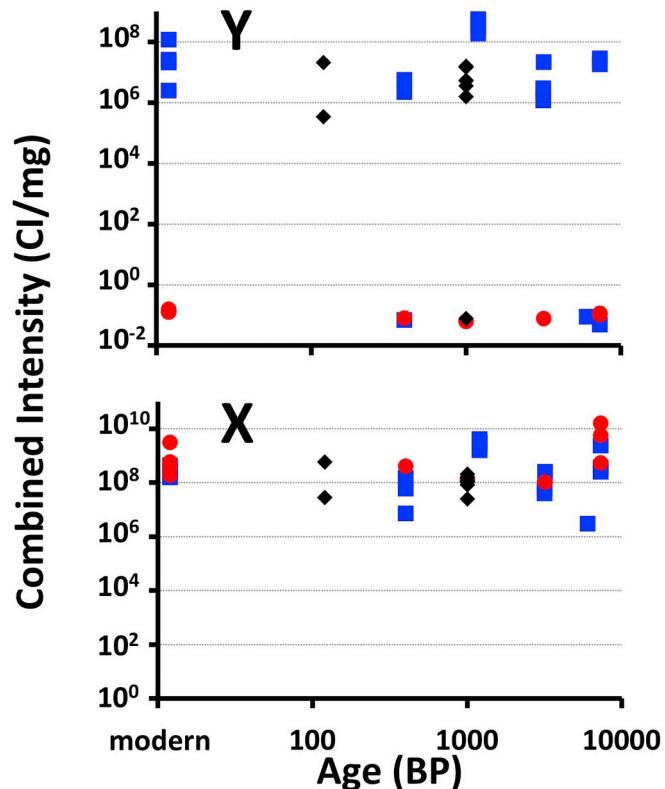


Fig. 5. Persistence of Amelogenin Peptides Over Time. The combined peak intensity (CI/mg) from peptides specific to either AMELY_HUMAN (Y) or AMELX_HUMAN (X), were measured and plotted as a function of sample age (BP). Estimated sex is indicated (male: blue square, female: red circle, unsexed: black diamond). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

signals, maximal peak intensities, from multiple protein specific peptides were combined bioinformatically. Both of these approaches increase the proteomic signal from an enamel sample. Because the individual assignment of a peptide sequences to a fragmentation spectrum was screened for false positive assignment, the specificity of each signal was also increased (Zhang et al., 2012). Thirdly, the approach used in this study was quantitative, whereas the Stewart study was binary and did not take into account the possibility of false negative assignment. The quantitative approach is important for two reasons. The signal of AMELY_HUMAN was one sixth to one eighth of the magnitude of AMELX_HUMAN protein, therefore in low signal samples

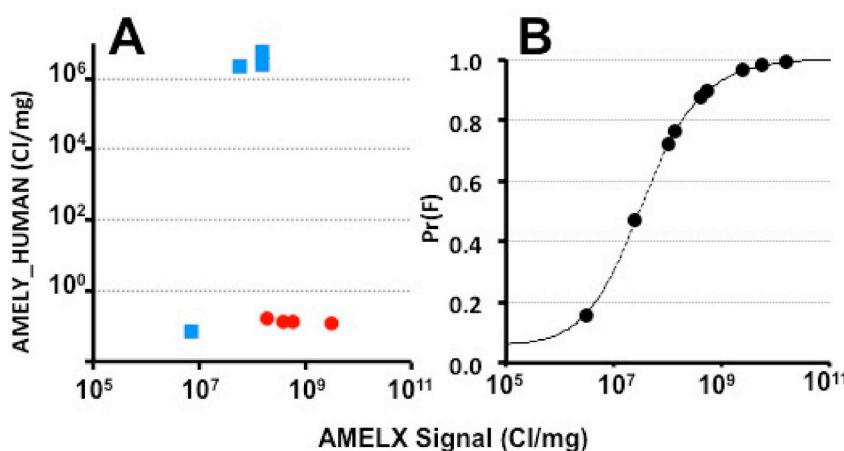


Fig. 4. Probability of Female Sex Estimation. (A) A calibration curve was constructed using only data from modern samples and samples with sex estimates based on DNA methodology (male samples, blue squares; female samples, red circles; $n = 11$). Remaining archaeological samples with no AMELY_HUMAN signal were applied to the calibration curve and probability of female sex ($Pr(F)$) estimated (B). A single unit was added to all measurements so that samples with no detected AMELY_HUMAN peptides would not have a null value. The resulting curve is symmetrical sigmoidal and a least-squares fit resulted in a formula of $y = 1.004627 + (0.05869012 - 1.004627)/(1 + (x/7.540008)^{13.99354})$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

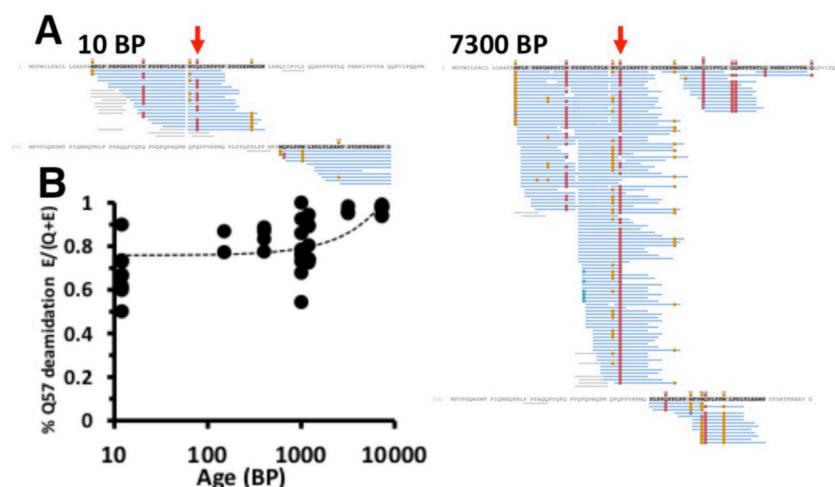


Fig. 6. Chemical Modification of AMELX_HUMAN Over Archaeological Time. (A) Glutamine 57 (red arrows) of AMELX_HUMAN deamidates to form glutamate (red squares). Amelogenin protein in modern samples contains relatively less glutamate relative to older samples. (B) The proportion of intensity from peptides containing glutamate at position 57 of AMELX_HUMAN was measured in each sample and a linear line of best fit ($y = ax + b$) calculated. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

it would drop out first resulting in a false positive estimate (Fincham et al., 1991). We observed variation in endogenous signal that ranged over two orders of magnitude. Both of these factors will result in some low signal male samples having no detectable AMELY_HUMAN (Fig. 3). Finally, because we always detected the AMELX_HUMAN protein, a subset of low signal male samples would be falsely assigned as female. The probability of a false negative sample however will decrease with an increase in amelogenin signal. Female sex should always be provided as a probability. Such a framework is not considered in the Stewart study. Nevertheless, the different approach in this study should not be interpreted as an advantage over the method described in the Stewart paper. Anthropologists requiring maintenance of the gross anatomical features of a tooth could employ the innovative method as outlined by Stewart.

The biogenesis of enamel is highly regulated and mediated through a group of enamel specific proteins that include both of the amelogenin proteins, ameloblastin, enamelin, and enamelysin, all of which were detected in this analysis (data are available via ProteomeXchange with identifier PXD009781, <http://www.proteomexchange.org>). The physical organization, calcium binding, and regulated removal of these proteins is necessary for the precise deposition of calcium apatite onto the growing faces of the microcrystal tufts to provide the right physical combination of hardness and flexibility in the enamel tissue. The mechanism is still the subject of research, but the organized and targeted degradation of these proteins by coordinated proteolysis is a necessary step (Carneiro et al., 2016; Kwak et al., 2016; Mazumder et al., 2014; Prajapati et al., 2016; Tarasevich et al., 2007). The result of this process is a range of overlapping non-tryptic peptides remaining in the tissue. By distributing the peptide signal through a wide variety of peptide masses, the signal from any one peptide is reduced. This is a concern when depending on one peptide as a source of sex information (Stewart et al., 2017; Stewart et al., 2016). The signal dilution is exacerbated by the high level of post-translational modifications particularly methionine oxidation and deamidation, which further increase the number of possible amelogenin peptide masses (Figs. 2 and 6). With time, other environmental chemistries, such as advanced glycation end-products, Maillard reactions, carbonyl oxidation and hydrolysis, will further diversify available peptide structures and reduce the signal from any one peptide species (Arena et al., 2014; Cabiscol et al., 2014). The use of trypsin in the proteomic sex assay has the potential to mitigate this slightly by reducing the diversity of sexually dimorphic peptides. These environmental processes are unlikely to account for the wide range of signal (combined peak intensities) associated with each gene product, which across the study is over 3 orders of magnitude, although less than 2 order of magnitude at any time point (Figs. 3 and 5).

Our results (Fig. 5) suggest that the quantified signal was stable over

7300 years, although this may be in part due to a reduction in the complexity of the enamel proteome, with the mineral-associated amelogenin comprising a greater proportion of the remaining peptides. All proteins and peptides within the sample will degrade over time, however the amelogenin peptides, associated with the mineral apatite surfaces will have less freedom of movement and therefore less degradative chemistry (Demarchi et al., 2016; Moradian-Oldak et al., 1998; Zhu et al., 2014). Over time the proportion of amelogenin in the proteome will increase and lower abundance peptides will contribute to the quantification of the protein.

A consequence of the high variation in combined peak ion intensity is that male false negative samples should be expected at some level requiring female sex estimation to be expressed as a probability (Fig. 4, Table 2). There are limitations to our approach however. We made an assumption that the probability of female sex as a function of total amelogenin signal would follow a symmetrical sigmoidal curve, following the premise that high signal samples would be less likely to be a low signal male false negative sample. The precise configuration of the sigmoidal probability curve (Fig. 4B) will change as more data from samples with known sex becomes available. One weakness of the current curve is that it depends too much on one low signal data point, a male sample with no detectable AMELY_HUMAN peptides. In theory, the lowest probability of female sex at the extreme end of low signal samples may well be higher than observed in Fig. 4B. Low signal female samples will also occur. The problem of insufficient low signal samples becomes greater if replicates from the same skeleton are averaged; a duplicate of the low signal sample contained a high level of AMELY_HUMAN peptides. A logistic regression is a more appropriate model for a discrete outcome, such as sex estimation, than regression models that have continuous outcomes (Agresti, 2013). A higher number of samples encompassing a range of signal levels will refine the parameters of the logistic regression curve and increase its accuracy. Alternatively if lower signal samples are not found, application of less material could be used to mimic low signal samples.

Protein deamidation is slower when the peptide backbone is less flexible, either because of protein structure or association with biomineral surfaces (Solazzo et al., 2014; Tyler-Cross and Schirch, 1991; van Doorn et al., 2012). Factors that increase peptide movement, such as temperature, will increase deamidation. The slower rate of deamidation at glutamine 57, as compared to glutamines 79, 86 and 87, could be due to increased binding to the calcium apatite surface or secondary protein structure, there is an adjacent tyrosine and tryptophan. The variation in the proportion of glutamine 57 deamidation at each time point however precludes this phenomenon from being used to estimate the archaeological age of the sample. Furthermore, the rate of deamidation increases at low pH and higher temperature, and the sample

protocol includes treatment in 1.2 M HCl at 56 °C for one hour. A high baseline is therefore expected and changes in the protocol to reduce deamidation during sample processing would be required to evaluate the effects of sample preparation on deamidation. While the degree of deamidation at glutamine 57 is not precise enough for age estimation, like DNA depurination it can be used to demonstrate consistency with what is expected from an ancient sample.

As with any new forensic and bioarchaeological method, sex estimation using proteomic identification of the Y-chromosome form of amelogenin (AMELY_HUMAN) will need to meet several criteria before it can be routinely employed. The chemical processing of samples needs to be optimized. The extent of peptide survivability and composition of the peptide signal over archaeological time needs to be demonstrated and the lower limits of detection and quantification determined. Different strategies can be used to further increase the sensitivity of mass spectrometry. An inclusion list is used in this study to focus on and increase detection of specific peptides. Peptides on this list are analyzed first before scanning for other peptides, as this increases sensitivity and ensures that relevant peptides proceed to fragmentation (Liebler and Zimmerman, 2013). With characterization of additional amelogenin peptides this list can be expanded, increasing sensitivity of proteomic sex estimation. This targeted approach can be expanded further using specialized techniques such as parallel reaction monitoring and data-independent analysis (Gallien et al., 2015; Li et al., 2016; Searle et al., 2015; Shi et al., 2016). In parallel reaction monitoring the instrumentation focuses exclusively on relevant peptides, increasing the sensitivity of detection, sometimes by orders of magnitude (Keshishian et al., 2007). In data independent acquisition, this approach is applied broadly across all peptides (Searle et al., 2015). The routine use of a targeted approach also reduces the cost and time of the protocol and allows for potential automation (Legg et al., 2017).

Proteomic detection of sex-chromosome specific amelogenins (AMELY_HUMAN and AMELX_HUMAN) in enamel from archaeological and forensic contexts forms the basis of protein-based sex estimation. As with DNA-based methods, the protein-based sex estimation assay could be applied just as easily to deciduous or permanent teeth in subadult individuals, and to skeletons with missing, degraded, or ambiguous osteological sex markers. Additionally, the use of proteomic methods in

sex estimation complements other proteomic applications in archaeological contexts, such as species identification, phylogenetics, and paleopathology. Further development of these methods, particularly if using a targeted data acquisition strategy, could be applicable over a range of mass spectrometry platforms, provide good levels of sensitivity, and be time and cost efficient. It provides a statistical and proteomic framework that can complement existing osteological and DNA-based sex estimation methods and extend the potential of sex estimation to additional archaeological assemblages, including subadult skeletons.

Disclaimers

A patent based on the concept and some data presented in this study has been awarded (US 8,877,455 B2, Australian Patent 2011229918, Canadian Patent CA 2794248, and European Patent EP11759843.3, GJP inventor). The patent is owned by Parker Proteomics LLC. Protein-Based Identification Technologies LLC (PBIT) has an exclusive license to develop the intellectual property and is co-owned by Utah Valley University and GJP. This ownership of PBIT and associated intellectual property does not alter policies on sharing data and materials. These financial conflicts of interest are administered by the Research Integrity and Compliance Office, Office of Research at the University of California, Davis to ensure compliance with University of California Policy.

Acknowledgements

The authors wish to thank Dr. Bethany Livermore-Turner (Georgia State University) for samples and advice on study design and Drs. Robert H Rice and Tammy Bounasera for advice on editing. For the California samples, we thank Most Likely Descendants Ramona Garibay and Katherine Perez for supporting the work, and William Self-Associates, Far Western Anthropological Research Group, and the UC Davis Anthropology Museum for providing access to the samples. GJP acknowledges the support of the National Institutes of Justice/Department of Justice (DN-BX-K065). JWE acknowledges the support of the National Science Foundation (#BCS-1318532).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jas.2018.08.011>.

Table A. 1

Inclusion List for Mass Spectrometry Data Acquisition. The above mass over charge values ($Z = 2$ and 3) with corresponding peptide sequences were used as the basis of an inclusion list in acquiring mass spectrometry data.

		m/z	
		M + 2	M + 3
AMELY	46 - 54	sequence	
		IALVLTPLK	484.33
		WYQSMIRPPY	670.83
		WYESMIRPPY	671.32
		WYQSmIRPPY	678.82
	55 - 65	WYESM(ox)IRPPY	679.32
		WYQSMIRPPYS	714.34
		WYESMIRPPYS	714.83
		WYQSM(ox)IRPPYS	722.34
		WYESM(ox)IRPPYS	722.83
AMELX	55 - 66	WYQSMIRPPYSS	757.86
		WYESMIRPPYSS	758.35
		WYQSMIRPPYSS	765.86
		WYESMIRPPYSS	766.35
46 - 54	TALVLTPLK	478.31	319.21

(continued on next page)

Table A. 1 (continued)

		m/z	
		M + 2	M + 3
55 - 63	WYQSIRPPY WYESIRPPY	605.31 605.80	403.87 404.20
55 - 64	WYQSIRPPYP WYESIRPPYP	653.83 654.32	436.22 436.55
55 - 65	WYQSIRPPYPS WYESIRPPYPS	697.35 697.84	465.23 465.56
55 - 66	WYQSIRPPYPSY WYESIRPPYPSY	778.88 779.37	519.59 519.92
	WYQSIRPPYPSYG WYESIRPPYPSYG	807.39 807.88	538.60 538.92

Gene Name	Splice Variant	AA#	Uniprot#
AMELX	X1 M G T W I L F A C L L G A A F A M P L P P H P G H P G Y I N F S Y E ----- V L T P L K W Y Q S - I	45	Q99217-1
	X2 M G T W I L F A C L L G A A F A M P ----- V L T P L K W Y Q S - I	29	Q99217-2
	X3 M G T W I L F A C L L G A A F A M P L P P H P G H P G Y I N F S Y E N S H S Q A I N V D R T A L V L T P L K W Y Q S - I	59	Q99217-3
AMELY	Y2 M G T W I L F A C L V G A A F A M P L P P H P G H P G Y I N F S Y E N S H S Q A I N V D R I A L V L T P L K W Y Q S M I	60	Q99218-2
	Y1 M G T W I L F A C L V G A A F A M P L P P H P G H P G Y I N F S Y E ----- V L T P L K W Y Q S M I	46	Q99218-1
AMELX	X1 R P P Y P S Y G Y E P M G G W L H H Q I I P V L S Q Q H P P T H T L Q P H H H I P V V P A Q Q P V I P Q Q P M M P V P G	105	
	X2 R P P Y P S Y G Y E P M G G W L H H Q I I P V L S Q Q H P P T H T L Q P H H H I P V V P A Q Q P V I P Q Q P M M P V P G	89	
	X3 R P P Y P S Y G Y E P M G G W L H H Q I I P V L S Q Q H P P T H T L Q P H H H I P V V P A Q Q P V I P Q Q P M M P V P G	119	
AMELY	Y2 R P P Y S S Y G Y E P M G G W L H H Q I I P V V S Q Q H P L T H T L Q S H H H I P V V P A Q Q P R V R Q Q A L M P V P G	120	
	Y1 R P P Y S S Y G Y E P M G G W L H H Q I I P V V S Q Q H P L T H T L Q S H H H I P V V P A Q Q P R V R Q Q A L M P V P G	106	
AMELX	X1 Q H S M T P I Q H H Q P N L L P P A Q Q P Y Q P Q P V Q P Q P H Q P M Q P Q P P V H P M Q P L P Q P P L P P M F P M Q	165	
	X2 Q H S M T P I Q H H Q P N L L P P A Q Q P Y Q P Q P V Q P Q P H Q P M Q P Q P P V H P M Q P L P Q P P L P P M F P M Q	149	
	X3 Q H S M T P I Q H H Q P N L L P P A Q Q P Y Q P Q P V Q P Q P H Q P M Q P Q P P V H P M Q P L P Q P P L P P M F P M Q	179	
AMELY	Y2 Q Q S M T P T Q H H Q P N L L P L P A Q Q P F Q P Q P V Q P Q P H Q P M Q P Q P P V Q P M Q P L L P Q P P L P P M F P L R	180	
	Y1 Q Q S M T P T Q H H Q P N L L P L P A Q Q P F Q P Q P V Q P Q P H Q P M Q P Q P P V Q P M Q P L L P Q P P L P P M F P L R	166	
AMELX	X1 P L P P M L P D L I T L E A W P S T D K T K R E E V D	191	
	X2 P L P P M L P D L I T L E A W P S T D K T K R E E V D	175	
	X3 P L P P M L P D L I T L E A W P S T D K T K R E E V D	205	
AMELY	Y2 P L P P I L P D L I H L E A W P A T D K T K Q E E V D	206	
	Y1 P L P P I L P D L I H L E A W P A T D K T K Q E E V D	192	

Fig. A.1. Amino Acid Variation in Amelogenin Sex-Specific Isoforms. Amelogenin genes are transcribed in 3 (AMELX: isoform 1:NM_001142.2, isoform 2:NM_182681.1, isoform 3:NM_182680.1) and 2 (AMELY: isoform 2:XP_016885531.1, isoform 1:NP_001134.1) transcripts respectively (Salido et al., 1992; Simmer, 1995). Amino acid variation between the two proteins are indicated in blue squares. The putative signal peptide is indicated in green. UNIPROT protein accession numbers are indicated.

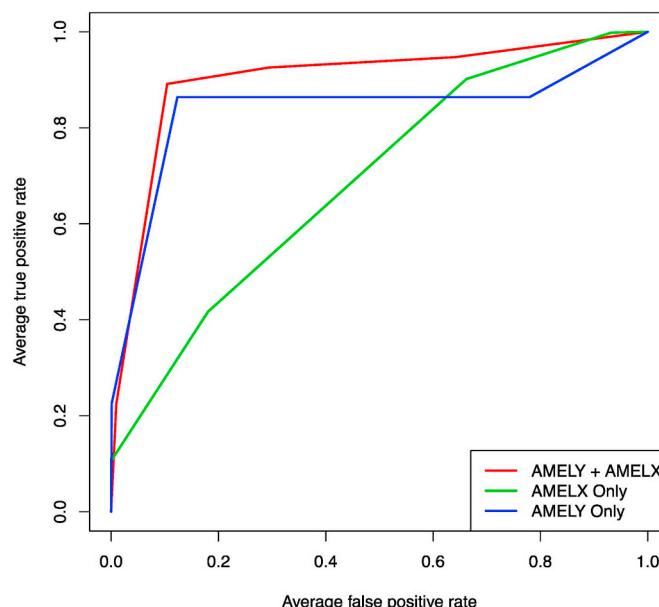


Fig. A.2. Receiver Operator Curve (ROC) for models with log AMELY plus log AMELX, log AMELY only, and log AMELX only. X-axis shows the false negative rate (1 – specificity) and the y-axis shows the true positive rate (sensitivity). Values for the model with log AMELX and log AMELY are shown in red, values for the model with log AMELY only are shown in blue, and values for the model with log AMELX only are shown in green.

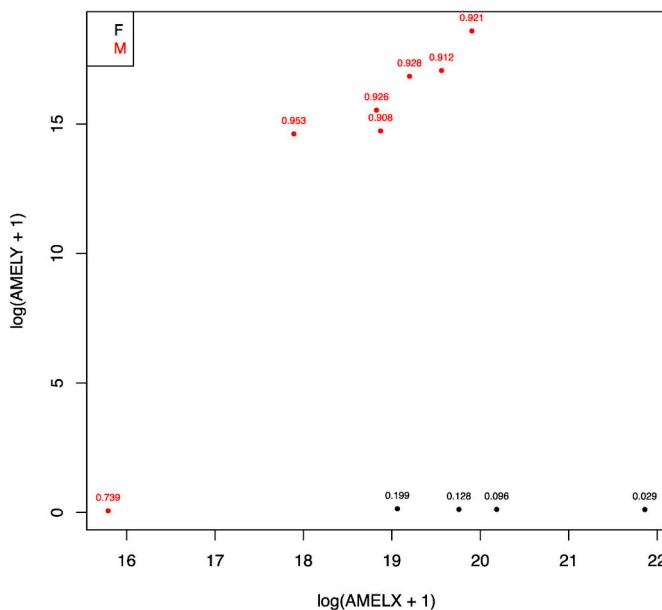


Fig. A.3. Predicted probabilities of male sex from model with log AMELY and log AMELX, by log AMELX and log AMELY. The x-axis shows log(AMELX + 1) and the y-axis shows log(AMELY + 1). For each observation, the predicted probability of male sex is shown above the point. Samples from female subjects are shown in black and samples from male subjects are shown in red.

References

- Aebersold, R., Mann, M., 2016. Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355. <https://doi.org/10.1038/nature19949>.
- Agresti, A., 2013. *Categorical Data Analysis*, third ed. John Wiley and Sons.
- Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaniego, J.A., Gilbert, M.T., Willerslev, E., Zhang, G., Scofield, R.P., Holdaway, R.N., Bunce, M., 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Biol. Sci.* 279, 4724–4733. <https://doi.org/10.1098/rspb.2012.1745>.
- Arena, S., Salzano, A.M., Renzone, G., D'Ambrosio, C., Scaloni, A., 2014. Non-enzymatic glycation and glycoxidation protein products in foods and diseases: an interconnected, complex scenario fully open to innovative proteomic studies. *Mass Spectrom. Rev.* 33, 49–77. <https://doi.org/10.1002/mas.21378>.
- Ballantyne, J., 2013. *DNA Profiling of the Semen Donor in Extended Interval Post-Coital Samples*. National Institute of Justice.
- Buckley, M., 2016. Species identification of bovine, ovine and porcine type 1 collagen; comparing peptide mass fingerprinting and LC-based proteomics methods. *Int. J. Mol. Sci.* 17, 445. <https://doi.org/10.3390/jims17040445>.
- Buikstra, J.E., Ubelaker, D.H., 1994. *Standards for Data Collection from Human Skeletal Remains*. Arkansas Archaeological Survey Press Fayetteville.
- Cabisco, E., Tamarit, J., Ros, J., 2014. Protein carbonylation: proteomics, specificity and relevance to aging. *Mass Spectrom. Rev.* 33, 21–48.
- Cappellini, E., Jensen, L.J., Szklarczyk, D., Ginolhac, A., da Fonseca, R.A., Stafford, T.W., Holen, S.R., Collins, M.J., Orlando, L., Willerslev, E., Gilbert, M.T., Olsen, J.V., 2012. Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J. Proteome Res.* 11, 917–926. <https://doi.org/10.1021/pr020721u>.
- Carneiro, K.M., Zhai, H., Zhu, L., Horst, J.A., Sitlin, M., Nguyen, M., Wagner, M., Simpliciano, C., Milder, M., Chen, C.L., Ashby, P., Bonde, J., Li, W., Habelitz, S., 2016. Amyloid-like ribbons of amelogenins in enamel mineralization. *Sci. Rep.* 6, 23105. <https://doi.org/10.1038/srep23105>.
- Castiblanco, G.A., Rutishauser, D., Ilag, L.L., Martignon, S., Castellanos, J.E., Mejia, W., 2015. Identification of proteins from human permanent erupted enamel. *Eur. J. Oral Sci.* 123, 390–395. <https://doi.org/10.1111/eos.12214>.
- Clarke, S., 1987. Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. *Int. J. Pept. Protein Res.* 30, 808–821.
- 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., Dominguez-Rodrigo, M., MacPhee, R., Kwekason, A., Ecker, M., Kolska Horwitz, L., Chazan, M., Kroger, R., Thomas-Oates, J., Harding, J.H., Cappellini, E., Penkman, K., Collins, M.J., 2016. Protein sequences bound to mineral surfaces persist into deep time. *eLife* 5, e17092. <https://doi.org/10.7554/eLife.17092>.
- Deutsch, E.W., Csordas, A., Sun, Z., Jarnuzczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D.S., Bernal-Llinares, M., Okuda, S., Kawano, S., Moritz, R.L., Carver, J.J., Wang, M., Ishihama, Y., Bandeira, N., Hermjakob, H., Vizcaino, J.A., 2017. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* 45, D1100–D1106. <https://doi.org/10.1093/nar/gkw936>.
- Doerr, A., 2013. Mass spectrometry-based targeted proteomics. *Nat. Methods* 10, 23.
- Eerkens, J.W., Barfod, G.H., Jorgenson, G.A., Peske, C., 2014. Tracing the mobility of individuals using stable isotope signatures in biological tissues: “locals” and “non-locals” in an ancient case of violent death from Central California. *J. Archaeol. Sci.* 41, 474–481.
- Eerkens, J.W., Carlson, T., Malhi, R.S., Blake, J., Bartelink, E.J., Barfod, G.H., Estes, A., Garibay, R., Glessner, J., Greenwald, A.M., Lentz, K., Li, H., Marshall, C.K., 2016a. Isotopic and genetic analyses of a mass grave in central California: implications for precontact hunter-gatherer warfare. *Am. J. Phys. Anthropol.* 159, 116–125. <https://doi.org/10.1002/ajpa.22843>.
- Eerkens, J.W., Harold, L., Perez, K., Murga, E., Kajankoski, P., Rosenthal, J.S., Greenwald, A.M., 2016b. Strontium isotopes identify locals and non-locals at CA-SCL-919 and CA-SCL-928 in milpitas, California. *Proc. Soc. Calif. Archaeol.* 30, 281–290.
- Eerkens, J.W., Washburn, E., Greenwald, A.M., 2017. Weaning and Early Childhood Diets at Two Early Period Sites: Implications for Parental Investment and Population Growth in Central California. *California Archaeology*, vol. 9, pp. 199–222.
- Fincham, A.G., Bessem, C.C., Lau, E.C., Pavlova, Z., Shuler, C., Slavkin, H.C., Snead, M.L., 1991. Human developing enamel proteins exhibit a sex-linked dimorphism. *Califif. Tissue Int.* 48, 288–290.
- Firth, D., 1993. Bias reduction of maximum likelihood estimates. *Biometrika* 80, 27–38.
- Gallien, S., Kim, S.Y., Domon, B., 2015. Large-scale targeted proteomics using internal standard triggered parallel reaction monitoring (IS-PRM). *Mol. Cell. Proteomics* 14, 1630–1644. <https://doi.org/10.1074/mcp.O114.043968>.
- Garvin, A.M., Fischer, A., Schnee-Griese, J., Jelinski, A., Bottinelli, M., Soldati, G., Tubio, M., Castella, V., Monney, N., Malik, N., Madrid, M., 2012. Isolating DNA from sexual assault cases: a comparison of standard methods with a nuclelease-based approach. *Invest. Genet.* 3, 25. <https://doi.org/10.1186/2041-2223-3-25>.
- Geiger, T., Clarke, S., 1987. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J. Biol. Chem.* 262, 785–794.
- Gordon, C.C., Buikstra, J.E., 1981. Soil pH, bone preservation, and sampling bias at mortuary sites. *Am. Antiqu.* 46, 566–571.
- Greenwald, A.M., Eerkens, J.W., Bartelink, E.J., 2016. Stable isotope evidence of juvenile foraging in prehistoric Central California. *J. Archaeol. Sci.: Report* 7, 146–154. <https://doi.org/10.1016/j.jasrep.2016.04.003>.
- Haas, R., Stefanescu, I.C., Garcia-Putnam, A., Aldenderfer, M.S., Clementz, M.T., Murphy, M.S., Llave, C.V., Watson, J.T., 2017. Humans permanently occupied the Andean highlands by at least 7 ka. *R. Soc. Open Sci.* 4, 1–11. <https://doi.org/10.1098/rsos.170331>.
- Haas, W.R., Viviano Llave, C., 2015. Hunter-gatherers on the eve of agriculture: investigations at Soro mik'aya Patjxa, lake titicaca basin 8000 - 6700 BP. *Antiquity* 89, 1297–1312.
- Iuliano-Burns, S., Hopper, J., Seeman, E., 2009. The age of puberty determines sexual dimorphism in bone structure: a male/female co-twin control study. *J. Clin. Endocrinol. Metab.* 94, 1638–1643. <https://doi.org/10.1210/jc.2008-1522>.
- Keshishian, H., Addona, T., Burgess, M., Kuhn, E., Carr, S.A., 2007. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* 6, 2212–2229. <https://doi.org/10.1074/mcp.M600610mcp0600610>.

- 1074/mcp.M700354-MCP200.
- Kosmidis, I., 2017. Brglm: Bias Reduction in Binary-response Generalized Linear Models. R Package Version 0.6.1.
- Krishan, K., Chatterjee, P.M., Kanchan, T., Kaur, S., Baryah, N., Singh, R.K., 2016. A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Sci. Int.* 261, 165. <https://doi.org/10.1016/j.forsciint.2016.02.007>. e161-168.
- Kwak, S.Y., Yamakoshi, Y., Simmer, J.P., Margolis, H.C., 2016. MMP20 proteolysis of native amelogenin regulates mineralization in vitro. *J. Dent. Res.* 95, 1511–1517. <https://doi.org/10.1177/0022034516666281>.
- LeDell, E., Petersen, M., van der Laan, M., 2014. cvAUC: Cross-validated Area under the ROC Curve Confidence Intervals. R Package Version 1.1.0.
- Legg, K.M., Powell, R., Reisdorff, N., Reisdorff, R., Danielson, P.B., 2017. Verification of protein biomarker specificity for the identification of biological stains by quadrupole time-of-flight mass spectrometry. *Electrophoresis* 38, 833–845. <https://doi.org/10.1002/elps.201600352>.
- Li, S., Zhang, Y., Wang, J., Yang, Y., Miao, C., Guo, Y., Zhang, Z., Cao, Q., Shui, W., 2016. Combining untargeted and targeted proteomic strategies for discrimination and quantification of cashmere fibers. *PLoS One* 11, e0147044. <https://doi.org/10.1371/journal.pone.0147044>.
- Liebler, D.C., Zimmerman, L.J., 2013. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* 52, 3797–3806. <https://doi.org/10.1021/bi400110b>.
- Lindahl, T., 1996. The Croonian Lecture, 1996: endogenous damage to DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 351, 1529–1538. <https://doi.org/10.1098/rstb.1996.0139>.
- Loreille, O., Ratnayake, S., Bazinet, A.L., Stockwell, T.B., Sommer, D.D., Rohland, N., Mallick, S., Johnson, P.L.F., Skoglund, P., Onorato, A.J., Bergman, N.H., Reich, D., Irwin, J.A., 2018. Biological sexing of a 4000-year-old Egyptian mummy head to assess the potential of nuclear DNA recovery from the most damaged and limited forensic specimens. *Genes (Basel)* 9. <https://doi.org/10.3390/genes9030135>.
- Madel, M.B., Niederstetter, H., Parson, W., 2016. TriXY-Homogeneous genetic sexing of highly degraded forensic samples including hair shafts. *Forensic Sci. Int. Genet.* 25, 166–174. <https://doi.org/10.1016/j.jfsgen.2016.09.001>.
- Mallick, P., Kuster, B., 2010. Proteomics: a pragmatic perspective. *Nat. Biotechnol.* 28, 695–709. <https://doi.org/10.1038/nbt.1658>.
- Mazumder, P., Prajapati, S., Lokappa, S.B., Gallon, V., Moradian-Oldak, J., 2014. Analysis of co-assembly and co-localization of ameloblastin and amelogenin. *Front. Physiol.* 5, 274. <https://doi.org/10.3389/fphys.2014.00274>.
- McFadden, C., Oxenham, M., 2016. Revisiting the Phenice technique sex classification results reported by MacLaughlin and Bruce (1990). *Am. J. Phys. Anthropol.* 159, 182–183. <https://doi.org/10.1002/ajpa.22839>.
- Mitsiadis, T.A., Filatova, A., Papaccio, G., Goldberg, M., About, I., Papagerakis, P., 2014. Distribution of the amelogenin protein in developing, injured and carious human teeth. *Front. Physiol.* 5, 477. <https://doi.org/10.3389/fphys.2014.00477>.
- Mittnik, A., Wang, C.C., Svoboda, J., Krause, J., 2016. A molecular approach to the sexing of the triple burial at the upper paleolithic site of dolni vestonice. *PLoS One* 11, e0163019. <https://doi.org/10.1371/journal.pone.0163019>.
- Moradian-Oldak, J., Tan, J., Fincham, A.G., 1998. Interaction of amelogenin with hydroxyapatite crystals: an adherence effect through amelogenin molecular self-association. *Biopolymers* 46, 225–238. [https://doi.org/10.1002/\(SICI\)1097-0282\(19981005\)46:4<225::AID-BIP4>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-0282(19981005)46:4<225::AID-BIP4>3.0.CO;2-R).
- Nesvizhskii, A.I., 2010. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics* 73, 2092–2123. <https://doi.org/10.1016/j.jprot.2010.08.009>.
- Ottoni, C., Bekaert, B., Decorte, R., 2017. DNA degradation: current knowledge and progress in DNA analysis. In: Schotsmans, E.M.J., Márquez-Grant, N., Forbes, S.L. (Eds.), *Taphonomy of Human Remains: Forensic Analysis of the Dead and the Depositional Environment*. Wiley and Sons Ltd., pp. 65–80.
- Phenice, T.W., 1969. A newly developed visual method of sexing the os pubis. *Am. J. Phys. Anthropol.* 30, 297–301. <https://doi.org/10.1002/ajpa.1330300214>.
- Porto, I.M., Laure, H.J., de Sousa, F.B., Rosa, J.C., Gerlach, R.F., 2011a. Techniques for the recovery of small amounts of mature enamel proteins. *J. Archaeol. Sci.* 38, 3596–3604.
- Porto, I.M., Laure, H.J., Tykot, R.H., de Sousa, F.B., Rosa, J.C., Gerlach, R.F., 2011b. Recovery and identification of mature enamel proteins in ancient teeth. *Eur. J. Oral Sci.* 119 (Suppl. 1), 83–87. <https://doi.org/10.1111/j.1600-0722.2011.00885.x>.
- Prajapati, S., Tao, J., Ruan, Q., De Yoreo, J.J., Moradian-Oldak, J., 2016. Matrix metalloproteinase-20 mediates dental enamel biomineralization by preventing protein occlusion inside apatite crystals. *Biomaterials* 75, 260–270. <https://doi.org/10.1016/j.biomaterials.2015.10.031>.
- R Core Team, 2018. R: a language and environment for statistical computing. In: Computing, R.F.F.S. (Ed.), (Vienna, Austria).
- Regan, L.A., 2006. Isotopic Determination of Region of Origin in Modern Peoples: Applications for Identification of U.S. War-dead from the Vietnam Conflict [dissertation], Anthropology. UNIVERSITY OF FLORIDA, Gainesville.
- Robinson, N.E., 2002. Protein deamidation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5283–5288. <https://doi.org/10.1073/pnas.082102799>.
- Salido, E.C., Yen, P.H., Koprivnikar, K., Yu, L.C., Shapiro, L.J., 1992. The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am. J. Hum. Genet.* 50, 303–316.
- Schweitzer, M.H., Zheng, W., Organ, C.L., Avci, R., Suo, Z., Freemark, L.M., Lebleu, V.S., Duncan, M.B., Vander Heiden, M.G., Neveu, J.M., Lane, W.S., Cottrell, J.S., Horner, J.R., Cantley, L.C., Kalluri, R., Asara, J.M., 2009. Biomolecular characterization and protein sequences of the Campanian hadrosaur *B. canadensis*. *Science* 324, 626–631. <https://doi.org/10.1126/science.1165069>. 324/5927/626 [pii].
- Searle, B.C., Egertson, J.D., Bollinger, J.G., Stergachis, A.B., MacCoss, M.J., 2015. Using data independent acquisition (DIA) to model high-responding peptides for targeted proteomics experiments. *Mol. Cell. Proteomics* 14, 2331–2340.
- Shi, T., Song, E., Nie, S., Rodland, K.D., Liu, T., Qian, W.J., Smith, R.D., 2016. Advances in targeted proteomics and applications to biomedical research. *Proteomics* 16, 2160–2182. <https://doi.org/10.1002/pmic.201500449>.
- Simmer, J.P., 1995. Alternative splicing of amelogenins. *Connect. Tissue Res.* 32, 131–136.
- Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. *Bioinformatics* 21, 3940–3941. <https://doi.org/10.1093/bioinformatics/bti623>.
- Skoglund, P., Storå, J., Götherström, A., Jakobsson, M., 2013. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* 40, 4477–4482. <https://doi.org/10.1016/j.jas.2013.07.004>.
- 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 alpha-keratin peptides and application to archaeological wool textiles. *Anal. Chem.* 86, 567–575. <https://doi.org/10.1021/ac4026362>.
- Spradley, M.K., Jantz, R.L., 2011. Sex estimation in forensic anthropology: skull versus postcranial elements. *J. Forensic Sci.* 56, 289–296. <https://doi.org/10.1111/j.1556-4029.2010.01635.x>.
- Steen, H., Mann, M., 2004. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 5, 699–711. <https://doi.org/10.1038/nrm1468>.
- Stewart, N.A., Gerlach, R.F., Gowland, R.L., Gron, K.J., Montgomery, J., 2017. Sex determination of human remains from peptides in tooth enamel. *Proc. Natl. Acad. Sci. U. S. A.* <https://doi.org/10.1073/pnas.1714926115>.
- Stewart, N.A., Molina, G.F., Mardegan, J.P., Nathan, I., Yates, A., Sosovicka, M., Vieira, A.R., Roberto, S., Line, P., Montgomery, J., Gerlach, R.F., 2016. The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction. *RSC Adv.* 6, 61673–61679.
- Tarasevich, B.J., Howard, C.J., Larson, J.L., Snead, M.L., Simmer, J.P., Paine, M., Shaw, W.J., 2007. The nucleation and growth of calcium phosphate by amelogenin. *J. Cryst. Growth* 304, 407–415. <https://doi.org/10.1016/j.jcrysgro.2007.02.035>.
- Tyler-Cross, R., Schirch, V., 1991. Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. *J. Biol. Chem.* 266, 22549–22556.
- van Doorn, N.L., Wilson, J., Hollund, H., Soressi, M., Collins, M.J., 2012. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* 26, 2319–2327. <https://doi.org/10.1002/rcm.6351>.
- Wadsworth, C., Buckley, M., 2014. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun. Mass Spectrom.* 28, 605–615. <https://doi.org/10.1002/rcm.6821>.
- Waldron, T., 1987. The relative survival of the human skeleton: implications for paleopathology. In: Boddington, A., Garland, A., Janaway, R. (Eds.), *Death, Decay and Reconstruction: Approaches to Archaeology and Forensic Science*. Manchester University Press, Manchester, pp. 55–64.
- 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., 2015. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature* 522, 81–84. <https://doi.org/10.1038/nature14249>.
- Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B.M., Stewart, J.R., Paabo, S., Collins, M.J., Hublin, J.J., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Chatelperronian at the Grotte du Renne. *Proc. Natl. Acad. Sci. U. S. A.* 113, 11162–11167. <https://doi.org/10.1073/pnas.1605834113>.
- White, T.D., Black, M.T., Folkens, P.A., 2012. *Human Osteology*, third ed. Academic Press, Burlington, MA.
- Wilson, J., van Doorn, N.L., Collins, M.J., 2012. Assessing the extent of bone degradation using glutamine deamidation in collagen. *Anal. Chem.* 84, 9041–9048. <https://doi.org/10.1021/ac301333t>.
- Woods, A.G., Sokolowska, I., Ngounou Wetie, A.G., Wormwood, K., Aslebagh, R., Patel, S., Darie, C.C., 2014. Mass spectrometry for proteomics-based investigation. *Adv. Exp. Med. Biol.* 806, 1–32. https://doi.org/10.1007/978-3-319-06068-2_1.
- 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. *Mol. Cell. Proteomics* 11<https://doi.org/10.1074/mcp.M111.010587>. M111 010587.
- Zhu, L., Liu, H., Witkowska, H.E., Huang, Y., Tanimoto, K., Li, W., 2014. Preferential and selective degradation and removal of amelogenin adsorbed on hydroxyapatites by MMP20 and KLK4 in vitro. *Front. Physiol.* 5, 268. <https://doi.org/10.3389/fphys.2014.00268>.