**Article:**
Blacka, Charllotte, Dowle, Adam, Lisowski, Mik et al. (4 more authors) (2025) Rapid proteomic amelogenin sex estimation of human and cattle remains using untargeted Evosep-timsTOF mass spectrometry. Rapid communications in mass spectrometry. e10022. ISSN 1097-0231

https://doi.org/10.1002/rcm.10022

**Title:** Rapid proteomic amelogenin sex estimation of human and cattle remains using untargeted Evosep-timsTOF mass spectrometry

**Short Title:** Rapid amelogenin sex estimation using untargeted mass spectrometry

**Authors:** Charllotte Blacka[1,2], Adam Dowle[3], Mik Lisowski[1], Michelle Alexander[1], Jessica Hendy[1], Kirsty Penkman[2], Jackie Mosely[2]

**Affiliations:** [1]Department of Archaeology, University of York, YO10 5DD, UK; [2]Department of Chemistry, University of York, YO10 5DD, UK; [3]Bioscience Technology Facility, Department of Biology, University of York, YO10 5DD, UK.

**Correspondence:** Charllotte Blacka, Department of Archaeology, University of York, York YO10 5DD, UK; Email: csb548@york.ac.uk; Telephone: 01904 324472.

**Funding information:** This project is part of the Natural Environment Research Council (NERC) Doctoral Training Partnership – Adapting to the Challenges of a Changing Environment (ACCE) [grant number NE/S00713X/1] .

**Keywords:** amelogenin, palaeoproteomics, sex estimation, LC-MS/MS, TIMS-TOF



## Abstract

**Rationale:** Sex estimation by analysis of amelogenin peptides in archaeological and fossil material has recently been gaining great traction within the fields of archaeology and palaeontology. Current widely used proteomic amelogenin sex estimation methods are hindered by relatively long mass spectrometric run times, or targeting peptides specific to human amelogenin proteins. Untargeted, high-throughput amelogenin sexing would be invaluable for a range of applications, from sex estimation of remains at mass grave sites to broadening the application of rapid amelogenin sexing to non-hominin species for husbandry and evolutionary studies.

**Methods:** A new acid etch protocol followed by Evosep-LC-TIMS-TOF mass spectrometry is presented for amelogenin analysis, providing global peptide data through rapid mass spectrometric methods in under 20 minutes per sample (including sample preparation, mass spectrometric acquisition and data processing). This sampling protocol was developed on modern cattle (*Bos taurus*) teeth, before Evosep-timsTOF partial validation with archaeological cattle and human (*Homo sapiens*) teeth, demonstrating the potential of straightforward application of this rapid amelogenin sexing method to a range of taxa.

**Results:** The rapid Evosep-LC-TIMS-TOF mass spectrometry methods gave comparable peptide counts to conventional long untargeted methods, while maintaining similar (or faster) acquisition times to those reported in methods targeting specific human amelogenin peptides. Implementation of the novel acid etch sampling approach also streamlined sample preparation without compromising peptide counts.

**Conclusions:** Rapid, untargeted Evosep-LC-TIMS-TOF mass spectrometry was successfully implemented in sex estimation of modern and archaeological material from *Bos taurus* and *Homo sapiens* teeth. This demonstrates an advancement in low-cost, high-throughput amelogenin sex estimation, for both human and zooarchaeological applications.

## 1. Introduction

Biological sex estimation of humans and animals is a key facet of archaeology, contributing to our understanding of the treatment of the dead, and relationships between sex and gender. For zooarchaeology, accurate sex estimation is vital in studies of sex-selective husbandry or hunting practices[1], and even ideology[2], while in palaeontology it can help resolve herd dynamics, subsistence strategies[3] and identify sexual dimorphism[4]. Sex estimation can be achieved by osteological assessment, the detection of x and y chromosomes using ancient DNA (aDNA), and more recently, mass spectrometric (MS) analysis of amelogenin peptides. Amelogenin peptide sex estimation presents some advantages over the former methods, with peptides surviving for much longer than aDNA[5,6], and enabling sex estimation of non-adult remains, which is not possible via osteological analysis. Peptidic sex estimation is achieved through the identification of the enamel-specific amelogenin proteins. Amelogenins are highly conserved and are the most abundant proteins in enamel, constituting more than 90% of its organic fraction[7]. They are encoded on the sex chromosomes of many Eutherian mammals, resulting in expression of sexually dimorphic X- and Y-isoforms of amelogenin, termed AMELX and AMELY, respectively[8,9]. The small differences in protein sequence between these sexually dimorphic proteins can be identified using MS techniques, enabling proteomic sex estimation.

Most palaeoproteomic amelogenin studies to date employ high resolution mass spectrometry (namely Orbitrap MS) for protein analysis, with analytical run times of approximately 60 minutes[10–22], although faster methods have recently been developed[23,24] (Figure S.1). When working with large assemblages (such as mass graves or population studies), these long MS run times (not including the accompanying preparative and data analysis times), and resulting high costs, preclude estimation of the sex of all individuals within a horizon/site, and so fully understanding population demographics is impossible. This inability to study large sample sets also means that there is limited applicability of amelogenin analysis to zooarchaeological sites, as highlighted by the comparative sparsity of zooarchaeological amelogenin publications[3,25–27], a problem that is also compounded by the lack of reliable amelogenin reference sequences for mammalian taxa in UniProt[28]. Previous approaches to develop high-throughput amelogenin sex estimation methods for human remains have employed targeted mass spectrometry methods, where the MS instrument is programmed to analyse a few peptides only, and the rest of the ions generated are disregarded[23,24]. These targeted

approaches significantly shorten run times and data complexity, but result in the loss of all protein/peptide information unrelated to these target peptides, meaning studying non-amelogenin enamel proteins or peptide modifications (such as deamidation for estimating the extent of degradation[29]) is not possible. Furthermore, these methods cannot be directly applied to non-hominin remains, since the targeted peptides are unique to hominins.

Here an untargeted, but rapid, proteomic amelogenin sex estimation method is presented, allowing for low-cost analysis of many taxa. The ease of application of the same method to different taxa is showcased through application to both cattle and human teeth. Trapped ion mobility separation-time of flight mass spectrometry (TIMS-TOF MS) is employed to achieve this, demonstrating the first reported use of ion mobility MS for amelogenin sexing. Although yet rarely used in palaeoproteomics, modern proteomic studies acclaim the suitability of TIMS-TOF MS for rapid analysis of low concentration samples[30–32], akin to those analysed in bioarchaeological studies. TIMS-TOF MS differs from other forms of MS usually utilised in amelogenin sex estimation by implementing a trapped ion mobility separation (TIMS) after the liquid chromatography (LC) separation. TIMS utilises the balancing of two opposing forces experienced by peptide ions in the TIMS cell: the drag from a gas flow heading into the instrument and an increasing opposing DC current that repels the ions[33]. Once these forces balance for a specific ion, the ion becomes trapped in space, before the current is reduced in a stepwise fashion, sequentially releasing 'packets' of ions into the MS. Ions with a larger rotationally averaged collisional cross-section (CCS; size or shape) experience more drag and travel further through the TIMS cell before being trapped, compared to ions with the same *m/z* but a smaller CCS. The timsTOF range of mass spectrometers use two sequential TIMS cells to enable parallel accumulation-serial fragmentation (PASEF), whereby the first cell is used to trap and hold ions as bulk 'packets' (increasing sensitivity), while the second cell steps down through iterations of DC current to release ions into the mass spectrometer proportionally to their CCS (increasing separation and dynamic range). The quadrupole, where fragmentation occurs, is operated intelligently and aligned with the ions leaving the second TIMS cell, to maximise speed of sampling enabling hundreds of MS/MS events per second at full sensitivity[34]. Ion mobility is implemented in this study as a means to separate ions based on their rotationally averaged CCS post chromatography and pre MS analysis. This enables rapid targeting of ions for MS/MS analysis.

Method development consisting of both sample preparation development and partial validation of the applicability of Evosep-timsTOF mass spectrometry to proteomic amelogenin sex estimation is presented herein. The sample preparation method was developed using modern cattle (*Bos taurus*) teeth, before testing on different rapid Evosep-timsTOF MS methods. Then, to test the applicability of this MS method two archaeological case studies were analysed: archaeological cattle teeth from the 46-54 Fishergate excavations, York[35]; and teeth from 18th-19th century human individuals from Cross Street Unitarian Chapel, Manchester, and Hazel Grove, Church Street, Manchester[36].

## 2. Materials and methods

### *2.1 Sample Information*

#### *2.1.1. Modern and Archaeological Cattle Teeth*

Analysis was undertaken on four modern cattle teeth of known sex, and six archaeological cattle mandibles with teeth of unknown sex from the excavations at 46-54 Fishergate, York, dating to between the 11th and 19th century CE[35] (Table S.1).

#### *2.1.2. Archaeological human teeth*

Human teeth from eight adults from Cross Street Unitarian Chapel, Manchester (CSM), and one adult and one non-adult from Chapel Street, Hazel Grove, Manchester (HGM), dating to 1760-1840 CE, were analysed. The teeth were from named individuals and these names have been used to infer biological sex (Table S.2). Seven of the individuals were well preserved, while the remaining three, including the non-adult individual, were poorly preserved (full description in Table S.2)[37,38].

### *2.2. Sample Preparation*

A full description of the sample preparation and MS methods used on each sample is provided in Table S.3 (summarised in Figure 1).

#### *2.2.1. Chemicals and consumables*

Chemicals used were purchased from Thermo Fisher Scientific Inc. (Loughborough, UK) (methanol, acetonitrile, hydrochloric acid, formic acid, 0.1% formic acid in water, 0.1% formic acid in acetonitrile, isopropanol), Sigma-Aldrich (St. Louis, USA) (ammonium hydroxide), and VWR (Rosny-sous-Bois, France) (trifluoroacetic acid, acetonitrile, hydrogen peroxide). All chemicals were analytical grade or higher. ZipTips used were Pierce™ 100 µL C18 Tips purchased from Thermo Fisher Scientific Inc. (Rockford, USA), and Evotips were purchased from Fintiede Solutions (Glasgow, UK). PCR tubes were purchased from Starlab (Hamburg, Germany) and all LoBind and centrifuge tubes were purchased from Eppendorf™ (Hamburg, Germany). All chemicals and consumables were stored following recommendations from the manufacturer.

#### *2.2.2. Preparation of Acid Etched Samples*

Conventional enamel acid etches were prepared according to the Stewart *et al.* (2016) protocol[10]. These etches were subjected to ZipTip clean-up. The ZipTip was conditioned as follows: the tip was cleaned with 100 $\mu$L MeOH, then 100 $\mu$L AT80 (80% 0.1% *v/v* TFA in ACN: 20% 0.1% *v/v* TFA in $H_2O$), then 100 $\mu$L 0.1% *v/v* TFA in $H_2O$. The etch was then passed through the tip 20 times before the tip was washed twice with 100 $\mu$L 0.1% *v/v* TFA in $H_2O$. The peptides were then eluted into a 1.5 mL LoBind tube with 100 $\mu$L 40% *v/v* ACN in $H_2O$. The eluted peptide solution was dried down using a rotary vacuum concentrator and the

dried peptides were stored in the freezer before reconstitution in 30 μL 0.1% *v/v* TFA in water prior to MS analysis. The Evosep LC injects samples directly from the Evotip as the use of an Evosep LC requires samples to be cleaned using Evotips. Neutralisation of acid-etched samples, followed by direct Evoptip clean-up was explored to bypass the use of both ZipTips, thus avoiding a two-step clean-up that would likely lead to peptide loss. This led to the testing of two acid etching methods: one with a ZipTip clean-up, followed by Evotip clean-up (referred to herein as the conventional acid etch method), and one with acid etch neutralisation followed by Evotip clean-up (referred to herein as the neutralised acid etch method) (Figure 2).

The conventional acid etch method was performed as follows: the teeth were washed with 3% *v/v* $H_2O_2$ in water, then rinsed with ultrapure water (18.2 MΩ-cm). 60 μL of 5% *v/v* HCl in water was pipetted into the cap of a 0.2 mL PCR tube, leaving a convex meniscus of HCl in the cap. The tooth was lowered into the HCl and contact with the acid was maintained for 2 minutes. The first etch was discarded and a second etch was performed by lowering the same area of the tooth into 60 μL of 5% *v/v* HCl in water once more for 2 minutes. The second etch was retained and cleaned up with ZipTips using the same protocol presented prior. MS analysis was performed on a 6 μL fraction of the 0.1% *v/v* TFA in water reconstituted conventional acid etch.

The novel neutralised acid etch method implemented a neutralisation step after acid etching. The teeth were washed with 3% *v/v* $H_2O_2$ in water, then rinsed with ultrapure water (18.2 MΩ-cm). 60 μL of 5% *v/v* HCl in water was pipetted into the cap of a 0.2 mL PCR tube, leaving a convex meniscus of HCl in the cap. The tooth was lowered into the HCl and contact with the acid was maintained for 2 minutes. The first etch was discarded and a second etch was performed by lowering the same area of the tooth into 60 μL of 5% *v/v* HCl in water once more for 2 minutes. The second etch was retained and 45 μL 5% *v/v* $NH_4OH$ in water was added, resulting in a white precipitate forming. The sample was briefly mixed with a vortex mixer before being centrifuged at 13,250 rcf for 3-4 minutes to separate the precipitate and the liquid. The supernatant was transferred to a new 0.2 mL PCR tube and stored in the freezer until MS analysis was performed. Where analytical replicate analysis was performed, a 320 μL acid etch was performed in the cap of a 1.5 mL centrifuge tube and split into 60 μL aliquots in 0.2 mL PCR tubes, followed by treatment of the etches as described prior for the 60 μL etches.

### *2.2.3. Preparation of Powdered Enamel Samples*

An adapted method from Cappellini *et al.*[25] was employed to sample amelogenin peptides from powdered enamel. Approximately 40 mg of enamel chips from each sample were crushed into a powder using an agate pestle and mortar. The enamel powder was suspended in 1.5 mL of 10% *v/v* HCl in water solution and refrigerated overnight. The samples were centrifuged for 10 min and the supernatant (S1) was removed and transferred to a new 1.5 mL LoBind tube before being frozen. The enamel samples were resuspended in 10% *v/v* HCl in water and refrigerated overnight. The LoBind tubes were centrifuged for 10 min and the supernatant (S2) was removed and transferred to a new 1.5 mL LoBind tube. 750 μL of S1

and 750 $\mu$L of S2 were then combined in a 2 mL LoBind tube to create S3. S3 was subjected to Ziptip clean-up, dried down, and reconstituted in 0.1% TFA as described in Section 2.2.2 before analysis.

### *2.3. LC-MS/MS Data Acquisition*

MS acquisition was performed using both a nanoLC-Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific Inc.) and EvosepLC-timsTOF HT (Bruker Daltonics GmbH).

#### *2.3.1. Orbitrap Fusion MS Analysis*

The powdered enamel peptides were loaded onto an M-Class nanoflow UPLC system (Waters Corporation) fitted with a nanoEaze M/S Symmetry 100Å $C_{18}$, 5 µm 180 µm x 20 mm trap column (Waters Corporation) and a PepMap, 2 µm, 100Å $C_{18}$ EasyNano nanocapillary 75 µm x 500 mm column (Thermo Fisher Scientific Inc.). 0.05% *v/v* TFA in water was used as a trap wash solvent and the trapping flow rate was 15 µL/min. The trap was washed for 5 min prior to the flow travelling to the capillary column. The injection volume for each sample was 6 $\mu$L and elution was achieved using gradient elution: aqueous 0.1% *v/v* formic acid in water (mobile phase A) and acetonitrile containing 0.1% *v/v* formic acid in water (mobile phase B). Solvent composition was altered as follows: 3-10% mobile phase B over 7 mins; 10-35% mobile phase B over 30 mins; 35-99% mobile phase B over 5 mins; and then a wash with 99% mobile phase B for 4 min. The capillary column flow rate was 300 nL/min and the column temperature was 40°C. After each run, the column was re-equilibrated for 15 mins prior to injection of the next sample.

The nanoLC system was interfaced with an Orbitrap Fusion Tribrid mass spectrometer with an EasyNano ionisation source (Thermo Fisher Scientific Inc.). Positive ESI-MS and $MS^2$ spectra were acquired using Xcalibur software (version 4.0, Thermo Fisher Scientific Inc.). The instrument source settings were: ion spray voltage, 1,900 V; sweep gas, 0 Arb; ion transfer tube temperature 275°C. $MS^1$ spectra were acquired on the Orbitrap MS with the following parameters: 120,000 resolution; scan range *m/z* 375-1,500; AGC target $4e^5$; max fill time 100 ms. Data dependent acquisition was performed in topN mode, selecting the 12 most intense precursors with charge states >1. easy-IC was used for internal calibration. Dynamic exclusion was performed for 20 s post precursor selection and a minimum threshold for fragmentation was set at $5e^3$. $MS^2$ spectra were acquired on the Orbitrap with the following parameters: 30,000 resolution; max fill time 100 ms; HCD; activation energy 32 NCE. Blanks were analysed after every sample and data was processed using the same methodology as the amelogenin extract samples. All blanks showed no carry over of AMELX or AMELY specific peptides across samples.

#### *2.3.2. timsTOF HT IMS MS Analysis*

Prior to MS analysis, the samples were subjected to Evotip clean-up. The Evotip procedure was as follows: the EvoTip (Evosep) was rinsed with 20 $\mu$L of solvent B (0.1% *v/v* formic acid in acetonitrile) and centrifuged at 800 rcf for 60 s. The tips were then soaked in isopropanol until the stationary phase turned pale white (approximately 20 s) and the soaked

tip was equilibrated with 20 $\mu$L of solvent A (0.1% *v/v* formic acid in water) and centrifuged again at 800 rcf for 60 s. For powdered enamel, and acid etching following the conventional acid etch sampling method, 6 $\mu$L of the 0.1% TFA reconstituted Ziptip sample was made up to 20 $\mu$L with solvent A and added to the tip, and for neutralised acid etch analysis, the entire supernatant was transferred to the Evotip. The tip was centrifuged at 800 rcf for 60 s. The tips were washed three times with 50 $\mu$L solvent A and centrifuged at 800 rcf for 60 s after each wash, before 100 $\mu$L of solvent A was added and centrifuged at 800 rcf for 10 s.

The wet Evotips were loaded onto an Evosep One (EV-1000) HPLC system (Evosep) equipped with either an Evosep performance column (EV-1064: 3 µm i.d., 100 µm x 8 cm) (for 24.0 or 14.4 minute method) or an Evosep Endurance column (EV-1107: 1.9 µm i.d., 150 µm x 4 cm) (for 4.8 or 2.9 minute method). The full description of the methods (as available in Evosep One Hystar Driver 2.3.57.0, Bruker Daltonics GmbH) is given in Table S.4. Solvents used in analysis were 0.1% *v/v* formic acid in water (mobile phase A) and 0.1% *v/v* formic acid in acetonitrile (mobile phase B).

The Evosep LC system was interfaced with a timsTOF HT mass spectrometer with a Captive Spray ionisation source (Bruker Daltonics GmbH). Positive ESI-MS/MS spectra were acquired using Compass HyStar 6.2 (Bruker Daltonics GmbH). The instrument source settings were as follows: ion spray voltage 1,500 V; ion transfer tube temperature, 180 °C; dry gas flow 3.0 L/min. Data dependent acquisition was performed with scan range: *m/z* 100-1,700 using parallel accumulation serial fragmentation (PASEF) scan mode. The trapped ion mobility settings were as follows: ramp time of 100 ms; ramp rate of 9.42 Hz; the ion mobility window ($1/k_0$ range), PASEF ramps, and total cycle time varied depending on method used: 24.0 minute and 14.4 minute used the standard 1.1 second cycle time method, while the 4.8 minute and 2.9 minute methods used the short 0.5 second cycle time method (Table S.5). MS2 data was acquired for precursors above 2500 intensity. HCD collision cell energy was 10 eV. Active exclusion (dynamic exclusion) was set to release after 0.4 min. Blanks were analysed after every sixth sample and data was processed using the same methodology as the amelogenin extract samples. All blanks showed no carry over of AMELX or AMELY specific peptides across samples. MGF files were generated from Bruker .d files in Bruker Compass DataAnalysis ver 6.0 using standard parameters (ProteinAnalysis tool) to collate MS/MS spectra and match mobility values to precursor ions.

### *2.4. MS Data Processing*

#### *2.4.1. $MS^2$ Data Searching in Mascot*

$MS^2$ data from the analysis of cattle teeth was searched against two FASTA libraries: one containing the whole bovine proteome and the cRAP contaminants library (119 contaminant entries), and the other containing bovine enamel and dentine sequences from SwissProt and UniProt[28] (amelogenin-X (UniProt accessions P02817 and P02817-2), amelogenin-Y (UniProt accessions Q99004/B7TCH9), matrix metalloproteinase-20, enamelin, tuftelin, ameloblastin, amelotin, collagen). Similarly, human data was searched against two FASTA libraries: one containing the whole human proteome and cRAP, and the other containing

human enamel and dentine protein sequences from SwissProt[28] (amelogenin-X (UniProt accessions Q99217-1, Q99217-2, and Q99217-3), amelogenin-Y (UniProt accessions Q99218-1 and Q99218-2), matrix metalloproteinase-20, enamelin, tuftelin, ameloblastin, amelotin, collagen) as well as cRAP. Mascot version 2.8, with the parameters as follows: enzyme, none; peptide tolerance was set to ±10.0 ppm; peptide charges 2+, 3+, and 4+ were searched; variable modifications of deamidation (NQ), oxidation (M), phosphorylation (STY) were chosen as they were the most frequently observed modifications and performed best in a series of tests where different combinations of modifications were employed (Table S.6). Male sex was assigned where two or more high-scoring unique AMELY peptides, with good quality MS/MS spectra, were identified (example spectra shown in Figures 3 and S.2).

Use of the whole cattle/human proteome gave the same sex estimations as use of the smaller, enamel and dentine proteome searches (Table S.7). Full proteome searches took approximately an hour or longer, while enamel and dentine proteome searches took a few minutes. Moreover, while a larger FASTA library can act as a form of quality assurance, it is also possible that a larger FASTA library could potentially result in more false female assignments, due to peptide scores needing to be higher due to the inflated score acceptance threshold for peptide identification, resulting in AMELY false negatives. Therefore, results from the enamel and dentine proteme searches are discussed herein to showcase data from a more rapid $MS^2$ data processing pipeline.

#### *2.4.2. Data Searching in FragPipe*

Peptide identification was performed using FragPipe following a similar protocol outlined in Cleland *et al.*[22]. Raw data files were searched with a LFQ-MBR workflow in FragPipe 22.0 with MSFragger ver 4.1[39,40], against a library of bovine enamel and dentine proteins or human enamel and dentine proteins (outlined in Section 2.4.1), and MSFragger decoys and contaminants, for bovine (286 protein entries) or human (292 protein entries) enamel samples, respectively. The MSFragger search parameters were as follows: 20 ppm precursor mass window, mass calibration on, isotope on, enzyme was set to nonspecific, peptide lengths were limited to 7-50 amino acids and 500-5000 Da. Up to 3 of each selected variable modification: oxidation (M), phosphorylation (STY) and deamidation (NQ) were allowed. No fixed modifications were selected. Minimum peaks were set to 15 and top N peaks were set to 150. A minimum of 4 fragment ions were required for peak matching, and a minimum spectral count of 3 was required for each peptide ion. An FDR threshold of 0.01 was set for MBR and peptide quantification. Intensities of amelogenin isoform specific peptide signals were calculated by summing the intensity of peptide ions within the SMIRHPY/NMLRPY (bovine) and SIRPPY/SMIRPPY (human) regions of all isoforms of amelogenin proteins.

#### *2.4.3. $MS^1$ Data Processing in Skyline*

$MS^1$ data was processed in Skyline (ver 23.1.0.380). Skyline search parameters were as follows: structural modifications allowed, oxidation (ox); maximum number of modifications, 3; ion match tolerance, 0.001 ppm. High quality, recurrent human unique AMELX and AMELY peptides, identified from $MS^2$ Mascot searches, SIRPPYPSYG (*m/z* 568.790), SIRPPYPSYGYEPM(ox)G (*m/z* 865.400) and SM(ox)IRPPY (*m/z* 440.223), were

scanned for in all samples in human samples, generating extracted ion chromatograms for each peptide. No conserved unique AMELX and AMELY peptides were identified in the cattle samples and therefore $MS^1$ data processing in Skyline was not possible for cattle data.

## 3. Results and Discussion

### *3.1. Comparison of established nanoLC-MS/MS to nanoLC-IMS-MS/MS for proteomic sex estimation*

Established nanoLC-Orbitrap MS/MS was compared to Evosep-timsTOF MS/MS (nanoLC-IMS-MS/MS) as a preliminary test of the applicability of Evosep-timsTOF MS/MS for proteomic sex estimation. Enamel samples of 1 male and 1 female *Bos taurus* were analysed by conventional 60 minute Orbitrap MS method and three different Evosep-timsTOF methods: a 24.0 minute method, a 14.4 minute method, and the 4.8 minute method to gauge the applicability of Evosep-timsTOF MS to amelogenin proteomic sex estimation.

All four compared methods (Figure 4a) were able to identify sufficient numbers of peptides for sex estimation, with the 24.0 and 14.4 minute Evosep-timsTOF methods giving the highest AMELX and AMELY peptide counts. The number of peptides covering the SMIRHPY region of bovine AMELX was used as a proxy for unique peptide count throughout, due to the presence of two AMELX isoforms in the FASTA library, resulting in the reported unique peptide count not being the true unique peptide count. True AMELX unique peptide counts will therefore be slightly higher. Peptides covering the SMIRPHY region of bovine AMELX were selected as a proxy for unique AMELX peptide counts, since the SMIRPHY peptide sequence is unique to bovine AMELX and not found in any other known bovine protein sequence; therefore these peptides are guaranteed to be AMELX peptides. Despite a reduction in peptide counts, percentage coverage of amelogenin was not severely reduced when the 4.8 minute method was implemented (Figure 4b). All four methods also successfully identified non-amelogenin enamel proteins (Figure S.3).

Since comparisons of ion intensities across different MS acquisition methods is not viable, ratios of AMELX/AMELY signal intensities in the male sample were used to compare the Orbitrap and timsTOF methods, where a higher value for the ratio means lower AMELY signal, relative to the AMELX signal. For the Orbitrap method, the AMELX/AMELY ratio was 16.9, while for the timsTOF methods, the ratios were 9.5, 11.7 and 31.7 for the 24.0 minute, 14.4 minute and 4.8 minute methods, respectively. There was also an approximately 10 times reduction in signal intensity when using the 4.8 minute method, compared to the 14.4 and 24.0 minute method.

Although the 4.8 minute Evosep-timsTOF method did give the lowest unique peptide counts and summed peptide ion intensities out of the four compared methods, the data yielded from this analysis was still sufficient for sex assignments, giving estimates that matched the known sex of the samples. While still enabling sex estimation, the 4.8 minute method allowed a dramatic decrease in run time (Figure 5); however, the applicability of the 4.8 minute method

to extremely low signal samples requires further study, and the longer methods discussed here may be necessary in those cases. A more detailed comparison of the four methods is available in S.1.

### *3.2. Comparison of novel neutralisation acid etch sampling to conventional acid etch sampling*

Acid etch sampling of enamel for proteomic sex estimation is now more commonly practised for enamel sampling over powdering enamel following the publication of the Stewart *et al.*[10] method (comparison of data from these two sample preparation methods can be found in SI.2, Figure S.4). Here, a novel neutralised acid etch method was tested to streamline the sample preparation method ahead of Evotip clean-up by removing the ZipTip step of sample preparation (since samples must be injected into the Evosep from an Evotip). The neutralised etch showed higher amelogenin peptide counts than the conventional etching method in both bovine and human samples in timsTOF MS analysis (Figure 6a). The neutralised etch also had higher confidence scores for protein identifications, and an increased number of unique peptides (especially AMELY peptides) were identified, which can increase confidence in the assignment of sex (by decreasing the number of false female assignments due to missed AMELY peptides). The human amelogenin peptide counts were recorded in a similar manner to the cattle amelogenin peptide counts, where the number of peptides covering the SIRRPY and SMIRPPY regions of human AMELX and AMELY, respectively, were used as a proxy for unique peptide count in the human samples. This was chosen since the presence of multiple AMELX and AMELY isoforms in the human enamel proteome complicates the identification of all unique AMELX or AMELY peptides in each sample, similarly to bovine AMELX. For this reason, the true number of unique human AMELX and AMELY peptides was higher than reported here. Peptides covering SIRPPY and SMIRPPY protein sequences were used as proxies for unique AMELX and AMELY peptides, respectively since those peptide sequences are not found in any other known human protein sequence and therefore SIRPPY and SMIRPPY peptides are guaranteed to be human amelogenin peptides. These proxies for AMELX and AMELY unique peptide counts are used throughout. Ameloblastin or enamelin peptide counts were similar or higher in the neutralised samples (Figure S.5), enabling the use of non-amelogenin peptide counts as a parameter for scrutinising amelogenin data quality. Summed amelogenin peptide ion intensities (Figure 6b) were also either improved by using the neutralised method, or were similar across the two etching methods. It should be noted that in the FragPipe searching of two cattle teeth that were etched using the conventional method, AMELX peptides were identified, but no peptides within the SMIRPHY region of AMELX (which the reported peptides were filtered for) were identified, and therefore their AMELX intensities are reported as 0 here.

Comparison of the two etching methods on both cattle and human material highlights the discrepancy between the two taxa in both unique peptide counts and peptide ion intensities, especially for AMELY. Having such drastic differences in data quality between the two taxa undoubtedly highlights a need for further investigation into the current applicability of these methods to bovine samples, especially upon application to more degraded material.

The difference in data quality between the conventional and neutralised etching methods in this experiment may partly be explained by the conventional acid etch method being subjected to two clean-up processes (ZipTip and Evotip), unlike the neutralised etches that were only subjected to Evotip clean-up, which could result in increased losses in the conventional acid etch preparation. Moreover, for the neutralised acid etching method, the full supernatant was transferred to the Evotip, while in the conventional method, a fraction of the reconstituted sample was transferred to the Evotip, so the increased peptide counts in the neutralised acid etches may also be due to the transfer of more material to the Evotips. However, based on this data, conventional acid etch protocol (using ZipTips) is not recommended for Evosep-timsTOF analysis, due to decreased peptide counts, and similar or decreased peptide ion intensities when compared to the neutralised method. Direct Evotip clean-up of untreated acid etches without the ZipTip clean-up stage would also be a viable preparative method for Evosep-timsTOF analysis, but there are some additional benefits associated with use of neutralisation acid etching method. Firstly, addition of base to acid that has enamel suspended in it results in the formation of a precipitate from the calcium phosphate component of enamel[41]; this precipitate acts as a trap for debris (e.g. sediment from burial) upon centrifugation, removing the need for the surface of the teeth to be completely cleaned before etching, or for filtration of the etch. This trapping of the debris/dirt also prevents Evotips from becoming clogged, which could lead to blockages and overpressure on the LC. Moreover, use of neutralised acid in expensive laboratory equipment (such as centrifuges) is more desirable than concentrated acid, as it decreases the risk of corrosive damage. Employing a neutralisation step also means that samples do not have to be evaporated to dryness at any point, further decreasing the overall time taken for sample preparation and removing a potential point of sample loss from the preparation protocol. Analysis of neutralised etches at different time points over a four week time period also demonstrated the etches were generally stable over short storage periods, with best results when the etches are stored in LoBind tubes, demonstrating the base is not adversely affecting the sexually dimorphic peptides within the etch (S.I 3; Figures S.6 and S.7).

### *3.3. Rapid Evosep-timsTOF MS methods for proteomic sex estimation*

The neutralised acid etch method was implemented in the sex estimation of archaeological cattle and human remains to determine the validity of rapid Evosep-timsTOF MS methods to proteomic sex estimation.

#### *3.3.1. Application of rapid Evosep-timsTOF MS methods to archaeological cattle remains*

The neutralised sampling method was applied to archaeological cattle remains of unknown sex from mediaeval Fishergate, York, dating to approximately 11th to 19th century CE[35]. These samples were analysed using the 14.4 minute, 4.8 minute and 2.9 minute Evosep-timsTOF methods (Figure 7).

All three of the applied Evosep-timsTOF MS methods yielded amelogenin peptide counts sufficient for sex estimation. As would be expected, the 14.4 minute method gives the highest number of peptide identifications, however the estimated sexes were not changed when 4.8 minute and 2.9 minute methods were used, demonstrating no advantage for sex estimation in using the 14.4 minute method for these samples. There was less of a discrepancy between the peptide counts from the 2.9 minute and 4.8 minute methods, so use of the 4.8 minute method over the 2.9 minute method was deemed superfluous. Indeed, the additional peptides observed in the 4.8 method were commonly variants of the same peptides seen in the 2.9 minute method, with different variations of modifications. While this means that the confidence that a particular region of the protein was covered by peptides with high quality spectra was likely to be higher in the 4.8 minute analysis, it also means that there were few/no new regions of the protein that were being identified when the longer of the two methods was used. This was further demonstrated by amelogenin percentage protein coverage, where no notable consistent difference was observed between the two acquisition lengths. Non-amelogenin peptide counts were also decreased when the shorter methods were used (Figure S.8). Similar trends to those seen in the peptide counts of these samples were observed in the peptide ion intensities (Figure S.9), where minimal differences were noted between the 4.8 and 2.9 minute methods, while the 14.4 minute method gave increased intensities when compared to the two shorter methods. These comparisons show that an almost doubled analytical run time (4.8 minutes vs 2.9 minutes), and therefore cost of analysis, did not seem to provide sufficiently higher quality data to justify use of the longer of the two methods for this dataset.

From these analyses, individuals *3352* and *4002* were assigned as males, since they had multiple high scoring unique AMELY peptides present, with coverage in the NMRLPPY region of AMELY across all three MS methods. The remaining 4 samples were assigned female due to the lack of AMELY identification across all analyses. Across the various analyses of these bovine samples, the amelogenin peptides identified were varied, and no single AMELX or AMELY peptide was repeatedly observed in all samples; therefore no peptide that could be used as a biomarker peptide for targeted $MS^1$ data processing was identified.

### *3.3.2. Application of rapid Evosep-timsTOF MS methods to archaeological human remains*

The neutralised acid etch method was implemented with MS analysis using the 4.8 minute and 2.9 minute MS methods for sex estimation of human remains from named individuals (Table S.2), dating from 1760-1840 CE from urban Manchester. A subset consisting of one tooth from four individuals was analysed on the 4.8 minute Evosep-timsTOF method for confirmation of the suitability of the 2.9 minute method when compared to the 4.8 minute method (Figure S.10). Peptide counts from the 4.8 minute method and 2.9 minute method showed similar trends as seen in analysis of the Fishergate cattle teeth, with no notable increase in protein coverage, meaning increased peptide counts from 4.8 minute analyses were a result of multiple peptidoforms from the same regions of the amelogenin proteins. The

full set of 10 known sex individuals were analysed using the 2.9 minute Evosep-timsTOF method (Figure 8a).

Detection of amelogenin peptides was successful in the 2.9 minute methods with 100% alignment with the sex assignments based on the names of the individuals, and consistent identification of amelogenin, enamelin and ameloblastin peptides. AMELX identifications were observed in all individuals, and AMELY identifications were only observed in assumed males (based on names). While the Evosep-timsTOF method was successful in estimating the sex of the non-adult remains (HGM 28), the amelogenin peptide counts in this sample were notably lower than the peptide counts in the adults; non-amelogenin peptide counts were also reduced in this sample compared to the adults (Figure S.11). The remains of HGM 28 were physically poorly preserved, which could have contributed to this reduced peptide counts. However both CSM 2.15 and CSM 4.53 also had poorly preserved remains but still showed high SIRPPY peptide counts in the analysis, so it is likely the immaturity of the enamel played a role in the decreased peptide counts[13,22,42]. Similarly, the summed peptide ion intensities for the non-adult tooth were notably lower than the adult teeth, with an approximately 10 times reduction in summed intensities (Figure S.12). As only one nonadult individual was analysed, exploring this observation of reduced counts and intensities in non-adults compared to adult individuals is beyond the scope of this study.

Conserved human AMELX and AMELY peptides (i.e. peptides that were identified in all of the analyses of all relevant human samples) were identified through $MS^2$ data searching (compiled in Tables S.8 and S.9). These peptides were tested for their applicability as protein biomarker peptides in rapid $MS^1$ data searching through processing in Skyline. Peptides SIRPPYPSYG (*m/z* 568.790) and SIRPPYPSYGYEPM(+15.99)G (*m/z* 865.400) were identified as the most suitable AMELX targets, while SM(+15.99)IRPPY (*m/z* 440.223) was identified as the most suitable AMELY target (Figure 8b). All human data was then processed through Skyline, targeting the aforementioned peptides. The identification of these peptides fully aligned with identification of its related protein in the $MS^2$ data searching (Table S.10), demonstrating the success of these peptides as biomarkers for their respective amelogenin proteins in Evosep-timsTOF MS analysis of human teeth. Both $MS^1$ and $MS^2$ data processing are rarely implemented simultaneously for amelogenin proteomic sex estimation, with most studies reporting the use of one or the other[3,11–13,18,20,21,24,27,42–48]. Employing both data processing techniques is beneficial to help ensure accurate sex estimations, as $MS^2$ data processing allows for the identifications of multiple peptides from a single protein, while $MS^1$ data processing can identify peptides that may not have been selected for fragmentation and are consequently not identified in the $MS^2$ data processing stage. $MS^1$ data processing can also be used as a rapid, non-complex data processing method to further decrease workflow time for proteomic sex estimation. Furthermore, use of ameloblastin and enamelin peptide counts can act as a further validation tool to aid scrutinisation of amelogenin peptide data, allowing a more global understanding of the success of peptide identification in the sample. When using the 2.9 minute method, the highest ameloblastin and enamelin peptide counts in the adult samples were 165 and 137, while the lowest were 83 and 53, respectively. All of the adult samples gave sex estimations

that matched with their assumed sex (based on names), with peptide intensities in the range of $10^6$-$10^7$(Figure S.12). Therefore human samples with non-amelogenin peptide counts of approximately 50 or more seem to correlate to high AMELX and AMELY peptide counts that yield reliable sex estimations. Despite this threshold, reliable male sex estimations seem to still be possible below this level, as seen in the duplicate analyses of non-adult HGM 28. For HGM 28, only 7-40 unique ameloblastin or enamelin unique peptides were identified, yet male sex estimation was still possible based on the number of unique AMELX/AMELY peptides, and the presence of the biomarker peptides in the $MS^1$ data. It should be noted that female sex estimates are the greatest false positive risk from low quality data, as this assignment requires confidence in the absence of AMELY. As such, when assigned female, it is recommended to more rigidly stick to the stringent threshold.

**Conclusions**

Here, a number of key methodological developments have enabled the untargeted, rapid identification of amelogenin peptides for sex estimation in cattle and human archaeological remains. Based on the result presented, it has been concluded that neutralisation acid etch sample preparation[49] followed by analysis by the 2.9 minute Evosep-timsTOF MS method is a viable, novel, streamlined, rapid method for high-throughput amelogenin proteomic sex estimation. Longer Evosep-timsTOF MS methods showed higher unique amelogenin peptide counts and peptide ion intensities for both modern and archaeological material, however sex assignments across all compared methods were the same. Further study would be required to understand the impact of the reduced peptide ion intensities and peptide counts on extremely low signal material. This method presents advantages over alternative targeted MS approaches to rapid amelogenin identification, enabling global peptide analysis in similar run times, and easy application to non-hominin remains. Collection of global peptide data enables observation of peptide modifications (such as deamidation) and enables non-amelogenin peptides to be studied, allowing their peptide counts to be employed as an additional validation of data quality. If targeted analysis was preferred, a targeted data processing step using Skyline could be implemented using the applicable AMELX and AMELY biomarker peptides identified here, allowing for a rapid workflow for sex estimation, while still retaining the global protein information obtained from the untargeted MS analysis. Quantitative analysis can also be achieved through implementation of software such as the open access softwares, FragPipe and MSFragger. Where possible, it is recommended that both $MS^1$ and $MS^2$ data processing are implemented, so that sex estimation is not limited by either the presence or absence of a few select peptides, nor the selection of potentially low intensity ions for MS/MS fragmentation. Moreover, implementation of both data processing methods gives two separate sex estimations that can validate one another. The neutralised acid etches were also demonstrated to be generally stable for a 4-week storage period in a LoBind tube, deeming the sampling method established here suitable for short term storage before analysis.

Application of the developed approach to 18th-19th century human individuals gave impressive peptide counts, yielding a 100% match between AMELY identification and male sex as inferred by names. Although peptide counts were consistently lower in cattle extracts

when compared to human extracts, the prospect of rapid sex estimation of cattle remains using Evosep-timsTOF mass spectrometry is still promising. Further investigations of how to improve confidence in cattle sex estimations, by increasing signal intensities and unique amelogenin peptide counts are required before rapid proteomic cattle sex estimation methods can be recommended. Buckley et al[27] also communicated concerns surrounding the reliability of proteomic-based cattle sex estimation.

The method outlined here presents a promising advance in rapid proteomic amelogenin sexing, allowing untargeted enamel data to be obtained in very short MS run times. The time and cost savings realised through this approach could therefore theoretically enable sexing of large sets of remains of any applicable Eutherian species, currently limited only by available amelogenin protein sequences.

**Acknowledgements:** The authors thank the NERC ACCE DTP for funding, and Dr Chloë Baldreki, Dr Samantha Presslee, and Fazeelah Munir for their assistance with laboratory work. The authors also thank Malin Holst for providing human materials and context information. The York Centre of Excellence in Mass Spectrometry was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from EPSRC (EP/K039660/1; EP/M028127/1).

**Data availability:** The MS datasets generated during this study are available in the ProteomeXchange Consortium (dataset identifier PXD054643) (https://www.proteomexchange.org/) via the MassIVE partner repository (dataset identifier MSV000095527).

**CRediT authorship contribution statement:**
**C. Blacka:** Investigation, Methodology, Formal analysis, Visualisation, Writing - original draft, Writing - review & editing. **A. Dowle:** Formal analysis, Writing - review & editing. **M. Lisowski:** Resources, Writing - review & editing. **M. Alexander**: Resources, Writing - review & editing. **K. Penkman**: Conceptualisation, Funding acquisition, Visualisation, Supervision, Methodology, Writing - review & editing. **J. Hendy**: Conceptualisation, Funding acquisition, Visualisation, Supervision, Methodology, Writing - review & editing. **J. Mosely**: Visualisation, Supervision, Methodology, Writing - review & editing.

**Declaration of competing interests:** The authors declare that there are no competing interests.

**References:**

1. Shirazi S, Broomandkhoshbacht N, Oppenheimer J, et al. Ancient DNA-based sex determination of bison hide moccasins indicates Promontory cave occupants selected female hides for footwear. *J Archaeol Sci*. 2022;137:105533. doi:10.1016/j.jas.2021.105533

2. Fages A, Seguin-Orlando A, Germonpré M, Orlando L. Horse males became over-represented in archaeological assemblages during the Bronze Age. *J Archaeol Sci Rep*. 2020;31:102364. doi:10.1016/j.jasrep.2020.102364

3. Berezina N, Ziganshin R, Kolobova K, et al. Bison sex matters: the potential of proteomic tooth enamel analysis for determination of ancient human subsistence strategies. *Archaeol Anthropol Sci*. 2024;16(9):142. doi:10.1007/s12520-024-02053-x

4. Madupe PP, Koenig C, Patramanis I, et al. Enamel proteins reveal biological sex and genetic variability within southern African Paranthropus. *bioRxiv*. Published online July 3, 2023:2023.07.03.547326. doi:10.1101/2023.07.03.547326

5. Demarchi B, Hall S, Roncal-Herrero T, et al. Protein sequences bound to mineral surfaces persist into deep time. *Elife*. 2016;5. doi:10.7554/eLife.17092

6. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. *Nat Rev Genet*. 2001;2(5):353-359. doi:10.1038/35072071

7. Fincham AG, Moradian-Oldak J, Simmer JP. The structural biology of the developing dental enamel matrix. *J Struct Biol*. 1999;126(3):270-299. doi:10.1006/jsbi.1999.4130

8. Lau EC, Mohandas TK, Shapiro LJ, Slavkin HC, Snead ML. Human and mouse amelogenin gene loci are on the sex chromosomes. *Genomics*. 1989;4(2):162-168. doi:10.1016/0888-7543(89)90295-4

9. Nakahori Y, Takenaka O, Nakagome Y. A human X-Y homologous region encodes "amelogenin." *Genomics*. 1991;9(2):264-269. doi:10.1016/0888-7543(91)90251-9

10. Stewart NA, Molina GF, Issa JPM, et al. The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction. *RSC Adv*. 2016;6(66):61673-61679. doi:10.1039/C6RA05120K

11. Stewart NA, Gerlach RF, Gowland RL, Gron KJ, Montgomery J. Sex determination of human remains from peptides in tooth enamel. *Proc Natl Acad Sci U S A*. 2017;114(52):13649-13654. doi:10.1073/pnas.1714926115

12. Lugli F, Di Rocco G, Vazzana A, et al. Enamel peptides reveal the sex of the Late Antique "Lovers of Modena." *Sci Rep*. 2019;9(1):13130. doi:10.1038/s41598-019-49562-7

13. Parker GJ, Yip JM, Eerkens JW, et al. Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel. *J Archaeol Sci*. 2019;101:169-180. doi:10.1016/j.jas.2018.08.011

14. Wasinger VC, Curnoe D, Bustamante S, et al. Analysis of the Preserved Amino Acid Bias in Peptide Profiles of Iron Age Teeth from a Tropical Environment Enable Sexing of Individuals Using Amelogenin MRM. *Proteomics*. 2019;19(5):e1800341. doi:10.1002/pmic.201800341

15. Lugli F, Figus C, Silvestrini S, et al. Sex-related morbidity and mortality in non-adult individuals from the Early Medieval site of Valdaro (Italy): the contribution of dental enamel peptide analysis. *J Archaeol Sci Rep*. 2020;34:102625.

doi:10.1016/j.jasrep.2020.102625

16. Rebay-Salisbury K, Janker L, Pany-Kucera D, et al. Child murder in the Early Bronze Age: proteomic sex identification of a cold case from Schleinbach, Austria. *Archaeol Anthropol Sci*. 2020;12(11):265. doi:10.1007/s12520-020-01199-8

17. Cintas-Peña M, Luciañez-Triviño M, Montero Artús R, et al. Amelogenin peptide analyses reveal female leadership in Copper Age Iberia (c. 2900-2650 BC). *Sci Rep*. 2023;13(1):9594. doi:10.1038/s41598-023-36368-x

18. Granja R, Araújo AC, Lugli F, Silvestrini S, Silva AM, Gonçalves D. Unbalanced sex-ratio in the Neolithic individuals from the Escoural Cave (Montemor-o-Novo, Portugal) revealed by peptide analysis. *Sci Rep*. 2023;13(1):19902. doi:10.1038/s41598-023-47037-4

19. Kotli P, Morgenstern D, Nagar Y, et al. Antiquity validation by amelogenin peptidomics profile: Sex determination in an ascetic burial. *Research Square*. Published online August 24, 2023. doi:10.21203/rs.3.rs-3279131/v1

20. Olszewski J, Hall RA, Kootker LM, et al. Osteological, multi-isotope and proteomic analysis of poorly-preserved human remains from a Dutch East India Company burial ground in South Africa. *Sci Rep*. 2023;13(1):14666. doi:10.1038/s41598-023-41503-9

21. Gamble JA, Spicer V, Hunter M, et al. Advancing sex estimation from amelogenin: Applications to archaeological, deciduous, and fragmentary dental enamel. *J Archaeol Sci Rep*. 2024;54:104430. doi:10.1016/j.jasrep.2024.104430

22. Cleland TP, McGuire SA, Beatrice JS, Moran KS, France CAM. SPEED-E: A modified version of the sample preparation by Easy extraction and Digestion(-free) protocol for enamel-based sex estimation in archaeological remains. *J Archaeol Sci*. 2024;168:106006. doi:10.1016/j.jas.2024.106006

23. Casas-Ferreira AM, del Nogal-Sánchez M, Arroyo ÁE, Vázquez JV, Pérez-Pavón JL. Fast methods based on mass spectrometry for peptide identification. Application to sex determination of human remains in tooth enamel. *Microchem J*. 2022;181:107645. doi:10.1016/j.microc.2022.107645

24. Koenig C, Bortel P, Paterson RS, et al. Automated high-throughput biological sex identification from archaeological human dental enamel using targeted proteomics. *bioRxiv*. Published online February 22, 2024:2024.02.20.581140. doi:10.1101/2024.02.20.581140

25. Cappellini E, Welker F, Pandolfi L, et al. Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. *Nature*. 2019;574(7776):103-107. doi:10.1038/s41586-019-1555-y

26. Kotli P, Morgenstern D, Horwitz L, Khalaily H, Bocquentin F, Boaretto E. A peptidomics method for assessing sex from modern and ancient bovine tooth enamel. *Research Square*. Published online January 29, 2024. doi:10.21203/rs.3.rs-3766361/v1

27. Buckley M, Pigière F, Pal Chowdhury M, Kitchener AC, Smyth J. Proteomic sexing of archaeological cattle remains at Neolithic Kilshane. *J Archaeol Sci*.

2024;172(106102):106102. doi:10.1016/j.jas.2024.106102

28. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45(D1):D158-D169. doi:10.1093/nar/gkw1099

29. van Doorn NL, Wilson J, Hollund H, Soressi M, Collins MJ. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun Mass Spectrom*. 2012;26(19):2319-2327. doi:10.1002/rcm.6351

30. Aballo TJ, Roberts DS, Melby JA, Buck KM, Brown KA, Ge Y. Ultrafast and Reproducible Proteomics from Small Amounts of Heart Tissue Enabled by Azo and timsTOF Pro. *J Proteome Res*. 2021;20(8):4203-4211. doi:10.1021/acs.jproteome.1c00446

31. Akkurt Arslan M, Kolman I, Pionneau C, et al. Proteomic Analysis of Tears and Conjunctival Cells Collected with Schirmer Strips Using timsTOF Pro: Preanalytical Considerations. *Metabolites*. 2021;12(1). doi:10.3390/metabo12010002

32. Meier F, Park MA, Mann M. Trapped Ion Mobility Spectrometry and Parallel Accumulation–Serial Fragmentation in Proteomics. *Mol Cell Proteomics*. 2021;20. doi:10.1016/j.mcpro.2021.100138

33. Ridgeway ME, Lubeck M, Jordens J, Mann M, Park MA. Trapped ion mobility spectrometry: A short review. *Int J Mass Spectrom*. 2018;425:22-35. doi:10.1016/j.ijms.2018.01.006

34. Meier F, Beck S, Grassl N, et al. Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J Proteome Res*. 2015;14(12):5378-5387. doi:10.1021/acs.jproteome.5b00932

35. O'Connor TP. Bones from 46-54 Fishergate. In: Addyman PV, Black VE, eds. *The Animal Bones*. Vol 15. The Archaeology of York. Council for British Archaeology for York Archaeological Trust; 1991:91.

36. Chidimuro B, Holst M, Newman S, Keefe K, Collins MJ, Alexander M. Grim up North? Exploring the diet of urban populations in post-medieval Greater Manchester, England, using stable isotope analysis. *Archaeol Anthropol Sci*. 2023;15(11):174. doi:10.1007/s12520-023-01881-7

37. Newman S, Holst M. *Osteological Analysis, Chapel Street, Hazel Grove, Greater Manchester*. York Osteoarchaeology; 2016.

38. Keefe K, Holst M. *Osteological Analysis, Cross Street Unitarian Chapel. Manchester*. York Osteoarchaeology; 2017.

39. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods*. 2017;14(5):513-520. doi:10.1038/nmeth.4256

40. Yu F, Haynes SE, Teo GC, Avtonomov DM, Polasky DA, Nesvizhskii AI. Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Mol Cell*

*Proteomics*. 2020;19(9):1575-1585. doi:10.1074/mcp.TIR120.002048

41. Dickinson MR, Lister AM, Penkman KEH. A new method for enamel amino acid racemization dating: A closed system approach. *Quat Geochronol*. 2019;50:29-46. doi:10.1016/j.quageo.2018.11.005

42. Gowland R, Stewart NA, Crowder KD, et al. Sex estimation of teeth at different developmental stages using dimorphic enamel peptide analysis. *Am J Phys Anthropol*. 2021;174(4):859-869. doi:10.1002/ajpa.24231

43. Marta DZB, Gonzalo AJ, Margarita SR, et al. Female sex bias in Iberian megalithic societies through bioarchaeology, aDNA and proteomics. *Sci Rep*. 2024;14(1):21818. doi:10.1038/s41598-024-72148-x

44. Kimsis J, Petersone-Gordina E, Poksane A, et al. Application of natural sciences methodology in archaeological study of Iron Age burials in Latvia: pilot study. *Forensic Sci Med Pathol*. Published online November 8, 2022. doi:10.1007/s12024-022-00553-7

45. Gasparini A, Lugli F, Silvestrini S, et al. Biological sex VS. Archaeological Gender: Enamel peptide analysis of the horsemen of the Early Middle age necropolises of Campochiaro (Molise, Italy). *J Archaeol Sci Rep*. 2022;41:103337. doi:10.1016/j.jasrep.2021.103337

46. Greco E, Gennaro AM, Piombino-Mascali D, et al. Dental Proteomic Analyses and Raman Spectroscopy for the Determination of the Biological Sex and Age of Human Remains from the Greek Cemetery of San Giorgio Extra, Reggio Calabria (Italy). *Reggio Calabria*. Published online July 26, 2023. doi:10.2139/ssrn.4509097

47. Mays S, Parker G, Johns C, et al. Sex identification of a Late Iron Age sword and mirror cist burial from Hillside Farm, Bryher, Isles of Scilly, England. *Journal of Archaeological Science: Reports*. Published online July 27, 2023:104099. doi:10.1016/j.jasrep.2023.104099

48. Froment C, Hourset M, Sáenz-Oyhéréguy N, et al. Analysis of 5000 year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides. *J Proteomics*. 2020;211:103548. doi:10.1016/j.jprot.2019.103548

49. Blacka C. Neutralisation acid etching of enamel for proteomic amelogenin sex estimation. Protocols.io. doi:10.17504/protocols.io.14egn6qr6l5d/v1

## Figures

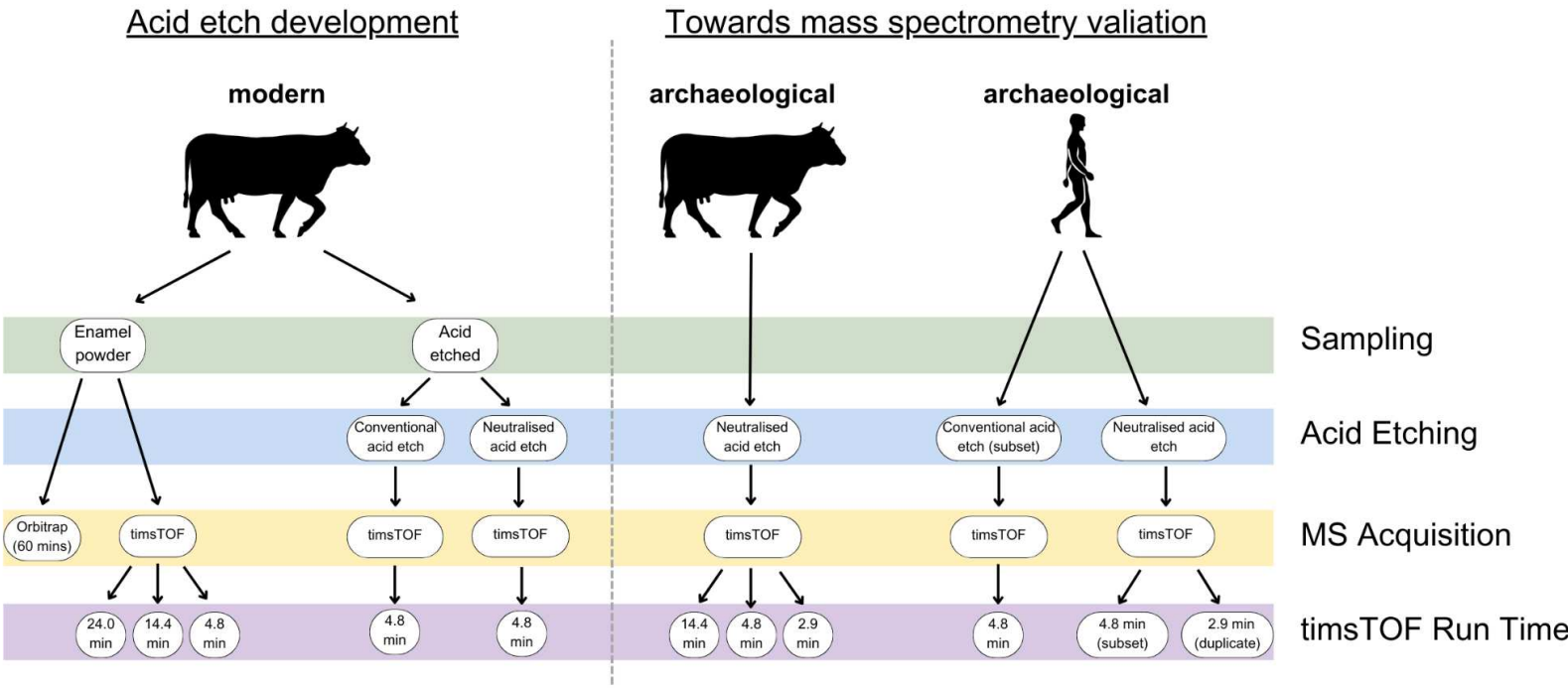
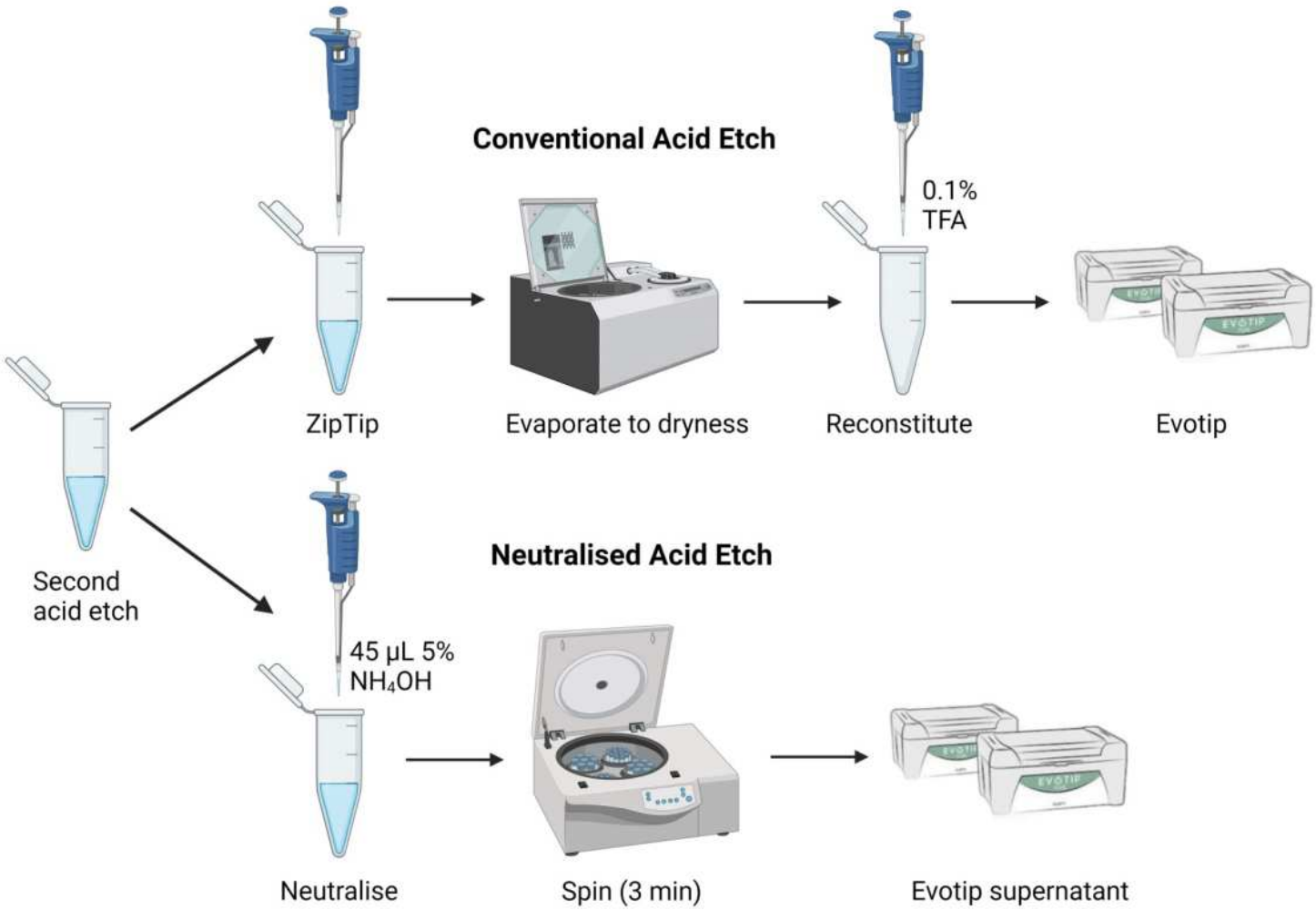



**Figure 1.** Summary of methodological and analytical variables applied to the three different sets of sample material discussed here. Each coloured background represents differences in variables: in green, sampling; blue, acid etching approach; yellow, MS instrument; purple, timsTOF run length.

812




813 Figure 2. a) Summary of the conventional acid etch[10] and the novel neutralised acid etch
814 method when Evosep LC separation is implemented (figure created with BioRender.com).
815

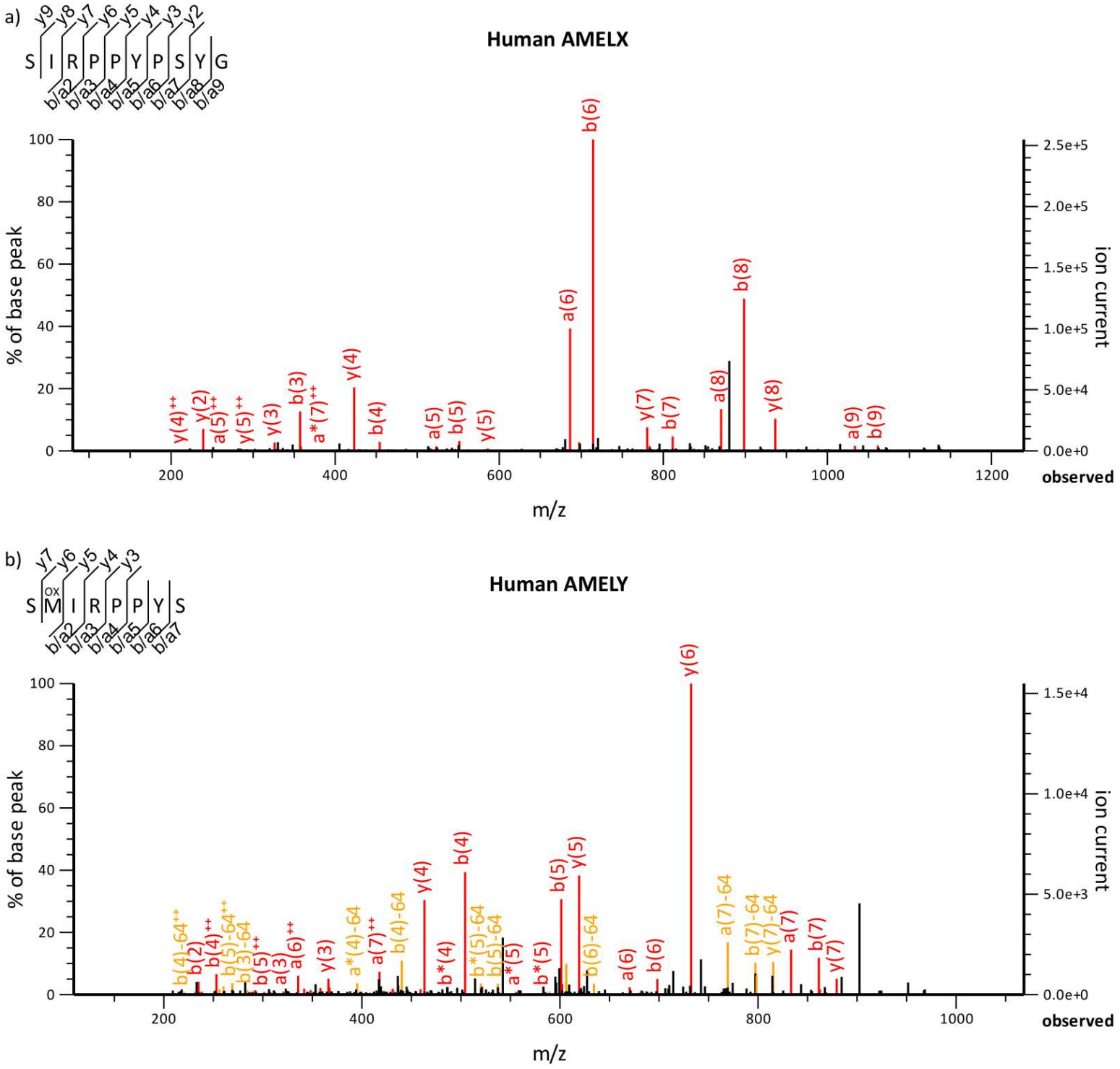



Figure 3. Example MS/MS spectra of a) SIRPPYPSYG (human AMELX) and b) SM(ox)IRPPYS (human AMELY). Spectra were annotated in Mascot.

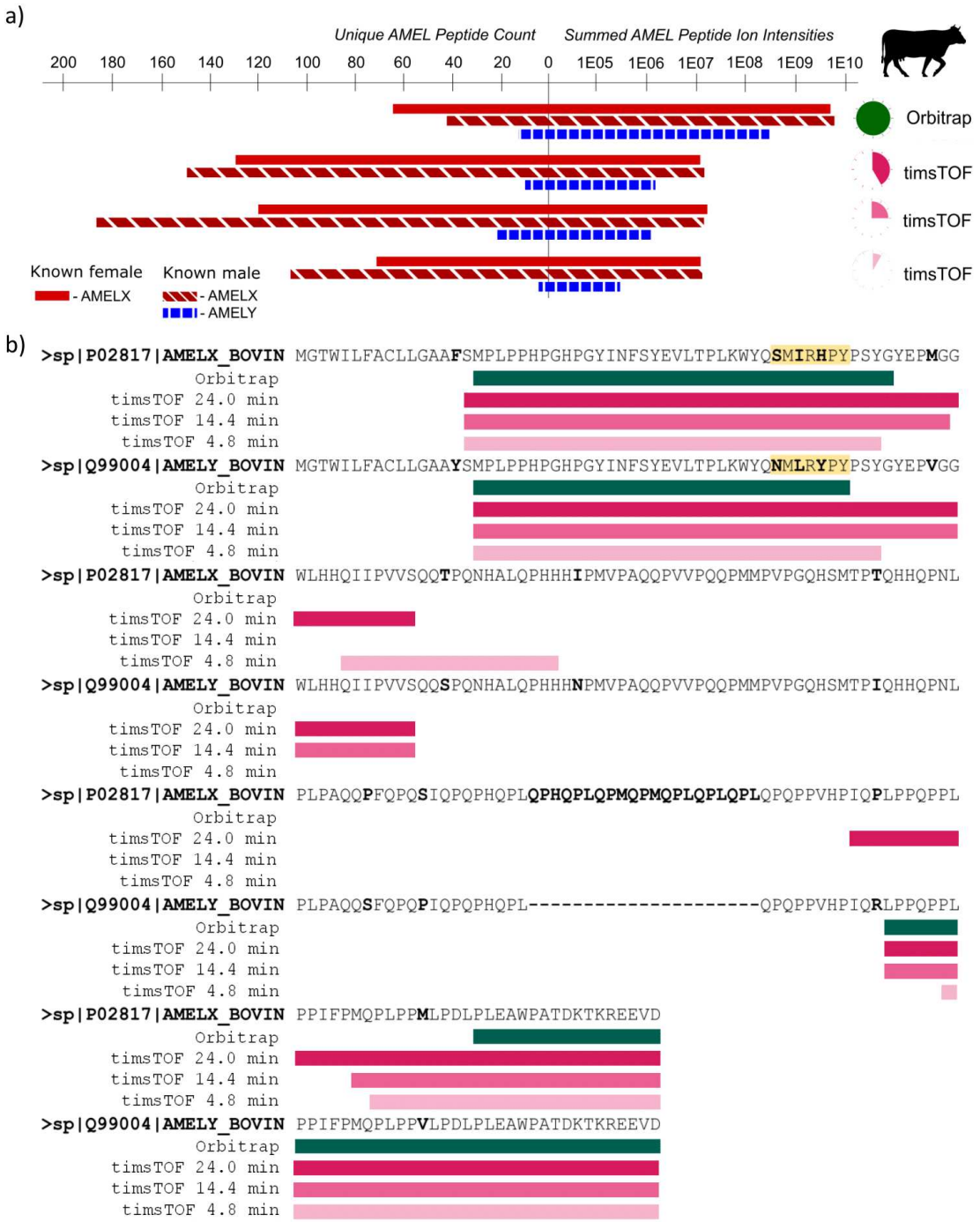



Figure 4. a) Unique AMELX (red) and AMELY (blue) peptide counts (left) and summed AMELX (red) and AMELY (blue) peptide ion intensities (right) from powdered enamel from a modern female (filled bars) and male (hashed bars) *Bos taurus* tooth when analysed by Orbitrap Fusion MS, and three different Evosep-timsTOF MS methods (24.0 minute, 14.4 minute, and 4.8 minute). b) Aligned bovine AMELX (Uniprot accession number P02817) and AMELY (Uniprot accession number Q99004) protein sequences, showing the protein coverage observed in the modern male tooth across the different MS methods. Bold letters indicate where there are differences in the AMELX and AMELY sequence. The yellow highlighted regions indicate the regions where most dimorphic peptides are identified in bovine samples.

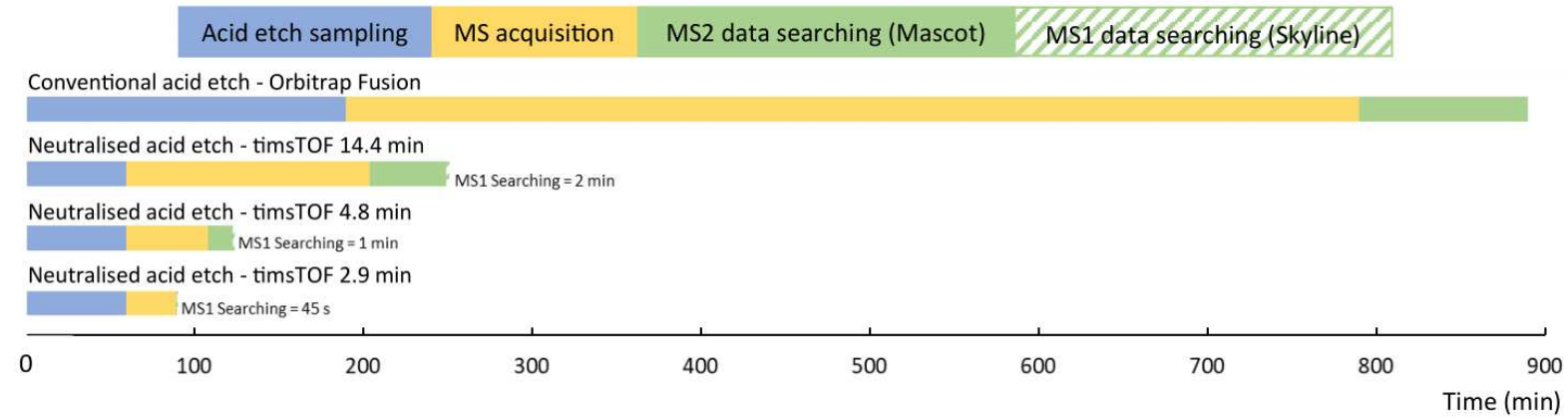



Figure 5. Comparisons of the time taken to perform peptidic amelogenin sex estimation for 10 samples with the main methods discussed in this paper. From top to bottom: conventional acid etching followed by Orbitrap Fusion MS; neutralised acid etching followed by 14.4 min Evosep-timsTOF MS; neutralised acid etching followed by 4.8 min Evosep-timsTOF MS; neutralised acid etching followed by 2.9 min Evosep-timsTOF MS. Note: $MS^1$ data searching time is omitted from the Orbitrap Fusion time scale as $MS^1$ data searching was only possible with the human samples, and no human teeth were analysed by Orbitrap Fusion MS.

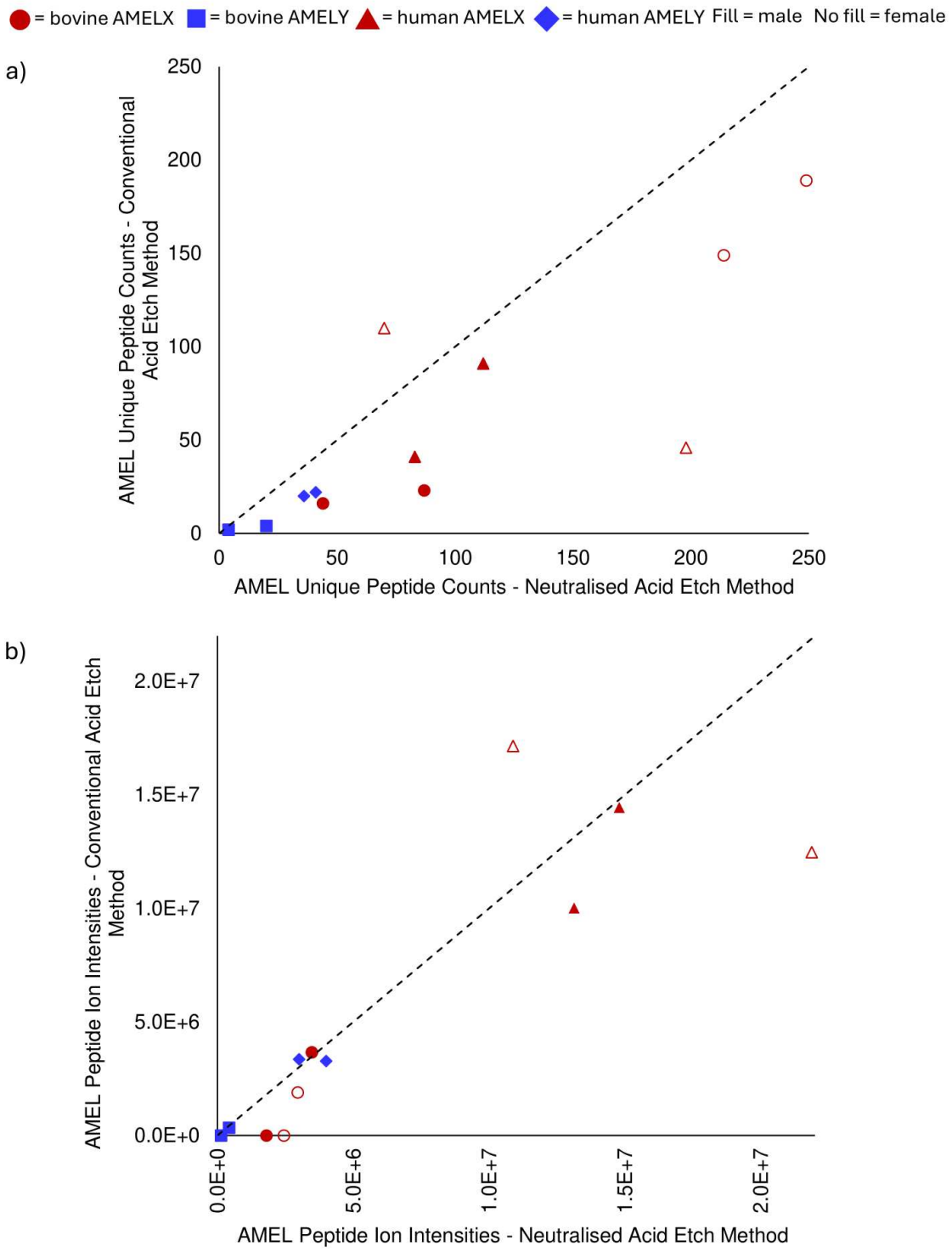



Figure 6. a) Unique AMELX (red) and AMELY (blue) peptide counts when the conventional acid etch approach is implemented versus the neutralised acid etch approach for sampling of cattle teeth and human teeth. b) AMELX (red) and AMELY (blue) peptide ion intensities, calculated from summing the intensities of the peptide ions in the SMIRPHY/NMLRPY (bovine) and SIRPPY/SMIRPPY (human) regions of AMELX/AMELY when the conventional acid etch approach is implemented versus the neutralised acid etch approach for sampling of cattle teeth and human teeth. MS analysis was performed using the 4.8 minute method.

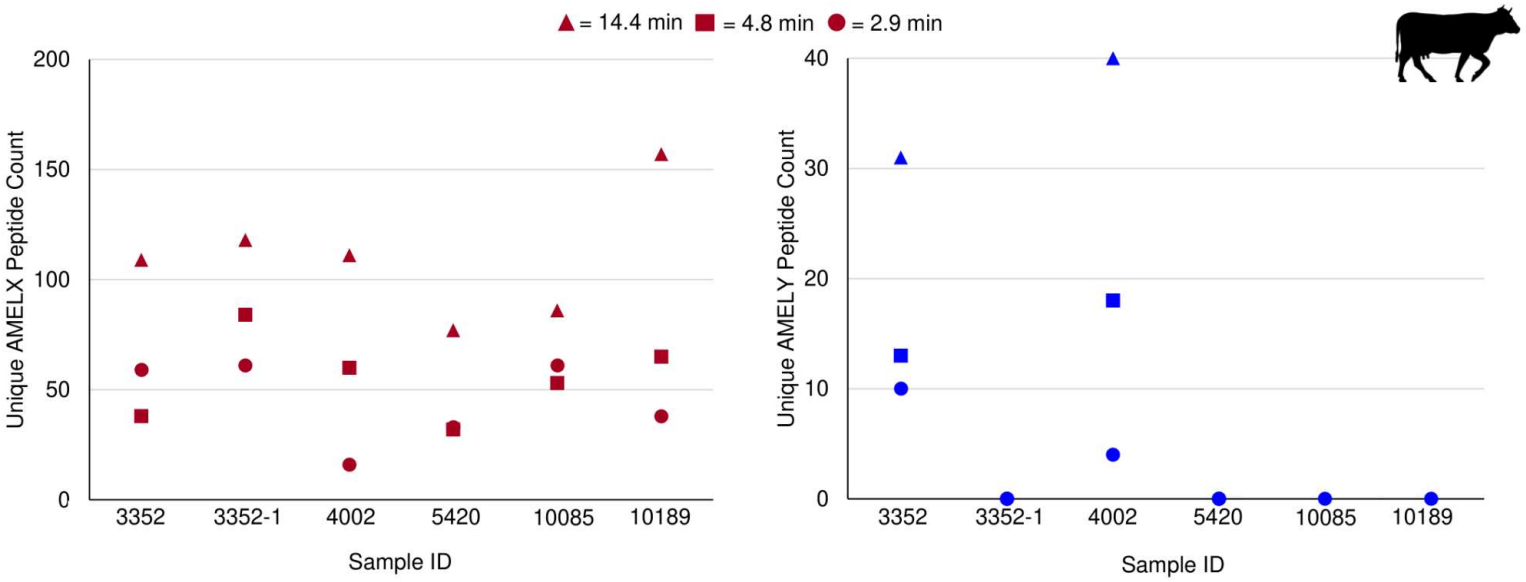



Figure 7. AMELX (left; red) and AMELY (right; blue) unique peptide counts values from analysis of the neutralised acid etched mediaeval Fishergate cow mandible teeth using the 14.4 min (triangles), 4.8 min (squares), and 2.9 min (circles) Evosep-timsTOF methods. Note the difference in scale in the y-axes.

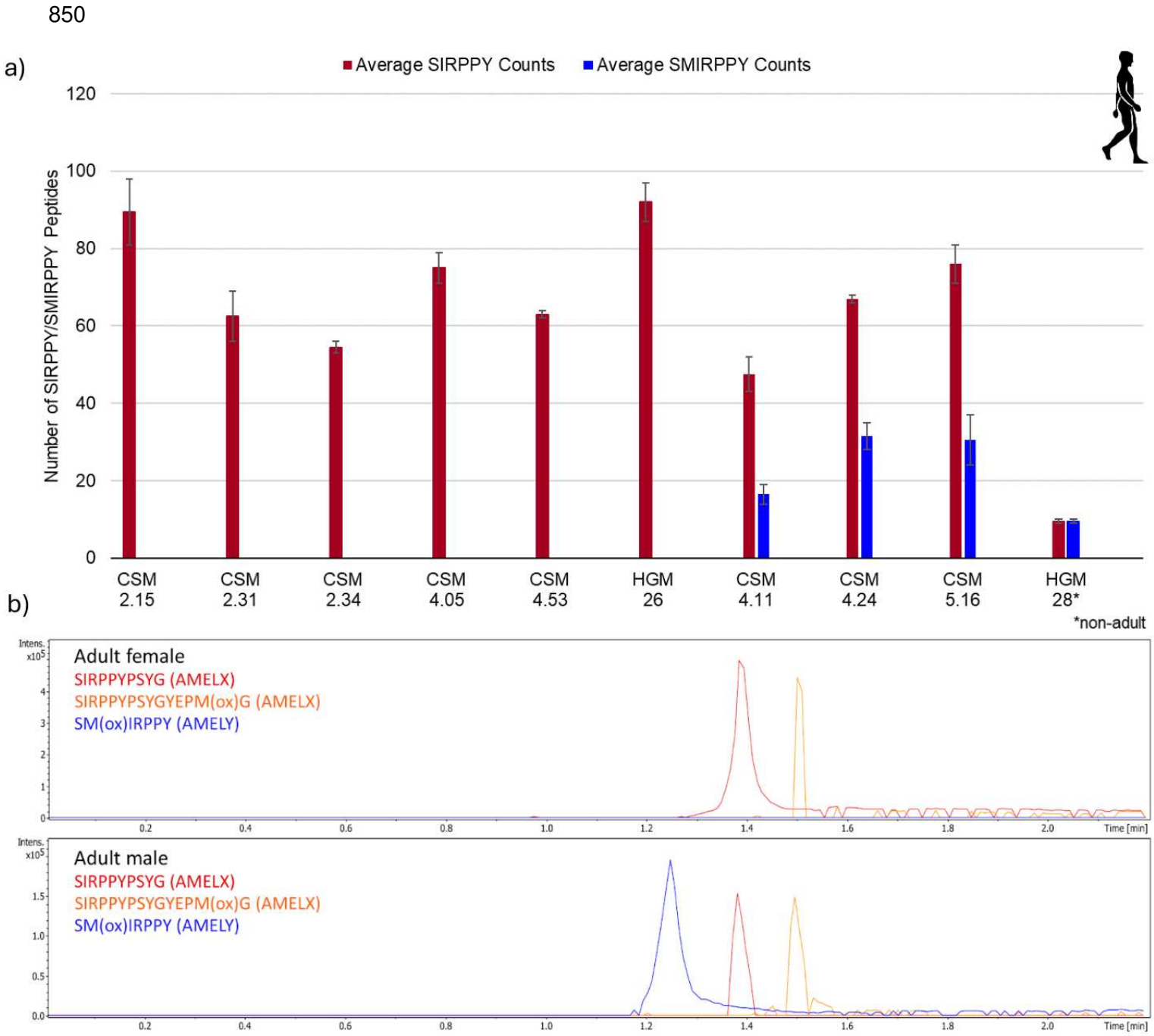



Figure 8. a) The mean number of peptides in the SIRRPY (red) and SMIRPPY (blue) regions of human AMELX and AMELY, respectively, of ten neutralised acid etched 18th-19th century human teeth, analysed with the 2.9 minute Evosep-timsTOF method. Error bars represent 1 standard deviation of 2 technical replicates. b) overlaid EICs of unique AMELX (red, SIRPPYPSYG; orange, SIRPPYPSYGYEPM(+15.99)G) and AMELY (blue, SM(+15.99)IRPPY) peptide *m/z* values in an adult female (top) and an adult male (bottom). Analysis was performed with the 2.9 minute timsTOF MS method. Error tolerance for each EIC was: +/- 0.001 *m/z*.