



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/224363/>

Version: Accepted Version

Article:

Blacka, Charllotte, Dowle, Adam, Lisowski, Mik et al. (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>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

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

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

3
4 **Short Title:** Rapid amelogenin sex estimation using untargeted mass spectrometry

5
6 **Authors:** Charllotte Blacka^{1,2}, Adam Dowle³, Mik Lisowski¹, Michelle Alexander¹, Jessica
7 Hendy¹, Kirsty Penkman², Jackie Mosely²

8
9 **Affiliations:** ¹Department of Archaeology, University of York, YO10 5DD, UK;
10 ²Department of Chemistry, University of York, YO10 5DD, UK; ³Bioscience Technology
11 Facility, Department of Biology, University of York, YO10 5DD, UK.

12
13 **Correspondence:** Charllotte Blacka, Department of Archaeology, University of York, York
14 YO10 5DD, UK; Email: csb548@york.ac.uk; Telephone: 01904 324472.

15
16 **Funding information:** This project is part of the Natural Environment Research Council
17 (NERC) Doctoral Training Partnership – Adapting to the Challenges of a Changing
18 Environment (ACCE) [grant number NE/S00713X/1].

19
20 **Keywords:** amelogenin, palaeoproteomics, sex estimation, LC-MS/MS, TIMS-TOF

21
22 For the purpose of open access, the author has applied a Creative Commons Attribution (CC
23 BY) licence to any Author Accepted Manuscript version arising for this submission.

24 **Abstract**

25
26
27 **Rationale:** Sex estimation by analysis of amelogenin peptides in archaeological and fossil
28 material has recently been gaining great traction within the fields of archaeology and
29 palaeontology. Current widely used proteomic amelogenin sex estimation methods are
30 hindered by relatively long mass spectrometric run times, or targeting peptides specific to
31 human amelogenin proteins. Untargeted, high-throughput amelogenin sexing would be
32 invaluable for a range of applications, from sex estimation of remains at mass grave sites to
33 broadening the application of rapid amelogenin sexing to non-hominin species for husbandry
34 and evolutionary studies.

35
36 **Methods:** A new acid etch protocol followed by Evosep-LC-TIMS-TOF mass spectrometry
37 is presented for amelogenin analysis, providing global peptide data through rapid mass
38 spectrometric methods in under 20 minutes per sample (including sample preparation, mass
39 spectrometric acquisition and data processing). This sampling protocol was developed on
40 modern cattle (*Bos taurus*) teeth, before Evosep-timsTOF partial validation with
41 archaeological cattle and human (*Homo sapiens*) teeth, demonstrating the potential of
42 straightforward application of this rapid amelogenin sexing method to a range of taxa.

43

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

49

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

54

55 1. Introduction

56 Biological sex estimation of humans and animals is a key facet of archaeology, contributing
57 to our understanding of the treatment of the dead, and relationships between sex and gender.
58 For zooarchaeology, accurate sex estimation is vital in studies of sex-selective husbandry or
59 hunting practices¹, and even ideology², while in palaeontology it can help resolve herd
60 dynamics, subsistence strategies³ and identify sexual dimorphism⁴. Sex estimation can be
61 achieved by osteological assessment, the detection of x and y chromosomes using ancient
62 DNA (aDNA), and more recently, mass spectrometric (MS) analysis of amelogenin peptides.
63 Amelogenin peptide sex estimation presents some advantages over the former methods, with
64 peptides surviving for much longer than aDNA^{5,6}, and enabling sex estimation of non-adult
65 remains, which is not possible via osteological analysis. Peptidic sex estimation is achieved
66 through the identification of the enamel-specific amelogenin proteins. Amelogenins are
67 highly conserved and are the most abundant proteins in enamel, constituting more than 90%
68 of its organic fraction⁷. They are encoded on the sex chromosomes of many Eutherian
69 mammals, resulting in expression of sexually dimorphic X- and Y-isoforms of amelogenin,
70 termed AMELX and AMELY, respectively^{8,9}. The small differences in protein sequence
71 between these sexually dimorphic proteins can be identified using MS techniques, enabling
72 proteomic sex estimation.

73

74 Most palaeoproteomic amelogenin studies to date employ high resolution mass spectrometry
75 (namely Orbitrap MS) for protein analysis, with analytical run times of approximately 60
76 minutes¹⁰⁻²², although faster methods have recently been developed^{23,24} (Figure S.1). When
77 working with large assemblages (such as mass graves or population studies), these long MS
78 run times (not including the accompanying preparative and data analysis times), and resulting
79 high costs, preclude estimation of the sex of all individuals within a horizon/site, and so fully
80 understanding population demographics is impossible. This inability to study large sample
81 sets also means that there is limited applicability of amelogenin analysis to zooarchaeological
82 sites, as highlighted by the comparative sparsity of zooarchaeological amelogenin
83 publications^{3,25-27}, a problem that is also compounded by the lack of reliable amelogenin
84 reference sequences for mammalian taxa in UniProt²⁸. Previous approaches to develop high-
85 throughput amelogenin sex estimation methods for human remains have employed targeted
86 mass spectrometry methods, where the MS instrument is programmed to analyse a few
87 peptides only, and the rest of the ions generated are disregarded^{23,24}. These targeted

88 approaches significantly shorten run times and data complexity, but result in the loss of all
89 protein/peptide information unrelated to these target peptides, meaning studying non-
90 amelogenin enamel proteins or peptide modifications (such as deamidation for estimating the
91 extent of degradation²⁹) is not possible. Furthermore, these methods cannot be directly
92 applied to non-hominin remains, since the targeted peptides are unique to hominins.

93

94 Here an untargeted, but rapid, proteomic amelogenin sex estimation method is presented,
95 allowing for low-cost analysis of many taxa. The ease of application of the same method to
96 different taxa is showcased through application to both cattle and human teeth. Trapped ion
97 mobility separation-time of flight mass spectrometry (TIMS-TOF MS) is employed to
98 achieve this, demonstrating the first reported use of ion mobility MS for amelogenin sexing.
99 Although yet rarely used in palaeoproteomics, modern proteomic studies acclaim the
100 suitability of TIMS-TOF MS for rapid analysis of low concentration samples³⁰⁻³², akin to
101 those analysed in bioarchaeological studies. TIMS-TOF MS differs from other forms of MS
102 usually utilised in amelogenin sex estimation by implementing a trapped ion mobility
103 separation (TIMS) after the liquid chromatography (LC) separation. TIMS utilises the
104 balancing of two opposing forces experienced by peptide ions in the TIMS cell: the drag from
105 a gas flow heading into the instrument and an increasing opposing DC current that repels the
106 ions³³. Once these forces balance for a specific ion, the ion becomes trapped in space, before
107 the current is reduced in a stepwise fashion, sequentially releasing ‘packets’ of ions into the
108 MS. Ions with a larger rotationally averaged collisional cross-section (CCS; size or shape)
109 experience more drag and travel further through the TIMS cell before being trapped,
110 compared to ions with the same m/z but a smaller CCS. The timsTOF range of mass
111 spectrometers use two sequential TIMS cells to enable parallel accumulation-serial
112 fragmentation (PASEF), whereby the first cell is used to trap and hold ions as bulk ‘packets’
113 (increasing sensitivity), while the second cell steps down through iterations of DC current to
114 release ions into the mass spectrometer proportionally to their CCS (increasing separation and
115 dynamic range). The quadrupole, where fragmentation occurs, is operated intelligently and
116 aligned with the ions leaving the second TIMS cell, to maximise speed of sampling enabling
117 hundreds of MS/MS events per second at full sensitivity³⁴. Ion mobility is implemented in
118 this study as a means to separate ions based on their rotationally averaged CCS post
119 chromatography and pre MS analysis. This enables rapid targeting of ions for MS/MS
120 analysis.

121

122 Method development consisting of both sample preparation development and partial
123 validation of the applicability of Evosep-timsTOF mass spectrometry to proteomic
124 amelogenin sex estimation is presented herein. The sample preparation method was
125 developed using modern cattle (*Bos taurus*) teeth, before testing on different rapid Evosep-
126 timsTOF MS methods. Then, to test the applicability of this MS method two archaeological
127 case studies were analysed: archaeological cattle teeth from the 46-54 Fishergate excavations,
128 York³⁵; and teeth from 18th-19th century human individuals from Cross Street Unitarian
129 Chapel, Manchester, and Hazel Grove, Church Street, Manchester³⁶.

130

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 CE³⁵ (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) protocol¹⁰. These etches were subjected to ZipTip clean-up. The ZipTip was conditioned as follows: the tip was cleaned with 100 µL MeOH, then 100 µL AT80 (80% 0.1% v/v TFA in ACN: 20% 0.1% v/v TFA in H₂O), then 100 µL 0.1% v/v TFA in H₂O. The etch was then passed through the tip 20 times before the tip was washed twice with 100 µL 0.1% v/v TFA in H₂O. The peptides were then eluted into a 1.5 mL LoBind tube with 100 µL 40% v/v ACN in H₂O. The eluted peptide solution was dried down using a rotary vacuum concentrator and the

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

184

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

194

195 The novel neutralised acid etch method implemented a neutralisation step after acid etching.
196 The teeth were washed with 3% *v/v* H_2O_2 in water, then rinsed with ultrapure water (18.2
197 $\text{M}\Omega\text{-cm}$). 60 μL of 5% *v/v* HCl in water was pipetted into the cap of a 0.2 mL PCR tube,
198 leaving a convex meniscus of HCl in the cap. The tooth was lowered into the HCl and contact
199 with the acid was maintained for 2 minutes. The first etch was discarded and a second etch
200 was performed by lowering the same area of the tooth into 60 μL of 5% *v/v* HCl in water
201 once more for 2 minutes. The second etch was retained and 45 μL 5% *v/v* NH_4OH in water
202 was added, resulting in a white precipitate forming. The sample was briefly mixed with a
203 vortex mixer before being centrifuged at 13,250 rcf for 3-4 minutes to separate the precipitate
204 and the liquid. The supernatant was transferred to a new 0.2 mL PCR tube and stored in the
205 freezer until MS analysis was performed. Where analytical replicate analysis was performed,
206 a 320 μL acid etch was performed in the cap of a 1.5 mL centrifuge tube and split into 60 μL
207 aliquots in 0.2 mL PCR tubes, followed by treatment of the etches as described prior for the
208 60 μL etches.

209

210 2.2.3. Preparation of Powdered Enamel Samples

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

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

222

223 2.3. LC-MS/MS Data Acquisition

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

227

228 2.3.1. Orbitrap Fusion MS Analysis

229 The powdered enamel peptides were loaded onto an M-Class nanoflow UPLC system
230 (Waters Corporation) fitted with a nanoEaze M/S Symmetry 100 \AA C₁₈, 5 μm 180 μm x 20
231 mm trap column (Waters Corporation) and a PepMap, 2 μm , 100 \AA C₁₈ EasyNano
232 nanocapillary 75 μm x 500 mm column (Thermo Fisher Scientific Inc.). 0.05% v/v TFA in
233 water was used as a trap wash solvent and the trapping flow rate was 15 $\mu\text{L}/\text{min}$. The trap
234 was washed for 5 min prior to the flow travelling to the capillary column. The injection
235 volume for each sample was 6 μL and elution was achieved using gradient elution: aqueous
236 0.1% v/v formic acid in water (mobile phase A) and acetonitrile containing 0.1% v/v formic
237 acid in water (mobile phase B). Solvent composition was altered as follows: 3-10% mobile
238 phase B over 7 mins; 10-35% mobile phase B over 30 mins; 35-99% mobile phase B over 5
239 mins; and then a wash with 99% mobile phase B for 4 min. The capillary column flow rate
240 was 300 nL/min and the column temperature was 40°C. After each run, the column was re-
241 equilibrated for 15 mins prior to injection of the next sample.

242

243 The nanoLC system was interfaced with an Orbitrap Fusion Tribrid mass spectrometer with
244 an EasyNano ionisation source (Thermo Fisher Scientific Inc.). Positive ESI-MS and MS²
245 spectra were acquired using Xcalibur software (version 4.0, Thermo Fisher Scientific Inc.).
246 The instrument source settings were: ion spray voltage, 1,900 V; sweep gas, 0 Arb; ion
247 transfer tube temperature 275°C. MS¹ spectra were acquired on the Orbitrap MS with the
248 following parameters: 120,000 resolution; scan range m/z 375-1,500; AGC target 4e⁵; max
249 fill time 100 ms. Data dependent acquisition was performed in topN mode, selecting the 12
250 most intense precursors with charge states >1. easy-IC was used for internal calibration.
251 Dynamic exclusion was performed for 20 s post precursor selection and a minimum threshold
252 for fragmentation was set at 5e³. MS² spectra were acquired on the Orbitrap with the
253 following parameters: 30,000 resolution; max fill time 100 ms; HCD; activation energy 32
254 NCE. Blanks were analysed after every sample and data was processed using the same
255 methodology as the amelogenin extract samples. All blanks showed no carry over of AMELX
256 or AMELY specific peptides across samples.

257

258 2.3.2. timsTOF HT IMS MS Analysis

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

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

270

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

278

279 The Evosep LC system was interfaced with a timsTOF HT mass spectrometer with a Captive
280 Spray ionisation source (Bruker Daltonics GmbH). Positive ESI-MS/MS spectra were
281 acquired using Compass HyStar 6.2 (Bruker Daltonics GmbH). The instrument source
282 settings were as follows: ion spray voltage 1,500 V; ion transfer tube temperature, 180 $^{\circ}\text{C}$;
283 dry gas flow 3.0 L/min. Data dependent acquisition was performed with scan range: m/z 100-
284 1,700 using parallel accumulation serial fragmentation (PASEF) scan mode. The trapped ion
285 mobility settings were as follows: ramp time of 100 ms; ramp rate of 9.42 Hz; the ion
286 mobility window ($1/k_0$ range), PASEF ramps, and total cycle time varied depending on
287 method used: 24.0 minute and 14.4 minute used the standard 1.1 second cycle time method,
288 while the 4.8 minute and 2.9 minute methods used the short 0.5 second cycle time method
289 (Table S.5). MS₂ data was acquired for precursors above 2500 intensity. HCD collision cell
290 energy was 10 eV. Active exclusion (dynamic exclusion) was set to release after 0.4 min.
291 Blanks were analysed after every sixth sample and data was processed using the same
292 methodology as the amelogenin extract samples. All blanks showed no carry over of AMELX
293 or AMELY specific peptides across samples. MGF files were generated from Bruker .d files
294 in Bruker Compass DataAnalysis ver 6.0 using standard parameters (ProteinAnalysis tool) to
295 collate MS/MS spectra and match mobility values to precursor ions.

296

297 *2.4. MS Data Processing*

298

299 *2.4.1. MS² Data Searching in Mascot*

300 MS² data from the analysis of cattle teeth was searched against two FASTA libraries: one
301 containing the whole bovine proteome and the cRAP contaminants library (119 contaminant
302 entries), and the other containing bovine enamel and dentine sequences from SwissProt and
303 UniProt²⁸ (amelogenin-X (UniProt accessions P02817 and P02817-2), amelogenin-Y
304 (UniProt accessions Q99004/B7TCH9), matrix metalloproteinase-20, enamelin, tuftelin,
305 ameloblastin, amelotin, collagen). Similarly, human data was searched against two FASTA
306 libraries: one containing the whole human proteome and cRAP, and the other containing

307 human enamel and dentine protein sequences from SwissProt²⁸ (amelogenin-X (UniProt
308 accessions Q99217-1, Q99217-2, and Q99217-3), amelogenin-Y (UniProt accessions
309 Q99218-1 and Q99218-2), matrix metalloproteinase-20, amelogenin, tuftelin, ameloblastin,
310 amelotin, collagen) as well as cRAP. Mascot version 2.8, with the parameters as follows:
311 enzyme, none; peptide tolerance was set to ± 10.0 ppm; peptide charges 2+, 3+, and 4+ were
312 searched; variable modifications of deamidation (NQ), oxidation (M), phosphorylation (STY)
313 were chosen as they were the most frequently observed modifications and performed best in a
314 series of tests where different combinations of modifications were employed (Table S.6).
315 Male sex was assigned where two or more high-scoring unique AMELY peptides, with good
316 quality MS/MS spectra, were identified (example spectra shown in Figures 3 and S.2).

317

318 Use of the whole cattle/human proteome gave the same sex estimations as use of the smaller,
319 enamel and dentine proteome searches (Table S.7). Full proteome searches took
320 approximately an hour or longer, while enamel and dentine proteome searches took a few
321 minutes. Moreover, while a larger FASTA library can act as a form of quality assurance, it is
322 also possible that a larger FASTA library could potentially result in more false female
323 assignments, due to peptide scores needing to be higher due to the inflated score acceptance
324 threshold for peptide identification, resulting in AMELY false negatives. Therefore, results
325 from the enamel and dentine proteome searches are discussed herein to showcase data from a
326 more rapid MS² data processing pipeline.

327

328 *2.4.2. Data Searching in FragPipe*

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

344

345 *2.4.3. MS¹ Data Processing in Skyline*

346 MS¹ data was processed in Skyline (ver 23.1.0.380). Skyline search parameters were as
347 follows: structural modifications allowed, oxidation (ox); maximum number of
348 modifications, 3; ion match tolerance, 0.001 ppm. High quality, recurrent human unique
349 AMELX and AMELY peptides, identified from MS² Mascot searches, SIRPPYPSYG (*m/z*
350 568.790), SIRPPYPSYGYEPM(ox)G (*m/z* 865.400) and SM(ox)IRPPY (*m/z* 440.223), were

351 scanned for in all samples in human samples, generating extracted ion chromatograms for
352 each peptide. No conserved unique AMELX and AMELY peptides were identified in the
353 cattle samples and therefore MS¹ data processing in Skyline was not possible for cattle data.

354

355 **3. Results and Discussion**

356

357 *3.1. Comparison of established nanoLC-MS/MS to nanoLC-IMS-MS/MS for* 358 *proteomic sex estimation*

359

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

366

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

380

381 Since comparisons of ion intensities across different MS acquisition methods is not viable,
382 ratios of AMELX/AMELY signal intensities in the male sample were used to compare the
383 Orbitrap and timsTOF methods, where a higher value for the ratio means lower AMELY
384 signal, relative to the AMELX signal. For the Orbitrap method, the AMELX/AMELY ratio
385 was 16.9, while for the timsTOF methods, the ratios were 9.5, 11.7 and 31.7 for the 24.0
386 minute, 14.4 minute and 4.8 minute methods, respectively. There was also an approximately
387 10 times reduction in signal intensity when using the 4.8 minute method, compared to the
388 14.4 and 24.0 minute method.

389

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

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

398

399 *3.2. Comparison of novel neutralisation acid etch sampling to conventional acid etch* 400 *sampling*

401

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

433

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

439

440 The difference in data quality between the conventional and neutralised etching methods in
441 this experiment may partly be explained by the conventional acid etch method being
442 subjected to two clean-up processes (ZipTip and Evotip), unlike the neutralised etches that
443 were only subjected to Evotip clean-up, which could result in increased losses in the
444 conventional acid etch preparation. Moreover, for the neutralised acid etching method, the
445 full supernatant was transferred to the Evotip, while in the conventional method, a fraction of
446 the reconstituted sample was transferred to the Evotip, so the increased peptide counts in the
447 neutralised acid etches may also be due to the transfer of more material to the Evotips.
448 However, based on this data, conventional acid etch protocol (using ZipTips) is not
449 recommended for Evosep-timsTOF analysis, due to decreased peptide counts, and similar or
450 decreased peptide ion intensities when compared to the neutralised method. Direct Evotip
451 clean-up of untreated acid etches without the ZipTip clean-up stage would also be a viable
452 preparative method for Evosep-timsTOF analysis, but there are some additional benefits
453 associated with use of neutralisation acid etching method. Firstly, addition of base to acid that
454 has enamel suspended in it results in the formation of a precipitate from the calcium
455 phosphate component of enamel⁴¹; this precipitate acts as a trap for debris (e.g. sediment
456 from burial) upon centrifugation, removing the need for the surface of the teeth to be
457 completely cleaned before etching, or for filtration of the etch. This trapping of the debris/dirt
458 also prevents Evotips from becoming clogged, which could lead to blockages and
459 overpressure on the LC. Moreover, use of neutralised acid in expensive laboratory equipment
460 (such as centrifuges) is more desirable than concentrated acid, as it decreases the risk of
461 corrosive damage. Employing a neutralisation step also means that samples do not have to be
462 evaporated to dryness at any point, further decreasing the overall time taken for sample
463 preparation and removing a potential point of sample loss from the preparation protocol.
464 Analysis of neutralised etches at different time points over a four week time period also
465 demonstrated the etches were generally stable over short storage periods, with best results
466 when the etches are stored in LoBind tubes, demonstrating the base is not adversely affecting
467 the sexually dimorphic peptides within the etch (S.I 3; Figures S.6 and S.7).

468

469 *3.3. Rapid Evosep-timsTOF MS methods for proteomic sex estimation*

470

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

474

475 *3.3.1. Application of rapid Evosep-timsTOF MS methods to archaeological* 476 *cattle remains*

477

478 The neutralised sampling method was applied to archaeological cattle remains of unknown
479 sex from mediaeval Fishergate, York, dating to approximately 11th to 19th century CE³⁵.
480 These samples were analysed using the 14.4 minute, 4.8 minute and 2.9 minute Evosep-
481 timsTOF methods (Figure 7).

482

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

505

506 From these analyses, individuals 3352 and 4002 were assigned as males, since they had
507 multiple high scoring unique AMELY peptides present, with coverage in the NMRLPPY
508 region of AMELY across all three MS methods. The remaining 4 samples were assigned
509 female due to the lack of AMELY identification across all analyses. Across the various
510 analyses of these bovine samples, the amelogenin peptides identified were varied, and no
511 single AMELX or AMELY peptide was repeatedly observed in all samples; therefore no
512 peptide that could be used as a biomarker peptide for targeted MS¹ data processing was
513 identified.

514

515 *3.3.2. Application of rapid Evosep-timsTOF MS methods to archaeological* 516 *human remains*

517

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

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

529

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

546

547 Conserved human AMELX and AMELY peptides (i.e. peptides that were identified in all of
548 the analyses of all relevant human samples) were identified through MS² data searching
549 (compiled in Tables S.8 and S.9). These peptides were tested for their applicability as protein
550 biomarker peptides in rapid MS¹ data searching through processing in Skyline. Peptides
551 SIRPPYPSYG (*m/z* 568.790) and SIRPPYPSYGYEPM(+15.99)G (*m/z* 865.400) were
552 identified as the most suitable AMELX targets, while SM(+15.99)IRPPY (*m/z* 440.223) was
553 identified as the most suitable AMELY target (Figure 8b). All human data was then
554 processed through Skyline, targeting the aforementioned peptides. The identification of these
555 peptides fully aligned with identification of its related protein in the MS² data searching
556 (Table S.10), demonstrating the success of these peptides as biomarkers for their respective
557 amelogenin proteins in Evosep-timsTOF MS analysis of human teeth. Both MS¹ and MS²
558 data processing are rarely implemented simultaneously for amelogenin proteomic sex
559 estimation, with most studies reporting the use of one or the other^{3,11–13,18,20,21,24,27,42–48}.
560 Employing both data processing techniques is beneficial to help ensure accurate sex
561 estimations, as MS² data processing allows for the identifications of multiple peptides from a
562 single protein, while MS¹ data processing can identify peptides that may not have been
563 selected for fragmentation and are consequently not identified in the MS² data processing
564 stage. MS¹ data processing can also be used as a rapid, non-complex data processing method
565 to further decrease workflow time for proteomic sex estimation. Furthermore, use of
566 ameloblastin and enamelin peptide counts can act as a further validation tool to aid
567 scrutinisation of amelogenin peptide data, allowing a more global understanding of the
568 success of peptide identification in the sample. When using the 2.9 minute method, the
569 highest ameloblastin and enamelin peptide counts in the adult samples were 165 and 137,
570 while the lowest were 83 and 53, respectively. All of the adult samples gave sex estimations

571 that matched with their assumed sex (based on names), with peptide intensities in the range of
572 10^6 - 10^7 (Figure S.12). Therefore human samples with non-amelogenin peptide counts of
573 approximately 50 or more seem to correlate to high AMELX and AMELY peptide counts
574 that yield reliable sex estimations. Despite this threshold, reliable male sex estimations seem
575 to still be possible below this level, as seen in the duplicate analyses of non-adult HGM 28.
576 For HGM 28, only 7-40 unique ameloblastin or enamel unique peptides were identified, yet
577 male sex estimation was still possible based on the number of unique AMELX/AMELY
578 peptides, and the presence of the biomarker peptides in the MS¹ data. It should be noted that
579 female sex estimates are the greatest false positive risk from low quality data, as this
580 assignment requires confidence in the absence of AMELY. As such, when assigned female, it
581 is recommended to more rigidly stick to the stringent threshold.

582

583 **Conclusions**

584 Here, a number of key methodological developments have enabled the untargeted, rapid
585 identification of amelogenin peptides for sex estimation in cattle and human archaeological
586 remains. Based on the result presented, it has been concluded that neutralisation acid etch
587 sample preparation⁴⁹ followed by analysis by the 2.9 minute Evosep-timsTOF MS method is
588 a viable, novel, streamlined, rapid method for high-throughput amelogenin proteomic sex
589 estimation. Longer Evosep-timsTOF MS methods showed higher unique amelogenin peptide
590 counts and peptide ion intensities for both modern and archaeological material, however sex
591 assignments across all compared methods were the same. Further study would be required to
592 understand the impact of the reduced peptide ion intensities and peptide counts on extremely
593 low signal material. This method presents advantages over alternative targeted MS
594 approaches to rapid amelogenin identification, enabling global peptide analysis in similar run
595 times, and easy application to non-hominin remains. Collection of global peptide data enables
596 observation of peptide modifications (such as deamidation) and enables non-amelogenin
597 peptides to be studied, allowing their peptide counts to be employed as an additional
598 validation of data quality. If targeted analysis was preferred, a targeted data processing step
599 using Skyline could be implemented using the applicable AMELX and AMELY biomarker
600 peptides identified here, allowing for a rapid workflow for sex estimation, while still
601 retaining the global protein information obtained from the untargeted MS analysis.
602 Quantitative analysis can also be achieved through implementation of software such as the
603 open access softwares, FragPipe and MSFragger. Where possible, it is recommended that
604 both MS¹ and MS² data processing are implemented, so that sex estimation is not limited by
605 either the presence or absence of a few select peptides, nor the selection of potentially low
606 intensity ions for MS/MS fragmentation. Moreover, implementation of both data processing
607 methods gives two separate sex estimations that can validate one another. The neutralised
608 acid etches were also demonstrated to be generally stable for a 4-week storage period in a
609 LoBind tube, deeming the sampling method established here suitable for short term storage
610 before analysis.

611

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

615 when compared to human extracts, the prospect of rapid sex estimation of cattle remains
616 using Evosep-timsTOF mass spectrometry is still promising. Further investigations of how to
617 improve confidence in cattle sex estimations, by increasing signal intensities and unique
618 amelogenin peptide counts are required before rapid proteomic cattle sex estimation methods
619 can be recommended. Buckley et al²⁷ also communicated concerns surrounding the reliability
620 of proteomic-based cattle sex estimation.

621

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

627

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

635

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

640

641 **CRedit authorship contribution statement:**

642 **C. Blacka:** Investigation, Methodology, Formal analysis, Visualisation, Writing - original
643 draft, Writing - review & editing. **A. Dowle:** Formal analysis, Writing - review & editing. **M.**
644 **Lisowski:** Resources, Writing - review & editing. **M. Alexander:** Resources, Writing -
645 review & editing. **K. Penkman:** Conceptualisation, Funding acquisition, Visualisation,
646 Supervision, Methodology, Writing - review & editing. **J. Hendy:** Conceptualisation,
647 Funding acquisition, Visualisation, Supervision, Methodology, Writing - review & editing. **J.**
648 **Mosely:** Visualisation, Supervision, Methodology, Writing - review & editing.

649

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

652

653 **References:**

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

- 658 2. Fages A, Seguin-Orlando A, Germonpré M, Orlando L. Horse males became over-
659 represented in archaeological assemblages during the Bronze Age. *J Archaeol Sci Rep.*
660 2020;31:102364. doi:10.1016/j.jasrep.2020.102364
- 661 3. Berezina N, Ziganshin R, Kolobova K, et al. Bison sex matters: the potential of
662 proteomic tooth enamel analysis for determination of ancient human subsistence
663 strategies. *Archaeol Anthropol Sci.* 2024;16(9):142. doi:10.1007/s12520-024-02053-x
- 664 4. Madupe PP, Koenig C, Patramanis I, et al. Enamel proteins reveal biological sex and
665 genetic variability within southern African Paranthropus. *bioRxiv.* Published online July
666 3, 2023:2023.07.03.547326. doi:10.1101/2023.07.03.547326
- 667 5. Demarchi B, Hall S, Roncal-Herrero T, et al. Protein sequences bound to mineral
668 surfaces persist into deep time. *Elife.* 2016;5. doi:10.7554/eLife.17092
- 669 6. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. *Nat Rev Genet.*
670 2001;2(5):353-359. doi:10.1038/35072071
- 671 7. Fincham AG, Moradian-Oldak J, Simmer JP. The structural biology of the developing
672 dental enamel matrix. *J Struct Biol.* 1999;126(3):270-299. doi:10.1006/jsbi.1999.4130
- 673 8. Lau EC, Mohandas TK, Shapiro LJ, Slavkin HC, Snead ML. Human and mouse
674 amelogenin gene loci are on the sex chromosomes. *Genomics.* 1989;4(2):162-168.
675 doi:10.1016/0888-7543(89)90295-4
- 676 9. Nakahori Y, Takenaka O, Nakagome Y. A human X-Y homologous region encodes
677 "amelogenin." *Genomics.* 1991;9(2):264-269. doi:10.1016/0888-7543(91)90251-9
- 678 10. Stewart NA, Molina GF, Issa JPM, et al. The identification of peptides by nanoLC-
679 MS/MS from human surface tooth enamel following a simple acid etch extraction. *RSC*
680 *Adv.* 2016;6(66):61673-61679. doi:10.1039/C6RA05120K
- 681 11. Stewart NA, Gerlach RF, Gowland RL, Gron KJ, Montgomery J. Sex determination of
682 human remains from peptides in tooth enamel. *Proc Natl Acad Sci U S A.*
683 2017;114(52):13649-13654. doi:10.1073/pnas.1714926115
- 684 12. Lugli F, Di Rocco G, Vazzana A, et al. Enamel peptides reveal the sex of the Late
685 Antique "Lovers of Modena." *Sci Rep.* 2019;9(1):13130. doi:10.1038/s41598-019-
686 49562-7
- 687 13. Parker GJ, Yip JM, Eerkens JW, et al. Sex estimation using sexually dimorphic
688 amelogenin protein fragments in human enamel. *J Archaeol Sci.* 2019;101:169-180.
689 doi:10.1016/j.jas.2018.08.011
- 690 14. Wasinger VC, Curnoe D, Bustamante S, et al. Analysis of the Preserved Amino Acid
691 Bias in Peptide Profiles of Iron Age Teeth from a Tropical Environment Enable Sexing
692 of Individuals Using Amelogenin MRM. *Proteomics.* 2019;19(5):e1800341.
693 doi:10.1002/pmic.201800341
- 694 15. Lugli F, Figus C, Silvestrini S, et al. Sex-related morbidity and mortality in non-adult
695 individuals from the Early Medieval site of Valdaro (Italy): the contribution of dental
696 enamel peptide analysis. *J Archaeol Sci Rep.* 2020;34:102625.

697 doi:10.1016/j.jasrep.2020.102625

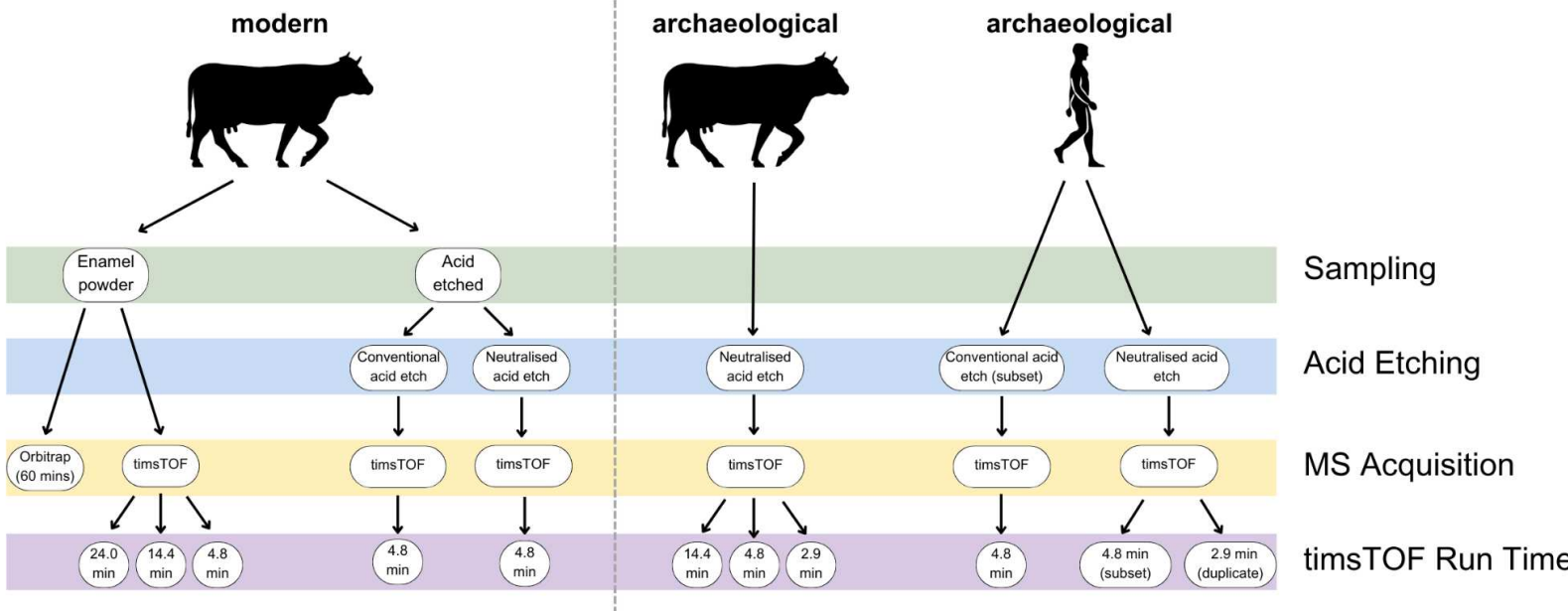
- 698 16. Rebay-Salisbury K, Janker L, Pany-Kucera D, et al. Child murder in the Early Bronze
699 Age: proteomic sex identification of a cold case from Schleinbach, Austria. *Archaeol*
700 *Anthropol Sci.* 2020;12(11):265. doi:10.1007/s12520-020-01199-8
- 701 17. Cintas-Peña M, Luciañez-Triviño M, Montero Artús R, et al. Amelogenin peptide
702 analyses reveal female leadership in Copper Age Iberia (c. 2900-2650 BC). *Sci Rep.*
703 2023;13(1):9594. doi:10.1038/s41598-023-36368-x
- 704 18. Granja R, Araújo AC, Lugli F, Silvestrini S, Silva AM, Gonçalves D. Unbalanced sex-
705 ratio in the Neolithic individuals from the Escoural Cave (Montemor-o-Novo, Portugal)
706 revealed by peptide analysis. *Sci Rep.* 2023;13(1):19902. doi:10.1038/s41598-023-
707 47037-4
- 708 19. Kotli P, Morgenstern D, Nagar Y, et al. Antiquity validation by amelogenin peptidomics
709 profile: Sex determination in an ascetic burial. *Research Square.* Published online
710 August 24, 2023. doi:10.21203/rs.3.rs-3279131/v1
- 711 20. Olszewski J, Hall RA, Kootker LM, et al. Osteological, multi-isotope and proteomic
712 analysis of poorly-preserved human remains from a Dutch East India Company burial
713 ground in South Africa. *Sci Rep.* 2023;13(1):14666. doi:10.1038/s41598-023-41503-9
- 714 21. Gamble JA, Spicer V, Hunter M, et al. Advancing sex estimation from amelogenin:
715 Applications to archaeological, deciduous, and fragmentary dental enamel. *J Archaeol*
716 *Sci Rep.* 2024;54:104430. doi:10.1016/j.jasrep.2024.104430
- 717 22. Cleland TP, McGuire SA, Beatrice JS, Moran KS, France CAM. SPEED-E: A modified
718 version of the sample preparation by Easy extraction and Digestion(-free) protocol for
719 enamel-based sex estimation in archaeological remains. *J Archaeol Sci.*
720 2024;168:106006. doi:10.1016/j.jas.2024.106006
- 721 23. Casas-Ferreira AM, del Nogal-Sánchez M, Arroyo ÁE, Vázquez JV, Pérez-Pavón JL.
722 Fast methods based on mass spectrometry for peptide identification. Application to sex
723 determination of human remains in tooth enamel. *Microchem J.* 2022;181:107645.
724 doi:10.1016/j.microc.2022.107645
- 725 24. Koenig C, Bortel P, Paterson RS, et al. Automated high-throughput biological sex
726 identification from archaeological human dental enamel using targeted proteomics.
727 *bioRxiv.* Published online February 22, 2024:2024.02.20.581140.
728 doi:10.1101/2024.02.20.581140
- 729 25. Cappellini E, Welker F, Pandolfi L, et al. Early Pleistocene enamel proteome from
730 Dmanisi resolves Stephanorhinus phylogeny. *Nature.* 2019;574(7776):103-107.
731 doi:10.1038/s41586-019-1555-y
- 732 26. Kotli P, Morgenstern D, Horwitz L, Khalaily H, Bocquentin F, Boaretto E. A
733 peptidomics method for assessing sex from modern and ancient bovine tooth enamel.
734 *Research Square.* Published online January 29, 2024. doi:10.21203/rs.3.rs-3766361/v1
- 735 27. Buckley M, Pigière F, Pal Chowdhury M, Kitchener AC, Smyth J. Proteomic sexing of
736 archaeological cattle remains at Neolithic Kilshane. *J Archaeol Sci.*

- 737 2024;172(106102):106102. doi:10.1016/j.jas.2024.106102
- 738 28. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids*
739 *Res.* 2017;45(D1):D158-D169. doi:10.1093/nar/gkw1099
- 740 29. van Doorn NL, Wilson J, Hollund H, Soressi M, Collins MJ. Site-specific deamidation
741 of glutamine: a new marker of bone collagen deterioration. *Rapid Commun Mass*
742 *Spectrom.* 2012;26(19):2319-2327. doi:10.1002/rcm.6351
- 743 30. Aballo TJ, Roberts DS, Melby JA, Buck KM, Brown KA, Ge Y. Ultrafast and
744 Reproducible Proteomics from Small Amounts of Heart Tissue Enabled by Azo and
745 timsTOF Pro. *J Proteome Res.* 2021;20(8):4203-4211.
746 doi:10.1021/acs.jproteome.1c00446
- 747 31. Akkurt Arslan M, Kolman I, Pionneau C, et al. Proteomic Analysis of Tears and
748 Conjunctival Cells Collected with Schirmer Strips Using timsTOF Pro: Preanalytical
749 Considerations. *Metabolites.* 2021;12(1). doi:10.3390/metabo12010002
- 750 32. Meier F, Park MA, Mann M. Trapped Ion Mobility Spectrometry and Parallel
751 Accumulation–Serial Fragmentation in Proteomics. *Mol Cell Proteomics.* 2021;20.
752 doi:10.1016/j.mcpro.2021.100138
- 753 33. Ridgeway ME, Lubeck M, Jordens J, Mann M, Park MA. Trapped ion mobility
754 spectrometry: A short review. *Int J Mass Spectrom.* 2018;425:22-35.
755 doi:10.1016/j.ijms.2018.01.006
- 756 34. Meier F, Beck S, Grassl N, et al. Parallel Accumulation–Serial Fragmentation (PASEF):
757 Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion
758 Mobility Device. *J Proteome Res.* 2015;14(12):5378-5387.
759 doi:10.1021/acs.jproteome.5b00932
- 760 35. O'Connor TP. Bones from 46-54 Fishergate. In: Addyman PV, Black VE, eds. *The*
761 *Animal Bones.* Vol 15. The Archaeology of York. Council for British Archaeology for
762 York Archaeological Trust; 1991:91.
- 763 36. Chidimuro B, Holst M, Newman S, Keefe K, Collins MJ, Alexander M. Grim up North?
764 Exploring the diet of urban populations in post-medieval Greater Manchester, England,
765 using stable isotope analysis. *Archaeol Anthropol Sci.* 2023;15(11):174.
766 doi:10.1007/s12520-023-01881-7
- 767 37. Newman S, Holst M. *Osteological Analysis, Chapel Street, Hazel Grove, Greater*
768 *Manchester.* York Osteoarchaeology; 2016.
- 769 38. Keefe K, Holst M. *Osteological Analysis, Cross Street Unitarian Chapel. Manchester.*
770 York Osteoarchaeology; 2017.
- 771 39. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI.
772 MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-
773 based proteomics. *Nat Methods.* 2017;14(5):513-520. doi:10.1038/nmeth.4256
- 774 40. Yu F, Haynes SE, Teo GC, Avtonomov DM, Polasky DA, Nesvizhskii AI. Fast
775 quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Mol Cell*

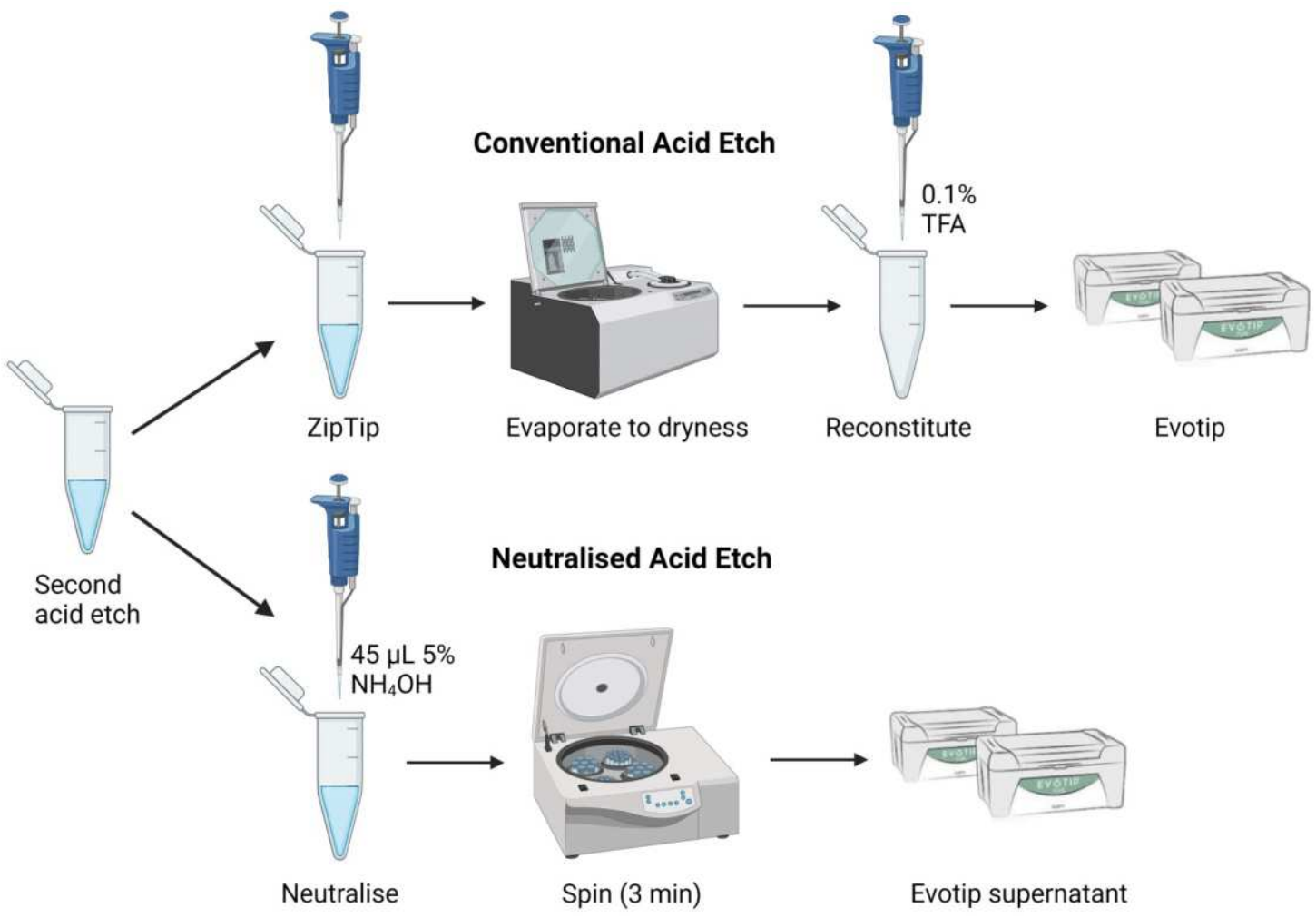
- 776 *Proteomics*. 2020;19(9):1575-1585. doi:10.1074/mcp.TIR120.002048
- 777 41. Dickinson MR, Lister AM, Penkman KEH. A new method for enamel amino acid
778 racemization dating: A closed system approach. *Quat Geochronol*. 2019;50:29-46.
779 doi:10.1016/j.quageo.2018.11.005
- 780 42. Gowland R, Stewart NA, Crowder KD, et al. Sex estimation of teeth at different
781 developmental stages using dimorphic enamel peptide analysis. *Am J Phys Anthropol*.
782 2021;174(4):859-869. doi:10.1002/ajpa.24231
- 783 43. Marta DZB, Gonzalo AJ, Margarita SR, et al. Female sex bias in Iberian megalithic
784 societies through bioarchaeology, aDNA and proteomics. *Sci Rep*. 2024;14(1):21818.
785 doi:10.1038/s41598-024-72148-x
- 786 44. Kimsis J, Petersone-Gordina E, Poksane A, et al. Application of natural sciences
787 methodology in archaeological study of Iron Age burials in Latvia: pilot study. *Forensic*
788 *Sci Med Pathol*. Published online November 8, 2022. doi:10.1007/s12024-022-00553-7
- 789 45. Gasparini A, Lugli F, Silvestrini S, et al. Biological sex VS. Archaeological Gender:
790 Enamel peptide analysis of the horsemen of the Early Middle age necropolises of
791 Campochiaro (Molise, Italy). *J Archaeol Sci Rep*. 2022;41:103337.
792 doi:10.1016/j.jasrep.2021.103337
- 793 46. Greco E, Gennaro AM, Piombino-Mascalì D, et al. Dental Proteomic Analyses and
794 Raman Spectroscopy for the Determination of the Biological Sex and Age of Human
795 Remains from the Greek Cemetery of San Giorgio Extra, Reggio Calabria (Italy).
796 *Reggio Calabria*. Published online July 26, 2023. doi:10.2139/ssrn.4509097
- 797 47. Mays S, Parker G, Johns C, et al. Sex identification of a Late Iron Age sword and mirror
798 cist burial from Hillside Farm, Bryher, Isles of Scilly, England. *Journal of*
799 *Archaeological Science: Reports*. Published online July 27, 2023:104099.
800 doi:10.1016/j.jasrep.2023.104099
- 801 48. Froment C, Hourset M, Sáenz-Oyhéréguy N, et al. Analysis of 5000 year-old human
802 teeth using optimized large-scale and targeted proteomics approaches for detection of
803 sex-specific peptides. *J Proteomics*. 2020;211:103548. doi:10.1016/j.jprot.2019.103548
- 804 49. Blacka C. Neutralisation acid etching of enamel for proteomic amelogenin sex
805 estimation. *Protocols.io*. doi:10.17504/protocols.io.14egn6qr6l5d/v1
- 806

Acid etch development

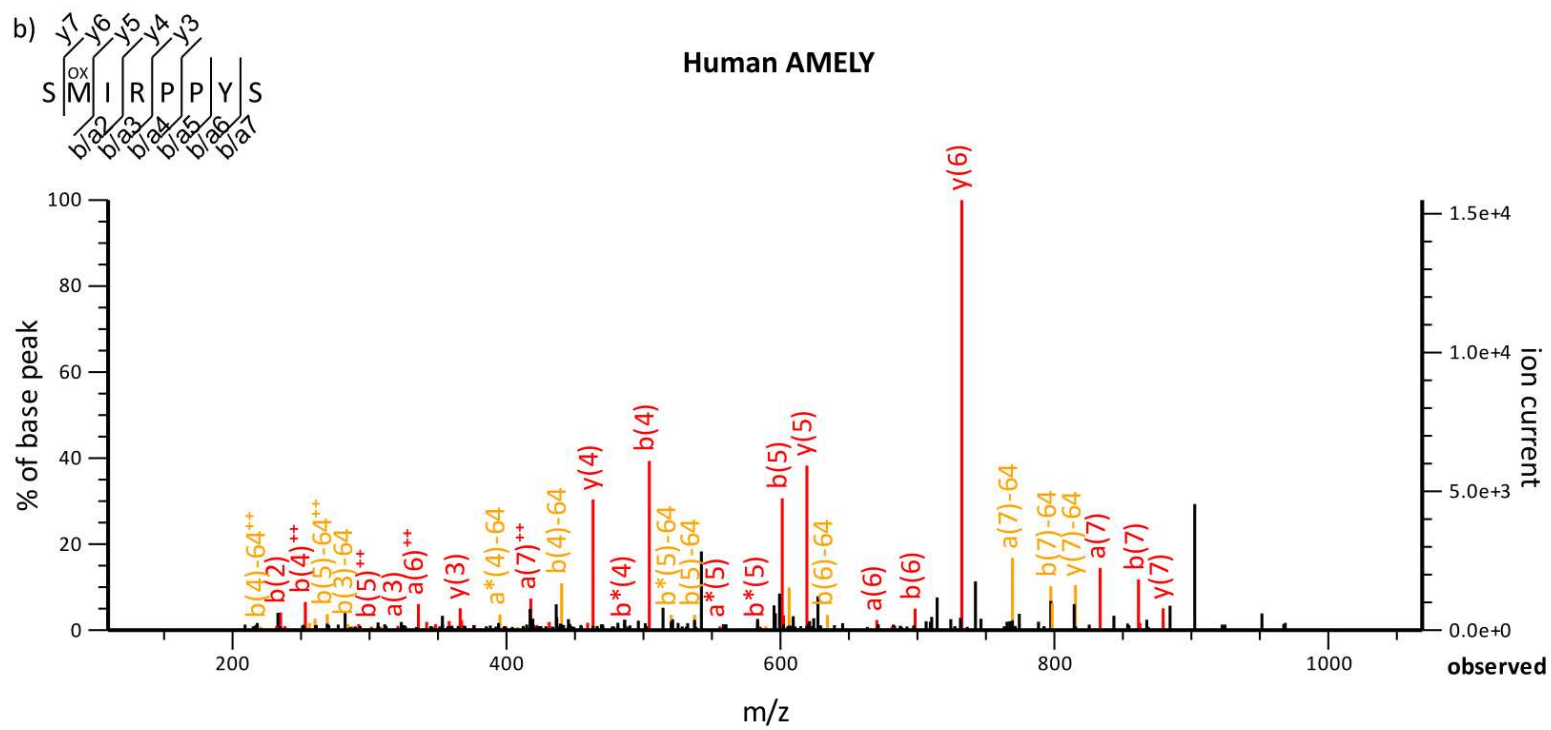
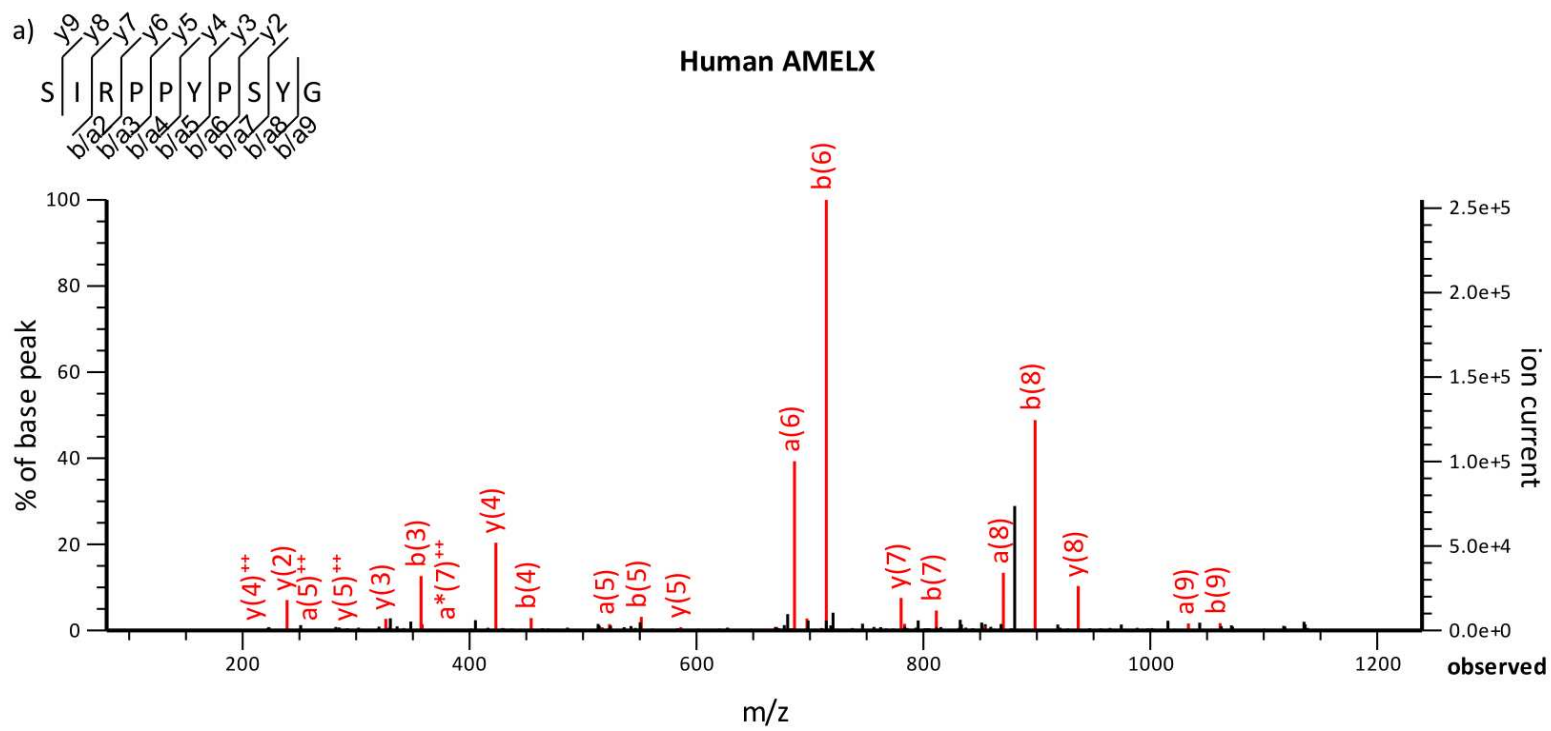
Towards mass spectrometry valuation



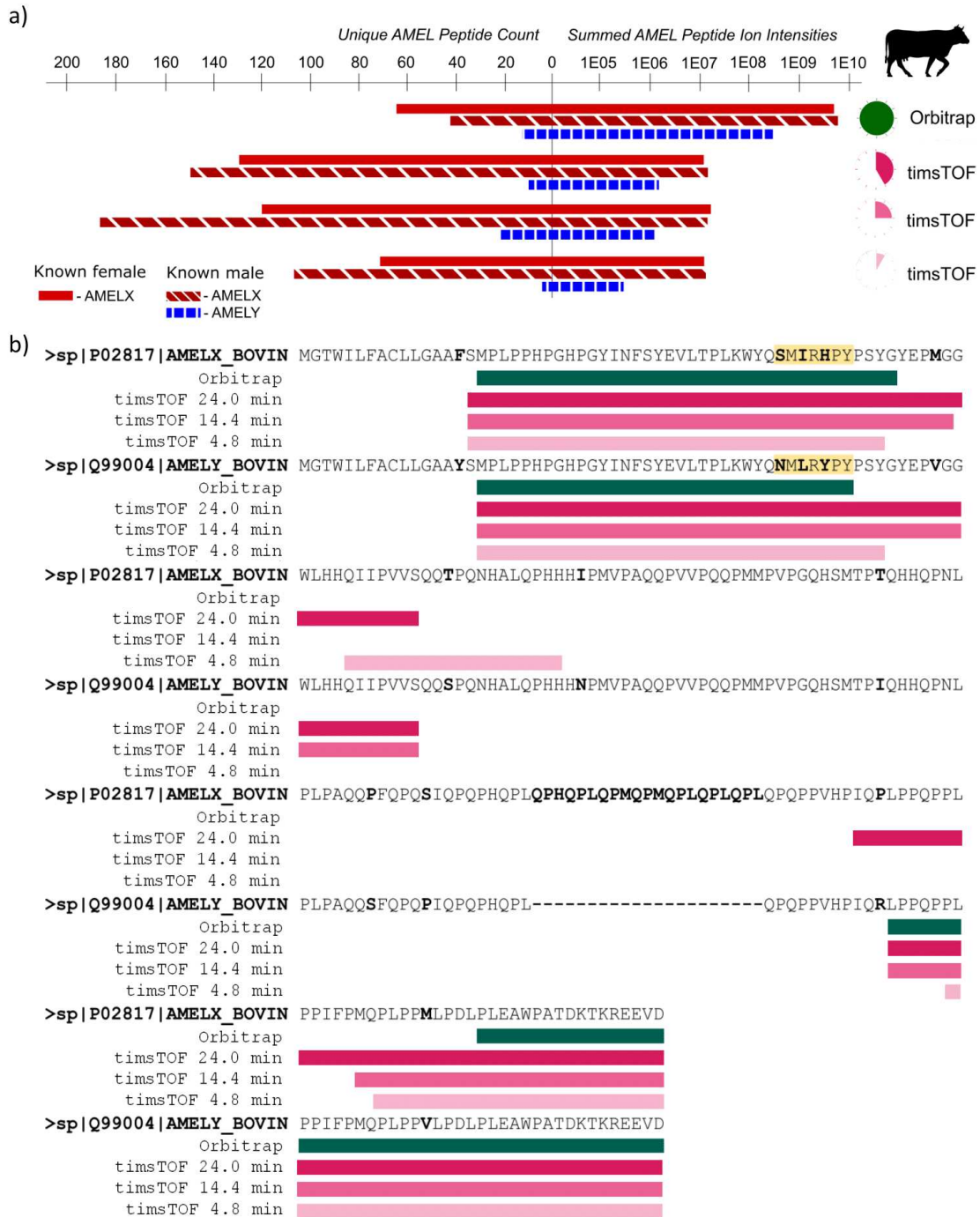
808 **Figure 1.** Summary of methodological and analytical variables applied to the three different
 809 sets of sample material discussed here. Each coloured background represents differences in
 810 variables: in green, sampling; blue, acid etching approach; yellow, MS instrument; purple,
 811 timsTOF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



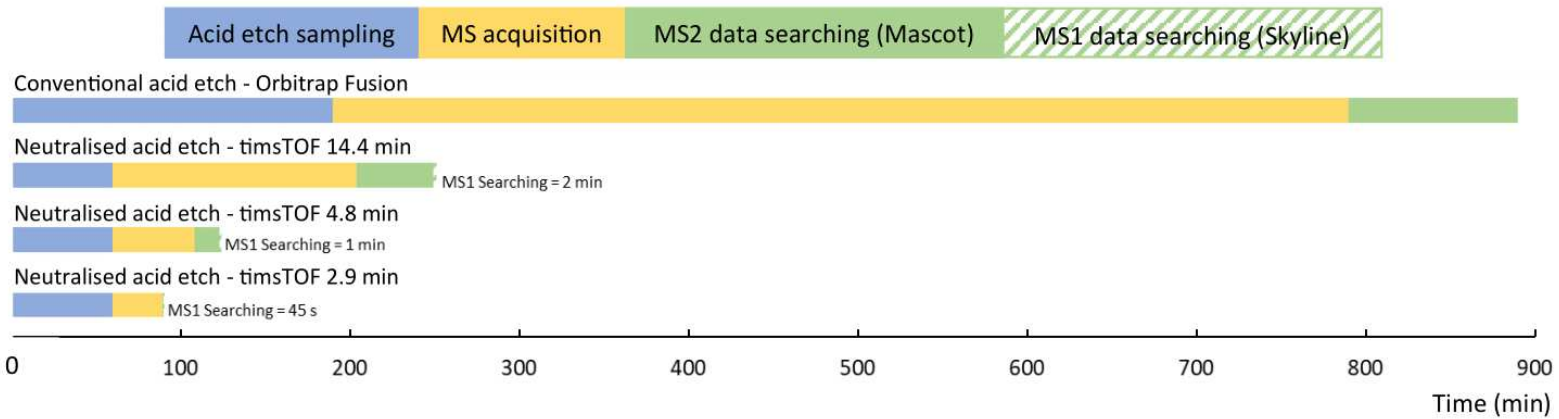
813 Figure 2. a) Summary of the conventional acid etch¹⁰ and the novel neutralised acid etch
814 method when Evosep LC separation is implemented (figure created with BioRender.com).
815



816 Figure 3. Example MS/MS spectra of a) SIRPPYPSYG (human AMELX) and b)
 817 SM(ox)IRPPYS (human AMELY). Spectra were annotated in Mascot.
 818

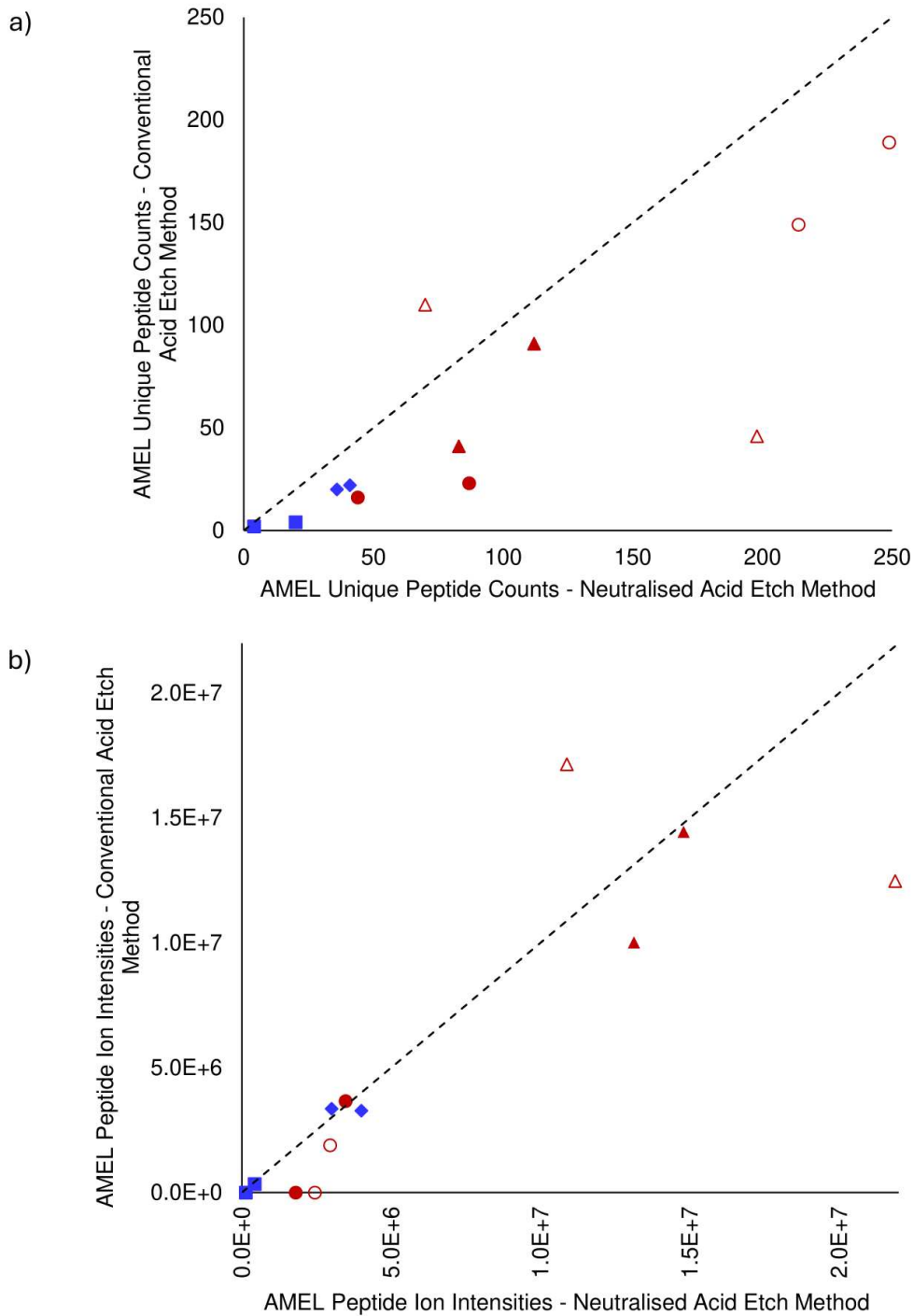


819 Figure 4. a) Unique AMELX (red) and AMELY (blue) peptide counts (left) and summed
820 AMELX (red) and AMELY (blue) peptide ion intensities (right) from powdered enamel from
821 a modern female (filled bars) and male (hashed bars) *Bos taurus* tooth when analysed by
822 Orbitrap Fusion MS, and three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

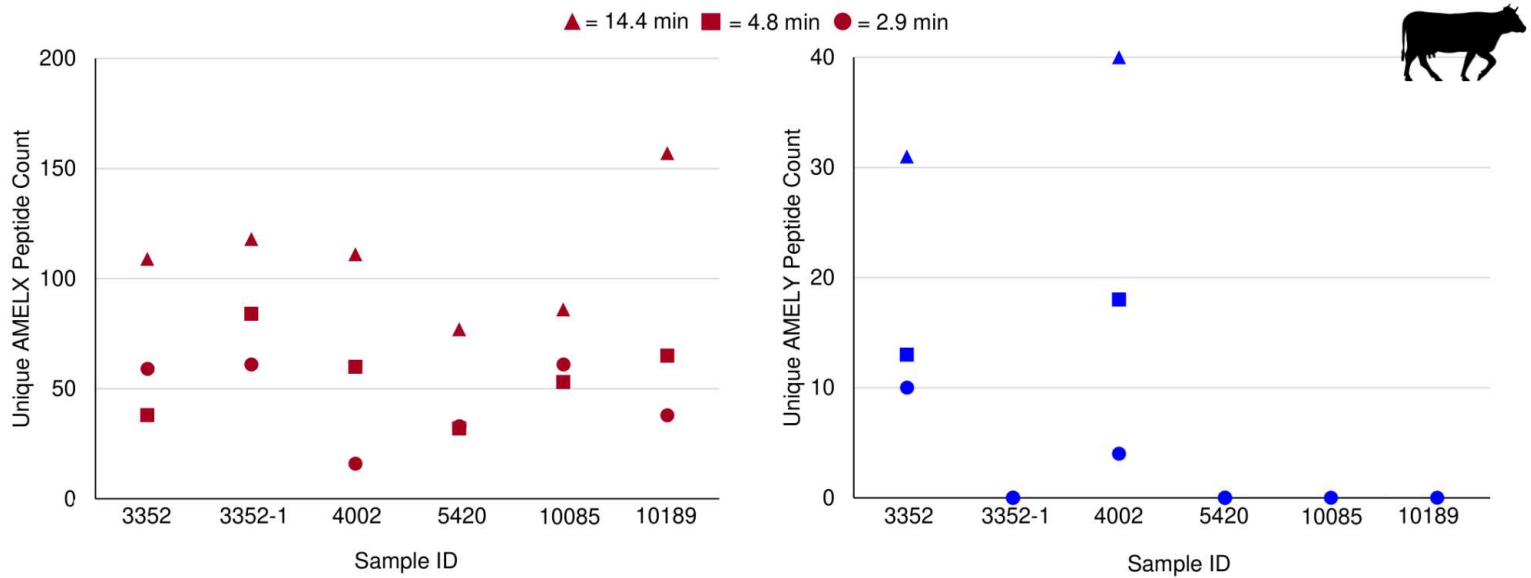


829 Figure 5. Comparisons of the time taken to perform peptidic amelogenin sex estimation for
830 10 samples with the main methods discussed in this paper. From top to bottom: conventional
831 acid etching followed by Orbitrap Fusion MS; neutralised acid etching followed by 14.4 min
832 Evosep-timsTOF MS; neutralised acid etching followed by 4.8 min Evosep-timsTOF MS;
833 neutralised acid etching followed by 2.9 min Evosep-timsTOF MS. Note: MS¹ data searching
834 time is omitted from the Orbitrap Fusion time scale as MS¹ data searching was only possible
835 with the human samples, and no human teeth were analysed by Orbitrap Fusion MS.
836

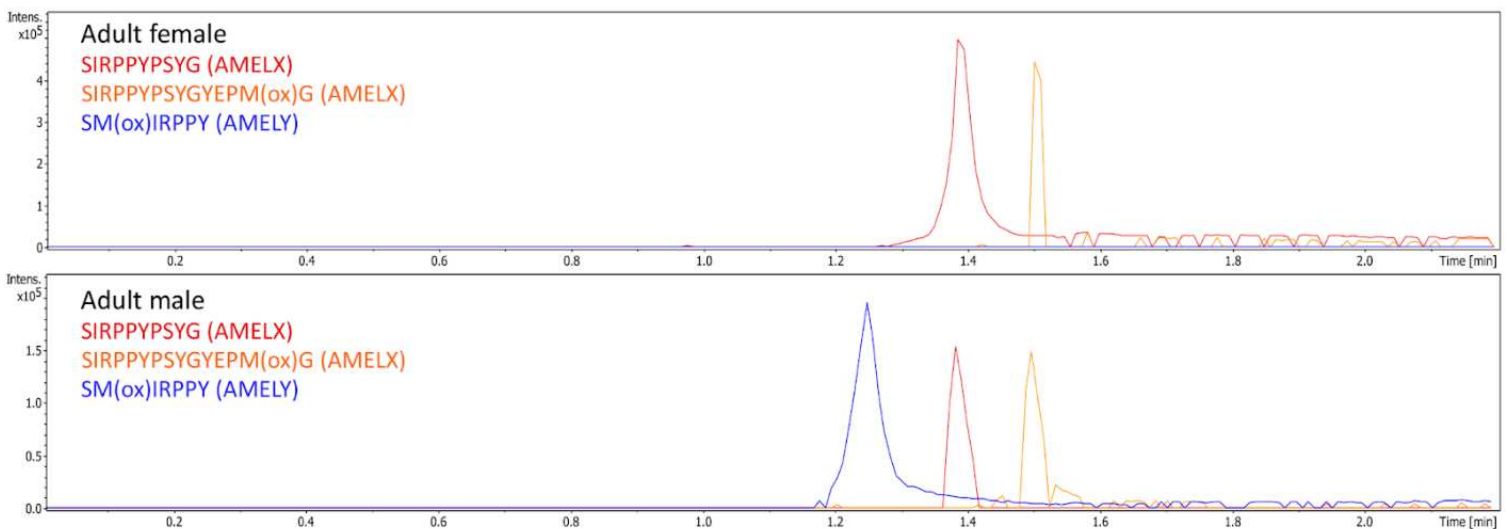
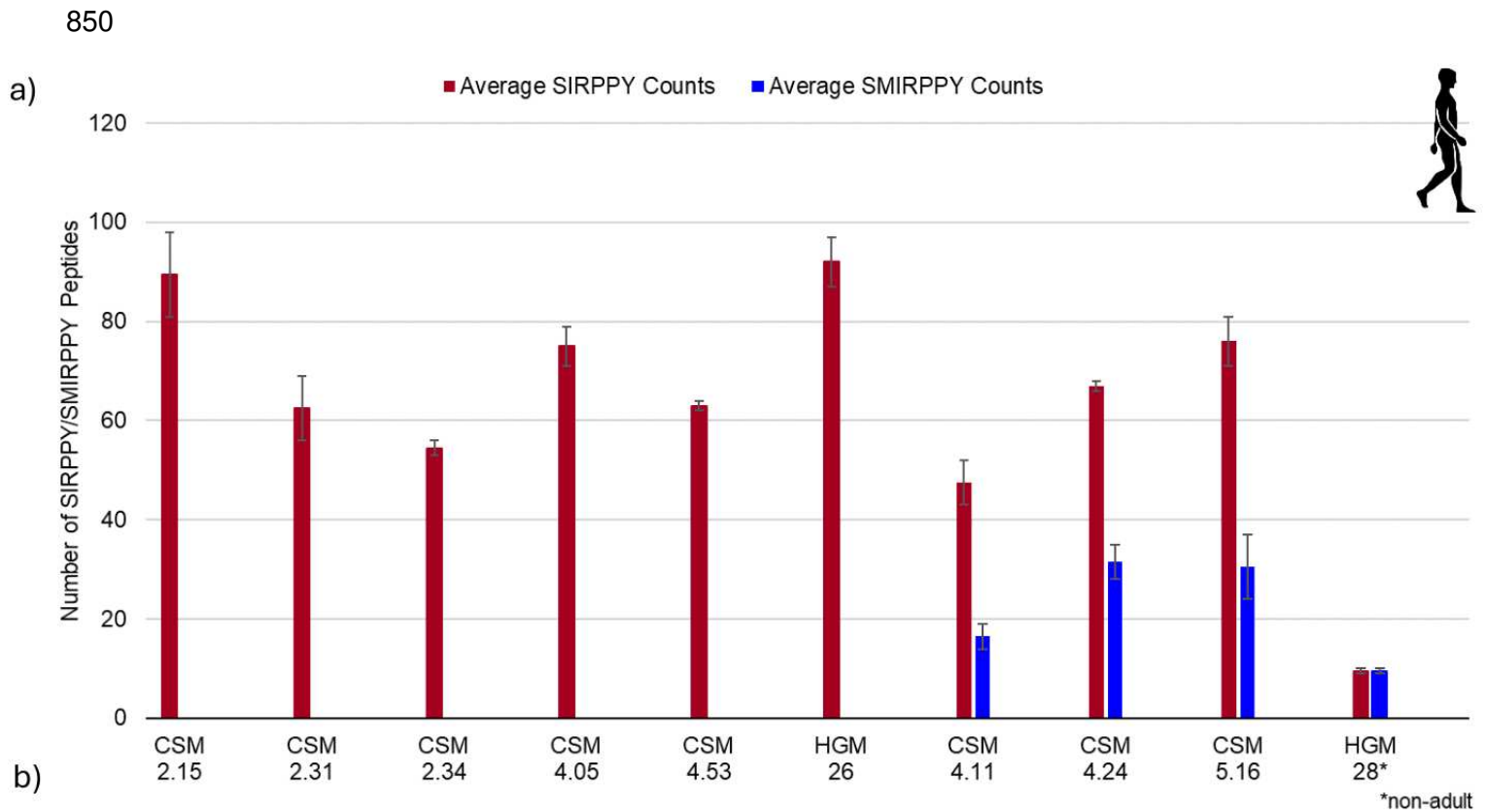
● = bovine AMELX ■ = bovine AMELY ▲ = human AMELX ◆ = human AMELY Fill = male No fill = female



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



846 Figure 7. AMELX (left; red) and AMELY (right; blue) unique peptide counts values from
 847 analysis of the neutralised acid etched mediaeval Fishergate cow mandible teeth using the
 848 14.4 min (triangles), 4.8 min (squares), and 2.9 min (circles) Evosep-timsTOF methods. Note
 849 the difference in scale in the y-axes.



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