Microchemical Journal 221 (2026) 116915

Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc

Fast ultrasonic-based extraction of peptides from teeth for forensics and archaeological purposes

Catarina A.P. André [a,1], Raquel Fonseca [a,1], André Q. Figueiredo [a,b], Carlos Lodeiro [a,b], Stefano Benazzi [c], Maria Giovanna Belcastro [d], Federico Lugli [e], Giulia Di Rocco [f], Hugo M. Santos [a,b], José L. Capelo [a,b,*]

[a] OMICS and Analytical Development Group, BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829–516 Caparica, Portugal
[b] PROTEOMASS Scientific Society, 2825–466 Caparica, Portugal
[c] Department of Cultural Heritage, University of Bologna, Via degli Ariani 1, 48121 Ravenna, Italy
[d] Department of Biological, Geological and Environmental Sciences, University of Bologna, Via Selmi, 3, 40126 Bologna, Italy
[e] Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Via Campi 103, 41125, Modena, Italy
[f] Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 103, 41125, Modena, Italy

ARTICLE INFO

Keywords:
Peptidomics
Dental
Amelogenin
Sex determination
Archaeological
Forensics

ABSTRACT

This study presents a rapid, ultrasound-assisted solid-liquid extraction protocol for recovering peptides from the tooth surface, while preserving the macroscopically visible integrity of the dental morphology. The methodology described herein enables the concurrent use of a single peptide extraction for multiple analytical objectives, encompassing: (i) biological sex determination via detection of sex-dimorphic amelogenin isoforms; (ii) dietary reconstruction based on the identification of food-derived protein residues; (iii) oral microbiome profiling to infer historical health status and disease burden; and (iv) an assessment of age-at-death (AAD) estimation through quantification of asparagine and glutamine deamidation kinetics. Peptide extraction is achieved through an ultrasound-assisted solid–liquid interface, facilitating the release of biomolecules embedded within the dental tissues' matrix without producing macroscopically detectable damage to the sample. The resulting peptide fraction undergoes purification using hydrophilic-lipophilic balanced (HLB) cartridges and is subsequently analysed via high-resolution mass spectrometry. The protocol offers notable advantages, including speed, reproducibility, high throughput, and alignment with the principles of analytical minimalism. Collectively, this strategy represents a substantial advancement in non-invasive dental peptidomics, providing a robust and minimally destructive analytical framework for forensic and bioarchaeological investigations.

1. Introduction

Human skeletal remains are among the most valuable sources of information in palaeoanthropology, as well as in archaeological and forensic investigations. In particular, the study of the tooth peptidome has emerged as a powerful tool for uncovering critical details about individuals, including biological sex, diet, the composition of the oral microbiome, and exposure to bacteria and age-at-death (AAD) [1,2].

Teeth, composed primarily of mineralised tissues such as enamel and dentin, but also cementum and dental plaque/calculus, provide a highly durable environment for the preservation of proteins and peptides over millennia, offering an alternative source of information to DNA, which is often absent or degraded [1,3,4].

Biological sex is one of the most significant aspects of information obtainable through paleoproteomics and peptidomics. This characteristic is associated with the most predominant enamel protein, amelogenin, a molecular biomarker encoded by the dimorphic gene AMEL located on the sex chromosomes. Males possess both AMELX and AMELY, while females have only AMELX. Researchers have demonstrated that sequence differences between AMELX and AMELY can be

* Corresponding author at: OMICS and Analytical Development Group, BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal.
E-mail address: jlcm@fct.unl.pt (J.L. Capelo).
[1] Equal contribution.

https://doi.org/10.1016/j.microc.2026.116915
Received 9 November 2025; Received in revised form 23 December 2025; Accepted 8 January 2026
Available online 9 January 2026
0026-265X/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

analysed using mass spectrometry, allowing for the determination of biological sex from dental samples [1,5–7].

In addition to biological sex, dietary habits, and the oral microbiome can also be explored through peptidomic analysis. Dietary proteins and peptides integrate into dental structures, such as calculus (commonly referred to as tartar). Through mass spectrometry and bioinformatics, different dietary peptides can be identified, providing insights into ancient diets [8,9]. Studies have identified plant, ruminant, oat, and milk peptides in ancient human teeth, revealing important details about the dietary practices of past populations. The dental microbiome has also been investigated to understand health conditions in ancient populations. Microbial peptides originating from bacteria and viruses can become embedded in dental calculus and pulp, allowing the identification of diseases. For instance, in ancient dental samples, proteins and peptides from bacteria associated with periodontitis, such as *Olsenella* and *Fretibacterium*, have been detected, along with commensal species like *Lautropia mirabilis*, *Neisseria* spp., *Streptococcus* spp., and *Cardiobacterium spp* [8,10–12].

AAD estimation represents a critical component in the identification of human remains. Previous studies have proposed that AAD can be inferred through the analysis of deamidation, a spontaneous, non-enzymatic post-translational modification (PTM) that converts asparagine (Asn, N) and glutamine (Gln, Q) residues into aspartic acid (Asp, D) and glutamic acid (Glu, E), respectively. Given that the deamidation rate of asparagine is generally higher than that of glutamine, it has been hypothesised that asparagine deamidation serve as a more reliable marker for AAD estimation. This modification results in a mass increase of +0.984 Da, making it detectable by mass spectrometry and offering a promising analytical approach for the estimation of AAD [13,14].

Studies using mass spectrometry for dental peptidome analysis often rely on protocols that are destructive, time-consuming, involve multiple steps, and have limited reproducibility [8,9,14,15]. To address these limitations, we developed a non-destructive, reproducible, and rapid ultrasonic-based protocol to extract the peptidome embedded in the dental surface. Our protocol preserves tooth structure and addresses (i) biological sex, (ii) dietary habits, (iii) oral microbiome and bacterial exposure, and (iv) current limits of AAD based on peptide deamidation, offering comprehensive insights into the lifestyles and health conditions of ancient populations, as well as aiding in individual identification.

2. Materials and methods

2.1. Experimental study design

The samples were part of the Documented Human Osteological Collection (DHOC) of the Certosa Cemetery of Bologna (Emilia Romagna, northern Italy), housed at the University of Bologna. This collection consisted of 425 well-preserved skeletons of individuals (from foetuses/newborn to 91 years old) who died in Bologna between 1898 and 1944. The sex is known and in 93% of the cases also the cause of death. This collection represents an important scientific resource for the study of human skeletal remains and to test and validate methods and techniques in the field of bioarchaeology, forensics, and palaeoanthropology [16–18]. In this study, a subset of 20 samples (labelled 1 to 20) was selected for analysis. The biological sex, AAD (in years), and documented cause of death for each individual are summarized in Table 1.

2.2. Teeth peptidome extraction

The peptidome extraction protocol used in this study was adapted from previously published methods [5,19], with modifications to improve compatibility with downstream mass spectrometry analysis. Specifically, Trifluoroacetic acid (TFA), a volatile MS-compatible acid, was used in place of the non-volatile acids employed in previous works [5,19]. Individual tooth specimens were fully immersed in 3 mL of 15% (*v*/v) TFA and subjected to ultrasonication using a Q700 Sonicator® Microplate Horn System (QSonica). Sonication was carried out in a 4-min cycle (30 s on, 15 s off) at 25% ultrasonic amplitude and a frequency of 20 kHz. The resulting supernatants, enriched in endogenous peptides, were aliquoted and stored at −80 °C until further analytical processing. Blank extractions (no-sample controls) were processed in parallel with the archaeological specimens to monitor protein contamination. Each blank underwent the complete extraction and purification workflow, including solid-phase extraction and LC-MS/MS analysis.

Table 1
Description of archaeological dental samples. All individuals are part of the Documented Human Osteological Collection (DHOC) curated at the Certosa Cemetery in Bologna (Emilia-Romagna, northern Italy). Biological sex, age at death, and documented cause of death are reported. AAD – age-at-death.

Sample	Biological Sex	AAD (years)	Cause of Death
1	Male	5 and 8 months	Acute enteritis
2	Female	5	Acute meningitis
3	Female	42	Tuberculosis
4	Male	60	Cerebral haemorrhage
5	Female	64	Cerebral haemorrhage
6	Female	51	Unknown
7	Female	32	Mitral stenosis
8	Female	43	Mitral insufficiency, cardiac stenosis
9	Female	31	Grade II and III burns (septicaemia)
10	Female	25	Mitral defect, circulatory insufficiency
11	Male	39	Accidental running over by an electric train
12	Male	23	Intestinal perforation
13	Male	20	Metapneumococcal empyema
14	Female	66	Gastric neoplasm
15	Male	66	Pott's disease (extrapulmonary tuberculosis)
16	Male	45	Cerebral haemorrhage
17	Male	24	Suicide by gunshot
18	Female	44	Acute anaemia (stab wound to the left elbow region)
19	Male	61	Accidental death due to drowning
20	Male	44	Epileptiform accesses

2.3. Peptidome purification with HLB columns

For the purification process, 0.5 mL of the extract obtained as described in Section 2.2 was diluted 1:1 with 0.5 mL of LC-MS grade water. The diluted samples were then subjected to solid-phase extraction (SPE) using Oasis HLB 1 cc Vac Cartridges (Waters, SKU: WAT094225), operated under vacuum with a Supelco Visiprep™ DL SPE Vacuum Manifold (Cat. No. 57044). Before SPE, the cartridges were conditioned with 1 mL of methanol, followed by 1 mL of 0.1% (*v*/v) TFA, to activate the hydrophilic-lipophilic balanced sorbent and equilibrate the stationary phase for optimal peptide retention. Samples were then loaded onto the preconditioned cartridges. To remove non-specifically bound contaminants and low-affinity compounds, each cartridge was washed twice with 1 mL of a solution containing 5% (*v*/v) methanol in 0.1% (*v*/v) TFA. Peptides were then eluted with 1 mL of 100% (v/v) methanol and collected for downstream analysis. After elution, the samples were dried using a SpeedVac concentrator and reconstituted in 10 µL of 3% (v/v) acetonitrile with 0.1% (v/v) FA. The reconstituted samples were then sonicated in an ultrasonic bath at 100% amplitude for 10 min to ensure complete solubilization.

2.4. High-resolution mass spectrometry analysis

LC-MS/MS analysis was performed using an EASY nLC II chromatographer from Bruker Daltonics, coupled to an Ultra High-Resolution Quadrupole Time-of-Flight (UHR-QTOF) IMPACT HD mass spectrometer from Bruker Daltonics equipped with a CaptiveSpray nanoBoosterTM using acetonitrile as a dopant. Each sample was analysed in

technical duplicate. One μL of the sample was loaded onto a μPAC™ Trapping column (Thermo Scientific™) and desalted for 5 min with 0.1% (*v*/v) aqueous formic acid (FAaq) at a flow rate of 7 μL/min. The trap column was online connected for peptide separation on an analytical column (μPAC™ 50 cm Thermo Scientific™) at a flow rate of 500 nL/min using a linear gradient. The mobile phases consisted of 0.1% *v*/v aqueous FA (phase A) and 90% v/v acetonitrile with 0.1% v/v FA (phase B). The composition of the mobile phase changed over time as follows: from 0% to 5% (B) in 2 min, followed by a linear increase from 5% to 17% over 36 min. The gradient was then continued from 17% to 25% B over 14 min, and subsequently from 25% to 35% B over 9 min. Thereafter, the proportion of solvent B was ramped to 90% within 5 min and maintained isocratically at 90% for an additional 20 min. Mass spectrometry (MS) acquisition was configured with an MS scan rate of 2 Hz, followed by MS/MS scans at 8–32 Hz, maintaining a cycle time of 3.0 s. Active exclusion was applied, with exclusion after one spectrum and release after 2 min. Precursors were reconsidered if their intensity increased threefold over the previous intensity, with a fragmentation intensity threshold set at 2500 counts. The instrument was routinely calibrated using ESI-L Low Concentration tuning mix (G1969–85000, Agilent) before data acquisition. System suitability and analytical performance were monitored by analysing reference digests (HeLa cell lysate) at regular intervals. Quality control metrics included consistency in peptide-spectrum matches, protein identifications, and retention time stability.

2.5. Bioinformatic data analysis

LC-MS/MS data were analysed using PEAKS Studio 12 (Bioinformatics Solutions Inc.), utilising the PEAKS DB with an in-depth de novo assisted search, database matching, PEAKS PTM, and Spider algorithm. For data analysis, technical replicates were combined under a single sample annotation. Search parameters included a precursor mass tolerance of ≤25 ppm, a fragment ion, mass tolerance of ≤0.05 Da, and non-specific cleavage at both termini. Variable modifications considered were methionine oxidation, N-terminal acetylation, serine/threonine/tyrosine phosphorylation, and asparagine/glutamine deamidation. De novo–derived peptide tags were searched against a custom database containing *Homo sapiens*, *Sus scrofa*, *Bos taurus*, *Capra hircus*, *Gallus gallus*, *Ovis aries*, *Triticum aestivum*, *Oryza sativa*, *Zea mays*, *Nicotiana tabacum*, *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Staphylococcus aureus*, *Klebsiella pneumoniae,* and *Pseudomonas aeruginosa*. Peptide identifications were filtered at a peptide-spectrum match false discovery rate (FDR) of ≤1%, a minimum peptide length of 6 amino acids, and a confidence assignment accuracy (CAA) ≥80%. To ensure taxonomic specificity, only peptides uniquely assignable to a single taxon were retained, and non-unique matches in the custom database were excluded.

3. Results and discussion

3.1. Biological sex determination

Biological sex was successfully determined for all 20 specimens through the detection of sex-specific amelogenin peptides via high-resolution mass spectrometry. The peptide markers employed in this analysis were AMELX-1 (residues 44–52: SIRPPYPSY; $[M + 2H]^{2+}$ = 540.279649 *m/z*) and AMELY-2 (residues 58–64: SM(*ox*)IRPPY; $[M + 2H]^{2+}$ = 440.224600 m/z), the latter incorporating a methionine sulfoxide (oxidised methionine) modification. These peptides have been previously validated as reliable molecular indicators of chromosomal sex in proteomic studies of dental enamel by Stewart et al. and Parker et al. [4,5].

In all specimens identified as originating from biological males (samples 1, 4, 11–13, 15–17, 19, and 20), both AMELX-1 (residues 44–52: SIRPPYPSY) and AMELY-2 (residues 58–64: SM(*ox*)IRPPY) peptides were detected, thereby confirming the presence of the Y-chromosome-specific AMELY isoform. In contrast, the individuals identified as biological females (samples 2, 3, 5–10, 14, and 18) showed only the AMELX-1 peptide and no detectable signal for AMELY-2. This result aligns with the expected absence of amelogenin expression from the Y chromosome in females. Representative extracted ion chromatograms and tandem mass spectra for male and female profiles are shown in Fig. 1A and B, respectively (further information in Supplementary material, SM1 and Supplementary Fig. 1, SF1). To expand the repertoire of peptide biomarkers available for chromosomal sex determination, we applied deep learning-based de novo sequencing to interrogate the mass spectrometry dataset more comprehensively [20]. This approach enabled the identification of a wide range of previously unreported amelogenin-derived peptides, thereby expanding the molecular toolkit for biological sex estimation beyond the conventional targeted peptide markers. The newly identified peptides exhibit high sequence diversity yet retain conserved structural motifs, particularly the shared N-terminal segment YEVLTPLKWYQ(+0.98) (see SM1, SF2A, and SF2B). This region appears to function as a stable core scaffold across both AMELX and AMELY peptide families. Chromosome-specific distinctions are primarily observed in the C-terminal regions, which allow confident differentiation between isoforms. AMELX-derived peptides frequently terminate in SIRPPY, SIRPPYP, or their extended forms such as SIRPPYPSY. In contrast, AMELY-associated peptides consistently feature the SMIRPPY or SMIRPPYS motifs, often accompanied by methionine oxidation (+15.99 Da). Additionally, several AMELY peptides exhibited distinct N-terminal extensions (e.g., SYEVLTPLKWY…), which were not particularly observed among AMELX variants. These differences arise from isoform-specific proteolytic events, further enhancing the discriminatory power of these sequences. In parallel, several novel AMELX peptides were identified, including YEVLTPLKWY, YEVLTPLKWYQ(+0.98), and YEVLTPLKWYQ(+0.98)SIRPPYPSYG, many bearing characteristic modifications such as deamidation (+0.98 Da). These peptides proved valuable for sex estimation, particularly in cases of molecular degradation or low peptide abundance. Notably, sex determinations based on these newly identified sequences were fully concordant with those obtained using canonical amelogenin markers, confirming their specificity and relevance. This expanded peptide panel enhances analytical robustness and complements standard markers in forensic and archaeological applications, where degradation or limited material can hinder the detection of standard markers.

To assess the diagnostic utility of AMELY peptides, we performed receiver operating characteristic (ROC) curve analysis (Fig. 2A). Notably, the AMELY peptide YEVLTPLKWYQ(+0.98)SM(+15.99)IRPPY was consistently detected in all ten male samples, yielding a sensitivity of 100% within this cohort (Fig. 2B). This strong performance supports its potential role as a high-confidence biomarker for Y-chromosome presence. It reinforces its applicability in future sex estimation workflows where preservation of canonical peptides be compromised [21,22].

3.2. Peptide signatures provide dietary and environmental insights

To assess the potential of dental peptidomics for dietary reconstruction, we re-analysed the dataset using an expanded composite reference database including *Homo sapiens*, common livestock (*Sus scrofa*, *Bos taurus*, *Capra hircus*, *Gallus gallus*, *Ovis aries*), major crops (*Triticum aestivum*, *Oryza sativa*, *Zea mays*), as well as *Nicotiana tabacum*, *Arabidopsis thaliana*, and *Chlamydomonas reinhardtii*. This broader taxonomic scope enabled the detection of peptide traces reflecting both dietary input and environmental exposure. A total of 30,375 peptides were initially identified. After applying stringent filtering criteria, including a peptide-spectrum match FDR of ≤1%, a minimum peptide length of six amino acids, and a CAA of ≥80%, the dataset was refined to high-confidence identifications. CAA refers to the predicted probability of correct amino acid assignment in de novo peptide sequences, as

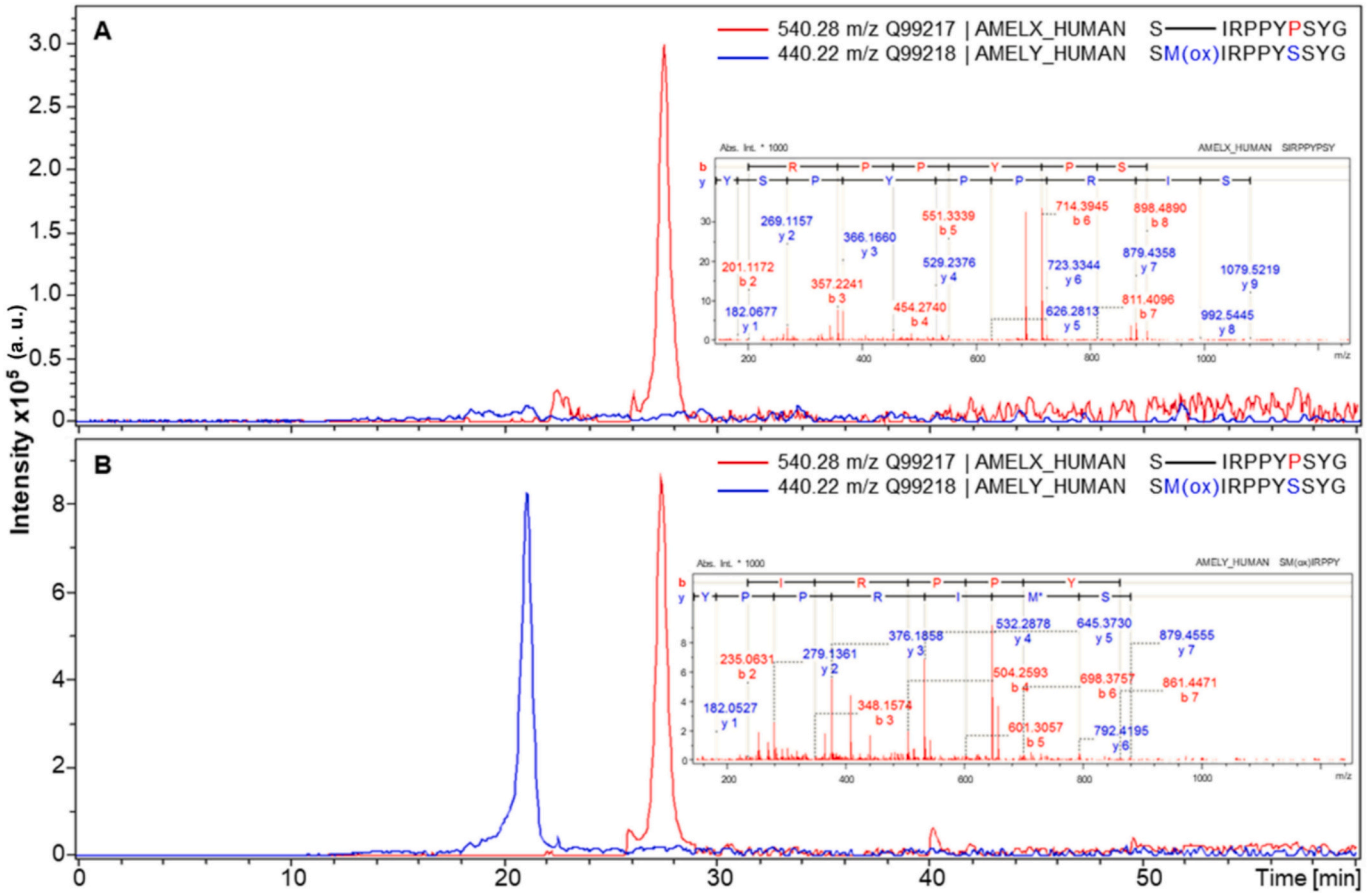
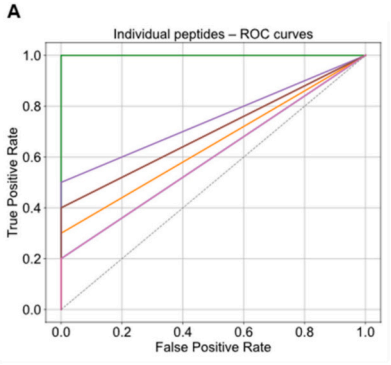


Fig. 1. Extracted Ion Chromatograms (EIC) and MS/MS spectra of the ion type [M + 2H]2+ of the peptides SIRPPYPSYG (AMELX) – *m/z* = 540.28, z = 2+ and retention time = 28 min; and SM(*ox*)IRPPYSSYG (AMELY) – m/z = 440.22, z = 2+ and retention time = 21 min. **(A)** EIC of biological female specimens, with the AMELX peptide (m/z = 540.28, red) peak. The inset displays the corresponding MS/MS spectrum confirming the AMELX peptide sequence. **(B)** EIC of a biological male sample, with both AMELX (red) and AMELY (m/z = 440.22, blue) peptides' peaks. The inset shows the MS/MS spectrum validating the AMELY peptide sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



B

Colour (ROC Curve)	Peptide	Area under the curve (AUC)
	SYEVLTPLKWYQ(+0.98)SM(+15.99)IRPPY	0.60
	SYEVLTPLKWYQ(+0.98)SMIRPPY	0.65
	YEVLTPLKWYQ(+0.98)SM(+15.99)IRPPY	1.00
	YEVLTPLKWYQ(+0.98)SM(+15.99)IRPPYS	0.70
	YEVLTPLKWYQ(+0.98)SMIRPPY	0.75
	YEVLTPLKWYQ(+0.98)SMIRPPYS	0.70
	YEVLTPLKWYQSM(+15.99)IRPPYS	0.60

Fig. 2. ROC curve analysis of AMELY-derived peptides identified by de novo sequencing. **(A)** ROC curves for individual peptides, showing True Positive Rates (detections in male samples) vs. False Positive Rates (detections in female samples). **(B)** Table listing the peptides from Panel A with corresponding curve colours and Area Under the Curve (AUC) values. Higher AUCs indicate greater accuracy in identifying male samples.

estimated by the PEAKS software. To ensure taxonomic reliability, peptides assigned to humans or bacteria were excluded, and only sequences uniquely matching a single taxon in the UniProt database were retained. After all filtering steps, the dataset was reduced to 656 high-confidence, species-specific peptides (see SM2). Among these, peptides assigned to domesticated plants and animals (Fig. 3A) suggest exposure to both cereal crops and animal-derived foods.

Wheat (*T. aestivum*) was the most confidently identified plant, with 56 unique peptides mapped to 54 distinct proteins, indicating strong taxonomic representation in the peptidomic dataset. These include enzymes such as starch synthase and trehalose-6-phosphate phosphatase, which are involved in carbohydrate metabolism and stress responses [23,24]. Maize (*Z. mays*) was represented by 41 unique peptides, including one protein confirmed by two peptide matches, consistent with its dietary presence via trade or cultivation [25–27]. The detection of proteins from *Nicotiana tabacum* (26 unique peptides), including

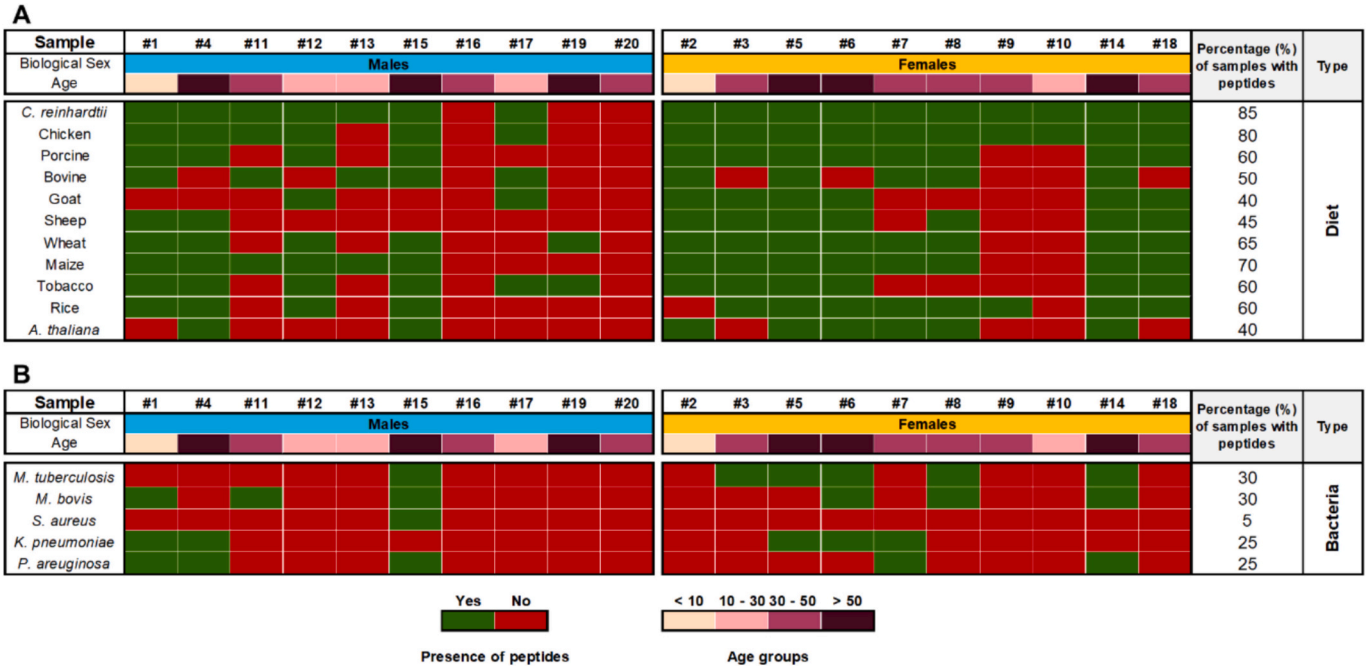


Fig. 3. Representation of the presence of identified peptides from various species across all specimens. Each cell indicates whether peptides from a given species were present (Yes - green) or absent (No - red) in a specific sample. Samples are organized by biological sex (males – blue; females – yellow) and by biological age: beige: < 10 years; light pink: 10–30 years; pink: 30–50 years; dark pink: > 50 years. **(A)** Shows dietary peptides, along with their abundance across all samples (expressed as a percentage). **(B)** Presents the bacterial peptides identified in the samples, with their overall abundance of detection in all samples (in percentage). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

MYB122-like and F-box proteins, is consistent with the presence of tobacco, which has been associated with medicinal and ritual use in various pre-contact contexts [28,29]. Among the 12 individuals with tobacco peptides, age groups ranged from infants to the elderly, and both sexes were equally represented. The presence of these peptides in young individuals supports a medicinal use hypothesis [30].

Animal-derived proteins further address dietary practices. Chicken (*G. gallus*) was the second most abundant source, with 44 peptides, including kidney-specific sphingosine kinase 1 (SPHK1) [31,32]. This is consistent with historical records: during the 19th–20th centuries, chicken was more accessible than pork or beef and likely a dietary staple among lower-income groups. The individuals in this study, from a documented skeletal collection in Bologna, were of low socioeconomic status, with evidence of poor health and limited resources (Mariotti et al., 2015; Belcastro et al., 2017) [17,33]. Pig (*S. scrofa*) proteins (33 peptides) included obscurin, associated with meat quality, and GNMT, a liver-specific enzyme, indicating consumption of both muscle and organ meats [34–37]. Cow (*B. taurus*) dairy consumption was inferred from 26 peptides, notably being identified β-casein (CSN2), a key milk protein [38–40]. Its presence supports dietary exposure and aligns with previous identifications of milk proteins in ancient calculus [9,41]. Goat (*C. hircus*) and sheep (*O. aries*) consumption was supported by 13 and 16 unique peptides, respectively. Goat peptides included TANC2, a protein expressed in neurons and glial cells, while a sheep-derived peptide from MYBPC3, involved in cardiac muscle contraction, suggests cerebral and heart tissues consumption [42,43].

Proteins from *C. reinhardtii*, a unicellular freshwater green alga commonly found in moist terrestrial and aquatic environments, were also detected, with 352 unique peptides mapped to 309 proteins. Of these, 30 proteins were identified by more than one unique peptide, increasing confidence in their identification. Given its terrestrial habitat, the presence of *C. reinhardtii* likely reflects soil contamination from the burial environment [44]. Tandem mass spectra for identified peptides are provided in SM1, SF3A, and SF3B.

3.3. Teeth microbiome and bacterial peptides

An initial screen of the peptidomic dataset revealed four dominant bacterial taxa across the 20 individuals: *Mycobacterium* spp., *Staphylococcus aureus*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa*. To explore potential bacterial exposure, we compared our dataset against a curated reference database comprising these species. A total of 3503 bacterial peptides were identified, which, after applying stringent filters (CAA) ≥ 80%, species-level specificity, and exclusion of peptides annotated as human, animal, or plant) were reduced to 28 high-confidence, species-specific bacterial peptides (SM2). Cross-referencing with the expanded Human Oral Microbiome Database (*e*HOMD) [45] confirmed the presence of all four taxa, with the addition of *Mycobacterium bovis*, which is not listed in *e*HOMD (Fig. 3B). Corresponding tandem mass spectra are available in SM1 and SF4.

Approximately 50% of the bacterial peptides originated from *Mycobacterium* strains. *Mycobacterium tuberculosis* was confidently identified by 11 unique peptides, several mapping to PPE proteins from the PE/PPE family, known to mediate immune evasion, antibiotic resistance, and intracellular survival [46]. In total, peptides from *M. bovis* were identified in 6 of the 20 individuals. Given that *M. bovis* is a zoonotic member of the *Mycobacterium tuberculosis* complex (MTBC) capable of causing tuberculosis in humans, and that known transmission routes include ingestion of unpasteurised dairy products, this indicate dietary exposure [47]. Supporting this, several *M. bovis* peptides were associated with proteins involved in pathogenicity and resistance, such as alpha/beta hydrolases and MmpL family proteins [48,49]. Two individuals had documented tuberculosis-related causes of death. Sample 3 belonged to a person who died of pulmonary tuberculosis, and sample 15 to one diagnosed with Pott's disease (spinal tuberculosis), an extrapulmonary form of the disease [50]. *Mycobacterium*-derived peptides were found in both cases, with sample 15 exhibiting both *M. tuberculosis* and *M. bovis*, and sample 3 only one peptide from *M. tuberculosis*. Regarding *Staphylococcus aureus*, a single peptide from the extracellular matrix-binding protein Ebh was detected. This protein

is associated with host colonisation and reflect transient exposure rather than infection [51,52].

The presence of bacterial peptides in dental-associated tissues and plaque is a growing area of research [53]. While bacterial proteins are known to be part of the dental pellicle, a surface protein layer, the extent to which they penetrate deeper dental structures remains unclear. Mass spectrometry-based studies have identified bacterial proteins in dental pulp, likely due to vascular dissemination [54]. Although enamel and dentin are highly mineralised and resistant to infiltration, structural defects such as microcracks or dentinal tubules permit bacterial ingress [55,56]. Our findings align with these observations and contribute new evidence for the potential of dental-based peptidomics in exploring historical oral microbiomes and microbial exposure.

3.4. Peptide deamidation for natural ageing estimation

In this study, we investigated the relationship between biological age and (i) asparagine deamidation, (ii) glutamine deamidation, and (iii) asparagine and glutamine simultaneous deamidation. All these analyses were performed by calculating the deamidation percentage of each sample, both by peptide intensity and number. Additionally, the analyses were considered for both (i) all identified peptides in each sample, and (ii) AMELX identified peptides only.

Neither asparagine nor glutamine deamidation alone displays a clear increase or decrease with biological age. Instead, both exhibit random distribution across all 20 samples, whether analysing all deamidated glutamine and asparagine residues (Fig. 4) or AMELX peptides exclusively (see SF5). The assessment of the simultaneous N- and Q deamidation percentage also revealed a random distribution. Additionally, we searched for peptides present simultaneously in their N-deamidated and non-deamidated (native) forms. We identified five such peptides in five or more samples. For each sample in which both forms were present, we calculated the N-deamidated/non-deamidated ratio. Again, no trend was found (see SF6).

The absence of a consistent correlation suggests that peptide deamidation cannot be reliably interpreted as a proxy for biological age-at-death in this archaeological context. This result is not unexpected, as deamidation rates are influenced by a range of factors beyond endogenous ageing, including protein sequence and structure, burial environment, pH, temperature, and post-mortem diagenetic processes [57,58]. In archaeological samples, these extrinsic factors likely dominate the deamidation signal and mask any subtle contributions from endogenous ageing processes.

4. Conclusions

This study introduced a rapid, ultrasound-assisted peptide extraction protocol for dental tissues that preserves macroscopically visible structural integrity. The method enables simultaneous recovery of diverse molecular signals from a single extraction. The resulting peptide fractions are compatible with high-resolution peptidomics analyses.

Biological sex estimation was achieved in all individuals using canonical and newly identified amelogenin-derived peptides. The incorporation of de novo sequencing expanded the peptide-based toolkit for chromosomal sex estimation, enhancing analytical robustness, particularly in degraded or size-limited samples. Dietary reconstruction revealed a diverse intake consistent with historical socioeconomic patterns. Peptides from domesticated animals (chicken, pig, cow, goat, sheep) and cultivated plants (wheat, maize, rice) suggest consumption of a varied diet, including both plant- and animal-derived foods such as

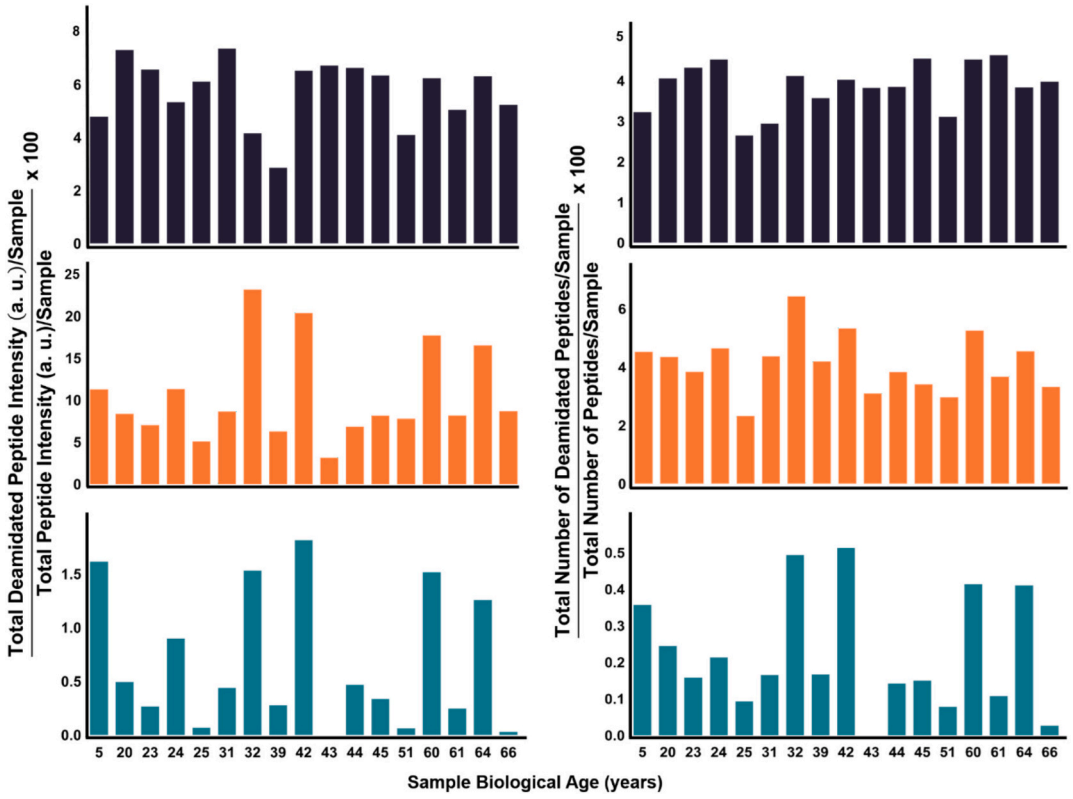


Fig. 4. Analysis of the deamidated peptides across all specimens. Demonstration of the relationship between the individuals' biological age (in years) and the deamidation percentage of each sample, both calculated by peptide intensity or peptide number. The top graphs demonstrate the correlation of N-deamidated peptides (black), the middle graphs represent the correlation of Q-deamidated peptides (orange), and the bottom graphs show the correlation of simultaneously N- and Q-deamidated peptides (greenish blue). **Abbreviations:** a. u. – arbitrary units. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dairy and organ meats. The detection of tobacco-derived peptides in both adults and children suggests non-dietary, likely medicinal or ritual, use. Environmental signals from the burial environment were captured through the detection of *C. reinhardtii*, interpreted as soil contamination. In addition, bacterial peptides, particularly from *M. tuberculosis* and *M. bovis,* were identified in several individuals, including two with documented tuberculosis-related causes of death. Although peptide deamidation was investigated as a potential proxy for biological age-at-death, the data did not reveal a consistent correlation. This lack of correlation reflect the confounding influence of post-mortem diagenesis on peptide deamidation, limiting its applicability as a direct proxy for chronological age in archaeological samples. Interpretations related to diet, microbiome, and environmental contact are based on high-confidence, species-specific peptides. While these findings are taxonomically robust, they are best understood as indicative of potential exposure, rather than definitive evidence of ingestion, infection, or direct usage. Overall, the approach presented here highlights the value of dental-based peptidomics as a powerful, integrative tool in bio-archaeological and forensic sciences. It enables the recovery of biological, cultural, and environmental information from a single source with minimal material loss, offering a new standard for molecular investigations of ancient human remains.

CRediT authorship contribution statement

Catarina A.P. André: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Raquel Fonseca:** Visualization, Investigation, Data curation. **André Q. Figueiredo:** Writing – review & editing, Visualization, Formal analysis, Data curation. **Carlos Lodeiro:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Stefano Benazzi:** Writing – review & editing, Validation, Resources, Conceptualization. **Maria Giovanna Belcastro:** Validation, Resources, Project administration, Conceptualization. **Federico Lugli:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition. **Giulia Di Rocco:** Writing – review & editing, Writing – original draft, Validation, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Hugo M. Santos:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **José L. Capelo:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by national funds from FCT – Fundação para a Ciência e a Tecnologia, I.P., under the scope of the Project UID/50006/2023 of the Associate Laboratory for Green Chemistry – LAQV REQUIMTE. Details here: https://laqv.requimte.pt/a/74-funding. H.M.S. acknowledges the Associate Laboratory for Green Chemistry – LAQV REQUIMTE (LA/P/0008/2020; DOI:10.54499/LA/P/0008/2020), funded by FCT I.P. for his research contract. A.Q.F. acknowledges FCT I.P. for the PhD grant (reference 2023.00528.BD). PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry – *Isabel Moura* (#PM001/2024, #PM001/2019, and #PM003/2016).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2026.116915.

Data availability

Data will be made available on request.

References

[1] J.A. Gamble, et al., Advancing sex estimation from amelogenin: applications to archaeological, deciduous, and fragmentary dental enamel, J. Archaeol. Sci. Rep. 54 (2024) 104430, https://doi.org/10.1016/j.jasrep.2024.104430.
[2] B.H.A. Mai, M. Drancourt, G. Aboudharam, Ancient dental pulp: Masterpiece tissue for paleomicrobiology, Mol Genet Genomic Med 8 (6) (2020), https://doi.org/10.1002/mgg3.1202.
[3] C. Wadsworth, et al., Comparing ancient DNA survival and proteome content in 69 archaeological cattle tooth and bone samples from multiple European sites, J. Proteome 158 (2017) 1–8, https://doi.org/10.1016/j.jprot.2017.01.004.
[4] G.J. Parker, et al., Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel, J. Archaeol. Sci. 101 (2019) 169–180, https://doi.org/10.1016/j.jas.2018.08.011.
[5] N.A. Stewart, R.F. Gerlach, R.L. Gowland, K.J. Gron, J. Montgomery, Sex determination of human remains from peptides in tooth enamel, Proc. Natl. Acad. Sci. 114 (52) (2017) 13649–13654, https://doi.org/10.1073/pnas.1714926115.
[6] C. Froment, et al., Analysis of 5000 year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides, J. Proteome 211 (2020) 103548, https://doi.org/10.1016/j.jprot.2019.103548.
[7] F. Lugli, et al., Enamel peptides reveal the sex of the late antique "lovers of Modena", Sci. Rep. 9 (1) (2019) 13130, https://doi.org/10.1038/s41598-019-49562-7.
[8] R.R. Jersie-Christensen, et al., Quantitative metaproteomics of medieval dental calculus reveals individual oral health status, Nat. Commun. 9 (1) (2018) 4744, https://doi.org/10.1038/s41467-018-07148-3.
[9] J. Hendy, et al., Proteomic evidence of dietary sources in ancient dental calculus, Proc. R. Soc. B Biol. Sci. 285 (1883) (2018) 20180977, https://doi.org/10.1098/rspb.2018.0977.
[10] V.D. La, G. Aboudharam, D. Raoult, M. Drancourt, 'Dental pulp as a tool for the retrospective diagnosis of infectious diseases', in *paleomicrobiology*, Berlin, Heidelberg: springer, Berlin Heidelberg (2008) 175–196, https://doi.org/10.1007/978-3-540-75855-6_11.
[11] O. Bender, et al., The hidden secrets of the dental Calculus: calibration of a mass spectrometry protocol for dental Calculus protein analysis, Int. J. Mol. Sci. 23 (22) (2022) 14387, https://doi.org/10.3390/ijms232214387.
[12] Y. Lin, et al., Omics for deciphering oral microecology, Int. J. Oral Sci. 16 (1) (2024) 2, https://doi.org/10.1038/s41368-023-00264-x.
[13] N. Procopio, A.T. Chamberlain, M. Buckley, Exploring biological and geological age-related changes through variations in intra- and Intertooth proteomes of ancient dentine, J. Proteome Res. 17 (3) (2018) 1000–1013, https://doi.org/10.1021/acs.jproteome.7b00648.
[14] M. Pal Chowdhury, M. Buckley, Trends in deamidation across archaeological bones, ceramics and dental calculus, Methods 200 (2022) 67–79, https://doi.org/10.1016/j.ymeth.2021.08.004.
[15] A.M. Casas-Ferreira, M. del Nogal-Sánchez, Á.E. Arroyo, J.V. Vázquez, J.L. Pérez-Pavón, Fast methods based on mass spectrometry for peptide identification. Application to sex determination of human remains in tooth enamel, Microchem. J. 181 (2022) 107645, https://doi.org/10.1016/j.microc.2022.107645.
[16] M.G. Belcastro, A. Pietrobelli, T. Nicolosi, M. Milella, V. Mariotti, Scientific and ethical aspects of identified skeletal series: the case of the documented human osteological collections of the University of Bologna (northern Italy), Forensic Sciences 2 (2) (2022) 349–361, https://doi.org/10.3390/forensicsci2020025.
[17] M.G. Belcastro, B. Bonfiglioli, M.E. Pedrosi, M. Zuppello, V. Tanganelli, V. Mariotti, The history and composition of the identified human skeletal collection of the Certosa cemetery (Bologna, Italy, 19th–20th century), Int. J. Osteoarchaeol. 27 (5) (2017) 912–925, https://doi.org/10.1002/oa.2605.
[18] R. Sorrentino, A. Pietrobelli, D. Mameli, V. Mariotti, T. Nicolosi, M.G. Belcastro, The Virtual Database of the Documented Human Osteological Collection (DHOC) of the Certosa Cemetery of Bologna (Italy, 19th–20th Century), American Journal of Biological Anthropology 186 (2) (2025), https://doi.org/10.1002/ajpa.25065.
[19] D.R. Green, F. Schulte, K.-H. Lee, M.K. Pugach, M. Hardt, F.B. Bidlack, Mapping the tooth enamel proteome and Amelogenin phosphorylation onto mineralizing porcine tooth crowns, Front. Physiol. 10 (2019), https://doi.org/10.3389/fphys.2019.00925.
[20] N.H. Tran, et al., Deep learning enables *de novo* peptide sequencing from data-independent-acquisition mass spectrometry, Nat. Methods 16 (1) (2019) 63–66, https://doi.org/10.1038/s41592-018-0260-3.
[21] I. Štamfelj, Sex estimation based on the analysis of enamel peptides: false assignments due to AMELY deletion, J. Archaeol. Sci. 130 (2021) 105345, https://doi.org/10.1016/j.jas.2021.105345.
[22] G.J. Parker, et al., AMELY deletion is not detected in systematically sampled reference populations: a reply to Štamfelj, J. Archaeol. Sci. 130 (2021) 105354, https://doi.org/10.1016/j.jas.2021.105354.

[23] A. Kumari, et al., 'Characterization of the starch synthase under terminal heat stress and its effect on grain quality of wheat', *3*, Biotech 10 (12) (2020) 531, https://doi.org/10.1007/s13205-020-02527-4.
[24] M.A. Islam, et al., In silico and transcription analysis of Trehalose-6-phosphate phosphatase gene family of wheat: Trehalose synthesis genes contribute to salinity, drought stress and leaf senescence, Genes (Basel) 12 (11) (2021) 1652, https://doi.org/10.3390/genes12111652.
[25] Y. Zhao, et al., The evening complex promotes maize flowering and adaptation to temperate regions, Plant Cell 35 (1) (2023) 369–389, https://doi.org/10.1093/plcell/koac296.
[26] Y. Xie, et al., *ZmELF3.1* integrates the *RA2-TSH4* module to repress maize tassel branching, New Phytol. 241 (1) (2024) 490–503, https://doi.org/10.1111/nph.19329.
[27] X. Quan, et al., Genome-wide and transcriptome analysis of Jacalin-related lectin genes in barley and the functional characterization of HvHorcH in low-nitrogen tolerance in Arabidopsis, Int. J. Mol. Sci. 24 (23) (2023) 16641, https://doi.org/10.3390/ijms242316641.
[28] F.-B. Zhang, S.-X. Ji, J.-G. Yang, X.-W. Wang, W.-H. Han, Genome-wide analysis of MYB family in Nicotiana benthamiana and the functional role of the key members in resistance to Bemisia tabaci, Int. J. Biol. Macromol. 235 (2023) 123759, https://doi.org/10.1016/j.ijbiomac.2023.123759.
[29] Q. Li, X. Zhao, J. Wu, H. Shou, W. Wang, The F-box protein TaFBA1 positively regulates drought resistance and yield traits in wheat, Plants 13 (18) (2024) 2588, https://doi.org/10.3390/plants13182588.
[30] S. Mishra, M. Mishra, Tobacco: its historical, cultural, oral, and periodontal health association, J Int Soc Prev Community Dent 3 (1) (2013) 12, https://doi.org/10.4103/2231-0762.115708.
[31] H. Chen, et al., Expression profile of sphingosine kinase 1 isoforms in human Cancer tissues and cells: importance and clinical relevance of the neglected 1b-isoform, J. Oncol. 2022 (2022) 1–12, https://doi.org/10.1155/2022/2250407.
[32] P. Guerre, M. Matard-Mann, P. Nyvall Collén, Targeted sphingolipid analysis in chickens suggests different mechanisms of fumonisin toxicity in kidney, lung, and brain, Food Chem. Toxicol. 170 (2022) 113467, https://doi.org/10.1016/j.fct.2022.113467.
[33] V. Mariotti, et al., Skeletal evidence of tuberculosis in a modern identified human skeletal collection (Certosa cemetery, Bologna, Italy), Am. J. Phys. Anthropol. 157 (3) (2015) 389–401, https://doi.org/10.1002/ajpa.22727.
[34] M. López-Pedrouso, J.M. Lorenzo, A. Cittadini, M.V. Sarries, M. Gagaoua, D. Franco, A proteomic approach to identify biomarkers of foal meat quality: a focus on tenderness, color and intramuscular fat traits, Food Chem. 405 (2023) 134805, https://doi.org/10.1016/j.foodchem.2022.134805.
[35] I.H. Malgwi, V. Halas, P. Grünvald, S. Schiavon, I. Jócsák, Genes related to fat metabolism in pigs and intramuscular fat content of pork: a focus on Nutrigenetics and nutrigenomics, Animals 12 (2) (2022) 150, https://doi.org/10.3390/ani12020150.
[36] N.V.L. Serão, et al., Candidate gene expression and intramuscular fat content in pigs, J. Anim. Breed. Genet. 128 (1) (2011) 28–34, https://doi.org/10.1111/j.1439-0388.2010.00887.x.
[37] J. Wang, et al., Diquat determines a deregulation of lncRNA and mRNA expression in the liver of Postweaned piglets, Oxidative Med. Cell. Longev. 2019 (2019) 1–9, https://doi.org/10.1155/2019/9148535.
[38] M. Miluchová, M. Gábor, J. Candrák, The effect of the genotypes of the CSN2 gene on test-day Milk yields in the Slovak Holstein cow, Agriculture 13 (1) (2023) 154, https://doi.org/10.3390/agriculture13010154.
[39] V. Bisutti, et al., The β-casein (CSN2) A2 allelic variant alters milk protein profile and slightly worsens coagulation properties in Holstein cows, J. Dairy Sci. 105 (5) (2022) 3794–3809, https://doi.org/10.3168/jds.2021-21537.
[40] N. Amalfitano, L.F. Macedo Mota, M. GuilhermeJ, A. Cecchinato Rosa, G. Bittante, Role of CSN2, CSN3, and BLG genes and the polygenic background in the cattle milk protein profile, J. Dairy Sci. 105 (7) (2022) 6001–6020, https://doi.org/10.3168/jds.2021-21421.
[41] A. Kensche, S. Pötschke, C. Hannig, A. Dürasch, T. Henle, M. Hannig, Efficacy of mouthrinses with bovine milk and milk protein isolates to accumulate casein in the in situ pellicle, Clin. Oral Investig. 24 (11) (2020) 3871–3880, https://doi.org/10.1007/s00784-020-03253-0.
[42] D. Gudra, et al., Genetic characterization of the Latvian local goat breed and genetic traits associated with somatic cell count, Animal 18 (5) (2024) 101154, https://doi.org/10.1016/j.animal.2024.101154.
[43] C.Y. Doh, A.V. Schmidt, K. Chinthalapudi, J.E. Stelzer, Bringing into focus the central domains C3-C6 of myosin binding protein C, Front. Physiol. 15 (2024), https://doi.org/10.3389/fphys.2024.1370539.
[44] S. Sasso, H. Stibor, M. Mittag, A.R. Grossman, From molecular manipulation of domesticated *Chlamydomonas reinhardtii* to survival in nature, Elife 7 (2018), https://doi.org/10.7554/eLife.39233.
[45] HOMD :: Human Oral Microbiome Database, Accessed: Mar. 31 (2025) [Online]. Available: https://www.homd.org/.
[46] C. D'Souza, U. Kishore, A.G. Tsolaki, The PE-PPE family of *Mycobacterium tuberculosis*: proteins in disguise, Immunobiology 228 (2) (2023) 152321, https://doi.org/10.1016/j.imbio.2022.152321.
[47] C. Kanipe, M.V. Palmer, *Mycobacterium bovis* and you: a comprehensive look at the bacteria, its similarities to *Mycobacterium tuberculosis*, and its relationship with human disease, Tuberculosis 125 (2020) 102006, https://doi.org/10.1016/j.tube.2020.102006.
[48] G. Johnson, The α/β hydrolase fold proteins of *Mycobacterium tuberculosis*, with reference to their contribution to virulence, Curr. Protein Pept. Sci. 18 (3) (2017) 190–210, https://doi.org/10.2174/1389203717666160729093515.
[49] A. Viljoen, V. Dubois, F. Girard-Misguich, M. Blaise, J. Herrmann, L. Kremer, The diverse family of MmpL transporters in mycobacteria: from regulation to antimicrobial developments, Mol. Microbiol. 104 (6) (2017) 889–904, https://doi.org/10.1111/mmi.13675.
[50] I. Glassman, K. Nguyen, J. Giess, C. Alcantara, M. Booth, V. Venketaraman, Pathogenesis, diagnostic challenges, and risk factors of Pott's disease, Clin Pract 13 (1) (2023) 155–165, https://doi.org/10.3390/clinpract13010014.
[51] N. Khan, et al., The giant staphylococcal protein Embp facilitates colonization of surfaces through Velcro-like attachment to fibrillated fibronectin, Elife 11 (2022), https://doi.org/10.7554/eLife.76164.
[52] S.R. Clarke, L.G. Harris, R.G. Richards, S.J. Foster, Analysis of Ebh, a 1.1-Mega-dalton Cell Wall-associated fibronectin-binding protein of *Staphylococcus aureus*, Infect. Immun. 70 (12) (2002) 6680–6687, https://doi.org/10.1128/IAI.70.12.6680-6687.2002.
[53] J. Enax, B. Ganss, B.T. Amaechi, E. Schulze Zur Wiesche, F. Meyer, 'The composition of the dental pellicle: an updated literature review', *Frontiers*, Oral Health 4 (2023), https://doi.org/10.3389/froh.2023.1260442.
[54] R. Barbieri, et al., Paleoproteomics of the dental pulp: the plague paradigm, PLoS One 12 (7) (2017) e0180552, https://doi.org/10.1371/journal.pone.0180552.
[55] J.R. Kina, J. Kina, E.F.U. Kina, M. Kina, A.M.P. Soubhia, Presence of bacteria in dentinal tubules, J. Appl. Oral Sci. 16 (3) (2008) 205–208, https://doi.org/10.1590/S1678-77572008000300008.
[56] T.G. Fagrell, P. Lingström, S. Olsson, F. Steiniger, J.G. Norén, Bacterial invasion of dentinal tubules beneath apparently intact but hypomineralized enamel in molar teeth with molar incisor hypomineralization, Int. J. Paediatr. Dent. 18 (5) (2008) 333–340, https://doi.org/10.1111/j.1365-263X.2007.00908.x.
[57] Y. Jin, Y. Yi, B. Yeung, Mass spectrometric analysis of protein deamidation – a focus on top-down and middle-down mass spectrometry, Methods 200 (2022) 58–66, https://doi.org/10.1016/j.ymeth.2020.08.002.
[58] H. Tonie Wright, D.W. Urry, Nonenzymatic Deamidation of Asparaginyl and Glutaminyl residues in protein, Crit. Rev. Biochem. Mol. Biol. 26 (1) (1991) 1–52, https://doi.org/10.3109/10409239109081719.