



Kent Academic Repository

Madupe, Palesa P., Taurozzi, Alberto J., Koenig, Claire, Patramanis, Ioannis, Munir, Fazeelah, Dickinson, Marc R., Mackie, Meaghan, Troché, Gaudry, Parker, Glendon, Kyriakidou, Pelagia and others (2026) *Proteomic analysis of dental enamel from 20 Homo naledi individuals show no male markers*. Cell, 189 (4). p4415-4424.e11. ISSN 0092-8674.

Downloaded from

<https://kar.kent.ac.uk/115663/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.cell.2026.05.044>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

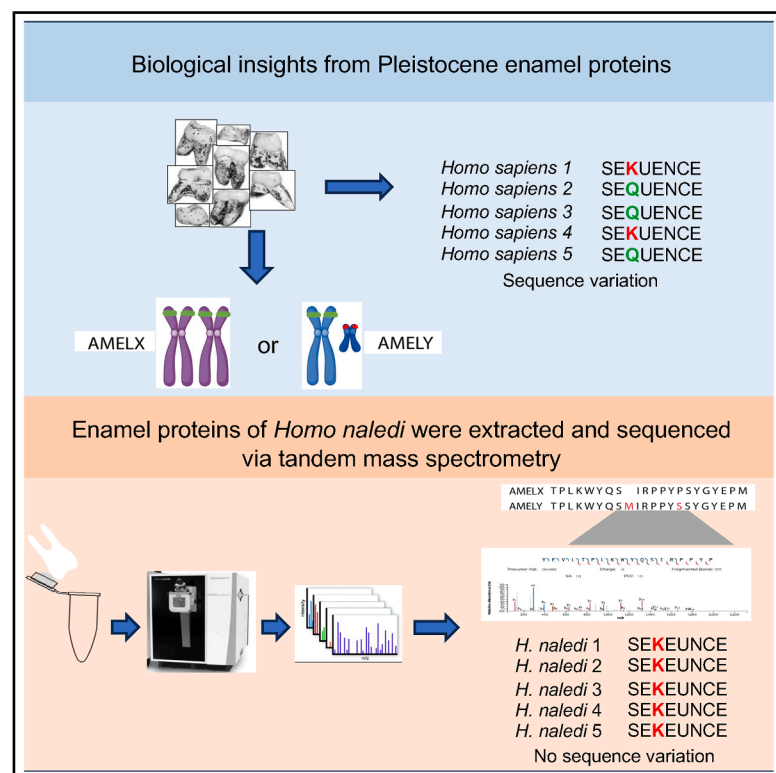
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Proteomic analysis of dental enamel from 20 *Homo naledi* individuals shows no male markers

Graphical abstract



Authors

Palesa P. Madupe, Alberto J. Taurozzi, Claire Koenig, ..., John Hawks, Lee Berger, Enrico Cappellini

Correspondence

palesa.madupe@palaeoproteomics.org (P.P.M.),
 lrberger@ngs.org (L.B.),
 ecappellini@sund.ku.dk (E.C.)

In brief

The morphological homogeneity of the middle Pleistocene hominin *Homo naledi* is reflected in its dental proteome, with no evidence supporting confident identification of male individuals, and the detection of both derived and ancestral amino acid substitutions in *Homo naledi*.

Highlights

- Out of 20 *H. naledi* individuals screened, no male individuals were confidently identified
- There is no variation in the protein sequences recovered
- *H. naledi* has an Amelogenin X unique amino acid within the Hominin group
- *H. naledi* COL1A17 shows an ancestral variant shared with *Paranthropus robustus*



Article

Proteomic analysis of dental enamel from 20 *Homo naledi* individuals shows no male markers

Palesa P. Madupe,^{1,2,16,21,*} Alberto J. Taurozzi,¹ Claire Koenig,^{3,17} Ioannis Patramanis,¹ Fazeelah Munir,⁴ Marc R. Dickinson,⁴ Meaghan Mackie,^{1,3,18,19} Gaudry Troché,¹ Glendon Parker,⁵ Pelagia Kyriakidou,⁶ Patrick Mahoney,^{7,8} Gina McFarlane,^{7,20} Bernhard Zipfel,⁹ Jürgen Cox,⁶ Kirsty Penkman,⁴ Lauren Schroeder,^{2,10} Rebecca R. Ackermann,^{2,11} Jesper V. Olsen,³ John Hawks,^{12,13} Lee Berger,^{13,14,15,21,*} and Enrico Cappellini^{1,21,*}

¹Globe Institute, University of Copenhagen, Copenhagen K, Denmark

²Human Evolution Research Institute (HERI), University of Cape Town, Cape Town, South Africa

³Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen N, Denmark

⁴Department of Chemistry, University of York, York YO10 5DD, UK

⁵Department of Environmental Toxicology, University of California, Davis, Davis, CA, USA

⁶Computational Systems Biochemistry, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

⁷Skeletal Biology Research Centre, School of Anthropology and Conservation, University of Kent, Kent, UK

⁸School of Natural Sciences, University of Kent, Kent, UK

⁹Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

¹⁰Department of Anthropology, University of Toronto Mississauga, Toronto, Canada

¹¹Department of Archaeology, University of Cape Town, Cape Town, South Africa

¹²Department of Anthropology, University of Wisconsin-Madison, Madison, WI, USA

¹³Centre for the Exploration of the Deep Human Journey, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

¹⁴The National Geographic Society, 1145 17th Street NW, Washington, D.C., USA

¹⁵The Carnegie Institution for Science, 5251 Broad Branch Rd NW, Washington, D.C., USA

¹⁶Present address: Max Planck for Evolutionary Anthropology, Leipzig, Germany

¹⁷Present address: Max Delbrück Center, Berlin, Germany

¹⁸Present address: Department of Archaeology, University College Dublin, Dublin, Ireland

¹⁹Present address: Archaeobiomics, Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy

²⁰Present address: Anthropology, Faculty of Arts and Education, University of Auckland, Auckland, New Zealand

²¹Lead contact

*Correspondence: palesa.madupe@palaeoproteomics.org (P.P.M.), lrberger@ngs.org (L.B.), ecappellini@sund.ku.dk (E.C.)

<https://doi.org/10.1016/j.cell.2026.05.044>

SUMMARY

The Rising Star cave system excavations resulted in a high number of well-preserved skeletal specimens from multiple individuals of *Homo naledi*, showing a high degree of morphological homogeneity, including dental variation possibly consistent with a single-sex sample. Here, we report the paleoproteomic analysis of dental enamel proteins extracted via micro-destructive acid etching from 23 *H. naledi* specimens belonging to a minimum of 20 individuals. After excluding the possibility of technical bias, no convincing evidence supporting the confident identification of male individuals was detected in any of the investigated samples. We also detect no variability in the recovered proteome, and we observe two amino acid substitutions: a derived one in amelogenin X compared with *Homo*, and an ancestral one in COL17A1, also present in *Paranthropus robustus*. Our results further support the homogeneity of *H. naledi* fossils and show how to sustainably investigate extinct hominins.

INTRODUCTION

Beginning in 2013, fossils assigned to the species *Homo naledi* were recovered from the Dinaledi Chamber inside the Rising Star cave system in the Cradle of Humankind, South Africa (Figures 1A–1C). Based on dental remains, this fossil record represents at least 15 individuals.² Since then, additional *H. naledi* remains have been found in the same and in other chambers

of the cave system, ranging from infants to older adults,^{3,4} further expanding the sample of individuals known for this species. Some of these remains have been dated to 335–236 ka (kilo annum/thousand years),^{5,6} although the estimated range for some specimens may be a wider interval (± 70 ka).⁵ Among hominins, *H. naledi* is unusual for several reasons. The morphology of this species is a mosaic, combining a small brain, shoulders, and trunk similar to early hominins such as *Australopithecus*,



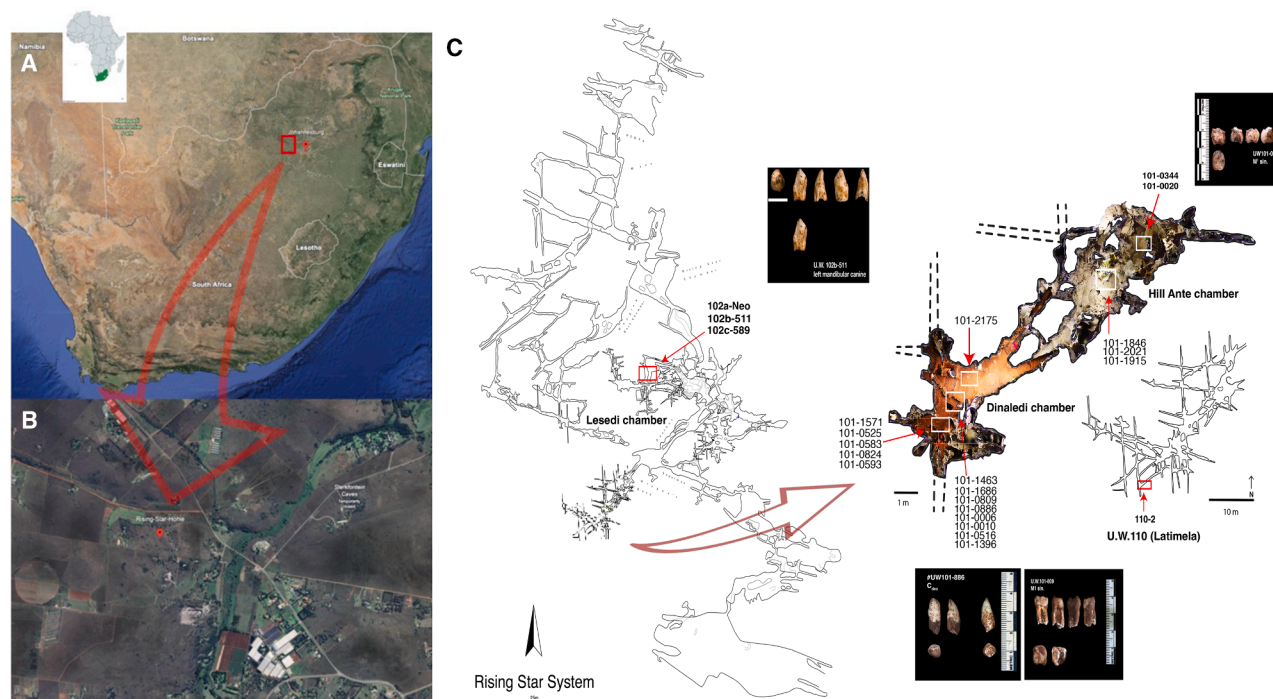


Figure 1. Location and layout of the Rising Star cave system

(A) Map of South Africa zoomed in (insert), showing the position of the area known as the Cradle of Humankind, approximately 50 km northwest of Johannesburg, where the Rising Star cave system is located.

(B) The Rising Star cave system within the Cradle of Humankind.

(C) Layout of the Rising Star System and the Dinaledi subsystem and Lesedi Chamber, where all the specimens were recovered, and the photos of the four *H. naledi* specimens initially micro-destructively sampled by acid etching, then sectioned for enamel growth analysis¹ and subsequently sampled destructively.

with a more *Homo*-like face, hands, and lower limbs.² Intriguingly, the collection is also remarkably homogeneous, with little skeletal size or morphological variation among the adult remains.^{2,3,7,8} A recent study of dental morphological variation also found that the homogeneity of *H. naledi* material may indicate a single sex, as evidenced by low size variation in teeth.⁹ However, the accuracy of morphology-based sex estimation can be greatly influenced by sex ratios and by the level of intersexual variability within species, factors that are difficult to establish for extinct taxa due in part to small sample sizes.¹⁰ The challenges of estimating the skeletal sex of ancient individuals, especially those in deep time, are further compounded by the degradation and incompleteness of skeletal remains.

Sex determination using ancient DNA (aDNA) has been successful for individuals with both high and low genomic coverage.^{11,12} However, aDNA preservation is limited by diagenesis, making this approach generally unfeasible in warmer climates across extended timescales.^{13–15} The only biomolecular solution currently viable to sex fossil specimens unamenable to aDNA analysis uses mass spectrometry (MS)-based paleoproteomic analysis of dental enamel and focuses on the detection of peptides unique to amelogenins X and Y (AMELX and AMELY).^{16–20} The sequences of human AMELX and AMELY differ for 23 diagnostic amino acid substitutions, plus a methionine (M) insertion in AMELY position 59 (Figure S1). AMELX is expressed in both male and female individuals, while ordinarily

AMELY is expressed in biological males,²¹ at a rate that is approximately 10% that of AMELX.²²

Enamel sampling for paleoproteomics requires either cutting off a fragment of the biomineralized material or, much less destructively, acid etching its surface.^{23,24} Proteomics has been successfully used not only to sex Holocene specimens,^{25–27} but also Pleistocene material,^{17,19,28} even by barely etching the enamel surface.^{18,20,29} While, to the best of the authors' knowledge, no ancient biomolecule analysis of *H. naledi* has been published to date, here we use paleoproteomics to investigate the sex of 23 *H. naledi* dental enamel specimens originating from a minimum of 20 different individuals (STAR Methods section "sample selection and minimum number of individuals"; Table S1; Figure S2). The *H. naledi* individuals studied here are from four different areas: Dinaledi and Lesedi Chambers, Hill Antechamber, and the area designated as U.W. 110 (Figure 1C), all within the Rising Star cave system, South Africa.^{2–4,30} To sample the broadest potential range of variation in *H. naledi*, the material examined in this study includes specimens of different physiological ages and from all published localities within the Rising Star cave system.

RESULTS

An initial round of analysis was performed on a set of four *H. naledi* specimens, U.W. 101-020, U.W. 102b-511, U.W.

101-809, and U.W. 101-886 (Figure 1C; Table S1), selected for destructive sampling in a previous work on the periodicity of enamel formation,¹ while maximizing the derived scientific output, as reported here, and in future studies. We conducted intra-crystalline chiral amino acid analysis to survey protein preservation. The extent of racemization of free and total hydrolyzable amino acids (FAA and THAA, respectively) indicates that the enamel from all four *H. naledi* specimens behaved as a closed system retaining endogenous peptides³¹ (Figure S3A). U.W. 102b-511, the potential juvenile individual,¹ exhibited higher D/L values in both the FAA and THAA fractions relative to the other three samples. The extent of protein degradation (assessed through AA composition, %FAA, and D to L conversion) in the *H. naledi* specimens is lower than in other extinct hominin tooth enamel samples where peptides have previously been recovered using MS (Figure S3B),^{18,19,28} supporting the feasibility of paleoproteomic analysis on these specimens. The lower level of amino acid degradation, compared with the published samples from the nearby fossil hominin site of Swartkrans, is consistent with their younger geological age.¹⁹

To test whether the minimally destructive etching methodology^{23,24} can be effectively used on *H. naledi*, the four teeth were first etched, then enamel samples weighing 28–48 mg were collected for paleoproteomic analysis based on the extraction protocol previously described.²⁸ From each sample, 10 mg were used for single-shot tandem MS analysis, and the remaining amount was analyzed via manual offline high-pH reverse-phase fractionation,¹⁹ in an attempt to improve the depth and breadth of sequence recovery.

The two different sampling approaches yielded comparable results, with a combined total number of 12,243 peptide spectrum matches (PSMs) obtained with conventional sampling, versus 18,709 PSMs obtained when using acid etching. We observed partial sequence recovery of 6–7 proteins per sample with destructive sampling and 5–11 proteins with acid etching (Figure S4). Notably, enamelin (ENAM), AMELX, and collagen 17 alpha 1 (COL17A1) were recovered in all specimens and replicates (Figure S4). The total number of amino acid positions confidently covered with conventional sampling ranged between 451 and 494, while with etching it ranged between 502 and 747. A total of 451 amino acid positions were shared across all specimens and replicates (Table 1). Fractionation yielded a minimal improvement (Figure S5), probably because it was performed using only 18–38 mg of starting material, considerably less than what was previously used to fractionate other hominin enamel extracts (>60 mg).¹⁹

AMELX was identified by multiple partially overlapping and nonoverlapping peptide sequences affected by extended terminal hydrolysis, confidently matched to multiple MS/MS spectra displaying almost complete b and y fragment ion series (Figure S6A). The identification of AMELY-specific peptides cannot be considered reliable beyond a reasonable doubt. Specimens from individual U.W. 102b-511, still developing at death,¹ showed no AMELY-specific evidence at all. Meanwhile, all the other three individuals (U.W.101-020, U.W.101-809, and U.W.101-886) displayed poor AMELY-specific PSMs with large gaps in their b and y fragment ion series, leaving the amino acids surrounding the AMELY-specific methionine residue in position

59 (M59; Figure S6B1–S6B3) uncovered, likely originating from a misidentification of those sequences by the search engine.

As no male markers were confidently identified from either the macro-destructive fragment or etching methods, and the etching method produced recoveries comparable to those obtained with conventional destructive sampling, we proceeded to etch the enamel of 19 more *H. naledi* teeth (STAR Methods section “peptide extraction”; Table S1; Figures S2 and S6C). No more than five specimens were processed at a time, with the intent to progressively screen all the available specimens. Only one sample (U.W.101-824) displayed a low-quality AMELY-specific PSM, with no confidently identified b and y fragment ions around the diagnostic M59 insertion (Figure S6B4).

The etching of the additional 19 *H. naledi* teeth (STAR Methods section “peptide extraction”; Table S1) resulted in 194 to 4774 PSMs, covering 153 to 695 total amino acid positions, depending on the sample (Table 1). Specimens U.W. 101-1915 and U.W. 101-1846 are potentially from the same individual (STAR Methods section “fossil samples and their provenance”), while the extracts from U.W. 101-1571 and U.W. 101-593, yielding very short sequence coverages of 153 and 322 amino acids, respectively (Table 1), were excluded from further analysis. The remaining 17 acid-etched specimens displayed an average amino acid sequence coverage equaling 523 positions. The combined dataset obtained from both conventional sampling and acid etching represented 20 *H. naledi* individuals and yielded a sum of 92,578 confidently identified PSMs from the MaxQuant software,³² while 116,264 and 145,447 PSMs were confidently identified using pFind³³ and Novor.cloud,³⁴ respectively. These PSMs partially covered between 5 and 11 proteins and an average of 577 amino acid positions (Figure S6C). A total of 441 amino acid positions were shared between all 20 *H. naledi* individuals, including partial protein sequences of AMELX, ENAM, and AMBN (Figure S6C).

All samples showed advanced levels of spontaneous amino acid modifications due to prolonged diagenesis, including deamidation of glutamine and asparagine, tryptophan conversion into kynurenine, and the conversion of arginine to ornithine (Figure S7). Additionally, as previously observed in other paleo-anthropological material,^{17–20,28,35} the observed peptide length distribution was skewed toward shorter chains, indicating advanced hydrolysis (also supported by the high %FAA detected) (Figure S8). Multiple lines of evidence, therefore, support the authenticity and antiquity of the recovered peptide sequences.

The analysis of all 20 *H. naledi* individuals did not yield any reliable evidence supporting the identification of AMELY, even though AMELX peptides were detected at intensities that are compatible with the concomitant detection of AMELY in specimens previously assigned to male *H. antecessor*¹⁷ and *P. robustus*¹⁹ (Figure 2). This indicates that the absence of AMELY-specific markers in the mass spectrometric data was not due to a lack of sensitivity of the experimental setup or to the inherent stochasticity of the data acquisition strategy.

We estimated the probability (Pr) of positively identifying each of these individuals as female based on the relative expression and preservation of AMELX and AMELY gene products (STAR Methods section “proteomic estimation of female biological

Table 1. An overview of the study samples and results

Sample ID U.W.	Type of extraction (fragment versus etch)	Total amino acids recovered	Excluded due to low coverage	Female confirmation based on site of AMELY59 intensity as a function of the site intensity of AMELX60	Predicted sex
101-020	both	908	no	yes	F
102b-511	both	672	no	yes	F
101-809	both	857	no	yes	F
101-886	both	796	no	yes	F
101-006	etch	463	no	yes	F
101-010	etch	467	no	yes	F
101-344	etch	486	no	yes	F
101-516	etch	587	no	yes	F
101-525	etch	466	no	yes	F
101-583	etch	460	no	yes	F
101-593	etch	153	yes	no	N/A
101-824	etch	695	no	yes	F
101-1396	etch	562	no	yes	F
101-1463	etch	441	no	yes	F
101-1571	etch	322	yes	no	N/A
101-1686	etch	565	no	yes	F
101-1846	etch	598	no	yes	F
101-1915	etch	560	no	yes	F
101-2021	etch	504	no	yes	F
101-2175	etch	467	no	no	F ^a
102a-Neo	etch	509	no	yes	F
102c-589	etch	527	no	yes	F
110-2	etch	528	no	yes	F

The total number of *H. naledi* samples studied is 23, two samples are potentially from the same individual (U.W.101-1846 and U.W.101-1915) (STAR Methods section “fossil samples and their provenance”), four individuals were etched and then cut, 19 teeth were only etched, two individuals have poor protein coverage, and lastly, for 3 individuals we could not confirm female based on AMELY59 as a function of AMELX60 site intensities (Figure 2).

^aConsistent with female identification.

sex”). MS datasets were analyzed using PEAKS³⁶ from all the *H. naledi* specimens, supplemented with 15 males from *H. sapiens*,²⁷ two males from *P. robustus*,¹⁹ a male *A. africanus*,¹⁸ a male Denisovan,²⁰ and a male *H. antecessor*¹⁷ for positive controls. This analysis was based on the following assumptions: (1) the ratio of AMELX to AMELY signal is comparable between *H. sapiens* and *H. naledi*, (2) the linear regression in logarithmically transformed data results in a normal and homoscedastic distribution of error, and finally, (3) the AMELY-specific sequences, otherwise consistent in Hominidae, are maintained in *H. naledi*. Of the 20 *H. naledi* individuals, 19 were confidently identified as females with a probability above 95% ($\text{Pr}(\text{Female} [F]) > 0.95$), while the signal from the remaining individual was consistent with a female attribution with a probability between 50% and 95% ($0.5 < \text{Pr}[F] < 0.95$) (Table S2). The results were in line with the analysis presented in Figure 2 and provided an additional confirmation of the absence of male specimens.

To assess inter-specific sequence variation, the assembled protein sequences from all the *H. naledi* individuals were aligned to their orthologs from extant and ancient hominin species. We observe a total of 18 confidently identified informative single

amino acid polymorphisms (SAPs) for Hominidae (Table S3). Two of these SAPs differentiate *H. naledi* from present-day humans, Neanderthals, and Denisovans. One of them, with a phenylalanine (F) in AMELX position 141, was detected in 15 *H. naledi* individuals and is also present in *Strepsirrhini* and *Cercopithecidae* (Figure 3A). This SAP is parsimoniously interpreted as a substitution of human AMELX 141 tyrosine (Y) (canonical transcript, Ensembl: ENST00000380712.7) to F141 in *H. naledi*, and not as modern human AMELY 142F (canonical transcript, Ensembl: ENSG00000099721), because we found no PSM supporting the identification of the human form of AMELX carrying the Y141. All the software tools we used to match MS/MS spectra consistently corroborated this observation (Data S1). The other SAP, featuring proline (P) in the position corresponding to 635 in the *H. sapiens* COL17A1 ortholog (canonical transcript, Ensembl: ENST00000648076), was detected in five *H. naledi* individuals by both MaxQuant and pFind, but not Novor.cloud (Data S1). P635 represents the ancestral allelic state displayed in extant non-human primates (Figure 3B). In contrast, *H. sapiens*, Neanderthals, and Denisovans carry the derived state alanine (A) 635. This site has not been covered so far in *H. antecessor*, *H. erectus*, and *A. africanus*,^{17,18,29} while the

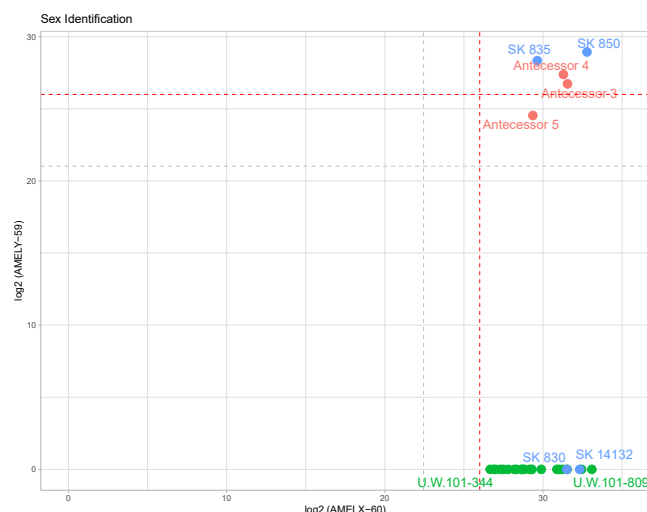


Figure 2. Site intensities of AMELY-59 as a function of the site intensities of AMELY-60

To calculate site intensities, the sum of all precursor intensities covering the amino acid position is used. The stochasticity limits are represented by red dashed lines, and below these lines, the acquisition method (i.e., data dependent acquisition [DDA]) is expected to be stochastic (i.e., <90% of the PSMs are confidently identified in two replicates). The peptide detection limits are represented by gray dashed lines, and they indicate the minimum intensity of a precursor covering the given amino acid position in a male individual for the studied data set. Precursors lying above the line can be detected in the MS. Plotted are *H. naledi* individuals (represented in green) and the previously studied *P. robustus*¹⁹ samples representing different individuals, two males and two probable females (blue), while the *H. antecessor*¹⁷ represents one unique male individual (red) over three different MS injections.

ancestral P635 state was also observed in two *P. robustus* individuals.¹⁹

The recovered partial protein sequences of the 20 *H. naledi* individuals show no variation within the sample set. To investigate the probability of failing to identify variants (clonality) in a sample set of 20 random individuals, we simulated the same sampling scheme, using modern human genomes for comparison, as no large-scale protein data exist for modern humans. We found that the frequency at which we observed no variants in the recovered protein regions after repeatedly sampling 20 individuals from the 1,000 Genomes Project dataset³⁷ was zero (Table S4). However, when performing the same sampling operation on panels from specific geographic areas, we observed that, for some panels with low genetic diversity at these recovered protein-coding regions, it is possible to sample no variation, but with a very low probability (Figure S9). In population panels where diversity at these recovered regions is high, such as populations from Africa, the probability of seeing no variation among 20 individuals was zero, even after resampling 1,000 times (Figure S10). Accordingly, the clonality of all the proteome sequences retrieved appears inconsistent with a sample of a large and diverse population. A less diverse population could support this result, but even then, the probability of sampling no variation in present-day human populations is low (1%–12%, depending on the population). Other explanations, such as prolonged isolation and/or inbreeding, affecting the *H. naledi* individuals

analyzed here, may also be consistent with these results. The genealogical relationship of the individuals is unknown, although the remains come from different locations in the cave system, separated by as much as 145 m, and were not deposited at a single time,^{3,5} making it unlikely that they are a single-family group.

Due to the limited length of the reconstructed partial protein sequences and the number of informative sites observed, the derived phylogenetic analysis of *H. naledi* should be considered tentative (STAR Methods section “phylogenetic analysis”; Figure S11). In the future, a more reliable phylogenetic reconstruction could be achieved by expanding the paleoproteomic analysis to *H. naledi* bone and dentine, or by aDNA sequencing, if biomolecular preservation allows.

DISCUSSION

Based on the results presented here and elsewhere,^{18,20,29} acid etching has been demonstrated to be a promising micro-destructive sampling approach to estimate the sex of Plio-Pleistocene dental enamel specimens, or to preliminarily screen protein recovery before proceeding with more destructive sampling strategies. Its use could make the palaeoproteomic investigation of particularly rare or precious specimens, so far incompatible with more conventional and invasive sampling practices, acceptable. While not systematically tested, enamel surface acid etching recovers slightly more PSMs than fragment cutting, in apparent contradiction with previous results.¹⁷ This could be plausible, as etching can retrieve the higher peptide diversity previously observed in maturation and erupted enamel,³⁸ by solubilizing peptides close to the enamel surface. Similarly, when the amount of the enamel fragment removed by cutting is progressively smaller, the recoveries of the two approaches become increasingly comparable, even reaching the point when etching recovery is slightly higher. Furthermore, enamel fragment demineralization exposes ancient peptides to acid and to its peptide terminal hydrolysis action for a minimum time of approximately 48 h. On the contrary, peptides retrieved using etching are suspended in acid for less than 1 h. This could lead to lower hydrolysis and slightly to the recovery of longer and potentially more informative peptide sequences. Nonetheless, the versatility and reliability of peptide recovery by acid etching still need to be systematically validated across space and time and on larger and more diverse sample sets exposed to a wide array of diagenetic factors.

Our results show no conclusive evidence supporting amelogenin heterozygosity, nor the reliable identification of any male, in a relatively large set of *H. naledi* individuals that also includes U.W. 102a (nicknamed “Neo”), so far considered a likely adult male due to its relative robusticity and cranial capacity.³ While paleoproteomics has successfully identified male hominin individuals from South Africa up to two million years old,^{18,19} indicating that such identification is methodologically possible, the evidence provided here proves beyond a reasonable doubt that the absence of any evidence supporting the identification of male individuals in our *H. naledi* sample set is not due to methodological limitations. Specifically, all the *H. naledi* AMELY signals we measure have an intensity compatible with the reliable observation of AMELY MS/MS spectra, if they were present. This excludes the

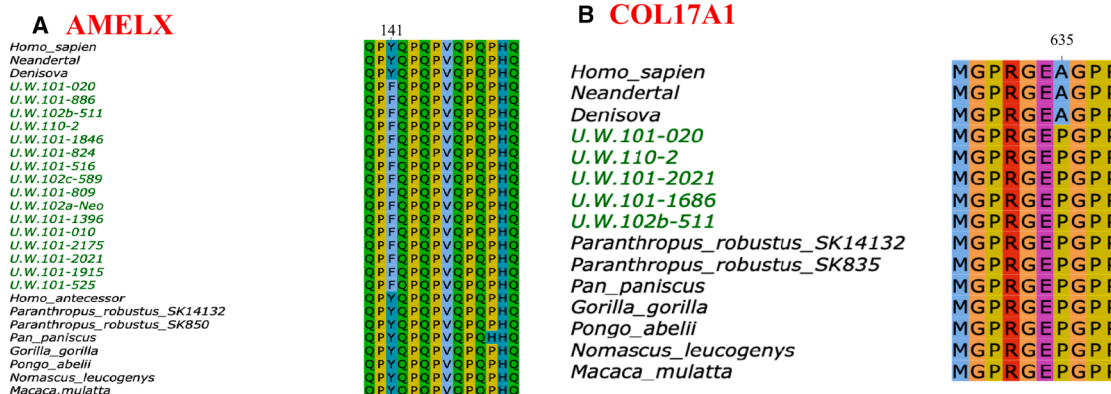


Figure 3. Recovered protein sequence alignment and present variation

(A) Alignment of AMELX sequences showing the variable site at position 141: *H. naledi* specimens carry an F while other hominins carry a Y in that position. (B) Sequence alignment for COL17A1 shows the variable site 635 with *H. naledi* specimens carrying a P, while *Homo sapiens*, Neanderthals, and Denisovans carry an A in that position.

possibility that the conventional AMELY-specific peptides generally used as markers to identify male individuals went undetected below the instrumental detection limits (Figure 2 and STAR Methods section “proteomic estimation of female biological sex”). Furthermore, our positive female probability analysis indicates that 19 out of the 20 individuals are confidently identified as females, while the proteomic profile of the remaining one is consistent with a female attribution (Table S2). All these considerations parsimoniously lead us to conclude that the *H. naledi* specimens we analyzed do not carry the conventional male-specific AMELY markers, either because they exclusively originate from female individuals, or because of a deletion or mutation present in all the male *H. naledi* individuals we may have possibly sampled. Either scenario would have very interesting implications for the biology and evolution of *H. naledi*. The homozygosity of AMEL supports the absence of AMELY, and the identification of substitution AMELX F141, rather than the observation of AMELY-specific position F142. Together with the documented low levels of morphological variation observed in *H. naledi*^{2,7} these two distinct lines of evidence independently suggest that the *H. naledi* individuals currently known from the Rising Star cave system are indeed biased⁹ toward—and may be exclusively—females.

While our results are compatible with the hypothesis that all the specimens analyzed belong to female individuals (Table 1), it is also possible, though unlikely, that our sample set includes male *H. naledi* individuals whose AMELY gene was mutated or deleted. Currently, paleoproteomics can only identify male individuals by detecting peptides covering publicly available AMELY-specific hominid sequences, or minor variants of these sequences using homology search, and by discovering previously unknown sequences by *de novo* sequencing. However, since the *H. naledi* AMELX consensus sequence showed high similarity to modern humans, it would be unlikely to observe AMELY sequences substantially divergent from modern humans or other living primates. If the entire AMELY gene is deleted in *H. naledi*, as previously observed in a Neanderthal individual,³⁹ then males of *H. naledi* would not be detectable with the

approach used here, as protein profiles of males would be indistinguishable from those of females. The deletion of the entire AMELY gene has been observed in extant human populations and has no known fitness consequences.⁴⁰ In modern humans, this mutation is population-dependent and very rare, occurring with a frequency ranging from 0.018% in an Austrian population⁴¹ to 8% in a Pakistani population.⁴² However, the rate of AMELY deletion reported in these studies should be interpreted cautiously, given sampling bias. When using large sets of high-quality genomes from the 1,000 Genomes Project and the Exome Sequencing Project,^{43,44} the deletion of AMELY could not be detected, leading us to conclude that the AMELY deletion is rare.⁴⁵

Given a sample set of 20 distinct individuals, and assuming a population sex ratio of 1:1, the odds of finding all female individuals are extremely low and represent 0.0000954%. However, the hypothesis of a single-sex accumulation of *H. naledi* in the Rising Star cave is supported by a recent study⁹ concluding that, based on dental sizes, the currently known *H. naledi* fossil record could consist of individuals with a sex ratio of 9:1 or even 1:0. Single-sex accumulation could be compatible with the possibility that *H. naledi* adopted cultural mortuary practices, as previously proposed.^{46,47} This reconstruction, however, has been challenged, and further evidence may test the hypothesis of the deliberate burial of the dead by *H. naledi*.⁴⁸ While accessing parts of the Rising Star cave system could be easier for smaller-bodied individuals, usually females among primates,⁴⁹ the areas of the cave system with *H. naledi* remains have varied entry conditions and are a significant constraint only to larger-bodied humans today. The sample includes 10 juvenile individuals at or prior to second molar eruption, an age range for which the body size of neither sex would exceed adult female body size.

Proteomics-based single-sex bias, like the one observed in *H. naledi* has not been documented in ancient and modern *Homo* populations. The closest known dominance of female human remains, observed at the Edcoul Cave, Montemor-o-Novo (Portugal), shows a 1:0.5 (F: M) ratio and is hypothesized to be due to socio-cultural factors, possibly related to funerary practices.⁵⁰ Similarly, female dominance in Iberian megalithic

societies is also observed from burial sites of Panoria populations, with the highest ratio being 1:0.42 (F:M),⁵¹ due to funerary practices based on matrilineal descent. However, these examples refer to Neolithic human populations, which may have operated very differently from *H. naledi* or from other mid-to-late Pleistocene *Homo* species. Some species of non-human primates have groups with a high adult female-to-male sex ratio, so that females and juveniles may predominate within a single social group. However, none have a similarly biased sex ratio for juvenile individuals, so this kind of social structure does not fit the *H. naledi* evidence.

Finally, the archaic COL17A1 P635 variant we observe in 5 *H. naledi* individuals is derived in Denisovans, Neanderthals, and modern humans, and it has not been covered in *H. erectus*, from either Dmanisi (Georgia) or China, and *Homo antecessor*.^{17,29} Interestingly, though, it has been observed in two *P. robustus* specimens, namely SK835 and SK14132. Determining which variant is present in this and other informative positions, in other extinct hominin species, such as *H. erectus* and *Australopithecus africanus*, will help clarify the phylogenetic position of *H. naledi*.

Previous paleoproteomic studies of Pleistocene enamel material have largely entailed destructive sampling of important and irreplaceable samples.^{17,19,28,35,52} This study, and other recent ones,^{18,20,29} show more sustainable, minimally destructive solutions to analyze Pleistocene material using paleoproteomics.^{18,20,24} We show that both more and less destructive methods can produce comparable results, and that etching could be used as a screening method to survey sexual dimorphism and test protein preservation. Thanks to its minimally destructive nature, etching could even be used to investigate the sex of rare and archaeologically valuable Plio-Pleistocene hominin fossils, such as type specimens, further advancing our understanding of morphological variation in ancient taxa.

Limitations of the study

The specimens sampled in this study, much larger in number than in any other previous paleoproteomic study on Pleistocene hominins, cover all the known *H. naledi* individuals excavated so far, except for those for which no teeth are preserved; we cannot exclude the presence of male individuals in the latter group. We assume that the *H. naledi* individuals had a typical karyotype and were not affected by syndromes that can result in non-binary expression, such as Turner or Klinefelter syndrome in humans.^{53–55} Our methodology cannot detect non-binary expression of biological sex, if present in the *H. naledi* population we studied. The approach we described here focuses exclusively on the characterization of the enamel proteome. While more destructive, the analysis of dentine and/or bone proteomes could most probably provide a richer and more informative phylogenetic signal.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Palesa P. Madupe (palesa.madupe@paleoproteomics.org).

Materials availability

This study did not generate new, unique reagents. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁵⁶ partner repository with the dataset identifier PXD054549.

Data and code availability

- For the purpose of open access, a Creative Commons Attribution (CC BY) license is applied to any author accepted manuscript version arising from this submission. Amino acid data from this study will be made available through the NOAA repository upon publication: <https://www.ncei.noaa.gov/pub/data/paleo/aar/>.
- The code for protein sequence reconstruction is available at <https://github.com/palesamadupe-palaeoprot/PalaeoProteinSequenceReconstruction>.
- The code for the sampling simulations is available at https://github.com/johnpatramanis/Hnld_Genetic_Diversity_Analysis.

ACKNOWLEDGMENTS

All the *Homo naledi* fossil specimens analyzed for this study are part of the permanent collection of Evolutionary Studies Institute at the University of the Witwatersrand, Johannesburg, South Africa. Permits for temporary export and sampling were issued by the South African Heritage Resources Agency (SAHRA permit IDs: 3405 and 3983). Access to all the specimens can be requested through the Evolutionary Studies Institute at the University of the Witwatersrand.

The work was funded by the European Union's Marie Skłodowska-Curie training network "PUSHH" under the Horizon 2020 research and innovation program, grant no. 861389 (P.P.M., C.K., I.P., P.K., F.M., J.C., K.P., J.V.O., and E.C.); the European Research Council (ERC) under the European Union's Horizon 2020, grant agreements no. 101021361 "BACKWARD" (P.P.M., I.P., A.J.T., M.M., G.T., L.S., J.V.O., R.R.A., and E.C.); the Novo Nordisk Foundation, grant agreement no. NNF14CC0001 (C.K. and J.V.O.); the National Research Foundation of South Africa, grant agreement no. AOP240509218040 (R.R.A.); the National Research Foundation of South Africa, grant agreement no. 136512 (R.R.A.); the Natural Environment Research Council (NERC), grant agreement no. NE/S010211/1 (M.R.D. and K.P.); and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada, grant agreement no. RGPIN-2020-04159 (L.S.). Funding for the recovery of the specimens and transport of the specimens, as well as certain aspects of morphological analysis, was funded by grants from the National Geographic Society and the Hill Exploration Project, as well as the LRB Foundation for Exploration.

AUTHOR CONTRIBUTIONS

Conceptualization, E.C., P.P.M., A.J.T., L.B., and J.H.; methodology, P.P.M., A.J.T., and E.C.; investigation, P.P.M., A.J.T., E.C., M.M., G.T., F.M., M.R.D., G.P., G.M., and P.M.; formal analysis, P.P.M., A.J.T., P.K., F.M., M.R.D., G.P., C.K., and I.P.; resources, B.Z., L.B., and J.H.; visualization, P.P.M., F.M., M.R.D., C.K., L.B., J.H., G.P., and I.P.; software, P.P.M., F.M., M.R.D., C.K., G.P., and I.P.; funding acquisition, E.C., J.V.O., L.B., and K.P.; project administration, E.C., L.B., and J.H.; supervision, E.C., R.R.A., J.V.O., J.C., and K.P.; validation, G.P. and P.K.; writing—original draft, P.P.M., A.J.T., E.C., J.H., I.P., and R.R.A.; writing—review & editing, E.C., P.P.M., A.J.T., L.B., J.H., R.R.A., L.S., M.M., M.R.D., C.K., I.P., B.Z., G.P., G.M., and P.M.

DECLARATION OF INTERESTS

G.J.P. has a patent on the use of amelogenin peptides for sex estimation (USPTO 8877455B2, AU2011229918B2, CA2794248, and EP11759843.3) and a provisional patent on the statistical framework in this study: Methods for statistical inference from proteomic data to estimate probability of a biological condition (US Prov. App. Ser. No. 63/964,307).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Fossil Samples and their Provenance
- METHOD DETAILS
 - Biomolecular preservation
 - Criteria of determination of minimum number of individuals (MNI)
 - Peptide extraction
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2026.05.044>.

Received: August 8, 2025

Revised: February 23, 2026

Accepted: May 29, 2026

Published: June 24, 2026

REFERENCES

1. Mahoney, P., McFarlane, G., Taurozzi, A.J., Madupe, P.P., O'Hara, M.C., Molopyane, K., Cappellini, E., Hawks, J., Skinner, M.M., and Berger, L. (2024). Human-like enamel growth in *Homo naledi*. *Am. J. Biol. Anthropol.* 184, e24893. <https://doi.org/10.1002/ajpa.24893>.
2. Berger, L.R., Hawks, J., de Ruiter, D.J., Churchill, S.E., Schmid, P., Delezene, L.K., Kivell, T.L., Garvin, H.M., Williams, S.A., DeSilva, J.M., et al. (2015). *Homo naledi*, a new species of the genus *Homo* from the Dinaledi Chamber, South Africa. *eLife* 4, e09560. <https://doi.org/10.7554/eLife.09560>.
3. Hawks, J., Elliott, M., Schmid, P., Churchill, S.E., Ruiter, D.J. de, Roberts, E.M., Hilbert-Wolf, H., Garvin, H.M., Williams, S.A., Delezene, L.K., et al. (2017). New fossil remains of *Homo naledi* from the Lesedi Chamber, South Africa. *eLife* 6, e24232. <https://doi.org/10.7554/eLife.24232>.
4. Elliott, M.C., Makhubela, T.V., Brophy, J.K., Churchill, S.E., Peixotto, B., Feuerriegel, E.M., Morris, H., Hunter, R., Tucker, S., Van Rooyen, D., et al. (2021). Expanded Explorations of the Dinaledi Subsystem, Rising Star Cave System, South Africa. <https://doi.org/10.48738/2021.ISS1.68>.
5. Dirks, P.H., Roberts, E.M., Hilbert-Wolf, H., Kramers, J.D., Hawks, J., Dosseto, A., Duval, M., Elliott, M., Evans, M., Grün, R., et al. (2017). The age of *Homo naledi* and associated sediments in the Rising Star Cave, South Africa. *eLife* 6, e24231. <https://doi.org/10.7554/eLife.24231>.
6. Robbins, J.L., Dirks, P.H.G.M., Roberts, E.M., Kramers, J.D., Makhubela, T.V., Hilbert-Wolf, H.L., Elliott, M., Wiersma, J.P., Placzek, C.J., Evans, M., et al. (2021). Providing context to the *Homo naledi* fossils: Constraints from flowstones on the age of sediment deposits in Rising Star Cave, South Africa. *Chem. Geol.* 567, 120108. <https://doi.org/10.1016/j.chemgeo.2021.120108>.
7. Garvin, H.M., Elliott, M.C., Delezene, L.K., Hawks, J., Churchill, S.E., Berger, L.R., and Holliday, T.W. (2017). Body size, brain size, and sexual dimorphism in *Homo naledi* from the Dinaledi Chamber. *J. Hum. Evol.* 111, 119–138. <https://doi.org/10.1016/j.jhevol.2017.06.010>.
8. Holloway, R.L., Hurst, S.D., Garvin, H.M., Schoenemann, P.T., Vanti, W.B., Berger, L.R., and Hawks, J. (2018). Endocast morphology of *Homo naledi* from the Dinaledi Chamber, South Africa. *Proc. Natl. Acad. Sci. USA* 115, 5738–5743. <https://doi.org/10.1073/pnas.1720842115>.
9. Delezene, L.K., Scott, J.E., Irish, J.D., Villaseñor, A., Skinner, M.M., Hawks, J., and Berger, L.R. (2024). Sex-biased sampling may influence *Homo naledi* tooth size variation. *J. Hum. Evol.* 187, 103490. <https://doi.org/10.1016/j.jhevol.2023.103490>.
10. Plavcan, J.M. (1994). Comparison of four simple methods for estimating sexual dimorphism in fossils. *Am. J. Phys. Anthropol.* 94, 465–476. <https://doi.org/10.1002/ajpa.1330940403>.
11. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* 40, 4477–4482. <https://doi.org/10.1016/j.jas.2013.07.004>.
12. Tierney, S.N., and Bird, J.M. (2015). Molecular sex identification of juvenile skeletal remains from an Irish medieval population using ancient DNA analysis. *J. Archaeol. Sci.* 62, 27–38. <https://doi.org/10.1016/j.jas.2015.06.016>.
13. Vincent, F., Hollund, H., de Voogt, A., Altena, E., and Fallet, C. (2013) Limits of Ancient dna Extraction from Teeth: the Case of Sudanese Nubia. *Nyame*.
14. Hedges, R.E.M. (2002). Bone diagenesis: an overview of processes. *Archaeometry* 44, 319–328. <https://doi.org/10.1111/1475-4754.00064>.
15. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709–715. <https://doi.org/10.1038/362709a0>.
16. Nielsen-Marsh, C.M., Stegemann, C., Hoffmann, R., Smith, T., Feeney, R., Toussaint, M., Harvati, K., Panagopoulou, E., Hublin, J.-J., and Richards, M.P. (2009). Extraction and sequencing of human and Neanderthal mature enamel proteins using MALDI-TOF/TOF MS. *J. Archaeol. Sci.* 36, 1758–1763. <https://doi.org/10.1016/j.jas.2009.04.004>.
17. Welker, F., Ramos-Madrugal, J., Gutenbrunner, P., Mackie, M., Tiwary, S., Rakownikow, Jersie-Christensen, R., Chiva, C., Dickinson, M.R., Kuhlwillm, M., de Manuel, M., et al. (2020). The dental proteome of *Homo antecessor*. *Nature* 580, 235–238. <https://doi.org/10.1038/s41586-020-2153-8>.
18. Madupe, P.P., Munir, F., Dickinson, M., Taurozzi, A.J., Mackie, M., Tawane, M., Mollereau, C., Hlazo, N., Penkman, K., Schroeder, L., et al. (2025). Results from an *Australopithecus africanus* dental enamel fragment confirm the potential of palaeoproteomics for South African Plio-Pleistocene fossil sites. *S. Afr. J. Sci.* 121. <https://doi.org/10.17159/sajs.2025/18571>.
19. Madupe, P.P., Koenig, C., Patramanis, I., Rütger, P.L., Hlazo, N., Mackie, M., Tawane, M., Krueger, J., Taurozzi, A.J., Troché, G., et al. (2025). Enamel proteins reveal biological sex and genetic variability in southern African *Paranthropus*. *Science* 388, 969–973. <https://doi.org/10.1126/science.adt9539>.
20. Tsutaya, T., Sawafuji, R., Taurozzi, A.J., Fagernäs, Z., Patramanis, I., Troché, G., Mackie, M., Gakuhari, T., Oota, H., Tsai, C.-H., et al. (2025). A male Denisovan mandible from Pleistocene Taiwan. *Science* 388, 176–180. <https://doi.org/10.1126/science.ads3888>.
21. Fincham, A.G., Bessem, C.C., Lau, E.C., Pavlova, Z., Shuler, C., Slavkin, H.C., and Snead, M.L. (1991). Human developing enamel proteins exhibit a sex-linked dimorphism. *Calcif. Tissue Int.* 48, 288–290. <https://doi.org/10.1007/BF02556382>.
22. Salido, E.C., Yen, P.H., Koprivnikar, K., Yu, L.C., and Shapiro, L.J. (1992). The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am. J. Hum. Genet.* 50, 303–316.
23. Porto, I.M., Laure, H.J., Tykot, R.H., de Sousa, F.B., Rosa, J.C., and Gerlach, R.F. (2011). Recovery and identification of mature enamel proteins in ancient teeth. *Eur. J. Oral Sci.* 119, 83–87. <https://doi.org/10.1111/j.1600-0722.2011.00885.x>.
24. Stewart, N.A., Molina, G.F., Mardegan Issa, J.P.M., Yates, N.A., Sosovicka, M., Vieira, A.R., Line, S.R.P., Montgomery, J., and Gerlach, R.F. (2016). The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction. *RSC Adv.* 6, 61673–61679. <https://doi.org/10.1039/C6RA05120K>.
25. Parker, G.J., Yip, J.M., Eerkens, J.W., Salemi, M., Durbin-Johnson, B., Kiesow, C., Haas, R., Buikstra, J.E., Klaus, H., Regan, L.A., et al. (2019). Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel. *J. Archaeol. Sci.* 101, 169–180. <https://doi.org/10.1016/j.jas.2018.08.011>.

26. Gron, K.J., Meiklejohn, C., Pedersen, K.B., Stewart, N.A., Alexandersen, V., Sørensen, L., and Montgomery, J. (2022). Sex and gender in the Mesolithic: Adults and children from the strøby egede burial, Køge Bugt, Denmark. *Proc. Prehist. Soc.* 88, 1–23. <https://doi.org/10.1017/ppr.2022.7>.
27. Koenig, C., Bortel, P., Paterson, R.S., Rendl, B., Madupe, P.P., Troché, G.B., Hermann, N.V., Martínez de Pinillos, M., Martínón-Torres, M., Mularczyk, S., et al. (2024). Automated high-throughput biological sex identification from archeological human dental enamel using targeted proteomics. *J. Proteome Res.* 23, 5107–5121. <https://doi.org/10.1021/acs.jproteome.4c00557>.
28. Cappellini, E., Welker, F., Pandolfi, L., Ramos-Madrigal, J., Samodova, D., Rütger, P.L., Fotakis, A.K., Lyon, D., Moreno-Mayar, J.V., Bukhsianidze, M., et al. (2019). Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature* 574, 103–107. <https://doi.org/10.1038/s41586-019-1555-y>.
29. Fu, Q., Wu, Z., Bennett, E.A., Xing, S., Ji, Q., Dong, Z., Rao, H., Gu, X., Dang, Y., Xing, J., et al. (2026). Enamel proteins from six *Homo erectus* specimens across China. *Nature*, 1–7. <https://doi.org/10.1038/s41586-026-10478-8>.
30. Brophy, J.K., Elliott, M.C., De Ruiter, D.J., Bolter, D.R., Churchill, S.E., Walker, C.S., Hawks, J., and Berger, L.R. (2021). Immature hominin craniodental remains from a new locality in the Rising Star Cave system, South Africa. *PaleoAnthropol.* <https://doi.org/10.48738/2021.ISS1.64>.
31. Penkman, K.E.H., Kaufman, D.S., Maddy, D., and Collins, M.J. (2008). Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. *Quat. Geochronol.* 3, 2–25. <https://doi.org/10.1016/j.quageo.2007.07.001>.
32. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. <https://doi.org/10.1038/nbt.1511>.
33. Li, D., Fu, Y., Sun, R., Ling, C.X., Wei, Y., Zhou, H., Zeng, R., Yang, Q., He, S., and Gao, W. (2005). pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics* 21, 3049–3050. <https://doi.org/10.1093/bioinformatics/bti439>.
34. Ma, B. (2015). Novor: real-time peptide de novo sequencing software. *J. Am. Soc. Mass Spectrom.* 26, 1885–1894. <https://doi.org/10.1007/s13361-015-1204-0>.
35. Kubat, J., Paterson, R., Patramanis, I., Barker, G., Demeter, F., Filoux, A., Kullmer, O., Mackie, M., Marques-Bonet, T., Huong, N.T.M., et al. (2023). Geometric morphometrics and paleoproteomics enlighten the paleodiversity of Pongo. *PLOS One* 18, e0291308. <https://doi.org/10.1371/journal.pone.0291308>.
36. Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., and Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 17, 2337–2342. <https://doi.org/10.1002/rcm.1196>.
37. Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* 48, D941–D947. <https://doi.org/10.1093/nar/gkz836>.
38. Green, D.R., Schulte, F., Lee, K.-H., Pugach, M.K., Hardt, M., and Bidlack, F.B. (2019). Mapping the Tooth Enamel Proteome and Amelogenin Phosphorylation Onto Mineralizing Porcine Tooth Crowns. *Front. Physiol.* 10, 925. <https://doi.org/10.3389/fphys.2019.00925>.
39. Skov, L., Peyrégne, S., Popli, D., Iasi, L.N.M., Deviese, T., Slon, V., Zavala, E.I., Hajdinjak, M., Sömer, A.P., Grote, S., et al. (2022). Genetic insights into the social organization of Neanderthals. *Nature* 610, 519–525. <https://doi.org/10.1038/s41586-022-05283-y>.
40. Butler, E., and Li, R. (2014). Genetic Markers for Sex Identification in Forensic DNA Analysis. *J. Forensic Investig.* 2, 10.
41. Steinlechner, M., Berger, B., Niederstätter, H., and Parson, W. (2002). Rare failures in the amelogenin sex test. *Int. J. Leg. Med.* 116, 117–120. <https://doi.org/10.1007/s00414-001-0264-9>.
42. Santos, F.R., Pandya, A., and Tyler-Smith, C. (1998). Reliability of DNA-based sex tests. *Nat. Genet.* 18, 103. <https://doi.org/10.1038/ng0298-103>.
43. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>.
44. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216–220. <https://doi.org/10.1038/nature11690>.
45. Parker, G.J., Buonasera, T., Yip, J.M., Eerkens, J.W., Salemi, M., Durbin-Johnson, B., Haas, R., Buikstra, J.E., Klaus, H., Rocke, D.M., et al. (2021). AMELY deletion is not detected in systematically sampled reference populations: A Reply to Štamfelj. *J. Archaeol. Sci.* 130, 105354. <https://doi.org/10.1016/j.jas.2021.105354>.
46. Berger, L.R., Makhubela, T.V., Molopyane, K., Krüger, A., Randolph-Quinney, P., Elliott, M., Peixotto, B., Fuentes, A., Tafforeau, P., Beyrand, V., et al. (2025). Evidence for deliberate burial of the dead by *Homo naledi*. *eLife* 12, RP89106. <https://doi.org/10.7554/eLife.89106>.
47. Agustín, F., Marc, K., Penny, S., Keneiloe, M., John, H., and Berger, L.R. (2023). Burials and engravings in a small-brained hominin, *Homo naledi*, from the late Pleistocene: contexts and evolutionary implications. *eLife* 12. <https://doi.org/10.7554/eLife.89125>.
48. Martínón-Torres, M., Garate, D., Herries, A.I.R., and Petraglia, M.D. (2023). No scientific evidence that *Homo naledi* buried their dead and produced rock art. *J. Hum. Evol.* 195, 103464. <https://doi.org/10.1016/j.jhevol.2023.103464>.
49. Lockwood, C.A., Menter, C.G., Moggi-Cecchi, J., and Keyser, A.W. (2007). Extended male growth in a fossil hominin species. *Science* 318, 1443–1446. <https://doi.org/10.1126/science.1149211>.
50. Granja, R., Araújo, A.C., Lugli, F., Silvestrini, S., Silva, A.M., and Gonçalves, D. (2023). Unbalanced sex-ratio in the Neolithic individuals from the Escoural Cave (Montemor-o-Novo, Portugal) revealed by peptide analysis. *Sci. Rep.* 13, 19902. <https://doi.org/10.1038/s41598-023-47037-4>.
51. Díaz-Zorita Bonilla, M., Aranda Jiménez, G., Sánchez Romero, M., Fregel, R., Rebay-Salisbury, K., Kanz, F., Vilchez Suárez, M., Robles Carrasco, S., Becerra Fuelle, P., Ordóñez, A.C., et al. (2024). Author Correction: Female sex bias in Iberian megalithic societies through bioarchaeology, aDNA and proteomics. *Sci. Rep.* 14, 27958. <https://doi.org/10.1038/s41598-024-78427-x>.
52. Welker, F., Ramos-Madrigal, J., Kuhlilm, M., Liao, W., Gutenbrunner, P., de Manuel, M., Samodova, D., Mackie, M., Allentoft, M.E., Bacon, A.-M., et al. (2019). Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. *Nature* 576, 262–265. <https://doi.org/10.1038/s41586-019-1728-8>.
53. Tallaksen, H.B.L., Johannsen, E.B., Just, J., Viuff, M.H., Gravholt, C.H., and Skakkebaek, A. (2023). The multi-omic landscape of sex chromosome abnormalities: current status and future directions. *Endocr. Connect.* 12, e230011. <https://doi.org/10.1530/EC-23-0011>.
54. Zinn, A.R., Page, D.C., and Fisher, E.M. (1993). Turner syndrome: the case of the missing sex chromosome. *Trends Genet.* 9, 90–93. [https://doi.org/10.1016/0168-9525\(93\)90230-F](https://doi.org/10.1016/0168-9525(93)90230-F).
55. Visootsak, J., and Graham, J.M., Jr. (2006). Klinefelter syndrome and other sex chromosomal aneuploidies. *Orphanet J. Rare Dis.* 1, 42. <https://doi.org/10.1186/1750-1172-1-42>.
56. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022:

- A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50, D543–D552. <https://doi.org/10.1093/nar/gkab1038>.
57. Skinner, M.F. (2019). Developmental stress in South African hominins: Comparison of recurrent enamel hypoplasias in *Australopithecus africanus* and *Homo naledi*. *S. Afr. J. Sci.* 115. <https://doi.org/10.17159/sajs.2019/5872>.
 58. Skinner, M.F., Delezene, L.K., Skinner, M.M., and Mahoney, P. (2024). Linear enamel hypoplasia in *Homo naledi* reappraised in light of new Retzius periodicities. *Am. J. Biol. Anthropol.* 184, e24927. <https://doi.org/10.1002/ajpa.24927>.
 59. Bolter, D.R., Hawks, J., Bogin, B., and Cameron, N. (2018). Palaeodemographics of individuals in Dinaledi Chamber using dental remains. *S. Afr. J. Sci.* 114, 6. <https://doi.org/10.17159/sajs.2018/20170066>.
 60. Delezene, L.K., Skinner, M.M., Bailey, S.E., Brophy, J.K., Elliott, M.C., Gurtov, A., Irish, J.D., Moggi-Cecchi, J., de Ruiter, D.J., Hawks, J., et al. (2023). Descriptive catalog of *Homo naledi* dental remains from the 2013 to 2015 excavations of the Dinaledi Chamber, site U.W. 101, within the Rising Star cave system, South Africa. *J. Hum. Evol.* 180, 103372. <https://doi.org/10.1016/j.jhevol.2023.103372>.
 61. Bolter, D.R., and Cameron, N. (2020). Utilizing auxology to understand ontogeny of extinct hominins: A case study on *Homo naledi*. *Am. J. Phys. Anthropol.* 173, 368–380. <https://doi.org/10.1002/ajpa.24088>.
 62. Dirks, P.H.G.M., Berger, L.R., Roberts, E.M., Kramers, J.D., Hawks, J., Randolph-Quinney, P.S., Elliott, M., Musiba, C.M., Churchill, S.E., de Ruiter, D.J., et al. (2015). Geological and taphonomic context for the new hominin species *Homo naledi* from the Dinaledi Chamber, South Africa. *eLife* 4, e09561. <https://doi.org/10.7554/eLife.09561>.
 63. Dickinson, M.R., Lister, A.M., and Penkman, K.E.H. (2019). A new method for enamel amino acid racemization dating: A closed system approach. *Quat. Geochronol.* 50, 29–46. <https://doi.org/10.1016/j.quageo.2018.11.005>.
 64. Kaufman, D.S., and Manley, W.F. (1998). A new procedure for determining dl amino acid ratios in fossils using reverse phase liquid chromatography. *Quat. Sci. Rev.* 17, 987–1000. [https://doi.org/10.1016/S0277-3791\(97\)00086-3](https://doi.org/10.1016/S0277-3791(97)00086-3).
 65. UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. <https://doi.org/10.1093/nar/gky1049>.
 66. Mackie, M., R  ther, P., Samodova, D., Di Gianvincenzo, F., Granzotto, C., Lyon, D., Pegg  , D.A., Howard, H., Harrison, L., Jensen, L.J., et al. (2018). Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angew. Chem. Int. Ed. Engl.* 57, 7369–7374. <https://doi.org/10.1002/anie.201713020>.
 67. Taurozzi, A.J., R  ther, P.L., Patramanis, I., Koenig, C., Sinclair Paterson, R., Madupe, P.P., Harking, F.S., Welker, F., Mackie, M., Ramos-Madr  gal, J., et al. (2024). Deep-time phylogenetic inference by paleoproteomic analysis of dental enamel. *Nat. Protoc.* 19, 2085–2116. <https://doi.org/10.1038/s41596-024-00975-3>.
 68. Dalgaard, P. (2010). *R: a Language and Environment for Statistical Computing* (R Development Core Team).
 69. Pag  s, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2017) Biostrings (Bioconductor). <https://doi.org/10.18129/B9.BIOC.BIOSTRINGS>.
 70. Katoh, K., and Standley, D.M. (2014). MAFFT: iterative refinement and additional methods. *Methods Mol. Biol.* 1079, 131–146. https://doi.org/10.1007/978-1-62703-646-7_8.
 71. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
 72. M  lder, F., Jablonski, K.P., Letcher, B., Hall, M.B., van Dyken, P.C., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Vieira, F.G., Meesters, C., et al. (2021). Sustainable data analysis with Snakemake. *F1000Res* 10, 33. <https://doi.org/10.12688/f1000research.29032.3>.
 73. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biol.* 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>.
 74. Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>.
 75. Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duch  ne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., K  hnert, D., De Maio, N., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* 15, e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>.
 76. Shapiro, S.S., and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>.
 77. Buonasera, T., Eerkens, J., de Flamingh, A., Engbring, L., Yip, J., Li, H., Haas, R., DiGiuseppe, D., Grant, D., Salemi, M., et al. (2020). A comparison of proteomic, genomic, and osteological methods of archaeological sex estimation. *Sci. Rep.* 10, 11897. <https://doi.org/10.1038/s41598-020-68550-w>.
 78. Keene, O.N. (1995). The log transformation is special. *Stat. Med.* 14, 811–819. <https://doi.org/10.1002/sim.4780140810>.
 79. St  mfelj, I. (2021). Sex estimation based on the analysis of enamel peptides: False assignments due to AMELY deletion. *J. Archaeol. Sci.* 130, 105345. <https://doi.org/10.1016/j.jas.2021.105345>.
 80. Rodr  guez Palomo, I., Nair, B., Chiang, Y., Dekker, J., Dartigues, B., Mackie, M., Evans, M., Macleod, R., Olsen, J.V., and Collins, M.J. (2024). Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins. *Peer Community J.* 4, e107. <https://doi.org/10.24072/pcjournal.491>.

STAR★METHODS

KEY RESOURCES TABLE

RESOURCE or REAGENT	SOURCE	IDENTIFIER
Biological samples		
<i>H. naledi</i>	This study	U.W.101-020
<i>H. naledi</i>	This study	U.W.102b-511
<i>H. naledi</i>	This study	U.W.101-809
<i>H. naledi</i>	This study	U.W.101-886
<i>H. naledi</i>	This study	U.W.101-006
<i>H. naledi</i>	This study	U.W.101-010
<i>H. naledi</i>	This study	U.W.101-344
<i>H. naledi</i>	This study	U.W.101-516
<i>H. naledi</i>	This study	U.W.101-525
<i>H. naledi</i>	This study	U.W.101-583
<i>H. naledi</i>	This study	U.W.101-593
<i>H. naledi</i>	This study	U.W.101-824
<i>H. naledi</i>	This study	U.W.101-1396
<i>H. naledi</i>	This study	U.W.101-1463
<i>H. naledi</i>	This study	U.W.101-1571
<i>H. naledi</i>	This study	U.W.101-1686
<i>H. naledi</i>	This study	U.W.101-1846
<i>H. naledi</i>	This study	U.W.101-1915
<i>H. naledi</i>	This study	U.W.101-2021
<i>H. naledi</i>	This study	U.W.101-2175
<i>H. naledi</i>	This study	U.W. 102a Neo
<i>H. naledi</i>	This study	U.W.102c-589
<i>H. naledi</i>	This study	U.W.110-2
<i>Paranthropus robustus</i>	Madupe et al. ¹⁹	SK 830
<i>Paranthropus robustus</i>	Madupe et al. ¹⁹	SK 835
<i>Paranthropus robustus</i>	Madupe et al. ¹⁹	SK 850
<i>Paranthropus robustus</i>	Madupe et al. ¹⁹	SK 14132
<i>Homo antecessor</i>	Welker et al. ¹⁷	<i>Homo antecessor</i>
<i>Australopithecus africanus</i>	Madupe et al. ¹⁸	Sts63
Human	Madupe et al. ¹⁹	Human
Chemicals, peptides, and recombinant proteins		
Molecular biology-grade water	ThermoFisher Scientific	Catalog number A57775
Trifluoroacetic acid (TFA)	ThermoFisher Scientific	Catalog number 85183
Acetonitrile	ThermoFisher Scientific	Catalog number 047138.K2
Methanol	ThermoFisher Scientific	Catalog number 022909.K2
Stagetips	ThermoFisher Scientific	Catalog No. 13-110-016
Ammonia	ThermoFisher Scientific	Catalog No. AC205840010
Formic acid	ThermoFisher Scientific	Catalog No. A1171-AMP
ReproSil C ₁₈ beads	Dr. Maisch, Germany	MZ0901-250046
Software and algorithms		
MaxQuant (V.1.6.3.4)	Cox et al. ³²	https://maxquant.org/download_asset/maxquant/latest

(Continued on next page)

Continued

RESOURCE or REAGENT	SOURCE	IDENTIFIER
pFind	Li et al. ³³	http://i.pfind.net/
Novor.cloud	Ma ³⁴	https://novor.cloud/
Deamidation	Mackie et al. ⁶⁶	https://github.com/dblyon/deamidation
Protein Sequence Reconstruction	This study	https://github.com/palesamadupe-palaeoprot/PalaeoProteinSequenceReconstruction
R	Dalgaard ⁶⁸	https://www.r-project.org/about.html
Biostrings	Pagès et al. ⁶⁹	https://bioconductor.org/packages/release/bioc/html/Biostrings.html
Mafft	Katoh and Standley ⁷⁰	https://mafft.cbrc.jp/alignment/software/
VCF tools	Danecek et al. ⁷¹	https://vcftools.github.io/index.html
Snakemake	Mölder et al. ⁷²	https://snakemake.readthedocs.io/en/stable/
Ensemble's Variant Effect Predictor	McLaren et al. ⁷³	https://www.ensembl.org/info/docs/tools/vep/index.html
Code for genetic sampling	This study	https://www.ensembl.org/info/docs/tools/vep/index.html
MrBayes	Huelsenbeck and Ronquist ⁷⁴	https://nbisweden.github.io/MrBayes/
BEAST	Bouckaert et al. ⁷⁵	https://www.beast2.org/
PEAKS ³⁶ XPro	Ma et al. ³⁶	https://www.bioinform.com
Excel	Microsoft Excel	https://www.microsoft.com/en-gb/microsoft-365/excel
Other		
Hominid enamel proteins	This study, Koenig et al. ²⁷ , Welker et al. ¹⁷ , Madupe et al. ¹⁹ , Madupe et al. ¹⁸ and Tsutaya et al. ²⁰	https://www.ebi.ac.uk/pride/ ; PXD054549, PXD049326, PXD014342, PXD040221, PXD054431, and PXD054412
1000 genomes	1000 genomes project ⁴³	https://www.internationalgenome.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Fossil Samples and their Provenance

All the specimens analyzed in this study had antimeres in order to limit the loss of morphological information (SAHRA permit ID: 3405 & 3983). The sampling strategy employed here was intended to have at least a representative sample from most sites where *H. naledi* is deposited within the Rising Star cave.

Sample selection and minimum number of individuals

At the initiation of this project, four *H. naledi* teeth were selected based upon several criteria to minimize destructive sampling of rare fossil material and to maximize the information return from several different methodologies. One key aim was preparation of thin sections of the tooth crown, in order to determine periodicity of enamel formation, directly observe histological characteristics of the crown, and for downstream analyses of developmental biology and environmental exposure. The resulting enamel periodicity data were published by Mahoney et al.¹ For these purposes, the priority for the four sampled teeth was to select relatively complete crowns with minimal dental wear, that clearly represent different individuals. Permanent canines and first molars were selected to maximize comparability with the data already gathered on hypoplasia and enamel defects^{57,58} and to extend the coverage of enamel development from before birth to the end of canine crown formation. Additionally, it was considered desirable to represent multiple *H. naledi* localities within the cave system, so in addition to two juvenile individuals from the Dinaledi Chamber, one individual each from the Hill Antechamber and the Lesedi Chamber were included in this initial sample. For the purposes of individuation, sample selection relied upon previous work by Bolter et al.⁵⁹ and Delezenne et al.,⁶⁰ as well as new observations of material that was not described in these previous studies.

U.W. 101–809 left mandibular first molar. This tooth is relatively complete and represents a late juvenile developmental stage. It was chosen because it is the likely antimeres of the right first molar in the U.W. 101–377 mandible, which has been attributed to the DH7 individual⁶¹ or dental individual 4.⁵⁹ This was included in Association 2 by Delezenne et al.⁶⁰

U.W. 101–886 right mandibular canine. This tooth is a complete crown with partially complete root, representing an early juvenile developmental stage. It is the antimeres of U.W. 101–1126, and is associated with a complete set of mandibular incisors, giving it the

possibility of representing an individual that is represented by several teeth within the sample. This individual must be different from DH7 represented by the U.W. 101–377 mandible, because that mandible includes an articulating right canine, U.W. 101–1014.

U.W. 101–020 left maxillary first molar. This tooth was the first specimen recovered from the Hill Antechamber. At the time this specimen was recovered, the Hill Antechamber and Dinaledi Chamber were not recognized as distinct depositional situations.^{2,62} Later work showed that these spaces are distinct and that passive transport from the Hill Antechamber into the Dinaledi Chamber is not feasible.^{4,46} U.W. 101–020 has slight occlusal wear and represents an individual of early juvenile developmental stage. It has been excluded as an antimeric of any RM² from the Dinaledi Chamber.

U.W. 102b–511 left mandibular canine. The 102b locality within the Lesedi Chamber has a mandibular fragment and five isolated teeth from a juvenile *H. naledi* individual.³ It is not clear whether 102b represents a different depositional space from the 102a area; however all other dental material in the Lesedi Chamber comes from adult individuals and this is the only dental sample from an individual of immature developmental stage.

The initial proteomic sampling of these four teeth demonstrated that no male-specific AMELY markers were present. Sampling four female individuals by chance within a mixed-sex sample is not, however, extremely unlikely. Because each of the four initial samples represent immature developmental stage individuals, and only one of these is clearly associated with additional mandibular and skeletal material, it is not possible to assess secondary sex characteristics from these fossils. Therefore, it was not possible to test whether the proteomic assessment of sex conflicts with any morphological data from these individuals. To test this possibility, the team decided to select an adult individual for which morphology had suggested male sex, based on body size, cranial size, and cranial morphological characteristics. Proteomic sampling of this individual and all additional specimens was carried out by the acid etching protocol and no cutting or drilling of material was performed.

LES1 right mandibular third molar. This individual from the Lesedi Chamber was assessed as the most probable male within the *H. naledi* sample based upon the greatest quantity of skeletal and cranial evidence.³

After sampling, the results from the LES1 tooth also showed no evidence of AMELY markers, suggesting female sex or a mutation in AMELY gene for this individual. At this stage of the research, the potential for reliable results from additional individuals using the minimally destructive approach was clear. The finding that the LES1 skeleton may represent a female individual suggested that a morphological reassessment of the entire sample was desirable, and that proteomic evidence of chromosomal sex would be essential to this reassessment.

The team adopted a strategy of identifying a series of teeth of *H. naledi* from all recovery areas that could be clearly established as different individuals. This survey consisted of multiple steps. The following list is presented in order of these steps, to make clear the evidence that the 23 teeth selected in the total sample represent a minimum number of 21 individuals.

First, every additional right maxillary first molar from the Dinaledi Chamber was included, with the exception of U.W. 101–999, which cannot be excluded as representing the same individual as the U.W. 101–886 mandibular canine. This added four distinct individuals to the sample set:

U.W. 101–1463 right maxillary first molar. This tooth from the Puzzle Box area of the Dinaledi Chamber is the antimeric of the left maxillary M1 that is in the U.W. 101–1277 maxilla. This specimen excavated from the Puzzle Box area of the Dinaledi Chamber is attributed to the DH1 individual, which represents an adult life history stage. On the basis of cranial size and some aspects of cranial morphology, this was suggested to be a male individual.^{2,7} This is dental individual 6 of Bolter et al.⁵⁹ and Association 7 of Deleuzene et al.⁶⁰

U.W. 101–525 right maxillary first molar. This tooth was recovered from surface area 4SE of the Dinaledi Chamber and represents an adult individual with occlusal wear facets but no significant dentine exposure. This tooth was included in Association 5 by Deleuzene et al.⁶⁰

U.W. 101–583 right maxillary first molar. This tooth was collected from surface area 4SH of the Dinaledi Chamber. With similar wear to U.W. 101–525, this is also inferred to be a young adult or adolescent individual.

U.W. 101–1396 right maxillary first molar. This tooth from the Puzzle Box area represents an adult life history stage. This tooth is part of Association 6 noted by Deleuzene et al.⁶⁰

From the Hill Antechamber, three additional right maxillary first molars have been recovered and were added to the sample set. None of these three teeth were compatible in morphology or dental wear stage to the U.W. 101–020 left maxillary first molar (discussed above), which represents an individual of early juvenile developmental stage. As noted above, the spatial distinctness of the Hill Antechamber supports the hypothesis that the individuals represented in this space are different from those in the Dinaledi Chamber; additionally, the following teeth duplicate right maxillary first molars represented in the Dinaledi Chamber.

U.W. 101–344 right maxillary first molar. This tooth was recovered from the surface of the Hill Antechamber at the base of the sloping floor. It was included in Berger et al.² prior to the recognition that the Hill Antechamber and Dinaledi Chamber are different depositional spaces. This tooth represents an adult individual.

U.W. 101–1846 right maxillary first molar. This tooth was recovered at the floor surface of the S150W150 excavation unit.² It represents an adult individual with substantial occlusal wear including the exposure of a dentine pool on the protocone, and a refitting interproximal facet with the U.W. 101–1845 right maxillary M².

U.W. 101–2021 right maxillary first molar. This tooth was recovered from sieving of sediments from the S100W100 excavation unit. It is an unworn first molar crown that does not match the morphology of U.W. 101–020 and cannot be an antimeric of this tooth.

The U.W. 110 locality is within the fissure network extending toward the west past the Dinaledi Chamber, and a fragmented cranium of an individual attributable to the early juvenile developmental stage was recovered from this locality, including a partial maxillary dentition. This dentition includes a germ of a right maxillary first molar that demonstrates that this partial cranium is a different individual from the early juveniles sampled from the Dinaledi Chamber. A right maxillary fourth premolar germ was selected for sampling:

U.W. 110–2 right maxillary fourth premolar. As noted above this represents an early juvenile that is distinct from all Dinaledi Chamber individuals.

From the Lesedi Chamber (designated U.W. 102), a mandibular fragment of one additional adult individual with two *in situ* teeth has been collected, which cannot represent LES1 or the early juvenile represented by U.W. 102b–511.

U.W. 102c–589 mandibular fragment with LM₁ and LM₂. This individual is an adult with moderate occlusal wear on M₂.

This survey brought the total number of distinct individuals to 14, using only the information from different localities and the duplication of right maxillary first molars as criteria. To add additional individuals, the survey then incorporated information on the developmental stage of teeth from the Dinaledi Chamber.

U.W. 101–1686 right mandibular deciduous second molar. This tooth is the antimere of the left mandibular deciduous second molar within the U.W. 101–1400 infant mandible, designated as dental individual 7 by Bolter et al.⁵⁹ and Association 1 by Deleuzene et al.⁶⁰ This individual represents the infant developmental stage. The individual also has a germ of a right maxillary first molar (U.W. 101–1688), but the rdm₂ was assessed to be a better candidate for enamel sampling by the acid etching approach.

Four early juvenile stage individuals are represented within the Dinaledi Chamber.⁵⁹ The early juvenile stage initiates at first molar eruption and ends at second molar eruption.⁵⁹ At this point in the survey, only one individual in this developmental stage was included, by the permanent mandibular left canine U.W. 101–886. Two of the remaining three are younger, but still within the same developmental stage, and are documented by left mandibular deciduous canines. The other individual is older within this stage and is represented by several teeth that duplicate the maxillary teeth proposed to represent the individual represented by U.W. 101–886.

U.W. 101–1571 left mandibular deciduous canine. This tooth was recovered from the surface of area 4SB of the Dinaledi Chamber. It exhibits substantial occlusal and distal wear and has a complete root. It is thus inconsistent with the development of the permanent U.W. 101–886 canine. It has the left mandibular deciduous canine rooted within the U.W. 101–1400 mandible, which itself has almost no occlusal wear.

U.W. 101–824 left mandibular deciduous canine. This tooth was recovered from area 4SC on the surface of the Dinaledi Chamber. It has a nearly complete root and substantial occlusal and distal wear. It duplicates U.W. 101–1571 and for similar reasons cannot represent the same individual as U.W. 101–886.

U.W. 101–593 right maxillary permanent second molar. This second molar germ is unworn and was recovered from the surface of area 4SJ in the Dinaledi Chamber. It is proposed as the antimere of U.W. 101–505.⁶⁰ U.W. 101–593 duplicates U.W. 101–1006 and is very near the same extent of crown development. U.W. 101–1006 is proposed as the same individual as the U.W. 101–886 permanent canine, making U.W. 101–593 (collected in a different location) a second individual of the same approximate age. Both these teeth exhibit greater crown development than would be expected for the individuals represented by the U.W. 101–1571 or U.W. 101–824 deciduous canines.

Three additional partial mandibles can be identified as adults within the Dinaledi Chamber sample, each of which exhibits greater wear than the individuals represented by the DH1 mandible or the U.W. 101–1142 mandibles, which are both represented in the sample above. Two of these three remaining adult individuals are represented in the sample by additional teeth, while one is only represented by the mandibular fragment with one tooth *in situ*. These partial mandibles each are assessed to represent different individuals than the right maxillary first molars U.W. 101–525 or U.W. 101–583, which were included above, on the basis of inconsistent occlusal wear. This is particularly evident for U.W. 101–516 and U.W. 101–006, both associated with first molars that are extremely worn. This mismatch in occlusal wear stage is also inferred for U.W. 101–010, for which the P₃ and C wear is much higher than would be consistent with the much less worn U.W. 101–525 and U.W. 101–583.

U.W. 101–516 left mandibular third molar. This tooth was recovered from the Puzzle Box area and is the antimere of the right mandibular third molar within the U.W. 101–001 mandible. U.W. 101–516 exhibits occlusal wear to the point of loss of most occlusal detail with dentine exposure on the cusps.

U.W. 101–006 right maxillary third molar. This tooth was found on the surface, in the area of the Puzzle Box. Only an enamel rim with a large excavated dentine pool remains of this crown, and the mesial portion of the rim is completely worn away. This is the oldest individual within the Dinaledi Chamber sample, inferred to be the same individual as the U.W. 101–361 mandible attributed to DH3.² This is dental individual 3 of Bolter et al.⁵⁹

U.W. 101–010 mandibular fragment with right mandibular canine and right mandibular third premolar. This mandibular fragment was found on the surface in area 1 of the Dinaledi Chamber. It is an adult individual with obliterating occlusal wear on the remaining premolar. It cannot be the same individual as any other partial mandible within the sample and exhibits substantially greater occlusal wear than would be consistent with U.W. 101–525 or U.W. 101–583.

The fourth additional mandible is represented within the area designated as Dinaledi Feature 1⁴⁶.

U.W. 101–2175 left mandibular fourth premolar. This tooth from Dinaledi Feature 1 belongs to a mandibular dentition that duplicates all mandibles from the Dinaledi Chamber and is therefore a different individual from these. This dentition exhibits substantial

occlusal wear on the premolars and first mandibular molar, making it an older adult individual than represented by the U.W. 101–525 or U.W. 101–583 maxillary first molars.

Including all of the above teeth yields a minimum number of 22 individuals from the combined sample. These 22 individuals are differentiated in several ways including recovery location, with four distinct localities represented, and within each locality by duplication or clear evidence of different developmental stage.

One additional tooth was also sampled. This tooth belongs to an individual that is already represented within the sample listed above. The result from this tooth proved to be consistent with the other tooth included from the same individual.

U.W. 101–1915 right maxillary third molar. This tooth was excavated from the S150W150 unit at a depth of 6 cm. It represents an adult individual with moderate occlusal wear. A mesial interproximal facet on this tooth is a good fit to the distal interproximal facet on the U.W. 101–1845 right maxillary M², suggesting that this tooth and U.W. 101–1846 are the same individual.

METHOD DETAILS

Biomolecular preservation

Chiral amino acid analysis was undertaken on the intra-crystalline fraction of U.W. 101–020, U.W. 101–809, U.W. 101–886 and U.W. 102b–511. The enamel specimens were processed following Dickinson et al.⁶³ After powdering and bleaching, the specimens were divided into two fractions: one for determining the free amino acids (FAA) and one for the total hydrolysable amino acids (THAA). The concentration of the intra-crystalline amino acids and their extent of racemization (D/L value) was then quantified using RP-HPLC (Agilent 1100 series HPLC fitted with Hypersil C18 base deactivated silica column (5 µm, 250 x 3 mm) and fluorescence detector) following a modified method of Kaufman and Manley.⁶⁴ D and L ratios were measured for the following amino acids: aspartic acid and asparagine (Asx); glutamic acid and glutamine (Glx); serine (Ser), alanine (Ala), valine (Val), phenylalanine (Phe), isoleucine (Ile), leucine (Leu), threonine (Thr), arginine (Arg) and tyrosine (Tyr) to provide the estimation of intra-crystalline protein degradation.

Criteria of determination of minimum number of individuals (MNI)

To determine MNI, we followed the method described in Bolter et al.⁵⁹ (detailed in [STAR methods](#) section “fossil samples and their providence”). In summary, teeth checked for duplication based on type, side, wear and eruption, and age class of the teeth. As a result, we determined that MNI is 22 because U.W. 101–1846 and U.W. 101–1915 could feasibly be from the same individual.

Peptide extraction

Destructive analysis

U.W. 101–020, U.W. 101–809, U.W. 101–886 and U.W. 102b–511 were sampled via the destructive method. The sample surface was first cleaned using molecular biology-grade water. Between 10 mg and 38 mg of dental enamel was used for single shot and fractionation experiments. An enamel sample was removed from the specimen using a motorized cutting disk (Dremel), carefully avoiding sampling the dentine. The enamel subsample was partially fragmented by crushing it with a pestle and mortar before the subsample was demineralized using 500 mL (single shot) or 1.5 mL (fractionation) 10% v/v trifluoroacetic acid (TFA) for 48 h at 4°C. Any remaining sample debris was pelleted by centrifugation at 10,000 x g for 15 mins and the supernatant was transferred to a fresh Eppendorf Protein loBind tube. During this time C₁₈ StageTips were prepared as described in Cappellini et al.²⁸ The StageTips were loaded (1 sample per StageTip), cleaned and eluted as described in Cappellini et al.²⁸ An extraction blank per batch was prepared simultaneously with samples.

Etching

All *H. naledi* specimens were analyzed using the micro-destructive method. This protocol was modified from Stewart et al.²⁴ Briefly, the sample surface was first cleaned using molecular biology-grade water and low dust laboratory tissue (Kimtech). 130 mL of 10% v/v TFA was placed into the cap of a 0.5 mL Eppendorf protein lo-bind tube and the sample was manually held so the surface of the sample was in contact with the acid solution for an initial 15 sec. The tube and acidic solution were discarded, a fresh one was prepared, and the sample surface that was in contact with the acid was cleaned once more, as described above. The sample was re-introduced to the new 10% TFA in the new tube, and contact was maintained for a total of 10–15 minutes with visual inspection every 3–5 minutes. These times were sample-dependent, and the aim was to avoid over-sampling and penetration of the enamel, which would result in exposing dentine to the etching solution. The acidic solution (sample extract) was removed from the 0.5 mL tube cap and placed into a fresh Protein loBind Eppendorf tube and the cap was washed with 100 mL 10% TFA and combined with the sample extract. C₁₈ StageTip peptide concentration/cleanup was performed as described for the samples analyzed by destructive extraction above, including extraction blanks.

Peptide clean-up and fractionation

Single shot analysis was performed on all the acid etched samples and the four samples from which 10 mg of enamel was sampled for destructive analysis. For the single shot analysis, the peptides were eluted from StageTips using 30 µL of 40% acetonitrile (ACN), 60% water + 0.1% TFA, the ACN was evaporated, and the peptides were re-suspended prior to injection. The obtained extracts were concentrated to around 1 µL and reconstituted in a 10 µL solution containing 0.1% TFA, 5% ACN in water.

For the fractionated samples (18–38 mg destructive extraction), the pH of the extract was adjusted to 10 by addition of 30% Ammonium hydroxide solution (NH₃OH). Following this, the sample was loaded onto a StageTip and washed with 200 µL of 20 mM NH₃ in

water. The elution phases were conducted as outlined below: 1- 96% 20 mM NH₃ in H₂O / 4% 20 mM NH₃ in ACN, 2- 92% 20 mM NH₃ in H₂O / 8% 20 mM NH₃ in ACN, 3- 88% 20 mM NH₃ in H₂O / 12% 20 mM NH₃ in ACN, 4- 80% 20 mM NH₃ in H₂O / 20% 20 mM NH₃ in ACN, and 5- 20% 20 mM NH₃ in H₂O / 80% 20 mM NH₃ in ACN. The resultant extracts were evaporated to about 1 μ L and reconstituted in a 10 μ L mixture of 0.1% TFA, 5% ACN in water.

LC-MSMS

Peptides were separated using nanoflow liquid chromatography over 77 minutes. The mass spectrometer was an Exploris 480 (Thermo Fisher Scientific, Waltham, MA, USA) operating data dependent acquisition (DDA) in top 10 mode.

The liquid chromatography system employed was the EasyLC 1200 system from Thermo Fisher Scientific (Waltham, MA, USA), using mobile phases A (5% ACN, 0.1% formic acid in water) and B (80% ACN, 0.1% formic acid in water) with a flow rate of 250 nL/min. The loading of 4 μ L occurred at a constant pressure of 500 bars. Then, two linear sections constituted the gradient 5%–30% B in 50 minutes and 30%–45% B in 10 minutes. This was followed by a 5-minute wash at 80% B and a 5-minute re-equilibration at 5% B. The separation was executed at 40°C. For the separation, an in-house pulled and packed column was used. The packing particles were ReproSil C₁₈ beads (Dr. Maisch, Germany) with a diameter of 1.9 μ m and 120 Å pores in a silica tube measuring 20 cm in length and 0.75 mm in inner diameter, equipped with an integrated electrospray ionization (ESI) tip.

The Exploris 480 mass spectrometer was employed in positive ionization mode with spray voltage at 2 kV, and ion transfer tube temperature at 275°C. In MS1, a scan was conducted across the m/z range of 350 to 1400 at an Orbitrap resolution of 120,000 at m/z 200. The subsequent MS2 experiment operated at a resolution of 60,000 within the m/z range of 100 to precursor m/z +20, utilizing 60 V collision energy at 1e–5 bar. Ions underwent fragmentation for the top 10 ions that met the criteria of intensity exceeding 2e4, a charge state between 2 and 6, and the exclusion of isotopologues. Dynamic exclusion of 20 seconds was employed.

Database search strategy

Raw data was searched using MaxQuant (V.1.6.3.4),³² pFind,³³ and novor.cloud.³⁴ A custom-built database, containing 343 entries of enamel specific proteins collected from UniProt⁶⁵ corresponding to *Homo sapiens*, *Pan*, *Pongo*, *Nomascus*, *Papio* and *Microcebus*, was used. This database was supplemented with the *Gigantopithecus*, *Homo antecessor*, and Denisovan enamel proteomes and some Neanderthal enamel specific sequences.^{17,52} Peptide identification was performed by setting the digestion parameter to “unspecific” since aspecific digestion occurred *in vivo* and through hydrolysis, and the minimum length for unspecific peptides was set to 7 amino acids. In the main search, the peptide mass tolerance was 4.5 ppm, and the fragment mass tolerance was 20 ppm. For MaxQuant, the Andromeda threshold score for both unmodified and modified proteins was set to 40, to filter out peptide spectral matches (PSM) with a low-quality score and for both pFind and Novor.cloud we left the threshold default score. No fixed post translational modifications were set. Glutamine and asparagine deamidation, oxidation of methionine, oxidation of proline, phosphorylation of serine/threonine/tyrosine, ornithine conversion from asparagine, and N-terminal pyroglutamic acid from glutamic and aspartic acids were all included as possible variable modifications. Proteins included in the database of common contaminants provided by MaxQuant, pFind and Novor.cloud, for example proteinaceous laboratory reagents and human skin keratins, as well as reverse sequences were manually removed and not considered any further. Similarly, proteins detected in the laboratory blank were not further considered.

Post-translational modifications analysis

The extent of deamidation was estimated using the python script from Mackie et al.⁶⁶ Other modifications were assessed by PSM counting. For a given modification, the ratio between the number of amino acids that could be modified to the number of amino acids modified was calculated. The counts were normalized by the PMS count.

Criteria for positively identifying and validating informative sites

The methodology is based on and adapted from Taurozzi et al.,⁶⁷ briefly a combination of these; high scoring PSMs, a complete or almost complete y and b fragment ion series, multiple PSMs and multiple overlapping peptide sequences.

To validate potential PSMs matched to amelogenin-Y peptides, we employed a manual validation approach rather than relying exclusively on algorithmic scoring methods, which are primarily designed for modern biological samples with larger datasets. Thus, spectra used for protein sequence reconstruction were manually inspected to evaluate fragment ion coverage, with particular attention to the methionine (M) residue in SMIRPPY peptides. Additionally, experimental spectra were compared with predicted spectra (<https://www.proteomicsdb.org/use/>) to assess the validity of peptide identifications.

To determine whether PSMs with phenylalanine (F) at position 141 for AMELX (present in Strepsirrhini and Cercopithecidae) or at position 142 for human AmelY (canonical Ensembl transcript ENSG00000099721) were present, multiple lines of evidence were evaluated. Peptides corresponding to the human AMELX 141 tyrosine (Y) variant were not detected at this position. Noteworthy as reported by Salido et al.,²² AmelY expression is approximately 10% that of AMELX, making it improbable that AMELX variant would be detected in the absence of the corresponding AMELX peptides. In addition, a greater number of PSMs matched the amelogenin-X protein sequence compared with amelogenin-Y.

PSMs matching amelogenin-Y were excluded if the supporting evidence was ambiguous: (1) A PSM matching AmelY in specimen U.W.101-1463 contained a peptide with potential AmelY-109 arginine (R) and AmelY-110 valine (V), but this finding was supported by only a single PSM with no overlapping peptides, and the RV site was not resolved by a complete or almost complete y and b fragment ion series. (2) Specimens U.W.101-006, U.W.101-1686, U.W.101-824, and U.W.102b-511 exhibited PSMs matching AmelY at position 167 leucine (L); however, no overlapping peptides were identified, and this peptide also matched a sequence from a soil-dwelling bacterium (see BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi?alnHdr_WP_431964074) (Figure S6 D).

Protein sequence reconstruction

Once the contaminants had been removed, the manually validated sequences were reconstructed with an in-house R⁶⁸ script which can be found in this GitHub repository (<https://github.com/palesamadupe-palaeoprot/PalaeoProteinSequenceReconstruction>). First, the tab-separated text file "evidence.txt" from the MaxQuant³² is read into a data frame. Next, the script filters and organizes the data for each of the specified genes (ALB, AMELX, AMELY, AMBN, AHSG, AMTN, COL17A1, COL1A1, COL1A2, ENAM, ODAM, MMP20). For each gene, the script filters rows that match the gene name in the "Proteins" column. It then creates a new data frame by selecting and combining the MS/MS scan number, Experiment, Score, and Sequence columns into a single column named "id" using an underscore as a separator. After filtering and organizing the data, the script writes the organized data to a tab-separated text file for each gene (e.g., "alb_peptides.txt"). Each file contains the combined "id" column and peptide sequences. The script then generates FASTA files using the package Biostrings⁶⁹ for the peptide sequences for each gene. Finally, the script aligns the peptides from each FASTA file against reference sequences using Mafft⁷⁰ suite. From the aligned peptides sequences, we obtained the consensus sequence for each gene for all the *H. naledi* specimens.

Sampling simulations

The full dataset of the 1000 genomes⁴³ was obtained in VCF file format.⁷¹ The dataset was subdivided into population specific vcf files, based on the population label of each individual. A copy of the complete, global population with all individuals was also maintained. All the vcf files were processed with a custom workflow based on snakemake.⁷² Sites that were recovered in all 20 highest coverage, *H. naledi* individuals were identified, resulting in a subset of 509 amino acids. The vcf file of each population was then reduced to only variants falling within the range of the coding regions corresponding to these 509 amino acids. The vcf files were further filtered to remove silent mutations and keep only variants that cause an amino acid change, based on Ensemble's Variant Effect Predictor.⁷³ Each filtered vcf file was sampled 1000 times, each time isolating 20 individuals and enumerating the number of variants present in the sample set along with calculating other genetic diversity metrics. The total number of times no variants were identified was divided by the 1000 repetitions, generating the probability of the observed *H. naledi* sampling for each population (Figure S9). All code for this analysis is available on GitHub (https://github.com/johnpatramanis/Hnld_Genetic_Diversity_Analysis) but requires the user to download and subdivide the 1000 Genomes vcf files.

Phylogenetic analysis

Sequences were divided into individual protein datasets, which were aligned using Mafft⁷⁰ and then corrected for leucine and isoleucine isobaric positions as described in¹⁷. The individual protein datasets were concatenated and converted to each of the appropriate formats for phylogenetic analysis. Multiple phylogenetic trees were generated for the concatenation of all the available proteins, using different software as demonstrated in Madupe et al.¹⁹ For the phylogenetic reconstruction using MrBayes,⁷⁴ we followed the detailed instructions in Madupe et al.¹⁹ and those in Welker et al.¹⁷ for Beast2⁷⁵ tip-dated analysis. A set of two individuals per subspecies (or one for the consensus tree) was randomly selected from the concatenated 'Diversity' dataset. For the Tip dating reconstruction, the ages of each ancient individual were included in this analysis and are as follows: Altai Neanderthal – 112 ka, Chagyrskaya Neanderthal – 80 ka, Altai Denisovan – 82 ka, four *P. robustus* individuals – 1.9 Ma,¹⁹ a *H. antecessor* 900 ka individual¹⁷ and 21 of the *H. naledi* individuals.

Statistical estimation of female biological sex

Male sex estimation is dependent on unambiguous detection of amelogenin Y-chromosome isoform specific peptide sequences (AMELY). Absence of AMELY however can be due to two hypotheses: female sex, or failed detection of AMELY. The probability of false negative male detection will decrease as data quality increases. Female sex therefore can be expressed as a probability. This supplement outlines a statistical framework for the probabilistic estimation of female sex in *H. naledi*. To mitigate non-parametric and heteroscedasticity effects, this framework was developed fully in logarithmic space.

The framework addressed three factors that were needed to accommodate individual sex estimates into a probabilistic framework. Firstly, the ratio of AMELX and AMELY needed to be estimated to allow AMELY values to be modelled as a function of AMELX signal. Secondly, the lower limit of quantitation (LLOQ) needed to be estimated so that the ratios could be predictable. Finally, the measurements of both the X/Y ratio and LLOQ needed to be distributed normally and be parametric, which was achieved when the measurements were transformed logarithmically. When the mean log(LLOQ) estimate was added to the mean log(X/Y) estimate the combined value described the AMELX level (logX) at which there was a 50% cumulative probability that the AMELY would be detected in the quantitative range ($\leq \log 50\%Y$). Since the deviation of both log(X/Y) and log(LLOQ) was normally distributed, the number of standard deviations between log50%Y and an individual logX value (Z-score) could then be used to give a cumulative probability of female sex (Pr(F)).

The lower limit of quantitation (logLLOQ) was estimated in 35 of the 75 *H. naledi* datasets to be 5.190 ± 0.467 . To estimate the ratio of AMELX to AMELY in male enamel a total of 15 known archaeological and paleo enamel samples were processed from 5 hominin species. This included archaeological *H. sapiens* from Jordan ($n = 10$),²⁷ a *H. antecessor* ($n = 1$)¹⁷; *P. robustus* ($n = 2$),¹⁹ *A. africanus*¹⁸ ($n = 1$), and Denisovan ($n = 1$)²⁰ samples. Fourteen individuals had detectable levels for both proteins (logX, logY) with a mean ratio (logX/Y) of 1.04 ± 0.24 . The logX value with a 50% probability of AMELY detection in the quantitative range was 6.23 ± 0.52 . Individual estimates of both logLLOQ and log(X/Y) did not show significant departures from normality.⁷⁶ Because of this Z-scores between individual logX measurements and the log50%Y could be used to accurately obtain cumulative probabilities of female sex (Pr(F)) in the absence of AMELY detection.

A total of 75 mass spectrometry datasets were analyzed in PEAKS³⁶ from 23 *H. naledi* specimens and of these 68 datasets contained detectable AMELX sequence. Twenty were definitively female ($\text{Pr}(\text{F}) > 0.95$), two were consistent with female ($\text{Pr}(\text{F}) > 0.5, < 0.95$), and one was indeterminate ($\text{Pr}(\text{F}) < 0.5$) (Table S2). Aside from the *Homo sapiens* enamel sample, no unambiguous AMELY – specific peptides were detected in the sample. This analysis assumed that the ratio of AMELX to AMELY signal in *Homo sapiens* was maintained in *H. naledi*, that linear regression in logarithmically transformed data results in a normal and homoscedastic distribution of error, and that amelogenin-Y specific sequences consistent in Hominidae, were maintained in *H. naledi*.

The framework made both biological and statistical assumptions. Biologically, it assumed the expression levels of AMELX and AMELY in *Homo sapiens* apply to *Homo naledi* and that the respective sequences in *Homo naledi* were the same as in other hominins. It assumed no structural deletion of the AMELY locus. The biological assumptions were reasonable or validated elsewhere in the study. Statistically, the study assumed that the AMELX values could be used to model AMELY levels, a reasonable assumption given that both were above the LLOQ level and the relative ratio of AMELX to AMELY signal was well measured. The study also assumed that any Z-scores between individual cumulative AMELX signal values and the 50%-AMELY threshold could be used for cumulative probability estimation. The assumption was likewise validated by the Shapiro Wilk tests for normal distribution of the logarithmically transformed LLOQ value and AMELX/AMELY ratio.

Mass spectrometry datasets (RAW format) were processed using PEAKS³⁶ XPro (Bioinform.com, PEAKS Studio 10.6 build 20201221), in Database mode (PEAKS-DB) with default settings except: Parent Mass Error Tolerance: 10.0 ppm, Fragment Mass Error Tolerance: 0.02 Da, Variable modifications: Oxidation (M): 15.99, deamidation (NQ): 0.98, Phosphorylation (STY): 79.97, Oxidation or Hydroxylation: 15.99. A maximum of 3 PTMS per peptide. Digestion in Unspecific mode. The reference database was a validated custom FASTA file of primate enamelome proteins, 'Hominidae_IP_17Oct22SMIRPPY_noMice.txt', available on the ProteomeXchange data repository (PXD054549). A decoy database was used and a 5% FDR cut-off applied to all peptide spectra matches. Resulting peptide data from the '>pan_troglodytes_AMELX-201' gene product, identical to the N-terminal region of *Homo sapiens* '> NP_001133.1 amelogenin, X isoform 1 precursor [*Homo sapiens*]' was copied and peptides containing the AMELX-specific "SI" motif (amino acid positions 44 and 45) were identified and collated. Matching peptide area (ion current) values were transformed logarithmically (base 10) and ranked to create a ranked peptide curve. Linear regression was applied to the middle 50% of ranked transformed peptide values, and residuals calculated for all AMELX-specific peptides. The point of negative deviation away from the regression was identified, where possible, and the rank was used obtain an estimate for LLOQ for the dataset ($\log_{10}$ -LLOQ). All transformed estimates of $\log_{10}$ (LLOQ) were used to calculate a mean and standard deviation. A W-statistic was also calculated following Shapiro and Wilk,⁷⁶ and compared with the Shapiro Wilk reference value to confirm that the W-statistic was greater and the hypothesis of non-normal distribution could be rejected (significance level = 0.05). The ratio of cumulative AMELX-specific peptide signals to AMELY-specific peptide signals was calculated from the datasets of all known hominin male enamel samples and logarithmically transformed (base 10). The $\log_{10}$ (AMELY) values ($\log Y$) was subtracted from the $\log_{10}$ (AMELX) values ($\log X$) for each male dataset to generate a mean and standard deviation (sd). The original $\log_{10}$ (AMELX) and the ' $\log X - \log Y$ ' values were tested for non-normal distribution.⁷⁶ The means for $\log$ LLOQ and ' $\log X - \log Y$ ' values were summed to estimate a value of $\log X$ where there was a 50% probability of AMELY being detected at the $\log$ LLOQ (50% AMELY threshold, $\log 50\% Y$) if the sample was male. The standard deviation of this threshold was calculated as the square root of the sum of the squares of the matching standard deviations (50% AMELY threshold sd, $\sigma_{50\% Y}$). The number of standard deviations (Z-score) for each dataset was calculated as the difference between individually observed $\log X$ values and the $\log 50\% Y$ threshold, divided by the standard deviation for the threshold ($\sigma_{50\% Y}$). The cumulative probabilities for the standard normal (Z) distribution, probability of female biological sex ($\text{Pr}(\text{F})$), was calculated in Excel using the "NORM.S.DIST("Z-score", True)" function. This study defines definitive female biological sex (F) as $\text{Pr}(\text{F}) > 0.95$, contingent 'consistent with female biological sex' (F*) as ($0.5 < \text{Pr}(\text{F}) < 0.95$), indeterminate sex (IND) as $0 \leq \text{Pr}(\text{F}) \leq 0.5$. Definitive male sex was defined as unambiguous detection of any AMELY – specific peptide regardless of signal. Please note: All simple log notations are in base 10.

The population of peptides specific for AMELX gene product (from the >pan_troglodytes_AMELX-201 gene product) ranged from 1 to 130 peptides in the *H. naledi* sample. The signal of any one peptide is the product of concentration and ionization efficiency. When the ion current for each peptide was logarithmically transformed (base 10) and ranked the resulting plot (Ranked Peptide Plot, Figure S12) had a distinctive profile with three segments. At the upper end of the range, the population of peptides was a consistent subset of high signal peptides, and the ranking resulted in a curve with a sharp decline over a limited number of peptides. The major and middle portion of the curve appeared as linear, and the third segment negatively deviated away from linearity. The distribution of peptides in the linear portion of logarithmic curve reflects a distribution of peptides with a mix of ionization properties and concentrations. As the recovery of peptides decreased, the concentration and signal of this population dropped into the zone where detection of individual peptides began to fail.

This study defined the lower limit of quantitation (LLOQ) as the point where signal began to fail due to non-detection (Figure S12, red marker) and fall away from the linear portion of the curve. This analysis assumed that the rate of detection failure would increase the closer the peptide signal approached the LLOD. Sampling the LLOD from peptide detection was therefore problematic, since sampling was small and stochastic effects were maximal. Sampling of LLOQ was more consistent and represented a portion of the peptide rank curve where modelling would likewise be consistent. A fixed reference point for modeling that could be estimated from each dataset. To assist in mapping the point of departure due to low concentration, the estimated LLOQ, a linear regression was calculated using the linear middle 50% of the curve (Figure S12A). The point of departure of observed peptide signals from the linear regression was sharper and easier to estimate when judging from the residual plot (Figure S12B). An overlay of ranked peptide plots

illustrated that AMELX peptides detected between the LLOQ and LLOD, also occurred in the linear range when concentration was increased, and ionization efficiency should not change appreciably (Figure S12C). The total estimated logLLOQ levels occurred across 2 orders of magnitude and have normal distribution (Figure S12D). LogLLOQ levels did not change at lower peptide recoveries (peptX).

The logLLOQ mean and standard deviation of 5.190 ± 0.467 was obtained from 35 of the 80 datasets where a clear point of deviation away from linearity could be discerned. The Shapiro-Wilk test did not show a significant departure from normality, $W(35) = .96$, $p = .297$.

These estimates were measured based on the sub-population of peptides that were specific for AMELX, since this was the sub-population that was informative for sex estimation. LLOQ however should be instrumental and so other peptide populations in the dataset, or all peptides, may possibly be used and theoretically be more precise since a higher density of ranked peptide signals should provide greater resolution of LLOQ, although precision of LLOQ with peptide number was not supported by Figure S12D. LLOQ estimation was also contingent on detection of at least one peptide being observed between the LLOQ and LLOD. LLOQ estimation had intrinsic errors and biases. Each estimate was based off the first peptide above the LLOQ point, an intrinsic conservative bias that theoretically would be minimized with more peptide observations, although, again, this was not supported by Figure S12D. When regressions were estimated, and the variation was symmetric and normal, 50% of the residuals will also be below the regression line. If these occurred adjacent to the LLOQ, it was difficult to resolve negative residuals from expected homoscedastic variation around a regression from residuals that result from failed peptide detection.

The underlying assumption of this framework is that the pattern of amelogenin, X isoform 1 (AMELX) specific peptides, can be used to model, or predict, the pattern of amelogenin, Y-isoform (AMELY) specific peptides. This was based on homology at the peptide level and the assumption that the ratio of expression of AMELX (NP_001133.1) and AMELY (NP_001134.1) was consistent within hominin individuals and populations, an assumption validated by these data. The ratio of AMELY to AMELX in *Homo sapiens* has been reported to be at a level of 10 to 1 at the transcript^{21,22} and 12 to 1, or 13 to 1 at the protein^{25,77} level. Linear regression of non-transformed values however has a non-normal distribution and therefore cannot be used to obtain a Z-score that could be used for accurate estimation of cumulative probability. This problem was addressed by conducting the analysis in logarithmic space.⁷⁸

The signal of AMELX- and AMELY-specific peptides was measured in enamel from 15 male hominin individuals (archaeological *H. sapiens* from Jordan, $n = 10^{27}$; *H. antecessor*, $n = 1^{17}$; *P. robustus*, $n = 2^{19}$; *A. africanus*, $n = 1^{18}$; and Denisovan, $n = 1^{20}$) AMELX-specific peptides were detected in at least one datafile from all individuals, with matching AMELY detected in 14 individuals, the exception being one *P. robustus* individual (SK835). Logarithmically transformed values (logX, logY) were calculated for each *Homo sapiens* dataset and the highest matching signal reported for each individual (Table S5) and plotted (Figure S13). The logarithm of the AMELX/AMELY ratio transforms to logX-logY. The mean and standard deviation (sd) were calculated at 1.040 ± 0.283 . The Shapiro-Wilk test did not show a significant departure from normality, $W(14) = .93$, $p = .281$.⁷⁶

Delineation of both LLOQ and the AMELX/AMELY ratio allowed AMELY (logY) levels to be modelled as a function of AMELX peptide signal and to do so in a way that conformed to a probabilistic framework. The mean of log(LLOQ) was added to the mean of the log(AMELX/AMELY) (log(X/Y)), to create a threshold value of log(AMELX) (logX) where AMELY would be predicted to be at the LLOQ. Because this value is a mean with a normal distribution, the cumulative probability that AMELY would be at the LLOQ would be 50%. The combined means therefore were defined as the '50% AMELY threshold' (50%Y). Mathematically this can be simplified as:

$$\bar{X}_{\log_{10}(50\%Y)} = \bar{X}_{\log_{10}(LLOQ)} + \bar{X}_{\log_{10}(X/Y)}$$

Or

$$\text{Log}_{10}(50\%Y) = 5.19 + 1.04 = 6.23$$

When described in exponential space, when the cumulative ion current (area) for all AMELX-specific peptides was 1.7×10^6 ($10^{6.23}$) the probability that the AMELY signal would be at the LLOQ ($10^{5.19} = 142,000$) would be 50%. If we assume the individual was male, the mutually exclusive lack of AMELY detection at the 50%Y threshold would represent a male false negative and this would also have a 50% probability:

$$\text{Pr}(\text{male false negative}) + \text{Pr}(\text{male true positive}) = 1.$$

Since AMELY detection failure at the LLOQ could be due to the two hypotheses: that of a male false negative and female sex (i.e. female true positive) and since these were also mutually exclusive, they could also be described as

$$\text{Pr}(\text{male false negative}) + \text{Pr}(\text{female true positive}) = 1.$$

Male and female true positive detection was contingent on the appearance or otherwise of AMELY detection. As the AMELX signal increases above the 50% AMELY threshold the probability of a male false negative would decrease, and the alternate probabilities

Pr(male true positive) and Pr(female true positive) would increase equally with male or female sex assignment contingent on AMELY detection.

The equivalency of Pr(M) and Pr(F) assumed that the Pr(F) can be modelled on the probability that AMELY would be detected above the LLOQ point for an observed AMELX value in a male individual dataset. The same probability would apply to the probability of female sex (Pr(F)) in the absence of AMELY detection.

At the practical level the definition of male was a simple binary detection of AMELY in an individual's enamel. This difference in sex assignment criteria, that potentially may occur below the LLOQ, would be expected to impose a bias toward male sex estimation. This model also does not accommodate the sub-population of higher signal peptides in the leading part of the ranked peptide curve. This population would potentially persist above the LLOQ even as overall concentration of all peptides decreases in low and lower signal datasets. The calculated Pr(F) values are therefore an underestimate and could be considered overly conservative.

As the quality of enamelome data increases, as measured by increased logX values, the probability of a male false positive would decrease, and female true positive would increase, in the absence of any AMELY signal. Since both the log(LLOQ) and log(X/Y) values had normal distributions, the log(50%Y) threshold also would have a normal distribution. Because of this, any difference between an observed logX value and the log(50%Y) threshold, specifically the number of standard deviations between these values (Z-score), can be used to estimate cumulative probability. The standard deviation of the log(50%Y) value would equal the square root of the sum of the squared standard deviations of logLLOQ and log(X/Y).

$$sd_{\log(50\%Y)} = \sqrt{sd_{\log(LLOQ)}^2 + sd_{\log(X/Y)}^2}$$

The estimated log₁₀(LLOQ) was 5.190 ± 0.467 (mean±SD) and the logarithm (base 10) of the ratio of AMELX over AMELY was 1.040 ± 0.243 (mean±SD). The standard deviation of log(50%Y) was therefore:

$$sd_{\log(50\%Y)} = \sqrt{0.467^2 + 0.243^2} = 0.521.$$

The log₁₀(50%Y) = 6.23 ± 0.521 (mean ± sd).

The Z-score or number of standard deviations between log(50%Y) and individual logX values was calculated as:

$$Z = (\log X - \log_{10}(50\%Y)) / sd_{50\%Y}$$

Definitive female sex (F) (Table S6) was defined as Pr(F) > 0.95, consistent with female sex (F*) defined as (0.5 < Pr(F) < 0.95), indeterminate (IND) as Pr(F) < 0.5, or no detection. Any unambiguous AMELY peptide was defined as definitively male (M). Of 75 *H. naledi* datafiles from 23 individuals, Z-scores and cumulative probabilities (Pr(F)) were estimated. Of 75 datafiles, 53 resulted in definitive female sex estimation (F), 10 with contingent female sex estimation (F*), and 12 indeterminate datafiles (IND). Of the 23 *H. naledi* individuals, 20 had at least one datafile with definitive female sex estimates, 2 individuals had contingent female sex estimates (U.W. 101–1571 and U.W. 101–2175) with Pr(F) values of 0.92 and 0.90 respectively, and a single indeterminate sample (U.W. 101–593). There was no unambiguous AMELY-specific peptides detected in any *H. naledi* sample.

A visual representation of *Homo naledi* female sex probability distribution is plotted below on a log-log plot, along with bioarchaeological male hominin individuals used for estimation of the AMELX/AMELY ratio. These were overlaid with logarithmically transformed LLOQ and 50% AMELY threshold values (±sd) (Figure S14A). A density plot of AMELX values (logX) for each dataset (Figure S14B) and probability of female sex values (Pr(F)) (Figure S14C) were also plotted.

This framework makes several biological and statistical assumptions. Biochemically it assumed that the relative expression and preservation of AMELX and AMELY gene products was consistent between other hominin species and *Homo naledi*. It also assumed that the sequence for AMELY in both species was stable and that there was no population-wide structural variation, AMELY deletion or recombination to the AMELX sequence, in the Y-chromosome of *H. naledi*.^{45,79} These assumptions were considered reasonable given the stability and detection of these sequences, as well as a consistent AMELX to AMELY ratio (Figure S14) in *H. sapiens* subpopulations and related hominin species.^{17–20,27}

The other assumption was the choice of peptide populations to estimate LLOQ. Only peptides specific to the AMELX gene product were used since these would best match the properties of AMELY-specific peptides and were considered necessary for this framework, given that AMELX was used to predict AMELY. However, LLOQ is primarily a property of the instrument, and a higher density of peptides would maximize sampling of log peptide values around the LLOQ, potentially increasing accuracy and reducing variability. It is possible that using all amelogenin peptides, not just those specific for either of the paralogous genes, or even all peptides present in the dataset would improve LLOQ accuracy and reduce variability. Also, this study confined data analysis to peptides identified in the Database search mode of PEAKS XPro. Potential use of the subsequent PEAKS³⁶ PTM-mode of peptide spectral matching roughly doubled the number of matches and increases overall ion current collection by about 10% to 25%. This study conducted its analysis using peptides matched using a custom reference proteome based on the limited enamel proteome with orthologous proteins across a range of primate species. The use of the custom, highly concise reference proteome increased the peptide signal 3.7-fold (p=0), the LLOQ was effectively unchanged. This was consistent with the noted increase in sensitivity when using smaller customized reference proteomes.⁸⁰

The AMELX-specific peptides with the highest signals also deviate from the linear regression based off the middle 50% of the peptide rank curve. A thorough analysis of whether the disproportionate signals of this population are due to higher concentration, higher

ionization efficiency, or a mixture of both has not been established. The effect of this peptide population on detected signal in the zone between the LLOQ and LLOD is also not analyzed, and more sophisticated models of AMELY prediction promise to give more precise and accurate cumulative probability estimates.

Finally, this framework assumes a normal distribution of measurements that were used to calculate estimates of logLLOQ and log(X/Y). Use of the Z-score to estimate cumulative probability of female sex assumes normal distribution, and replication of this framework on other datasets and samples should include normality tests such as the Shapiro Wilk test.

QUANTIFICATION AND STATISTICAL ANALYSIS

- Software. PEAKS 35 XPro for mass spectrometry processing; Excel for cumulative normal probabilities; The Shapiro-Wilk test was conducted using an online calculator (www.statskingdom.com/shapiro-wilk-test-calculator.html).
- Analyses: log10 transformation; ranking of peptide signals; linear regression on the middle 50% of ranked log-transformed peptide values; residual analysis to identify LLOQ; calculation of mean $\pm$ Standard Deviation (SD) for logLLOQ and log(X/Y); Shapiro-Wilk normality tests; combination of means to form log50%Y and propagation of SD by root-sum-of-squares; Z-score calculation and conversion to cumulative probability (Pr(F)) was obtained using NORM.S.DIST.
- Number of individuals: *H. sapiens* from Jordan, $n = 10^{27}$; *H. antecessor*, $n = 1^{17}$; *P. robustus*, $n = 2^{19}$; *A. africanus*, $n = 1^{18}$; and Denisovan, $n = 1^{20}$ and *H. naledi*, $n=22$.
- Means and SDs were 5.190 ± 0.467 for logLLOQ; 1.040 ± 0.283 for log(X/Y).
- Tested the distributions with the Shapiro-Wilk test (W-statistic) and report W and p -values to assess normality for logLLOQ and log(X/Y).
 - Referenced in [STAR methods](#): Proteomic Estimation of Female Biological Sex, [Figures S12–S14](#) and summarized in [Tables S5–S6](#).