Extended Data Table 2 | Proteome composition and coverage

Specimen	Protein Name	Sequence length	Razor and unique peptides	Matched spectra*	Coverage after MaxQuant searches (%)	Final coverage after MaxQuant and PEAKS searches (%)	Final coverage (aa)
16628	Collagen alpha-1(I)	1158	5	8	3.2	3.2	37
16629	Amelogenin X	209	79	190	36.8	36.8	77
	Ameloblastin	440	51	84	25.0	25.0	110
	Enamelin	1129	58	133	6.2	6.5	73
	Collagen alpha-1(I)	1453	3	3	2.0	2.0	29
	Collagen alpha-1(III)	1464	2	3	1.4	1.4	20
	Amelotin	212	2	2	4.7	4.7	10
16630	Enamelin	1129	180 | 3	530 | 5	11.8 | 2.7	15.4	174
	Ameloblastin	440	105	231	30.9	31.4	138
	Amelogenin X	213	116	529	62.0	62.9	134
	Amelogenin Y	192	4	9	13.0	22.9	44
	Amelotin	212	5	6	8.0	8.0	17
16631	Enamelin	916	175	751	11.0	11.7	107
	Amelogenin X	213	156	598	48.8	61.5	131
	Amelogenin Y	90	5	18	15.6	25.6	23
	Ameloblastin	440	71	133	24.1	25.2	111
	MMP20	482	2	2	3.9	3.9	19
16632	Enamelin	1144	401	2160	17.9	19.1	219
	Amelogenin X	192	280	960	84.4	84.4	162
	MMP20	424	49	67	33.3	33.3	141
	Serum albumin	607	11	18	6.1	6.1	37
	Collagen alpha-1(I)	1513	4	4	2.6	2.6	40
16634	Amelogenin X	185	68	157	53.5	53.5	99
	Ameloblastin	440	47	58	23.4	23.4	103
	Enamelin	920	33	87	4.5	4.5	41
	MMP20	483	4	4	5.6	5.6	27
16635	Amelogenin X	206	394 | 3	2793 | 5	73.8 | 7.8	85.9	177
	Enamelin	1150	382 |2	2966 | 2	18.3 | 1.6	25.1	289
	Ameloblastin	442	131	463	31.3	39.3	166
	Amelotin	267	26	148	9.9	9.9	20
	Serum albumin	607	34	64	18.5	24.5	149
	MMP20	483	15	25	11.8	15.3	74
16637	Collagen alpha-1(I)	1453	2	2	1.7	1.7	25
	Collagen alpha-1(II)	1421	2	2	1.9	1.9	27
	Collagen alpha-1(III)	1464	2	2	1.6	1.6	23
16638	Enamelin	1129	235 | 7	1155 | 13	11.8 | 4.7	12.9	146
	Amelogenin X	192	185 | 3	734 | 5	52.0 | 10.9	60.4	116
	Ameloblastin	440	64 | 2	120 | 4	30.0 | 5.7	36.4	160
	MMP20	481	6	7	8.1	9.1	44
16639	Enamelin	1129	202	726	12.0	12.6	142
	Amelogenin X	213	167	624	59.2	67.6	144
	Ameloblastin	440	88	155	26.8	30.5	134
	Amelogenin Y	192	13	13	18.8	18.8	36
16641	Amelogenin X	213	91	251	64.3	65.3	139
	Ameloblastin	440	69	122	28.9	28.9	127
	Enamelin	1129	24	75	7.8	7.8	88
	Amelotin	212	3	3	7.1	7.1	15
16642	Amelogenin X	185	89	245	42.7	42.7	79
	Enamelin	733	14	19	2.5	2.5	18
	Ameloblastin	421	3	3	7.1	7.1	30
	MMP20	483	2	2	3.5	3.5	17
16856	Amelogenin X	209	66 | 4	365 | 25	38.8	45.5	95
	Enamelin	916	58 | 13	153 | 70	8.2	10.2	93
	Ameloblastin	440	21	31	14.8	14.8	65
	Collagen alpha-1(I)	1047	8 | 10	9 | 11	14.5	16.9	177
	Collagen alpha-2(I)	1054	4 | 8	5 | 9	10.6	10.6	112
	Serum albumin	583	0 | 8	0 | 12	16.6	16.6	97
	Amelogenin Y	90	3	7	10.0	10.0	9
16857	Collagen alpha-1(I)	1047	18 | 14	24 | 18	21.7	23.4	245
	Collagen alpha-2(I)	1274	16 | 11	17 | 11	17.7	24.3	310
16860	Amelogenin X	192	46	98	30.7	32.3	62
	Ameloblastin	440	19	37	9.1	9.1	40
	Enamelin	900	15	25	3.8	3.8	34
16861	Amelogenin X	185	14	15	36.8	38.9	72
	Ameloblastin	343	2	2	4.4	4.4	15
	Enamelin	915	2	2	1.2	1.2	11
Neg. Contr. Gr. 1: 235, 275, 706	ND						
Neg. Contr. Gr. 2: 630, 875, 889	ND						
Neg. Contr. Gr. 3: 1214, 1218	Amelogenin X	122	5	7	18.0	18.0	22

Aggregated data from different extraction methods and/or tissues from the same specimen are shown. In table cells that report two values separated by the | symbol, the left value refers to MaxQuant searches performed selecting unspecific digestion and the right value refers to MaxQuant searches performed selecting trypsin digestion. For those cells that include one value only, this value refers to MaxQuant searches performed selecting unspecific digestion. Final amino acid (aa) coverage, incorporating both the MaxQuant and PEAKS searches, is reported in the final column. Extended Data Table 1 provides the tissue sources per specimen, and the CGG and GNM specimen numbers.
*Supporting all peptides.

Corresponding author(s): Enrico Cappellini
Jesper V. Olsen
Eske Willerslev

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Mass spectrometric data were acquired using the Xcalibur™ Software, controlling the Thermo Scientific™ LC-MS systems.
Data analysis	MaxQuant (versions 1.5.3.30, for main searches, and 1.6.0.16, for the dependent peptides searches) PEAKS (version 7.5) Geneious (version 5.4.4) ANGSD (version 0.915) ProSplign MAFFT Phangorn (R package) PHyML (version 3.1) MrBayes (version 3.2.6) CASAVA (version 1.8.2) PALEOMIX (version 1.2.6) AdapterRemoval (version 1.5) BWA backtrack (versions 0.5.10, 0.7.12 and 0.7.15) BWA aln (version 0.7.7) MarkDuplicate (http://picard.sourceforge.net/) HiSeq Control Software 2.0.12.0/RTA 1.17.21.3 SAMtools (version 0.1.19)

SeqPrep (https://github.com/jstjohn/SeqPrep)
PRINSEQ-lite (v0.20.4)
BEDTools (version 2.25)
mapDamage2 (version 2.0.5)
Exonerate (version 2.2)
IceLogo (version 1.3.8)
MS2PIP (version 20190312)
IPSA tools (version 1.0)
MSConvert tool, part of ProteoWizard (version 3.0)
In-house developed R-script used to align the sequences identified by PEAKS (available upon request to the corresponding authors)
deamidation.py (publicly available at: https://github.com/dblyon/deamidation)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD011008. Genomic BAM files used for Rhinocerotidae protein sequence translation and protein sequence alignments used for phylogenetic reconstruction are included in the compressed archive named "Supplementary_Data_1.zip".

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was required. All available faunal specimen samples (23) were analyzed. Sample size includes numerous bone, dentine, and enamel samples, that therefore collectively allow us to estimate proteome survival in each of these tissues at the Dmanisi site.
Data exclusions	No data was excluded from the study.
Replication	Phylogenetic trees were reproduced using three different algorithms, and found consistent results (see Methods and SI). Proteomic results were replicated for several samples using repeated LC-MS/MS runs, and we observed consistent results within and between samples.
Randomization	Samples were injected in the LC-MS/MS system in randomised order.
Blinding	Ancient samples and control blanks were anonymised before the operator injected them in the LC-MS/MS system.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology

Specimen provenance	Studied specimens derive from the Dmanisi archaeological/palaeontological site in Georgia (see Methods). Export of specimens to the Centre of GeoGenetics, Natural History Museum of Denmark, University of Copenhagen was regulated by approval of D. Lordkipanidze, Director of the Georgian National Museum and co-author.
Specimen deposition	Specimens are available upon request to E. Willerslev, E. Cappellini (Natural History Museum of Denmark), or D. Lordkipanidze, (Georgian National Museum).
Dating methods	No new dates obtained.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.