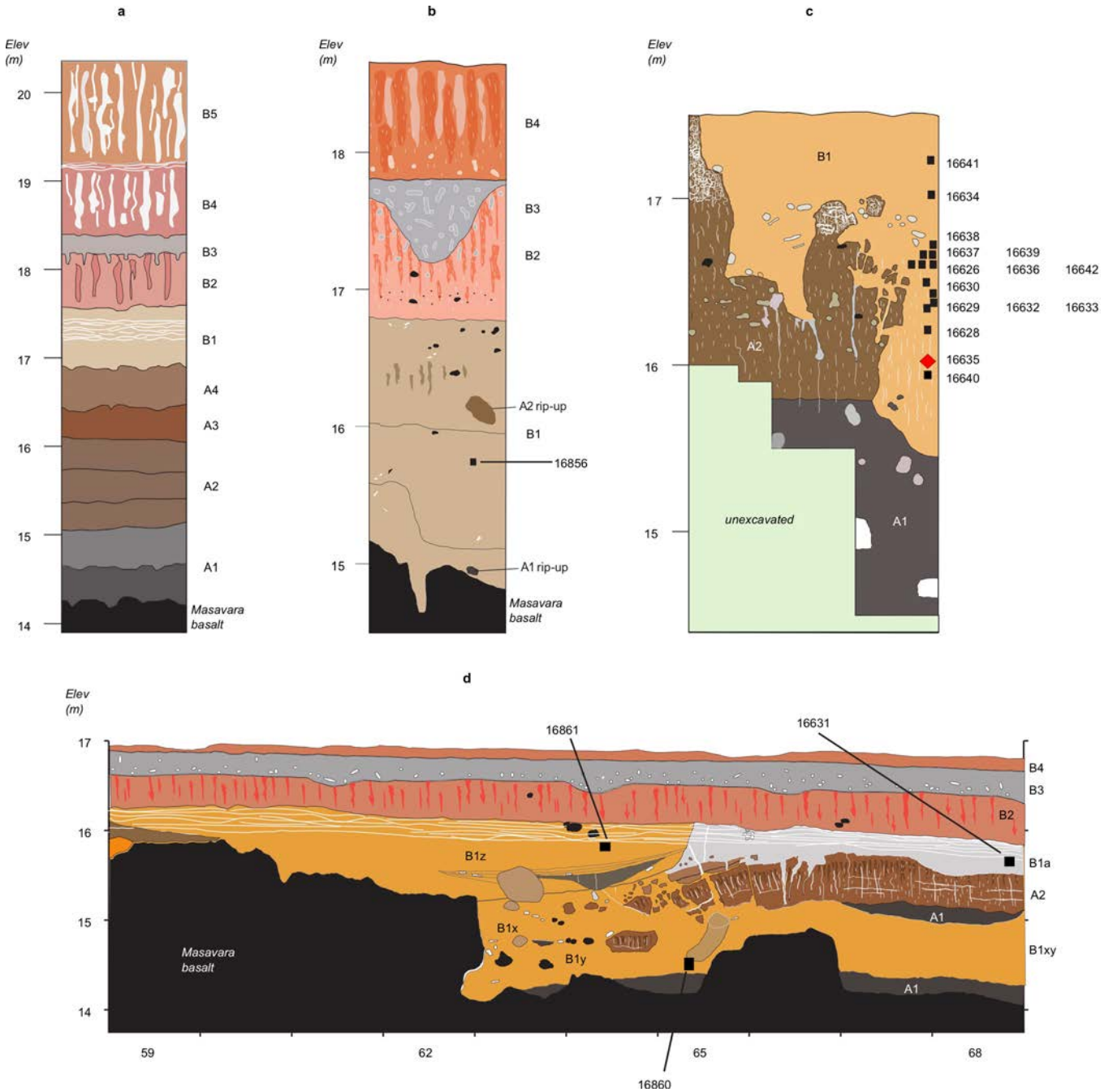
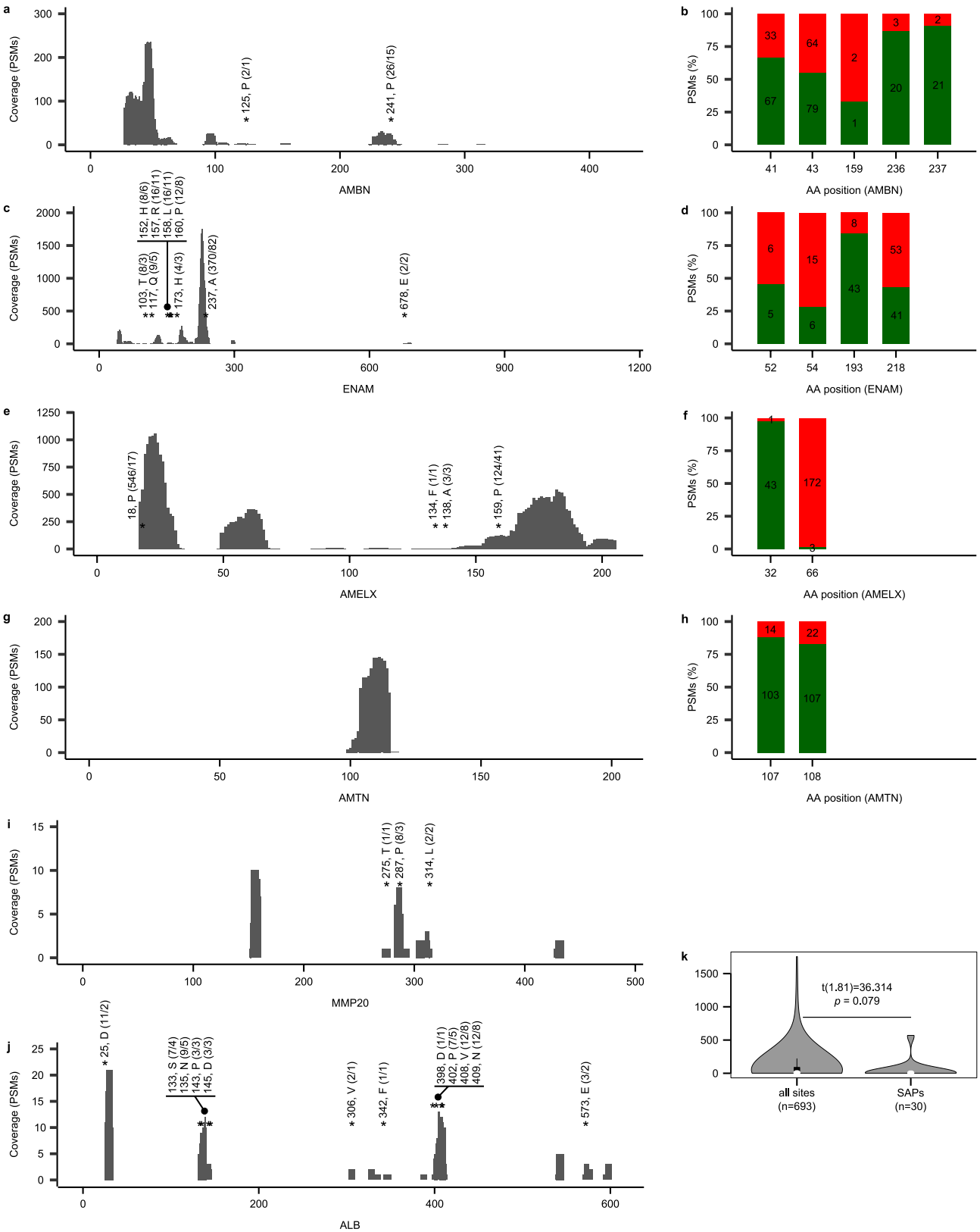
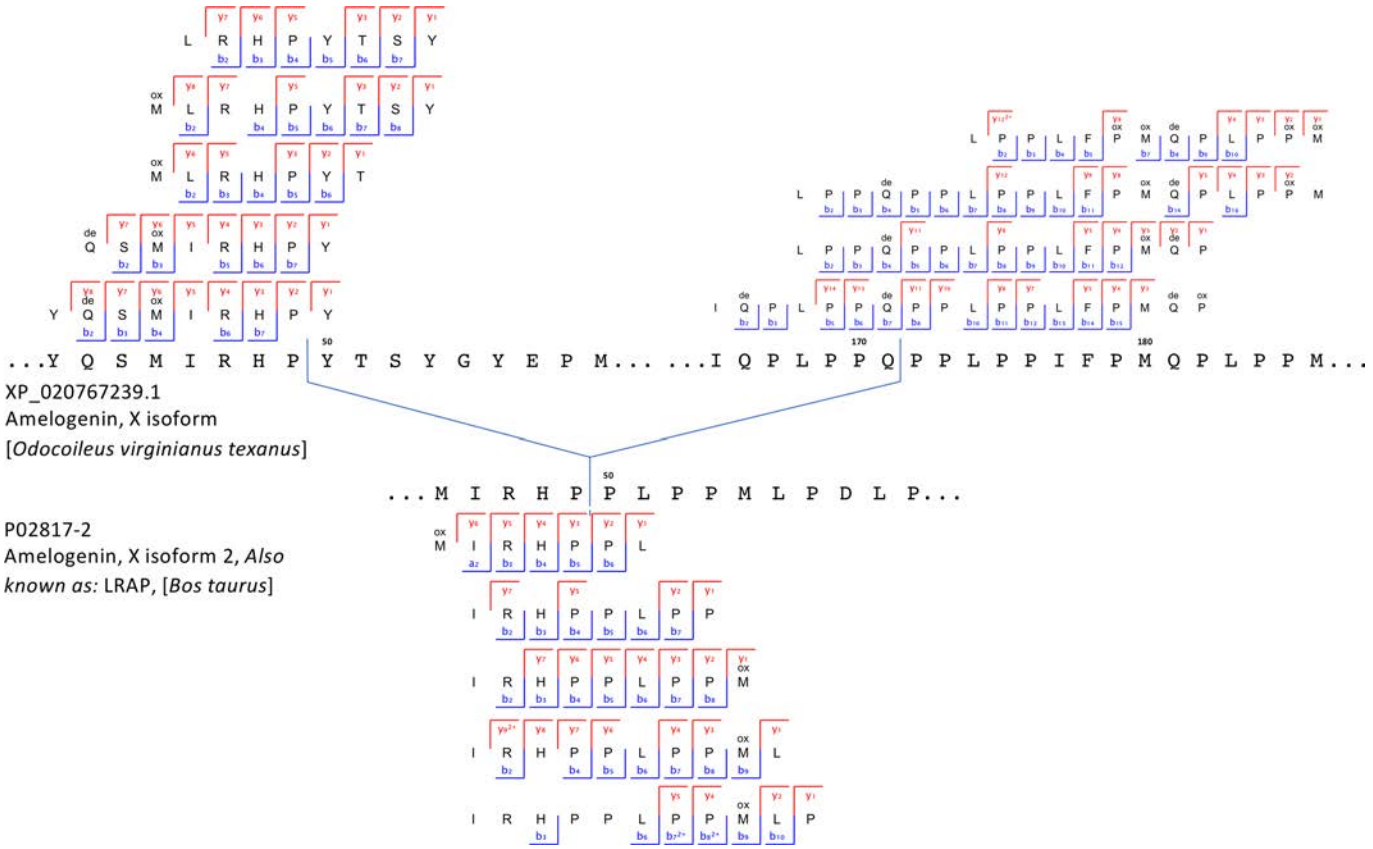


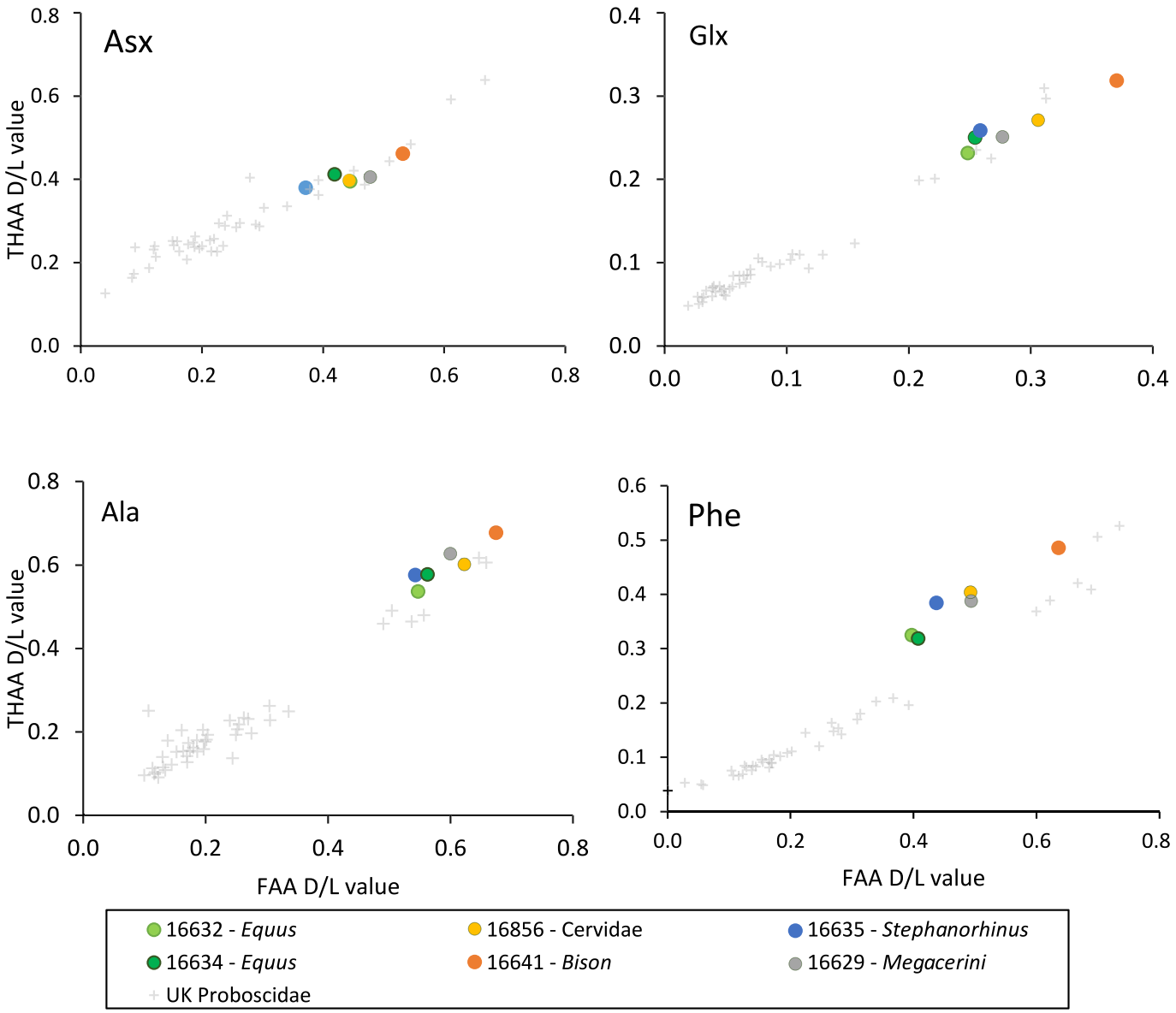
Extended Data Fig. 1 | Generalized stratigraphic profiles for Dmanisi, indicating origins of the specimens. a, Type section of the Dmanisi M5 excavation block. **b**, Stratigraphic profile of excavation area M6. M6 preserves a larger gully associated with the pipe-gully phase of stratigraphic–geomorphic development in stratum B1. The thickness of the stratum B1 gully fill extends to the basalt surface but includes 'rip-ups' of strata A1 and A2, showing that the deposits in stratum B1 post-date those of stratum A. **c**, Stratigraphic section of excavation area M17. Here, Stratum B1 was deposited after the erosion of stratum A deposits. The stratigraphic position of specimen Dm.5/157–16635 is highlighted with a red diamond. The Masavara basalt is about 50 cm below the base of the profile shown. **d**, Northern section of block 2. Following the collapse of a pipe and erosion to the basalt, the deeper part of this area was filled with local gully fill of strata B1x, B1y and B1z. Note the uniform burial of all stratum B1 deposits by strata B2, B3 and B4. The sampled specimens are indicated by the five-digit CGG numbers. Extended Data Table 1 provides both the CGG and GNM specimen numbers.



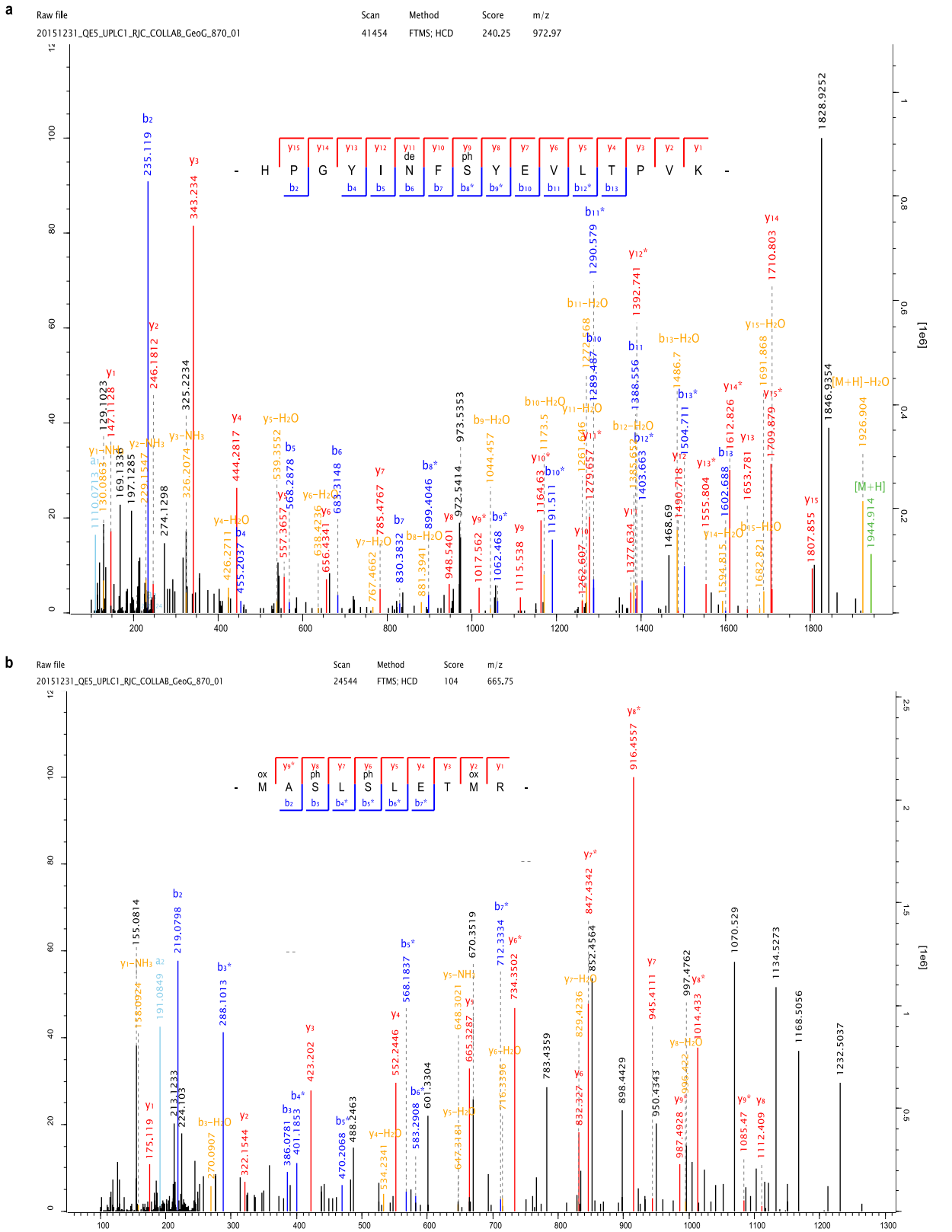
Extended Data Fig. 2 | Proteome-sequence coverage for specimen Dm.5/157-16635. a, **c**, **e**, **g**, **i**, **j**, Peptide–spectrum match (PSM) sequence coverage of the proteins AMBN (**a**), ENAM (**c**), AMELX (**e**), AMTN (**g**), MMP20 (**i**) and ALB (**j**). Annotations include 'amino acid position, amino acid called in that position (number of PSMs and peptides covering that position)' for the phylogenetically informative single-amino-acid polymorphisms within Rhinocerotidae. **b**, **d**, **f**, **h**, Frequency (per cent) of phosphorylated (green) and unphosphorylated (red) PSMs per amino acid position for AMBN (**b**), ENAM (**d**), AMELX (**f**) and AMTN (**h**). Numbers within the bars provide the PSM counts. **k**, Violin plot of distribution of PSM coverage for all covered sites ($n = 693$), and for sites of phylogenetic relevance (single-amino-acid polymorphisms, $n = 30$). The box plots define the range of the data, with whiskers extending to 1.5× interquartile range, boxes denoting the 25th and 75th percentiles and dots indicating the median. All panels are based only on MaxQuant search results. The Supplementary Data contains examples of MS/MS spectra, and fragment-ion series alignments for each of the marked single-amino-acid polymorphisms.



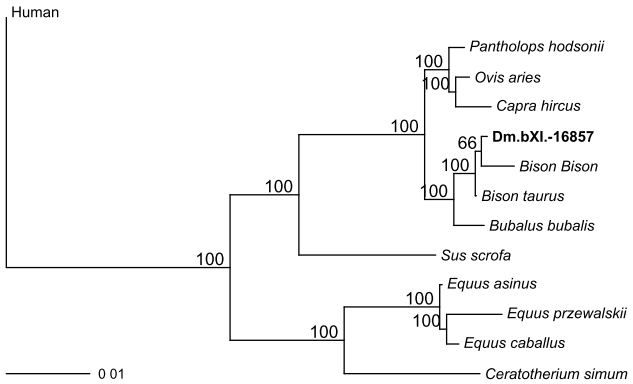
Extended Data Fig. 3 | Peptide and fragment-ion coverage of AMELX isoform 1 and isoform 2 from specimen Dm.M6/7.II.296–16856. Peptides specific to AMELX isoform 1 and isoform 2 appear in the top and bottom parts of the figure, respectively. No AMELX isoform 2 is currently reported in public databases for the Cervidae group. Accordingly, the AMELX-isoform-2-specific peptides were identified by MaxQuant spectral matching against bovine (*Bos taurus*) AMELX isoform 2 (UniProt accession number P02817-2). AMELX isoform 2 (also known as leucine-rich amelogenin peptide (LRAP)) is a naturally occurring isoform of AMELX from the translation product of an alternatively spliced transcript.



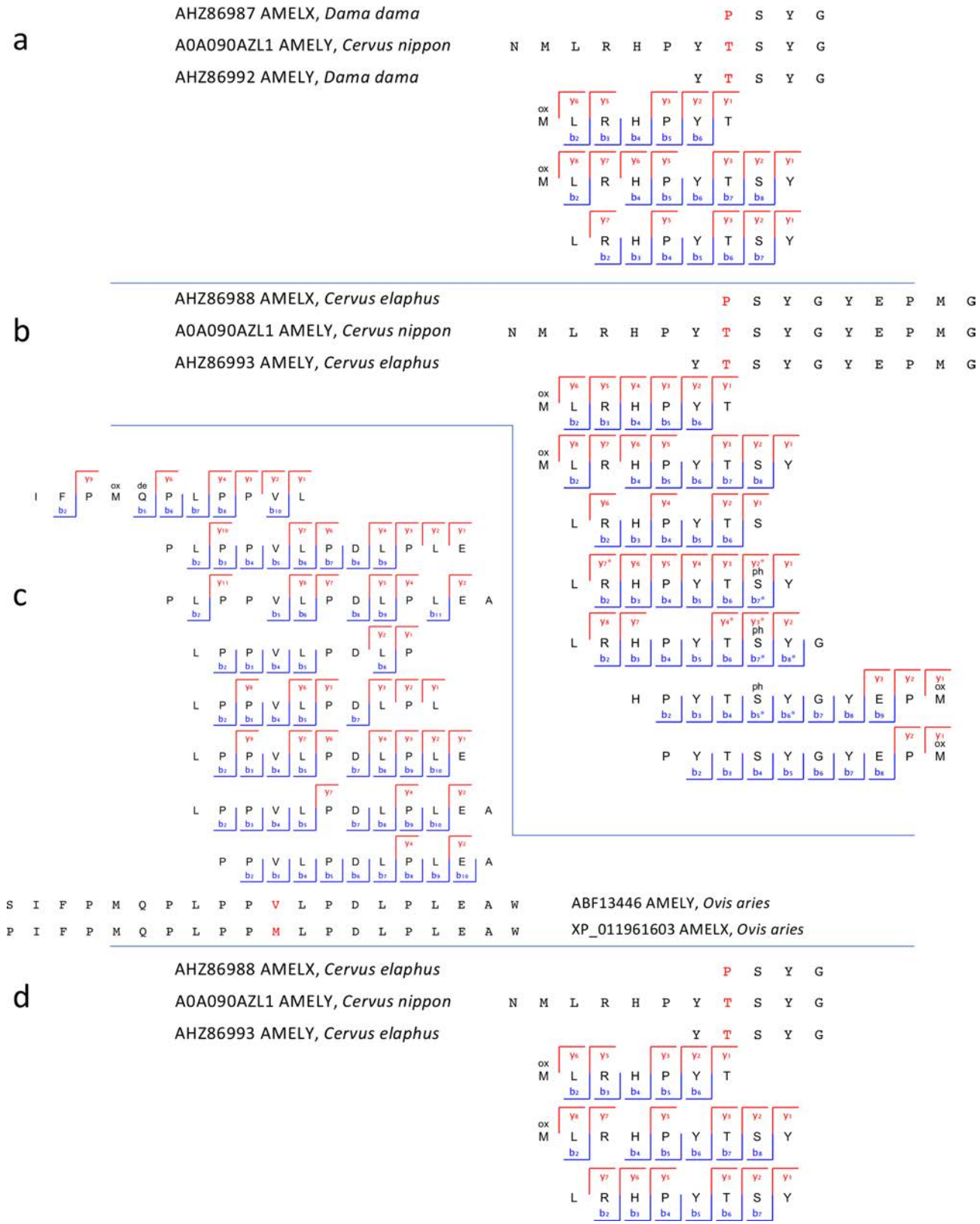
Extended Data Fig. 4 | Amino acid racemization. Extent of intra-crystalline racemization in enamel for the free amino acid (FAA, *x* axis) fraction and the total hydrolysable amino acids (THAA, *y* axis) fraction for four amino acids (Asp plus Asn (here denoted Asx), Glu plus Gln (here denoted Glx), Ala and Phe). Note the differences in axis scale. Intra-crystalline data from Proboscidea enamel from a range of sites in the UK[64] have been shown for comparison (grey crosses). Taxa from both Dmanisi and the UK exhibit a similar relationship between FAA and THAA racemization, and R^2 values have been calculated on the basis of a polynomial relationship (order = 2, all > 0.93).



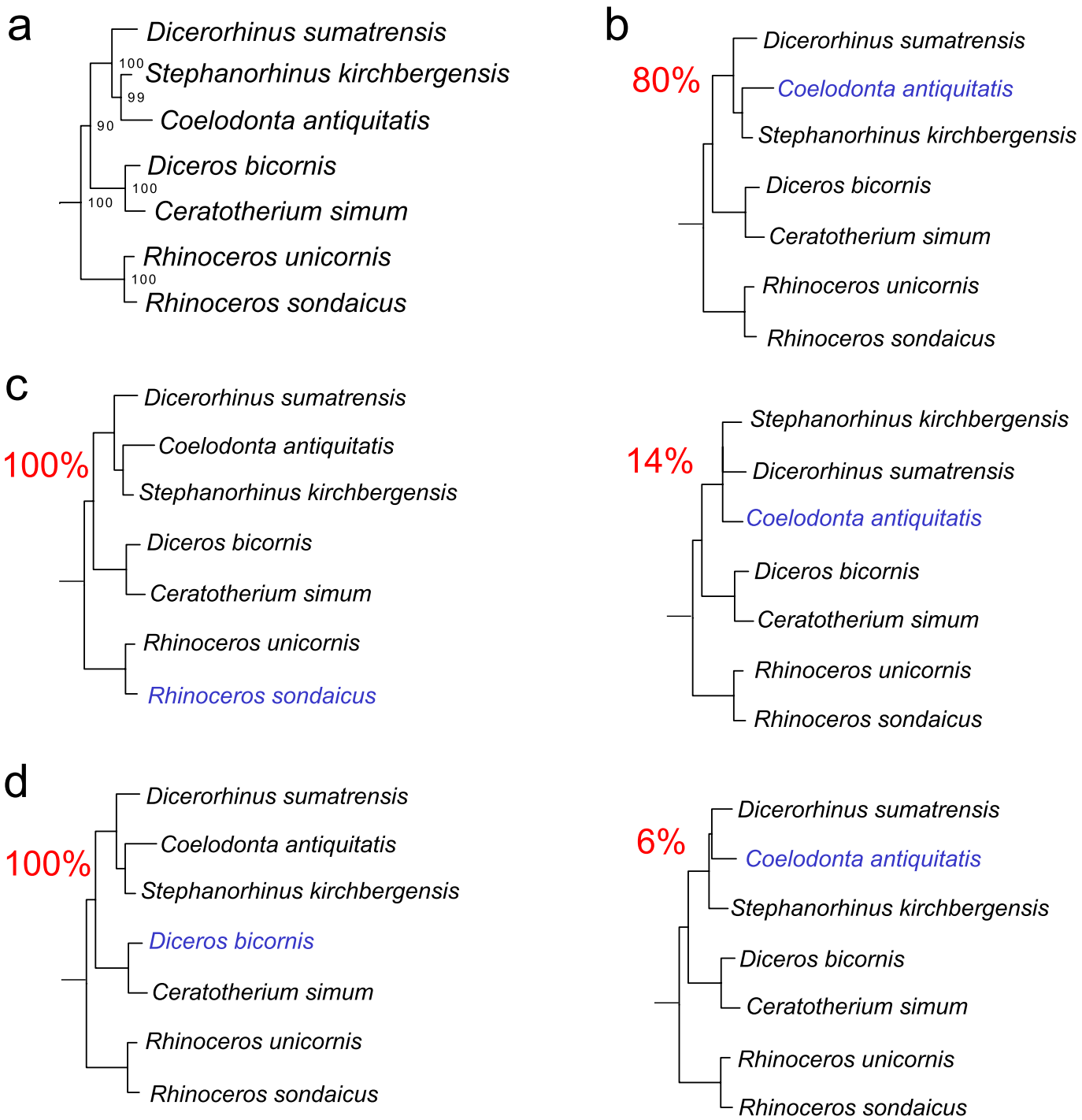
Extended Data Fig. 5 | Phosphorylation in the proteome of ancient enamel. Annotated spectra including phosphorylated (here denoted ph) serine (S). **a**, Phosphorylation in the S-X-E motif of AMELX. **b**, Phosphorylation in the S-X-phosphorylated S motif of AMBN. Phosphorylation was independently observed in all three separate analyses of Dm.5/157–16635, including multiple spectra and peptides (Extended Data Fig. 2).



Extended Data Fig. 6 | Phylogenetic relationships between the comparative reference dataset and specimen Dm.bXI–16857. Consensus tree from Bayesian inference. The posterior probability of each bipartition is shown as a percentage to the left of each node.



Extended Data Fig. 7 | AMELY-specific matches. a, Specimen Dm.6/151.4.A4.12–16630. **b**, Specimen Dm.69/64.3.B1.53–16631. **c**, Specimen Dm.8/154.4.A4.22–16639. **d**, Specimen Dm.M6/7.II.296–16856. Note the presence of deamidated glutamine (deQ) and asparagine (deN), oxidated methionine (oxM) and phosphorylated serine (phS).



Extended Data Fig. 8 | Effect of the missingness in the tree topology. **a**, Maximum-likelihood phylogeny obtained using PhyML and the protein alignment that excludes Dm.5/157–16635. **b**, Topologies obtained from 100 random replicates of the woolly rhinoceros (*C. antiquitatis*). In each replicate, the number of missing sites was similar to that observed for the Dm.5/157–16635 specimen (72.4% missingness). The percentage shown for each topology indicates the number of replicates in which that particular topology was recovered. **c**, As in **b**, but for the Javan rhinoceros (*R. sondaicus*). **d**, As in **b**, but for the black rhinoceros (*D. bicornis*).

Extended Data Table 1 | Genome and proteome survival in 23 specimens of fossil fauna from Dmanisi

CGG ref. numb.	GNM specimen number	Morphological identification*	Anatomy	Ancient DNA	Protein extr. Method A	Protein extr. Method B	Protein extr. Method C	Phylogenetic analysis
16486	Dm.bXI.sqA6.V._.	*Canis etruscus*	P4 sin.				○E+D	
16626	Dm.6/154.2/4.A4.17	Artiodactyla	tibia sin.			○B		
16628	Dm.7/154.2.A2.27	Cervidae	mc III&IV dex.			●B†		
16629	Dm.5/154.3.A4.32	Cervidae	hemimandible sin. with dp2, dp3, dp4, m1			○B	●E+D	
16630	Dm.6/151.4.A4.12	*Pseudodama nestii*	hemimandible dex. with p2-m3			○B	○D, ●E	
16631	Dm.69/64.3.B1.53	Cervidae	maxilla sin. with P3			○B	○D, ●E	
16632	Dm.5/154.2.A4.38	*Equus stenonis*	i3 dex.				●E+D	Fig. S10
16633	Dm.5/153.3.A2.33	*Equus stenonis*	mc III & mc II sin.				○B	
16634	Dm.7/151.2.B1/A4.1	*Equus stenonis*	m/1 or m/2 dex.				○D, ●E	
16635	Dm.5/157.profile cleaning	*Stephanorhinus* sp.	m/1 sin.	○			○D, ●E	Fig. 4, Fig. S11
16636	Dm.6/153.1.A4.13	Rhinocerotidae	tibia dex.			○B		
16637	Dm.7/154.2.A4.8	Bovidae	mt III&IV sin.			●B†		
16638	Dm.5/154.1.B1.1	Bovidae	hemimandible dex. with p3-m3			○B	○D, ●E	Fig. S12
16639	Dm.8/154.4.A4.22	Bovidae	maxilla dex. with P2-M2				○D, ●E	Fig. S13
16640	Dm.6/151.2.A4.97	*Bison georgicus*	mt III&IV sin.			○B		
16641	Dm.8/152.3.B1.2	*Bison georgicus*	m3 dex.				○D, ●E	Fig. S14
16642	Dm.8/153.4.A4.5	*Canis etruscus*	hemimandible sin. with p1-m2				○D, ●E	
16856	Dm.M6/7.II.296	Cervidae	m2 sin.	○	●D†	○D, ●E	●E+D	
16857	Dm.bXI.profile cleaning	Indet.	long bone fragment of a herbivore	○	●B†	○B	○B	Fig. S15, EDF6
16858	Dm.bXI.North.B1a. collection	Cervidae	metapodium fragment		○B	○B	○B	
16859	D4.collection	Indet.	fragments of pelvis and ribs of a large mammal	○	○B	○B	○B	
16860	Dm.65/62.1.A1. collection	Cervidae	P4 sin.	○		○D, ●E	○D, ●E	
16861	Dm.64/63.1.B1z. collection	*Equus stenonis*	fragment of an upper tooth			○D, ●E	○D, ●E	
Neg. contr. (blank)					NC	NC	NC	

The CGG reference number and the GNM specimen field number are reported for each specimen. B, bone; D, dentine; E, enamel. Extractions of enamel might include some residual dentine. Accordingly, both tissues are either listed separately (in cases with no collagen preservation) or together (in cases with collagen preservation). Open circles indicate no molecular preservation; closed circles indicate molecular preservation.

*Or the narrowest possible taxonomic identification achievable using comparative anatomy methods.

†Only collagens survive.

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.