
Supplementary information

Enamel proteome shows that *Gigantopithecus* was an early diverging pongine

In the format provided by the authors and unedited

Frido Welker, Jazmín Ramos-Madrigal, Martin Kuhlwilm, Wei Liao, Petra Gutenbrunner, Marc de Manuel, Diana Samodova, Meaghan Mackie, Morten E. Allentoft, Anne-Marie Bacon, Matthew J. Collins¹, Jürgen Cox, Carles Lalueza-Fox, Jesper V. Olsen, Fabrice Demeter, Wei Wang, Tomas Marques-Bonet & Enrico Cappellini

SUPPLEMENTARY INFORMATION FOR

Enamel proteome shows that *Gigantopithecus* was an early diverging pongine

Frido Welker^{1*}, Jazmín Ramos-Madrigal¹, Martin Kuhlwilm², Wei Liao^{3,4}, Petra Gutenbrunner⁵, Marc de Manuel², Diana Samodova⁶, Meaghan Mackie^{1,6}, Morten E. Allentoft⁷, Anne-Marie Bacon⁸, Matthew J. Collins^{1,9}, Jürgen Cox⁵, Carles Lalueza-Fox², Jesper V. Olsen⁶, Fabrice Demeter^{7,10}, Wei Wang^{11*}, Tomas Marques-Bonet^{2,12,13,14*}, Enrico Cappellini^{1*}

¹ Evolutionary Genomics Section, Globe Institute, University of Copenhagen, Copenhagen, Denmark.

² Institute of Evolutionary Biology (UPF-CSIC), University Pompeu Fabra, Barcelona, Spain.

³ School of Earth Sciences, China University of Geosciences, Wuhan, China.

⁴ Anthropology Museum of Guangxi, Nanning, China.

⁵ Computational Systems Biochemistry, Max Planck Institute of Biochemistry, Martinsried, Germany.

⁶ Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark.

⁷ Lundbeck Foundation GeoGenetics Centre, Globe Institute, University of Copenhagen, Copenhagen, Denmark.

⁸ CNRS FRE 2029 BABEL, Université Paris Descartes, Faculté de Chirurgie Dentaire, Paris, France.

⁹ Department of Archaeology, University of Cambridge, Cambridge, UK.

¹⁰ UMR7206 Eco-anthropologie, Muséum national d'Histoire naturelle, Musée de l'Homme, Paris, France.

¹¹ Institute of Cultural Heritage, Shandong University, Qingdao, China.

¹² Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain.

¹³ Centre for Genomic Regulation (CNAG-CRG), Barcelona Institute of Science and Technology, Barcelona, Spain.

¹⁴ Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain.

*Corresponding authors: E. Cappellini (ecappellini@bio.ku.dk), T. Marques-Bonet (tomas.marques@upf.edu), F. Welker (frido.welker@bio.ku.dk), and W. Wang (wangw@sdu.edu.cn).

Humans from the Simons Genome Diversity Project.

The following 19 humans sequenced as part of the Simons Genome Diversity Panel were used to construct the human reference panel used for phylogenetic alignment (“Full alignments”):

Homo_sapiens-Africa_20ERR1419150	Homo_sapiens-Africa_21ERR1419161
Homo_sapiens-Africa_22ERR1419175	Homo_sapiens-Africa_23ERR1419183
Homo_sapiens-Africa_24ERR1347655	Homo_sapiens-Africa_25ERR1419136
Homo_sapiens-Africa_26ERR1419145	Homo_sapiens-Africa_27ERR1419179
Homo_sapiens-Africa_28ERR1419171	Homo_sapiens-WestEurasia_0ERR1395604
Homo_sapiens-WestEurasia_1ERR1419103	Homo_sapiens-WestEurasia_2ERR1419088
Homo_sapiens-WestEurasia_3ERR1419098	Homo_sapiens-WestEurasia_4ERR1419111
Homo_sapiens-WestEurasia_5ERR1419121	Homo_sapiens-WestEurasia_6ERR1419141
Homo_sapiens-WestEurasia_7ERR1419091	Homo_sapiens-WestEurasia_8ERR1419102
Homo_sapiens-WestEurasia_9ERR1419117	

Supplementary Table S1. Proteome characteristics of Dm.5/157–16635 (*Stephanorhinus* sp., Dmanisi, Georgia, main text reference ¹⁷) and CF-B-16 (*Gigantopithecus blacki*, Chuifeng Cave, China, this study).

	<i>Stephanorhinus</i>	<i>Gigantopithecus</i>
Chronological age (million years ago)	1.77	1.9
Total number of spectra acquired	117,516	38,082
Total number of spectra matched	6,393	1,398
% Spectra matched	5.44	3.67
Number of proteins identified	6	6
Number of peptides identified	987	409
Number of amino acids covered	875	456
Average peptide length	10.89	10.20
Average deamidation rate N*	96.34 ± 1.08	92.85 ± 3.29
Average deamidation rate Q*	96.57 ± 0.57	92.30 ± 2.19

*mean ± 1 standard deviation obtained after 1,000 bootstrap replicates of MaxQuant proteomic data analysis (endogenous proteins only, see Methods).

Supplementary Table S2. Entry names of publicly available protein sequences utilized in database construction and phylogenetic analysis.

Protein	<i>Homo sapiens</i>	<i>Pan paniscus</i>	<i>Pan troglodytes</i>	<i>Gorilla gorilla</i>	<i>Pongo abelii</i>	<i>Nomascus leucogenys</i>	<i>Macaca Mulatta</i>
COL1α1	CO1A1_HUMAN	XP_003817507	H2QDE6_PANT R	G3RBN8_GORGO	H2NVM9_PONAB	XP_012358721	H9Z595_MACMU
COL1α2	CO1A2_HUMAN	XP_003809763	H2QUY2_PANT R	G3QT97_GORGO	H2PMW7_PONAB	G1RZZ2_NOMLE	H9Z2D1_MACMU
ALB	ALBU_HUMAN	XP_003832390	H2RBT1_PANT R	G3S791_GORGO	ALBU_PONAB	G1R8T8_NOMLE	ALBU_MACMU
AMBN	AMBN_HUMAN	XP_003809040	H2R148_PANT R	G3RCU1_GORGO	H2PDI5_PONAB	G1R841_NOMLE	F7HLX4_MACMU
AMELY	AMELY_HUMAN	(absent)	AMELY_PANTR	C3UJP7_9PRIM	(absent)	(absent)	A0A1D5RDA1_MACMU
AMELX	AMELX_HUMAN	XP_003805726	A5JJS6_PANTR	G3SDK0_GORGO	H2PUX0_PONAB	G1RCS3_NOMLE	A5JJS8_MACMU
ENAM	ENAM_HUMAN	B2L7U5_PANPA	H2QPM0_PANTR	B2L7U8_9PRIM	H2PDI6_PONAB	G1R843_NOMLE	F7H832_MACMU
TUFT1	TUFT1_HUMAN	XP_003817293	K7CQG4_PANT R	G3QY68_GORGO	H2N5V2_PONAB	G1RGY4_NOMLE	G7MDK9_MACMU
KLK4	KLK4_HUMAN	XP_003813692	XP_009434410	G3QU55_GORGO	A0A2J8U913_PONAB	G1R1C5_NOMLE	G7NMD1_MACMU
MMP20	MMP20_HUMAN	XP_003828430	H2Q4M8_PANT R	G3QLA8_GORGO	H2NF32_PONAB	G1R6B1_NOMLE	F7GQW6_MACMU
AMTN	AMTN_HUMAN	XP_003809041	H2QPL9_PANT R	G3RVJ5_GORGO	H2PDI4_PONAB	G1R825_NOMLE	F6VN65_MACMU
ODAM	ODAM_HUMAN	XP_003809049	ODAM_PANTR	G3QY18_GORGO	H2PDH6_PONAB	G1R7Z0_NOMLE	ODAM_MACMU
AHSG	FETUA_HUMAN	XP_008953975	FETUA_PANTR	E1U7Q5_9PRIM	H2PC98_PONAB	G1R4B1_NOMLE	F6VZ47_MACMU

Entry names either refer to UniProt or GenBank.

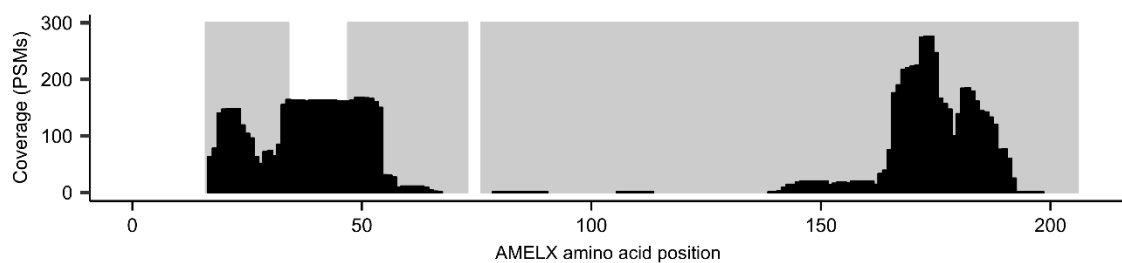
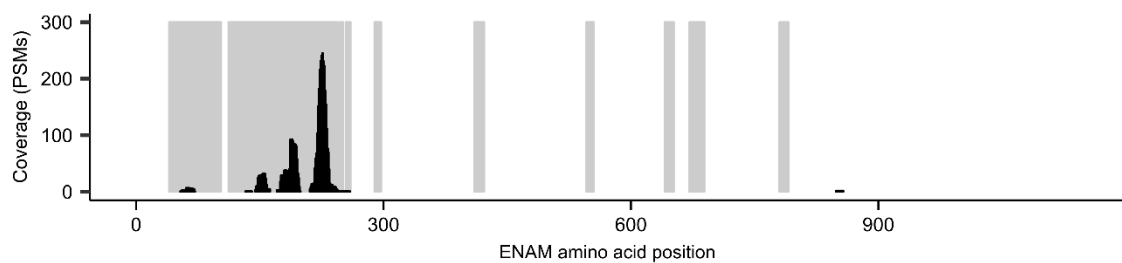
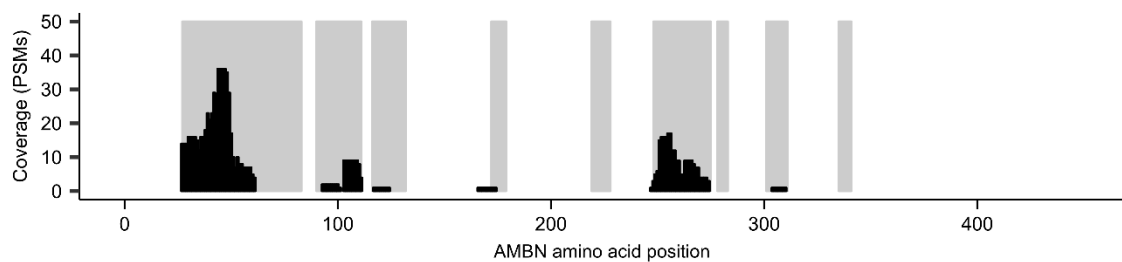
Supplementary Table S3. Description of the protein alignments used in the study.

Protein	Total sites	Non-missing sites	Unique sites ¹	Single-individual alignment		Full alignment	
				Polymorphic sites ²	Polymorphic sites & non-singletons ²	Polymorphic sites ²	Polymorphic sites & non-singletons ²
AMBN	447	116	0	64 (7)	19 (3)	74 (11)	28 (5)
AMELX	206	145	2	9 (4)	2 (0)	17 (11)	3 (1)
AMTN	209	18	0	26 (0)	13 (0)	39 (1)	24 (1)
ENAM	1,142	186	1	161 (29)	54 (12)	189 (31)	87 (16)
MMP20	483	9	0	35 (3)	11 (0)	41 (3)	20 (0)

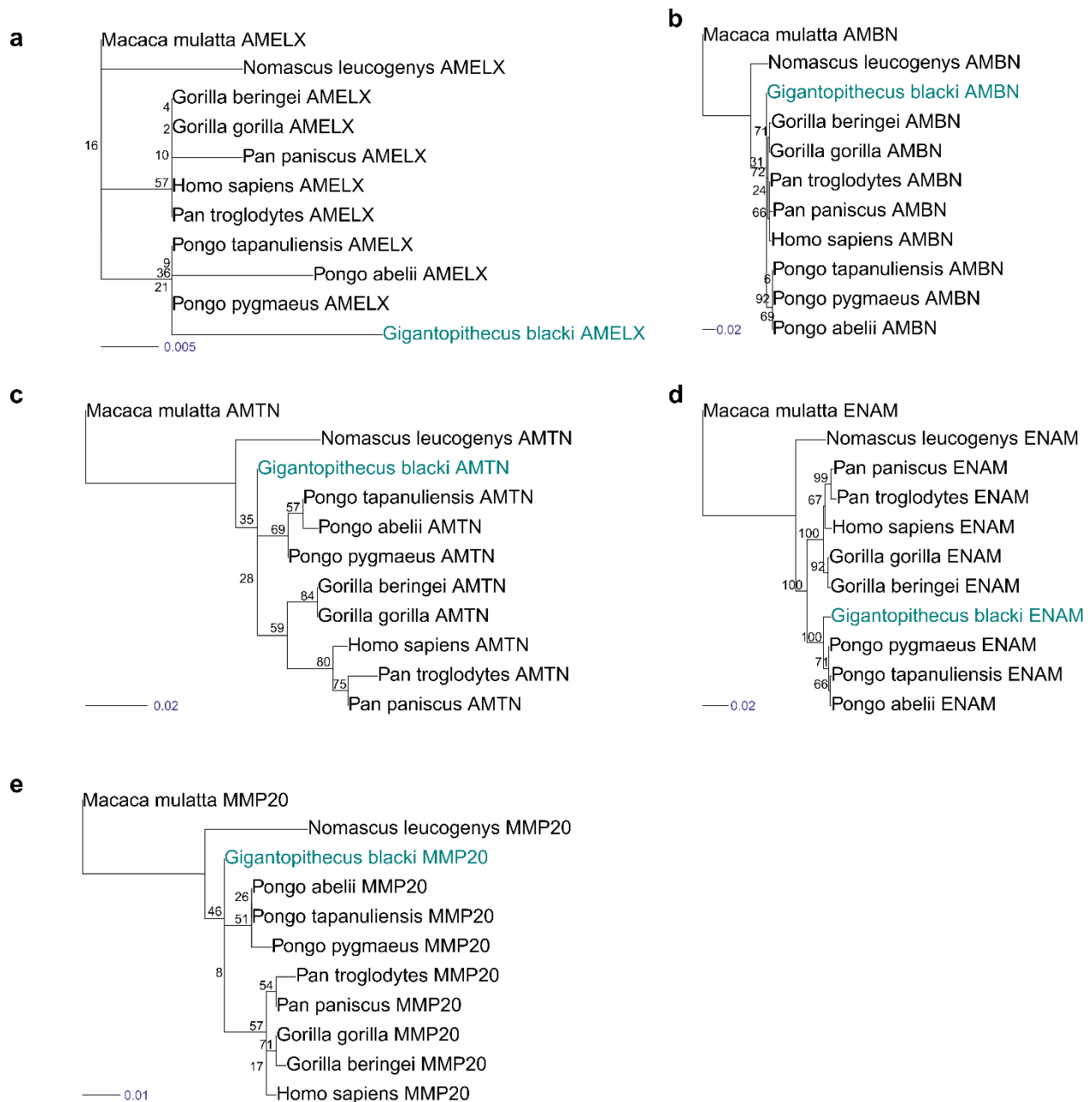
Single individual alignment: alignment built using a single individual per species. This alignment was used to estimate

divergence times for each node. Full alignment: alignment built using all individuals in the reference dataset. ¹:

Private mutations/polymorphism found in the ancient sample. ²: The number of sites where the ancient sample is non-missing are indicated in brackets.



Previous page. **Supplementary Figure S1. MaxQuant PSM coverage plot for the six endogenous enamel proteins recovered from the *Gigantopithecus* enamel.** Grey background shading indicates regions with peptide sequence coverage in a previously published Early Pleistocene enamel proteome from Dmanisi (*Stephanorhinus* sp.; ¹⁹). Although differences are expected, due to differences in primary sequence, recovered peptide sequences from *Gigantopithecus* generally fall within sequence regions identified in the thermally younger *Stephanorhinus* sp. specimen from Dmanisi.



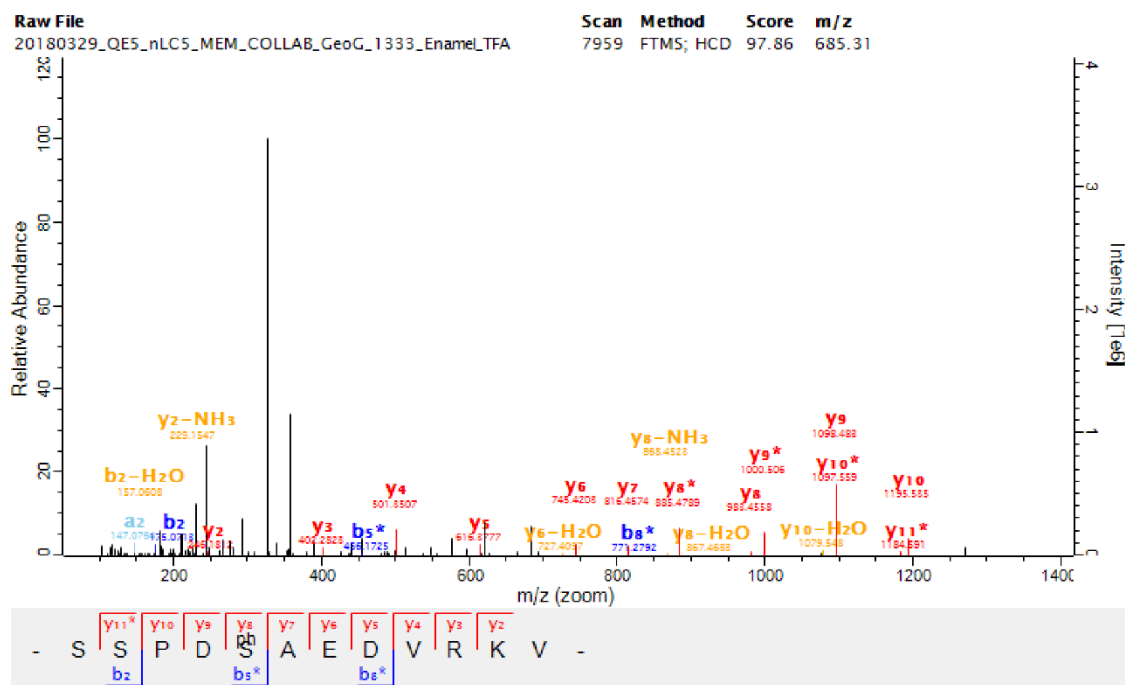
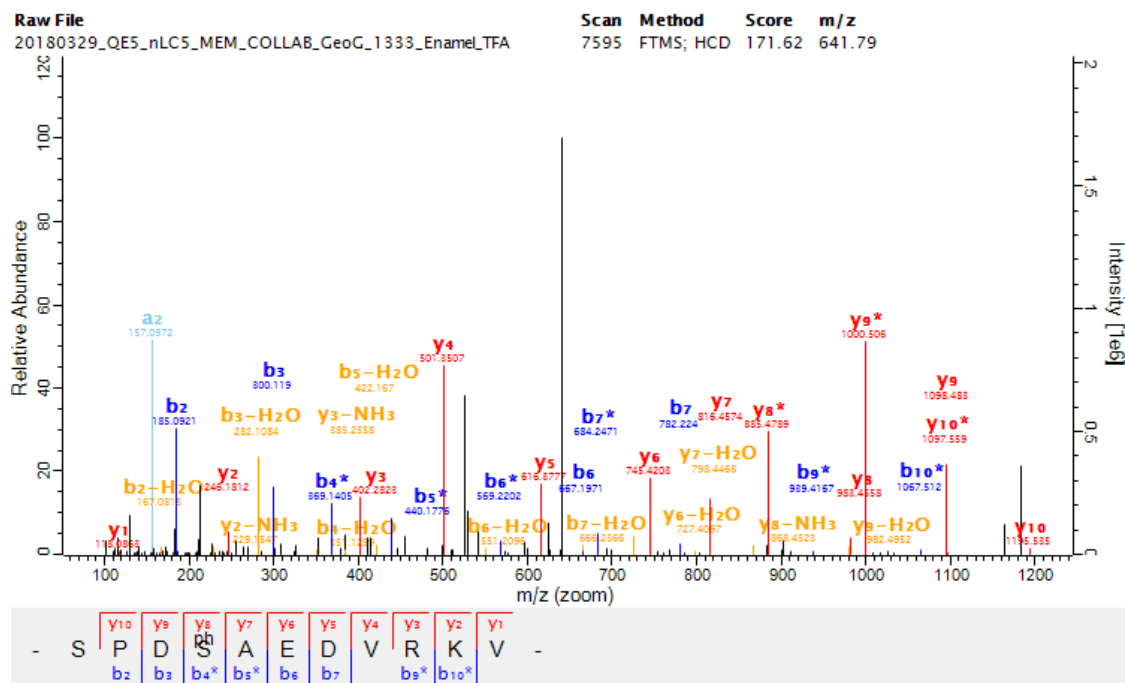
Supplementary Figure S2. PHYML individual protein trees for the five endogenous proteins recovered from the *Gigantopithecus blacki* enamel proteome (AHSG excluded). a, AMELX. b, AMBN. c, AMTN. d, ENAM. e, MMP20. Nodal support values are after 100 bootstraps.

Note on Supplementary Figures S3 to S12.

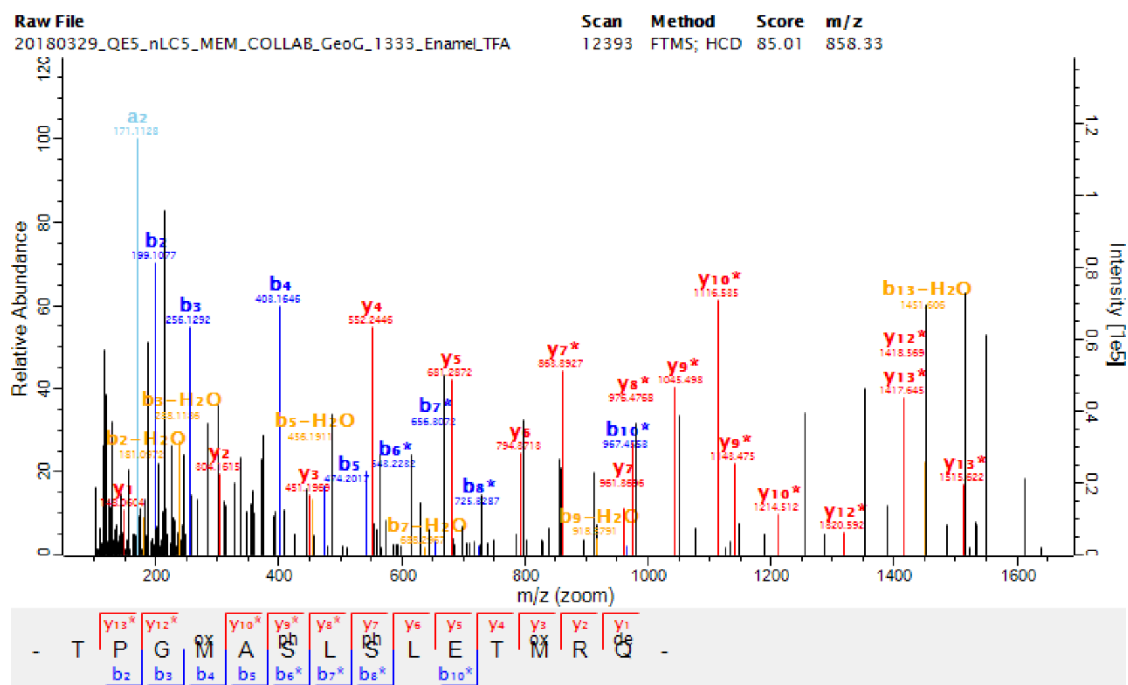
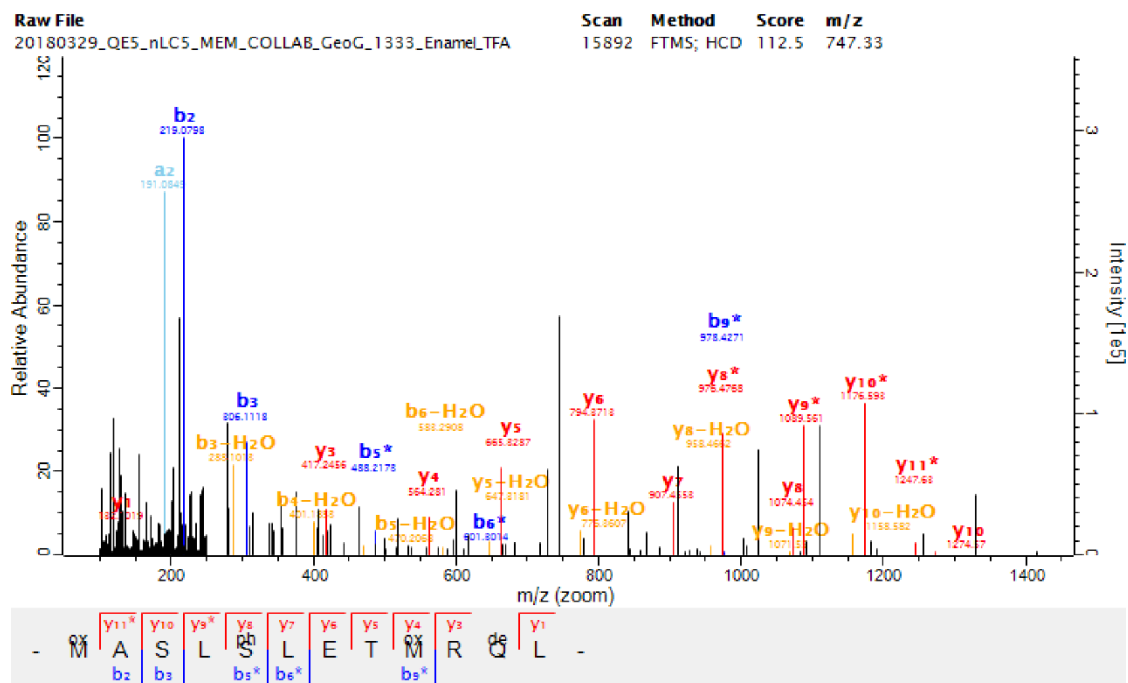
Supplementary figures S3 to S12 contain MaxQuant-derived examples of annotated MS/MS spectra. Fragment ions are labelled in the “Standard” annotation mode provided in MaxQuant. For each panel, both the annotated MS/MS spectrum and an amino acid sequence, covered by the relevant b- and y-ions, are shown. For the latter, additional PTM annotations on relevant amino acid positions include “*ph*” (phosphorylation), “*de*” (deamidation), and “*ox*” (oxidation). In some of these cases, additional PSMs were also identified by MaxQuant (Supplementary Figure S1) and PEAKS.

Supplementary figures S3 to S8 include unique spectra matching to peptide sequences representing one of the six endogenous proteins identified in the *Gigantopithecus* proteome. Supplementary Figures S9 to S12 include several examples of MS/MS spectra overlapping SAPs of phylogenetic interest.

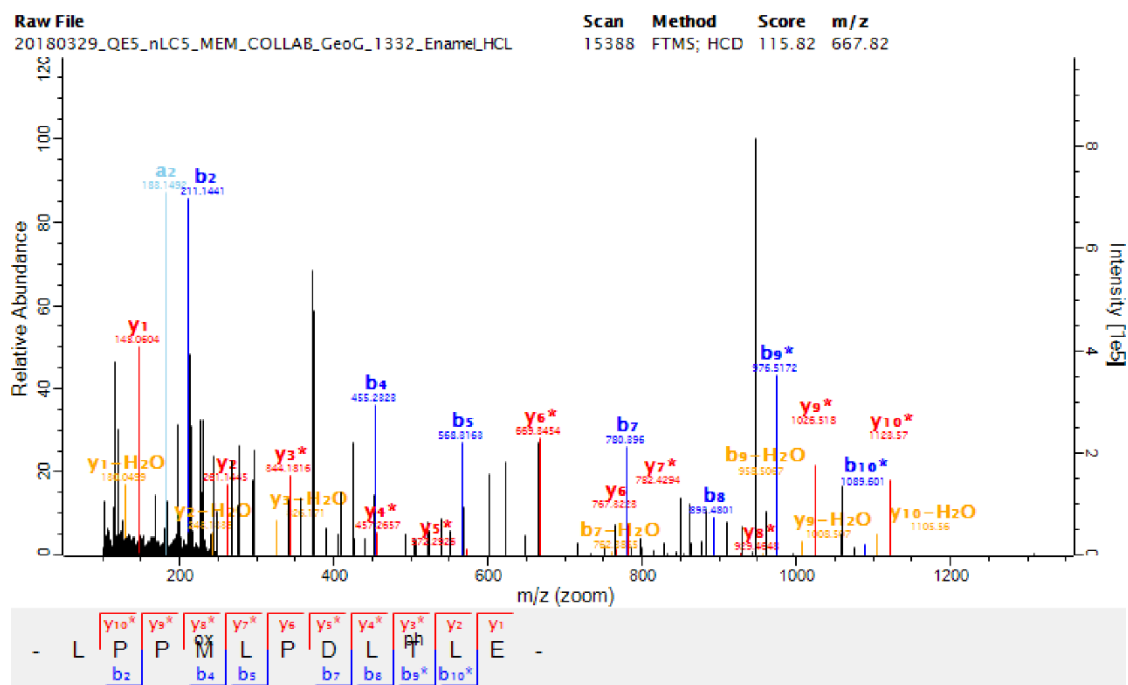
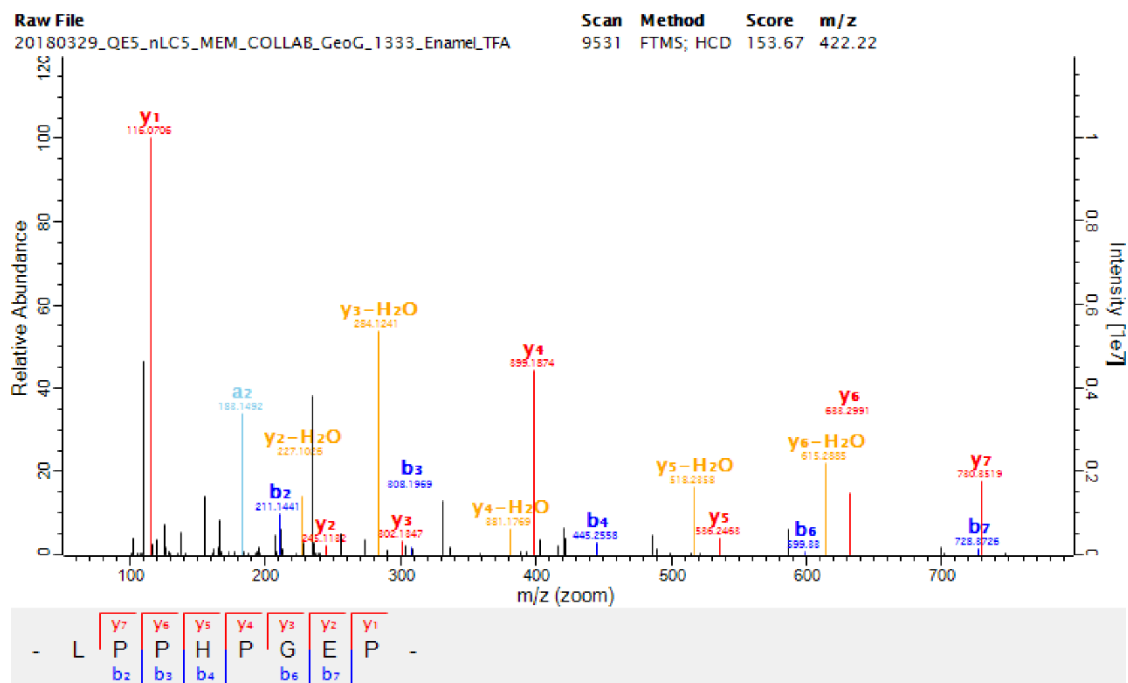
All the mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the data set identifier PXD013838.

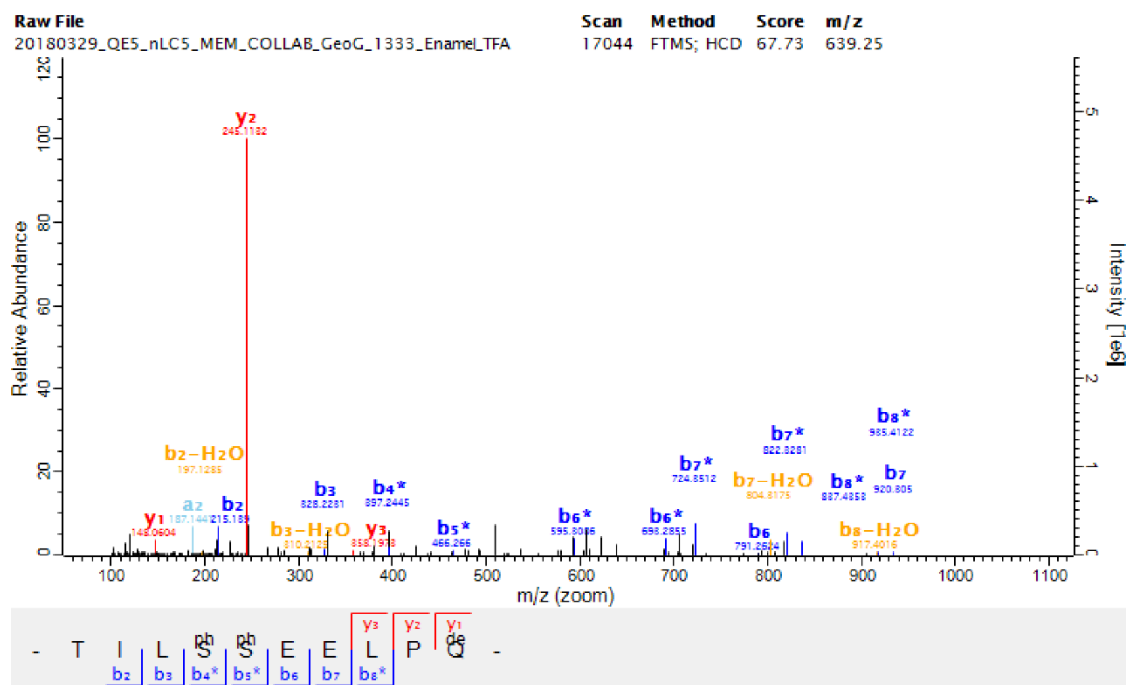
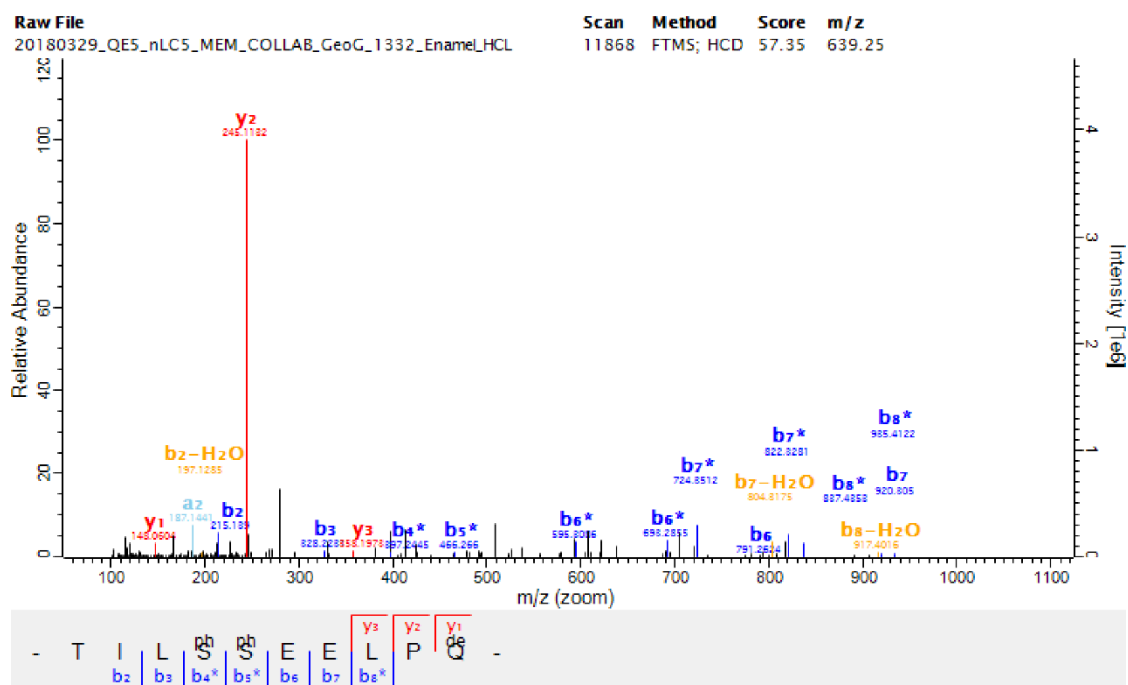


Supplementary Figure S3. Examples of annotated MS/MS spectra unique to AHSF. *Top*, amino acid positions 135-145. ***Bottom***, amino acid positions 134-145. Coordinates refer to UniProt entry H2PC98_PONAB.

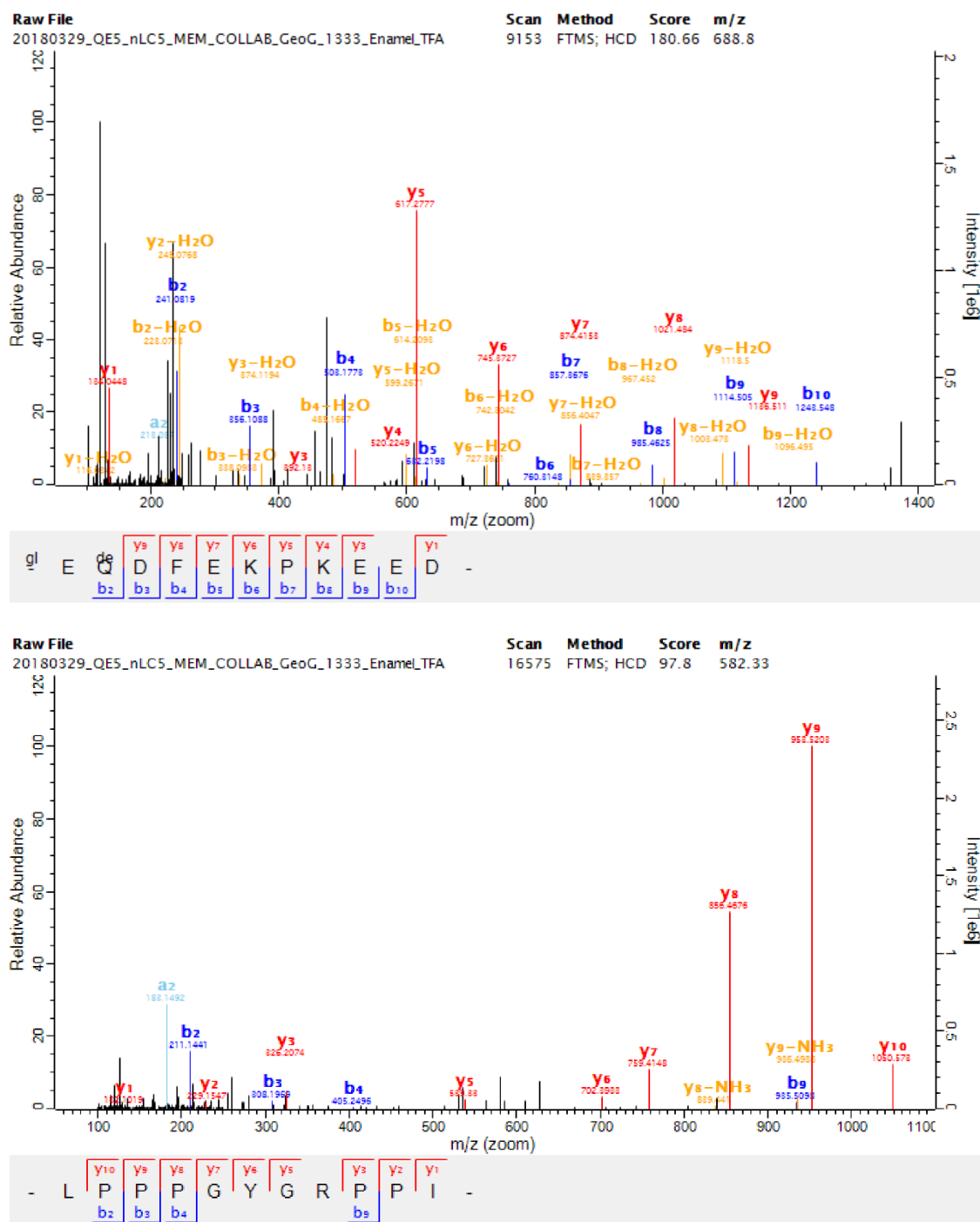


Supplementary Figure S4. Examples of annotated MS/MS spectra unique to AMBN. *Top*, amino acid positions 39-50. ***Bottom***, amino acid positions 36-49. Coordinates refer to UniProt entry H2PDI5_PONAB.

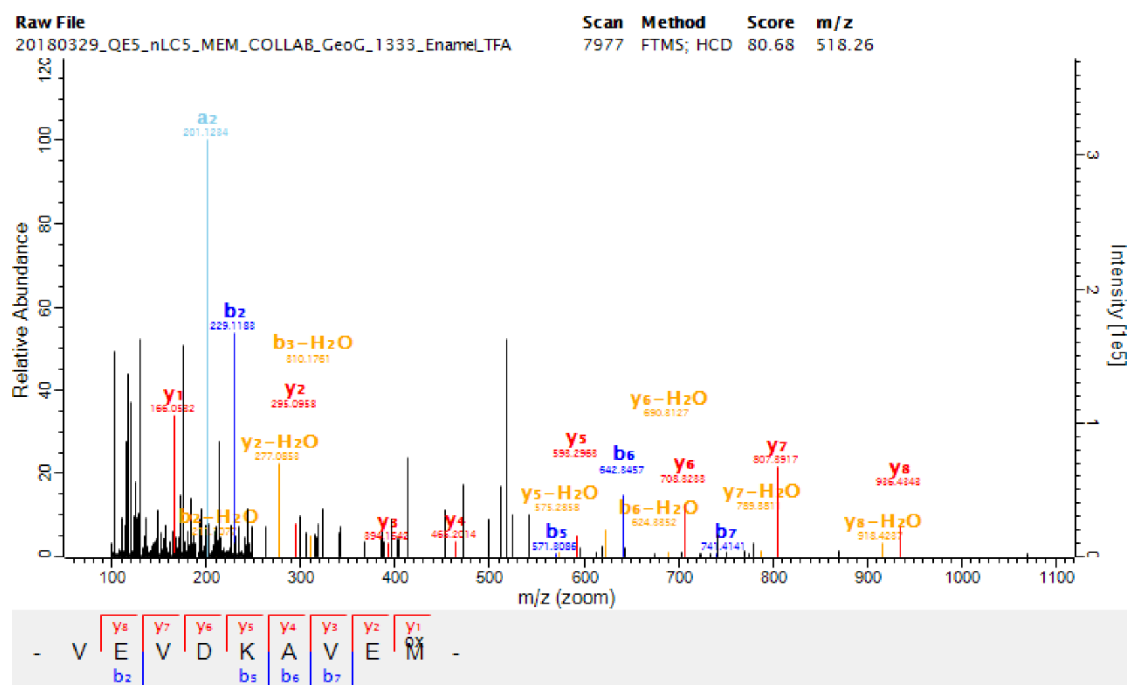




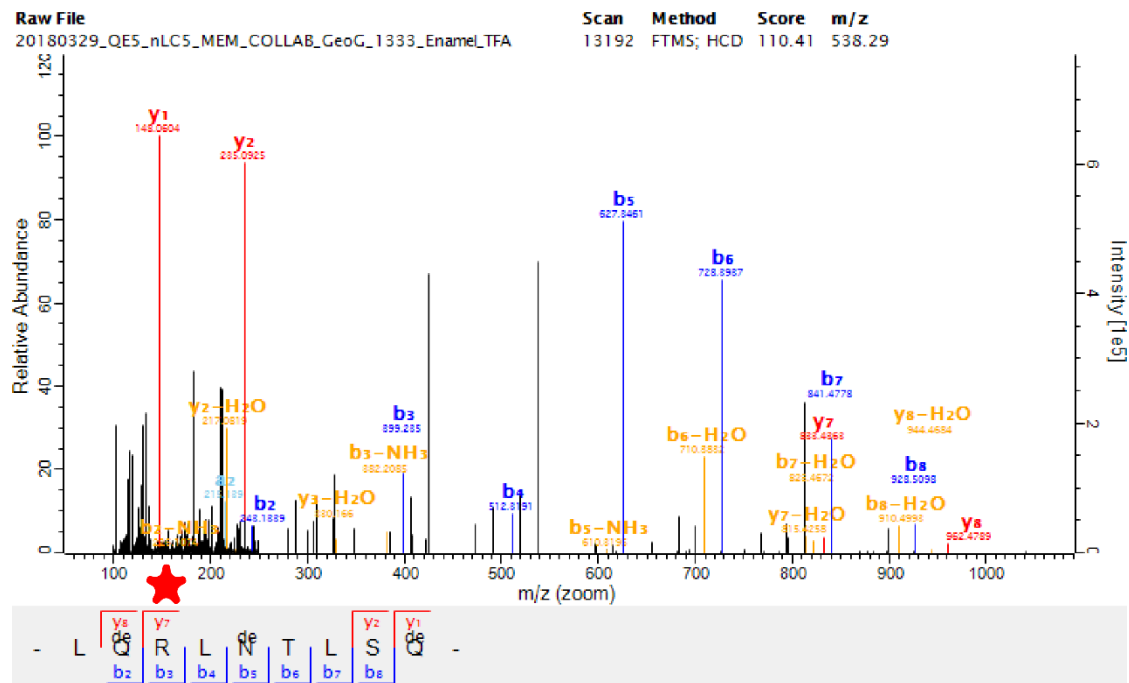
Supplementary Figure S6. Examples of annotated MS/MS spectra unique to AMTN. *Top*, amino acid positions 112-121. ***Bottom***, amino acid positions 112-121. Coordinates refer to UniProt entry H2PDI4_PONAB.



Supplementary Figure S7. Examples of annotated MS/MS spectra unique to ENAM. *Top*, amino acid positions 221-231. ***Bottom***, amino acid positions 180-190. Coordinates refer to UniProt entry H2PDI6_PONAB.



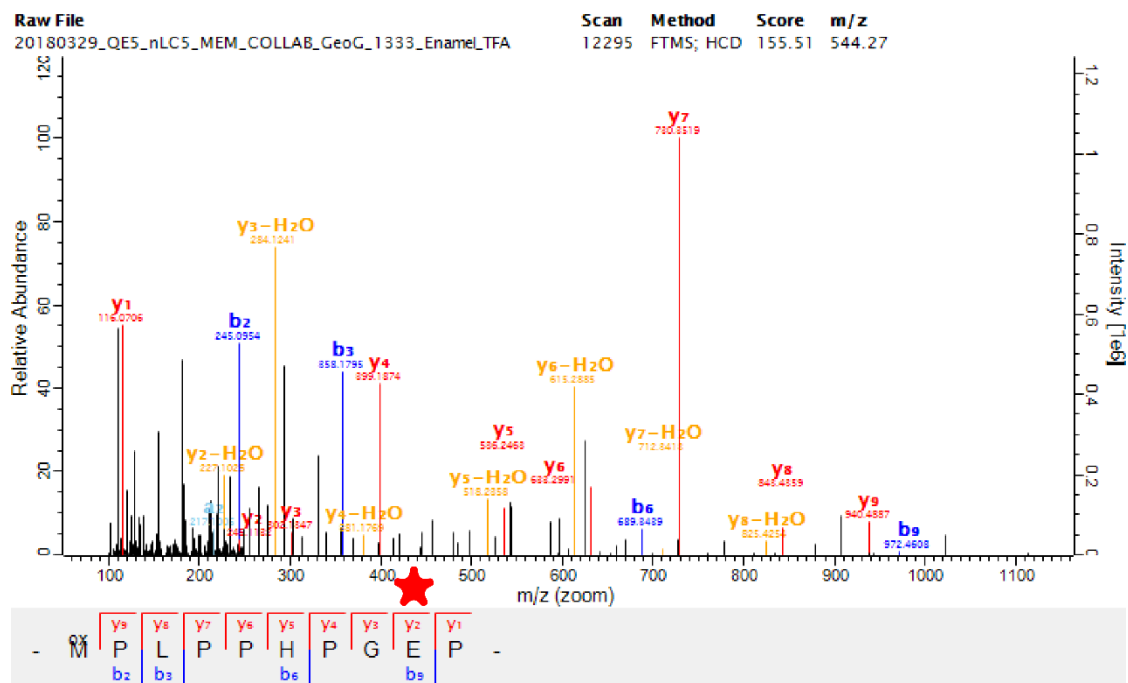
Supplementary Figure S8. Example of an annotated MS/MS spectrum unique to MMP20. Amino acid positions 136-144. Coordinates refer to UniProt entry H2NF32_PONAB.



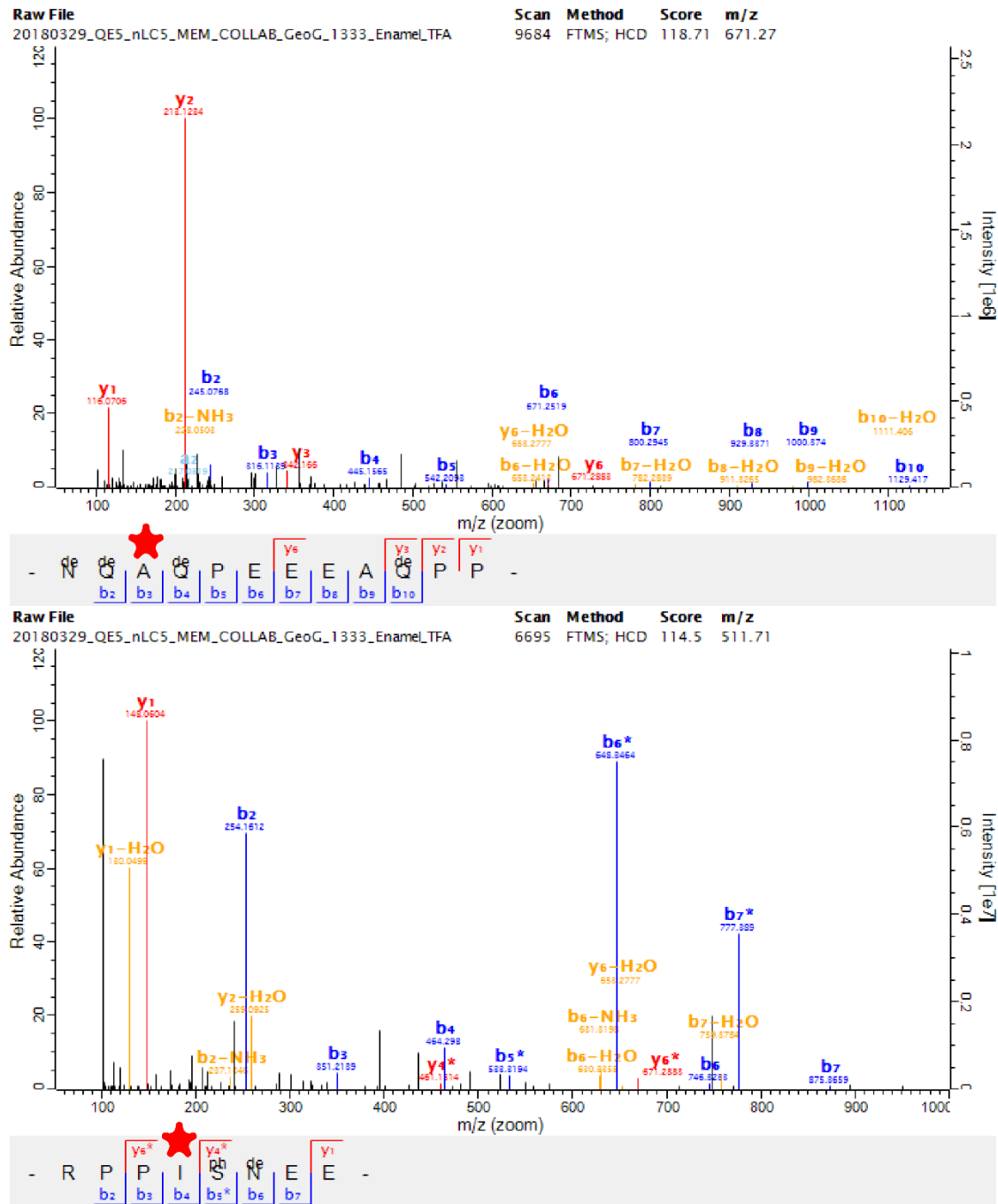
Supplementary Figure S9. Example an of annotated MS/MS spectrum of a phylogenetically informative

SAP in AMBN. Amino acid positions 53-61, with a Hominidae-derived “R” on amino acid position 55.

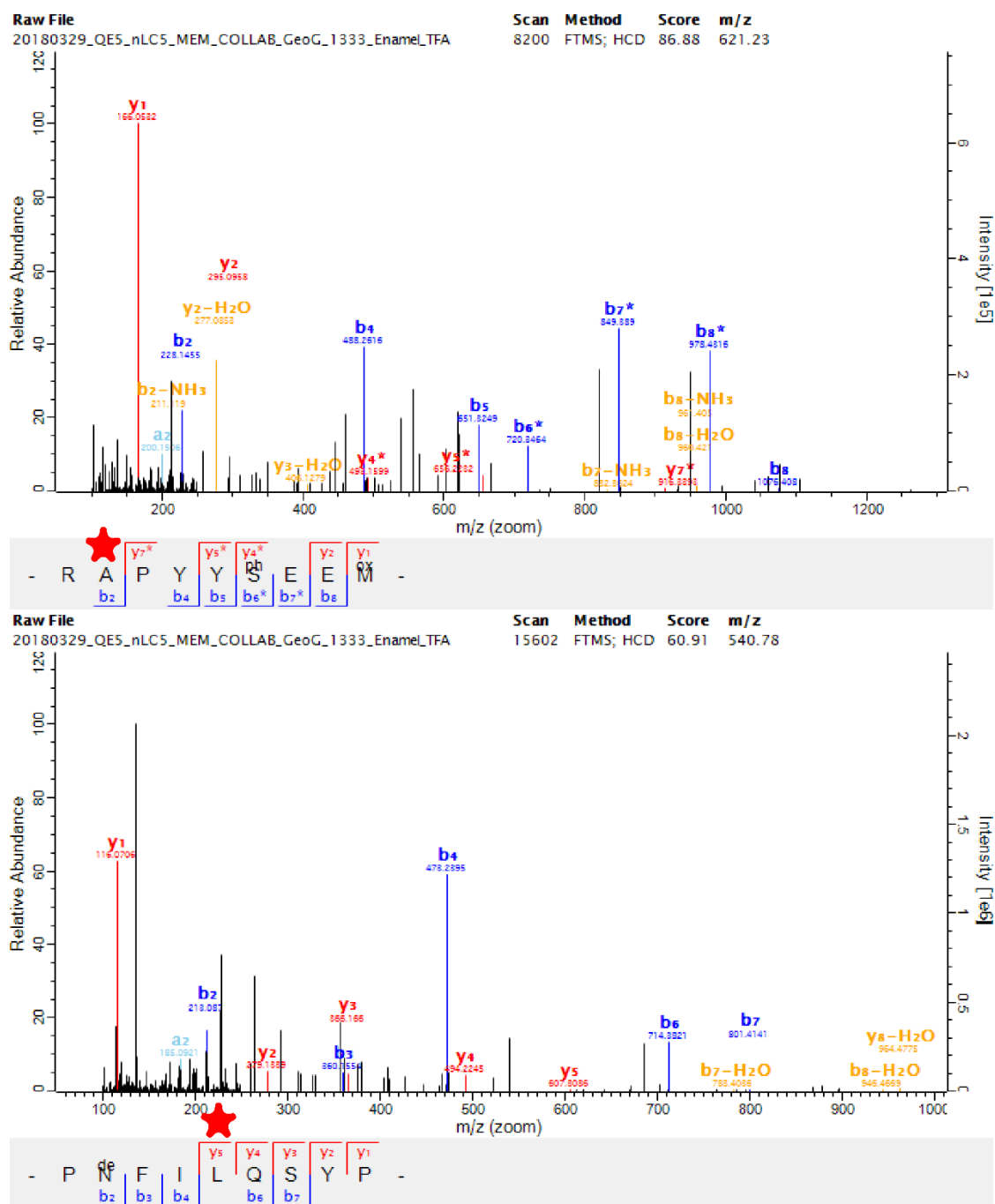
Coordinates refer to UniProt entry H2PDI5_PONAB.



Supplementary Figure S10. Examples of annotated MS/MS spectra of phylogenetically informative SAPs in AMELX. Amino acid positions 18-27, with a *Gigantopithecus*-derived “E” on amino acid position 26. Coordinates refer to UniProt entry H2PUX0_PONAB.



Supplementary Figure S11. Examples of annotated MS/MS spectra of phylogenetically informative SAPs in ENAM. *Top*, amino acid positions 145-156, with an ancestral “A” on amino acid position 147 (substituted to “P” in *Gorilla*, *Pan*, and *Homo*). ***Bottom***, amino acid positions 187-194, with a Hominidae-derived “I/L” on amino acid position 190. Coordinates refer to UniProt entry H2PDI6_PONAB.



Supplementary Figure S12. Examples of annotated MS/MS spectra of phylogenetically informative SAPs in ENAM. *Top*, amino acid positions 211-219, with a *Pongo*+*Gigantopithecus* derived “A” on amino acid position 212 (“A” in other Hominoids). ***Bottom***, amino acid positions 849-857, with a *Gigantopithecus*-derived “L” on amino acid position 853. Coordinates refer to UniProt entry H2PDI6_PONAB.