Protein sequence comparison of human and non-human primate tooth proteomes

Carine Froment, Clément Zanolli, Mathilde Hourset, Emmanuelle Mouton-Barbosa, Andreia Moreira, Odile Burlet-Schiltz, Catherine Mollereau

Protein sequence comparison of human and non-human primate tooth proteomes

Carine Froment[1], Clément Zanolli[2], Mathilde Hourset[3,4], Emmanuelle Mouton-Barbosa[1], Andreia Moreira[3], Odile Burlet-Schiltz[1] and Catherine Mollereau[3]

[1] Institut de Pharmacologie et Biologie Structurale (IPBS), Université de Toulouse, CNRS, UPS, Toulouse, France.

[2] Laboratoire PACEA, UMR 5199 CNRS, Université de Bordeaux, Pessac, France.

[3] Laboratoire d'Anthropobiologie Moléculaire et Imagerie de Synthèse (AMIS), UMR 5288 CNRS, Université de Toulouse, UPS, Toulouse, France

[4] Faculté de chirurgie dentaire de Toulouse, Université de Toulouse, UPS, Toulouse, France.

Corresponding authors:

Dr Catherine Mollereau catherine.mollereau-manaute@ipbs.fr

Laboratoire AMIS

Faculté de médecine, 37 allées Jules Guesde

31073 Toulouse Cedex 03, France

Tel : 33 561 14 55 13

and

Dr Odile Burlet-Schiltz, odile.schiltz@ibps.fr

IPBS

205, Route de Narbonne BP 64182

31077 Toulouse Cedex 04, France

Tel: 33 561 17 55 47

Keywords: Palaeoproteomics; Tooth; Primates; nanoLC-MS/MS; Taxonomy

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We are sincerely grateful to J. Braga (AMIS, Toulouse France), S. Jiquel (ISEM, Montpellier France) and R. Macchiarelli (HNHP, Poitiers France) for giving us access to their primate collections. We also thank the two colleagues that gave their surgically extracted molars. The work was supported by the French CNRS (PEPS blanc INEE 2016 and DefiXlife 2018-2019), in part by the Région Occitanie, European funds (Fonds Européens de Développement Régional, FEDER), Toulouse Métropole, and by the French Ministry of Research with the

Investissement d'Avenir Infrastructures Nationales en Biologie et Santé program (ProFI, Proteomics French Infrastructure project, ANR-10-INBS-08).

Abstract

In the context of human evolution, the study of proteins may overcome the limitation of the high degradation of ancient DNA over time to provide biomolecular information useful for the phylogenetic reconstruction of hominid taxa. In this study, we used a shotgun proteomics approach to compare the tooth proteomes of extant human and non-human primates (gorilla, chimpanzee, orangutan and baboon) in order to search for a panel of peptides able to discriminate between taxa and further help reconstructing the evolutionary relationships of fossil primates. Among the 25 proteins shared by the five genera datasets, we found a combination of peptides with sequence variations allowing to differentiate the hominid taxa in the proteins AHSG, AMBN, APOA1, BGN, C9, COL11A2, COL22A1, COL3A1, DSPP, F2, LUM, OMD, PCOLCE and SERPINA1. The phylogenetic tree confirms the placement of the samples in the appropriate genus branches. Altogether, the results provide experimental evidence that a shotgun proteomics approach on dental tissue has the potential to detect taxonomic variation, which is promising for future investigations of uncharacterized and/or fossil hominid/hominin specimens.

1. Introduction

In the last ten years, major advances have been achieved in the field of human evolution with the increasing recovery of Late Pleistocene human genomes. These advances have unveiled the existence of a new human group, the Denisovan hominins, an extinct relative of Neanderthals [1, 2], in addition to recurrent interbreeding between archaic and modern humans during their period of cohabitation in Eurasia [3, 4]. However, despite the exceptional retrieval of 0.4 million-year (My) old hominin genomic data from Sima de los Huesos in Spain [5, 6], the use of a DNA-based approach for older specimens, beyond the Middle Pleistocene [7], remains limited due to a high degree of fragmentation of ancient DNA over time [8, 9]. Therefore, the next challenging step is the biomolecular characterization of fossils further back in time, in particular for the Early Pleistocene-Late Pliocene key period (3-2 My ago) corresponding to the emergence of the genus *Homo*. Such invaluable missing data are essential for a more accurate reconstruction of the phylogenetic relationships between the different hominin taxa [10].

As an alternative to DNA, proteins appear to be more resistant to post-mortem damage, potentially allowing biomolecular investigation of the deeper time [11, 12]. Even in unfavorable conditions proteins survive longer than DNA, as demonstrated by the recovery of > 3 My old peptides entrapped in ostrich egg shell in the warm African environment [13] and more recently, of dental proteins from a subtropical Early Pleistocene specimen (1.9 My ago) representative of the extinct Asian hominid *Gigantopithecus blacki* [14]. Given the ever increasing performance of mass spectrometers, in terms of sensitivity, sequencing speed and resolution, it is now possible to study the protein content of ancient samples, using small amount of precious paleontological and paleoanthropological material. This opens the way to a new field of research: Palaeoproteomics. Proteins have been sequenced from Late to Early Pleistocene faunal and human fossils [7, 14-17], providing invaluable functional information on past human (patho)physiology [18-20] and ancient diets [21, 22]. In addition, phylogenetic reconstruction and taxonomic placement based on protein identification have also been possible for extinct species and/or when DNA was no longer available [16, 23-27]. The two main examples below on debated hominin fossils, have recently shed light on the relationships between human lineages during evolution. On the basis of a collagen variant, Chen et al. [24] were able to assign the mandible of Xiahe from the Tibetan plateau to a Denisovan individual. For the first time, the presence of this hominin outside the Siberian Altaï area was therefore demonstrated together with its anatomical relationship with other archaic hominin fossils in China. By analyzing the enamel proteome of a representative of *Homo antecessor* from Atapuerca (Spain) dated to 0.8-0.9 My ago, Welker et al. [27] have

suggested a very close affinity between this species and the last common ancestor of modern humans, Neanderthals and Denisovans.

Collagen sequences are widely used to identify or characterize specimens [9, 28, 29], but they are sometimes not enough informative because of a high conservation between species limiting accurate phylogenetic reconstruction [30]. It is therefore necessary to identify other proteins with sufficient taxonomic variability. In this context, the tooth proteome is expected to be more informative than the bone proteome [9, 14, 16, 27]. Indeed, in addition to collagens, it also contains a variety of specific non-collagenous proteins with taxonomic interest, including the X and Y forms of amelogenin able to provide information about the sex of ancient individuals [27, 31-34].

In the present study we used a shotgun proteomics approach to compare the tooth proteomes of modern humans from the 21th century and non-human primates (gorilla, chimpanzee, orangutan, baboon) from the 19th-mid 20th century, in order to search for a panel of peptides able to discriminate between taxa. We focused on the analysis of the proteins shared by the five genera datasets to allow for comparison. We found in 14 proteins a combination of peptides with taxonomic variation potentially useful for future studies on uncharacterized and/or older hominid/hominin specimens.

2. Materials and Methods

2.1 Samples

Non-human primate teeth (2 individuals per genus) from 19th-mid 20th century collections were obtained from the University of Poitiers with the agreement of Pr R. Macchiarelli (samples T1: *Gorilla,* T2: *Papio*, T5: *Pan.*, T6: *Pan*), from the University of Toulouse with the agreement of Pr J. Braga (samples T3: *Papio*, T4: *Gorilla*), and from the University of Montpellier with the agreement of S.Jiquel (samples T7: *Pongo*, T8: *Pongo*). Except for the two chimpanzee specimens that are attributed to *Pan troglodytes*, the species of the other non-human primate samples is not known. Human teeth were obtained from two colleagues who kindly donated their own M3 that had been surgically extracted three (LOS2) and fifteen (LKS2) years ago. Description of the teeth is given in Table S1. Teeth were handled under a laminar flow hood and the operator was equipped with disposable clothes. The surface of the teeth was carefully cleaned with a solution of 1% SDS and abundantly rinsed with sterile pure water. The tooth area to be sampled, (preferably chosen on the cervical part of the crown so that the damaged area is minimally visible after repositioning the tooth on the jaw) was first manually drilled with a micromotor STRONG 207-106 (Pouget-Pellerin, France) to abrade the surface. Then a spherical carbide drill (1-1.2 mm in diameter) was used to pierce the enamel through the dentine tissue, over a sterile microtube to collect the powder. Sample amounts ranged from 2 to 26 mg for non-human primates and 60 mg for

humans, respectively (Table S1). Collected powder was immediately processed for protein extraction.

2.2. Protein extraction and Trypsin digestion

Samples were prepared as five independent series, each including an extraction blank with no material (Blk) that was processed in the same way as the tooth samples. Protein extraction was performed by using the filter-aided sample preparation (FASP) protocol described in [31]. It was slightly modified to include an additional step for recovering the flow-through content of the first Amicon™ Ultra-4 (10kDa) filter unit (Merck Millipore) after centrifugation which may contain material of interest (Fig. 1). After a demineralization step in 0.5 M EDTA pH 8 for 18 h at room temperature under rotation, the pellet was extracted in 0.1 M Tris pH 8, 0.1 M DTT, 4% SDS for 2h at 60 °C. The supernatants from the demineralization and extraction steps were mixed with 8 M urea in 0.1 M Tris pH 8 and ultra-filtered through an Amicon™ Ultra-4 (10kDa) centrifugal filter unit (4000g, swinging rotor, room temperature). The flow-through (except for LOS2) was collected and ultra-filtered through an Amicon™ Ultra-4 (3kDa) centrifugal filter unit to recover smaller protein fragments excluded from the 10kDa filtration. The two filtration units (giving at end samples referred as T and Tpep, respectively) were then similarly processed. After a wash with 2 ml of 8 M urea in 0.1M Tris pH 8, protein alkylation (50 mM 2-Chloroacetamide in 8M urea, 0.1 M Tris pH8) was performed on the filter units for 20-30 min at room temperature in the dark. The units were then washed (2 x 1 ml) with 8M urea in 0.1 M Tris pH 8, followed by 50 mM ammonium bicarbonate washes (1 x 1 ml, 1 x 0.5 ml). Proteins (T samples) and peptides (Tpep samples) retained on the filter were dissolved in 50 mM ammonium bicarbonate and an aliquot was harvested for quantification using the Qubit protein assay kit (Thermo Fisher Scientific). They were digested by overnight incubation at 37 °C with 2 µg sequencing grade modified porcine trypsin (Promega). The digestion was prolonged the next day for 4-6 h with 2 µg additional trypsin. The tryptic peptide mixtures were recovered by centrifugation over a new tube. The centrifugates were then transferred to microtubes, dried by using a centrifugal vacuum concentrator and kept at -20°C until mass spectrometry analysis.

EDTA, Tris, and SDS were purchased from Invitrogen, urea and ammonium bicarbonate from Acros Organics, chloroacetamide from Sigma-Aldrich.

2.3. nanoLC-MS/MS analysis

The dried peptides were resuspended with 0.05% trifluoroacetic acid in 2% acetonitrile at an estimated concentration of 1µg/µl based on protein quantification, and then analyzed by online nanoLC using an UltiMate® 3000 RSLCnano LC system (Thermo Scientific, Dionex) coupled to an Orbitrap Fusion Tribrid™ mass spectrometer (Thermo Scientific, Bremen, Germany). 1µl of the samples were loaded on a 300 µm ID x 5 mm PepMap C18

pre-column (Thermo Scientific, Dionex) at 20 µl/min in 2% acetonitrile, 0.05% trifluoroacetic acid. After 5 minutes of desalting, peptides were on-line separated on a 75 µm ID x 50 cm C18 column (in-house packed with Reprosil C18-AQ Pur 3 µm resin, Dr. Maisch, and equilibrated in 95% of buffer A (0.2% formic acid)) with a gradient of 5 to 25% of buffer B (80% acetonitrile, 0.2% formic acid) for 80 min then 25% to 50% for 30 min at a flow rate of 300 nL/min.

The instrument was operated in the data-dependent acquisition (DDA) mode using a top-speed approach (cycle time of 3s). The survey scans MS were performed in the Orbitrap over m/z 350-1550 with a resolution of 120,000 (at 200 m/z), an automatic gain control (AGC) target value of 4e5, and a maximum injection time of 50 ms. Most intense ions per survey scan were selected at 1.6 m/z with the quadrupole and fragmented by Higher Energy Collisional Dissociation (HCD). The monoisotopic precursor selection was turned on, the intensity threshold for fragmentation was set to 50,000 and the normalized collision energy was set to 35%. The resulting fragments were analyzed in the Orbitrap with a resolution of 30,000 (at 200 m/z), an automatic gain control (AGC) target value of 5e4, and a maximum injection time of 60 ms. The dynamic exclusion duration was set to 30 s with a 10 ppm tolerance around the selected precursor and its isotopes. For internal calibration the 445.120025 ion was used as lock mass.

Each sample was subjected to two independent LC-MS/MS runs (TR1, TR2) for assessing the identification reproducibility. To control for carry-over contamination, the MS workflow process included a washing step followed by two blank MS runs using gradient conditions similar to those of the samples and performed before and after each sample MS run (including the blank samples).

2.4. Bioinformatics analysis of nanoLC-MS/MS data

All raw mass spectrometry files were processed in parallel using two different protein identification softwares: Proteome Discoverer™ software 2.3.0.523 (Thermo Fischer Scientific) with Mascot 2.6.2 (Matrix Science, London, UK) combined with the Percolator algorithm (version 2.05) for PSM search optimization, and PEAKS™ Studio 10.0 software (Bioinformatics Solutions Inc., Waterloo, ON, Canada) using the full set of available processes PEAKS de novo > PEAKS DB > PEAKS PTM> PEAKS SPIDER [35, 36]. For both softwares, data obtained from T and Tpep samples were searched against the UniProtKB Swiss-Prot and TrEMBL protein databases (including canonical and isoform sequences, and supplemented with frequently observed contaminants) corresponding to their own taxon: Uniprot_isoF_Human database released 2019_09 with *Homo sapiens* taxonomy (195349 sequences), Uniprot_isoF_Gorilla database released 2019_07 with *Gorilla* taxonomy (46070 sequences), Uniprot_isoF_Pan database released 2019_06 with *Pan* (chimpanzees)

taxonomy (154055 sequences), Uniprot_isoF_Pongo released 2019_02 with *Pongo* (orangutan) taxonomy (96058 sequences), Uniprot_isoF_Papio released 2019_02 with *Papio* (baboons) taxonomy (46692 sequences). Data obtained from the blank samples were searched against the five databases.

For Proteome Discoverer analysis, Mascot database searches were performed individually for each raw file using a processing workflow consisting of the following parameters: mass tolerances in MS and MS/MS were set to 10 ppm and 0.02 Da, respectively. Carbamidomethylation of cysteine was set as a fixed modification. The enzyme specificity was selected as semi-tryptic, with a maximum of three missed cleavages. The main protein modifications commonly observed in damaged and ancient proteins were set as variable modifications: deamidation (N, Q), oxidation (M, P), carbamylation (K, N-terminal protein), and conversion to pyro-glutamic acid (N-terminal Q). The Percolator algorithm was used to validate PSMs and peptides based on Posterior Error Probability (PEP) values at a FDR ≤ 1% [37, 38]. FDR was estimated by a target-decoy approach using the reversed database. Afterwards, the processing workflow results (.msf files) were combined into sample (technical replicates TR1 + TR2), genus (technical replicates TR1 + TR2, T and Tpep samples) or extraction blank multiconsensus reports. Each resulting dataset was then filtered using a consensus workflow consisting of the following parameters: Only PSMs with rank 1 and Mascot ion score ≥ 20 were considered. Peptide identifications were grouped into proteins according to the law of parsimony and filtered to 5% FDR.

For PEAKS analysis, all the raw files belonging to the same genus (technical replicates TR1 + TR2, T and Tpep samples) or to the extraction blanks were loaded into a single identification workflow per genus or per extraction blank and processed using the following parameters: mass tolerances in MS and MS/MS were set to 10 ppm and 0.02 Da, respectively. Carbamidomethylation of cysteine was set as a fixed modification, deamidation (N, Q) and oxidation (M, P) as variable modifications. A maximum of 3 modifications per peptide was allowed. In the PEAKS PTM module, all the 313 modifications and also a maximum of 3 modifications per peptide were considered, and the validation was based on an average local confidence (ALC) score ≥ 15%. Trypsin with semi-specific digest mode and a maximum of three missed cleavages were selected. Finally, each resulting dataset per genus and per extraction blank obtained from the PEAKS SPIDER module, was filtered and exported using the following threshold values: Peptide score of -10lgP ≥ 20-22 adjusted to obtain a FDR ≤ 1% for PSMs and peptides, Protein score of -10lgP ≥ 25-49 adjusted to obtain a FDR ≤ 5% without taking into account the criterion of unique peptide and only considering significant peptides, and *de novo* only ALC (%) ≥ 50. FDR was estimated by the PEAKS “decoy-fusion” approach.

For each analysis, the proteins marked as contaminant or found in the extraction blanks were excluded from the datasets analyzed. All the MS/MS spectra of the taxon-specific peptides were inspected manually. The species specificity of peptides with taxonomic interest were checked by a protein Blast (BlastP) search in Uniprot (https://www.uniprot.org/) and NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

All the RAW data files, the output files corresponding to Proteome Discoverer™ (.msf files) and PEAKS (.csv files and HTML reports) analyses, and all the fasta files used for the database searches have been deposited on the ProteomeXchange Consortium [PMID 24727771] via the PRIDE partner repository and can be accessed with the dataset identifier PXD018933.

2.5. Proteomic data analysis

To be able to compare proteomes between genera, the accession numbers of the proteins identified in the different taxa were converted to human gene names by using the Retrieve ID functions of Uniprot (https://www.uniprot.org/), or the db2db conversion tool of bioDBnet (https://biodbnet-abcc.ncifcrf.gov/), and/or by BLAST alignment and manual inspection. Classification of the identified proteins into functional categories according to GO terms was performed by using the GSEA software at https://www.gsea-msigdb.org/gsea/msigdb/annotate.jsp [39]. Graphic representations and statistical analysis of the data were performed by using Prism 7 (GraphPad Software Inc., USA). Online tool found at http://bioinformatics.psb.ugent.be/webtools/Venn/ was used to produce the Venn diagrams.

2.6 Phylogenetic analysis

The phylogenetic tree was based on the alignment of 14 proteins (AHSG, AMBN, APOA1, BGN, C9, COL11A2, COL22A1, COL3A1, DSPP, F2, LUM, OMD, PCOLCE, SERPINA1). The concatenated protein sequences identified in each genus sample were aligned to the concatenated corresponding protein sequences from 8 hominoids (*Homo sapiens*, *Gorilla gorilla*, *G.gorilla gorilla*, *Pan troglodytes*, *Pan paniscus, Pongo abelii*, *Pongo pygmaeus*, *Nomascus leucogenys*) and 3 Papionini (*Papio anubis*, *Papio hamadryas*, *Mandrillus leucophaeus*) by using the Catfasta2phyml tool available at https://github.com/nylander/catfasta2phyml. The concatenated alignment is given in the supplementary Phylotreealign.fasta file. The best-fit model for generating the phylogenetic tree was selected based on the Akaike's information criterion corrected (AICc) implemented in jModelTest 2 [40]. The tree was built using RAxML-NG [41] with a bootstrapping procedure (X1000) as statistical test for branch support.

3. Results

3.1. Comparative analysis of human, gorilla, chimpanzee, orangutan and baboon tooth proteomes.

For each genus, two types of samples (T and Tpep) were prepared (Fig. 1). The T samples correspond to proteins retained on the filter unit of the FASP protocol and usually analyzed in proteomic experiments. The Tpep samples were recovered from the elution of the filter units that is usually discarded but might also contain material of interest such as peptides or protein fragments present in the demineralized/lysed extracts. The samples from 2 individuals per genus (*Homo*, *Gorilla*, *Pan*, *Pongo* and *Papio)* were analyzed in duplicate by using a nanoLC-MS/MS data-dependent analysis of the tryptic digestions. The Proteome Discoverer software was used with the Mascot search engine for protein identification in genus-specific protein databases. This led to identification of 32 to 172 proteins per T samples, with only a few additional ones identified in Tpep samples (Fig. S1A and S1B).

The total number of proteins identified per taxon by Proteome Discoverer ranged from 33 in chimpanzee (*Pan*) to 228 in human (Fig. S1C and Table S2 for the raw list of proteins). To allow for the comparison between the five proteomes, the protein accessions were converted into the corresponding human gene names. This resulted in a reduced number of identification (Fig.S1C) because of redundant proteins corresponding to multiple accession numbers (isoforms, incomplete sequence, etc.) for a same gene. The analysis of the distribution of proteins between the five proteomes (Fig. 2A) indicates that 25 proteins are common to the five genera (Table 1). Less than 15% of proteins are unique to each taxon, with the exception of *Homo* (45%) and *Papio* (25%). A gene set overlap analysis according to GO annotation terms indicates that the Top 20 most significant gene sets are different between the common and the exclusive pooled proteins (Fig.2B). While shared proteins are annotated with terms related to extracellular matrix organization/ossification and blood/wound healing, 20% to 46% of the proteins unique in the *Homo*, *Gorilla*, *Pongo* and *Papio* samples are significantly associated with immunity, in addition to extracellular matrix organization and coagulation (Fig.2B, Table S3). Identifying proteins involved in the immune system in the variable proteome is not unexpected since it may reflect the different sanitary status between taxa and/or individuals, in addition to a possible genus-specific expression of proteins [42].

3.2. Analysis of the proteins of interest for taxonomic discrimination

A comparative analysis of the peptide lists from the proteins common to the five primate proteomes was performed to search for the presence of taxon-specific peptides that could be useful markers for specimen identification. Compared to the only use of protein database search engines, *de novo* peptide sequencing algorithms that do not require a

protein sequence database [35], or error-tolerant search algorithms that utilize protein sequence databases while allowing sequence deviation [43], represent powerful approaches in palaeoproteomics to allow for the identification of novel amino acid substitutions. Therefore, to potentially improve the identification of taxon-specific peptides, an additional round of bioinformatic data analysis was performed using PEAKS software which uses a *de novo*-assisted database search algorithm to maximize the peptide identification efficiency [35, 36]. As shown in Fig.S3A, PEAKS yielded slightly more protein identifications than Proteome Discoverer (Fig. S1C), with a 40-72% overlap of the identifications between the two datasets (Fig. S3B) if considering gene name correspondence rather than protein accessions to avoid protein redundancy.

The 25 proteins found in the five primate samples, especially the collagens, were mainly identified by peptides covering regions of highly conserved sequences. However, a number of peptides showing amino acid variations between taxa were identified in 14 proteins (AHSG, AMBN, APOA1, BGN, C9, COL11A2, COL22A1, COL3A1, DSPP, F2, LUM, OMD, PCOLCE, SERPINA1). The lists of all the peptides identified for each protein in each dataset per genus with Proteome Discoverer or PEAKS softwares are given in Tables S4 to S17. The phylogenetic tree based on the concatenated alignment of the 14 proteins confirms the placement of the five samples in the appropriate genus branches (Fig. 3). Because of missing protein sequences in the database for some species, in particular *G. gorilla*, *P. pygmaeus*, *P. hamadryas*, and as no peptide strictly specific to a species was identified in the dataset, with the exception of *H. sapiens* (Table 2), the taxonomic discrimination at the level of the species is less guaranteed. The main representative peptides showing amino acid substitutions able to discriminate between the hominids taxa are presented in Table 2, after the validation of their specificity by a BlastP search. The HCD MS/MS spectra are shown in Fig. 4 and 5 and Fig. S4. Peptides specific to the taxa *Homo*, *Gorilla* and *Pongo* were identified. No *Pan*-specific peptides were detected, probably because of the lower protein coverage in this sample (see discussion). Despite the phylogenetic distance and a high number of amino acid variations with respect to the four other primates, no *Papio*-specific peptide was found (Tables S4-S17).

The protein displaying the most diversity is alpha-2-HS-glycoprotein/Fetuin A (AHSG) which is well covered in the five genera (Table 1). A number of discriminating peptides were identified in this protein (Tables 2 and S4). Peptides covering AHSG-[72-99] with a lysine residue at the position 99 are specific to *Gorilla* (Fig. 4A) while the peptides covering AHSG-[104-117] with a lysine residue at the position 117 and AHSG-[328-337] with a leucine residue at the position 329 are specific to *Pongo* (Fig. 4B and 4C). Interestingly, a particular combination of amino acids in the peptides covering AHSG-[318-337] specifically differentiates *Homo* (P329/V333) and *Pongo* (L329/V333) from the two other taxa

(P329/A333) (Fig. 4C, and Table 2). In addition, the position 45 and 52 in the peptide AHSG-[29-57] allows to distinguish *Homo* and *Gorilla* (I45/L52) from *Pan* and Pongo (I45/H52) and from *Papio* (V45/L52) (Fig.4D and Tables 2 and S4). Among the other proteins, *Homo*-specific peptides are identified in the F2 protein (F2-[199-217]), *Gorilla*-specific peptides are identified in the proteins COL3A1 (COL3A1-[351-368] and -[902-923]), DSPP (DSPP-[403-411]) and SERPINA1 (SERPINA1-[126-149]), and *Pongo*-specific peptides are identified in the proteins APOA1 (APOA1-[185-195]) and BGN (BGN-[220-239]) (Table 2 and Fig. S4). Beside the taxon-specific peptides, peptides with a variable signature or a combination of variable amino acids among taxa, may be also informative. For example, peptides covering PCOLCE-[305-320] distinguish *Homo* and *Pan* (S309) from *Gorilla* and *Pongo* (T308) (Table 2 and Fig. 5). Peptides C9-[232-242] bearing A238, C9-[546-558] bearing I546/E554/N557 and DSPP-[362-370] bearing T365/A367 are specific to the Homininae taxon (Table 2 and Fig. S4). The peptide DSPP-[56-66] with a leucine residue at the position 60 is specific to *Homo* and *Pan*. Taken together, the data demonstrate therefore that it is possible by using a MS-based proteomics approach on dental tissue to identify a combination of peptides enabling the distinction between members of the four hominid genera (*Homo*, *Gorilla*, *Pan* and *Pongo*) in accordance with the phylogenetic tree (Fig. 3).

4. Discussion

The present study is, to our knowledge, the first comparative analysis of tooth proteomes from five living primate genera, including one cercopithecidae (*Papio*) and the four extant hominids (*Homo*, *Pan*, *Gorilla*, *Pongo*). It is noteworthy that we had to deal with the incomplete genomic annotation of non-human primates to convert protein accessions from each species into the corresponding canonical (human) gene names, a prerequisite for allowing comparison between taxa. However, even if a few mis-conversions or unrecognitions might still remain, a total of 312 proteins corresponding to the pooled proteins from all samples across all genera after removing of duplicates, were identified (Fig. S5A). 67% belong to dentine tissue, compared to the recently reported comprehensive human dentine proteome [44]. The other 30% proteins include the amelogenin protein specific to the enamel tissue, diverse collagens, in addition to constituents of the extracellular matrix and immune system proteins (Fig S5B). The later components may reflect the dynamic and heterogeneous part of dentine, a tissue rich in diverse bioactive peptides involved in host defense, regenerative process, angiogenesis, growth and differentiation [44-47]. These are expected to vary between species or individuals depending on health status and/or traumatic injury [42, 47, 48]. The number of proteins identified was lower in chimpanzee compared to the other specimens, and this was not due to a smaller tooth sampling (same amount as in

orangutan, Table S1). Without excluding a possible poor preservation of proteins in the chimpanzee specimens, or a less efficient protein extraction, the low number of proteins could also reflect a different level of protein expression in the dental tissue in this taxon [49]. Indeed, the chimpanzee has a particularly thin enamel which is more prone to tooth damage than in the other great apes [50]. Another explanation could be related to the tooth types that were sampled. While for all the other specimens, enamel and dentine powder were collected from permanent teeth, for the two chimpanzees we could only sample their deciduous teeth. The nature and degree of protein expression might differ between permanent and deciduous teeth, although this does not appear to be the case in humans [51] and this will need to be investigated further.

All the proteins common to the five proteomes, with the exception of COL10A1 and COL23 A1, have already been identified in human dentine extracts [44, 46, 52], although AMBN may also be derived from the enamel [45, 53]. COL22A1 and F2 have been detected at the enamel-dentine junction (EDJ), an interface between the mineralized tissues involved in mechanical load [54]. The analysis of the peptides in the proteins common to the five proteomes allowed for the identification of sequence variations in AHSG, AMBN, APOA1, BGN, C9, COL11A2, COL22A1, COL3A1, DSPP, F2, LUM, OMD, PCOLCE and SERPINA1 enabling a taxonomic placement at the genus level (Fig.3 and Table 2). However, no species marker was detected, excepted for the species *H.sapiens*. Interestingly, nine of the proteins have been detected in 5000-year-old bovine dentin samples [55]. These include AHSG also identified in the 1.9 My old enamel of the extinct primate *Gigantopithecus blacki* [14] and COL3A1 identified in the dentine from the Xiahe specimen attributed to Denisova [24]. Therefore, some of the proteins described here and showing a potential taxonomic interest survive in time and to fossilization processes.

AHSG has already been reported to be resistant to degradation and to display enough sequence variation to be of interest for phylogenetic studies [9, 12]. Here, although the protein was identified by 3 peptides in only 2 out of 20 blank replicates (Table S2), it was identified by a larger number of peptides in the samples (Table 1) suggesting a probable endogenous origin of the protein in the samples. AHSG was therefore kept in the analysis to make a comparison with the data from the literature on ancient specimens. Among the peptides identified in AHSG, those bearing amino acids K-99 or L-329 are specific to *Gorilla* and *Pongo*, respectively. In addition, the combination of P-329 and V-333 is specific to the *Homo* taxon. In *Gigantopithecus blacki* [14] only one peptide was identified in a highly conserved region of the protein (AHSG-[133-145]) susceptible to contamination in our analysis (Table S4). AMBN was poorly covered in our samples (Tables 1 and S5) and not detected in the dentine of the Xiahe Denisovan [24], in contrast to the other paleontological specimens sampled from dental enamel [14, 27]. This suggests that using enamel tissue to

extract this protein is more appropriate than using dentine alone, or both dentine and enamel. Similarly to ancient samples (*Homo antecessor*, *Homo erectus*, *Gigantopithecus blacki*), the N-terminal part of the AMBN including the substitutions S34T and R55G that distinguish *Papio* from the other hominid taxa were covered in our samples (Table S5). However, no peptides were identified in the region overlapping a combination of amino acid substitutions that differentiates *Pongo* (V264/G269) and *Papio* (M265/G170) from the others taxa (V265/E270), and which indeed helps to affiliate *Gigantopithecus* to the pongine clade. The dentine proteome of the Denisovan specimen from Xiahe exclusively contains collagens. Interestingly, all the peptides identified in the ancient COL3A1 were also identified in our samples (Table S11). In particular, they include the peptides covering the substitution A364V specific to *Gorilla*, and the substitution S796G that differentiates *Homo*/*Gorilla*/*Pan* (S) from *Pongo*/*Papio* (G).

A number of taxon-specific peptides were identified in the other dentine proteins (Table 2), with the exception of *Pan* due to the low protein coverage of the chimpanzee samples, and probably also because of the high sequence homology with *Homo* (lower bootstrap at the hominin node, Fig.3). Peptides with taxonomic specificity were generally identified together with the corresponding peptides in the other taxa proteins (Fig. 4, Fig. 5, Fig. S4 and Tables S4-S17). As some positions were also covered in ancient samples, the results support the potential of a MS-based proteomics approaches for protein identification and taxonomic discrimination of extant and fossil primates from tooth samples.

In conclusion, the present comparison between human and non-human primates tooth proteomes shows that a shotgun proteomics approach on dental tissue has the potential to discriminate between the hominid taxa *Homo*, *Gorilla*, *Pan* and *Pongo,* despite a high protein sequence homology (Fig.3). The results also suggest that dentine proteins offer informative variability. However the data highlight the limitation of the method to differentiate individual species. A targeted MS-based approach using a combination of the peptides identified in this study, especially in AHSG, APOA1, BGN, COL3A1, DSPP, F2 and PCOLCE, could be applied for further in-depth taxonomic investigations of ancient samples, as previously done with amelogenin peptides for sex estimation [31]. In light of the recent papers on Pleistocene specimens [14, 27] a promising way is open to characterize the taxonomic attribution and phylogenetic relationships of fossil hominid remains, notably for those older than the Middle Pleistocene for which DNA information may not be preserved or not retrievable with the currently available methods.

References

[1] Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012;338(6104):222-6. Epub 2012/09/01. doi: 10.1126/science.1224344. PubMed PMID: 22936568; PubMed Central PMCID: PMC3617501.

[2] Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505(7481):43-9. Epub 2013/12/20. doi: 10.1038/nature12886. PubMed PMID: 24352235; PubMed Central PMCID: PMC4031459.

[3] Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature. 2015;524(7564):216-9. doi: 10.1038/nature14558. PubMed PMID: 26098372; PubMed Central PMCID: PMCPMC4537386.

[4] Slon V, Mafessoni F, Vernot B, de Filippo C, Grote S, Viola B, et al. The genome of the offspring of a Neanderthal mother and a Denisovan father. Nature. 2018. doi: 10.1038/s41586-018-0455-x. PubMed PMID: 30135579.

[5] Meyer M, Arsuaga JL, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature. 2016;531(7595):504-7. doi: 10.1038/nature17405. PubMed PMID: 26976447.

[6] Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga JL, et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature. 2014;505(7483):403-6. doi: 10.1038/nature12788. PubMed PMID: 24305051.

[7] Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature. 2013;499(7456):74-8. Epub 2013/06/28. doi: 10.1038/nature12323. PubMed PMID: 23803765.

[8] Cappellini E, Collins MJ, Gilbert MT. Biochemistry. Unlocking ancient protein palimpsests. Science. 2014;343(6177):1320-2. Epub 2014/03/22. doi: 10.1126/science.1249274. PubMed PMID: 24653025.

[9] Welker F. Palaeoproteomics for human evolution studies. Quaternary Science Reviews. 2018;190:137-47.

[10] Herries AIR, Martin JM, Leece AB, Adams JW, Boschian G, Joannes-Boyau R, et al. Contemporaneity of Australopithecus, Paranthropus, and early Homo erectus in South Africa. Science. 2020;368(6486). doi: 10.1126/science.aaw7293. PubMed PMID: 32241925.

[11] Schweitzer MH, Schroeter ER, Goshe MB. Protein molecular data from ancient (>1 million years old) fossil material: pitfalls, possibilities and grand challenges. Anal Chem.

2014;86(14):6731-40. Epub 2014/07/02. doi: 10.1021/ac500803w. PubMed PMID: 24983800.
[12] Wadsworth C, Buckley M. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. Rapid communications in mass spectrometry : RCM. 2014;28(6):605-15. Epub 2014/02/13. doi: 10.1002/rcm.6821. PubMed PMID: 24519823.
[13] Demarchi B, Hall S, Roncal-Herrero T, Freeman CL, Woolley J, Crisp MK, et al. Protein sequences bound to mineral surfaces persist into deep time. Elife. 2016;5. doi: 10.7554/eLife.17092. PubMed PMID: 27668515; PubMed Central PMCID: PMCPMC5039028.
[14] Welker F, Ramos-Madrigal J, Kuhlwilm M, Liao W, Gutenbrunner P, de Manuel M, et al. Enamel proteome shows that Gigantopithecus was an early diverging pongine. Nature. 2019;576(7786):262-5. doi: 10.1038/s41586-019-1728-8. PubMed PMID: 31723270; PubMed Central PMCID: PMCPMC6908745.
[15] Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RA, Stafford TW, et al. Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. Journal of proteome research. 2012;11(2):917-26. Epub 2011/11/23. doi: 10.1021/pr200721u. PubMed PMID: 22103443.
[16] Cappellini E, Welker F, Pandolfi L, Ramos-Madrigal J, Samodova D, Ruther PL, et al. Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. Nature. 2019;574(7776):103-7. doi: 10.1038/s41586-019-1555-y. PubMed PMID: 31511700; PubMed Central PMCID: PMCPMC6894936.
[17] Nielsen-Marsh CM, Richards MP, Hauschka PV, Thomas-Oates JE, Trinkaus E, Pettitt PB, et al. Osteocalcin protein sequences of Neanderthals and modern primates. Proc Natl Acad Sci U S A. 2005;102(12):4409-13. Epub 2005/03/09. doi: 10.1073/pnas.0500450102. PubMed PMID: 15753298; PubMed Central PMCID: PMC555519.
[18] Jersie-Christensen RR, Lanigan LT, Lyon D, Mackie M, Belstrom D, Kelstrup CD, et al. Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. Nature communications. 2018;9(1):4744. doi: 10.1038/s41467-018-07148-3. PubMed PMID: 30459334; PubMed Central PMCID: PMCPMC6246597.
[19] Maixner F, Overath T, Linke D, Janko M, Guerriero G, van den Berg BH, et al. Paleoproteomic study of the Iceman's brain tissue. Cellular and molecular life sciences : CMLS. 2013;70(19):3709-22. Epub 2013/06/07. doi: 10.1007/s00018-013-1360-y. PubMed PMID: 23739949.
[20] Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, et al. Pathogens and host immunity in the ancient human oral cavity. Nature genetics. 2014;46(4):336-44. Epub 2014/02/25. doi: 10.1038/ng.2906. PubMed PMID: 24562188; PubMed Central PMCID: PMC3969750.

[21] Hendy J, Warinner C, Bouwman A, Collins MJ, Fiddyment S, Fischer R, et al. Proteomic evidence of dietary sources in ancient dental calculus. Proc Biol Sci. 2018;285(1883). doi: 10.1098/rspb.2018.0977. PubMed PMID: 30051838; PubMed Central PMCID: PMCPMC6083251.
[22] Warinner C, Hendy J, Speller C, Cappellini E, Fischer R, Trachsel C, et al. Direct evidence of milk consumption from ancient human dental calculus. Scientific reports. 2014;4:7104. Epub 2014/11/28. doi: 10.1038/srep07104. PubMed PMID: 25429530; PubMed Central PMCID: PMC4245811.
[23] Brown S, Higham T, Slon V, Paabo S, Meyer M, Douka K, et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. Scientific reports. 2016;6:23559. Epub 2016/03/30. doi: 10.1038/srep23559. PubMed PMID: 27020421; PubMed Central PMCID: PMC4810434.
[24] Chen F, Welker F, Shen CC, Bailey SE, Bergmann I, Davis S, et al. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. Nature. 2019. doi: 10.1038/s41586-019-1139-x. PubMed PMID: 31043746.
[25] Rybczynski N, Gosse JC, Harington CR, Wogelius RA, Hidy AJ, Buckley M. Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. Nature communications. 2013;4:1550. Epub 2013/03/07. doi: 10.1038/ncomms2516. PubMed PMID: 23462993; PubMed Central PMCID: PMC3615376.
[26] Welker F, Collins MJ, Thomas JA, Wadsley M, Brace S, Cappellini E, et al. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. Nature. 2015;522(7554):81-4. Epub 2015/03/25. doi: 10.1038/nature14249. PubMed PMID: 25799987.
[27] Welker F, Ramos-Madrigal J, Gutenbrunner P, Mackie M, Tiwary S, Rakownikow Jersie-Christensen R, et al. The dental proteome of Homo antecessor. Nature. 2020;580(7802):235-8. doi: 10.1038/s41586-020-2153-8. PubMed PMID: 32269345.
[28] Buckley M, Collins M, Thomas-Oates J, Wilson JC. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid communications in mass spectrometry : RCM. 2009;23(23):3843-54. Epub 2009/11/10. doi: 10.1002/rcm.4316. PubMed PMID: 19899187.
[29] Welker F, Hajdinjak M, Talamo S, Jaouen K, Dannemann M, David F, et al. Palaeoproteomic evidence identifies archaic hominins associated with the Chatelperronian at the Grotte du Renne. Proc Natl Acad Sci U S A. 2016. Epub 2016/09/18. doi: 10.1073/pnas.1605834113. PubMed PMID: 27638212.
[30] Cappellini E, Prohaska A, Racimo F, Welker F, Pedersen MW, Allentoft ME, et al. Ancient Biomolecules and Evolutionary Inference. Annu Rev Biochem. 2018;87:1029-60. doi: 10.1146/annurev-biochem-062917-012002. PubMed PMID: 29709200.

[31] Froment C, Hourset M, Saenz-Oyhereguy N, Mouton-Barbosa E, Willmann C, Zanolli C, et al. Analysis of 5000year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides. J Proteomics. 2020;211:103548. doi: 10.1016/j.jprot.2019.103548. PubMed PMID: 31626997.
[32] Stewart NA, Gerlach RF, Gowland RL, Gron KJ, Montgomery J. Sex determination of human remains from peptides in tooth enamel. Proc Natl Acad Sci U S A. 2017;114(52):13649-54. doi: 10.1073/pnas.1714926115. PubMed PMID: 29229823; PubMed Central PMCID: PMCPMC5748210.
[33] Wasinger VC, Curnoe D, Bustamante S, Mendoza R, Shoocongdej R, Adler L, et al. Analysis of the Preserved Amino Acid Bias in Peptide Profiles of Iron Age Teeth from a Tropical Environment Enable Sexing of Individuals Using Amelogenin MRM. Proteomics. 2019;19(5):e1800341. doi: 10.1002/pmic.201800341. PubMed PMID: 30650255.
[34] Zanolli C, Hourset M, Esclassan R, Mollereau C. Neanderthal and Denisova tooth protein variants in present-day humans. PLoS One. 2017;12(9):e0183802. doi: 10.1371/journal.pone.0183802. PubMed PMID: 28902892; PubMed Central PMCID: PMCPMC5597096.
[35] Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid communications in mass spectrometry : RCM. 2003;17(20):2337-42. doi: 10.1002/rcm.1196. PubMed PMID: 14558135.
[36] Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics. 2012;11(4):M111 010587. doi: 10.1074/mcp.M111.010587. PubMed PMID: 22186715; PubMed Central PMCID: PMCPMC3322562.
[37] Kall L, Storey JD, MacCoss MJ, Noble WS. Posterior error probabilities and false discovery rates: two sides of the same coin. Journal of proteome research. 2008;7(1):40-4. doi: 10.1021/pr700739d. PubMed PMID: 18052118.
[38] Sinitcyn P, Rudolph JD, Cox J. Computational methods for understanding mass spectrometry-based shotgun proteomic data. Annula Review of Biomedical Data Science. 2018;1:28. doi: https://doi.org/10.1146/annurev-biodatasci-080917-013516.
[39] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. doi: 10.1073/pnas.0506580102. PubMed PMID: 16199517; PubMed Central PMCID: PMCPMC1239896.

[40] Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9(8):772. doi: 10.1038/nmeth.2109. PubMed PMID: 22847109; PubMed Central PMCID: PMCPMC4594756.
[41] Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453-5. doi: 10.1093/bioinformatics/btz305. PubMed PMID: 31070718; PubMed Central PMCID: PMCPMC6821337.
[42] Haley PJ. Species differences in the structure and function of the immune system. Toxicology. 2003;188(1):49-71. doi: 10.1016/s0300-483x(03)00043-x. PubMed PMID: 12748041.
[43] Welker F. Elucidation of cross-species proteomic effects in human and hominin bone proteome identification through a bioinformatics experiment. BMC Evol Biol. 2018;18(1):23. doi: 10.1186/s12862-018-1141-1. PubMed PMID: 29463217; PubMed Central PMCID: PMCPMC5819086.
[44] Widbiller M, Schweikl H, Bruckmann A, Rosendahl A, Hochmuth E, Lindner SR, et al. Shotgun Proteomics of Human Dentin with Different Prefractionation Methods. Scientific reports. 2019;9(1):4457. doi: 10.1038/s41598-019-41144-x. PubMed PMID: 30872775; PubMed Central PMCID: PMCPMC6418255.
[45] Jagr M, Eckhardt A, Pataridis S, Broukal Z, Duskova J, Miksik I. Proteomics of human teeth and saliva. Physiol Res. 2014;63 Suppl 1:S141-54. PubMed PMID: 24564654.
[46] Park ES, Cho HS, Kwon TG, Jang SN, Lee SH, An CH, et al. Proteomics analysis of human dentin reveals distinct protein expression profiles. Journal of proteome research. 2009;8(3):1338-46. doi: 10.1021/pr801065s. PubMed PMID: 19193101.
[47] Smith AJ, Scheven BA, Takahashi Y, Ferracane JL, Shelton RM, Cooper PR. Dentine as a bioactive extracellular matrix. Arch Oral Biol. 2012;57(2):109-21. doi: 10.1016/j.archoralbio.2011.07.008. PubMed PMID: 21855856.
[48] Bahar FG, Ohura K, Ogihara T, Imai T. Species difference of esterase expression and hydrolase activity in plasma. J Pharm Sci. 2012;101(10):3979-88. doi: 10.1002/jps.23258. PubMed PMID: 22833171.
[49] Horvath JE, Ramachandran GL, Fedrigo O, Nielsen WJ, Babbitt CC, St Clair EM, et al. Genetic comparisons yield insight into the evolution of enamel thickness during human evolution. Journal of human evolution. 2014;73:75-87. doi: 10.1016/j.jhevol.2014.01.005. PubMed PMID: 24810709.
[50] Lee JJ, Morris D, Constantino PJ, Lucas PW, Smith TM, Lawn BR. Properties of tooth enamel in great apes. Acta Biomater. 2010;6(12):4560-5. doi: 10.1016/j.actbio.2010.07.023. PubMed PMID: 20656077.

[51] Wright JT, Hall K, Yamauchi M. The protein composition of normal and developmentally defective enamel. Ciba Found Symp. 1997;205:85-99; discussion -106. doi: 10.1002/9780470515303.ch7. PubMed PMID: 9189619.
[52] Jagr M, Eckhardt A, Pataridis S, Miksik I. Comprehensive proteomic analysis of human dentin. Eur J Oral Sci. 2012;120(4):259-68. doi: 10.1111/j.1600-0722.2012.00977.x. PubMed PMID: 22813215.
[53] Castiblanco GA, Rutishauser D, Ilag LL, Martignon S, Castellanos JE, Mejia W. Identification of proteins from human permanent erupted enamel. Eur J Oral Sci. 2015;123(6):390-5. doi: 10.1111/eos.12214. PubMed PMID: 26432388.
[54] Jagr M, Ergang P, Pataridis S, Kolrosova M, Bartos M, Miksik I. Proteomic analysis of dentin-enamel junction and adjacent protein-containing enamel matrix layer of healthy human molar teeth. Eur J Oral Sci. 2019;127(2):112-21. doi: 10.1111/eos.12594. PubMed PMID: 30466169.
[55] Procopio N, Chamberlain AT, Buckley M. Exploring Biological and Geological Age-related Changes through Variations in Intra- and Intertooth Proteomes of Ancient Dentine. Journal of proteome research. 2018;17(3):1000-13. doi: 10.1021/acs.jproteome.7b00648. PubMed PMID: 29356547.

Table 1: List of the proteins common to the five proteomes.

The list of proteins was retrieved from the comparison of the protein lists obtained from each dataset per genus with Proteome Discoverer (PD) or PEAKS (PX) softwares. For each protein, Uniprot accession numbers correspond to the master protein and the first protein of the protein group identified by PD and by PX, respectively. The number of total identified peptides (and unique peptides) that allowed for protein identification is indicated. * In the case of proteins identified with only 1 unique peptide, the HCD MS/MS spectra of unique peptides are presented in Fig. S2. nd (not detected)

			Homo		***Gorilla***		***Pan***		***Pongo***		***Papio***	
Gene name	**Protein name**	**Ident Software**	**Uniprot Acession**	**# Peptides (unique)**	**Uniprot Acession**	**# Peptides (unique)**	**Uniprot Acession**	**# Peptides (unique)**	**Uniprot Acession**	**# Peptides (unique)**	**Uniprot Acession**	**# Peptides (unique)**
AHSG	Alpha 2-HS glycoprotein	PD	P02765	27 (27)	E1U7Q5	47 (44)	A0A2R9ADR9	5 (5)	H2PC98 E1U7Q6	24 (2)	A0A096NPS0	36 (36)
		PX	B7Z8Q2	51 (19)	E1U7Q5	89 (85)	A0A2R9ADR9	29 (29)	H2PC98 E1U7Q6	59 (4)	B9MSS3	71 (71)
AMBN	Ameloblastin	PD	Q9NP70	6 (6)	G3RCU1	2 (2)	A0A2R9CKL8	2 (2)	A0A2J8UTQ2	2 (2)	A0A096NFU6	3 (3)
		PX	Q9NP70	8 (8)	nd		A0A2J8PF83	2 (2)	nd		A0A096NFU6	1(1)
APOA1	Apolipoprotein A1	PD	A0A024R3E3	12 (12)	G3QY98	7 (7)	K7D1U8	5 (5)	P0DJG1	9 (9)	P68293	12 (12)
		PX	P02647	17 (15)	G3QY98	11 (11)	P0DJG0	8 (8)	A0A2J8X1C8	12 (12)	P68293	17 (16)
BGN	Biglycan	PD	B4DDQ2	11 (11)	A0A2I2YJ91	15 (15)	H2R1R5	10 (10)	H2PX51	14 (14)	A0A096NEE7	27 (27)
		PX	A6NLG9	17 (17)	A0A2I2YJ91	38 (37)	H2R1R5	20 (19)	H2PX51	17 (17)	A0A096NEE7	61 (59)
C3	Complement C3	PD	P01024	6 (6)	G3RBJ0	7 (7)	A0A2R9B9K1	4 (4)	A0A2J8R6I7	3 (3)	A0A0A0MUD9	3 (3)
		PX	V9HWA9	18 (17)	G3RBJ0	17 (17)	K7CUE1	11 (11)	A0A2J8R6I7	4 (4)	A0A0A0MUD9	13 (13)
C9	Complement C9	PD	A0A024R035	3 (3)	G3RIM1	12 (12)	A0A2R9BNI9	2 (2)	H2PFE3	4 (4)	A0A096N4A4	12 (12)
		PX	P02748	6 (6)	G3RIM1	26 (26)	A0A2R9BNI9	10 (10)	H2PFE3	6 (6)	A0A096N4A4	23 (23)
CLEC11A	C-type lectin domain family 11A	PD	M0R081	3 (3)	G3S6C9	2 (2)	A0A2J8J1C9	2 (2)	A0A2J8U8Y9	1 (1)*	A0A2I3M1B2	5 (5)
		PX	Q5U0B9	4 (2)	G3S6C9	4 (4)	H2QGY4	3 (0)	nd		A0A2I3M1B2	8 (8)
COL10A1	Collagen alpha-1(X) chain	PD	Q5QPC7	1 (1)*	G3S3J2	3 (2)	A0A2R9A6P1	2 (1)*	A0A2J8U152	1 (1)*	A0A096NA96	2 (2)
		PX	nd		A0A2I2YVF5	4 (3)	nd		A0A2J8U152	1 (1)	A0A096NA96	5 (4)
COL11A1	Collagen alpha-1(XI) chain	PD	D3DT71	9 (6)	A0A2I2Z7V8	20 (18)	A0A2I3TI23	6 (5)	A0A2J8VCH4	7 (7)	A0A2I3LKH0	19 (17)
		PX	P12107-4	12 (9)	G3QG18	24 (23)	A0A2I3RPT1	10 (9)	A0A2J8VCG6	8 (8)	A0A096MQI3	27 (24)
COL11A2	Collagen alpha-2(XI) chain	PD	A0A1U9X7I9	30 (27)	G3R2X9	34 (32)	H2R4E0	8 (7)	A0A2J8Y2T4	7 (7)	A0A096NHF5	33 (31)
		PX	P13942	34 (30)	G3R2X9	52 (50)	A0A2J8NYG5	15 (0)	A0A2J8Y2T2	11 (11)	A0A2I3NDC2	48 (47)
COL22A1	Collagen alpha-1(XXII) chain	PD	Q8NFW1	7 (7)	G3R3K3	19 (19)	H2QWR6	5 (5)	A0A2J8SAY6	9 (9)	A0A096NYZ3	9 (9)
		PX	Q8NFW1-2	10 (1)	G3R3K3	27 (27)	H2QWR6	9 (2)	A0A2J8SAY6	7 (1)	A0A2I3NDC7	16 (16)
COL23A1	Collagen alpha-	PD	Q86Y22	1 (1)*	G3QVJ2	2 (2)	A0A2J8JTY4	4 (3)	A0A2J8SD64	1 (1)*	A0A096NDQ8	2 (2)

	1(XXIII) chain	PX	nd		nd		A0A2R9C687	2 (1)	nd		nd	
COL3A1	Collagen alpha-1(III) chain	PD	P02461	90 (31)	G3RK87	38 (26)	K7D7I8	45 (15)	A0A2J8WW41	50 (33)	A0A096N2I8	59 (41)
		PX	P02461	92 (69)	G3RK87	46 (28)	K7D7I8	53 (6)	A0A2J8WW41	48 (10)	A0A096N2I8	84 (58)
COL5A1	Collagen alpha-1(V) chain	PD	B2ZZ86	17 (14)	G3R760	15 (13)	K7CMZ9	11 (10)	A0A2J8UIU7	13 (13)	A0A096NWJ0	18 (17)
		PX	Q59EE7	17 (13)	G3R760	25 (24)	K7D7I8	22 (3)	A0A2J8UIT7	18 (18)	A0A096NWJ0	35 (32)
COL5A2	Collagen alpha-2(V) chain	PD	P05997	39 (37)	G3RDT1	37 (33)	H2R6B8	18 (14)	H2P838	34 (31)	A0A096MVP2	42 (41)
		PX	P05997	39 (37)	G3RDT1	60 (14)	A0A2R9AJC2	32 (28)	H2P838	32 (29)	A0A096MVP2	65 (62)
DSPP	Dentin sialophosphoprotein	PD	Q9NZW4	7 (7)	G3SE58	11 (11)	A0A2J8MA47	2 (2)	A0A2J8V572	4 (4)	A0A2I3M6W8	4 (4)
		PX	Q9NZW4	11 (11)	G3SE58	21 (21)	A0A2J8MA47	6 (6)	A0A2J8V572	9 (9)	A0A2I3M6W8	9 (9)
F2	Prothrombin	PD	P00734	20 (3)	G3QVP5	25 (25)	A0A2R9C6X1	6 (6)	Q5R537	8 (8)	A0A096N4Z1	11 (11)
		PX	P00734	31 (4)	G3QVP5	35 (7)	H2Q3I2	13 (13)	Q5R537	16 (16)	A0A096N4Z1	24 (24)
HSPG2	Heparan sulfate proteoglycan	PD	P98160	2 (2)	A0A2I2YAC3	8 (8)	H2PY96	3 (3)	H2N8P5	6 (6)	A0A096N531	3 (3)
		PX	A0A024RAB6	5 (5)	A0A2I2YAC3	10 (10)	A0A2R9C5E1	4 (4)	H2N8P5	6 (6)	A0A096N531	5 (5)
LUM	Lumican	PD	P51884	10 (10)	G3S376	9 (9)	A0A2R8ZXH3	5 (5)	Q5RFG1	3 (3)	A0A096MQ49	9 (9)
		PX	Q53FV4	10 (10)	G3S376	13 (13)	H2Q6L3	7 (7)	H2NI87	6 (6)	A0A096MQ49	15 (15)
OMD	Osteomodulin	PD	B2R7N9	7 (7)	A0A2I2YQC5	7 (7)	H2QXG5	2 (2)	H2PSP3	2 (2)	A0A096P237	6 (6)
		PX	B2R7N9	11 (11)	A0A2I2YQC5	11 (11)	H2QXG5	6 (6)	H2PSP3	4 (4)	A0A096P237	11 (11)
PCOLCE	Procollagen C-endopeptidase enhancer1	PD	A4D2D2	8 (8)	G3R5A8	11 (11)	A0A2R9C060	3 (3)	A0A2J8XUJ5	3 (3)	A0A096NNW4	11 (11)
		PX	Q15113	10 (10)	G3R5A8	16 (16)	H2QV35	5 (5)	A0A2J8XUJ5	4 (4)	A0A096NNW4	14 (14)
SERPINA1	Alpha-1-antitrypsin	PD	A0A384MDQ7	9 (9)	S4UFD6	4 (4)	A0A2J8QMJ5	2 (2)	Q5RCW5	5 (5)	P01010	3 (3)
		PX	E9KL23	24 (24)	G3QXZ8	16 (16)	A0A2J8QMJ5	4 (4)	Q5RCW5	8 (8)	P01010	5 (5)
SERPINC1	Antithrombin-III	PD	P01008	9 (9)	G3S9Q7	6 (5)	A0A2R9CFX9	5 (5)	Q5R5A3	4 (4)	A0A096N0R9	7 (7)
		PX	A0A0K0Q2Z1	13 (10)	G3S9Q7	9 (8)	A0A2R9CFX9	13 (9)	Q5R5A3	5 (5)	A0A096N0R9	19 (11)
SPARC	Secreted protein acidic rich in C/Osteonectin	PD	P09486	5 (5)	G3RJ76	13 (13)	A0A2R9BZI6	6 (6)	Q5R767	10 (10)	A0A096MNJ1	15 (15)
		PX	D3DQH8	7 (1)	G3RJ76	27 (27)	H2QRU3	17 (17)	Q5R767	11 (11)	A0A096MNJ1	48 (48)
VTN	Vitronectin	PD	P04004	8 (8)	G3R679	7 (7)	H2QCH3	2 (2)	H2NT31	2 (2)	A0A096P388	6 (6)
		PX	D9ZGG2	18 (18)	G3R679	11 (11)	A0A2R9BDP7	6 (6)	A0A2J8TMB3	3 (3)	A0A096P388	9 (9)

Table 2: Proteins with the main representative peptides showing a taxonomic variation among the hominids.

The list of peptides was retrieved from each dataset obtained per genus with Proteome Discoverer (PD) or PEAKS (PX) softwares (the complete lists of peptides is given in Tables S4-S17, with more detailed information). For each peptide, a Blast search on protein (BlastP) was performed in Uniprot and NCBI to check the species specificity (100% identity and Query cover).

Red bold high size character indicates a position showing an amino acid variation specific to the taxon. *The ancestral/derived amino substitution is provided with respect to the position in the human Uniprot protein accession indicated in brackets. Black bold high size character indicates a discriminative variation between groups of hominids. The spectra of the peptides are shown in Fig. 4, Fig. 5 and Fig. S4.

Gene name	***Homo* sample** Uniprot accession	Position	Ancestral /derived*	BlastP result
AHSG	**P02765**			
	QPNCDDPETEEAALVA**I**DYINQN**L**PWGYK	29-57		*H. sapiens, G. gorilla gorilla, G. gorilla*
	HTFMGVVSLGS**P**SGE**V**SHPR	318-337		***H. sapiens***
	LGS**P**SGE**V**SHPR	326-337		***H. sapiens***
DSPP	**Q9NZW4**			
	ESGV**L**VHEGDR	56-66		*H. sapiens, P. troglodytes, P. paniscus*
	ESE**T**H**A**VGK	362-370		*H. sapiens, G. gorilla gorilla (na G. gorilla), P. troglodytes, P. paniscus*
F2	**P00734**			
	SEGSSVNLSPP**L**EQCVPDR	199-217	S210L	***H. sapiens***
	N.LSPP**L**EQCVPDR.G	206-217		***H. sapiens***
PCOLCE	**A4D2D2/Q15113**			
	TEE**S**PSAPDAPTCPK	306-320		*H. sapiens, P. troglodytes, P. paniscus*
Gene name	***Gorilla* sample** Uniprot accession	Position	Ancestral /derived*	BlastP result
AHSG	**E1U7Q5/A0A2I2ZQ06**		(P02765)	
	QPNCDDPETEEAALVA**I**DYINQN**L**PWGYK	29-57		*H. sapiens, G. gorilla gorilla, G. gorilla*
	QPSGELFEIEIDTLETTCHVLDPTPVA**K**	72-99	R99K	***G. gorilla gorilla, G. gorilla***
	VLDPTPVA**K**	91-99		***G. gorilla gorilla, G. gorilla***
	H**T**FMGVV**SL**GS**P**SGE**A**SHPR	318-337		*G. gorilla gorilla, G. gorilla, P.troglodytes, P. paniscus*
	LGS**P**SGE**A**SHPR	326-337		*G. gorilla gorilla, G. gorilla, P.troglodytes, P. paniscus*
C9	**G3RIM1**			
	TSNFNA**A**ISLK	232-242		*H. sapiens, G. gorilla gorilla (na G. gorilla), P. troglodytes, P. paniscus*
	ISEGLPAL**E**FP**N**E	546-558		*H. sapiens, G. gorilla gorilla (na G. gorilla), P. troglodytes, P. paniscus*
COL3A1	**G3RK87**		(P02461)	
	GEVGPAGSPGSNG**V**PGQR	351-368	A364V	***G. gorilla gorilla (na G. gorilla)***

	DGP**A**GPAGNTGAPGSPGVSGPK	902-923	P405A	***G. gorilla gorilla (na G. gorilla)***
DSPP	**G3SE58**		(Q9NZW4)	
	ESE**T**H**A**VGK	362-370		*H.sapiens, gorilla gorilla, P. troglodytes, P. paniscus*
	GQHGMIL**S**K	403-411	G410S	***G. gorilla gorilla (na G. gorilla)***
F2	**G3QVP5**			
	K.**G**QPSVLQVVNLPIVER.P	518-533		*G.gorilla gorilla (na G. gorilla), H. sapiens, P. troglodytes, P. paniscus*
PCOLCE	**G3R5A8**			
	TEE**T**PSAPDAPTCPK	306-320		*G. gorilla gorilla (na G. gorilla), P. abelii (na P. hamadryas)*
SERPINA1	**S4UFD6/G3QXZ8**		(P01009)	
	TLNQPDSQLQLTTG**S**GLFLSEGLK	126-149	N140S	***G. gorilla gorilla, G. gorilla***

Gene name	***Pongo* sample** Uniprot accession	Position	Ancestral /derived*	BlastP result
AHSG	**H2PC98/ E1U7Q6**		(P02765)	
	QPNCDDPETEEAALVA**I**DYINQN**H**PWGYK	29-57		*P abelii, P. pygmaeus, P. troglodytes, P. paniscus*
	QLKEHAVEGDCDF**K**	104-117	Q117K	***P. abelii, P. pygmaeus***
	S**L**SGE**V**SHPR	328-337	P329L	***P. abelii, P. pygmaeus***
APOA1	**P0DJG1/A0A2J8X1C8**		(P02647)	
	THLAPY**T**DELR	185-195	S191T	***P. abelli, P.pygmaeus***
BGN	**H2PX51**		(P21810)	
	ELHLDNNKLA**G**VPSGLPDLK	220-239	R291G	***P. abelii (na P. pygmaeus)***
	LA**G**VPSGLPDLK	228-239		***P. abelii (na P. pygmaeus)***
PCOLCE	**A0A2J8XUJ5**			
	TEE**T**PSAPDAPTCPK	305-319		*G. gorilla gorilla (na G. gorilla), P. abelii (na P. pygmaeus)*

Legend to figures

Figure 1: Schematic of the FASP protocol describing the steps for the preparation of the T samples (tryptic peptides issued from protein digestion) and the Tpep samples (tryptic peptides issued from the digestion of peptides and/or fragmented proteins already present in the demineralized/lysed extract). See Materials and Methods for details.

Figure 2: Comparison of human (*Homo*); gorilla (*Gorilla*) chimpanzee (*Pan*), orangutan (*Pongo*) and baboon (*Papio*) tooth proteomes.
(A) Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) showing the distribution between genera of the proteins identified in each resulting dataset per genus using Proteome Discoverer, and converted into the corresponding human gene names for allowing comparison. (B) Gene set overlap analysis according to GO annotation terms of the proteins common to the five taxa (Common), or uniquely identified in each genus, by using the GSEA tool at https://www.gsea-msigdb.org/gsea/msigdb/annotate.jspdb. The bar graph shows the pattern of distribution of the proteins sorted with significant FDR q-values <0.05 in the Top 20 GO Gene Sets. The distribution was calculated as the ratio of the protein counts in each category (blood/wound-healing, extracellular matrix/ossification, immune response and others the rest of the less represented GO terms) to the total protein counts in the Top 20 gene sets, and was represented as stacked bars for each sample.

Figure 3: Phylogenetic tree based on the concatenated alignment of 14 proteins (AHSG, AMBN, APOA1, BGN, C9, COL11A2, COL22A1, COL3A1, DSPP, F2, LUM, OMD, PCOLCE, SERPINA1).
The position of the sample datasets is visualized in bold characters.
The tree was built using RAxML-NG. Node values (%) correspond to 1000 bootstraps and branch length indicates the rate of amino acid substitution.

Figure 4: HCD MS/MS spectra of AHSG peptides with taxonomic interest.
(A) AHSG-[91-99] *Gorilla*-specific peptide VLDPTPVAK (doubly charged precursor ion, MH2+, at m/z 470.2787; scan 17775; OFCCF180623_18_SP01_CCF01530_T4_TR1_Gorilla.raw). (B) AHSG-[104-117] *Pongo*-specific peptide QLKEHAVEGDCDFK (triply charged precursor ion, MH3+, at m/z 559.5901; scan 12157; OFCCF200122_13_SP03_CCF01646_T8_TR1_Pongo.raw). (C) AHSG-[318-337] *Homo*-specific peptide HTFMGVVSLGSPSGEVSHPR (triply charged precursor ion, MH3+, at m/z 694.3470; scan 57611;

OFCCF180912_14_S3_CCF01452_LOS2_TR1_Human.raw), AHSG-[318-337] *Gorilla* peptide HTFMGVVSLGSPSGEASHPR (triply charged precursor ion, MH3+, at m/z 685.0026; scan 57284; OFCCF180623_18_SP01_CCF01530_T4_TR1_Gorilla.raw) and AHSG-[328-337] *Pongo*-specific peptide SLSGEVSHPR (doubly charged precursor ion, MH2+, at m/z 534.7749; scan 6900; OFCCF200122_31_SP03_CCF01646_T8_TR2_Pongo.raw). (D) AHSG-[29-57] *Homo*- and *Gorilla*- specific peptide QPNcDDPETEEAALVAIDYINQNLPWGYK (triply charged precursor ion, MH3+, at m/z 1121.8558; scan 65473; OFCCF180623_24_SP01_CCF01530_T4_TR2_Gorilla.raw) and the corresponding AHSG-[29-57] *Pongo*-specific peptide QPNcDDPETEEAALVAIDYINQNHPWGYK (triply charged precursor ion, MH3+, at m/z 1129.8445; scan 54151; OFCCF200122_31_SP03_CCF01646_T8_TR2_Pongo.raw). Series of y- and b-ions are highlighted in blue and red, respectively. c: carbamidomethylated cysteine residue.

Figure 5: HCD MS/MS spectra of PCOLCE peptides with taxonomic interest.
PCOLCE-[306-320] *Homo* peptide TEESPSAPDAPTcPK (doubly charged precursor ion, MH2+, at m/z 793.8555; scan 14888; OFCCF180912_22_S3_CCF01452_LOS2_TR2_Human.raw) and, the corresponding *Gorilla* PCOLCE-[306-320] and *Pongo* PCOLCE-[305-319] peptides TEETPSAPDAPTcPK (doubly charged precursor ion, MH2+, at m/z 800.8607; scan 19435; OFCCF180623_24_SP01_CCF01530_T4_TR2_Gorilla.raw, and doubly charged precursor ion, MH2+, at m/z 800.8613; scan 14388; OFCCF200122_13_SP03_CCF01646_T8_TR1_Pongo.raw, respectively).
Series of y- and b-ions are highlighted in blue and red, respectively. c: carbamidomethylated cysteine residue.

Figure 1: Schematic of the FASP protocol describing the steps for the preparation of the T samples (tryptic peptides issued from protein digestion) and the Tpep samples (tryptic peptides issued from the digestion of peptides and/or fragmented proteins already present in the demineralized/lysed extract). See Materials and Methods for details.

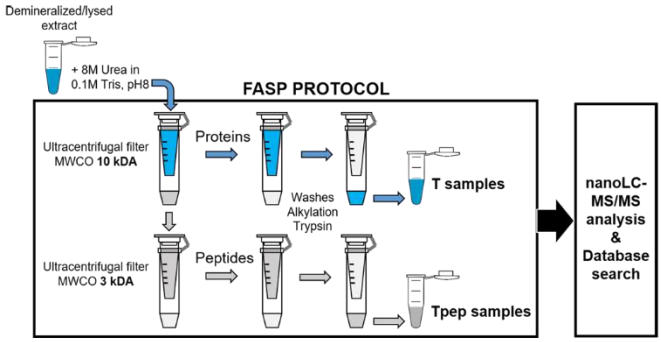

Figure 2: Comparison of human (*Homo*); gorilla (*Gorilla*) chimpanzee (*Pan*), orangutan (*Pongo*) and baboon (*Papio*) tooth proteomes.

(A) Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) showing the distribution between genera of the proteins identified in each resulting dataset per genus using Proteome Discoverer, and converted into the corresponding human gene names for allowing comparison. (B) Gene set overlap analysis according to GO annotation terms of the proteins common to the five taxa (Common), or uniquely identified in each genus, by using the GSEA tool at https://www.gsea-msigdb.org/gsea/msigdb/annotate.jspdb. The bar graph shows the pattern of distribution of the proteins sorted with significant FDR q-values <0.05 in the Top 20 GO Gene Sets. The distribution was calculated as the ratio of the protein counts in each category (blood/wound-healing, extracellular matrix/ossification, immune response and others the rest of the less represented GO terms) to the total protein counts in the Top 20 gene sets, and was represented as stacked bars for each sample.

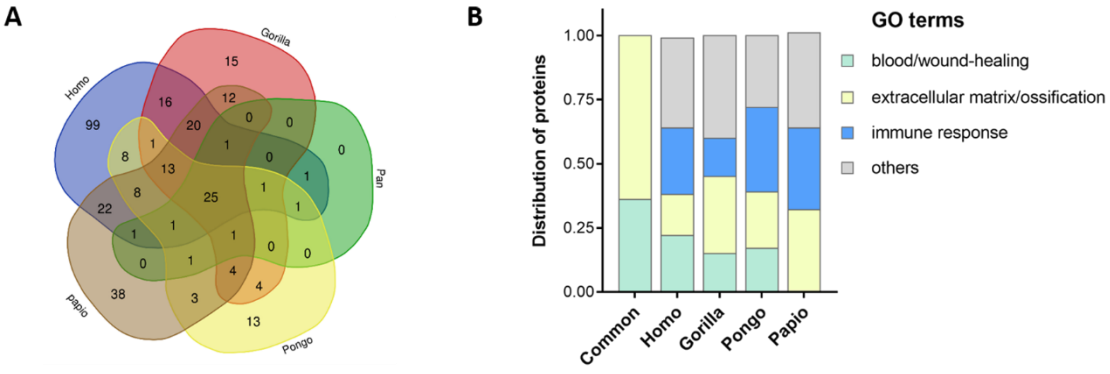

Figure 3: Phylogenetic tree based on the concatenated alignment of 14 proteins (AHSG, AMBN, APOA1, BGN, C9, COL11A2, COL22A1, COL3A1, DSPP, F2, LUM, OMD, PCOLCE, SERPINA1).

The position of the sample datasets is visualized in bold characters.

The tree was built using RAxML-NG. Node values (%) correspond to 1000 bootstraps and branch length indicates the rate of amino acid substitution.

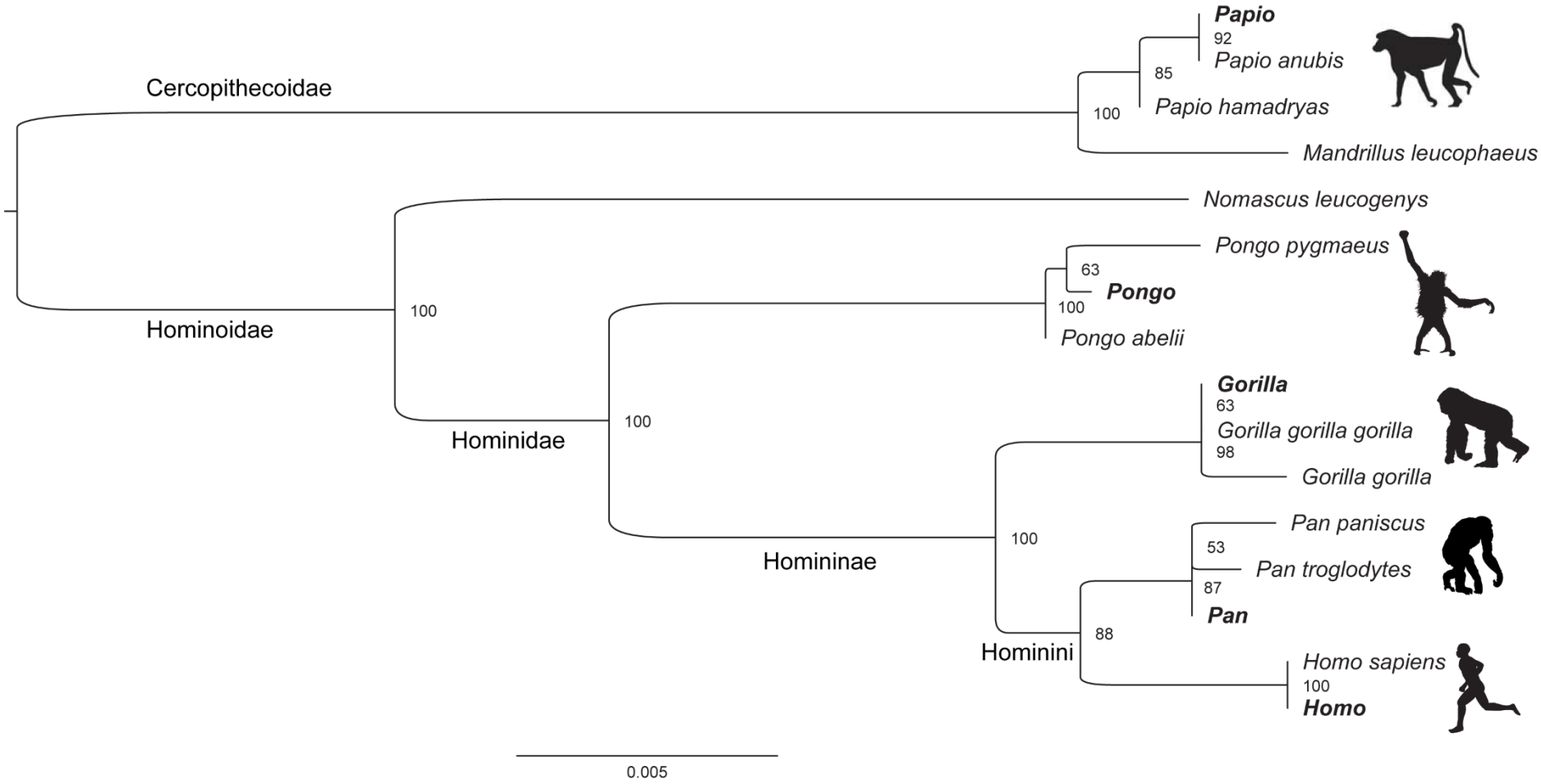

Figure 4: HCD MS/MS spectra of AHSG peptides with taxonomic interest.

(A) AHSG-[91-99] *Gorilla*-specific peptide VLDPTPVAK (doubly charged precursor ion, MH2+, at m/z 470.2787; scan 17775; OFCCF180623_18_SP01_CCF01530_T4_TR1_Gorilla.raw). (B) AHSG-[104-117] *Pongo*-specific peptide QLKEHAVEGDCDFK (triply charged precursor ion, MH3+, at m/z 559.5901; scan 12157; OFCCF200122_13_SP03_CCF01646_T8_TR1_Pongo.raw). (C) AHSG-[318-337] *Homo*-specific peptide HTFMGVVSLGSPSGEVSHPR (triply charged precursor ion, MH3+, at m/z 694.3470; scan 57611; OFCCF180912_14_S3_CCF01452_LOS2_TR1_Human.raw), AHSG-[318-337] *Gorilla* peptide HTFMGVVSLGSPSGEASHPR (triply charged precursor ion, MH3+, at m/z 685.0026; scan 57284; OFCCF180623_18_SP01_CCF01530_T4_TR1_Gorilla.raw) and AHSG-[328-337] *Pongo*-specific peptide SLSGEVSHPR (doubly charged precursor ion, MH2+, at m/z 534.7749; scan 6900; OFCCF200122_31_SP03_CCF01646_T8_TR2_Pongo.raw). (D) AHSG-[29-57] *Homo*- and *Gorilla*- specific peptide QPNcDDPETEEAALVAIDYINQNLPWGYK (triply charged precursor ion, MH3+, at m/z 1121.8558; scan 65473; OFCCF180623_24_SP01_CCF01530_T4_TR2_Gorilla.raw) and the corresponding AHSG-[29-57] *Pongo*-specific peptide QPNcDDPETEEAALVAIDYINQNHPWGYK (triply charged precursor ion, MH3+, at m/z 1129.8445; scan 54151; OFCCF200122_31_SP03_CCF01646_T8_TR2_Pongo.raw). Series of y- and b-ions are highlighted in blue and red, respectively. c: carbamidomethylated cysteine residue.

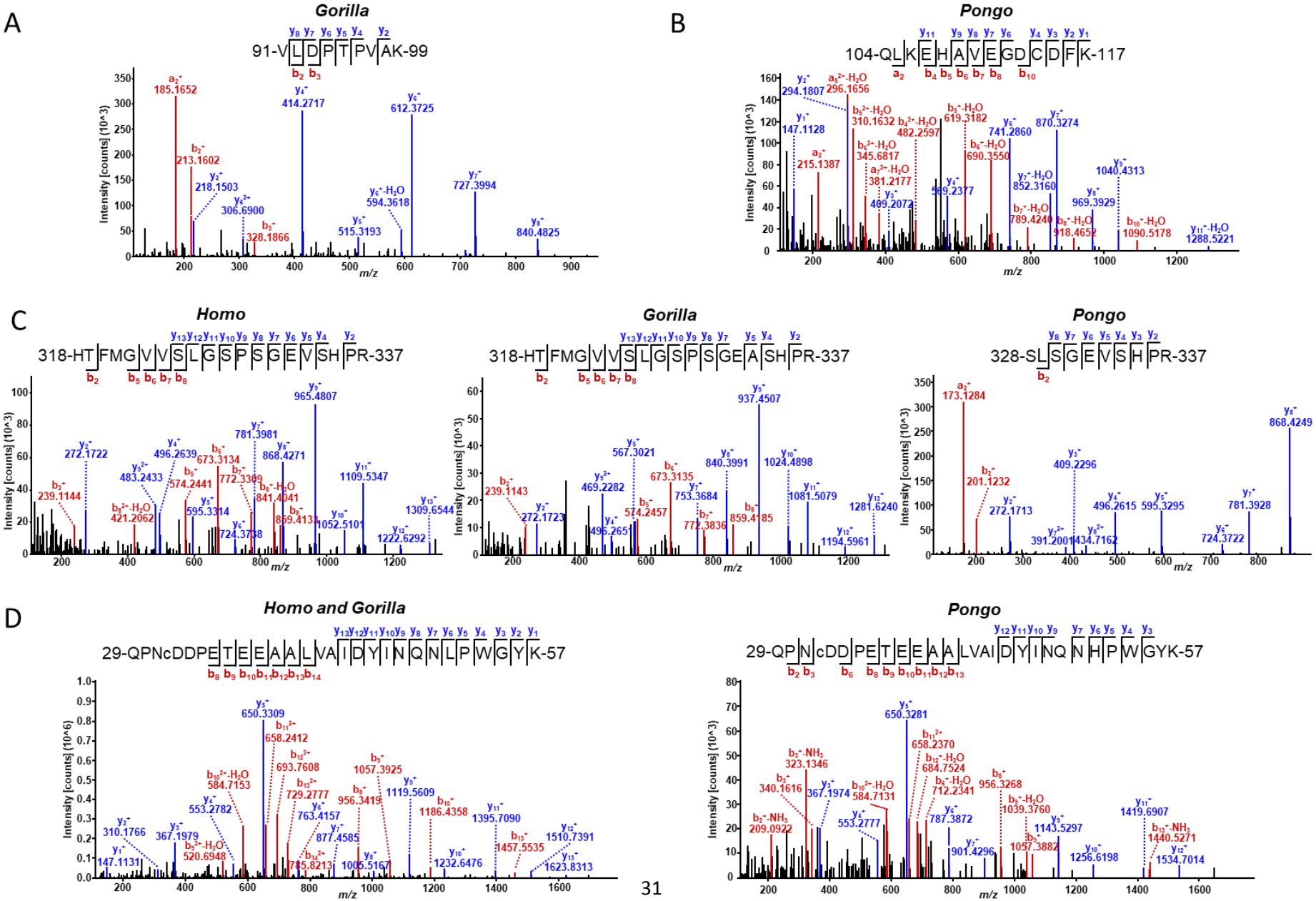
A
Gorilla
91-VLDPTPVAK-99
B
Pongo
104-QLKEHAVEGDCDFK-117
C
Homo
318-HTFMGVVSLGSPSGEVSHPR-337
Gorilla
318-HTFMGVVSLGSPSGEASHPR-337
Pongo
328-SLSGEVSHPR-337
D
Homo and Gorilla
29-QPNcDDPETEEAALVAIDYINQNLPWGYK-57
Pongo
29-QPNcDDPETEEAALVAIDYINQNHPWGYK-57
Intensity [counts] (10^3)
Intensity [counts] (10^6)
m/z

Figure 5: HCD MS/MS spectra of PCOLCE peptides with taxonomic interest.

PCOLCE-[306-320] *Homo* peptide TEESPSAPDAPTcPK (doubly charged precursor ion, MH2+, at m/z 793.8555; scan 14888; OFCCF180912_22_S3_CCF01452_LOS2_TR2_Human.raw) and, the corresponding *Gorilla* PCOLCE-[306-320] and *Pongo* PCOLCE-[305-319] peptides TEETPSAPDAPTcPK (doubly charged precursor ion, MH2+, at m/z 800.8607; scan 19435; OFCCF180623_24_SP01_CCF01530_T4_TR2_Gorilla.raw, and doubly charged precursor ion, MH2+, at m/z 800.8613; scan 14388; OFCCF200122_13_SP03_CCF01646_T8_TR1_Pongo.raw, respectively).

Series of y- and b-ions are highlighted in blue and red, respectively. c: carbamidomethylated cysteine residue.