



HAL
open science

Untangling the fibre ball: Proteomic characterization of South American camelid hair fibres by untargeted multivariate analysis and molecular networking

Clara Azémard, Elise Dufour, Antoine Zazzo, Jane Wheeler, Nicolas Goepfert, Arul Marie, Séverine Zirah

► To cite this version:

Clara Azémard, Elise Dufour, Antoine Zazzo, Jane Wheeler, Nicolas Goepfert, et al.. Untangling the fibre ball: Proteomic characterization of South American camelid hair fibres by untargeted multivariate analysis and molecular networking. *Journal of Proteomics*, 2021, 231, pp.104040. 10.1016/j.jprot.2020.104040 . hal-03209364

HAL Id: hal-03209364

<https://hal.science/hal-03209364>

Submitted on 21 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Untangling the fibre ball: proteomic characterisation of South American camelid hair fibres**
2 **by untargeted multivariate analysis and molecular networking**

3

4 Clara Azémard^{1,2}, Elise Dufour², Antoine Zazzo², Jane C. Wheeler⁴, Nicolas Goepfert³, Arul
5 Marie¹, Séverine Zirah^{1,*}

6 ¹ Unité Molécules de Communication et Adaptations des Microorganismes (MCAM),
7 Muséum national d'Histoire naturelle, CNRS, CP 54, 63 rue Buffon, 75005 Paris, France.

8 ² Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnements (AASPE),
9 Muséum national d'Histoire naturelle, CNRS, CP 56, 55 rue Buffon, 75005 Paris, France.

10 ³ Archéologie des Amériques, UMR 8096, CNRS – Université Paris 1 Panthéon-Sorbonne,
11 MSH Mondes, 21 allée de l'université, 92023 Nanterre, France

12 ⁴ CONOPA - Instituto de Investigación y Desarrollo de Camélidos Sudamericanos, Av.
13 Reusche M4, Pachacamac, Lima 19, Peru

14 *Corresponding author

15

16 **ABSTRACT**

17 The proteomic analysis of hairs, yarns or textiles has emerged as a powerful method to
18 determine species of origin, mainly used in archaeozoological research and fraud control.
19 Differentiation between the South American Camelid (SAC) species (the wild guanaco and
20 vicuña and their respective domesticates the llama and alpaca) is particularly challenging
21 due to poor database information and significant hybridization between species. In this
22 study, we analysed 41 modern and 4 archaeological samples from the four SACs species.
23 Despite strong similarities with Old World Camelidae, we identified 7 peptides specific to
24 SACs assigned to keratin K86 and the keratin-associated proteins KAP13-1 and KAP11-1.
25 Untargeted multivariate analysis of the LC-MS data permitted to distinguish SAC species and
26 propose discriminant features. MS/MS-based molecular networking combined with
27 database-assisted *de novo* sequencing permitted to identify 5 new taxonomic peptides
28 assigned to K33a, K81 and/or K83 keratins and KAP19-1. These peptides differentiate the two
29 wild species, guanaco and vicuña. These results show the value of combining database
30 search and untargeted metabolomic approaches for paleoproteomics, and reveal for the first
31 time the potential of molecular networks to highlight deamidation related to diagenesis and

32 cluster highly similar peptides related to interchain homologies or intra- or inter-specific
33 polymorphism.

34

35 **Keywords**

36 Paleoproteomics, South American camelids, hair fibres, keratins, molecular network

37

38 **Significance**

39 This study used an innovative approach combining multivariate analysis of LC-MS data
40 together with molecular networking and database-assisted *de novo* sequencing to identify
41 taxonomic peptides in palaeoproteomics.

42 It constitutes the first attempt to differentiate between hair fibres from the four South
43 American Camelids (SACs) based on proteomic analysis of modern and archaeological
44 samples. It provides different proteomic signatures for each of the four SAC species and
45 proposes new SAC taxonomic peptides of interest in archaeozoology and fraud control.

46 SACs have been extensively exploited since human colonization of South America but have
47 not been studied to the extent of their economic, cultural and heritage importance. Applied
48 to the analysis of ancient Andean textiles, our results should permit a better understanding
49 of cultural and pastoral practices in South America.

50 The wild SACs are endangered by poaching and black-market sale of their fibre. For the first
51 time, our results provide discriminant features for the determination of species of origin of
52 contraband fibre.

53

54 **1. Introduction**

55 South American camelids (SACs) played a central role in the social, cultural and economic life
56 of pre-Hispanic societies [1, 2]. Modern SACs include two wild species, the guanaco (*Lama*
57 *guanicoe*) and the vicuña (*Vicugna vicugna*) and two domestic species, the llama (*Lama*
58 *glama*) and the alpaca (*Vicugna pacos*) [3]. SACs evolved in South America approximately 2
59 million years ago (Ma) from Hemiauchenia, a North American immigrant that arrived 3.3 Ma
60 [4]. The SACs are morphologically similar with inter-specific differences in coat colour and
61 size. They are adapted to the diverse climatic conditions of the Andes, ranging from the

62 extreme aridity of sea level deserts to moist tundra (*punas* at 3500-5,000 m a.s.l) and the
63 pampa grasslands with differences in specific latitudinal and vertical distribution. Two
64 subspecies each of northern and southern vicuña and guanaco have been identified [5-7].

65 The four SACs have been extensively exploited since human colonization of the continent to
66 provide goods (meat, fibre, leather, etc.) and services [1, 2]. Guanaco and vicuña were the
67 principal source of meat during the Late Pleistocene and Early Holocene [8, 9]. The alpaca
68 and llama descend from the northern vicuña (*V.v. mensalis*) and guanaco (*L.g. cacsilensis*)
69 subspecies [10] and were domesticated in the central and South-central Andes [11, 12]. They
70 were the only large domestic animals present before the Spanish Conquest in the 16th
71 century. Both domestic camelids were the principal animals used in rituals in the central
72 Andean region [1, 13]. As early as 3000 years ago camelid fibre was used to manufacture
73 textiles at archaeological sites in the Atacama desert [14] and by Inca time the products of
74 state controlled textile production comprised the currency of the empire. These textiles,
75 woven from animal and/or vegetal fibres, are markers of social rank, prestige, and power as
76 well as vectors of ideology and beliefs in South America, and thus are goods with high
77 symbolic significance [15, 16].

78 The role of domestic camelids narrowed after the 16th century in relation to the major
79 social and economic upheaval caused by the European Conquest. Spanish documents record
80 the disappearance of 90% of the domestic camelids together with a similar percentage of
81 the human population within the first century [17-19]. Nowadays, small poorly managed
82 herds dominate, producing meat largely for household consumption and fibre for sale on the
83 market. The wild vicuña and guanaco possess short, extremely fine fibre [19], and the
84 domestics have similar production quality potential but the characteristics vary according to
85 species, variety or even the altitude where the individuals live [20]. Domestic camelids
86 present a diverse range of coat colour patterns [2, 9] and different phenotypes can be
87 distinguished based on fibre characteristics for both species. There are two phenotypes of
88 alpacas, with approximately 90 % exhibiting shorter, crimped fibres of the *huacaya* type
89 while the remainder are of the *suri* type with long wavy, silky fibres [2]. At least four
90 phenotypes or varieties of llamas have been described from the central Andes [19]. Most
91 belong to the *q'ara* or non-woolly type, which is characterized by relatively sparse fibre
92 growth. In comparison, the *ch'aku* or woolly llama has a heavier fleece. Today, alpaca

93 dominates the production of SAC fibre [21, 22] while the rarity of high-priced vicuña fleece
94 makes it an expensive luxury product in international markets [23, 24].

95 The taxonomic identification of animal fibres and yarns is a major challenge in diverse
96 fields including fraud control, fashion, and forensics [25]. In archaeology, the identification of
97 ancient fibres constitutes the first step to investigate the origin of raw materials used for
98 artefact manufacturing and the cultural choices underlying this practice by ancient
99 populations [26]. Morphological characterization of the fibre can be obtained through
100 scanning electron microscopy (SEM) [27-29], but identification based on morphology can be
101 difficult, given high fibre heterogeneity. Differential scanning calorimetry (DSC) identification
102 methods have been reported, but appear unsuitable for yarns from degraded textiles or
103 processed fibres [28-30]. Molecular analyses of DNA remains have been developed as
104 complementary approaches [12, 31]. However, this approach is limited when DNA is
105 degraded due to taphonomic and diagenetic processes.

106 Proteomics has emerged as a complementary method to identify animal textile fibres
107 [32-36]. The development of paleoproteomics [37-40] applied to archaeological fibre
108 samples has permitted insight into the weaving of textiles and clothing in past societies [41-
109 48]. For human samples, hair shaft proteomics has applications in forensic as well as
110 bioarchaeology [43, 49, 50] and has recently been proposed to distinguish gender and
111 ethnicity [51]. The application of (paleo)proteomics to hair fibres faces two challenges. First,
112 fibres and fleeces are made of a series of filament-forming trichocyte keratin proteins [52,
113 53]. Keratins form coiled coil structures assembled into heteropolymeric filaments,
114 associated with a network of keratin-associated proteins (KAPs). Keratins are classified into
115 two families: type I (K31 to K40, 40–55 kDa, acidic), and type II (K81 to K87, 55–70 kDa, basic
116 or neutral). The important similarities in sequences observed for keratins within a same type
117 make it difficult to assign a peptide to a specific protein. A second challenge arises from the
118 absence of sequenced genomes for uncommon species (either wild or domestic), which
119 limits identification based on database search to cross-species proteomics [54]. This is
120 particularly true for SAC fibres, as the NCBI database contains only sequences from one SAC
121 (Tribe Lamini), the alpaca (*V. pacos*). It also contains sequences from the three Old World
122 camelid species (Tribe Camelini): the Bactrian camel (*Camelus bactrianus*), the wild Bactrian
123 camel (*Camelus ferus*) and the dromedary (*Camelus dromaderus*). Additionally, the sequence

124 information for keratins and KAPs of *C. dromaderus* and *V. pacos* is partial. Another difficulty
125 for SAC is the? extensive hybridization between the two domestic species [12]. Few studies
126 have explored hair fibre proteins from camelids, either through MALDI peptide fingerprint
127 [55] or shotgun LC-MS/MS analysis [42, 44], providing a few diagnostic peptides associated
128 to the Camelidae family.

129 Untargeted multivariate analysis of LC-MS data [56] and molecular networking, which
130 clusters ions based on the similarities in their MS/MS spectra [57, 58] have emerged as
131 powerful methods to investigate complex mixtures, mainly in metabolomics. These methods
132 have never been applied in palaeoproteomics. Here we propose to use these methods for
133 the first time to address the notoriously complex issue of fibre identification of SACs.

134

135 **2. Material and methods**

136 **2.1. Modern and archaeological samples**

137 Modern hair samples of 45 specimens (41 modern and 4 archaeological) representing
138 the four species were collected from animals in France and Peru. In total our sample includes
139 25 alpacas (57 %), 6 llamas (13%), 7 guanacos (15 %) and 7 vicuña (15%), (Table 1). Within
140 this sample, llama L1 has a llama-like phenotype and is classified as a llama, but is recognised
141 by its owner as a llama x alpaca hybrid.

142 Modern samples were taken from live animals at zoos located in France (Parc Zoologique de
143 Paris (MNHN) in Vincennes, Ménagerie du Jardin des Plantes (MNHN) in Paris, Montpellier
144 zoo); from personal fibre collections (Elise Dufour, Nicolas Goepfert and Christophe
145 Moulherat) and from the CONOPA Alpaca Genetic Diversity Project (Fig. 1; Table 1). When
146 available, information about age, sex, breed and location were recorded (Table 1).
147 Archaeological hair samples from 4 specimens of naturally desiccated mummies (2 alpacas
148 and 2 llamas) were also collected (Fig. 1; Table 1). They come from the Late Intermediate
149 (AD 950-1350) site of El Yaral [59, 60] located in the extremely arid Moquegua valley of
150 southern Peru at an elevation of 1000 m a.s.l. The mummies were identified on the basis of
151 phenotypic traits (body conformation, fibre distribution, incisor morphology) [59] and
152 detailed analysis of their fibre characteristics revealed greater fineness and uniformity than
153 in present day animals [3]. The samples were exported under Resolución Suprema No. 075-

154 96-ED of 26 September 1996.

155



156

157 **Fig. 1.** Clockwise: Hair collection on alpaca A16 Paratia, Puno, Peru and representative hair fibre
158 samples collected in this study: llama SB1 Santa Barbara, Cuzco; archaeological llama EY66 El Yaral,
159 Moquegua; alpaca A16 Paratia, Puno.

160

161 **Table 1**

162 List of the hair samples used in this study. The archaeological samples are highlighted in grey.

Species	Sample	Sex	Age	phenotype	Origin
Alpaca	A1	male		huacaya	France Ménagerie du Jardin des Plantes, Paris
<i>Vicugna pacos</i>	A2	male		huacaya	France Ménagerie du Jardin des Plantes, Paris
	A3	male		huacaya	France Ménagerie du Jardin des Plantes, Paris
	A16_PR-01	male	juvenile	huacaya	Peru Paratia, Puno
	A17_PR-20	female	juvenile	suri	Peru Paratia, Puno
	A18_PR-21	male		huacaya	Peru Paratia, Puno
	A19_PR-29	female		suri	Peru Paratia, Puno
	A20_PA-02	female		huacaya	Peru Patambuco, Puno
	A21_PA-15	female		huacaya	Peru Patambuco, Puno
	A22_PA-18	female		huacaya	Peru Patambuco, Puno
	A23_PA-28	female		suri	Peru Patambuco, Puno
	A24_CN-02	male		suri	Peru Conduriri, Puno
	A25_CN-12	male		suri	Peru Conduriri, Puno
	A26_CN-20	male		suri	Peru Conduriri, Puno
	A27_CN-28	female		suri	Peru Conduriri, Puno
	A28_MZ-07	female		suri	Peru Santa Rosa de MazoCruz, Puno
	A29_MZ-15	female		suri	Peru Santa Rosa de MazoCruz, Puno
	A30_MZ-19	male		huacaya	Peru Santa Rosa de MazoCruz, Puno
	A31_MZ-29	female		suri	Peru Santa Rosa de MazoCruz, Puno
	A32_SR-09	female		suri	Peru Santa Rosa Melgar, Puno
	A33_PS-01	male		huacaya	Peru Pisacoma, Puno

	A34_CP-15	female		huacaya	Peru	Capazo, Puno
	A35_CO-16	male		suri	Peru	Cojata, Puno
	EY54				Peru	El Yaral (AD 950-1350), Moquegua
	EY104				Peru	El Yaral (AD 950-1350), Moquegua
Llama	L2				Peru	Huaca Pucllana, Lima
<i>Lama glama</i>	L3	male	5-6 y		Peru	Huaca Pucllana, Lima
	SB1				Peru	Santa Barbara, Canchis, Cuzco
	EY66				Peru	El Yaral (AD 950-1350), Moquegua
	EY69				Peru	El Yaral (AD 950-1350), Moquegua
	L1	female		hybrid	Peru	Huaca Pucllana, Lima
Vicuña	M06189	female	10.5 y		France	Montpellier Zoo
<i>Vicugna vicugna</i>	M08089	female	9.5 y		France	Montpellier Zoo
	M15117	male	1.5 y		France	Montpellier Zoo
	ZB5218	female	21 y		France	Parc zoologique de Paris, Vincennes
	ZB5219	female	7 y		France	Parc zoologique de Paris, Vincennes
	ZB5220	male	7 y		France	Parc zoologique de Paris, Vincennes
	V1				Peru	Puno (MNHN collection Paris)
Guanaco	G1				Patagonia	San Fernando
<i>Lama guanicoe</i>	ZB4359	male	4.5 y		France	Parc zoologique de Paris, Vincennes
	ZB4360	female	2.5 y		France	Parc zoologique de Paris, Vincennes
	ZB4654	female	8 y		France	Parc zoologique de Paris, Vincennes
	ZB4656	female	13 y		France	Parc zoologique de Paris, Vincennes
	ZB5416	female	18 y		France	Parc zoologique de Paris, Vincennes
	ZB5418	female	16 y		France	Parc zoologique de Paris, Vincennes

163

164 2.2. Protein extraction and treatment

165 For each individual, about 20 mg hair was washed with milliQ water + 2% Decon™ 90
166 (Decon Laboratories) under stirring, washed 2 × 15 min in a MeOH/CHCl₃ (1:2 v/v) under
167 stirring and then washed 15 min in milliQ water under sonication. The washed fibres were
168 treated following the procedure described in [44], adapted from [35]. Briefly, the fibres were
169 extracted upon incubation under stirring for 3 days at 50°C in 3 mL 25mM Tris-HCl, 2.4 M
170 thiourea, 5M urea (pH 8.5), and dithiothreitol (DTT, 100 mM). After filtration, a fraction of
171 the protein extract (100 µL) was reduced upon 1 h incubation in the presence of DTT (5 mM)
172 and alkylated upon 1 h incubation in the dark in the presence of iodoacetamide (20 mM).
173 Finally, 1 µL of trypsin gold® (1 g/L in 50 mM acetic acid) was added, the solution was
174 completed to 1 mL with milliQ water and incubated at 37 °C for 16 h.

175 The digests were treated by solid phase extraction on a C8 cartridge (Sep-Pak C8 Plus,
176 400 mg sorbent, 37–55 µm particle size, Waters). After conditioning the cartridge with 8 mL
177 HPLC-grade CH₃CN and 4 mL milliQ-water + 0.1% formic acid (FA), the sample was loaded
178 and the cartridge was washed with 1 mL H₂O + 0.1% FA and eluted with 2 mL H₂O + 0.1% FA/

179 CH₃CN 20:80 (v/v). The samples were dried under vacuum and resuspended in 1 mL H₂O +
180 0.1% FA/ CH₃CN 70:30 (v/v).

181

182 **2.3. Analysis of the digests by mass spectrometry**

183 The tryptic digests were analysed by ultra-high-performance liquid chromatography -
184 mass spectrometry (UHPLC-MS) on an Ultimate 3000-RSLC system (Thermo Scientific)
185 connected to an electrospray ionization - quadrupole–time of flight (ESI-Q-TOF) instrument
186 (Maxis II ETD, Bruker Daltonics) following the conditions described in [44]. Briefly, the
187 separation was achieved using a RSLC Polar Advantage II Acclaim column (2.2 μm, 120 Å, 2.1
188 × 100 mm, Thermo Scientific) with a linear gradient of mobile phase A (milliQ water + FA
189 0.1%) and B (LC-MS grade CH₃CN + FA 0.08%), at 300 μL/min. The MS detection was carried
190 out in positive mode in the range *m/z* 100–2000. Analyses were performed using collision-
191 induced dissociation in data dependent auto-MS/MS mode. The MS instrument was
192 calibrated at each run start using a sodium formate solution. The injection volume was
193 adjusted (between 1 and 12 μL) in order to obtain a similar MS total ion chromatogram
194 intensity.

195

196 **2.4. Data analysis**

197 The LC-MS/MS data were analysed and converted to mgf files using Data Analysis 4.4
198 software (Bruker Daltonics). Database search was carried out with Mascot
199 (www.matrixscience.com, in-house licence), against the NCBI and SwissProt protein
200 databases. We restricted the search to Mammalia, allowed 2 missed cleavages and
201 considered Cys carbamidomethylation as a fixed modification and Asp and Glu deamidation
202 together with Met oxidation as variable modifications. Charge states 1+, 2+, and 3 + were
203 considered with 10 ppm tolerance for precursor ions and 0.05 Da tolerance for MS/MS
204 fragments. The statistic decoy was applied to minimise false-positive match. In addition, we
205 performed error-tolerant search against NCBI protein database restricted to Camelidae
206 sequences. Combined Mascot results were generated with ProteinScape Version 4.0.2
207 (Bruker Daltonics). The LC-MS/MS data and Mascot results have been deposited in the
208 ProteomeXchange Consortium via the PRIDE [61] partner repository with the dataset
209 identifier PXD018980 and 10.6019/PXD018980. The LC-MS/MS data were also analysed

210 using PEAKS DB Studio software version 7.5 (BioInformatics Solutions) [62], which performs
211 database search - assisted *de novo* sequencing. PEAKS searches (including *de novo*, PEAKS
212 PTM and SPIDER algorithms) were performed against the restricted Camelidae NCBI protein
213 database, using the same search parameters as in Mascot. The *de novo* scores were set to
214 1% FDR. We used the last proposed nomenclatures for keratins [63] and KAPs [64]. For
215 keratins, the correspondence with the previous classifications is provided in [44].

216 A list of diagnostic peptides potentially distinguishing between Camelidae species was
217 made based on the literature, alignments of the available sequences, Mascot and PEAKS
218 results. The peptides were selected based on specificity (checked by peptide alignments with
219 blastp, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>), length (longer than 6 amino acids) and
220 scores.

221 The LC-MS/MS data of a subset of representative samples from the four SAC species
222 (30 samples belonging to 17 alpacas, 5 guanacos 3 llamas and 7 vicuñas) were treated with
223 MetaboScape 3.0 software (Bruker Daltonics) to perform feature detection and quantitation
224 from MS spectra, together with feature matching across samples. A feature is defined as the
225 de-isotoped *m/z* signals belonging to the same molecule. This treatment permitted to
226 generate a multivariate matrix of 30 rows and 3062 features. The generated data matrix
227 treated was submitted to multivariate statistical analysis: principal component analysis
228 (PCA), partial least squares discriminant analysis (PLS-DA) [65] and sparse PLS-DA (sPLS-DA)
229 [66] using the package mixOmics [67] in the freely available R environment version 3.6.1
230 (www.r-project.org). sPLS-DA was carried out for 3 components with 50 variables kept per
231 component.

232 The molecular network was constructed with the Global Natural Products Social
233 Molecular Networking (GNPS) open source platform [58] using a single xml file exported
234 from MetaboScape as input, permitting to take into account retention times. The following
235 set-up was used: parent ion mass tolerance of 0.05 Da, fragment ion mass tolerance 0.05 Da,
236 cluster minimal size 1, minimum matched peaks 4, minimum cosine similarity score 0.5, MS
237 Cluster off. The resulting network was visualized using Cytoscape open source software
238 version 3.7.1 [68].

239

240 **3. Results**

241 **3.1. Targeted taxonomic peptide search**

242 First, we examined the best protein hits obtained for each of the 45 SAC samples through
243 database search (supporting information Table S1, Figures S1-S3). The best sequence
244 coverage was found for type I keratins (best hit for K33a with 41 to 74% coverage), then type
245 II keratins (best hit K83, with 25 to 69 % coverage) and finally KAPs (best hit for KAP13-1 with
246 10 to 51 % coverage). The best-hit taxonomy was always *Camelus* sp. for K33a, *Camelus* sp.
247 or *V. pacos* for K83 and always *V. pacos* for KAP13-1, which only provided unambiguous
248 identification at the family level. The available sequence referenced for K33a from *V. pacos*
249 (XP_015107542.1) is truncated (see alignment provided as supporting information Figure
250 S4), which precludes its identification in the alpaca samples.

251 The taxonomic peptide markers already reported in the literature for Camelidae [42,
252 44] (supporting information Table S2) and keratin sequence alignments (supporting
253 information Table S3) provided a list of potential diagnostic peptides for the differentiation
254 of alpaca (and potentially all the Lamini tribe) from the Camelini tribe. Two diagnostic SAC
255 peptides from keratin K86 and three from KAP13-1 were detected in our study (supporting
256 information Table S4). K86 peptide EYQEVLSNK was found mainly in the extracts from the
257 domestic species (alpaca and llama) as well one vicuña while LCEGLGAVNVCVSSSR was
258 detected in all four SAC species. KAP13-1 peptides SSFDSPNYFSSR and SWQPASFQPIFR were
259 detected in the four SAC species while CQTFQVESSPCQSSCYR was detected only in certain
260 alpaca samples. In addition, error-tolerant search permitted to propose a diagnostic peptide
261 sequence carrying a Cys to Trp substitution in KAP11-1, TYQQSWVSSCR, detected in all but
262 one of the vicuña samples and in two alpaca samples. A total of 7 diagnostic peptides were
263 identified based on database search (Table 2, supporting information Table S4). Their
264 specificity was proposed based on distribution within the samples.

265
266
267

268 **Table 2**

269 Diagnostic peptides proposed for the proteomics identification of Camelidae fibres and % occurrence
 270 per SAC species (A: alpaca, n = 25; L: llama, n = 6; V: vicuña, n = 7; G: guanaco, n = 6). Peptides in
 271 black are proposed from database search. The number of observations is based on the presence of
 272 MS/MS spectra for each sample (supporting information Table S4). Peptides in blue are proposed
 273 from untargeted approaches and molecular networking combined with *de novo* sequencing (see
 274 section 3.2). Their occurrence is based on peak detection in the LC-MS data (supporting information
 275 Table S5). Correspondence between the two modes is shown for peptide TYQQSWVSSCR from
 276 KAP11-1.

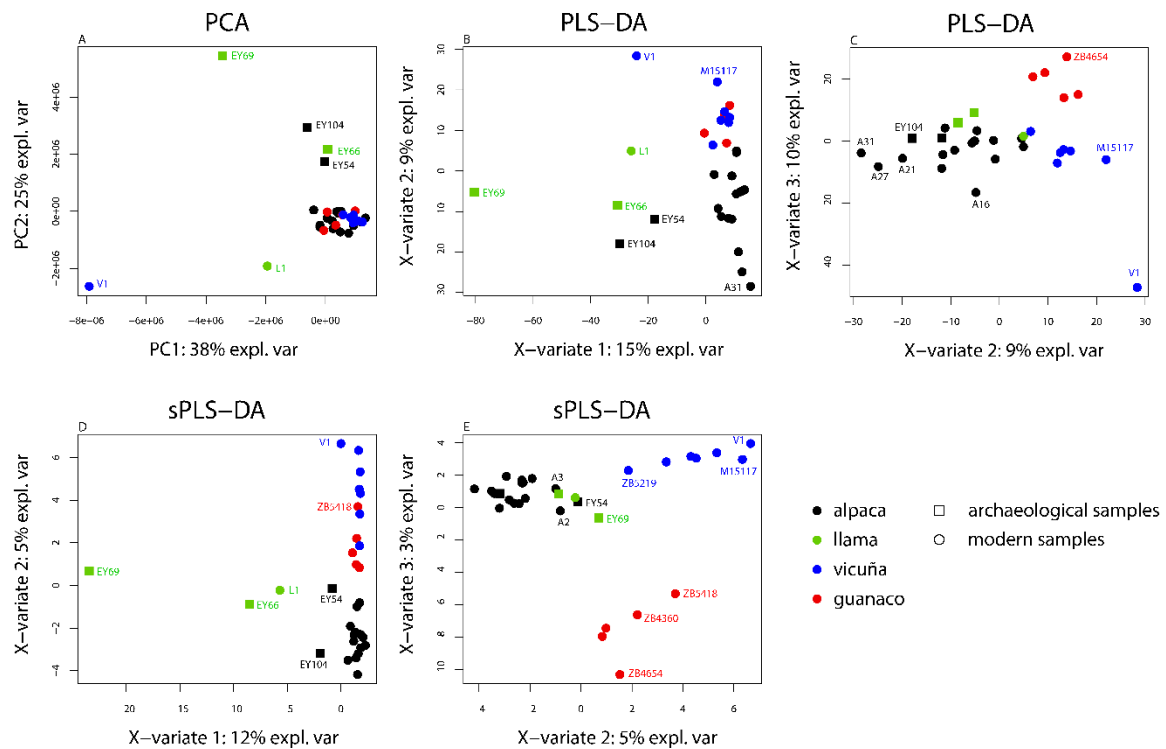
277 *These sequences can be used to differentiate between Camelidae but are shared with other
 278 mammals.

	Protein	% occurrence per SAC species	Proposed specificity	Ref.
		A L V G		
Type I keratin diagnostic peptides				
YGSQLSQVQGLITNVEHQLAEIR	K31,33,34		Camelidae.	[42, 44]
QTEELNKQVVSSSEQLQSNQAEIHELRR	K31,33,34		Camelidae. Absent in guanaco.	[44]
QVVSSEQLQSNQAEIHELRR	K31,33,34		Camelidae. Absent in guanaco.	[42]
CGPCNTFMR	K33a		Camelini.	[44]
NSLENTLTETEAR*	K33a		Camelidae. Absent in alpaca (1 exception).	This study
Type II keratin diagnostic peptides				
EYQEVLSK*	K86		Mainly domestic SAC. Absent in guanaco.	This study
LCEGLGAVNVCVSSSR	K86		SAC.	This study
DLNLDICVAIEIKEYDDIAR	K81,83,86		Camelidae.	[44]
LEAAVTQAEQQGEATVNDAR	K81,86		Camelini.	[42]
KYEEEVTLR*	K81,83		Camelini, vicuña, alpaca & hybrid llama. Absent in guanaco.	This study
KKYEEETTLR	K83		SAC. Vicuña, alpaca & hybrid llama. Absent in guanaco.	This study
KAP diagnostic peptides				
TSTLSRPCQTTYSGSLGFGSR	KAP13-1		<i>Camelus ferus</i> .	[44]
AFSSVR	KAP13-1		Camelidae.	This study
CQTFQVESSPCQSSCYR	KAP13-1		SAC. Certain alpacas.	This study
SSFDSPNYFSSR	KAP13-1		SAC.	This study
SWQPASFQPIFR	KAP13-1		SAC.	This study
TYQQSWVSSCR	KAP11-1		SAC. Vicuña, alpaca & hybrid llama. Absent in guanaco.	This study
LGHGFGFSGYGSGSGSFR	KAP19-1		SAC. Absent in vicuña.	This study
LGHGFGFSGYFGSGSGSFR	KAP19-1		Certain vicuñas.	This study

279

280 **3.2. Untargeted taxonomic peptide search**

281 The peptide markers found in the targeted approach are good indicators that
282 individual species-specific markers may exist. In order to identify the signals involved in the
283 group segregation, we performed an untargeted search and analysed the global proteomic
284 signature of a set of representative samples from each SAC species (30 samples in total). The
285 LC-MS data were converted to a multivariate matrix containing intensities of the features
286 detected (3062 variables), and multivariate analysis was performed to assess the structure
287 of the dataset and highlight the features discriminating among SAC species. The first two
288 components of the PCA explained 38 and 25% of variance, respectively (Fig. 2A). The score
289 plot showed mainly a segregation of archaeological *versus* modern samples, visible on the
290 second component. Most of the modern samples shared very close proteomic profiles, with
291 the exception of two outliers, the vicuña V1 and the hybrid llama L1 samples. The PLS-DA
292 improved the discrimination between SAC species (Fig. 2B-C). While the first component
293 (15% of explained variance) mainly segregated the archaeological and two outliers from the
294 rest, the second component (9% of explained variance) separated the wild (guanaco, vicuña)
295 from the domesticated (llama, alpaca) samples. The third component (10% of explained
296 variance) tended to separate between guanaco and vicuña. Finally, s-PLS-DA was used to
297 identify the most discriminant features between species (Fig. 2D-E). Selection of the 50 most
298 discriminant variables per component separated llama samples from all the other samples
299 along the first component, wild and domesticated individuals on the second one and
300 guanaco and vicuña on the third one.



301

302 **Fig. 2.** Bottom-up proteomic profiling of SAC hair protein extracts. Score plot of the: PCA (A), PLS-DA
 303 (B,C) and s-PLS-DA (D,E) performed from the LC-MS data. The archaeological samples are underlined.

304

305 This analysis permitted generation of a list of features of taxonomic interest

306 (supporting information Table S5). Two main types of differences were observed: (i)

307 between modern and archaeological samples and (ii) related to the taxonomy. To help in the

308 feature assignment, we generated a molecular network, which clusters ions based on the

309 similarities in their MS/MS spectra [57, 58] (Fig. 3). The ions of interest were annotated

310 through database-assisted *de novo* sequencing. This approach was useful to identify features

311 selectively detected in the archaeological samples, which consisted in highly-deamidated

312 peptides (Fig. 3A-B). In addition, it permitted to highlight important sequence similarities

313 between keratins, with peptides from type I or type II keratins forming clusters (Fig. 3 A-C).

314 Most importantly, molecular networking permitted to cluster sequences showing high

315 similarities and differentiating between species (Fig 3. D-F). As an example, the $[M+3H]^{3+}$

316 ions at m/z 660.96 and 666.29 appeared as interesting diagnostic peptides to differentiate

317 vicuña from the other SACs (Fig 3.D). Automatic database-assisted *de novo* sequencing did

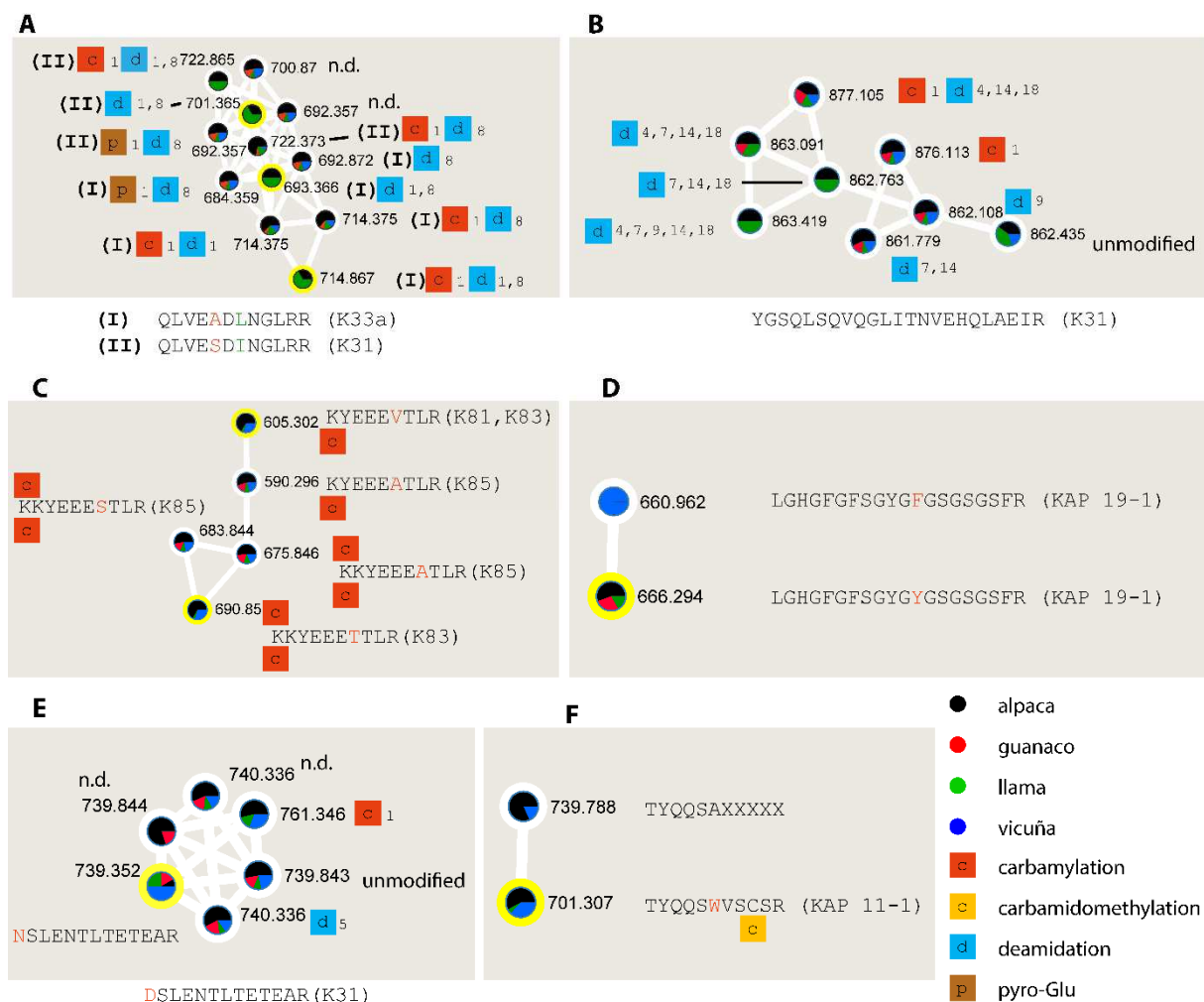
318 not provide reliable sequences. However, the similarities in the MS/MS spectra revealed by

319 the network permitted manual *de novo*-sequencing, providing two sequences varying by one

320 amino acid, attributed to KAP19-1 (Fig. 4). A substitution (Asp to Asn) was also highlighted in
 321 the peptide DSLENTLTETEAR from K31 (Fig 3.E).

322 We represented the distribution per species in the molecular network nodes based
 323 on detection of the ions in the LC-MS data (intensity above a threshold corresponding to
 324 background signal) and not based on occurrence of MS/MS data. This permitted to revise
 325 visually the distribution of diagnostic peptides proposed from database search (Table 2). For
 326 example, TYQQSWVSSCR from KAP11-1 ($[M+2H]^{2+}$ at m/z 701.31), proposed as vicuña and
 327 alpaca diagnostic peptide, was confirmed to be specific to these two species, but also
 328 appeared in sample L1, corresponding to the llama hybrid (Fig. 3F). This peptide clustered
 329 with a $[M+2H]^{2+}$ ion at m/z 739.79, also specific to vicuña and alpaca, assigned to the partial
 330 sequence TYQQA by *de novo* sequencing.

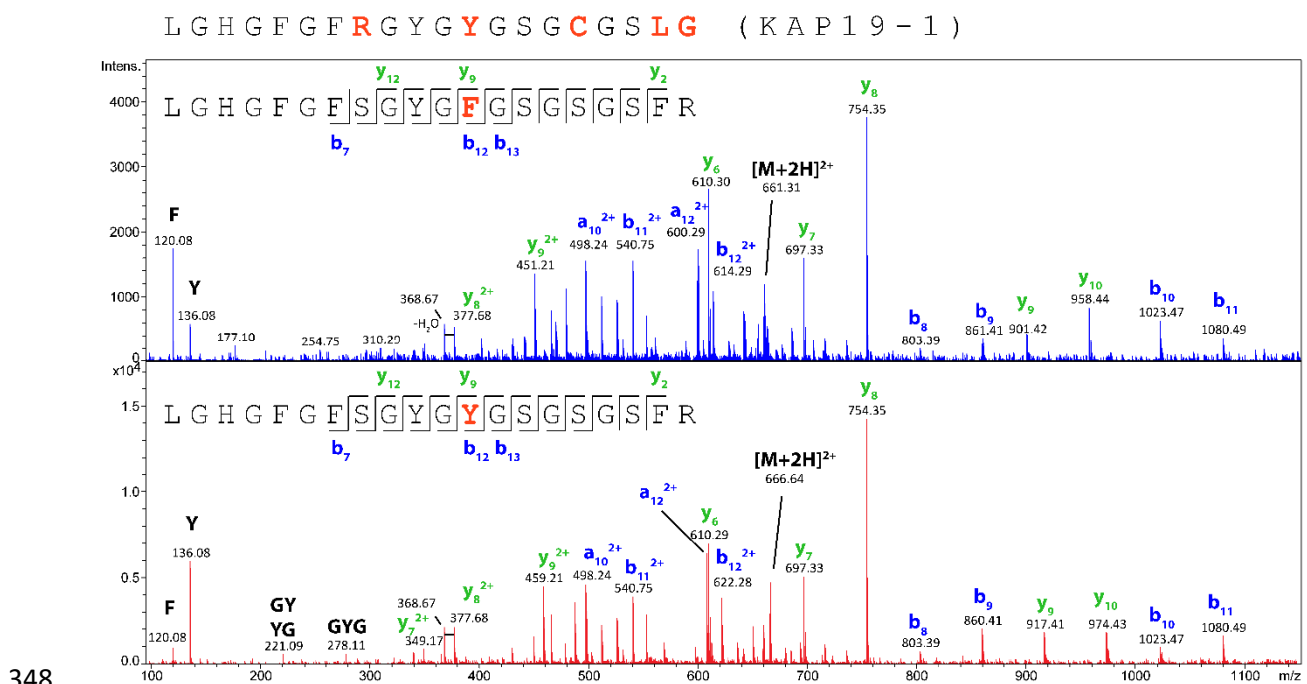
331



332

333 **Fig. 3.** Molecular network of bottom-up proteomic data of SAC hair protein extracts. Selected
 334 clusters of features selected from s-PLS-DA (circled in yellow) or molecular network analysis,
 335 attributed to differences between modern and archaeological samples (A,B), inter-chain keratin

336 sequence similarities (A,C) or differences related to taxonomy (D, E, F). The nodes are annotated with
 337 peptide sequences and modifications. (A) Cluster assigned to sequences QLVEADLNGLR from K33a
 338 and QLVESDINGLR from K31, carrying different modifications; (B) Cluster assigned to sequence
 339 YGSQLSQVQLITNVEHQLAEIR from K31 harbouring different extent of deamidation. In (A,B), the
 340 deamidated peptides are over-represented in the alpaca and llama archaeological samples. (C)
 341 Cluster assigned to highly similar peptides from type-2 keratins; (D) Cluster of features discriminating
 342 vicuña from the three other SAC species, assigned to sequences attributed to KAP 19-1, determined
 343 by manual *de novo* sequencing; (E) Cluster assigned to peptide DSLENTLTETEAR from K31, permitting
 344 to propose an Asp to Asn substitution mainly in non-alpaca species; (F) Cluster assigned to a peptide
 345 from KAP 11-1 with a Cys to Trp in substitution and related sequence determined by *de novo*
 346 sequencing, over-represented in alpaca and vicuña.
 347



349 **Fig. 4.** MS/MS spectra of the $[M+2H]^{2+}$ species at m/z 660.96 (vicuña sample M06189) and m/z
 350 666.29 (guanaco sample ZB5418) clustered in Fig. 3D, used for manual *de novo* sequencing. The N-
 351 terminal segment was proposed from the sequence of KAP 19-1 from *V. pacos*, shown above the
 352 spectra.
 353

354 Combining untargeted analysis, molecular networking and identification through
 355 database-assisted *de novo* sequencing permitted proposal of 5 new diagnostic peptides
 356 (Table 2). In addition to TYQQSWVSSCR from KAP11-1, 4 peptides permitted differentiation
 357 between the two wild species, guanaco and vicuña. KYEEVTLR and KKYEEETTLR from type II
 358 keratins were absent in guanaco, LGHGFGFSGYGYGSGSGSFR from KAP1-91 was absent in
 359 vicuña and the Tyr to Phe substituted LGHGFGFSGYGFSGSGSFR was detected in 2 out of 7
 360 vicuña samples. Note that the diagnostic peptides KYEEVTLR and KKYEEETTLR from type II

361 keratins and TYQQSWVSSCR from KAP11-1 were not detected in llama samples with the
362 exception of the hybrid llama.

363

364 **4. Discussion**

365 **4.1. Combining targeted and untargeted approaches in paleoproteomics**

366 The species-specific identification of SAC hair keratin proved challenging. Database-
367 search approaches, which are usually used in paleoproteomics [45], permitted identification
368 of diagnostic peptides informative at the tribe level, *i.e.*, differentiating Camelini and Lamini
369 but failed to differentiate within Lamini. This illustrates the difficulty of proteomic-based
370 identification at the species level when information in the databases is missing or partial.
371 The sequence information available for keratins and KAPs is particularly poor for SAC, as
372 compared to other clades (supporting information Table S6). In this context, the use of
373 untargeted approaches is relevant because it permitted us to highlight discriminant signals,
374 some of which could be subsequently identified by *de novo* sequencing.

375 The untargeted multivariate analysis of proteomics data is known to be a powerful
376 method to hunt for biomarkers [69] or perform high-throughput analysis of post-
377 translational modifications in proteins [70]. For the differentiation between animal hair
378 fibres, hierarchical clustering [48], PCA and PLS based on MALDI-TOF peptide fingerprint [55]
379 as well as quantitative proteomics [36] can provide signatures discriminating between
380 species. Here we confirm the potential of such approaches for (archaeo)zoological
381 applications on SAC fibres. One limitation of this approach is the difficulty of identifying the
382 discriminant signals. In addition, for paleoproteomics, the mass shifts resulting from
383 degradation such as deamidation generates multiple signals for a single peptide, resulting in
384 different signatures between altered and unaltered samples. This trend is illustrated here
385 with the PCA that segregates mainly modern and archaeological samples. In this context,
386 molecular networking provides pertinent information for the identification of discriminant
387 features and the clustering of a given peptide with its modified forms.

388 Molecular networking based on MS/MS data has emerged in the last ten years as a
389 powerful method to investigate complex mixtures, applied mainly to metabolomics and
390 natural product discovery [58, 71]. It has been used for proteo(peptido)genomic approaches
391 for the discovery of peptidic natural products [72, 73], but never for paleoproteomics. Here

392 we show that it serves to highlight modifications related to diagenesis, such as deamidation,
393 confirming its potential to detect peptide degradation products in complex biomatrices [74].
394 This circumvents the difficulty of interpreting multivariate data analysis of samples with
395 different degrees of alteration. In addition, this method permits clustering of peptides with
396 high sequence similarities. The clustered peptides can be highly similar fragments from
397 different keratin chains or result from amino acid substitutions related to intra- or inter-
398 specific polymorphism (Fig. 3). As for natural product discovery, molecular networking
399 applied to bottom-up proteomics is very useful to identify ions of interest. We were able to
400 perform manual *de novo* sequencing based on similarities in different MS/MS spectra
401 highlighted by the molecular network (Fig. 4). Using multiple MS/MS data from highly similar
402 peptides allows *de novo* sequencing to be greatly simplified through matching of the b- and
403 y-series between spectra. An automation of this process would benefit palaeoproteomic
404 research.

405

406 **4.2. Characterization at the family and tribe level**

407 Our work demonstrates the existence of genetic markers in Camelidae and the
408 potential of keratin proteomics to retrieve phylogenetic information. The existence of
409 genetic differences between the Camelini and Lamini is not surprising given their
410 evolutionary history and ancestor separation in the Late Oligocene - Late Miocene (~16-17
411 Ma) [75]. The Camelidae originated in North America about 45 Ma and collagen
412 paleoproteomics has helped to infer the phylogenetic relationships of these extinct [76].
413 Approximately 3 Ma, some species of Lamini migrated to South America (67) where they
414 gave rise to the SACs [77]. The presence of peptide markers revealed here suggests that
415 proteomics might be useful to explore past diversity that was larger than at present and
416 taxonomic relationships among SACs. However, this approach relies on the survival of
417 molecules in ancient remains and because keratin has less chance of surviving in the fossil
418 record than collagen and other bone proteins keratin proteomics might be restricted to
419 recent and modern applications.

420 Our proteomics results show for the first time specific markers for the two wild
421 species of Lamini. This is in agreement with the mitochondrial and nuclear DNA evidence
422 that documents two separate genera [4, 78, 79] with no evidence for hybridization between

423 the wild guanaco and vicuña even when individuals of both genera are sympatric [78].
424 Regarding the domestic species, although we identified markers that allow to distinguish
425 some of the llamas from the alpacas, we also found that the two species share various
426 specific sequences and thus are not clearly separated. This is in keeping with the lack of
427 agreement arising from multivariate analysis as well as morphological observations leading
428 to contradictory hypotheses regarding the phylogeny of domestic species [3]. For instance in
429 1988 a molecular study of antibodies suggested alpaca and llama were closer to each other
430 [80] than to their respective wild ancestor identified via nuclear DNA [4]. Confusion might
431 partly result from low genetic divergence. All SACs share the same karyotype ($2n = 74$) and
432 can produce fertile hybrids under natural or human controlled breeding. Although there is a
433 general difference in size (the smallest is the vicuña and the largest the llama), species-
434 specific identification from skeletal examination is acknowledged to be difficult, due to
435 proximity in morphological conformation of llama and alpaca [11]. Fibre morphology and
436 diameter present general distinctive characteristics but they tend to overlap among species
437 [28].

438 Overall, hybridization might be the most important factor accounting for the long
439 standing debate about llama and alpaca ancestry and also for the results of the present
440 study. Prior to the conquest, controlled breeding produced high quality fibre producing
441 alpacas and llamas [19, 81], but today 90% of Peruvian alpacas and most llamas are kept in
442 mixed herds of fewer than 100 animals and selective reproduction is generally not practiced.
443 DNA analysis documents that >80% of alpacas and >40% of llamas in the Andes are hybrids
444 [79]. Crossbreeding between alpacas and llamas is practiced by some herders to produce
445 hybrids with better fibre quality or higher fleece weight to increase income. Visual
446 identification of hybrids is uncertain, although Quechua speaking herders refer to them as
447 wari, distinguishing between llama wari and alpaca wari based on appearance or known
448 breeding history [2]. One of our study animals (L1) was identified as a llama wari and was the
449 only “llama” sampled with peptides KYEEVTLR, KKYEETTLR, and TYQQSWVSSCR that may
450 be specific to alpacas and vicuñas. Therefore, these peptides could be markers of the
451 *Vicugna* genus. The second peptide is particularly interesting as it was found in all the alpaca
452 and vicuña samples. It would be important to analyse more llama samples, especially DNA
453 tested non-hybrids if possible, to test this hypothesis.

454 The proposed diagnostic peptides were only rarely detected for every sample
455 assigned to a given species. This could be related to a limit of detection in the LC-MS/MS
456 analysis or result from intraspecific variability or hybridization. A wide sampling of
457 individuals from the four SAC species would be important to validate the proposed
458 diagnostic markers and search for ones related to subspecies and/or phenotypes, in
459 particular within the wild species.

460

461 **4.3. Consideration for future SAC fibre analysis**

462 The taxonomic identification of fibres is primarily used in fraud control and
463 archaeozoology. The bottom-up proteomic signature (Fig. 2) and diagnostic peptides
464 proposed in this study for the differentiation of Camelidae (Table 2) should be of great
465 interest in these two contexts. In every case, species identification should rely on several
466 markers to prevent false results resulting from hybridization between the domestic species.

467 All SACs have high quality fibre, but vicuña fibres are recognized as the best by the
468 textile industry. Fibre price depends on its quality (fineness, absence of prickle, length) as
469 well as fleece characteristics (colour, single or double coat type) [22] and productivity.
470 Because vicuña produces the finest fibre it was poached almost to extinction, before
471 receiving protection [82]. Even now, despite regulated legal shearing of live vicuña and sale
472 of this fibre, illegal hunting and black-market sale of vicuña fibre continues to flourish.
473 Guanaco fleece is double-coated with undercoat hair nearly as fine as vicuña. It has also
474 been threatened by extinction due to overhunting. Commerce in both vicuña and guanaco
475 fibre and high end luxury products is regulated by CITES, but enforcement is difficult given
476 the elevated value and flourishing illegal trade. The diagnostic peptides identified in this
477 study to differentiate guanaco and vicuña fibre should be highly valuable in this context. In
478 particular, the diagnostic peptides LGHGFGFSGYGFGSGSGSFR / LGHGFGFSGYGFGSGSGSFR
479 from K19-1 should allow absolute identification of vicuña fibre to ensure honesty in labelling
480 of finished products, and to document illegal movement of mislabelled fibre shipments
481 where poached vicuña fleece is passed as alpaca fibre.

482 The specific determination of hair found in archaeological contexts will constitute a
483 major step forward in understanding the use and role of SACs in the past without actualistic
484 preconceived ideas. It is generally assumed that the present distribution and production

485 characteristics of SACs were the same in the past. Some scholars have claimed that in the
486 past only alpaca from the highlands were able to produce high quality fibre [20], but this
487 assertion has been reassessed. The analysis of fibre from naturally mummified pre-Inca
488 alpacas and llamas reared at 1,000 m a.s.l in the Moquegua valley at El Yaral, has revealed
489 the existence of both alpacas and llamas with higher quality fibre than is generally found in
490 modern ones (3). This shows, as well as for other aspects of SAC uses and human practices,
491 such as pastoralism [83], that the old paradigm needs to be questioned. The loss of verbally
492 transmitted camelid breeding knowledge and marginalization of the few surviving llamas
493 and alpacas in favour of European origin livestock led to the relegation of traditional herding
494 to remote, inhospitable regions and loss of livestock quality. The four archaeological samples
495 studied in the present study were well preserved and could be used to approach such
496 questions despite deamidation. This is an encouraging step for the future study of
497 manufactured products (clothing, coca bags, etc.) that are found in many Andean
498 archaeological contexts. Culminating with the Inca, textiles produced from SAC fibre defined
499 the economy and influenced hunting and herding practices and priorities. Paleoproteomics
500 will be useful for determining which SAC species were exploited to produce the yarns.
501 Moreover, the archaeological material constitutes very precious reference samples for the
502 search of taxonomic Lamini markers as they are not affected by genetic consequences of
503 post-Conquest hybridization.

504

505 **5. Conclusion**

506 Hair fibre palaeoproteomics is still a developing research field and presents
507 difficulties compared to that of bone. In particular, the multiple homologous keratin
508 sequences make protein inference difficult, and for the SAC, poor sequence information
509 available in databases and important hybridization between species further complicate
510 interpretation. Here, a combination of different approaches including database search,
511 multivariate data analysis, molecular networking and database-assisted *de novo* sequencing
512 of modern and archaeological samples allowed us to circumvent some of these issues and
513 provide insights for future research.

514 Our study is the first to examine all four SAC species at the same time. It provides a
515 bottom-up proteomic signature of the four species as well as new diagnostic peptides for the

516 identification of Camelidae. Good sequence coverage obtained from archaeological hair
517 opens the possibility of studying fibres collected from fleeces, yarns and ancient textiles
518 preserved thanks to the climatic conditions found in some regions of South America. Our
519 work also suggests that more attention should be paid to KAPs to differentiate between
520 closely related species. Although KAPs are less abundant in fibres than keratins, they were
521 detected in all of our samples and provided various taxonomic peptides. Although these
522 proteins are rarely investigated, they have been shown to be linked to the mechanical
523 properties of wool [84] and to survive long periods of burial in specific conditions [85],
524 allowing work on degraded fibres.

525 Finally, this study highlights the potential of molecular networks for
526 (paleo)proteomics. This method can be very valuable to analyse proteomics data from
527 organisms with no genomic information, to evaluate non-silent polymorphism and to
528 understand post-translational modifications and alterations related to diagenesis.

529

530 **Declaration of Competing Interest**

531 The authors declare no conflict of interest.

532

533 **Funding**

534 This work was supported by a grant from Agence Nationale de la Recherche under the LabEx
535 ANR-10-LABX-0003-BCDiv, in the program “Investissements d'avenir” n\u0001 ANR-11-IDEX-
536 0004-02 and an Action transversale du Muséum grant (ATM blanche, project DCAPG - 2016)
537 from the MNHN. The Alpaca Genetic Diversity project was supported by the Cloud Hollow
538 Educational Fund, Maine Community Foundation, and research on the El Yaral mummies by
539 US National Science Foundation grant 9321377.

540

541 **Acknowledgments**

542 We are grateful to all the persons who helped in the hair collection: Norin Chai and his team
543 at the Ménagerie du Jardin des plantes, Alexis Lécu, Nathanie Dogna and their colleagues at
544 the Parc Zoologique de Paris, Baptiste Chenet and his team at the zoo of Montpellier, Isabel
545 Flores (Directora Huaca Pucllana project), Christophe Moulherat (Musée du Quai Branly) and
546 Pr. Mike Bruford (Cardiff University). We thank the Analytical Platform of the Muséum
547 national d'Histoire naturelle (Paris, France) for access to the mass spectrometry facilities. We

548 acknowledge Mélanie Laborde for the graphical abstract drawing. Finally, we thank the
549 Labex BCDiv (ANR) and the ATM for financial support. We also thank Bob Weintraub (Cloud
550 Hollow Educational Fund) for making possible the alpaca research in Puno.

551

552 **References**

- 553 [1] D. Bonavia, Los camélidos sudamericanos: Una introducción a su estudio, Institut français
554 d'études andines, Lima, 1996.
- 555 [2] J.A. Flores Ochoa, K. MacQuarrie, J. Blassi, J. Portús Pérez, J. Blassi, Oro de los Andes: Las llamas,
556 alpacas, vicuñas y guanacos de Sudamérica, BPR Publishers, Barcelona, 1994.
- 557 [3] J.C. Wheeler, South American camelids - past, present and future, *Journal of Camelid Science* (5)
558 (2012) 1-24.
- 559 [4] M. Kadwell, M. Fernandez, H.F. Stanley, R. Baldi, J.C. Wheeler, R. Rosadio, M.W. Bruford, Genetic
560 analysis reveals the wild ancestors of the llama and the alpaca, *Proceedings. Biological sciences*
561 268(1485) (2001) 2575-84.
- 562 [5] C.S. Casey, P. Orozco-terWengel, K. Yaya, M. Kadwell, M. Fernandez, J.C. Marín, R. Rosadio, L.
563 Maturrano, D. Hoces, Y. Hu, J.C. Wheeler, M.W. Bruford, Comparing genetic diversity and
564 demographic history in co-distributed wild South American camelids, *Heredity* 121(4) (2018) 387-
565 400.
- 566 [6] J.C. Marín, C.S. Casey, M. Kadwell, K. Yaya, D. Hoces, J. Olazabal, R. Rosadio, J. Rodriguez, A.
567 Spotorno, M.W. Bruford, J.C. Wheeler, Mitochondrial phylogeography and demographic history of
568 the vicuna: implications for conservation, *Heredity* 99(1) (2007) 70-80.
- 569 [7] B.A. González, R.E. Palma, B. Zapata, J.C. Marín, Taxonomic and biogeographical status of guanaco
570 *Lama guanicoe* (Artiodactyla, Camelidae), *Mammal Review* 36(2) (2006) 157-178.
- 571 [8] G. Custred, Hunting technologies in Andean culture, *Journal de la Société des américanistes* 66
572 (1979) 7-19.
- 573 [9] J.M. Capriles, N. Tripcevich, *The Archaeology of Andean Pastoralism*, UNM Press 2016.
- 574 [10] J.C. Marín, K. Romero, R. Rivera, W.E. Johnson, B.A. González, Y-chromosome and mtDNA
575 variation confirms independent domestications and directional hybridization in South American
576 camelids, *Animal genetics* 48(5) (2017) 591-595.
- 577 [11] G. Mengoni Gonalons, H. Yacobaccio, The domestication of South American Camelids - a view
578 from the South-Central Andes, in: M.A. Zeder (Ed.), *Documenting domestication: new genetic and*
579 *archaeological paradigms*, University of California Press, Berkeley, 2006, pp. 228-244.
- 580 [12] J.C. Wheeler, L. Chikhi, M.W. Brudford, Genetic Analysis of the Origins of Domestic South
581 American Camelids, in: M.A. Zeder, D.G. Bradley, E. Emshwiller, B.D. Smith (Eds.), *Documenting*
582 *Domestication: New Genetic and Archaeological Paradigms*, University of California Press 2006, pp.
583 329-341.
- 584 [13] N. Goepfert, New Zooarchaeological and Funerary Perspectives on Mochica Culture (100–800
585 A.D.), Peru, *Journal of Field Archaeology* 37 (2012) 104-120.
- 586 [14] P.Z. Dransart, Llamas, herders and the exploitation of raw materials in the Atacama Desert,
587 *World Archaeology* (22) (1991) 304-319.
- 588 [15] P. Eeckhout, Textiles and burial customs, in: S.L. (Dir) (Ed.), *INCA. Textiles and ornaments of the*
589 *Andes*, Ludion, Brussels, 2018, pp. 71-90.
- 590 [16] E. Philipps, Garment traditions of the Andes: construction, gender and identity, in: S.L. (Dir) (Ed.),
591 *INCA. Textiles and ornaments of the Andes*, Ludion, Brussels, 2018, pp. 71-90.
- 592 [17] J.A. Flores Ochoa, Classification et dénomination des camélidés sud-américains, *Annales*
593 *Economies Sociétés Civilisations* 33(5-6) (1978) 1006-1016.

594 [18] S. Wachtel, *La vision des vaincus. Les Indiens du Pérou devant la Conquête espagnole*, Nathan,
595 Paris, 1992.

596 [19] J.C. Wheeler, *Superando la Conquista, el caso de los Camélidos Sudamericanos*, in: R.B. J Avila
597 Moreno, Eds (Ed.), *Antropología Aplicada en el Perú de Hoy*, Colegio Profesional de Antropólogos del
598 Perú, Lima, 2016, pp. 85-99.

599 [20] T.L. Topic, T.H. McGreevy, J.R. Topic, *A Comment on the Breeding and Herding of Llamas and*
600 *Alpacas on the North Coast of Peru*, *American Antiquity* 52(4) (1987) 832-835.

601 [21] B.A. McGregor, *The quality of fibre grown by Australian alpacas : part 1 : the commercial quality*
602 *attributes and value of alpaca fibre*, *IAI 1997 : Shaping the future : Proceedings of the International*
603 *Alpaca Industry 1997 seminar* (1997) 43-48.

604 [22] E.N. Frank, M.V.H. Hick, C.D. Gauna, H.E. Lamas, C. Renieri, M. Antonini, *Phenotypic and genetic*
605 *description of fibre traits in South American domestic camelids (llamas and alpacas)*, *Small Ruminant*
606 *Research* 61(2-3) (2006) 113-129.

607 [23] M.A. Gardetti, S.S. Muthu, *Handbook of Sustainable Luxury Textiles and Fashion*, Springer2015.

608 [24] E. Posnett, *Harvest: The Hidden Histories of Seven Natural Objects*, The Bodley Head, London,
609 2019.

610 [25] M.M. Houck, *Identification of textile fibers*, Reprint ed., CRC Press, Boca Raton, FL, 2010.

611 [26] I. Good, *Archaeological textiles: A review of current research*, *Annu. Rev. Anthropol.* 30 (2001)
612 209-226.

613 [27] R.R. Franck, *Silk, Mohair, Cashmere and Other Luxury Fibres*, Elsevier2001.

614 [28] A. Rast-Eicher, *Fibres: microscopy of archaeological textiles and furs*, *Archaeolingua*, Budapest,
615 2016.

616 [29] A.B. Wildman, *The Microscopy of Animal Textile Fibres: Including Methods for the Complete*
617 *Analysis of Fibre Blends. 235 Half-tone and LI Colour Photomicrographs, 88 Line Drawings*, *Wool*
618 *Industries Research Association*1954.

619 [30] C. Vineis, A. Aluigi, C. Tonin, *Outstanding traits and thermal behaviour for the identification of*
620 *speciality animal fibres*, *Textile Research Journal* 81(3) (2011) 264-272.

621 [31] CIS, *Camelid Identification System*. <http://cisdna.com/>.

622 [32] J.E. Plowman, *The proteomics of keratin proteins*, *Journal of chromatography. B, Analytical*
623 *technologies in the biomedical and life sciences* 849(1-2) (2007) 181-9.

624 [33] S. Clerens, C.D. Cornellison, S. Deb-Choudhury, A. Thomas, J.E. Plowman, J.M. Dyer, *Developing*
625 *the wool proteome*, *J Proteomics* 73(9) (2010) 1722-31.

626 [34] J.E. Plowman, S. Deb-Choudhury, S. Clerens, A. Thomas, C.D. Cornellison, J.M. Dyer, *Unravelling*
627 *the proteome of wool: towards markers of wool quality traits*, *J Proteomics* 75(14) (2012) 4315-24.

628 [35] S. Paoella, M. Bencivenni, F. Lambertini, B. Prandi, A. Faccini, C. Tonetti, C. Vineis, S. Sforza,
629 *Identification and quantification of different species in animal fibres by LC/ESI-MS analysis of keratin-*
630 *derived proteolytic peptides*, *Journal of mass spectrometry : JMS* 48(8) (2013) 919-26.

631 [36] S. Li, Y. Zhang, J. Wang, Y. Yang, C. Miao, Y. Guo, Z. Zhang, Q. Cao, W. Shui, *Combining*
632 *Untargeted and Targeted Proteomic Strategies for Discrimination and Quantification of Cashmere*
633 *Fibers*, *PLOS ONE* 11(1) (2016) e0147044.

634 [37] E. Cappellini, M.J. Collins, M.T.P. Gilbert, *Biochemistry. Unlocking ancient protein palimpsests*,
635 *Science (New York, N.Y.)* 343(6177) (2014) 1320-1322.

636 [38] S. Dallongeville, N. Garnier, C. Rolando, C. Tokarski, *Proteins in Art, Archaeology, and*
637 *Paleontology: From Detection to Identification*, *Chemical Reviews* 1(116) (2016) 2-79.

638 [39] T.P. Cleland, E.R. Schroeter, *A comparison of common mass spectrometry approaches for*
639 *paleoproteomics*, *J Proteome Res* 17(3) (2018) 936-945.

640 [40] J. Hendy, F. Welker, B. Demarchi, C. Speller, C. Warinner, M.J. Collins, *A guide to ancient protein*
641 *studies*, *Nature Ecology & Evolution* 2(5) (2018) 791-799.

642 [41] C. Solazzo, S. Heald, M.W. Ballard, D.A. Ashford, P.T. DePriest, R.J. Koestler, M.J. Collins,
643 *Proteomics and Coast Salish blankets: A tale of shaggy dogs?*, *Antiquity* 85(330) (2011) 1418-1432.

644 [42] C. Solazzo, M. Wadsley, J.M. Dyer, S. Clerens, M.J. Collins, J. Plowman, *Characterisation of novel*
645 *α-keratin peptide markers for species identification in keratinous tissues using mass spectrometry:*

646 Novel α -keratin peptide markers for species identification, *Rapid Communications in Mass*
647 *Spectrometry* 27(23) (2013) 2685-2698.

648 [43] M. Fresnais, P. Richardin, M. Sepúlveda, E. Leize-Wagner, A. Charrié-Duhaut, Omics for precious
649 rare biosamples: Characterization of ancient human hair by a proteomic approach, *OMICS: A Journal*
650 *of Integrative Biology* 21(7) (2017).

651 [44] C. Azémard, A. Zazzo, A. Marie, S. Lepetz, C. Debaine-Francfort, A. Idriss, S. Zirah, Animal fibre
652 use in the Keriya valley (Xinjiang, China) during the Bronze and Iron Ages: A proteomic approach,
653 *Journal of Archaeological Science* 110 (2019) 104996.

654 [45] C. Solazzo, Characterizing historical textiles and clothing with proteomics, *Conservar Património*
655 31 (2019) 97-114.

656 [46] K. Hollemeyer, W. Altmeyer, E. Heinzle, Identification and quantification of feathers, down, and
657 hair of avian and mammalian origin using matrix-assisted laser desorption/ionization time-of-flight
658 mass spectrometry, *Anal Chem* 74(23) (2002) 5960-8.

659 [47] K. Hollemeyer, W. Altmeyer, E. Heinzle, C. Pitra, Species identification of Oetzi's clothing with
660 matrix-assisted laser desorption/ionization time-of-flight mass spectrometry based on peptide
661 pattern similarities of hair digests, *Rapid Communications in Mass Spectrometry* 22(18) (2008) 2751-
662 2767.

663 [48] K. Hollemeyer, W. Altmeyer, E. Heinzle, C. Pitra, Matrix-assisted laser desorption/ionization
664 time-of-flight mass spectrometry combined with multidimensional scaling, binary hierarchical cluster
665 tree and selected diagnostic masses improves species identification of Neolithic keratin sequences
666 from furs of the Tyrolean Iceman Oetzi: Species identification: Neolithic keratin sequences, *Rapid*
667 *Communications in Mass Spectrometry* 26(16) (2012) 1735-1745.

668 [49] Y.J. Lee, R.H. Rice, Y.M. Lee, Proteome analysis of human hair shaft: from protein identification
669 to posttranslational modification, *Molecular & cellular proteomics : MCP* 5(5) (2006) 789-800.

670 [50] G.J. Parker, T. Leppert, D.S. Anex, J.K. Hilmer, N. Matsunami, L. Baird, J. Stevens, K. Parsawar,
671 B.P. Durbin-Johnson, D.M. Rocke, C. Nelson, D.J. Fairbanks, A.S. Wilson, R.H. Rice, S.R. Woodward, B.
672 Bothner, B.R. Hart, M. Leppert, Demonstration of Protein-Based Human Identification Using the Hair
673 Shaft Proteome, *PLOS ONE* 11(9) (2016) e0160653.

674 [51] N. Mohamed Nasir, J. Hiji, J.J. Jayapalan, O.H. Hashim, Potential use of human hair shaft keratin
675 peptide signatures to distinguish gender and ethnicity, *Peer J.* 8 (2020) e8248.

676 [52] R.C. Marshall, D.F.G. Orwin, J.M. Gillespie, Structure and biochemistry of mammalian hard
677 keratin, *Electron Microscopy Reviews* 4(1) (1991) 47-83.

678 [53] J.E. Plowman, D.P. Harland, S. Deb-Choudhury, *The hair fibre: Proteins, structure and*
679 *development*, Springer, Singapore, 2018.

680 [54] J.C. Wright, R.J. Beynon, S.J. Hubbard, Cross species proteomics, *Methods Mol Biol* 604 (2010)
681 123-35.

682 [55] Y. Izuchi, M. Tokuhara, T. Takashima, K. Kuramoto, Peptide profiling using matrix-assisted laser
683 desorption/ionization-time-of-flight mass spectrometry for identification of animal fibers, *Mass*
684 *Spectrometry* 2(1) (2013).

685 [56] A.C. Schrimpe-Rutledge, S.G. Codreanu, S.D. Sherrod, J.A. McLean, Untargeted metabolomics:
686 Strategies, challenges and emerging directions, *Journal of the American Society for Mass*
687 *Spectrometry* 27(12) (2016) 1897-1905.

688 [57] J.Y. Yang, L.M. Sanchez, C.M. Rath, X. Liu, P.D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. de
689 Felicio, A. Fenner, W.R. Wong, R.G. Linington, L. Zhang, H.M. Debonsi, W.H. Gerwick, P.C. Dorrestein,
690 Molecular networking as a dereplication strategy, *Journal of natural products* 76(9) (2013) 1686-99.

691 [58] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A.
692 Kapon, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.T. Liu, M.
693 Crusemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R.D. Kersten, L.A. Pace, R.A. Quinn,
694 K.R. Duncan, C.C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrew, T. Northen, R.J. Dutton, D. Parrot, E.E.
695 Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel,
696 B.T. Murphy, L. Gerwick, C.C. Liaw, Y.L. Yang, H.U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R.
697 Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A.B. P, D. Torres-Mendoza, D.J.

698 Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich,
699 E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt,
700 K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N.
701 Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman,
702 V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodriguez, A. Lamsa, C. Zhang, K.
703 Dorrestein, B.M. Duggan, J. Almaliti, P.M. Allard, P. Phapale, L.F. Nothias, T. Alexandrov, M. Litaudon,
704 J.L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller,
705 K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B.O. Palsson, K. Pogliano, R.G. Lington,
706 M. Gutierrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, Sharing and
707 community curation of mass spectrometry data with Global Natural Products Social Molecular
708 Networking, *Nature biotechnology* 34(8) (2016) 828-837.

709 [59] J.C. Wheeler, A.J.F. Russel, H. Redden, Llamas and Alpacas: Pre-conquest breeds and post-
710 conquest hybrids, *Journal of Archaeological Science* 22(6) (1995) 833-840.

711 [60] M.C. Lozada, J.E. Buikstra, G. Rakita, J.C. Wheeler, Camelid herders: The forgotten specialists in
712 the Coastal Señorío of Chiribaya, Southern Peru, in: E. J. Marcus and P.R. Williams (Ed.), *Andean*
713 *Civilization: A Tribute to Michael E. Moseley*, UCLA Cotsen Institute of Archaeology Press 2009, pp.
714 351-364.

715 [61] J.A. Vizcaino, A. Csordas, N. Del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol,
716 F. Reisinger, T. Ternent, Q.W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its
717 related tools, *Nucleic acids research* 44(22) (2016) 11033.

718 [62] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, PEAKS: Powerful software
719 for peptide de novo sequencing by MS/MS. , *Rapid Comm. Mass Spectrom.* 17(20) (2003) 2337-2342.

720 [63] J. Schweizer, P.E. Bowden, P.A. Coulombe, L. Langbein, E.B. Lane, T.M. Magin, L. Maltais, M.B.
721 Omary, D.A.D. Parry, M.A. Rogers, M.W. Wright, New consensus nomenclature for mammalian
722 keratins, *The Journal of Cell Biology* 174(2) (2006) 169-174.

723 [64] H. Gong, H. Zhou, G.W. McKenzie, Z. Yu, S. Clerens, J.M. Dyer, J.E. Plowman, M.W. Wright, R.
724 Arora, C.S. Bawden, Y. Chen, J. Li, J.G.H. Hickford, An updated nomenclature for keratin-associated
725 proteins (KAPs), *International Journal of Biological Sciences* 8(2) (2012) 258-264.

726 [65] M. Barker, W. Rayens, Partial least squares for discrimination, *Journal of Chemometrics* 17(3)
727 (2003) 166-173.

728 [66] K.-A. Lê Cao, S. Boitard, P. Besse, Sparse PLS discriminant analysis: biologically relevant feature
729 selection and graphical displays for multiclass problems, *BMC Bioinformatics* 12(1) (2011) 253.

730 [67] K.-A. Lê Cao, P.G.P. Martin, C. Robert-Granié, P. Besse, Sparse canonical methods for biological
731 data integration: application to a cross-platform study, *BMC Bioinformatics* 10(1) (2009) 34.

732 [68] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T.
733 Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction
734 networks, *Genome research* 13(11) (2003) 2498-504.

735 [69] P. Sinitcyn, J.D. Rudolph, J. Cox, Computational methods for understanding mass spectrometry-
736 based shotgun proteomics data, *Annual Review of Biomedical Data Science* 1(1) (2018) 207-234.

737 [70] V. Schwammle, T. Verano-Braga, P. Roepstorff, Computational and statistical methods for high-
738 throughput analysis of post-translational modifications of proteins, *J Proteomics* 129 (2015) 3-15.

739 [71] R.A. Quinn, L.F. Nothias, O. Vining, M. Meehan, E. Esquenazi, P.C. Dorrestein, Molecular
740 networking as a drug discovery, drug metabolism, and precision medicine strategy, *Trends in*
741 *pharmacological sciences* 38(2) (2017) 143-154.

742 [72] H. Mohimani, A. Gurevich, A. Mikheenko, N. Garg, L.F. Nothias, A. Ninomiya, K. Takada, P.C.
743 Dorrestein, P.A. Pevzner, Dereplication of peptidic natural products through database search of mass
744 spectra, *Nature chemical biology* 13(1) (2017) 30-37.

745 [73] N. Bandeira, W.T. Liu, M. Ghassemian, T.L. Simmons, W.H. Gerwick, R. Lington, P.C. Dorrestein,
746 P.A. Pevzner, Dereplication and de novo sequencing of nonribosomal peptides, *Nature methods* 6(8)
747 (2009) 596-9.

748 [74] S. Naimi, S. Zirah, R. Hammami, B. Fernandez, S. Rebuffat, I. Fliss, Fate and Biological Activity of
749 the Antimicrobial Lasso Peptide Microcin J25 Under Gastrointestinal Tract Conditions, *Frontiers in*
750 *microbiology* 9 (2018) 1764.

751 [75] A. Hassanin, F. Delsuc, A. Ropiquet, C. Hammer, B. Jansen van Vuuren, C. Matthee, M. Ruiz-
752 Garcia, F. Catzeflis, V. Areskoug, T.T. Nguyen, A. Couloux, Pattern and timing of diversification of
753 Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of
754 mitochondrial genomes, *Comptes rendus biologiques* 335(1) (2012) 32-50.

755 [76] M. Buckley, C. Lawless, N. Rybczynski, Collagen sequence analysis of fossil camels, *Camelops* and
756 c.f. *Paracamelus*, from the Arctic and sub-Arctic of Plio-Pleistocene North America, *J Proteomics* 194
757 (2019) 218-225.

758 [77] S.D. Webb, Late Cenozoic Mammal Dispersals between the Americas, in: F.G. Stehli, S.D. Webb
759 (Eds.), *The Great American Biotic Interchange*, Springer US, Boston, MA, 1985, pp. 357-386.

760 [78] H.F. Stanley, M. Kadwell, J.C. Wheeler, Molecular evolution of the family Camelidae: a
761 mitochondrial DNA study, *Proceedings. Biological sciences* 256(1345) (1994) 1-6.

762 [79] J. Wheeler, M.W. Bruford, Genetic Analysis of the Origins of Domestic South American Camelids,
763 in: M.A. M.A. Zeder, Daniel G. Bradley, Eve Emshwiller and Bruce G. Smith, Eds (Ed.), *Documenting*
764 *Domestication: New Genetic and Archaeological Paradigms*, University of California Press 2006, pp.
765 329-341.

766 [80] M.C. Penedo, M.E. Fowler, A.T. Bowling, D.L. Anderson, L. Gordon, Genetic variation in the blood
767 of llamas, *Llama glama*, and alpacas, *Llama pacos*, *Animal genetics* 19(3) (1988) 267-76.

768 [81] J. Wheeler, Evolution and present situation of the South American camelidae, *Biological Journal*
769 *of the Linnean Society* 54(3) (1995) 271-295.

770 [82] P. Acebes, J. Wheeler, J. Baldo, P. Tuppia, G. Lichtenstein, D. Hoces, W.L. Franklin, *Vicugna*
771 *vicugna*. The IUCN Red List of Threatened Species 2018: e.T22956A18540534, International Union for
772 Conservation of Nature, 2018.

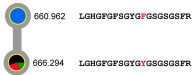
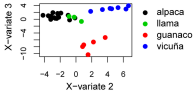
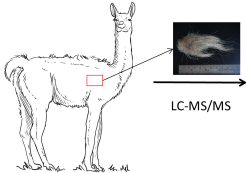
773 [83] E. Dufour, N. Goepfert, Past Andean Pastoralism: A Reconsidered Diversity. Introduction to the
774 Special Issue, *Environmental Archaeology* (2019) 1-5.

775 [84] J.E. Plowman, S. Deb-Choudhury, W.G. Bryson, S. Clerens, J.M. Dyer, Protein expression in
776 orthocortical and paracortical cells of merino wool fibers, *Journal of agricultural and food chemistry*
777 57(6) (2009) 2174-80.

778 [85] C. Solazzo, J.M. Dyer, S. Clerens, J. Plowman, E.E. Peacock, M.J. Collins, Proteomic evaluation of
779 the biodegradation of wool fabrics in experimental burials, *International Biodeterioration and*
780 *Biodegradation* 80 (2013) 48-59.

781

782



South American Camelids