

# Your horse is a donkey! Identifying domesticated equids from Western Iberia using collagen fingerprinting

## Supplementary Information

The raw data has been uploaded to the following locations for open access:

- MS/MS Data:
  - DOI: doi:10.25345/C5T727K8H
  - ProteomeXchange through MassIVE: PXD035509
  - MassIVE Record Number: MSV000089943
- ZooMS Spectra:
  - DOI: 10.5281/zenodo.6878868
  - Zenodo Record Number: 6878868

Also uploaded with the ZooMS Spectra to Zenodo under the same DOI are:

1. Aligned FASTA files for the mature peptides of *col1 $\alpha$ 1* and *col1 $\alpha$ 2* of all sequences used in the manuscript.

## Included in this Supplemental Information PDF

### Supplemental Figures

Figure S1: MS/MS examples for the distinguishing chymotryptic peptide in horse and donkey

### Supplemental Tables

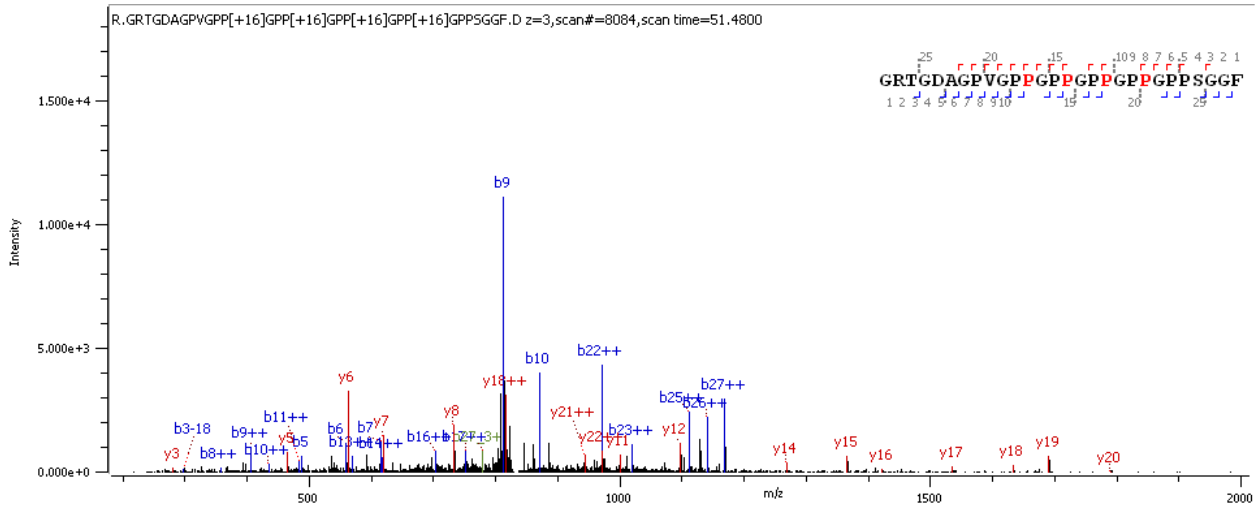
Table S1: List of published collagen markers for species from the Equidae family.

Table S2: Number of proteins in proteome search and coverage of collagen for confirmation search digested with chymotrypsin.

# SI Figure 1: MS/MS Sequence Identification of Biomarkers

The MS/MS spectra of the peptides for the diagnostic marker COL1A1 991-1018 from horse (LARC.265) and donkey (LARC.1498). Both peptides have 4 proline oxidations.

## Horse: m/z 2497



## Donkey m/z 2511

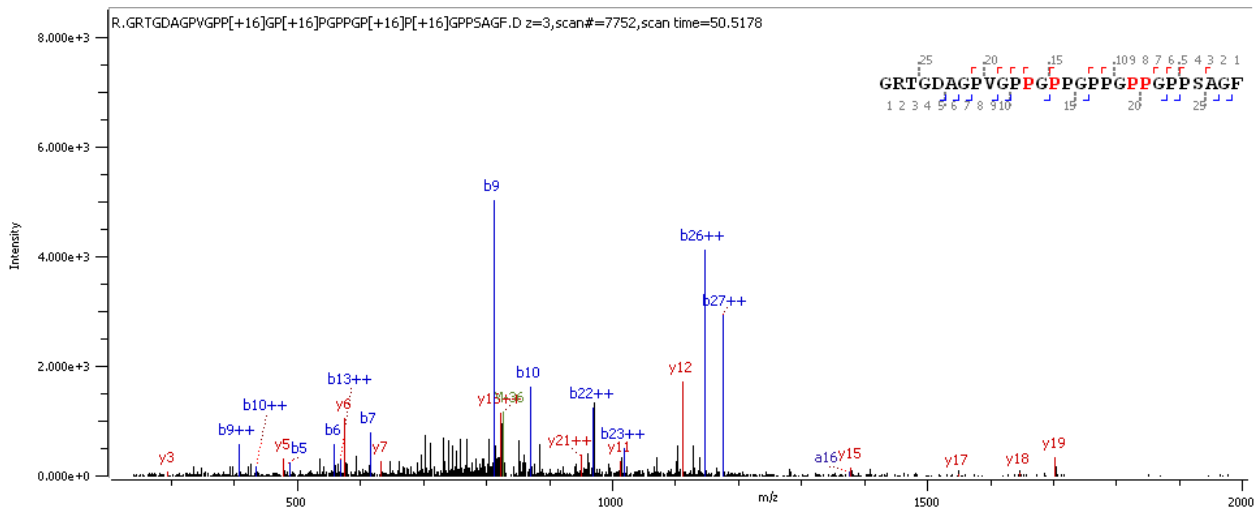


Table 1: List of published ZooMS collagen markers for species from the Equidae family using trypsin.

Scientific Name	COL1A1 508-519	COL1A1 586-618	COL1A1 586-618 (+16)	COL1A2 978-990	COL1A2 978-990 (+16)	COL1A2 484-498	COL1A2 502-519	COL1A2 292-309	COL1A2 793-816	COL1A2 454-483	COL1A2 757-789	COL1A2 10-42	References
<i>Equus grevyi</i>	1105.6	2883.4	2899.4	1182.6	1198.6	1427.7	1550.8	1649.8	2145.1	2820.4	2983.4	2999.4	Welker et al. (2016)
<i>Equus quagga</i>	1105.6	2883.4	2899.4	1182.6	1198.6	1427.7	1550.8	1649.8	2145.1	2820.4	2983.4	2999.4	Welker et al. (2016)
<i>Equus caballus</i>	1105.6	2883.4	2899.4	1182.6	1198.6	1427.7	1550.8	1649.8	2145.1	2820.4	2983.4	2999.4	Welker et al. (2016); Buckley et al. (2009); Buckley and Collins (2011); Kirby et al. (2013); Buckley et al. (2017)
<i>Equus asinus</i>	1105.6	2883.4	2899.4	1182.6	1198.6	1427.7	1550.8	1649.8	2145.1	2820.4	2983.4	2999.4	Welker et al. (2016)
<i>Equus hemionus khur</i>	1105.6	2883.4	2899.4	1182.6	1198.6	1427.7	1550.8	1649.8	2145.1	2820.4	2983.4	2999.4	Welker et al. (2016)
<i>Equus hemionus hydruntinus</i>	1105.6	2883.4	2899.4	1182.6	1198.6	1427.7	1550.8	1649.8	2145.1	2820.4	2983.4	2999.4	Welker et al. (2016)
<i>Equus caballus</i>	1105.6	2883.4	2899.4	1182.6	1198.6	1427.7	1550.8	1649.8	2145.1	2820.4	2983.4	2999.4	Welker et al. (2016)

*Note:*

The nomenclature of the markers follow the scheme recommended in Brown et al. (2021).

Table 2: Number of proteins in proteome search and coverage of collagen for confirmation search digested with chymotrypsin.

Sample	Taxonomic ID	Sample Type	# proteins <sup>a</sup>	COL1A1 best hit <sup>b</sup>	% COV <sup>b</sup>	COL1A2 best hit <sup>b</sup>	% COV <sup>b</sup>
LARC.265	<i>Equus caballus</i>	Modern	7	<i>Equus caballus</i>	93	<i>Equus caballus</i>	95
LARC.1498	<i>Equus asinus</i>	Modern	6	<i>Equus asinus</i>	78	<i>Equus asinus</i>	89

<sup>a</sup> Filters: 2 or more unique pepts, log prob > 3, run against database of SwissProt<sup>TM</sup>, horse proteome, donkey proteome.

<sup>b</sup> from the semi-specific Byonic<sup>TM</sup> runs with the limited database.

Table 3: The scaled percentage change of observation of cut sites to the total percentage of the amino acids in the collagen sequence in comparison to the cut site specificity profile of chymotrypsin. The primary and secondary preferred chymotrypsin cut sites are indicated in bold. Arginine is indicated in red. Amino acids are grouped into samples with increased percent change and decreased percent change.

Amino Acid	Cut sites in LARC.265 <sup>a</sup>	Cut sites in LARC.1498 <sup>a</sup>	Total cut sites <sup>a</sup>	% amino acid in horse sequence <sup>b</sup>	Scaled percent change (%)
<b>F</b>	650	375	1025	1.35	1907.00
<b>M</b>	190	44	234	0.48	1183.00
<b>Y</b>	53	49	102	0.29	832.00
<b>L</b>	469	329	798	2.65	696.00
<b>R</b>	490	280	770	5.21	291.00
<b>H</b>	61	14	75	0.58	243.00
N	56	53	109	1.59	81.00
Q	91	42	133	2.51	40.00
T	51	35	86	1.74	31.00
K	87	26	113	3.23	-0.08
S	20	29	49	3.67	-0.65
A	55	44	99	10.81	-0.76
E	19	11	30	4.49	-0.82
I	5	2	7	1.06	-0.83
D	11	4	15	2.80	-0.86
G	68	51	119	33.30	-0.91
V	3	3	6	2.70	-0.94
P	8	1	9	21.53	-0.99
C	0	0	0	0.00	0.00
W	0	0	0	0.00	0.00

<sup>a</sup> From LC-MS/MS data with Total being the sum of cut sites in both samples.

<sup>b</sup> From the mature COL1A1 and COL1A2 proteins identified from Protein Calculator (Anthis and Clore, 2013).

## References

- Anthis, N.J., Clore, G.M., 2013. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Science* 22, 851–858. <https://doi.org/10.1002/pro.2253>
- Brown, S., Douka, K., Collins, M.J., Richter, K.K., 2021. On the standardization of ZooMS nomenclature. *Journal of Proteomics* 235, 104041. <https://doi.org/10.1016/j.jprot.2020.104041>
- Buckley, M., Collins, M.J., 2011. Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British archaeological and paleontological sites. *Antiqua* 1, e1–e1. <https://doi.org/10.4081/antiqua.2011.e1>
- Buckley, M., Collins, M., Thomas-Oates, J., Wilson, J.C., 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 23, 3843–3854. <https://doi.org/10.1002/rcm.4316>
- Buckley, M., Harvey, V.L., Chamberlain, A.T., 2017. Species identification and decay assessment of Late Pleistocene fragmentary vertebrate remains from Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting. *Boreas* 46, 402–411. <https://doi.org/10.1111/bor.12225>
- Kirby, D., Buckley, M., Promise, E., A. Trauger, S., Rose Holdcraft, T., 2013. Identification of collagen-based materials in cultural heritage. *Analyst* 138, 4849–4858. <https://doi.org/10.1039/C3AN00925D>
- Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B., Stewart, J., Pääbo, S., Collins, M., Hublin, J.-J., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings of the National Academy of Sciences* 113, 11162–11167. <https://doi.org/10.1073/pnas.1605834113>