A novel strategy for the discrimination of gelatinous Chinese medicines based on enzymatic digestion followed by nano-flow liquid chromatography in tandem with orbitrap mass spectrum detection

Huan Yang1,2, Yuping Shen1, Ying Xu1, Aida Serra Maqueda2, Jie Zheng1, Qinan Wu1, James P Tam3

1Department of Chinese Medicine, Nanjing University, Jiangsu, People’s Republic of China. 2Division of Structural Biology and Biochemistry, School of Biological Sciences, College of Science, Nanyang Technological University, Singapore. 3Department of Chinese Medicine, Nanjing, Jiangsu, People’s Republic of China

Correspondence: Qinan Wu, Department of Chinese Medicine, Nanjing University, Jiangsu, 38 Xianlin Avenue, 210023, People’s Republic of China. Tel: +86 25 8581 1507, Fax: +86 25 8581 1153, Email: qnwyjs@163.com

James P Tam, Division of Structural Biology and Biochemistry, School of Biological Sciences, College of Science, Nanyang Technological University, 60 Nanyang Drive, 637551 Singapore. Tel: +65 6316 7042, Fax: +65 6791 3856, Email: jptam@ntu.edu.sg

Abstract: Gelatinous Chinese medicines made from mammalian skin or horn or reptile shell are a very important type of animal-derived Chinese medicine. They have been extensively used either as both hemopoietic and hemostatic agents to treat vertigo, palpitation, hematuria, and insomnia in traditional Chinese medicine clinics; consumed as a popular tonic for weaker persons such as the elderly or women after giving birth; or further manufactured to health supplements for certain populations. However, they cannot be discriminated from each other by only using the routine approach in the Chinese Pharmacopoeia, as it lacks enough specificity and, consequently, the requirements can be met even by adding assayed ingredients. In this study, our efforts to differentiate three gelatinous Chinese medicines, Asini Corii Colla, Cervi Cornus Colla, and Testudinis Carapacis ET Plastri Colla, are presented, and a novel strategy based on enzymatic digestion followed by nano-flow liquid chromatography in tandem with orbitrap mass spectrum detector analysis is proposed herein. Fourteen diagnostic fragments identified from the digests of these medicines were exclusively selected for their discrimination. By taking advantage of the favorable features of this strategy, it is feasible and convenient to identify enzymatic-digested peptides originated from signature proteins in each medicine, which thus could be employed as potential biomarkers for their form of raw medicinal material, and the pulverized and the complex especially, that being the direct basis for authentication purpose.

Keywords: traditional Chinese medicine, animal-derived Chinese medicines, potential biomarkers, authentication, diagnostic fragments

Introduction

Traditional Chinese medicine (TCM), built upon a complete system of theory, is a broad range of medicinal practices generally incorporating common concepts which have been well developed and are based on experiences accumulated from different ethnic groups for 2000 years in the People’s Republic of China, involving various forms of application of natural products and other therapeutic methods. Typical of a number of herbs and mineral medicinals applied in TCM, the harness of animal-derived medicines also has a long history from the use of hirudo, ostreae concha, etc, documented in two ancient Chinese medical books of doctrine: Yellow Emperor’s Inner Canon and Treatise on Cold Damage. In addition, in the prevailing Chinese Pharmacopoeia (ChP), there are 51 raw animal-derived Chinese medicines (ACMs), 61
processed products from 36 of them, as well as 365 Chinese proprietary medicines composed of ACMs, accounting for 22.0% of all 2,165 Chinese medicines, which indicates their irreplaceable importance in the present day.\textsuperscript{4,5}

Gelatinous Chinese medicines (GCMs), made from mammalian skin or horn or reptile shell, are a very important type of ACM. They are extensively used as both hemopoietic and hemostatic agents to treat vertigo, palpitation, hematuria, and insomnia in TCM clinics; consumed as a popular tonic for weaker persons such as the elderly or women after giving birth; or further manufactured to health supplements for certain populations.\textsuperscript{6–13} In the ChP, there are three GCMs under direct regulation of the China Food and Drug Administration (CFDA) by the quality control scheme illustrated in Figure 1; the details of their animal origins, medicinal parts, etc, are summarized in Figure 1 and Table 1.

However, ACMs that are rich in protein and other molecules of biological information differ greatly from the herbs containing abundant secondary metabolites of small molecules or favorable cystine-rich peptides (cyclotides) discovered more and more extensively in recent decades.\textsuperscript{14–17} Which has greatly increased the difficulty of discriminating between them, and therefore the exploration of their discrimination was much fewer than other Chinese medicines so far, especially it is lack of specifically designed strategies and means for this purpose, and some cutting-edge MS technologies such as matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) and orbitrap extensively used for proteomics research have been rarely employed in the study of ACMs until the present. In this paper, our efforts to differentiate three GCMs from each other are presented, and, as shown in Figure 2, a novel strategy based on nano-flow liquid chromatography in tandem with orbitrap mass spectrum detection (NanoLC-orbitrap MSD) is proposed for their authentication, with some evidential differences explored on the GCM samples digested by specific enzyme. NanoLC-orbitrap MSD is distinguished by high resolution; great mass accuracy, specificity, sensitivity for analysis, and rapid conduction of MS profiling for the peak of interest, making this technology one of the most powerful tools in current proteomics research.\textsuperscript{18} By taking advantage of these favorable features, it is feasible and convenient to identify the enzymatic-digested peptides originated from signature proteins present in each GCM, which thus could be employed as the potential biomarkers for ACMs, including their form of raw medicinal material, the pulverized material, and Chinese Proprietary medicines for authentication purposes.

![Figure 1](https://www.dovepress.com/)

**Figure 1** Quality control scheme of GCMs in the ChP.

**Abbreviations:** ACC, Asini Corii Colla; CCC, Cervi Cornus Colla; ChP, Chinese Pharmacopoeia; GCM, gelatinous Chinese medicine; TCPC, Testudinis Carapacis ET Plastri Colla; TLC, thin layer chromatography.
Table 1 Gelatinous Chinese medicines in the Chinese Pharmacopoeia

<table>
<thead>
<tr>
<th>Latin name</th>
<th>Common names</th>
<th>Animal origins</th>
<th>Medicinal parts</th>
<th>Dosage (g/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asini Corii Colla</td>
<td>Donkey-hide glue, E Jiao</td>
<td>Equus asinus L.</td>
<td>Skin</td>
<td>3–9</td>
</tr>
<tr>
<td>Cervi Cornus Colla</td>
<td>Deer-horn glue, Lu Jiao Jiao</td>
<td>Cervus elaphus Linnaeus, Cervus nippon Temminck</td>
<td>Ossified antlers or horn base (naturally detach after cutting pilose non-ossified antlers from deer)</td>
<td>3–6</td>
</tr>
<tr>
<td>Testudinis Carapacis ET Plastri Colla</td>
<td>Tortoise-shell glue, Gui Jia Jiao</td>
<td>Chinemys reevesi (Gray)</td>
<td>Carapace and plastron</td>
<td>3–9</td>
</tr>
</tbody>
</table>

Materials and methods

Materials and chemicals

Asini Corii Colla (ACC) (B/N: 110916, 091204, 100235; Shandong Dong-E E-Jiao Co., LTD. Shandong, People’s Republic of China); Cervi Cornus Colla (CCC) (B/N: 20130922, 20120810, 20121201; Jilin Zhong-Ding Pharmaceuticals Co. Jilin, People’s Republic of China); and Testudinis Carapacis ET Plastri Colla (TCPC) (B/N: 20131018, 20100529, 20120721; Shandong Dong-E E-Jiao Co., LTD). The voucher specimens were deposited and authenticated by Associate Professor Hongxia Chen, Pharmacognosy Research Center, Jiangsu University (Zhenjiang, People’s Republic of China). These raw materials were individually smashed into powder form by mortar in liquid nitrogen and passed through a 40-mesh sieve. The fine powder of each sample was then collected and stored in a desiccator at room temperature after lyophilization for 24 hours to remove the remaining moisture.

Tris base of molecular biology grade was obtained from Promega Corporation (Fitchburg, WI, USA). HCl solution (37%), acetonitrile, formic acid (FA), glacial acetic acid, NaCl, CaCl₂, anhydrous Na₂CO₃, NH₄HCO₃, NaHCO₃, and NaOH were purchased from EMD Millipore (Billerica, MA, USA). Ethylene glycol was provided by Alfa Aesar (Ward Hill, MA, USA). Ten percentage sodium dodecyl sulfate (SDS) solution, ammonium persulfate, tricine, 2× Laemmli sample buffer, glycine, 40% acrylamide/bis (19:1), 40% acrylamide/bis (29:1), 30% acrylamide/bis (29:1), and N,N,N′,N′-tetramethylethylenediamine TEMED are electrophoresis purity reagents purchased from Bio-Rad Laboratories Inc. (Hercules, CA, USA). Bovine serum albumin (BSA) (fraction V) was a product of GE Healthcare Bio-Sciences Corp. (Piscataway, NJ, USA). Exact-Pro broad range (10–245 kDa) precasted protein ladder (Bio-5150) was from First BASE Laboratories Sdn Bhd (Seri Kembangan, Selangor, Malaysia). AgNO₃, Na₂S₂O₃, N-[Tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid, and DTPA were supplied by Sigma-Aldrich (St. Louis, MO, USA). N,N,N′,N′-tetramethylethylenediamine (TEMED) was purchased from Life Technologies (Grand Island, NY, USA). 4-Chloro-1-naphthol (4-CN) was from Fisher Scientific (Pittsburgh, PA, USA).

Figure 2 Novel strategy proposed for the discrimination of GCMs.

Abbreviations: GCM, gelatinous Chinese medicine; MS⁺, multi-stage mass spectrometry; NanoLC-orbitrap MSD, nano-flow liquid chromatography in tandem with orbitrap mass spectrum detection.
acid (TES), CuSO$_4$·5H$_2$O, Coomassie brilliant blue G-250 and R-250, H$_3$PO$_4$ solution (49%–51%), bicinechonic acid (BCA) solution, formaldehyde solution (37%, in 10%–15% MeOH), 2-mercaptoethanol, gelatin from cold water fish skin, and collagenase from Clostridium histolyticum (type I) were all purchased from Sigma-Aldrich Co. (St Louis, MO, USA).

**Total protein determination**

**Preparation of sample solutions**

For the fine powder of each raw material, 0.100 g was precisely weighed into a beaker and 5.00 mL of distilled H$_2$O was added to dissolve the sample in an 85°C water bath for 20 minutes, and the solution was completely transferred into a 10 mL volumetric flask. The solution was then centrifuged at 4,000 rpm for 30 minutes and the supernatant was collected, 1.00 mL of which was diluted to 10.0 mL with H$_2$O for subsequent analysis by two conventional methods including both Bradford assay and BCA assay for comparison purposes.

**BCA assay and Bradford assay**

BCA assay was conducted on a UV transparent 96-well microplate and the absorbency was scanned at 560 nm using a microplate reader. Bradford assay was carried out in test tubes and the absorbency measurement was taken at 595 nm.\(^{19}\)

**Calculation of total protein content**

The percentage of total protein content was calculated according to the following equation:

\[
\text{% total protein content} = \frac{\text{The amount of protein}}{\text{The weight of sample}} \times 100\% \quad (1)
\]

**SDS polyacrylamide gel electrophoresis analysis**

A Bio-Rad electrophoresis system was equipped with a PowerPac universal power supply and a Mini-PROTEAN Tetra cell for SDS polyacrylamide gel electrophoresis (SDS–PAGE) analysis.\(^{20}\) Glycine SDS–PAGE (5% stacking gel, 10% resolving gel, 1.0 mm) was performed at 80 V for 15 minutes followed by 110 V for another 75 minutes, and tricine SDS–PAGE (5% stacking gel, 18% resolving gel, 0.75 mm) was performed at 80 V for 10 minutes followed by 200 V until the loading dye front reached the bottom of the resolving gel. Then, the gel sandwich was disassembled and subject to either Coomassie brilliant blue staining or silver staining.

**Preparation of sample solutions**

**Pretreatment of samples**

One hundred milligrams of each GCM sample was dissolved in 10 mL Milli-Q water by vortex and sonication and then centrifuged at 12,000 rpm at 4°C for 30 minutes. The supernatant was collected for subsequent membrane ultrafiltration by centrifugal filter unit (5 kDa molecular weight cut-off; EMD Millipore) at 3,000 rpm at 4°C, according to the manufacturer’s instructions, to remove most of the small molecules existing in the solution. The remaining concentrated brown sample solution in the upper vessel of the unit was completely transferred to a glass container and subject to further lyophilization, resulting in a pale and light block mass which was then stored in a desiccator under room temperature for the following experiments.

**Digestion by enzyme**

Ten milligrams of individual pretreated GCM and gelatin were dissolved in 2.0 mL of 50 mM TES buffer (0.36 mM CaCl$_2$, pH 7.4) for collagenase digestion to give the final concentration at 5.0 mg/mL. After sonication and vortexing to promote the dissolution, the samples were visually inspected for the presence of residual solid material and subject to centrifugation at 12,000 rpm for 30 minutes. Then, 100 μL of supernatant was denatured at 95°C for 10 minutes and cooled down to room temperature, and 2.0 μL of enzyme solution (0.10 mg/mL collagenase in TES buffer) was then added before the full digestion performed at 37°C over 24 hours.\(^{21,22}\) Few efforts were made to carefully enhance the activity of the enzyme, since the only purpose of this step was to digest the present proteins thoroughly, which could be simply achieved by the addition of excess enzymes, and then the completeness of the above digestions was ensured by means of SDS–PAGE with silver staining.

**NanoLC-orbitrap MSD analysis conditions**

Enzymatic-digested peptides were analyzed by NanoLC-MS/MS on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) coupled with a Dionex UltiMate 3000 UHPLC system from Thermo Fisher Scientific. Spraying was done using a Michrom’s Thermo Captive Spray nanoelectrospray ion source (Bruker-Michrom Inc., Auburn, CA, USA) at 250°C capillary temperature and 1.5 kV source voltage. Separation was performed in a reverse-phase Acclaim PepMap RSL column (75 μm internal diameter×15 cm, 2 μm particles), also from Thermo Fisher Scientific, which was maintained at 35°C throughout the analysis. Separation was done with a flow rate of 300 nL/min using 0.1% FA aqueous solution as eluent.
A and 0.1% FA in 90% acetonitrile aqueous solution as eluent B. A 60-minute gradient elution was used starting at 3% of eluent B for 1 minute, followed by 3%-35% of eluent B within 47 minutes, increased to 50% of eluent B within 4 minutes, again increased to 80% within 0.1 min, and kept isocratic for 1.3 minutes. Then, the mobile phases were returned to 3% of eluent B within 0.1 minute and maintained for 6.5 minutes for equilibrium.

Data acquisition was done in positive ion mode using the LTQ Tune Plus software (Thermo Fisher Scientific) alternating between full MS (350–1,600 m/z, resolution 60,000, with 1 μscan per spectrum) and FT-MS/MS (150–2,000 m/z, resolution 15,000, with 1 μscan per spectrum) for the ten most intense ions (with a 500-count threshold). Fragmentation was performed in high-energy collisional dissociation mode at 32% normalized collision energy. Automatic gain control target for both full MS and FT-MS/MS was set to 1e+06, and precursor ion charge state screening was activated. Dynamic exclusion list was enabled with an exclusion duration of 30 seconds and an exclusion list size of 500.

De novo sequencing of each peptide was conducted by PEAKS software automatically. Local confidence is the confidence (presented as a percentage) that a particular amino acid is present in the de novo peptide at a particular position. And the sum of the local confidence scores (0 to 1) from each amino acid in the peptide sequence was divided by the number of amino acids, which was presented to be average local confidence (ALC) used to assess the accuracy of the interpretation.

Results and discussion
Protein content
As shown in Table 2, high correlation coefficients ($r^2$) of calibration curve were achieved in both assays, and the protein content of different GCMs was increased in the following order: TCPC < CCC < ACC. Meanwhile, the contents determined by BCA assay were much higher than in the Bradford assay, suggesting that the proteins in GCMs do not contain a large number of arginine and/or aromatic residues in their chemical structure.

Table 2 Calibration curve and protein content

<table>
<thead>
<tr>
<th>Assays</th>
<th>Regression equations</th>
<th>$r^2$</th>
<th>Linearity ranges</th>
<th>Protein content (w/w)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACC</td>
</tr>
<tr>
<td>BCA</td>
<td>Y = 0.0010X + 0.0651</td>
<td>0.9993</td>
<td>62.50–2,000 μg/mL</td>
<td>44.2%</td>
</tr>
<tr>
<td>Bradford</td>
<td>Y = 0.0056X + 0.729</td>
<td>0.9995</td>
<td>20.00–100.0 μg/mL</td>
<td>19.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCA</td>
<td>Y = 0.0010X + 0.0651</td>
<td>0.9993</td>
<td>62.50–2,000 μg/mL</td>
<td>40.7%</td>
</tr>
<tr>
<td>Bradford</td>
<td>Y = 0.0056X + 0.729</td>
<td>0.9995</td>
<td>20.00–100.0 μg/mL</td>
<td>15.9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th>TCPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCA</td>
<td>Y = 0.0010X + 0.0651</td>
<td>0.9993</td>
<td>62.50–2,000 μg/mL</td>
<td>36.4%</td>
</tr>
<tr>
<td>Bradford</td>
<td>Y = 0.0056X + 0.729</td>
<td>0.9995</td>
<td>20.00–100.0 μg/mL</td>
<td>14.4%</td>
</tr>
</tbody>
</table>

Abbreviations: ACC, Asini Corii Colla; BCA, bicinchoninic acid; CCC, Cervi Cornus Colla; TCPC, Testudinis Carapae ET Plastri Colla; $r^2$, correlation coefficient; w/w, weight/weight.
fragments obtained were all subject to de novo sequencing for further analysis by PEAKS Studio 7 (Bioinformatics Solutions Inc., Waterloo, ON, Canada). The cut-off ALC was set to be 90% for good peptide matching – much higher than the 55% ALC commonly recognized in usual practice, which must not ensure that the whole sequence is completely correct. In this way, 14 diagnostic fragments identified from the digests of these medicines were exclusively selected for their discrimination and are summarized in Table 3. All the fragments selected were within the range of 500–800 m/z of multiple charges at $2^+ \text{ or } 3^+$, which were composed of 10–19 amino acid residues. Among these, glycine contributed to more than one-third of the sequence in almost all the fragments, and proline was always at the C-terminal site of the sequence. In addition, some of the amino acids abundant in collagen was also incorporated in the sequence of the fragments, eg, glutamic acid, arginine, alanine, and threonine.

In the present study, the term “potential biomarkers” was tentatively employed to acknowledge their potential value and at the same time to indicate their uncertainty, and more GCMs including the products made from other animal sources are yet to be analyzed furthermore.

Potential biomarkers’ sequence alignment and analysis
All of the sequences of the potential biomarkers were compared with those of other known proteins based on an online Basic Local Alignment Search Tool (BLAST) analysis, available on the National Center for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi), and were used as an input sequence for a Position-Specific Iterated BLAST (PSI-BLAST) search program against the non-redundant protein sequences databases (nrdb). As a consequence, a high degree of similarity was found to the sequences of predicted or previously published proteins, and most of them were collagens from mammals.

As summarized in Table 4, the highest sequence identity (82%) was found in the partial sequences of collagens from animals of the Equus genus, including Equus asinus L., which is the only medicinal animal origin for ACC, and collagen alpha-1 type I chain aligned with the sequence of the potential biomarker was one of 599 proteins discovered in this animal. The result accordingly suggested the certain reliability of the novel approach newly proposed in the current study. Although other potential biomarkers have been well aligned with the collagens from various animals, they could not be matched with the sequences of the proteins from

NanoLC-orbitrap MSD analysis
It was observed from the NanoLC-orbitrap MSD total ion chromatogram of the three GCMs’ digests, as shown in Figure 4, that most of the fragments generated were intensively eluted out of the capillary column from approximately 8 to 20 minutes. Also, the profiles of the peak clusters differed from each other slightly, although not much effort was made to optimize the liquid chromatography conditions for a better performance, as our major aim of these analyses was to identify the diagnostic biomarkers from the peptides for the differentiation of individual GCMs, which does not necessarily require a fine separation of the sample.

The exact mass of the peptide ions was calculated according to the precursor mass and isotope pattern displayed in the MS spectrum at each time point, and the MS and MS/MS

---

**Notes:** Lane 1: ladder. Lane 2: collagenase after incubation. Lane 3: ACC. Lane 4: ACC after in-solution digestion by collagenase. Lane 5: CCC. Lane 6: CCC after in solution digestion by collagenase. Lane 7: TcPC. Lane 8: TcPC after in solution digestion by collagenase. Lane 9: gelatin after in solution digestion by collagenase. Lane 10: gelatin.

**Abbreviations:** ACC, Asini Corii Colla; CCC, Cervi Cornus Colla; MW, molecular weight; SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TcPC, Testudinis Carapacis ET Plastri Colla.

---

**Figure 3** SDS–PAGE analysis of gelatinous Chinese medicines and their digests (1.0 mm, 5% stacking gel, 10% resolving gel, silver staining).

**Figure 4** NanoLC-orbitrap total ion chromatogram of the three GCMs’ digests, as shown in Figure 4, that most of the fragments generated were intensively eluted out of the capillary column from approximately 8 to 20 minutes. Also, the profiles of the peak clusters differed from each other slightly, although not much effort was made to optimize the liquid chromatography conditions for a better performance, as our major aim of these analyses was to identify the diagnostic biomarkers from the peptides for the differentiation of individual GCMs, which does not necessarily require a fine separation of the sample.

The exact mass of the peptide ions was calculated according to the precursor mass and isotope pattern displayed in the MS spectrum at each time point, and the MS and MS/MS
Figure 4 NanoLC-orbitrap MSD TIC of gelatinous Chinese medicines' digests.

Notes: NanoLC-orbitrap MSD TIC of (A) ACC (B/N: 110916), (B) CCC (B/N: 20130922), and (C) TCPC (B/N: 20131018).

Abbreviations: ACC, Asini Corii Colla; CCC, Cervi Cornus Colla; NanoLC-orbitrap MSD, nano-flow liquid chromatography in tandem with orbitrap mass spectrometer detection; TCPC, Testudinis Carapacias ET Plastri Colla; TIC, total ion chromatogram; B/N, batch number; NL, normalized level; RT, retention time.
their animal origins, namely *Cervus elaphus*, *Cervus nippon*, and *Chinemys reevesii*, which have only 914 proteins, 866 proteins, and 305 proteins incorporated under each entry in The National Centre for Biotechnology Information (NCBI) database, respectively. This could be largely caused by the very limited proteins discovered for each of the animal origins used for the production of the GCMs, which has greatly increased the difficulty level for further study.

**Conclusion**

It can be concluded that these three GCMs could not be well discriminated by using either the approach enforced in ChP for the quality control of those GCMs required by CFDA, as the methods lack sufficient specificity, or by using protein content determination and SDS–PAGE analysis as carried out as the first attempts in this study, since no crucial visual differences were observed.

The novel strategy proposed and developed in this study represents a simple and convenient method for the purpose of the discrimination of some GCMs by using a powerful proteomics tool, NanoLC-orbitrap MSD, after the enzymatic digestion by bacterial-sourced collagenases. This method retains the advantages of the highly specific method of qualitative analysis while avoiding the associated problems.

**Acknowledgments**

The research was supported by the Singapore National Research Foundation grant NRF-CRP8-2011-05; the Jiangsu Nature Science Foundation (BK2012290 and SBK2015042622); the Program for Graduate’s Innovative Research of Jiangsu (CXLX13_691); the Student Research Program of Jiangsu University (13A178); and the National Natural Science Foundation of China (81303174).

### Table 3 List of potential biomarkers for individual GCMs

<table>
<thead>
<tr>
<th>Codes</th>
<th>Potential biomarkers</th>
<th>GCMs</th>
<th>NanoLC</th>
<th>Orbitrap MSD</th>
<th>PEAKS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t&lt;sub&gt;s&lt;/sub&gt; (min)</td>
<td>Charges</td>
</tr>
<tr>
<td>A1</td>
<td>GFGAVGAKGEGGGAP</td>
<td>ACC</td>
<td>18.07</td>
<td>2+</td>
<td>664.3849</td>
</tr>
<tr>
<td>A2</td>
<td>FGSDLKADGGAGAP</td>
<td>ACC</td>
<td>26.13</td>
<td>2+</td>
<td>660.3151</td>
</tr>
<tr>
<td>A3</td>
<td>LAGPGFRAGP</td>
<td>ACC</td>
<td>16.87</td>
<td>2+</td>
<td>523.2746</td>
</tr>
<tr>
<td>A4</td>
<td>FGSGLDGSGAGAGAP</td>
<td>ACC</td>
<td>19.72</td>
<td>2+</td>
<td>631.8045</td>
</tr>
<tr>
<td>A5</td>
<td>QTGPGAGGAGVGPP</td>
<td>ACC</td>
<td>17.53</td>
<td>2+</td>
<td>680.3351</td>
</tr>
<tr>
<td>C1</td>
<td>GPPGSGVFPGAGVGAGKP</td>
<td>CCC</td>
<td>14.26</td>
<td>3+</td>
<td>539.2655</td>
</tr>
<tr>
<td>C2</td>
<td>QPGTTPPPGLTQPQP</td>
<td>CCC</td>
<td>17.17</td>
<td>2+</td>
<td>719.8500</td>
</tr>
<tr>
<td>C3</td>
<td>VAGPAAADGLPGGAG</td>
<td>CCC</td>
<td>15.41</td>
<td>2+</td>
<td>684.3293</td>
</tr>
<tr>
<td>C4</td>
<td>FSGGLDGSSAGGAGAP</td>
<td>CCC</td>
<td>22.90</td>
<td>2+</td>
<td>631.8044</td>
</tr>
<tr>
<td>T1</td>
<td>QTGPGAAAGRVPP</td>
<td>TCPC</td>
<td>17.93</td>
<td>2+</td>
<td>680.3376</td>
</tr>
<tr>
<td>T2</td>
<td>QPGGPPGAGPAGGGAP</td>
<td>TCPC</td>
<td>15.99</td>
<td>2+</td>
<td>692.8358</td>
</tr>
<tr>
<td>T3</td>
<td>QVGPAQNAGSAGGAGAP</td>
<td>TCPC</td>
<td>16.00</td>
<td>2+</td>
<td>783.9045</td>
</tr>
<tr>
<td>T4</td>
<td>LGAPGFRQP</td>
<td>TCPC</td>
<td>17.28</td>
<td>2+</td>
<td>523.2755</td>
</tr>
<tr>
<td>T5</td>
<td>FSGGLDGSDAGGAGAP</td>
<td>TCPC</td>
<td>24.64</td>
<td>2+</td>
<td>660.3176</td>
</tr>
</tbody>
</table>

**Abbreviations:** ACC, Asini Corii Colla; ALC, average local confidence; CCC, Cervi Cornus Colla; GCM, Gelatinous Chinese medicine; MSD, mass spectrum detection; NanoLC, nano-flow liquid chromatography; TCPC, Testudinis Carapacis ET Plastri Colla; t<sub>s</sub>, retention time.

### Table 4 List of sequences producing significant alignments by PSI-BLAST against nrdb

<table>
<thead>
<tr>
<th>GCM</th>
<th>Potential biomarker</th>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>Sequence identity</th>
<th>Accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>GFGAVGAKGEGGGAP</td>
<td>Predicted: collagen alpha-1(I) chain (Equus caballus)</td>
<td>37.1</td>
<td>79.5</td>
<td>100%</td>
<td>82%</td>
<td>XP_005597538.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Predicted: collagen alpha-1(I) chain (Equus przewalski)</td>
<td>37.1</td>
<td>97.6</td>
<td>100%</td>
<td>82%</td>
<td>XP_008516208.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collagen alpha-1 type I chain (Equus asinus L)</td>
<td>37.1</td>
<td>120</td>
<td>100%</td>
<td>82%</td>
<td>ACM24774.1</td>
</tr>
</tbody>
</table>
Disclosure
The authors report no conflicts of interest in this work.

References