



Published in final edited form as:

J Proteome Res. 2017 February 03; 16(2): 920–932. doi:10.1021/acs.jproteome.6b00873.

Expansion for the *Brachylophosaurus canadensis* Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein

Elena R. Schroeter^{†,*}, Caroline J. DeHart[‡], Timothy P. Cleland[§], Wenxia Zheng[†], Paul M. Thomas[‡], Neil L. Kelleher[‡], Marshall Bern^{||}, and Mary H. Schweitzer^{†,⊥}

[†]Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina 27695, United States

[‡]National Resource for Translational and Developmental Proteomics, Northwestern University, Evanston, Illinois 60208, United States

[§]Department of Chemistry, University of Texas-Austin, Austin, Texas 78712, United States

^{||}Protein Metrics, San Carlos, California 94070, United States

[⊥]North Carolina Museum of Natural Sciences, Raleigh, North Carolina 27601, United States

Abstract

Sequence data from biomolecules such as DNA and proteins, which provide critical information for evolutionary studies, have been assumed to be forever outside the reach of dinosaur paleontology. Proteins, which are predicted to have greater longevity than DNA, have been recovered from two nonavian dinosaurs, but these results remain controversial. For proteomic data derived from extinct Mesozoic organisms to reach their greatest potential for investigating questions of phylogeny and paleobiology, it must be shown that peptide sequences can be reliably and reproducibly obtained from fossils and that fragmentary sequences for ancient proteins can be increasingly expanded. To test the hypothesis that peptides can be repeatedly detected and validated from fossil tissues many millions of years old, we applied updated extraction

*Corresponding Author: easchroe@ncsu.edu.

ORCID

Elena R. Schroeter: 0000-0003-4314-2976

Paul M. Thomas: 0000-0003-2887-4765

Neil L. Kelleher: 0000-0002-8815-3372

The authors declare no competing financial interest.

Raw data, including MS acquisition files (.RAW files), PEAKS, and Byonic raw export files (.MZID files), have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD005087.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00873.

Peptide alignments for *B. canadensis* (Figure S1), annotated spectra for all peptides reported in this study (Figures S2–S16), and a phylogenetic topology of the extant species used for phylogenetic analysis (Figure S17). (PDF)

Table S1. Accession numbers for collagen I alpha 1 and 2 of extant species used in phylogenetic analysis. (XLSX)

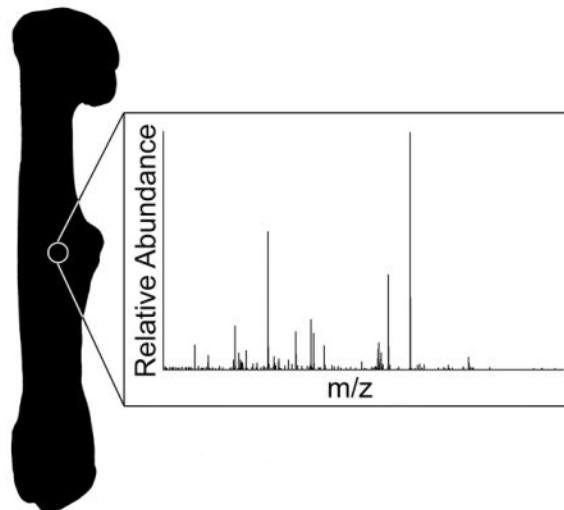
Table S2. PEAKS peptide export lists. (XLSX)

Table S3. Metrics for MS scans, MS/MS scans, and PSM identifications from all searches. (XLSX)

Table S4. List of the detected precursor error, largest fragment error, and average fragment error for each reported spectra in both PEAKS and Byonic. (XLSX)

methodology, high-resolution mass spectrometry, and bioinformatics analyses on a *Brachylophosaurus canadensis* specimen (MOR 2598) from which collagen I peptides were recovered in 2009. We recovered eight peptide sequences of collagen I: two identical to peptides recovered in 2009 and six new peptides. Phylogenetic analyses place the recovered sequences within basal archosauria. When only the new sequences are considered, *B. canadensis* is grouped more closely to crocodylians, but when all sequences (current and those reported in 2009) are analyzed, *B. canadensis* is placed more closely to basal birds. The data robustly support the hypothesis of an endogenous origin for these peptides, confirm the idea that peptides can survive in specimens tens of millions of years old, and bolster the validity of the 2009 study. Furthermore, the new data expand the coverage of *B. canadensis* collagen I (a 33.6% increase in collagen I alpha 1 and 116.7% in alpha 2). Finally, this study demonstrates the importance of reexamining previously studied specimens with updated methods and instrumentation, as we obtained roughly the same amount of sequence data as the previous study with substantially less sample material. Data are available via ProteomeXchange with identifier PXD005087.

Graphical Abstract



Keywords

paleoproteomics; collagen I; bone; *Brachylophosaurus canadensis*; phylogenetics; Archosauria

INTRODUCTION

Extrapolation from kinetic experiments (e.g., Collins et al.¹) predicts a half life for endogenous biomolecules of less than a million years. As a result, paleoproteomic investigations seeking to identify and characterize protein sequences from fossil material have predominantly focused on fossils from the mid-Pleistocene (~1.5 Ma) and younger.^{2–22} However, preservation of significantly older proteins has been demonstrated (e.g., egg shell proteins ~3.8 Ma²³), and multiple lines of evidence suggest that soft tissues and the proteins comprising them can persist in fossil specimens as old as the Cretaceous (>66 Ma).^{24–31} To

date, peptide sequences of collagen I, the most abundant protein in bone,^{32,33} have been detected from the bone matrix of *Tyrannosaurus rex* (MOR 1125, 66 Ma)^{27,34} and *Brachylophosaurus canadensis* (MOR 2598, 80 Ma)²⁸ using mass spectrometry (MS). These collagen I peptides have been consistently identified by MS in bone extracts, but not laboratory reagents or entombing sediments subjected to identical MS parameters, and have been supported through demonstration of immunological reactivity both in situ and in extracts of the fossil tissue.²⁸ Despite the consistency of these data, the endogeneity of these sequences has been questioned^{35,36} and remains controversial. This is in part because of the ages of the specimens, which are well beyond the proposed theoretical limits for protein survival in bone that have been extrapolated from chemical models.^{35,37}

Recently, the persistence of vertebrate peptides in these dinosaur skeletal elements has been further supported by additional immunological and MS testing on osteocytes³⁰ and blood vessels³¹ isolated from *B. canadensis* (MOR 2598), which revealed peptide sequences for various vertebrate proteins that are also present in modern vessels and which cannot be produced by bacteria (e.g., actin, tubulin, myosin, histones). Multiple mechanisms have been proposed or experimentally demonstrated that may result in early stabilization of protein molecules, potentially resulting in preservation over geological time. These have included: (1) a role for iron/oxygen chemistry;³⁸ (2) association with bone mineral that may either provide protection from enzymatic activity^{39–42} or prevent molecular swelling that exposes reactive sites;^{43,44} or (3) preferential preservation of specific collagen I peptides identified in *T. rex* and *B. canadensis* because of their physically shielded location within the collagen fibril.⁴⁵

For proteomic sequences derived from extinct Mesozoic organisms to reach their greatest potential for investigating questions of phylogeny, paleobiology, evolutionary relationships, or the acquisition of evolutionary novelties (e.g., feathers), it must be shown that peptides can be reliably and reproducibly obtained. Furthermore, it is vital to maximize sequence coverage for ancient proteins, as small fragmented sequences (predicted for ancient remains) may be too conserved between related organisms to provide sufficient resolution for phylogenetic analyses. Thus it is crucial that fossil samples be periodically re-evaluated in light of technological advancements, expanding molecular databases of both extant and extinct organisms, and extraction method innovations to increase the sequence coverage and number of proteins recovered from ancient proteomes.⁴⁶

The original MS analysis of bone matrix from the hadrosaur *B. canadensis* (MOR 2598)²⁸ employed multiple bulk chemical extractions of fossil bone and a combination of ion trap and Orbitrap-ion trap analyses over the course of 1 year. Here we present data derived from additional MS analyses on this same dinosaur using new methods of extraction coupled to high-resolution MS techniques. These analyses differed from the original 2009 report (Schweitzer et al.²⁸) in that we employed: (1) a different chemical extraction technique; (2) a different MS sample preparation technique (i.e., in-gel digestion); (3) a different mass spectrometer (12T Fourier Transform-Ion Cyclotron Resonance mass spectrometer [FT-ICR]); (4) high-resolution and high mass accuracy detection for both precursor and fragment ions (FT-FT); (5) a different proteomics laboratory (National Resource for Translational and Developmental Proteomics at Northwestern University); (6)

different bioinformatics software (e.g., PEAKS and Byonic); and (7) more stringent false discovery rate (FDR) settings. Virtually the only component in common with Schweitzer et al.²⁸ was the dinosaur specimen itself. Thus data reported herein independently test the hypothesis that peptides can be repeatedly detected and validated from fossil tissues many millions of years old. Additionally, we show that with advances in proteomics technology and an increasing library of extant genomes, known protein sequences of extinct organisms may be expanded, potentially increasing their value for phylogenetic hypothesis testing.

MATERIALS AND METHODS

Specimen

MOR 2598 (~80 Ma) was excavated from sandstone of the Judith River Formation, Eastern Montana, USA.²⁸ Detailed information on the excavation and handling in the field is available in Schweitzer et al.²⁸

Sample Handling and Anticontamination Procedures

Protein extractions were conducted at North Carolina State University, in a laboratory dedicated to the analyses of fossilized tissues in which tissues of extant organisms are not permitted. All instruments, solutions, and reagents were isolated, and all workers wore personal protective equipment to prevent contaminant transfer. Silver staining, in-gel digestion (IGD), and mass spectrometry were conducted at the National Resource for Translational and Developmental Proteomics at Northwestern University. At this location, all sample handling and molecular techniques (e.g., electrophoresis, gel-band excision, destaining, tryptic digestion) were conducted in a laminar flow hood in which no extant bone tissue had previously been analyzed, with reagents, buffers, pipette tips, and centrifuge tubes used solely for these fossil specimens and isolated from other lab supplies. Prior to every use of the hood, all surfaces and tools therein (e.g., pipettes) were triple cleaned with 100% methanol. Prior to analysis of the isolated samples, the ion optics of the 12T FT-ICR mass spectrometer were removed, dismantled, and cleaned by sonication in 100% Optima-grade methanol for 10 min. The ion transfer capillary leading to the ion optic flight path was subjected to additional cleaning by sonication in 20% (v/v) nitric acid for 15 min prior to washing in Optima-grade water and sonication in 100% Optima-grade methanol for 10 min. Samples were introduced to the mass spectrometer via ultrahigh performance liquid chromatography (UHPLC) using a new, self-packed trap, analytical column, and spray tip, which were conditioned with bovine serum albumin (BSA) peptides and had never been used for any previous analysis.

Protein Extraction

The experimental design is graphically represented in Figure 1. Cortical fragments of the *B. canadensis* femur were ground to the consistency of coarse sand in a mortar and pestle (previously sterilized by soaking in nitric acid for 2 days, then autoclaving). Sediment samples from directly adjacent to the bone were ground separately with a mortar and pestle identically sterilized. Dinosaur bone or sediment powders (1 g) were incubated in 40 mL of 0.5 M EDTA for 4 days at 4 °C, with agitation. An additional tube with no bone or sediment also received 40 mL of 0.5 M EDTA and was treated in tandem with bone and sediment

samples to serve as a control for exogenous contamination from within the laboratory environment. Samples were then centrifuged at RT for 10 min at 10 000 rpm, and supernatants were collected and stored at 4 °C until dialysis. Pellets were resuspended in 40 volumes (40 mL) of 0.05 M ammonium bicarbonate (ABC) and incubated overnight at 65 °C, with agitation. Samples were again centrifuged for 10 min at 10 000 rpm, and supernatants were collected and stored at 4 °C. Pellets were resuspended a final time in 40 volumes (40 mL) of 4 M guanidine hydrochloride (GuHCl) in 0.05 M Tris (pH 7.4) and incubated overnight with agitation at 65 °C. Samples were centrifuged, the final supernatants were collected, and pellets were discarded. ABC and GuHCl supernatants were centrifuged for 10 min at 10 000 rpm to pellet any debris, then transferred to 3500 MWCO SnakeSkin dialysis tubing (ThermoScientific) and dialyzed against 18.2 MΩ water (4L) for 4 days at 4 °C, exchanging dialysis water twice daily. Dialyzed supernatants were lyophilized (FreeZone Freeze-Dry System; Labconco) to completion (2 to 3 days). Each lyophilized sample was solubilized in 5 mL of 0.05 M ABC, then divided into 1 mL aliquots in 1.5 mL tubes (Eppendorf) and frozen at -80 °C.

Silver Stain

ABC and GuHCl extraction aliquots (1 mL each, or the extraction product from 200 mg bone or sediment) of *B. canadensis*, sediment, and buffer blank were thawed and concentrated to near dryness by speed vacuum, then resolubilized in 37.5 μL of PBS and combined with 12.5 μL of 4× Laemmli buffer (Bio-Rad) with 9% β-mercaptoethanol (BME). Samples were heated for 5 min at 95 °C, then loaded onto a Mini-Protean TGX 4–20% gel (Bio-Rad) along with 5 μL of Precision Plus Protein Dual Color Standard (Bio-Rad). Proteins were electrophoresed for ~1 h at a constant 30 mA. The gel was transferred to a new sterile Petri dish, rinsed twice in 18.2 MΩ water, fixed with 50% methanol/5% acetic acid in 18.2 MΩ water for ~2 h with rocking, and then stained using a Pierce Silver Stain Kit for Mass Spectrometry (Thermo Scientific) following manufacturer's protocols (all buffers made in 18.2 MΩ water and all steps at RT). In brief, the gel was washed (10% ethanol, twice for 5 min each), sensitized (50 μL sensitizer +25 mL water for 1 min), rinsed (18.2 MΩ water, twice for 1 min each), stained (250 μL enhancer +25 mL silver solution, 5 min with rocking), rinsed (18.2 MΩ water, twice for 20 s each), and developed (250 μL enhancer + 25 mL developer, for ~3 min). Development was terminated with 5% acetic acid, and the gel was stored in fresh 5% acetic acid overnight at 4 °C. An image of the stained gel is provided in Figure 2A.

Gel Band Excision

Band excision was performed in a laminar flow hood while wearing a head covering, a face mask, and extended cuff gloves over the sleeves of a full-length lab coat. To avoid cross-contamination during band excision, lanes from dinosaur, sediment, and blank GuHCl extracts were excised separately, transferred to new sterile Petri dishes, and cut into 10 gel sections, as shown in (Figure 2B). Each section was diced into ~1 mm³ cubes using a new, fresh razor rinsed in Optima-grade methanol (Fisher Scientific), followed by rinsing with Optima-grade water (Fisher Scientific) for *each* individual gel section. Each diced gel section was then transferred to a new 1.5 mL tube (LoBind, Eppendorf), using tweezers rinsed in a stream of methanol followed by a stream of water before *every* manipulation.

Excised gel bands were then destained using the Pierce Silver Stain Kit for Mass Spectrometry, following manufacturer's protocols (all buffers made in Optima-grade water and all steps at RT). In brief, pieces were incubated in 200 μL of destain solution (twice for 15 min each, with gentle agitation), then washed in 300 μL of each of the following for 15 min with agitation: 100 mM ammonium bicarbonate in Optima-grade water (ABC); 50% 100 mM ABC, 50% Optima-grade acetonitrile (ACN); and 100% ACN. Dehydrated sections were dried to completion by speed vacuum (~10 min) and stored at $-80\text{ }^{\circ}\text{C}$.

In-Gel Digestion

Gel sections were reduced in 100 μL of 10 mM DTT in 100 mM ABC for 30 min at RT, then alkylated with 100 μL of 300 mM iodoacetamide in 100 mM ABC (final concentration 150 mM) for 30 min at RT in the dark. Sections were then incubated with agitation for 15 min at RT in 300 μL of each of the following: 100 mM ABC; 50% 100 mM ABC, 50% acetonitrile (ACN); and 100% ACN. Dehydrated sections were dried by speed vacuum (~10 min). Two μL (500 ng) of Trypsin Gold (Promega), activated for 15 min at $30\text{ }^{\circ}\text{C}$, was then added to each tube and allowed to rehydrate into gel sections with 40 μL of ice-cold 100 mM ABC for 30 min on ice. 20–30 μL of 100 mM ABC was then added, and rehydration was allowed to continue for 1 h on ice. Gel sections were then digested overnight at $30\text{ }^{\circ}\text{C}$. Digestion was stopped with 25 μL of Optima-grade 5% formic acid (FA; Fisher Scientific), and the supernatants were transferred to new 1.5 mL collection tubes (LoBind, Eppendorf). The remaining excised gel sections were dehydrated by incubation in 100 μL of each of the following for 10 min with agitation at RT, followed by supernatant collection: 50% ACN, 5% FA; 80% ACN, 5% FA; 100% ACN. For each excised gel section, all supernatants were combined into one 1.5 mL collection tube, speed-vacuumed to dryness, and stored at $-80\text{ }^{\circ}\text{C}$ to await MS analysis.

Nanocapillary Liquid Chromatography and Mass Spectrometry

Extracted peptides from all 10 excised gel sections of *B. canadensis*, sediment, and buffer GuHCl extracts were resuspended in 18 μL of "Buffer A" (95% Optima grade water, 5% Optima grade ACN, 0.2% FA) and centrifuged at 21 000g for 20 min at $4\text{ }^{\circ}\text{C}$ to pellet any debris. Nine μL of each supernatant was then transferred to a clean Microsolv RSA AQ autosampler vial. "Blank" and sediment negative controls were run first both to detect any initial protein contamination within the instrument and to prevent any potential bone protein carryover from fossil samples. Six μL of each supernatant was injected onto a self-packed C18 Aqua (Phenomenex) trap column (3 cm L, 150 μm ID, 3 μm d_p , 125 Å pore size) using a Dionex UltiMate 3000 UHPLC System. Peptides were washed and desalted in Buffer A for 12 min at a rate of 3 $\mu\text{L}/\text{min}$, then transferred to and eluted from a self-packed analytical column (C18 Aqua, 15 cm L, 75 μm ID, 3 μm d_p , 125 Å pore size) and spray emitter (12 cm L, 15 μm ID, New Objective, self-packed with 2 mm L C18 Aqua resin) with the following gradient: 5% B at 0 min, 5% B at 12 min, 40% B at 75 min, 85% B at 78 min, 85% B at 81 min, 5% B at 84 min, and 5% B at 100 min. (B: 95% Optima grade ACN, 5% Optima grade water, 0.2% FA). Spectra were acquired using a custom 12T LTQ-Velos FT-ICR mass spectrometer (Thermo Fisher Scientific) for tandem MS/MS. Full-scan FT MS1 spectra were obtained with a 400–1800 m/z scan range at resolving power 85.7k. The top eight most abundant peaks per MS1 scan were selected for fragmentation by collision-induced

dissociation (CID), and MS2 scans were performed in the ICR cell (FT/FT) in centroid mode with the following parameters: isolation window of 4 m/z , normalized collision energy of 35%, activation q of 0.25, and duration of 15 ms at a resolving power of 42.9k. Dynamic exclusion was enabled with the following parameters: repeat count = 2, repeat duration = 45 s, exclusion list size = 500, and exclusion duration = 30 s.

Data Analysis

Spectra from all 10 gel sections from *B. canadensis* GuHCl extract and 20 sections of negative control GuHCl extracts (sediment and buffer) were searched in PEAKS⁴⁷ (version 8.0, Bioinformatics Solutions) using a 10 ppm mass tolerance for precursors and 0.05 Da for fragment ions. Up to three missed cleavages and nonspecific cleavage at one end of the peptide were allowed. Carbamidomethylated cysteine was set as a fixed modification, and oxidation [M], oxidation or hydroxylation [RYFPNKD], [G] @C, carboxymethyl [KW, X@N], and deamidation [NQ] were set as variable modifications. Spectra were searched against the full SwissProt database (including sequences for common laboratory contamination) and also against NCBI Archosauria (Aves + Crocodylia). In both searches, PEAKS PTM⁴⁸ and SPIDER were enabled to account for unspecified PTMs and mutations. Results were filtered using $\leq 1\%$ FDR for peptide spectral matches (PSMs) and a protein score of $\geq -\log_{10} p 30$, plus at least one unique peptide.

Spectra were also searched in Byonic⁴⁹ 2.7.105 (Protein Metrics, San Carlos, CA) using similar parameters as for PEAKS above, with some variations because of the differing capabilities of the search programs. Three searches were performed against a database containing ~600 target proteins, including bird collagens and hemoglobins and common contaminants such as keratins, dermcidin, trypsin, and albumin. The first search considered semitryptic peptides with up to three missed cleavages, a mass accuracy of 0.05 Da for fragments, and, along with the PTMs listed for PEAKS searches above, added ammonia loss [@N-terminal Cys], Gln > pyro-Glu [@N-terminal Gln], Glu > pyro-Glu [@N-terminal Glu], dioxidation [@Trp], and acetylation [@Protein N-terminus] as variable PTMs. The second search considered fully tryptic peptides with up to two missed cleavages, along with at most one amino acid substitution from the database sequence. The third search was a wildcard search, which allowed a fragment mass accuracy of 30.0 ppm and a wildcard in the range -40 to +100 Da. A wildcard search considers “all” mass deltas within the range by setting the modification mass to be the difference between the candidate peptide mass and the mass of the precursor ion. This search finds unanticipated sequence variants and modifications; however, it is less sensitive to exact matches because of the increased size of the search space. Proteins were accepted as true that met the criteria of 2+ above the log probability score of the first decoy hit in each search.

Phylogenetic Analysis

Peptides were aligned using Seaview 4.5.4 x64 against 28 Archosauria, Testudine, and Lepidosauria collagen I alpha 1 and alpha 2 sequences derived from NCBI number (Table S1). A basic consensus method was used to generate two different *B. canadensis* collagen I alpha 1 and alpha 2 sequences. The first sequence contained only peptides detected here (Figure S1), and the second contained the peptides detected here and the peptides previously

detected.²⁸ After manual concatenation, MrBayes files were generated in Mesquite 3.04.⁵⁰ Using MrBayes v 3.2.2 x64,⁵¹ the collagen I matrix was searched with the following parameters: outgroup *Thamnophis sirtalis*, prset aamodelpr = mixed, 1 250 000 generations with sampling frequency 200, and burnin of 312 500 generations. Trees were output in FigTree version 1.4.0.

RESULTS AND DISCUSSION

Silver Staining

The GuHCl fraction from *B. canadensis* showed distinct bands on the gel that were not observed in the GuHCl extracts from sediment or buffer controls (Figure 2A) or any of the ABC extractions. We observed four relatively concentrated bands in the lane containing *B. canadensis* GuHCl extract at ~250, ~60, ~45, and ~17 kDa and a lighter band at ~120 kDa, with varying degrees of smearing in between. Silver stains of ancient protein extracts have typically shown a higher degree of smearing than observed here,^{27,28,52,53} and it has been hypothesized that such smearing was the result of peptide breakages at various locations, producing a cascade of fragments rather than peptides of uniform length that converge in tight bands.^{54,55} However, we suggest the possibility that smearing of these extracts in electrophoresis may be reduced relative to other reports from fossils because the protocol that we employed resulted in a steep reduction of residual EDTA in the GuHCl extracts when compared with previous studies. EDTA has been shown to cause smearing in SDS-PAGE gels of extant bone extractions⁵³ and is difficult to remove by dialysis.⁵⁶ The extraction performed on MOR 2598 in Schweitzer et al.²⁸ combined EDTA supernatants from demineralization directly with GuHCl supernatants prior to dialysis and subsequent gel electrophoresis. In our protocol, EDTA supernatants were kept separate from subsequent fractions. Although the pellet was not washed after EDTA supernatants had been removed, the inclusion of an ABC extraction between the EDTA and GuHCl incubations would substantially reduce the amount of EDTA salts in the subsequent GuHCl fraction (see Figure 1).

Although staining within the *B. canadensis* GuHCl lane and absence of staining in the control lanes supports the hypothesis that proteinaceous material is exclusively present in the fossil sample, silver staining of a SDS-PAGE assay alone is not sufficiently specific or conclusive to argue for the recovery of endogenous proteins from fossils. Thus we conducted an in-gel digestion and mass spectrometric analysis of the *B. canadensis*, sediment, and buffer GuHCl extract gel lanes (Figure 2B) to characterize any proteins present within these bands.

Mass Spectrometry

Since Schweitzer et al.,²⁸ a large number of extant archosaur genomes (e.g., Green et al.,⁵⁷ Prum et al.⁵⁸) have been completed; this increase in diversity allows for more extensive detection of collagen sequences. Eight peptides of collagen I were identified by tandem mass spectrometry of the *B. canadensis* GuHCl extract (Table 1), including five sequences of collagen I alpha 1 (Figures 3 and 4, Figures S2–S12) and three of collagen I alpha 2 (Figure 5, Figures S13–S16). All eight sequences were recovered from gel section 2 (~250 kDa),

which was one of the most heavily stained regions of the gel lane (Figure 2B). No collagen I sequences were observed in the negative controls (sediment and buffer-only samples). However, both the *B. canadensis* samples and negative controls contained abundant human keratin and trypsin peptides (Table S2). These common and well-recognized laboratory contaminants accounted for the majority of the identified PSMs in all samples, along with minute amounts of other contamination from human skin and the laboratory environment (e.g., dermcidin, desmoplankin, bovine casein). *B. canadensis* samples and sediment samples also contained an overlapping environmental signal, including fungal proteins (e.g., 40S and 60S ribosomal proteins) and bacterial proteins (e.g., trehalose-binding lipoprotein) that were not observed in the buffer-only negative control. This indicates that these sequences derive from the burial environment and not from the laboratory or reagents used. Metrics for MS/MS scans and PSM identifications from all samples are provided in Table S3. Although the search parameters we employed used error tolerance parameters (10 ppm precursor/0.05 Da fragments) wider than the accuracy capabilities of the most recent generation of mass spectrometers, the actual precursor and fragment errors we observe for all identified collagen PSMs are almost exclusively within a very narrow range. Specifically, all precursor errors are <4.5 ppm, and we report only five fragment ions with an error greater than 20 ppm in PEAKS and three with an error greater than 0.01 Da in Byonic across all collagen I spectra. A list of the detected precursor error, largest fragment error, and average fragment error for each reported spectra in both PEAKS and Byonic is provided in Table S4.

Of the five peptides recovered for collagen I alpha 1, two were previously identified from this specimen by Schweitzer et al.²⁸ GSAGPP(OH)GATGFP(OH)GAAGR (Figure 3A) and GATGAP(OH)GIAGAP(OH)GFP(OH)GAR (Figure 3B). This represents a replication of 24.2% of the original collagen I alpha 1 sequence data. Recovery of sequences identical to those reported in Schweitzer et al.,²⁸ despite multiple alterations to the experimental method and analyses in a different laboratory setting, demonstrates reproducibility of the previous results and confirms that these peptides are derived from the fossil sample itself and not modern contamination from the laboratory environment. Furthermore, these two peptides were associated with the most individual PSMs of all identified sequences; four to five PSMs were recovered for each peptide, whereas many of the new sequences are known from one or two spectra (Table 1). The fact that the two most abundant peptides in this *B. canadensis* extraction were also those identified in Schweitzer et al.²⁸ suggests that these may represent some of the most abundant preserved peptides in the fossil specimen as a whole and supports the hypothesis that these peptides may have a particularly high preservation potential because of their physically shielded location in the collagen fibril.⁴⁵ In addition to the two reidentified sequences, we detected three new peptides of collagen I alpha 1 (Table 1), which represent a combined length of 38 amino acid residues. These include GF(OH)GADGIAGP(OH)K, GFP(OH)GLP-GP(OH)SGEPGK, and GQAGVMGFP(OH)GPK (Figure 4). When considered together with the previously identified *B. canadensis* collagen I alpha 1 sequence data (113 residues on UniProt, P86289), this represents an increase of 33.6% in known sequence length (an increase from 10.7 to 14.3% of the entire mature sequence when compared with chicken collagen I alpha 1, P02457). Three new collagen I alpha 2 peptides were also identified (Table 1), which have a combined length of 42 amino acid residues. These include EGPVGF(OH)GADGR (Figure

5A), GATGLP(OH)GVAGAP(OH)GLP(OH)GPR, and GEP(OH)GNIGFP(OH)GPK (Figure 5B). Collagen I alpha 2 sequence data were more poorly represented in Schweitzer et al.²⁸ (36 residues on UniProt, P86290) than alpha 1; the addition of sequence data from the three new peptides presented here represents an increase of 116.7% (an increase from 3.5 to 7.6% of the entire mature sequence when compared with chicken collagen I alpha 2, P02467). These increases are important because they demonstrate the need for continued reassessment of fossil specimens as advances in technology and updates or alterations in methodology become available. Additionally, the modified protocols allowed us to obtain these new data using less sample material and instrument time, and although the original sequences were obtained over many protein extraction assays and MS injections over the course of months,²⁸ we acquired roughly the same amount of sequence data in a single experiment.

Proteomic analyses of archeological fossil specimens have been able to identify numerous PTMs, both biological (e.g., glycosylation, acetylation, methylation)^{13,21,55} and diagenetic (e.g., deamidation, carboxymethylation [CML]);^{13,14,21} however, the eight collagen I sequences identified in this study display only two PTMs: hydroxylation of proline and oxidation of methionine. The oxidation of methionine is variably observed on one sequence (GQAGVM(OX)GFP(OH)GPK, Figure S9) and is known to occur as an artifact of sample preparation.⁵⁹ Hydroxylation of proline, however, is a crucial *in vivo* PTM for the structure and function of collagen I, as hydroxylated prolines play an integral role in the formation of its triple helical tertiary structure^{60,61} and cannot be produced by bacteria.⁶² Thus the presence of this modification in collagen sequences derived from fossils supports an endogenous origin. The collagen I sequences identified in this study each contain between one and three hydroxylated prolines, and that number remains consistent for each peptide across multiple spectra.

Consistent with Schweitzer et al.²⁸ we do not observe some of the peptide modifications characteristic of degradation on the identified collagen sequences, such as truncation, CML, or variability in quantities of hydroxylated prolines.^{13,21,55} Deamidation, which is widely observed in archeological specimens,^{13,14,21,55,63,64} is also absent; however, it should be noted that there are only three asparagine and glutamine residues (1 Asp, 2 Gln) present across all of these collagen I spectra from all peptides (encompassing 280 residues). Thus it is likely that the apparent lack of deamidation is a factor of effective sampling bias as opposed to a true signal, particularly because vessel proteins detected from this same *B. canadensis* showed deamidation that was variable and incomplete (~20–85%QN).³¹ Regardless, deamidation does not give a reliable signal for age⁶⁵ and is therefore an inappropriate criteria of authenticity for this data set.

Beyond deamidation, the lack of degradation markers within the eight sequences (e.g., truncation, variable hydroxylproline) is consistent with Schweitzer et al.²⁸ It is possible that given their repeated identification and absence of degradation, these sequences represent the most durable regions of the collagen I molecule, preserved precisely *because* they are least susceptible to degradation. As observed in the sequences identified in Schweitzer et al.,²⁸ the peptides identified in this analysis possess few acidic residues (aspartic acid and glutamic acid), which has been hypothesized to limit their solubility and potential for proteolytic degradation.⁴⁵ Just five acidic residues are present across all peptides reported here, which is

less than predicted when compared with the average number in equivalent lengths of human collagen (2 of 8.4 predicted for 73 collagen I alpha 1 residues, and 3 of 4.7 predicted for 42 collagen I alpha 2 residues),^{45,66} and four of the eight recovered peptides do not contain any acidic residues (Table 1).

Because the eight identified peptides lack distinct markers for degradation, it is necessary to explicitly consider the possibility that they have arisen from modern contamination and to establish criteria other than deamidation to support endogeneity. The collagen I sequences recovered here are present in minute amounts, primarily comprising two peptides previously identified from the fossil and occur exclusively in the *B. canadensis* sample amidst a clear environmental signal that matches that of the sediment. Furthermore, none of the samples contain anomalous proteins besides those that can be explained by sample preparation (e.g., trypsin), human skin particles (e.g., keratin, dermicidin, desmoplankin), common laboratory contaminants (e.g., bovine casein), or homologous overlap with fungal and bacterial proteins (Table S2). When considered in light of the rigorous anticontamination protocols practiced during sample and LC–MS preparation (see Methods), it is not parsimonious to propose that these collagen I sequences are derived from modern contamination. Although these measures and data make it highly unlikely that the sequences we report are exogenously derived, we conducted phylogenetic analyses to further confirm that they were biologically consistent with what is predicted based on hypothesized morphology- and osteology-based evolutionary relationships of *B. canadensis* to other extant and extinct dinosaurs and their relatives.

Phylogenetic Analyses

We first conducted phylogenetic analyses solely upon modern archosaur collagen sequences from existing databases to establish their phylogenetic topology free of any potential polarization by the highly fragmentary fossil sequences (Figure S17). We incorporated into our analyses all NCBI archosaur sequences for which both the alpha 1 and alpha 2 chains of collagen are known (21 species) and eight outgroups of snakes, lizards, and turtles (Table S1). The resulting topology places the paleognathans (*Struthio camelus australis* and *Tinamus guttatus*) as more derived than the galliformes (*Gallus gallus* and *Coturnix japonica*), essentially placing Paleognathae within basal Neognathae instead of as a separate sister clade, the latter a relationship robustly supported by DNA and morphological analyses.^{58,67} However, despite the interchange of the basal-most clades outside of Neoaves (Paleognathae and Galloanserae) and the more minor topological deviations within subclades of Neoaves,⁵⁸ the phylogeny constructed from collagen I sequences (both alphas 1 and 2) is broadly reflective of phylogenies for Archosauria using other molecular and morphological data,⁶⁸ basal birds are distinguished from more derived neoavians, and more distantly related species of saurians (e.g., turtles, geckos, anoles, and snakes) are resolved in a manner consistent with DNA and other molecular studies.^{69,70} While this topology does not directly reflect other phylogenetic hypotheses of birds, it provides a framework with which to elucidate the molecular phylogenetic placement of *B. canadensis* within Archosauria and to exclude the recovered peptides as contamination if they are derived from a non-Archosaurian source.

Phylogenetic analyses of the newly obtained fossil sequences for collagen I alpha 1 and alpha 2, both when considered separately and when combined with the previous sequences, resulted in placement of *B. canadensis* in basal archosauria (Figure 6). When analyzed separately, the new collagen I sequences place *B. canadensis* between birds and crocodylians, similar to hypothesized placements for Ornithischia based upon morphological data (e.g., Brochu⁶⁸) but closer to the *Alligator* clade than bird clades (Figure 6A). This is likely caused by the presence of one collagen I alpha 2 peptide that shares 100% identity with crocodylians *Alligator sinensis* and *Alligator mississippiensis* and only 94% identity with any aves species, GATGLP(OH)GVAGAP(OH)GLP(OH)GPR (Figures S15 and S16). However, when the new sequences are combined with those reported in Schweitzer et al.,²⁸ *B. canadensis* is grouped into the basal most clade of birds (in this analysis), in a polytomy with *Gallus gallus* and *Coturnix japonica* (Figure 6B). The ambiguous placement of *B. canadensis* between the two groups indicates two important things: (1) the sequences obtained from MOR 2598 are not contaminants derived from the specific taxa it shares similarities with in this analysis (e.g., *Alligator mississippiensis* or *Gallus gallus*) because they are not completely homologous with either species and (2) the sequences derived from the *B. canadensis* have similarities with *both* crocodylians and basal birds, which are alternately more prominent depending on the part of the sequence that is being analyzed. These findings are consistent with what would be predicted from fragmentary ornithischian sequences without a complete, known sequence to match spectra against: close similarities with the two nearest extant phylogenetic bracketing groups,⁷¹ without complete overlap of either. Thus the results of the phylogenetic analyses strongly support an endogenous origin and greatly reduce the possibility that these eight peptide sequences may be the result of contamination.

To be consistent with previous studies,^{13,31,55} we have used the sequences as they are identified by bioinformatics software in our phylogenetic analyses. However, it must be noted that without detection of a complete ion series, the identity of certain residues in poorly fragmented regions may be ambiguous. For example, the differentiation of [...]GAP(OH) and [...]GSP cannot be confidently resolved without at least one fragment ion covering the region between the second and third residues. Because these ambiguities are limited by surrounding fragment ions, it is not in question whether these peptides are correctly identified as collagen I. We can therefore be confident that their sequences are generally accurate and that their placement within Archosauria is correct. However, very small differences in residues can affect phylogenetic placement among closely related species, especially when a protein sequence is fragmentary, making *exact* placement of *B. canadensis*, or any fossil specimen, tentative. This issue illustrates the next frontier in paleomolecular (>1 Ma) research: confident sequence validation, beyond simple identification and confirmation of authenticity. To construct the most accurate phylogenetic topologies possible, future studies must investigate the variability of amino acid sequences in sequence tags where fragmentation is poor, in addition to exploring techniques to increase fragmentation of ancient, potentially altered proteins.

CONCLUSIONS

We describe herein the recovery of eight peptide sequences of collagen I from the nonavian dinosaur *Brachylophosaurus canadensis* (MOR 2598), previously shown to contain protein fragments. Here we used different methods of sample preparation, MS instrumentation, and data analysis, and conducted these analyses in a different laboratory space, separated by several years from Schweitzer et al.²⁸ These peptides included two sequences that were recovered in both studies, making it highly unlikely that these identifications arose from contamination. Furthermore, phylogenetic analyses placed these sequences within Archosauria and revealed similarities to both crocodylians and basal birds, depending on which parts of the sequence were being analyzed. These findings: (1) offer further, independent support that peptides can persist in specimens tens of millions of years old; (2) bolster the validity of Schweitzer et al.,²⁸ and (3) substantially expand the collagen sequence coverage of *B. canadensis* collagen I compared with what was previously characterized. The reidentification of two previously obtained peptides from this specimen, which are also the two most abundant in this study, suggests that these peptides are preferentially preserved, thus supporting the hypothesis that protein/collagen fiber structure plays a role in preservation.⁴⁵ Additionally, this study demonstrates the utility of reexamining previous specimens with updated methods and instrumentation, as we were able to obtain roughly the same amount of sequence data as Schweitzer et al.²⁸ with far less sample material.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Neil Kelleher Lab Group and Haylee Thomas for logistical support, Bob Harmon, Carrie Ancell, and the MOR field crew, and Jack Horner for permitting destructive analyses of this specimen. This work was funded under an NSF INSPIRE grant to M.H.S., E.R.S., W.Z., and M.B., with postdoctoral support from NIH K12GM102745 to T.P.C., and with support from NIH P41GM108569, for the National Resource for Translational and Developmental Proteomics based at Northwestern University, to C.J.D., P.M.T., and N.L.K.

References

1. Collins MJ, Waite ER, van Duin ACT. Predicting protein decomposition: the case of aspartic-acid racemization kinetics. *Philos Trans R Soc, B*. 1999; 354:51–64.
2. Ostrom PH, Gandhi H, Strahler JR, Walker AK, Andrews PC, Leykam J, Stafford TW, Kelly RL, Walker DN, Buckley M, Humpala J. Unraveling the sequence and structure of the protein osteocalcin from a 42 ka fossil horse. *Geochim Cosmochim Acta*. 2006; 70(8):2034–2044.
3. Buckley M, Anderung C, Penkman K, Raney BJ, Gotherstrom A, Thomas-Oates J, Collins MJ. Comparing the survival of osteocalcin and mtDNA in archaeological bone from four European sites. *Journal of Archaeological Science*. 2008; 35:1756–1764.
4. Buckley M, Collins M, Thomas-Oates J, Wilson JC. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*. 2009; 23(23):3843–3854. [PubMed: 19899187]
5. Buckley M, Kansa SW, Howard S, Campbell S, Thomas-Oates J, Collins MJ. Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *Journal of Archaeological Science*. 2010; 37:13–20.
6. Buckley M, Larkin N, Collins M. Mammoth and Mastodon collagen sequences; survival and utility. *Geochim Cosmochim Acta*. 2011; 75:2007–2016.

7. Buckley M, Wadsworth C. Proteome degradation in ancient bone: diagenesis and phylogenetic potential. *Palaeogeogr, Palaeoclimatol Palaeoecol.* 2014; 416:69–79.
8. Wadsworth C, Buckley M. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun Mass Spectrom.* 2014; 28:605–615. [PubMed: 24519823]
9. Richter KK, Wilson J, Jones AKG, Buckley M, van Doorn N, Collins MJ. Fish ‘n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments. *Journal of Archaeological Science.* 2011; 38(7):1502–1510.
10. Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RAR, Stafford TW, Holen SR, Collins MJ, Orlando L, Willerslev E, Gilbert MTP, Olsen JV. Proteomic Analysis of a Pleistocene Mammoth Femur Reveals More than One Hundred Ancient Bone Proteins. *J Proteome Res.* 2012; 11(2):917–926. [PubMed: 22103443]
11. Humpala JF, Ostrom PH, Gandhi H, Strahler JR, Walker AK, Stafford TW Jr, Smith JJ, Voorhies MR, George Corner R, Andrews PC. Investigation of the protein osteocalcin of *Camelops hesternus*: Sequence, structure and phylogenetic implications. *Geochim Cosmochim Acta.* 2007; 71(24):5956–5967.
12. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson PLF, Fumagalli M, Vilstrup JT, Raghavan M, Korneliusen T, Malaspina A-S, Vogt J, Szklarczyk D, Kelstrup CD, Vinther J, Dolocan A, Stenderup J, Velazquez AMV, Cahill J, Rasmussen M, Wang X, Min J, Zazula GD, Seguin-Orlando A, Mortensen C, Magnussen K, Thompson JF, Weinstock J, Gregersen K, Roed KH, Eisenmann V, Rubin CJ, Miller DC, Antczak DF, Bertelsen MF, Brunak S, Al-Rasheid KAS, Ryder O, Andersson L, Mundy J, Krogh A, Gilbert MTP, Kjaer K, Sicheritz-Ponten T, Jensen LJ, Olsen JV, Hofreiter M, Nielsen R, Shapiro B, Wang J, Willerslev E. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature.* 2013; 499(7456):74–78. [PubMed: 23803765]
13. Cleland TP, Schroeter ER, Schweitzer MH. Biologically and diagenetically derived peptide modifications in Moa collagens. *Proc R Soc London, Ser B.* 2015; 282(1808):20150015.
14. Welker F, Collins MJ, Thomas JA, Wadswley M, Brace S, Cappellini E, Turvey ST, Reguero M, Gelfo JN, Kramarz A, Burger J, Thomas-Oates J, Ashford DA, Ashton PD, Rowsell K, Porter DM, Kessler B, Fischer R, Baessmann C, Kaspar S, Olsen JV, Kiley P, Elliott JA, Kelstrup CD, Mullin V, Hofreiter M, Willerslev E, Hublin J-J, Orlando L, Barnes I, MacPhee RDE. Ancient proteins resolve the evolutionary history of Darwin’s South American ungulates. *Nature.* 2015; 522(7554): 81–84. [PubMed: 25799987]
15. Brown S, Higham T, Slon V, Pääbo S, Meyer M, Douka K, Brock F, Comeskey D, Procopio N, Shunkov M, Derevianko A, Buckley M. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci Rep.* 2016; 6:23559. [PubMed: 27020421]
16. Buckley M. A Molecular Phylogeny of *Plesiorcycteropus* Reassigns the Extinct Mammalian Order ‘Bibymalagasia’. *PLoS One.* 2013; 8(3):e59614. [PubMed: 23555726]
17. Waters MR, Stafford TW, McDonald HG, Gustafson C, Rasmussen M, Cappellini E, Olsen JV, Szklarczyk D, Jensen LJ, Gilbert MTP, Willerslev E. Pre-Clovis Mastodon Hunting 13,800 Years Ago at the Manis Site, Washington. *Science.* 2011; 334(6054):351–353. [PubMed: 22021854]
18. Nielsen-Marsh CM, Ostrom PH, Gandhi H, Shapiro B, Cooper A, Hauschka PV, Collins MJ. Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka. *Geology.* 2002; 30(12):1099–1102.
19. Nielsen-Marsh CM, Richards MP, Hauschka PV, Thomas-Oates J, Trinkaus E, Pettitt PB, Karavanic I, Poinar H, Collins MJ. Osteocalcin protein sequences of Neanderthals and modern primates. *Proc Natl Acad Sci U S A.* 2005; 102(12):4409–4413. [PubMed: 15753298]
20. Buckley M, Kansa S. Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. *Archaeological and Anthropological Sciences.* 2011; 3(3):271–280.
21. Hill RC, Wither MJ, Nemkov T, Barrett A, D’Alessandro A, Dzieciatkowska M, Hansen KC. Preserved proteins from extinct bison *Latifrons* identified by tandem mass spectrometry; hydroxylysine glycosides are a common feature of ancient collagen. *Mol Cell Proteomics.* 2015; 14(7):1946–1958. [PubMed: 25948757]

22. Buckley M. Ancient collagen reveals evolutionary history of the endemic South American 'ungulates'. *Proc R Soc London, Ser B.* 2015; 282:20142671.
23. Demarchi B, Hall S, Roncal-Herrero T, Freeman CL, Woolley J, Crisp MK, Wilson J, Fotakis A, Fischer R, Kessler BM, Rakownikow Jersie-Christensen R, Olsen JV, Haile J, Thomas J, Marean CW, Parkington J, Presslee S, Lee-Thorp J, Ditchfield P, Hamilton JF, Ward MW, Wang CM, Shaw MD, Harrison T, Domínguez-Rodrigo M, MacPhee RDE, Kwekason A, Ecker M, Kolska Horwitz L, Chazan M, Kröger R, Thomas-Oates J, Harding JH, Cappellini E, Penkman K, Collins MJ. Protein sequences bound to mineral surfaces persist into deep time. *eLife.* 2016; 5:e17092. [PubMed: 27668515]
24. Avci R, Schweitzer MH, Boyd RD, Wittmeyer JL, TeranArce F, Calvo JO. Preservation of Bone Collagen from the Late Cretaceous Period Studied by Immunological Techniques and Atomic Force Microscopy. *Langmuir.* 2005; 21(8):3584–3590. [PubMed: 15807605]
25. Schweitzer MH, Wittmeyer JL, Horner JR, Toporski JK. Soft-Tissue Vessels and Cellular Preservation in *Tyrannosaurus rex*. *Science.* 2005; 307(5717):1952–1955. [PubMed: 15790853]
26. Schweitzer MH, Wittmeyer JL, Horner JR. Soft tissue and cellular preservation in vertebrate skeletal elements from the Cretaceous to the present. *Proc R Soc London, Ser B.* 2007; 274:183–197.
27. Schweitzer MH, Suo Z, Avci R, Asara JM, Allen MA, Arce FT, Horner JR. Analyses of Soft Tissue from *Tyrannosaurus rex* Suggest the Presence of Protein. *Science.* 2007; 316(5822):277–280. [PubMed: 17431179]
28. Schweitzer MH, Zheng W, Organ CL, Avci R, Suo Z, Freimark LM, Lebleu VS, Duncan MB, Vander Heiden MG, Neveu JM, Lane WS, Cottrell JS, Horner JR, Cantley LC, Kalluri R, Asara JM. Biomolecular Characterization and Protein Sequences of the Campanian Hadrosaur *B. canadensis*. *Science.* 2009; 324(5927):626–631. [PubMed: 19407199]
29. Lindgren J, Uvdal P, Engdahl A, Lee AH, Alwmark C, Bergquist K-E, Nilsson E, Ekström P, Rasmussen M, Douglas DA, Polcyn MJ, Jacobs LL. Microspectroscopic Evidence of Cretaceous Bone Proteins. *PLoS One.* 2011; 6(4):e19445. [PubMed: 21559386]
30. Schweitzer MH, Zheng W, Cleland TP, Bern M. Molecular analyses of dinosaur osteocytes support the presence of endogenous molecules. *Bone.* 2013; 52:414–423. [PubMed: 23085295]
31. Cleland TP, Schroeter ER, Zamdborg L, Zheng W, Lee JE, Tran J, Bern M, Duncan MB, Lebleu VS, Ahlf D, Thomas PM, Kalluri R, Kelleher NL, Schweitzer MH. Mass spectrometry and antibody-based characterization of blood vessels from *Brachylophosaurus canadensis*. *J Proteome Res.* 2015; 14:5252–5262. [PubMed: 26595531]
32. Brinckman J. Collagens at a Glance. *Top Curr Chem.* 2005; 247:1–6.
33. Fratzl, P. Collagen: Structure and Mechanics, and Introduction. In: Fratzl, P., editor. *Collagen: Structure and Mechanics.* Springer Science +Business Media, LLC; New York: 2008. p. 1-13.
34. Asara JM, Garavelli JS, Slatter DA, Schweitzer MH, Freimark LM, Phillips M, Cantley LC. Interpreting Sequences from Mastodon and *T. rex*. *Science.* 2007; 317:1324–1325. [PubMed: 17823333]
35. Buckley M, Walker A, Ho SYW, Yang Y, Smith C, Ashton P, Oates JT, Cappellini E, Koon H, Penkman K, Elsworth B, Ashford D, Solazzo C, Andrews P, Strahler J, Shapiro B, Ostrom P, Gandhi H, Miller W, Raney B, Zylber MI, Gilbert MTP, Prigodich RV, Ryan M, Rijdsdijk KF, Janoo A, Collins MJ. Comment on Protein Sequences from Mastodon and *Tyrannosaurus rex* Revealed by Mass Spectrometry. *Science.* 2008; 319(5859):33.
36. Pevzner PA, Kim S, Ng J. Comment on Protein Sequences from Mastodon and *Tyrannosaurus rex* Revealed by Mass Spectrometry. *Science.* 2008; 321(5892):1040b.
37. Collins MJ, Riley MS, Child AM, Turner-Walker G. A basic mathematical simulation of the chemical degradation of ancient collagen. *Journal of Archaeological Science.* 1995; 22:175–183.
38. Schweitzer MH, Zheng W, Cleland TP, Goodwin MB, Boatman E, Theil E, Marcus MA, Fakra SC. A role for iron and oxygen chemistry in preserving soft tissues, cells and molecules from deep time. *Proc R Soc London, Ser B.* 2014; 281(1775):20132741.
39. Trueman CNG, Martill DM. The long-term survival of bone: the role of bioerosion. *Archaeometry.* 2002; 44(3):371–382.

40. Butterfield N. Exceptional Fossil Preservation and the Cambrian Explosion. *Integr Comp Biol*. 2003; 43:166–177. [PubMed: 21680421]
41. Butterfield N. Organic preservation of non-mineralizing organisms and the taphonomy of the Burgess Shale. *Paleobiology*. 1990; 16(3):272–286.
42. Turner-Walker, G. *The Chemical and Microbial Degradation of Bones and Teeth*. Wiley; New York: 2008. p. 389
43. Schweitzer MH. Soft tissue preservation in terrestrial Mesozoic vertebrates. *Annu Rev Earth Planet Sci*. 2011; 39:187–216.
44. Collins MJ, Nielsen-Marsh CM, Hiller J, Smith CI, Roberts JP, Prigodich RV, Wess TJ, Csapo J, Millard AR, Turner-Walker G. The survival of organic matter in bone. *Archaeometry*. 2002; 44(3): 383–394.
45. San Antonio JD, Schweitzer MH, Jensen ST, Kalluri R, Buckley M, Orgel JPRO. Dinosaur Peptides Suggest Mechanisms of Protein Survival. *PLoS One*. 2011; 6(6):e20381. [PubMed: 21687667]
46. Cleland T, Stoskopf M, Schweitzer M. Histological, chemical, and morphological reexamination of the heart of a small Late Cretaceous. *Naturwissenschaften*. 2011; 98(3):203–211. [PubMed: 21279321]
47. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G. PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2003; 17:2337–2342. [PubMed: 14558135]
48. Han X, He L, Xin L, Shan B, Ma B. PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *J Proteome Res*. 2011; 10(7):2930–2936. [PubMed: 21609001]
49. Bern M, Kil YJ, Becker C. Byonic: Advanced peptide and protein identification software. *Current Protocols in Bioinformatics*. 2012; 40:13.20.10–13.20.14.
50. Maddison, WP., Maddison, DR. *Mesquite: A Modular System for Evolutionary Analysis*, version 3.04. The Mesquite Project; Vancouver, British Columbia: 2015. <http://mesquiteproject.org>
51. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19:1572–1574. [PubMed: 12912839]
52. Tuross N. Alterations in fossil collagen. *Archaeometry*. 2002; 44(3):427–434.
53. Cleland TP, Voegelé K, Schweitzer MH. Empirical Evaluation of Bone Extraction Protocols. *PLoS One*. 2012; 7(2):e31443. [PubMed: 22348088]
54. Tuross N, Hare PE. *Collagen in Fossil Bone*. Carnegie Institution of Washington: Yearbook. 1978; 77:891–895.
55. Cleland TP, Schroeter ER, Feranec RS, Vashishth D. Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proc R Soc London, Ser B*. 2016; 283:20160593.
56. Pereira-Mouries L, Almeida MJ, Ribeiro C, Peduzzi J, Barthélemy M, Milet C, Lopez E. Soluble silk-like organic matrix in the nacreous layer of the bivalve *Pinctada maxima*: a new insight in the biomineralization field. *Eur J Biochem*. 2002; 269:4994–5003. [PubMed: 12383258]
57. Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, Vandeweghe MW, StJohn JA, Capella-Gutiérrez S, Castoe TA, Kern C, Fujita MK, Opazo JC, Jurka J, Kojima KK, Caballero J, Hubley RM, Smit AF, Platt RN, Lavoie CA, Ramakodi MP, Finger JW, Suh A, Isberg SR, Miles L, Chong AY, Jaratlerdsiri W, Gongora J, Moran C, Iriarte A, McCormack J, Burgess SC, Edwards SV, Lyons E, Williams C, Breen M, Howard JT, Gresham CR, Peterson DG, Schmitz J, Pollock DD, Haussler D, Triplett EW, Zhang G, Irie N, Jarvis ED, Brochu CA, Schmidt CJ, McCarthy FM, Faircloth BC, Hoffmann FG, Glenn TC, Gabaldón T, Paten B, Ray DA. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science*. 2014; 346(6215): 1254449. [PubMed: 25504731]
58. Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. 2015; 526:569–573. [PubMed: 26444237]

59. Liu H, Ponniah G, Neill A, Patel R, Andrien B. Accurate determination of protein methionine oxidation by stable isotope labeling and LC-MS analysis. *Anal Chem.* 2013; 85(24):11705–11709. [PubMed: 24200102]
60. Hulmes, DJS. Collagen Diversity, Synthesis and Assembly. In: Fratzl, P., editor. *Collagen: Structure and Mechanics.* Springer Science +Business Media, LLC; New York: 2008. p. 15-47.
61. Engel J, Bachinger HP. Structure, Stability and Folding of the Collagen Triple Helix. *Top Curr Chem.* 2005; 247:7–33.
62. Rasmussen M, Jacobsson M, Björck L. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem.* 2003; 278(34):32313–32316. [PubMed: 12788919]
63. van Doorn NL, Wilson J, Hollund H, Soressi M, Collins MJ. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun Mass Spectrom.* 2012; 26:2319–2327. [PubMed: 22956324]
64. Wilson J, van Doorn NL, Collins M. Assessing the extent of bone degradation using glutamine deamidation in collagen. *Anal Chem.* 2012; 84:9041–9048. [PubMed: 23030643]
65. Schroeter ER, Cleland TP. Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Commun Mass Spectrom.* 2016; 30:251–255. [PubMed: 26689157]
66. Miller, EJ. Chemistry of the Collagens and Their Distribution. In: Reddi, AH., Pieze, KA., editors. *Extracellular Matrix Biochemistry.* Elsevier Science Publishing Co., Inc; New York: 1984. p. 41-81.
67. Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T. A Phylogenomic Study of Birds Reveals Their Evolutionary History. *Science.* 2008; 320(5884):1763–1768. [PubMed: 18583609]
68. Brochu CA. Progress and future directions in archosaur phylogenetics. *J Paleontol.* 2001; 75(6): 1185–1201.
69. Field DJ, Gauthier JA, King BL, Pisani D, Lyson TR, Peterson KJ. Towards consilience in reptile phylogeny: miRNAs support an archosaurs, not lepidosaur, affinity for turtles. *Evol Dev.* 2014; 16(4):189–196. [PubMed: 24798503]
70. Hedges SB. Amniote phylogeny and the position of turtles. *BMC Biol.* 2012; 10(1):64. [PubMed: 22839753]
71. Witmer, LM. The Extant Phylogenetic Bracket and the Importance of Reconstructing Soft Tissues in Fossils. In: Thomason, J., editor. *Functional Morphology in Vertebrate Paleontology.* Cambridge University Press; Cambridge, U.K: 1995. p. 19-33.

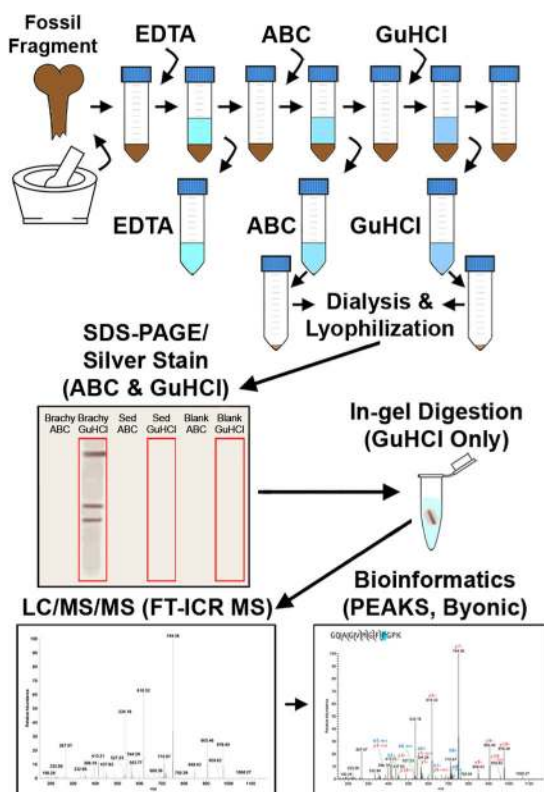


Figure 1. Graphical depiction of experimental workflow. Ground fossil was incubated (sequentially) in EDTA, ABC, and GuHCl. ABC and GuHCl extracts were separated by gel electrophoresis, and GuHCl lanes were in-gel digested, followed by LC–MS/MS and analysis of acquired spectra with two bioinformatics software packages.

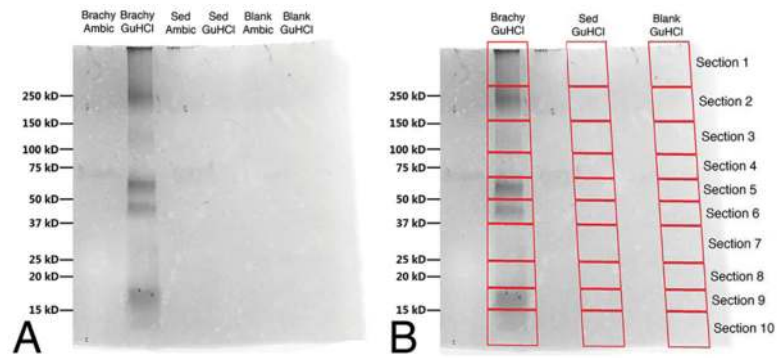


Figure 2. Silver stained gel of ABC and GuHCl extracts from *B. canadensis*, sediment, and blank samples. (A) Distinct banding is observed only in the *B. canadensis* GuHCl lanes. (B) GuHCl lanes for dinosaur, sediment, and blank samples were each cut into 10 sections, as shown for in-gel digestion.

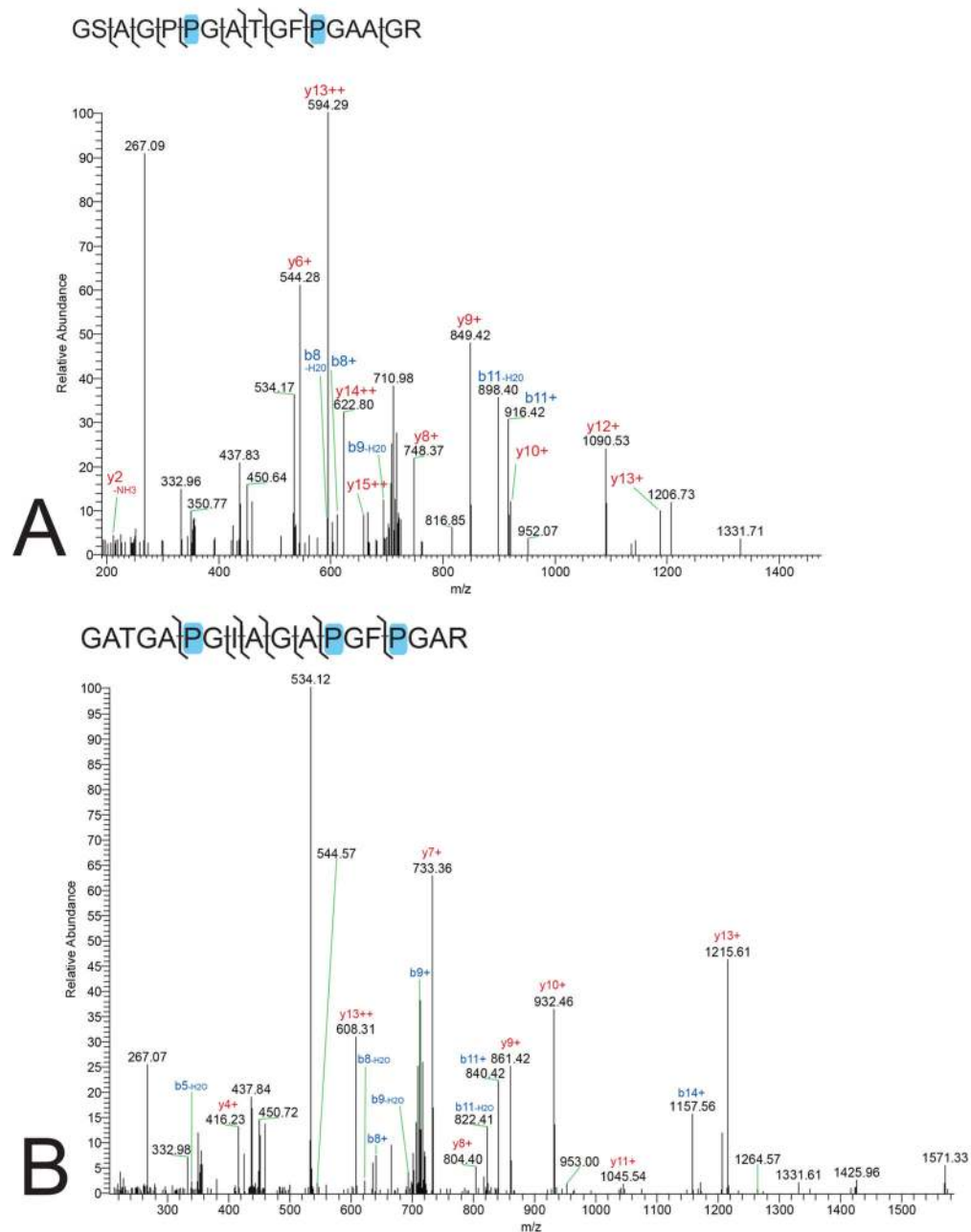


Figure 3. Annotated spectra of *B. canadensis* peptides identified in both the current analysis of MOR 2598 and Schweitzer et al.²⁸ Blue highlights denote a hydroxylation of a proline residue. (A) Collagen I alpha 1 peptide GSAGPPGATGFPGAAGR (20141017_Brachy_G_02.RAW; scan #1308). (B) Collagen I alpha 1 peptide GATGAPGIAGAPGFPGAR (20141017_Brachy_G_02.RAW; average of scans #1906, #1958, #1940, #1885).

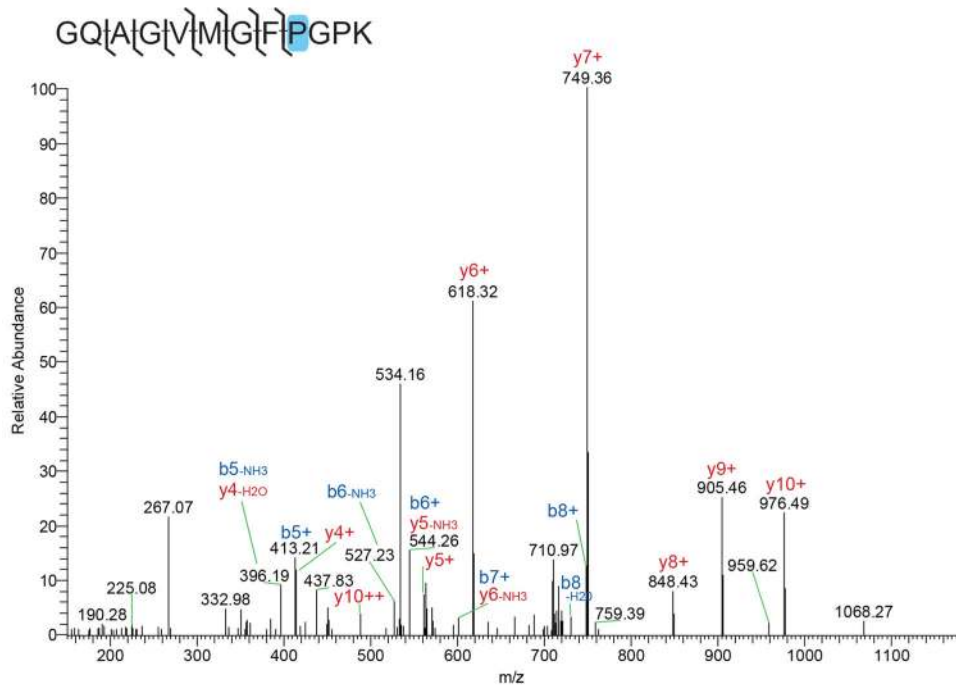


Figure 4. Annotated spectra of a new collagen I alpha 1 peptide of *B. canadensis* identified in this study, GQAGVMGFPGPK (20141017_Brachy_G_02.RAW; scan #2402). Blue highlight denotes a hydroxylation of a proline residue.

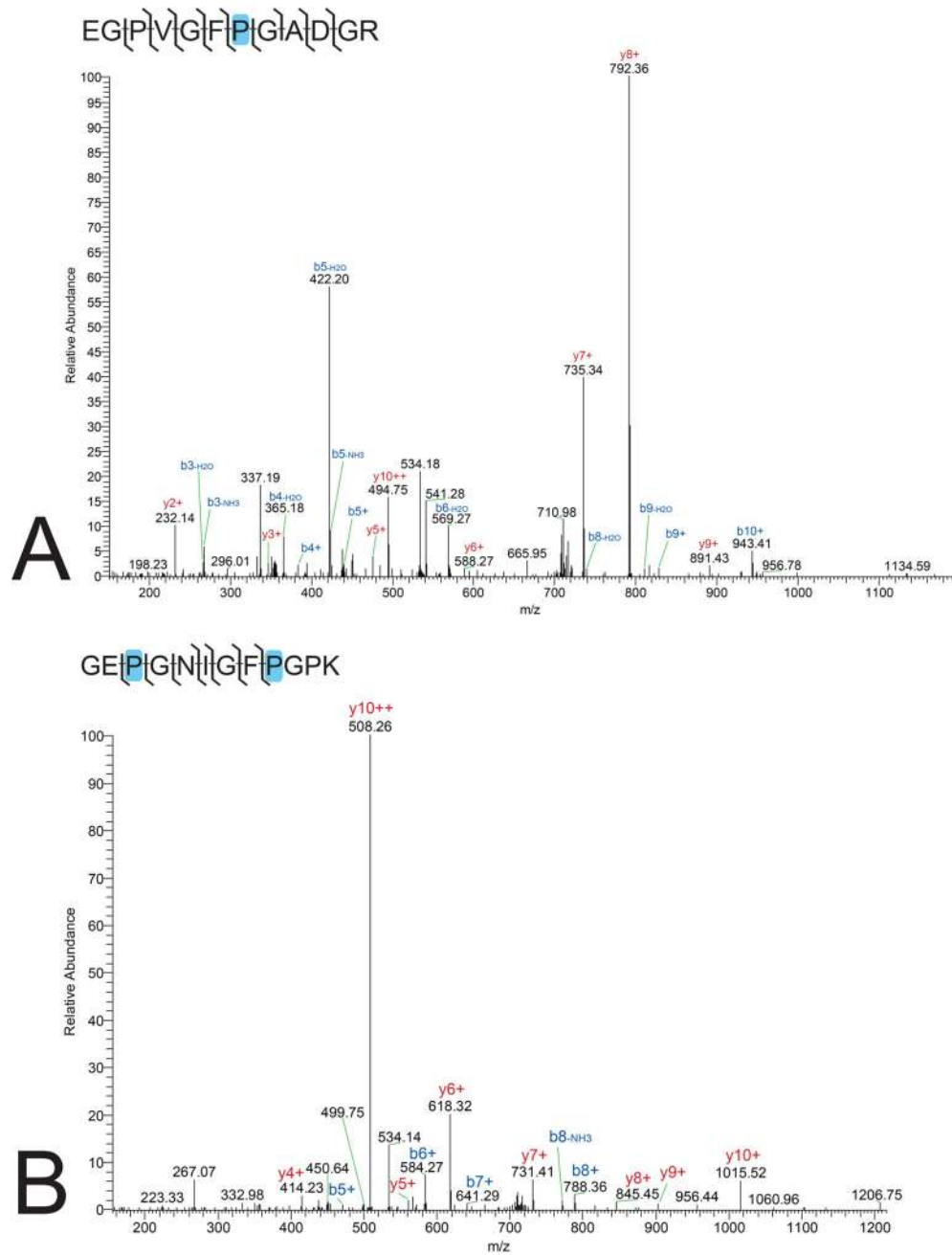


Figure 5. Annotated spectra of new collagen I alpha 2 peptides of *B. canadensis* identified in this study. Blue highlights denote hydroxylation of a proline residue. (A) EGPVGFPGADGR (20141017_Brachy_G_02.RAW; average of scans #1624 and #1636) and (B) GEPGNIGFPGPK (20141017_Brachy_G_02.RAW; scan #1635).

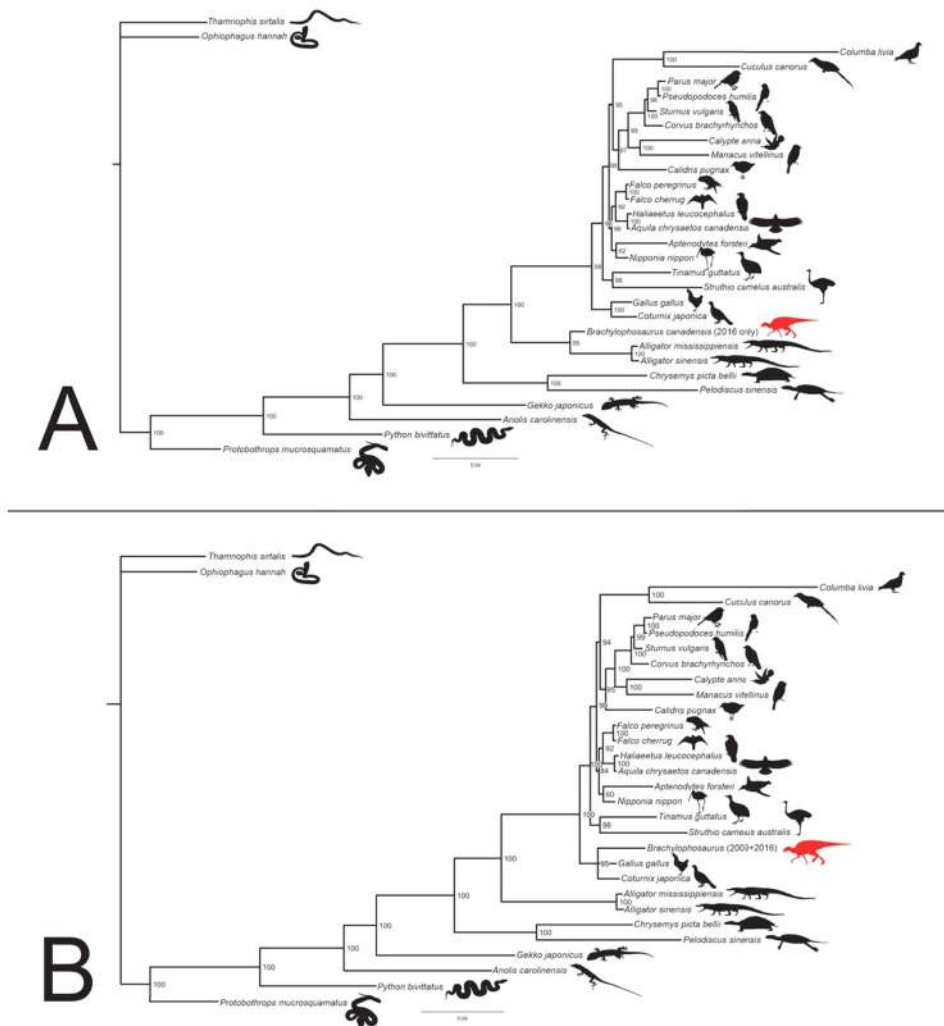


Figure 6. Phylogenetic trees obtained by analysis of (A) only the new *B. canadensis* peptides identified in this study and (B) a combined alignment of the new *B. canadensis* peptides with those previously identified.²⁸ (A) Analysis of the new peptides only (see Table 1) results in a topology that places *B. canadensis* closer to *Alligator* than aves. (B) When all known peptides are considered, *B. canadensis* is placed within the basal most (in this tree) clade of aves. Both trees place *B. canadensis* in basal archosauria, consistent with evolutionary hypothesis based on morphology (e.g., ref⁶⁸). We thank the following people for access to silhouettes and images for adaptation: Martein Brand (*Cyanopica cooki*; CCA 3.0 unported); John Gould & T. Michael Keeseey (*Topaza pella*; PDM 1.0); Rebecca Groom (*Falco peregrinus peregrinus*; CCA 3.0 unported); Scott Hartman (Alligatoridae, *Aspideretoides*, *Glyptops utahensis*, *Brachylophosaurus canadensis*; CCA-NC-SA 3.0 unported); Neil Kelley (*Aptenodytes forsteri*; CCA-SA 3.0 unported); Liftbarn (*Falco*, CCA-SA 3.0 unported); George Edward Lodge (*Tinamus major*; PDM 1.0); Matt Martyniuk and Michael Keeseey (*Struthio camelus australis*, CCA-SA 3.0 unported); Gareth Monger (*Sturnus vulgaris*; CCA 3.0 unported); Elisabeth Östman (Phasianidae; PDM 1.0); Peileppe (*Corvus*

brachyrhynchus; PDD 1.0); Ferran Sayol (*Parus atricapillus*; PDD 1.0); L. Shyamal (adapted *Trimeresurus malabaricus*; GFDL and *Aquila*; CCA 3.0 unported); Andrew Smith (adapted *Python natalensis* (1840); Public Domain); Steven (*Gekko gecko*; PDD 1.0); Gart Stolz (adapted *Notechis ater*; PDD 1.0); Aaand Titus and Geeta Pereira (*Ophiophagus Hannah*; CCA-SA 3.0 unported), Steven Traver (*Gallus gallus*, *Nipponia nippon*; PDD 1.0); Luc Viatour (adapted *Columba livia domestica*; CCA-SA 3.0 unported); Sarah Werning (*Anolis carolinensis*; CCA 3.0 unported); Emily Willoughby (*Sayornis phoebe*; CCA-SA 3.0 unported); Elaine R. Wilson (*Arenaria interpres*; CCA-SA 3.0 unported); and Lip Kee Yap (*Phaenicophaeus curvirostris*; CCA-SA 3.0 unported).

Table 1

Peptides Recovered in This Study^a

	PEAKS UniProt $-\log_{10} p$	PEAKS NCBI-Archo $-\log_{10} p$	Byonic Search 1 Log Prob	Byonic Search 2 Log Prob	Byonic Wildcard Log Prob	Spectra
collagen I, alpha 1	70.98	69.49	23.14	20.73	21.04	13
GATGAP _(OH) GIAGAP _(OH) GFP _(OH) GAR ^b	26.77	29.84	6.41	6.94	6.83	5
GFP _(OH) GADGIAGP _(OH) K	N/A	24.91	3.80	3.09	3.82	1
GFP _(OH) GLPCP _(OH) SGEPGK	N/A	N/A	3.31	2.99	4.02	1
GQAGVMGFP _(OH) GPK	39.41	36.72	7.38	6.39	6.36	2
GSAGPP _(OH) GATGFP _(OH) GAAGR ^b	37.59	33.18	7.41	5.36	6.30	4
collagen I, alpha 2	31.12	56.98	9.68	11.16	12.20	5
EGPVGFP _(OH) GADGR	N/A	35.56	6.50	6.63	6.68	2
GATGLP _(OH) GVAGAP _(OH) GLP _(OH) GPR	N/A	24.25	N/A	N/A	2.69	2
GEP _(OH) GNIGFP _(OH) GPK	31.12	26.67	7.10	7.32	7.35	1

^aPeptide scores (PEAKS) and log probability scores (Byonic) are given for each protein and peptide as well as the number of PSMs detected.^bIndicates previously detected peptides.²⁸