Available online at www.sciencedirect.com
ScienceDirect
Geochimica et Cosmochimica Acta 71 (2007) 5956–5967
Geochimica et Cosmochimica Acta
www.elsevier.com/locate/gca

Investigation of the protein osteocalcin of *Camelops hesternus*: Sequence, structure and phylogenetic implications

James F. Humpula [a], Peggy H. Ostrom [a,*], Hasand Gandhi [a], John R. Strahler [b], Angela K. Walker [b], Thomas W. Stafford Jr. [c], James J. Smith [d], Michael R. Voorhies [e], R. George Corner [e], Phillip C. Andrews [b]

[a] *Department of Zoology, Michigan State University, East Lansing, MI 48824, USA*
[b] *Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA*
[c] *Stafford Research Labs, 200 Acadia Avenue, Lafayette, CO 80026, USA*
[d] *Department of Entomology and Lymann Briggs College, Michigan State University, East Lansing, MI 48824, USA*
[e] *University of Nebraska State Museum, Lincoln, NE 68588, USA*

Received 14 June 2007; accepted in revised form 10 September 2007; available online 22 September 2007

Abstract

Ancient DNA sequences offer an extraordinary opportunity to unravel the evolutionary history of ancient organisms. Protein sequences offer another reservoir of genetic information that has recently become tractable through the application of mass spectrometric techniques. The extent to which ancient protein sequences resolve phylogenetic relationships, however, has not been explored. We determined the osteocalcin amino acid sequence from the bone of an extinct Camelid (21 ka, *Camelops hesternus)* excavated from Isleta Cave, New Mexico and three bones of extant camelids: bactrian camel (*Camelus bactrianus)*; dromedary camel (*Camelus dromedarius*) and guanaco (*Llama guanacoe*) for a diagenetic and phylogenetic assessment. There was no difference in sequence among the four taxa. Structural attributes observed in both modern and ancient osteocalcin include a post-translation modification, Hyp_9, deamidation of Gln_{35} and Gln_{39}, and oxidation of Met_{36}. Carbamylation of the N-terminus in ancient osteocalcin may result in blockage and explain previous difficulties in sequencing ancient proteins via Edman degradation.

A phylogenetic analysis using osteocalcin sequences of 25 vertebrate taxa was conducted to explore osteocalcin protein evolution and the utility of osteocalcin sequences for delineating phylogenetic relationships. The maximum likelihood tree closely reflected generally recognized taxonomic relationships. For example, maximum likelihood analysis recovered rodents, birds and, within hominins, the *Homo-Pan-Gorilla* trichotomy. Within Artiodactyla, character state analysis showed that a substitution of Pro_4 for His_4 defines the *Capra-Ovis* clade within Artiodactyla. Homoplasy in our analysis indicated that osteocalcin evolution is not a perfect indicator of species evolution. Limited sequence availability prevented assigning functional significance to sequence changes. Our preliminary analysis of osteocalcin evolution represents an initial step towards a complete character analysis aimed at determining the evolutionary history of this functionally significant protein. We emphasize that ancient protein sequencing and phylogenetic analyses using amino acid sequences must pay close attention to post-translational modifications, amino acid substitutions due to diagenetic alteration and the impacts of isobaric amino acids on mass shifts and sequence alignments.

1. INTRODUCTION

Due to their taxonomic specificity and ability to address evolutionary questions, nucleic acids could be considered the ultimate biomarker. Yet, relative to other biomolecules,

* Corresponding author. Fax: +1 517 432 2789.
E-mail address: Ostrom@msu.edu (P.H. Ostrom).

0016-7037/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.gca.2007.09.003

DNA is not stable, suggesting temporal limits on our window into the past (Collins et al., 2000). Pyrosequencing techniques, aimed at identifying short nucleotide fragments, assist us in emerging from this time trap (Poinar et al., 2005). In addition, mineral associated proteins, arguably more stable than DNA, may also extend the fossil genetic record (Collins et al., 2000). In the past 5 years, literature on ancient protein sequences has increased with the first complete sequence for the bone protein osteocalcin for a >55 ka fossil reported in 2002 (Nielsen-Marsh et al., 2002) and sequences of collagen peptides from a 68 Ma dinosaur presented in 2007 (Asara et al., 2007). An important next step is to implement protein sequencing in a phylogenetic analysis to enhance our understanding of evolutionary relationships of extinct taxa. Consequently, we examined the ability of osteocalcin to provide phylogenetic information for an extinct camelid and modern members of the family Camilidae.

The family Camelidae (order Artiodactyla) originated during the early Tertiary of North America. Camelids are characterized by a lack of horns, long slender necks, and long limbs with a much-reduced ulna and fibula (Kurtén and Anderson, 1980). Most camelid evolution occurred in North America, with both camels and llamas being common during the Pleistocene (Kurtén and Anderson, 1980; Lange, 2002). North American camelids became extinct at the end of the Pleistocene, ca. 11,000–12,000 years BP. The extant family Camelidae comprises two tribes, Lamini and Camelini, both of which split from a common ancestor approximately 11 Ma ago (Webb, 1974). Lamini and Camelini comprise four New World species and two Old World species, respectively (Webb, 1974).

The genus *Camelops*, contains five separate species that existed in northern North America between 2 million and 11,000–12,000 years ago (Savage, 1951; Stuart, 1991). These species are grouped into two subsets based on size and dental characters. The larger of the two groups includes the extinct species *Camelops hesternus* and *Camelops heurfanensis* (Savage, 1951). *C. hesternus*, or Yesterday's Camel, first appeared 300,000 years ago, and was abundant throughout the western United States, the Yukon, and Alaska (Kurtén and Anderson, 1980). For unknown reasons, it went extinct between 12,600 and 10,800 years ago. Similar in relative proportion to the modern dromedary camel, *Camelus dromedarius*, the legs of *C. hesternus* were approximately 20% longer than *C. dromedarius*. *C. heurfanensis* is very similar to *C. hesternus*, and is differentiated from the other large *Camelops* species by the location of the post-palatine foramen and the arrangement of the placement of the lower border of the mandible from the symphysis to a point below the last molar (Savage, 1951).

Osteocalcin, also known as Bone Gla Protein or BGP, is a 46–50 amino acid long acidic protein whose potential for survival is attributed to its high affinity for hydroxyapatite (Hauschka et al., 1989; Lian et al., 1989; Cancela et al., 1995). Osteocalcin's conserved region contains a single disulfide bridge connecting amino acids 23 and 29 and, in all but one species, has three γ-carboxyglutamic acid (Gla) residues at positions 17, 21, and 24 (Hauschka et al., 1989). The marine fish *Argyrosomus regius* also contains Gla_{25} (Frazao et al., 2005). The Gla residues and the disulfide bridge are key factors in the stabilization of osteocalcin and in maintaining its association with the hydroxyapatite mineral phase (Hoang et al., 2003).

Evolutionary comparisons between osteocalcin sequences have been explored, but very little work has been done to determine osteocalcin's taxonomic level of resolution (Nielsen-Marsh et al., 2002; Laizé et al., 2005). Furthermore, the strength of the hypothesized phylogenetic relationships derived from osteocalcin sequences is unknown. To address these issues, a phylogenetic analysis of derived osteocalcin sequences should be compared to a robust, well-supported DNA-based phylogeny. Zardoya and Meyer (1996) determined the level of performance of all mitochondrial protein-coding regions in returning two expected topologies. Cytochrome *b* was one of the highest performing mtDNA-coding sequences found, and therefore appropriate for evaluating the phylogenetic efficacy of using osteocalcin protein sequences.

The osteocalcin sequence for *C. hesternus*, modern camels (*Camelus bactrianus* and *C. dromedarius*) and guanaco (*Llama guanacoe*) were determined by MALDI-MS (matrix-assisted laser desorption ionization mass spectrometry). The evolution of osteocalcin within the Order Artiodactyla was investigated and a phylogenetic analysis of osteocalcin based on 24 species of vertebrates was performed. Comparisons between the phylogenies based on osteocalcin sequence data and phylogenies based on Nd4, Nd5, and Cytochrome *b* allow an assessment of the degree of phylogenetic resolution and examination of specific phylogenetic relationships exhibited in the osteocalcin tree.

2. MATERIALS AND METHODS

2.1. Sample description

A metapodial from a *Camelus bactrianus* (modern bactrian camel) (MSU-9669) and the right femur of *Llama guanacoe* (UM 157201) were obtained from Michigan State University Museum and University of Michigan, respectively. A vertebral fragment of dromedary camel (*Camelus dromedarius*) (AMNH 10734) was obtained from the American Museum of Natural History. The University of Nebraska State Museum provided a phalanx of *Camelops hesternus* from Isleta Cave, New Mexico, which was AMS ^{14}C dated as $21{,}190 \pm 110$ RC years (CAMS-22182) by Stafford Research Laboratories. The Isleta Cave site consists of two caves located within a Quaternary lava flow approximately 13 km west of Isleta, New Mexico, at an elevation of 1716 m. The fossil assemblage within the caves consists of 42 mammal genera and 6 reptile genera (Harris and Findley, 1964).

2.2. Purification/digestion

The mechanically cleaned and powdered sample (20–50 mg; −190 °C, CertiPrep 6750 SPEX freezer/mill) was demineralized with sodium EDTA (0.5 M; pH 8.0; 4 h at 25 °C), centrifuged (14,000 rpm, 10 min) and the resulting supernatant applied to a fresh C_{18} gravity column (60 Å,

Fisher, 0.5 × 2 cm). The column was eluted with eight, 1 mL aliquots of a mixture of solvent A (94.9% water, 5% acetonitrile, 0.1% TFA) and solvent B (89.9% acetonitrile, 10% water, 0.1% TFA), with increasing concentration of solvent B (20%, 25%, 30%, 32%, 34%, 36%, 38%, and 40%). The eluent from each of the eight fractions was collected, concentrated (Speedvac, Eppendorff), and reconstituted using 10 μL of 1% *n*-octyl-β-D-glucopyranoside (OGP, Sigma) in 50 mM Tris–HCl, pH 8.0. The reconstituted eluent was diluted 1:10 with solvent A and 0.5 μL was spotted on a MALDI target, with 0.5 μL of a 1% solution of trifluoroacetic acid (TFA, Sigma) and 0.5 μL of a saturated solution of CHCA (α-cyano-4-hydroxycinnamic acid) matrix in 1:1 acetonitrile and 0.1% TFA. MALDI-MS spectra were acquired with an ABI-STR operated in linear mode. Spectra were used to identify gravity column fractions that contained a peak consistent with the m/z of osteocalcin (~5.7 kDa) (Hauschka et al., 1989).

Osteocalcin was purified using reversed phase high performance liquid chromatography (rpHPLC) fitted with a C_{18} trap (1 × 10 mm, Michrom BioResources) and a 1 × 150 mm C_{18} column (300 Å, 5 μm, Reliasil, Michrom BioResources) (Nielsen-Marsh et al., 2002). The C_{18} trap was equilibrated with 20% solvent B, the sample injected, the trap washed with 300 μL of solvent A, and the flow from the trap diverted to the column. Osteocalcin was eluted with a gradient of 20% solvent B, increased to 30% solvent B over 15 min, held at 30% solvent B for 5 min, increased to 35% solvent B over 15 min, held at 35% solvent B for 10 min, increased to 95% solvent B over 5 min, and held for 5 min. All peaks were collected and MALDI-MS was used to identify putative osteocalcin.

Peaks from the rpHPLC containing putative osteocalcin were dried, reconstituted with 10 μL of 1% OGP in 50 mM Tris–HCl, pH 8.0, and digested with Trypsin (Promega) or ASP-N (Sigma). The digest products were purified using rpHPLC (gradient: 5% solvent B, increased to 40% solvent B over 40 min, held at 40% solvent B for 10 min, increased to 95% solvent B over 1 min, held at 95% solvent B for 5 min, and decreased to 5% solvent B over 1 min) and analyzed using MALDI-MS.

Peptide disulfide bonds were reduced with tricarboxyethyl phosphine (TCEP, 5 mM; 1 h at 60 °C) and cysteines blocked with methyl methane thiosulfonate (MMTS,10 mM; 10 min, room temperature). Peptide hydroxyl groups were formylated by incubation with 98% formic acid at room temperature for 24 h. Oxidized methionine was reduced with ethanedithiol and bromotrimethylsilane.

2.3. MALDI-MS and Edman sequencing

MALDI-MS was performed on intact osteocalcin and digest products using an ABI-STR (linear mode) and an ABI-4800 TOFTOF (reflectron mode). Tandem mass spectrometry (MS/MS) was performed on digest products using the ABI-4800 TOFTOF (ABI/MDS Sciex). Ions from the ABI-STR are average m/z and reported as nominal mass while those from the ABI-4800 TOFTOF are monoisotopic and reported to one decimal place. Spectra of intact osteocalcin obtained on the ABI-STR were externally calibrated using the 1+ and 2+ peaks of bovine insulin (average mass 5734.6 and 2867.8, respectively; Sigma). Resolution was ca. 530 and mass accuracy at m/z of 2800 was better than 353 ppm (range 4–353 ppm).

MALDI-MS spectra obtained on the ABI-4800 TOF-TOF were acquired using 1500–3000 laser shots. Spectra of intact osteocalcin were externally calibrated using the $[M+H]^{+1}$ and $[M+2H]^{+2}$ of bovine insulin (monoisotopic mass 5730.6 and 2865.8, respectively, Sigma) and its β chain (monoisotopic mass 3495.9). Resolution was ca. 13,000 and mass accuracy was better than 34 ppm (range 9–34 ppm). MS/MS spectra were obtained using 2000–8000 laser shots, and were calibrated using fragmentation of angiotensin II (human). External calibration of MS/MS spectra was done using the Mass Standards Kit for the 4700 Proteomics Analyzer (ABI/MDS Sciex). Atmosphere was used as the collision gas (pressure: 6×10^{-7} torr, collision energy: 1 kV). The raw data for mass spectra that appear as figures in the text or in electronic annexes are in Electronic Annex 1 (EA-1) and are also housed at http://www.proteomecommons.org/data.jsp. These may be viewed using MS Expedite, also contained in EA-1. The observed and expected masses to charge ratio (m/z) of y and b ions in each of the MS/MS spectra presented as figures or discussed in this manuscript are provided in EA-2.

Peptide sequences were determined by *de novo* sequencing using MS Expedite (http://www.proteomecommons.org/current/530/MSViewerApp.jnlp). Fragment ion masses for putative peptide sequences were generated using MS-Product (prospector.ucsf.edu). MS/MS spectra were searched using Mascot (v2.1 MatrixScience, London, UK) against a database consisting of NCBInr (Feb 22, 2007) and the putative camel osteocalcin sequence. Edman sequencing was performed on an Applied Biosystems 494 CLc.

2.4. Phylogenetic analysis

The data set of osteocalcin sequences analyzed in this study is composed of 25 complete protein sequences. Sequences were aligned with ClustalX (Thompson et al., 1997) using the Gonnet series matrix. For the pair-wise parameters, the gap opening penalty was set to 35.00 and the gap extension penalty was set to 0.75. For the alignment parameters, the gap opening penalty was set to 15.00 and the gap extension penalty was set to 0.30. The alignment was unaffected by changing the parameters and needed no further modification.

The aligned osteocalcin sequences were analyzed by maximum likelihood analysis, neighbor joining, and maximum parsimony, with fish *Takifugu rubripes* and *Tetradon nigroviridis* sequences as outgroups. Maximum likelihood analyses were performed using Tree-Puzzle 5.0 (Schmidt et al., 2002), which employs a quartet puzzling algorithm to search the tree space. Puzzling steps (10,000) were performed with parameter estimates set to "exact" using the JTT (Jones et al., 1992) model of amino acid substitution.

Neighbor joining and maximum parsimony protein sequences were performed using PAUP* 4.0b10 (Swofford, 2002). For neighbor joining analysis, uncorrected

p-distances were used, with ties between trees broken randomly. Neighbor joining bootstrap analysis (1000 replicates) was carried out using the default settings. The maximum parsimony analysis was done using a heuristic search employing TBR branch swapping, with uninformative characters excluded and the initial tree found using random taxon addition. Maximum parsimony bootstrapping for osteocalcin employed 100 replicates (due to time constraints).

The mtDNA data set consists of 24 complete sequences from the Cytochrome *b* protein-coding region of mtDNA. Taxa used to create the mtDNA and osteocalcin tree were the same with the exception of *Setonix*, which does not yet have a complete Cytochrome *b* genome sequence (Table 1). Sequences were aligned individually using ClustalX with default settings then manually corrected in MacClade 4.06 (Maddison and Maddison, 2000). The aligned sequences were then analyzed by maximum likelihood, maximum parsimony, and neighbor joining, again using methods with *T. rubripes* and *T. nigroviridis* as outgroups.

For the Cytochrome *b* analyses, neighbor joining, maximum parsimony, and maximum likelihood were carried out using PAUP* 4.0b10 (Swofford, 2002). Modeltest 3.7 (Posada and Crandall, 1998) was used to determine the appropriate DNA substitution model (GTR+I+G) for the maximum likelihood analysis. The maximum likelihood tree was found with a heuristic search using TBR branch swapping. Maximum likelihood bootstrapping was performed using NNI branch swapping and starting with neighbor joining trees. For the neighbor joining analysis maximum likelihood distances were used, and bootstrapping was accomplished with 1000 replicates. The maximum parsimony analysis consisted of a heuristic search with TBR branch swapping, uninformative characters excluded, and the initial tree found using random addition. Parsimony bootstrapping was accomplished using 1000 replicates.

In both data sets, the maximum likelihood tree best represents the accepted tree topology (ncbi.nlm.nih.gov). The maximum parsimony bootstrap and neighbor joining bootstrap are included as support.

Character mapping was accomplished by mapping the amino acid character states of osteocalcin residues 4, 5, 19, 48, and 49 onto the Cytochrome *b* maximum likelihood tree, pruned to remove all taxa except the artiodactyls and a chosen outgroup, *Equus caballus*, and modified to resolve the sister relationship of *Sus* and *Camelus* according to Matthee et al. (2001). Resolution of ambiguous changes in the tree was achieved using the DELTRAN option in MacClade, which delays evolutionary transformations as long as possible within the tree. For each of these five amino acid residues, character consistency indices (CI) and retention indices (RI) were calculated using MacClade.

3. RESULTS

3.1. Osteocalcin sequencing

Gravity column fractions of *C. hesternus* show $[M+H]^+$ ions at m/z 5625 and 5668 that are within the range of osteocalcin from modern vertebrates ($M_r = 5210–5889$) (Hauschka et al., 1989). The 43 Da difference between these two analytes is consistent with carbamylation (see below) as previously observed in fossil osteocalcin (Ostrom et al.,

Table 1
Sequence reference numbers for osteocalcin and Cytochrome *b* for all taxa in this study

Scientific name	Common name	Osteocalcin reference number	Cytochrome *b* Reference number
Bison bison	Bison	P83498	AY689186
Bos taurus	Cow	NP_776674	NC_006853
Camelus dromedarius	Dromedary Camel	Sequence included	X56281
Canis familiaris	Dog	P81455	NC_002008
Capra hircus	Goat	POC225	NC_005044
Dromaius novaehollandiae	Emu	P15504	NC_002784
Equus caballus	Horse	Ostrom et al. (2006)	NC_001640
Felis catus	House cat	P02821	NC_001700
Gallus gallus	Chicken	NP_990718	NC_001323
Gorilla gorilla	Gorilla	P84349	NC_001645
Homo sapiens	Human	NP_954642	AC_000021
Macaca sp.	Macaque	P02819	AJ_309865
Mus musculus	House mouse	AAA29856	NC_005089
Oryctolagus cuniculus	European rabbit	P39056	NC_001913
Ovis aries	Sheep	ABD83814	NC_001941
Pan troglodytes	Chimpanzee	P84348	NC_001643
Pongo pygmaeus	Orangutan	Q5RDP6	NC_001646
Rana sp.	Bull frog	BAD16774	AF205094
Rattus norvegicus	European rat	NP_038200	NC_001665
Setonix sp.	Quokka	1005180C	
Sus scrofa	Pig	AAN73020	NC_000845
Takifugu rubripes	Fugu puffer	AAO24898	NC_004299
Tetraodon nigroviridis	Spotted green puffer	AA024897	NC_007176
Xenopus laevis	African clawed frog	AAB36024	NC_001573
Xenopus tropicalis	Western clawed frog	NP_001006689	NC-006839

2006). Because no sequences exist for osteocalcin from extant camelids, interpretations of mass spectra were made by comparison to bison osteocalcin (Nielsen-Marsh et al., 2005) (bison, YLDHGLGAuA PYPDPLEPKR EVCELNPDCD ELADHIGFQE AYRRFYGPV: with decarboxylation of Gla_{17}, Gla_{21}, Gla_{24}; Hyp_9; disulfide bond between Cys_{23}, and Cys_{29}).

We hypothesized that the tryptic digest of the *m/z* 5668 analyte from *C. hesternus* contained peptides representing residues 1–19 (*m/z* 2067.7), 20–43 (*m/z* 2846.2), and 44–49 (*m/z* 744.4) (Fig. 1a) of osteocalcin and that the predominant peak in the ASP-N digest was 34–49 (*m/z* 1987.7) of osteocalcin. The peak at *m/z* 2110.9, 43 Da greater than putative 1–19, likely represents 1–19 modified by carbamylation (Fig. 1a, 1–19a). MS/MS results resolved the location of carbamylation and sequence information for each of the digest products.

The MS/MS of putative 1–19a produces a fragmentation pattern that identifies residues 1–19 with a substitution of Val_{13} for Pro_{13} relative to bison and carbamylation of Tyr_1 (Fig. 2a). Carbamylation of Tyr_1 (Tyr = 163.1) is supported by a difference of 205.8 Da between y_{18} (*m/z* 1905.1) and the parent ion (*m/z* 2110.9). A series of peaks 43 Da greater than the *m/z* expected for b ion fragments, denoted b^*, are consistent with N-terminal carbamylation. Several fragment ions (b_8, b_{10}, and b_{14}) are consistent with loss of an N-terminal carbamyl group and cleavage on the N-terminal side of proline and hydroxyproline. The difference in *m/z* between y_3 and y_5 (210.1) is consistent with Pro_{15} and Leu_{16}, as in bison.

The peptide mass fingerprint (PMF) of *C. hesternus* showed two peaks differing by 17 Da (*m/z* 2829.2 and 2846.2) that were candidates for 20–43 (Fig. 1a). MS/MS data for these peaks (1) allows assignment of residues 20–22 and 30–43 of these peptides with substitution of Gln_{35} for His_{35} and Met_{36} for Ile_{36} relative to bison and (2) explains the 17 Da difference between the two (Fig. 2b; Electronic Annex EA-3). MS/MS of peptide *m/z* 2846.2 shows a characteristic neutral loss of 64 from the parent ion in numerous y ions (denoted y'_n, e.g. y'_8, y'_9, y'_{12}, y'_{13}) indicating methionine sulfoxide (Metox) in the peptide (Fig. 2b). Treatment of the peptide fraction containing *m/z* 2846.2 and 2829.2 to reduce oxidized methionine resulted in a 16 Da decrease of the peak at *m/z* 2846.2. MS/MS of *m/z* 2829.2 shows pairs of fragment ions 1 Da apart, suggesting partial deamidation at two positions: Gln_{35} and Gln_{39} (EA-3). Taken together the spectra indicate that the 17 Da difference between the peptides results from oxidation of Met_{36}(+16) and deamidation (+1) of either Gln_{35} or Gln_{39} in peptide 2846.2. A series of fragment ions in peptide *m/z* 2846.2 (putative b_{10}, b_{11}, b_{12}, b_{14}, b_{15}, b_{16}, b_{20}, b_{21}, b_{23}, and y_7 to y_{13}) is consistent with this interpretation (Fig. 2b).

The tryptic peptide *m/z* 744.4 (putative 44–49) allows assignments of residues 44–49 with substitutions of Thr_{48} and Thr_{49} for Pro_{48} and Val_{49} of bison (Fig. 2c). The MS/MS data for ASP-N digest product *m/z* 1987.7 (putative 34–49) confirms assignments for residues 34–44 and residue 49 with $Metox_{36}$ and deamidation of Gln_{35} and Gln_{39} (Fig. 2d). Deamidation of Gln_{39} was identified by a 1 Da increase in the *m/z* of y_{11}, y_{12}, and y_{13} relative to that predicted. Deamidation of Gln_{39} and oxidation of Met_{36} is identified by a 17 Da increase in the *m/z* of y_{14} relative to that expected and a characteristic neutral loss of 64 Da in y_{14} and y_{15}. $Metox_{36}$ and deamidation of Gln_{35} and Gln_{39} is indicated by a 18 Da increase in the *m/z* of y_{15} relative to that expected.

Based on comparison to bison, our data for *C. hesternus* defines residues 1–22 and 30–49 (YLDHGLGAuAPYVDPLEPKREV_____ DELADQMGFQEAYRRFYGTT) (with decarboxylation of Gla_{17}, Gla_{21}; Hyp_9). Owing to limited availability of the fossil, the remaining ambiguity was resolved by sequencing three modern camelids, *C. bacterianus*, *C. dromedarius*, and *L. guanacoe* by PMF, tandem mass spectrometry and partial Edman sequencing. This also allowed a taxonomic comparison between the modern and ancient taxa. MALDI-MS sequence determinations of modern taxa were based on a comparison to *C. hesternus* and bison osteocalcin.

A partial Edman sequence for ASP-N peptide 34–49 of *C. bacterianus* identified Asp_{34}, Gln_{35}, Met_{36}, Phe_{37}, and Gln_{38}. The *m/z* of 44–49 (744.4) is the same for *C. hesternus*, *C. bactrianus*, and *L. guanacoe* (Fig. 1). The *m/z* of 1–19 of osteocalcin from the modern taxa (*C. dromedarius* *m/z* 2068.7; *C. bactrianus* *m/z* 2068.4; *L. guanacoe* *m/z* 2068.4) are 43 Da lower than 1–19a (*m/z* 2110.9) of *C. hesternus*, a difference consistent with carbamylation; a modification that appears to occur during fossilization (Fig. 1a–d). The PMF showed two candidate peaks for 20–43 that differ by 16 Da in *C. bactrianus* and *L. guanacoe* (Fig. 1c and d; *m/z* 2829.7 and 2845.7) or 17 Da in *C. dromedarius* (Fig. 1b; *m/z* 2829.8 and 2846.7) and *C. hesternus* (Fig. 1a; *m/z* 2829.2 and 2846.2). The *m/z* of ASP-N peptide 34–49 of *C. bacterianus* and *L. guanacoe* are the same (*m/z* 1970.2) but that of *C. dromedarius* (*m/z* 1986.4) and *C. hesternus* (*m/z* 1987.7) are 16 and 18 Da higher, respectively. Mass differences observed in peptides 20–43 and 34–49 are explained through investigation of MS/MS data (below).

Complete sequence coverage for tryptic peptide 1–19 of *C. dromedarius* is provided by MS/MS (Fig. 3a). Amino acids Leu, Ile, and hydroxyproline are isobaric. This ambiguity at residue 9 was resolved by formylation of peptide 1–19 which resulted in a 28 Da increase in y and b ions that contain residue 9 consistent with formation of formate ester on hydroxyproline at residue 9. As with *C. hesternus*, Val_{13} is observed in *C. dromedarius* and the difference in *m/z* between b_{14} and b_{16} identifies Pro_{15} and Leu_{16}. The MS/MS of 1–19 of osteocalcin from *C. bacterianus* identifies residues 1–19 and from *L. guanacoe* identifies residues 3–10, 13–19 (Electronic Annexes EA-4 and EA-5) with Val_{13} in both taxa. These data suggest that the osteocalcin sequence for 1–19 from all taxa is identical.

The PMF of *C. dromedarius* showed two peaks differing by 17 Da that were candidates for 20–43 (Fig. 1b) at *m/z* 2829.8 and *m/z* 2846.7, respectively. Complete sequence coverage of peptide *m/z* 2846.7 was obtained following reduction and cysteine blockage with MMTS (Fig. 3b). We define Pro_{27} and Asp_{28} because it is the only dipeptide pair that exists in other organisms at positions 27 and 28 and its mass (212.1) is consistent with the 212.1 Da differ-

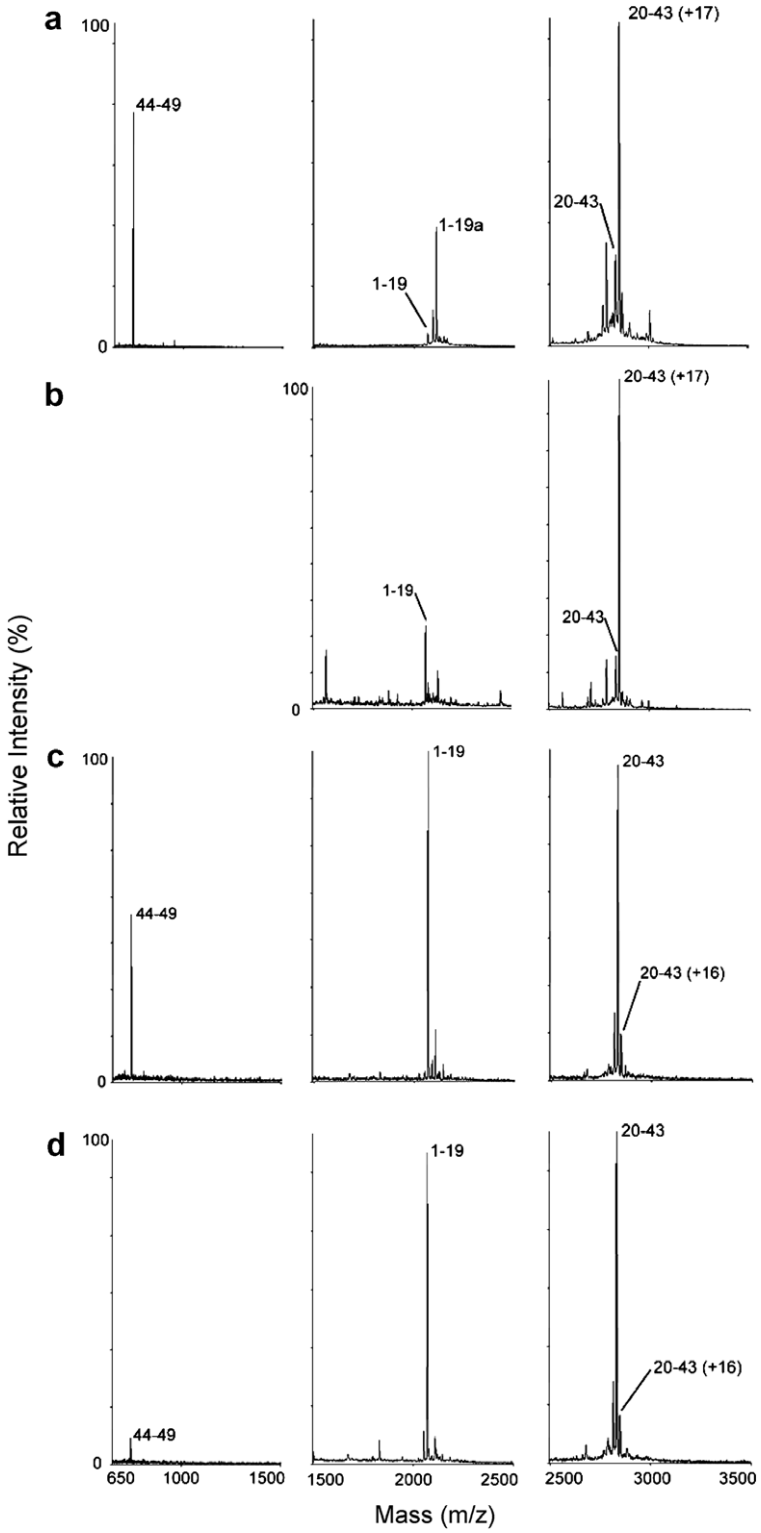

Fig. 1. Peptide mass fingerprint for osteocalcin tryptic digest from: (a) 21 ka *Camelops hesternus*; (b) *Camelus dromedarius*; (c) *Camelus bactrianus*; and (d) *Llama guanacoe* produced from an ABI-4800 TOFTOF (confirmed on an ABI-STR). Peptide 1–19a, 43 Da greater than 1–19, is osteocalcin modified by carbamylation. Ions denoted (+16) have methionine sulfoxide in position 36 ($Metox_{36}$). Ions denoted (+17) have $Metox_{36}$ and deamidation of Gln_{35}.

ence between b_7 and b_9. Dipeptide $Leu_{32}Ala_{33}$ is inferred by comparison to *C. bactrianus* and *L. guanacoe* whose y ions identify these residues (EA-4 and EA-5). The 147.0 Da difference in the m/z of y_7 and y_8 and a characteristic neutral loss of 64 Da in numerous ions denoted y'_n (e.g. y'_8, y'_9, y'_{12}, y'_{13}) is consistent with Metox in position 36.

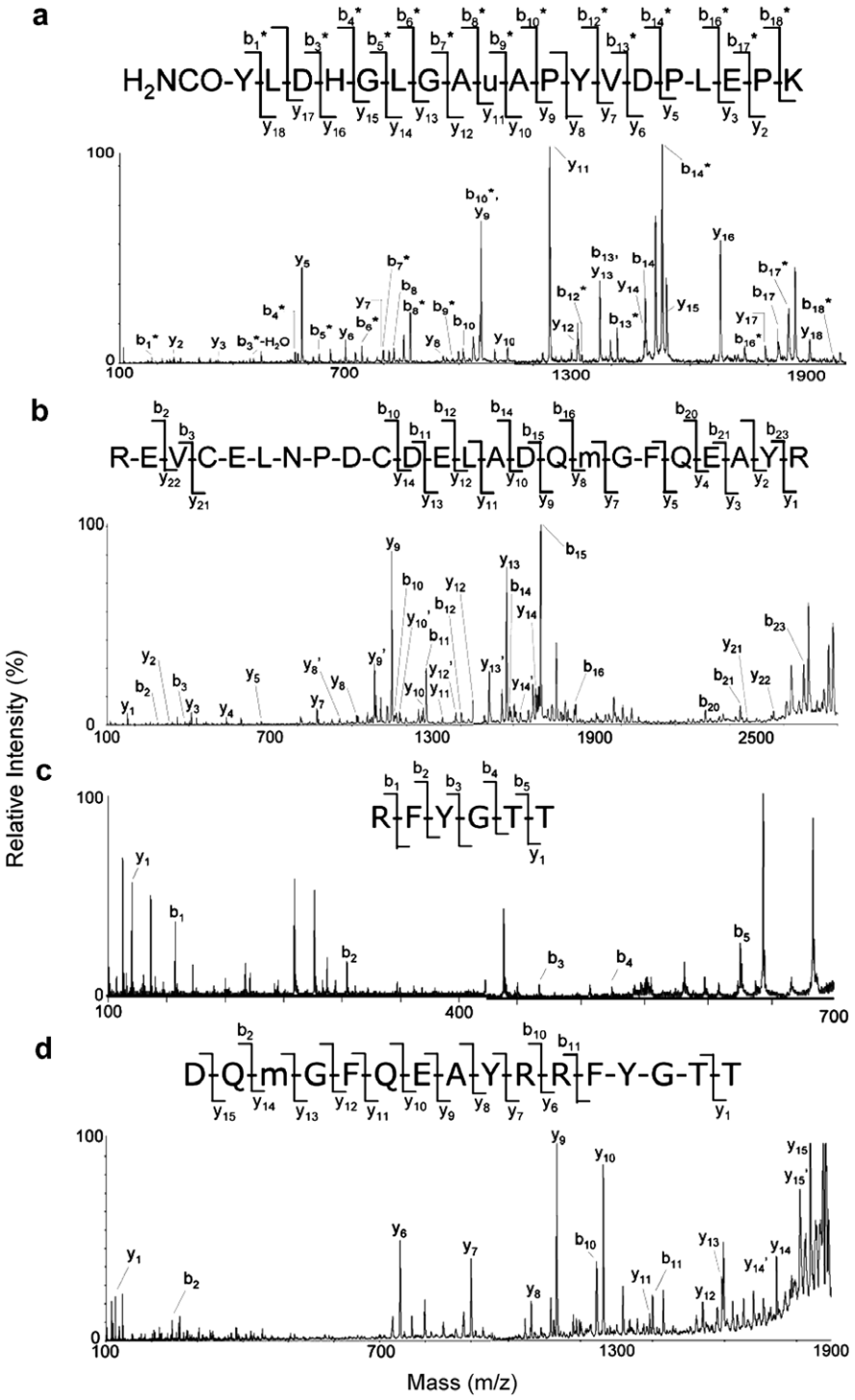

Fig. 2. MS/MS spectra for tryptic peptides (a) 1–19, (b) 20–43 (+17), and (c) 44–49 and ASP-N peptide (d) 34–49 of *Camelops hesternus* produced on an ABI-4800 TOFTOF. Sequence coverage after reference to modern camelids is 1–22, 30–49. Carbamylated b ions are identified as b*. Hydroxyproline is ''u'' and methionine sulfoxide is ''m''. Ions containing methionine sulfoxide with a characteristic neutral 64 Da loss are denoted y'_n. Panel (a) is scaled relative to the largest ion in the spectrum (b_{14}*). Panel (d) contains deamidation at Gln_{35} and Gln_{39}. Observed and expected *m/z* for ions from these MS/MS analysis are presented in Electronic Annex EA-2.

Deamidation of Gln_{35} is indicated by a 1 Da increase in y ions containing this residue (e.g. y_8, y_9, y_{12}). The MS/MS data indicate that the 17 Da difference between the two peaks in the PMF for 20–43 in *C. dromedarius* represent one with Met_{36} and another containing $Metox_{36}$ and deamidation of Gln_{35}.

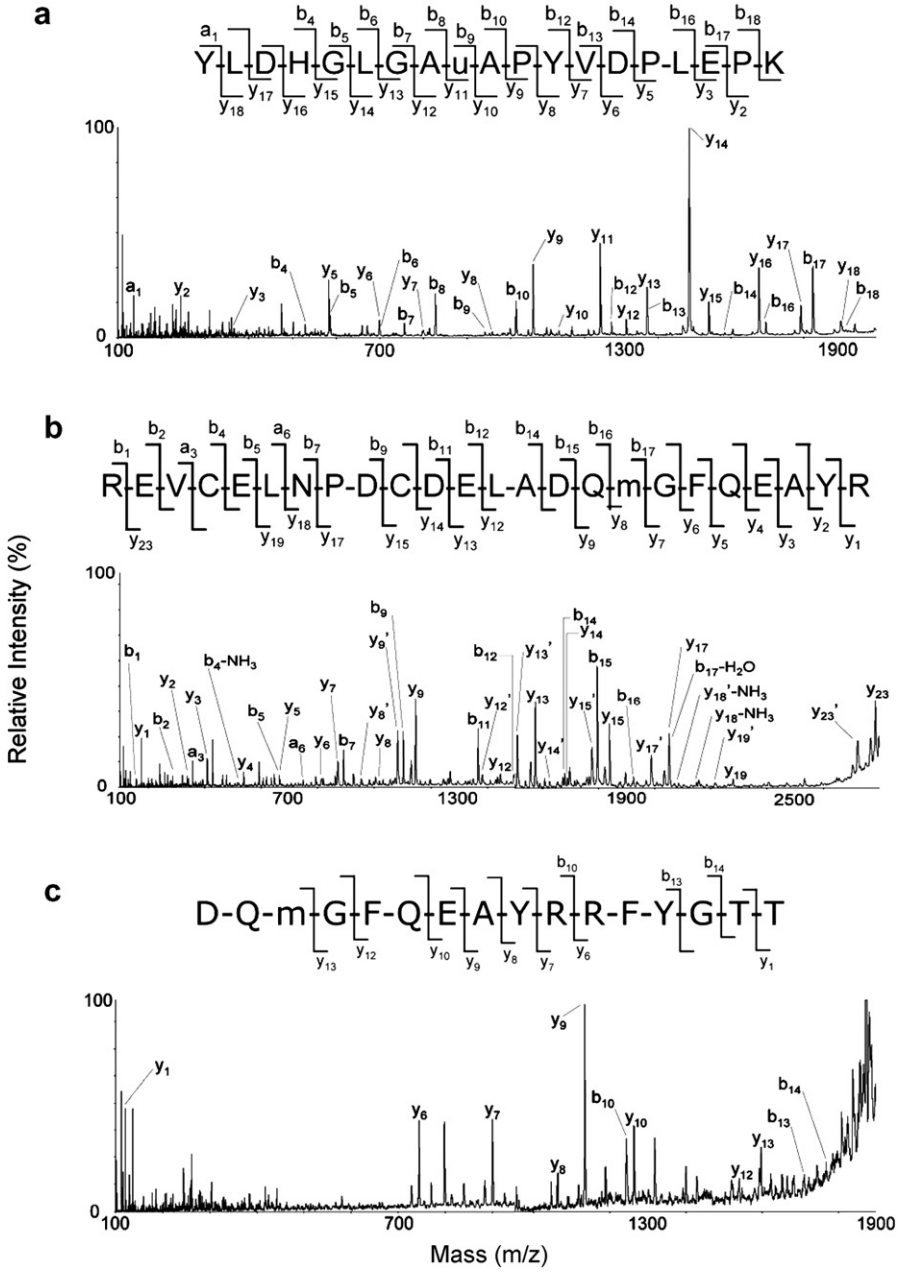

Fig. 3. MS/MS spectra for tryptic peptides (a) 1–19 and (b) MMTS-derivatized 20–43, and (c) ASP-N peptide 34–49 of *Camelus dromedaries* produced on an ABI-4800 TOFTOF. By inference to *Camelus bactrianus* and *Llama guanacoe*, there is complete sequence coverage as detailed in the text. Hydroxyproline is denoted ''u'' and methionine sulfoxide is denoted ''m''. Ions containing methionine sulfoxide with a characteristic neutral 64 Da loss are denoted y'_n. Panel (d) contains deamidation at Gln_{35}. Observed and expected m/z for ions from these MS/MS analysis are presented in Electronic Annex EA-2.

In *C. bactrianus*, the MS/MS of tryptic peptide 20–43 provides coverage for 20–24 and 29–43 and corroborates Met_{36} identified by Edman sequencing (EA-4). The MS/MS of 20–43 of *L. guanacoe* provides coverage for 20–24, 29–34, and 37–42 (EA-5). The 20–43 (+16) peptide observed in the PMF of both *C. bactrianus* and *L. guanacoe* is 16 Da higher than the expected mass and likely contains $Metox_{36}$ (Fig. 1c and d). The MS/MS and PMF data suggest that the sequences for 20–43 from all taxa are identical with modifications in position 35 (Gln_{35} or Glu_{35}), position 39 (Gln_{39} or Glu_{39}), and 36 (Met_{36} or $Metox_{36}$).

The MS/MS of ASP-N peptide 34–49 from osteocalcin of *C. dromedarius* confirms residues 37–43 and 47–49 (Fig. 3c). A 16 Da increase in the observed m/z of a and b ions relative to that expected suggests $Metox_{36}$. Complete sequence coverage for peptide 34–49 was obtained for *C. bactrianus* and *L. guanacoe* with Met_{36} in both taxa. MS/MS data for peptides 34–49 and 44–49 of *C. bactrianus* and *L. guanacoe* (EA-4 and EA-5) suggests that with the

exception of Met_{36} vs. $Metox_{36}$, there is no difference in sequence between residues 34–49 among taxa.

Data from the modern camelids allow us to address remaining sequence ambiguity for *C. hesternus*. Because there is no difference in the sequence of the modern camelids and the *m/z* of y_{21}, y_{22}, and b_{10} of tryptic peptide 20–43 from *C. hesternus* and the modern camelids are the same, we define residues 23–29 in *C. hesternus* as CELNPDC. In summary, with the exception of oxidation of Met_{36} and deamidation of Gln_{35} or Gln_{39}, the MS/MS results and the PMF indicate that the sequence of the modern and ancient camelids is the same.

3.2. Phylogenetic analysis

The maximum likelihood (ML) tree for the 25 osteocalcin amino acid sequences in our data set with maximum parsimony bootstrap and neighbor joining bootstrap included as support is shown in Fig. 4a. The ML tree was chosen for presentation rather than maximum parsimony (MP) or neighbor joining (NJ) trees because the maximum likelihood ML tree more closely reflected generally recognized taxonomic relationships (Colbert et al., 1991; ncbi.nlm.nih.gov). For example, neither the MP analysis nor the NJ tree grouped *Rattus* with *Mus* or *Ovis* with *Bison*, while the ML tree recovered both of these relationships. In addition, in both the MP and NJ trees, the birds rendered the amphibians paraphyletic, while the ML recovered the amphibians as a monophyletic group. The ML also recovered the *Pan-Homo-Gorilla* clade. It is noteworthy that the *Pan-Homo-Gorilla* clade was also recovered by MP and NJ, the MP tree had *Pongo* as sister to this clade and the NJ tree had *Macaca* sister to Pan, Homo, Gorilla, and Pongo, as generally recognized.

For comparative purposes, a mitochondrial Cytochrome *b* tree was constructed using DNA sequences from the same set of taxa used for the OC analysis (n.b., a sequence from the marsupial, *Setonix.* was not available for this analysis). The ML tree (Fig. 4b) is in agreement with most of the generally recognized relationships of the taxa in our data set, with the notable exception of placing *Camelus* as sister to *Sus*. In contrast, the MP and NJ trees contain a polytomy that includes *Equus*, *Camelus*, *Sus*, *Canis*, *Felis*, *Oryctolagus*, and the *Bos-Bison-Capra-Ovis* clade observed in the ML tree. Also different from the ML tree and recognized relationships is placement of the *Pan-Homo-Gorilla* clade as a sister group to *Rattus*, *Mus*, and the polytomy.

4. DISCUSSION

We isolated and sequenced osteocalcin from a 21 Kyr bone from *Camelops hesternus*, as well as from three modern camelid species (*Camelus bactrianus*, *Camelus dromedarius*, and *Llama guanacoe*). Sequence coverage for *C. hesternus* was obtained via PMF and tandem mass spectrometry. This marks the first molecular data for the genus *Camelops*. The data indicate that there is no sequence difference among the ancient and modern camelid species.

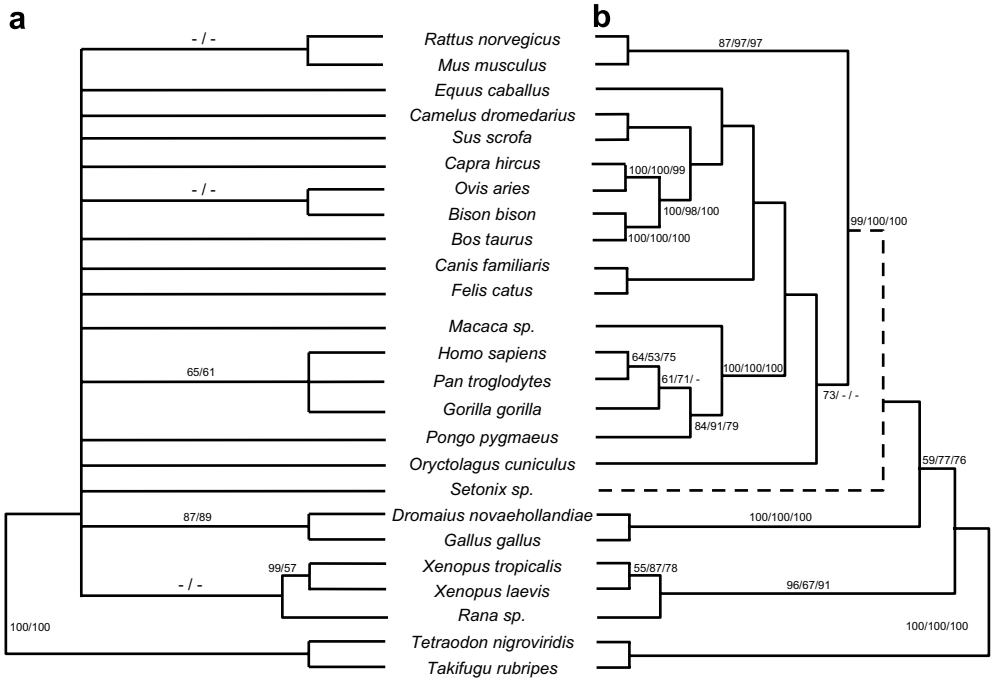

Fig. 4. Maximum likelihood (ML) trees for (a) osteocalcin protein sequence and (b) Cytochrome *b*-coding region of mtDNA. Cytochrome *b* sequence for *Setonix sp.* was not available. Values on branches of trees represent neighbor joining bootstrap/maximum parsimony bootstrap/maximum likelihood values. Cytochrome *b* ML tree created using PAUP* 4.0b10 with model generated using Modeltest 3.7. Osteocalcin ML tree created using Tree-Puzzle 5.0. Maximum parsimony and neighbor joining bootstrap values obtained using PAUP* 4.0b10. Tree length for tree A = 148, CI = 0.655, RI = 0.523. Tree length for tree B = 3267, CI = 0.350, RI = 0.406. Uninformative characters were excluded. Branch support was assessed by bootstrap analysis using both maximum parsimony and neighbor joining methods.

Modifications to the primary structure of osteocalcin were observed in modern and ancient osteocalcin. A modification of Met_{36} to $Metox_{36}$ was observed in *C. hesternus* and *C. dromedarius*. Met, like Cys and Trp, is one of the most easily oxidized amino acids in proteins (Vogt, 1995). This modification is reversed in vivo, and is hypothesized to regulate protein function. Oxidation of methionine may arise during diagenesis over geologic time or as an artifact of sample handling. The presence of EDTA and iron can greatly enhance the oxidation of methionine, especially in an acidic environment. Thus, conclusions about the depositional environment of the *C. hesternus* specimen can not be inferred based on this modification. Despite this, the presence of Met_{36} is notable. The side chain of Met is hydrophobic and its oxidation may influence the hydrophobicity of osteocalcin. Because the function of osteocalcin is unknown (e.g. mediation of hydroxyapatite crystal morphology), we do not know if oxidation of Met_{36} restricts the function of the protein in vivo (Mundy and Poser, 1983; Vogt, 1995; Dowd et al., 2003). However, if Met_{36} is involved in binding of a partner or is involved in the creation of a reactive site its oxidation may result in loss of binding or functionality (Vogt, 1995).

Osteocalcin from *C. hesternus* exhibited deamidation of Gln_{35} and Gln_{39} in ASP-N digest product 34–49. Deamidation of Gln_{35} was also observed in *C. dromedarius* and in osteocalcin from a fossil horse (Ostrom et al., 2006) but not in osteocalcin from other modern camelids. Because of its occurrence in the modern sample, we can not distinguish the role of post-depostional processes from sample handling in deamidation of ancient osteocalcin (Ostrom et al., 2006).

Carbamylation was observed in osteocalcin of *C. hesternus* and previously in osteocalcin of a 42 ka horse but not in modern osteocalcin. In solution, urea and ammonium cyanate (tautomeric with isocyanic acid) are in equilibrium. Carbamylation of an amino group occurs when isocyanic acid reacts with the amino group of a protein ($HNCO + H_2N\text{-}R \rightarrow H_2NCOH_2N\text{-}R$) (Stark, 1965). Urea in dry caves within the southwestern United States is abundant due to its accumulation into amberat by pack rats (Betancourt et al., 1990). Because of its ability to block the N-terminus, carbamylation may be linked to previous difficulties in sequencing ancient proteins. N-terminal blockage is cited as the mechanism responsible for failed attempts to sequence ancient proteins by Edman degradation (Robbins et al., 1993).

Although the phylogenetic relationships of camelids could not be determined using osteocalcin sequences, primarily due to a lack of sequence differences, evolutionary relationships between camelids and other artiodactyls were explored. To do this, we mapped the evolution of specific osteocalcin amino acid residues onto the Cytochrome *b* tree, pruned to contain only artiodactyls (with *Equus* as outgroup) and modified to resolve the sister relationship of *Sus* and *Camelus* (cf. Matthee et al., 2001) (Fig. 5). Five OC protein characters (amino acid residues 4, 5, 19, 48, and 49) are variable within this subset of taxa. These characters are all located in the N- or C-terminal regions of OC, which are considered to be either variable or less structurally constrained than the α-helical regions (Hauschka et al., 1989; Hoang et al., 2003).

Of these five osteocalcin amino acid characters, only character 4 evolves consistently in the Cytochrome *b*

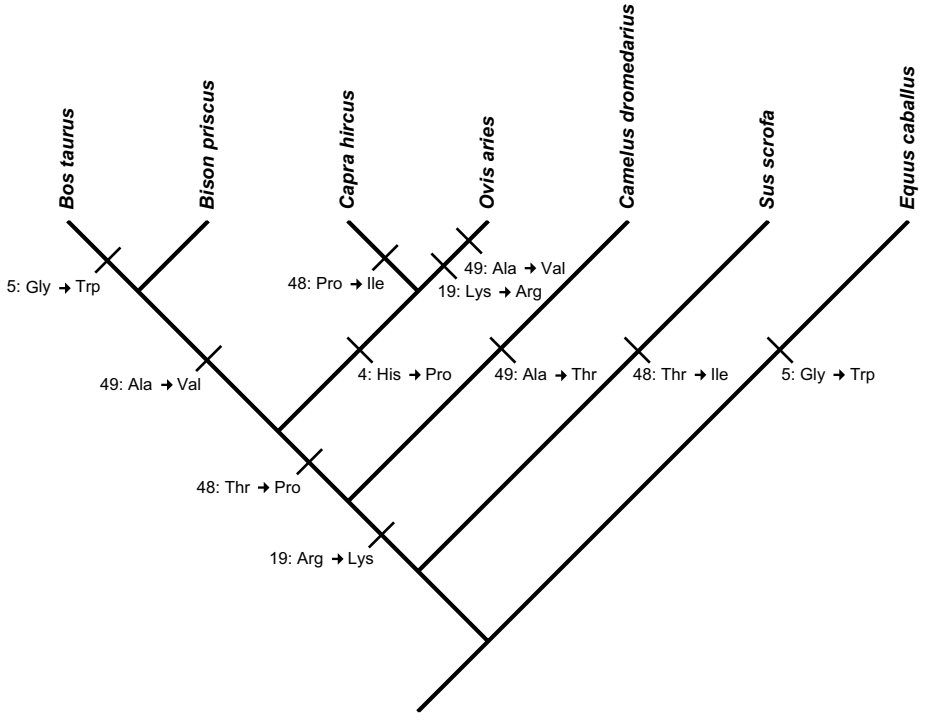

Fig. 5. Osteocalcin amino acid residues 4, 5, 19, 48, and 49 mapped onto a pruned version of the Cytochrome *b* maximum likelihood tree pruned to resolve sister relationship of *Sus* and *Camelus*.

Table 2
Evolution of OC within the Artiodactyla. CI is the consensus index and RI is retention index

Character	CI	RI	Character evaluation
4	1.00	1.00	His → Pro defines the *Ovis aries* – *Capra hircus* clade
5	0.50	0.00	Gly → Trp in both *Equus* and *Bos*
19	0.50	0.50	Arg → Lys places Camelus with Bovids; with reversion (Lys → Arg) in *Ovis*
48	0.67	0.50	Thr → Ile in Sus, Thr → Pro defines Bovids, with Pro → Ile in Capra
49	0.67	0.50	Ala → Thr in Camelus; Ala → Val in *Bos* and *Bison*, *Ovis*

maximum likelihood tree (without homoplasy; RI = CI = 1.00). Characters at all of the other four amino acid positions evolve with varying degrees of homoplasy, and these evolutionary changes are summarized in Table 2. These data indicate that within Artiodactyla osteocalcin has not evolved in such a way to reflect species relationships. Our pruned data set shows clear examples of both parallel evolution and convergence (character 5 in *Bos* and *Equus* and character 48 in *Sus* and *Capra*, respectively). Further work will need to be done to determine if these amino acid changes are due to positive selection or genetic drift. If the amino acid substitutions resulted from selective rather than neutral drift, it implies an event with functional significance for osteocalcin.

We also examined the utility of osteocalcin sequences for exploring phylogenetic relationships at higher taxonomic levels. The objective was to understand whether phylogenetic inferences based on OC sequences (Fig. 4a) reflected taxonomic relationships delineated in the Cytochrome *b* tree (Fig. 4b). In the ML tree of osteocalcin, amphibians were grouped together without bootstrap support and several characters support the monophyly of the ingroup. Rodents were recovered, placing *Rattus* with *Mus*, albeit without bootstrap support. The *Homo-Pan-Gorilla* trichotomy within hominins was also recovered (NJ/MP bootstrap = 65/61), as were the birds (*Dromaius* and *Gallus* grouped with NJ/MP bootstrap = 87/89).

Evolutionary changes in osteocalcin among primates, particularly the absence of post-translational modification of Pro_9, are related to enzymatic requirements for prolyl-4-hydroxylase (Nielsen-Marsh et al., 2005). The prolyl-4-hydroxylase enzyme requires vitamin C and recognizes the consensus sequence $Leu_6Gly_7Ala_8Pro_9Ala_{10}Pro_{11}Tyr_{12}$ but hydroxylation appears to be resisted in five primates when Val_{10} substitutes for Ala_{10}.

A preliminary evaluation of character states among fish and tetrapods showed that the osteocalcin sequence of fish is clearly unique from that of other taxa. Gln_{15} and Ile_{27} in fish are replaced by Pro_{15} and Pro_{27} in amphibians, birds and mammals. Ser_{18} in fish and amphibians is replaced by Ala_{18} in birds and Pro_{18} in mammals. Furthermore, Frazao et al. (2005) identified a fourth Gla at position 31 in the teleost fish *Argyromsomus regius*. They argued that the additional carboxylated glutamic acid could be related to actions associated with interactions of the carboxylase enzyme with the propeptide sequence. The activity of the carboxylase is mediated by a conserved 18 amino acid propeptide sequence. Its affinity and, hence carboxylation, is impaired when mutations occur at position −15, −9, and −6. Whereas only one of these residues is conserved in tetrapods, two are conserved in the *A. regius* sequence.

In addition to substitutions at individual positions, teleosts lack a conserved sequence present at the C-terminus of most tetrapods (RFYGPV) and retain an extended and conserved motif at the N-terminus (AYXXY/FYGP) that is truncated in tetrapods (Nishimoto et al., 2003; Cancela et al., 1995). Given the limited number of osteocalcin sequences available and lack of understanding of osteocalcin's functional attributes, assigning significance to sequence substitutions is difficult. Such interpretations would be significant as mineralized tissue is an innovation of vertebrates providing the framework for phenotypic variation (endoskeleton, predatory teeth).

While the conservative nature of osteocalcin appears to make phylogenetic analysis of its amino acid sequence amenable for recovering a limited number of relationships, character analysis of the osteocalcin sequence is useful for identifying specific taxa. For example, within the order Artiodactyla, character 4 defines the *Capra-Ovis* clade with a substitution of Pro_4 for His_4. Thus, in the absence of other distinguishing morphological or molecular characteristics for an artiodactyls skeletal element, osteocalcin sequencing could be used to constrain taxonomic affinity.

To develop an understanding of osteocalcin evolution and its utility for phylogenetic reconstruction, sequence data from additional taxa are required, especially among reptiles, birds, and within the orders Carnivora and Perisodactyla. The preliminary analysis of osteocalcin evolution provided here is an initial step toward a complete character analysis specifically aimed at determining the evolutionary history of osteocalcin.

Our efforts mark the first phylogenetic analysis of an ancient protein. As was done here, future studies applying mass spectrometric based approaches to sequencing and phylogenetic analysis must be cognizant of the impacts of post-translational modifications, the possibility of isobaric amino acids (e.g. Leu and Ile) and the influence of diagenetic changes (e.g. deamidation of Gln to Glu and dehydroxylation of Ser to Ala) on sequence alignments and/or mass shifts. As noted here, the fact that protein sequence databases are limited and do not retain information for extinct species, presents obstacles for interpreting sequence data from ancient taxa. While diagenetic peptides of indigenous proteins are presumably taxonomically informative, sequence information from such fragmentary material are not amenable to standard phylogenetic analyses. Developing approaches delineating phylogentic relationships for such material is one of the next challenges to unraveling the secrets of the paleoproteome.

ACKNOWLEDGMENTS

We thank Barbara Lundrigan, Michigan State University and Ross MacPhee, American Museum of Natural History for bones of modern camelids used in this research. We thank two unnamed

referees for contributing helpful comments on this manuscript and those of the Associate Editor, Jay Brandes. This work was supported by the National Science Foundation (NSF) Grant EAR-0309467 (to P.H.O.), the National Resource for Proteomics and Pathways, NCRR Grant P41-18627 (to P.C.A.) and the Michigan Proteome Consortium.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gca.2007.09.003.

REFERENCES

Asara J. M., Schweitzer M. H., Freimark L. M., Phillips M. and Cantley L. C. (2007) Protein sequences from mastodon and *Tyrannosaurus Rex* revealed by mass spectrometry. *Science* **316**, 280–285.

Betancourt J. L., Van Devender T. R. and Martin P. S. (1990) *Packrat Middens: The Last 40,000 years of Biotic Change*. University of Arizona Press, Tucson, Arizona.

Cancela M. L., Williamson M. K. and Price P. A. (1995) Amino-acid sequence of bone Gla protein from the African clawed toad *Xenopus laevis* and the fish *Sparus aurata*. *Int. J. Pept. Protein Res.* **46**, 419–423.

Colbert E. H., Morales M. and Minkoff E. C. (1991) *Colbert's Evolution of the Vertebrates: A History of the Backboned Animals Through Time*, fifth ed. Wiley, Canada.

Collins M. J., Gernaey A. M., Nielsen-Marsh C. M., Vermeer C. and Westbroek P. (2000) Slow rates of degradation of osteocalcin: green light for fossil bone protein? *Geology* **28**, 1139–1142.

Dowd T. L., Rosen J. F., Li L. and Gundberg C. M. (2003) The three-dimensional structure of bovine calcium ion-bound osteocalcin using H NMR spectroscopy. *Biochemistry* **42**, 7769–7779.

Frazao C., Simes D. C., Coelho R., Alves D., Williamson M. K., Price P. A., Cancela M. L. and Carrondo M. A. (2005) Structural evidence of a fourth Gla residue in fish osteocalcin: biological implications. *Biochemistry* **44**, 1234–1242.

Harris A. H. and Findley J. S. (1964) Pleistocene-recent fauna of Isleta Caves Bernalillo County New Mexico. *Am. J. Sci.* **262**, 114–120.

Hauschka P. V., Lian J. B., Cole D. E. C. and Gundberg C. M. (1989) Osteocalcin and matrix Gla protein—vitamin K-dependent proteins in bone. *Physiol. Rev.* **69**, 990–1047.

Hoang Q. Q., Sicheri F., Howard A. J. and Yang D. S. C. (2003) Bone recognition mechanism of porcine osteocalcin from crystal structure. *Nature* **425**, 977–980.

Jones D. T., Taylor W. R. and Thornton J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.

Kurtén B. and Anderson E. (1980) *Pleistocene Mammals of North America*. Columbia University Press, New York.

Laizé V., Martel P., Viegas C. S. B., Price P. A. and Cancela M. L. (2005) Evolution of matrix and bone γ-carboxyglutamic acid proteins in vertebrates. *J. Biol. Chem.* **280**, 26659–26668.

Lange I. M. (2002) *Ice Age Mammals of North America: A Guide to the Big, the Hairy, and the Bizarre*. Mountain Press, Missoula, Montana.

Lian J., Stewart C., Puchacz E., Mackowiak S., Shalhoub V., Collart D., Zambetti G. and Stein G. (1989) Structure of the rat osteocalcin gene and regulation of vitamin D-dependent expression. *Proc. Natl. Acad. Sci. USA* **86**, 1143–1147.

Maddison D. R. and Maddison W. P. (2000) *MacClade version 4 Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Matthee C., Burzlaff A. J. D., Taylor J. F. and Davis S. K. (2001) Mining the Mammalian Genome for artiodactyl systematics. *Syst. Biol.* **50**, 367–390.

Mundy G. R. and Poser J. W. (1983) Chemotactic activity of the γ-carboxyglutamic acid containing protein in bone. *Calcif. Tissue Int.* **35**, 164–168.

Nielsen-Marsh C. M., Ostrom P. H., Gandhi H., Shapiro B., Cooper A., Hauschka P. V. and Collins M. J. (2002) Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka. *Geology* **30**, 1099–1102.

Nielsen-Marsh C. M., Richards M. P., Hauschka P. V., Thomas-Oates J. E., Trinkaus E., Pettitt P. B., Karavanic I., Poinar H. and Collins M. J. (2005) Osteocalcin protein sequences of Neanderthals and modern primates. *Proc. Natl. Acad. Sci. USA* **102**, 4409–4413.

Nishimoto S. K., Waite J. H., Nishimoto M. and Kriwacki R. W. (2003) Structure, activity, and distribution of fish osteocalcin. *J. Biol. Chem.* **278**, 11843–11848.

Ostrom P. H., Gandhi H., Strahler J. R., Walker A. K., Andrews P. C., Leykam J., Stafford T. W., Kelly R. L., Walker D. N., Buckley M. and Humpula J. (2006) Unraveling the sequence and structure of the protein osteocalcin from a 42 ka fossil horse. *Geochim. Cosmochim. Acta* **70**, 2034–2044.

Poinar H. N., Schwarz C., Qu J., Shapiro B., MacPhee R. D. E., Buigue B., Tikhonov A., Huson D., Tomsho L. P., Auch A., Rampp M., Miller W. and Schuster S. C. (2005) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394.

Posada D. and Crandall K. A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Robbins L. L., Muyzer G. and Brew K. (1993) Macromolecules form living and fossil biominerals: Implications for the establishment of molecular phylogenies. In *Organic Geochemistry: Principles and Applications* (eds. M. H. Engle and S. A. Macko). Plenum, New York, pp. 799–816.

Savage D. E. (1951) Late Cenozoic vertebrates of the San Francisco bay region. *Univ. Calif. Publ. Geol. Sci.* **18**, 215–314.

Schmidt H. A., Strimmer K., Vingron M. and von Haeseler A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504.

Stark G. R. (1965) Reactions of cyanate with functional groups of proteins. Inertness of aliphatic hydroxyl groups. Formation of carbamyl- and acylhydantoins. *Biochemistry* **4**, 363–2367.

Stuart A. J. (1991) Mammalian extinctions in the late Pleistocene of northern Eurasia and North America. *Biol. Rev. Camb. Philos. Soc.* **66**, 453–562.

Swofford D. L. (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods). Version 4.0b10*. Sinauer Associates, Sunderland, Massachusetts.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F. and Higgins D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882.

Vogt W. (1995) Oxidation of methionyl residues in proteins—tools, targets, and reversal. *Free Radic. Biol. Med.* **18**, 93–105.

Webb S. D. (1974) Pleistocene llamas of Florida, with a brief review of the Lamini. In *Pleistocene mammals of Florida* (ed. S. D. Webb). University Press of Florida, Gainesville, Florida, pp. 170–214.

Zardoya R. and Meyer A. (1996) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.* **13**, 933–942.

Associate editor: Jay A. Brandes