Vana asja uurimisest kaasaegsete meetoditega

Aare Abroi
31. jaan. 2014

Swedish Museum of Natural History
The fetus used by Carl Linnaeus as the type specimen of the Asian elephant.

Proteins help solve taxonomy riddle
Proteomic technique proves that 300-year-old Linnaean elephant was wrongly classified.
Nature 07 November **2013**

Linne arvas, et tegemist on Aasia elevandiga (*Elephas maximus)* aga paljud zooloogid kahtlustasid, et tegemist on hoopis Aafrika elevandiga (*Loxodonta africana)*.

DNA osutus liigi määramiseks liiga lagunenuks.

Prooviti proteoomika lähenemist söögitoru proovidele. Ja oh imet – õnnestuski identifitseerida mitmeid valke mis ei ole nendel kahel liigil identsed ning tõestada, et tegemist on Aafrika elevandiga.

Mass-spektromeetria

Mass-spektromeetria on kaasaegne kõrgtehnoloogiline meetod mis toodab palju andmeid ja annab tööd paljudele bioinformaatikutele.

Mõõdab osakese mass/laengu suhet. Iseenesest vana meetod, kuid biomolekulide korral oli kaua aega probleemiks kuidas saada biomolekulid tervetena gaasifaasi. Umbes 25-30 aastat tagasi esimesed lahendused sellele probleemile.

Väga suur lahutusvõime. Sobilikud piigid (m/z suhtega ainesed) saab saata järgnevale uuringule (MS-MS) kus valkude puhul saab määrata mõned otsmised aminohapped (heal juhul mõnikümmend).

MS lahutusvõimest

Prooton on elektronist 1836 korda raskem.
Neutron on elektronist 1838,5 korda raskem.

ESI-MS mõõdab tavaliselt (peptiide) 1000 kuni 4000 m/z –ga (ehk 5 kuni 50 aa pikkuseid peptiide).

St. kui meil on peptiid massiga 2000 Da siis lahutusvõime 1 ppm (parts per miljon) on 0,002 Da, elektroni mass 0,0005 Da.

Kui osakese molekulmass on 200 Da siis 1 ppm = 0,0002 Da, st. eristab osakest massiga 200 Da ning laeng -1 (üks lisa elektron) (m/z 200,0005) osakesest millel on mass 400 Da ja laeng -2 (2 lisa elektroni) (m/z 200,0010).

Aminohapete molekulmassid Gly 75 kuni Trp 204

Proteomic Analysis of a Pleistocene Mammoth Femur Reveals More than One Hundred Ancient Bone Proteins.

Journal of Proteome Research **2012**

http://commons.wikimedia.org/wiki/File:Mammoth_femur.jpg

Igikeltsast leitud ~43 000 aasta vanusest karvase mammuti (*Mammuthus primigenius*) reieluust võetud 2mm diameetriga proovid. Luukoe proovi umbes 70 mg.

Iga valgu kohta vähemalt kaks peptiidi. Saadi üle 120 valgu.

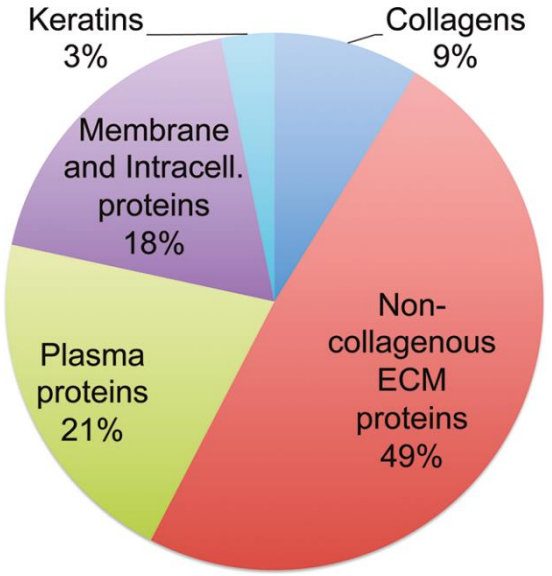
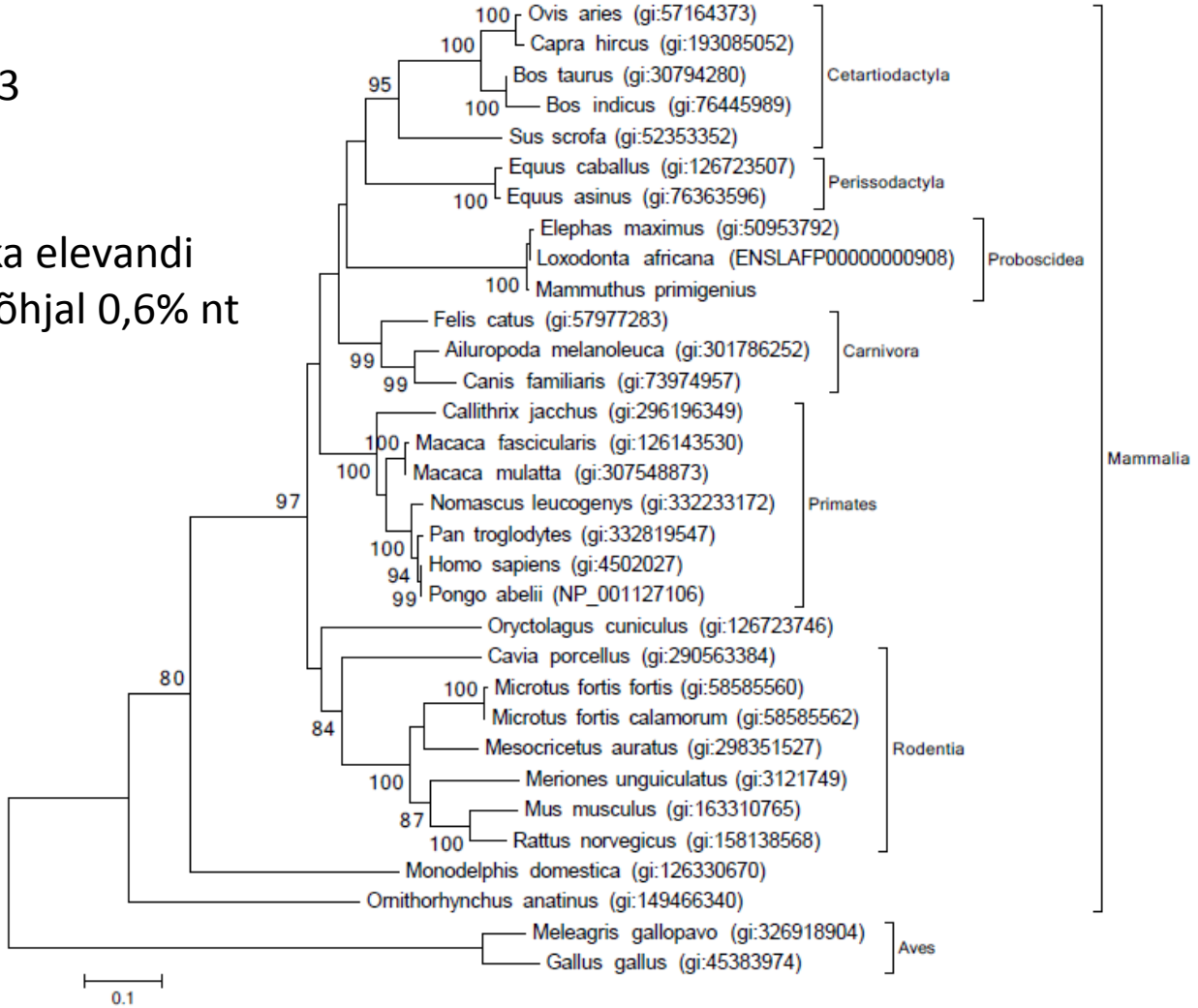


Figure 1. Composition of the proteome from a 43 000 year old mammoth bone. Secreted extracellular matrix proteins (**ECM**), collagenous and non-collagenous, are prevalent. Plasma proteins, associated with the hematopoietic role of bone, are broadly represented as well.

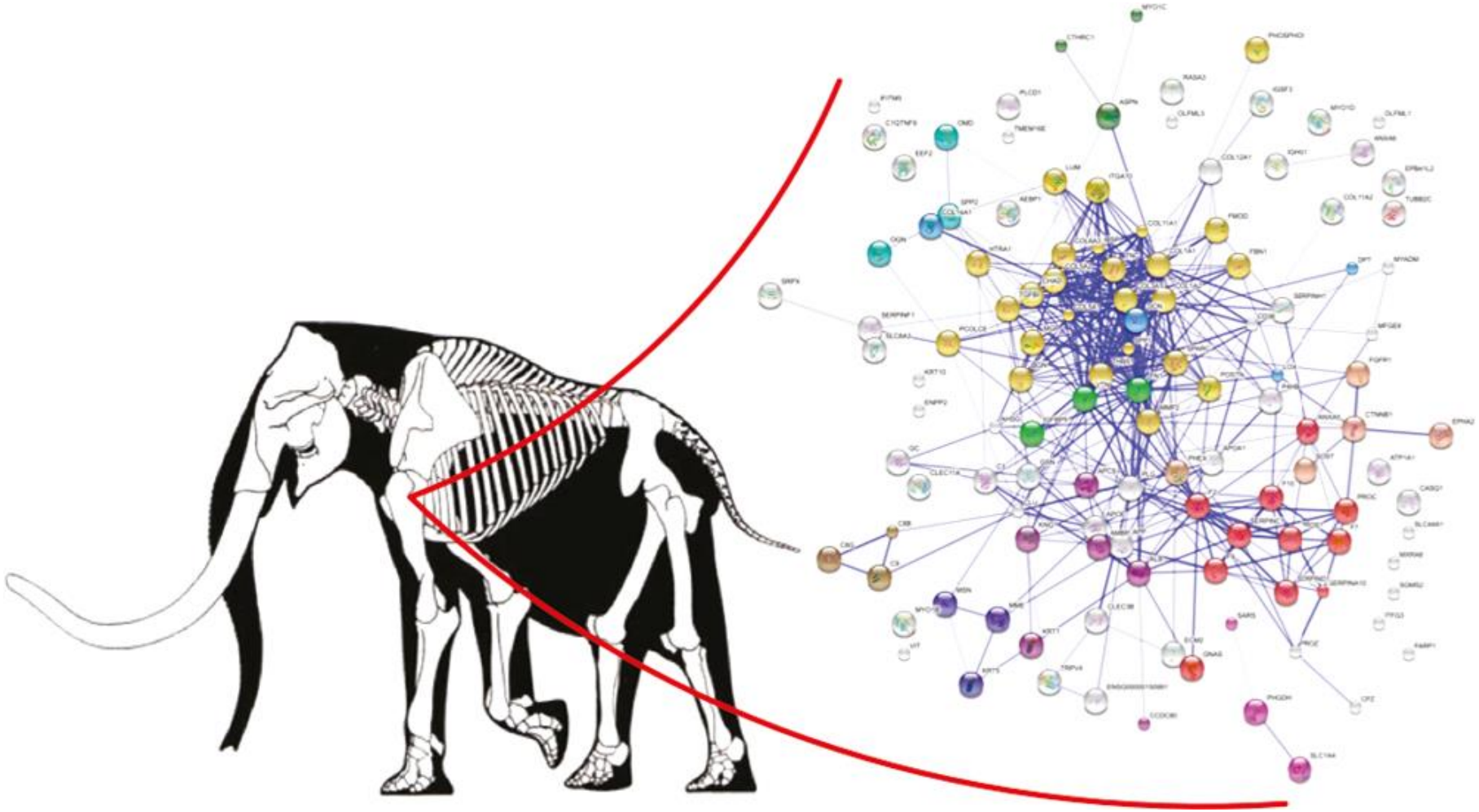
Kollageen COL1A1 ja COL1A2 valkudele saadi umbes 70% katvus, st. 70% oletatavatest aa identifitseeriti.

Umbes 17 000 aa kohta ~13 mutatsiooni ehk 0,076%

Karvase mammuti ja Aafrika elevandi erinevus tuumagenoomi põhjal 0,6% nt ja 0,22% aa tasemel.



Seerum albumiini põhine ML puu



Cappellini et al. Journal of Proteome Research **2012**

Protein Sequences from Mastodon and Tyrannosaurus Rex Revealed by Mass Spectrometry

John M. Asara, Mary H. Schweitzer, Lisa M. Freimark, Matthew Phillips, Lewis C. Cantley

Science 13 April 2007:
Vol. 316 no. 5822 pp. 280-285

Fossilized bones from extinct taxa harbor the potential for obtaining protein or DNA sequences that could reveal evolutionary links to extant species. We used mass spectrometry to obtain protein sequences from bones of a 160,000- to 600,000-year-old extinct mastodon (Mammut americanum) and a 68-million-year-old dinosaur (Tyrannosaurus rex). The presence of T. rex sequences indicates that their peptide bonds were remarkably stable. Mass spectrometry can thus be used to determine unique sequences from ancient organisms from peptide fragmentation patterns, a valuable tool to study the evolution and adaptation of ancient taxa from which genomic sequences are unlikely to be obtained.

Tavapäraselt otsitakse vastavalt MS andmetele millisele varem kirjeldatud (andmebaasis olevale) peptiidile saadud andmed vastavad.

Andmebaasis olevad peptiidid võivad olla ka DNA pealt in silico transleeritud (st. oletatavad, mitte reaalsed peptiidid).

Esmalt võrreldi saadud MS andmeid teadaolevate valkudega.

Et leida väljasurnud liigile omaseid peptiide tekitati võimalike peptiidide valim arvestades konserveerumust.

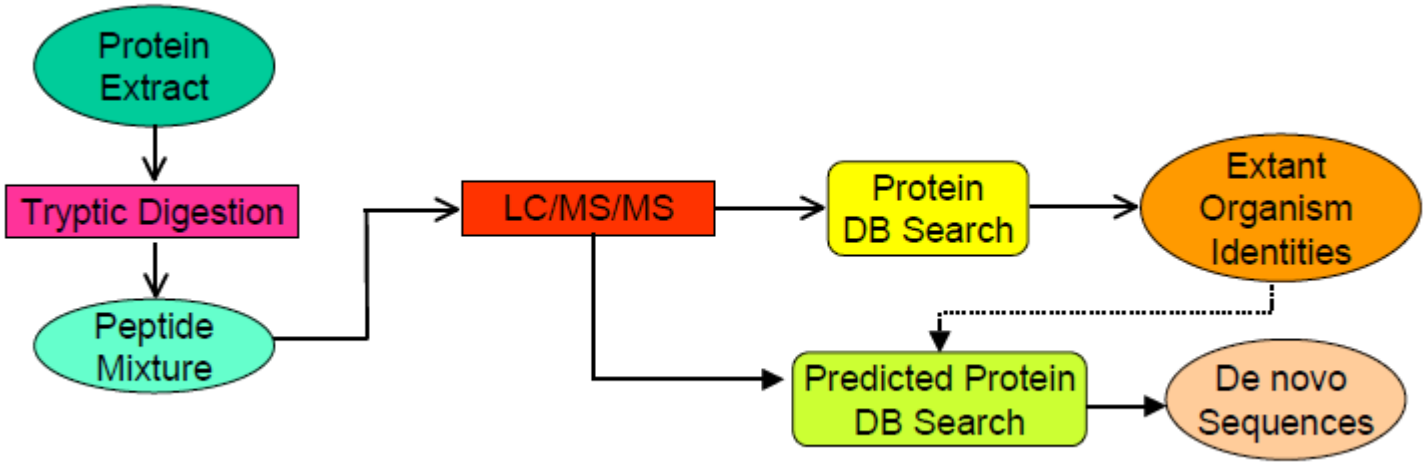
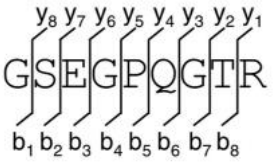


Figure S1. Workflow for sequencing peptides from previously unsequenced taxa by LC/MS/MS. Peptide sequences from extinct taxa are determined by digesting the protein extract with trypsin and the resulting peptide mixture is then analyzed by microcapillary LC/MS/MS. The fragmentation (MS/MS) spectra are then searched against available protein databases to identify peptide sequences that match identically to extant related organisms of known sequence. The identified protein(s) of interest from related taxa are then used to generate a theoretical database of predicted protein sequences potentially covering the unsequenced taxon of interest. The unmatched MS/MS spectra are then searched against these predicted protein sequences. Peptide sequence matches to the predicted database entries are novel sequences that are unique to the unsequenced taxon of interest.
Asara JM et al. Science 2007

A

Peptide Sequences Unique to Mastodon	Protein	Xcorr	Sp
GSEGPQGTR	collagen α1t1	2.22	1187
GAPGPQGP*G*GAP*GPK	collagen α1t1	3.44	1017
EGAPGSEGAPG*RDGAIGPK	collagen α1t1	2.86	580
GLTGPIGPP*GPAGAP*GDKGEG*GPSGPAGPTGAR	collagen α1t1	5.63	715



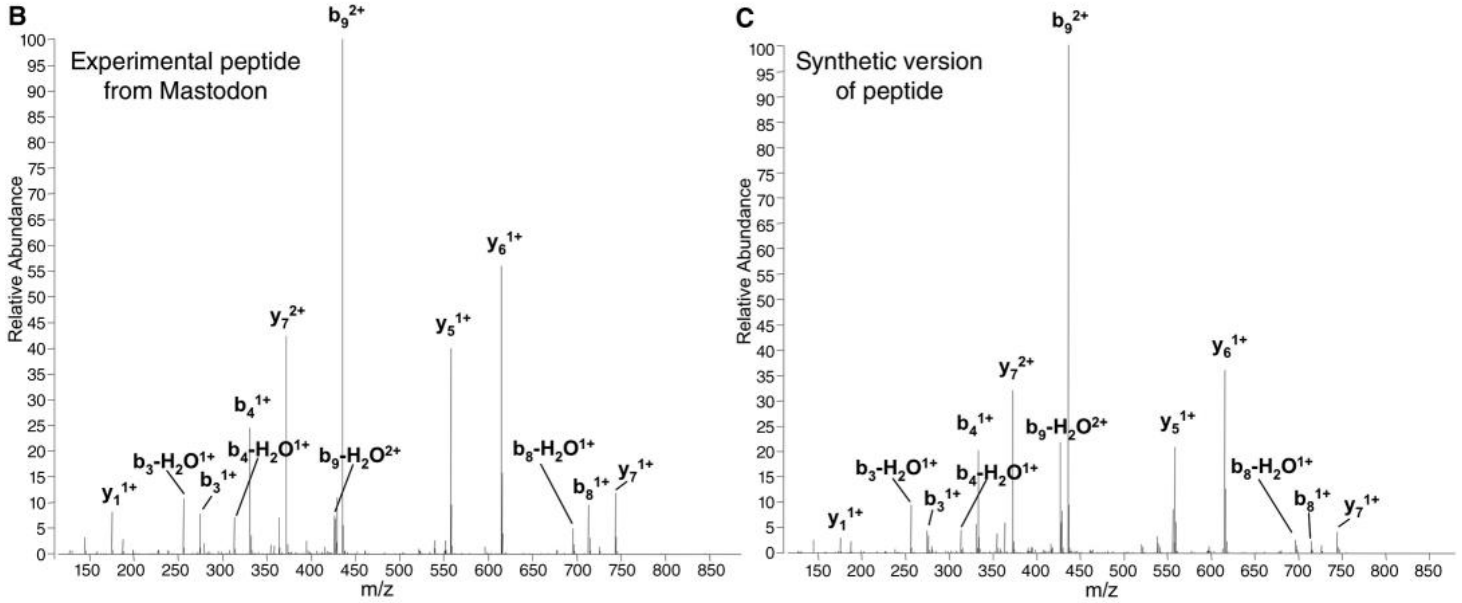
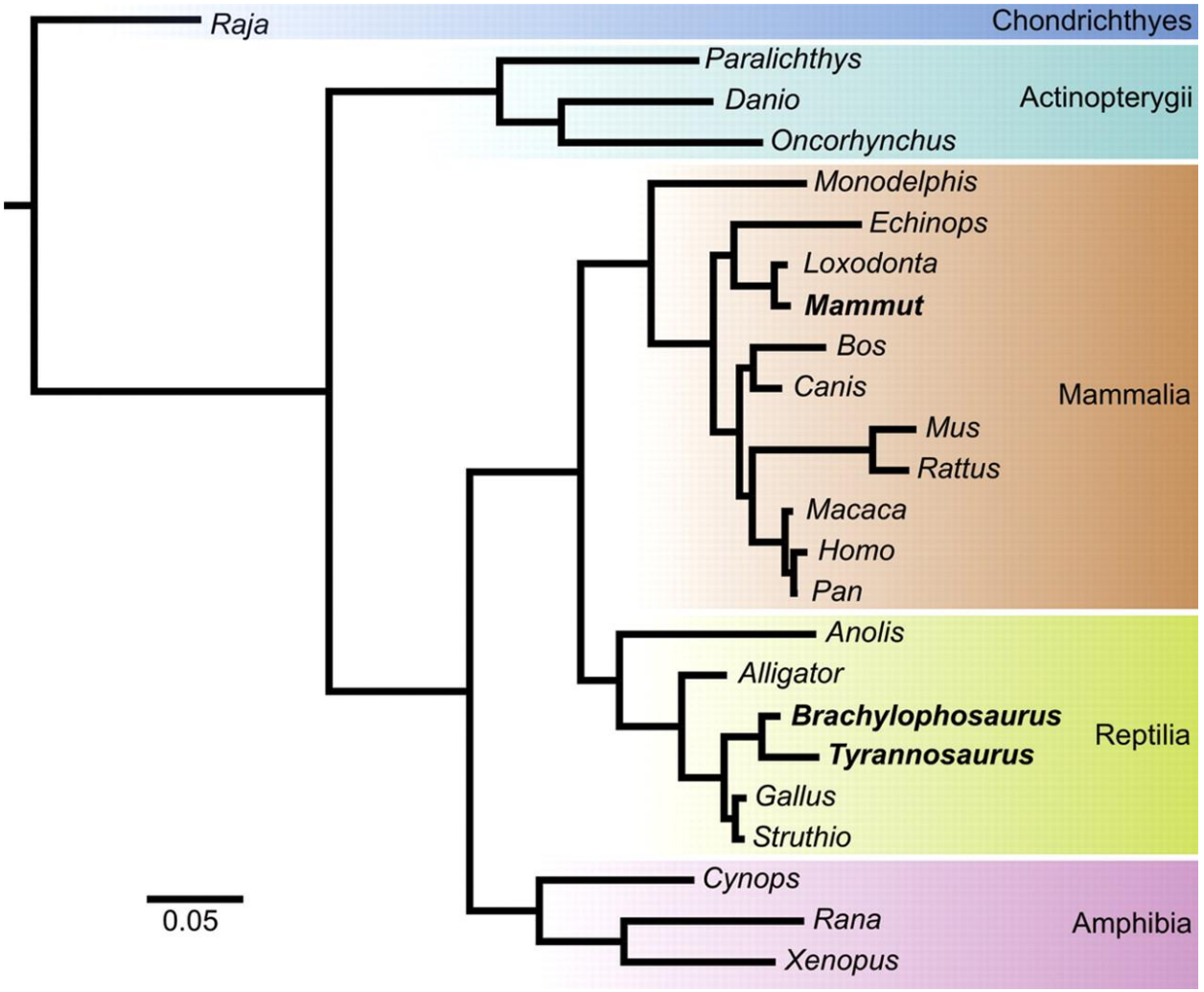


Fig. 2. Collagen peptide sequences unique to extinct mastodon identified by LC/MS/MS. (A) The four collagen a1t1 peptide sequences found by the approach that are unique to ancient mastodon. Xcorr (cross-correlation score) and Sp (preliminary score) represent the scores resulting from database searching against protein databases using Sequest. The asterisk represents the hydroxylation site after the posttranslationally modified residue. (B) An example of the experimental MS/MS spectrum of a doubly charged tryptic peptide for the collagen a1t1 peptide sequence GSEGPQGTR from the LC/MS/MS analysis of mastodon fossilized bone extract identified from a Sequest search against a theoretical collagen protein database. (C) The synthetic version of the same peptide sequence. Allmajor ions from the experimental spectrum align very well with the ions from the synthetic version, validating the sequence.
Asara JM et al. Science 2007

After passing the scoring thresholds, all MS/MS were then manually inspected rigorously to be sure that all b- (fragment ions resulting from amide bond breaks from the peptide's N-terminus) and y- ions (fragment ions resulting from amide bond breaks from the peptide's C-terminus) aligned with the assigned protein database sequence including shifts due to post-translational modifications.

Fig. 4 Consensus of the posterior distribution of phylogenetic trees including M. americanum (Mastodon, MOR 605) and the extinct dinosaurs B. canadensis (MOR 2598) and T. rex (MOR 1125, in bold).



M H Schweitzer et al. Science 2009;324:626-631

Published by AAAS

Biomolecular Characterization and Protein Sequences of the Campanian Hadrosaur *B. canadensis*

Science 1 May 2009

Mary H. Schweitzer, Wenxia Zheng, Chris L. Organ, Recep Avci, Zhiyong Suo, Lisa M. Freimark, Valerie S. Lebleu, Michael B. Duncan, Matthew G. Vander Heiden, John M. Neveu, William S. Lane, John S. Cottrell, John R. Horner, Lewis C. Cantley, Raghu Kalluri, John M. Asara

Molecular preservation in non-avian dinosaurs is controversial. We present multiple lines of evidence that endogenous proteinaceous material is preserved in bone fragments and soft tissues from an **80-million-year-old** Campanian hadrosaur, *Brachylophosaurus canadensis* [Museum of the Rockies (MOR) 2598]. Microstructural and immunological data are consistent with preservation of multiple bone matrix and vessel proteins, and phylogenetic analyses of *Brachylophosaurus* collagen sequenced by mass spectrometry robustly support the bird-dinosaur clade, consistent with an endogenous source for these collagen peptides. These data complement earlier results from *Tyrannosaurus rex* (MOR 1125) and confirm that molecular preservation in Cretaceous dinosaurs is not a unique event.

http://www.wikidino.com/?attachment_id=2347

80-million-year-old Campanian hadrosaur, *Brachylophosaurus canadensis*

MS ja immunoloogiliste meetoditega näidati et dinosauruse kollageen on säilunud.

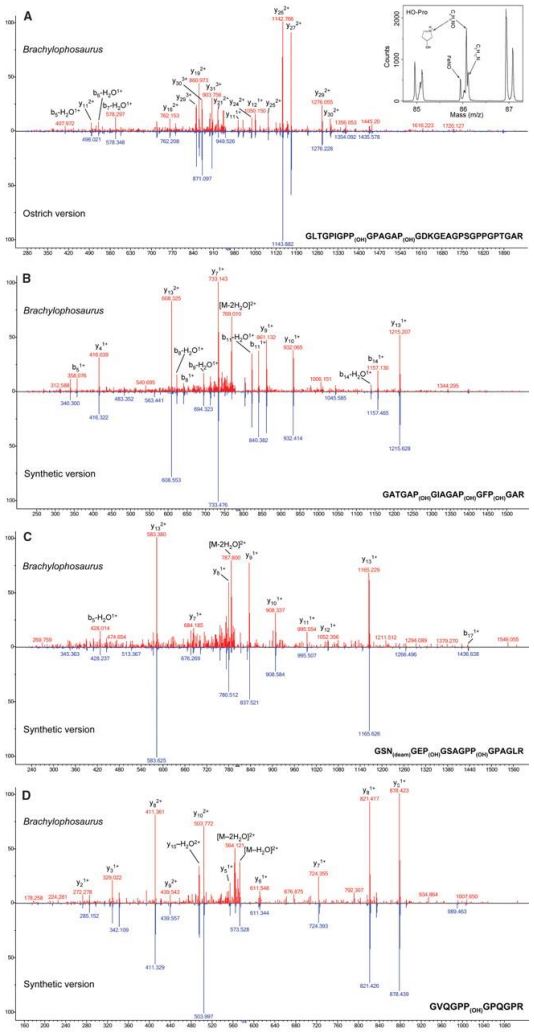
Hüdroksüproliin mida bakterites väidetavalt ei ole kuid kollageenis on konserveerunud.

Liivakivist leitud reieluu.

8 peptiidi, kokku 149 aminohapet

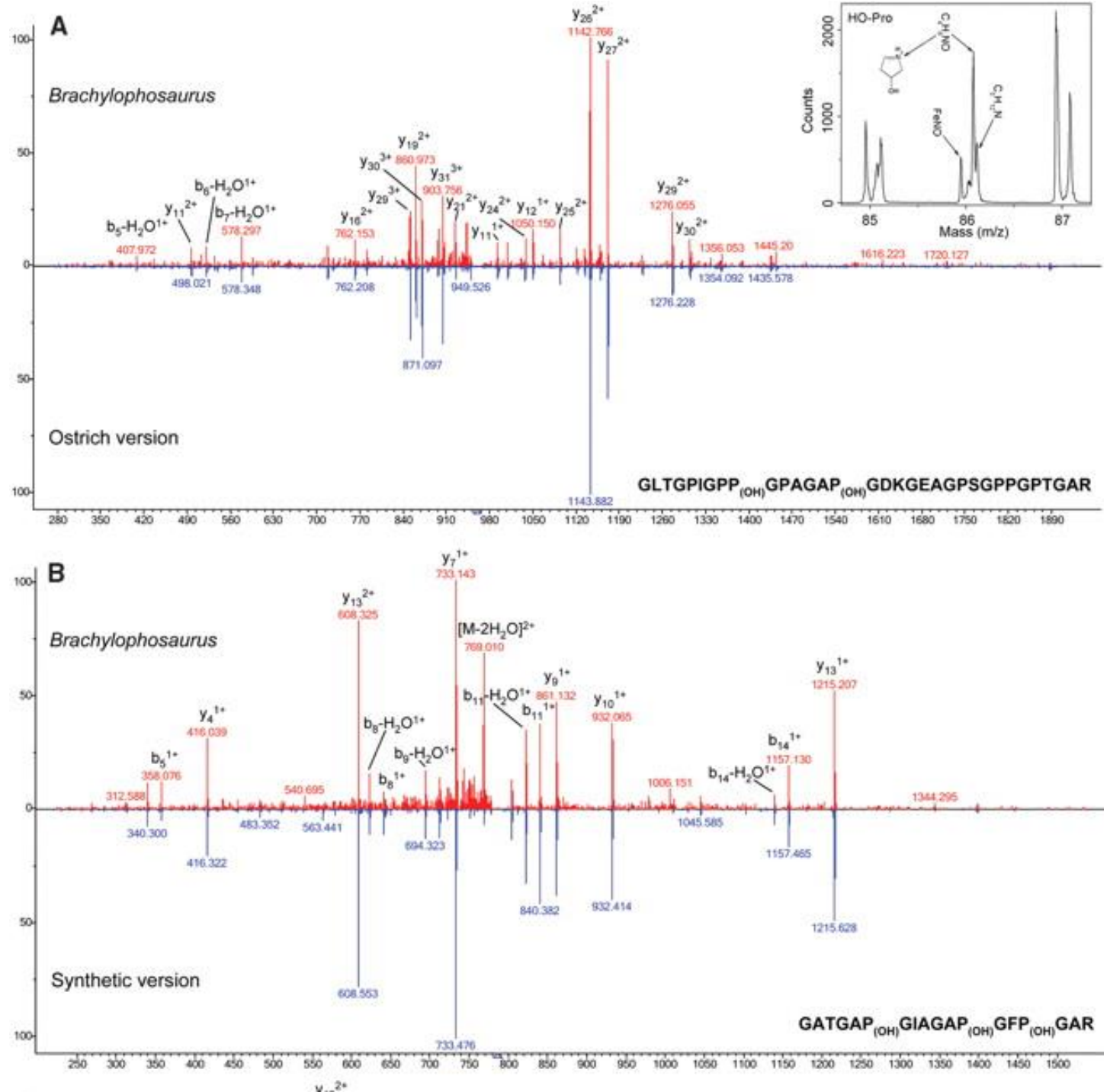
Ühel järjestusel identifitseeriti aspargiini deamidatsioon ja kõikides peptiidides esinas vähemalt üks hüdroksüproliin.

Fig. 3 Mass spectrometry analyses of B. canadensis tissues and extracts.



M H Schweitzer et al. Science 2009;324:626-631

Published by AAAS


A
Brachylophosaurus
Ostrich version
GLTGPIGPP(OH)GPAGAP(OH)GDKGEAGPSGPPGPTGAR
HO-Pro
Counts
Mass (m/z)
B
Brachylophosaurus
Synthetic version
GATGAP(OH)GIAGAP(OH)GFP(OH)GAR

DNA põhine	**Proteoomika põhine**
saab amplifitseerida	ei saa amplifitseerida
	väga suur lahutusvõime
	Laguproduktid on võimalik kõrvale jätta.
copy nr madal	copy nr kõrge (albumiin, kollageen)

Huvilised võivad lugeda veel
http://www.sciencemag.org/content/343/6177/1320.full.pdf

Paleoproteomics and ancestral protein reconstruction ??

Toiduks (kui maitseained välja jätta) kasutame taimede puhul enamasti osiseid kus valk/DNA suhe on suur.

The organic bone matrix, 90% of which is collagen, represents approximately 25% of the bone mass and is believed to persist over long periods of time, supposedly even over the Cretaceous/Tertiary boundary.

The **monoisotopic mass** is the sum of the masses of the atoms in a molecule using the unbound, ground-state, rest mass of the principal (most abundant) isotope for each element instead of the isotopic average mass. The term is designed for measurements in mass spectrometry primarily with smaller molecules. It is not typically useful as a concept in physics or general chemistry