

FROM FOSSILS TO PHYLOGENIES PART 3: HOW DINOSAURS FIT INTO THE EVOLUTIONARY TREE OF LIFE

[Written by: Baylee Goodwin, Dane Besser, and Stephen Ramsey]

Vocabulary Words	
Phylogenetics	Most Recent Common Ancestor (MRCA)
Taxa	Descendants
Node	Sister Clades
Speciation event	Outgroup

Background

During the Mass Spectrometry and BLAST activities, you were given amino acid sequences that had been recovered from a fossilized bone specimen from a *Tyrannosaurus rex* (as well as sequences from a Hadrosaur and a Mastodon). You learned how to input the *T. rex* amino acid sequence into BLAST to identify what present-day animals are most closely related to *T. rex*. In this activity you will learn how to use a computer to analyze related amino acid sequences from a variety of animals to gain insight on their evolutionary relationships.

Phylogenetics is the study of evolutionary relationships among a set of taxa, where taxa is another name for groups of organisms, like plants and animals. In phylogenetics, evolutionary relationships are laid out on a phylogenetic tree (fig 1 & 3). The root of the tree is the start of the evolutionary lineage being depicted. As you move from the left to the right, you are moving forward in time. As time passes you can see how nodes diverge in two directions, this is a speciation event. Which is when a group of animals separates and evolves into two brand new groups of animal. The nodes also mark where the most recent common ancestor (MRCA) is. For example, in figure 3, A and B are groups of animals that diverged from a MRCA found at the node that join the two. The leaves at the end of the tree mark the descendants of the ancestors. Phylogenetic trees are a useful way to compare

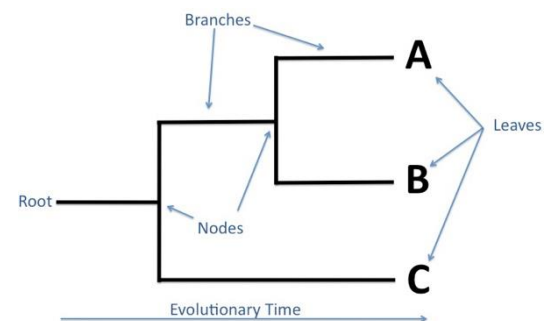


Figure 1: This is an example of a phylogenetic tree showing how to read it.

	Hun.	Prin.	Anc.
Human	ATGAACGCATGC		
Chimp.	ATGCACGCATGC		
Gorilla	ATGCATGCATGC		

Figure 2: This is an example of DNA sequences from multiple species lined up together. Species who share mutations that others do not have are more closely related. This is how molecular biology can help determine evolutionary relationships.

how animals are related to one another. In figure 3, animals from group A and B are more closely related to each other than they are to animals in group C. Therefore, A and B would be considered sister clades, since they are the most closely related. Group C would be considered the outgroup since they are the most distantly related.

There are two main methods on how to determine the evolution of a set of taxa: morphology vs molecular data. Morphology uses physical features of animals to determine how they evolved, whereas molecular data uses DNA/amino acid sequences to determine where speciation events occurred. Molecular data is typically more accurate since mutations in DNA are the driving force for evolution.

When mutations arise in the DNA sequence of an organism, they can result in changes to the translated amino acid sequence of a protein. For example, the original DNA sequence in a small portion of a gene might have read ATAAGT, but after the mutation it reads ATAACT (i.e., a G was replaced with a C). This changes the amino acid in the sequence from leucine to a stop codon (signaling the end of the protein), which results in the cell making a shortened protein whose function may substantially differ from the original full-length protein. When a mutation is present in an organism's cell, it can be passed on from the organism to its offspring, which is how animals evolve on a molecular scale.

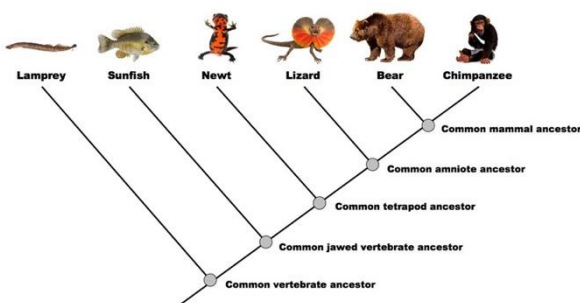


Figure 3: This is an example of a basic phylogenetic tree. It highlights where the most recent common ancestors (MRCA) are found on the tree, and which animal groups evolved from the ancestor.

The genetic differences between two species, such as a bird and a lizard, represent the accumulation of billions of mutations over many millions of years. The differences in the DNA (or, as we will study today, protein) sequences among a set of representative species can be used to determine how the species are related. As we will discover, the more closely related organisms will have more similar protein sequences, and the more distantly related organisms will have more dissimilar protein sequences.

In order to create a phylogenetic tree, the first step is to obtain protein sequence data from a set of animal species that we want to compare. We will be searching for the "alpha-2 type 1 collagen" protein sequence since that is the what scientists were able to extract from the fossilized femur bone of the *T. rex*. Collagen is evolutionarily rather well-conserved across species, which is why it is a good choice for using amino acid sequences to build a phylogenetic tree. When a protein is "well-conserved" it means that the protein is found in multiple species that are distantly related, collagen is a well-conserved protein found in all animals with true bone. In order to find the collagen sequence, you will conduct a search in an online database called GenBank. The alpha-2 type 1 collagen protein sequences have already been collected for you for most of the animals, however you still need to collect the appropriate amino acid sequence for the *T. rex*.

The animal species for which you will be building a phylogenetic tree are chicken, rainbow trout, human, dog, cattle, toxodon (*Toxodon platensis*), mastodon (*Mammot americanus*), salamander, frog, and *T. rex*. (Note: these species were selected for this activity because they have alpha-2 type 1 collagen protein sequences available in the database). These animals will allow you to analyze where the *T. rex* fits in the tree relative to birds, mammals, amphibians, and fish (no reptile data were available), which are four out of the five major vertebrate taxonomic groups. It also allows you to see how other extinct animals like mastodon and toxodon relate to present-day animals.

Learning Objectives

- Understand how amino acid sequences can be compared using a computer program in order to reconstruct a phylogenetic tree
- Learn how to obtain protein or peptide sequence data in the correct formatting
- Understand how to interpret a phylogenetic tree

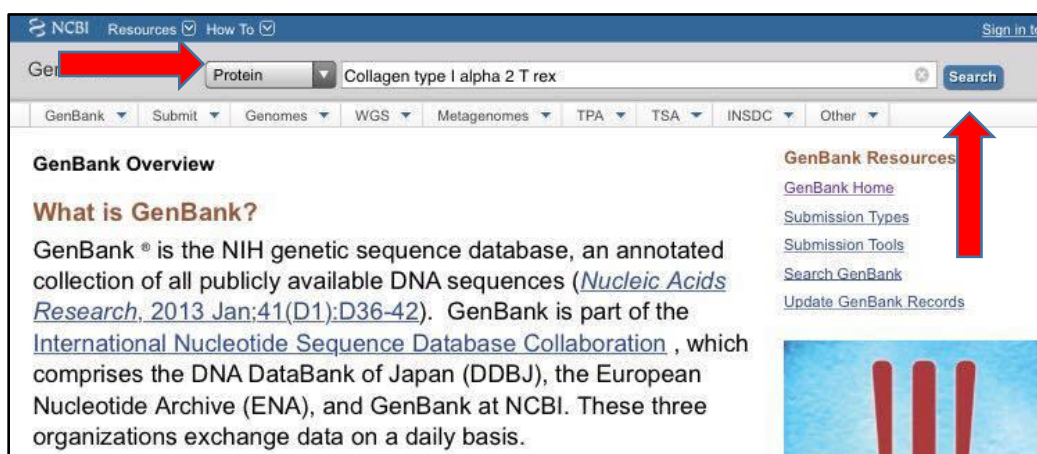
Procedure

Obtain *T. rex* amino acid sequence data:

1. Navigate to the following link: <http://www.ncbi.nlm.nih.gov>

(NCBI stands for the National Center for Biotechnology Information, which is a branch of the National Library of Medicine that hosts the GenBank database). The NCBI website is free for the public to access, and it contains libraries of genomic, genetic, and biomedical data. We will be using it to access protein sequences in GenBank. GenBank contains the sequences of many genes and their protein products, for hundreds of thousands of different species.

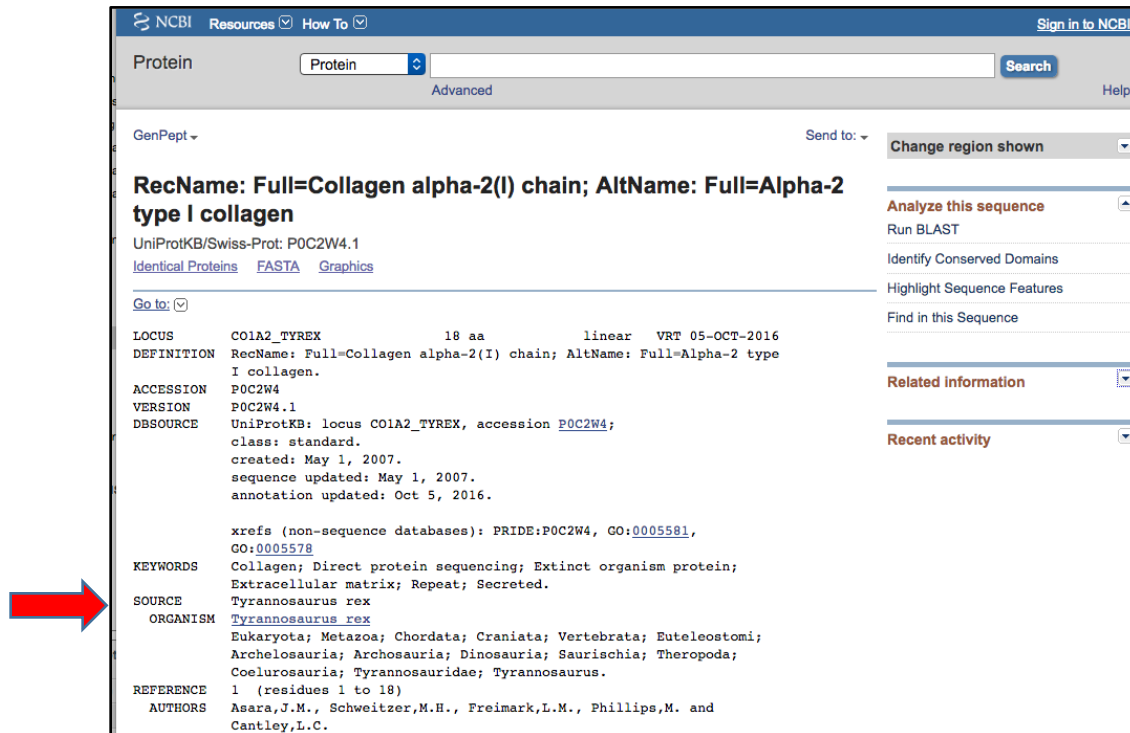
2. In the search bar at the top of the web page type in “collagen type I alpha 2 T rex” or “α2t1 collagen T rex” and select the protein database from the drop-down menu, do not include the quotations in the search (The “type I” is a capital I, not the number 1).



- Click on the blue Search button.
- It will provide a list with the top results relevant to your search. It should load 3 to 4 items, be sure to click on the result that says alpha-2(I) chain, **not** alpha-1(I).

The results display every known protein sequence that matches with the key words alpha 2, type 1, and *T. rex*. It will show results of things that do not precisely match your search, so be sure to fully read the names of the results. If you were to broaden your search to “alpha 2 collagen” it will result in hundreds of matches, rather than only three or four.

- Once you select the correct result, it will open up this page (pictured below). In order to make sure that you have selected the correct result, look at the column on the left hand side of the page. The fourth item down should say “source organism”, and the organism should be *Tyrannosaurus rex*. If it does not say it, hit the back button and retype the search query exactly as shown in Step 3.



NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

GenPept Send to: Change region shown

RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen

UniProtKB/Swiss-Prot: P0C2W4.1
[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS	COL1A2_TYREX	18 aa	linear	VRT 05-OCT-2016
DEFINITION	RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen.			
ACCESSION	P0C2W4			
VERSION	P0C2W4.1			
DBSOURCE	UniProtKB: locus COL1A2_TYREX, accession P0C2W4 ; class: standard. created: May 1, 2007. sequence updated: May 1, 2007. annotation updated: Oct 5, 2016.			
xrefs (non-sequence databases): PRIDE:P0C2W4, GO: 0005581 , GO: 0005578				
KEYWORDS	Collagen; Direct protein sequencing; Extinct organism protein; Extracellular matrix; Repeat; Secreted.			
SOURCE	Tyrannosaurus rex			
ORGANISM	Tyrannosaurus rex Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Tyrannosauridae; Tyrannosaurus.			
REFERENCE	1 (residues 1 to 18)			
AUTHORS	Asara,J.M., Schweitzer,M.H., Freemark,L.M., Phillips,M. and Cantley,L.C.			

Analyze this sequence
Run BLAST
Identify Conserved Domains
Highlight Sequence Features
Find in this Sequence

Related information
Recent activity

6. Once you reach the correct protein record page, click on the [FASTA](#) button underneath the protein's name in black bold writing.

The screenshot shows the NCBI protein record page for P0C2W4.1. The 'FASTA' button is circled in red. The page includes a search bar, a 'Send to' dropdown, and a 'Change region shown' dropdown. The main content area displays the protein name, UniProtKB/Swiss-Prot accession number, and various database links. The 'FASTA' button is located below the protein name and above the 'GenPept' link.

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

GenPept Send to: Change region shown

RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen

UniProtKB/Swiss-Prot: P0C2W4.1

[Identical Proteins](#) **FASTA** [Graphics](#)

Go to: (v)

LOCUS C01A2_TYREX 18 aa linear VRT 05-OCT-2016

DEFINITION RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen.

ACCESSION P0C2W4

VERSION P0C2W4.1

DBSOURCE UniProtKB: locus C01A2_TYREX, accession [P0C2W4](#); class: standard. created: May 1, 2007. sequence updated: May 1, 2007. annotation updated: Oct 5, 2016.

xrefs (non-sequence databases): PRIDE:P0C2W4, GO:[0005581](#), GO:[0005578](#)

KEYWORDS Collagen; Direct protein sequencing; Extinct organism protein; Extracellular matrix; Repeat; Secreted.

SOURCE Tyrannosaurus rex

ORGANISM [Tyrannosaurus rex](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Tyrannosauridae; Tyrannosaurus.

REFERENCE 1 (residues 1 to 18)

AUTHORS Asara,J.M., Schweitzer,M.H., Freemark,L.M., Phillips,M. and Cantley,L.C.

Analyze this sequence Run BLAST Identify Conserved Domains Highlight Sequence Features Find in this Sequence

Related information Recent activity

GenBank should then display a FASTA record page, like this:

The screenshot shows the NCBI FASTA record page for P0C2W4.1. The page includes a search bar, a 'Send to' dropdown, and a 'FASTA' button. The main content area displays the protein name, UniProtKB/Swiss-Prot accession number, and various database links. The 'FASTA' button is located below the protein name and above the 'GenPept' link.

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

FASTA Send to: (v)

RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen

UniProtKB/Swiss-Prot: P0C2W4.1

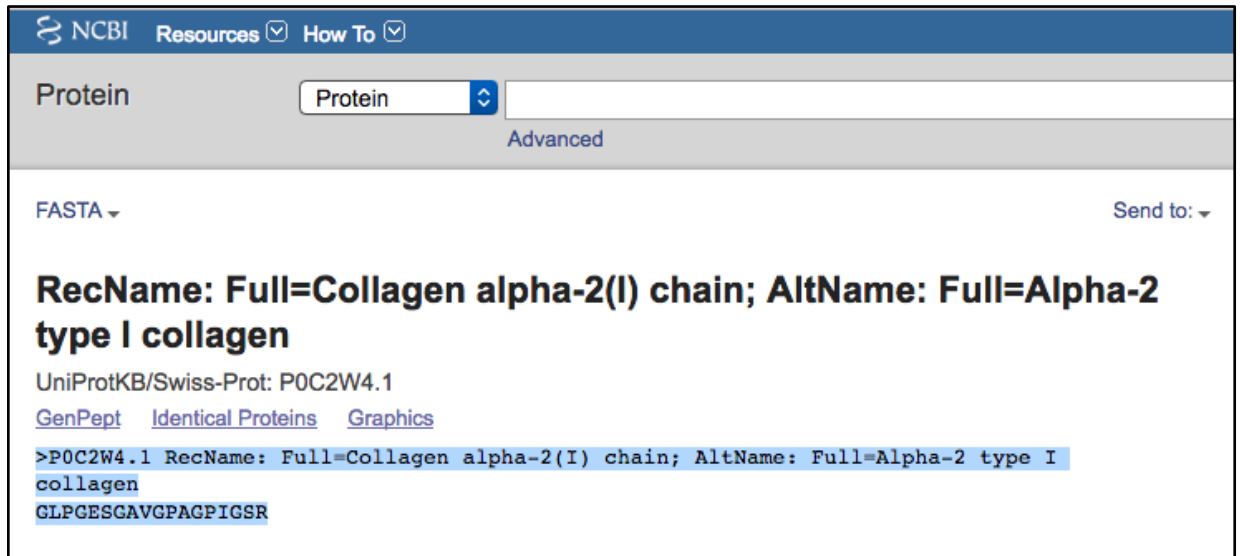
[GenPept](#) [Identical Proteins](#) [Graphics](#)

>P0C2W4.1 RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen

GLPGESGAVGPACPIGSR

"FASTA" (an abbreviation for "Fast-All") is the simple text-based file format that is often used to transmit DNA or amino acid sequences from one computer program to another. In a FASTA file, the DNA nucleotides or protein amino acids are represented by individual letter codes. The FASTA file format begins with a ">" (greater than) character followed by a description, which is then followed by lines of sequence data.

7. On the FASTA record page, select and **copy all of the text from the “>” all the way to the end of the amino sequence.**



NCBI Resources How To

Protein Protein Advanced

FASTA Send to:

RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen

UniProtKB/Swiss-Prot: P0C2W4.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

>P0C2W4.1 RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen
GLPGESGAVGPAGPIGSR

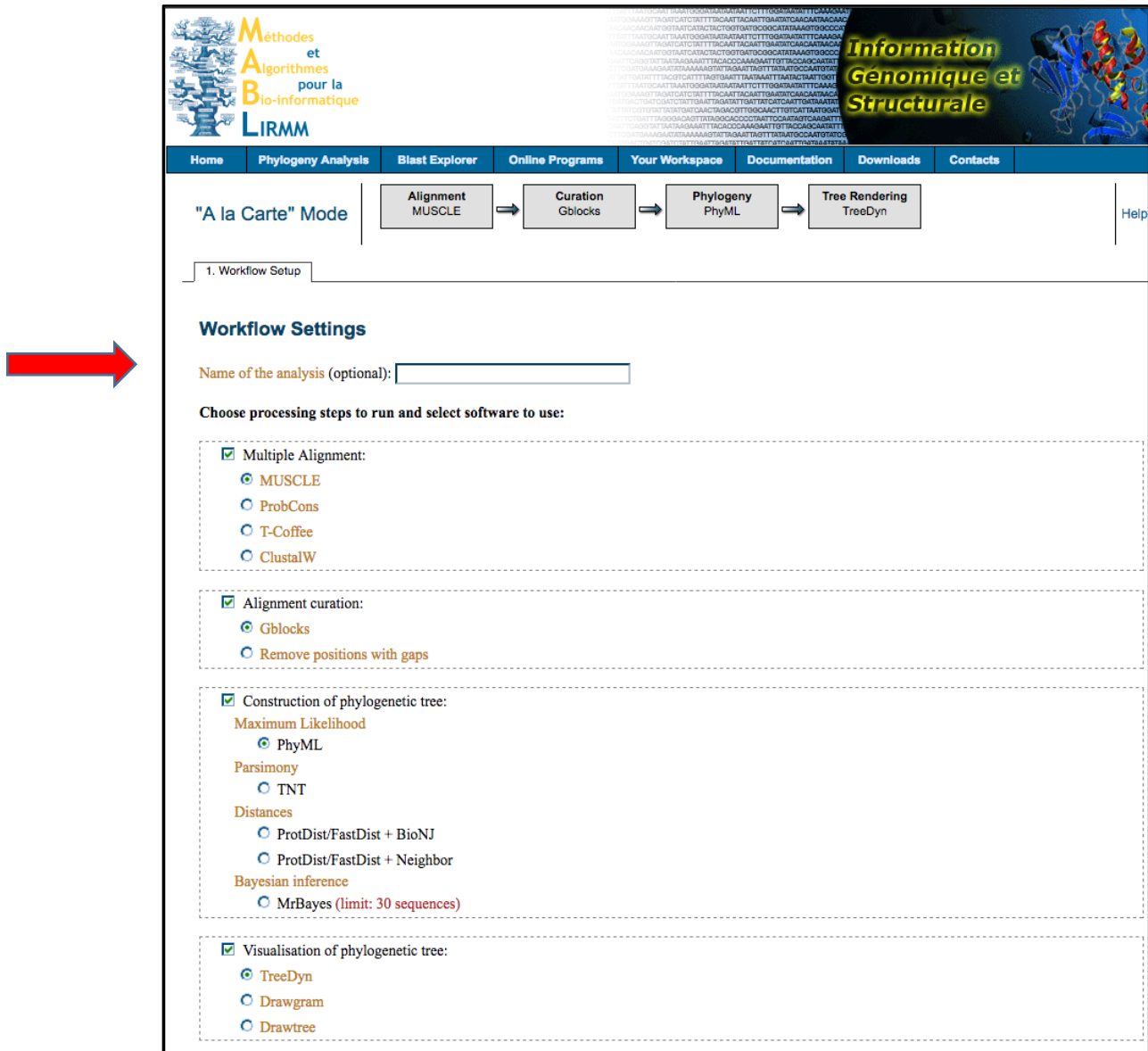
8. You are going to paste this sequence into the phylogeny building program in order to include the *T. rex* in your phylogenetic tree.

Constructing a phylogenetic tree using MAB:

1. With your web browser, navigate to the following web page: <http://www.phylogeny.fr/alacarte.cgi>

This page is the MAB (Methods and Algorithms for Bioinformatics) Phylogeny Analysis tool, which you will use to generate a phylogenetic tree. (Note: most of the MAB website is in French, but the form that you will use to run the Phylogeny Analysis tool is in English.)

2. This link will open up directly to “A la carte” mode. Under “Workflow Settings” insert a name for your analysis.



The screenshot shows the LIRMM website interface. At the top, there is a navigation bar with links: Home, Phylogeny Analysis, Blast Explorer, Online Programs, Your Workspace, Documentation, Downloads, and Contacts. Below this is a workflow diagram with steps: Alignment MUSCLE, Curation Gblocks, Phylogeny PhyML, and Tree Rendering TreeDyn. The 'A la Carte' mode is selected. The 'Workflow Settings' section is active, showing a 'Name of the analysis (optional):' field. Below this, there are four sections for selecting processing steps:

- ☒ Multiple Alignment:
 - ☒ MUSCLE
 - ☐ ProbCons
 - ☐ T-Coffee
 - ☐ ClustalW
- ☒ Alignment curation:
 - ☒ Gblocks
 - ☐ Remove positions with gaps
- ☒ Construction of phylogenetic tree:
 - Maximum Likelihood
 - ☒ PhyML
 - Parsimony
 - ☐ TNT
 - Distances
 - ☐ ProtDist/FastDist + BioNJ
 - ☐ ProtDist/FastDist + Neighbor
 - Bayesian inference
 - ☐ MrBayes (limit: 30 sequences)
- ☒ Visualisation of phylogenetic tree:
 - ☒ TreeDyn
 - ☐ Drawgram
 - ☐ Drawtree

3. Scroll to the bottom of the page and select “create workflow”. Do not change any of the settings, they are already set to the correct options for creating your phylogenetic tree.



The screenshot shows the bottom of the LIRMM website. The 'Run workflow:' section has two radio buttons: ☒ all at once and ☐ step by step. Below these is a button labeled 'Create workflow', which is circled in red.

4. MAB should open up to the second browser tab, “data and settings”. This tab gives you the option to upload your file or paste the sequence. Copy and paste the *T. rex* sequence into the big text box below "Input Data" in the MAB browser window. You need to change the description to say "T-rex". **Delete** “P0C2W4.1 RecName: Full=Collagen alpha-2(I) chain; AltName: Full=Alpha-2 type I collagen” and **replace** it with “T-rex”. (Be sure to leave the > otherwise it will not recognize the format. This step is important, now instead of the tree reading the full protein name, it will read the name of the animal.)

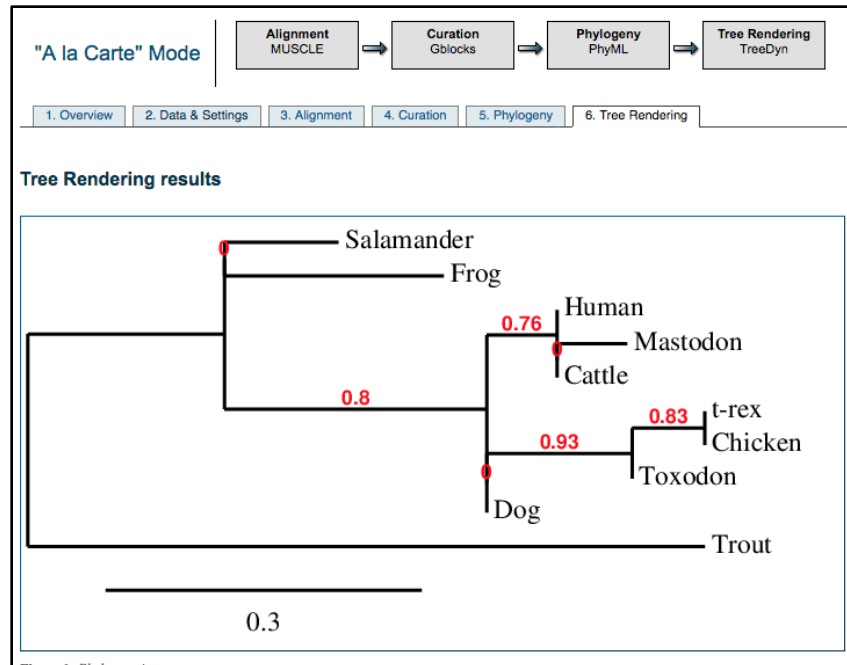
[illegible]

- Now you will need to paste the sequences for all of the other animals being compared in your phylogenetic tree. Scroll down to the end of this PDF and copy everything under the heading “Collagen Sequence Data (Copy and past everything below, including the “>”):”. Paste all of the sequences into the text box below the *T. rex* sequence.
- Scroll down to the bottom and enter your email address if you wish to be emailed your tree, if not select “submit”. Do not change any settings before hitting submit.

After clicking the Submit button, MAB will display a brief animation of a phylogenetic tree. During this time, MAB is aligning the sequences and then comparing them.¹

¹ This program uses a common method for aligning the sequences, called MUSCLE (Multiple Sequence Comparison by Log-Expectation). This step is important because it uses an algorithm to align each peptide sequence in order to accurately predict where mutations occurred that signal how the animals evolved. If the sequences are not aligned they can not be used to generate a phylogenetic tree.

7. MAB Phylogeny Analysis can take anywhere between 1 to 5 minutes to construct the phylogenetic tree. Once it loads, it should look like this:

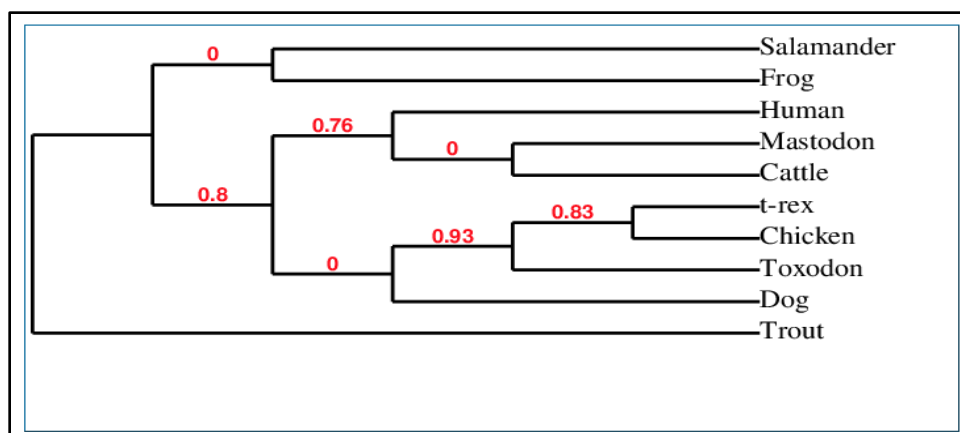


8. Scroll down to the "Tree Styles" section toward the bottom of the web page: Click on the radio button for "Cladogram" (in this context, "cladogram" is telling MAB to show a phylogenetic tree without scaling the length of tree branches based on degree of dissimilarity).

Tree style:

- ☒ Phylogram
- ☐ Cladogram (ignore branch lengths)
- ☐ Radial (by Drawtree)
- ☐ Radial (by TreeDyn)
- ☐ Circular

This will make it easier to read and understand the evolutionary relationships. The tree should now look like this:

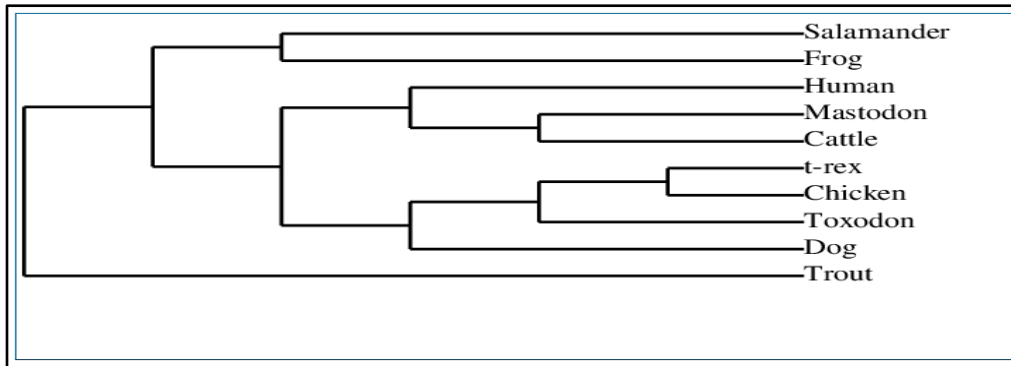


9. The final setting that needs to be adjusted is under “Display”, change the setting from “Branch support values” to “none”.

Display:

- ☒ Branch support values ☐ Branch lengths ☒ None color:
- ☐ Display branch support values in %
- ☒ Display legend

10. You now have your finished phylogenetic tree. It should look like this:



11. If you want to save your tree, you can click on the "PNG" or "PDF" links underneath the tree:

⇒ Download the tree: [PNG](#) - [PDF](#) - [SVG](#) - [TGF \(Treedyn format\)](#) - [Newick](#) - [Text](#)

Analyzing results

1. Which of the species that you analyzed, is the *T. rex* most closely related to? Does this match with the BLAST results from Session 2?
2. Which pairs of animal species are "sister species"? (i.e., which animals are most closely related?)
3. What species is the "out-group" (i.e., the least related to the rest of the species) in this phylogenetic tree?
4. Can you find anything puzzling with the relationships depicted in this phylogenetic tree? (hint, look at dog). Do you suppose this might reflect the fact that only a very short amino acid sequence from a single gene was analyzed?

Evaluating results

1. Why is it important to understand evolutionary relationships among animals?
2. Why is it important to learn more about extinct animals?
3. Why is it important for scientists to publish their findings, such as genetic sequences, in public databases?
4. What other questions could these same techniques be used to answer?

Collagen Sequence Data (Copy and past everything below, including the ">"):

>Human

MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTGPPGPPGPPG
PPGLGGNFAAQYDGKGVGLGPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEPGEPGQTGPAGARGPA
GPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPG
APGENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGAV
GNAGPAGPAGPRGEVGLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGA
TGARGLVGEPGPAGSKGESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPPGPPGLRGSPGSR
GLPGADGRAGVMGPPGSRGASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEGPVGLPG
IDGRPGPIGPAGARGEPPNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGV
QGGKGEQGPPGPPGFQGLPGPSGPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIG
SRGPSGPPGPDGNKGEPGVVGAVGTAGPSGPSGLPGERGAAGIPGGKGEKGEPLRGEIGNPGRDGA
RGAPGAVGAPGPAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGE
RGAKGPKGENGVVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPP
GPPGPAGKEGLRGPRGDQGPVGRTEVGAVGPPGFAGEKGPSGEAGTAGPPGTPGPQGLLGAPGILGL
PGSRGERGLPGVAGAVGEPGLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGNDGPPGRDGQP
GHKGERGYPGNIGPVGAAGAPGPHGPVGPAGKHGHRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKG
EPGEKGPRGLPGLKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRGTGHPGTVGP
AGIRGPQGHQGPAGPPGPPGPPGPPGVSGGGYDFGYDGDIFYRADQPRSAPSLRPKDYEVDATLKSINN
QIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIPAK
NWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLANYASQNITYHCKNSIAYMDEE
TGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGA
DQEFFVDIGPVCFK

>Chicken

MLSFVDTRILLLLAVTSYLATSQHLLFQASAGRKGPARGDKGPQGERGPPGPPGRDGEDGPPGPPGPPGPP

GLGGNFAAQYDPSKAADFPGPMGLMGPRGPPGASGPPGPPGFQGVPGEPGEPGQTGPQGPGRGPPGP
PGKAGEDGHPGKPRPGERGVAGPQGARGFPOTPGLPGFKGIRGHNGLDGQKGQPGTPGTKGEPGAP
GENGTPGQPGARGPLPGERGRIGAPGPAGARGSDGSAGPTGPAGPIGAAGPPGFPGAPGAKGEIGPAGN
VGPTGPAGPRGEIGLPGSSGPVGPPGNPGANGLPGAAGLPGVAGAPGLPGPRGIPGPPGPAGPSG
ARGLVGEPGPAGAKGESGNKGEPGAAGPPGPPGPSGEEGKRGSNGEPGSAGPPGPAGLRGVPGSRGL
PGADGRAGVMGPAGNRGASGPVGAKGPNGDAGRPGEPGLMGPRGLPGQPGSPGPAGKEGPVGFPGA
DGRVGPPIGPAGNRGEPGNIGFPGPKGPTGEPGKPGKGNVGLAGPRGAPGPEGNNGAQQPPGVTGNQ
GAKGETGPAGPPGFQGLPGPSGPAGEAGKPGERGLHGFEFGVPGPAGPRGERGLPGESGAVGPAGPIGS
RGPSGPPGPDGNKGEPGNVGPAGAPGPAGPGGIPGERGVAGVPGGKGEKGAPGLRGDTGATGRDGA
RGLPGAIGAPGPAGGAGDRGEGGPAGPAGPAGARGIPGERGEPGPVGPSGFAGPPGAAGQPGAKGER
GPKGPKGETGPTGAIGPIGASGPPGPVGAAGPAGPRGDAGPPGMTGFPGAAGRVPVGPAGITGPPGP
PGPAGKDGPRGLRGDVGVPVGRTEQGIAGPPGFAGEKGPSGEAGAAGPPGTPGPQGILGAPGILGLPG
SRGERGLPGIAGATGEPGPLGVSGPPGARGPSGPVGSPPNGAPGEAGRDGNPGNDGPPGRDGAPGF
KGERGAPGNPGPSGALGAPGPHGQVGPSGKPGNRGDPGPVGPVGPAGAFGPRGLAGPQGPGRGEKGEP
GDKGHRGLPGLKGHNGLQGLPGLAGQHGDQGPNGNNGPAGPRGPPGPSGPPGKDGRNGLPGPIGPA
GVRGSHGSQGPAGPPGPPGPPGPPGPNNGGGYEVGFDAEYYRADQPSLRPKDYEVDATLKTLLNNQIETLL
TPEGSKKNPARTCRDLRLSHPEWSSGFYWIDPNQGCTADAIRAYC
DFATGETCIHASLEDIPTKTWYVSKNPKDKKHIWFGETINGGTQFEYNGEGVTTKDMATQLAFMRLLAN
HASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELRAEGNSRFTFSVLVDGCSKKNNKWGKTIIEY
RTNKPSRLPILDIAPLDIGGADQEFGLHIGPVCFK

>Trout

MLSFVDNRILLLLAVTSLLASCQSGGLKGPRGAKGPRGDRGPQGPNGRDGKAGLPGIAGPPGPPGLGG
NFAAQFDGGKGSDPGPGPMGLMGSRGPNPPGAPGPQGFTGHAGEPGEPGQTGSIGARGPTGSAGKP
GEDGNNGRPGKPGDRGGPGTQGARGFPOTPGLPGMKGHRGYNGLDGRKGESGTAGAKGETGAHGA

NGSPGPAGSRGLNGERGRAGPAGPAGARGADGSTGPAGPAGPLGAAGPPGFPGAPGPKGEIGGAGSN
GPSGPQGGRGEPGINGAVGPVGPVGNPGNNGINGAKGAAGLPGVAGAPGFPGPRGGPGPQGPQGST
GARGLGGDPPGSGQKGDGSAKGEPGHSGVQGAAGPAGEEGKRGSTGEVGATGPAGLRGARGGAGTR
GLPGLEGRGGPIGMPGARGATGPGGIRGAPGDAGRAGESGLTGARGLPGNSSGQGGPPGKEGPPGAAG
LDGRTGPPGPTGPRGQPGNIGFPGPKGPGGEAGKGGDKGPTGATGLRGGPGADGNNGAPGPAGVVG
NTGEKGEQGPAGAPGFQGLPGPAGPAGEAGKAGNQGMPGDQGLPGPAGVKGERGNSGPAGSAGSQG
AIGARGPAGTPGPDGGKGEPGSGVIVGAAGHQGPGGMPGERGAGGTPGPKGEKGEGGHTRGLGNMG
RDGARGGPGPSGPPGPSGANGEKGESGSFGPAGPAGLRGPSGERGEGGPAGLPGFAGPPGSDGQSGP
RGEKGPAGGKGDVGPAGPAGPSGQSGPSGASGPAGPPGGRGDAGPSGLTGFPGAAGR VGGPGPAGI
AGPPGSAGPAGKDGPRGLRGDPGPGGPQGEQGVVGPAGISGDKGPSGESGPPGAPGTAGPQGVLGPS
GFVGLPGSRGDKGLPGGPGAVGEPGRLGPAGASGPRGPAGNIGMPGMTGTQGEAGREGNSGNDGPP
GRPGAAGFKGDRGEPGSPGALGSSGQPGPNGPAGSAGRPGNRGESGPTGNGGPVGAVGARGAPGPA
GPRGEKGGAGEKGDRGMKGLRGHGGLQGMPGPNGPSGETGSAGITGPAGPRGPAGPHGPPGKDGRA
GGHGAIGPVGHRGSPGHLGPAGPPGSPGLPGPAGPAGGGYDQSGGYDEYRADQPSFRAKDYEVDATI
KSLNSQIENLLTPEGSKKNPARTCRDIRLSHPDWSSGFYWIDPNQGGCIADAIKAYCDFSTGHTCIHPHE
SIARKNWYRSSSENKKHVWFGETINGGTEFAYNDETLSPQSMATQLAFMRLLANQATQNITYHCKNSVA
YMDGENGNLKKAVLLQGSNDVELRAEGNSRFTFNVLEDGCTRHTGQWSKTVIEYRTNKPSRLPILDIAP
LDIGEADQEFGLDIGPVCFK

>Dog

MLSFVDTRTLLLLAVTSCLATCQSLQEATARKGPTGDRGPRGERGPPGPPGRDGGDIPGPPGPPGPPG
PPGLGGNFAAQYDGKGVGLGPGPMGLMGRGPPGASGAPGPQGFQGPAGEPGEPGQTGPAGARGPPG
PPGKAGEDGHPGPKPRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGA
PGENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGPVG
NPGPAGPAGPRGEVGLPGVSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGAT
GARGIVGEPGPAGSKGESGNKGEPGSAGAQQPPGPSGEEGKRGPNGEAGSAGPSGPPGLRGSPGSRG

LPGADGPAGVMGPPGPRGATGPAGVRGPNGDSGRPGEPGLMGPRGFPGAPGNVGPAGKEGPMGLPGI
DGRPGPIGPAGARGEPNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGV
QGGKGEQGPAGPPGFQGLPGPAGTAGEVGKPGERGLPGEFGLPGPAGPRGERGPPGESGAAGPSGPIG
SRGPSGPPGPDGNKGEPVGLGAPGTAGASGPGGLPGERGAAGIPGGKGEKGETGLRGEIGNPGRDGA
RGAPGAMGAPGPAGATGDRGEAGPAGPAGPAGPRGTPGERGEVGPAGPNGFAGPAGAAGQPGAKGE
RGTKGPKGENGPVGPTGPIGSAGPSGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPSGITGPPG
PPGAAGKEGLRGPRGDQGPVGRRTGETGASGPPGFTGEKGPSGEPGTAGPPGTPGPQGLLGAPGILGLP
GSRGERGLPGVAGSVGEPGPLGIAGPPGARGPPGAVGAPGVNGAPGEAGRDGNPGNDGPPGRDGQA
GHKGERGYPGNIGPVGAVGAPGPHGPVGPTGKHGNRGEPPGAGSVGPVGAVGPRGPSGPQGIRGDKG
EPGEKGPRGLPGLKGHNGLQGLPGLAGQHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGQP GTVG
PAGIRGSQGSQGPAGPPGPPGPPGPPGPSGGGYDFGYEGDFYRADQPRSPPSLRPKDYEVDATLKSLN
NQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIP
AKNWYRNSKVKKHIWLGETINGGTQFEYNVEGVTTKEMATQLAFMRLLANHASQNITYHCKNSIAYMD
EETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWRKTIIEYKTNKPSRLPILDIAPLDIGD
ADQEFRVDVGPVCFK

>Cattle

MLSFVDTRTLLLLAVTSCLATCQSLQEATARKGPSGDRGPRGERGPPGPPGRDGGDGIPGPPGPPGPPG
PPGLGGNFAAQFDAKGGGPGPMGLMGPRGPPGASGAPGPQGFQGPPEPGEQGTPAGARGPPGPP
GKAGEDGHPGKPRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPG
ENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGELGPVGN
PGPAGPAGPRGEVGLPGLSGPVGPPGNPGANGLPGAAGAAGLPVAGAPGLPGRGIPGPVGAAGATG
ARGLVGEPGPAGSKGESGNKGEPGAVGQPGPPGPSGEEGKRGSTGEIGPAGPPGPPGLRGNPGSRGLP
GADGRAGVMGPAGSRGATGPAGVRGPNGDSGRPGEPGLMGPRGFPGSPGNIGPAGKEGPVGLPGIDG
RPGPIGPAGARGEPNIGFPGPKGPSGDPGKAGEKGHAGLAGARGAPGPDGNNGAQGPPGLQGVQGG

KGEQGPAGPPGFQGLPGPAGTAGEAGKPGERGIPGEFGLPGPAGARGERGPPGESGAAGPTGPIGSRG
PSGPPGPDGNKGEPGVVGAPGTAGPSGPSGLPGERGAAGIPGGKGEKGETGLRGDIGSPGRDGARGA
PGAIGAPGPAGANGDRGEAGPAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGTK
GPKGENGPVGPTGPVGAAGPSGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPSGISGPPGPPGP
AGKEGLRGPRGDQGPVGRSGETGASGPPGFVGEKGPSGEPGTAGPPGTGPQGLLGAPGFLGLPGSRG
ERGLPGVAGSVGEPGPLGIAGPPGARGPPGNVGNPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKG
ERGYPGNAGPVGAAGAPGPQGPVGPVGKHGNGRGEPPAGAVGPAGAVGPRGPSGPQGIRGDKGEPPG
DKGPRGLPGLKGHNGLQGLPGLAGHHGDQGAPGAVGPAGPRGPAGPSGPAGKDGRIGQPGAVGPAGI
RGSQGSQGPAGPPGPPGPPGPPGPSGGGYEFGFDGDFYRADQPRSPTSLRPKDYEV DATL KSLNNQIE
TLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPEDIPVKNW
YRNSKAKKHVWVGETINGGTQFEYNVEGVTTKEMATQLAFMRLLANHASQNITYHCKNSIAYMDEETG
NLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWQKTIIEYKTNKPSRLPILDIAPLDIGGADQ
EIRLNIGPVCFK

>Frog

MLSFVDLRSVLLLA VTLYLVTCQEVRRGPRGDKGPPGEQGPPGIPGRDGEDGLPGLPGPPGVPGLGGNF
AAQYDPSKSAEPGQQGIMGPRGPPGPPGSPGSQGFQGLPGENGEPGQTGPVGSRGPSGAPGKAGEDG
HPGKSGRPGERGPVGPQGARGFPGTPGLPGFKGIRGHTGSDGQKGAPGAAGVKGENGANGDNNGSPG
QAGARGLPGERGRIGPAGSAGSRGSDGSSGPVGPAGPIGSAGAPGLPGAPGAKGELGPAGNNGPTGA
AGGRGEPGPPGSLGPAGPPGNPGTNGVNNGAKGTAGLPGVGGAPGLPGGRGIPGPAGPAGPSGARGLA
GDPGIAGGKGDTGSKGEPGSGVQQQGPAGPSGEEGKRGPNGEAGSSGPSGNAGIRGVPGTRGLPGPD
GRAGGIGPAGSRGSSGPPGARGPNGDAGRPGEPLLGARGLPGFSGSNGPQGKEGPAGPQGIEGRSG
AAGPAGARGEPPGAIGFPGPKGPNGEPGKNGDKGNQGPSGNRGAPGPDGNNGAQQPAGLGGATGEKG
EQGPSGAPGFQGLPGPGGPPGEVGKPGERGAPGDFGPPGSAGTRGERGAPGESGGAGPHGPSGSRGP
SGAPGPDGQKGEPGAAGLNGGLGPSGPAGIPGERGTAGTPGTKGEKGDAGNSGDYGNPGRDGARGP
AGAAGAPGPAGGPGDRGESGPAGPSGVAGPRGAPGERGEAGPAGPTGFAGPPGAAGHTGAKGDRGA

KGPKGEAGSPGPLGAHGSAGPAGPNGPAGSTGARGDAGPSGATGFPGPAGRAGAPGPPGNVGPSPGT
GHPGKDGSRGPRGDSGPVGRPGEQGQHGPVGLAGDKGPSGEAGPAGPPGAAGPSGVLGARGILGLP
GTRGERGLPGGPGSNGEPGPSGLAGSSGPRGPPGSVGSPPGVGHSGEAGRDGHGPGNDGPPGRDGLP
GAKGERGYPGNTGPSGLAGAPGPAGSAGPAGKSGNRGEGGPSGPAGITGPSGPRGPAGPQGVRGDKG
EAGERGARGLDGRKGHNGLSGLPGPSGTPGETGPSGSVGPVGPGRGPSGPSGPPGKEGRSGHHPGAMGP
VGPRGPAGFTGPAGPPGPPGPPGHAGPSGGGYDGGDGGGEYYRADQPERKPKDYEVDATLKSLNQQIEV
ILTPEGSRKNPARTCRDLRLSHPEWTSGFYWIDPNQGCTSDAIRVFCDFSSGETCIHANPDEITQKNWY
INTSNKDKKHLWFGEILNGGTQFEYHDEGLTAKDMATQLAFMRLLANQASQNITYHCKNSIAYMDEET
GNLKKAVILQGSNDVELRAEGNTRFTYSVLEDGCTKHTGEWGKTVIEYRTNKPSRLPI
LDIAPLDIGGHDQEIGFEIGPVCFK

>Toxodon

GPMGIMGPRGPPGASGAPGPAGEPGEPGQTGPAGARGPPGPPGKAGEDGHPGKPRPGERGVVGPQG
ARGFPGTPGIPGFKGIRGHNGIDGIKGQPGAPGVKGEPGAPGENGTPGQAGARGIPGERGRVGAPGA
GARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGPVGNPGPAGPAGPRGEVGIPGVSGPVGP
PGANGITGAKGAAGIPGVAGAPGIPGPRGIPGPVGAAGATGARGIVGEPGPAGSKGESGNKGEPGSAG
PQGPPGPAGEEGKRGPNGEAGSTGPTGPPGIRGSRGIPGADGGSRGATGPAGVRGDSGRPGEPGIMG
PRGFPGSPGNIGPAGKEGPVGIPGIDGRPGPTGPAGARGEPPGNIGFPGPKGPTGDPGKNGDKGHAGIA
GARGPAGPPGFQGIPGPAGTAGEVGKPGERGIPGEFGIPGPAGARGERGPPGESGAVGPAGPIGSRGPS
GPPGPDGNKGEPGNIGAIGTAGPSGPSGIPGERGAAGIPGGKGEKGETGIRRGAPGAIGAPGPAGANG
DRGEAGPAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGTKGPKGENGPVGPTGP
VGAAGPAGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPAGITGPPGPPGAAGKEGIRGPRGDQ
GPVGRSGETGASGIPGFAGEKGPAGEPGTAGIPGTGPQGIIGAPGIIGIPGSRGERGIPGVAGSIGEPG
PIGIAGPPGARGPPGAVGNPGVNGAPGEAGRHNHNRGEPGPAGSVGPAGAVGPRGPSGPQGIRGDKGE
PGDKGPRGIPGIKGHNIGIQQIPGIAGQHGDQGAPGAVGPAGPRGPAGPSGPAGKDGRIGHPGTVGPA

GIRGSQGSQGPAGPPGPPGPPGPPGPS

>Mastodon

QYDAKGVGLGPGPMGLMGPRGPPGATGPPGSPGFQGGPEGPEGQTGPAGSRGPAGPPGKAGEDGH
PGKPGRPGERGVVGPQQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQI
GARGLPGERGRVGGPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGPVGNPGPSGPAGP
RGEAGLPGVSGPVGPPGNPGANGLAGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGARGIVGEPG
PAGSKGESGSKGEPGSAGPQGGPPGPSGEEGKRGPNGEAGSAGPAGPPGLRGGPGSRGLPGADGRAGV
MGPPGSRGASGPAGVRGPSGDSGRPGEPGVMGPRGLPGSPGNVGPAGKEGPAGLPIDGRPGPIGPA
GARGEPPNIGFPGPKGPAGDPGKNNGDKGHAGLAGPRGAPGPDGNNGAQGGPGLQGVQGGKGEQGPA
GPPGFQGLPGPSGTAGEAGKPGERGIPGEFGLPGPAGPRGERGPPGQSGAAGPTGPIGSRGPSGPPGP
DGNKGEPGVVGAPGTAGPSGPVGLPGERGAAGIPGGKGEKGETGLRGDTGNTGRDGARGAPGAVGAP
GPAGATGDRGEAGPAGSAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQAGAKGERGTKGPKGEN
GPVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPAGITGPPGPPGAAGKEGL
RGPRGDQGPVGRGTGETGASGPPGFAGEKGSSGEPGTAGPPGAPGPQGILGPPGILGLPGSRGERGLPG
VAGAVGEPGLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGSDGPPGRDGLPGHKGERGYPGN
AGPVGTAGAPGPQGPLGPAGKHGNGRGEPPGAGSVGPVGAVGPRGPSGPQGARGDKGEAGDKGPRGL
PGFKGHNGLQGLPGLAGQHGDQGSPPSVGPAGPRGPAGPSGPVGKDGRPGHAGAVGPAGVRGSQGS
QGPSGPPGPPGPPGPPGPPGPSGGGYDFGYDGDIFYRA

>Salamander

MLSFVDTRIVLLLAVTSSLATCQYNYEANRGPRGYKGPQGDQGGPPGAPGRDGVDPGPPGPAGPPGPPGP
SGLGGNFAAQYDGGKSDPGPGPMGMMGPRGPPGPSGSPGAQGFQGLPEGPEGQTGPVGSRGPTG
PPGKSGEDGSPGKSGRPGERGTVGTQGARGFPGTPGLPGFKGLRGHNGFDGVKGAAGSQQAKGETG
ANGENGSPGQAGARGLPGERGRVGGAGPGARGSDGSAGPSGPAGPIGSAGAPGLPGAPGAKGEIGSA
GNNGPSGPAGSRGDPGLPGSVGPVGPAGNPGSNGVSGAKGAAGLPGVGGAPGLPGPRGIPGPQGASG
AAGARGLAGDPGSPGGKGDGSGSKGEPGSAGQQGNAGPSGEEGKRGPNGEPGSSGPAGPAGIRGVPG

TRGLPGPDGRAGGMGPPGSRGSSGPAGVRGPSGDAGRPGEPGLLGQRGLPGFPGNTGPVGKEGPAGP
AGIEGRTGAAGPTGARGEPSIGFPGPKGPGGEPGKNGDKGSAGPSGARGAPGPDGNNGAQGPPGVV
GNTGEKGEQGPAGAPGFQGLPGPGGAAGEAGKVGDRGMPGDFGPPGPAGVRGERGAPGESGSAGPL
GPVGSRGPSGPPGPDGTKGEPGVAGLAGAVGPSGSGGSPGERGGAGTPGPKGEKGEAGNRGEYGNQ
GRDGARGPAGASGAPGPSGGPGDRGESGPSGPAGPAGSRGAPGERGEHGPGGPTGFGGPPGAAGHT
GVKGERGEKGPKGELGPQGPVGASGASGPAGPNGPAGAPGSRGEVGPAGATGFPGPAGRTGGPGPAG
MGGPPGPSGHAGKDGPRGPRGDSGPVGRPGEQGGGLGPQGISGEKGPSGEPGTAGPPGSSGPSVGLG

ARGILGLPGTRGERGLPGGPGGNGEPGATGPTGTAGSRGAPGPVGSAGMNGPAGEAGRDGNPNDGP
PGRDQQAGAKGERGYPGNTGGVGHAGAPGPHGSVGPAGKSGNRGEPGPSGSQGPAGLPGARGPAGP
AGSRGDKGESGEKGGRGLDGRKGHNGLQGLPGLPGTSGEAGSAGPSGPSGPRGPAGPSGPPGKDGH
SGQPGPVGPAGVRGSPGHQGPAGPPGSPGAPGPAGPSGGGYDGGFEGGEFYRADQPSLRPKDYEVD
TLKTLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGFYWIDPNQGCTADAIRVYCDFSTGETCIHSN
PETISAKTSYVNKNPKDKKHVVWGEVLNGGTQFEYNEEGVTTKDMATQFAFMRLLANHASQNITYHCK
NSIAYMDGETGNLKKAVLLQGSNDVELRAEGNSRFTFSVLEDSCTKHTGEWGRTVMEYRTNKPSRLPIL
DIAPMDIGGAEQEFRVDIGPVCFK