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

[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
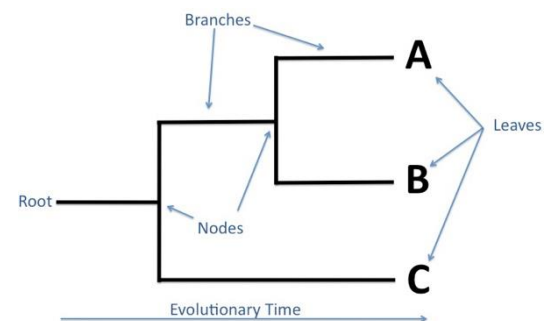

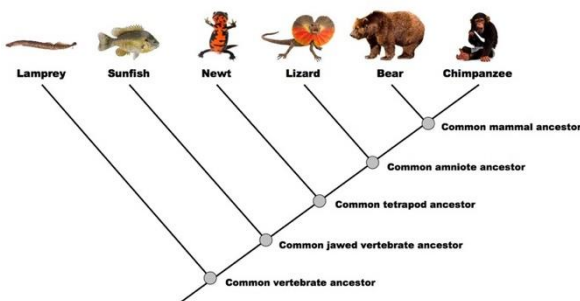Figure 1: This is an example of a phylogenetic tree showing how to read it.

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 <u>sister clades</u>, since they are the most closely related. Group C would be considered the <u>outgroup</u> 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 (*Mammut 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.
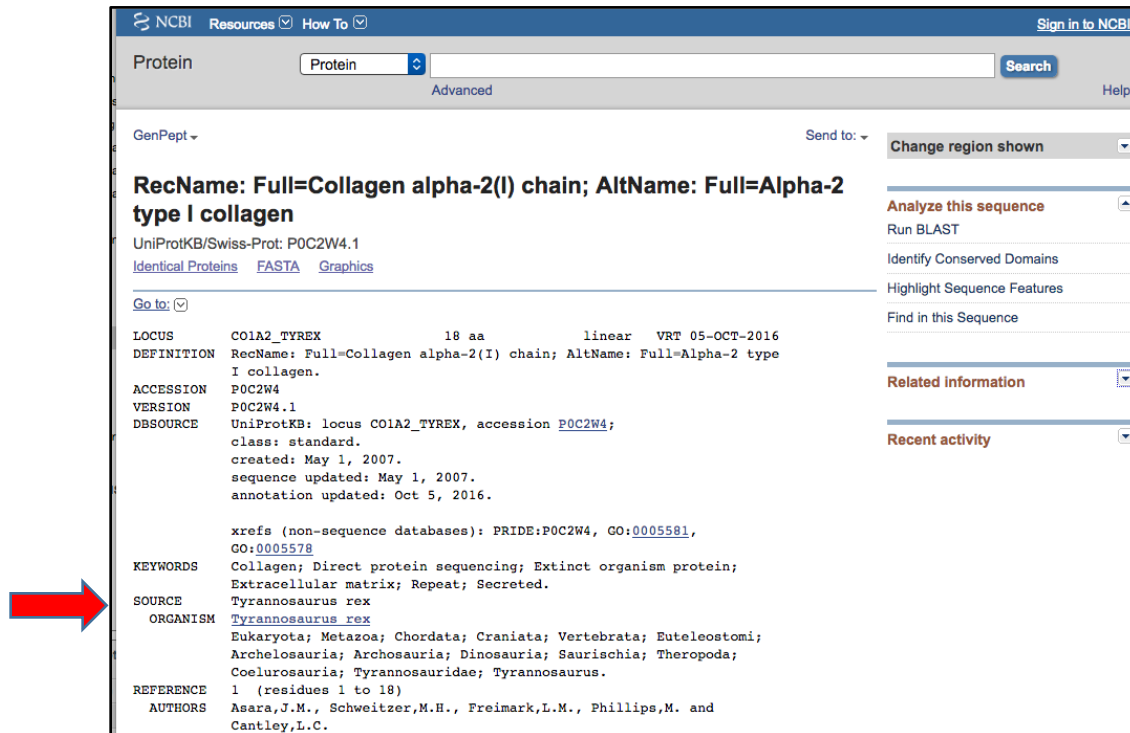
2. In the search bar at the top of the web page type in "collagen type I alpha 2 T rex" or "α2t1 collagen T rex" and select the protein database from the drop-down menu, do not include the quotations in the search (The "type I" is a captial I, not the number 1).
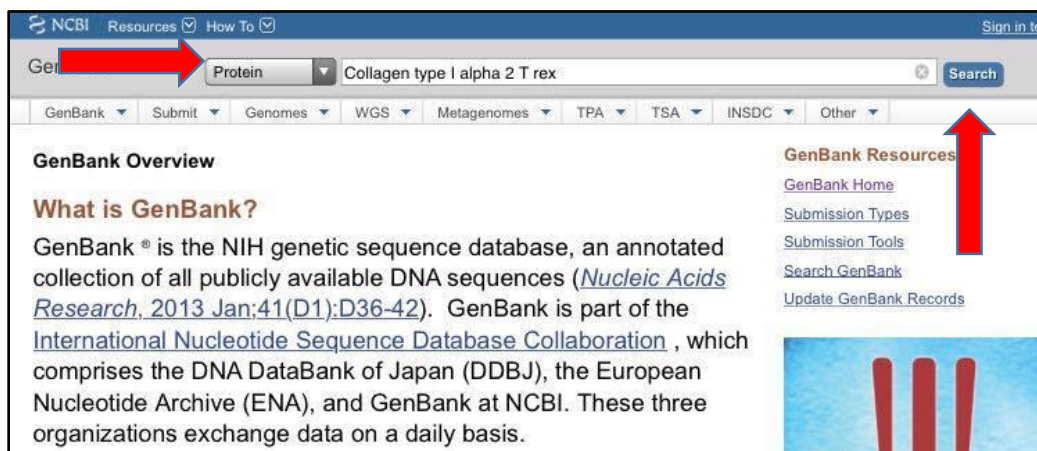
3. Click on the blue Search button.
4. It will provide a list with the top results relevant to your search. It should load 3 to 4 items, be sure to click on the result that says alpha-**2**(I) chain, **not** alpha-1(I).

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.

5. 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.

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

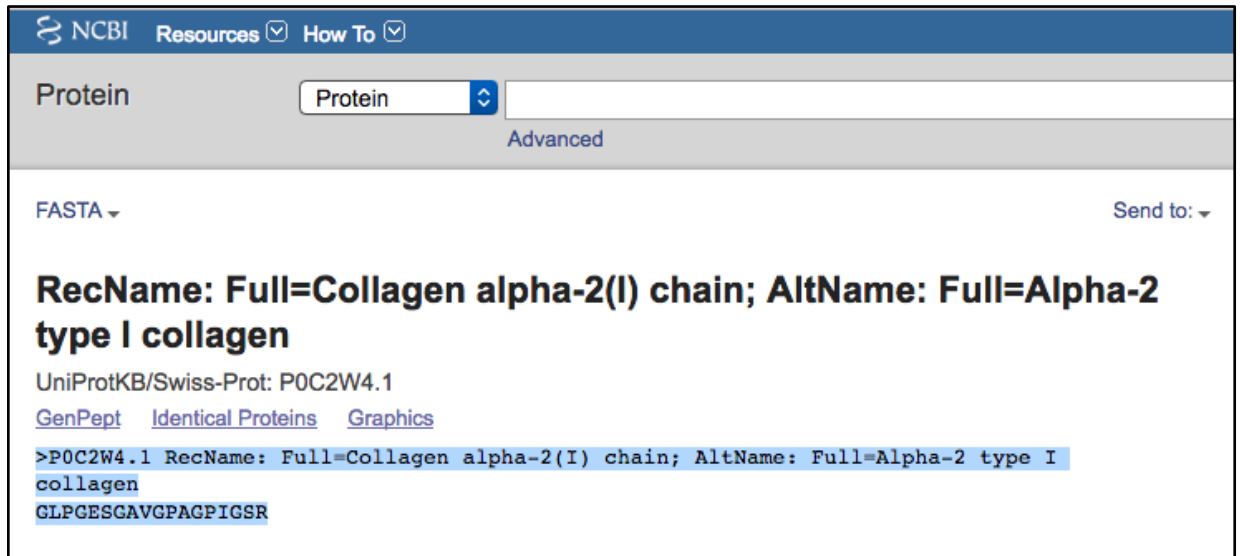GenBank should then display a FASTA record page, like this:

"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.**

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.

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

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.)

5. 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.

6. Scroll down to the bottom and enter your email address if you wish to be emailed your tree, if not select "submit". Do <u>not</u> 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:

Figure 1: Phylogenetic tree
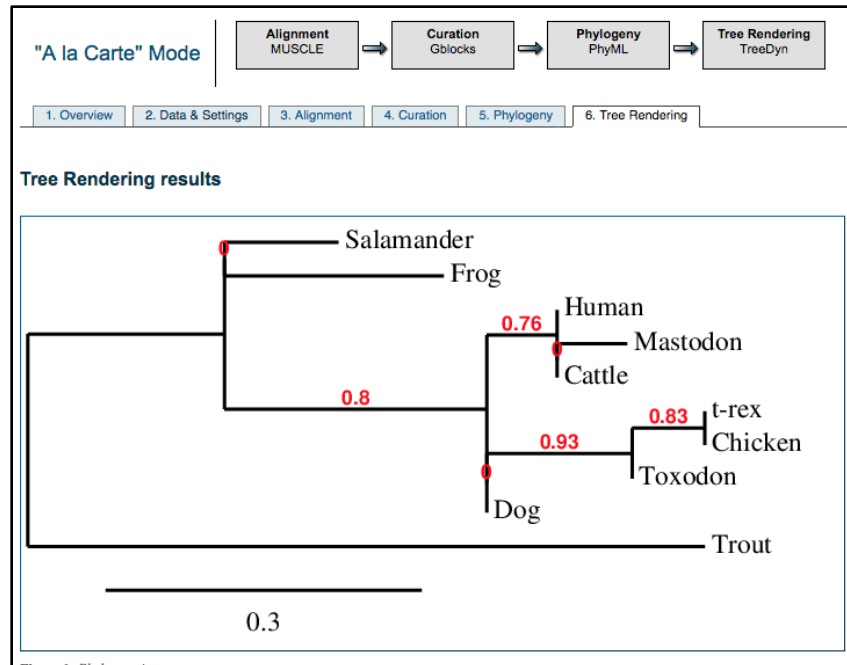
8. Scroll down to the "Tree Styles" section toward the bottom of the web page: Click on the radio button for "Cladogram" (in this context, "cladogram" is telling MAB to show a phylogenetic tree without scaling the length of tree branches based on degree of dissimilarity).

This will make it easier to read and understand the evolutionary relationships. The tree should now look like this:
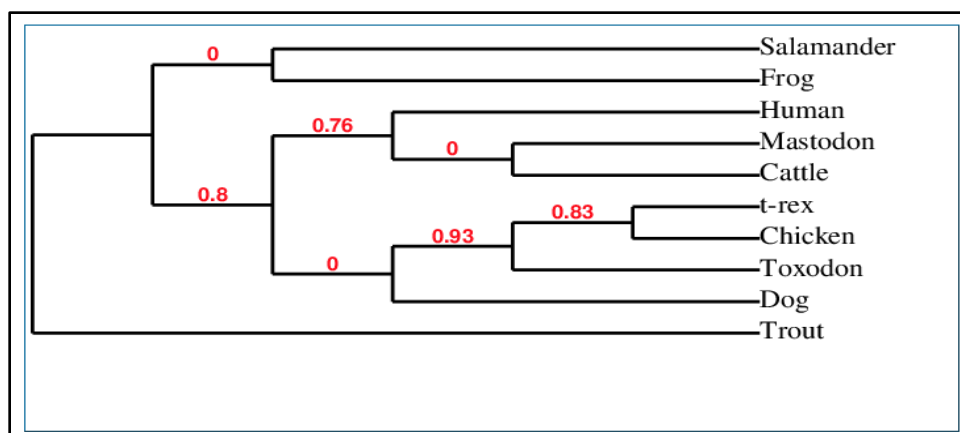
9. The final setting that needs to be adjusted is under "Display", change the setting from "Branch support values" to "none".

Display:
- ◉ Branch support values ○ Branch lengths ○ None   *color:* red
- ☐ Display branch support values in %
- ☑ Display legend

10. You now have your finished phylogenetic tree. It should look like this:
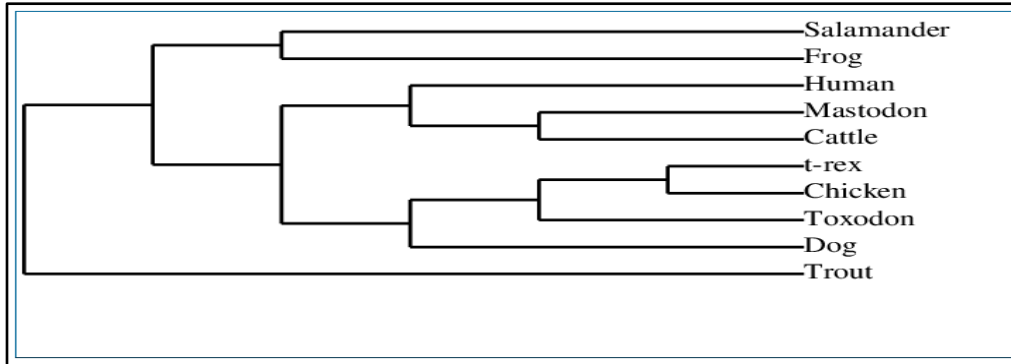
11. If you want to save your tree, you can click on the "PNG" or "PDF" links underneath the tree:

==> 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?
    Chicken ("bird"); this should be the primary match that appeared on BLAST.

2. Which pairs of animal species are "sister species"? (i.e., which animals are most closely related?)

    T. rex and chicken; Mastodon and cattle; salamander and frog.

3. What species is the "out-group" (i.e., the least related to the rest of the species) in this phylogenetic tree?
    Trout; this species is least related to the other species.

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?

Some things students might find puzzling are: humans evolving before mastodon, even though the mastodon has been extinct for several thousand years; and the fact that chicken evolved from mammals (the dog and toxodon came before the chicken), implying that the birds evolved from mammals. However, birds evolved long before mammals did. This is due to the very short amino acid sequence that was recovered from the *T rex*; most amino acid sequences are thousands of amino acids long, rather than 18. If a more complete protein was recovered, the tree would be more precise.

Evaluating results

1. Why is it important to understand evolutionary relationships among animals?

   It is important to understand phylogenetics because it is the study of the classification of life. It explains how evolution of organisms has occurred, and has expanded our knowledge on the animals we share this planet with. The more about other animals we learn, the more we learn about the evolution of humans.

2. Why is it important to learn more about extinct animals?

   Learning the biology and evolution behind extinct animals can help us learn more about how present day animals evolved. By sequencing the *T rex* protein, we have scientifically proven that birds evolved from dinosaurs. As we learn more about the

3. Why is it important for scientists to publish their findings, such as genetic sequences, in public databases?

   If scientists do not publish their findings, then other scientists would have to go through the process of sequencing an animal's genome again (which can cost thousands of dollars) in order to get the same results. Sharing data saves money, time, and can help build libraries of genetic data for scientists and the public to use access.

4. What other questions could these same techniques be used to answer?

   There are a variety of answers students can give here. Here are a few examples that I think are interesting uses of molecular data analysis: using these same methods, unknown animals and insects can be identified, along with its ancestry; when scientists find a new species that has never been described before, they can use bioinformatics to analyze the genome of the animal and identify what it is most closely related to in order to properly name and place it among all other animals; this process also works to identify organisms in the field if an animal/plant specialist is not available.

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

>Human

MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTGPPGPPGPPG

PPGLGGNFAAQYDGKGVGLGPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEPGEPGQTGPAGARGPA

GPPGKAGEDGHPGKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPG

APGENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGAV

GNAGPAGPAGPRGEVGLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGA

TGARGLVGEPGPAGSKGESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPPGPPGLRGSPGSR

GLPGADGRAGVMGPPGSRGASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEGPVGLPG

IDGRPGPIGPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGV

QGGKGEQGPPGPPGFQGLPGPSGPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIG

SRGPSGPPGPDGNKGEPGVVGAVGTAGPSGPSGLPGERGAAGIPGGKGEKGEPGLRGEIGNPGRDGA

RGAPGAVGAPGPAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGE

RGAKGPKGENGVVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPP

GPPGPAGKEGLRGPRGDQGPVGRTGEVGAVGPPGFAGEKGPSGEAGTAGPPGTPGPQGLLGAPGILGL

PGSRGERGLPGVAGAVGEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGNDGPPGRDGQP

GHKGERGYPGNIGPVGAAGAPGPHGPVGPAGKHGNRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKG

EPGEKGPRGLPGLKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVGP

AGIRGPQGHQGPAGPPGPPGPPGPPGVSGGGYDFGYDGDFYRADQPRSAPSLRPKDYEVDATLKSLNN

QIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIPAK

NWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLANYASQNITYHCKNSIAYMDEE

TGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGA

DQEFFVDIGPVCFK

>Chicken

MLSFVDTRILLLLAVTSYLATSQHLFQASAGRKGPRGDKGPQGERGPPGPPGRDGEDGPPGPPGPPGPP

GLGGNFAAQYDPSKAADFGPGPMGLMGPRGPPGASGPPGPPGFQGVPGEPGEPGQTGPQGPRGPPGP

PGKAGEDGHPGKPGRPGERGVAGPQGARGFPGTPGLPGFKGIRGHNGLDGQKGQPGTPGTKGEPGAP

GENGTPGQPGARGLPGERGRIGAPGPAGARGSDGSAGPTGPAGPIGAAGPPGFPGAPGAKGEIGPAGN

VGPTGPAGPRGEIGLPGSSGPVGPPGNPGANGLPGAKGAAGLPGVAGAPGLPGPRGIPGPPGPAGPSG

ARGLVGEPGPAGAKGESGNKGEPGAAGPPGPPGPSGEEGKRGSNGEPGSAGPPGPAGLRGVPGSRGL

PGADGRAGVMGPAGNRGASGPVGAKGPNGDAGRPGEPGLMGPRGLPGQPGSPGPAGKEGPVGFPGA

DGRVGPIGPAGNRGEPGNIGFPGPKGPTGEPGKPGEKGNVGLAGPRGAPGPEGNNGAQGPPGVTGNQ

GAKGETGPAGPPGFQGLPGPSGPAGEAGKPGERGLHGEFGVPGPAGPRGERGLPGESGAVGPAGPIGS

RGPSGPPGPDGNKGEPGNVGPAGAPGPAGPGGIPGERGVAGVPGGKGEKGAPGLRGDTGATGRDGA

RGLPGAIGAPGPAGGAGDRGEGGPAGPAGPAGARGIPGERGEPGPVGPSGFAGPPGAAGQPGAKGER

GPKGPKGETGPTGAIGPIGASGPPGPVGAAGPAGPRGDAGPPGMTGFPGAAGRVGPPGPAGITGPPGP

PGPAGKDGPRGLRGDVGPVGRTGEQGIAGPPGFAGEKGPSGEAGAAGPPGTPGPQGILGAPGILGLPG

SRGERGLPGIAGATGEPGPLGVSGPPGARGPSGPVGSPGPNGAPGEAGRDGNPGNDGPPGRDGAPGF

KGERGAPGNPGPSGALGAPGPHGQVGPSGKPGNRGDPGPVGPVGPAGAFGPRGLAGPQGPRGEKGEP

GDKGHRGLPGLKGHNGLQGLPGLAGQHGDQGPPGNNGPAGPRGPPGPSGPPGKDGRNGLPGPIGPA

GVRGSHGSQGPAGPPGPPGPPGPPGPNGGGYEVGFDAEYYRADQPSLRPKDYEVDATLKTLNNQIETLL

TPEGSKKNPARTCRDLRLSHPEWSSGFYWIDPNQGCTADAIRAYC

DFATGETCIHASLEDIPTKTWYVSKNPKDKKHIWFGETINGGTQFEYNGEGVTTKDMATQLAFMRLLAN

HASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELRAEGNSRFTFSVLVDGCSKKNNKWGKTIIEY

RTNKPSRLPILDIAPLDIGGADQEFGLHIGPVCFK

>Trout

MLSFVDNRILLLLAVTSLLASCQSGGLKGPRGAKGPRGDRGPQGPNGRDGKAGLPGIAGPPGPPGLGG

NFAAQFDGGKGSDPGPGPMGLMGSRGPNGPPGAPGPQGFTGHAGEPGEPGQTGSIGARGPTGSAGKP

GEDGNNGRPGKPGDRGGPGTQGARGFPGTPGLPGMKGHRGYNGLDGRKGESGTAGAKGETGAHGA

NGSPGPAGSRGLNGERGRAGPAGPAGARGADGSTGPAGPAGPLGAAGPPGFPGAPGPKGEIGGAGSN

GPSGPQGGRGEPGINGAVGPVGPVGNPGNNGINGAKGAAGLPGVAGAPGFPGPRGGPGPQGPQGST

GARGLGGDPGPSGQKGDSGAKGEPGHSGVQGAAGPAGEEGKRGSTGEVGATGPAGLRGARGGAGTR

GLPGLEGRGGPIGMPGARGATGPGGIRGAPGDAGRAGESGLTGARGLPGNSGQGGPPGKEGPPGAAG

LDGRTGPPGPTGPRGQPGNIGFPGPKGPGGEAGKGGDKGPTGATGLRGGPGADGNNGAPGPAGVVG

NTGEKGEQGPAGAPGFQGLPGPAGPAGEAGKAGNQGMPGDQGLPGPAGVKGERGNSGPAGSAGSQG

AIGARGPAGTPGPDGGKGEPGSVGIVGAAGHQGPGGMPGERGAGGTPGPKGEKGEGGHRGLEGNMG

RDGARGGPGPSGPPGPSGANGEKGESGSFGPAGPAGLRGPSGERGEGGPAGLPGFAGPPGSDGQSGP

RGEKGPAGGKGDVGPAGPAGPSGQSGPSGASGPAGPPGGRGDAGPSGLTGFPGAAGRVGGPGPAGI

AGPPGSAGPAGKDGPRGLRGDPGPGGPQGEQGVVGPAGISGDKGPSGESGPPGAPGTAGPQGVLGPS

GFVGLPGSRGDKGLPGGPGAVGEPGRLGPAGASGPRGPAGNIGMPGMTGTQGEAGREGNSGNDGPP

GRPGAAGFKGDRGEPGSPGALGSSGQPGPNGPAGSAGRPGNRGESGPTGNGGPVGAVGARGAPGPA

GPRGEKGGAGEKGDRGMKGLRGHGGLQGMPGPNGPSGETGSAGITGPAGPRGPAGPHGPPGKDGRA

GGHGAIGPVGHRGSPGHLGPAGPPGSPGLPGPAGPAGGGYDQSGGYDEYRADQPSFRAKDYEVDATI

KSLNSQIENLLTPEGSKKNPARTCRDIRLSHPDWSSGFYWIDPNQGCIADAIKAYCDFSTGHTCIHPHPE

SIARKNWYRSSENKKHVWFGETINGGTEFAYNDETLSPQSMATQLAFMRLLANQATQNITYHCKNSVA

YMDGENGNLKKAVLLQGSNDVELRAEGNSRFTFNVLEDGCTRHTGQWSKTVIEYRTNKPSRLPILDIAP

LDIGEADQEFGLDIGPVCFK

>Dog

MLSFVDTRTLLLLAVTSCLATCQSLQEATARKGPTGDRGPRGERGPPGPPGRDGDDGIPGPPGPPGPPG

PPGLGGNFAAQYDGKGVGLGPGPMGLMGPRGPPGASGAPGPQGFQGPAGEPGEPGQTGPAGARGPPG

PPGKAGEDGHPGKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGA

PGENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGPVG

NPGPAGPAGPRGEVGLPGVSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGAT

GARGIVGEPGPAGSKGESGNKGEPGSAGAQGPPGPSGEEGKRGPNGEAGSAGPSGPPGLRGSPGSRG

LPGADGPAGVMGPPGPRGATGPAGVRGPNGDSGRPGEPGLMGPRGFPGAPGNVGPAGKEGPMGLPGI

DGRPGPIGPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGV

QGGKGEQGPAGPPGFQGLPGPAGTAGEVGKPGERGLPGEFGLPGPAGPRGERGPPGESGAAGPSGPIG

SRGPSGPPGPDGNKGEPGVLGAPGTAGASGPGGLPGERGAAGIPGGKGEKGETGLRGEIGNPGRDGA

RGAPGAMGAPGPAGATGDRGEAGPAGPAGPAGPRGTPGERGEVGPAGPNGFAGPAGAAGQPGAKGE

RGTKGPKGENGPVGPTGPIGSAGPSGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPSGITGPPG

PPGAAGKEGLRGPRGDQGPVGRTGETGASGPPGFTGEKGPSGEPGTAGPPGTPGPQGLLGAPGILGLP

GSRGERGLPGVAGSVGEPGPLGIAGPPGARGPPGAVGAPGVNGAPGEAGRDGNPGNDGPPGRDGQA

GHKGERGYPGNIGPVGAVGAPGPHGPVGPTGKHGNRGEPGPAGSVGPVGAVGPRGPSGPQGIRGDKG

EPGEKGPRGLPGLKGHNGLQGLPGLAGQHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGQPGTVG

PAGIRGSQGSQGPAGPPGPPGPPGPPGPSGGGYDFGYEGDFYRADQPRSPPSLRPKDYEVDATLKSLN

NQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIP

AKNWYRNSKVKKHIWLGETINGGTQFEYNVEGVTTKEMATQLAFMRLLANHASQNITYHCKNSIAYMD

EETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWRKTIIEYKTNKPSRLPILDIAPLDIGD

ADQEFRVDVGPVCFK

>Cattle

MLSFVDTRTLLLLAVTSCLATCQSLQEATARKGPSGDRGPRGERGPPGPPGRDGDDGIPGPPGPPGPPG

PPGLGGNFAAQFDAKGGGPGPMGLMGPRGPPGASGAPGPQGFQGPPGEPGEPGQTGPAGARGPPGPP

GKAGEDGHPGKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPG

ENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGELGPVGN

PGPAGPAGPRGEVGLPGLSGPVGPPGNPGANGLPGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATG

ARGLVGEPGPAGSKGESGNKGEPGAVGQPGPPGPSGEEGKRGSTGEIGPAGPPGPPGLRGNPGSRGLP

GADGRAGVMGPAGSRGATGPAGVRGPNGDSGRPGEPGLMGPRGFPGSPGNIGPAGKEGPVGLPGIDG

RPGPIGPAGARGEPGNIGFPGPKGPSGDPGKAGEKGHAGLAGARGAPGPDGNNGAQGPPGLQGVQGG

KGEQGPAGPPGFQGLPGPAGTAGEAGKPGERGIPGEFGLPGPAGARGERGPPGESGAAGPTGPIGSRG

PSGPPGPDGNKGEPGVVGAPGTAGPSGPSGLPGERGAAGIPGGKGEKGETGLRGDIGSPGRDGARGA

PGAIGAPGPAGANGDRGEAGPAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGTK

GPKGENGPVGPTGPVGAAGPSGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPSGISGPPGPPGP

AGKEGLRGPRGDQGPVGRSGETGASGPPGFVGEKGPSGEPGTAGPPGTPGPQGLLGAPGFLGLPGSRG

ERGLPGVAGSVGEPGPLGIAGPPGARGPPGNVGNPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKG

ERGYPGNAGPVGAAGAPGPQGPVGPVGKHGNRGEPGPAGAVGPAGAVGPRGPSGPQGIRGDKGEPG

DKGPRGLPGLKGHNGLQGLPGLAGHHGDQGAPGAVGPAGPRGPAGPSGPAGKDGRIGQPGAVGPAGI

RGSQGSQGPAGPPGPPGPPGPPGPSGGGYEFGFDGDFYRADQPRSPTSLRPKDYEVDATLKSLNNQIE

TLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPEDIPVKNW

YRNSKAKKHVWVGETINGGTQFEYNVEGVTTKEMATQLAFMRLLANHASQNITYHCKNSIAYMDEETG

NLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWQKTIIEYKTNKPSRLPILDIAPLDIGGADQ

EIRLNIGPVCFK

>Frog

MLSFVDLRSVLLLAVTLYLVTCQEVRRGPRGDKGPPGEQGPPGIPGRDGEDGLPGLPGPPGVPGLGGNF

AAQYDPSKSAEPGQQGIMGPRGPPGPPGSPGSQGFQGLPGENGEPGQTGPVGSRGPSGAPGKAGEDG

HPGKSGRPGERGPVGPQGARGFPGTPGLPGFKGIRGHTGSDGQKGAPGAAGVKGENGANGDNGSPG

QAGARGLPGERGRIGPAGSAGSRGSDGSSGPVGPAGPIGSAGAPGLPGAPGAKGELGPAGNNGPTGA

AGGRGEPGPPGSLGPAGPPGNPGTNGVNGAKGTAGLPGVGGAPGLPGGRGIPGPAGPAGPSGARGLA

GDPGIAGGKGDTGSKGEPGSVGQQGPAGPSGEEGKRGPNGEAGSSGPSGNAGIRGVPGTRGLPGPD

GRAGGIGPAGSRGSSGPPGARGPNGDAGRPGEPGLLGARGLPGFSGSNGPQGKEGPAGPQGIEGRSG

AAGPAGARGEPGAIGFPGPKGPNGEPGKNGDKGNQGPSGNRGAPGPDGNNGAQGPAGLGGATGEKG

EQGPSGAPGFQGLPGPGGPPGEVGKPGERGAPGDFGPPGSAGTRGERGAPGESGGAGPHGPSGSRGP

SGAPGPDGQKGEPGAAGLNGGLGPSGPAGIPGERGTAGTPGTKGEKGDAGNSGDYGNPGRDGARGP

AGAAGAPGPAGGPGDRGESGPAGPSGVAGPRGAPGERGEAGPAGPTGFAGPPGAAGHTGAKGDRGA

KGPKGEAGSPGPLGAHGSAGPAGPNGPAGSTGARGDAGPSGATGFPGPAGRAGAPGPPGNVGPSGPT

GHPGKDGSRGPRGDSGPVGRPGEQGQHGPVGLAGDKGPSGEAGPAGPPGAAGPSGVLGARGILGLP

GTRGERGLPGGPGSNGEPGPSGLAGSSGPRGPPGSVGSPGPVGHSGEAGRDGHPGNDGPPGRDGLP

GAKGERGYPGNTGPSGLAGAPGPAGSAGPAGKSGNRGEGGPSGPAGITGPSGPRGPAGPQGVRGDKG

EAGERGARGLDGRKGHNGLSGLPGPSGTPGETGPSGSVGPVGPRGPSGPSGPPGKEGRSGHPGAMGP

VGPRGPAGFTGPAGPPGPPGPPGHAGPSGGGYDGGDGGEYYRADQPERKPKDYEVDATLKSLNQQIEV

ILTPEGSRKNPARTCRDLRLSHPEWTSGFYWIDPNQGCTSDAIRVFCDFSSGETCIHANPDEITQKNWY

INTSNKDKKHLWFGEILNGGTQFEYHDEGLTAKDMATQLAFMRLLANQASQNITYHCKNSIAYMDEET

GNLKKAVILQGSNDVELRAEGNTRFTYSVLEDGCTKHTGEWGKTVIEYRTNKPSRLPI

LDIAPLDIGGHDQEIGFEIGPVCFK

>Toxodon

GPMGIMGPRGPPGASGAPGPAGEPGEPGQTGPAGARGPPGPPGKAGEDGHPGKPGRPGERGVVGPQG

ARGFPGTPGIPGFKGIRGHNGIDGIKGQPGAPGVKGEPGAPGENGTPGQAGARGIPGERGRVGAPGPA

GARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGPVGNPGPAGPAGPRGEVGIPGVSGPVGPPGN

PGANGITGAKGAAGIPGVAGAPGIPGPRGIPGPVGAAGATGARGIVGEPGPAGSKGESGNKGEPGSAG

PQGPPGPAGEEGKRGPNGEAGSTGPTGPPGIRGSRGIPGADGGSRGATGPAGVRGDSGRPGEPGIMG

PRGFPGSPGNIGPAGKEGPVGIPGIDGRPGPTGPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGIA

GARGPAGPPGFQGIPGPAGTAGEVGKPGERGIPGEFGIPGPAGARGERGPPGESGAVGPAGPIGSRGPS

GPPGPDGNKGEPGNIGAIGTAGPSGPSGIPGERGAAGIPGGKGEKGETGIRRGAPGAIGAPGPAGANG

DRGEAGPAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGTKGPKGENGPVGPTGP

VGAAGPAGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPAGITGPPGPPGAAGKEGIRGPRGDQ

GPVGRSGETGASGIPGFAGEKGPAGEPGTAGIPGTPGPQGIIGAPGIIGIPGSRGERGIPGVAGSIGEPG

PIGIAGPPGARGPPGAVGNPGVNGAPGEAGRHGNRGEPGPAGSVGPAGAVGPRGPSGPQGIRGDKGE

PGDKGPRGIPGIKGHNGIQGIPGIAGQHGDQGAPGAVGPAGPRGPAGPSGPAGKDGRIGHPGTVGPA

GIRGSQGSQGPAGPPGPPGPPGPPGPS

>Mastodon

QYDAKGVGLGPGPMGLMGPRGPPGATGPPGSPGFQGPPGEPGEPGQTGPAGSRGPAGPPGKAGEDGH

PGKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQI

GARGLPGERGRVGGPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGPVGNPGPSGPAGP

RGEAGLPGVSGPVGPPGNPGANGLAGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGARGIVGEPG

PAGSKGESGSKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPAGPPGLRGGPGSRGLPGADGRAGV

MGPPGSRGASGPAGVRGPSGDSGRPGEPGVMGPRGLPGSPGNVGPAGKEGPAGLPGIDGRPGPIGPA

GARGEPGNIGFPGPKGPAGDPGKNGDKGHAGLAGPRGAPGPDGNNGAQGPPGLQGVQGGKGEQGPA

GPPGFQGLPGPSGTAGEAGKPGERGIPGEFGLPGPAGPRGERGPPGQSGAAGPTGPIGSRGPSGPPGP

DGNKGEPGVVGAPGTAGPSGPVGLPGERGAAGIPGGKGEKGETGLRGDTGNTGRDGARGAPGAVGAP

GPAGATGDRGEAGPAGSAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQAGAKGERGTKGPKGEN

GPVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPAGITGPPGPPGAAGKEGL

RGPRGDQGPVGRTGETGASGPPGFAGEKGSSGEPGTAGPPGAPGPQGILGPPGILGLPGSRGERGLPG

VAGAVGEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGSDGPPGRDGLPGHKGERGYPGN

AGPVGTAGAPGPQGPLGPAGKHGNRGEPGPAGSVGPVGAVGPRGPSGPQGARGDKGEAGDKGPRGL

PGFKGHNGLQGLPGLAGQHGDQGSPGSVGPAGPRGPAGPSGPVGKDGRPGHAGAVGPAGVRGSQGS

QGPSGPPGPPGPPGPPGPSGGGYDFGYDGDFYRA

>Salamander

MLSFVDTRIVLLLAVTSSLATCQYNYEANRGPRGYKGPQGDQGPPGAPGRDGVDGPPGPAGPPGPPGP

SGLGGNFAAQYDGGKSDPGPGPMGMMGPRGPPGPSGSPGAQGFQGLPGEPGEPGQTGPVGSRGPTG

PPGKSGEDGSPGKSGRPGERGTVGTQGARGFPGTPGLPGFKGLRGHNGFDGVKGAAGSQGAKGETG

ANGENGSPGQAGARGLPGERGRVGGAGPGARGSDGSAGPSGPAGPIGSAGAPGLPGAPGAKGEIGSA

GNNGPSGPAGSRGDPGLPGSVGPVGPAGNPGSNGVSGAKGAAGLPGVGGAPGLPGPRGIPGPQGASG

AAGARGLAGDPGSPGGKGDSGSKGEPGSAGQQGNAGPSGEEGKRGPNGEPGSSGPAGPAGIRGVPG

TRGLPGPDGRAGGMGPPGSRGSSGPAGVRGPSGDAGRPGEPGLLGQRGLPGFPGNTGPVGKEGPAGP

AGIEGRTGAAGPTGARGEPGSIGFPGPKGPGGEPGKNGDKGSAGPSGARGAPGPDGNNGAQGPPGVV

GNTGEKGEQGPAGAPGFQGLPGPGGAAGEAGKVGDRGMPGDFGPPGPAGVRGERGAPGESGSAGPL

GPVGSRGPSGPPGPDGTKGEPGVAGLAGAVGPSGSGGSPGERGGAGTPGPKGEKGEAGNRGEYGNQ

GRDGARGPAGASGAPGPSGGPGDRGESGPSGPAGPAGSRGAPGERGEHGPGGPTGFGGPPGAAGHT

GVKGERGEKGPKGELGPQGPVGASGASGPAGPNGPAGAPGSRGEVGPAGATGFPGPAGRTGGPGPAG

MGGPPGPSGHAGKDGPRGPRGDSGPVGRPGEQGGLGPQGISGEKGPSGEPGTAGPPGSSGPSGVLG

ARGILGLPGTRGERGLPGGPGGNGEPGATGPTGTAGSRGAPGPVGSAGMNGPAGEAGRDGNPGNDGP

PGRDGQAGAKGERGYPGNTGGVGHAGAPGPHGSVGPAGKSGNRGEPGPSGSQGPAGLPGARGPAGP

AGSRGDKGESGEKGGRGLDGRKGHNGLQGLPGLPGTSGEAGSAGPSGPSGPRGPAGPSGPPGKDGH

SGQPGPVGPAGVRGSPGHQGPAGPPGSPGAPGPAGPSGGGYDGGFEGGEFYRADQPSLRPKDYEVDS

TLKTLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGFYWIDPNQGCTADAIRVYCDFSTGETCIHSN

PETISAKTSYVNKNPKDKKHVWVGEVLNGGTQFEYNEEGVTTKDMATQFAFMRLLANHASQNITYHCK

NSIAYMDGETGNLKKAVLLQGSNDVELRAEGNSRFTFSVLEDSCTKHTGEWGRTVMEYRTNKPSRLPIL

DIAPMDIGGAEQEFRVDIGPVCFK