



Supporting Online Material for

Molecular Phylogenetics of Mastodon and *Tyrannosaurus rex*

Chris L. Organ, Mary H. Schweitzer, Wenxia Zheng, Lisa M. Freimark, Lewis C. Cantley, John M. Asara*

*To whom correspondence should be addressed. E-mail: jasara@bidmc.harvard.edu

Published 25 April 2008, *Science* **320**, 499 (2008)

DOI: 10.1126/science.1154284

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 and S2
References
Appendices 1 and 2

Supporting Online Material (SOM)

Figure S1 and S2

Materials and Methods

Extra text

References

Appendix 1: Sequence alignment file

Appendix 2: List of alligator collagen peptide sequences

Figure S1

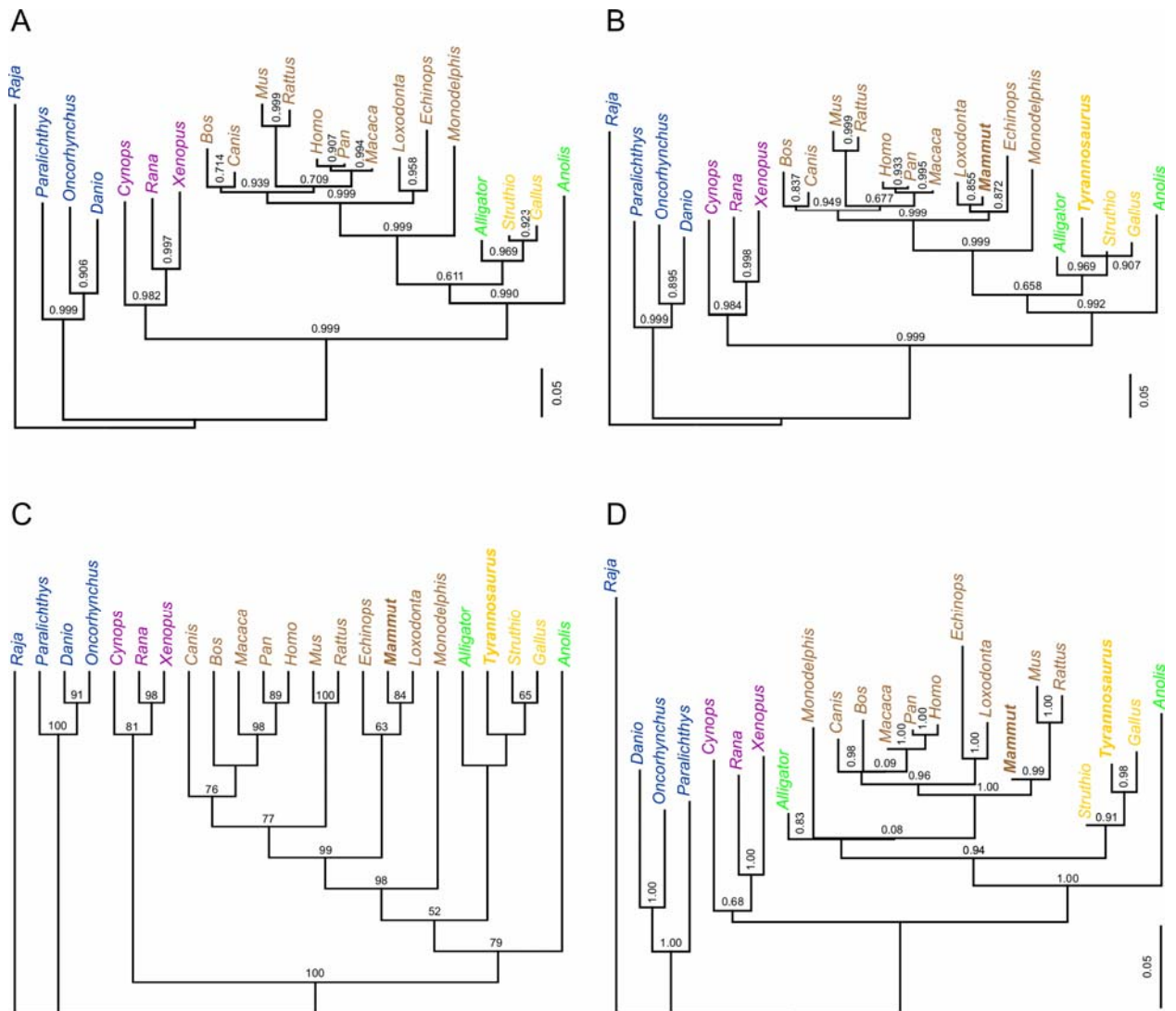


Figure S1. Alternate phylogenetic trees with color coded taxa: “fish” (blue), lissamphibians (purple), mammals (brown), non-dinosaurian reptiles (green), dinosaurs (gold). The extinct species, *Mammuth* and *Tyrannosaurus*, are bolded. **(A)** Maximum likelihood tree generated in PhyML v2.4.5 (Guindon and Gascuel 2003) including only

the extant species of this study (see below) with branches supported by the approximate likelihood-ratio test (Anisimova and Gascuel 2006). **(B)** A second maximum likelihood tree including sequences from *Mammut* and *Tyrannosaurus*. **(C)** 50% majority-rule consensus of 4 most-parsimonious ultrametric trees (consistency index: 0.756) generated in PAUP* v4.0b10 (Swofford 2003) using maximum parsimony, with bootstrap values supporting branches (500 replicates). **(D)** Neighbor joining tree generated in MEGA v 4.0 supported by an interior branch test - 500 bootstrap replicates (Tamura, Dudley et al. 2007).

Figure S2

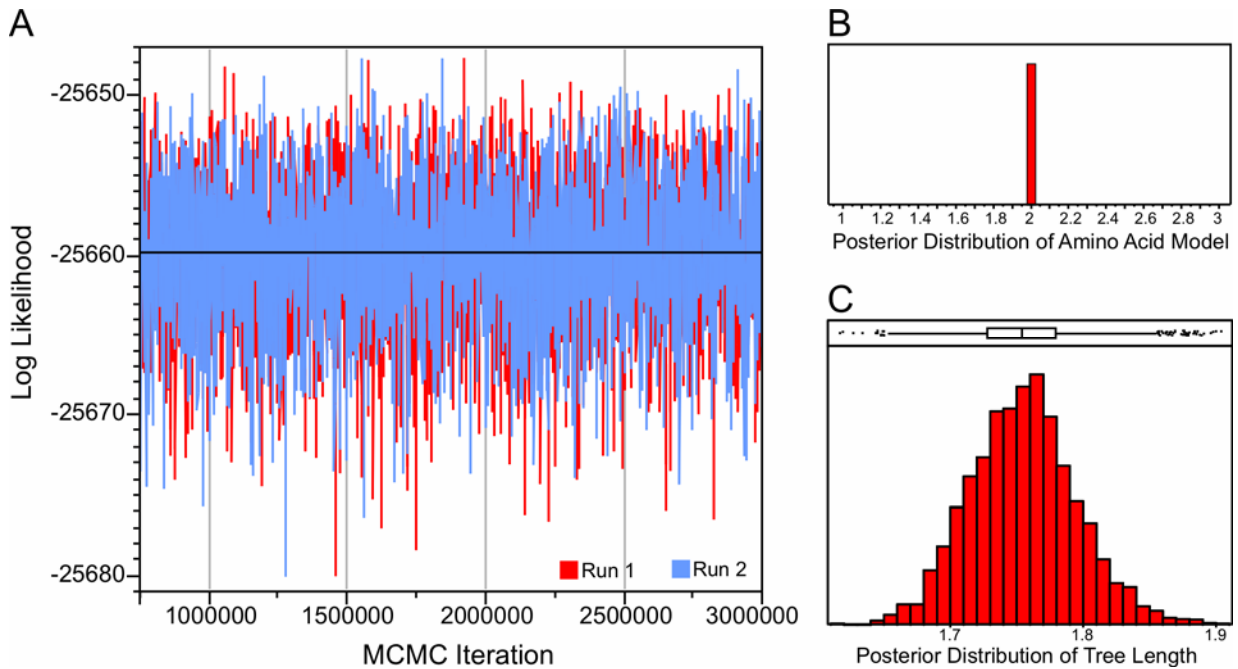


Figure S2. Posterior distribution statistics from the Bayesian phylogenetic analysis visualized in the program Tracer v1.4 (Rambaut and Drummond 2007) with graphs generated using JMP v6.03 (SAS_Institute 2006). **(A)** Trace of the posterior parameter log likelihoods demonstrate converged stationary phases for runs one and two (Bayes Factor=0.374, calculated in Tracer). **(B)** Posterior distribution of amino acid models shows that model 2 (Dayhoff) best fits the data (SD=0). **(C)** Posterior distribution of tree length (mean=1.75, SD=0.04; the center box in the outlier box plot represents the interquartile range).

Materials and Methods

Acquisition of collagen sequences from extant organisms

Full length collagen $\alpha 1(I)$ protein sequences were obtained from the NCBI non-redundant protein database or the ENSEMBL genome browser for the following organisms (including corresponding accession numbers – ENSEMBL numbers begin with “EN”): chicken P02457 (*Gallus gallus*), African clawed frog NP_001080821 (*Xenopus laevis*), Bullfrog BAA29028 (*Rana catesbeiana*), Japanese newt BAA36973 (*Cynops pyrrhogaster*), rainbow trout BAB55661 (*Oncorhynchus mykiss*), zebra fish ENSDARP00000003177 (*Danio rerio*), skate BAD98524 (*Raja kenoi*), flounder BAD77968 (*Paralichthys olivaceus*), Norway rat ENSRNOP00000005311 (*Rattus norvegicus*), mouse ENSMUSP00000001547 (*Mus musculus*), rhesus monkey ENSMMUP00000001955 (*Macaca mulatta*), dog ENSCAFP00000025056 (*Canis lupus familiaris*), cow ENSBTAP00000017420 (*Bos taurus*), chimpanzee ENSPTRP00000015952 (*Pan troglodytes*), human ENSP00000225964 (*Homo sapiens*), opossum XP_001367735 (*Monodelphis domestica*), green anole gencode on scaffold_353:1172992-1215567 (*Anolis carolinensis*), tenrec ENSETEP00000014000 (*Echinops telfairi*) and elephant ENSLAFP00000002863 (*Loxodonta africana*). Ostrich (*Struthio camelus*) and Alligator (*Alligator mississippiensis*) collagen $\alpha 1(I)$ and $\alpha 2(I)$ proteins were sequenced by mass spectrometry (LC/MS/MS).

Collagen $\alpha 2(I)$ protein sequences were obtained from the Ensembl genome browser: *Danio rerio* ENSDARP00000042989, *Canis familiaris* ENSCAFP00000003058, *Bos taurus* ENSBTAP00000017920, *Pan troglodytes* ENSPTRP00000033204, *Echinops telfairi* ENSETEP00000009581, *Homo sapiens* ENSP00000377927, *Gallus gallus* ENSGALP00000015687, *Mus musculus* ENSMUSP00000031668, *Rattus norvegicus* ENSRNOP00000016423, *Loxodonta africana* ENSLAFP00000014382, *Monodelphis domestica* ENSMODP00000027344. Collagen $\alpha 2(I)$ protein sequence for *Anolis carolinensis* (GENSCAN00000056727) were obtained from an *ab initio* gencode on Ensembl (*Anolis* scaffold_272:764,480-824,308). Note that the two collagen genes used in this study are paralogues.

Sample preparation and enzymatic digestion of alligator collagen

A metacarpal was harvested from a 13 year old frozen *Alligator mississippiensis*, ground to a powder and 0.8 mg was aliquoted to 4 50 mL tubes (0.2 g/tube). Bone was demineralized overnight with 10mL of 0.5M EDTA (pH=8.0)/tube, and centrifuged. Supernatant was decanted, and the remaining pellet extracted as follows: To each tube, 5ml of 0.5 M EDTA (pH 8.0, 0.22 μ m filtered) was added, and tubes were gently agitated for 3 days at room temperature. Supernatant was removed and stored at 4 °C, (hereafter referred to as “EDTA super”). To the remaining pellet, 5mL of 6M GuHCl, 0.1 M Tris pH 7.4 (0.22 μ m filtered) was added and sample agitated at 60°C for 20 hours, centrifuged (6,000 rcf, 15 min), and supernatant removed and added to “EDTA super”, referred to as ‘first extraction’. This was dialyzed against e-pure 18.3 mega ohm water (4-L beaker), at 4°C, stirring for 4 days with 2 changes/day. The dialysate was lyophilized to completion producing a precipitate that was white, fibrous and fluffy, consistent with extant collagen. Resulting total yield was 23.2mg. Approximately 1 mg of this white lyophilate was proteolyzed with 5 μ g of modified trypsin (Promega Corp., Madison, WI) in 150 μ L of 50mM ammonium bicarbonate (pH=8.3) solution overnight at 37°C with shaking. The next day, the digestion was stopped with 50 μ L of 1% trifluoroacetic acid (TFA) (Fluka,

St. Louis, MO) and stored frozen. Contents were transferred to a 200 μ L autosampler vial (Waters Corp., Bedford, MA) for mass spectrometry analyses.

Microcapillary LC/MS/MS shotgun sequencing analysis

A 2 μ L aliquot (approximately low pmol amount of peptides) was injected onto the LC/MS/MS system for sequencing. The μ LC/MS/MS setup consisted of a 75 μ m id x 10cm length microcapillary column (New Objective Inc., Woburn, MA) self-packed with Magic C₁₈ (Michrom Bioresources, Auburn, CA) operated at a flow-rate of ~275 nL/min. The HPLC gradient was 5% B to 38% B over 50 minutes followed by a 5 minute wash at 95% B and a 20 minute post-run equilibration at 0% B (A: 99% water/0.9% acetonitrile/0.1% acetic acid; B: 99% acetonitrile/0.9% water/0.1% acetic acid). The microcapillary LC system is coupled directly to a ThermoElectron LTQ 2D linear ion trap (ThermoScientific, San Jose, CA) mass spectrometer operated in positive ion mode for data-dependent acquisitions (Top 8: 1 MS survey scan followed by eight MS/MS scans) of peptide fragmentation (MS/MS) spectra by collision-induced dissociation using helium gas. The spray tip voltage was 2.5kV and capillary voltage was 35V.

Peptide/protein sequence identification and validation

All collected MS/MS fragmentation spectra were searched against the reversed NCBI non-redundant protein database (National Institutes of Health, Bethesda, MD), reversed Swiss-Prot protein database (Swiss Institute of Bioinformatics, Geneva, Switzerland) and reversed predicted collagen sequence database generated in-house (Asara *et. al.* 2007) using the Sequest search engine in Proteomics Browser software (ThermoScientific, San Jose, CA). Differential post-translational modifications (PTM) such as hydroxylation (+16 Da) of the amino acids proline and lysine were included in the database searches as well as oxidation of methionine (+16 Da), a common in vitro modification that occurs during sample processing. Deamidation as a PTM was not accounted for due to the ion trap's low resolution and mass accuracy; however, the precursor peptide mass error tolerance window (+/-2.5Da) included deamidation events during database searches. Peptide sequences were initially accepted if they matched the forward databases and passed the following Sequest scoring thresholds: 2+ ions, Xcorr \geq 2.0, Sf \geq 0.4, P \geq 5; 3+ ions, Xcorr \geq 2.65, Sf \geq 0.5, P \geq 5. Peptides with gas phase charges of 1+ and 4+ were not accepted as valid due to difficulty of interpretation of such ions. After passing the initial 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 sequence. Sequences were also determined using the Paragon search algorithm from Protein Pilot software (Applied Biosystems, CA) using the same reversed databases. Paragon searches all known PTMs and allows for up to two amino acid substitutions per peptide sequence to assist in finding unique alligator sequences. Searches were run with a 95% confidence interval for protein identifications. Peptide confidence score cut-off >5 was used prior to manual inspection if other peptides from the identified protein contained peptide confidence scores greater than 95. For both searching algorithms, the peptide identification false positive rates were estimated at less than 2% based on the number of forward to reversed peptide hits. The primary peptide

sequences and corresponding protein are reported here for sequence alignment and phylogenetic analyses.

Phylogenetic analyses and extra text

Molecular data were managed using BioEdit v 7.0.8 (Hall 1999). Collagen α 1(I) and Collagen α 2(I) protein sequences were obtained from NCBI and ENSEMBL and aligned using ClustalX v 2.0 (Thompson, Gibson *et al.* 1997) using the default amino acid alignment settings. Polypeptide fragments from *Struthio*, *Alligator*, *Tyrannosaurus*, and *Mammot* were organized separately in multi-FASTA format and aligned into contigs using ClustalX in profile alignment mode. A consensus sequence was then created using SeaView v2.0 (Galtier, Gouy *et al.* 1996) and non-alignable fragments were manually examined to identify additional profile alignments (Supplementary Appendix 1).

MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) was used for Bayesian inference of phylogenetic relationships. The prior for the amino acid model set to mixed, in which the MCMC chain explores 10 fixed-rate amino acid matrices (Poisson, Jones, Dayhoff, Mtrev, Mtmam, Wag, Rtrev, Cprev, Vt, and Blossum). When the MCMC chain converges, the proportion of the posterior probability of each model represents the relative contribution to the posterior results. The analysis was performed using three million generations in four chains, sampling posterior trees every 1000 generations. A replicate search was performed using four million generations in four chains, sampling posterior trees every 1000 generations with identical results to the first run. The first 25% of the posterior trees were discarded as burn-in during parameter and tree summarization. The program Tracer v1.4 (Rambaut and Drummond 2007) was used to examine the log likelihoods, ensuring that they were in the stationary “fury caterpillar” phase (Supporting Fig. 2). The majority-rule consensus tree shown in the primary text (Fig. 1) contains nodal values that represent the percentage of times a particular clade is present in the posterior distribution of trees. The Bayesian tree is identical to the likelihood tree and highly similar to the parsimony tree as well (see below). An additional tree (not shown) was inferred in which all sites (for all taxa) were removed that contained gaps (missing data) in the *Tyrannosaurus* sequence, resulting in a dramatically smaller dataset. The resulting tree was poorly resolved in multiple areas due to the lack of informative sites for all taxa, though the *Tyrannosaurus* still grouped with other amniotes. Another tree (not shown) was inferred in which regions containing gaps (missing data) in the alligator, ostrich, and *T. rex* sequences were removed. This reduced the amount of missing data in the set and produced the same topology, but it did not appreciably change the support for archosaurian nodes.

Phylogenetic trees were generated using Maximum likelihood (Sup. Fig. 1A and B) in PhyML v2.4.5 (Guindon and Gascuel 2003). The following settings were used: Dayhoff protein substitution model; the proportion of invariable sites was estimated; four substitution rate categories; gamma distribution parameter was estimated; the starting tree was a BIONJ distance-based tree; and the tree topology was optimized. Maximum likelihood trees were evaluated by the approximate likelihood-ratio test (Anisimova and Gascuel 2006), which collapses a branch to length 0, compares the likelihood of the

second most likely tree to the most likely tree and then computes a p-value. We used the most conservative method to calculate the significance of the aLRT test by using the minimum of the parametric χ^2 -based and the nonparametric Shimodaira–Hasegawa-like (SH-like) procedures. The maximum likelihood results agree with other phylogenetic studies (Pisani, Yates *et al.* 2002; Rest, Ast *et al.* 2003; Springer, Murphy *et al.* 2003) and are concordant with the Bayesian tree. As with the Bayesian and other trees produced from the two collagen proteins, the *Anolis* sequence is misplaced as a basal amniote instead of a diapsid and sister taxon to the archosaurs (birds and crocodilians).

Phylogenetic trees were also generated using maximum parsimony (heuristic search at default settings) in PAUP* v 4.0b10 (Swofford 2003) using the PaupUP v1.0.3.1 interface (Calendini and J.-F. 2005). The trees were rooted by defining the fish *Raja kenoei*, *Oncorhynchus mykiss*, *Danio rerio*, and *Paralichthys olivaceus* as members of the outgroup. Gaps were treated as missing data. A heuristic search was used with the following settings: best trees kept and set to automatically increase by 100; random stepwise addition with 100 replicates and one tree held at each step; TBR branch swapping on best trees only with no more than 2 trees with a score of 5 saved. A full heuristic bootstrap analysis was then performed with 500 replicates. The 50% majority-rule consensus of four most-parsimonious trees (consistency index: 0.756) (Sup. Fig. 1C) agrees with the Bayesian and likelihood trees with the exception that the primates do not group with the rodents, which also contradicts other published studies (Springer, Murphy *et al.* 2003; Springer, Madsen *et al.* 2004). Importantly, the *Tyrannosaurus* and mastodon are placed in expected locations of the tree, though bootstrap values do not provide strong support (which is not surprising given the amount of missing data in these taxa).

MEGA v4.0 (Tamura, Dudley *et al.* 2007) was used to infer a neighbor joining tree with the Dayhoff protein substitution model. Gaps were excluded as necessary using the Pairwise-deletion option and the rate among sites were gamma distributed using a gamma parameter of 1. Tree support was evaluated by the interior branch test with 500 bootstrap replicates (Tamura, Dudley *et al.* 2007). Neighbor joining performs the worst of the four methods, misplacing *Anolis*, *Alligator*, mastodon, and the *Tyrannosaurus* (though the *Tyrannosaurus* still groups with the other dinosaurs (birds)).

Missing data can be a significant obstacle in phylogenetic reconstruction (Wiens 1998; Kearney 2002; Sanderson, Driskell *et al.* 2003). Computer simulations have shown that including taxa with extensive missing data reduces phylogenetic accuracy (Huelsenbeck 1991), though if more characters are added to such a simulation, the problem of including taxa that extensively lack data is mitigated (Wiens 2003). Therefore, support for inferred phylogenetic relationships among taxa is more a function of the amount and quality of data brought to bear in an analysis, than the amount of missing data (Wiens, Fetzner *et al.* 2005). Furthermore, the poor performance of neighbor joining compared with the Bayesian, likelihood, and parsimony results are expected given that distance methods perform poorly compared with likelihood-based approaches for datasets with large amounts of missing data (Wiens 2003).

To evaluate the influence and order of species during the alignment that included *T. rex*, the living archosaur species (*Gallus*, *Struthio*, and *Alligator*) were removed from the dataset and the sequences were realigned in ClustalX. The resulting alignment was then used to infer a phylogenetic tree (results not shown). The Bayesian approach correctly placed the sole archosaur (*Tyrannosaurus*) as the sister group to mammals with very high support (0.99 posterior probability). Parsimony also correctly placed the *Tyrannosaurus*. Neighbor joining performed nearly as well, grouping the *Tyrannosaurus* and *Monodelphis* together as a sister group to the rest of mammals, but still outside of Eutheria within the amniotes.

The Punctuated Evolution program (Venditti, Meade *et al.* 2006) can detect punctuated evolution in molecular phylogenetic trees, deviation from a molecular clock caused by punctuated evolution, and the node density artifact (which is the underestimation of branch lengths – and thus evolutionary rates - in portions of the tree with few taxa). The program was used to analyze the Bayesian and likelihood trees by removing the outgroup, *Raja*, and using the web interface at <http://www.evolution.reading.ac.uk/pe/index.html>. The test found no evidence for punctuated evolution or node density artifact (the number of significant β and the number of significant $\delta < 1$ and β were both 0).

Supporting Online References

- Anisimova, M. and O. Gascuel (2006). "Approximate likelihood ratio test for branches: A fast, accurate and powerful alternative." *Systematic Biology* **55**(4): 539-552.
- Asara, J.M. et. al. (2007). "Protein sequences from Mastodon and Tyrannosaurus rex revealed by mass spectrometry." *Science* **316**(5822): 280-285.
- Calendini, F. and M. J.-F. (2005). PaupUP v1.0.3.1 A free graphical frontend for Paup* Dos software.
- Galtier, N., M. Gouy, et al. (1996). "SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny." *Computer Applications in the Biosciences* **12**: 543-548.
- Guindon, S. and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." *Systematic Biology* **52**(5): 696-704.
- Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." *Nucleic Acids Symposium Series* **41**: 95-98.
- Huelsenbeck, J. P. (1991). "When are fossils better than extant taxa in phylogenetic analysis?" *Systematic Zoology* **40**: 458-469.
- Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogeny." *Bioinformatics* **17**: 754-755.
- Kearney, M. (2002). "Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions." *Systematic Biology* **51**: 369-381.
- Pisani, D., A. M. Yates, et al. (2002). "A genus-level supertree of the Dinosauria." *Proceedings of the Royal Society of London, Series B* **269**(1494): 915-921.

- Rambaut, A. and A. J. Drummond (2007). Tracer v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Rest, J. S., J. C. Ast, et al. (2003). "Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome." Molecular Phylogenetics and Evolution **29**: 289–297.
- Sanderson, M. J., A. C. Driskell, et al. (2003). "Obtaining maximal concatenated phylogenetic data sets from large sequence databases." Molecular Biology and Evolution **20**: 1036–1042.
- SAS_Institute (2006). JMP, Version 6.03. Cary, NC, SAS Institute Inc.
- Springer, M. S., W. J. Murphy, et al. (2003). "Placental mammal diversification and the Cretaceous-Tertiary boundary." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 1056-61.
- Springer, M. S. S., M. J., O. Madsen, et al. (2004). "Molecules consolidate the placental mammal tree." Trends in Ecology and Evolution **19**(8): 430-438.
- Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, Massachusetts., Sinauer Associates.
- Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. ." Molecular Biology and Evolution **10.1093/molbev/msm092**.
- Thompson, J. D., T. J. Gibson, et al. (1997). "The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Research **25**: 4876-4882.
- Venditti, C., A. Meade, et al. (2006). "Detecting the node-density artifact in phylogeny reconstruction " Systematic Biology **55**: 637-643.
- Wiens, J. J. (1998). "Does adding characters with missing data increase or decrease phylogenetic accuracy?" Systematic Biology **47**: 625–640.
- Wiens, J. J. (2003). "Missing data, incomplete taxa, and phylogenetic accuracy." Systematic Biology **52**(4): 528–538.
- Wiens, J. J., J. W. Fetzner, et al. (2005). "Hylid frog phylogeny and sampling strategies for speciose clades." Systematic Biology **54**: 719-748.

Supporting Online Material Appendix 1: Sequence Alignment

```
>Danio
MFSFVDIRLALLLSATVLLARGQGE-----DDRTGGSCSLDQGQVYNDRDVWKPEPCQI
CVCDSGTVMCDEVICEDTS-DCPNPVI PHDECCPVCP-----DDDFQEPSVEGPRGSPGD
KGERGPAGPPGNDGIPGQPGLPGPPGPPGPPG---LGGNFSPQMSGGFD-----EKSSP-M
AVPGPMGPMGPRGAPGPPGSPGQFPGTGPPEGEAGAP----GPMGPRGAAGPPGKNGE
DGESGKPRPGERGPPGQARGFPPTPLPGIKGHRGFSGLDGAAGDAGPAGPKGEPGA
PGENGTGAMGPRGLPGERGRAGPPGAAGARGNDGAAGAAGPPGPTGPAGPPGFPGGPGS
KGEVGPQGSRGAEGPQARGEAGNPGPAGPAGPAGNNGADGAPGAKGAPGAPGIAGAPGF
PGPRGPPGAAGAAGAPGPKGNTGEAGAPGAKGEAGAKGEAGAQQVQGPPEEGKRGF
RGEFAGGARGPTGERGAPGARGFPADGAAGPRGAPGERGGPGVVGPKGATGEPGRNGE
PGMPGSKGMTGSPGSPGPDGKTGPGGAPGQDGRPGPPGPGVARGQPGVMGFPGPKGAAGE
AGKPGERVMGAI GATGAPGKGDV GAPGAPGAPGAPGAPGERGEQGAAGPPGFQGLPGPQGA
TGEPGKSQGEQAPGEAGAPGSPGSRGDRGFPGERAPGPAGPVGARGSPGSAGNDGAKGE
SGAAGAPGAQGPGLQGMPGERGAAGLPGLKGRDQDQAKGADGAAGKDGIRGMTGPIGP
PGPAGAPGDKGESGAQGLVGTGARGPPGERGETGAPGPAGFAGPPGADGLPGAKGEPGD
NGAKGDAGAPGPAGATGAPGPQGPVGTGPKGARGAAGPPGATGFPGAAGRVPVPPGSGN
```

SGPPGPPG PAGKEGQKGNRGETGPAGRTGEVGAAGPPGAPGEKGNPGAEGATGPAGIPGP
QGIGGQRGIVGLPGQRGERGFPLPGPSGEIGKQGPSGSPGERGPPGPMGPPGLAGPPGE
PGREGTPGNEGSAGR DGAAGPKGDRGETGPSGT PGAPGPPGAAGPIGPAGKTGDRGETGP
AGVPGPAGPSGPRGSPGAGARGDKGETGEAGERGMKGHRGFTGMPPGPPGPPGSPGESGP
AGASGPAGPRGPAGSAGSAGKDGMSGLPGPIGPPGPRGRNGEIGPAGPPGPPGPPGAPGP
SGGGFDIGFIAQP-QEKAPDPFRHFRADDANVMRDRDLEVDTTLKSLSQQIESIISPDTG
KKNPARTCRDLKMCHPDWKSGEYWIDPDQGCNQDAIKVYCNMETGETCVNPTESAI PKKN
WYTSKNIKEKKHVWFGEAMTDGFQFEYGSE-GSKPEDVNIQLTFLRLMSTEASQNIITYHC
KNSIAYMDQASGNLKKALLLQGSNEIEIRAEGNSRFTYSVTEDGCTSHTGAWGKTVIDYK
TTKTSRLPIIDIA PMDVGAPNQEFG--IEVGPVCFL-MLS FVDTRILLLLAVTSYLASCQ
C-----GLKGPKG---PRGERGPKGPDGKPRP---GLPGPAG---PPGPP-
--GLGGNFAAQYDGA KGPDPGPGMGLMGPRGSPGSPGAPGAQGLQGHAGEPGEPEGQAGA
IGARGPPGPPGKNGEDGNNGRPGKPGDRGVLGAQGARGFPGT PGLPGMKGHRGYNGIDGR
KGEPGAAGAKGENGAAGSNGT PGQRGGRGLPGERGRVGPAGPAGARGADGNTGPAGPAGP
LGSAGPPGFPAGPGPKGELGPAGPTGPSGAQQORGEPPNGAVGVPVGPNGNPGANGINGA
KGAAGLPGIAGAPGFPGRGGPQGPSGASGPRGLGGDPGPVGVKGD SGVKGEPSGAGP
QGPPGSPGEEGKRSTGEQGTGPLGLRGRGAAGTRGLPGLAGRS GPMGMPGPRGGVGA
PGARGPPGDAGRAGEAGLVGARGLPGSPGSSGPPGKEGPGAAGQDGRTPGPPGPTGPRGQ
PGNIGFPGPKGPSGEAGKPGKPGVGTGLRGS PGDPGNNGPAGPVGLAGAPGEKGEQGP
SGAPGFQGLPGPAGPVGEAGKPGDRGI PGDQGVSGPAGVKGERGNP GAGAAGAQQPIGA
RGPSGT PGPDGNKGEPGA VGPAGAPGPPQAAGMPGERGAAGTPGAKGEKGEAGYRGLEGN
AGKDGARGAPGSPGPPGAGANGDKGETGSFGPPG PAGPRGAPGERGESG PAGPSGFAGP
PGADGQTGPRGEKGPAGGKGDAGPAGPAGPAGNTGPLGSPGVPVGPARGDSGPTGLTGF
PGAPGRVGPVGPAGIVGPAGLTGPAGKDGPRGPRGDVGPAGPPG ENGMIGPLGLAGEKGP
PGEAGAPGAPGAPGQQLGSQGFNGLPGSRGDRGLPGIPGSVGEPRVGPAGAPGARGP
GGNIGMPGMTGPQGEAGREGSPGNDGPPGRPGAAGIKGDRGEPGSPGTAGPVGAPGPNP
SGAVGRPNRGE SGPSPGPTGAVGPAGARGAPGAPRGEKGVAGEKGDGRGMKGLRGHPGL
QGMPPGPNGPSGSPAGIAGPSGPRGAPGPNPAGKDGSNMMPGAI GPPGHRGPAGHVGP
AGPPGSPGLPGPPGSPGGYDTSGGYD--EYRADQA----SLRAKDYEV DATIKSLNTQI
ENLLSPEGSKKNPARTCRDIRLSHP EWSS-GFYWIDPNQGCTMDAIKAFCD FSTGQTCIH
PHPESI PRKNWY--RSSQEKHTWFGETINSGTEFAYNDETLS PQSMATQLAFMRL LANQ
AVQNIITYHC KNSIAYMDAENGLKKAVLLQGSNDVELRAEGNSRFTFSVLE DGC SRHTGQ
WSKTVIEYRTNKPSRLPILDIA PLDIGGADQEFGLDIGPVCFK

>Oncorhynchus

MFSFVDIRLALLLSATVLLARGQGE-----DDRTAGSCTLDGQFYNDRDVWKPEPCQI
CVCDSGTVMCDEVICEDTS-DCPNPVI PHDECCPICP-----DDGFQEPKVEGPQGD RGA
KGEPGPAGFPNDGIPGQPLPGPPGPPGPPG---LGGNFSPQMSGGFD----EKSGGGM
SMPGPMGMPGPRGPPGPPGSSGPPQGF TGPPEPGEAGSS----GPMGPRGPAGPPKNGD
DGESGKPRPGERGASGPQGARGFPGT PGLPGIKGHRGFSGLDGAKGESGPAGPKGEGGA
SGENGAAGAMGPRGLPGERGRAGPNGAAGARGNDGAAGAAGPPGPTGPAGAPGFPGGPGA
KGEVGAQGARGGEGPQGSRGEAGNPGPAGAAGPAGNNGADGNPGTKGAPGSSG IAGAPGF
PGPRGPPGPPQAGGAPGPKGNTGEV GANAKGEAGAKGESGPAGVQGPAGPAGEEGKRGG
RGEPGGAGARGAPGERGAPGSRGFP GSDGASGPKGGPGERGGAGVAGAKGNTGEPGRNGE
PGMPGSKGMTGSPGSPGPDGKTGPSGAGGQDRPGPPGVPVARGQPGVMGFP GPKGAAGE
GGKPGERGVMGSPS GAVGAPGKDG DVGAPGAPVAGPSGERGEQAGGPPGFQGLSGPQGA
IGETGKPGEQQLPGE GGPAGSAGSRGDRGF PGERGAPGSPG PAGARGSPG SAGNDGGKGE
AGAAGAPGGQGPPLQGMPGERGAGGLPGLKGD RGDQGVK GADGAGGKDGVRGMTGPIGP
NGPAGSPGDKGETGAPGAVGSPGARGAPGERGESGAPGAGFAGPPGGDGQPGAKGEAGD
NGAKGDGGAQGPAGPTGAPGPPGAGNTGAKGARGAAGPPGATGFPGAAGRFGPPGSPGN
NGPPGTGPPGGKEGQKGNRGETGPAGRP GELGAAGPPGPKGEKQPPGGDPNGPSGT PGP
QGIGGQRGIVGLPGQRGERGFPLAGQLGEPGKQGP GPFGERGPPGPMGPPGLAGAPGE
PGREGTPGNEGS SRDGAAGPKGERGESG VAGASGAPGPPGAPGAVGPAGKSGDRGESGP
AGPAGIAGPAGPRGSPGAGARGDKGESGEAGERGMKGHRGFTGMQGP GPPGSPGQSGESGP
AGASGPAGPRGSPGSAGAAGKDGMSGLPGPIGPPGPRGRSGEMGSPGTGPPGPPGPPGPPG
PGGGFDMGFIAQPAQEKAPDPFRHFRADDANVMRDRDLEVDTTLKSLSQQIENIRSPGT
KKNPARTCRDLKMCHPDWKSGEYWIDPDQGCNQDAIKVYCNMETGETCVYPT EADIPKKS
WYTSKNIKEKKHVWFGEAMTDGFQFEYGSE-GSNAKDVNIQLTFLRLMATEASQNIITYHC
KNSIAYMDQASGNLKKALLLQGSNEIEIRAEGNSRFTYSVTEDGCTSHTGAWGKTVIDYK
TTKTSRLPIIDIA PMDVGAPNQEFG--IEVGPVCFL-----

>Paralichthys

MFSFVDLRLALLLSAAVLLVRAQGE-----DDRAGGSCTLDGQVFADRDVWKPEPCQI
CVCDSGTVMCDEVICEDTT-DCPNPVI PHDECCPICP-----DDGFQDSQVEGPKGERGP
KGDRLPGPPGNDGIPGQNLPGPPGPPGPPG---LGGNFSPQMSSGGYD----DKSPG-M
SVPGPMGPMGPRGPPG PAGSSGPQGFTGPPGEPGEAGAS---GPMGPRGAAGPPGKNGE
DGESGKPRGGERGPGSGPQGARGFPPTPLPGIKGHRGFSGLDGAKGDSGPAGPKGEAGT
PGENGT PGAMGPRGLPGERGRTGATGSAGARGNDGAAGAAGPPGPTGPAGPPGFPPGPGA
KGE GGPQGSRGPEG PAGARGE PGSAGPAGAAGPAGNPGTDGAPGAKGLPGSAGVAGAPGF
PGPRGPPGAQGAAGATGPKGNTGEAGAPGAKGESGAKGESGVAGVQGPFGPSGEEGKRG
RGEPGAAGVRGPPGERGGPGGRGFP GSDGSAGAKGAPGERGAPGLLGPKGSTGEPGRTGE
AGLPGAKGMTGSPGSPGPDGKMAAGAPGDGRPGPPG SVGARGQPGVMGFPGPKGAAGE
GGKPGERGVMGPTGPVGA PGKDG DVGAQGPSGSPGAGERGEQGAAGGPGFQGLPGPQGA
VGETGKPGEQGLPGEAGATGPAGARGDRGFPGER GAPG INGPAGARGSPGSGNDGAKGD
AGAPGAPGAQGP PGLQGMPGERGSAGLPGLRGDRGDQGAKGADGAPGKDGPRGMTGAI GL
PGAAGATGDKGEPGAPGIVGPAGARGGPGERGESGPPGPAGFAGPPGADGQPGAKGEAGD
NGAKGDAGPPGSPGPTGAAGPGGPVGNTPGPKGARGPAGPPGATGFPGSAGRVGPPGSPGN
AGPAGPPGATGKEGQKGNRGETGPAGRPGEMGAAGLPGPSGEKGNPGAEGAPGSSGIPGP
QGINGGRGIVGLPGQRGERGF PGLPGPSGEVKGHGSPGN GERGPPGPMGPPGLAGAPGE
AGREGAPGNEGSAGRDGAAGPKGDRGESGPAGASGAPGPPGAPGVPGPAGKSGDRGESGP
AGPAGSAGPAGPRGPAGAPGLRGDKGESGEAGERGMKGHRGFTGMQGP PG PAGASGDQGP
AGAAGPAGPRGSPSGSAGSAGKDGMSGLPGPTGPPGPRGRSGEMGPAGPPGPPGPAGAPGA
PGGGFDLGFIAQP- QEKAPDPYRMFRADDANVLRDRDLEVDSTLKSLSQQIEQIRSPDGT
RKNPARTCRDLKMCHPDWKS GEYWIDPDQGCTQDAIKVYCNMETGETCVSPSQREVAKKN
WYVSKNIKEKKHVWFGEAMNDGFQFEYGESE- GSAPEDVNIQMTFLRLMSTEASQNVTYHC
KNSVAYMDASVGNLKKALLLQGSNEIEIRAEGNSRFTYSVLEDGCTSHGTGTWGKTVIDYK
TSKTSRLPIIDIA PMDVGAPDQEFG--LEVGPVCFL-----

>Raja

MFSFVDTRLLLLLAATLVVGRCAE-----DDANEKGCICKDGQTYSDKDIWKPEPCKI
CVCDDGGQVLCDEIMCDEIN-DCPNAEIPFGECCAI CPDGTTYPTGPRIQPEGPKGEPGP
KGDRGLPGLPGRDGI PGTPGRDGP PPGPLPG--TAGGNFLPQMSGGYEGGYSEKTSGGI
PMPGPMGPMGPRGAPGPAGSPGPQGFQGRQGETGEAGAP---GPMGPRGPAGPPGKNGD
DGEHGKAGRPGDRGPPGPQGARGFPPTPLPGIKGHRGVNGLDGAKGDSGPAGPKGETGS
PGENGAPGAMGPRGIAGERGRPGATGPAGVRGNDGASGAAGPPGAVGSPGPPGFPAGPA
KGEAGPAGSRGADGATGARGEAGNPGPAGPTGAPGSAGSDGAPGAKGGVGPAGISGAPGF
PGPRGSPGPGQAVGATGPKGNNGDSGAPGAKGEAGAKGDSGIAGPQGAPGSPGEEGKRS
RGEVGVQGLPGSPGERGSAGNRGFPPEGLVGAAGIPGERGSPGPAGPKGATGESGRTGE
PGLPGTRGLTGS PGSPGADGKPGSPGAGQDGRPGPAGAI GQRGQPGVMGFPKGTAGE
SGKTGERGIVGPAGLVGLPGKDGDI GAPGATGPAGPAGERGEQGITGAPGFQGGPPQVGA
AGEHGKSGEQIPGAVGPPGSTGPRGERGFTGERGSPGAMGPTGARGSPGSSGSEKAGD
AGPPGAPGGQGAPGLQGMPGERGASGLPGPKGARGESGVKGADGAPKDGMRGMTGPIGP
PGHAGVVGEKGEAGPPGSPGVPVARGSTGDRGEPGAPGAGFAGPPGADGQPGAKGEAGD
TGAKGEAGPAGPTGAVGIAGPPGPTGATGAKGARGAAGSPGATGFPGGVGRVGP PPSGN
AGPTGLPGPPGKEGAKGIRGETGAVGRPGEPGIAGPAGPAGEKGPAGSDGSPGAPGLPGP
QGLSGTRGVLTGMRRGERGNMGPQQTGEAGKQGPVGPAGERGPPGPMGPPGLAGPSGE
AGREGSPGAEGAPGRDGA VAKGDRGEAGATGASGAPGAPGAPGVPVPSGKTGDRGEPGA
AGPAGSPGPMGPRGANGLTGPRGDRGESGENGARGIKGHRGFPGAQGLPGPPGPPGEGQ
SGPSGASGPRGPAGPSGSSGKDGASGMPGPIGPPGPRGRTGEIGVGP PPGPPGIPGPPGP
SHGGFDFQFAAPQIE-KGPDPLRYFRADDAGAVADRDIQVDTTLKS LTQQIESIRSPGT
RKNPARTCRDLKMCHADWKS GDYWDPNQGCILDVIRVYCNMETGETCVYSSPQSI PQKN
WYTTKNPKKHHVWFGENMDGGFQFGYGGDDGVLASDVAIQITFLRLMSSEATQNITYHC
KNSIAYMDEEAGNLKKA VLLQGSNEIEIRAEGNSRFTYSVTEDGCTTHTGDWGRVTIEYK
TTKTSRLPIIDIA PMDVGGADQEFG--VDIGPVCFL-----

>Rana

MFSFVDTRLLLLVIAATILVAKCQGE-----DDLGYSGCVVDGRTYNDKDVWKPEACQI
CVCDEGTILCDEVICEDIG-DCPNPEIPMGECCPVCG-E--GQYQTGSVVEGPKGETGP
RGERGPPGAPGRDGI PGQPGIPGPPGPPGAG---LGGNFAPQMSYGYD---EKSAG-I
SMPGPMGPMGPRGPPGSPGSPGPQGFQGPPEPEPEGAA---GAMGPRGPPGPPGKNGE
DGEAGKPRGPRGERGPPGPQGARGLPGTAGLPGMKGHRGFNGLDGAKGDTGPAGPKGEPGN
PGENGAPQVGPRLPGERGRPGSPGAGARGNDGTPGAAGPPGPTGPTGPPGFPVGP
KGDAGPQGSRGPDGPQARGEPEGAPQAGPAGSAGNPGTDGQPGAKGATGAPGIAGAPGF
PGARGAPGPQGGSPGPKGNNGEPAQGNKGEPGAKGESGPAGSQGPPGPPGEEGKRS
RGEVGPSPGPPGAGERGAPSRGFPAGDAGGPKGPPGERGPVGSAGPKGSPGESRPG
PGLPGAKGLTGS PGSPGPDGKTGPAGAPGQDGRPGPPGPPGARGQSGVMGFPKGAAGE
PGKPGERGVAGPPGAVGAPGKDGEAGA QGPPGAPGAGERGEQGPAGPPGFQGLPGSPGA

>Xenopus

MFSFVDTRTLLLLIAATILVARCQGE-----HDVQTSDCVQHGITYSNRDVWKPEDCQI
CVCDNGNILCDEVMCEDA--DCPNPVIIVPGECCPVCT-DND-AQYSEVTGVEGPKGDVGP
KGDKGIAGPPGRDGI PGQPGI PGPQP PPGPPG---LGGNFAPQMSYGYD----EKSAG-I
SMPGPMGPMGPRGPPPGSPGSPGPGQFQGPPEGEPGAS---GAMGPRGSSGPPGKNGE
DGEAGKPRPGERGPPGPGQARGLPGTAGLPGMKGHRGFNGLDGAKGDSGPAGPKGEPGS
PGENGAPQVGPRLSGERGRPGSPGAGARGNDGAPGAAGPPGSTGSPGPPGFPGGVGP
KGDAGPQGSRGSDGPQGRGEPGAPGQAGAAGPSGNPGSDGQPGAKGATGAPGIAGAPGF
PGARGAPGAQGGSPGPKGNNGEPGAQGNKGEAGAKGEPGAGVQGPSPGSEEGKRGS
RGEF-PAGPPGPAGERGGPGSRGFPGSDGASGPKGAPGERGPVGPAGPKGSSGESGRPGE
PGLPGAAGLTPGSPGSDGKTGPAGAPGQDGRAGPPGPPGARGQSGVMGFPGPKGAAGE
PGKNGEKGAVGPPGAVGLPGKDGDAQAQPPGAPGAGERGEQGPAGGPGFQGLPGSPGP
AGESGKPGEQGVPGDVGSPGAPARGERGFPGERGAQGPQGPQARGSNAPGNDGAKGE
AGAAGAPGGQGSPLQGMGERGSSGLPGAAGDRGDQGVKGS DGTGPKDGVRLTGP IGP
PGPGGAPGDKGEAGPAGPAGPTGARGAPGERGEPAPPGPAGFAGPPGADGQPGAKGEQGD
SGAKGDAGPSGLPGPTGAPGAPALGSPGPKGARGAPGPPGATGFPGAAGRVP GPPGSGN
AGPPGSPGAPAGKEGAKGPRGETGPAGRSGEPGAAGPPGPPGEGKSPGSDGPAGAPG IGP
QGVAGSRGTVGLPGMRGERGFSGLPGAPGEPGKQGSSGSPGERGPPGSPGPPGLGGPPGE
SGREGAPGSEGAPGRDGA VGPKGDRGE--AAGPPGAPGAPGAPGVPAGKSGDRGETGP
SGPAGPAGTAGARGPAGPQGPGRGDKGEAGEQGERGMKGHRGFNGPSGPPGPPGSSGEQGP
SGASGPAGPRGPPGSSGNPGKDGSNGLPGLIGPPGPRGRTGDVGPAGPPGP-GPPGPPGQ
SGGGFDFSFMPQPPEKSHDG-RFYRADDANVMRDRDLEVDSTLKSLSKQIENIRSPGT
KKNPARTCRDLKMCHSDWKSGEYWIDPNQGCILDAIKVYCNMETGETCIYPTQSSIPQKS
WYTSKNLREKKHVWFGEAMSDGFQFEYGESE-GSNSADVTIQLTFLRLMATEASQNITYHC
KNSVAYMDQATGNLKKALLLQGSNEIEIRAEGNSRFTYSVVEDGCTQHTGEWGKTVIDYK
TTKTSRLPITDVAPMDIGAPDQEFQFGIDIGPVCFV-----

>Canis

MFSFVDLRLLLLLAATALLTHGQE---EGQEEDI---TCVQNGRLRYDRDVWKPEACRI
CVCNDNGNVLCDDDVICDETK-NCPGAQVPPGECCPVCPSPLASPTDQETTGVGPKGDTGP
RGPRGPAGPPGRDGI PGQPGLPGPPGPPGPPGPPGLGGNFAPQMSYGYD----EKSTGGI
SVPGPMGSPGPRGLPGPPGAPGPQGFQGGPPGEPGEPGAS----GPMGPRGPPGPPGKNGD
DGEAGKPRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGS
PGENGAPQMGPRLPGERGRPGAPGAPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGA
KGEAGPQGARGSEGPQVVRGEPGPPGAPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGF
PGARGSPGQPSGPPGPKGNSGEPGAPGNKGDGTGAKGEPGPTGIQGGPPGAGEEGKRG
RGEPTGLPGPPGERGGPSRFGADGVAGPKGPAGERGSPGAPGKGSPEAGRPE
AGLPGAKGLTGSPPGPDGKTGPPGAPGDGRPGPPGPPGARGQAGVMGFPGPKGAAGE
PGKAGERGVPVPPGAVGPAGKDGEGAQGGPPGAPAGERGEQGPAGSPGFQGLPGPAGP
PGEAGKPEQGVPGDLGAPGSPGARGERGFPGERGVQGGPPGAPRGANGAPGNDGAKGD
AGAPGAPGSQAGPLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGP
PGPAGAPGDKGEAGSPGAPGTGARGAPGDRGEPGPPGAPGAPGPPGADGQPGAKGEPGD
AGAKGDAGPPGAPGTGPPGPIGNVAPGPKGARGGAPGATGFGAAGRVPVPPGSGN
AGPPGPPGAPGKEGGKARGETGPAGRPEVGPVPPGPPGAGEKGS PGADGPAGAPGTGPP
QGIAGQRGVVGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGE
SGREGSPGAEGSPGRDGS PGPKGDRGETGPAGPPGAPGAPGAPVGPAGKNGDRGETGP
AGPAGPIGPVARGPAGPQGPGRDKGETGEQGDRIKGRGFSGLQGGPPGPPGSPGEPQGP
SGASGPAGPRGPPSAGSPGKDLNGLPGPIGPPGPRGRTGDAGVGPVPPGPPGPPGPPG
PSGGFDFSLPQPPQEKADGGRYRADDANVVRDRDLEVDTTLLKSLSQIENIRSEGS
RKNPARTCRDLKMHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPQVAQKN
WYISKNPKEKRHVWYGESMTDGFQFEYGGQ-GSDPADVAIQLTFLRLMSTEASQNITYHC
KNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYSVTYDGTSTHTGAWGKTVEYK
TTKTSRLPIIDVAPLDVGPADQEFQ--MDIGPVCFL-MLSFVDTRTLLLLAVTSLCATCQ
CKCLRLVWARRGRKGTGDRG---PRGERGPPGPPGRDGD---GIPGPPG---PPGPPG
PPGLGGNFAGQYD-GKGVGLGPGPMGLMGPGRGPPGASGAPGQGFQGGPAGEPEGPQGTGP
AGARGPPGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPLPGFKGIRGHNLGDL
KGQPGAPGVKGEPPAGENGTPGQTGARGLPGERGRVGPAGPAGARGSDGSPVGPVGPAGP
IGSAGPPGFPAGPAPKGEIGPVGNPAGPAGPRGEVGLPGVSGVGPVGNPANGLTGA
KGAAGLPGVAGAPLPGPRGIPGPVGAAGATGARGLVGEPGAGSKGESGNKGEPPGSA
QGGPPSGEEGKRGPNGEAGSAGSPGPPGLRGSPPSRGLPGADGRAGVMGPPGPRGATGP
AGVRGPPGDSGRPEPGLMGPGRFPAGPNVGPAGKEGPMGLPGIDGRPGPIGPAGARGE
PGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGGPPGQGVQGGKGEQGP
AGPPGFQGLPGPAGTAGEVGPGERGLPGEFGLPGPAGPRGERGPPGESGAAGPSGPIGS
RGPSGPPGPDGNKGEVGLGAPGTAGASGPPGLPGERGAAGIPGGKGEKGETGLRGEIGN
PGRDGARGAPGAMGAPGAPGATGDRGEAGPAGPAGPRGSPGERGEVGPAGPNGFAGP
AGAAGQPGAKGERGTGPKGKENGVPVGTGPIGSAGPSGPNPAGPAGSRGDGGPPGATGF
PGAAGRTGPPGPGSITGPPGPPGAAGKEGLRGPGRDQGPVGRGTGETGASGPPGFTGEKGP
SGEPGTAGPPGTGPPQGLLGPAGILGLPGSRGERGLPGVAGSVGEPGPLGIAGPPGARGP
PGAVGAPGVNGAPGEAGRDGNPNDGPPGRDQAGHKGERGYPGNIGPVGAVGAPGPHGP
VGPTGKHGNRGEPGAGSVGPVAVGPRGSPGQIRGDKGEPGEKGPRLPGLKGNHNL
QGLPGLAGQHGQDQAGPSVGPAGPRGAPGSPGPKDGRGTGQPGTVGPAGIRGSQGSQGP
AGPPGPPGPPGPPGPPSGGGYDF--GYEGDFYRADQPRSPSLRPKDYEV DATLKSLNNQI
ETLLTPEGSRKNPARTCRDLRLSHPEWSSEGYWIDPNQGCMTDAIKVYCDFSTGETCIR
AQPENIPAKNHY--RNSKVKKHIWLGETINGGTQFEYNVEGVTTKEMATQLAFMRLLANH
ASQNITYHCNLSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNE
WRKTIIEYKTKNPSRLPILDIAPLDIGDADQEFRVDVGPVCFK

>Bos

MFSFVDLRLLLLLAATALLTHGQEEGQEEGQEEI PPVTCVQNGRLRYHDRVWKPVPCQI
CVCNDNGNVLCDDDVICDELK-DCPNAKVPTDECCPVCPEGQESPTDQETTGVGPKGDTGP
RGPRGPAGPPGRDGI PGQPGLPGPPGPPGPPGPPGPPGLGGNFAPQLSYGYD----EKSTG-I
SVPGPMGSPGPRGLPGPPGAPGPQGFQGGPPGEPGEPGAS----GPMGPRGPPGPPGKNGD
DGEAGKPRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGS
PGENGAPQMGPRLPGERGRPGAPGAPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGA
KGEAGPQGPARGSEGPQVVRGEPGPPGAPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGF

PGARGPSGPQGPSPPGPKGNSGEPGAPGSKGDTGAKGEPGPTGIQGPPGPAGEEGKRGAR
RGEPPGAGLPGPPGERGGPGSRGFPAGDGVAGPKGPAGERGAPGAPGPKGSPGEAGRPE
AGLPKAGKLTGSPGSPGPDGKTGPPGAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGE
PGKAGERVPGPPGAVGPAKDGEGAQGGPPGAPAGERGEQGPAGSPGFQGLPGPAGP
PGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGGPPGAPRGANGAPGNDGAKGD
AGAPGAPGSQGAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGP
PGPAGAPGDKGEAGSPGAPGTGARGAPGDRGEPGPPGAPGAGPPGADGQPGAKGEPGD
AGAKGDAGPPGAPGAPGPPGIIGNVGAPGPKGARGGAPGATGFPGAAGRVP
AGPPGPPGAPGKEGSKGPRGETGPAGRPGEVGP
QGIAGQRGVVGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGE
SGREGAPGAEGSPGRDGS PGAKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKSGDRGETGP
AGPAGPIGPVARGPAGPQGPGRGDKGETGEQGDRIKGRHGFSGLQGPPGPPGSPGEQGP
SGASGPAGPRGPPGSAGSPGKDGLNGLPGPIGPPGPRGRTGDAGPAGPPGPPGPPGPPG
PSGGYDL SFLPQPPQEKAHDGGRYRADDANVVRDRDLEVDTTLKSLSQQIENIRSP
RKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKN
WYISKNPKEKRHVWYGESMTGGFQFEYGGQ - GSDPADVAIQTLFLRLMSTEASQNI
KNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYSVTYDGTSTHTGAWGKT
TTKTSRLPIIDVAPLDVAGPDQEF - - FDVGPACFL - MLSFVDTRTL
C - - - - KCLQLVSEGPSGDRG - - - PRGERGPPGPPGRDGD - - - GIPGPPG - - - PPGPPG
PPGLGGVRLSVS - LTVSCL - SSSQGLMGRGPPGASGAPGQGFQGGPPGEPGEPGQTGP
AGARGPPGPPGKAGEDGHPGKPRGERGVVGPQGARGFPPTPLPGFKGIRGHNL
KGRQPGAPGVKGEAGENGTPGQTGARGLPGERGVGAPGAPGARGSDGSPVGPVGPAGP
IGSAGPPGVGAPGPKGELGPGVNGPPGAPGAPGPRGEVGLPGLSGVGP
KGAAGLPGVAGAPLPGPRGIPGPVGAAGATGARGLVGEPGAGSKGESGNKGE
PGPPGSPGEEGKRGSTGEIGPAGPPGPPGLRGNPGRGLPGADGRAGVMGPAGSRGATGP
AGVRGPNGDSGRPGEPGLMGRGFPGPSGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGE
PGNIGFPGPKGPSGDPGKAGEKGHAGLAGARGAPGPDGNNQAQGGPPGLQGVQGGKGEQGP
AGPPGFQGLPGPAGTAGEAGKPGERGIPEGFGLPGPAGARGERGPPGESGAAGPTGP
RGSPPGPDGNGKGEPPGVGAPGTAGPSGSPGLPGERGAAGIPGGKGEKGETGLR
PGRDGARGAPGAI GAPGAPGANGDRGEAGPAGPAGPAGPRGSPGERGEVGPAGP
AGAAGQPGAKGERGTTKGPKEGNGPVGPTGPVGAAGPSGPPGPPGAGSRG
PGAAGRTGPPGPGSISGPPGPPGAPGKEGLRGRGDQGPVGRSGETGASGPPG
SGEPGTAGPPGTPGPQGLLGAAGFLGLPGSRGERGLPGVAGSVGEPGLGIAGPPG
PGNVNPGVNGAPGEAGRDGNPNDGPPGRDQPGHKGERGYPGNAGPVAAGAPG
VGPVGNHGNRGEPPGAPGAVGAPGAVGPRGSPGQGI RGDGKGEKGPGRGLPGLK
QGLPGLAGHGDQGAPGAVGAPGPRGAPGSPGPKDGRIGQPGAVGAPGIRGSQGSQGP
AGPPGPPGPPGPPGSPGGGYEF - - GFDGDFYRADQPRSPTSLRPKDYEV
ETLLTPEGSRKNPARTCRDLRLSHPEWSS - GYYWIDPNQGCNLDAIKVYCF
AQPEDIPVKNWY - - RNSKAKKHVWVGETINGGTQFEYNVEGVTTKEMATQLAF
ASQNIYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKT
WQKTII EYKTNKPSRLPILDIAPLDIGGADQEI RLNIGPVCFK

>Macaca

MFSFVDLRLLLLLLAATALLTHGQEEGQVEGQDEDI PPIITCVQNGRLRYH
CVCDNGKVL CDDVI CDETK - NCPGAEVPEGECCPVC PDGSE SPTDQETT
RGPRGPAGPPGRDGI PGQPGLPGPPGPPGPPGPPGLGGNFAPQLSYGYD - - - - EKSTGGI
SVPGPMGSPGRGLPGPPGAPGPPGQGFQGGPPGEPGEPGAS - - - - GPMGPRGPPGPPGKNGD
DGEAGKPRGPRGERGPPGQARGLPGTAGLPGMKHGRGFSGLDGAAGDAGPAGPKGEPGS
PGENGAPQMGPRGLPGERGRPGAPGAPGARGNDGATGAAGPPGPTGPAGPPGFP
KGEAGPQGPGRGSEGPQGVREGEPPGAPGAGPAGNPGADGQPGAKGANGAPG
PGARGPSGPQGGPPGPKGNSGEPGAPGSKGDTGAKGEPGVPVQGGPPGAGEEGKRGAR
RGEPPGPTGLPGPPGERGGPGSRGFPAGDGVAGPKGPAGERGSPGAPGPKGSPGEAGR
AGLPKAGKLTGSPGSPGPDGKTGPPGAPGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGE
PGKAGERVPGPPGAVGPAKDGEGAQGGPPGAPAGERGEQGPAGSPGFQGLPGPAGP
PGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGGPPGAPRGANGAPGNDGAKGD
AGAPGAPGSQGAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGP
PGPAGAPGDKGETGSPGAPGTGARGAPGDRGEPGPPGAPGAGPPGADGQPGAKGEPGD
AGAKGDAGPPGAPGAPGPPGIIGNVGAPGPKGARGGAPGATGFPGAAGRVP
AGPPGPPGAPGKEGKGRGETGPAGRPGEVGP
QGIAGQRGVVGLPGQRGERGFPGNPGPS - - - - -
- - - - - GDRGETGPAGPPGAPGAPGAPGVPVGPAGKSGDRGETGP
AGPAGVPVGPARGPAGPQGPGRGDKGETGEQGDRIKGRHGFSGLQGPPGPPGSPGEQGP
SGASGPAGPRGPPGSAGTPGKDGLNGLPGPIGPPGPRGRTGDAGPVP
PSGGFD SFLPQPPQEKAHDGGRYRADDANVVRDRDLEVDTTLKSLSQQIENIRSP
RKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKN

WYISKNPDKKRHVWFGESMTDGFQFEYGGG - GSDPADVAIQLTFLRLMSTEASQNITYHC
KNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTVIEYK
TTKTSRLPIIDVAPLDVGPADQEFQ - - FDVGPVCFL - - - - -

>Pan

MFSFVDLRLLLLLLAATALLTHGQEEGQVEGQDEDI PPIITCVQNGRLRYHDRDVKPEPCRI
CVCDNGKVL CDDVICDETK - NCPGAEVPEGECCPVC PDGSESPTDQETTGVGPKGDTGP
RGPRGPAGPPGRDGI PGQPGLPGPPGPPGPPGPPGLGGNFAPQLSYGYD - - - - EKSTGGI
SVPGPMGSPGRPLPGPPGAPGPQGFQGPPEGEPGAS - - - - GPMGPRGPPGPPGKNGD
DGEAGKPRPGERGPPGQGARGLPGTAGLPGMKGHRGFSGLDGAAGDAGPAGPKGEPGS
PGENGAPQMGRPLPGERGRPGAPGAPARGNDGATGAAGPPGPTGPAGPPGFPAGVA
KGEAGPQGRGSEGPQVREGEPPGAPAGAGPAGNPGADGQPGAKGANGAPGIAGAPGF
PGARGSPGPPGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPAGGEEGRGA
RGEPPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGAPGKSPGEAGRPE
AGLPGAKGLTGSPPGPDGKTGPPGAPQDGRPPGPPGARGQAGVMGFPGPKGAAGE
PGKAGERGVPPGAVGPAGKDGEAGAQQPPGAPGAGERGEQGPAGSPGFQGLPGPAGP
PGEAGKPEQGVPGDLGAPGSPGARGERGFPPGERGVQGPAGPRGANGAPGNDGAKGD
AGAPGAPGSQAGPLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGP IGP
PGPAGAPGDKGESGSPGAPGPTGARGAPGDRGEPGPPGAPGAGPPGADGQPGAKGEPGD
AGAKGDAGPPGAPGAPPPGPIGNVGPAGKARGSPGATGFPGAAGRVP GPPGSGN
AGPPGPPGAPGKGGKPRGETGPAGRPGEVGP GPPGAPGKSPGADGPAGAPGTGPP
QGIAGQRGVVGLPQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGE
SGREGAPGAEGSPGRDGSAGKDRGETGPAGPPGAPGAPGAPGVPAGKSGDRGETGP
AGPAGVPVGPARGPAGPQGRDGETGEQDRIKGRGFSGLQGP GPPGSPGEPQGP
SGASGPAGPRGPPGAGAPGKDLNGLPGPIGPPGPRGRTGDAGVP GPPGPPGPPG
PSAGFDFSLPQPQEKAHDGGRYRADDANVVRDRDLEVDTTLKSLSQIENIRSPGGS
RKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKN
WYISKNPDKKRHVWFGESMTDGFQFEYGGG - GSDPADVAIQLTFLRLMSTEASQNITYHC
KNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTVIEYK
TTKTSRLPIIDVAPLDVGPADQEFQ - - FDVGPVCFL - MLSFVDTRTLLLLLAVTLCLATCQ
S - - - - - LQEETVRKGPAGDRG - - - - PRGERGPPGPPGRDGED - - - - GPTGPPG - - - - PPGPPG
PPGLGGNFAAQYD - GKGVGLGPGPMGLMGRGPPGAAGAPGQGFQGPAGEPEGEPQTGP
AGARGPAGPPGKAGEDGHPGKPRPGERGVVGPQARGFPPTPLPGFKGIRGHNL DGL
KGQPGAPGVKGEPPAGENGTPGQTGARGLPGERGRVGPAGARGSDGSVGPVGPAGP
IGSAGPPGFPAGPPKGEIGAVGNAGPAGPAGPRGEVGLPGLSGVPVPPGNPGANGLTGA
KGAAGLPGVAGAPLPGPRGIPGPVGAAGATGARGLVGEPGAGSKGESGNKGEPPGAGP
QGPAGPSGEEGKRPNGEAGSAGPPGPPGLRGSPPSRGLPGADGRAGVMGPAGSRGASGP
AGVRGPNGDAGRPEGLMGRPLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGE
PGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNGAQQPPGQGVQGGKGEQGP
AGPPGFQGLPGPSGPTGEVVKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGS
RGPSGPPGPDGNKGEPPVAVGTAGPSGSPGLPGERGAAGIPGGKGEKGEPLRGEIGN
PGRDGRGAPGAVGAPGAPGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGP

AGAAGQPAGKGERGAKGPKGENGVVGTGPVGAAGPAGPNGPPGAGSRGDGGPPGMTGF
PGAAGRTGPPGPSGISGPPGPPGAGKEGLRGRGDQGPVGRTEVAVGPPGFAGEKGP
SGEAGTAGPPGTGPPQQLLGAAGILGLPGSRGERGLPGVAVAVGEPGLGIAGPPGARGP
PGAVGSPVNGAPGEAGRDGNPNNDGPPGRDQPGHKGERGYPNIGPVGAAGAPGPHGP
VGPAGKHGNRGETGSPGVPVAVGPRGSPGQGIKGEKGPGRGLPGLKGHNGL
QGLPGIAGHGDQAGPSVGPAGPRGPAGPSGPAGKDRGTGHPGTVPAGIRGPQGHQGP
AGPPGPPGPPGPPGVSGGYDF--GYDGFYRADQPRASPLRPKDYEV DATLKLNNQI
ETLLTPEGSRKNPARTCRDLRLSHPEWSS-GYYWIDPNQGCTMDAIKVYCDFSTGETCIR
AQPENIPAKNWF--RSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLAN
ASQNTIYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNE
WGKTIIEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK

>Echinops

-----PSLSCVQDGLKYKNRDVWKPEPCRI
CVCDKGTVLCDKVTCPDND-KCPGAIKNPGECCAI CPDGEASPTDQETTGVGPKGDTGP
RGPRGPAGPAGRDGIPGQPG-----
-----GPSGPRGLPGPPGAPGPPQGFQGPAGEPEGEPGAS----GPMGPRGPPGPPGKNGD
DGEAGKPRAGERGPPGPPQAGRLPGTAGLPGMKGHR-----

-GESCPQGTTRGSEGPQCVRGEVPPP--AGAA-PARNPGDDEQPGGKANGAPGIAGAPGF
PGARGSPGPPGSPGAPGPKNSGEPGAPGNKGDAGAKGEPGPTGVQGPAGGEEGKRG
RGEPPGSLPGPPGERGGPGSRGFPGSDGVAGPKGPAGERGSPGAGPKGSPGESGRPGE
AGLPGAK-----

-----GVPDGLGAPGSPGARGERGFPGERGVQPPGAPGPRGSNGAPGNDGAKGD
AGAPGAPGSQAGPLQGMPPGERGAAGLPGPKGDRGDAGPKGADG-----
-----GESGSPGAPPTGARGAPDRGEPGAPGAGFAGPPGADQPGAKGEPGD
AGAKGDAGLSGPAGPTGAPPIGNVAVGPKGARGSPGPPCATG-PGAA-RVAPPS----

-----GP
AGPAGPIGPVAVGARGSPGPPQAGARGDKGETGEQDGRGMKGHRGFSGLQGPPLPGSPGEQGP
SGASGPAGPRGPPGSPGSPGKDGSLGPGPIGPPGPRGRTGDAGVAVGPPGPPGPPGPPG
PSGPFDFFSFMPQPPQEKAHDDGGYYRADDANVVRDRDLEVDTTLKSLSQOIEINIRSPG
RKNPARTCRDLKMHSDWKSGEYWIDPNQGCTLDAIKVFCNMETGETCVYPTQPSVAQKN
WYISKPNKESHVWYGESMTDGFQFEYGGG-GSDPADVAIQLTFLRLMSTEASQNTIYHC
KNSIAYMDQQTGNLRLKALLLQGSNEIEIRAEGNSRFTYSVIVSDGCTSHTGAWGKTIIEYK
TTKTSRPLIIDVAPLDIGAPDQEF--FDISPVCFL-MLSFVDMRTLLLLAVTSLCLATCQ
S-----LQTETARKGPTGDRG---PRGDRGPAGPPGRDGED---GIPGPPG---PPGPPG
PPGLGGNFAAQYD-GKGVGLGPPMGLMGRGPPGANGPPGPPGFNGPAGEPEGEPGQTGP
AGSRGPAGPPGKAGEDGHPGNPERSGERGVV--GARGFPPTPLPGFKIRGHNGLDGL
KGQPGAAGVKGDAEFTGQ-----GARGLPGERGRVGGSPGARGSDGSPVAVGPPG
LGSAGPPGFPAGPPKAGELGPPVGNPSPGPPGPRGEVGLPGVSGVPPGNPANGLAGA
KGAAGLPGVAVAGPLPGPRGIPGAGSAGATGARGLVGEPGAGSKGESGSKGEPGSAVP
QG-----PA---GSPGSRGLPGADGRAGVMGPPGNRGASGP
AGSRGSPGSDSRPGEPLMGPRGLPGSPGNVGPAGKEGLGGLPGIDGRPGPTGPAGARGE
PGNIGFPGPKGPTGDAGKPKDKGHAGLAGPRGAPGPDGNGAQQGPPGPPGQVQGGKGEQGP
AGPPGFQGLPGPAGPTGEVVKPGERGLHGEFGLPGPAGPRGERGPPGQSGAAGPTGSI
RGPSGPPGPDGKGEVAVGAPGTAGASGPPGGLPGEWGAACVPGGKGEKGETGLRGEIGN
TGRDARGAVGAPGAPGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNFAGP
AGAAGQAGPKGERGTTKPKGENGAVGPTGPIGSAGSPGPPGPPGAGSRGDGGPP-----
-----GITGPPGPPGAAGKEGLRGRGDQGPVGRTEVAVGPPGFAGEKGP
SGEPGTAGPPGTGPPQILGPPGILGLPGSRGERGLPGVAGSLE-PGPLGISGPPGARG--
-GAVGNPGVNGA-----GNPDSGPPGRDGLPGHKGERGYPNAGPVGNAAGAPGPHGS
VGPAGKYGNRGEPPGAVGPPVAVGPRGSPGQGIKGEKGPGRGLTGFKGHNGL
QGLPGLAGQHDQGSPTVGPAGPRGPAGPSGPAGKDRGSGHPGAVGAVGVRGSSQGSQGP
SGPPGPPGPPGPPGVSGGYDF--GYDGFYRADQPRASPLRAKDHEVDATLKLNNQI
ETLLTPEGSKKNPARTCRDLRLSHPEWSS-GYYWIDPNQGCTMDAIKVYCDFSTGETCIR
AQPESIPVKNWF--QNSKVKKHVVWGETINGGSQ-----

-----RKTNE

WGKTIIEYKTNKPSRLPILDIAPLDIGGADQELIVDIGPACFK

>Mammut

-----GPPGPPGKNGD
DGEAGKPRPGERG--PQGARGLPGTAGLPGMKGHR-----GDAGPAGPKGEPGS
PGENGAPQMGR-----GNDGATGAAGPPGPTGPAGPPGFPGAVGA
KGEAGPQGARGSEGPGVRGEPGPPGAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGF
PGARGPSGPQGSPGAPGPKNSGEPGAQGNKGDGTGAKGEPGPVGIQGPPGPAGEEGKR--
-GEPGPTGLPGPPGER-----GFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPE
AGLPGAKGLTGSPPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTAGE
PGK----GVPGPPGAVGPAGKDGEAGAQQPPGPAGPAGER-----
-----GVQGPPGPAGPR-----GD
TGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGPRGLTGPIGP
PGPAGAPGDKGEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAK-----
----GDAGPPGPAGPTGAPGPIGNVGASGPK--GSAGPPGATGFPGAAGR-----
-----GPRGETGPAGRPEVGPVGPVGPAGEKG-----
-----GVVGLPGQR--GFPLPGPSGEPGKQGPSGSSGER-----
--EGAPGAEGSPGRDGSPPGAKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKSGDRGETGP
AGPAGPAGPAGVVRGAPGQPRGDKGETGEQDRL--GFSLQGPPGPPGSPGEQGP
SGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPR--TGDAGPVGPVGPVGPVGPVGP
P-----

>Tyrannosaurus

-----GATGAPGIAGAPGF
PGARGAPGPQGSPGAPGPK-----
-----GVQGPPGPQGR-----
-----GSAGPPGATGFPGAAGR-----
-----GVVGLPGQR-----

PGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGGPPGQGVQGGKGEQGP
PGPPGFQGLPGPSGPAGEVVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGS
RGPSGPPGPDGKNGKEPGVVGAVGTAGPSGSPGLPGERGAAGIPGGKGEKGEPLRGEIGN
PGRDARGAPGAVGAPGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGP
AGAAGQPGAKGERGAKGPKGENGVVGTGPVGAAGPAGPNGPPGAGSRGDGGPPGMTGF
PGAAGRTGPPGPGSISGPPGPPGAGKEGLRGPRGDQGPVGRGTGEVAVGPPGFAGEKGP
SGEAGTAGPPGTGPGQGLLGPAGILGLPGSRGERGLPGVAVGAVGEPGLGIAGPPGARGP
PGAVGSPGVNAPGEAGRDGNPNDGPPGRDQPGHKGERGYPGNIGPVGAAGAPGPHGP
VGPAGKHGNRGETGSPGVPVGPAGAVGPRGSPGQIRGDKGEPGEKGPRLPGLKGNHGL
QGLPGIAGHHGDQGAPGSPGVPAGPRGPAGSPGAGKDGRTGHPGTVPAGIRGPPQHQP
AGPPGPPGPPGPPGVSAGGYDF--GYDGFYRADQPRSPSLRPKDYEV DATLKSLNNQI
ETLLTPEGSRKNPARTCRDLRLSHPEWSS--GYWIDPNQGCTMDAIKVYCDFSTGETCIR
AQPENIPAKNWI--RSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLANH
ASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNE
WGKTII EYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK

>Loxodonta

-----GPKGDTGP
RGRPGPAGPPGRDGI PGQPGLPGPPGPPGPPGPPGLGGNFAPQLSYGYD----EKSAGGI
SVPGPM-PSGLRGL-----

-----GAPGPTGPAGPPGFPGAVGA
KGEAGPQGARGSEGPQGVREPGPPGAPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGF
PGARGPAGPQGPSAGAPGPKNSGEPGAPGSKGDAGAKGEPGVPVGIQGGPPGAGEEGKRG
RGEPTGLPGPPGERGGPGSRGFPADGVAGPKGPAGERGSPGAPGKGSPEAGRPGE
AGLPGAKGLTGS PGSPGPDGKTGPPVPAGQDGRPGPPVP--GARGQAGVMGFPGPKGAAG--

-----VPGDLGAPGSPGARGERGFPGERGVQGGPPGAPRGSNGAPGNDGAKGD
AGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGPRGLTGPIGP
PGPAGAPGDKGEAGPSGPAGPTGARGAPGDRGEPGPPGAPGAGPPGADGQPGAKGEPGD
AGAKGDAGPPGAPGPTGAPGPIGNVGAAGPKGARGSPGATGFPGAAGRVP GPPGSPGN
AGLSGPPGSAGKECGKGRGETG-----

-----GEPGKQAGPSSGDRGPPGVPGLTGPAGE
PGREGPAG-----

--PAGPAGPAGVRGPAGPQGPGRDKGETGEQDRGLKGHRGFSGLQGGPPGPPGSPGEQGP
SGASGPAGPRGPPGSAGAPGKDLNGLPGPIGPPGPRGRTGDAGVP GPPGPPGPPGPPG
PSGAFDFSFLPQPPEKAHDGGRYYRADDANVVRDRDLEVDTTLKSLSQIENIRSPG
RKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCESMTDE-----

-----QFEYGGGE--GSDPADVAIQTLFLRLMSTEASQNITYHC
KNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYSVTEDGCTSHGTWGTII EYK
TTKTSRLLPIIDVAPLDVAGAPDQEF--FDIGPVCFL-----

-----ATTRKGP GDRG--PRGDRGPTGPPGRDGED--GIPGPPG--PPGPPG
PPGLGGNFAAQYD--AKGIGLGPMPGLMGPRGPPGATGPPGSPGFQGGPPGEPGEPGTGP
AGSRGPAGPPGKAGEDGHPGKPRGPRGERGVVGPQGARGFPPTGGLP-----GHNGLDGL
KGQPGAPGVKGEPPGAPGENTPGQIGARGLPGERGRVGGPGPAGARGSDGSVGPVGPAGL
IGSAGSPGFSGS--GLK-----GNPGANGLAGA
KGAAGLPGVAGAPGLGPRGIPGPVGAAGATGARGLVGEPGPAGSKGESGSKGEPGSAGP
QGPPGSPGEEGKRGSSGEAGSAGPAGPPGLRGGPGSRGLPGADGRAGVMGPPGSRGASGP
AGVRGSPGSDSRGPEPGVMGPRGLPGSPGNVGPAGKEGPAGLPGIDGRPGPIGPAGARGE
PGNIGFPGPKGPA----KNGDKGHAGLAGPRGAPGPDGNNGAQGGPPGLQGVQGGKGEQGP
AGPPGFQGLPGPSGTAGEAGKPGERGI PGEFGLPGPAGPRGERGPPGQSGAAGPTGPIGS
RGPSGPPGPDGKNGKEPGVVGAPGTAGPSGPGGLPGERGAAGIPGGKGEKGETGLRGDTGN
TGRDARGAPGAVGAPGATGDRGEAGPAGSAGPAGPRGSPGERGEVGPAGPNGFAGP
AGAAGQAGAKGERGKGPKEGNGPVGPTGPVGAAGPAGPNGPPGAGSRGDGGPPGATGF
PGAAGRTGPPGPGITGPPGPPGAAGKEGLRGPRGDQGPVGRGTGETGASGPPGFAGEKGS
SGEPGTAGPPGTGPGQILGPPGILGLPGSRGERGLPGVAVGAVGGLGPLGIAGPPGARGP
PGAVGSPGVNAPGEAGRDGNPNDGPPGRDGLPGHKGERGYPGNAGPVGTAGAPGPPG
LGPAGKHGNRGEPGAGSVGPVAVGPRGSPS-----

-----GQHGDQGS PGSPGPRGPPGSPGVPKDGPRGHAGAVGPAGVRSQGSQGP
SGPPGPPGPPGPPGSPGGGYDF--GYDGFYRADQPRSPSLRPKDYEV DATLKSLNNQI
ETLLTPEGSKKNPARTCRDLRLSHPEWSS--GYWIDPNQGCTMDAIKVYCDFSTGETCIR
AQPENIPAKNWI--RSSKAKKHIWFGGETINGGTQIEYNEEGVTTKDMATQLAFMRLLANH
ASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCS-----

>Monodelphis

MFSFVDPRLLLLLLAVTAVLTHGQD-----EEDIPEGTCIQNAIKYNNGEVWPKKICQI
CVCDNGSILCDEIICEDVS-NCPNVQYKDDECCPSC--AVASPSLSTEIGVEGPKGDTGP
RGERGPPGLPGKDGKPGNPGIQGPPGLPGIPG--LGGNFASQMSYGYD----EKSGGGM
SVPGPMGSPGRGLPGPPGNPGPQGFQGGPPGEPGEPGAS----GPMGPRGPAGPPGKNGD
DGEAGKPRGPRGERGPPGQARGLPGTAGLPGMKGHRGFSGLDGAKGDSGPAGPKGEPGS
PGENGAPQMGRGLPGERGRPGPPGARGNDGATGAAGPPGPTGPAGPPGFPGAVGA
KGEAGPQGSRGSEGPQGVREPGPPGAGAAGPSGNPGADGQPGAKGANGAPGIAGAPGF
PGARGPSGPPGSPGAPGPKNSGEPGAPGNKGDPAKGEPPVGVQGGPPPAGEEGKRGS
RGEPPSGLPGPAGERGGPSRGFPAGDGVAGPKGAPGERGAPGAPGKSPGEAGRPE
AGLPGAKGLTGSPPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPGPKGAAGE
PGKAGERGVPPGAVGAAGKDGEAGAQQPPGAPAGERGEQGPAGSPGFQGLPGPAGP
PGEAGKPEQGVPGDAGAPGSPGARGERGFPPGERGVQGGPPGQPRGSNGAPGNDGAKGD
AGAPGAPGGQGPGLQGMPPGERGAAGLPGAKGDRGDAGPKGADGAAGKDGVRGLTGPIGP
PGPAGPTGDKGESGSPGVPGTGARGAPGERGEPGPPGAGFAGPPGADGQPGAKGEPGD
AGAKGDAGPPGAPGTGAPGAGNVGAPGPKGARGNAGPPGATGFPAGAARVGGPPGSPGN
AGPPGPPGAPGKEGGKPRGETGPIGRPEVGGPPGPPGSPGSEKGS PGADGPAGAPGTGPP
QGIAGQRGVVGLPGQRGERGFPLPGSPGEPGKQGPSGISGERGPPGAPPPGLAGPPGE
SGREGSPAEGSPGRDGSPPGKDRGETGPAGPPGAPGAPGAPGVPAGKSGDRGETGP
AGPAGVPVGTGARGPSGPPGPRGDKGETGEQDGRGMKGHRGFSGLQGGPPGPPGSPGEPGP
SGASGAPGRGPPGSPGASGKDLNGLPGPIGPPGPRGRTGDAGPAGPPGPPGAPGPPG
PSGGFDFSLRQPPPEKAHDSGRYYRADDANV-RDRDLEVDTTLKSLLTQQIENIRSEPT
RKNPARTCRDLRMCHSDWKSSEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPKVPLKN
WYINKNPKDKKHVWFGESEMTDGFQFEYGGG-GSDPADVAIQMTFLRLMSTEASQNITYHC
KNSVAYMDQQTGNLKKALLLQGSNEIEIRAEGNSRFTYGVTEGCTSHGTGNWGTIVIEYK
TTKTSRLPIIDVAPMDIGAPNQEF--CDVGPVCF-----
-----SLLKGPQGDGR--PRGPRGPPGLPGRDGED--GPQGPPG--PPGLPG
LGG--NFAAQYDASKGLDMGPGPM-----

-----GP
LGSAGPPGFPAGPKVSEGGPPAPRLHHQLPPRPSLGLPGMTGPVGPAGNPGANGLTGA
KGAAGLPGVAGAPGLPGPRGIPGAGAAGASGPRGLAGEPGPAGSKGESGNKGEPPGSA
QGGPPGNGEEGKRGPNGEPPGSSGPPGPPGLRGVPGSRGLPGADGRAGGMGPPGNGRSGP
AGVPPNGDAGRPPGEPGLMGRGLPGSPGNSGPTGKEGPAGLPGADGRPGPTGPAGNRGE
PGNIGFPGPKGPTGDPGKSGEKHAGLAGARGAPGPDGNNGAQGGPPGAVQGGKGEQGP
AGPPGFQGLPGSPGAGEGKVGGERGLAGEFGLPGPAGPRGERGPPGESGAVGPTGSIGS
RGPSGPPGPDGKGEPPVVGAPGNAGPAGSGVPPGERGAAGVPGGKGDGETGPRGEFGN
PGRDGR-----GSPGPPGPIGGRGDRGPPGATGF
PGAAGRTGSPGAPGAPWSPPG-----
-----GPPGSSGPPGILGAPGILGLPGSRGERGLPGVSGSLGEPGLGIAGPPGARGP
PGAVGSPGVNAPGEAGRD-----

-----KDMATQLFFMRLLANH
ASQNLTYHCKNSVAYMDAESGLKKAIVILQGSNDVELTAEGNSRFTFSVLEEDGCSRKTNE
WGKTIIIEYKTNKPSRLPFLDIAPLDIGGADQEFYLDVGPVCFK

>Alligator

-----GDTGP
RGPRGPAGPPGR-----SAG-V
TVPGPMGPAGPR-----QGPPGEPGEPGAS----GPMGPRGPAGPPGKNGD
DGEAGKPRGPRGQRPSPGQARGLPGTAGLPGMK--GFSGLDGAKGDAGPAGPKGEPGS
PGENGAPQMGR-----GRPGSPGARGNDGAPGAAGPPGPTGPAGPPGFPGAVGA
KGEAGPQGARGSEGPQGV-----GATGAPGIAGAPGF
PGARGPSGPPGSPGAPGPKNSGEPGAPGNKGDGAKGEPPGAVQGGPPPAGEEGKR--
-GEPGPAGLPGPAGERGGPSRGFPAGDGIAGPKGPAGERGSPGAVGPKGSTGESGRPE
PGLPGAKGLTGSPPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPGPKGAAGE
PGKAGERGAPGPPGAVGAAGK-----GPAGPAGER-----PGFQGLPGPAGP
-----PGNAGAPGPAGAR-----GVQGGPPGAPPRGANGAPGNDGAKGD
AGAPGAPGNEGPPGLEGMP--GAAGLPGAK-----GLTGPIGP
PGPAGAPGDKGEAGPPGAPGTGARGAPDRGEPGPPGAGFAGPPGADGQPGAKGESGD

AGPKGDAGPPGAGPTGAPGPAGNVGAPGPKGPRGAAGPPGATGFPGAAGR VGPVPPGPPSGN
AGPPGPPGAGKEGGKGRGETGPAGRPGEPGAGPPGPPGEGKSGADGPAGAPGTPGP
QGIAGQRGVVGLPGQR--GFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGE
AGREGAPGAEGAPGRDGAAGAKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKNGDRGETGP
AGPAGPAGPSGVRGPAGPQGPRGDKGETGEQGDRIKGHRGFSGLQGPPGPPGSPGEQGP
SGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGR-----

-----ALLLQGANIEIR-----

-----GSNGEPGSAGPPGPAGLR-----

-----GLPGESGAVGPVGPIGS

R-----

>Struthio

--PGPMGPAGPR-----GPAGPPGKNGD
DGEAGKPRPGQRGPPGQGAR-----GEPGS
PGENGAPQMGPR-----GRPGSPGAGARGNDGAPGAAGPPGPTGPAGPPGFPGAAGA
KGEAGPQARGSEGPGVRGEPGPPGAGAAGPAGNPGADGQPGAKGATGAPGIAGAPGF
PGARGPSGPQGPSGAPGPKGNSGEPGAPGNKGDGTAK-----QGPPGPAGEEGKRG
RGEPGPAGLPGPAGER-----PGADGIAGPKGPPGERGSPGAVGPKGSPGEAGR--
-----GLTGSPPGSPGPDGKTGPPGPAGQDGR-----GQAGVMGFPGPKGAAGE
PGKPGER-----PGPAGPAGER-----
-----GERGFPGERGVQGPPGQGP-----
-----GMPGERGAAGLPGAKGDRGDPGPKGADGAPGKDGLR-----
-----GSAGPPGATGFPGAAGR-----
-----EGGKGARGETGPAGRPGEPGPAGPP-----
-----GVVGLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGL-----
---EGAPGAEGAPGRDGAAGPKGDRGETGPAGPPGAPGAP-----NGDRGETGP
AGPAGPPGAGARGPAGPQGPRGDKGETGEQGDRIKGHRGFSGLQGPPGPPGAPGEQGP
SGASGPAGPRGPPGSAGAAGKDGLNGLPGPIGPPGPRGR-----

-----ALLLQGSNEIEIR-----

-----GSNGEPGSAGPPGPAGLR-----

-----GLPGESGAVGPAGPIGS
R-----

>Gallus
MFSFVDSRLLLLIAATVLLTRGEGE-----EDIQTGSCVQDGLTYNDKDVWKPEPCQI
CVCDSGNILCDEVI CEDTS - DCPNAEIPFGECCPICPDVDASPVYPESAGVEGPKGDTGP
RGDRGLPGPPGRDGI PGQPGLPGPPGPPGPPG---LGGNFAPQMSYGYD----EKSAG-V
AVPGMMPAGPRGLPGPPGAPGPQGFQGGPPGEPGEPGAS---GPMGPRGPAGPPGKNGD
DGEAGKPRPGQRGPPPGQARGLPGTAGLPGMKGHRGFSGLDGAKGQPGPAGPKGEPGS
PGENGAPQMGPRGLPGERGRPGSPGARGNDGAPGAAGPPGPTGPAGPPGFPGAAGA
KGETGPPQARGSEGPQSRGEPGPPGAGAAGPAGNPGADGQPGAKGATGAPGIAGAPGF
PGARGPSPGQPSGAPGPKNSGEPGAPGNKGDGTAKGEPGPAGVQGGPPGAGEEGKRGA
RGEPGPAGLPGPAGERGAPGSRGFPGADGIAGPKGPPGERSPGAVGPKGSPGEAGRPGE
AGLPGAKGLTGSPPGPDGKTGPPGAPGDGRPGPAGPPGARGQAGVMGFPGPKGAAGE
PGKPGERGAPPPPAGVAAGKDGEAGAQQPPGPTGPAGERGEQGPAGAPGFQGLPGPAGP
PGEAGKPGEQGVPGNAGAPGPAGARGERGFPGERGVQGGPPGQPRGANGAPGNDGAKGD
AGAPGAPGNEGPPGLEGMPPGERGAAGLPGAKGDRGDPGPKGADGAPGKDGLRGLTGP
PGPAGAPGDKGEAGPPGPAGPTGARGAPGDRGEPGPPGAGFAGPPGADGQPGAKGETGD
AGAKGDAGPPGAPGTGAPGAGNVGAPGPKGARGSPAGATGFPGAAGRVGPPGPGSGN
IGLPGPPGAGKEGSKGPRGETGPAGRPGEPPGPPGPPGEGKSPGADGPIGAPGTPGP
QGIAGQRGVVGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGE
AGREGAPGAEGAPGRDGAAGPKGDRGETGPAGPPGAPGAPGAPGVPAGKNGDRGETGP
AGPAGPPGPAGARGPAGPQGPGRGDKGETGEQGDGMKGRGFSGLQGGPPGPPGAPGEQGP
SGASGADPRGPPGSGAAGKDGLNGLPGPIGPPGPRGRTGEVGPVGGPPGPPGPPGPPG
PSGGFDFSLPQPQEKADHGGRYRADDANVMRDRDLEVDTTLKSLSQQIENIRSPEGT
RKNPARTCRDLKMGCHGDWKSGEYWDPNQGCNLDIAIKVYCNMETGETCVYPTQATIAQKN
WYLSKNPKKHHVWFGETMSDGFQFEYGGE-GSNPADVAIQLTFLRLMSTEATQNVTYHC
KNSVAYMDHDTGNLKKALLLQGANIEIEIRAEGNSRFTYGVTEGCTSHTGAWGKTVIEYK
TTKTSRLPIIDLAPMDVGPADQEFQ---IDIGPVCFL-MLSFVDTRILLLLAVTSYLSQ
H---VSEASAGRKGRDGDG---PQGERGPPGPPGRDGED-----GPPG---PPGPPG
PPGLGGNFAAQYDPSKAADFGPMPGLMGPRGPPGASGPPGPPGFQGVVPEGEPGQGTGP
QGPRGPPGPPGKAGEDGHPGKPRPGERGVAGPQARGFPPTPGLPGFKGIRGHNGLDGQ
KGQPGTPTGKGEPPGAPGNGTTPGQPGARGLPGERGRIGAPGPAGARGSDGSAGPTGPAGP
IGAAGPPGFPAGPAKGEIGPAGNVGPTGPAGPRGEIGLPGSSGVPVGNPANGLPGA
KGAAGLPVAGAPLPGRGIPGPPGPAGPSGARGLVGEPGPAGAKGESGNKGEPPGAAGP
PGPPGPSGEEGKRGSGNGEPGSAGPPGPAGLRGVPGSRGLPGADGRAGVMGPAGNRGASGP
VGAKGPNGDAGRPEGLMGPRGLPGQPGSPGAGKEGPVGFPGADGRVGP
PGNIGFPKGPPTGEPGKPEKGNVGLAGPRGAPGPEGNNGAQQPPGVGTGNQGAKEGTGP
AGPPGFQGLPGPSGAGEAGKPGERGLHGEFGVVPAGPRGERGLPGESGAVGPAGPIGS
RGPSGPPGPDGNKGEPPNVGPAGAPGPAGGGIPGERGVAGVPGGKGEKGA
PLRGDTGATGRDGARGLPGAIGAPGPAGGADRGEPPGAPGARGIPGERGEPPVGPSPGFAGP
PGAAGQPGAKGERGPKGPKGETGPTGAIPIGASGPPGPVGAAGPAGPRGDAGPPGMTGF
PGAAGRVGPPGAGITGPPGPPGAGKDGPRGLRGDVPVGRTEQQIAGPPGFAGEKGP
SGEAGAAGPPGTPGPPQILGAPGILGLPGSRGERGLPIAGATGEPGPLGVSGPPGARGP
SGPVGSPGPNAPGEAGRDGNPNDGPPGRDGPAGFKGERGAPGNPSPGALGAPGPHGQ
VGPSPKPNRGDPVGPVGPAGAFGPRGLAGPQGPGEKGEKPDGKHRGLPGLKGNGL
QGLPGLAQHGDDQPPGNNGPAGPRGPPGSPGPPKDRNGLPGPIGPAGVRGSHGSQGP
AGPPGPPGPPGPPGPPGPPGNGGYEV--GFDAEYRADQP---SLRPKDYEV DATLKTLLNQI

ETLLTPEGSKKNPARTCRDLRLSHPEWSS - GFYWIDPNQGCTADAIRAYCDFATGETCIH
ASLEDIPTKTWYVSKNPKDKKHIWFGETINGGTQFEYNGEGVTTKDMATQLAFMRLLANH
ASQNTITYHCKNSIAYMDEETGNLKKAVILQGSNDVELRAEGNSRFTFSVLVDGCSKKNK
WGKTIIEYRTNKPSRLPFLDIAPLDIGGADQEFGLHIGPVCFK

>Anolis

MFSFVDSRLLLLIAATVLLTNGQGE-----EDIPSKGSCTQDGLRYNDKDVWKPEPCQI
CVCDSGNILCDSMVCEEDL-NCENPEIPFGECCPVCPESGASVPYPET-GVE-----
-----NFAPQLSYGYD-----EKSGG-L
SLPGPMGPAGPRGLPGPPGAPGPGQGFQGGPPGEPGEAGASLTVVKHIGSGIGLDFLFFSSFS
KGEAGKPRPGERGAPGPGQARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGS
PGENGAPGVGPRGLPGERGRPGPAGARGNDGAAGAAGPPGPTGPAGPPGFPGAAGP
KGETGAQGSRGSEGPQGARGEPGPPGAGSAGPSGNPGSDGQPGAKGATGAPGISGAPGF
PGARGPAGPQGPSGAPGPKGNSGEPGAQGNKGDAGAKGEPGAPGIQGGPPGPPGEEGKRGS
RGEPPAGLPGPAGERGAPGSRGFPGSDGIAGPKGPPGERGSPGPAGAKGSPGESRPGE
PGLPGVKGLTGSPPGPDGKTGPTGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGAAGE
PGKPGERGAPGAVGAPGKDGGETGAQGGPPGPTGPAGERGEQGPAGAPGFQGLPGPAGA
PGESGKPGEQGVPGDVGAPGSPGARGERGFPGERGVQGPAGPAGPRGANGAPGNDGAKGD
AGAPGAPGNQGGPGLQGMPPGERGAAGLPGAKGDRG-----
-----EAGPSGPAGPTGARGAPGDRGEPGAPGAGFAGPPGADGQPGAKGESGD
AGAKGDAGAPGAGATGPPGAPGAVGAPGPKGSRGSAGPPGATGFPGAAGRVPVGGPSGN
IGLPGPPGPPSGKEGAKGPRGETGPAGRAGEAGPVGPPGPPGEEKGSPGGDGPAGAPGTGPG
QGIAGQRGVVGLPGRGERGFPGLPGSPGEPGKQGPSGNPGERGPPGPMGPPGLAGPPGE
AGREGSPGAEGAPGRDGAAGPKGDRGETGPAGPPGAPGAPGAPGPIGPAGKNGDRGETGP
AGPAGPAGPARGPNGPQGPRGDKGETGEQDGRGMKGRGFSGLQGGPPGPPGSPGEQGP
SGSSGPAGPRGPPGSAGAAGKDGLNGLPGPIGPPGPRGRTGDVGPAGPPGPPGPPGPPG
PSGGFDFSFMPQPPQEKAHHDGGRYYRADDANVMRDRDLEVDTTLKSLSMQIENIRSPEGT
RKNPARTCRDLKMCHEGDKWKSGEYWIDPNQGCNLDAIKVYCNMETGQTCVHASQPNIAQKN
WYISKNPKEKMHVWFGETMTDGFQFEYGESE-GSNPADVNIQLTFLRLMSTEASQNTITYHC
KNSVAYMDQQSGNLKKALTLQGSNEIEIRAEGNSRFTYGVIEDGCTSHTGSWGKTVIEYK
TTKTSRLPIIDVAPMDVGPADQEF--LEIGPACFL-----
-----GPAGEPRDGED---GPPGPPG---PPGLSG
-----NFQAQYDGSKSADLPGPMGIMGPRGPPGLTGPVGGSPGYQGPAGEPEPEPQTGP
QGARGPMGPPGKPGEDGHPGKPRGSRGERV-----SGPQGHNLGDGR
KGEPGAQGTKEGEPAGENGTPGQSGARGLPGERGRVGPAGPVGARGSDGSNGPSGPAGP
IGSAGSPGFPAGPAGKEVGSAGLTGSPGAPRGEAGLPGASGPVGGPPGNPGANGINGA
KGAAGLPGVAGAPGLPGPRGIPGSPGAG-----AAGTRGLTGSAGA
AGNPGPSGEEGKRGSTGEPSSGSPGAGLRGVPGSRGLPGSDGRAGAIGLAGNRGATGA
AGAKGPNGDAGRPEPGLMGPRGLPGQSGSPGAPGKEGVSGLPGPDGRPGVGPAGPRGE
PGNIGFPGPKGPNGEPPGKPGDRGHVGVGTGNRGAPGPDGNGAQQGPAGVIGNTGSKGETGP
AGPPGFQGLPGSPGVPGEAGKPGERGLHGDGFLPGPAGARGERGAPGESGAVGSPGPIGN
RGPSGPPGPDGNKGEPPGNAGLNGGAGSPGSPGIPGERGIAGVPGGKGEKGPQGFGRDAGT
TGRDGSRGAPGAVGAPGAPGAGSAGERGEAGPSGQAGPAGPRGIPGERGEVGPAGPPGFAGP
PGAAGQPGAKGERGQRGPKGDVGPSSAIGVVGNAGPAGPVGPSGAAGARGDSGPPGATGF
PGNAGRPPGPPGAGITGPAGPPGAPGKEGIRGPRGDTGPVGRTEAGIVGPPGFRGEKGP
SGESGAAGPPGTPGPQGLLGAAPGILGLPGSRGERGLPGIAGALGDHGPVGIAGPPGARGP
SGPIGNPVGNGAHGEAGRDGNPNSDGSPPRGDAPGYKGERGAPGNTGSHGALGAPGPHGA
VGPVGTGSRGEPGVPVGPVGTGAPAGARGLAGPQGPGRGEKGTVDGKGRGLPGLKGNL
QGLPGIAGHGDQGSPPGATGPAGPRGAPGSPGAPKDGDRNGLPGVGPAGVRSQGSQGP
SGPPGSPGLPGPPGVNNGGYEI--GYDMEYYRADQP---SLRPKDYEV DATLKS LNNQI
ETLLTPEGSRKNPARTCRDLRLSHPDWTS - GFYWIDPNQGCTLDAIRVYCDFATAETCIY
ANAESIPSKNWNKNSAEKKPIWLGETINGGTQFEYNEEGVTTKDMATQLAFMRLLANH
ASQNTITYHCKNSVAYMDEETGNLKKAVVLQGSNDVELRAEGNSRFTYSVLEDGCSKKSNE
WGKTVIEYRTNKPSRLPFLDIAPLDIGGADQEI RLDVGP----

>Mus

MFSFVDLRLLLLLLGGATALLTHGQE-----DIPEVSCIHNGLRVPNGETWKPEVCLI
CI CHNGTAVCDDVQCNEEL-DCPNPQRREGECCAFCEPEYVSP-NSEDVGVGEPKGDGPP
QGPRGPVGPVGRDGI PGQPGLPGPPGPPGPPGPPGLGGNFASQMSYGYD----EKSAG-V
SVPGPMGSPGRGLPGPPGAPGPGQGFQGGPPGEPGEGSS----GPMGPRGPPGPPGKNGD
DGEAGKPRPGERGPPGPGQARGLPGTAGLPGMKGHRGFSGLDGAKGADAGPAGPKGEPGS
PGENGAPGMGPRGLPGERGRPGPPGTAGARGNDGAVGAAGPPGPTGPTGPPGFPGA
KGEAGPQARGSEGPQVVRGEPGPPGAPGAGPAGNPGADGQPGAKGANGAPGIA GAPGF
PGARGSPGPPGPPGPKGNSGEPGAPGNKGDGTAKGEPGATGVQGGPPGAGEEGKRGA
RGEPPGSPGLPGPPGERGGPGRGFPADGVAGPKGSPGERGAPGAPGPKGSPGEAGRPE
AGLPGAKGLTGSPPGPDGKTGPPGAPGQDGRPGPAGPPGARGQAGVMGFPGPKGTAGE

PGKAGERGLPGPPGAVGPAGKDGEGAQAGPAGPAGERGEQGPAGSPGFQGLPGPAGP
PGEAGKPGEQGVPGDLGAPGPGSARGERGFPPGERGVQGGPPGAGPRGNNAPGNDGAKGD
TGAPGAPGSQAGPLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGARGLTGP IGP
PGPAGAPGDKGEAGSPGPPGPTGARGAPGDRGEAGPPGAGFAGPPGADGQPGAKGEPGD
TGVKGDAGPPGAPGAPPPGPIGNVAGAPGPKGPRGAAGPPGATGFPGAAGRVPVPPGSGN
AGPPGPPGVPVKEGGKGRGETGPAGRPGEVGPVPPGPPGAGEKGS PGADGPAGSPGTTPG
QGIAGQRGVVGLPGQRGERGFPLPGPSGEPGKQGPSGSSGERGPPGPMGPPGLAGPPGE
SGREGSPGAEGSPGRDGAAGKDRGETGPAGPPGAPGAPGAPGVPVGPAGKNGDRGETGP
AGPAGPIGPAGARGPAGPQGRGDKGETGEQGDRIKGRHGFSGLQGGPPGSPGSPGEQGP
SGASGPAGPRGPPGAGSPGKDG LNLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPG
PSGGYDFSLFPQPPQEKSDGGRYRADDANVVRDRDLEVDTTLKSLSQQIENIRSP
RKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVYCNMETGQTCVFPPTQPSVPQKN
WYISPNPKEKKHVWFGE SMTDGFQFEYGSE - GSDPADVAIQLTFLRLMSTEASQNITYHC
KNSVAYMDQQTGNLKKALLLQGSNEIELRGE GNSRFTYSTLV DGTSTHTGTWGKTVEYK
TTKTSRPLIIDVAPLDIGAPDQEF - LDIGPACFV - MLFSFVDRTRLLLLAVTSLATCQ
Y - - - - LQSGSVRKGTGTDRG - - - PRGQRGAPGPRGRDGV DGPMPGPPGPPGSPGPPGSA
PPGLTGNFAAQYS - DKGVS SGPMPGLMGRGPPGAVGAPGPPGQGFQGPAGEPGEPTGP
AGPRGPAGSPGKAGEDGHPGKPRPGERGVVGPQGARGFPPTPGLPGFKGVKHSMDGL
KGQPGAQGVKGEPPAGENGTPGQAGARGLPGERGRV GAGPAGARGSDG SVGPVGPAGP
IGSAGPPGFPAGPAGPKGELGPVGNP GAGPAGPRGEVGLPGLSGPVGPPGNP GTNGLTGA
KGATGLPGVAGAPGLPGRGIPG PAGAAGATGARGLVGEPGAGSKGESGNKGEPPG SVGA
QGGPPGSGEEGKRGSPGEAGSAGPAGPPGLRGSPPSRGLPGADGRAGVMGPPGNRGSTGP
AGIRGPNGDAGRPEPGLMGRGLPGSPGNVGP SGKEGVPVGLPGIDGRPGPIGPAGPRGE
AGNIGFPGPKGPSGDGPKPGERGHPGLAGARGAPGPDGNNGAQGPPGQGVQGGKGEQGP
AGPPGFQGLPGPSGTTGEVVKPGERGLPGEFGLPGPAGPRGERGTPGESGAAGPSGPIGS
RGPSGAPGPDGNKGEAGAVGAPGSAGASGPGGLPGERGAAGIPGGKGEKGETGLRGDTGN
TGRDARGIPGAVGAPGAGASGDRGEAGAAGPSGPAGPRGSPGERGEVGPAGPNGFAGP
AGAAGQPGAQGEKGTGKPKGENGIVGPTGSVGAAGPSGPNP GPPVGSRGDGGPPGMTGF
PGAAGRTGPPGPGSIAGPPGPPGAAGKEGIRGPRGDQGPVGR TGETGASGPPGPFVGEKGP
SGEPGTAGAPGTAGPQGLL GAPGILGLPGSRGERGLPIAGALGEPGPLGISGPPGARGP
PGAVGSPGVNGAPGEAGRDGNPGSDGPPGRDQPGHKGERGYPGSIGPTGAAGAPGPHGS
VGPAGKHGNRGEPPGAGSVGPVAVGPRGSPGQIRGDKGEPGDKGHRGLPGLKGYSL
QGLPGLAGLHGDQGAPGVPVGPAGPRGAGPSGVPVKDGRSGQGPVGPAGVRSQGSQGP
AGPPGPPGPPGPPGSGGGYDF - - GFEGDFYRADQPRSQPSLRPKDYEV DATLKS LNNQI
ETLLTPEGSRKNPARTCRDLRLSHPEWNS - DYYWIDPNQGCTMDAIKVYCDFSTGETCIQ
AQPVNTPAKNSY - - SRAQANKHVWLGETINGGSQFEYNVEGVSSKEMATQLAFMRL LANR
ASQNITYHC KNSIAYLDEETGSLNKAVLLQGSNDVELVAEGNSRFTYSVLVDGCSKKTNE
WGKTIIEYKTNKPSRPLFLDIAPLDIGGADQEF RVEVGPVCFK

>Rattus

MFSFVDRLLLLL LGATALLTHGQE - - - - - DIPEVSCIHNGLRVPNGETWKP DVCLI
CICHNGTAVCDGVLCCKEDL - DCPNPQKREGECCPFCPEEYVSP - DAEVIGVEGPKGDPGP
QGPRGPVGPVGGQDGI PGQPGLPGPPGPPGPPGPPGLGGNFASQMSYGYD - - - - EKSAG - V
SVPGPMGSPGPRGLPGPPGAPGPPGQGFQGGPPGEPGEPGAS - - - - GPMGPRGPPGPPGKNGD
DGEAGKPRPGERGPPGPPGARGLPGTAGLPGMKGHRGFSGLDGA KGDTPAGPKGEPGS
PGENGAPQMGRGLPGERGRPPGPPG SAGARGNDGAVGAAGPPGPTGPTGPPGFPGAAGA
KGEAGPQARGSEGPQGVREGEPPGPPGAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGF
PGARGSPGPPGSPGAPGPKNSGEPGAPGNKGDGAKGEPGAGVQGGPPGAGEEGKRG
RGEPPGSPGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGAGPKGSPGEAGRPE
AGLPGAKGLTGS PGSPGPDGKTGPPG PAGQDGRPGPAGPPGARGQAGVMGFPGPKGTAGE
PGKAGERGVPPGAVGPAGKDGEGAQAGPAGPAGERGEQGPAGSPGFQGLPGPAGP
PGEAGKPGEQGVPGDLGAPGPGSARGERGFPPGERGVQGGPPGAGPRGNNAPGNDGAKGD
TGAPGAPGSQAGPLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGP IGP
PGPAGAPGDKGETGSPGAGPTGARGAPGDRGEPGPPGAGFAGPPGADGQPGAKGEPGD
TGVKGDAGPPGAPGAPPPGPIGNVAGAPGPKGSRGAAGPPGATGFPGAAGRVPVPPGSGN
AGPPGPPGVPVKEGGKGRGETGPAGRPGEVGPVPPGPPGAGEKGS PGADGPAGSPGTTPG
QGIAGQRGVVGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGE
SGREGSPGAEGSPGRDGAAGKDRGETGPAGPPGAPGAPGAPGVPVGPAGKNGDRGETGP
AGPAGPIGPAGARGPAGPQGRGDKGETGEQGDRIKGRHGFSGLQGGPPGSPGSPGEQGP
SGASGPAGPRGPPGAGSPGKDG LNLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPG
PSGGYDFSLFPQPPQEKSDGGRYRADDANVVRDRDLEVDTTLKSLSQQIENIRSP
RKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVYCNMETGQTCVFPPTQPSVPQKN
WYISPNPKEKKHVWFGE SMTDGFQFEYGSE - GSDPADVAIQLTFLRLMSTEASQNITYHC
KNSVAYMDQQTGNLKKALLLQGSNEIELRGE GNSRFTYSTLV DGTSTHTGTWGKTVEYK
TTKTSRPLIIDVAPLDIGAPDQEF - - MDIGPACFV - MLFSFVDRTRLLLLAVTSLATCQ

C-----WRGGH-----PDLNRGKHLHPSFHARGASGPRGRDGVDPVGPVPPGPPGAPGPPGPPG
PPGLTGNFAAQYS-DKGVSAAGPGMGLMGPRGPPGAVGAPGPQGFQGPAGEPGEPPGQTGP
AGSRGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPPTPGLPGFKGIRGHNGLDGL
KGQPGAQGVKGEPPGAPGENTPGQAGARGLPGERGRVGPAGPAGARGSDGSGVGPVGPAGP
IGSAGPPGFPAGPGKGEGLGPVGNPAGPAGPRGEAGLPGLSGPVGPPGNPGANGLTGA
KGATGLPGVAGAPGLPGPRGIPGPVGAAGATGPRGLVGEPPGAGSKGETGNKGEPPGASAGA
QGPVPPSGEEGKRGSPGEPGASAGPAGPPGLRGSPPSRGLPGADGRAGVMGPPGNRGSTGP
AGVRGPNGDAGRPEPGLMGPRGLPGSPGNVGPAGKEGPVGLPGIDGRPGPIGPAGPRGE
AGNIGFPKPGPSGDPGKPGKHPGLAGARGAPGPDGNNGAQGPPGPQGVQGGKGEQGP
AGPPGFQGLPGPSGTAGEVGPGERGLPGEFGLPGPAGPRGERGPPGESGAAGPSGPIGS
RGPSGAPGPDGNKGEAGAVGAPGSAGASGPGGLPGERGAAGIPGGKGEKGETGLRGEIGN
PGRDARGAPGAI GAPGPAGASGDRGEAGAAGPSGPAGPRGSPGERGEVGPAGPNGFAGP
AGSAGQPGAKGEKGTGKPGKGENGIVGPTGPVGAAGPSGPNPAGPAGSRDGGPPGMTGF
PGAAGRTGPPGPGSITGPPGPPGAAGKEGIRGPRGDQGPVGRTEIGASGPPGFAGEKGP
SGEPGTTGPPGTAGPQGLLGAAPGILGLPGSRGERGLPGIAGALGEPGLGIAGPPGARGP
PGAVGSPGVNGAPGEAGRDGNPGSDGPPGRDGPQGHKGERGYPNIGPTGAAGAPGPHGS
VGPAGKHGNRGEPPGAGSVGPVAVGPRGSPGQGI RGDKGEPPGDKGARGLPGLKGNGL
QGLPGLAGLHGDQGAPGPVGPAGPRGPAGPSGPIGKDGRSGHPGPVGPAGVRSQGSQGP
AGPPGPPGPPGPPGVSGGGYDF--GFEGDFYRADQPRSQPSLRPKDYEV DATLKLNNQI
ETLLTPEGSRKNPARTCRDLRLSHPEWKS-DYYWIDPNQGCTMDAIKVYCDFSTGETCIQ
AQPVNTPAKNAY--SRAQANKHVWLGETINGGSQFEYNAEGVSSKEMATQLAFMRLLANR
ASQNITYHCKNSIAYLDEETGRLNKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNE
WDKTIIEYKTNKPSRLPFLDIAPLDIGGTNQEFRVEVGPVCFK

Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPPGLAGPPGESGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	ALLLQGSNEIEIR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	ALLLQGSNEIELR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAPGPAGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GVVGLPGQR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGEPGKPGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGADGIAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPSGPAGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGSPGEQGPSGASGPAGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDPGPQQPRGPVGPGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSPGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	QGSPGSSGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GLPGTAGLPGMK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GHRGFSGLQGPPGSPGSPGEEQGPSGASGPAGP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPSGPQGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDKGETGEQGDR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPAGPAGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSEGPQGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPPGPPGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GFPGERGVQGGPPGAGPRGNNAP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGPAGVQGGPPGAGEEGKRGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPAGPQGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GNDGAVGAAGPPGPTGPTGPPGFPGAAGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GRPGAPGPAGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	SAGISVPGPMGPMGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GETGPAGPAGPIGP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GETGPAGRPGEPPGAGPPGPPGEK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDAGPPGPAGPTGAPGPAGZVGAPGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GNDGAPGAAGPPGPTGPAGPP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	EGAPGAEGAPGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPPGPQGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDTGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGLPGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	ALLLQGANEIEIR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GETGPAGPPGAPGAPGAPGPVGPAGKSGDR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPTGPAGPPGFPGAVGAKGEAGPQGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GNNGEPGAQQNKGEPAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	EGSPGSEGAPGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDAGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GGPGRGFPGADGVAGPKGPAGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GLTGSPGSPGPDGKTGPAGAAGQDGHPP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDRGETGPAGPPGAPGAPGAPGPVGPAGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPPGSAGSPGKDGLNGLPGPIGPPGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGPAGPPGPAGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPSGEPGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAPGPQGPAGAPGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	SAGVSVPGPMGSPGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GANGAPGNDGAKGDAGAPGAPGNEGPPGLEGM
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	AGPPGADGQPGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGPQGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAPGPAGPKGS
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GKRGFPGLPGPSGEPGKQGPSGASGER

Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEGGPSGPAGPTGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSVGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGAPGPAGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGPSGPPGPTGARGAPGDRGEAGP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	AGAPGDKGEAGPSGPAGPTGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GANGAPGNDGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDNGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GFPGADGISGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSPGSDGPAGAPGIPGPQGIAGQR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPIGPPGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GNDGATGAAGPPGPTGPAGPPGFPGAVGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	TGPPGPAGQDGRPGPPGPPGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GESGPSGPAGPTGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPSGPQGARGL
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGEPGKS
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GFSGLQGGPPG
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	AVLLQGSNEIIR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAPGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GESGDAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGLPGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGSPGENGTPGQMGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGAAGAPGGRGPPGLQGMPGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GLTGPPIGPPGAPAGPDKGEAGPSGPPGPTGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAPGPVGPAGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGEQGERGMK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPQGIAGQR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDPGPQGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGPPGATGFPGAAGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGPAGVQGGPPGSGEEGKR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGSPGPDGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDTGPRGRGPAGPPGRDGIPGQPGLPGP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGPAGVQGGPPGSGEEGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGRVPVGGPSGNAGPPGPPGPAGKEGGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDAGPAGPKGEPGSP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GETGPAGPAGPIGPAGARGPAGPQGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GGPGRGFPGADGVAGPKGPPSGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GANGAPGIAGAPGFPGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	SLLLQGSNEIELR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	NGDRGETGPAGPAGPAGPSGVR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGEPGKSGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGEPGKAGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGPVGVQGGPPGPAGEEGKR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSTGESGRPGEPGL
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDAGAPGAPGSQGAPGL
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSEGPQGSR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSEGPQGVR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GGPGRGFPGSDGASGPKGAPGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GQAGVMGFPGPKGTAGEPGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GAAGEPGKA
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GQPGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	DGAPGAKGDR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPAGPPGSPGEK

Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	IGNVGAPGPKGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPPGLAGPPGEAGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GVQGGPPGAPGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	EGAPGAEGSPGRDGGSPGAKGDR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGPPGPAGPTGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	QGPPSGASGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGPSGPPGPTGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPSGPQQGPPGPPGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGATGVQGGPPGPAGEEGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	EGGKGPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGFQGLPGPAGPPGEAGKPGEQGV
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSPGPAGPKGS
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPSGPAGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGAAGRLGPPGPSGNAGPPGPPGPGGKEGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GLTGPIGPPGPSGAPGDK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GATGSPGIAGAPGFPGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDAGAPGAPGSQQGAPGLQGMPGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GVQGGPPGAPGRGANGAPGNDGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GNSGEPGAPGNKGDGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPIGPPGPRGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GIKGHRGFSGLQGGPPGPPGSPGEQGGPSGASGP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GLPGGPKNGDDGESGKPRPGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	EGSPGAEGSPGRDGGAPGAKGDR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDAGAPGAPGNEGPPGLEMPGER
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGPPGPAGPTGARGAPGDR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGPPGPAGAAGPAGNPGADGQPGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GETGPQGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GSPGADGPIGAPGTPGPQGIAGQR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDAGPQGSR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGATGVQGGPPGPAGEEGKR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GNVGAPGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEPGSPGENGAPQMGRGLPGERGRP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GNNGEPGAQGNK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPAGERGSPGPAGPK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GDRGDAGPKGADGSPGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPAGPPGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGERGGQGPAGAQQPR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPSGAPGDKGEGGSPGAPPTGAR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	IGPPGPRGR
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GEAGPPGPAGPTGARGAPGDRGEPGP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GHRGFSGLQGP
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	PGPAGVQGGPPGPSGEEGK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	QGVREPGPPGPAGAAGPAGNPGADGQPGAK
Collagen alpha-1(I) chain precursor (Alpha-1 type I collagen)	GPPGAVGPAGK
Collagen alpha-2(I) chain precursor (Alpha-2 type I collagen)	GLPGESGAVGPGPIGSR