PAFAH1B1 and *NDEL1* in the Cetartiodactyla superorder: Investigating adaptive evolution in two cerebral gyrification genes in *Orcinus orca* and *Tursiops truncatus*

David Haas [1,2,3]

Marine Genomics Research Experience
Spring 2013

[1] University of Washington, Friday Harbor Laboratories, Friday Harbor, WA 98250
[2] Department of Biology, University of Washington, Seattle, WA 98195
[3] Cetacean Research Partners, Richmond Beach, WA 98177

Contact information:
David Haas
Cetacean Research Partners
http://cetaceanresearch.org
dave@cetaceanresearch.org

Keywords: gyrencephaly · cetartiodactyla · cetaceans · delphinidae · *Orcinus orca* · *Tursiops truncatus* · *PAFAH1B1* · *NDEL1* · adaptive evolution · positive selection · purifying selection

ABSTRACT

Cerebral gyrencephaly, distinguished by the convoluted folds and grooves in the cerebral cortex, occurs in many animals in Class Mammalia. The highest rates of cerebral gyrification occur in Order Cetacea within the Superorder Cetartiodactyla. Of all cetaceans, two members of Family Delphinidae, *Orcinus orca* (killer whales) and *Tursiops truncatus* (Atlantic bottlenose dolphins) have exceptionally high indices of cerebral gyrification. Research has shown that mutation in two genes, *PAFAH1B1* and *NDEL1*, both located on human chromosome 17, disrupt gyrencephaly normally observed in human brains. Based on this research, *PAFAH1B1* and *NDEL1* were evaluated for positive selection in *Orcinus orca* and *Tursiops truncatus,* to test the hypothesis that these genes are undergoing adaptive evolution against the background of purifying selection. The results of multiple selection tests supported a set of null hypotheses, indicating that these genes were under purifying, not positive, selection, which supports the hypothesis that delphinid cerebral gyrification is not under the control of either *PAFAH1B1* or *NDEL1.*

INTRODUCTION

There is a high degree of variation in cerebral cortex complexity in Class Mammalia. One measure of this complexity is the degree to which the mammalian cerebral cortex has undergone gyrification (i.e.: convoluted folding) during development (Hofman, 1985). Some researchers (Mayhew et al.1996) have noted order-level similarities within mammals, suggesting a possible shared genetic mechanism that can result in a relative lack of cerebral gyrification (lissencephaly, as seen in Order Sirenia, the dugongs and manatees) or very well-defined cerebral gyrification (gyrencephaly, as seen in Order Proboscidea, the elephants). Superorders like Cetartiodactyla (which includes whales, dolphins, porpoises and all terrestrial even-hoofed ungulates) and orders like primates also show a wide range gyrification variation (Marino 2002). Cetaceans, especially animals in Family Delphinidae (the dolphins), have the highest degree of cerebral gyrencephaly, as measured by the gyrencephalic index described in Manger (2006) and Manger et al. (2012).

Two main hypotheses have been posited to explain the physiological mechanisms of cerebral gyrencephaly in mammals: (1) that specific growth processes during cerebral cortex development lead to gyrification (the grey matter hypothesis); and (2) that axon elongation and retraction as a result of cortical tension thresholds lead to gyrification (the mechanical tension hypothesis). Both hypotheses are summarized in the recent work of Ziles et al. (2013), but this work also suggests the need for further exploration of the genetic and molecular basis for gyrification in the mammalian brain. Research focusing on the cerebral gyrification genetics has thus far largely focused on loss of gyrification in humans (Hirotsune et al. 1998). This condition, known as classic lissencephaly, occurs when chromosome sites 17p13.1 and 17p13.3 have undergone mutation (Cardosa et al. 2002). Mutations at these sites have been shown to disrupt normal protein synthesis of the gene products of *NDEL1* and *PAFAH1B1*, respectively (Cardosa et al. 2002, Avela et al. 2011). The *PAFAH1B1* gene codes for the LIS1 protein, and it has been hypothesized that defective LIS1 may result in developing neurons failing to migrate to their final locations (cellular mispositioning), ultimately resulting in lissencephaly (Kerjan and Gleeson 2007). Further studies by Yamada et al. (2008), Youn et al. (2009) and Torisawa et al. (2011) have also shown that the genetic product of *NDEL1*, the NDEL1 protein, forms a complex with LIS1 and the dynein heavy chain (DHC), working together as functional unit to facilitate the migration of neuronal cells to the cerebral cortex via microtubules array alongside radial glial cell scaffolding. These results strongly suggest a key role for both *PAFAH1B1 and NDEL1* in mammalian cerebral gyrification.

While the genetic and molecular role of *PAFAH1B1* and *NDEL1* in human cerebral gyrification is becoming better understood, the question of whether cerebral gyrification is under direct control of these genes in cetaceans and other mammals is not. There has also been no investigation into whether animals with unusually high levels of cerebral gyrification show signs of adaptive evolution (measured as positive selection, as in Furlong et al. 2008, Yang et al. 2010, and McGowen et al. 2011) in either *PAFAH1B1* or *NDEL1*. This research takes up those questions, using *PAFAH1B1* and *NDEL1* gene sequences from *Orcinus orca, Tursiops truncatus* and dozens of other mammals to ask the following questions: (1) Do *PAFAH1B1* and *NDEL1* maximum likelihood and Bayesian trees lend support to other well-supported phylogenetic relationships among extant mammalian taxa? (2) Are *PAFAH1B1* and *NDEL1* under positive selection in killer whales and bottlenose dolphins? (3) Are *PAFAH1B1* and *NDEL1* under positive selection in other gyrencephalic mammals? (4) Is there evidence to support the hypothesis that gyrification of the mammalian cerebral cortex is under *PAFAH1B1* and/or *NDEL1* control?

MATERIALS AND METHODS

Gathering gene sequences:

Using the NCBI GenBank database, a series of searches were performed for *PAFAH1B1* and *NDEL1* in *Orcinus orca* and *Tursiops truncatus.* An additional search for *NEUROG1* in *Orcinus orca* and *Tursiops truncatus* was added for purposes of including a negative control (i.e.: a gene which is highly conserved across the animal kingdom and undergoing purifying selection (Swalla, pers. comm.). A comparison was made between each of these genes in *O. orca* and *T. truncatus* to ensure a highly level of congruence; in all three genes, the sequences were effectively identical. Based on this finding, the killer whale (*Orcinus orca*) sequences for *PAFAH1B1*, *NDEL* and *NEUROG1* were used to as the standard for orthologous sequence queries with GenBank's reciprocal BLAST search tool. Three BLAST searches yielded at least thirty mammalian species for *PAFAH1B1*, *NDEL1* and *NEUROG1*, respectively, with e-values greater than $1x10^{-5}$. The FASTA-formatted DNA sequences of these taxa were copied and pasted into TextWrangler for Mac OS X, along with the sequences of one or more appropriate outgroup animals (see Table 1 for the final list of assembled species and their respective *PAFAH1B1*, *NDEL1* and *NEUROG1* GenBank Accession Numbers). Sequence data for the positive control gene, *MCPH1,* was acquired from online supplemental materials used in the work of McGowen et al. (2011) (see Supplemental Note 1 for download instructions).

Aligning gene sequences:

For each gene, FASTA-formatted gene sequence text files were opened in MEGA 5.1 for Mac OS X. All perspective taxa sequences were first evaluated for analytic suitability. Where obvious sequencing errors or large gaps or omissions occurred within a given FASTA sequence (e.g.: hundreds of base pairs), those taxa were eliminated. The remaining taxa's sequences were then aligned using MEGA's CLUSTAL and/or MUSCLE alignment algorithms. The resulting alignments were manually inspected for additional problems per Davies et al. (2012), Kirwan et al. (2013) and Kocot (pers. comm.) until suitable alignments were produced. As a check, these DNA alignments were converted into their amino acid coding products and compared with GenBank's predicted amino acid products (see amino acid alignments derived from *PAFAH1B1, NDEL1, NEUROG1* and *MCPH1* genes in Supplemental Notes 2 through 5, respectively). Final alignments for each of these four genes were saved and exported as FASTA and NEXUS formats for subsequent analysis. An online tool was also used to generate a PHYLIP-formatted alignment file, using final NEXUS alignments exported from MEGA.

Generation of phylogenetic trees:

PAFAH1B1 and *NDEL1* alignments were subjected to phylogenetic reconstruction using two methods:

(1) RAxML maximum likelihood with 5,000 thorough bootstrap replicates (Stamatakis, 2006). This was performed using raxmlGUI v1.3 for Mac OS X; and

(2) Bayesian analysis with 200,000 posterior probability replicates (Ronquest and Huelsenbeck, 2003), generated with MrBayes v3.2.1 for Mac OS X.

A maximum likelihood bootstrap consensus tree and Bayesian analysis posterior probability consensus tree was created using FigTree for Mac.

See Supplemental Note 1 for the specific commands and parameters used to generate these results and for information about downloading the applications used in this study, as well as the data output and final results.

Tests for positive selection:

Tests for positive selection of *PAFAH1B1*, *NDEL1*, *NEUROG1* and *MCPH1* measured as selection intensity were performed for all sampled taxa, using site models in the CODEML package in PAML 4.7 (Yang 2007, McGowan et al. 2011, and Kirwan et al. 2013). Alignment files generated for use in MrBayes analysis were opened in HyPhy for Mac OS X, and trees were created and saved using a PHYLIP-compatible (parenthesis tree notation) interleaved-sequence format to ensure compatibility with PAML's CODEML program. The alignment files used to generate these trees were subsequently edited to include the letter "i" on the first line and to ensure that two spaces were present between the number of taxa, the number of sequences, and the names of taxa which preceded the sequences, in accordance with PAML's data parsing requirements. Once PAML-compatible alignment and tree file formats were created, seven different CODEML codeml.ctl were generated and executed to evaluate a series of evolutionary site analysis models. See Supplemental Note 1 for the specific commands and parameters used in CODEML positive selection analysis. The Supplemental Note 1 also contains details about site model tests, how nesting site models correspond with null and alternative hypotheses, and how these nested models are used to evaluate adaptive evolution in a statistically significant way.

RESULTS

Consensus order-level phylogenies, as described in recently supported mammalian phylogenetic analysis (e.g.: Lindblad-Toh et al. 2011) were largely upheld in both maximum likelihood and Bayesian analysis consensus trees for *PAFAH1B1* and *NDEL1* (see Figures 1 through 4). The maximum likelihood consensus tree for *PAFAH1B1* recovered Family Delphinidae *(Orcinus orca* and *Tursiops truncatus)* 100 percent of the time, and Superorder Cetartiodactyla was recovered in 96 percent of the bootstrap replicates. Similarly, the Bayesian analysis consensus tree for *PAFAH1B1*, generated from 200,000 replicates, recovered Family Delphinidae and Superorder Cetartiodactyla with posterior probabilities of 100 at both branches. *PAFAH1B1* maximum likelihood and Bayesian analysis consensus trees were also largely congruent with each other, with some minor differences at the tips of the primate order. The maximum likelihood consensus tree for *NDEL1* recovered Family Delphinidae 100 percent of the time, and the Superorder Cetartiodactyla was recovered in 80 percent of the bootstrap replicates. However the consensus maximum likelihood tree showed soft support for *Sus scrofa* (pigs) as the basal branch of the terrestrial artiodactyls. The Bayesian analysis consensus tree for *NDEL1* recovered Family Delphinidae and Superorder Cetartiodactyla with posterior probabilities of 100, but the addition of *Sus scrofa* here resulted in a soft polytomy within the Cetartiodactyla.

In the CODEML nested model tests for positive selection, three genes (*PAFAH1B1*, *NDEL1* and *NEUROG1*) were all negative for positive selection, with average omega values (ω) much less than 1 for all site model tests (see Figure 1). For *PAFAH1B1,* all site models (m0 through m8) computed extremely low average omega values (none greater than 0.02), strongly suggestive of purifying selection at all codon sites in this gene, with the nested m0 vs. m8 pair being statistically significant. For *NDEL1*, all site models computed extremely low average omega values as well, with none greater than 0.055, also strongly suggesting that this gene is under powerful purifying selection, with two nested models (m0 vs m3 and m1a vs. m2a) showing statistical significance. Like *PAFAH1B1* and *NDEL1,* the positive control *NEUROG1* also produced very low average omega values across all CODEML site model tests (ω = 0.08 to 0.17), suggestive of purifying selection, with the m1a vs. m2a nested model being statistically significant. Only the positive control gene *MCPH1* differed significantly from the three previously mentioned genes, with multiple omega values exceeding 1 (ω > 1). All four nested models reported numerous sites under positive selection, with two specific amino acid sites (55 S and 319 L) having statistically significant Naive Estimated Bayesian and Bayes Estimated Bayesian (BEB) probabilities (i.e.: the probability that $\omega > 1$ is greater than 95 percent), and with all four nested model tests showing strong statistical significance. This largely confirms the broader positive selection result reported in McGowen et al. (2011), though the exact amino acid sites reported as being under positive selection (89 and 362) differ from the results reported here.

DISCUSSION

The series of maximum likelihood and Bayesian phylogenetic analyses conducted on *PAFAH1B1* and *NDEL1* are largely consistent with the cetartiodactylan superorder hypotheses put forward by Thewissen et al. (2007) and O'Leary and Gatesy (2008). Unfortunately there was not sufficient taxonomic or genetic sampling in this research to support either hypothesis with respect to cetacean and artiodactylan branching within the cetartiodactylan superorder.

The results of the tests for selection were, however, quite definitive, strongly rejecting the hypothesis that *PAFAH1B1* and *NDEL1* were under positive selection. On the contrary, such low average omega values for *PAFAH1B1* ($\omega = 0.0132$) and *NDEL1* ($\omega = 0.0469$) strongly indicate purifying selection in these genes. This is true of the two cetaceans species examined here, but applies equally to all animals in Class Mammalia. The strength or purifying selection in these genes is even stronger than the observed values for the negative control gene, *NEUROG1* ($\omega = 0.1231$) and contrasts strongly with the average omega value for the positive control gene, MCPH1 ($\omega = 0.6372$).

Given such strong purifying selection, it is clear that neither *PAFAH1B1* or *NDEL1* is undergoing adaptive evolution in cetaceans or any other animals in the mammalian order. This lends strong support to the overarching null hypothesis of this research, specifically that neither *PAFAH1B1* or *NDEL1* are the genes directly responsible for the extreme cerebral gyrification seen in *Orcinus orca* or *Tursiops truncatus.* While these genes may play important roles in the ultimate fate of cortical neuronal cells, this research strongly suggests that neither of these genes, singly or together, are the drivers for the unusual gyrencephalic adaptive evolution observed in modern cetaceans.

FUTURE DIRECTIONS

The focus of this research has been on *PAFAH1B1* and *NDEL1,* in the context of how those two genes influence gyrencephaly in cetaceans, humans, and, most generally, mammals. The LIS1-NDEL1-dynein complex has been shown to transport neuronal cells through the six tissue layers of the brain and is a crucially important mechanical driver for cortical folding. That said, this research strongly suggests that the genes that code for the LIS1 enzyme and NDEL1 protein are not likely the genes controlling mammalian gyrification. Future research should focus on locating the genes which are actually responsible for this. An area of primary interest should be the developmental genes responsible for creating and activating the axial glial cells, along which neuronal cells migrate as they traverse the cerebral cortex tissue layers during gyrencephalic development. It may be possible that these axial glial cells are responsible for producing the microtubules along which the LIS1-NDEL1-dynein complex walks, pulling the neuronal cells into position as it moves. By focusing on the tracks instead of the train, it may be possible to finally elucidate the genes that control gyrencephaly and to test whether those genes are undergoing intense positive selection in highly gyrencephalic animals like cetaceans.

ACKNOWLEDGMENTS

I would like to thank Dr. Billie J. Swalla for her ideas and guidance and for her strong support to pursue research of great personal interest. Thanks also go to Dr. Kevin Kocot for his help with understanding phylogenetic analysis and bioinformatics tools and to Elliott Jacobsen-Watts for great feedback on presentations and this document. Special thanks go out to my fellow Marine Genomics Researchers, all of whom helped me with this research in ways large and small. I also wish to extend my gratitude to the FHL Spring 2013 faculty, teaching assistants, staff and students (the Three Seas folks, the OAers, the SedHeads and the NEREos). Big thanks to my wife for supporting my decision to come back to the University of Washington to finish my undergraduate degree and for her encouragement to return to FHL a second time. Finally, thanks to the UW-Friday Harbor Labs and to the Mary Gates Endowment for the funding which helped make this research possible.

REFERENCES

Avela, K., Aktan-Collan, K., Horelli-Kuitunen, N., Knuutila, S., & Somer, M. (2011). A Microduplication on Chromosome 17p13.1p13.3 Including the PAFAH1B1 (LIS1) Gene. *American Journal of Medical Genetics Part A*, *155A*(4), 875–879. doi:10.1002/ajmg.a.33944

Bi, W., Sapir, T., Shchelochkov, O. A., Zhang, F., Withers, M. A., Hunter, J. V., et al. (2009). Increased LIS1 expression affects human and mouse brain development. *Nature Genetics*, *41*(2), 168–177. doi:10.1038/ng.302

Capra, V., Mirabelli-Badenier, M., Stagnaro, M., Rossi, A., Tassano, E., Gimelli, S., & Gimelli, G. (2012). Identification of a rare 17p13.3 duplication including the BHLHA9 and YWHAE genes in a family with developmental delay and behavioural problems. *Bmc Medical Genetics*, *13*. doi: 10.1186/1471-2350-13-93

Cardoso, C., Leventer, R. J., Dowling, J. J., Ward, H. L., Chung, J., Petras, K. S., et al. (2002). Clinical and molecular basis of classical lissencephaly: Mutations in the LIS1 gene (PAFAH1B1). *Human Mutation*, *19*(1), 4–15.

Davies, K. T. J., Cotton, J. A., Kirwan, J. D., Teeling, E. C., & Rossiter, S. J. (2012). Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. *Heredity*, *108*(5), 480–489. doi:10.1038/hdy.2011.119

Hirotsune, S., Fleck, M. W., Gambello, M. J., Bix, G. J., Chen, A., Clark, G. D., et al. (1998). Graded reduction of Pafah1b1 (Lis1) activity results in neuronal migration defects and early embryonic lethality. *Nature Genetics*, *19*(4), 333–339.

Hofman, M. A. (1985). Neuronal correlates of corticalization in mammals: A theory. *Journal of theoretical biology*, *112*(1), 77–95. doi:10.1016/S0022-5193(85)80117-X

Kerjan, G., & Gleeson, J. G. (2007). Genetic mechanisms underlying abnormal neuronal migration in classical lissencephaly. *Trends in genetics : TIG*, *23*(12), 623–630. doi:10.1016/j.tig.2007.09.003

Kirwan, J. D., Bekaert, M., Commins, J. M., Davies, K. T. J., Rossiter, S. J., & Teeling, E. C. (2013). A phylomedicine approach to understanding the evolution of auditory sensory perception and disease in mammals. *Evolutionary Applications*, *6*(3), 412–422. doi:10.1111/eva.12047

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, *478*(7370), 476–482. doi:doi:10.1038/nature10530

Manger, P. R. (2006). Cambridge Journals Online - Biological Reviews - Fulltext - An examination of cetacean brain structure with a novel hypothesis correlating thermogenesis to the evolution of a big brain. *Biological Reviews*.

Manger, P. R., Prowse, M., Haagensen, M., & Hemingway, J. (2012). Quantitative analysis of neocortical gyrencephaly in African elephants (Loxodonta africana) and six species of cetaceans: Comparison with other mammals. *The Journal of Comparative Neurology*, *520*(11), 2430–2439. doi:10.1002/cne.23046

Marino, L. (2002). Convergence of Complex Cognitive Abilities in Cetaceans and Primates. *Brain, Behavior and Evolution*, *59*(1-2), 21–32. doi:10.1159/000063731

Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., et al. (2007). Cetaceans Have Complex Brains for Complex Cognition. *PLOS Biology*, *5*(5), e139. doi: 10.1371/journal.pbio.0050139

Mayhew, T. M. G. L. M. V. D. S. W. (1996). The gyrification of mammalian cerebral cortex: quantitative evidence of anisomorphic surface expansion during phylogenetic and ontogenetic development. *Journal of Anatomy*, *188*(Pt 1), 53.

McGowen, M. R., Montgomery, S. H., Clark, C., & Gatesy, J. (2011). Phylogeny and adaptive evolution of the brain-development gene microcephalin (MCPH1) in cetaceans. *Bmc Evolutionary Biology*, *11*(1), 98. doi:10.1111/j.1469-7998.1973.tb04656.x

O'Leary, M. A., & Gatesy, J. (2008). Impact of increased character sampling on the phylogeny of Cetartiodactyla (Mammalia): combined analysis including fossils - O'Leary - 2007 - Cladistics - Wiley Online Library. Cladistics.

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models.

Silvestro, Michalak (2012) - raxmlGUI: a graphical front-end for RAxML. Organisms Diversity and Evolution 12, 335-337. DOI: 10.1007/s13127-011-0056-0

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690. doi:10.1093/bioinformatics/btl446

Thewissen, J. G. M., Cooper, L. N., Clementz, M. T., Bajpai, S., & Tiwari, B. N. (2007). Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature*, *450*(7173), 1190–1194. doi:10.1038/nature06343

Torisawa, T., Nakayama, A., Furuta, K., Yamada, M., Hirotsune, S., & Toyoshima, Y. Y. (2011). Functional Dissection of LIS1 and NDEL1 Towards Understanding the Molecular Mechanisms of Cytoplasmic Dynein Regulation. *Journal of Biological Chemistry*, *286*(3), 1959–1965. doi: 10.1074/jbc.M110.169847

Yamada, M., Toba, S., Yoshida, Y., Haratani, K., Mori, D., Yano, Y., et al. (2008). LIS1 and NDEL1 coordinate the plus-end-directed transport of cytoplasmic dynein. *The EMBO Journal*, *27*(19), 2471–2483. doi:10.1038/emboj.2008.182

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. doi:10.1093/molbev/msm088

Youn, Y. H., Pramparo, T., Hirotsune, S., & Wynshaw-Boris, A. (2009). Distinct Dose-Dependent Cortical Neuronal Migration and Neurite Extension Defects in Lis1 and Ndel1 Mutant Mice. *Journal of Neuroscience*, *29*(49), 15520–15530. doi:10.1523/JNEUROSCI.4630-09.2009

Zhou, X., Xu, S., Yang, Y., Zhou, K., & Yang, G. (2011). Phylogenomic analyses and improved resolution of Cetartiodactyla. *Molecular phylogenetics and evolution*, *61*(2), 255–264. doi: 10.1016/j.ympev.2011.02.009

Zilles, K., Palomero-Gallagher, N., & Amunts, K. (2013, February 15). Development of cortical folding during evolution and ontogeny. *Trends in neurosciences*. doi:10.1016/j.tins.2013.01.006

	GenBank Accession Numbers		
Taxa Sampled	***PAFAH1B1***	***NDEL1***	***NEUROG1***
Ailuropoda melanoleuca	---	XM_002921148.1	XM_002912966.1
Anolis carolinensis	GAFZ01046655.1	---	---
Bos taurus	NM_174663.2	NM_001191246.3	XM_585336.2
Callithrix jacchus	---	XM_003732891.1	XM_002744176.1
Canis lupus	XM_849389.2	XM_844953.2	XM_538637.2
Cavia porcellus	XM_003469673.1	XM_003466226.1	XM_003473262.1
Ceratotherium simum	XM_004433291.1	XM_004433082.1	XM_004420175.1
Cricetulus griseus	XM_003507268.1	XM_003497167.1	XM_003498161.1
Dasypus novemcinctus	---	---	XM_004475887.1
Equus caballus	XM_001918354.2	XM_003362452.1	---
Felis catus	NM_001114340.1	XM_003996200.1	---
Gallus gallus	---	NM_001030696.1	---
Gorilla gorilla	---	XM_004058561.1	XM_004042539.1
Homo sapiens	NM_000430.3	NM_030808.4	NM_006161.2
Loxodonta africana	XM_003416987.1	XM_003416918.1	XM_003404505.1
Macaca fascicularis	---	NM_030808.2	---
Macaca mulatta	NM_001258130.1	XM_002800284.1	XM_001110622.2
Monodelphis domestica	---	XM_001363909.1	---
Mus musculus	AY189217.1	NM_023668.2	NM_010896.2
Nomascus leucogenys	---	XM_003274660.1	XM_003266406.1
Odobenus rosmarus	XM_004404134.1	XM_004398450.1	XM_004405470.1
Orcinus orca	XM_004267042.1	XM_004266859.1	XM_004282101.1
Oryctolagus uniculus	---	NM_001082020.1	XM_002710221.1
Ornithorhynchus anatinus	XM_001508375.2	---	---
Otolemur garnettii	XM_003799075.1	XM_003791105.1	XM_003782176.1
Ovis aries	XM_004012556.1	XM_004012688.1	XM_004023187.1
Pan paniscus	XM_003809281.1	XM_003816018.1	XM_003829259.1
Pan troglodytes	---	XM_001166675.2	XM_003310829.1
Papio anubis	XM_003912082.1	XM_003919672.1	XM_003900121.1
Pongo abelii	NM_001131381.1	NM_001132583.1	XM_002815902.1
Rattus norvegicus	NM_031763.3	NM_133320.1	NM_019207.1
Saimiri boliviensis	XM_003933082.1	XM_003929225.1	XM_003920567.1
Sarcophilus harrisii	XM_003770123.1	XM_003768720.1	---
Sus scrofa	---	NM_001243862.1	---
Taeniopygia guttata	XM_002199771.2	---	---
Trichechus manatus	XM_004376143.1	XM_004375994.1	XM_004384702.1
Tursiops truncatus	XM_004315500.1	XM_004321197.1	XM_004327133.1
Xenopus laevis	NM_001090465.1	---	---

TABLE 1. Final list of taxa used in *PAFAH1B1*, *NDEL1*, *NEUROG1* phylogenetic analysis and tests for positive selection. DNA sequences were downloaded from GenBank's web site using the respective GenBank accession numbers. FASTA-formatted sequence information from GenBank was copied and pasted into TextWrangler for Mac OS X and edited to for use in MEGA 5.1 for Mac OS X, an alignment and phylogenetic analysis program. MEGA's CLUSTAL and/or MUSCLE alignment algorithms were executed against these taxa. Final DNA alignments were converted into their equivalent protein coding amino acids and compared against GenBank's predicted amino acid sequences to ensure that alignments were maximally correct.

Model	ln L	Average ω	parameters	Sites under positive selection (p > 0.95, BEB)
All mammals				
m0	-5218.7	0.0115	54	none
m3	-5218.7	0.0114	60	none
m1a	-5210.55	0.0152	55	none
m2a	-5210.67	0.0152	57	none
m7a	-5211.44	0.0123	55	none
m8a	-5210.00	0.0147	56	none
m8	-5211.44	0.0123	57	none
Nested Model Test	**LRT**	**df**	**X^2 test p-val**	**Conclusion [is p < 0.001 ?]**
m0 v. m3	2ΔlnL = 2e-6	6	p = 2.0e-19	**PURIFYING SELECTION = statistically significant**
m1a vs. m2a	2ΔlnL = 0.253258	2	p = 0.1189	purifying selection = not statistically significant
m7a vs. m8	2ΔlnL = 0.00819	1	p = 0.0721	purifying selection = not statistically significant
m8a vs. m8	2ΔlnL = 2.87622	1	p = 0.0899	purifying selection = not statistically significant

TABLE 2. Results of CODEML site model tests for selection, using the *PAFAH1B1* maximum likelihood bootstrap consensus tree derived from raxmlGUI v1.3. Three replicates of seven model tests, labelled m0 through m8, were run, and the resulting log likelihood values, average values of omega (w) and numbers of parameters were averaged across the three replicates. Nested model likelihood ratio tests were computed (LRT = 2 · ΔlnL), and degrees of freedom were determined by computing the difference between compared site model parameters. Chi-squared confidence intervals were computed for each nested model based on these value, producing four confidence intervals for each of the four nested model tests for positive selection. Confidence intervals with p < 0.001 are considered highly significant.

Model	**ln L**	**Average ω**	**parameters**	**Sites under positive selection (p > 0.95, BEB)**
All mammals				
m0	-5800.67	0.0433	64	none
m3	-5800.67	0.0433	70	none
m1a	-5794.25	0.0516	65	none
m2a	-5794.25	0.0516	67	none
m7a	-5778.52	0.0462	65	none
m8a	-5778.52	0.0462	66	none
m8	-5778.52	0.0463	67	none
Nested Model Test	**LRT**	**df**	**X^2 test p-val**	**Conclusion [is p < 0.001 ?]**
m0 v. m3	2ΔlnL =1e-05	6	p = 2.08e-17	**PURIFYING SELECTION = statistically significant**
m1a vs. m2a	2ΔlnL = 2e-06	2	p = 1.0e-06	**PURIFYING SELECTION = statistically significant**
m7a vs. m8	2ΔlnL = 0.0006616	1	p = 0.064827	purifying selection = not statistically significant
m8a vs. m8	2ΔlnL = 0.003962	1	p = 0.050189	purifying selection = not statistically significant

TABLE 3. Results of CODEML site model tests for selection, using the *NDEL1* maximum likelihood bootstrap consensus tree derived from raxmlGUI v1.3. Three replicates of seven model tests, labelled m0 through m8, were run, and the resulting log likelihood values, average values of omega (w) and numbers of parameters were averaged across the three replicates. Nested model likelihood ratio tests were computed ($LRT = 2 \cdot \Delta lnL$), and degrees of freedom were determined by computing the difference between compared site model parameters. Chi-squared confidence intervals were computed for each nested model based on these value, producing four confidence intervals for each of the four nested model tests for positive selection. Confidence intervals with $p < 0.001$ are considered highly significant.

Model	**ln L**	**Average ω**	**parameters**	**Sites under positive selection (p > 0.95, BEB)**
All mammals				
m0	-4734.63	0.0851	54	none
m3	-4633.78	0.1003	60	none
m1a	-4666.67	0.1705	55	none
m2a	-4666.67	0.1705	57	none
m7a	-4631.88	0.01119	55	none
m8a	-4631.87	0.01117	56	none
m8	-4631.89	0.01119	57	none
Nested Model Test	**LRT**	**df**	**X^2 test p-val**	**Conclusion [is p < 0.001 ?]**
m0 v. m3	2ΔlnL = 201.6896	6	p = 1	purifying selection = not statistically significant
m1a vs. m2a	2ΔlnL = 0.000308	2	p = 0.000154	**PURIFYING SELECTION = statistically significant**
m7a vs. m8	2ΔlnL = 0.004796	1	p = 0.055212	purifying selection = not statistically significant
m8a vs. m8	2ΔlnL = 0.03012	1	p = 0.137782	purifying selection = not statistically significant

TABLE 4. Results of CODEML site model tests for selection, using the *NEUROG1* maximum likelihood bootstrap consensus tree derived from raxmlGUI v1.3. Three replicates of seven model tests, labelled m0 through m8, were run, and the resulting log likelihood values, average values of omega (w) and numbers of parameters were averaged across the three replicates. Nested model likelihood ratio tests were computed (LRT = 2 · ΔlnL), and degrees of freedom were determined by computing the difference between compared site model parameters. Chi-squared confidence intervals were computed for each nested model based on these value, producing four confidence intervals for each of the four nested model tests for positive selection. Confidence intervals with $p < 0.001$ are considered highly significant.

Model	ln L	Average ω	parameters	Sites under positive selection (p > 0.95)
All mammals				
m0	-4448.93	**1.00094**	76	--
m3	-4430.62	**1.0703**	82	NEB: 55 S, w = 7.87, p = 0.991
				NEB: 319 L, w = 9.065, p = 1.000
m1a	-4441.6	0.83082	77	--
m2a	-4430.88	**1.0254**	79	BEB: 55 S, w = 4.495 ± 1.513, Pr(w>1) = 0.980
				BEB: 319 L, w = 4.549 ± 1.451, Pr(w>1) = 0.998
m7a	-4442.08	0.782	77	--
m8a	-4441.60	0.8308	78	--
m8	-4431.4148	0.9914	79	BEB: 55 S, w = 3.809 ± 1.326, Pr(w>1) = 0.984
				BEB: 319 L, w = 3.845 ± 1.288, Pr(w>1) = 0.998
Nested Model Test	**LRT**	**df**	**X^2 test p-val**	**Conclusion [is p < 0.001 ?]**
m0 v. m3	2ΔlnL = 37.22788	6	p = 1.59e-06	**POSITIVE SELECTION = statistically significant**
m1a vs. m2a	2ΔlnL = 21.43377	2	p = 2.22e-05	**POSITIVE SELECTION = statistically significant**
m7a vs. m8	2ΔlnL = 21.3274	1	p = 3.87e-06	**POSITIVE SELECTION = statistically significant**
m8a vs. m8	2ΔlnL = 20.3609	1	p = 6.41e-06	**POSITIVE SELECTION = statistically significant**

TABLE 5. Results of CODEML site model tests for selection, using the *MCPH1* maximum likelihood bootstrap consensus tree derived from raxmlGUI v1.3. Three replicates of seven model tests, labelled m0 through m8, were run, and the resulting log likelihood values, average values of omega (w) and numbers of parameters were averaged across the three replicates. Nested model likelihood ratio tests were computed (LRT = 2 · ΔlnL), and degrees of freedom were determined by computing the difference between compared site model parameters. Chi-squared confidence intervals were computed for each nested model based on these value, producing four confidence intervals for each of the four nested model tests for positive selection. Confidence intervals with $p < 0.001$ are considered highly significant.

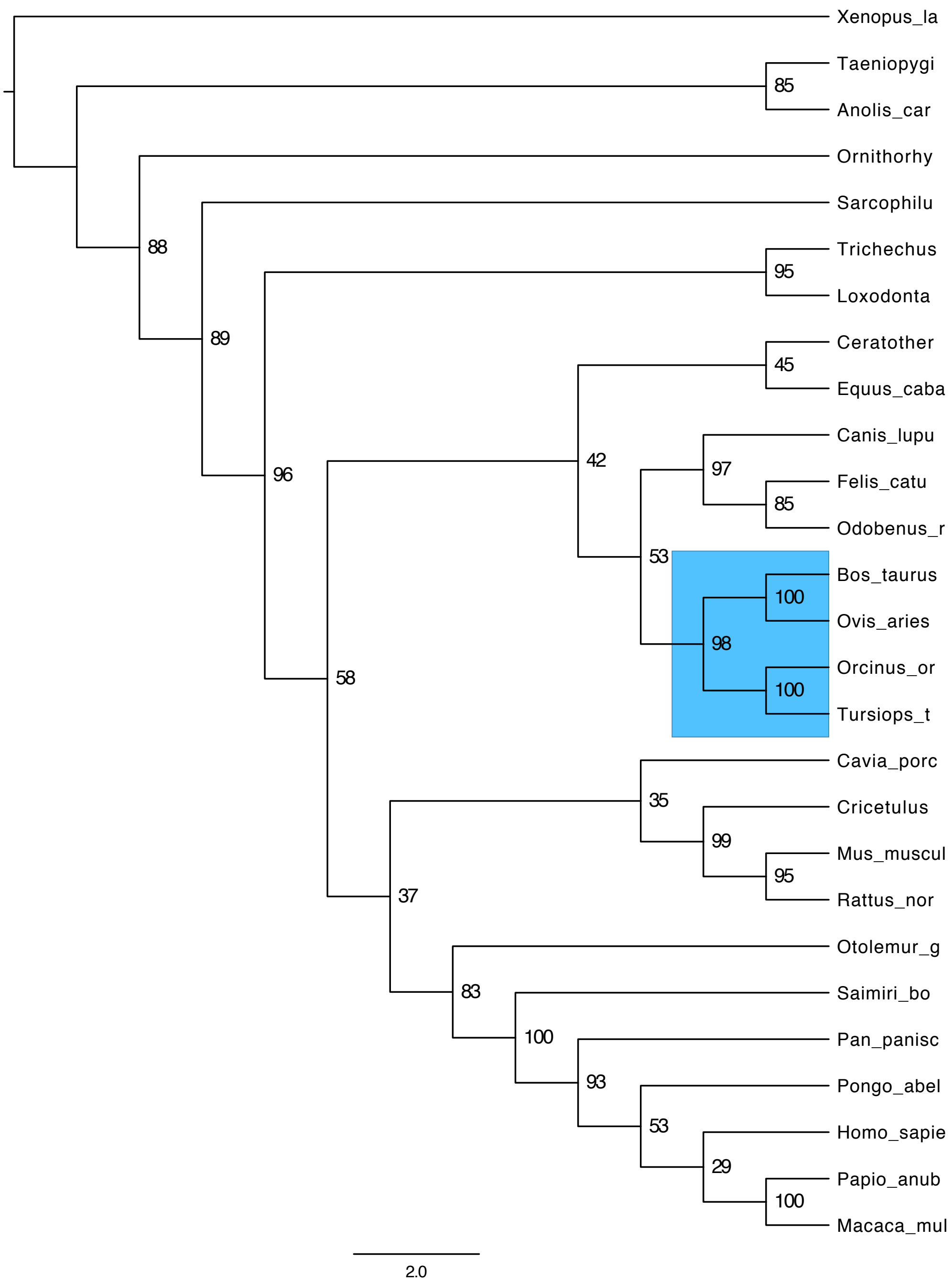

Figure 1. Maximum likelihood consensus tree for *PAFAH1B1.* These phylogenetic relationships were computed using raxmlGUI's maximum likelihood analysis engine with 5,000 thorough bootstrap replicates. Branch numbers are bootstrap values, which are a measure of how many maximum likelihood replicates recovered this branch out of the 5,000 replicates performed. Family Delphinidae *(Orcinus orca* and *Tursiops truncatus)* was recovered in 100 percent of bootstrap replicates, and Superorder Cetartiodactyla (indicated in blue) was recovered in 96 percent of the bootstrap replicates.

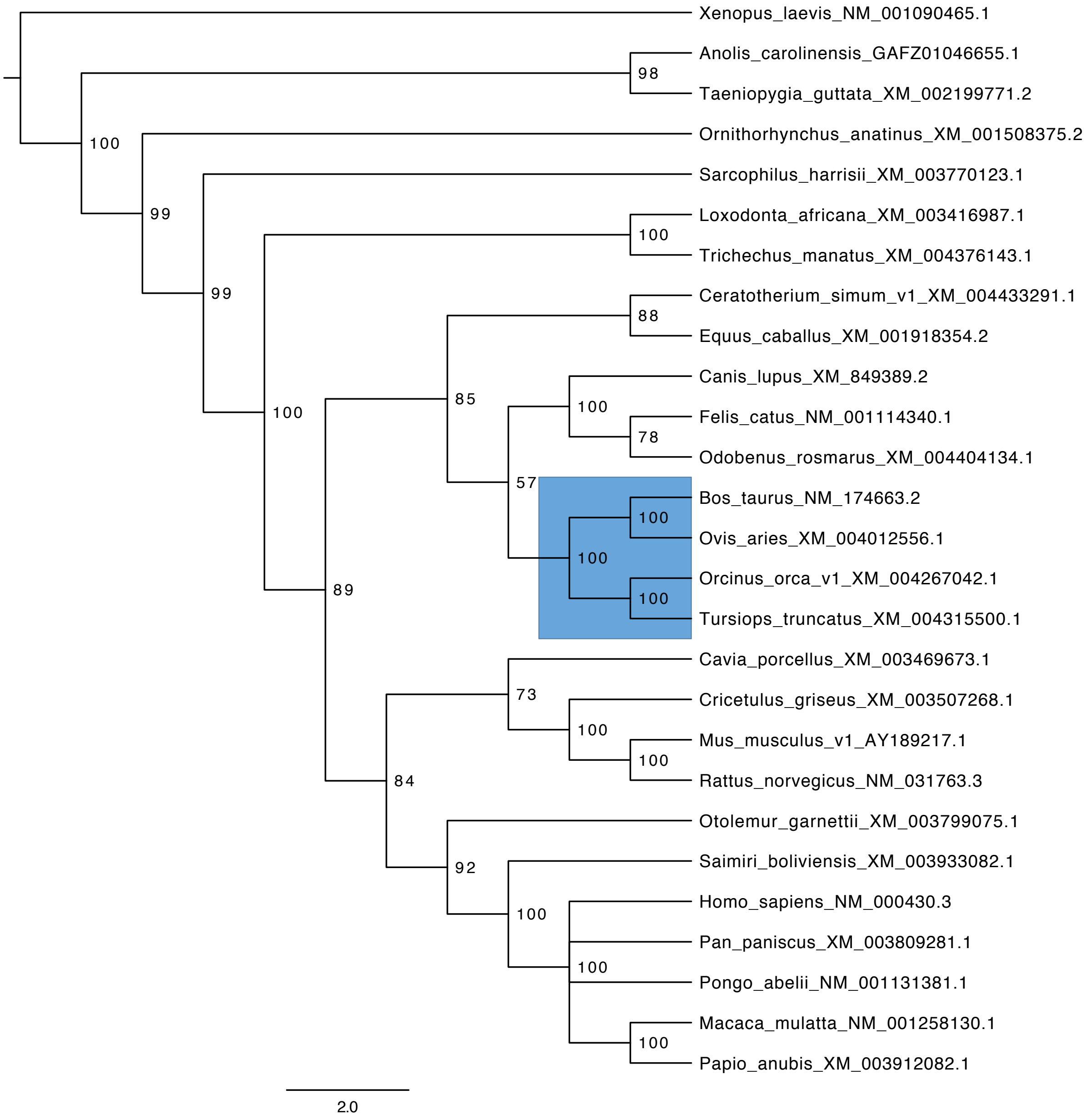

Figure 2. Bayesian analysis consensus tree for *PAFAH1B1.* These phylogenetic relationships were computed with MrBayes, using 200,000 Bayesian analysis replicates. Branch numbers are posterior probability values and represent a confidence interval (maximum = 100) based on all available evidence. Family Delphinidae *(Orcinus orca* and *Tursiops truncatus)* and Superorder Cetartiodactyla (indicated in blue) were recovered with posterior probabilities of 100.

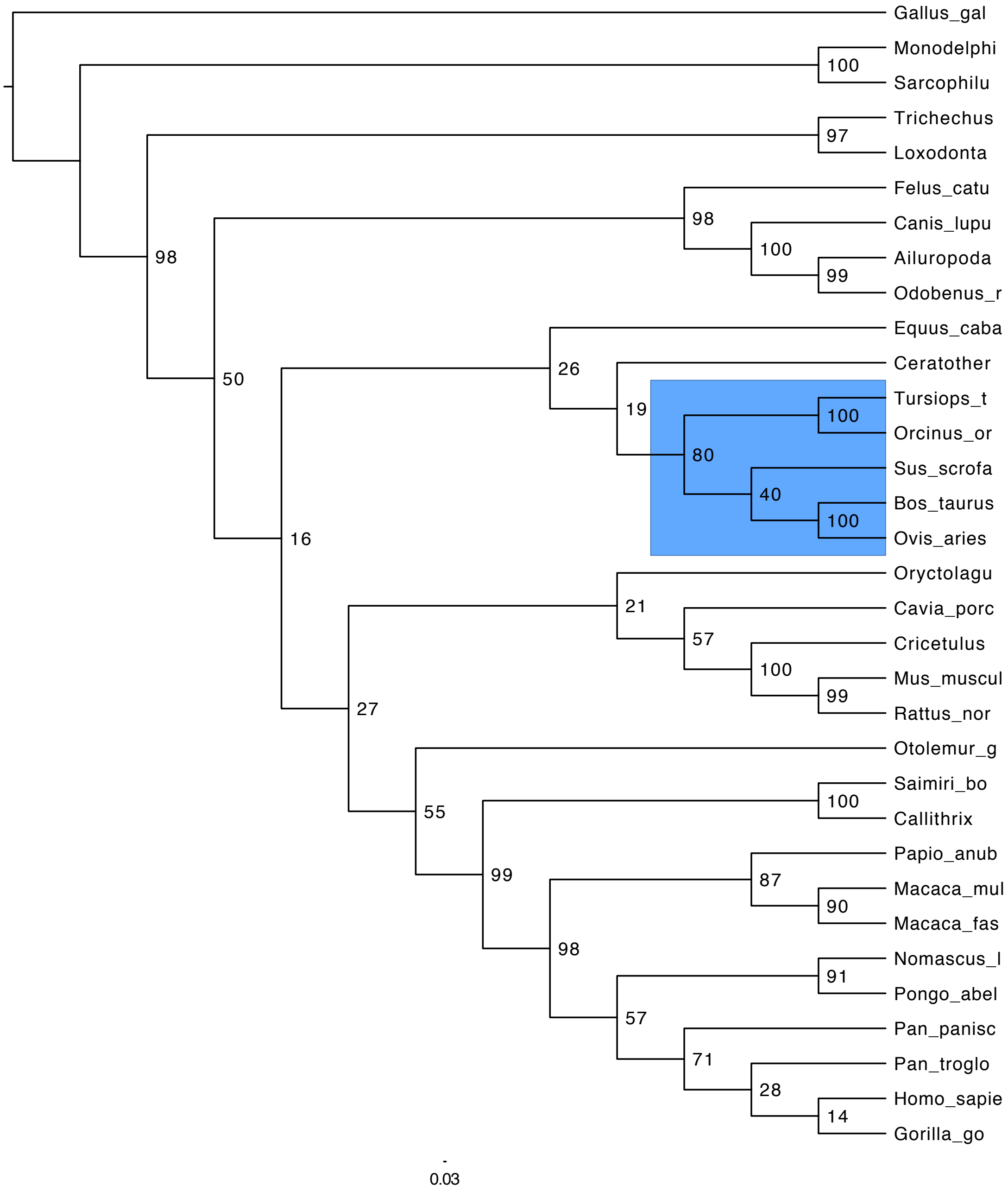

FIGURE 3. Maximum likelihood consensus tree for *NDEL1.* These phylogenetic relationships were computed using raxmlGUI's maximum likelihood analysis engine with 5,000 thorough bootstrap replicates. Branch numbers are bootstrap values, which are a measure of how many maximum likelihood replicates recovered this branch out of the 5,000 replicates performed. Family Delphinidae was recovered 100 percent of the time, and Superorder Cetartiodactyla (indicated in blue) was recovered in 80 percent of the bootstrap replicates. There is weak support for the early cetacean branching hypothesis, with *S. scrofa* basally branched with respect to other the other artiodactyls *(Bos taurus and Ovis aries)*.

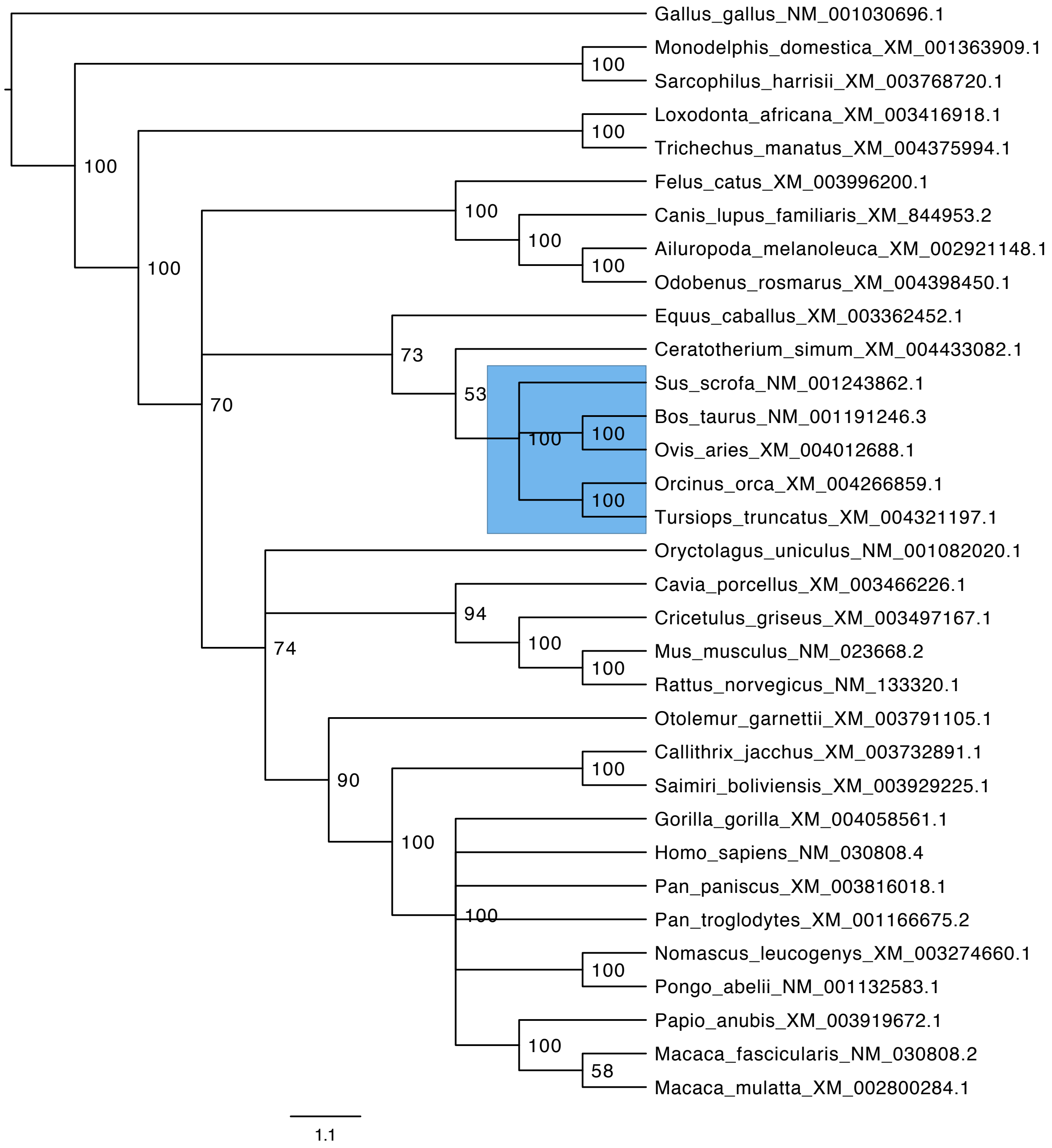

FIGURE 4. Bayesian analysis consensus tree for *NDEL1.* These phylogenetic relationships were computed with MrBayes, using 200,000 Bayesian analysis replicates. Branch numbers are posterior probability values and represent a confidence interval (maximum = 100) based on all available evidence. Family Delphinidae *(Orcinus orca* and *Tursiops truncatus)* and Superorder Cetartiodactyla (indicated in blue) were recovered with posterior probabilities of 100, but with *Sus scrofa* forming a soft Cetartiodactylan polytomy.

SUPPLEMENTAL NOTE 1

Files, Applications and Results / Data

The raw and aligned sequence files, the applications used to generate results, and resulting data sets of this research are available at http://cetaceanresearch.org/research/gyrencephaly_2013.

The specific parameters used in each program to generate results are detailed below:

RAxML

To generate maximum likelihood analyses with thorough bootstraps, the most recent version of raxmlGUI v1.3 was downloaded from http://sourceforge.net/projects/raxmlgui/. The following options were specified in raxmlGUI, which is equivalent to running the most recent version of RAxML with the following parameters:

```
ramxmlGUI options:
Add alignment file = <gene_name>_aligned.phy
Outgroup = unspecified; let raxmlGUI make its best assessments
Run mode = ML + thorough bootstrap
Run = 1
Replicates = 5000
BS brL = selected
Model = GTRGAMMA
```

RAxML command-line equivalence for these options is:

```
$raxmlHPC-PTHREADS-SSE3-Mac —T 2 —f c —m GTRGAMMA —s /Users/dave/
Dropbox/__MarGen/<gene_name>/<gene_name>_aligned.phy —n
<gene_name>_aligned.phy_red —w /Users/dave/Dropbox/__MarGen/
<gene_name>/″ -0
```

Once raxmlGUI maximum likelihood analysis and bootstraps were completed, a further step was taken in raxmlGUI to make the consensus bootstrap tree compatible with FigTree for Mac OS X, which appends node and branch bootstrap values to the tree file(s). This was done from raxmlGUI using the following user-interface seqeuence:

Utilites → Convert tree file to FigTree format

MrBayes

MrBayes version 3.2.1 for 64-bit Mac OS X systems was downloaded from http://mrbayes.sourceforge.net. MrBayes analysis was run from the command line using iTerm (http://iterm.sourceforge.net), using the following command sequences for all Bayesian analyses and posterior probability generations:

For each respective gene, where the gene's name = <*gene_name*>:

```
$ mb

> execute <gene_name>_aligned.nexus
> lset nst=6 rates=invgamma
> outgroup=X
> mcmc ngen=200000 samplefreq=100 printfreq=100 diagnfreq=1000
```

If analysis split frequency was not less than 0.01, an additional 200,000 replicates would have been performed. However, in all cases, 200,000 replicates was sufficient.

Once analysis was completed for *PAFAH1B1*, *NDEL1* and *NEUROG1*, results were saved to files using the following two commands:

```
> sump
> sumt
```

It is worth noting that many of the extended NEXUS file formats are not easily interpreted by MrBayes. As such this specific format of the NEXUS file should be used:

```
--

#NEXUS
begin data;
        dimensions ntax=XX nchar=YYY;
        format datatype=dna missing=? gap=-;
matrix
Orcinus_orca_XM_004282101.1           ATGCCAGCC…
< … all remaining taxa sequences… >
;
end;

--
```

PAML and CODEML

PAML, which includes the CODEML module, was downloaded from http://abacus.gene.ucl.ac.uk/software/paml.html. In order to make PAML's CODEML module work with the .phy files used in raxmlGUI, a copy of the file used in raxmlGUI was made, then edited to ensure that (1) each taxa had two spaces after its name, and (2) that the letter "i" for interleaves or the letter "s" for sequential was appended to the end of the first line, with two spaces between the number of taxa, the number of DNA sequences in the alignment, and the added letter. This is a peculiarity of paml's expectations for .phy files, but these edits made paml work perfectly with these slightly modified .phy files.

H(0)
m0 (one-ratio) -- assumes one w (=dN/dS) for all codons in the sequence
to compare m0 vs. m3
where P << 0.001 rejects m0
with df = 4

```
    seqfile = <genename>_aligned.phy  * NOTE: edit to include i on
first line
    treefile = <genename>_alltaxa.tree * NOTE: generate from HyPhy
    outfile = <genename>_m0_results.txt
    noisy = 3
    verbose = 1
    runmode = 0
    seqtype = 1
    CodonFreq = 2
    model = 0
    NSSites = 0
    icode = 0
    fix_kappa = 0
    kappa = 2
    fix_omega = 0
    omega = 5
```

--

H(A)
m3 (discrete) -- uses an unconstrained discrete distribution with all three site
    classes estimated from the data, with w(0) < 1 and w(1) = 1
to compare m3 vs. m0

```
    seqfile = <genename>_aligned.phy  * NOTE: edit to include i on
first line
    treefile = <genename>_alltaxa.tree * NOTE: generate from HyPhy
    outfile = <genename>_m3_results.txt
    noisy = 3
    verbose = 1
    runmode = 0
    seqtype = 1
```

```
    CodonFreq = 2
    model = 0
    NSSites = 3
    icode = 0
    fix_kappa = 0
    kappa = 2
    fix_omega = 0
    omega = 5


--------------------------------

H(0)
m1a (nearly neutral) -- assumes two site classes estimated with data,
with
    w(0) < 1 and w(1) = 1
to compare m1a vs. m2a --  -- tests whether or not the analyzed region
evolves under
    positive selection, using comparisons to their nested neutral
models
where P < 0.001 rejects m1a
with df =  2

    seqfile = <genename>_aligned.phy  * NOTE: edit to include i on
first line
    treefile = <genename>_alltaxa.tree * NOTE: generate from HyPhy
    outfile = <genename>_m1a_results.txt
    noisy = 3
    verbose = 1
    runmode = 0
    seqtype = 1
    CodonFreq = 2
    model = 0
    NSSites = 1
    icode = 0
    fix_kappa = 0
    kappa = 2
    fix_omega = 0
    omega = 5


--

H(A)
m2a (positive selection - alternative hypothesis model) -- adds a
third class of sites
    to m1a, with w(2) > 1
to compare m1a vs. m2a
```

```
    seqfile = <genename>_aligned.phy  * NOTE: edit to include i on
first line
    treefile = <genename>_alltaxa.tree * NOTE: generate from HyPhy
    outfile = <genename>_m2a_results.txt
    noisy = 3
    verbose = 1
    runmode = 0
    seqtype = 1
    CodonFreq = 2
    model = 0
    NSSites = 2
    icode = 0
    fix_kappa = 0
    kappa = 2
    fix_omega = 0
    omega = 5


--------------------------------

H(0)
m7 (beta) -- a flexible null model, in which the w ratio for a codon
is a random draw
    with a beta distribution with 0 < w < 1
to compare m7 vs. m8 -- tests whether or not the analyzed region
evolves under
    positive selection, using comparisons to their nested neutral
models
with df = 2

    seqfile = <genename>_aligned.phy  * NOTE: edit to include i on
first line
    treefile = <genename>_alltaxa.tree * NOTE: generate from HyPhy
    outfile = <genename>_m7_results.txt
    noisy = 3
    verbose = 1
    runmode = 0
    seqtype = 1
    CodonFreq = 2
    model = 0
    NSSites = 7
    icode = 0
    fix_kappa = 0
    kappa = 2
    fix_omega = 0
    omega = 5


--
```

H(A)

m8 (beta and w) -- adds an extra class site to model m7, with a proportion of w(s) > 1 estimated from the data

to compare m7 and m8a vs. m8

```
    seqfile = <genename>_aligned.phy  * NOTE: edit to include i on
first line
    treefile = <genename>_alltaxa.tree * NOTE: generate from HyPhy
    outfile = <genename>_m8_results.txt
    noisy = 3
    verbose = 1
    runmode = 0
    seqtype = 1
    CodonFreq = 2
    model = 0
    NSSites = 8
    icode = 0
    fix_kappa = 0
    kappa = 2
    fix_omega = 0
    omega = 5
```

--

H(0)

m8a (beta and w(s)=1) -- introduced by Swanson et al.; similar to m8 except that the category w(s) is fixed at w(s) = 1, specified in CODEML using NSSite = 8

to compare m8 vs. m8a -- tests for evidence of positive selection while eliminating the potential identification of relaxed purifying selection with df = 1

```
    seqfile = <genename>_aligned.phy  * NOTE: edit to include i on
first line
    treefile = <genename>_alltaxa.tree * NOTE: generate from HyPhy
    outfile = <genename>_m8a_results.txt
    noisy = 3
    verbose = 1
    runmode = 0
    seqtype = 1
    CodonFreq = 2
    model = 0
    NSSites = 8
    icode = 0
```

```
    fix_kappa = 0
    kappa = 2
    fix_omega = 1
    omega = 1

--------------------------------
```

After installing paml to /usr/local/paml, codeml is run for each model using codeml.ctl files above, as follows:

```
$ /usr/local/paml/bin/codeml ./<model_number>.ctl
```

Assign significance of detection of positive selection on the selected branch, as follows:

Retrieve likelihood values lnL(H(A)) and lnL(H(0)) from alternative and null hypothesis results files generated above.

Then reconstruct the Likelihood Ratio Test (LRT), as follows:

deltaLRT = 2 · (lnL(H(A)) - lnL(H(0)) (e.g.: 2 * ((-5710) - (-5712)) = 4)

In the above line, if deltaLRT = 4, and if chi^2 curve has one degree of freedom (check the results of "$grep lnL *.results" for np: XX values of respective tests), so p-value for chi^2 test = some small value under chi^2, so result is significant.

In cases where the result is significant, it is possible to retrieve the sites under positive selection using Bayes Empirical Bayes (BEB) method, which is described here: http://dx.doi.org/10.1093/molbev/msi097

e.g.:

```
Positive sites for foreground lineages Prob(w>1):
    36 K 0.971*
    159 C 0.993**
```

Amino acids K and C refer to the first sequence in the alignment. Position 36 has a high probability (97.1%) of being under positive selection. Position 159 has a very high probability (99.3%) of being under positive selection.

See Table 4 for results that show this happening.

SUPPLEMENTAL NOTE 2: Lissencephaly 1 amino acids coded by *PAFAH1B1* in selected taxa.

Anolis
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDVNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIASASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYATISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDISTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Bos
MVLSQRQRDELNRAIADYLRSNGYEAAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVEDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Canis
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Cavia
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Cerato
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHTGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Cricetul
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Equus
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQ??VMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE

TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Felis
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Homo
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDVNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Loxodonta
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDFKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Macaca
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLTSCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHLVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Mus
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Odobenus
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Orcinus
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSCISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK

TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Ornitho
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Otolemur
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Ovis
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Pan
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Papio
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLTSCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHLVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Pongo
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSIQDIPFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Rattus
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

```
Saimiri
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Sarcophil
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYSLSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSSISEATGSE
TKKSSKPGPILLSGSRDKTIKMWDVITGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Taeniop
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELHVNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHTGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSTISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDISTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDFKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Trichechus
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSTISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Tursiops
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDKKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPLGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVSASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECIRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSCISEATGSE
TKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNWVRGVLFHSGGKFILSCADDK
TLRVWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*

Xenopus
MVLSQRQRDELNRAIADYLRSNGYEEAYSVFKKEAELDMNEELDIKYAGLLEKKWTSVIR
LQKKVMELESKLNEAKEEFTSGGPIGQKRDPKEWIPRPPEKYALSGHRSPVTRVIFHPVF
SVMVTASEDATIKVWDYETGDFERTLKGHTDSVQDISFDHSGKLLASCSADMTIKLWDFQ
GFECLRTMHGHDHNVSSVAIMPNGDHIVSASRDKTIKMWEVQTGYCVKTFTGHREWVRMV
RPNQDGTLIASCSNDQTVRVWVVATKECKAELREHEHVVECISWAPESSYSTISDATGSE
TKRSGKPGPFLLSGSRDKTIKMWDISIGMCLMTLVGHDNWVRGVQFHPGGKFILSCADDK
TIRIWDYKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSVDQTVKVWECR*
```

SUPPLEMENTAL NOTE 3: nudE-like 1 amino acids coded by *NDEL1* in selected taxa.

```
Gallus_gallus_NM_001030696.1
MDSEEIPTFSSPKEETAYWKELSLKYKQSFQEAREELAEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVETLKEKLEHQYAQSYKQVSLLEDDLSQTRAIKDQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDDKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGSENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNANSSMMSSNG
TKYPHPGHTSFFDKGAVNGFDQGTPGL?-?ASRPSSAPGMLPLSV*

Monodelphis_domestica_XM_001363909.1
MDGEEIPDFSSLKEETAYWKGLSLKYKQSFQEAQEELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKGIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGLVNSNS
TKFSRSGHTSFFDKGTVNGLDQVPPT-GLGSSRPSSAPGMLPLSV*

Sarcophilus_harrisii_XM_003768720.1
MDGEEIPDFSSLKEETAYWKELSLKYKQSFREAQEELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFSSPKGIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGLVNSNS
TKFSRSGHTSFFDKGTVNGLDQVPPT-GLGSSRPSSAPGMLPLSV*

Ailuropoda_melanoleuca_XM_002921148.1
MDGEDIPDFSSLKEETAYWKELSMKYKQSFQEARDELAEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNRRLRCDVEALKEKLELQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
ASPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVTNGSG
TKFPRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Bos_taurus_NM_001191246.3
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENNFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Callithrix_jacchus_XM_003732891.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRTGHTSFFDKGAVNGFDPAPPP-GLGSSRPSSAPGMLPLSV*

Canis_lupus_familiaris_XM_844953.2
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
ASPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Cavia_porcellus_XM_003466226.1
MDGEDIPNFSSLKEENAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLEQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQAARKSYISGNVNCGVMNSNG
```

```
TKFSRSGHTSFFDKGAVNGFDPAPPPPSLGSSRPSSAPGMLPLSV*

Ceratotherium_simum_XM_004433082.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Cricetulus_griseus_XM_003497167.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKPYIPGSVNCGVMNSNG
PECPRSGRGTFFHKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Equus_caballus_XM_003362452.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Felus_catus_XM_003996200.1
MDGEDIPNFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYLSGNVNCGLMNSDS
TKFPRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Gorilla_gorilla_XM_004058561.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Homo_sapiens_NM_030808.4
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Loxodonta_africana_XM_003416918.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSCSGHSSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Macaca_fascicularis_NM_030808.2
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
```

TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Macaca_mulatta_XM_002800284.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Mus_musculus_NM_023668.2
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYVPGSVNCGVMNSNG
PECPRSGRATFFHKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Nomascus_leucogenys_XM_003274660.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGMENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Odobenus_rosmarus_XM_004398450.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
ASPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Orcinus_orca_XM_004266859.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPASPPPGLGSSRPSSAPGMLPLSV*

Oryctolagus_uniculus_NM_001082020.1
MDGEDIPDFSSLKEETAYWKELSLKYKQTFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPLSAPGMLPLSV*

Otolemur_garnettii_XM_003791105.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Ovis_aries_XM_004012688.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG

```
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Pan_paniscus_XM_003816018.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARTSALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Pan_troglodytes_XM_001166675.2
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Papio_anubis_XM_003919672.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKVAPDGLR
AILLNLSSKTSTLGVCCCLL*IFHALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Pongo_abelii_NM_001132583.1
MDSEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Rattus_norvegicus_NM_133320.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYVPGSVNCGVMNSNG
PECPRSGRATFFHKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Saimiri_boliviensis_XM_003929225.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENTFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVLNGNG
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Sus_scrofa_NM_001243862.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
AKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Trichechus_manatus_XM_004375994.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
```

```
TKFSRSGHTSFFDKGAVNGFDPAPPPPGLGSSRPSSAPGMLPLSV*

Tursiops_truncatus_XM_004321197.1
MDGEDIPDFSSLKEETAYWKELSLKYKQSFQEARDELVEFQEGSRELEAELEAQLVQAEQ
RNRDLQADNQRLKYEVEALKEKLEHQYAQSYKQVSVLEDDLSQTRAIKEQLHKYVRELEQ
ANDDLERAKRATIVSLEDFEQRLNQAIERNAFLESELDEKESLLVSVQRLKDEARDLRQE
LAVRERQQEVTRKSAPSSPTLDCEKMDSAVQASLSLPATPVGKGTENSFPSPKAIPNGFG
TSPLTPSARISALNIVGDLLRKVGALESKLAACRNFAKDQASRKSYISGNVNCGVMNSNG
TKFSRSGHTSFFDKGAVNGFDPASPPPGLGSSRPSSAPGMLPLSV*
```

SUPPLEMENTAL NOTE 4: Neurogenin-1 amino acids coded by *NEUROG1* in selected taxa.

```
Orcinus_or
MPAPLETCLSDLDCASTSDLSGFLTDEEDCARLQQPTSSSGPPMTARKGSPGIPGASDTP
RAQDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLQPQCAPCLPGPPSPASD
AESWGSGAAASPCAAAASPLSDPSSPAASEDFTYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Tursiops_t
MPAPLETCLSDLDCASTSDLSGFLTDEEDCAR??PLSSCSKLSVPVLASSPAASEDSYGP
G---DPQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCSPCLPGPPSPASD
AESWGSGAAASPCAAAASPLSDPSSPAASEDFTYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Ailuropoda
MPARLETCLSDLDCTSSSDLSGFLTDEEDCARLQPPASASGTPVPARRGAPGVPAASDTP
RAQDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGG-----------------------
------GAAASPCAAAASPLSDPSSPAASEDFGYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Bos_taurus
MPATLETCLSDLDCASTSDLSGFLTDEEDCARLQQPASATGPPVTVRRGAPGIPGAPDTP
RAPDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGTRERLLPPQCAPCLPGPPSPASD
AESWASGAAASPCAAAASPLSDPSSPVASEDFTYGPGDPLFSFPGLPKDLLHTTPRFIPY
H*

Callithrix
MPAPLETCISDLDCASSSDLSGFLTDEEDCARLQPPASASGPPAPVRRGAPSTSRASEVP
VAQDEEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGSGARERLLSPQCVPCLPGPPSPASD
AESWGSGAAASPSAAAASPLSDPSSPAASEDFTYGPGDPIFSFPGLPKDLLHTTPCFIPY
H*

Canis_lupu
MPAPLETCLSDLDCASSSDLSGFLTDKEDCARLPTPAPASGTPVPVRRAVPGVPAAPDTP
RAQDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCAPCLPGPPSPSSD
AESWGSGAAASPCATAASPLSDPSSPAASEDFAYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Cavia_porc
MPAPLETCLSDLDCANSSELSGFLTDEEDCARFPSPASASGPPAPARRGARGMPAATDVP
VAQDEEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGSGARERLLSPQCAPCLPRPPSPASD
AESWGSGAAASPCAAAASPLSDPSSPAASEDFPYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Ceratother
MPASLDTRLSDLDCASSSDLSGFLTDEEDCARLQPPGSASGPPVPACRGAPGIPGASDTS
RAQDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCAPCLPGPPSPASD
SESWGSGAASSPCTAAASPLSDPSSPAASEDFAYGPGDPLFSFPGLPKDLLHTTPCFIPY
Q*

Cricetulus
MPAPLETCLSDLDCASSSDLSKFLTDEEDCARLQPLASTSGLSVPARRSAPTIS---NVP
GAQDEEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCAPCLPGPPSPASD
```

```
AESWGSGAASSPCATVASPLSDPSSPSAPDDFTYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Dasypus_no
MPAPLETCLSDLDCASSSDLSGFLTDEEDCARLQPPASASDPPAPPRRSAPGIPGGSDAS
GAQDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGAGARERLLPPQCAPCLPGPASPASD
SESWGSGAASSPCAAAASPLSDPSSPAASEDFSYGPGDSLFSFPGLPKDLFHRAPCFAPY
H*

Gorilla_go
MPAPLETCISDLDCASGSDLSGFLTDEEDCARLQQAASASGPPAQARRGAPNISRASEVP
GAQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCVPCLPGPPSPASD
AESWGSGAAA------ASPLSDPSSPAASEDFTYRPGDPVFSFPSLPKDLLHTTPCFIPY
H*

Homo_sapie
MPARLETCISDLDCASGSDLSGFLTDEEDCARLQQAASASGPPAPARRGAPNISRASEVP
GAQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCVPCLPGPPSPASD
AESWGSGAAA------ASPLSDPSSPAASEDFTYRPGDPVFSFPSLPKDLLHTTPCFIPY
H*

Loxodonta_
MPASLETCLSDLDCSSSSDLSGFLTDEEDCARLQPPGSASGPPAPPRRGPPGIPGAPDAP
GAQD-EQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCVPCLPGPPSPASD
AESWGSGAAASPCAAATSPLSDPSSPAASEDFAYGPSDPLFPFPGLSKDLLHTTPCFMPY
H*

Macaca_mul
MPAPLETCISDLDCASGSDLSGFLTDEEDCTRLQQAASASGPPVPARRGAPNISRASEVP
GAQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGSARERLLPPQCVPCLPGPPSPASD
AESWGSGAAA------ASPLSDPSSPAASEDFTYGPSEPAFSFSSLPKDLLHTTPCFIPY
H*

Mus_muscul
MPAPLETCISDLDCSSSSDLSSFLTDEEDCARLQPLASTSGLSVPARRSAPALSGASNVP
GAQDEEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGSARERLLPPQCVPCLPGPPSPASD
TESWGSGAAASPCATVASPLSDPSSPSASEDFTYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Nomascus_l
MPAPLETCISDLDCASGSDLSGFLTDEEDCARLQQAASASGPPAPARRGAPNISRASEVP
GAQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGGRERLLPPQCVPCLPGPPSPASD
TESWGSGAAA------ASPLSDPSSPAASEDFTYGPGDPVFSFPSLPKDLLHTTPCFIPY
H*

Odobenus_r
MPAPLETCLSDLDCASSSDLSGFLTDEEDCARLQPPASASGTPVPVRRGAPGVPAASDTP
RAQEDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCAPCLPGPPSPSSD
AESWGSGAAASPCAAAASPLSDPSSPAASEDFAYGPGDPLFSFPSLPKDLLHATPCFIPY
H*

Oryctolagu
MPAPLETCLSDLDCASGGHLSGFLTDEDDCARLQPPVSASGPPAPARRSVPGVSGAPGVA
DAQDHEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKFTKIETLRFAYNYIWALAETLRLADQGLSGGGARERLLSPQCAPCLPGPPSPASD
AESWGSAPAASPCAAAASPLSDPSSPATSDDFAYAPGDPLFSFPSLSKDFLHTTPGFIPY
```

```
H*

Otolemur_g
MPAPLETCLSDLDCASSSDLSGFLTDEEDCAKLPPPASASGPPAPARRNTPGISGASDVP
GAQDEEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLSGGSARERLLPPQCAPCLPGPPSPASD
AESWGSGAAASPCAAAASPISDPSSPATSEDFAYGPGDPLFSFPGLPKDLLHTTPCFIPY
H*

Ovis_aries
MPATLETCLSDLACASTSDLSGFLTDEEDCARLQQPASASGPPVTARRGASGIPGAPDTP
RAPDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLA-------------------------PSPASD
AESWASGAAASPCAAAASPLSDPSSPVASEDFTYGPGDPLFSFPGLPKDLLHTTPRFIPY
H*

Pan_panisc
MPAPLETCISDLDCASGSDLSGFLTDEEDCARLQQAASASGPPAPARRGAPNISRASEVP
GAQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCVPCLPGPPSPASD
AESWGSGAAA------ASPLSDPSSPAASEDFTYRPGDPVFSFPSLPKDLLHTTPCFIPY
H*

Pan_troglo
MPAPLETRISDLDCASGSDLSGFLTDEEDCARLQQAASASGPPAPARRGAPNISRASEVP
GAQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCVPCLPGPPSPASD
AESWGSGAAA------ASPLSDPSSPAASEDFTYRPGDPVFSFPSLPKDLLHTTPCFIPY
H*

Papio_anub
MPAPLETCISDLDCASGSDLSGFLTDEEDCARLQQAASASGPPVPARRGAPNISRASEVP
GAQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGSARERLLPPQCVPCLPGPPSPASD
AESWGSGAAA------ASPLSDPSSPAASEDFTYGPSEPAFSFPSLPEDLLHTTPCFIPY
H*

Pongo_abel
MPAPLETCISDLDCASGSDLSGFLTDEEDCARLQQAASALGPPAPARRGAPNISRASEVP
GTQDDEQERRRRRGRTRVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCVPCLPGPPSPASD
AESWGSGAAA------ASPLSDPSSPAASEDFTYGPGDPVFSFPSLPKDLLHTTPCFIPY
H*

Rattus_nor
MPAPLETCLSDLDCASGSDLSSFLTDEEDCARLQPLASTSGLSVPARRSAPTLSGASNVP
GGQDEEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQCVPCLPGPPSPASD
TESWGSGAAASPCATVASPLSDPSSPSASEDFTYGPGGPLFSFPGLPKDLLHTTPCFIPY
H*

Saimiri_bo
MPAPLETCISDLDCASSSDLSGFLTDEEDCARLQPPASASGPPAPVRRGAPSTSRASEVP
GSQDEEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGSGARERLLPQQCVPCLPGPPSPASD
AESWGSGAAASPSAAAASPLSDPSSPAASEDFTYGLGDPVFSFPGLPKDLLHTTPCFIPY
H*

Trichechus
MPASLETCLSDLDCSSNSDLSGFLTDEEDCARLQPPDSASGPPAPPRRGAPGISGAPDAP
GAQDDEQERRRRRGRARVRSEALLHSLRRSRRVKANDRERNRMHNLNAALDALRSVLPSF
PDDTKLTKIETLRFAYNYIWALAETLRLADQGLPGGGARERLLPPQSAPCLPGPPSPISD
AESWGSGAAASPCAAATSPLSDPSSPAASEDFAYGPGDPLFPFPGLSKDLLHTTPCFIPY
H*
```

SUPPLEMENTAL NOTE 5: Microcephalin amino acids coded by *MCPH1* in selected taxa.

```
Orcinus1_o
DSFAGVLRLSFDDLCGSSGYGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAERVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPTPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSARGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAKARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Steno_bred
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPTPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDRPRAEARGWGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKST?KEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Pseudorca_
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLFPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSRVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Ldelphis1_
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPTGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDITSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Lrhyn1_alb
DSFAGVLHLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPTPRLQLENSLKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Grampus_gr
DSFAGVLRLSFDD?CGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVCPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLEKSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLFPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Peponoceph
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLFPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGF?LQKPANYGSDPGLSSVTSKETPAAEETPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELP?G
```

```
Globicepha
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRDLSPTPSVTNEPAG
GPSCPQSPSATRRRVSVNSSPPPEERLSKRRCLGRPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLFPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGSDPGLSSVTSKETPAAEETPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Sousa_chin
DSFAGVLRLSFDDLCGSSGRGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPATPGPLG
QLTLHQPPGSLSKGDTRRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPTPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Turs1_adun
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDEIKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSATPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Stenella_c
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESK?AACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
G?SCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFS?IGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCP?AEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDK?NFVGDWDVEKSAEAMEELPRG

Stenella_a
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPANPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSS?VETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Turs2_trun
DSFAGVLRLSFDDLCGSSGCGTQQRR?GGFLDESKSAACASSAVL?TRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLS?TPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSC?GR
SPRSAGG?DSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Stenella_l
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLG?
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPGAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG
```

```
Dnus1_cape
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
G?SCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGV?PPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAG?TDSTVKAGFRLQKPANYGAD?GLSSVTSKETPAAEGTPGDCPGAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Dnus2_cape
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
G?SCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENS?KCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGV?PPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAG?TDSTVKAGFRLQKPANYGAD?GLSSVTSKETPAAEGTPGDCPGAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Lissodelph
DFFAGVLHLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPATPGPLG
QLTLHQPPGSLSKGDT?RQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKC?SSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETRAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTP?KGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTP-KMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Cephalorhy
DSFAGVLHLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPATPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSVNSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETRAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTP-KMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Dnus_delph
DSFAGVLRLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKTRSVRPPASPGPLG
QLTLHQP?GSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKR?CLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGV?PPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAG?TDSTVKAGFRLQKPANYG?D?GLSSVTSKETPAAEGTPGDCPGAEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVRGWIDITSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Lrhyn2_acu
DSFAGVLHLSFDDLCGSSGRGTQQRRLGGFLDESKSAACASSAVLKARGVRPPASPGPLG
QLTPHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVRPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEARGGGDAR
PRGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Lrhyn3_obl
DSFAGVLHLSFDDLCGSSGCGTQQRRLGGFLDESKSAACASSAVLKARSVRPPATPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTNEPA?
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPRVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETRAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTLHGLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTP-KMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Lrhyn4_obs
```

```
DSFAGVLHLSFDDLCGSSGCGTQQRRLGGFLD?SKSAACASSAVLKARSVRPPATPGPLG
QLTLHQPPGSLSKGDTHRQRVAAGQVIT?DVKPAREPAEGVFDEKRGLSPTPSVTNEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLL??VQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETRAAEGTPGDCPRAEARGGGDAR
PGGSDSPHTL?GLTPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTP-KMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Sotalia_fl
DSFAGVLRLSFDDLCGSSGRGTQQRRLGGFLDESKSAACASSAVLKTRSVRQPATPGPLG
QLTLHQPPGSLSKGDTRRQRVAVGQVITPDVKPAREPAEGVFDEKRGLSPMPSVTNEPAG
GPPCPQSPSATRRRVSANSSPPPEERLSKRRCLGWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSGSLLPGVQPPAGPALLHCMRSLSRRERTSLLGRADFSCIGR
SPRSAGGTDSTVKAGFHLQKPANYGAGPGLSSVTSKETPAAEGTPGDCPRVEARGGGDAR
PGGSDSPHTLNGLTPQKGLRGDSTSLKGSNKEVRGWVDTTSTQKGDTPSEMVNSPESEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Phocoena_p
DSFAGVLHLSFDDLCGSSGCGTQQRKLGGFLDESKSAACASSAVLKTRSVRPPASPGSLG
QLTLHQSPGSLSKGDTHRQGVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTKEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRSLRWPPAPRLQLESSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSESLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCP?AEARGGEDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVEGWIDIKSTQKEDTPSKMVNSPEGEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Phocoenoid
DSFAGVLHLSFDDLC?SSGCGTQQRKLGGFLD?S?SAACASSAVLKTRSVRPPASPGSLG
QL?LHQSPGSLSKGDTHRQGVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTKEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRSLRWP?APRLQLESSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSESLL?GVQPPAGPALLHCMRSLSRRER?SLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGD?PRAEARGGEDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVEGWIDIKSTQKEDTPSKMVNSPEGEAQ
SDDKLNFVGDWDVEKSAEAMEELPRG

Monodon_mo
DSFAGVLHLSFDDLCGSSGCGTQQRKLGGFLDESKSAACASSAVLKTRSVRPPASPGSLG
QLTLHQSPGSLSKGDTHRQGVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTKEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRSLRWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSESLLPGMQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPWAEARGGEDTR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVKGRIDIKSTQKEDTPSKMVNSPEGEAQ
SDDKLNFVGDWDVEESAEAMEELP*G

Neophocaen
DSFAGVLHLSFDDLCGSSGCGTQQRKLGGFLDESKSTACASSAVLKTRSVRPPASPGSLG
QLTLHQSPGSLSKGDTHRQGVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTKEPAR
GPSCPQSPSATRRRVSANSSPPPEERLSKRRSLRWPPAPRLQLESSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSESLLPGVQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADLGLSSVTSKETPAAEGTPGDCPRAEARGGEDAR
PGGSDSPHTLNGLTPQKGLRGDFTSLKGSNKEVEGWIDIKSTQKEDTPSKMVNSPEGEAQ
SDDKLNFVGDWDVEKSAEAMEEL?RG

Delphinapt
DSFAGVLHLSFDDLCGSSGCGTQQRKLGGFLDASKSAACASSAVLKTRSVRPPASPGSFG
QLTLHQSPGSLSKGDTHRQGVAAGQVITPDVKPAREPAEGVFDEKRGLSPTPSVTKEPAG
GPSCPQSPSATRRRVSANSSPPPEERLSKRRSLRWPPAPRLQLENSPKCASSSAVETLGR
GESSYDDYFSPDNLKERDSESLLPGMQPPAGPALLHCMRSLSRRERTSLLERADFSCIGR
SPRSAGGTDSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPWAEARGGEDTR
PGGSDSPHTLNGLSPQKGLRGDFTSLKGSNKEVKGWIDIKSTQKEDTPSKMVNSPECEAQ
SDDKLNFVGDWDVEESAEAMEELP*G

Berardius_
DSFAGVLHLSFEGLCGSSGCGTQQRKLGGFLDESKSAACASSAVLKTISVRPPASPGSLG
```

```
QLTLHQSPGSLSRGDTHRQRGAAGQVITPDAKPVRELAEGVFDEKRGLSPTPSVTKEPPG
SPSCPQSPSATRRRVSANSSPPLEKRLSKRRSLRWPPAPRLQPENSPKCASSSAVETLGC
GESSYDDYFSPDNLKERSSESLLAGVQPPAGHALLHCMRSLSRRERTSLLGRADFSCIGR
SPRSADGTDSAVKAGFRLQKPANCRADPGLSSVTSEETPAAEGTPGDCPRAEAQRPGDAC
PGGSDSPHTLDGLT??KGSRGDFTPLKGSNKEVKGWVDIKSTQKEDTPSKMVNSPEGEAQ
SDDKLNFVGDCDVEKSTEAREEFPRG

Platanista
DSFAGVLHLSFDDLCGSSGCGTQQRKLGGLLDESKSAACASSAVLKTSSVRPPASPGSLG
QLTLHQSPGSLSKGDTPQQRGAAGRVITPDAKPAREPAEGVLDEKRGLSPTPSVTKESPG
GPSSPQSPSATRRRVSANSSPPPGERWSK----GWPPTPRLQPENSPECASSSAVETLGR
GGSSYDDYFSPDNLKERDLESLLPGVRPPAGPALLHCTRSLSRRERTSLLERADFSCIGR
SPRSADGTHSTVKAGFRLQKPANYGADPGLSSVTSKETPAAEGTPGDCPRAEAQRRGGAH
PGGSDSPHTLDGLTPQKGCRGDFTPLKGSNKEVKGWIDIKSTQKEDTPYKMVNSPEGEAQ
SDDKLNFVGNWVVENSTEAREELPRG

Eschrichti
DFFAGVLHLSFDDLCGSSGCGTQQRKLGGFLDESKSAACASSAVLKTSSVRPPASPGSLG
QLTLHQSPGSLSKGDTHGQRGAAGQAITPDAKPAWEPARGVFDGKRGLSPTSSVTKEPPR
GPSCPQSSSATGRRVSANSRPPPEERLRKRRSLGWPPAPRLQPENSPKCASRSAVETLGR
GESSYDDYFSPNNLKERDSESLLPGVQPPAGPALLHCTRSLSRRERTSLLERADFSCIGR
NPRSADGTYSTVKAGFRLQKPANYGADLGLSSVTSKETPAAKGTLGDCPRAEAQRRGDAR
LGGSDSPHTLDGLTPQKGHRGDFTPLKGSNKEMKGWIDIKSTQKEDTPSKMVNSPEGEAL
SDYKLNC-------------------

Balopt1_ph
DFFAGVLHLSFDDLCGSSGCGTQQRKLGRFLDESKSAACASSAVLKTSSVRPPASPGSLG
QLTLHQSPGSLSKGDTHGQRGAAGQAITPDAKPAWEPARGVFDGKRGLSPTSSVTKEPPR
GPSCPQSSSATGRRVSANSRPPPEERLRKRWSLGWPPAPRLQPENSPKCASRSAVETLGR
GESSYDDYFSPNNLKERDSESLLPGVQPPAGPALLHCTRSLSRRERTSLLERADFSCIGR
NPRSADGTYSTVKAGFRLQKPANYGADLGLSSVTSKETPAAEGTPGDCPRAEAQRRGDAR
PGGSDSPHTLDGLTPQKGRRGDFTPLKGSNKEMKGWIDIKSTQKEDTPSKMVNSPEGEAL
SDYKLNC-------------------

Physeter_c
DSFAGVLHLSFDDLCGSSGCGTQQRKLGGFLDESKSAACASSAVLKASSVRPPVSPGSLG
QLTLHQSPGSLSKGDTHPQRGTAGQVITPDAKPAREPAEGVFDEKRGLPPTPSVTKERPG
GPSCPQSPSATRRWVSASSSPPPGERLSRR----RPPAPRLPPQSSPQRASGSAVGTLGR
GGSPYDDYFSPDNLKERDSESLLPGVQPPASPALLHCMRSLSRRERTCLLERADFSCIGR
SPRSADGTDSTVKAVFRLQKPANYAADPGLSSVTSKETPAAEGTPGDCPRAEAQRRGGAR
PGGSGSPHTLDGLTPQKGRPGDFTPLKGSSEEVKGWIDVGSTQKEDTPSKMVNSPEGEAQ
SDDKLNFVGDCDLEKSAEAREELPRG

Megaptera_
DFFAGVLHLSFDDLCGSSGCGTQQRKLGGFLDESKSAAWASSAVLKTGSVRPPAPPGSLG
QLTLHQAPGSLSKGDTHGQRGAAGRAIAPDVKPAWEPARGVFDGKRGLSPTSSVTKEPPS
GPSCPQSSSATGRRVSANSRPPPEERLRKRRSLGWPPAPRLQPENSPKCASRSAVETLGR
GESSYDDYFSPNNLEERDSESLLPGVRPPAGPALLHCTRSLSRRERTSLLQRADFSCLGR
NPRSADGTYSTVKAGFRLQKPANYGADLGLSSVTSKETPAAEGTPGDCPRAEAQRRGDAR
PGRSDSPHTLDGLTPQKGQRGDFTPLKGSNKEMKGWIDIKSTQKEDTPSKMVNSPEGEAL
SDYKLNC-------------------

Ziphius_ca
------------------------------------------------------------
------------------------------------------------------?KEPPG
SPSCPQSPSATRRRVSAISSPPPEERLSKRRSLRWPPAPWLQRDNSLKCASSSAVETLGR
GESSYDDYFSPDNLKERNSESLLAGVQPPAGHALLHCMRSLSRRERTSLLGRADFSCIGR
SPRSADGTDSAVKAGFRLQKPANCRADPGLSSVTSKETPAAEGTPGDCPRAEAQRPGDAC
PGGSDSPHTLDGLTPQKGSRGDFTPLKGSNKEVKGWVDIKSTQKEDTPSKMVNSPEGEAQ
SDDKLNF-------------------

Mesoplodon
------------------------------------------------------------
------------------------------------------------------?KEPPG
```

```
SPSCPQSPSATRRRVSAISSPPPEERLSKRRSLRWPPAPRLQPENSPKCASSSTVETLGR
GESSYDDYFSPDNLKERNSESLLAGVQPPAGHALLHCMRSLSRRERTSLLGRADFSCIGR
SPRSADGTDSAVKAGFCLQKPANCRADPGLSSVTSKETPAAEGTPGDCHRAEAQRPGDAC
PGGSDSPHTLDGLTPQKGSRGDFTPLKGSNKEVKGWVDIKSTQKEDTPSKMVNSPEGEAQ
SDD?LNF-------------------

Kogia_sima
------------------------------------------------------------
------------------------------------------------------?KERPR
GPSCPQSPSATRRRVSVSSSPPPGERLSRRGPLGRPPAPRLLPESSPQHASGSAVGTVGR
GGSPYDDYFSPDNLKERDLESLLPGVQPPASPALFHCMRSLSRRERTCLLERADFSCIGR
SLRSADGTDSTVKAGFCLQKPANYAADPGLSSVTSKETPATEGTPGDCPRAEAQRRGGAR
PGGSGSPHTLDGLTPQKGHRGDFTPLKGSSKEVKGWIDGKSTQKEDTPSKMVNSPEGEAQ
SDDKLNF-------------------

Hippopotam
DSFADVLHASFGDLCGSSGCGTQERKLGGFLDESNSAVCASSPVLKTSSVCLPASPGSLG
QFTPQKSASTLFEGGVHGQADAAG??ITPDAKPSPGQAQGVFDEKRDL----SATKGPPG
GHSRPQSSSAKRQRTWETCSPPLGERLKKRRSRGRSPAPRLWLEKSPRRVSSPAVQALGL
EESSYDDYFSPDNLTERHPESLLPGAWPPAGPALLPCRRGLSRRERTSLLESADFSCIGR
SPRA?-----?AKTAFHLQKPANCGADVGLSSRASEEMPAAEETPADCPQAEAQRRGDAC
LGGGASTHVLDGFTPQKGCRGDFTPLEGSSKEMKGWMDVKGTQKEDASSKAVNSSEGEAH
SDYKLNFVGDCYVEKSTEEREELPRG
```