

**Center Initiated Proposal:**

**Sequencing and analysis of ~200 additional mammals to identify functional constraint in the human genome at roughly single base resolution.**

Kerstin Lindblad-Toh & Eric Lander,

Broad Institute of MIT and Harvard

(& the G10k mammalian community)

**Project Summary:** Comparative genomics offers a powerful tool to identify the functional elements in the human genome, while allowing the study of general genome evolution and the evolution of functional elements specifically. We propose to generate *de novo* genome assemblies from ~200 carefully selected mammals chosen to maximize the power to detect evolutionary constraint on functional elements. With close to 250 mammals total, the dataset should provide the ability to detect evolutionary constraint to the level of roughly single base pair resolution. By searching for evolutionary signatures and overlap with other genomic datasets such as RNA-Seq and chromatin states, we plan to assign a candidate function to the majority of constraint bases. In addition, this dense mammalian dataset will allow several types of analyses examining lineage specific evolution.

## **Background:**

Promising findings using mammalian comparative analysis: Comparative analysis of mammalian genome sequences has been proven to be a powerful approach for identifying functional elements. Our recent comparative analysis of 29 eutherian genomes demonstrates the power of mammalian comparative analysis for identifying and interpreting functional elements in the human genome based on their evolutionary signatures. In particular:

- Detection of constraint: We find that at least 5.5% of the human genome, defined at a resolution of 12 nucleotides, is under constraint. We identify more than 3.6 million constrained elements, covering 4.2% of the human genome, a dramatic increase in both quantity and resolution over what could be reliably detected previously. We find strong correlation between sites under constraint during mammalian evolution and those under constraint in the human population, with similarity between the patterns of constraint as well.

- Protein-coding genes: Although the vast majority of protein-coding sequence has already been identified, our analyses allow the prediction of an additional ~3,800 novel protein-coding exons. Analysis of unusual patterns of constraint within protein-coding sequence identifies four genes for which translational readthrough likely occurs. We identify ~10,000 regions of overlapping functional constraints within protein-coding exons, including splicing regulatory elements, A-to-I editing sites, miRNA target sites and potential enhancers.

- Regulatory elements: We find 220 families of conserved RNA structural elements; these are enriched in Untranslated Regions (UTRs) and likely play diverse regulatory roles. Promoters display three distinct patterns of conservation, each associated with different functional classes of genes. We identify 2.1 million conserved instances of regulatory motifs for 335 regulatory proteins, providing a median of 4,277 predicted target sites per regulator and together covering 8.5% of the constrained nucleotides. Of the 2.7 million constrained elements falling in intergenic and intronic regions away from known promoters, we can suggest a possible function for ~27% based on correlation with chromatin states, such as promoters and enhancers.

- Recent evolution: We detect regions of accelerated evolution, revealing sets of genes under distinct types of positive selection. We identify more than 280,000 instances of creation of novel conserved elements through the exaptation of transposable elements. We identify hundreds of regions under accelerated evolution, both in the human lineage and across primates, and find that they are associated with different functional classes of genes.

Power considerations: The ability to accurately pinpoint functional elements in the human genome by comparative analysis is related to the average number of neutral substitutions in orthologous sites in the evolutionary tree of the species used. The 29 eutherian genomes analyzed provide a maximum branch length of ~4.5 substitutions per sites in the evolutionary tree, permitting detection of constraint at the level of 12 bp. (When the tree has a total branch length of ~4.5 substitutions; the probability that a genomic sequence not under purifying selection will remain fixed across all the species is  $P_1 \sim 0.01$  for single bases and  $P_{12} < 10^{-22}$  for 12-mers. This was a considerable improvement in comparison to the earlier human-mouse-rat-dog (HMRD) analysis which provided a branch length of ~0.68 substitutions, with  $P_1 \sim 0.50$  and  $P_{12} \sim 10^{-3}$ .) To reach complete power of single base resolution, we estimate that close to 25 substitutions per site would be necessary ( $P_1 < 10^{-10}$ ), while the power for dinucleotides would be  $P_2 < 10^{-21}$ .

When selecting mammals from the mammalian tree one eventually reaches diminishing returns in terms of novel branch length added as the remaining species become more and more closely related. By picking roughly one species per family for an additional ~100 mammals, we estimate – by extrapolation from the branch lengths seen within the 29 mammals data set – that one could achieve roughly 15 substitutions per site. This would give a power of  $P_1 < 10^{-6}$  for single bases and  $P_2 < 10^{-15}$  for dinucleotides.

**Proposed Plan:**

We propose the *de novo* genome sequencing and assembly of ~200 additional mammals, but suggest that this project is performed in three phases. In this CIP we lay out the whole project, but based on the current ongoing grant renewal, we only request permission to perform phase I at this time point.

**Phase I:** We have identified 27 high priority eutherian mammals selected to generate the maximum branch-length within the eutherian tree (**Table S1**). The selected species are distributed across the tree based on branch length and organismal diversity within each of the major branches, including 2 primates (a separate proposal for primates is forthcoming from BCM shortly), 9 non-primate Euarchontoglires (including chinchilla, hamster, beaver, porcupine and various rodents), 10 Laurasiatherians (including a number of bats, hippopotamus, rhinocerus, and seal), 4 Afrotherians (including aardvark and manatee) and 2 Xenarthra (including anteater). Several of these species are also biomedical model organism, making their sequencing an additional high priority. This set is expected to generate a total branch length of  $\geq 10$  substitutions per site together with the 40 mammals currently sequenced or already in progress (**Table 1**). This set also includes three high-priority marsupial genomes. A high-quality tissue or DNA sample source has been identified for each of these species and a substantial fraction of the samples are currently at the Broad. All the remaining samples scheduled to arrive at Broad before June 1, 2011 (**Table S1**).

**Table 1: Summary of currently sequenced and phase I proposed species**

Clade/grouping	Current Public set		Proposed		Total	
	Species	Branch Length (Subs/site)	Species	Branch Length (Subs/site)	Species	Branch Length (Subs/site)
Euarchontoglires (Primates)	14	1.31	2	0.1	16	1.41
Euarchontoglires (Non-primate)	8	1.6	9	0.72	17	2.32
Laurasiatheria (non-Chiroptera)	11	1.8	6	0.6	17	2.4
Laurasiatheria (Chiroptera)	2	0.35	4	0.4	6	0.75
Afrotheria	3	0.68	4	0.8	7	1.48
Xenarthra	2	0.28	2	0.24	4	0.52
<b>EUTHERIAN TOTAL</b>	<b>40</b>	<b>6.02</b>	<b>27</b>	<b>3.16</b>	<b>67</b>	<b>9.38</b>
Marsupial	1	1.02	3	0.42	4	1.44
<b>TOTAL</b>	<b>41</b>	<b>7.04</b>	<b>30</b>	<b>3.58</b>	<b>71</b>	<b>10.82</b>

**Phase II:** We will generate *de novo* genome assemblies from ~ 100 additional mammals, carefully selected to complement those chosen in Phase I as well as ongoing efforts. On average, we will select one species per mammalian family. The selected species will include ~20 primates (including vervet, tamarin and howler monkey), ~30 non-primate Eutherians, ~20 Laurasiatherians, ~10 Afrotherians, ~5 Xenarthra, ~15 marsupials (including wombat and various kangaroos and possums). We expect the phase II to yield a total branch length of 15-20 substitutions per site together with previously sequenced genomes.

At the end of phase II the relative improvement from additional genome will be evaluated based on the achieved branch length and element detection power across Eutherians, Marsupials and important clades.

**Phase III:** If additional species are desirable to reach single base resolution or increased power for certain clades, we plan to sequence up to and 80 additional mammals yielding a total of 250 mammals all in all.

**Species selection:** Samples will be available through the Genome 10k project and organismal communities. The recent G10k meeting, March 16-18, 2011 in Santa Cruz served as an excellent venue for the mammalian communities to meet and develop a candidate list of mammals aiming to sample the diversity of the mammalian tree for maximum branch-length, while selecting species of biological and evolutionary relevance to the research community. An emphasis was also placed on sample procurement feasibility, with each selected species, requiring an available sample or a major lead on sample source. Table S1 contains three subsets of species: a) Phase I species securely identified and on their way to the Broad, b) a selection of species where a major source/lead for sample source is available and c) desirable samples where sample acquisition is a challenge but should be attempted.

**Sequencing strategy:** We plan to employ the standard library mix for large genomes: fragment libraries (50x-coverage), short jumping libraries (20x), long jumping libraries (10x), and Fosmid jumping libraries (2x). The assembly will be performed using ALLPATHS-LG and validated using a number of analysis and quality metrics. We expect contig N50 sizes to be > 50 kb in size and scaffold N50s to be ~10Mb in size.

The raw data will be deposited in the appropriate public data bases All assemblies will be made available on the Broad website and deposited in public databases as always.

**Analysis plan:** All genomes generated as part of this effort will be analyzed together in batches; initially all phase I genomes together with the existing 40 genomes will be analyzed. Due to the size of phase II and phase III efforts we plan to analyze the data reiteratively after roughly every 30 genomes added.

The analysis will follow largely the principle of the 29 mammals project; starting with alignments to the human genome followed by constraint detection using SiPhy omega and pi. Additional features including synonymous constraint elements, RNA structures and motif

binding sites will be detected. When possible a candidate function will be assigned to constraint based on overlap with experimental data such as RNA-Seq, chromatin states, transcription factor chip-Seq or promoter bashing experiments.

The evolution of elements within different clade will be examined using existing methodologies; such as those previously applied to find codon specific selection. HARs and PARs. In addition to using the human centered alignments, we will also generate and analyze a series of other alignments centered on important clades, utilizing a high-quality genome of an important organism as the center for these alignments. We plan to generate rodent focused alignments and constraint detection using mouse as a reference, dog will be used as a references for Laurasiatherian, elephant will be used to reference both Afrotheria and Xenarthra together and opossum will serve as the reference for marsupials.

**Ancillary data:**

As is typically our mandate we will generate RNA-Seq and diversity data for a subset of our projects where they are useful for annotation or organismal research communities.

**Timeline:** This project may be launched immediately, as sample acquisition has been ongoing for some time.

**Capacity plan:** The Phase I project will only constitute <10% of the Broad capacity for 2011.

TABLE S1 Selected species for phase I and additional potential species for phase II and III.

PHASE I

#	SPECIES	COMMON NAME	Clade	ORDER	FAMILY	GENOME SIZE (Gb)	Tissue samples ready WHEN ?	Sample source, comments
<b>Samples for sequencing in 2011</b>								
1	<i>Alouatta palliata</i>	Golden-mantled howling monkey	Euarchontoglires	Primates	Atelidae	3.6	5/27/11	Dr. Oliver Ryder, San Diego Zoo
2	<i>Lemur catta</i>	Ring-tailed lemur	Euarchontoglires	Primates	Lemuridae	~3.2	5/27/11	Dr. Oliver Ryder, San Diego Zoo
3	<i>Heterocephalus glaber</i>	Naked mole rat	Euarchontoglires	Rodentia	Bathyergidae	2.8	Now	At Broad
4	<i>Chinchilla laniger</i>	Chinchilla	Euarchontoglires	Rodentia	Chinchillidae	3.9	Now	At Broad
5	<i>Cricetulus griseus</i>	Chinese hamster	Euarchontoglires	Rodentia	Cricetidae	~3.2	Now	At Broad
6	<i>Mesocricetus auratus</i>	Syrian/Golden Hamster	Euarchontoglires	Rodentia	Cricetidae		5/20/11	Dr. Vera Gorubnova, U. Rochester
7	<i>Jaculus jaculus</i>	Lesser Egyptian jerboa	Euarchontoglires	Rodentia	Dipodidae (sf Dipodinae)	No data	5/1/2011	Dr. Kimberly Cooper, Harvard Medical School
8	<i>Erethizon dorsatum</i>	North American porcupine	Euarchontoglires	Rodentia	Erethizontidae	No data	5/20/11	Dr. Vera Gorubnova, U. Rochester
9	<i>Hystrix cristata</i>	Crested porcupine	Euarchontoglires	Rodentia	Hystriidae	2.8*	5/27/11	Dr. Oliver Ryder, San Diego Zoo
10	<i>Octodon degus</i>	Degu	Euarchontoglires	Rodentia	Octodontidae	4.2	6/1/2011	Dr. Theresa Lee, University of Michigan
11	<i>Pedetes capensis</i>	Springhare	Euarchontoglires	Rodentia	Pedetidae	No data	5/27/11	Dr. Oliver Ryder, San Diego Zoo
12	<i>Leptonychotes weddellii</i>	Weddell seal	Laurasiatheria	Carnivora	Phocidae	No data	Now	At Broad
13	<i>Hippopotamus amphibius</i>	Hippopotamus	Laurasiatheria	Cetartiodactyla	Hippopotamidae	No data	4/22/11	Dr. Oliver Ryder, San Diego Zoo
14	<i>Hyemoschus aquaticus</i>	Water chevrotain	Laurasiatheria	Cetartiodactyla	Tragulidae	No data	5/27/11	Dr. Oliver Ryder, San Diego Zoo
15	<i>Tadarida brasiliensis</i>	Mexican free-tailed bat	Laurasiatheria	Chiroptera	Molossidae	2.9	6/1/2011	Dr. Robert Baker, Texas Tech University
16	<i>Macrotus californicus</i>	California leaf-nosed bat	Laurasiatheria	Chiroptera	Phyllostomidae	No data	6/1/2011	Dr. Robert Baker, Texas Tech University
17	<i>Rhinolophus ferrumequinum</i>	Greater horseshoe bat	Laurasiatheria	Chiroptera	Rhinolophidae	No data	6/1/2011	Dr. Steve Rossiter or Dr. Walter Metzner
18	<i>Eptesicus fuscus</i>	Big brown bat	Laurasiatheria	Chiroptera	Vespertilionidae	~2.3	5/1/2011	Dr. Thomas Kunz, Boston University
19	<i>Condylura cristata</i>	Star nosed mole	Laurasiatheria	Eulipotyphla	Talpidae (sf Scalopiniae)	No data	6/1/2011	Dr. Ken Catania, Vanderbilt University
20	<i>Ceratotherium simum</i>	White rhinoceros	Laurasiatheria	Perissodactyla	Rhinocerotidae	No data	4/22/11	Dr. Oliver Ryder, San Diego Zoo
21	<i>Tapirus indicus</i>	Malayan tapir	Laurasiatheria	Perissodactyla	Tapiridae	3.7*	4/22/11	Dr. Oliver Ryder, San Diego Zoo
22	<i>Chrysochloris asiatica</i>	Cape golden mole	Afrotheria	Afrosoricida	Chrysochloridae	6.2	6/1/2011	Dr. Terrence J. Robinson, Stellenbosch University, South Africa
23	<i>Elephantulus rufescens</i>	Long-eared (Rufous) elephant shrew	Afrotheria	Macroscelidea	Macroscelididae	5.8*	6/1/2011	Dr. Terrence J. Robinson, Stellenbosch University, South Africa
24	<i>Trichechus manatus</i>	West Indian manatee	Afrotheria	Sirenia	Trichechidae	4.6	Now	At Broad
25	<i>Orycteropus afer</i>	Aardvark	Afrotheria	Tubulidentata	Orycterioidea	5.8	4/22/11	Dr. Oliver Ryder, San Diego Zoo
26	<i>Chaetophractus vellerosus</i>	Screaming hairy armadillo	Xenarthra	Xenarthra	Dasyproctidae (sf Euphractinae)	4.4	5/27/11	Dr. Oliver Ryder, San Diego Zoo
27	<i>Myrmecophaga tridactyla</i>	Giant anteater	Xenarthra	Xenarthra	Myrmecophagidae	No data	4/22/11	Dr. Oliver Ryder, San Diego Zoo
28	<i>Phascolarctos cinereus</i>	Koala	Marsupialia	Diprotodontia	Phascolarctidae	3.6	5/27/11	Dr. Oliver Ryder, San Diego Zoo
29	<i>Macropus giganteus</i>	Eastern grey kangaroo	Marsupialia	Diprotodontia	Macropodidae	~4.0	5/20/11	Dr. Marilyn Renfree, University of Melbourne, Australia
30	<i>Tarsipes rostratus</i>	Honey Possum (Noolbenger)	Marsupialia	Diprotodontia	Tarsipedidae	No data	5/20/11	Dr. Marilyn Renfree, University of Melbourne, Australia

PHASE II& III

<b>Confirmed Samples Sourced</b>								
31	<i>Aotus nancymae</i>	Owl monkey	Euarchontoglires	Primates	Aotidae	No data		Dr. Oliver Ryder, San Diego Zoo
32	<i>Cebus albifrons</i>	White-fronted capuchin	Euarchontoglires	Primates	Cebidae	3.9		Dr. Oliver Ryder, San Diego Zoo
33	<i>Cercopithecus sp.</i>	Guenons	Euarchontoglires	Primates	Cercopithecidae	~4.4		Dr. Oliver Ryder, San Diego Zoo
34	<i>Theropithecus gelada</i>	Gelada	Euarchontoglires	Primates	Cercopithecidae	No data		Dr. Oliver Ryder, San Diego Zoo
35	<i>Trachypithecus sp.</i>	Langurs, leaf monkeys	Euarchontoglires	Primates	Cercopithecidae	No data		Dr. Oliver Ryder, San Diego Zoo
36	<i>Mandrillus leucophaeus</i>	Drill	Euarchontoglires	Primates	Cercopithecidae	3.4*		Dr. Oliver Ryder, San Diego Zoo
37	<i>Pygathrix sp.</i>	Douc langurs	Euarchontoglires	Primates	Cercopithecidae	No data		Dr. Oliver Ryder, San Diego Zoo
38	<i>Erythrocebus patas</i>	Patas monkey	Euarchontoglires	Primates	Cercopithecidae	~3.6		Dr. Oliver Ryder, San Diego Zoo
39	<i>Presbytis sp.</i>	Surili monkeys	Euarchontoglires	Primates	Cercopithecidae	No data		Dr. Oliver Ryder, San Diego Zoo
40	<i>Colobus sp.</i>	Black and white colobus monkeys	Euarchontoglires	Primates	Cercopithecidae	3.5		Dr. Oliver Ryder, San Diego Zoo
41	<i>Nasalis larvatus</i>	Proboscis monkey	Euarchontoglires	Primates	Cercopithecidae (sf Colobinae)	4.2		Dr. Oliver Ryder, San Diego Zoo
42	<i>Cheirogaleus medius</i>	Fat-tailed dwarf lemur	Euarchontoglires	Primates	Cheirogaleidae	No data		Dr. Anne Yoder, Duke University Primate Center
43	<i>Daubentonia madagascariensis</i>	Aye-aye	Euarchontoglires	Primates	Daubentoniidae	No data		Dr. Anne Yoder, Duke University Primate Center
44	<i>Nycticebus coucang</i>	Slow loris	Euarchontoglires	Primates	Lorisidae (sf Perodicticinae)	3.5		Dr. Oliver Ryder, San Diego Zoo
45	<i>Pithecia pithecia</i>	White faced saki	Euarchontoglires	Primates	Pitheciidae	No data		Dr. Oliver Ryder, San Diego Zoo
46	<i>Calomyscus sp.</i>	Mouse-like hamster	Euarchontoglires	Rodentia	Calomyscidae	3.1*		Dr. Graphodatsky, Russian Academy of Sciences
47	<i>Castor canadensis</i>	North american beaver	Euarchontoglires	Rodentia	Castoridae	No data		Dr. Vera Gorubnova, U. Rochester
48	<i>Dicrostonyx torquatus</i>	Arctic lemming	Euarchontoglires	Rodentia	Cricetidae (sf Arvicolinae)	No data		Dr. Graphodatsky, Russian Academy of Sciences
49	<i>Ctenomys haigi</i>	Haig's tuco-tuco	Euarchontoglires	Rodentia	Ctenomyidae	~3.5*		Dr. Eileen Lacey, University of California, Berkeley
50	<i>Cuniculus paca</i>	Lowland paca	Euarchontoglires	Rodentia	Cuniculidae/Agoutidae	No data		Dr. Vera Gorubnova, U. Rochester
51	<i>Dasyprocta punctata</i>	Central American (punctate) agouti	Euarchontoglires	Rodentia	Dasyproctidae	No data		Dr. Oliver Ryder, San Diego Zoo
52	<i>Dinomys branickii</i>	Pacarana	Euarchontoglires	Rodentia	Dinomysidae	No data		Dr. Oliver Ryder, San Diego Zoo
53	<i>Allactaga major</i>	Great jerboa	Euarchontoglires	Rodentia	Dipodidae (sf Allactaginae)	No data		Dr. Graphodatsky, Russian Academy of Sciences
54	<i>Sciasta betulina</i>	Northern birch mouse	Euarchontoglires	Rodentia	Dipodidae (sf Allactaginae)	No data		Dr. Graphodatsky, Russian Academy of Sciences
55	<i>Thomomys bottae</i>	Botta's pocket gopher	Euarchontoglires	Rodentia	Geomysidae	~5		Dr. Oliver Ryder, San Diego Zoo
56	<i>Graphiurus murinus</i>	Woodland dormouse	Euarchontoglires	Rodentia	Gliridae (sf Graphiurinae) (aka Myc)	No data		Dr. Graphodatsky, Russian Academy of Sciences
57	<i>Hydrochoerus hydrochaeris</i>	Capybara (carpincho)	Euarchontoglires	Rodentia	Hydrochaeridae	No data		Dr. Oliver Ryder, San Diego Zoo
58	<i>Acomys cahirinus</i>	Eastern spiny mouse	Euarchontoglires	Rodentia	Muridae (sf Deomyinae)	2.6		Dr. Graphodatsky, Russian Academy of Sciences
59	<i>Meriones unguiculatus</i>	Mongolian jird (gerbil)	Euarchontoglires	Rodentia	Muridae (sf Gerbillinae)	3.6		Dr. Vera Gorubnova, U. Rochester
60	<i>Mycocastor coypus</i>	Coypu (Sp.,) Nutria (NA/Asia)	Euarchontoglires	Rodentia	Mycocastoridae	3.5		Dr. Vera Gorubnova, U. Rochester
61	<i>Cricetomys emini</i>	Giant rat	Euarchontoglires	Rodentia	Nesomyidae (sf Cricetomyinae)	No data		Dr. Warren Johnson, NCI
62	<i>Cavia aprae</i>	Wild cavy	Euarchontoglires	Rodentia	Petromuridae	No data		Dr. Gorubnova has tissues
63	<i>Tamias striatus</i>	Eastern chipmunk	Euarchontoglires	Rodentia	Sciuridae			Dr. Vera Gorubnova, U. Rochester
64	<i>Sciurus carolinensis</i>	Eastern gray squirrel	Euarchontoglires	Rodentia	Sciuridae (sf Sciurus)	5.1		Dr. Vera Gorubnova, U. Rochester
65	<i>Ondatra zibethicus</i>	Muskrat	Euarchontoglires	Rodentia	Sciuridae (sf Sciurus)			Dr. Vera Gorubnova, U. Rochester
66	<i>Myosapalax myosapalax</i>	Siberian zokor	Euarchontoglires	Rodentia	Spalacidae	No data		Dr. Graphodatsky, Russian Academy of Sciences
67	<i>Crocuta crocuta</i>	Spotted hyena	Laurasiatheria	Carnivora	Hyaenidae	No data		Dr. Oliver Ryder, San Diego Zoo
68	<i>Odobenus rosmarus</i>	Walrus	Laurasiatheria	Carnivora	Odobenidae	No data		Dr. Brent Stewart, Sea World, San Diego

69	<i>Phoca vitulina</i>	Harbor seal	Laurasiatheria	Carnivora	Phocidae	2.9*		Dr. Brent Stewart, Sea World, San Diego
70	<i>Balaena mysticetus</i>	Bowhead whale	Laurasiatheria	Cetartiodactyla	Balaenidae	No data		Dr. Per Palsboll, Stockholm University, Sweden
71	<i>Orcinus orca</i>	Killer whale	Laurasiatheria	Cetartiodactyla	Delphinidae			Dr. Brent Stewart, Sea World, San Diego
72	<i>Berardius bairdii</i>	Baird's beaked whale	Laurasiatheria	Cetartiodactyla	Ziphiidae	No data		Dr. Terrie Williams, UC Santa Cruz
73	<i>Saccopteryx bilineata</i>	Greater Sac-winged Bat	Laurasiatheria	Chiroptera	Emballonuridae	No data		Dr. Robert Baker, Texas Tech University
74	<i>Pteronotus davyi</i>	Davy's Naked-backed Bat	Laurasiatheria	Chiroptera	Mormoopidae	No data		Dr. Robert Baker, Texas Tech University
75	<i>Rousettus</i> sp.	Fruit bat	Laurasiatheria	Chiroptera	Pteropodidae	~2.2*		Dr. Emma Teeling, University College, Dublin
76	<i>Talpa altaica</i>	Siberian mole	Laurasiatheria	Eulipotyphla	Talpidae (sf Talpinae)	2.4*		Dr. Warren Johnson, NCI
77	<i>Amblysomus hottentotus</i>	Hottentot golden mole	Afrotheria	Afrosoricida	Chrysochloridae	No data		Dr. Terrence J. Robinson, Stellenbosch University, South Africa
78	<i>Heterohyrax brucei</i>	Hoggar/yellow-spotted hyrax	Afrotheria	Hyracoidea	Rhynchocoridae	No data	6/1/2011	Dr. Oliver Ryder, San Diego Zoo
79	<i>Dugong dugon</i>	Dugong	Afrotheria	Sirenia	Dugongidae	No data		Dr. Janet Lanyon, University of Queensland, Australia
80	<i>Petaurus breviceps</i>	Sugar glider	Marsupialia	Diprotodontia	Petauridae	No data		Dr. Clare Holleley, NCI
81	<i>Trichosurus vulpecula</i>	Silver-grey brushtail possum	Marsupialia	Diprotodontia	Phalangeridae	2.9		Dr. Marilyn Renfree, University of Melbourne, Australia
82	<i>Pseudocheirus peregrinus</i>	Common ring-tailed possum	Marsupialia	Diprotodontia	Phalangeridae	No data		Dr. Marilyn Renfree, University of Melbourne, Australia
83	<i>Vombatus ursinus</i>	Common wombat	Marsupialia	Diprotodontia	Vombatidae	3.9		Dr. Marilyn Renfree, University of Melbourne, Australia
84	<i>Dromiciops gliroides</i>	Monito del Monte	Marsupialia	Microbiotheria	Microbiotheriidae	No data		Dr. Warren Johnson (via Chilean colleagues)
85	<i>Isodon macrourus</i>	Northern brown bandicoot	Marsupialia	Peramelemorphia	Peramelidae	No data	6/1/2011	Dr. Rob Miller, University of New Mexico
<b>Species which samples are still being sought</b>								
86	<i>Leontopithecus rosalia</i>	Golden lion tamarin	Euarchontoglires	Primates	Callitrichidae	No data		Hector Sueanez, Universidade Federal do Rio de Janeiro
87	<i>Saguinus midas</i>	Cottontop tamarin	Euarchontoglires	Primates	Callitrichidae	No data		Hector Sueanez, Universidade Federal do Rio de Janeiro
88	<i>Cebus xanthosternus</i>	Buff-headed capuchin	Euarchontoglires	Primates	Cebidae	~3.7*		Hector Sueanez, Universidade Federal do Rio de Janeiro
89	<i>Mirza coquereli</i>	Coquerel's giant mouse lemur	Euarchontoglires	Primates	Cheirogaleidae	No data		Anne Yoder, Duke University
90	<i>Indri indri</i>	Indri	Euarchontoglires	Primates	Indridae	No data		Unlikely, but pursue
91	<i>Propithecus verreauxi</i>	Verreaux's sifaka	Euarchontoglires	Primates	Indridae	No data		Anne Yoder, Duke University
92	<i>Eulemur</i> sp.	Brown lemurs	Euarchontoglires	Primates	Lemuridae	~3.1		Anne Yoder, Duke University
93	<i>Lepilemur</i> sp.	Sportive Lemur	Euarchontoglires	Primates	Lepilemuridae	3.2*		Julie Horvath, Duke University
94	<i>Callicebus nigrifrons or personatus</i>	Titi monkey	Euarchontoglires	Primates	Pitheciidae	~2.2*		Hector Sueanez, Universidade Federal do Rio de Janeiro
95	<i>Abrocoma cinerea</i>	Ashy chinchilla rat	Euarchontoglires	Rodentia	Abrocomidae	3.5*		No obvious source yet
96	<i>Anomalurus beecrofti</i>	Beecroft's scaly-tailed squirrel	Euarchontoglires	Rodentia	Anomaluridae	No data		Suggested replacements from different family: red squirrel, chipmunk, woodchuck (all three species are available from Dr. Gorbunova as tissues and cell lines)
97	<i>Apodonta rufa</i>	Mountain beaver	Euarchontoglires	Rodentia	Apodontidae	No data		No obvious source, but highly unlikely
98	<i>Capromys pilorides</i>	Desmarest's hutia	Euarchontoglires	Rodentia	Capromyidae	No data		Al Roca, University of Illinois
99	<i>Dolichotis patagonum</i>	Patagonian mara	Euarchontoglires	Rodentia	Caviidae (sf Dolichotis)	3.6		Warren J. will follow up; Dr. Gorbunova has contact in Brazil and could possibly get it to US Antonio Augusto Ribeira; kept as pets in Brazil
100	<i>Ctenodactylus gundi</i>	Northern gundi	Euarchontoglires	Rodentia	Ctenodactylidae	No data		Dr. Richard in France has cell culture and might have tissues, Polina will contact her
101	<i>Laonastes aenigmamus</i>	Laonastes	Euarchontoglires	Rodentia	Diatomyidae	No data		Dr. Richard in France has cell culture and might have tissues, Polina will contact her
102	<i>Proechimys semispinosus</i>	Gorgona spiny rat	Euarchontoglires	Rodentia	Echimyidae	3.8		No obvious source, but highly unlikely, asked Bailey about contacts in Panama
103	<i>Dactylomys</i> sp.	Coro-coro	Euarchontoglires	Rodentia	Echimyidae (sf Dactylomyinae)	No data		The whole branch of myomorphs is underrepresented. We could suggest some myomorphs as replacements from Dr. Graphodatsky lab: <i>Apodemus peninsulae</i> , <i>Microtus species</i> , <i>Tcherskia triton</i>
104	<i>Eliomys quercinus</i>	Garden dormouse	Euarchontoglires	Rodentia	Gliridae (sf Leithiinae)(aka Myoxi)	No data		No obvious source, but highly unlikely
105	<i>Dendromys</i> sp.	Climbing mouse	Euarchontoglires	Rodentia	Nesomyidae (sf Dendromurinae)	No data		No obvious source, but highly unlikely
106	<i>Tympanoctomys barrerae</i>	Plains/Red Viscacha Rat	Euarchontoglires	Rodentia	Octodontidae	8.3		Warren Johnson will follow up
107	<i>Thryonomys swinderianus</i>	Greater cane rat	Euarchontoglires	Rodentia	Thryonomyidae	No data		Dr. Dobigny will try to get tissues from cane rat; also contacted Dr. Robinson about tissues of greater cane rat; if no tissue found suggest to replace with rat from totally different branch of Rodents - Ondatra (muskrat) Dr. Gorbunova has tissues and cell line;
108	<i>Ptilocercus lowii</i>	Pen-tailed tree shrew	Euarchontoglires	Scandentia	Ptilocercidae (Tupaiaidae sister far)	No data		No obvious source yet
109	<i>Mephitis mephitis</i>	Striped skunk	Laurasiatheria	Carnivora	Mephitidae	No data		Warren Johnson and Klaus Koepli will contact Jerry Drago for sample
110	<i>Meles meles</i>	Eurasian badger	Laurasiatheria	Carnivora	Mustelidae	No data		Emma Teeling will contact for sample
111	<i>Paguma larvata</i>	Masked palm civet	Laurasiatheria	Carnivora	Viverridae	No data		Warren Johnson and Klaus Koepli will contact Philippe Gaubert for sample
112	<i>Physeter catodon</i>	Sperm whale	Laurasiatheria	Cetartiodactyla	Physeteridae	No data		Ask BGI to drop? Need to confirm with BGI; also no obvious source yet
113	<i>Hylomys suillus</i>	Short-tailed gymnure	Laurasiatheria	Eulipotyphla	Erinaceidae (sf Galericinae)	No data		Polina Perelman will follow up
114	<i>Solenodon cubanus</i>	Solenodon	Laurasiatheria	Eulipotyphla	Solenodontidae	No data		No obvious source yet, but Warren Johnson will pursue
115	<i>Suncus murinus</i>	Asian house shrew	Laurasiatheria	Eulipotyphla	Soricidae (sf Crocidurinae)	No data		Polina Perelman will follow up
116	<i>Manis temminckii</i>	Ground pangolin	Laurasiatheria	Pholidota	Manidae	No data		Warren Johnson will follow up
117	<i>Limnogale mergulus</i>	Web-footed (aquatic) tenrec	Afrotheria	Afrosoricida	Tenrecidae (sf Oryzoricinae)	No data		Will contact Field Museum
118	<i>Potamogale velox</i>	Giant otter shrew	Afrotheria	Afrosoricida	Tenrecidae (sf Potamogalinae)	No data		No obvious source yet
119	<i>Bradypus tridactylus</i>	Pale-throated three-toed sloth	Xenarthra	Xenarthra	Bradypodidae	4.1*		Unlikely, but pursue
120	<i>Cyclopes didactylus</i>	Silky anteater	Xenarthra	Xenarthra	Cyclopedidae	No data		No obvious source yet
121	<i>Cabassous unicinctus</i>	Southern naked-tail armadillo	Xenarthra	Xenarthra	Dasypodidae (sf Tolypeutinae)	3.9*		Hector Sueanez will pursue sample
122	<i>Myrmecobius fasciatus</i>	Numbat	Marsupialia	Dasyuromorphia	Myrmecobiidae	No data		Jenny Graves will pursue sample
123	<i>Caluromys philander</i>	Bare-tailed wooly opossum	Marsupialia	Didelphimorphia	Didelphidae (sf Caluromyinae)	No data		Hector Sueanez will pursue sample
124	<i>Cercartetus caudatus</i>	Long-tailed pygmy possum	Marsupialia	Diprotodontia	Burramyidae	3.4*		Jenny Graves and Mark Springer will follow up
125	<i>Bettongia tropica</i>	Northern bettong	Marsupialia	Diprotodontia	Potoroidae	No data		Jenny Graves and Mark Springer will follow up
126	<i>Notoryctes typhlops</i>	Marsupial mole	Marsupialia	Notoryctemorphia	Notoryctidae	No data		Jenny Graves and Mark Springer will follow up
127	<i>Caenolestes fuliginosus</i>	Silky shrew opossum	Marsupialia	Paucituberculata	Caenolestidae	No data		Jenny Graves and Mark Springer will follow up