

Supplement to BAC Library Construction Proposal

[see original proposal below]

This document is a supplement to the proposal to construct 18 BAC libraries from a diverse set of mammals submitted February 10 2002. Below are more detailed justifications for creating BAC libraries from five key representatives of mammalian diversity and phylogeny: African Elephant, Nine-banded Armadillo, Bat, European common shrew, and lesser tree shrew. Please note the bat species listed in this supplement is different than that originally proposed, and was changed to maximize research accessibility.

1. African elephant, *Loxodonta africana*/ Order Proboscidea/major clade Afrotheria

One of the four major placental lineages, Afrotheria was the first to diverge from the other major groups of placental mammals. Thus, from a phylogenetic standpoint, afrotherians represent a unique group of modern placental mammals whose origins date back to the deepest node of Placentalia. Elephants represent the most highly visible and best studied species of any afrotherian, in fields ranging from ecology to behavior and genetics. We have recommended the African savanna elephant (*Loxodonta africana*) versus the Asian elephant due to a greater presence of African elephant research and data in the PubMed and GenBANK databases. In addition we are aware of genomic mapping resources currently being developed for the African elephant. A morphologically extreme placental mammal (i.e. extremely modified from the ancestral mammalian body plan), elephants exhibit large body size, long life span, and complex behavior. Would be useful for understanding genomic alterations affecting body growth and macroevolutionary developmental changes.

2. Nine-banded armadillo, *Dasypus novemcinctus*/Cingulata/ major clade Xenarthra

Provides a representative of the second major placental clade to diverge, Xenarthra. The nine-banded armadillo was chosen as an initial xenarthran representative because it is the best known and studied of all armadillos, and perhaps all xenarthrans in the fields of physiology, genetics, and ecology. In addition the NIAID has previously established breeding colonies of nine-banded armadillos for leprosy research. Furthermore, this species displays a number of interesting reproductive characteristics, including single-sex quadruplets and delayed implantation of embryos, that would provide interesting models for comparative genetic analysis.

3. Bat, *Rhinolophus ferrumequinum*/Chiroptera/ major clade Laurasiatheria

Bats represent a morphologically extreme placental mammal with adaptations for powered flight, including morphological differentiation of digits to form wings. All bat genomes examined to date appear to have reduced genome sizes (as little as 1.7-2 billion bp). This characteristic would facilitate identification of functionally important regulatory sequences in noncoding DNA. In addition this species is known to have long lifespans, in spite of small body size and high metabolic rates, which is in stark contrast to the observed longevity in equivalent rodent models. The species selected is housed at an animal colony at the University of California Los Angeles therefore providing a useful and accessible research resource, where it is being utilized as a model for mammalian neurophysiology.

- 4. European common shrew, *Sorex araneus*/Eulipotyphla/ major clade Laurasiatheria**
Provides representation of the ancestral mammalian morphology within the major clade Laurasiatheria. Furthermore, the European common shrew is the only eulipotyphlan to have been characterized at the genomic level, using both somatic cell hybrid and chromosome painting techniques. In addition it exhibits an interesting feature among mammals, in that chromosome number varies considerably across individuals of the same population. The presence of XYY males would provide a unique genomic perspective on the evolution of the Y chromosome in mammals. Given previous genomic characterization and interesting biological attributes, this species seems the most appropriate eulipotyphlan candidate for BAC library construction.
- 5. Lesser tree shrew, *Tupaia minor*/Scandentia/ major clade Euarchontoglires**
Tree shrews, as members of the sister order to primates, provide an important and unique evolutionary perspective to better understand genomic sequence generated from primate genomes. As the closest outgroup to primates, tree shrews offer a key point of comparison critical to fully investigate the evolution of shared and unique characteristics of primate genomes. Further, tree shrews exhibit very primitive anatomical and behavioral characteristics, which make them an important reference point for studying developmental/genomic changes among mammals as they depart from the ancestral mammalian body plan and life history.

James W. Thomas
National Human Genome Research Institute
National Institutes of Health

William J. Murphy
Laboratory of Genomic Diversity
National Cancer Institute

Eduardo Eizirik
Laboratory of Genomic Diversity
National Cancer Institute

Mark S. Springer
Professor and Chairperson,
Department of Biological Sciences
University of California at Riverside

Stephen J. O'Brien, Chief
Laboratory of Genomic Diversity
National Cancer Institute

Construction of 18 BAC libraries from a diverse set of mammals

The sequence of the human genome and the ongoing efforts to systematically sequence biomedical model organisms (mouse, rat and zebrafish) are providing unprecedented resources for advancing the ongoing research in these species. As important as the intrinsic information these projects provide for each individual species is the tremendous additive value gained from the inter-species comparison of these genomes. The recent ability to compare vertebrate genomes at a sequence level underscores how comparative genomics will play an increasingly critical and common role in future biomedical research.

Past and planned BAC library representation of mammals to date has focused primarily on the importance of individual species in terms of biomedical research and agricultural or economic impact, but not on their value as a point of comparison. As a result, while the list of mammalian BAC libraries includes many species (i.e., mouse, rat, opossum, cow, pig, horse, cat, rabbit, dog, platypus, human, chimpanzee, baboon, macaque, orangutan, lemur, vervet and spider monkeys), these species are, in fact, not truly representative of mammals as a group. In particular, only 2 of the 4 placental mammal super clades and 8 of the 26 mammalian orders are represented. To realize the depth of knowledge that can be gained from access to the full range of mammalian phenotypic and genomic diversity, we propose the following list of 18 mammalian species for BAC library construction. The species are ranked in order of priority, justified by the arguments below.

- | <u>Species</u> | <u>Order</u> | <u>Clade</u> |
|---|--------------------|-------------------|
| 6. African elephant, <i>Loxodonta Africana</i> | Proboscidea | Afrotheria |
| Provides a representative of one of the four major placental lineages, Afrotheria, which was the first to diverge from the other groups. A morphologically extreme placental mammal (i.e. extremely modified from the ancestral mammalian body plan), elephants exhibit large body size, long life span, and complex behavior. Would be useful for understanding genomic alterations affecting body growth and macroevolutionary developmental changes. | | |
| 7. 9-banded armadillo, <i>Dasyus novemcinctus</i> | Cingulata | Xenarthra |
| Provides a representative of the second major placental clade to diverge, Xenarthra. This species serves as the only animal model for leprosy, while also displaying a number of interesting reproductive characteristics, including single-sex quadruplets and delayed implantation of embryos. | | |

- 8. Tammar wallaby, *Macropus eugenii* Diprotodontia Australidelphian marsupial**
Like the American marsupial, *Monodelphis domestica*, for which BAC libraries are currently being constructed, the tammar wallaby is the most intensively studied Australasian marsupial on a genomic level. Successful breeding colonies exist for this species, which provides a model for studies of marsupial reproductive biology and investigations relating to mammalian sex determination, particularly as these traits contrast with placental mammal development and reproductive biology.
- 9. Brown bat, *Myotis lucifugus* Chiroptera Laurasiatheria**
Morphologically extreme placental mammal with adaptations for powered flight, including morphological differentiation of digits to form wings. Most bat genomes examined appear to have as little as 1.7-2 billion bp—this characteristic would facilitate identification of functionally important regulatory sequences in noncoding DNA. The species selected is known to have lifespans in excess of 30 years, in spite of small body size and high metabolic rates, which is in stark contrast to the observed longevity in equivalent rodent models.
- 10. Lesser tree shrew, *Tupaia minor* Scandentia Euarchontoglires**
As an immediate outgroup to primates, tree shrews provide an important evolutionary perspective to better understand genomic sequence generated from primate genomes. Tree shrews also exhibit very primitive mammalian anatomical and behavioral characteristics, which makes them an important reference point to study developmental/genomic changes among mammals.
- 11. Humpback whale, *Megaptera novaeangeliae* Cetartiodactyla Laurasiatheria**
Morphologically extreme placental mammal, exhibiting large body size, modified limbs into fins, and numerous other morphological adaptations to aquatic life. This species also has a very long life span (as great as 95 years). Would be useful for understanding genomic alterations affecting extreme morphological departure from the "standard" mammalian developmental program, particularly by comparison to close evolutionary relatives, cows and pigs, for which large scale genomic initiatives are currently in place.
- 12. Lesser hedgehog tenrec, *Echinops telfairi* Afrosoricida Afrotheria**
Another divergent member of Afrotheria, this species contrasts with elephants by exhibiting a number of ancestral mammalian morphological features, including low & variable body temperature, a cloaca, and undescended testicles.
- 13. Gray squirrel, *Sciurus carolinensis* Rodentia Euarchontoglires**
Provides molecular sequence data for a rodent phylogenetically distant from mouse and rat. Differs from these latter two species by having a conserved genomic organization and lower rates of nucleotide evolution, facilitating genetic comparisons with other mammals.
- 14. European common shrew, *Sorex araneus* Eulipotyphla Laurasiatheria**
Provides representation of the ancestral mammalian morphology within the major clade Laurasiatheria. Exhibits a unique feature among mammals, in that chromosome number varies in individuals of *Sorex araneus* ($2n=21-27$ males, 20-25

females). The presence of XYY males would provide a unique genomic perspective on the evolution of Y chromosome in mammals.

- 15. West Indian manatee, *Trichechus manatus* Sirenia Afrotheria**
Morphologically extreme placental mammal, exhibiting large body size, modified limbs into fins, and numerous other morphological adaptations to aquatic life. Would be useful for understanding genomic alterations affecting extreme morphological departure from the "standard" mammalian developmental program, and provides a comparison for the convergent evolution of aquatic adaptations relative to whales.
- 16. Two-toed sloth, *Choloepus didactylus* Pilosa Xenarthra**
A second representative from the major placental clade Xenarthra, sloths maintain low but variable body temperatures, low metabolic rates and variation in the number of cervical vertebrae (which is seven in virtually every other mammal).
- 17. Echidna, *Tachyglossus aculeatus* Monotremata Monotremata**
Along with platypus, echidnas are only other family in this most divergent (relative to humans) order of mammals. While similar to platypus in some regards, like the embryonic development unique to monotremes that has aspects similar to reptiles and marsupials, echidnas are morphologically and phylogenetically distinct from platypus and thus would provide another unique point of comparison.

The following species were selected as representatives of other placental orders, or in the case of the dasyruoid marsupial and flying fox, other distinct phylogenetic lineages that exhibit unique morphological, physiological and genomic characteristics, providing unique points of molecular sequence comparison.

- | | | |
|--|----------------|-------------------------------|
| 18. Chinese pangolin, <i>Manis pentadactyla</i> | Pholidota | Laurasiatheria |
| 19. Elephant shrew, <i>Elephantulus rufescens</i> | Macroscelidea | Afrotheria |
| 20. Rock hyrax, <i>Procavia capensis</i> | Hyracoidea | Afrotheria |
| 21. Dasyurid marsupial, <i>Sminthopsis crassicaudata</i> | Dasyuromorphia | Australidelphian
marsupial |
| 22. Flying fox, <i>Pteropus giganteus</i> | Chiroptera | Laurasiatheria |
| 23. Aardvark, <i>Orycteropus afer</i> | Tubulidentata | Afrotheria |

The rationale for choosing these species was to select a group of mammals such that each species represents a non-overlapping subset of mammalian phylogenetic and morphological diversity. In essence, this list represents an optimized set of mammals that are at unique positions in the mammalian radiation and therefore will provide unique and informative points of comparison. Within this list, the species are first prioritized to fill the major the gaps in mammalian phylogeny, that is at the major clade level, and successive species selected for representation of mammalian orders, while emphasizing lineages with unique morphological, physiological or genomic characteristics. There is

redundancy in this list for the orders of Rodentia, Monotremata, Cetartiodactyla and Chiroptera relative to available and proposed libraries. However, these choices were considered not to be duplicative but beneficial in terms of adequately representing the diversity (number of species, morphology and phylogeny) within those orders.

Since the primary criteria for choosing these species was to maximize diversity and species representation, not all are necessarily connected with a specific research community or other genomic resources. However, we believe that the biomedical research community at large would benefit from the wealth of comparative information that would be catalogued in these libraries. In particular, we believe these libraries will provide ready and affordable access to cloned DNA that contains specific genes or regions of interest. The actual clones themselves can then provide templates and resources necessary for both physical and cytogenetic comparative mapping, positional cloning, small-scale sequencing, and direct experimental assessment of gene function in vitro and in vivo.

We request that all libraries be constructed with a depth of clone coverage equal to at least 10X genome equivalents (the genome size of all mammals listed, except bats, are expected to be $\sim 3 \times 10^9$ bp), have an average clone insert size of ~ 150 kb and be made using standard BAC vectors, such as pTARBAC2.1 or pBACe3.6. These specifications are the standard for previous mammalian BAC libraries and therefore should be readily achieved in each of the mammals we propose. Libraries with these specifications will provide a comprehensive resource for the isolation of any given gene or region of the genome, allow the cloning of nearly any gene on a single BAC, facilitate the selection of a single or small set of clones for sequencing by using restriction enzyme fingerprints to assess clone integrity and overlaps, and be an accessible source of probes for fluorescent in situ hybridization (FISH).

All of the species were selected with sample availability and research accessibility in mind. In most cases representative species can be obtained from zoos or wildlife parks while the other species are readily available in the wild. The authors of this report have in their possession samples from or have professional contacts for access to these mammals, should NHGRI see these species fit for BAC library construction.

The priority in which we favor construction of BAC libraries from the proposed species follows the order in which the species are listed. Given the immediate utility of these libraries as representative mammals, we request that the first 5-10 of these libraries be available within the next 12 months, with the remaining BAC libraries made available within 24 months. We have not requested support for the construction of these or other mammalian BAC libraries from any other sources and to the best of our knowledge, no BAC libraries exist or are planned for the species listed.

This proposal for the construction of BAC libraries is directed at increasing the diversity and comparative value of future BAC library resources. Toward that aim, we have requested BAC libraries be constructed from 18 representative species from 17 mammalian orders. In addition, we would be absolutely supportive of other proposals with similar aims, especially in regards to maximizing representation of mammals. Combined with existing and planned BAC library resources, the set of species proposed here would establish access for the biomedical community to the genomes responsible for the tremendous diversity represented in 22 mammalian orders.

James W. Thomas
National Human Genome Research Institute
National Institutes of Health

William J. Murphy
Laboratory of Genomic Diversity
National Cancer Institute

Eduardo Eizirik
Laboratory of Genomic Diversity
National Cancer Institute

Mark S. Springer
Professor and Chairperson,
Department of Biological Sciences
University of California at Riverside

Stephen J. O'Brien, Chief
Laboratory of Genomic Diversity
National Cancer Institute