White paper proposal for sequencing the genome of *Peromyscus*

Rachel O'Neill[1], Gabor Szalai[2], Richard Gibbs[3], George Weinstock [3], Paul Vrana[4], Julie Glenn[2], Michael Dewey[2] and Michael Felder[2]

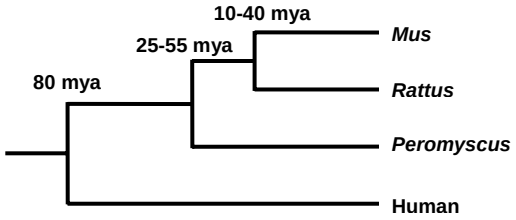
I. INTRODUCTION

Mice of the genus *Peromyscus* provide a rare opportunity to combine laboratory studies with natural genetic variation found in present-day wild populations. Research on this genus has been widespread across so many disciplines that the genus has aptly been referred to as "The *Drosophila* of North American Mammalogy" (Musser and Carleton 1993).

Range of *Peromyscus leucopus* Range of *Peromyscus maniculatus*

The genus *Peromyscus* contains the two most abundant native North American mammals (the deer mouse, *P. maniculatus*, and the white-footed mouse, *P. leucopus*) as well as one of the most endangered mammals in the United States (*P. polionotus trissyllepsis*). Members of the genus are found from Alaska to Central America and from the Atlantic to the Pacific. They occur in a wide range of habitats including sea-level wetlands and beaches, forests, prairies, deserts, and mountains of elevation up to 14,000 ft. Variation is seen in morphology, physiology, behavior, growth, coat-color, diet and habitat. Many of these differences exist within the *P. maniculatus* species complex, whose members exhibit sufficient interfertility to make genetic analysis possible.

Though superficially resembling laboratory mice (*Mus domesticus*) and rats (*Rattus norvegicus*), deer mice are not closely related to either species. Instead, *Mus* and *Rattus* share a much more recent common ancestor with each other than with *Peromyscus*. In phylogenetic terms *Peromyscus* not only aids researchers in understanding the *Mus/Rattus* lineage by serving as an outgroup, but also provides a species intermediate between the two major rodent genetic models and humans. There are other species in the genome-sequencing pipeline that are phylogenetically positioned between muroid rodents and human, such as the squirrel and guinea pig. However, none offer the capacity to study phenotypes produced through natural genetic variation and none provide numerous laboratory models for the study of human diseases, infectious disease vectors, and ecological niche adaptation.

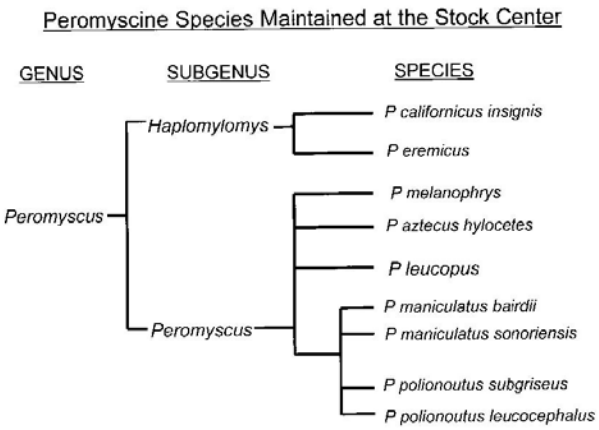


One of the major advantages of research with *Peromyscus* is their ready adaptability to colony conditions. Deer mice manifesting traits of interest in the field are easily moved into the laboratory where the trait can be studied under controlled conditions (reviewed in Joyner et al. 1998). Similarly, laboratory-based studies can easily be explored under field conditions. In recognition of this fact, and to facilitate laboratory studies with *Peromyscus,* the *Peromyscus* Genetic Stock Center (PGSC) was established at the University of South Carolina in 1985 with funding originally derived from NSF, and later, with additional funding from NIH.

[1]University of Connecticut, Molecular & Cell Biology, BH 323A, 354 Mansfield Road, Unit 2131, Storrs, CT 06269-2131; [2]*Peromyscus* Genetic Stock Center, Department of Biological Sciences, University of South Carolina, 715 Sumter St., Columbia, SC 29208; [3]Baylor College of Medicine. Human Genome Sequencing Center, Houston, TX 77030; [4]College of Medicine, University of California Irvine, Irvine, CA 92799

corresponding author: Michael Felder, Director, *Peromyscus* Genetic Stock Center; email: felder@biol.sc.edu, tel: (803)777-5135

Currently the PGSC maintains about 4000 animals representing seven species and two subspecies of *Peromyscus* along with 19 pure-breeding lines of single natural mutations affecting behavior, physiology, and coat color, as well as mutations that lead to diseases such as movement disorders, epilepsy and cancer. These outbred animals represent a more appropriate model for human pathology than inbred laboratory mice with induced mutations. In addition, while the typical lifespan of *Mus* is only two years, the lifespan of one Peromyscine, *P. leucopus*, is about eight years, putting these species in the center of aging research. Since Peromyscines are the most abundant mammals in North America, they are also naturally present at every toxic waste site and are primary reservoirs for two emerging infectious diseases, hantavirus pulmonary syndrome (HPS) and Lyme disease.



The main objective of this proposal is to have a high-quality, long-range assembly of the *Peromyscus maniculatus* genome. This species represents over 30% of the current PGSC stock requests and was the species used for the *Mus*/*Peromyscus* synteny map, thus providing useful anchors in scaffold assembly. Additionally, we propose to generate low coverage of three other *Peromyscus* species based on their broad phylogenetic representation within the genus, their usefulness in biomedical research, and the availability of genetic tools developed for these species. These three include: *P. californicus* (this species is extensively used in behavior studies and is a model for type 2 diabetes), *P. leucopus* (this species exhibits an extremely long life-span suitable for aging studies and is a model for tumor metastasis) and *P. polionotus* (a species that will aide in SNP discovery as our recently completed linkage map is based on a backcross between *P. maniculatus* and *P. polionotus*). Development of a SNP map through low-level sequencing of some of these species will aide in the identification of specific loci responsible for the phenotypes described below.

II. SPECIFIC BIOLOGICAL RATIONALES FOR THE *PEROMYSCUS* SEQUENCE DATA

II.1. Biomedical applications of *Peromyscus*

The Stock Center is leading a multi-institutional effort to develop resources to enable detailed genetic analyses of novel human disease models. Some of the wild-type and mutant stocks housed at the PGSC are used as models for understanding human physiology and disease. Several of these are outlined below.

II.1.1 Movement Disorders – Boggler. These animals exhibit a pronounced tremor and staggering, awkward gait. The tremor is expressed whenever the animal is awake or at rest. The awkward gait is due to uncoordinated movements of the hind limbs. The tremor involves the whole body, but head movements are more noticeable. Boggler deer mice are less adept at climbing than normal mice, but swimming ability is unimpaired. The trait cannot be recognized until after three months of age, and expression occasionally is delayed until the animal is more than a year old. These are typical characteristics of human patients suffering from Parkinson's disease. The PGSC is now studying the brain morphology of mutant animals to validate the use of these animals as models for Parkinson's disease.

II.1.2 Movement Disorders - Juvenile Ataxia. Ataxic behavior is observed in animals from about 15 days of age until 35 - 45 days. These mice are unable to right themselves when placed upon their backs, exhibit an ataxic gait and are unable to climb up their tails when held by the tail tip. Falling behavior occurs equally to the right or left. Swimming ability is unaffected. What is particularly

interesting about these mice is that they overcome their neurological difficulties as they age, making this pathology not only unusual but potentially informative of regenerative neurological activity. Currently the PGSC is investigating whether these animals can serve as a model for coenzyme-Q10 deficiency related ataxia.

II.1.3 Autism. Deer mice exhibit repetitive behaviors (i.e., hindlimb jumping, backward somersaulting) that occur at a high rate, persist across the life of the animal and appear relatively early in development. These behaviors are associated with standard laboratory housing and do not require isolation housing, specific cues or contexts, or a pharmacological agent for induction. These features, plus considerable heterogeneity in individual levels of expression, modulation by early experience, mediation by cortical-basal ganglia circuitry, and association with cognitive rigidity make *Peromyscus* an appealing model of restricted, repetitive behavior in autism (Lewis et al. 2007).

II.1.4 Epilepsy. The first rodent susceptible to audiogenic seizures (AGS) was reported for *Peromyscus* (Dice 1935) and was attributed to a single recessive unit, denoted as "*epl*" (Watson 1939). These animals differ from the *Mus* AGS models in several significant ways. First, they are sensitive to a low intensity noise (105 dB) as compared to the DBA/2 mouse (120 dB). Second, while AGS susceptibility seems to be a juvenile condition in the *Mus* models, *Peromyscus* were tested at the average age of 2.64 months and some retained the trait up to 11 months (Watson 1939). Additionally, after recovering from a seizure *epl*/*epl Peromyscus* displayed aggression (Chance and Yaxley 1949), a characteristic behavior of epilepsy patients. To map the *Peromyscus epl* gene the PGSC crossed the homozygous animals to *P. maniculatus sonoriensis,* a highly heterogeneous stock. When analyzing F2 intercross animals, we noted a different, previously undescribed AGS phenotype. Weanlings displayed wild-running followed by an "absence" period of about 30 seconds, then a second wild-running followed by clonus and tonus. The first two phases are lost at about 60 days of age. Further characterization of the *Peromyscus* epilepsy model will give new insights into the genetic, molecular, structural and behavioral aspects of this disease and perhaps initiate use in the pharmaceutical industry (as all antiepileptic compounds are tested on AGS susceptible mice or rats).

II.1.5. Cancer. A frequent occurrence of Harderian adenocarcinomas has been observed in two inbred *P. leucopus* strains (Parnell et al. 2005). The periocular tumors appeared to be highly malignant, with elevated mitotic indices, marked anaplasia, and metastases to regional lymph nodes and lungs. The tumors were readily transplantable to other animals of the same line. These animals may serve as the only spontaneous metastasis model in laboratory mammals.

II.1.6 Alopecia. Two recessive *P. maniculatus* mutants have been described for hair loss. Mature deer mice homozygous for the *hairless-1* mutation have no hair except vibrissae and sometimes scattered body hairs. The trait can be recognized when animals begin to shed the first coat, about 2-3 weeks of age. The claws tend to elongate and assume a spiral twist. The skin becomes somewhat corrugated in many instances (Sumner 1924). Adult deer mice homozygous for the *hairless-2* gene are virtually nude except for vibrissae and sensitive hairs, which exist as short stumps. Some individuals have normal shaped, but short vibrissae. The trait can be detected in newborn deer mice by the twisted appearance of the vibrissae. Animals begin to lose hair about twelve days of age, following a period during which the coat has a distinctive sheen. Juveniles have dark pigmented bodies with pink faces. Transient coats appear at subadult and adult molts. The deer mice are most readily distinguished from *hairless-1* by the absence of elongate claws. Scanning electromicroscopy reveals many hairs that are abnormally twisted or bilobar, and that are fragile and readily break off at the skin surface. Ears of homozygous *hairless-2* deer mice are abnormally thin and there is a general reduction in body weight (Egoscue 1962). Crossing the two mutant lines revealed that *hairless-1* and *hairless-2* are two separate genes and not alleles of the same locus.

II.1.7. Diabetes. California mice (*P. californicus*) develop diet-related type 2 diabetes mellitus (T2DM) when fed a diet containing 25.8% kcal from fat. Sections of pancreas from diabetic and prediabetic mice had pathologic changes consistent with T2DM. After six months of feeding a low-fat diet, mice

were normoglycemic, normotriglyceridemic, and normocholesterolemic. Some mice remained hyperinsulinemic (Krugner-Higby et al. 2000, 2006). Thus, California mice may be a useful animal model for human T2DM.

Another nascent model for the genetics of male T2DM susceptibility utilizes the *P. maniculatus bairdii* (BW) and *P. polionotus subgriseus* (PO) stocks at the PGSC. The males of these species differ greatly in their ability to regulate blood glucose levels after fasting via a standard glucose tolerance test (GTT). Blood glucose levels in BW males remained elevated and indicative of T2DM, while blood glucose levels in PO males stabilized to pre-injection levels. Consomic BW animals bred at the PGSC which carry only a PO Y chromosome (BW Y^{PO}) respond like the PO males, suggesting that these differences are mediated by differences in Y chromosome sequence. Intriguingly, an unmapped human locus on the Y chromosome affects blood pressure via a stress response.

II.1.8. Alcohol metabolism. Naturally occurring genetic variation within *Peromyscus* has been useful in elucidating mechanisms of metabolism and enzyme multiplicity. The primary site of alcohol metabolism in mammals is the liver. The observation that isolated microsomal fractions would metabolize alcohol (MEOS, microsomal ethanol oxidation system) was met with extreme skepticism because of possible contamination with alcohol dehydrogenase and/or catalase (Lieber and DeCarli 1970; Isselbacher and Carter 1970; Teschke et al. 1972). Efforts were made to isolate microsomes free of ADH and catalase contamination, but ADH class I is expressed at high levels in liver. The discovery of an ADH-negative allelic variant (Adh^n) in *Peromyscus* and production of homozygous animals lacking class I ADH activity (Burnett and Felder 1978a, 1978b, 1980) led to the observation that at least 50% of alcohol metabolism is mediated through pathways other than the ADH pathway. ADH-negative *Peromyscus* have also been used to determine the role of the enzyme in the "swift increase in alcohol metabolism" (Glassman et al. 1985) and the dependence of ADH on allyl alcohol hepatotoxicity (Belinsky et al. 1985). The ADH-negative phenotype was ultimately determined to be a gene deletion (Zheng et al. 1993), and the presence of the negative phenotype allowed molecular cloning of another class of ADH that has been found to occur within the ADH complex on chromosome 3 in the mouse (Szalai et al. 2002). Further research utilizing this unique ADH-negative mutant and additional enzyme mutants (i.e., A-polypeptide; Cattanach and Perez, 1969) will continue to inform biomedical research.

II.1.9. Ageing. In mammals, Sacher (1978) found a functional relation between longevity and some anatomical and physiological dimensions (body size, body weight, brain size, etc). As a notable exception, *P. leucopus* lives 2.5 times as long as *Mus musculus*, although both have similar body sizes. A recent study showed that the increased lifespan potential in *P. leucopus* is associated with a decreased generation of cellular reactive oxygen species and increased oxidative stress resistance, which is in accordance with the prediction of the oxidative stress hypothesis of aging (Csiszar et al. 2007 submitted). Recently the PGSC has seen an increased demand for these animals, as they are becoming a major model for ageing.

II.1.10. Public Health. Due to their broad geographic distribution and abundance in the United States, Peromyscines, particularly *P. maniculatus* and *P. leucopus,* have attracted attention and public health concern for their potential as reservoirs of infectious disease organisms. This is the case for both pathogens currently affecting human heath as well as emerging pathogens. Just in the past decade Peromyscines were found to be primary reservoirs for two infectious diseases, hantavirus pulmonary syndrome (HPS) and Lyme disease. *P. maniculatus* is the primary carrier of the Sin Nombre strain of Hantavirus responsible for the alarming 1993 outbreak of HPS. Though cases of HPS are not numerous, the major concern is its high rate of mortality; over 40% of affected individuals die after contracting HPS.

Peromyscus species also carry a number of other tick-borne diseases. Lyme disease was first recognized during the late 1970's in a cluster of cases in Lyme, Connecticut as a persistent malady characterized by fever and aches which progresses to neuralgia and arthritic symptoms. Though

seldom lethal, Lyme disease presents a significant public health concern; in 1999 over 16,000 cases were reported and from 2003-2005 the CDC recorded almost 65,000 cases from 46 states and the District of Columbia. Small mammals, particularly *P. leucopus*, constitute the major reservoir for the spirochete that leads to Lyme disease infection.

Other diseases for which deer mice are likely to play an epidemiological role include various forms of ehrlichiosis caused by the rickettsias *Ehrlichia chaffeensis, E. ewingii*, and one yet to be classified, as well as another disease, babesiosis, a malaria-like disease caused by *Babesia microti* a protozoan that infects erythrocytes. Evidence also suggests that deer mice are a reservoir for vesicular stomatitis virus, an agent that substantially affects the cattle and equine industries. Currently, studies are underway to determine whether *Peromyscus* may have a role in the epidemiology of West Nile Virus.

Since *Peromyscus* is a vector for many human pathogens, dissecting the various components of immune function in this group of species is of importance in understanding the lifecycle and etiology of these diseases. This is particularly the case in studies of the interaction of infectious disease organisms, such as Hantavirus or the Lyme disease spirochetes, with their Peromyscine hosts. Similarly, immunological endpoints would be enormously informative in physiological studies of animals experiencing environmental and other stresses such as low temperatures, high altitudes, arid environments, variations in photoperiods and various reproductive strategies (see below). Unfortunately, such studies are not yet possible in *Peromyscus* primarily due to the lack of immunological reagents. Because of the evolutionary gap separating *Mus* and *Peromyscus* most of the reagents developed for research in *Mus* fail to work in *Peromyscus*, be they antibodies recognizing self surface antigens on T cells, B cells and their subpopulations or antibodies recognizing the various cytokines. The availability of genome-wide *Peromyscus* sequences would quickly advance the development of *Peromyscus*-specific reagents and would spur the movement of the species into the forefront of infectious disease and physiological research.

II.1.11. Toxicology. Due to their nearly ubiquitous distribution throughout North America, *Peromyscus* are the most obvious choice for a naturally occurring mammalian model for toxicological research. *Peromyscus* species have been found on PAH-contaminated sites in northwest Washington state, pesticide-treated apple orchards in Pennsylvania, mercury-contaminated stream banks in California and PCB-contaminated dumpsites in South Carolina. These species provide a wealth of information on chemical uptake and toxicological response due to their relatively small home range and omnivorous dietary habits. Several biomarkers have been used successfully with *Peromyscus* and have documented cholinesterase depression with organophosphate pesticides, cytochrome P-450 induction by PAHs (Dickerson et al. 1994) and chlorinated aromatic compounds, porphyrin profiles, aminolevulinic acid dehydratase (ALAD), glutathione status, the Jerne plaque forming assay, B and T cell blastogenesis assays and biogenic amine profiles. All of these assays measure functional enzymes or systems responsive to specific contaminant exposures.

Heavy metals have also been shown to affect DNA methylation in plants, however, relatively little research has been done on their effects on mammals. DNA methylation is known to be important in maintaining pluripotency in embryonic stem cells (Ikegami et al. 2007; also see section II.4 Genomic Imprinting and Hybrid Dysgenesis below). At the same time, numerous tumors have been shown to exhibit aberrant DNA methylation (Jaenisch and Bird 2003). *Peromyscus* naturally occurring on several sites contaminated with heavy metals have already shown karyotypic abnormalities and morphological abnormalities (Husby et al. 1999). Thus, how these contaminants affect embryonic stem cells is an important emerging field of research.

II.1.12 Genomic Imprinting and Placentation. Comparative studies of placental development have been essential to our understanding of the ontogeny and physiology of this organ. Yet our understanding of the underlying cause of placental dysmorphies leading to pathologies such as intrauterine growth restriction (IUGR), pre-eclampsia and spontaneous abortion remains limited. As in many avenues of human disease research our advances in understanding these pathologies are

limited by the availability of informative animal models. *Peromyscus* hybrids have been well studied as examples of hybrid dysgenesis with particular involvement of the placenta in hybrid fetal anomalies. An especially fruitful model involves species seemingly at the initial stages of reproductive isolation. Crosses between the *P. maniculatus bairdii* (BW) and *P. polionotus subgriseus* (PO) stocks produce parent-of-origin effects on growth and development. BW females mated to PO males (bw x po) produce growth-retarded offspring, which are otherwise healthy and fertile (Dawson 1965). In contrast, PO females mated to BW males (PO x BW) produce overgrown and severely defective offspring and half of all these breedings end in complete death of the litter by mid-gestation. Those that survive to the latter part of gestation display numerous developmental defects, many reminiscent of multiple human diseases such as Gestational Trophoblast Disease and Silver-Russell Syndrome (Duselis and Vrana 2007). The placenta is particularly affected in these crosses. The bw x po placentas average half the weight of those from the parental strains while the PO x BW placentas weigh on average ca. three-fold that of the parental strains, appear very disorganized, and a portion of offspring lack obvious embryonic structures, analogous to human molar pregnancies. The most severe hybrid phenotypes correlate with an apparent genetic interaction between at least two imprinted regions as well as perturbations of imprinted gene expression which are independent of this interaction (Vrana et al. 2000). Fine-mapping of these two loci placed one in the same region responsible for a human placental growth abnormality (Loschiavo et al. 2007). Utilizing microarray technology scientists have been able to obtain a first glance at disrupted genetic pathways in hypertrophic and hypotrophic hybrid placentas. Several genes revealed as having altered expression patterns in the abnormal hybrid placentas fall into gene networks known to be critical in placental development and stem cell differentiation (Duselis et al. in press). Others represent novel pathways including several genes with as yet unknown functions. This model provides a valuable resource for furthering our understanding of placental development and the genetic causes of placental defects.

X chromosome inactivation is also altered in the hybrids. As in house mice and bovine placentas, the paternal X is silenced in extra-embryonic tissues. In somatic tissues, however, inactivation is severely skewed such that the BW X chromosome is preferentially expressed in both hybrid types. Genetic tests show that the skewing is mediated by the X inactivation center, where the *Xist* gene lies (Vrana et al. 2000). Thus, *Peromyscus* is an exciting model in which to study genomic imprinting, X-inactivation, and placental and pregnancy-related human diseases.

II.2. Behavior. *Peromyscus* has been the subject of numerous behavioral studies. *Peromyscus maniculatus* is notoriously prone to the development of stereotypical behavior (chronic repetitive movements) and is being studied as a model for repetitive movement disorders in humans (Presti et al. 2002). To understand the genetics underlying this predisposition, lines of *P. maniculatus* manifesting high vs low susceptibility to development of stereotypy are being developed (MH Lewis, University of Florida, personal communication).

With regard to partner fidelity, polygamy is the norm for most rodents (and mammals), but among Peromyscines two species, *P. polionotus* and *P. californicus,* provide models for monogamy (Gubernick and Teferi 2000; Ryan and Altmann 2001). Of particular interest to geneticists are the interfertile species *P. polionotus* (monogamous) and *P. maniculatus* (polygamous) which offer the opportunity to identify genes underlying differences in partner fidelity (Young and Hoekstra in prep). Initial studies in this area have revealed patterns of estrogen receptor α and vasopressin in the paraventricular nucleus of the hypothalamus that appear to be related to sociality and which were unexpected based on studies of *Mus* and *Rattus* (Kramer et al. 2005).

Another behavior is the fascinating interspecific variation in the structure of burrows between *P. maniculatus,* which builds shallow nests, and *P. polionotus,* which digs deep tunnels. A genetic analysis of the difference showed tunneling to be dominant, and that the difference between species appears to be controlled by *as few as one or two genes* (Dawson et al. 1988; Weber and Hoekstra submitted).

II.3. Habitat Adaptation: Coloration. *Peromyscus* provides a rich source of examples of intraspecific coat color variation enabling mice to blend in with their environment – a textbook example of adaptation. From work in *Mus,* the cell biology, biochemistry, and genetics of pigmentation are well understood (Silvers 1979) and are conserved across vertebrates (Hoekstra 2006). This provides an excellent system to determine which genes are *naturally* exploited, and what genetic changes result in the production of protective coloration (Hoekstra et al. 2006; Steiner et al. in press). Moreover, elucidating variation in the pigmentation pathway will help inform cancer biology, specifically the study of melanomas.

In addition, genetic dissection of spontaneous lab mutants is also of biomedical relevance. For example, *ashiness*, an age-dependent coat color mutation in the deer mouse (Teed et al. 1990) may serve as a model for melanocyte function, as these animals develop white hairs on the muzzle and at the base of the tail at age of six months. Some become virtually all white by 18 months. Implants of melanocyte-stimulating hormone induced production of pigment in depigmented portions of the coat, indicating that viable melanocytes were present.

II.4. Habitat Adaptation: Photoperiod Sensitivity. Like many species in temperate and boreal zones, *Peromyscus* do not breed the year round but are subject to seasonal breeding patterns. An adaptive manifestation of this is the gonadal regression stimulated by short photoperiods. Selective breeding resulted in the generation of photoperiod sensitive and resistant lines for both *P. leucopus* (Heideman et al. 1999) and *P. maniculatus* (Desjardins et al. 1986). These studies indicate that reproductive photoresponsiveness is significantly heritable. Other research showed an interaction between food and photoperiod that is not dependent upon body condition or energy balance (Reilly et al. 2006). Still more work revealed that short days not only reduce testosterone concentration in male white-footed mice (*P. leucopus*) compared with long days, but short days also impaired spatial learning and memory performance by a mechanism which does not directly involve the hippocampus (Pyter et al. 2006). Furthermore, *Peromyscus* living in the tropics do not show seasonal variation in immune response like their temperate-climate counterparts (Demas and Nelson 2003). Because *Peromyscus* occur over such a wide range of latitudes they are an ideal model for understanding natural variation in neuroendocrine traits and immune response and the role that variation plays in life-history evolution.

II.5. Habitat Adaptation: Altitude. *P. maniculatus* has the broadest altitudinal range of any North American mammal. This taxon has therefore attracted considerable attention as a model for studying the genetic basis of adaptation to high-altitude. Research in physiological genetics has revealed that adaptive variation in blood biochemistry among deer mice from different elevations is associated with a complex hemoglobin polymorphism (Chappell and Snyder 1984; Chappell et al. 1988; Snyder et al. 1988; Storz 2007; Storz et al. 2007; Storz et al. submitted). In high altitude deer mice from different mountain ranges across North America, adaptive variation in hypoxia tolerance is attributable to variation among duplicated globin genes as well as allelic variation that is segregating at each of the different genes (Storz et al. 2007; Storz et al. submitted). This system represents a unique case where fitness-related variation in whole-organism physiology can be related to a relatively simple biochemical phenotype (blood oxygen affinity) with a well-characterized genetic basis. It is also a prime example of natural genetic variation in *Peromyscus* promoting discoveries unlikely with laboratory strains of *Mus* or *Rattus*.

III. STRATEGIC PLAN FOR ACQUIRING GENOME SEQUENCE

III.1. The *Peromyscus* Genetic Stock Center. The Stock Center was established at the University of South Carolina in 1985 with funding originally derived from NSF, and later, with additional funding from NIH. One reviewer described the Stock Center as the "Jackson Laboratory of *Peromyscus*". Its main missions are 1) to furnish disease free, genetically defined, variant and normal, living, pedigreed *Peromyscus* to investigators and educators for research and instruction, and 2) to improve *Peromyscus* as a laboratory animal resource. Currently the Stock Center maintains about 4000

animals representing seven species and two subspecies of *Peromyscus* along with 19 pure-breeding lines of single mutations affecting behavior, physiology, and coat color. In the last four years approximately 4000 live animals were provided to over 90 different laboratories engaged in *Peromyscus* research. In the last four years there has been a 50% increase in the number of labs annually utilizing Stock Center resources. In addition to live animals, the Stock Center supplies biological materials including fresh, frozen, and preserved tissues and molecular probes and libraries. The Center also functions as a clearinghouse for information regarding this genus by sponsoring an internet database (*Pero*Base, http://wotan.cse.sc.edu/perobase/), a *Peromyscus* Genetic Stock Center web page (http://stkctr.biol.sc.edu), and the semiannual "*Peromyscus* Newsletter" which includes a listing of new publications in the field (~100/yr). The newsletter is distributed to ca. 550 individuals. Additionally, the PGSC web page lists reprints of more than 3000 article-length papers on *Peromyscus* held by the Stock Center. This resource is particularly valuable as it contains old literature not readily obtained elsewhere.

III.2. The Research Community. There are over one hundred researchers involved in both field- and lab-based *Peromyscus* research. The PGSC provides animals to 87 frequent users. Approximately half of these researchers are involved in epidemiology, infectious disease and toxicology research. The other half includes a diversity of researchers studying various aspects of *Peromyscus* biology, especially physiological and behavioral studies (see Appendix I). Another large group of researchers is conducting studies on wild-caught *Peromyscus* and include research in evolution, biogeography, ecology, and systematics. An increasing number of these researchers are utilizing molecular techniques as new resources become available, such as the linkage map and EST libraries. The *Peromyscus* genome sequence would provide a much needed resource for expounding these important biological and biomedical studies.

III.3 Demand for the *Peromyscus* sequence data, current research utilizing *Peromyscus*. We have received 22 letters from investigators who utilize *Peromyscus* in their research and who deem the acquisition of *Peromyscus* genome sequence data as essential to the further development and expansion of their own research capabilities (Appendix II). In addition to indicating the enthusiasm and magnitude of the community, these letters reveal the breadth of research activity that would be enhanced by the availability of a full *Peromyscus* genome sequence.

III. 4. Current Genomic Resources.

III.4.1 Genome size. The genome size of *P. maniculatus* has been shown to be similar to that of the house mouse. The genome is distributed over 46 autosomes plus the X and Y sex chromosomes (Deaven et al. 1977).

III.4.2 Genetic map. Through our mapping efforts, we have produced 6 libraries enriched for microsatellite DNA loci, yielding ~3000 clones, >1000 primer pairs, and >400 polymorphic loci. We have also sequenced ~6000 ESTs from 2 cDNA libraries (see below). We are currently genotyping the microsatellite loci and genes from the ESTs which occur at informative intervals on the *Rattus* and/or *Mus* maps. Our overall goal is to develop *Peromyscus* genomics for studies in which it is uniquely suited, while simultaneously taking advantage of comparative information provided by the gene-dense Human, *Mus* and *Rattus* genome projects and using these as reference species. We envision an intermediate genetic map will be assembled by October 1, 2007 (Ramsdell et al. 2007 submitted).

III.4.3 Cytogenetic and synteny map. Reciprocal cross species chromosome painting using whole chromosome paint probes from both *P. maniculatus* and *M. musculus* has provided us with a complete synteny map between these two species. This data has allowed for the derivation of chromosome homologies between *Peromyscus* and species for which *Mus* syntenies have been developed, including Human and *Rattus*. From this we have developed the ancestral Muroidea karyotype and defined breaks of synteny between *Peromyscus*, *Rattus* and *Mus* (Mlynarski et al. 2007

submitted). This map, combined with the newly derived linkage map (Ramsdell et al. 2007 submitted), will provide physical landmarks for the assignment of chromosomes for genome sequence annotation.

III.4.3 BAC libraries. Bacterial Artificial Chromosome (BAC) libraries have been constructed for *P. maniculatus* (Childrens' Hospital Oakland Research Institute, CHORI-233) and for *P. leucopus* (J. Storz, Univ. Nebraska). Each of these libraries provides ~11X coverage and has been screened several times. Individual BAC clones have been isolated that span the *α-* and *β-globin* gene families as well as several pigmentation loci. The sequencing, assembly, and annotation of these BAC clones (185-190 kb in length) have revealed patterns of conserved gene structure and synteny among the genomes of *Rattus*, *Mus*, and *Peromyscus* and have provided many important insights into the process of gene family evolution (Storz et al. submitted). The Vrana lab at U.C. Irvine probed the filters to identify several genomic regions of import to the *Peromyscus* research community. Initial regions coincided with the research interests of several laboratories: coat color genetics (Hoekstra lab, Harvard), genomic imprinting, X chromosome-inactivation, genome organization (O'Neill lab) and placental development (Vrana lab).

III.4.4 EST libraries. As an early attempt to expand the molecular resources available for *Peromyscus*, the Stock Center sequenced the 5' ends of 1510 placenta clones and 4798 testis clones. After removing low quality sequences, clustering, and contig assembly, there remained 904 unique placenta and 2002 unique testis sequences. These sequences are widely distributed across the chromosomes and have proven invaluable to the efforts to map the deer mouse genome. Despite widespread coverage, however, there remain areas for which we have no genetic information and therefore cannot map. Thus, we submitted a proposal to the Department of Energy's Joint Genome Institute to construct cDNA libraries from 10 *P. maniculatus bairdii* tissues (whole early-gestation embryo, whole mid-gestation embryo, whole newborn pups, ovary, epithelium, brain, liver, spleen, thyroid and bone marrow) and to sequence and annotate 10,000 ESTs from each of those libraries. That proposal was recently approved and the sequencing efforts are projected to reach completion by December 31, 2007.

III.5 The rationale for complete sequence of the organism. Broad comparisons of coding and non-coding regions will be necessary 1) for examining the abundances and distributions of members of gene families and repetitive elements, 2) to examine the roles of genome-wide sequence structures as foci of recombination, 3) to provide target material for the isolation and/or amplification of specific sequences for many different kinds of studies, and 4) to enable whole genome analyses of *P. maniculatus* genome structure in comparisons to other mammalian and more distant vertebrate species. Neither the targeting of specific coding regions, nor the piecemeal examination of BAC clones will have the power to reveal the whole genome changes that have accompanied the specializations of gene structure and function that occurred during the diversification of mammals. Microarrays based on whole genome sequence will be developed and produced at the Center for Applied Genetics and Technology (UCONN) and will have the advantage of including all genes in the genome. In addition, such arrays can be used in novel "fast-track" approaches to the positional cloning of QTLs that contribute to normal and abnormal physiologic and behavioral variation and are detected and localized by QTL linkage mapping approaches.

III.6 Sequencing strategy. We propose to obtain 6x genome coverage of *P. maniculatus*. This species is the most widely used in biological studies and was the species used for the *Mus*/*Peromyscus* synteny map, thus providing useful anchors in scaffold assembly. In addition, we propose to obtain 2x genome coverage for three additional species. *P. leucopus*, *P. californicus* are two species that represent divergent clades within Peromyscines and are broadly used in biomedical research. We also propose to obtain 2X coverage sequence for a third species, *P. polionotus*. This species will aide in SNP discovery as our recently completed linkage map is based on a backcross between *P. maniculatus* and *P. polionotus.* This last species is part of this initiative to increase the density of SNPs to provide identification of disease loci and traits in various Peromyscine models.

This strategy will allow for confirmation of SNPs by genotyping on the previously obtained backcross panel between *P. maniculatus* and *P. polionotus* (<100,000 years divergent).

We have consulted with the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) in order to identify their interest in the project and to discuss the appropriate strategy (see attached letter). They will guide us to the final methods for a complete genome sequence. This will involve traditional and new sequencing technologies (NST), which are rapidly changing, and likely will offer different options by the time this project is executed. Thus the methodology presented below is an example, but not necessarily the final approach to be taken in this initiative.

The goal of the project is a high quality draft sequence (e.g. 6x coverage) using the most straightforward approach to obtain a mammalian draft genome sequence: traditional 'Sanger' sequencing. The 454 and Solexa (and Applied Biosystems SOLiD) methods have not yet been used for large genomes for anything other than resequencing. So the most conservative approach to be considered is replacing some amount of the Sanger reads with those from a NST, the exact mix to be determined by cost and quality considerations.

However, with the promise from 454 of improvements in the near future, i.e. longer reads (500 bases), better read pair methods, more reads/bases per run (2 million reads, 1 Gb data), and reduced costs, by the time the project starts it may be possible to perform the entire WGS component using 454 technology, at a reduced cost over the mixed approach. Some limitations are the particular error modes of 454 sequences (particularly homopolymeric tracts) and the challenge in assembling the data (the 454 software does not handle such large data sets although the ATLAS assembler does). It is possible that the sequence can be upgraded by mixing 454 data with reads from another NST (e.g. Solexa or SOLiD) that has a different error mode and thus can 'correct' the 454 draft. This may be possible and still remain cost effective. We expect there to be advances in the NST field in the near term that will allow new approaches such as that described above to be realistic and cost-effective alternatives to the more conservative mixed approach. Similar strategies would apply to the lower coverage genomes proposed.

Finally, traditional Sanger sequencing can be used in a more targeted fashion for selective finishing of regions of interest or upgrading particularly difficult regions in the assembly. The availability of the BAC resources and genetic and physical mapping reagents will be particularly useful for these purposes and make useful complements to a whole genome sequencing strategy. The data will also include BAC end sequences representing approximately 10-fold genome coverage from the CHORI233 library. The combined data will be assembled using the ATLAS assembler developed at the BCM-HGSC and the resulting contigs and super-contigs will be associated with the appropriate chromosomal regions, first via the *Peromyscus* data, and then via the consideration of syntenic alignments to the mouse sequence.

The DNA to been sequenced will be chosen from an individual who carries a low heterozygosity in order to facilitate the ultimate DNA assembly. Candidate females have already been identified.

All data will be placed into the pubic databases according to the standard HGSC and NHGRI policies.

III.7. Financial statement. There are no other sources for funding this sequencing project available at this time.

Literature Cited

Belinsky SA, Bradford BU, Forman DT, Glassman EB, Felder MR, and Thurman RG. 1985. Hepatotoxicity due to allyl alcohol in deermice depends on alcohol-dehydrogenase. *Hepatology* 5:1179-1182.

Burnett KG and Felder MR. 1978a. Genetic regulation of liver alcohol-dehydrogenase in *Peromyscus. Biochemical Genetics* 16:443-454.

Burnett KG and Felder MR. 1978b. *Peromyscus* alcohol-dehydrogenase: lack of crossreacting material in enzyme negative animals. *Biochemical Genetics*. 16:1093-1105.

Burnett KG and Felder MR. 1980. Ethanol metabolism in *Peromyscus* genetically deficient in alcohol dehydrogenase. *Biochemical Pharmacology* 28:125-130.

Cattanach BM and Perez JN. 1969. A genetically determined variant of the A-subunit of lactic dehydrogenase in the deer mouse. *Biochemical Genetics* 3:499-506.

Chance, MRA. and Yaxley DC. 1949. New aspects of the behavior of *Peromyscus* under audiogenic hyper-excitement. *Behaviour* 2:96-105.

Chappell MA, Snyder LR. 1984. Biochemical and physiological correlates of deer mouse alpha-chain hemoglobin polymorphisms. *Proc Natl Acad Sci U S A*. 81(17):5484-8.

Chappel MA, Hayes JP, Snyder LRG 1988. Hemoglobin polymorphisms in deer mice (Peromyscus maniculatus), physiology of beta-globin variants and alpha-globin recombinants. Evolution 42 (4): 681-688

Csiszar A, Labinskyy N, Xiangmin Z, Hu F, Serpillon S, Huang Z, Ballabh P, Levy R, Hintze T, Wolin M, Austad SN, Podlutsky A and Ungvari Z. 2007 – *submitted* . Vascular $O_2^{\cdot-}$ and H_2O_2 production and oxidative stress resistance in two closely related rodent species with disparate longevity

Dawson WD. 1965. Fertility and size inheritance in a *Peromyscus* species cross. *Evolution* 19:44-55.

Dawson WD, Lake CE, Schumpert SS 1988. Inheritance of burrow building in *Peromyscus*. *Behav Genet* 18:371-82

Deaven LL, Vidal-Rioja L, Jett JH, and Hsu TC. 1977. Chromosomes of *Peromyscus* (Rodentia, Cricetidae). VI. The genomic size. *Cytogenetics and Cell Genetics* 19:241-249.

Demas GE and Nelson RJ. 2003. Lack of immunological responsiveness to photoperiod in a tropical rodent, *Peromyscus aztecus hylocetes*. *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology* 173:171-176.

Desjardins C, Bronson FH, and Blank JL. 1986. Genetic selection for reproductive photoresponsiveness in deer mice. *Nature* 322:172-173.

Dice, LR. 1935 Inheritance of waltzing and of epilepsy in mice of the genus Peromyscus. *J.Mammal*. 16: 25-35.

Dickerson RL, Hooper MJ, Gard NW, Cobb GP, Kendall RJ. 1994. Toxicological foundations of ecological risk assessment: biomarker development and interpretation based on laboratory and wildlife species. *Environ Health Perspect*. 102 Suppl 12:65-9

Duselis AR and Vrana PB. 2007. Assessment and disease comparisons of hybrid developmental defects. *Human Molecular Genetics* 16:808-819.

Duselis AR, Obergfell C, Mack JA, O'Neill MJ, Nguyen QK, O'Neill RJ, and Vrana PB. 2007 - *in press*. Cell-Cycle and extra-cellular matrix gene expression changes during deer mouse (*Peromyscus*) hybrid placental development. *Reproduction, Fertility and Development*.

Egoscue HJ. 1962. New hairless mutation in deer mice. *J Hered* 53 (4): 192-& 1962

Glassman EB, McLaughlin GA, Forman DT, Felder MR, and Thurman RG. 1985. Role of alcohol dehydrogenase in the swift increase in alcohol metabolism (SIAM). Studies with deermice deficient in alcohol dehydrogenase. *Biochemical Pharmacology* 34: 3523-3526.

Gubernick DJ and Teferi T. 2000. Adaptive significance of male parental care in a monogamous mammal. *Proceedings of the Royal Society of London Series B-Biological Sciences* 267:147-150.

Heideman PD, Bruno TA, Singley JW, and Smedley JV. 1999. Genetic variation in photoperiodism in *Peromyscus leucopus*: geographic variation in an alternative life-history strategy. *Journal of Mammalogy* 80:1232-1242.

Hoekstra HE. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97:222-234.

Hoekstra HE, Hirschman RJ, Bundey RA, Insel PA, and Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313:101-104.

Husby MP, Hausbeck JS, McBee K. 1999 Chromosomal aberrancy in white-footed mice (Peromyscus leucopus) collected on abandoned coal strip mines. *Environmental Toxicology and Chemistry* 18:919-925.

Ikegami K, Iwatani M, Suzuki M, Tachibana M, Shinkai Y, Tanaka S, Greally JM, Yagi S, Hattori N, and Shiota K. 2007. Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes to Cells* 12(1):1-11.

Isselbacher KJ and Carter EA. 1970. Ethanol oxidation by liver microsomes: evidence against a separate and distinct enzyme system. *Biochemical and Biophysical Research Communications* 39:530-537

Jaenisch R and Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* 33(3s):245-254.

Joyner CP, Myrick LC, Crossland JP, Dawson WD. 1998. Deer mice as laboratory mice. *ILAR Journal* 39:322-330.

Kramer KM, Yamamoto Y, Hoffman GE, and Cushing BS. 2005. Estrogen receptor alpha and vasopressin in the paraventricular nucleus of the hypothalamus in *Peromyscus*. *Brain Research* 1032:154-161.

Krugner-Higby L, Shadoan M, Carlson C, Gendron A, Cofta P, Marler C, Wagner J. 2000. Type 2 diabetes mellitus, hyperlipidemia, and extremity lesions in California mice (*Peromyscus californicus*) fed commercial mouse diets. *Comparative Medicine* 50:412-418.

Krugner-Higby L, Shelness GS, Holler A. Heritable, diet-induced hyperlipidemia in California mice (*Peromyscus californicus*) is due to increased hepatic secretion of very low density lipoprotein triacylglycerol. *Comp Med*. 56(6):468-75.

Lewis MH, Tanimura Y, Lee LW, and Bodfish JW. 2007. Animal models of restricted repetitive behavior in autism. *Behavioral Brain Research* 176:66-74.

Lieber CS and DeCarli LM. 1970. Hepatic microsomal ethanol-oxidizing system: in-vitro characteristics and adaptive properties in-vivo. *Journal of Biological Chemistry* 245:2505-2512

Loschiavo M, Nguyen QK, Duselis AR, and Vrana PB. 2007. Mapping and identification of candidate loci responsable for *Peromyscus* hybrid overgrowth. *Mammalian Genome* 18:75-85.

Mlynarski, E.E., Obergfell, C., Ramsdell, C., Dewey, M.J., O'Neill, M.J., and O'Neill, R.J. 2007 -submitted. *Peromyscus maniculatus* and *Mus musculus* Synteny Map Reveals Divergent Patterns of Breakpoint Reuse in Rodentia. *Genome Reserch*

Musser GG, Carleton MD (1993) *Peromyscus*. In: Wilson DE, Reeder DM (eds) Mammal Species of the World. Smithsonian Institution Press, Washington and London, p 728

Parnell PG, Crossland JP, Beattie RM, and Dewey MJ. 2005. Frequent Harderian gland adenocarcinomas in inbred white-footed mice (*Peromyscus leucopus*). *Comparative Medicine* 55:382-386.

Presti MF, Powell SB, and Lewis MH. 2002. Dissociation between spontaneously emitted and apomorphine-induced stereotypy in *Peromyscus maniculatus bairdii*. *Physiology and Behavior* 75:347-353.

Pyter LM, Trainor BC, and Nelson RJ. 2006. Testosterone and photoperiod interact to affect spatial learning and memory in adult male white-footed mice (*Peromyscus leucopus). European Journal of Neuroscience* 23:3056-3062.

Ramsdell, CM, Lewandowski AA, Weston Glenn JL, Vrana PB, O'Neill RJ, and Dewey MJ. 2007 - s*ubmitted*. Comparative genome mapping of the deer mouse (*Peromyscus maniculatus*) reveals greater similarity to rat (*Rattus norvegicus*) than to the lab mouse (*Mus musculus*). *Genome Research*.

Reilly SJ, Oum R, Heideman PD. 2006. Phenotypic plasticity of reproductive traits in response to food availability and photoperiod in white-footed mice (*Peromyscus leucopus*). *Oecologia* 150:373-382.

Ryan KK and Altmann J. 2001. Selection for male choice based primarily on mate compatibility in the oldfield mouse, *Peromyscus polionotus rhoadsi*. *Behavioral Ecology and Sociobiology* 50:436-440.

Sacher GA. 1978. Longevity and aging in vertebrate evolution. *Bioscience* 28 (8): 497-501 1978

Silvers WK 1979. The Coat Colors of Mice: a Model for Mammalian Gene Action and Interaction. Springer-Verlag, New York

Snyder LRG, Hayes JP, Chappell MA. 1988. Alpha-Chain Hemoglobin Polymorphisms are Correlated with Altitude in the Deer Mouse, *Peromyscus maniculatus Evolution* 42(4): 689-697

Storz JF. 2007. Hemoglobin function and physiological adaptation to hypoxia in high-altitude mammals. *Journal of Mammalogy* 88:24-31.

Storz JF, Sabatino SJ, Hoffman FG, Gering EJ, Moriyama H, Ferrand N, Monteiro B, and Nachman MWl. 2007. The molecular basis of high-altitude adaptation in deer mice. *PLOS Genetics* 3:448-459.

Storz JF, Hoffmann FG, Opazo JC, and Moriyama H. 2007 - *in review*. Adaptive functional divergence among triplicated alpha-globin genes in rodents. *Molecular Biology and Evolution*

Sumner FB 1924. Hairless mice. *Journal of Heredity* 15: 475+.

Steiner CC, Weber JN and Hoekstra HE. 2007- in press. Adaptive variation in beach mice caused by two interacting pigmentation genes. PLoS Biology.

Szalai G, Duester G, Friedman R, Jia HG, Lin SP, Roe BA, and Felder MR. 2002. Organization of six functional mouse alcohol dehydrogenase genes on two overlapping bacterial artificial chromosomes. *European Journal of Biochemistry* 269:224-232.

Teed SK, Crossland JP, and Dawson WD. 1990. Coat color genetics of *Peromyscus*. 1. Ashiness, an age-dependent coat color mutation in the deer mouse. *Journal of Heredity* 81:309-313.

Teschke R, Matsuzaki S, Ohnishi K, DeCarli LM, and Lieber CS.1977. Microsomal ethanol oxidizing system (MEOS): current status of its characterization and its role. *Alcoholism: Clinical and Experimental Research* 1:7-15.

Vrana PB, Fosella JA, Matteson P, del Rio T, O'Neill MJ, and Tilghman SM. 2000. Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in *Peromyscus*. *Nature Genetics* 25:120-124.

Watson, ML 1939. The inheritance of epilepsy and of waltzing in *Peromyscus. Contr. Lab. Of Vertebrate Genetics, Univ. Michigan*, No. 11.

Weber JN and Hoekstra HE. 2007 - in review. Evolution of burrowing behavior in deer mice (genus Peromyscus).

Zheng YW, Bey M, Liu H, and Felder MR. 1993. Molecular-basis of the alcohol dehydrogenase-negative deer mouse. Evidence for deletion of the gene for class I enzyme and identification of a possible new enzyme class. *Journal of Biological Chemistry* 268:24933-24939.

Research Applications and Institutional Usage of *Peromyscus*
Three years (2004-2006)

Research Application	Number of Specimens (%)	Number of Orders (%)
Basic Biology: Hybrid dysgenesis, reproduction, behavior, inbreeding depression, physiology, genetics, photoperiod, molecular biology, etc.	3316 (60.4)	120 (55.3)
Biomedical: Epidemiology, host-parasite, Lyme, Hanta, Borellia, pharmacokinetics, immunology.	832 (15.4)	53 (24.4)
Toxicology and Xenobiotic Metabolism	1022 (18.6)	25 (11.5)
Educational: Museums, teaching, demonstrations	323 (5.9)	19 (8.8)

Institution	Number of Specimens (%)
Academic	4711 (85.8)
Government	370 (6.7)
Private	412 (7.5)

Wild-Type Stock Utilization of *Peromyscus*
Three year (2004-2006)

Species/Subspecies (Stock)	Specimens	
	Number	PerCent
P maniculatus bairdii (BW)	1393	31.3%
P leucopus (LL)	1410	31.9%
P californicus (IS)	491	10.9%
P polionotus subgriseus (PO)	419	9.3%
P maniculatus sonoriensis (SM2)	272	6.0%
P eremicus (EP)	196	4.3%
P melanophrys xenerus (XZ)	162	3.6%
P aztecus (AM)	157	3.5%
P polionotus leucocephalus (LS)	10	0.2%
TOTAL	4510	