Proposal for the Construction of New BAC Libraries From Three Commonly Used Inbred Rat Strains

Howard Jacob, Ph. D.
Director, Human and Molecular Genetics Center
Professor, Department of Physiology
Warren P. Knowles, Chair in Genetics
Medical College of Wisconsin
8710 Watertown Plank Road
Milwaukee WI 53236
jacob@mcw.edu
Ph. (414) 456-4887
Fax (414) 456-6516

Importance of the organism and genomic resources: The importance of the laboratory rat in biomedical research is well established. Since 1966, there have been on average over 27,000 publications per year using rat (PubMed search, key word: rat). In the last 7 years (1996-2002) there have been on average 35,000 publications annually. The initiation of the rat genome project has yielded a tremendous wealth of genomic resources including a cytogenetic map; radiation hybrid (RH) cell lines and the associated RH maps (over 6,000 genetic markers and 16,000 genes and ESTs mapped); cDNA libraries generating more than 369,341 ESTs (with more being generated) clustered into over 63,000 UniGenes; over 10,000 genetic markers; and a draft (~6.5 X) sequence of the genome. The rat is primarily known as a physiological model and there has been some question over the years about the likelihood that the rat genomic tools would be fully utilized. Figure 1 show the number of users per week of the EBI genome browsers for mouse in its first 10 weeks versus the first 10 weeks since the rat genome was released. The rapid jump from week 9 to week 10 is likely to be the result of the community becoming aware of the data. The number hits on this browser, the number of visits to the Rat Genome Database (54,000 per month) and the hits to our rat physiology site (PhysGen, 2000 per month) suggest that the rat tools are being well used and are likely to be continued to be used.

One reason the rat is a dominant model in nutrition, pharmacology and physiology is that there are over 600 inbred strains of rats (http://rgd.mcw.edu) most of which were developed as models for complex, common diseases; furthermore, there are over 200 transgenic rats (Mary Shimoyama, RGD, personal communication). As examples, there are strains used in studies of transplantation and immunogenetics, cancer, cardiovascular diseases, behavior, growth and reproduction, various metabolic disorders, neurological and neuromuscular disorders, diseases of the skin and hair, aging, sleep apnea, pulmonary disease, nutrition, endocrinology, and toxicology. For the genetic tool set, the 48 most commonly used inbred strains have been characterized with respect to the allele sizes of 4,328 SSLP markers (Steen, Kwitek-Black et al. 1999). The large number of genetic markers characterized in this large number of strains, plus a dense radiation hybrid map consisting of 22,000 mapped elements has enabled the construction of detailed comparative maps for the rat, mouse and human (http://rgd.mcw.edu/VCMAP) thereby enabling rat physiology to be linked to mouse genetics and to human clinical questions. While genomic tools are advancing the fields of complex disease genetics, pharmacology and physiology, they also point up the need for constructing more rat BAC libraries, as have already been created for multiple mouse strains, in order to provide additional relevant genetic tools for rat disease model studies.

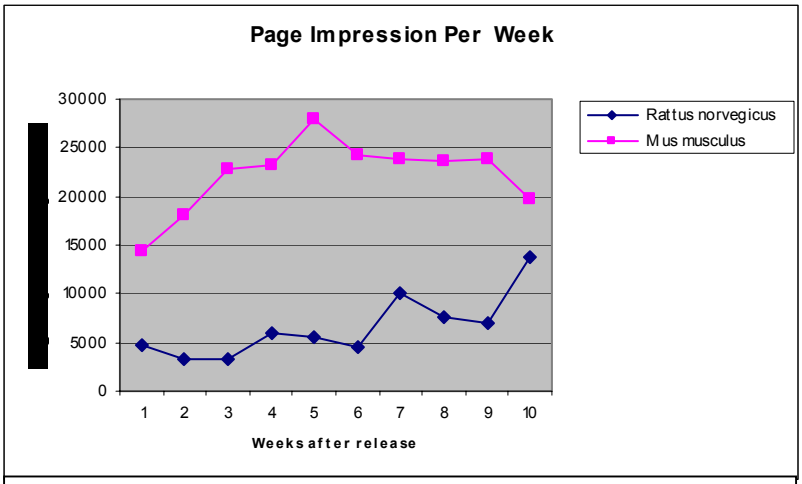

Figure 1: Comparison of Page Impressions Between Rat and Mouse for the first 10 weeks of release of the first draft. EBI provided this comparison of rat and mouse genome sequence "usage" on their genome browser.

Use of BAC libraries: The existing BAC library for rat is from the BN strain, developed to support the sequencing project. There is now a need for more "alleles" to be available for identifying single nucleotide polymorphisms, understanding conserved non-coding sequences (CNS), and for functional cloning via transgenic BAC rescue experiments.

The identification of genes involved in complex traits and phenotypes remains problematic, despite the availability of a wealth of genomic tools, powerful analytical strategies and computational algorithms. Localizing quantitative trait loci (QTLs) to specific, relatively small genetic intervals (2-10

cM) is a straightforward, albeit laborious task; identifying the gene(s) of interest within the interval is not as clear-cut (Nadeau and Frankel 2000; Glazier, Nadeau et al. 2002; Jacob and Kwitek 2002). Many positional cloning projects for common, complex diseases report that the QTLs contain more then one gene (Podolin, Denny et al. 1998; Legare, Bartlett et al. 2000; Morel, Blenman et al. 2001), and in these cases functional cloning (Symula, Frazer et al. 1999) may be the only way to demonstrate a causal relationship between a disease and a variant gene or CNS. To our knowledge, the first functional cloning experiment in rat using a BAC transgenic approach has only recently been completed. Figure 2 shows the functional rescue of lymphopenia using a BAC from the BN library (Dr. Michael Michalkiewicz, personal communication).

The availability of BAC libraries from 4 strains of rat would greatly facilitate the use of sequencing and the identification of single-nucleotide differences (SNPs) within candidate genes, because there would be, on average, 3 alleles (see below for evidence) for each gene. Since knock-in technology is not yet perfected, availability of 4 libraries (3 proposed in this white paper and the existing 1), and the availability of efficient site-directed mutagenesis in BACs using bacteria are critical tools for validating genes that have been positionally cloned in the rat (Lee, Yu et al. 2001). A final function of the libraries will be their enabling comparisons between conserved noncoding sequences in different inbred strains without the need to conduct site directed mutagenesis for each CNS.

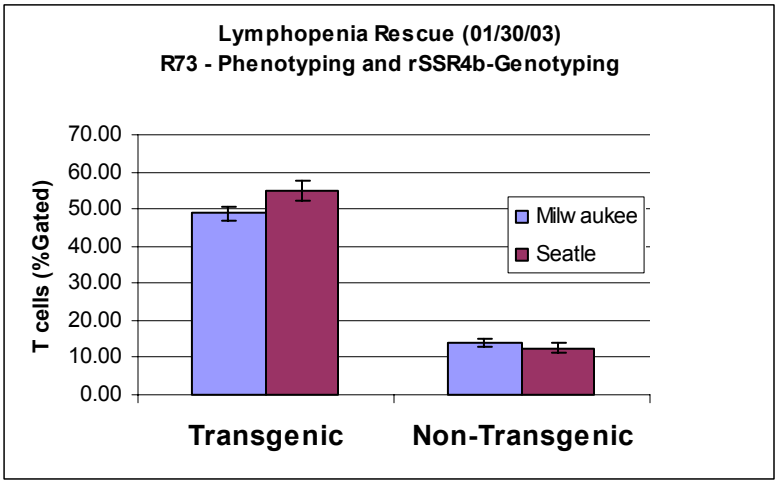

Figure 2: BAC Transgenic Rescue of Lymphopenia. The BAC contain IAN5 (MacMurray, Moralejo et al. 2002) which has been shown to be the gene was responsible for lymphopenia was inject into congenic F344 rats carrying the lymphopenia gene and expressing abnormally low numbers of T-cells. Seven transgenic and 8 non-transgenic litter mates were tested for T-cell numbers. Only the rats carrying the "wild type" IAN5 gene on a BAC from the BN library had normal T-cells. T-cells were determined in two locations, with investigators at both sites unware of the genotypes.

Selection of the rat strains. A rat BAC library already exists, based on the BN strains and developed as part of the rat genomic sequencing effort. Here, three additional BAC libraries are requested: F344, SS, and FHH. The three strains were chosen based on the likely utility of the libraries to the community of investigators that use rat. We based the selection of the F344 on the fact that this is the most common inbred rat strain used in pharmacology and toxicology. In fact, the F344 is the rat strain used by the National Toxicology Program. As cytochrome P450s are keys for metabolizing xenobiotics, the cytochrome P450s in this strain are arguably the best characterized of any species. Consequently, having a BAC library of this strain will facilitate the study of pharmacogenomics, toxicogenomics and other pharmaceutical studies.

The other two strains, SS/JrHsd/Mcwi (SS) and FHH/Eur (FHH) are part of the Programs for Genomics Applications (PGA) at the Medical College of Wisconsin. The goal of this PGA is to build two panels of chromosome substitution strains (consomics), one panel systematically replacing each chromosome (20 autosomes and the sex chromosomes) from the SS strain with that of the BN strain (the strain that was sequenced); for the other panel, each of the FHH chromosomes is replaced with its BN homolog. Each of the strains is then subject to rigorous phenotyping, generating an average of 8,165 data points per strain (146,974/ 18 strains completed to date). Over the next 18 months, all 44 consomic strains will undergo the same phenotyping protocols, yielding an unprecedented physiological data set, all of which has been genetically mapped to one or more chromosomes (**http://pga.mcw.edu**). All data generated by this project are freely available and can be downloaded from the web site. The parental

and consomic rats are being made publicly available through Charles River Laboratories. Not only is their baseline physiology well characterized, and mapped, but congenics can also be rapidly made from any consomic strain within 6 months, by an F2 intercross of the consomic with the background parental strain. These rats are likely to become a national resource for rat research. The website, which currently gets 2,000 visits per month (figure 3), enables investigators to download data or to use the analytical tools present on the website. The ability to generate BAC transgenics using the consomic rats would complete an extremely powerful collection of tools. A potential criticism of the selection of the SS and FHH is that these strains are primarily used by investigators from the National Heart Lung and Blood Institute. However, it is important to note that, collectively, the BN, FHH, and SS can be used to study alcoholism, asthma, cardiovascular disease, depression, endocrine disorders, insulin resistance, renal disease, reproductive fitness, and pathogen resistance. These three strains are currently being tested for a variety of other traits including (to name a few) behavior, cancer (breast, throat), neurohumoral axis, and anesthetic sensitivity . Finally, when the MCW investigators set out to develop these consomics, we investigated how closely related the strains were to each other, using the microsatellite allele characterization data. A recent paper by Thomas et al, formalized this analysis and demonstrates that the BN, F344, FHH and SS will cover 4 of the 7 clades that can be generated from the 48 most commonly used inbred strains(Thomas, Chen et al. 2003). In principle, making libraries from these strains will capture a significant portion of the known genetic variation in rat. As result, one can infer that a tremendous amount of the variation available in rat biology will also be captured. The development of these three new libraries will have a large impact on many investigators across a large number of fields.

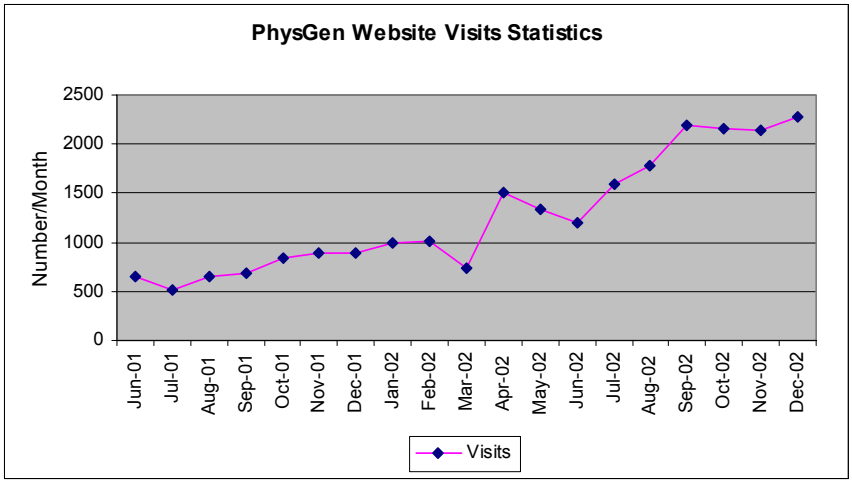

Figure 3: Average Monthly Number of Visits to the PhysGen site. The first data release was in June of 2001, since then the number of visits has doubled annually. A visit is defined as a series of pages without inactivity for 30 minutes

Research community: In contrast to the mouse community, the rat community is distributed across a large number of disease areas, with the biology, not the genetics, being the glue. A rat genomics community has developed over the last 5 years and now has an annual international meeting. Concurrently, as rat genomic and genetic tools have been developed, there has been an increase in the application of these tools in rat research. Visits to public rat resources continue to increase, as does the number of papers using genetic tools. Since 1991, when the first QTL was mapped in rat, there have been 320 papers published relating to genetic studies in the rat (PubMed search: quantitative trait AND rat), with nearly half being published in the last 3 years. These papers include investigations of the genetic basis of arthritis, copper metabolism, pituitary tumor growth, aerobic capacity, blood pressure and hypertension, diabetes, cardiovascular disease, ethanol tolerance, behavioral conditioning, anxiety, fat accumulation and chemical carcinogenesis as examples (PubMed abstracts).

Similarly, a CRISP search using the key words “rat and QTL” identified 27 NIH-funded grants on topics including alcohol, hypertension, breast and pituitary cancer and autoimmune disease. A number of inbred strains, of which BN, SS and SHR, WKY, FHH and F344 are the most commonly used, are employed in these projects. As a comparison, the equivalent searches for mouse-related research found 702 publications since 1991, and 92 funded grants. Therefore, while the rat does not get the same “air time” as the mouse, the genomic resources and reagents are being used. The availability of the genomic

sequence and these new BAC libraries will further enhance the tool kit for investigators using rats to study human disease.

Status of genomic sequencing: The rat genomic sequence project, based on a BAC library derived from the BN/SsN/Mcwi strain, has reached an approximate 6.5-fold coverage. The first draft assembly was released in November of 2002, and a second assembly is anticipated in February 2003. The rat genome size has been estimated to be ~3 x 10^9 bp.

Strains and DNA availability: All 3 strains, BN, FHH and SS, are from MCW and all have been sent to Charles River Laboratories for general distribution. MCW provides genetic monitoring of the strains and colonies at Charles River Laboratories. We collaborated with Dr. DeJong to generate the DNA for the BN BAC library, which has an average insert size of 210 kb.

Specifications: The depth of each library should be sufficient to represent >99% of the genome, such that there will be multiple clones covering any given locus. This ensures that most genes will be fully contained within at least one BAC, rather than being split at an end point. It is desirable for the BACs to be as large as feasible, i.e. 200 kb, to reduce the effort and cost required to generate and replicate the library: 5X coverage is adequate. No unusual vectors will be required.

Time frame: The large number of projects that make use of these strains in QTL mapping suggests that they will be used as soon as available.

Other support: No other BAC libraries are planned to my knowledge.

Literature Cited

Glazier, A. M., J. H. Nadeau, et al. (2002). "Finding genes that underlie complex traits." Science **298**(5602): 2345-9.

Jacob, H. J. and A. E. Kwitek (2002). "Rat genetics: attaching physiology and pharmacology to the genome." Nat Rev Genet **3**(1): 33-42.

Lee, E. C., D. Yu, et al. (2001). "A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA." Genomics **73**(1): 56-65.

Legare, M., F. n. Bartlett, et al. (2000). "A major QTL determined by multiple genes in epileptic EL mice." Genome Research **10**(1): 42 - 48.

MacMurray, A. J., D. H. Moralejo, et al. (2002). "Lymphopenia in the BB rat model of type 1 diabetes is due to a mutation in a novel immune-associated nucleotide (Ian)-related gene." Genome Res **12**(7): 1029-39.

Morel, L., K. R. Blenman, et al. (2001). "The major murine systemic lupus erythematosus susceptibility locus, Sle1, is a cluster of functionally related genes." Proc Natl Acad Sci U S A **98**(4): 1787-92.

Nadeau, J. and W. Frankel (2000). "The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs." Nature Genetics **25**(4): 381-384.

Podolin, P., P. Denny, et al. (1998). "Localization of two insulin-dependent diabetes (Idd) genes to the Idd10 region on mouse chromosome 3." Mamm Genome **9**(4): 283-286.

Steen, R. G., A. E. Kwitek-Black, et al. (1999). "A high-density intergrated genetic linkage and radiation hybrid map of the laboratory rat." Genome Research **9**: AP1-Ap8.

Symula, D. J., K. A. Frazer, et al. (1999). "Functional screening of an asthma QTL in YAC transgenic mice." Nature Genetics **23**(2):241-4.

Thomas, M. A., C. F. Chen, et al. (2003). "Phylogenetics of rat inbred strains." Mamm Genome **14**(1): 61-4.