White Paper for Complete Sequencing of the Common Marmoset

(*Callithrix jacchus*) genome

Keith Mansfield, Suzette Tardif, and Evan Eichler
Corresponding author:
Keith Mansfield, D.V.M.
Harvard Medical School
New England Regional Primate Research Center
PO Box 9102
Southborough, MA 01772-9012
Tel: 508-624-8183
Fax: 508-624-8190
Keith_Mansfield@HMS.Harvard.edu

Suzette Tardif, Ph.D.
Associate Director
Southwest Primate Research Center
PO Box 760549
San Antonio Texas, Tx 78245
210-258-9885
stardif@icarus.sfbr.org

Evan Eichler, Ph.D.
Associate Professor
University of Washington
Genome Sciences, HSB K336B
Box 357730
1705 NE Pacific St.
Seattle, WA  98195
phone: (206) 543-9526
fax:  (206) 685-7301
eee@gs.washington.edu
I. Introduction

The common marmoset (*Callithrix jacchus*) is a small New World primate native to eastern Brazil that has been used extensively in biomedical research models in both North America and Europe. New World primates (platyrrhines) represent a diverse group of animals encompassing 2 families (the Cebidae and Callitrichidae), and at least 76 distinct species. Separation of the platyrrhines occurred approximately 35 million years ago and this early separation coupled with adaptations to the neotropical environment has produced a number of distinct differences in physiology and disease susceptibility from Old World primates (catarrhines). Such differences present unique opportunities in model development not available in more traditional nonhuman primate species. Historically the common marmoset has been used in neuroscience, reproductive biology, infectious disease, and behavioral research. Recently the species has increasingly been utilized in drug development and safety assessment. Advantages relate to size, cost, husbandry and biosafety issues as well as unique physiological differences that may be utilized in model development. Availability and ease of breeding in captivity have spurred interest in the marmoset as an alternative species to more traditional nonhuman primates for use in biomedical research. A recent series of reviews summarizes many aspects of marmoset physiology, behavior and disease susceptibility as well as their use in biomedical research programs (1-6).

In contrast to many other laboratory animal species, utilization of nonhuman primates has increased in recent years and there currently exists a significant shortage of such animals for use in biomedical research. Although the rhesus monkey remains the most commonly used nonhuman primate in research vivaria in this country, the national supply of such animals has been unable to meet the current or projected demands of the research community. While efforts are underway to increase domestic production and to identify alternative foreign sources, this will unlikely alter short term availability. Despite these ongoing efforts, the shortage of well-defined research animals will likely negatively impact national biomedical and defense research initiatives for the foreseeable future. In addition to increasing production of rhesus macaques, one solution to this shortage is to provide alternative primate species that may be utilized by investigators in their research programs. Such was a recommendation from a recent workshop sponsored by the National Academy of Sciences entitled *International perspectives: the future of nonhuman primate resources* (7) and echoed in an Expert Panel’s *Recommendations for the Regional Primate Research Center Program* conducted for the National Center for Research Resources (8). The common marmoset was identified as a logical alternative primate species due to its size, availability and wide spread use in biomedical research. Key to this strategy is the development of molecular research tools that will allow investigators to take full advantage of unique attributes of such an alternative species.

In response to these recommendations the inaugural meeting of the *Marmoset Research Group of the Americas* was held on June 13-14, 2004 at the University of Wisconsin-Madison. This meeting brought together 61 scientists with diverse interests representing 32 different academic and private research organizations to discuss current needs and research opportunities (see appendix I). Development of
DNA genomic sequence of the common marmoset was identified as a high priority objective and was the impetus for submission of this document.

II. Rationale for a DNA genomic sequence of the Common marmoset (*Callithrix jacchus*).

This document summarizes the justification for producing a complete genomic DNA sequence for the common marmoset. Some notable characteristics of common marmosets that justify this project are presented below.

1) **The common marmoset is a widely used nonhuman primate species in biomedical research.** The common marmoset already supports diverse research programs and is used widely in studies involving aspects of infectious disease, stem cell research, neural and cognitive sciences, toxicology and drug development and reproductive biology. As discussed in detail below these programs are centered on understanding fundamental elements of human biology as well as the pathogenesis of important disease syndromes of man. Due to their close genetic, physiologic and metabolic similarity to humans this species will continue to serve as an important resource in such basic and biomedical research programs. *A genomic sequence would lead to the development of unique tools and research avenues in both existing and novel marmoset models of human disease.*

2) **The common marmoset has unique biological differences that make it useful for specific research programs.** There are a number of unique physiologic differences between Old World and New World primates that have been utilized in the development of novel models. For example, anatomic differences in placentation and frequent twinning leads to bone marrow chimerism of dizygotic twins. Thus twins may be used in adoptive transfer experiments to examine components of the cellular immune responses to a variety of antigens and in disease pathogenesis (9). Similar experiments are currently difficult or impossible to perform in Old World primates. Limited diversity at both MHC class I and II loci has been identified in both tamarins and marmosets and may be responsible for differences in disease susceptibilities that are exploited in infectious disease studies (10). The evolutionary forces that led to the limited MHC diversity and chimerism and the impact these adaptations had on the genome are not understood. *The availability of a DNA genomic sequence would assist in understanding these unique differences and adaptive responses.*

3) **The common marmoset is an attractive alternative to other nonhuman primate species used in biomedical research.** Common marmosets represent an attractive alternative nonhuman primate species for a variety of reasons. These small animals breed well in captivity with reproductive efficiencies that may exceed 150% (number of live born per year/ number of breeding females). Furthermore, sexual maturity is reached by 18 months of age allowing for rapid expansion of existing colonies. The small relative size of the common marmoset compared to macaque species may represent other potential advantages and translates into lower caging and feeding costs and reduced floor space requirements. Smaller size facilitates social housing of common marmosets and implementation of environmental enrichment programs. The marmoset is less destructive to its environment than larger nonhuman primates and this
makes possible the provision of a more complex and enriching environment. In drug development, reduced body mass may substantially decrease compound synthesis costs and time requirements. This coupled with lower purchase price of common marmosets compared to macaques, may translate into substantial cost savings when performing equivalent studies. For these practical reasons the marmoset has become a mainstay of biomedical research in Europe and has seen increased utilization in the United States. We anticipate that this trend will continue particularly as the demand for nonhuman primate studies increase with ongoing infectious disease and biodefense initiatives.

**Development of molecular tools is paramount to the continued refinement of these systems and further acceptance of the marmoset as an alternative primate species.**

4) **The severe shortage of macaque monkeys may in part be alleviated through increased utilization of other primate species.** In the past decade a severe shortage in the availability of rhesus macaques has developed and negatively impacted national biomedical research objectives. Despite on going programs to breed and import additional animals, this shortage persists. Alternative primate species such as the common marmoset provide an attractive alternative to more traditional primate species. The availability of research tools is a critical issue when evaluating species choice in model development. One of the reasons that rhesus macaques are attractive as an animal model is the availability of research tools developed by diverse groups that have application across scientific disciplines. While there are clearly not as many research tools available for marmosets as there are for macaque species, resources have now been developed in a number of key areas. *The availability of a DNA genomic sequence would add additional tools and speed the development of molecular and immunological assays that would positively impact diverse research programs.*

5) **The common marmoset represents a major arm of the primate evolutionary tree.** Separation of New World Primates occurred approximately 35 million years ago and predates the separation of Old World Primates from the human lineage by at least several million years. As indicated above this early separation coupled with adaptation to the neotropical environment has produced unique biological differences that have been used in animal model development and have lead to important insights into the mechanisms of disease. With completion of the rhesus macaque sequence consideration should be given to obtaining sequence from the platyrrhine primate lineage. *The availability of a DNA genomic sequence would permit comparative analysis of these differences and provide an outgroup for human-Old World primate comparisons.*

6) **The common marmoset is related in evolutionary and genetic terms to other nonhuman primates in biomedical research.** This organism is a member of the New World Monkeys (Superfamily Ceboidea), and is the anchor species of the callitrichine clade, one of seven anciently separated New World monkey clades that diverged from each other at least 18 mya. The marmoset is related to other New World Primate species including squirrel monkeys, owl monkeys and tamarins. While these other species represent an important biomedical resource, the common marmoset arguably supports more broad based and diverse research programs. The interspecies sequence differences for these species is likely to be in the range of 3-4% and sequence data from the marmoset would be valuable to researchers utilizing these species. *Access*
to genomic sequence data from marmoset may create new research opportunities for other commonly utilized nonhuman primate species.

III. Specific biological/biomedical rationale for the utility of sequence data

Historically the common marmoset has found use in neuroscience, reproductive biology, infectious disease, and behavioral research and more recently the species has increasingly been utilized in drug development and safety assessment. Highlighted below are major research areas that have made use of the common marmoset as well as how these areas would benefit from the availability of genomic DNA sequence from this species. These models are focused on understanding basic biological processes relevant to human health and disease and provide examples of how genomic sequence data will: 1) improve human health through the evaluation of novel therapeutic strategies; 2) inform human biology; 3) expand our understanding of basic biological processes relevant to human health and development; 4) provide additional surrogate systems for human experimentation by expanding existing model capabilities; and 5) facilitate the ability to do experiments in other primate systems.

1) Impact of Marmoset genomic data to improve human health and expand model capabilities

Infectious diseases. Common marmosets have been used extensively in infectious disease research due in large part to their unique sensitivity to a number of important human infectious agents. Research programs have been established to examine viral, bacterial and parasitic agents. In particular, marmosets are susceptible to a variety of herpes virus agents including gammaherpesviruses such as Epstein Barr virus (EBV) and Herpesvirus saimiri (HVS), both of which may result in lymphoproliferative disorders in this species. Common marmosets are also highly susceptible to alaphaherpesviruses such as Herpes simplex virus (HSV) and develop an acute disseminated disease with neurologic involvement. The species has been used in models of hepatitis A virus infection, Junin virus, malaria, measles and GB virus B, a surrogate model of human hepatitis C virus (table 1). The reason marmosets are highly susceptible to infection with many of these agents is not understood but limited diversity at MHC class I and II loci is hypothesized to play a role (10;11). Additional sequence data on MHC organization may shed light on this species unique susceptibilities.

Gammaherpesvirus models of acute oncogenesis. A distinct characteristic of gammaherpesviruses is their ability to establish latent infections in lymphoid cells and their association with abnormal lymphoid proliferation and cancer in humans and nonhuman primates (12). The first open reading frame of the primate gammaherpesviruses has been shown to directly contribute to viral associated pathogenesis and these gene products are capable of eliciting cellular signal transduction events resulting in cell growth transformation. The marmoset model has been used to evaluate the in vivo significance of oncogene deletion or substitution (13-15).

HVS inoculation of common marmosets is a well-established model of acute viral oncogenesis and has been used to investigate the molecular pathogenesis of viral induced lymphoma. Sequence divergence among HVS isolates is most extensive at the left end of the viral genomic DNA and is the basis for the classification of HVS into subgroups A, B, and C (16). Sequence variation in the saimiri transforming protein (STP) gene in this
region correlates with differences in viral capacity to induce \textit{in vitro} immortalization of T lymphocytes and for the induction of lymphoma in nonhuman primates (17;18). Both subgroup A and subgroup C viruses can immortalize common marmoset T lymphocytes to interleukin-2 (IL-2)-independent proliferation and highly oncogenic subgroup C strains are able to immortalize human, rabbit, and rhesus monkey lymphocytes. Inoculation of marmosets with HVS type C produces malignant lymphoma within 3-4 weeks. The malignant cells infiltrate a number of organs including the liver, spleen, kidney and gastrointestinal tract and leukemia can be visualized in peripheral blood smears.

Techniques have been developed to construct mutant HVS viruses and examine the subsequent effect of these deletions and viral gene substitutions on \textit{in vitro} viral replication and \textit{in vivo} disease course. In such experiments a number of outcomes may be measured including viral load, tumor induction and survival (19). Since recombinant HVS lacking STP can be repeatedly isolated from the peripheral blood of common marmosets for months or years, STP is not required for viral replication or persistence in \textit{vivo}, but it is essential for transformation in cell culture and for lymphoma induction in common marmosets. \textit{Host genomic data could be used to investigate how these viral oncogenes interact with host proteins.}

\textbf{Gammaherpesvirus models of persistent viral infection.} While the vast majority of gammaherpesvirus research in common marmosets has focused on the use of HVS, the pathogenesis of \textit{Herpesvirus ateles} and Epstein Barr virus (EBV) have also been studied (20). Experimental inoculation of common marmosets with EBV may induce a more indolent course with similarities to infectious mononucleosis and has been used to investigate the effect of immunization on viral shedding (21;22). Disease outcome may be dependant on viral strain used. Coinoculation of animals with \textit{Plasmodium brasilanum} and EBV has also been studied as an experimental model of Burkitt’s lymphoma (23).

More recently, an indigenous gammaherpesvirus related to EBV has been recognized in common marmosets and may serve as an additional model of viral oncogenesis and persistence (24). This virus termed marmoset lymphocryptovirus (LCV) or Callitrichine herpesvirus 3 was first identified in animals with spontaneous B cell lymphomas at the Wisconsin National Primate Research Center. Subsequent work has revealed that 40-60\% of captive marmosets are seropositive for this virus with most animals not revealing overt signs of clinical disease. Recently sequencing of the virus has been completed (25). Definition of the viral gene repertoire revealed collinear genomic organization and 60 open reading frames with homology to those seen in EBV and other primate LCVs. In addition to these conserved regions, a number of unique putative genes were recognized that are hypothesized to play a role in viral persistence. Future work to compare and contrast these closely related viruses may provide insight into gammaherpesvirus pathogenesis. \textit{Host genetic factors that influence viral latency is a major research focus and would be assisted by the availability of genomic sequence from this species.}

\textbf{GB virus B model of hepatitis.} GB virus B is a newly recognized infectious agent and member of the flaviviridae family related to hepatitis C (26). Hepatitis C virus (HCV) is the most common blood borne pathogen recognized in the United States and the incidence and health impact of this agent is expected to increase dramatically in the next
decade. HCV causes a persistent viral infection leading to chronic hepatitis and hepatocellular carcinoma. The only current animal model of HCV utilizes the chimpanzee. However, this model is faced with a number of drawbacks including ethical issues, availability and cost of housing such animals (27).

The GB agents are a group of closely related viruses initially recognized by investigators attempting to identify other infectious causes of non A-E hepatitis in man. GB virus A, another member of the flaviviridae family, may be found as a common asymptomatic infection of many species of New World primates. GB virus B is a hepatotropic virus that results in acute hepatitis when inoculated into New World primates (28). Unlike GB virus A, the natural host of GB virus B is unknown. GB virus B shares overall genomic organization with HCV and 25-30% homology with the HCV polyprotein (29). The putative envelop proteins E1 and E2 share structural motifs with HCV and similar function and specificity have been demonstrated for the NS3 serine protease in cleaving the viral polyprotein (30;31). GB virus B inoculation of several species of New World primates results in the development of acute hepatitis and shows promise as a novel surrogate animal model of HCV infection of man.

Marmosets inoculated with GB virus B develop multifocal nonsuppurative hepatitis with portal inflammatory cell infiltrates and piecemeal necrosis. Sequential hepatic biopsies may be obtained under ultrasonic guidance to examine morphologic and immunophenotypic changes within individual animals, a technique that should prove useful in elucidating cellular immune responses to viral infection. Immunophenotypically large numbers of CD3 CD8 positive lymphocytes can be found to infiltrate the hepatic parenchyma.

Recently an in vitro primary hepatocyte culture system has been developed that supports the growth of GB virus B (32). This coupled with the availability of infectious molecular clones of the virus and the ability to infect several species of New World primates promises to foster the development of a useful animal model with which to study the pathogenesis of chronic hepatic infections caused by this group of agents (33;34). The potential to create chimeric GB virus B-HCV molecular clones may allow examination of viral determinants of virulence and persistence in a small nonhuman primate model. These studies would be enhanced by the availability of additional molecular tools such as microarrays and quantitative PCR assays to investigate host immune responses.

These are only several of the many examples of how common marmosets have impacted research on infectious diseases. Availability of genomic sequence would be beneficial to many of these programs and would allow investigators to develop tools to investigate the unique susceptibility of marmosets to many infectious diseases and to examine the complex nature of the host’s immune responses.

Neuroscience. Marmosets have been used widely in neuroscience research programs in models of cerebral vascular disease, tardive dyskinesia, multiple sclerosis (MS) and neurodegenerative diseases such as Parkinson’s and Huntington’s disease (table 2) (35-37). In addition to use in models of human disease, marmosets have figured prominently in studies of normal neurophysiology (38-40).
Experimental allergic encephalitis of multiple sclerosis. The induction of experimental allergic encephalitis (EAE) in common marmosets has been used extensively as a model of human MS (41). MS is an important chronic disabling neurologic disorder of young adults with a prevalence of 1 in 1,000. The etiology is unknown but genetic and environmental factors have been indicated in its pathogenesis. While an infectious agent may have an initiating role, a number of autoantigens are also involved.

In common marmosets, EAE may be induced through injection of whole myelin, myelin extracts in complete Freund’s adjuvant (CFA) and human recombinant myelin oligodendrocyte glycoprotein (MOG) in CFA (42). Clinical signs more closely resemble MS than in rodent or other primate models and may be classified in such animals as having a relapsing-remitting and primary progressive course. The lesions are initially characterized by a multifocal nonsuppurative inflammatory cell infiltrate with differing degrees of demyelination and axonal degeneration (43). A number of measured outcomes have been developed and used in the EAE model including neurologic impairment scores, magnetic resonance imaging scores and histologic examination (44;45). Magnetic resonance imaging can confirm the multifocal and progressive nature of the lesion, a characteristic feature of MS in man (46).

Marmosets routinely give birth to genetically non-identical twins or triplets and placental anastomosis results in full bone marrow chimerism of littermates. Thus an alloimmune response is absent from fraternal twins following transplantation of tissue or marrow elements. This unique biologic feature of marmosets has been exploited to examine the role of antigen specific T cell subsets in disease pathogenesis. Adoptive and passive transfer experiments have shown that T cells and pathogenic antibody are required for disease recapitulation and the importance of myelin basic protein specific CD4 T cells in disease pathogenesis (47). Epitope mapping experiments have been undertaken with recombinant MOG peptides and antigenic determinants responsible for disease induction have been identified. Residues within the extracellular domain of MOG elicit disease producing T and B cell responses in the context of the major histocompatibility allele Caja-DRB*W1201 (48). More recently the EAE model has been used in the preclinical evaluation of novel immunotherapies such as testing new anti-inflammatory agents, tolerance based therapies and costimulated target therapy (49;50).

Marmoset models of Parkinson’s disease. Parkinson’s disease is a common neurodegenerative disorder affecting 1.5 million Americans. Progressive dysfunction of the nigrostriatal dopaminergic pathway results in striatal dopamine deficiency and the development of clinical signs including tremors, rigidity and bradykinesia. While the etiology of Parkinson’s disease is unknown, environmental and genetic factors may play a role with oxidative stress and mitochondrial dysfunction as common pathways.

Two methods of disease induction have been utilized in common marmosets. In the first, stereotaxic injections of 6-hydroxydopamine (6-OHDA) into the nigrostriatal bundle is performed. 6-OHDA is taken up directly by dopaminergic neurons resulting in their destruction in an anatomically localized fashion. In the second, 1-methyl-4-phenyl-
1,2,3,6 tetrahydropyridine (MPTP) is administered as a single or repeated dose and is metabolized to its toxic constituents before acting on target neurons within the central nervous system. Rodent models including 6-OHDA and MPTP treated rats and more recently transgenic mice expressing human alpha-synuclein have been used extensively as animal models of Parkinson’s disease and should be considered as alternative species. However, primate models may be particularly useful to investigate transplantation, gene therapy and biopharmaceutical therapeutic interventions. Sequence data could assist with target selection and validation in new therapeutic strategies for Parkinson’s disease.

**β-amyloid deposition in the brains of marmosets.** Several reports document the occurrence of amyloid plaques in aged marmosets suggesting that the species may represent a novel animal model of Alzheimer’s disease (52;53). An analysis of aged common marmosets greater than 7 years of age revealed a high incidence of β-amyloid deposits in brain tissue. β-amyloid immunoreactive plaques were demonstrated primarily within cortical zones and less frequently within paralimbic areas. Deposits were also found with in cortical and meningeal vessels in the form of an amyloid angiopathy. Staining for thioflavin S revealed the presence of β pleated sheet conformation in compact fibrillar plaques but immunohistochemistry for phosphorylated tau protein failed to demonstrate dystrophic neurites in association with these plaques. Functional deficits have yet to be correlated with these morphologic changes in aged animals. An experimental system has been developed to investigate the effect of reductions in cholinergic activity within the neocortex and hippocampus on amyloid precursor protein. In this model an immunoglobulin saporin conjugate was injected with stereotactic guidance to induce selective loss of cholinergic neurons and their axonal projections to the temporal and frontal cortex. This immunotoxin treatment resulted in an increase in expression of amyloid precursor protein as measured by immunohistochemistry and image analysis and supports a role for cholinergic dysfunction in the pathogenesis of Alzheimer’s disease. In most cases the etiology of Alzheimer’s disease in this species remains unknown and genetic data could be used to understand molecular basis of disease susceptibility.

**Toxicology and drug development.** Particularly in Europe, the marmoset has been used as a non-rodent second species in drug safety assessment and pharmaceutical toxicology (54). Because of the closer phylogenetic relationship of marmosets to man than other second species such as the dog, common marmosets may be more suitable for certain types of pharmacokinetic and toxicologic screening. Biotechnology-derived pharmaceuticals or biopharmaceuticals should be evaluated in a relevant animal species in which the test material is pharmacologically active. In cases in which the mechanism of action is dependant on specific receptors or epitopes, relevant species may be restricted to nonhuman primates and the utilization of nonhuman primates in biopharmaceutical development can be anticipated to increase. The small size of common marmosets may represent an additional advantage as it reduces the quantity of compound to be synthesized and may translate into shorter develop times.

The success of any such program is dependent on a comprehensive understanding of the background pathology within the species to be utilized and an appreciation of species-specific xenometabolic pathways. Recently the molecular and comparative characterization of the cytochrome p450 pathway has been undertaken in marmosets and
several other species and such work should facilitate species selection for toxicological testing (55). However, such comparative studies are in their infancy and would be greatly enhanced by the availability of a complete genomic sequence from this species. Such information would allow investigators to understand differences in drug metabolic pathways and assist in the most appropriate animal model for drug development studies. Furthermore such information would prove invaluable in assessing target validation in animal model selection.

Reproductive toxicology. The teratologic effect of experimental compounds have become an increasingly important concern. The young age of sexual maturity, high reproduction efficiency and the similarities in placentation between humans and marmosets, suggests that the marmoset may make an attractive model to investigate the teratologic potential of xenogenous compounds. In some instances the teratologic effects of compounds may differ substantially between rodent and primate species. Thalidomide administration during early gestation results in specific and dramatic limb defects in primates, an outcome not observed in laboratory rodents such as the rat and mouse. Indeed the marmoset has been used to comprehensively investigate the mechanism of thalidomide teratogenesis (56-59). Experimental work conducted in the marmoset has suggested that metabolites of thalidomide may cause the down-regulation of surface adhesion receptors thereby altering cell to cell and cell to extracellular matrix interactions within the developing limb bud (60).

For practical reasons the marmoset is an attractive model in which to study teratologic effects of experimental compounds. The availability of microarrays to examine early embryogenesis would assist and inform such studies and would provide novel avenues to pursue teratologic studies in a nonhuman primate species. The relevance of marmoset metabolic pathways and relationship to human pathways would be assisted by genomic sequence data.

Reproductive biology. The marmoset has found widespread use in studies on reproductive biology and embryology and has focused on areas of normal reproductive physiology such as the mechanism of luteal regression and regulation of reproductive behavior. Marmosets have also been used to investigate novel methods of immunocontraception and the first primate embryonal stem cells were derived from C. jacchus at the Wisconsin National Primate Research Center (61;62). The common marmoset ovarian cycle lasts approximately 28 days with ovulation occurring around day 10 and shows similarities to human and other nonhuman primates in terms of steroid hormone profiles. Luteolysis may be induced and cycles controlled through intramuscular injection of cloprostenol (63).

Similarities in reproductive biology between humans and marmosets, their small size and breeding capacity in captivity has facilitated the use of C. jacchus to investigate aspects of normal reproductive physiology. Luteal formation and regression is a complex differentiation process that involves both extrinsic and ovarian factors. The common marmoset has been used to examine the effect of luteinizing hormone and follicular stimulating hormone on granulosa cell development and steroidogenesis (64;65). The marmoset has been used to define the mechanisms of estradiol inactivation in the primate endometrium through the differential expression of hydroxysteroid dehydrogenase isotypes (66).
Immunocontraception. Immunocontraception has been proposed as a potential birth control method for humans and other animal species. Investigations have focused on antigens within the zona pellucida as a potential vaccine target as this glycoprotein surrounds the oocyte and plays a key role in both fertilization and early embryonic development. A problematic consequence of zona pellucida vaccines is the induction of immune mediated ovarian dysfunction resulting in a depletion of primordial follicles. Such an outcome is unacceptable for any vaccine slated for human use. The common marmoset has been used to investigate the effect of immunization with zona pellucida antigens on ovarian function and fertility (67;68). Mapping studies conducted in this species has revealed immunodominant epitopes and suggested that multivalent vaccines may be developed which consistently induce infertility without concurrent ovarian pathology (67). Molecular tools produced with knowledge of the marmoset genome could support these studies by enabling investigators to examine regulation of gene expression during oocyte maturation. A better understanding of MHC organization could assist in the design of these experiments.

2) Utility of Marmoset Genome Sequence Data to inform human biology.

Gene/Genome annotation. Based on comparative sequence analysis of 3 MB of comparative-grade marmoset BAC sequence from Genbank, we found that human and marmoset genomic sequence differ by 11.6 +/- 0.5 substitutions/100 bp (Kimura two-parameter, see Appendix). Analysis of 4800 of marmoset BAC-end sequence, found that 3842 (>75%) of these end-sequences could be unambiguously mapped to an optimal location against human (Eichler and Zhao in preparation). These data suggest that the majority of genomic sequence can be trivially aligned against the human genome. This is in sharp contrast to mouse-human comparisons where ~50% of genes and genomic sequence showed mutual-best 1:1mapping positions between man and mouse (Clark, 2003). High quality sequence from non-human primates will dramatically improve 1:1 gene/genome annotation. In addition, these data will allow a direct assessment of pseudogenization (Paabo, Gilman, 2003), gene expansion, deletion and neutral genomic divergence. A determination of what genes have been gained and lost during the course of human evolution is of critical biomedical and evolutionary relevance (Olson and Varki, 2002). Marmoset in conjunction with macaque genome sequence will delimit the relative contribution of mutational forces (gene conversion, duplication, rearrangement) and the rates of these events which confound 1:1 mapping of genes. Marmoset provides twice the evolutionary distance (when compared to macaque) from human and thereby greater power for these types of analyses, whilst still allowing high quality genomic sequence alignment of the majority of the DNA.

Ancestral State Determination of Human Segmental Duplication. 5.3% of the human euchromatin exists as blocks of segmental duplication with >90% sequence identity (She, 2004, Bailey, 2002). Although such regions are particularly problematic for whole-genome shotgun sequence, recent analyses have shown that segmental duplications are important for disease, chromosomal rearrangement, evolution and gene innovation among primate species. Analyses of chimpanzee and rhesus macaque genomes show comparable levels of segmental duplication to human (6.5% and 4.5%) respectively (based on multiFISH signals from randomly selected clones). In contrast, analyses of the marmoset showed that only 7/304 (2.3%) randomly selected BAC clones gave signals to
multiple chromosomes by FISH. This is a significantly reduced level of segmental duplication which is comparable to other non-primate mammalian genomes. The marmoset genome, thus, would represent an ideal species for determination of ancestral origin of great-ape and Old World monkey lineage specific expansions of segmental duplication

*Patterns of Non-neutral Selection.* A complete set of genes from humans, chimpanzee, orangutan and two outgroup species (macaque and marmoset) would allow the selection pattern of almost all human genes to be interrogated. Establishment of orthologous relationships among paralogues is virtually impossible among duplicate genes over long evolutionary differences (Dehal, 2001). Unambiguous determination of orthologues is NOT trivial even for trios involving chimpanzee—human and mouse as an outgroup (Only 38% of 7645/20,000 three-way comparisons, for example could be reliably established in recent study (Clark, 2003). Representation at each of the phylogenetic nodes of human divergence from non-human primates provides the required specificity and sensitivity for such studies. In addition, multiple sequence alignment from divergent yet alignable New World and Old-World monkey sequences provides considerable power for the detection of primate-specific regulatory elements (Bofelli, 2003). Such sequences can not be detected by comparative sequence analysis from other mammalian genomes (such as dog, rat and mouse) and require high quality primate genome sequence from multiple nodes in the phylogenetic tree.

**IV. Strategic issues in acquiring new sequence data**

1) **The demand for new sequence data.** The most commonly used marmoset subspecies in research is *Callithrix jacchus jacchus*. The animal is not endangered. Of number of large colonies exist at Academic and Private facilities within the United States including the major research institutions University of Wisconsin, Harvard Medical School, Southwest Foundation for Biomedical Research, Rutgers University, Johns Hopkins University and University of Massachusetts Medical School. Letter’s of support from investigators at each of these institutions are appended illustrating the demand and utility of sequence data to their research objectives. Sequencing of the marmoset genome was identified as a priority objective at the first annual Marmoset research Group of the Americas meeting held a the University of Wisconsin in the summer of 2004.

2) **The suitability for the organism for experimentation.** The common marmoset is a small non-endangered New World primate that is used extensively in biomedical research programs. The animals breed well in captivity and are already kept at a number of leading research institutions in this country. As detailed above these programs encompass diverse areas of biomedical research including infectious diseases, reproductive biology, toxicology and drug development, behavior and neurobiology. Availability of DNA genomic sequence would increase the utility and impact of this species and have an immediate and positive impact on a number of research programs. The small size of the marmoset and ease of breeding and husbandry make it particularly attractive to institutions that do not have preexisting primate programs.

3) **The rationale for the complete sequence of the organism.** There are two broad justifications for complete sequence of the marmoset.
Organismal biology: Most of the biomedical applications (i.e. development of gene expression arrays, genomic mapping of viral latency loci, characterization of immune-related genes, a reference genome for other experimental New World monkeys, etc.) require high quality draft sequence especially for the genic portions of the marmoset genome. Detailed analysis of a working draft version of the chimpanzee genome reveals that a 3.5 fold coverage in whole-genome shotgun sequence is insufficient for this purpose. At this level of coverage, less than 50% of the genes are complete (i.e. missing exons) and a gap is encountered once every 8 kb of sequence (PanGSC, in preparation; Eichler personal communication). This precludes the development of complete gene models for this species which is one of the primary motivations for its genome sequence. Low-level coverage of the marmoset genome, while helpful, would be cost-ineffective in the long-term, requiring non-specialists in genome sequencing to redouble efforts for the most biomedically relevant loci.

Human Genome Annotation: A complete marmoset genome would provide considerable power for a comprehensive annotation of the human genome. First, comparisons between mouse and human genomes have shown that only ~50% of the genes and less than 50% genome sequence can be reliably assigned to 1:1 orthologous groups (Clark, 2003, MGSC v.30), due to lineage-specific gene family expansions, pseudogenization, deletion and neutral sequence divergence. The evolutionary history of this sequence is effectively lost without complete genome sequence from intermediate non-human primate species with sufficient genetic distance. One of the grand challenges identified by NHGRI was to reconstruct the evolutionary history of every base of human DNA (Collins, 2003). This can not be achieved in the absence of high quality whole genome sequence from multiple primates representing discrete timepoints of separation from the human lineage. Marmoset represents a timepoint at 35-40 million years of separation. The genome sequence of the marmoset which shows only 11% divergence would dramatically improve 1:1 orthologous assignments allowing processes of gene-conversion, duplication, transposition and pseudogenization to be reliably modeled. In contrast to the rhesus macaque which shows only 5.9% neutral genomic sequence divergence, the marmoset genome sequence doubles the evolutionary distance while still affording orthologous alignment for the majority of the genome. With respect to paralogous genes, it should be noted that the marmoset genome has ½ the amount of recent segmental duplications when compared to rhesus and human (Eichler, in preparation). From the perspective of segmental duplications, it is, therefore the ideal genome for determination of ancestral state since almost all expansions occurred after its separation from a common primate ancestor. Finally, as a representative of one of the most prolific groups of primates (New World monkeys), the genomic sequence provides an ideal reference genome for comparative genome analyses among primates. Analysis of genomic sequence from 2 New World monkey species in conjunction with an Old-World monkey is sufficient to detect primate-specific regulatory elements by phylogenetic shadowing (Bofelli, 2003). These determinations require a high quality draft genome sequence with limited error and gaps in the sequence.

4) The cost of sequencing the genome and the state of readiness of the organism’s genome for sequencing. A male BAC library has been constructed in Pieter De Jong’s laboratory at the Children’s Hospital Oakland Research Institute (CHORI-259; http://bacpac.chori.org/marmoset259.htm ). Genomic sequence has been generated for
over 200 full-length BAC clones and more than 5000 BAC-end sequences (NISC and TIGR). Both of these analyses show the resource to be of exceptional quality (average insert size ~170 kb). The marmoset material originated from a single male individual from a large pedigree (Suzette Tardif; Southwest national Primate Research Center, San Antonio). Closely related females have been identified for the purpose of WGS library construction.

The estimated size of the marmoset genome is comparable to humans. Karyotype studies show no obvious expansions of heterochromatin (Fleagle, 1999). Analysis of existing genomic sequence from large-insert BAC clones shows comparable levels of repeat content. We, therefore, estimate the euchromatin to be 2.8-3.0 Gb in size. The marmoset has not been a model genetic organism, therefore neither extensive genetic nor physical maps exist. The low-level of neutral sequence divergence (10-11%), however, means that a high quality intermediate physical map could be rapidly generated for the marmoset research community and the genomic sequence community by alignment of paired-end sequence against the human reference.

In general, a six-fold whole genome shotgun sequence assembly is sought (6 × 3 Gb = 18 Gb of high quality raw sequence to be generated in three phases):

Paired-end sequence phase: In order to generate an intermediate high-quality physical map of the marmoset genome, we propose complete BAC-end sequencing of the existing marmoset library (CH259) of 192,000 BACs (12-fold) and an additional 6 fold physical coverage within a marmoset 40kb-insert fosmid library (500,000 fosmids @550 bp per end sequence). While this approach would provide only (761 Mb) 0.25 X of sequence coverage, the combined resource would yield ~18 fold clonal physical coverage of the marmoset genome with a sequence tag once every 5 kb based on alignment to human (estimates 25-30% loss due to unambiguous repeat placement within recent repeats). This would provide a rapid and cost-effective means for not only generating a physical map of the marmoset genome as well as generation of sample sequence for further quality assessment. In addition, clustered paired-end sequence discrepancies with respect to the human genome would flag particularly problematic regions for further characterization and assembly validation including sites of evolutionary rearrangements and lineage-specific duplication. BAC clones representing such problematic regions would be identified at this phase for subsequent BAC working draft sequencing (BAC-phase; see below). The physical coverage (based on paired-end placement) is sufficient to identify sites of recent large segmental duplication (>100 kb).

WGS phase: The bulk of whole-genome shotgun sequence (5.7-6 fold) would be generated via whole-genome shotgun libraries of 2 and 5 kb plasmid inserts. While theoretical calculations predict that 99.9% of the euchromatin would be represented, recent empirical analyses of the human genome suggest that similar sequence coverage effectively achieved only 93% coverage (Istrail, 2004). Nearly half of this loss of euchromatin is due to misassembly of segmental duplications (She, 2004). Since the marmoset genome shows reduced segmental duplication, we estimate that 35 million high quality sequence reads would be sufficient to assemble >96% of the euchromatin (excluding euchromatic portions of the Y chromosome).
BAC-phase. We propose to generate high quality working draft sequence or finished sequence from an additional 500-1000 Mb from large-insert BAC clones. These clones will be identified from either problematic regions of the genome (gene family clusters, segmental duplications, regions of chromosomal rearrangement) and/or areas of biomedical relevance with respect to the marmoset research community. Regions unlikely to be adequately assembled by strict application of WGS will receive priority.

5) Other sources of funding available for sequencing. No other funds are currently available or being sought for this project. The value of this model organism to stroke, neuroscience, pharmacogenomic, reproductive and infectious disease research will undoubtedly translate into broad inter-institutional support at the National Institutes of Health. Supplementary funding from such institutes, however, requires that this organism receive high priority from the NHGRI.

V. Summary

The marmoset is a commonly used New world primate in a variety of biomedical research programs. As described above sequencing of the marmoset genome would provide immediate benefits to investigators utilizing this species and new opportunities for model development. Is information would increase our understanding of human biology as well as lead to potential advances in disease prevention and therapeutic interventions. In addition to these applications to biomedical research models sequencing the common marmoset genome will provide considerable power in annotating the human genome.
Table 1: Examples of infectious disease models utilizing the common marmoset (*Callithrix jacchus*).

<table>
<thead>
<tr>
<th>Agent</th>
<th>Animal model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Viral</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Herpesvirus ateles</em></td>
<td>Acute oncogenesis</td>
<td>(87)</td>
</tr>
<tr>
<td>Herpes simplex virus</td>
<td>Vector safety assessment</td>
<td>(88)</td>
</tr>
<tr>
<td><em>Herpesvirus saimiri</em></td>
<td>Acute oncogenesis</td>
<td>(89)</td>
</tr>
<tr>
<td>Epstein Barr virus</td>
<td>Viral persistence</td>
<td>(90)</td>
</tr>
<tr>
<td>Hepatitis A virus</td>
<td>Acute hepatitis</td>
<td>(91)</td>
</tr>
<tr>
<td>GB virus B</td>
<td>Acute hepatitis</td>
<td>(92)</td>
</tr>
<tr>
<td>Measles virus</td>
<td>Pathogenesis</td>
<td>(93)</td>
</tr>
<tr>
<td>Junin virus</td>
<td>Hemorrhagic fever</td>
<td>(94)</td>
</tr>
<tr>
<td><strong>Parasitic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Brugia malayi</em></td>
<td>filariasis</td>
<td>(95)</td>
</tr>
<tr>
<td><em>Plasmodium sp.</em></td>
<td>malaria</td>
<td>(96)</td>
</tr>
<tr>
<td><strong>Bacterial</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mycoplasma genitalium</em></td>
<td>Urogenital infection</td>
<td>(97)</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bovine spongiform encephalopathy (BSE)/Scrapie</td>
<td>Prion disease</td>
<td>(98)</td>
</tr>
<tr>
<td>Disease</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Degenerative</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>1-methyl-4-phenyl-1,2,3,6 tetrahydropyridine administration (99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6-hydroxydopamine administration (100)</td>
<td></td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>Striatal transplantation (101;102)</td>
<td></td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>Spontaneous amyloid deposition (53)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental induction (103)</td>
<td></td>
</tr>
<tr>
<td><strong>Behavior</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear and anxiety</td>
<td>Conditioned and unconditioned response models (104)</td>
<td></td>
</tr>
<tr>
<td>Reproductive behavior</td>
<td>Social and endocrine influence on reproduction (105;106)</td>
<td></td>
</tr>
<tr>
<td>Child behavior and</td>
<td>Early deprivation (107;108)</td>
<td></td>
</tr>
<tr>
<td>development</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>Cerebral vascular occlusion (36)</td>
<td></td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>Experimental allergic encephalitis (109;110)</td>
<td></td>
</tr>
</tbody>
</table>
Reference List


25


### Appendix I

**Marmoset Research Group of Americas**  
**University of Wisconsin-Madison**  
**June 13-14th, 2004**

<table>
<thead>
<tr>
<th>Attendee</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott, David</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Almond, Roz</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Austad, Steve</td>
<td>University of Idaho</td>
</tr>
<tr>
<td>Bales, Karen</td>
<td>University of Illinois at Chicago</td>
</tr>
<tr>
<td>Benton, Charles</td>
<td>Harlan Teklad</td>
</tr>
<tr>
<td>Bertram, Rick</td>
<td>Glaxo Smith Kline</td>
</tr>
<tr>
<td>Biswas, Subhabrata</td>
<td>University of Massachusetts Medical School</td>
</tr>
<tr>
<td>Boe, Carla</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Bolton, Iris</td>
<td>Pfizer, Inc.</td>
</tr>
<tr>
<td>Brackee, Gordon</td>
<td>University of Texas SW</td>
</tr>
<tr>
<td>Burdett, Eric</td>
<td>University of Utah</td>
</tr>
<tr>
<td>Burkholder, Tanya</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>Butler, Ray</td>
<td>Novartis Pharmaceuticals</td>
</tr>
<tr>
<td>Cronin, Katherine</td>
<td>Lincoln Park Zoo</td>
</tr>
<tr>
<td>Curlee, Joseph</td>
<td>Harlan-Sprague Dawley, Inc.</td>
</tr>
<tr>
<td>Emerson, Carol</td>
<td>Lovelace Respiratory Research Institute</td>
</tr>
<tr>
<td>Estrada, Alejandro</td>
<td>University of Mexico</td>
</tr>
<tr>
<td>Ferrell, Thomas</td>
<td>Calvert Laboratories, Inc.</td>
</tr>
<tr>
<td>Ferris, Craig</td>
<td>University of Massachusetts Medical School</td>
</tr>
<tr>
<td>French, Jeff</td>
<td>University of Nebraska at Omaha</td>
</tr>
<tr>
<td>Genain, Claude</td>
<td>University of California, San Francisco</td>
</tr>
<tr>
<td>Ginther, Anita</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td>Hadzic, Zezira</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Humle, Tatyana</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Juame, Juan</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Jensen, Heather</td>
<td>University of Nebraska at Omaha</td>
</tr>
<tr>
<td>Jordan, Kay</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>Kurian, Aimee</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Layne, Donna</td>
<td>Southwest National Primate Research Center</td>
</tr>
<tr>
<td>Ludlage, Elisabeth</td>
<td>Harvard University, Harvard Medical School, NEPRC</td>
</tr>
<tr>
<td>MacGill, Tracy</td>
<td>Center for Drug Evaluation &amp; Research, FDA</td>
</tr>
<tr>
<td>Mansfield, Keith</td>
<td>Harvard University, Harvard Medical School, NEPRC</td>
</tr>
<tr>
<td>Maxwell, Heather</td>
<td>University of Nebraska at Omaha</td>
</tr>
<tr>
<td>Moloo, Badru</td>
<td>University Health Network</td>
</tr>
<tr>
<td>Moscovice, Lisa</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Nadon, Nancy</td>
<td>National Institute on Aging</td>
</tr>
<tr>
<td>Newman, John</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>Neilsen, Ronald</td>
<td>University of Utah</td>
</tr>
</tbody>
</table>
Power, Michael Smithsonian Institution
Power, Rachel Transgenics
Reeb, Jamie Harlan-Sprague Dawley, Inc.
Reilly, Sheila Novartis Pharmaceuticals
Roberts, Lucille National Institutes of Health
Rolf, Lester University of Pennsylvania
Ross, Corinna University of Nebraska-Lincoln
Rukstalis, Michael University of Nebraska at Omaha
Schultz-Darken, Nancy University of Wisconsin-Madison
Scott, Jill University of Wisconsin-Madison
Siani, Jennifer University of Maryland
Smucny, Darlene Southwest National Primate Research Center
Snowdon, Charles University of Wisconsin-Madison
Sousa, Maria Bernardete Universidade Federal do Rio Grande do Norte
Sramek, Mary Abbott Laboratories
Tardif, Suzette Southwest National Primate Research Center
Thorsen, Peter University of California, San Francisco
Van Belle, Sari University of Wisconsin-Madison
Wang, Xiaoqin Johns Hopkins University
Yamamoto, Marie Universidade Federal do Rio Grande do Norte
Zahed, Sophia University of Wisconsin-Madison
Zemba, Lindsay National Institutes of Health
Ziegler, Toni University of Wisconsin-Madison
Appendix II

Genetic distance between humans and non-human primates. Based on the optimal global alignment of available non-human primate genomic sequence from BAC clones aligned against the human genome sequence. The substitutions per site were determined using the Kimura 2-parameter model over 3 kb alignment windows. The mean and standard deviation are shown for chimpanzee, baboon (a representative Old World monkey), marmoset (a representative New World monkey) and Lemur (a representative prosimian) (Eichler, EE, in preparation. Not for public dissemination).