V12242009

Genome and other Sequence Information from Primate Models of HIV Infection

Jeffrey Rogers[1] , Carlos Bustamente[2], Richard, Gibbs[1], R. Paul Johnson[3], Welkin Johnson[3], Devin Locke[4] , Keith Mansfield[3], David O'Connor[5], Guido Silvestri[6], Ron Veazey[7]

1. Baylor College of Medicine, Baylor, TX. 2. Cornell University, Ithaca NY. 3. New England Primate Research Center, Harvard Medical School, Southborough, MA. 4. Washington University Medical School, St. Louis, MO. 5. University of Wisconsin, Madison, WI. 6. University of Pennsylvania, Philadelphia, PA. and The Yerkes National Primate Research Center, Atlanta, GA 7. Tulane University, New Orleans, LA

1. Introduction

"For this field to make progress, we need to become evolutionists and learn more about natural SIV hosts. We need to think about ways to harness the human immune system to adapt to the virus in the same way." (Dr. Warner Greene, Chair - Final Remarks of the NIAID Summit on AIDS Vaccines, Bethesda, March 25th 2008, http://www3.niaid.nih.gov/news/events/summitHIVVaccine.htm)

HIV infection, originally described in 1981, has significantly changed the way we view infectious disease. Prior to HIV-1, many believed that most infectious diseases would eventually be eradicated, either by antibiotics (bacterial infections) or vaccination (viruses). Unfortunately, both have proved false, as bacterial infections have demonstrated an unexpected and rapidly evolving resistance to antibiotics, and HIV infection has subverted all attempts at a vaccine. Unlike other viruses, HIV has an affinity for the very cells that initiate immune responses (see below) as well as an ability to mutate at rates much higher than any other virus. In fact, it has been estimated that intrapatient diversity in an HIV-1 infection eclipses the global diversity of seasonal influenza strains, which as we all know requires at least yearly adjustments for effective vaccination. These factors have thus far frustrated all attempts to

develop an effective method for preventing infection by HIV, and today there are 34 million people infected with HIV, with approximately 2.7 million new infections occurring each year.

However, nonhuman primates hold much promise for deciphering the host immune responses associated with resistance to disease, and perhaps even a cure or reliable vaccine against HIV infection. Among the nonhuman primates, a number of African species are able to live with infections of SIV and related viruses without progressing to disease, demonstrating that effective immune control is indeed possible. As a result of concerted efforts by a large research community, we now have a reasonably detailed understanding of molecular details of viral lifecycle. This has produced a series of effective antiviral drugs that are valuable in treating infection and improving both survivorship and quality of life.

Comparatively less is known about what will constitute an effective host immune response to HIV infection, and there is an increased urgency to understand host response mechanisms in order to inform efforts to create an effective anti-HIV vaccine. The recent failures of large-scale vaccine trials have highlighted the fact that there is no obvious path forward in terms of effective vaccine-based prevention of infection, or therapies that lead to permanent elimination of infection. It is clear that several aspects of the host response are relevant to understanding resistance to disease, including resistance or sensitivity to infection via various routes (e.g. parenteral, mucosal, mother-to-child), acute immune response, viral setpoint and others. Extensive research in human populations has demonstrated that genetic differences among people can influence their ability to resist infection or to delay progression to pathology and disease (Singh et al. 2008). Specific gene loci (i.e. CCR5, RANTES, CCR2, HLA-B, HLA-C) exhibit allelic variation within human populations that affects the ability of any given human host to resist infection or respond to viral exposure.

Nonhuman primate species provide outstanding animal models for the study of cellular mechanisms of viral infection and subsequent progression toward pathology. Nonhuman primates also provide valuable model systems for the study of host factors that affect resistance to infection and disease. In fact, HIV-1 and HIV-2 are both primate lentiviruses, a group that includes more than 30 viruses endemic to a wide variety of African nonhuman primates (Figure 1). Both viruses are the result of zoonotic transmissions of simian immunodeficiency viruses from apes (HIV-1) and old world monkeys (HIV-2). Seroprevalence studies suggest that humans may be frequently exposed to a variety of primate retroviruses, such that nonhuman primate populations represent a potential pool for future zoonoses.

Extensive study of the various SIVs and their effects on various nonhuman primates has shown that a number of African species are natural hosts, and tolerate lifelong infection without obvious compromise to health. There are also other closely related species that may not carry their own natural SIV infections, but are nevertheless also resistant to disease induced by SIV. However, a large set of nonhuman primates, especially macaques of the genus *Macaca*, exhibit various degrees of susceptibility to infection. Some can be readily infected with specific SIV viral isolates and succumb to disease quickly. Other macaques are somewhat more resistant but not entirely protected. Furthermore, within particular species some individuals will succumb to AIDS-like disease more slowly while others are able to control infection effectively and avoid disease altogether.

The evidence for both consistent species-level differences in response to SIV infection, and within-species variation among individuals in susceptibility and disease progression is now extensive and growing. We are beginning to understand some of the specific factors that influence these individual- and species-level differences, but much more remains to be learned. A detailed understanding of the

relevant host factors that affect resistance to infection as well as progression of disease after infection is certainly important for progress toward an effective vaccine against HIV and improved therapies that will delay or prevent the pathogenic symptoms of infection.

NHGRI has undertaken genome sequencing projects for a number of nonhuman primates that are valuable animal models for various types of studies related to HIV and AIDS. But there are four fundamental reasons, detailed within this white paper, for increasing the breadth and depth of these DNA sequencing efforts related to AIDS research by sequencing additional species and by developing more extensive catalogs of genetic variation within species:

a) Natural Hosts: Non-pathogenic, natural host infections essentially uncouple disease progression from viral infection. Whole genome DNA sequences for a series of natural hosts for SIV will allow researchers to dramatically expand their investigation of the host genetic factors that influence resistance or susceptibility to SIV infection and allow species to tolerate viral infection without progressing to AIDS

b) Non-Natural Hosts: whole genome DNA sequences for non-natural hosts (i.e. hosts in which infection results in pathogenic outcome) that can be used to pursue molecular study of infection and pathogenesis, including detailed genetic analyses of the host factors that influence inter-individual differences in susceptibility to infection or host immunological response are strongly justified. On-going studies in human populations are identifying genetic polymorphisms that influence susceptibility and/or disease progression. Once complete genomic sequences are available for nonhuman primate species that show individual variation in susceptibility, then

experiments can be designed in which specific hypotheses concerning the genetic mechanisms that alter risk of infection or protection against disease can be tested in controlled experiments. Similarly, comparisons of DNA sequences from natural and non-natural hosts will generate hypotheses concerning the potential influence of specific genes or pathways, and these hypotheses can be tested in prospective or retrospective studies of human cohorts. This will involve SNP variation in specific genes as well as copy number variation affecting genes within the immune system.

c) Expanded Analysis of Comparative Primate Immunology: comparative analysis of genome sequences will provide the essential catalog of gene content and gene sequences that is necessary to begin comparative analysis of the molecular architecture and functional organization of the immune systems in natural hosts versus non-natural hosts (susceptible species). While individual gene studies can identify and examine specific genes that alter risk of infection or disease, the larger goal of a complete understanding of the factors affecting AIDS will be best served by development of a broader gene network and systems biological approach that integrates information about a large number of loci that must interact to generate effective immune function and capacity. Whole genome sequences for natural and non-natural host species will lay the foundation for comprehensive analyses of primate immune function and the diversity of immunological systems that have evolved in distinct evolutionary lineages of primates.

d) Improved Tools and Resources for SIV Research: development of genomic sequence data will allow various types of manipulations of nonhuman primate experimental animals that will

facilitate progress in HIV/SIV research. Such manipulations will include selective breeding programs to generate sets of individuals with particularly informative multi-locus genotypes, further studies of RNAi and other cellular approaches to investigate host-virus interaction and potential preventive or therapeutic strategies.

While our understanding of primate natural hosts and susceptible species is not yet adequate, it is clear that each of these primate species constitutes an individual natural experiment in host-virus relationship and interaction. Different African primates use different mechanisms to achieve their tolerance of or resistance to SIV infection. Various closely related Asian primates exhibit different patterns of susceptibility, post-infection viral loads and eventual progression to an AIDS-like disease. This natural diversity in primate immunological response to infection can provide critical information that will benefit efforts to develop vaccines and therapies for humans. An expanded program in genomic and other sequencing will provide several types of novel and potentially remarkable data for the global effort to combat this epidemic.

This white paper is organized into the following sections:

1. Introduction
2. General rationales: target species and their phenotypes
    - Species-specific differences in disease progression—natural hosts of SIV
    - Within-species differences in disease progression---non-natural hosts
    - Additional chimpanzee sequencing

3. Additional rationales for sequencing
4. Sequencing strategy and study design
5. Appendices describing specific priorities

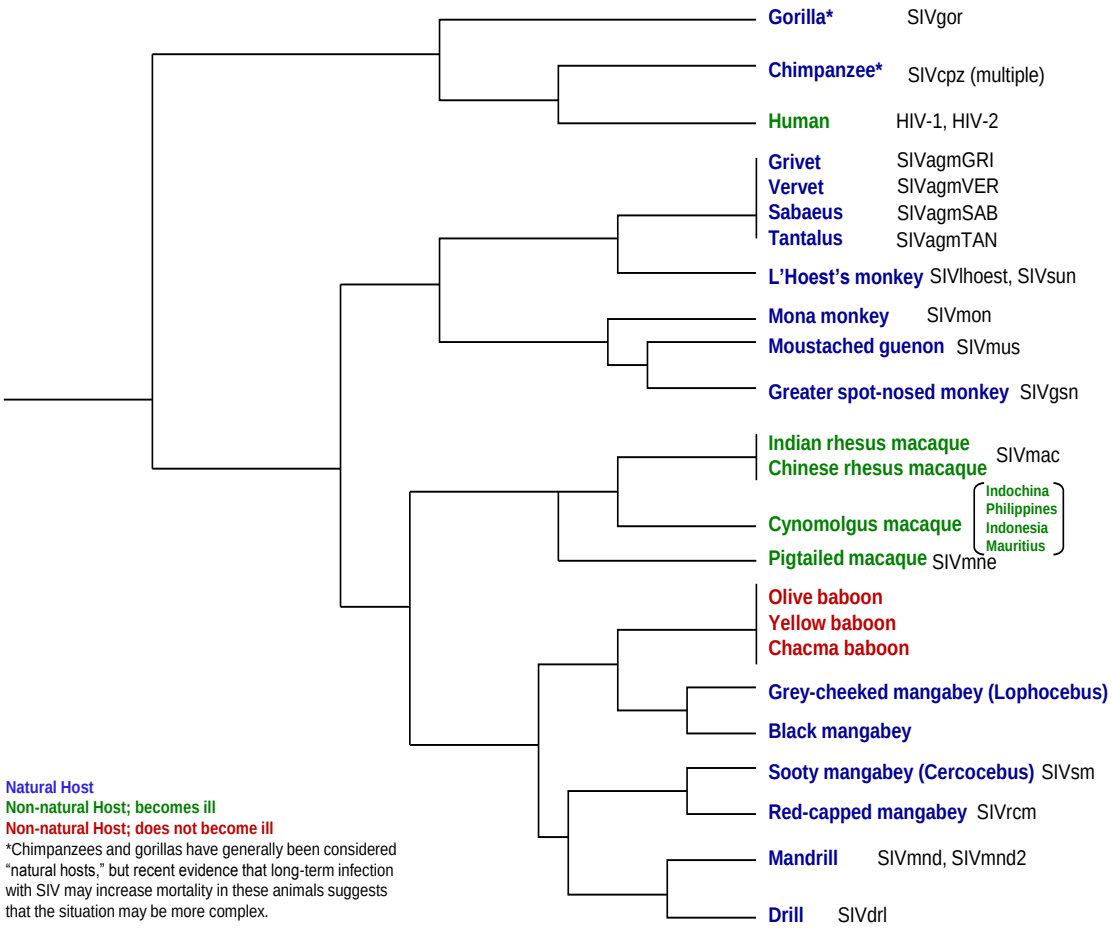

Figure 1. Phylogenetic relationships among species, and associated SIV strains discussed in this white paper. Natural hosts are shown in blue, non-natural hosts that develop simian AIDS are shown in green, and non-natural hosts that do not develop simian AIDS are in red.

2. General Rationales: Target Species and their phenotypes

2.1 Natural Hosts

A number of African primates are naturally infected with lentiviruses, the group of viruses that includes HIV and various SIVs. These "natural hosts," which include mandrills, sooty mangabeys and African green monkeys (among others), all resist progression to disease (AIDS) when infected with SIV. In sharp contrast, experimental SIV infection of macaques results in clinical illness (simian AIDS) similar to that observed in human HIV-infected individuals progressing to AIDS. The differences between natural hosts and closely related species that do suffer from disease after infection are of high interest. First, SIV infection in these non-progressing natural hosts results in persistent infection, and they maintain high levels of plasma viremia, equal to or higher than HIV-infected humans (and usually for the remainder of their natural lifespans), yet they do not develop AIDS. Moreover, natural hosts are immune from vertical (mother-to-child) transmission despite having high levels of viremia in mothers and in breastmilk (Gordon et al. 2005; Silvestri et al. 2007). Importantly, natural hosts have NOT evolved to control viral replication (reviewed in Silvestri et al., 2007), and it appears that non-progressing hosts naturally have fewer viral target cells (CD4+CCR5+ T cells) compared to progressing hosts (macaques and humans). These differences seem likely to be a consequence of long-term selection, non-progressors having been infected with these viruses for millennia, rather than a consequence of developing effective and/or standard "immune responses" to defeat this virus. The inability of mammalian immune responses to control a retrovirus that selectively infects and destroys the very cells responsible for initiating and regulating immune responses (CD4+ T cells) has selected instead for hosts that rely less on these cells for regulating immune responses. In essence, these non-progressing species have successfully uncoupled viral replication from disease.

The exact events that make SIVs non-pathogenic in their natural hosts, while inducing immunodeficiency and consequent disease in non-natural hosts (including humans) are still largely unknown. Detailed comparison of the immune systems of natural hosts versus susceptible species is a growing area of research. There are significant differences in the immune phenotypes between natural and non-natural hosts, and among different natural hosts. Observed common features of SIV infections in natural hosts include: (i) general lack of disease progression; (ii) lack of vertical transmission; (iii) high levels of virus replication throughout infection; (iv) preservation of peripheral CD4+ T cell homeostasis but depletion of mucosal CD4+ T cells; (v) low immune activation during chronic infection; (vi) reduced expression of CCR5 on CD4+ T cells; (vii) short *in vivo* lifespan of productively infected cells, suggesting *in vivo* virus cytopathicity (Silvestri et al. 2007; Sodora et al. 2009). Species-specific differences include: (i) lower baseline levels of CD4+ T cells in African green monkeys with increased fraction of unusual T cells (i.e., CD3+CD4+CD8+ and CD3+CD4-CD8-); (ii) presence of CXCR4-tropic viruses in African green monkeys; (iii) defective ICAM-2 signaling in sooty mangabeys; (iv) consistent but moderate peripheral CD4+ T cell depletion in mandrills; and various others (Paiardini et al. 2009; Sodora et al. 2009). The fact that natural SIV hosts show common features as well as species-specific differences emphasizes the importance of collecting as much information as possible for all the species for which there is currently limited data.

One important reason for having genome sequence data from as many of these species as possible is to provide the tools for basic comparative studies of the variety of immune responses, both between natural and non-natural hosts, and among different natural hosts. Each species is likely to present different components of the primate immune system that have changed to adapt to lentiviral infection. It is less

likely that direct sequence comparisons will provide immediate information as to the basis for these phenotypic differences. However, such comparisons may well provide clues that will direct additional experiments, for example by highlighting regions of the genome in immune system genes that have evolved at different rates in the last few million years in natural hosts, compared to non-natural hosts. Primate phylogeny is arguably one of the most robust species phylogenies available, and the phylogenetic relationships of all the species in question are well established (Figure 1); thus, genes that have undergone significant positive selection and/or significant sequence change can be identified because they will show inter-species differences that do not match the anticipated differences predicted from the phylogeny.

Natural SIV hosts, in particular the sooty mangabeys, also show considerable intra-species variability with respect to their response to infection (Dunham et al. 2006; Sumpter et al. 2007) . Preliminary but comprehensive pedigree studies conducted on the large colony of naturally SIV-infected SMs housed at the Yerkes National Primate Research Center in Atlanta suggest that a strong genetic component underlies these phenotypic differences in terms of response to SIV infection. We predict that the identification of the genetic basis for the inter-species (i.e., differences in orthologous genes) and intra-species (i.e., genetic polymorphisms) variability may provide useful clues to the mechanisms that ultimately determine resistance to AIDS. In addition, we hypothesize that inter- or intra-species differences in the course of SIV infection in natural hosts may be related to specific differences in the expression profile of specific cell types or tissues, and particularly those involved in the immune response to these viruses.

Comparisons of whole genome sequences across species that differ in immune response may provide clues to the underlying causative molecular differences. However, in the case of within-species variation in immune response, the methods for identifying functionally significant genes are more established--- a reference sequence plus a SNP resource will facilitate association studies for these interesting phenotypes. With access to whole genome sequences and large numbers of SNPs for the appropriate primate species, it will be possible to use either genetic linkage in pedigrees or association studies in unrelated animals to map genes that influence variation in viral replication, immune function and response to infection. This approach will supplement and extend similar analyses underway in humans, and coupled with the greater experimental control afforded by primate SIV models, will constitute an invaluable complement to human epidemiological studies.

In this context, we believe that studies of the genetic and genomic make-up of natural SIV hosts are key components of a multi-pronged approach to expand our understanding of how these animals have evolved to become resistant to AIDS. We propose three basic approaches: 1) Full genome sequence of a relatively large number of natural SIV hosts (See Appendices). The highest priority is for the African green monkey (already approved for sequencing) and the sooty mangabey due to their availability and because SIV infection has been very well studied in these organisms. 2) Variation (SNP) data for those species, i.e., sooty mangabeys and African green monkeys, for which significant phenotypic information is available, to support the discovery of gene polymorphisms that may be associated with individual variability in the course of infection; and, in the same two species; 3) Broad transcriptional surveys of multiple tissues to allow for the correlation of expression data with comparative results and allow analysis of gene expression differences between species, or variation within populations of a species. Additional details are provided in Section 4: **Sequencing Strategies and Study Design.**

2.2 Non-Natural hosts

The non-natural hosts discussed here fall into two classes: those that get ill after infection with a strain of virus from another primate (in this white paper, the macaques), and those that do not get ill (here, some species of baboon). Both are of interest. The non-natural hosts that get ill are the most direct models for human HIV infection. In studies of HIV in humans, the specific strain of the infecting virus or exact timing of infection is rarely known, and confounding variables such as behavior, nutrition, drug abuse, potential for co-infections, and treatment and/or compliance issues all limit our ability to study or interpret the pathogenesis of HIV infection. This is especially true for long-term non-progressors (LTNP). However, experimental studies of nonhuman primates carefully control for and reduce or eliminate these variables. For example, primates can be inoculated with known viruses (even molecular viral clones) and the timing, dose, and route of infection are precisely known. Using these models, we have made remarkable progress in our understanding of HIV pathogenesis. However, little progress has been made in understanding correlates of resistance to infection or control of HIV replication in patients, largely because of the aforementioned difficulties, and because the LTNP state in humans is very rare.

Remarkably, the various nonhuman primate models of infection demonstrate a marked variation in their response to disease between species, and even among different subspecies. It is especially important that even among individual animals of the same species or subspecies inoculated with identical viruses, there is often variation in immunologic control between individuals. Recent human and macaque studies have indicated correlations between LTNP and specific MHC alleles, but to date, these genotypes are not predictive of viral control, and immunologic correlates of viral control remain

unknown (Goulder and Watkins 2008) Nevertheless, the NHP data currently available suggest that host differences are indeed responsible for control of HIV infection, and that complete genotyping of macaques of different species that demonstrate differing levels of control of retroviruses may lead to better understanding of the basis of viral control, and eventually to better therapies and immunologic strategies to prevent and cure HIV infection in humans.

Macaques

Macaques (genus *Macaca*) are a broad and diverse yet closely-related group of 20 nonhuman primate species that exhibit remarkable diversity in their responses to experimental infection with SIV. SIVsm from a natural host (sooty mangabey), the same virus that gave rise to HIV-2 in humans, somehow invaded captive macaque colonies in the 1960's, emerging over the next decade as outbreaks of disease retrospectively recognized as AIDS (after identification of HIV-1 and the first SIVs in the 1980s). Thus, the emergence of SIV in macaques (SIVmac) bears striking paralleles to the emergence of HIV-2 in humans over about the same timeframe. To date, no macaque species has been documented to be "naturally" infected with SIV or a related virus in their native habitat. Thus, like the original HIV infected human patients originally described in 1981, experimentally infected macaques are "naïve" hosts.

No other animal model recapitulates the cellular, molecular and even the intrinsic pathogenesis of HIV infection in humans as well as the rhesus macaque model of SIV infection. The fact that these viruses are so closely related (estimated between 60-80% homology) as well as the fact that both result in virtually identical disease courses (including interindividual variation in susceptibility to infection and

disease progression, similarly rapid and selective depletion of mucosal CD4+ memory cells, progressive decline of blood CD4+ T cells, and an essentially identical spectrum of opportunistic infections) indicates that this is indeed the animal model of choice (Baroncelli et al. 2008; Lackner and Veazey 2007). The primary reason is that like humans, macaques are "naïve" to this immunotropic virus that selectively infects and destroys memory CD4+ T cells, which in these species, are essential for promoting and maintaining effective immune responses to these selective "opportunistic" infections. However, not all "macaques" respond similarly to HIV/SIV infection (Baroncelli et al. 2008), and these differences are an important source of fundamental information about the host response to lentiviral infection. These between-species differences actually make this genus uniquely valuable for critical investigation.

Three species of macaque are justified here for sequencing by virtue of their differential responses to SIV infection; these are 1) pigtail macaques (*M.nemestrina*), which are clearly more susceptible to SIV transmission and disease progression, 2) Chinese rhesus macaques (*M. mulatta*), because susceptibility to SIV infection differs in several respects from Indian origin animals of the same species, and 3) cynomolgus macaques (*M.fascicularis*), which appear to be more resistant to infection and the development of AIDS following experimental infection with SIV or HIV. Detailed justification for these species is included below and in the Appendices.

In addition to between-species phenotypic differences, macaques display significant differences in susceptibility to infection, disease progression and immune responses between subspecies and, like human HIV/AIDS cohorts, between individuals. Such variation can be exploited using genomic methods to identify host loci that contribute to variability and to pinpoint the genetic correlates of

susceptibility/resistance to disease . In particular, this motivates acquisition of SNP data for both natural hosts and non-natural hosts that display such phenotypic variation within species.

Within the broad species-level categories, there are geographic subpopulations of pigtailed, rhesus, and cynomolgus macaques that exhibit heterogeneity in their susceptibility to SIV. The best data demonstrating the importance of these subpopulation differences comes from studies of Indian (In-Rh) and Chinese (Ch-Rh) origin rhesus macaques. When inoculated with SIVmac, Ch-Rh demonstrate a much higher rate of viral control, and many (approximately 1/3) eventually control plasma viremia to undetectable levels and become long-term non-progressors (LTNP). Interestingly however, the majority of Ch-Rh maintain high plasma viremia and progress to AIDS in a manner almost indistinguishable from In-Rh or HIV-infected humans (Degenhardt et al. 2009; Trichel et al. 2002). This is somewhat perplexing since both subspecies are equally susceptible to vaginal SIV/SHIV transmission, and have similar peak plasma viral loads in early infection, suggesting that the resistance of Ch-Rh is an adaptive, rather than an innate immune response. Moreover, since 2/3 Ch-Rh progress to disease and 1/3 become LTNP, this also tells us that these differences are host, and not virus related. Combined, these studies provide the strongest evidence that host, and not viral factors, are associated with resistance to viral replication and progression to disease in macaques, and that comparative genotyping of Ch and In origin rhesus macaques, and in particular, comparing genes between progressors and non-progressors, will eventually delineate the genes and host factors responsible for resistance to AIDS.

Similarly, there are subpopulations of cynomolgus macaques from Indonesia, the Philippines, Vietnam, and Mauritius. These subpopulations, like In-Rh and Ch-Rh, have diverse major histocompatibility complex (MHC) genetics. For research purposes, the Mauritian cynomolgus macaque (MCM) has

particularly attractive genetic attributes. Unlike In-Rh and Ch-Rh, MCM descend from a very small founder population (less than 10 individuals) that colonized the island within the last 500 years. MCM are thus much more genetically homogeneous, an advantage when studying immune responses against an unpredictable pathogen like SIV.

Investigations have already begun to define some of the genetic differences that influence these immunological differences among individuals within species. Degenhardt et al. (2008) examined individual variation in the number of copies of the CCL3L locus in 57 Indian-origin and Chinese-origin rhesus macaques that had all been experimentally infected with SIV_{mac}. They found that copy number variation in CCL3L accounted for 18% of the variance in the time until animals were diagnosed with simian AIDS. A lower number of copies of CCL3L was associated with more rapid progression of disease. Indian-origin rhesus macaques have, on average, fewer copies of this gene, and thus this locus can explain some proportion of the observed difference in resistance to disease that has been described for Indian-origin versus Chinese-origin rhesus monkeys. More complete genome sequence data for various species and subspecies will undoubtedly produce additional discoveries of this type (the whole genome sequence of M. mulatta is derived from a single Indian-origin rhesus macaque, with no data from Chinese-origin animals).

The sequence data from these species will be highly complementary to ongoing studies in human HIV/AIDS cohorts that have uncovered potential associations between human allelic variants and susceptibility to HIV-1 infection and/or markers of disease progression (Singh et al. 2008). Such results are generally impossible to confirm experimentally, and the studies are thus susceptible to spurious associations. The analysis in macaque AIDS models of variation in genetic loci that affect disease

progression will complement human studies in three important ways. First, because the underlying spectrum of random variation across all loci is different in humans and macaques, similar positive correlations between a particular gene and SIV/SHIV infection in macaques constitutes independent evidence supporting a role for that gene/locus. Second, because the frequencies of variants with functional consequences will, by chance, be distributed differently among AIDS defining loci in humans and macaques, some loci may be more amenable to tests of association in macaques than humans (a possibility that may be enhanced by overall greater diversity among macaque species). And third, established associations/correlations can be used to generate specific hypotheses that can be tested experimentally in macaque models using prospectively typed cohorts or samples, an avenue that is impossible with human patients. In all cases, the starting point is a substantial discovery phase in which common variants at a locus are catalogued in order to determine allele frequencies and identify appropriate genotyping screens.

Finally, it is worth pointing out that while SIV-infected macaques are the most robust and widely used animal model for HIV/AIDS, the fact remains that no true animal model of HIV-1 infection and disease exists. However, recent identification of species-specific genetic barriers to HIV-1 infection have allowed significant progress in the rational design of so-called "simian-tropic" HIV-1 (stHIV) (Ambrose *et al.* 2007; Hatziioannou *et al.* 2006, 2009; Kamada *et al*. 2006). This work is a striking example of genetic discoveries being put to practical application. The potential for a true animal model of HIV infection is extremely exciting, and it is clear that understanding of interspecies genetic differences has permitted progress (Neil *et al*. 2008; Sheehy *et al*. 2002; Stremalu *et a.l*, 2004). Such models would permit direct testing of HIV-1 specific antiviral inhibitors, direct evaluation of HIV-derived immunogens (instead of SIV), and the impact of host genetic variation on both could also be evaluated. Progress in this area would be greatly enhanced by data and tools that accelerate discovery of

the remaining blocks to HIV-1 infection in macaques and, once replication in macaques is achieved, such tools would also be applied to understanding the impact of host genes on stHIV infection and disease.

Baboons

We also propose sequencing of several baboon species. As detailed in the appendices, these non-natural hosts do not carry native lentiviruses, but can be infected with SIV strains from other NHPs without getting ill. Thus they represent a third type of host-SIV interaction that differs from that of humans and macaques (pathogenic outcome), and the naturally-adapted hosts (endemic infection with nonpathogenic outcome).

Considerations for sequencing the proposed non-natural, "pathogenic" hosts are presented in Section 4, below. Because Indian-origin rhesus and cynomolgus macaques are already sequenced or approved for sequencing, in the context of this white paper we request that efforts focus on ensuring that assembled genomes are of high enough quality for analysis and that sufficient information regarding SNPs and other sources of variation (e.g., CNV) is available for uncovering the genetic basis for observed phenotypic differences.

2.3 Additional chimpanzee sequencing

We take the opportunity of this white paper to propose additional chimpanzee sequencing. Although this genome has been approved for genome sequencing at high quality, recent observations that chimpanzees can develop AIDS following both natural and experimental infection with SIV or HIV underscore the desirability of obtaining additional variation information from several subspecies (see Appendix 1).

3. Additional rationales for sequencing

3.1 Toward a more complete understanding of the diversity of primate immunological systems

Specific analyses of individual genes such as CCL3L or particular MHC Class I genes can determine whether within species polymorphisms or between species differences in these genes influence susceptibility to infection or progression of disease. Significant differences in overall immune system function (i.e. multiple variables describing immune function, not just single traits) have been described between natural hosts and non-natural hosts. Lesser but still multi-variate differences among natural hosts are also known. For example, the reduced numbers of viral target cells in natural hosts versus non-natural hosts, or the ability of natural hosts to resist mother-to-offspring transmission despite the susceptibility to infection through other routes, suggest that there are a number of immunological changes involved in the evolution of resistance to disease, not just one or two genetic changes in key molecules (Paiardini et al. 2009; Silvestri et al. 2007; Sodora et al. 2009). Consequently, it is possible that an understanding of the differences between natural and non-natural hosts may depend on a systems biological approach that investigates SIV infection and disease progression at the level of an integrated systemic description of host response. Such a multivariate systemic comparison of immune function in natural vs. non-natural hosts will certainly require whole genome information on gene sequences, gene copy number, gene expression in specific cell types and other forms of analysis that depend on whole

genome DNA sequences. Similarly, this field may eventually move to proteomic analysis of SIV infection, host response and the time course of viral replication or establishment of equilibrium viral load. Proteomic studies of the SIV-AIDS model in natural vs. non-natural hosts will also require whole genome DNA sequences.

3.2 Toward an expanded array of resources for SIV/HIV research

Increasing the power of existing animal models. Recently, Nelson and O'Brien have described a genetic propensity index (GPI), which allows investigators to estimate the effects of multiple genetic polymorphisms across individuals on the results of clinical vaccine trials (Nelson and O'Brien 2006). Given the call for more pre-clinical evaluation of vaccination strategies and intensified investigation of the immune response to SIV/SHIV infection, cataloguing variation in AIDS defining loci in the most commonly used animal model(s) is urgent. In light of the practical and financial limitations on NHP studies, a similar index will greatly improve the power of smaller studies, either by accounting for confounding variation retrospectively or through prospective genotyping and selection of individual animals for enrollment in study groups. It is already common practice to type animals for specific alleles of class I MHC in SIV/SHIV vaccine experiments. As data on other relevant loci accumulate, it will become feasible to type individual animals for critical variants at multiple loci.

Successful identification of the molecular factors and mechanisms that explain the resistance of natural hosts to AIDS-like disease will require a wide range of tools, resources and strategies. Whole genome DNA sequence data for a number of primate species including natural and non-natural hosts of SIV will facilitate the development of resources that will be unique to the primate models. For example, there is now clear evidence that genetic differences among people influence susceptibility to infection and/or

progression to disease. Therefore, in the primate models it may be valuable to generate, through controlled breeding, individual animals with specific multi-locus genotypes at functionally significant genes. This will allow explicit testing of genetic hypotheses concerning host resistance. Furthermore, the complex nature of the phenotypes involved suggests that the most valuable animals for some studies will be individuals with specific multi-locus genotypes covering MHC genes, T-cell receptors, chemokine loci and other types of genes. Whole genome sequencing will allow discovery of SNPs and other polymorphisms in a wide range of potentially significant loci. That will make it possible to establish controlled breeding programs that produce animals with specific desirable genetic constitution.

Genomic sequences for these primate species will also provide the basic information necessary for the development of reagents for RNAi experiments and other approaches designed to target specific genes and test hypotheses regarding their effects on susceptibility to viral infection. Such gene manipulations will be valuable in nonhuman primates because such experiments obviously cannot be conducted in humans. Nor can they be done in non-primate animals that do not respond similarly to lentiviruses. Other classes of genetic loci beyond protein coding genes may also be important in future studies. Whole genome sequencing of natural and non-natural hosts will provide the necessary background that facilitates identification and intensive study of microRNAs, other functional RNAs and other types of functional genomic elements.

4. Sequencing Strategies and Study Design

We describe the types of sequencing product that will be required for obtaining maximum value from the proposed species. Given the rapid pace of improvement in sequencing technologies, specific plans for sequencing each species on the target list are not outlined in detail.

4.1 – Reference Genome Assemblies. Previous proposals for primate genome sequencing (http://www.genome.gov/25521747) argued for high-quality whole genome shotgun (WGS) assemblies, in addition to sequencing ~1,000 BACs per species targeted to structurally difficult regions. The rationale for achieving high-quality assemblies has not changed since then. We request a similarly high quality assembly for all species proposed here. In order to detect subtle differences between closely related genomes that may play a role in complex phenotypes, highly accurate sequence will be required. High-quality assemblies will also allow accurate and efficient mapping of population genetic data and transcriptome data, as well as provide a general resource for the research community.

We anticipate next-generation sequencing platforms (Roche 454, Illumina GAII, ABI SOLiD) will become a suitable source of WGS coverage for high-quality assemblies. At this time, however, we do not have enough examples of *de novo* assemblies of mammalian-sized genomes with next-generation WGS data to predict potential limitations, or to estimate how much additional refinement will be required to meet quality targets. However, we expect that within the next 12 months, advancements in next-generation sequencing technologies and assembly algorithms will achieve reasonably high quality assemblies of mammalian-size genomes for $500k-1M[1] per genome. It is impractical to begin work on a broader set of species until this issue is further resolved. (Note that hybrid and pure 454 assemblies are possible, but more expensive.)

We also anticipate the investigation of HIV/SIV-related phenotypes will require the detailed study of structurally complex regions that have consistently proven refractory to WGS assembly. There are

[1] NHGRI has been given projections that Illumina instruments will produce 90Gb runs some time in the next 12 months. In addition, read lengths are approaching 100 bp with paired ends, which will make assembly possible. These projections are of course risky to rely on. They also do not include costs for, eg clone resources that may be needed.

many examples of highly duplicated gene loci associated with response to infection, immunity and vaccine development (i.e. alpha, beta defensins, chemokine ligand receptors, HLA, etc.) and drug detoxification (cytochrome P-450 gene families, glutathione S-transferases, carboxylesterases, etc.) that are polymorphic in the human population. It is therefore important to ensure that these regions are accurately represented.

While we anticipate future technological developments will greatly influence the methods used to accurately sequence complex regions, and reduce cost, the current standard is BAC-based clone-by-clone sequencing. With this in mind, the community will need to specify a finite list of the regions that are critical to resolve (the equivalent of the ~1,000 BACs of earlier primate proposals). It is likely that this list will contain biologically important regions that will be targeted across all species, but this list may also need to be partly customized for each species. Customization will occur either before work begins, based on biological insight, or after assembly, based on poor quality regions of the assembly or noticeable variation in depth of coverage indicative of copy number variation (CNV). In this manner, BAC sequences will serve to improve assembly quality, provide accurate sequence of complex duplicated regions, and resolve haplotypes, which will be critical for subsequent population-based analyses.

In some cases, it may actually be more important to target regions in many (40-50) species (e.g., to compare pathogenic and non-pathogenic hosts) rather than to sequence entire genomes. Whether this is desirable will depend on the relative costs of whole genome versus regional approaches. One approach to sequencing unique regions from many species at low cost that may prove effective is genomic capture. Currently used for cancer resequencing projects (TCGA, TSP), the principle of designing

probes from a reference sequence for the capture, through hybridization, of matching fragments from patient samples, could potentially be applied to survey sequencing of related primate species. At this time, the lower bound of sequence divergence that can be captured is not known, and current capture protocols have not been optimized for hybridization to distantly related targets. However, should capture prove unable to reliably obtain sequence across large spans of evolutionary time, capture methods can still provide a low-cost technique for obtaining sequences from many regions (hundreds to thousands of genes) simultaneously. Capture-based resequencing may then provide a low cost solution to genotyping many individuals of a population in lieu of designing costly custom SNP arrays.

4.1 Sequence variation—SNPs and CNVs/structural variations. Eventually, SNP data should be recovered for all of the species herein where there are large research colonies, though the highest priorities are the species that display phenotypic variation in SIV response: rhesus macaque (Indian vs. Chinese) and cynomolgus macaque, and to a slightly lesser extent the chimpanzee. The near-term aim is to provide sufficient information about variation in order to develop tools (genotyping reagents) to allow correlation of variation with differences in disease susceptibility and progression phenotypes. This is the most direct application of genome sequence to understanding disease phenotypes. It will be possible to use this resource, for example together with pedigree information, to obtain linkage information for some traits. Later, as the availability of genotyping reagents stimulates larger studies, association studies will become possible. It is also important to note that the significance of such a resource goes beyond HIV/SIV related phenotypes, as e.g. macaques, baboons and other primate species are used for a range of human disease studies.

Multiple ongoing projects (e.g. 1000 Genomes) have demonstrated the efficacy of using new sequencing platforms (Illumina, SOLiD) to obtain SNP information rapidly and cheaply. As this ability improves, it

becomes worthwhile to obtain SNP information for all species where there are relevant, varying phenotypes that can be used for disease studies. There may also be reason to obtain SNP information for species that are used in basic population genetics studies. Although not strictly relevant to NIH's mission, such data will also be useful for conservation and studies of evolutionary biology by other communities.

CNV's and other structural variations will also be important to assess. As of this writing, the new sequencing platforms are able to detect some classes of these, but may have difficulty detecting others. To the extent that these can be obtained cheaply and efficiently, they should be considered part of this white paper.

An appropriate number of candidate SNPs per species would be 1-2 million, with this broad range justified by our current lack of knowledge of the basic population structure of many of the species. Similarly the absolute number of CNVs to be tested could be empirically determined. Together these data could be combined in a single genotyping reagent, akin to the Affymetrix 6.0 human chip to allow both kinds of variation to be monitored in each species.

4.3. Expression analysis. Some expression analysis should be done for all species for which there is, or is proposed to be, a reference sequence.

The new sequencing platforms make this inexpensive enough for it to become a routine part of all *de novo* organismal genome projects. There are two types of data that are important to consider.

4.3.1 RNA sequencing to aid in annotation of a new genome. This is already standard practice for all new genomes.

4.3.2 A GTex-like effort: a catalog of expression from different tissues or even immune cell populations. We propose that comprehensive quantitative analysis of the transcriptome in selected tissues and immune cell types such as CD4+ T cells, CD8+ T cells, B cells, NK cells and macrophages should be pursued for rhesus (and later other) macaques, African green monkeys and sooty mangabeys. These species (especially rhesus) have an active research community, well-characterized phenotypes, and sufficient available tissue samples. For macaque, it will be very useful for the community to have, as a resource, high quality "baseline" expression data for multiple macaque tissues. This can serve as a basis for comparison to individual datasets generated by individual labs e.g., during the course of infection. Such a baseline, public dataset is more likely to be useful if it is produced at a sequencing center because of the amount of data needed, the high quality needed, and the desirability for public deposition. It will also be highly informative to be able to compare expression phenotypes between natural and non-natural hosts. For example, such information from natural and non-natural hosts will allow us to identify those conserved evolutionary features shared by natural hosts that may represent the core mechanisms underlying protection from disease following natural SIV infection. This will allow for studies in which *in vivo* manipulation of the immune system of non-natural hosts will be performed in order to achieve the same state of non-pathogenicity.

Appendix 1: Justifications for Specific Sequencing Targets

The following sections provide specific scientific justifications for the species and subspecies recommended for new whole genome sequencing projects. The overall scientific rationale for sequencing additional nonhuman primate taxa, based on the consensus conclusion that these sequencing projects will provide an improved understanding of HIV, SIV and susceptibility to lentivirus infection and/or subsequent disease, is presented in the main body of this White Paper. Here we present specific explanations for the individual species and subspecies that are proposed.

Papio Baboons: Yellow and Chacma Baboons

Baboons (genus *Papio*) are important laboratory primates that are significant in a wide variety of biomedical research fields. *Papio* baboons are regularly used for studies of infectious diseases, including analyses related to herpes viruses, cytomegalovirus and schistosomiasis. Over the past ten years, these animals have also been used as models for HIV infection and anti-HIV vaccine strategies (Locher et al. 2003; Locher et al. 2004), although baboons are not used in HIV/SIV research as widely as are macaques. The complete sequencing of the genome of olive baboons (*Papio anubis*) has been approved and is underway. Olive baboons were the first *Papio* species chosen for sequencing because they are the most numerous baboons in U.S. laboratory populations. The basic rationale for the sequencing of additional baboon species is that there is substantial genetic and phenotypic diversity among baboons. Some species that are not currently approved for sequencing have been shown to experience natural infection in the wild by SIV viruses originating in African green monkeys. Baboons are a non-natural host for SIV. But unlike macaques, baboons do not progress to simian AIDS after infection, and of course the genetic basis of this difference in disease outcome is unknown.

Across a number of regions within Africa, baboons are sympatric with African green monkeys and/or chimpanzees and/or sooty mangabeys, all of which are natural hosts for various strains of SIV. Consequently, baboons can be exposed to several different forms of SIV. Although surveys find that most baboons do not carry SIV viruses, strains of SIV have been isolated from wild yellow baboons, *P. cynocephalus* (Jin et al. 1994) and wild chacma baboons, *P. ursinus* (van Rensburg et al. 1998). The SIV strains isolated from these baboons came originally from African green monkeys, and the infected adult baboons did not exhibit any negative consequences of infection. Whole genome sequences from these species (yellow and chacma baboon) will contribute unique information to our understanding of the evolution of resistance to SIV-caused disease.

Papio baboons are closely related to rhesus macaques (see Figure One) but are more closely related to sooty mangabeys (genus *Cercocebus*), black mangabeys (genus *Lophocebus*) and mandrills (genus *Mandrillus*). These three genera are natural carriers of SIV, while *Papio* baboons themselves do not carry their own lentiviruses. A more complete genetic characterization of the genus *Papio*, along with characterization of these closely related natural hosts, will be valuable in understanding the evolution of susceptibility to infection, frequency or mechanism of transmission among conspecifics, and progression to clinical disease.

The genus *Papio* is generally subdivided into five morphotypes. These *Papio* morphotypes are considered species by some investigators and subspecies by others. Here we recommend whole genome genome sequencing and assembly of the yellow baboon (*P. cynocephalus*) and the chacma baboon (*P. ursinus*). Yellow baboons are currently available in research colonies, although not in the same numbers as olive baboons. The chacma baboon is frequently used in other countries, and can be imported to the US. Chacma baboons are the *Papio* species most divergent from the olive baboon (Newman et al. 2004).

Chinese Rhesus Macaque

Rhesus macaques (*Macaca mulatta*) are the most commonly used nonhuman primate in biomedical research, and the preeminent model for AIDS research. This species has a wide geographic distribution, stretching from Pakistan and India eastward to central and southern China. This is the largest geographic distribution of any primate species other than humans, and thus provides substantial opportunity for local and regional genetic differentiation. The initial whole genome DNA sequence for this species was produced from a single individual with original ancestry in India (Gibbs et al. 2007). Most rhesus macaques in US breeding and research colonies are derived from founding stocks imported from India. However, India ceased exporting macaques in the 1970's and the number of Chinese-origin animals used in research is now increasing steadily.

Numerous studies have shown that Chinese-origin animals are more resistant to infection with SIV and progress more slowly to AIDS-like disease than Indian-origin animals (e.g. (Ling et al. 2002; Trichel et al. 2002)). Genetic differences between Indian- and Chinese-origin animals are widely regarded as the major factor explaining this difference. Recently, researchers have shown that copy number variation in the CCL3L locus accounts for a portion of the observed variance in average disease response between populations (Degenhardt et al. 2009). Thus, it will be tremendously valuable to be able to compare the complete sequence from Chinese-origin rhesus monkeys to the existing sequence for Indian-origin animals.

In addition to the Indian-origin reference animal, the initial rhesus sequencing project also included targeted re-sequencing of 150kb from five ENCODE regions in both Chinese- and Indian-origin individuals (Hernandez et al. 2007) . This analysis found significant inter-population differences. Only 33% of all SNPs identified in the full dataset were shared between the two populations. Prior analyses of mtDNA, microsatellites and other genetic markers have also found significant genetic divergence separating these populations (Ferguson et al. 2007; Gibbs et al. 2007) . Based on the 150kb re-sequencing data, the two regional populations are estimated to have diverged 160,00 to 190,000 years ago. This molecular divergence is not surprising given the numerous anatomical, behavioral and physiological differences between Chinese and Indian rhesus.

The significant biological differences between these two populations, especially the well-replicated differences in response to SIV infection, justify thorough genomic analysis of both. Animals from each population are frequently used in U.S. research colonies, and the biological differences between them suggest there are many functionally significant molecular differences. The differential response to infection with SIV demonstrates unequivocally that there are differences in immune function that are influenced by genetic differences and are directly relevant to our understanding of differences in susceptibility among humans to infection with HIV and progression to disease. We propose here that a complete genome sequence and assembly should be produced for the Chinese-origin population, at least of sufficient quality for reliable comparisons to the current Indian macaque sequence. This may require resolving to high quality some regions of clear interest, for example *Kir* and MHC.We also propose additional investigation of genetic diversity among Chinese-ancestry animals themselves.

Pig-tailed macaques, *Macaca nemestrina*

Rhesus macaques (*M. mulatta*) and cynomolgus macaques (*M. fascicularis*) are the most commonly used nonhuman primates in research related to HIV, SIV and AIDS. However, pig-tailed macaques (*M. nemestrina*) are used in increasing numbers in a range of AIDS-related studies (e.g. Batten et al. 2006; Mankowski et al. 2008; Baroncelli et al. 2008; Lederer et al. 2009). Macaques of different species or populations significantly differ in their responses to and level of control of these viruses. Pigtailed macaques are now recognized as being a unique and valuable model for studies of susceptibility to infection and progression to disease following exposure to SIV (Baroncelli et al. 2008). Compared with rhesus macaques, pigtailed macaques appear to be more susceptible to SIV transmission, and once infected, they exhibit higher viral loads, and generally a more rapid progression to disease. Largely due to this increased susceptibility to infection and disease, pigtailed macaques have been used in studies of vaginal transmission, as well as SIV encephalitis, which appears to develop at a higher incidence in this species compared to rhesus (Mankowski et al. 2008). In addition, these animals actually support replication of HIV better than other nonhuman primate species. Although HIV-1 still replicates poorly in this model, the fact that these animals support HIV replication at all (Baroncelli et al. 2008) is considered important by many investigators, some of whom are trying to identify or engineer HIV strains that may eventually exhibit sustained viral replication in nonhuman primates. The ability to achieve sustained infection of nonhuman primates with HIV would obviously improve the macaque AIDS model and create numerous additional avenues for future study.

Given these prior results, examining and comparing genetic differences among humans, rhesus macaques and pigtailed macaques may help to decipher key genetic elements important in resistance to HIV-1 infection. Production of a whole genome assembly for the pigtailed macaque genome, along with initial analyses of SNP and CNV variation, will provide the HIV/SIV research community with the essential information necessary to begin those investigations. Researchers have already begun to quantify differences in gene expression among macaque and other primate species (Lederer et al. 2009). Comparative genomics at the level of gene sequences, gene copy number variation and quantitative analyses of transcription will benefit tremendously from the sequencing of this primate genome.

Additional Cynomolgus Macaques

Cynomolgus macaques (*Macaca fascicularis*) are widely used in simian immuno-deficiency virus (SIV) pathogenesis and vaccine research. Cynomolgus macaques can be infected mucosally with SIV, develop profound acute plasma viremia, maintain a chronic viral load 'steady state', and ultimately develop AIDS-defining opportunistic infections. The pathogenesis of SIV in cynomolgus macaques is similar to the pathogenesis of SIV in Indian rhesus macaques (*Macaca mulatta*). However, the genetics of Indian rhesus macaques are understood much better. Characterizing the genetics of cynomolgus macaques would enable the development of species-specific reagents (e.g., RNA microarrays, SNP arrays). Additionally, there may be important differences between rhesus and cynomolgus macaques in genes that interact with SIV and coordinate antiviral immune responses. For example, major histocompatibility complex (MHC) and killer immunoglobulin receptor alleles are rarely conserved between cynomolgus and rhesus macaques. These differences influence the quality of SIV-specific cellular immunity and consequently could have a major impact on the host's ability to control viral replication. Indeed, Mauritian-origin cynomolgus macaques, Chinese-origin rhesus macaques, and Indian-origin rhesus macaques exhibit marked differences in immunologic control of the pathogenic viruses SIVmac251 and SHIV89.6P (Reimann et al. 2005) .

The sequencing of a reference cynomolgus macaque genome is underway. This will enable detailed comparisons of rhesus and cynomolgus macaques and should facilitate the development of first-generation, cynomolgus macaque-specific genetic tools. Cynomolgus macaques are, however, widely distributed throughout Asia. Comprehensive understanding of cynomolgus macaque genetics will require genetic surveying of animals from Mainland Southeast Asia, Indonesia, the Philippines, and Mauritius (which has a sizable population of non-indigenous cynomolgus macaques). Little is known about the genetic differences among these populations. At one well-studied locus, alleles of major histocompatibility complex (MHC) genes are largely origin-specific. This provides preliminary evidence that regional populations of cynomolgus macaques are, in fact, genetically distinct. These genetic differences could cloud the interpretation of laboratory experiments with cynomolgus macaques and complicate the use of genetic tests in much the same way as tests validated in Indian origin rhesus macaques can be inappropriate for use in Chinese origin rhesus macaques (Vogel et al. 1995). Additionally, information concerning genetic variation could be used to manage primate resources: confirming animals' geographic origin, identifying hybrid macaques, and establishing priorities for breeding.

Compared to rhesus macaques, cynomolgus macaques are widely available for laboratory research. Thousands are imported annually, in stark contrast to Indian rhesus macaques whose importation has been banned since the late 1970s. As research demands for macaques grow, cynomolgus macaques are increasingly used in studies that previously utilized rhesus macaques. The ready availability of cynomolgus macaques also means that samples from feral and captive-bred animals are accessible for genetic surveys, including DNA sequencing analyses.

Sooty Mangabeys (*Cercocebus atys*)

Two simian immunodeficiency viruses, SIVcpz from chimpanzees and SIVsmm from the sooty mangabeys (*Cercocebus atys*), are the origin of HIV-1 and HIV-2, respectively. A key feature of natural SIV infection of sooty mangabeys (SMs) is the absence of clinical disease (Paiardini et al. 2009; Pandrea et al. 2008b; Silvestri et al. 2007). In sharp contrast, experimental SIV infection of non-natural host macaque species, such as the rhesus, results in clinical illness (simian AIDS) similar to that observed in HIV-infected humans. The mechanisms underlying the benign nature of SIV infection in SMs are still largely unknown; however, it is widely believed that elucidation of these "AIDS resistance" factors will provide important insights into the determinants of immune deficiency in HIV-infected humans (Kirchhoff 2009; Paiardini et al. 2009) . A key feature of non-pathogenic SIV infections in SMs is that these animals usually maintain normal or near normal peripheral CD4+ T cell counts despite levels of virus replication that are as high or higher than those of HIV-infected humans. From a vaccine perspective, the fact that the AIDS resistance of natural SIV hosts appears to be independent of effective control of virus replication emphasizes the tremendous challenge of artificially inducing a type of protective immunity that was not selected for by the evolutionary pressure posed by lentiviruses on the primate immune system (Silvestri et al. 2007). As such, it has been recently proposed that studies of SMs, as well as other natural SIV hosts species, may provide alternative strategies to develop an AIDS vaccine and prevention that do not rely solely on the induction of virus-specific adaptive immune responses (Sodora et al. 2009) .

There are several reasons why the sooty mangabeys represent a "unique" model for studies of AIDS pathogenesis. These reasons represent the rationale behind our proposal to generate a complete sequence of the sooty mangabey genome. First, the virus infecting SMs in the wild, i.e., SIVsmm, is the origin of both SIVmac (which infects Asian macaques and is the most commonly used primate model for studies of AIDS pathogenesis, therapy and vaccines) and HIV-2 (which is responsible for a significant epidemic). There has not been any sign of natural passage of SIVs from other small African primate species to humans. As such, it can be argued that sooty mangabeys represent the most relevant monkey model of natural SIV infection in terms of understanding human disease. Second, the only colony of naturally SIV-infected nonhuman primates available for research in the U.S. is the large colony of SIVsmm-infected SMs housed at the Yerkes Primate Research Center. There are no other colonies of naturally SIV-infected monkeys in the U.S. Third, there are significantly more immunological reagents available for the sooty mangabey than for any other African nonhuman primate species that is known to be a natural SIV host. Lastly, the CD4 cell biology of sooty mangabeys is more similar to that of humans or macaques than is the CD4 cell biology of African green monkeys. Typically, sooty mangabeys are similar to humans and macaques in that their circulating levels of CD4+ T cells, comprising 30-50% of all lymphocytes. In contrast, African green monkeys show lower levels of circulating and mucosal CD4+ T cells, and higher fraction of double positive CD4+CD8+ T cells. For this reason it can be argued that SMs represent a uniquely suited model to study the effects of SIV infection on CD4+ T cells. Thus, natural SIV infection of SMs represents a highly relevant and intensively studied model of natural SIV infection of African NHPs. The possibility of integrating the available virological and immunological information on SMs with a complete knowledge of the SM genome will represent an invaluable resource for the current AIDS research effort.

Black mangabey (*Lophocebus aterrimus*)

Black mangabeys are natural hosts of SIV. However, this species has been poorly studied in comparison with sooty mangabeys, African green monkeys and other natural hosts. The species-specific virus found in black mangabeys, designated SIVbkm, was discovered only recently (Takemura et al. 2005) and little is known about the phenotype of this infection either in black mangabeys themselves or in non-natural hosts such as rhesus macaques. Interestingly, one study of cross-species transmission of SIVsmm (the virus that infects sooty mangabeys) into three black mangabeys resulted in one case of progression to simian AIDS (Apetrei et al. 2004). This is the only case in which cross-species transmission of an SIV from one natural host to another natural host has caused disease (Apetrei et al. 2004). This experiment raises questions concerning species-specific features of the black mangabey genome that may make these animals more susceptible to HIV infection and AIDS. Researchers have long divided primates into natural hosts and non-natural hosts. This observation in black mangabeys, along with recent evidence of SIV-related mortality in chimpanzees (Keele et al. 2009) suggests that a rigid two-category classification may be too simple. Genomic analysis of black mangabeys and other species that do not fit this traditional dichotomous scheme may be quite significant for understanding resistance to infection and/or progression to disease.

Red-capped mangabey (*Cercocebus torquatus*)

There are two main reasons to conduct whole genome sequencing of red-capped mangabeys (RCMs) that are related to the two key features of this model of natural SIV infection. The first reason is that RCMs harbor a virus, SIVrcm, which has recombined with the SIV of the *Cercopithecus nictitans* (the greater-spotted nose monkey). The recombination of SIVrcm and SIVgsn produced the virus now found in chimpanzees (SIVcpz), and it is SIVcpz that is the precursor of HIV-1. While we do not know whether SIVrcm was "needed" to induce a virus that could be transmitted to chimpanzees, and eventually humans, this virological feature of SIV infection of RCMs makes these animals especially interesting. The second reason is that RCMs are quite unique, at least among natural SIV hosts studied to date, in that they frequently harbor a mutation in the CCR5 gene (Chen et al. 1998) that results in lack of surface expression of this molecule. CCR5 is commonly used by HIV and SIV as entry co-receptor. In addition, RCMs can be infected with SIVs that use CCR2b as co-receptors (Chen et al. 1998) . As such, RCMs may be interesting in that they may provide a model to study the impact of the absence of CCR5 on the overall immune system genetics and function.

Guenons: *Cercopithecus* species

We recommend approval for whole genome sequencing of four distinct primate species from the genus *Cercopithecus* (i.e., *Cercopithecus lhoesti* or L'Hoest's monkey, *Cercopithecus nictitans* or greater spot nosed monkey, *Cercopithecus mona* or mona monkey, and *Cercopithecus cephus* or mustached guenon). Three of these species (nictitans, mona, and cephus) are infected with distinct strains of SIV (i.e., SIVgsn, SIVmon, and SIVmus) that are of particular interest to HIV/AIDS research. Indeed, these three strains, unlike most SIV strains (such as SIVmac, SIVsmm, SIVagm, SIVrcm, SIVmnd-1 and –2, etc) are similar to both HIV-1 and SIVcpz (the virus from which HIV-1 originates) in that a) they contain a vpu gene and b) their nef gene product is unable to down-modulate the CD3-TCR complex from the surface of infected cells (herein referred to as "dysfunctional" nef). This latter finding is actually not surprising since it has been shown that SIVcpz (and thus HIV-1 as well) represents the result of an ancient recombination between the 5' pol-containing segment of the SIVrcm genome and the 3' env/nef-containing segment of the genome of SIVgsn. A central question in the HIV/AIDS field is whether this particular recombination was an event due to "non-virological" reasons (i.e., geographical co-localization, predation habits of chimpanzees, etc) or, alternatively, it was "necessary" from a virological and/or immunological standpoint for a successful cross-species transmission of SIV to chimpanzees (and possibly humans as well). In this context, it is unfortunate that virtually nothing is known about the phenotype of SIV infection in the three species of *Cercopithecus* that are infected with vpu-containing, nef-dysfunctional viruses. Of note, a recent survey of natural SIV infection conducted on wild *Cercopithecus* spp in Cameroon using a serological assay revealed that only a small minority of nictitans and cephus (9/859 and 9/864, respectively) are infected (Aghokeng et al. 2009). This finding is in sharp contrast with the much higher frequency of SIV infection (33-50%) in wild mandrills, mangabeys, or AGMs (Aghokeng et al. 2009). Such differences in prevalence suggest that perhaps nictitans and cephus employ different pathways for host adaptation to SIV viruses than do mandrills, mangabeys or green monkeys. The rationale for the inclusion of *Cercopithecus lhoesti* among the high priority primate species for whole genome sequencing is that they would represent a "control" species within the genus *Cercopithecus* that is infected with a more typical vpu-negative, nef-functional strain of SIV, yet are very closely related to the hosts carrying nef-dysfunctional viruses.

The vervet (*Chlorocebus aethiops*) was previously approved and is underway. Here we add a request to develop an expression resource for this organism (see section 4.3.2 of the main document).

Sequencing the Mandrill and Drill genomes: genus *Mandrillus*

The primary argument for sequencing the genomes of both drills (*Mandrillus leucophaeus*) and mandrills (*Mandrillus sphinx*) is that they are both natural hosts (for SIV_{drl} and SIV_{mnd}, respectively). As with all other natural hosts discussed in this White Paper, their genomes will provide additional perspective on adaptation to infection with SIV. The phylogeny is relevant here. Baboons (genus *Papio*) and macaques (genus *Macaca*) are not natural hosts. But baboons are more closely related to mandrills and drills than they are to macaques (see Figure One: phylogeny of primate species). Thus the status of natural and non-natural host does not map in a simple way onto the well-established phylogeny of the host species themselves. In this context, the phylogeny of the viruses is very useful in reconstructing the history of cross-species transmission (see Figure Two: phylogeny of viruses). Knowledge of the genome sequences of drills and mandrills will provide important additional information concerning the diversity found in the immunobiology of natural hosts.

Roughly 80 papers published in the last 10 years on these two species reflect research topics including SIV infection (22 publications), response to infection with filarial parasites (including vaccine development), cytomegalovirus, Ebola and other viral infection, genome evolution, immunology, dental research, primate evolution, and sexual dimorphism/circannual changes related to reproductive behavior. Although the volume of research on this genus is comparatively small, mandrills are, along with mangabeys and African green monkeys, perhaps the best studied natural hosts for SIV.

Between these two *Mandrillus* species, *M. sphinx* is the stronger candidate for genome sequencing. Specific responses of mandrills to infection are generally consistent with what has been found for other natural hosts. That is, active viral replication and high levels of virus are not associated with low or minimal levels of immune system activation in the host (Onanga et al. 2002; Onanga et al. 2006). Quite significantly, some cases have been reported of this species progressing to AIDS after long-term infection (Pandrea et al. 2001). This may indicate that mandrills are similar to chimpanzees in being able to tolerate SIV infection for long periods of time, but in some cases eventually developing clinical disease (Juompan et al. 2008; Keele et al. 2009). In addition, *M. sphinx* displays some important differences from what is known about other natural hosts (e.g. mangabey and African green monkey). Mandrills are infected by and adapted to two rather different SIV's (SIVmnd1 and SIVmnd2), which have independent origins. The former possibly originated from a virus found in *Cercopitehcus lhoesti* and the latter likely derived from a virus found in drills (*M. leucophaeus*). In addition, mandrills may experience more CD4 depletion post-infection than SM and AGMs (unpublished data, G. Sylvestri, personal communication). Some data from this species suggest a correlation between the low rates of vertical transmission in this species and CCR5 expression levels on CD4+T cells (Pandrea et al. 2008a). Although more is known about *M. sphinx* and the case for genome sequencing is thus stronger, ideally, both species should be sequenced. A genome sequence from *M. leucophaeus* would provide a close comparison with that of *M. sphinx.* Drills are infected with and adapted to relatively closely related SIV's (SIVdrl and SIVmnd2).

Sequencing Additional Chimpanzees

Chimpanzees are a unique and critical subject for genomic research. As our closest living evolutionary relative, the content and intra-species diversity of the chimpanzee genome is

important for understanding the human genome. The biological similarity of chimpanzees to humans also makes this an essential animal model for disease research. Chimpanzees have frequently been used to test vaccines against HIV, hepatitis and other pathogens. The close genetic similarity makes chimpanzees a potential reservoir for transmission of various infectious pathogens to humans. For example, chimpanzees are natural hosts of SIVcpz, widely recognized as the progenitor of HIV-1. It is thought that SIVcpz crossed into humans on at least two occasions, giving rise to the HIV-1 M and O lineages, respectively (Hahn et al. 2000).

While it has been widely assumed that lentiviral infections of wild chimpanzees, like endemic SIV infection in African monkeys, is apathogenic, two lines of evidence suggest that this may not be the case. First, a small cohort of captive chimpanzees experimentally infected with lab strains of HIV-1 in the 1980's has now developed AIDS or AIDS-like symptoms (Juompan et al. 2008; Novembre et al. 1997; O'Neil et al. 2000). Unlike other experimental models of AIDS, progression in these animals was variable and, like HIV-infection in many humans, unfolded slowly over a period of ten years or more (O'Neil et al. 2000). Field samples now suggest that naturally occurring SIVcpz infection in wild chimpanzees can result in AIDS-like symptoms and death (Keele et al. 2009). Thus, the virus-host relationship in chimpanzees may more closely resemble human HIV/AIDS than nonpathogenic infections in African monkeys or experimental disease models such as macaques. The potential of chimpanzees to be reservoirs of zoonotic transmission, their utility for reconstructing the direction and extent of evolution in human genes, and this newly appreciated relevance to understanding genetic influences on the progression of AIDS, makes the cataloguing of genetic variation within and between chimpanzee populations a research priority

The currently available genome sequence for this species is derived from a west African chimpanzee (subspecies *Pan troglodytes verus*). Three other subspecies are recognized: central African (*Pan t. troglodytes*), east African (*P. t. schweinfurthii*) and Nigerian-Cameroon (*P. t. vellerosus*). As part of the original chimpanzee genome project, four additional west African and three central African chimpanzees were sequenced at low coverage (Chimpanzee Sequencing and Analysis Consortium (2005), and heterozygosity estimated to be 8.0×10^{-4} in west African individuals, 17.6×10^{-4} among central African individuals and 19.0×10^{-4} between west and central African individuals. Subsequent analyses show that chimpanzees also exhibit substantial copy number variation (Perry et al. 2008), which may be particularly relevant to immunology and infectious disease, although little is known about the distribution of CNV diversity within and between subspecies. Overall, genetic diversity among west African chimpanzees appears to be roughly equivalent to levels observed in humans, while central and east African individuals are twice as variable and show substantial divergence from the west African individuals (Becquet et al. 2007; Marques-Bonet et al. 2009) . The significance of this species for understanding the history of the human genome, and the relationship between host genetics and HIV/SIV infection and disease argue strongly for additional genome analysis of central and eastern chimpanzees.

Appendix 2: Table of species and work proposed. Work that is already approved and done, underway or planned is shown for context.

Species	***Improved high quality draft***	***Regional finishing (may require clone resource)***	***Variation (SNPs, etc)***	***RNA seq for annotation***	***RNA seq for expression resource***
Papio anubis	*Underway*	Proposed	Planned	*Planned*	
P. cynocephalus	Proposed	Proposed	Proposed	Proposed	
P. ursinus	Proposed	Proposed	Proposed	Proposed	
Macacca mulatta (Indian)	*Underway*	*Planned*	*Underway*	*Planned*	Proposed
M. mulatta (Chinese)	Proposed. Sufficient quality for good comparison to Indian macaque	Proposed, as needed for high-quality comparisons in critical regions of interest	Essentially a high-quality variation project	Proposed	Proposed (in comparison with Indian macaque)
M. nemestrina	Proposed	Proposed	Proposed	Proposed	
M. fascicularis	*Underway*	Proposed	Proposed	Proposed	
Cercocebus atys	Proposed	Proposed	Proposed	Proposed	Proposed
Lophocebus aterrimus	Proposed	Proposed	If samples available	Proposed	
Cercocebus torquatus	Proposed	Proposed	*If samples available*	Proposed	
Chlorocebus aethiops	*Underway*	Proposed	*Planned*	*Planned*	Proposed
Cercopithecus lhoesti	Proposed	Proposed		Proposed	
C. nictitatus	Proposed	Proposed		Proposed	
C. mona	Proposed	Proposed		Proposed	
C. cephus	Proposed	Proposed		Proposed	
Mandrillus leukocaphalus	Proposed	Proposed	Proposed	Proposed	
M. sphinx	Proposed	Proposed	Proposed	Proposed	
Pan troglodytes subsp.: *versus, troglodytes, schweinfurthii*	*P. t. versus Completed*	*Underway*	Some done, more proposed: 1 *troglodytes*, multiple *schweinfurthii*	*Done* (*versus*)	

References Cited

2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437(7055):69-87.

Aghokeng AF, Ayouba A, Mpoudi-Ngole E, Loul S, Liegeois F, Delaporte E, and Peeters M. 2009. Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provides insights into risks for potential new cross-species transmissions. Infect Genet Evol.

Ambrose, Z., V. N. KewalRamani, P. D. Bieniasz, and T. Hatziioannou. 2007.HIV/AIDS: in search of an animal model. Trends Biotechnol 25:333-7.

Apetrei C, Gormus B, Pandrea I, Metzger M, ten Haaft P, Martin LN, Bohm R, Alvarez X, Koopman G, Murphey-Corb M and others. 2004. Direct inoculation of simian immunodeficiency virus from sooty mangabeys in black mangabeys (Lophocebus aterrimus): first evidence of AIDS in a heterologous African species and different pathologic outcomes of experimental infection. J Virol 78(21):11506-11518.

Baroncelli S, Negri DR, Michelini Z, and Cara A. 2008. Macaca mulatta, fascicularis and nemestrina in AIDS vaccine development. Expert Rev Vaccines 7(9):1419-1434.

Becquet C, Patterson N, Stone AC, Przeworski M, and Reich D. 2007. Genetic structure of chimpanzee populations. PLoS Genet 3(4):e66.

Chen Z, Kwon D, Jin Z, Monard S, Telfer P, Jones MS, Lu CY, Aguilar RF, Ho DD, and Marx PA. 1998. Natural infection of a homozygous delta24 CCR5 red-capped mangabey with an R2b-tropic simian immunodeficiency virus. J Exp Med 188(11):2057-2065.

Degenhardt JD, de Candia P, Chabot A, Schwartz S, Henderson L, Ling B, Hunter M, Jiang Z, Palermo RE, Katze M and others. 2009. Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus Macaques (Macaca mulatta). PLoS Genet 5(1):e1000346.

Dunham R, Pagliardini P, Gordon S, Sumpter B, Engram J, Moanna A, Paiardini M, Mandl JN, Lawson B, Garg S and others. 2006. The AIDS resistance of naturally SIV-infected sooty mangabeys is independent of cellular immunity to the virus. Blood 108(1):209-217.

Ferguson B, Street SL, Wright H, Pearson C, Jia Y, Thompson SL, Allibone P, Dubay CJ, Spindel E, and Norgren RB, Jr. 2007. Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (Macaca mulatta). BMC Genomics 8:43.

Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK and others. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316(5822):222-234.

Gordon SPI, Dunham R, Apetrei C, and al. E. 2005. The call of the wild: What can be learned from studies of SIV infection of natural hosts? In: Thomas FB, Leitner T, and Hahn BH, editors. HIV Sequence Compendium. Los Alamos, N.M.: Los Alamos National Laboratory. p 2-29.

Goulder PJ, and Watkins DI. 2008. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat Rev Immunol 8(8):619-630.

Hahn BH, Shaw GM, De Cock KM, and Sharp PM. 2000. AIDS as a zoonosis: scientific and public health implications. Science 287(5453):607-614.

Hatziioannou, T., Z. Ambrose, N. P. Chung, M. Piatak, Jr., F. Yuan, C. M. Trubey, V. Coalter, R. Kiser, D. Schneider, J. Smedley, R. Pung, M. Gathuka, J. D. Estes, R. S. Veazey, V. N. KewalRamani, J. D.

Lifson, and P. D. Bieniasz. 2009. A macaque model of HIV-1 infection. Proc Natl Acad Sci U S A 106:4425-9.
Hatziioannou, T., M. Princiotta, M. Piatak, Jr., F. Yuan, F. Zhang, J. D. Lifson, and P. D. Bieniasz. 2006. Generation of simian-tropic HIV-1 by restriction factor evasion. Science 314:95.
Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J and others. 2007. Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science 316(5822):240-243.
Juompan LY, Hutchinson K, Montefiori DC, Nidtha S, Villinger F, and Novembre FJ. 2008. Analysis of the immune responses in chimpanzees infected with HIV type 1 isolates. AIDS Res Hum Retroviruses 24(4):573-586.
Kamada, K., T. Igarashi, M. A. Martin, B. Khamsri, K. Hatcho, T. Yamashita, M. Fujita, T. Uchiyama, and A. Adachi. 2006. Generation of HIV-1 derivatives that productively infect macaque monkey lymphoid cells. Proc Natl Acad Sci U S A 103:16959-64.
Keele BF, Jones JH, Terio KA, Estes JD, Rudicell RS, Wilson ML, Li Y, Learn GH, Beasley TM, Schumacher-Stankey J and others. 2009. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460(7254):515-519.
Kirchhoff F. 2009. Is the high virulence of HIV-1 an unfortunate coincidence of primate lentiviral evolution? Nat Rev Microbiol 7(6):467-476.
Lackner AA, and Veazey RS. 2007. Current concepts in AIDS pathogenesis: insights from the SIV/macaque model. Annu Rev Med 58:461-476.
Ling B, Veazey RS, Luckay A, Penedo C, Xu K, Lifson JD, and Marx PA. 2002. SIV(mac) pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans. AIDS 16(11):1489-1496.
Locher CP, Fujimura S, Murthy KK, Brasky K, Leland M, and Levy JA. 2003. Expression patterns of phenotypic markers on lymphocytes from human immunodeficiency virus type 2-infected baboons. AIDS Res Hum Retroviruses 19(1):31-40.
Locher CP, Witt SA, Ashlock BM, Polacino P, Hu SL, Shiboski S, Schmidt AM, Agy MB, Anderson DM, Staprans SI and others. 2004. Human immunodeficiency virus type 2 DNA vaccine provides partial protection from acute baboon infection. Vaccine 22(17-18):2261-2272.
Marques-Bonet T, Ryder OA, and Eichler EE. 2009. Sequencing primate genomes: what have we learned? Annu Rev Genomics Hum Genet 10:355-386.
Neil, S. J., T. Zang, and P. D. Bieniasz. 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451:425-30.
Nelson GW, and O'Brien SJ. 2006. Using mutual information to measure the impact of multiple genetic factors on AIDS. J Acquir Immune Defic Syndr 42(3):347-354.
Novembre FJ, Saucier M, Anderson DC, Klumpp SA, O'Neil SP, Brown CR, 2nd, Hart CE, Guenthner PC, Swenson RB, and McClure HM. 1997. Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71(5):4086-4091.
O'Neil SP, Novembre FJ, Hill AB, Suwyn C, Hart CE, Evans-Strickfaden T, Anderson DC, deRosayro J, Herndon JG, Saucier M and others. 2000. Progressive infection in a subset of HIV-1-positive chimpanzees. J Infect Dis 182(4):1051-1062.
Onanga R, Kornfeld C, Pandrea I, Estaquier J, Souquiere S, Rouquet P, Mavoungou VP, Bourry O, M'Boup S, Barre-Sinoussi F and others. 2002. High levels of viral replication contrast with only transient changes in CD4(+) and CD8(+) cell numbers during the early phase of experimental infection with simian immunodeficiency virus SIVmnd-1 in Mandrillus sphinx. J Virol 76(20):10256-10263.

Onanga R, Souquiere S, Makuwa M, Mouinga-Ondeme A, Simon F, Apetrei C, and Roques P. 2006. Primary simian immunodeficiency virus SIVmnd-2 infection in mandrills (Mandrillus sphinx). J Virol 80(7):3301-3309.

Paiardini M, Pandrea I, Apetrei C, and Silvestri G. 2009. Lessons learned from the natural hosts of HIV-related viruses. Annu Rev Med 60:485-495.

Pandrea I, Onanga R, Rouquet P, Bourry O, Ngari P, Wickings EJ, Roques P, and Apetrei C. 2001. Chronic SIV infection ultimately causes immunodeficiency in African non-human primates. AIDS 15(18):2461-2462.

Pandrea I, Onanga R, Souquiere S, Mouinga-Ondeme A, Bourry O, Makuwa M, Rouquet P, Silvestri G, Simon F, Roques P and others. 2008a. Paucity of CD4+ CCR5+ T cells may prevent transmission of simian immunodeficiency virus in natural nonhuman primate hosts by breast-feeding. J Virol 82(11):5501-5509.

Pandrea I, Sodora DL, Silvestri G, and Apetrei C. 2008b. Into the wild: simian immunodeficiency virus (SIV) infection in natural hosts. Trends Immunol 29(9):419-428.

Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C and others. 2008. Copy number variation and evolution in humans and chimpanzees. Genome Res 18(11):1698-1710.

Reimann KA, Parker RA, Seaman MS, Beaudry K, Beddall M, Peterson L, Williams KC, Veazey RS, Montefiori DC, Mascola JR and others. 2005. Pathogenicity of simian-human immunodeficiency virus SHIV-89.6P and SIVmac is attenuated in cynomolgus macaques and associated with early T-lymphocyte responses. J Virol 79(14):8878-8885.

Sheehy, A. M., J. D. Gaddis, J. D. Choi, and M. H. Malim. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646-650.

Silvestri G, Paiardini M, Pandrea I, Lederman MM, and Sodora DL. 2007. Understanding the benign nature of SIV infection in natural hosts. J Clin Invest 117(11):3148-3154.

Singh P, Kaur G, Sharma G, and Mehra NK. 2008. Immunogenetic basis of HIV-1 infection, transmission and disease progression. Vaccine 26(24):2966-2980.

Sodora DL, Allan JS, Apetrei C, Brenchley JM, Douek DC, Else JG, Estes JD, Hahn BH, Hirsch VM, Kaur A and others. 2009. Toward an AIDS vaccine: lessons from natural simian immunodeficiency virus infections of African nonhuman primate hosts. Nat Med 15(8):861-865.

Stremlau, M., C. M. Owens, M. J. Perron, M. Kiessling, P. Autissier, and J. Sodroski. 2004. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature 427:848-53.

Sumpter B, Dunham R, Gordon S, Engram J, Hennessy M, Kinter A, Paiardini M, Cervasi B, Klatt N, McClure H and others. 2007. Correlates of preserved CD4(+) T cell homeostasis during natural, nonpathogenic simian immunodeficiency virus infection of sooty mangabeys: implications for AIDS pathogenesis. J Immunol 178(3):1680-1691.

Takemura T, Ekwalanga M, Bikandou B, Ido E, Yamaguchi-Kabata Y, Ohkura S, Harada H, Takehisa J, Ichimura H, Parra HJ and others. 2005. A novel simian immunodeficiency virus from black mangabey (Lophocebus aterrimus) in the Democratic Republic of Congo. J Gen Virol 86(Pt 7):1967-1971.

Trichel AM, Rajakumar PA, and Murphey-Corb M. 2002. Species-specific variation in SIV disease progression between Chinese and Indian subspecies of rhesus macaque. J Med Primatol 31(4-5):171-178.

Vogel T, Norley S, Beer B, and Kurth R. 1995. Rapid screening for Mamu-A1-positive rhesus macaques using a SIVmac Gag peptide-specific cytotoxic T-lymphocyte assay. Immunology 84(3):482-487.