LETTER

OPEN

doi:10.1038/nature25140

Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host

David Palesch[1]*, Steven E. Bosinger[1,2]*, Gregory K. Tharp[1], Thomas H. Vanderford[1], Mirko Paiardini[1,2], Ann Chahroudi[1,3], Zachary P. Johnson[1], Frank Kirchhoff[4], Beatrice H. Hahn[5], Robert B. Norgren Jr[6], Nirav B. Patel[1], Donald L. Sodora[7], Reem A. Dawoud[1], Caro-Beth Stewart[8], Sara M. Seepo[8], R. Alan Harris[9,10], Yue Liu[9], Muthuswamy Raveendran[9,10], Yi Han[9], Adam English[9], Gregg W. C. Thomas[11], Matthew W. Hahn[11], Lenore Pipes[12], Christopher E. Mason[12], Donna M. Muzny[9,10], Richard A. Gibbs[9,10], Daniel Sauter[4], Kim Worley[9,10], Jeffrey Rogers[9,10] & Guido Silvestri[1,2]

In contrast to infections with human immunodeficiency virus (HIV) in humans and simian immunodeficiency virus (SIV) in macaques, SIV infection of a natural host, sooty mangabeys (*Cercocebus atys*), is non-pathogenic despite high viraemia[1]. Here we sequenced and assembled the genome of a captive sooty mangabey. We conducted genome-wide comparative analyses of transcript assemblies from *C. atys* and AIDS-susceptible species, such as humans and macaques, to identify candidates for host genetic factors that influence susceptibility. We identified several immune-related genes in the genome of *C. atys* that show substantial sequence divergence from macaques or humans. One of these sequence divergences, a C-terminal frameshift in the toll-like receptor-4 (*TLR4*) gene of *C. atys*, is associated with a blunted *in vitro* response to TLR-4 ligands. In addition, we found a major structural change in exons 3–4 of the immune-regulatory protein intercellular adhesion molecule 2 (ICAM-2); expression of this variant leads to reduced cell surface expression of ICAM-2. These data provide a resource for comparative genomic studies of HIV and/or SIV pathogenesis and may help to elucidate the mechanisms by which SIV-infected sooty mangabeys avoid AIDS.

SIV infection of natural hosts, such as sooty mangabeys, is typically non-pathogenic despite high viraemia. This is in stark contrast to HIV infection in humans and experimental SIV infection in rhesus macaques (*Macaca mulatta*) that progress to AIDS unless treated with antiretroviral therapy. The main virological and immunological features of natural SIV infection in sooty mangabeys have been described over the past 15 years in studies that compared and contrasted this infection with the pathogenic infections of HIV and SIV in humans and rhesus macaques[1]. SIV-infected sooty mangabeys show several features that have been observed in pathogenic infections, including high viraemia, short *in vivo* lifespan of productively infected cells, depletion of mucosal $CD4^+$ T cells, strong type-I interferon response in the acute infection, and cellular immune responses that fail to control virus replication. However, in contrast to pathogenic infections, SIV-infected sooty mangabeys (i) have healthy $CD4^+$ T cell levels; (ii) do not experience mucosal immune dysfunction, avoiding depletion of T helper 17 (T_H17) cells, intestinal epithelial damage and microbial translocation; (iii) maintain low levels of immune activation during the chronic infection; and (iv) achieve compartmentalization of virus replication that preserves central-memory and stem-cell memory $CD4^+$ T cells as well as follicular T_H cells[1,2]. An additional notable feature of SIV infection in natural hosts is the low rate of mother-to-infant transmission that is related to low expression of CCR5 on circulating and mucosal $CD4^+$ T cells[3]. Although many aspects of the natural course of SIV infection in sooty mangabeys have now been described, the key molecular mechanisms by which these animals avoid AIDS remain poorly understood.

In this study, we sequenced the genome of a captive sooty mangabey and compared this genome to the genomes of AIDS-susceptible primates to look for candidate genes that may influence susceptibility to AIDS in SIV-infected hosts. We sequenced genomic DNA to a whole-genome coverage of about 180× using the Illumina HiSeq 2000 platform, and produced an initial assembly using ALLPATHS-LG, Atlas-Link and Atlas-GapFill (see Methods for details). The total size of the assembled *C. atys* genome (Caty_1.0; NCBI accession number GCA_000955945.1) is around 2.85 Gb, with a contig N50 size of 112.9 kb and scaffold N50 size of 12.85 Mb (Table 1). Genome annotation identified 20,829 protein-coding genes and 4,464 non-coding genes in the *C. atys* assembly, which is comparable to other available draft quality genomes of nonhuman primates (Table 1). These analyses demonstrate that the Caty_1.0 reference genome is of sufficient quality to facilitate population-scale whole-genome and transcriptome sequencing studies.

To identify novel immunogenetic factors specific to *C. atys* that may be involved in the ability of this species to avoid progression to AIDS, we established a bioinformatic pipeline for a comparative protein analysis (Fig. 1 and Extended Data Fig. 1, see Methods for details). Using this approach, we found 34 candidate immune-related genes with sequences that diverged between *C. atys* and *M. mulatta* (Table 1 and Extended Data Table 1). Although we cannot exclude a role of immune genes with minor differences in *C. atys* and *M. mulatta*, the highly divergent genes listed in Table 1 and Extended Data Table 1 constitute candidate genes involved in the outcomes of SIV infection in these two species.

Our screen identified sequence divergence in a number of proteins that are important during HIV infection, such as APOBEC3C (91.6%) and BST2 (also known as tetherin, 95.1%), as well as pattern-recognition receptors (MBL2, CLEC4A, CLEC4D and CLEC6A), the antiviral sensor cyclic GMP–AMP synthase (cGAS (also known as MB21D1)) and other immune mediators (Extended Data Table 1). Because CD4 and CCR5 are important for AIDS pathogenesis, we aligned the sequences of *Ca*CD4 and *Ca*CCR5 to *Mm*CD4 and *Mm*CCR5, respectively[4,5]. Neither gene showed any major structural changes in the wild-type variants, although CD4 was slightly below the 97%

[1]Emory Vaccine Center and Yerkes National Primate Research Center, Emory University, Atlanta, Georgia 30329, USA. [2]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia 30329, USA. [3]Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia 30329, USA. [4]Institute of Molecular Virology, Ulm University Medical Center, 89081 Ulm, Germany. [5]Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [6]Department of Genetics, Cell Biology and Anatomy, University of Nebraska, Medical Center, Omaha, Nebraska 68198, USA. [7]Center for Infectious Disease Research, formerly Seattle Biomedical Research Institute, Seattle, Washington 98109, USA. [8]Department of Biological Sciences, University at Albany-State University of New York, Albany, New York 12222, USA. [9]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [10]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. [11]Department of Biology and School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405, USA. [12]Department of Physiology and Biophysics, Weill Cornell Medical College, New York, New York 10065, USA.
*These authors contributed equally to this work.