INVESTIGATING AN ADAPTIVE EXPLANATION FOR NONPROGRESSOR IMMUNE RESPONSES TO SIV

by

KELSIE MCVEETY

A THESIS

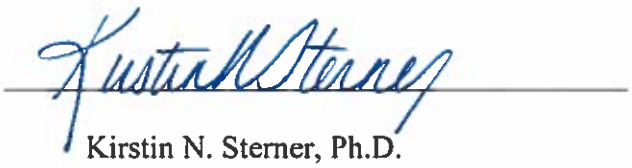
Presented to the Department of Anthropology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2015

An Abstract of the Thesis of

Kelsie McVeety for the degree of Bachelor of Science
in the Department of Anthropology to be taken June 2015

Title: Investigating an adaptive explanation for nonprogressor immune responses to SIV

Approved: 

Kirstin N. Sterner, Ph.D.

HIV/AIDS continues to be a widespread epidemic that is currently treated with immunotherapies that target the viral mechanisms, leaving room for the evolution of resistant strains. Most research into novel therapies that target the human immune response investigates the mechanisms that leave humans vulnerable to HIV infection. However, there is much to be gained by examining nonhuman primate species whose immune systems can naturally survive with high levels of replication of a related lentivirus, simian immunodeficiency virus (SIV), without the result of chronic immune activation or viral resistance. These species are known as nonprogressors, and by investigating an adaptive explanation for their tolerance of SIV this project sought to connect the evolutionary history of these species to variation in immune response to SIV, with the intent of identifying target genes and molecular pathways for future medical intervention and deepening our knowledge of how natural selection and coevolution with lentiviruses has shaped primate immune responses in general. Evolutionary genetic methods were used to test for positive selection, convergent evolution, and variants that distinguish nonprogressors from progressors in the protein-

coding and *cis*-regulatory regions of four genes (*CASP1, CD38, EEF1D,* and *SOCS1*) differentially expressed in progressors and nonprogressors during the chronic phase of infection (Bosinger *et. al*. 2009). Results indicate these genes are largely conserved in these primates and have experienced negative selection. Furthermore, nonprogressors do not share derived variants in the promoter regions. Future research should look to test the functional significance of distinguishing polymorphisms in regulatory and protein-coding regions of genes differentially expressed in nonprogressors and progressors.

Acknowledgements

I would like to extend my sincerest gratitude to my primary thesis advisor and mentor, Professor Kirstin Sterner, for her endless support, encouragement, and constructive criticism throughout this process. My thesis and I would not be near the quality we are today without her. I would also like to thank Professor Josh Snodgrass and Professor Helen Southworth for their emotional support and assistance in the editing process. Further thanks are due to everyone in the University of Oregon Molecular Anthropology Group--especially Professor Nelson Ting for his presentation advice, Mr. Noah Simons for his assistance with the data analysis, and Arleigh for her commiserations and pep talks regarding the seemingly endless thesis process.

Of course I must also thank my incredible Mom and Dad, Renie and Bruce McVeety, for everything they have ever said and sacrificed for my sake. Much gratitude also goes to my brother and sister, Ray and Megan, who kept my life bright over this last year. I also would like to thank my dog, Jackie, who passed away last March, for all the love and support she gave me in her long life.

Many thanks go out to all my friends all over the world, from Eugene, Oregon to Dublin, Ohio, and beyond. I want to extend special thanks to Abhishek Saxena, Emily Ross, Ashley Smith, and Courtney Cunningham, for consistently reminding me over the last ten years that I am capable of great things.

Last, but definitely not least, I want to acknowledge my high school biology teacher, Mrs. Sondra Snodgrass. Her I.B. Biology class is what fostered my early interest in anthropological genetics. I would not be here without her guidance.

Table of Contents

List of Figures

List of Tables

INTRODUCTION

The AIDS Epidemic

The acquired immune deficiency syndrome (AIDS) epidemic has undoubtedly been one of the most impactful medical tragedies of our time, and is far from over. The virus responsible, human immunodeficiency virus (HIV), has taken approximately 39 million lives worldwide since the disease was identified in 1982 (*UNAIDS*, 2015a). While the total deaths per year have decreased 35% since 2005, there were still 1.5 million lost in 2013 (*UNAIDs,* 2015a). In addition, 2.1 million were newly infected in 2013, totaling 35 million adults and children living with HIV in 2013 (*UNAIDS*, 2015b). Unfortunately only 23% of infected children and 37% of infected adults receive the treatment they need (*WHO*, 2015). This profound loss of human life has led to a succession of research into treatments and possible cures. Some are effective, and have lead to higher quality of life for HIV+ individuals, while many others have failed. Thus far the primary objective of most biomedical and public health research has been to understand the mechanisms responsible for the chronic innate immune responses associated with progression to AIDS in order to find possible points of medical intervention, but the search for quality treatments and preventative vaccines is still ongoing.

HIV Infection

When humans are infected with the retrovirus HIV, the virus preferentially bonds to the CD4 molecule on the outside of white blood cells called T cells, which are crucial to defending against infections and cancers (Fauci, 1988; Nielsen *et al.*, 2005).

In addition to the CD4 receptor, HIV also requires the presence of a chemokine (cell-signaling) co-receptor such as CXCR4 or CCR5 to properly bind with a T cell (König *et al.* 2008). After binding, the virion fuses with the T cell, releasing its contents (RNA, and the enzymes reverse transcriptase, integrase, and protease) into the cell (König *et al.* 2008). Once inside, reverse transcriptase begins to transcribe viral RNA into DNA and integrase integrates this viral DNA into the host DNA (Fauci, 1988). Through this process, HIV takes over the host cell's transcriptional machinery and uses it to produce copies of its RNA and messenger RNA (mRNA). The mRNA is translated into amino acids and protease snips the amino acid chain into functional proteins that, when combined with the produced viral RNA, will make up the newly produced virions (Fauci, 1988). These new copies of HIV then bud from the surface of the host cell and circulate in order to infect other cells, thus continuing the viral life cycle (Fauci, 1988).

This replication occurs throughout the two stages of HIV infection, the acute stage and the chronic stage. During the acute stage of HIV infection, the individual is often unaware of the quickly replicating virus, though they may experience some flu-like symptoms. If individuals receive antiretroviral therapy during the acute stage or the beginning of the chronic stage, and if they are active in taking care of their health, they can often live a normal lifespan without AIDS-related disease (Stages of HIV Infection, 2013). Regardless if a person receives treatment, the arrival of HIV in the body causes an extensive innate immune response during the acute stage of the infection that leads to depletion of T cells.

The innate immune response is the body's first line of defense against infiltration by foreign organisms, and is thought to be evolutionarily older than the

adaptive immune system (Levy, 2001). The innate immune response is initiated rapidly--in minutes to hours after initial infection. Though it does not provide long-lasting protective immunity, the innate immune system offers some low-level defense against many pathogens by recruiting immune cells to infection sites, stimulating the production of immune-signaling cytokines, activating molecular pathways to identify bacteria, activating cells, promoting clearance of dead cells, and triggering the adaptive immune response (Levy, 2001). The adaptive immune response involves the development of specific immune cells to remember and quickly respond to certain pathogens. Humans appear to be unable to develop an HIV-specific immune response, resulting in an unchecked innate immune response and high viral replication, eventually leading to complete immune dysregulation (Hazenberg *et al.* 2003). Immune dysregulation is an unrestrained, unregulated immune response that is either inappropriately intense or ineffectively weak, and involves the depletion of the body's T cells and causes susceptibility to various illnesses. This final stage of immune dysfunction is what is known as AIDS (Stages of HIV Infection, 2013). In the United States, a person is classified as having AIDS when their CD4+T cell count is less than 200 T cells per microliter (Buehler and Berkelman, 1990).

Current Treatments and Approaches

Past and current treatments for HIV have targeted several steps during infection in order to disrupt viral replication. The first effective treatment for HIV was tested by Fischl and colleagues (1987) in a double-blind, placebo-controlled trial. The drug, an antiretroviral called azidothymidine (AZT), was found to increase circulating immune T cells in HIV+ individuals, slowing down viral replication and disease progression. This

new treatment was quickly patented, approved, and dispersed. AZT was initially prescribed at extremely high doses, creating heavy pill burdens and side effects such as severe anemia, muscle weakness, and liver problems (*NIH: AIDSinfo,* 2014a).

Today, combination antiretroviral therapy is the most common treatment for HIV infection, and can involve many medications from over 20 drugs that fall into 6 distinct categories based on how they disrupt HIV replication (De Clercq, 2004; *NIH: National Institute of Allergy and Infectious Diseases*, 2013; **Table 1**). Each category targets the virus at a different step in the infection process. While these medications when taken together consistently improve quality of life and slow progression to AIDS, they do not cure HIV infection, can have severe side effects, and leave patients vulnerable to antiviral resistant strains because they interact directly with the quickly mutating virus.

Table 1: Current antiretroviral treatments.

Treatment	**Function**
CCR5 antagonists	Target HIV before it enters the cell by interfering with the CCR5 receptors to which the virus adheres.
Fusion inhibitors	Prevent fusion between the virus and host cell membrane.
Integrase strand transfer inhibitors	Inhibit integrase production.
Protease inhibitors	Prevent protease from creating functional HIV proteins from amino acid chains.
Nucleotide reverse transcriptase inhibitors	Appear to the cell and viral enzymes to be free building blocks of DNA but actually stunt a HIV DNA chain when attached. AZT was the first NRTI.
Non-nucleotide reverse transcriptase inhibitors	Directly attach to reverse transcriptase and prevent it from converting HIV RNA to HIV DNA.

Despite the development of many effective antiretrovirals, the drugs can often cause serious medical side effects and if not taken properly can become ineffective because of the development of viral resistance. Side effects include occasional dizziness, swelling of the mouth, tongue or liver damage, bone loss, and abnormal fat distribution (*NIH: National Institute of Allergy and Infectious Diseases*, 2009). When a combination of different HIV medicines or HIV medicine and another drug interfere with each other's potency, or a person does not take medication consistently and exactly as prescribed, the drugs can be only partially effective and wipe out most but not all of the virus (Clavel and Hance, 2004; *NIH: National Institute of Allergy and Infectious Diseases*, 2010). Some strains of HIV mutate very quickly, and all strains have severely short generations compared to humans, which allows the virus to become resistant to the medicine before humans can adapt to the virus. If a virus strain has a genetic mutation that confers resistance to the antiretroviral drug combination, that strain will not be wiped out and will continue to replicate, surpassing the other strains and making that drug combination powerless against the growing infection (Clavel and Hance, 2004; *NIH: AIDSinfo*, 2014b). The treatments medical researchers have developed are effective at slowing the progression of the disease, but are vulnerable to drug resistance at a high rate (50% of infected individuals carry a strain that is resistant to at least one drug) and mainly target viral replication post-infection (Clavel and Hance, 2004).

As an alternative or complement to tackling the virus, many researchers have investigated possible immunotherapies that stimulate the host's immune system to respond to HIV either before or after infection. Shiver and colleagues (2002) found that they could generate HIV-specific immune cells with their vaccine, which prevented

illness in nonhuman primate models up to a year after initial infection. However, in later human trials the vaccine was not found to be effective in preventing or controlling viral replication (Buchbinder *et al.*, 2008). There are several other types of immunotherapies which attempt to regulate the functioning of the host immune system that have been adapted from cancer treatments such as cytokine therapy, adoptive T cell therapy, and dendritic cell vaccines. While these three have not had great success in clinical trials (Lieberman *et al.*, 1997), one type of immunotherapy appears more promising. Some researchers are attempting to block two genes that inhibit immune cell activation in order to activate more immune cells for antiviral responses (Velu *et al.*, 2009). While this approach may help the immune system respond, the development of AIDS is closely tied to chronic immune activation and could also be triggered by this type of treatment (Smith and Housseau, 2015).

Through the long and arduous process of medical intervention research there have been and continue to be many setbacks, but those setbacks have brought advances in technology and have expanded knowledge of mechanisms of disease pathogenesis (Robb, 2011). Much of this research has sought to develop medical interventions by examining some nonhuman primates (NHPs) as models of human immune response (McChesney and Miller, 2013). NHPs such as rhesus macaques (*Macaca mulatta*), known as *progressors*, develop immunodeficiency and AIDS-like symptoms when infected with simian immunodeficiency virus (SIV), making them useful models for testing HIV therapies and vaccines (Letvin *et al.*, 1985). In contrast to progressors, there are four types of species that can evade progression to AIDS: *nonprogressor* species that don't experience chronic phase T cell loss or progression to AIDS (Silvestri *et al.*

2007), *tolerator* species that are susceptible to experimental infection with HIV-2 and experience T cell loss without progression to AIDS (Locher *et al.* 2002), species with SIV but no known progression to AIDS (Etienne *et al.* 2012), and species with no known SIV infection (**Appendix A**). While research on the cellular mechanisms of HIV and SIV infection in progressor species has informed the development of antiretroviral interventions, this research only explores the *proximate* mechanisms of infection and progression to AIDS (i.e. *how* it occurs). To explore the *ultimate* evolutionary explanations of vulnerability to infection and progression to AIDS (i.e. *why* it exists in some species), we can examine nonprogressor species, which can tolerate SIV infection without chronic immune activation or triggering viral resistance. Certain nonprogressor species such as sooty mangabeys (*Cercocebus atys*) are closely related to progressors like rhesus macaques.

Research into an adaptive explanation for the nonprogressor immune response could help identify mechanisms associated with nonprogressors' natural immunity to SIV and possibly lead to future HIV treatments. An adaptive explanation suggests there was at one time heritable variation in nonprogressor ancestors' ability to respond to lentiviruses, and that mutations arose that allowed some individuals to prevent progression to simian AIDS. These individuals would survive more and reproduce more than the individuals who succumbed to AIDS, eventually leading to the widespread nonprogressor immune type that has somehow evaded viral resistance. By comparing closely related species for variation in susceptibility to disease, we can identify physiological mechanisms that appear to be operating differently and may have allowed nonprogressor ancestors to differentially survive. We may be able to connect these

different proximate mechanisms to the specific gene mutations that control them by searching for signs of positive selection in genes that contribute to these immune mechanisms. Testing an adaptive explanation for the nonprogressor immune profile could help identify a genetic basis for their tolerance of SIV and further our understanding of how natural selection shapes immune responses in general.

Progressor versus Nonprogressor Responses

In order to investigate an adaptive explanation for nonprogressor tolerance to SIV, it is important to first understand the basic molecular mechanisms that distinguish progressor and nonprogressor immune responses. While all primates susceptible to HIV/SIV experience high amounts of viral replication, short lifespan of infected cells, loss of T immune cells, and extreme innate immune activation during the acute phase of infection, research has identified key differences in chronic disease progression (Chahroudi 2012). The most relevant difference is that even though nonprogressors present with high levels of virus replication, immune dysregulation is almost nonexistent during the chronic stage (Silvestri *et al.* 2007). In sooty mangabeys and African green monkeys (*Chlorocebus sabaeus*), innate and adaptive immune responses normalize after acute infection (Bosinger *et al.* 2009, Jacquelin *et al.* 2009). However, in humans, rhesus macaques, and chimpanzees, innate and adaptive immune reactions can remain elevated until death or intervention (Silvestri *et al.* 2007). By avoiding chronic immune activation, nonprogressors may maintain T cell levels, prevent bystander immunopathology, and preclude dysregulation of critical immune cell subsets and tissues. It is not yet clear which specific mechanisms prevent chronic immune activation in nonprogressors.

One possible mechanism may be through the regulation of interferon levels. Interferons (IFN) are a group of signaling proteins that are involved in antiviral immune responses. Though IFN appears to increase in both progressors and nonprogressors during the acute stage of infection, IFN sharply decreases in nonprogressors during the chronic stage of infection (Bosinger *et al.* 2009, Jacquelin *et al.* 2009). In contrast, progressor species seem to sustain this high level of IFN throughout the chronic phase, which may contribute to chronic immune activation and disease progression (Silvestri *et al.* 2007).

A second possible mechanism for preventing chronic immune activation could be the limitation of the movement of the virus from the intestine to full-body circulation, called microbial translocation. Nonprogressors are able to maintain their mucosal immune environment, which protects the integrity of the gut mucosal barrier. In acute SIV infection of sooty mangabeys and African green monkeys, mucosal helper T cell populations decrease about 50 to 90%, but this depletion does not continue. Though the blood and tissues of natural hosts usually contain only low levels of microbial products, researchers have been able to mimic microbial translocation by injecting bacteria in SIV-infected AGMs and induce increased immune activation (Chahroudi 2012).

Specific immune regulatory pathways might also be responsible for the rapid resolution of immune activation in nonprogressor species. These pathways could include those involved in programmed cell death, cell signaling, and microbial translocation. Mutations in the genes that control these pathways, which distinguish nonprogressors from progressors, could be the result of coevolution with ancestral

forms of SIV, and may help identify candidates for immunotherapy and other medical interventions. In order to investigate an adaptive explanation for the nonprogressor immune profile, we can search for signs of positive selection in genes that contribute to these pathways.

To do this we must first characterize the evolutionary history underlying the mechanisms that leave species vulnerable to HIV and SIV, as well as the evolutionary history of the mechanisms that confer immunity to SIV in some monkey species. Research indicates that the primary strain of immunodeficiency virus that infects humans, HIV-1, resulted from multiple cross-species transmissions of SIV from chimpanzees to humans (Compton *et al.* 2013). The initial transmissions of SIV from nonhuman primates to humans most likely occurred from blood to blood contact during bushmeat preparation (Marx *et al.* 2001). SIV can be found in over 30 species of African NHPs, and is nonpathogenic and non-lethal in most species that can be infected (Klatt *et al.* 2012). Notable nonprogressor species are sooty mangabeys, African Green monkeys (AGM), mandrills (*Mandrillus sphinx*), and Greater Spot-nosed monkeys (*Cercopithecus nictitans*), among others, while progressors include macaques and some chimpanzees (*Pan troglodytes*) (Klatt *et al.* 2012). Apetrei and colleagues (2005) originally hypothesized that the specific viral strains that trigger AIDS-like immune dysregulation in progressors are intrinsically pathogenic compared to SIV strains that infect nonprogressors. However, experimental infection of progressor rhesus macaques with sooty mangabey and African green monkey SIV strains (SIVsmm and SIVagm.sab) still resulted in progression to simian AIDS (Klatt *et al.* 2012). A more promising direction of research is the coevolution hypothesis, which suggests some Old

World monkeys have coevolved with lentiviruses ancestral to SIV for millions of years and may have evolved adaptations to allow them to coexist with the virus. Genetic research has indicated that macaques, African Green monkeys, and colobus monkeys (Colobinae) are just a few of the species whose genomes display signs of selection by environments inhabited by ancient lentiviruses ancestral to SIV (Compton *et al.* 2013). This research has dated the emergence of some simian lentiviruses to 10 million years ago, allowing nonprogressor species ample time to evolve immune adaptations. If nonprogressors have adapted to SIV strains, we may be able to use this evolutionary pattern to identify molecular mechanisms that confer immunity to SIV, furthering our knowledge of how natural selection shapes tolerance to viral infection through coevolution.

Research Objectives

Researchers have identified several proximate mechanisms that differ between progressors and nonprogressors and may contribute to differential survival, but have less thoroughly investigated an adaptive explanation for the nonprogressor immune response. By testing immune genes for signs of selection and variation that differentiates nonprogressors from progressors, it may be possible to identify the specific pathways responsible for the prevention of chronic immune dysregulation and further understand the role coevolution with viruses plays in the evolution of immune tolerance.

To begin this process, three key evolutionary methods were utilized to examine this variation in 4 immune genes: phylogenetic analyses, tests for positive selection in

the protein-coding regions by estimating the ratios of nonsynonymous to synonymous rates (dN/dS), and analysis of transcription factor binding sites in regulatory regions.

Phylogenetic analyses were used to infer gene trees from protein-coding regions of four genes (*CASP1, CD38, EEF1D, SOCS1*) for the purpose of identifying aberrant patterns in the gene's evolutionary history when compared with the species' evolutionary history (Tamura *et al.* 2007). If the gene tree did not match the species phylogeny, this might suggest that selection had influenced the gene's evolutionary history.

In order to further analyze the four genes for signatures of selection, ratios of nonsynonymous to synonymous rates (dN/dS) were computed for the protein-coding regions. The building blocks of DNA (nucleotides) that make up the protein-coding region are read in sets of three called codons. These codons are translated into amino acids to build proteins, directly affecting cell function. A nucleotide substitution can be beneficial, deleterious, or neutral, depending on how it changes the amino acid it codes for. Because there is a redundancy in the code, i.e. different combinations of nucleotides in a codon can produce the same amino acid, some substitutions do not change the amino acid and thus have a neutral effect on the encoded protein and its function; this is called a synonymous substitution. In contrast, some substitutions can change the amino acid, protein, and sometimes the function of the protein produced. These substitutions are known as nonsynonymous substitutions, and can be beneficial, neutral, or deleterious. Beneficial substitutions increase the reproductive fitness of an individual in a specific environment, and may become more common in the population due to natural selection. Neutral substitutions do not affect the reproductive fitness of an individual in

a specific environment, and can be fixed or lost in a population through neutral evolution. Deleterious substitutions decrease the fitness of individuals in a specific environment, and thus are selected against. A protein-coding sequence that has not been shaped by natural selection will have about the same amount of nonsynonymous and synonymous substitutions and thus will have a dN/dS close to 1. Alternatively, a protein-coding region shaped by positive selection will have more nonsynonymous substitutions than synonymous substitutions and thus will have a dN/dS greater than 1. A protein-coding region shaped by negative selection will have more synonymous than nonsynonymous substitutions and thus dN/dS will be less than 1. In order to assess signatures of positive selection in the 4 target genes, the maximum likelihood of two ratio models were compared. The fixed ratio model, M_0, assumed dN/dS (or ω) was the same across all branches of the species tree. M_1 is known as a free ratio model, as it allows ω to vary across branches. To assess whether selective pressures have differentially affected nonprogressor species, estimated ω values were plotted on species phylogeny and compared for each branch. By comparing the dN/dS values of immune genes *CASP1, CD38, EEF1D,* and *SOCS1* protein-coding regions in primate species, we may be able to detect sequences shaped by natural selection.

Finally, promoters (a type of regulatory region) were analyzed for variation in the transcription factor binding sites (TFBS). A promoter is a part of the genome that usually sits upstream of a gene, controlling how much of a gene is produced, or how much is "expressed." The promoters of the target genes may be indirectly affecting the differential immune profiles observed in primates by influencing how many copies of a gene are produced (gene expression) rather than directly changing the function of the

produced protein. Transcription factors are proteins that bind to a TFBS to initiate the process of transcribing DNA into mRNA so it later can be translated into protein. If there is variation in these sites of attachment, it could affect how the gene is expressed and possibly lead to distinct immune response profiles. Variation in the TFBS that impacts immune function and differentiates nonprogressors from progressors could indicate possible sites of positive selection that possibly play a role in nonprogressor tolerance to viral infection.

Though the physiological mechanisms that differentiate nonprogressors from progressors are well characterized, there has been less research exploring an adaptive explanation for the nonprogressor immune response. This approach stands to further elucidate the role lentiviruses have played in the development of primate immune systems. It may eventually be possible to identify candidate pathways for medical interventions, but the first step in the evolutionary approach is to search for signs of selection and mutations that distinguish nonprogressors from progressors in the genes that control important immune mechanisms.

Objectives and Hypotheses

Objective 1: In order to detect strong convergent evolution in nonprogressor genes, gene trees were inferred and compared to known species phylogenies.

- **H_1:** Nonprogressor species form a monophyletic clade excluding progressor species.
- **H_2:** Gene trees reflect known species relationships.
- **H_0:** Gene tree topology does not reflect known species relationships or progressor/nonprogressor status.

H_1 would indicate very strong convergent or parallel evolution in nonprogressor species. H_2 could suggest there were only a few sites that distinguished all nonprogressors from progressors, or that the protein-coding sequences are different due to species divergence and the genes are not appropriate targets. H_0 would suggest untested factors are responsible for differences between these sequences.

Objective 2: Fixed and free models of ratios of nonsynonymous to synonymous rates were compared in order to identify sequences with signs of positive selection.

- **H_1:** The free ratio model (M_1) better fits the data than the fixed ratio model (M_0), and all nonprogressor species have a dN/dS greater than 1 compared to progressor species.
- **H_2:** The free ratio model (M_1) better fits the data than the fixed ratio model (M_0), but variation in dN/dS does not reflect progressor/nonprogressor status.
- **H_3:** The fixed ratio model (M_0) better fits the data than the free ratio model (M_1).

H_1 would suggest the gene in question has evidence suggesting positive selection for in nonprogressor species as opposed to progressor species, supporting an adaptive explanation for nonprogressor immune type. H_2 would suggest the gene was not shaped by positive selection for nonprogressor immune response but instead was influenced by untested factors. H_3 would suggest natural selection shaped the gene similarly in all tested species. A dN/dS greater than 1 would be a sign positive selection shaped the gene. If dN/dS was fixed near 1, this would suggest neutral evolution shaped the gene rather than positive or negative selection. If dN/dS was fixed near 0, this would suggest that negative selection shaped the gene.

Objective 3: Promoter sequences were compared to identify substitutions that distinguish nonprogressor transcription factor binding sites from progressor TFBS.

- **H_1:** Mutations in TFBS distinguish nonprogressors from all progressors.
- **H_2:** Mutations in TFBS reflect known species relationships.
- **H_0:** Mutations in TFBS do not reflect known species relationships or progressor/nonprogressor status.

H_1 would indicate mutations in TFBS could be responsible for the differential gene expression and thus contrasting immune responses in progressors and nonprogressors. H_2 would indicate species divergence drove variations TFBS, not natural selection for nonprogressor immune response. Results consistent with H_0 would suggest variation in TFBS is the result of untested factors, not evolutionary history or natural selection for nonprogressor immune response.

METHODS

Gene Choice

SOCS1, EEF1D, CASP1, and *CD38* were chosen from Bosinger and colleagues (2009), which identified these genes as differentially expressed in nonprogressors and progressors during the chronic phase of SIV infection. Rotger *et al.* 2011 investigated the same genes in human progressors and viral nonprogressors (VNP), confirming the genes were also expressed differently in these two groups. Differential expression of protein-coding genes is usually controlled through regulatory regions like promoters and can alter the function of the protein and may affect fitness. The four genes of interest this study investigated are detailed in **Table 2.**

Table 2. Function and expression profile of target genes used in the present study (Bosinger *et al.* 2009).

Name	Expression profile in progressor (*M. mulatta*)	Expression profile in nonprogressor (*C. atys*)	Function
Caspase 1 (*CASP1*)	Increased	Reduced	Induces cell apoptosis
Cluster of differentiation 38 (*CD38*)	Increased	Reduced	Facilitates cell adhesion, signal transduction and calcium signaling
Eukaryotic Translation Elongation Factor 1 Delta (*EEF1D*)	Reduced	Increased	Following HIV infection, stimulates repression of translation of host cell proteins and enhances translation of viral proteins.
Suppressor of cytokine signaling 1 (*SOCS1*)	Reduced	Increased	Regulates T cell function and limits inflammatory responses

Each of these genes encodes a protein that performs a function in regulating or modulating immune responses. Two of the genes display reduced expression in progressors and increased expression in nonprogressors: *SOCS1* and *EEF1D* (Bosinger *et al.* 2009; **Table 2**). *SOCS1* plays a role in regulatory T cell function and limits inflammatory responses. If progressors have lower *SOCS1* expression they may have more trouble controlling the release and subsequent infection of immune cells (Takahashi *et al.* 2011). *EEF1D* was examined because of its role in HIV pathogenesis. The protein encoded by *EEF1D* represses host cell response and enhances HIV replication within individual cells (Warren *et al.* 2012). Thus it is counterintuitive that *EEF1D* expression was found to be increased in nonprogressors. If *EEF1D* functions similarly in progressors and nonprogressors when expressed, then increased expression would increase SIV replication, furthering disease progression. This would suggest there may be a mutation in nonprogressor *EEF1D*—as well as the gene's promoter region—that alters the function of the gene. An alternative scenario is that for some unknown reason increased *EEF1D* expression is beneficial to nonprogressors.

The other two genes, *CASP1* and *CD38*, display reduced expression in nonprogressors and increased expression in progressors (Bosinger *et al.* 2009; **Table 2**). *CASP1* was studied for variations between progressors and nonprogressors because of its role in immune exhaustion. *CASP1* encodes a protein that is involved in inducing T cell apoptosis in infected cells (Stasakova *et al.* 2005). The increased expression of *CASP1* thus may be partly responsible for the T cell exhaustion observed in SIV-infected progressor species. The final gene studied, *CD38*, is also expressed at higher levels in progressors. *CD38* produces a membrane protein that helps regulate

intracellular signaling, and the loss of this gene is associated with impaired immune responses and HIV pathogenesis (Malavasi et al. 2008). If reduced expression of *CD38* contributes to HIV pathogenesis in humans but nonhuman primate progressors display increased expression, then perhaps there are mutations in *CD38* that account for difference in immune response between closely related species.

Gene Alignments

The protein-coding sequences for *SOCS1, EEF1D, CASP1,* and *CD38* were obtained from a genome database, Ensembl release 79 (Cunningham *et al.* 2015); the longest transcripts were used. Only available primate and rodent taxa with complete or near-complete sequence data were included in the alignments. Based on these criteria for inclusion, ClustalW (Goujon *et al.* 2010; Kearse *et al*. 2012) version 2 (an alignment program) was used to create one alignment per gene for each publically available primate species. The protein-coding region of *CASP1* (1203 bp) was aligned in 9 species, the protein-coding region of *CD38* (933 bp) in 12 species, the protein-coding region of *EEF1D* (1983 bp) in 7 species, and the protein-coding region of *SOCS1* (648 bp) in 9 species (**Table 3**). The alignments were manually edited to enforce codon boundaries and remove stop codons.

Table 3: Promoter, protein-coding, and amino-acid sequence alignments.

(length in base pairs (bp) and amino acids (aa), *denotes nonprogressors, underline denotes progressors).

Name	Promoter	Protein-coding Region	Amino Acid
Caspase 1 (*CASP1*)	*Homo sapiens* *Pan troglodytes* *Pongo abelii* *Papio anubis* **Mandrillus leucophaeus* *Macaca fascicularis* **Chlorocebus sabaeus* **Cercocebus atys*	*Homo sapiens* *Pan troglodytes* *Gorilla gorilla* *Nomascus leucogenys* *Pongo abelii* *Papio anubis* **Chlorocebus sabaeus* *Tarsius tarsier* *Mus musculus*	*Homo sapiens* *Pan troglodytes* *Gorilla gorilla* *Nomascus leucogenys* *Pongo abelii* *Papio anubis* **Chlorocebus sabaeus* *Tarsius tarsier* *Mus musculus*
	438 bp	1203 bp	401 aa
Cluster of differentiation 38 (*CD38*)	*Homo sapiens* *Pan troglodytes* *Macaca mulatta* *Macaca fuscata* **Chlorocebus sabaeus* *Saimiri boliviensis*	*Homo sapiens* *Pan troglodytes* *Gorilla gorilla* *Pongo abelii* *Nomascus leucogenys* *Papio anubis* *Macaca mulatta* **Chlorocebus sabaeus* *Callithrix jacchus* *Otolemur garnettii* *Rattus norvegicus* *Mus musculus*	*Homo sapiens* *Pan troglodytes* *Gorilla gorilla* *Pongo abelii* *Nomascus leucogenys* *Papio anubis* *Macaca mulatta* **Chlorocebus sabaeus* *Callithrix jacchus* *Otolemur garnettii* *Rattus norvegicus* *Mus musculus*
	487 bp	933 bp	311 aa
Eukaryotic Translation Elongation Factor 1 Delta (*EEF1D*)	*Homo sapiens* *Pan troglodytes* *Gorilla gorilla* *Nomascus leucogenys* *Papio anubis* *Macaca nemestrina* **Cercocebus atys* **Chlorocebus sabaeus*	*Homo sapiens* *Pan troglodytes* *Gorilla gorilla* **Chlorocebus sabaeus* *Otolemur garnettii* *Rattus norvegicus* *Mus musculus*	*Homo sapiens* *Pan troglodytes* *Gorilla gorilla* **Chlorocebus sabaeus* *Otolemur garnettii* *Rattus norvegicus* *Mus musculus*
	328 bp	1983 bp	663 aa
Suppressor of cytokine signaling 1 (*SOCS1*)	*Homo sapiens* *Gorilla gorilla* **Chlorocebus sabaeus* *Callithrix jacchus* *Rattus norvegicus* *Mus musculus*	*Homo sapiens* *Pan troglodytes* *Pongo abelii* *Papio anubis* *Macaca mulatta* **Chlorocebus sabaeus* *Callithrix jacchus* *Otolemur garnettii* *Rattus norvegicus*	*Homo sapiens* *Pan troglodytes* *Pongo abelii* *Papio anubis* *Macaca mulatta* **Chlorocebus sabaeus* *Callithrix jacchus* *Otolemur garnettii* *Rattus norvegicus*
	116 bp	648 bp	216 aa

Phylogenetic Analyses

To test Objective 1, phylogenetic trees were inferred using Molecular Evolutionary Genetics Analysis (MEGA; Tamura *et al.* 2011) version 6.06 for *CASP1, CD38, EEF1D,* and *SOCS1* protein-coding sequences. Based on Bayesian Information Criterion (BIC), the best DNA substitution model for all the protein-coding sequence alignments was identified as Tamura's 3-parameter model (T92) (Lio and Goldman, 1998). Mouse (*Mus musculus*) and rat (*Rattus norvegicus*) were used as outgroups when available. Maximum likelihood methods were used to infer gene trees for each gene alignment with 10,000 bootstrap replicates. The bootstrap method estimates the confidence of each grouping of species (clade) on a phylogenetic tree by generating many different possible trees (replicates) and calculating the proportion of trees in which each clade appears (Soltis and Soltis, 2003). Inferred tree topologies were compared to known species phylogenies to assess variation in gene evolutionary history.

dN/dS Analyses

In order to test Objective 2, PAMLX (Xu *et al.* 2013; Yang 2007) v4.8's codeml program was used to estimate the ratios of nonsynonymous to synonymous rates (dN/dS) in the protein-coding regions of *CASP1, CD38, EEF1D,* and *SOCS1*. Known species phylogenies (Glazko and Nei, 2003) were converted to Newick format to be used as input trees (**Figure 1A-D**). In order to assess signatures of positive selection in the protein-coding regions of these genes, the maximum likelihood of the free ratio model (M_1) was compared to that of the fixed ratio model (M_0). To assess whether selective pressures have differentially affected nonprogressor species, estimated ω

values were plotted on species phylogeny and compared for each branch. An ω greater than 1 would indicate the gene was shaped by positive selection along that branch, less than 1 would indicate negative selection, and an ω equal to 1 would indicate neutral evolution shaped the gene in question.

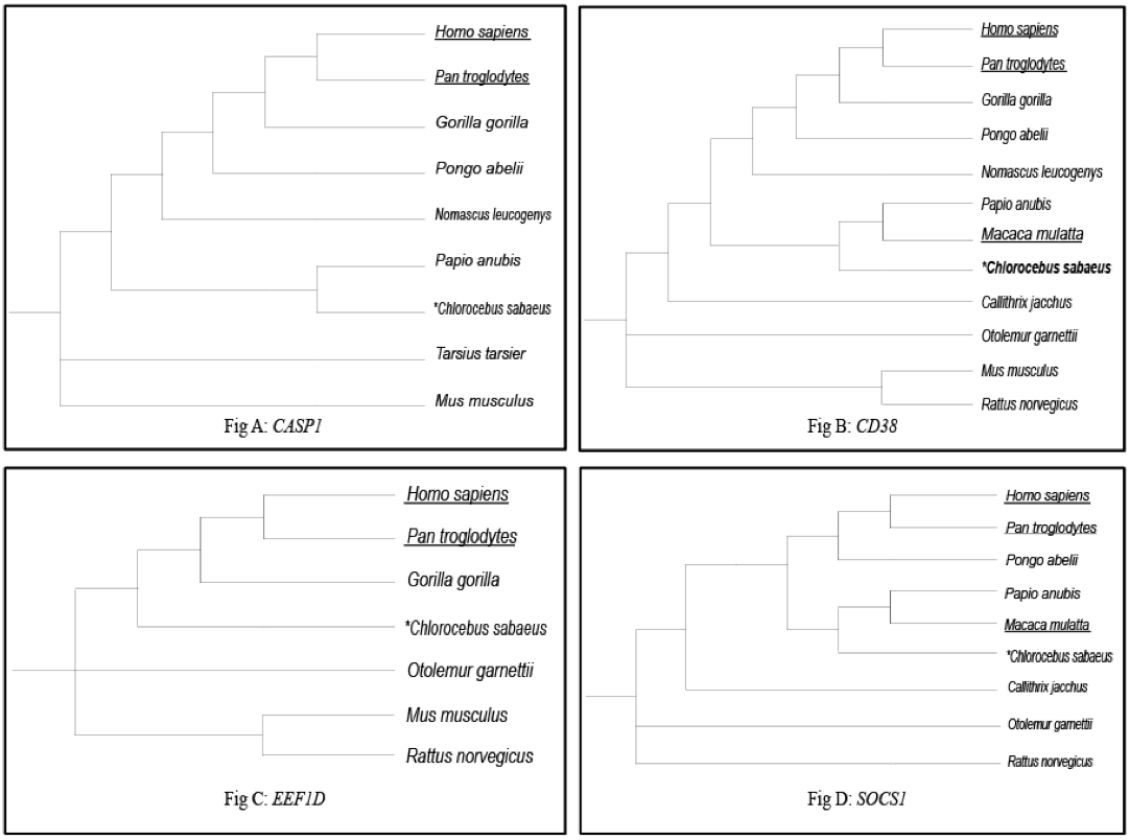


Figure 1: Known species phylogeny.

Phylogenetic trees demonstrating known species phylogeny (Glazko and Nei, 2003) were trifurcated for dN/dS analyses and compared with gene trees for phylogenetic analyses (*denotes nonprogressors, underline denotes progressors). (A) Phylogeny for CASP1 species. (B) Phylogeny for CD38 species. (C) Phylogeny for EEF1D species. (D) Phylogeny for SOCS1 species.

Promoter Alignments

Human promoter sequences for *CASP1*, *EEF1D*, and *SOCS1* were obtained from the genomics research supply company SwitchGear Genomics (accessed April

2015). In order to identify TFBS in *CD38* promoters, I used the chimpanzee promoter sequence published by Ferrero *et al.* 2004. These promoters were used to identify promoters from other species using NCBI's BLAST tool, which compares a query sequence to a genome database to identify similar sequences (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Only publically available primate and rodent species with near-complete orthologous sequences with high conservation were included in the alignments. Based on these criteria for inclusion, ClustalW version 2 (Goujon *et al.* 2010; Kearse *et al*. 2012) was used to create one alignment per gene for available primate species. The *CASP1* promoter (438 bp) was aligned in 8 species, the *CD38* promoter region (487 bp), the *EEF1D* promoter (328 bp) in 8 species, and the *SOCS1* promoter (116 bp) in 6 species (**Table 3**).

Promoter Analyses

To fulfill Objective 3, Consite was used to identify human transcription factor binding sites in promoters for *CASP1, EEF1D,* and *SOCS1* (Sandelin *et al.* 2004). Only TFBS with a cut-off score of at least 95% were used. TFBS in the isolated *CD38* chimpanzee promoter were found in Ferrero *et al.* 2004. For each gene, identified TFBS were compared across progressor and nonprogressor promoter alignments at the same position in each sequence.

Amino Acid Alignments

The protein-coding regions of the genes were translated using Geneious and the translations were aligned using the ClustalW program (Goujon *et al.* 2010; Kearse *et al*. 2012) to create one alignment per gene. The translated protein-coding region of *CASP1*

(401 aa) aligned in 9 species, *CD38* (311 aa) aligned in 12 species, *EEF1D* (633 aa) aligned in 7 species, and *SOCS1* (216 aa) aligned in 9 species (**Table 3**).

Amino Acid Analyses

Amino acid alignments were used to verify the validity of Ensembl's human protein-coding sequences by comparing them to UniProtKB's protein sequences for each gene (UniProt Consortium, 2014). UniprotKB's sequence version 1 was used for *CASP1,* sequence version 2 for *CD38,* sequence version 2 for *EEF1D,* and sequence version 1 for *SOCS1*.

RESULTS

Phylogenetic Analyses

In order to examine the evolutionary relationships of each gene in publically available primate species (Objective 1), gene trees were inferred and compared to known species trees.

CASP1

The data demonstrate that the *CASP1* gene tree (**Figure 2A**) reflects the species tree (**Figure 1A**). Catarrhines (*H. sapiens, G. gorilla, P. troglodytes, P. abelii, N. leucogenys, P. anubis,* and *C. sabaeus*) formed a monophyletic clade, and the bootstrap confidence intervals are all 85 or above. A bootstrap above 85 indicates high confidence in the inferred relationship between sequences (Soltis and Soltis, 2003). This supports H_2 and rejects the other hypotheses, suggesting the *C. sabaeus* sequence does not have a significant number of substitutions that distinguish it from nonprogressor sequences.

CD38

Phylogenetic analyses show that the *CD38* gene tree (**Figure 2B**) reflects the species tree (**Figure 1B**), with two caveats. First, while the position of *P. anubis* branch is consistent with known species relationships, the extant *P. anubis* branch is much longer than any of the other Catarrhine branches. Greater branch length indicates greater divergence from the last common ancestor of *P. anubis* and *M. mulatta*. This would suggest that the version of *CD38* found in *P. anubis* is not closely related to the form found in *M. mulatta*, which is notable as *P. anubis* tolerate SIV but *M. mulatta* are progressors (**Appendix A**). Furthermore, tolerator *P. Anubis* and nonprogressor *C.*

sabaeus do not form an exclusive clade, rejecting H_1. Second, the *Otolemur garnettii* sequence is more closely related to that of rodents than primates. Bootstrap confidence intervals were all above 70. Overall, the majority of this tree supports H_2, but because these two aberrations support H_0 it is likely that untested factors were responsible for variations from known species relationships.

EEF1D

Phylogenetic data demonstrate that the *EEF1D* gene tree (**Figure 2C**) reflects the species tree (**Figure 1C**). The Catarrhines form a monophyletic clade, as do the rodents. The bootstrap confidence intervals are all 100, aside from that of the *Pan troglodytes* and *Homo sapiens* clade, which was 46. This suggests that the chimpanzee and human clade may not be valid and one of the sequences may be more closely related to the gorilla sequence. This gene tree supports H_2, suggesting the nonprogressor *C. sabaeus* sequence does not have a significant about of derived substitutions distinguishing it from progressor sequences.

SOCS1

While phylogenetic data demonstrate the SOCS1 gene tree (Figure 2D) reflects the species tree (Figure 1D) in the ape clade and the rodent clade, there are a number of differences among other species. First, according to the gene tree, *C. jacchus* sequence diverged from the common ancestor with the ape clade after the clade diverged from the common ancestor with the Old World monkey clade. However *C. jacchus* is a Callitrichid and in the known species phylogeny is sister to the Old World monkey clade. This suggests the *C. jacchus* sequence is more closely related to the ape sequences than either are related to the Old World monkey sequences. However, the

confidence interval for this node was only 44, suggesting this inferred relationship may not be accurate. Next, the Old World monkey sequences formed a monophyletic clade with a high confidence interval of 99, but the analysis was unable to resolve the relationships within this group. According to the known species phylogeny, *P. anubis* and M. mulatta are more closely related to each other than either are to C. sabaeus (**Figure 1D**). These data support H_0, as the progressors and nonprogressor sequences appear to have a relationship that does not reflect the known species relationships nor progressor/nonprogressor status.

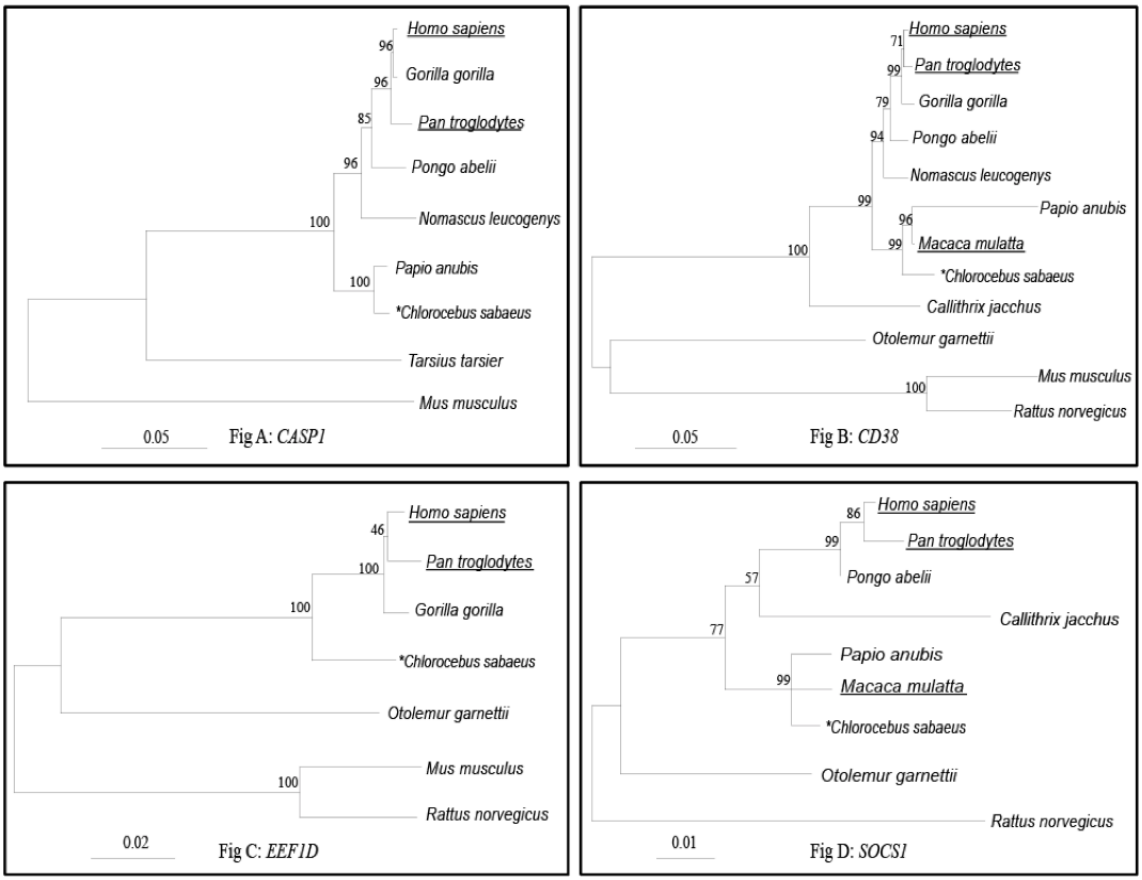


Figure 2: Gene trees inferred using maximum likelihood methods with 10,000 bootstrap replicates.

Bootstrap values are plotted on the branches and branch lengths are to scale (scale bar denotes nucleotide substitutions per site; *denotes nonprogressors; underline denotes progressors). (A) Inferred phylogeny for CASP1 protein-coding sequences. (B) Inferred phylogeny for CD38 protein-coding sequences. (C) Inferred phylogeny for EEF1D protein-coding sequences. (D) Inferred phylogeny for SOCS1 protein-coding sequences.

dN/dS Analysis

To test for signatures of positive selection in the protein-coding regions, the maximum likelihood of a fixed ratio (M_0) and free ratio (M_1) model of dN/dS variation were compared for each gene. The likelihood that the ω of *CASP1, CD38,* and *EEF1D* vary between species was significantly (p<0.05) greater than the likelihood of dN/dS

being the same across branches. In contrast, the likelihood that the ω of *SOCS1* was same across branches was not significantly different from the likelihood that ω varied across species (**Table 4**).

Table 4: Statistical analyses comparing maximum likelihood for each model in each gene.

Gene	Model	lnL	ω (dN/dS)	LRT=2(l1-l0)	df=np1-np2	*p* value
CASP1	M_0	-4150.761196	0.43620	27.39	14	**0.017119**
	M_1	-4137.065689	See Fig. 3A			
CD38	M_0	-3826.573694	0.56079	1634.63	21	**0.00001**
	M_1	-3009.260193	See Fig. 3B			
EEF1D	M_0	-5124.545955	0.13335	35.16	10	**0.000117**
	M_1	-5106.967187	See Fig. 3C			
SOCS1	M_0	-1295.005876	0.01711	13.58	14	0.481829
	M_1	-1288.218271	See Fig. 3D			

CASP1

The free ratio model (M_1) is a significantly (*p*=0.01712) better fit to *CASP1* than the fixed ratio model (M_0) (**Table 4**). All extant branches had ω less than 1, indicating the extant sequences showed signs of negative selection and supporting H_2 (**Figure 3A**). All of the common ancestors between ape species shared ω greater than 1, indicating positive selection occurred in the ancestral forms of these extant sequences. The common ancestor between *P. anubis*, which experiences T-cell loss but no progression to AIDS (Appendix A), and nonprogressor *C. sabaeus* also had a ω greater than 1, suggesting this ancestral sequence was shaped by positive selection.

CD38

The free ratio model (M_1) is a significantly (p=0.00001) better fit to *CD38* than the fixed ratio model (M_0) (**Table 4**). This indicates natural selection has shaped *CD38* differently in the study species. This sequence analysis had two extant branches with ω greater than 1*: P. anubis* (ω=1.1868) and *N. leucogenys* (ω=3.4707) (**Figure 3B**). This indicates positive selection may have shaped or still be acting upon *P. anubis* and *N. leucogenys CD38* sequences. Several ancestral nodes also had ω greater than 1, including the common ancestor between the nonprogressor *C. sabaeus* and the progressor *Papio/Macaca* clade, the common ancestor between *C. jacchus* and the Catarrhine clade, and the common ancestor between the Old World monkey clade and the ape clade. The high ω in the last common ancestor between *C. sabaeus* and the *Papio/Macaca* clade could be an indicator that nonsynonymous substitutions in *CD38* were positively selected for in a nonprogressor (*C. sabaeus*) and tolerator (*P. anubis*) lineage. Overall these results support H_2, indicating lineages were affected differently by natural selection, but not in a way that directly reflected progressor/nonprogressor status.

EEF1D

The free ratio model (M_1) is a significantly (p=0.000117) better fit to *EEF1D* than the fixed ratio model (M_0) (**Table 4**). All species and ancestral sequences had ω less than 0.6, indicating negative selection shaped this gene differently in all lineages and supporting H_2 (**Figure 3C**). The nonprogressor *C. sabaeus* had a ω of 0.3946, while progressor species *H. sapiens* and *P. troglodytes* had ω of 0.5685 and 0.1121,

respectively, indicating this gene was likely not being differentially selected for in nonprogressors compared to progressors.

SOCS1

The fit of the fixed ratio model (M_0) is not significantly different than that of the free ratio model (M_1) in *SOCS1* (**Table 4**). The ω of each extant species were all very close to 0, as were all the ancestral nodes except for the last common ancestor between *P. anubis* and *M. mulatta* (**Figure 3D**; **Table 4**). This suggests positive selection may have occurred during or before the divergence between *P. anubis* and *M. mulatta*. Overall these data support H_3, indicating that negative selection has similarly shaped all lineages except the *P. anubis* and *M. mulatta* clade, and that sequences were most likely not influenced by positive selection for nonprogressor status.

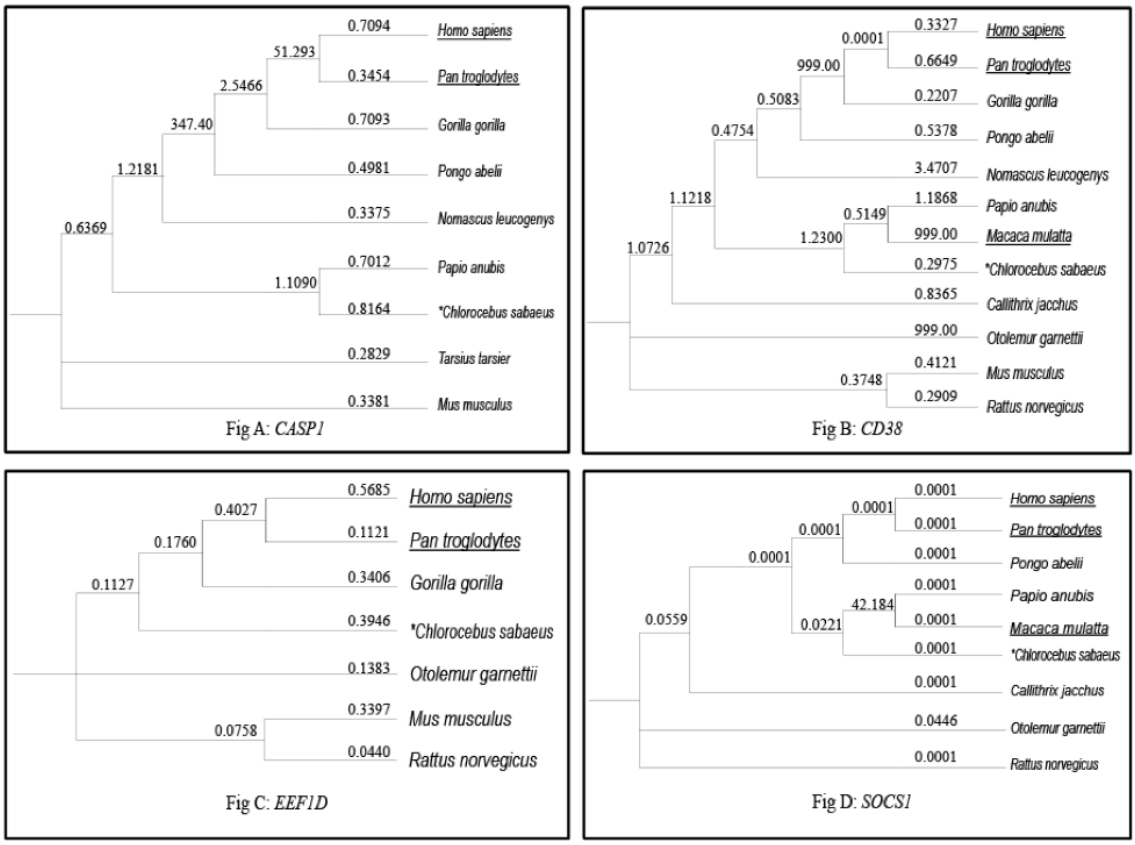


Figure 3: *ω* (dN/dS) values were estimated and plotted on known species phylogeny.

(*denotes nonprogressors, underline denotes progressors). (A) ω of CASP1 protein-coding sequences. (B) ω of CD38 protein-coding sequences. (C) ω of EEF1D protein-coding sequences. (D) ω of SOCS1 protein-coding sequences.

TFBS Analysis

Table 5A: TFBS identified in *CASP1* promoters with position (bp) in *H. sapiens* promoter (*denotes nonprogressors, underline denotes progressors, **bold underline** denotes difference from *H. sapiens*, - - - denotes unavailable sequence).

	SP-1	FREAC-3	GATA-2	GATA-2	SPI-1	FREAC-3	SPI-1	FREAC-3
Position	483-488	602-609	637-641	668-672	710-715	744-750	783-788	797-804
Homo sapiens	GGGAAC	TACTCTCC	TATCC	GGATA	GGGAAG	AGTGAGTA	AGGAAG	TACTTTCA
Pan troglodytes	GGGAAC	TACTCTCC	TATCC	GGATA	GGGAAG	AGTGAGTA	AGGAAG	TACTTTCA
Pongo abelii	GGGAAC	TACTCTCC	TATCC	GGATA	GGGAAG	AGTGAGTA	AGGAAG	TACTTTCA
Papio anubis	GGGAAC	TA**A**TCTC**T**	TATCC	**T**GATA	GGGAAG	AGTG**G**GTG	AGGAAG	TACTTTCA
****Mandrillus leucophaeus***	GGGAAC	TA**T**TCTCC	TATCC	GGATA	GG**A**AAG	AGT**AA**CT**G**	AGGAAG	TACTTTCA
Macaca fascicularis	GGGAAC	TA**T**TCTCC	TATCC	GGATA	GG**A**AAG	AGT**AA**GT**G**	AGGAAG	TACTTTCA
****Chlorocebus sabaeus***	GGGAAC	- - - - - - - -	- - - - - -	- - - - - -	- - - - - -	- - - - - - - -	- - - - - -	- - - - - - - -
****Cercocebus atys***	GGGAAC	TA**T**TCTCC	TATCC	GGATA	GG**A**AAG	AGT**AA**GTG	AGGAAG	TACTTTCA

Table 5B: TFBS identified in CD38 promoters with position (bp) in P. troglodytes promoter (*denotes nonprogressors, underline denotes progressors, **bold underline** denotes difference from H. sapiens, - - - denotes unavailable sequence).

	GMCSF	TCF-1	AP-2	PEA-3	IRF-1	E2A	PuF	TCF-1	a2(I)coll	PuF	IRF-1	NRE Box1	AP-2
Position	29-42	64-67	108-114	131-136	172-176	222-227	267-272	280-284	313-314	354-359	361-365	423-429	462-467
Homo sapiens	CATTT	CTTTT	GGCAGCCC	AGGAAGC	AAGGGA	CACCTGC	GGGTGGG	CACAG	ATTGG	GGGTGGG	AAGGGA	AGCCTCT	TCGCCC
Pan troglodytes	CATTT	CTTTT	GGCAGCCC	AGGAAGC	AAGGGA	CACCTGC	GGGTGGG	CACAG	ATTGG	GGGTGGG	AAGGGA	AGCCTCT	TCGCCC
Macaca fascicularis	CATTT	CTTTT	GGCAGCCC	AGGAAGC	AAGGGA	CA**T**CTGC	GGGTGGG	CACAG	ATTGG	GGGTGGG	AAGGGA	AGCCTCT	TCGCCC
Macaca mulatta	-----	CTTTT	GGCAGCCC	AGGAAGC	AAGGGA	CA**T**CTGC	GGGTGGG	CACAG	ATTGG	GGGTGGG	AAGGGA	AGCCTCT	TCGCCC
**Chlorocebus sabaeus*	CATTT	CTTTT	GGCAGCCC	AGGAAGC	AAGGGA	CATCTGC	GGGTGGG	CACAG	ATTGG	GGGTGGG	AAGGGA	AGCCTCT	TCGCCC
Saimiri boliviensis	CATTT	CTTTT	GGCAC**T**CC	AGGAAGC	AAGGGA	CATCTGC	GGGTGGG	CACAG	ATTGG	GGGTGGG	AAGGGA	**C**GCCC**T**G	TCGCCC

Table 5C: TFBS identified in EEF1D promoters with position (bp) in H. sapiens promoter (*denotes nonprogressors, underline denotes progressors, **bold underline** denotes difference from *H. sapiens*, - - - denotes unavailable sequence).

	SPI-1	MZF_1-4	MZF_1-4	AP2alpha	SPI-1
Position (bp)	652-657	283-288	789-794	801-809	880-885
Homo sapiens	GTTCCC	CGGGGA	TCCCCC	GCCCCCGGC	CTTCCC
Pan troglodytes	GTTCCC	CGGGGA	TCCCCC	GCCCCCGGC	CTTCCC
Gorilla gorilla	GTTCCC	C**A**GGGA	TCCCCC	GCCCCCGGC	CTTCCC
Nomascus leucogenys	GTTCCC	CGGGGA	TCCCCC	GCCCC**T**GGC	CTTCCC
Macaca nemestrina	GTTCCC	CG**CT**GA	TCCCCC	GCCCCCGGC	C**A**TCCC
Papio anubis	GT**G**CCC	CG**CT**GA	TCCCCC	GCCCCCGGC	C**A**TCCC
****Cercocebus atys***	GTTCCC	CG**TT**GA	TCCCCC	GCCCCCGGC	C**A**TCCC
****Chlorocebus sabaeus***	GTTCC**T**	CG**C**GGA	T**T**CCCC	GCCCCCGGC	C**A**TCCC

Table 5D: TFBS identified in *SOCS1* promoters with position (bp) in *H. sapiens* promoter (*denotes nonprogressors, underline denotes progressors, **bold underline** denotes difference from *H. sapiens*, - - - denotes unavailable sequence).

	AP2alpha	SPI-1
Position (bp)	998-1006	1014-1023
Homo sapiens	GCCCGCGGC	GCCCCGCCGC
Gorilla gorilla	GCCCGCGGC	GCCCCGCCGC
****Chlorocebus sabaeus***	GCCCGCGGC	**A**CCCCGCCGC
Callithrix jacchus	G**AG**CG**G**G**CG**	G**GGT**CG**A**CG**G**
Rattus norvegicus	**AG**CCGCGGC	G**T**CCCGCCGC
Mus musculus	**AG**CCGCGGC	G**T**CC**T**GCC - -

CASP1

Eight TFBS were identified in *CASP1* promoters: three SP1-1, three FREAC-3, and two GATA-2 (**Table 5A**). Nonprogressor species, *C. sabaeus* and *C. atys*, either share complete TFBS sequence with all other species (Position 483-488, 637-641, 783-788, 797-804), or share substitutions with progressor species (Position 602-609, 668-672, 710-715, 744-750). Positions for *CASP1, EEF1D,* and *SOCS1*, refer to bp from beginning of SwitchGear Genomics *H. sapiens* promoter. This evidence supports H_2, as substitutions are consistent with known species relationships. This suggests species divergence most likely shaped these sequences rather than convergent evolution on the nonprogressor immune type.

CD38

13 TFBS were identified in *CD38* promoters: GMCSF, two TCF-1, two AP-2, PEA-3, two IRF-1, E2A, two PuF, a2(I)coll, and NREboxl (**Table 5B**). The nonprogressor species, *C. sabaeus,* share complete TFBS sequence with all other species for 10 TFBS (Position 29-42, 64-67, 131-136, 172-176, 267-272, 280-284, 313-314, 354-359, 361-365, 462-467). Positions for *CASP1* refer to bp from beginning of *P. troglodytes* promoter sequence published by Ferrero and colleagues (2004). All other mutations in TFBS are in other species and reflected species phylogeny, supporting H_2. This might indicate species divergence shaped these sequences rather than convergent evolution on the nonprogressor immune type.

EEF1D

Five TFBS were identified *EEF1D* promoters: two SPI-1, two MZF_1-4, and AP2alpha (**Table 5C**). The nonprogressor species, *C. sabaeus* and *C. atys*, share no mutations in TFBS that distinguished them from progressors and other species. The mutations are consistent with known species relationships, supporting H_2 and suggesting species divergence may have shaped these sequences rather than convergent evolution on the nonprogressor immune type.

SOCS1

Two TFBS were identified in *SOCS1* promoters: AP2alpha, and SPI-1 (**Table 5D**). The nonprogressor species, *C. sabaeus*, share the complete sequence for AP2alpha with progressor *H. sapiens*. The *C. sabaeus* sequence has one unique mutation in SPI-1 that distinguishes it from all other species investigated. This tentatively supports H_1, but requires further investigation as only one nonprogressor species was sampled.

DISCUSSION

In 1947 the modern evolutionary synthesis of Charles Darwin's theory of evolution by natural selection with Gregor Mendel, Theodosius Dobzhansky, and other researchers' theories of genetic variation and inheritance laid the groundwork for the contemporary evolutionary genetic perspective on human susceptibility to disease (Dobzhansky 1937; Huxley 1942; Smocovitis, 1992). While researchers are attempting to use this perspective to identify genes that leave humans vulnerable to pathogens, an alternative approach seeks to identify genes in our closest living relatives that may contribute to their ability to coexist with similar pathogens (Nesse *et al.* 2010; Sironi *et al.* 2015). We can begin to search for a connection between these genes, past environmental pressures, and an advantageous phenotype by asking how evolutionary history has shaped the genotypes of extant species that can coexist with certain pathogens. This provides an evolutionary context for the fixation of certain alleles in these extant species. One hypothesis is that certain substitutions increased the reproductive fitness of species in environments shared with a pathogen and rose in frequency due to positive selection. The use of evolutionary methods to test functional regions of the genome allows us to target these genes that may play a role in tolerance to pathogens. If we can identify genotypes that were or are under positive selection and that are associated with advantageous phenotypes, we can go on to experimentally test the function of these genotypes *in vitro* or *in vivo* and perhaps someday manipulate the identified mechanisms as a medical intervention. In order to identify molecular mechanisms that contribute to nonprogressor tolerance of SIV, the present study used evolutionary genetic methods to test for convergent evolution, positive selection, and

distinguishing variants in the promoter and protein-coding regions of *CASP1*, *CD38*, *EEF1D*, and *SOCS1*.

Objective 1

In order to test for evidence of strong convergent evolution, evolutionary histories of genes in different species were compared to the evolutionary histories of the species themselves. By using the molecular clock theory and estimated mutation rates we can infer what mutations might have occurred over the evolution of the extant sequences and estimate what ancestral sequences might have been at different points in the past. We can then extrapolate a possible gene tree for comparison with a species tree, letting us infer when mutations might have become fixed in ancestral sequence compared to overall species divergence. This helps us investigate co-occurrence of an environmental factor (such as a pathogen like SIV) with the emergence of mutations that distinguish species with advantageous phenotypes (like nonprogressors) from species with other phenotypes (like progressors). When gene trees and species trees appear the same it indicates that the interspecific polymorphisms were shaped by known species relationships. Species trees and gene trees may appear different for a variety of reasons. The gene may be vital to survival or reproduction and thus remained mostly conserved over time, only accumulating a few mutations, due to negative selection. The genes may have also accumulated interspecific polymorphisms after the species divergence due to positive selection. This positive selection could take the form of convergent or parallel evolution with other species subjected to similar environmental pressures. I intended to test this last hypothesis by examining gene trees for deviation from species trees with respect to progressors and nonprogressors. I expected to see all

nonprogressor species form monophyletic clades, which would provide evidence of strong convergent evolution of a shared genotype and a phenotype. However this was not found to be the case in any of my gene trees, as only one orthologous nonprogressor species (*C. sabaeus*) and one tolerator species that doesn't progress to AIDS but still experiences chronic T cell depletion (*P. anubis*) were available for each gene. Furthermore, all gene trees demonstrated evidence that species relationships and untested factors were more likely responsible for the evolutionary histories of these genes than positive selection for the nonprogressor immune type. This suggests that there were few sites distinguishing the nonprogressor and tolerator species from the progressor species, and that these genes may not be ideal targets for functional investigation.

There were limitations to the ability of phylogenetic analyses to detect convergent evolution in nonprogressor species. Orthologous sequences for nonprogressor species were limited and only *C. sabaeus* was represented nonprogressors in these analyses. Thus, there were no other nonprogressors with which this species could form a monophyletic clade, immediately eliminating H_1. To further investigate positive selection as the driving force of nonprogressor sequence divergence, future research should look to sequence and genotype other progressor and nonprogressor species' protein-coding regions, and perhaps examine other genes identified as differentially expressed in progressors and nonprogressors. Another possible avenue of research could include investigation of convergent evolution in these genes in human viral nonprogressors (VNP). While there is a large amount of data on nonprogressor nonhuman primates, there is comparatively very little information on the

small subset of humans who also can survive rates of viral replication without medical intervention (Rotger *et al.* 2011). Fewer studies have been conducted on these so-called viremic non-progressors (VNP) or long-term non-progressors (LTNP) because the proportion of individuals is so small (~0.1% of humans), and because many of these individuals are unaware of their HIV status (Rotger et al. 2011). By investigating signatures of convergent evolution in immune genes of human and nonhuman nonprogressors we may be able to conclude whether these species have undergone convergent evolution in their development of an adaptation (or exaptation) to environments with ancestral lentiviruses.

Objective 2

In order to test for positive selection in the protein-coding regions, ratios of nonsynonymous to synonymous substitution rates (ω) were estimated. By inferring ancestral sequences through the use of known species trees and comparing the rates of nonysnonymous and synonymous substitutions in each lineage, it is possible to estimate ω across the lineages represented in the given phylogeny, demonstrating whether selective pressures have differently influenced lineages. An ω less than 1 on all branches would indicate negative selection in all lineages. This suggests the sequences are highly conserved and may have a function that is closely tied with fitness, leading to the loss of most nonsynonymous substitutions. In contrast, an ω of 1 would indicate neutral evolution and suggest the gene does not have a close relationship with reproductive fitness, leading to the equal fixation of nonsynonymous and synonymous substitutions. However if at some point positive selection fixed nonsynonymous mutations in specific lineages and not others, one would expect to see higher rates of

nonsynonymous mutations in those lineages, leading to an ω greater than 1. Thus, by comparing ω in the lineages that lead to extant species with a particular phenotype (such as nonprogressors) to rates in lineages leading to extant species without that phenotype (nonprogressors) we can investigate a connection between certain interspecific polymorphisms, phenotypic differences, and differences in selective pressures over time. My project did this by comparing the fit of a fixed ratio model (M_0) to that of a free ratio model (M_1) and by examining the ω values on each branch of the tree for signs of selection in nonprogressor lineages.

The results of the present study demonstrate that although ω varies across primates in *CASP1*, *CD38*, and *EEF1D*, there is little evidence of positive selection in nonprogressor lineages. The dN/dS analysis of the protein-coding regions of *SOCS1* indicate this region was shaped similarly by negative selection in all species, aside from possible positive selection in the last common ancestor of *P. anubis* and *M. mulatta.* There are a few limitations to the scope of these results. First, only one orthologous nonprogressor sequence (*C. sabaeus*) and one tolerator sequence (*P. anubis*) were available online and thus I was not able to draw a connection between shared nonprogressor phenotype and signatures of positive selection. This could be circumvented by, as discussed above, personally sequencing more orthologous samples in other progressor and nonprogressor species. Another limitation to this approach is that it assumes the nonprogressors have diverged from their common ancestors with progressors and coevolved with ancestral forms of SIV to develop these adaptations. While each nonprogressor lineage could have independently evolved mutations in these genes due to positive selection by an environment shared with SIV, it is also possible

that the last common ancestor of all Old World monkeys and apes could have evolved the adaptation within its own lineage, and then progressor species lost this adaptation due to drift or positive selection by novel environmental pressures. In order to fully characterize the evolutionary relationship between SIV, progressor/nonprogressor immune phenotypes, and genotypes, both these possibilities must be thoroughly explored.

Objective 3

Finally, *cis*-regulatory regions were examined for variation that distinguishes species with different levels of expression in the chronic phase of infection. Regulation of expression of immune genes is closely tied to the function of the encoded protein, and by searching the promoter regions for interspecific polymorphisms that distinguish species with different phenotypes we can identify possible connections between genotype and phenotype to be explored experimentally (Loisel *et al.* 2006). Although there is no analog for the codon system in non-coding regions to infer functionality of point mutations, we can examine TFBS loci in the promoter to identify variants associated with differential expression. If there is variation in these binding sites, it is possible that could affect the level of transcription and in turn affect the function of the produced protein. As Bosinger *et al.* 2009 identified *CASP1*, *CD38, EEF1D,* and *SOCS1* as being differentially expressed in progressor and nonprogressor species, these genes were ideal candidates for exploring polymorphisms in *cis*-regulatory regions that may be connected to this differential expression and possibly to differential disease progression.

While no polymorphisms were found that distinguished nonprogressors from all progressors in *CASP1, CD38,* and *EEF1D,* the *SOCS1* promoter region in *C. sabaeus* demonstrated a unique mutation in one TFBS, which distinguished it from all other species. Unfortunately there was only one orthologous nonprogressor promoter for that assay, and thus has ambiguous significance as a mutation distinguishing nonprogressors from progressors. This limitation could be circumvented by sequencing and identifying polymorphisms in other progressor and nonprogressor species' regulatory regions to better characterize between species variation. Furthermore, to truly eliminate selection for nonprogressor polymorphisms in the regulatory regions of the genes as a contributor to the nonprogressor immune type we must not only examine *cis*-regulatory regions but also *trans*-regulatory regions.

My project utilized putatively identified *cis*-regulatory regions, but ideally all functional regulatory regions would be identified and examined for variation that distinguished nonprogressors from progressors. Further research should also search for TFBS in different loci in alignments of orthologous regulatory regions, because though my alignments demonstrated relatively high conservation it is possible that TFBS do not occur at the same locus in each species.

Functional Significance

In order to identify possible adaptations, genotypes need to be tied to a functional phenotype and differential effects on fitness. For positive selection and thus adaptive evolution to take place, genomic variation must have an effect on the function of the produced protein and this function must provide a fitness benefit to the individual in a specific environment. While we can use phylogenetic analyses, dN/dS analyses,

and TFBS analyses to help us sort through an enormous amount of genetic data and identify target genes for further investigation, these genes must be tied to a functional difference in order to support an adaptive explanation. There are several ways functional effects of genetic polymorphisms can be investigated. First, we can examine possible functional changes caused by amino acid substitutions that distinguish nonprogressors from all progressors. Polyphen-2 version 2.2.2 can be used to predict the effects of transfecting nonprogressor single nucleotide polymorphisms into human sequences (Adzhubei, 2010). Further examination of how these substitutions change the polarity, charge, and size of a protein can help identify changes in the protein's function. The functional effects of these amino acid substitutions can also be investigated through experimental studies involving the genetic engineering of T cell lines with nonprogressor and progressor polymorphisms (Hajeer and Hutchinson, 2001). If a cell line with a genotype associated with nonprogressor immune response is found to produce a different protein than a cell line with a progressor genotype, it can be hypothesized that the nonprogressor polymorphisms may play a role in the mechanisms that contribute to tolerance to the virus. Functional effects of substitutions in regulatory regions can also be examined through experimental cell-line manipulation of pattern-matched TFBS, as well as knockout studies in lab mice (Loots, 2008).

One possible but untested method for identifying functional changes would involve the application of the new CRISPR/cas9 system of gene editing by altering a progressor sequence to include nonprogressor variants and challenging the subjects with SIV to see how the variants affect progression to AIDS (Mali *et al.* 2013). However examining function in live primates is ethically controversial and thus very hard to

complete. Nevertheless, by combining these function analyses with long-term monitoring of the co-occurrence of SIV infection and opportunistic infections in wild progressor and nonprogressor populations, it may eventually be possible to connect changes in genotype and phenotype with changes in individual fitness.

CONCLUSIONS

By searching for signs of positive selection in genes that contribute to the immune mechanisms that allow nonprogressors to coexist with SIV, this research attempted to connect the evolutionary history of these genes to an observable phenotypic difference, with the intent of identifying target pathways for future medical intervention. Furthermore, by testing an adaptive explanation for the nonprogressor immune profile I sought to further our understanding of how natural selection and coevolution with lentiviruses has shaped primate immune responses in general. While the results indicate that nonprogressor protein-coding and *cis*-regulatory regions of *CASP1, CD38, EEF1D,* and *SOCS1* do not provide evidentiary support of positive selection compared to progressor species, future research may shed insight on the functional significance of distinguishing polymorphisms in these and other regulatory and protein-coding regions of genes differentially expressed in nonprogressors and progressors.

APPENDIX

A. Categorization of nonhuman primate immune responses to HIV/SIV

Species	Progressor/nonprogressor Status	Reference
Callithrix jacchus	No known infection	
Cercocebus atys	Nonprogressor	Peeters *et al.* 1994
Cercopithecus mona	Nonprogressor	Peeters *et al.* 2002
Cercopithecus nictitans	Nonprogressor	Peeters *et al.* 2002
Chlorocebus aethiops	Nonprogressor	Peeters *et al.* 2002
Chlorocebus pygerythrus	Nonprogressor	Peeters *et al.* 2002
Chlorocebus sabaeus	Nonprogressor	Bosinger *et al.* 2009
Chlorocebus tantalus	Nonprogressor	Peeters *et al.* 2002
Colobus guereza	Nonprogressor	Peeters *et al.* 2002
Macaca fascicularis	Progressor (AIDS in lab)	Habis *et al.*1999
Macaca mulatta	Progressor (AIDS in lab)	Bosinger *et al.* 2009
Macaca nemestrina	Progressor (AIDS in lab)	Ho *et al.* 2009
Mandrillus leucophaeus	Nonprogressor	Peeters *et al.* 2002
Mandrillus sphinx	Nonprogressor	Greenwood *et al.* 2014
Miopithecus talapoin	Nonprogressor	Peeters *et al.* 2002
Nomascus leucogenys	No known SIV infection	
Otolemur garnettii	No known SIV infection	
Pan troglodytes	Progressor (high mortality, some AIDS)	Etienne *et al.* 2011
Papio anubis	Tolerator (T cell depletion, no AIDS)	Locher *et al.* 2002
Pongo abelii	No known SIV infection	
Gorilla gorilla	Naturally infected (Immune response unknown)	D'arc *et al.* 2015 Etienne *et al.* 2012
Saimiri boliviensis	No known SIV infection	
Tarsius tarsier	No known SIV infection	

REFERENCES

(2013). Types of HIV/AIDS Antiretroviral Drugs. NIH: National Institute of Allergy and Infectious Diseases.

(2014a). Zidovudine (Retrovir). *NIH: AIDSinfo.* http://aidsinfo.nih.gov/drugs/4/zidovudine/0/patienthttp://aidsinfo.nih.gov/drugs/4/zidovudine/0/patient

(2014b). Education Materials: HIV/AIDS Fact Sheets: Drug Resistance. *NIH: AIDSinfo.* http://aidsinfo.nih.gov/education-materials/fact-sheets/21/56/drug-resistancehttp://aidsinfo.nih.gov/education-materials/fact-sheets/21/56/drug-resistance

(2015a). Fact Sheet. *UNAIDS*. http://www.unaids.org/en/resources/campaigns/2014/2014gapreport/factsheet/

(2015b). AIDSinfo: World Overview. *UNAIDS.* http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature methods* 7, 4: 248-249.

Apetrei, C., Kaur, A., Lerche, N. W., Metzger, M., Pandrea, I., Hardcastle, J., ... & Marx, P. A. (2005). Molecular epidemiology of simian immunodeficiency virus SIVsm in US primate centers unravels the origin of SIVmac and SIVstm. *Journal of virology*, *79*(14), 8991-9005.

Bosinger, S.E., Li, Q., Gordon, S.N., Klatt, N.R., Duan, L., Xu, L., Francella, N., Sidahmed, A.,Smith, A.J., Cramer, E.M., Zeng, M., Masopust, D., Carlis, J.V., Ran, L., Vanderford, T.H., Paiardini, M., Isett, R.B., Baldwin, D.A., Else, J.G., Staprans, S.I., Silvestri, G., Haase, A.T., and Kelvin, D. J. (2009). Global genomic analysis reveals rapid control of a robust innate response in SIV-infected sooty mangabeys. *The Journal of clinical investigation* 119, 12: 3556.

Buchbinder, S. P., Mehrotra, D. V., Duerr, A., Fitzgerald, D. W., Mogg, R., Li, D., Gilbert, P.B., Lama, J.R., Marmor, M., del Rio, C., McElrath, M.J., Casimiro, D.R., Gottesdiener, K.M., Chodakewitz, J.A., Corey, L., and Robertson, M.N. (2008). Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *The Lancet* 372, 9653: 1881-1893.

Buehler, M.D., & Berkelman, M.R. L. (1990). 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults.

Chahroudi, A., Bosinger, S.E., Vanderford, T.H., Paiardini, M., & Silvestri, G. (2012). Natural SIV hosts: showing AIDS the door. *Science* 335, 6073: 1188–93.

Clavel, F., and Hance, A. J. (2004). HIV drug resistance. *New England Journal of Medicine* 350, 10: 1023-1035.

Compton A.A., Malik H.S., and Emerman M. (2013). Host gene evolution traces the evolutionary history of ancient primate lentiviruses. *Philosophical Transactions of the Royal Society B* 368: 1-10.

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Kähäri, A.K., Keenan, S., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Overduin, B., Parker, A.,Patricio, M., Perry, E., Pignatelli, M., Riat, H.S., Sheppard, D., Taylor, K.,Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Aken, B.L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S.M.J., Spudich, G.,Trevanion, S.J., Yates, A., Zerbino, D.R., and Flicek, P. (2015). Ensembl 2015. *Nucleic acids research* 43, D1: D662-D669.

D'arc, M., Ayouba, A., Esteban, A., Learn, G.H., Boué, V., Liegeois, F., Etiennea, L., Tagge, N., Leendertzf, F.H. Boeschg, C., Madinda, N.F., Robbinsg, M.M, Grayi, M., Cournila, A., Ooms, M. Letko, M., Simon, V.A., Sharpm, P.M., Hahn, B.H., Delaportea, E., Ngolen, E.M., and Peeters, M. (2015). Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proceedings of the National Academy of Sciences*, *112*(11), E1343-E1352.

De Clercq, E. (2004). Antiviral drugs in current clinical use. *Journal of Clinical Virology* 30, 2: 115-133.

Dobzhansky, T. (1937). *Genetics and the Origin of Species* (Vol. 11). Columbia University Press.

Etienne, L., Nerrienet, E., LeBreton, M., Bibila, G.T., Foupouapouognigni, Y., Rousset, D., Nana , A., Djoko, C.F., Tamoufe , U., Aghokeng, A.F., Mpoudi-Ngole6 , E., Delaporte, E., Martine Peeters1, M., Wolfe, N.D., and Ayouba, A. (2011). Characterization of a new simian immunodeficiency virus strain in a naturally infected Pan troglodytes troglodytes chimpanzee with AIDS related symptoms. *Retrovirology*, *8*(4).

Etienne, L., Locatelli, S., Ayouba, A., Esteban, A., Butel, C., Liegeois, F., Aghokeng, A., Delaporte, E., Ngole, E.M., and Peeters, M. (2012). Non-invasive follow-up of simian immunodeficiency virus infection in wild-living non-habituated western lowland gorillas in Cameroon.*Journal of virology*, JVI-01186.

Fauci, A. S. (1988). The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science*, *239*(4840), 617-622.

Ferrero, E., Orciani, M., Vacca, P., Ortolan, E., Crovella, S., Titti, F., Saccucci, F., and Malavasi, F. (2004). Characterization and phylogenetic epitope mapping of CD38 ADPR cyclase in the cynomolgus macaque. *BMC immunology* 5, 1: 21.

Fischl, M.A., RIchman, D.D., Greico, M.H., Gottlieb, M.S., Volberding, P.A., Laskin, O.L., Leedom, J.M., Groopman, J.E., Mildvan, D., Schooley, R.T., Jackson, G.G., Durack, D.T., King, D., and The AZT Collaborative Working Group. (1987). The Efficacy of Azidothymidine (AZT) in the Treatment of Patients with AIDS and AIDS-Related Complex. *New England Journal of Medicine* 317, 4:185-191.

Gene Cards: Promoter Reporter. *SwitchGear Genomics.* Accessed May 2015.

Glazko, G. V., & Nei, M. (2003). Estimation of divergence times for major lineages of primate species. *Molecular biology and evolution*, *20*(3), 424-434.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic acids research* 38, suppl 2: W695-W699.

Greenwood, E.J., Schmidt, F., Liegeois, F., Kondova, I., Herbert, A., Ngoubangoye, B., Rouet, F., Heeney, J.L. (2014). Loss of memory CD4+ T-cells in semi-wild mandrills (Mandrillus sphinx) naturally infected with species-specific simian immunodeficiency virus SIVmnd-1. *Journal of General Virology*, *95*(Pt 1), 201-212.

Habis, A. I. D. S., Baskin, G. B., Murphey-Corb, M., & Levy, L. S. (1999). Simian AIDS-associated lymphoma in rhesus and cynomolgus monkeys recapitulates the primary pathobiological features of AIDS-associated non-Hodgkin's lymphoma. *AIDS research and human retroviruses*, *15*(15), 1389-1398.

Hajeer, A.H., and Hutchinson, I.V. (2001). Influence of TNFα gene polymorphisms on TNFα production and disease. *Human immunology*, *62*(11), 1191-1199.

Hazenberg, M. D., Otto, S.A., van Benthem, B.H., Roos, M.T., Coutinho, R.A., Lange, J.M., Hamann, D., Prins, M., Miedema, F. (2003). Persistent immune activation in HIV-1 infection is associated with progression to AIDS. *Aids* 17: 13, 1881-1888.

Ho, O., Larsen, K., Polacino, P., Li, Y., Anderson, D., Song, R., Ruprecht, R.M., Hu, S. L. (2009). Pathogenic infection of Macaca nemestrina with a CCR5-tropic subtype-C simian-human immunodeficiency virus. *Retrovirology*, *6*(1), 65.

Huxley, J. (1942). Evolution. The Modern Synthesis. *Evolution. The Modern Synthesis.*

Jacquelin, B., Mayau, V., Targat. B., Liovat, A., Kunkel, D., Petitjean, G., Dillies, M., Roques, P., Butor, C., Silvestri, G., Giavedoni, L.D., Lebon, P., Barré-Sinoussi, F., Benecke, A., & Müller-Trutwin, M. (2009). Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. *Journal of Clinical Investigation* 119: 12, 3544-3555.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 12: 1647-1649.

Klatt, N. R., Silvestri, G., & Hirsch, V. (2012). Nonpathogenic simian immunodeficiency virus infections. *Cold Spring Harbor perspectives in medicine 2*, 1: a007153.

König R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M., Irelan, J. T., Chiang, C., Tu, B.P., De Jesus, P.D., Lilley, C.E., Seidel, S., Opaluch, A.M., Caldwell, J.S., Weitzman, M.D., Kuhen, K.L., Bandyopadhyay, S., Ideker, T., Orth, A.P., Miraglia, L.J., Bushman, F.D., Young, J.A., and Chanda, S.K. (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, *135*(1), 49-60.

Letvin, N.L., Daniel, M.D., Sehgal, P.K., Desrosiers, R.C., Hunt, R.D., Waldron, L.M., Mackey, J.J., Schmidt, D.K., Chalifoux, L.V., and King, N.W. (1985). Induction of AIDS-like disease in macaque monkeys with T-cell tropic retrovirus STLV-III. *Science* 230, 4721: 71-73.

Levy, J. A. (2001). The importance of the innate immune system in controlling HIV infection and disease. *Trends in immunology* 22, 6: 312-316.

Lieberman, J., Skolnik, P. R., Parkerson III, G. R., Fabry, J. A., Landry, B., Bethel, J., Kagen, J. and the DATRI 006 Study Team. (1997). Safety of autologous, ex vivo-expanded human immunodeficiency virus (HIV)-specific cytotoxic T-lymphocyte infusion in HIV-infected patients. *Blood* 90, 6: 2196-2206.

Lio, P., & Goldman, N. (1998). Models of molecular evolution and phylogeny. *Genome research*, 8, 12: 1233-1244.

Locher, C. P., Witt, S. A., Herndier, B. G., Tenner-Racz, K., Racz, P. and Levy, J. A. (2001), Baboons as an animal model for human immunodeficiency virus pathogenesis and vaccine development. *Immunological Review*s, 183: 127–140. doi: 10.1034/j.1600-065x.2001.1830111.x

Loisel, D. A., Rockman, M. V., Wray, G. A., Altmann, J., and Alberts, S. C. (2006). Ancient polymorphism and functional variation in the primate MHC-DQA1 5′ cis-regulatory region. *Proceedings of the National Academy of Sciences*, *103*(44), 16331-16336.)

Loots, G. G. (2008). Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. *Advances in genetics*, *61*, 269-293.

Malavasi, F., Deaglio, S., Funaro, A., Ferrero, E., Horenstein, A. L., Ortolan, E., Vaisitti, T., and Aydin, S. (2008). Evolution and function of the ADP ribosyl cyclase/CD38 gene family in physiology and pathology. *Physiological reviews* 88, 3: 841-886.

Mali, P., Esvelt, K. M., & Church, G. M. (2013). Cas9 as a versatile tool for engineering biology. *Nature methods*, *10*(10), 957-963.

Marx, P. A., Alcabes, P. G., & Drucker, E. (2001). Serial human passage of simian immunodeficiency virus by unsterile injections and the emergence of epidemic human immunodeficiency virus in Africa. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356* 1410: 911-920.

McChesney, M.B., & Miller, C.J. (2013). New directions for HIV vaccine development from animal models. *Current opinion in HIV and AIDS* 8, 5: 376.

Nesse, R.M., Bergstrom, C.T., Ellison, P. T., Flier, J.S., Gluckman, P., Govindaraju, D.R., Omennh, G.S., Perlmani, R.L., Schwartzj, M.D., Thomask, M.G., Stephen C. Stearnsl, and Valle, D. (2010). Making evolutionary biology a basic science for medicine. *Proceedings of the National Academy of Sciences* 107, suppl 1: 1800-1807.

Nielsen, M. H., Pedersen, F. S., & Kjems, J. (2005). Molecular strategies to inhibit HIV-1 replication. *Retrovirology 2*, 10: 1-20.

Peeters, M., & Courgnaud, V. (2002). Overview of primate lentiviruses and their evolution in non-human primates in Africa. *HIV sequence compendium*, 2-23.

Robb, M. (2011). HIV vaccine development: a new beginning. *Expert review of vaccines* 10, 7: 925-927.

Rotger, M., Dalmau, J., Rauch, A., McLaren, P., Bosinger, S.E., Martinez, R., Sandler, N.G., Roque, A., Liebner, J., Battlegay, M., Bernasconi, E., Descombes, P., Erkizia, I., Fellay, J., Hirschel, B., Miro, J.M., Palou, E., Hoffman, M., Massanella, M., Blanco, J., Woods, M., Gunthard, H.F., de Bakker, P., Douek, D.C., Silvestri, G., Martinez-Picado, J., and Telenti, A. (2011). Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. *The Journal of clinical investigation* 121, 6: 2391.

Sandelin, A., Wasserman, W.W., and Lenhard, B. (2004). ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic acids research* 32, suppl 2: W249-W252.

Shiver, J.W., Fu, T.M., Chen, L., Casimiro, D.R., Davies, M.E., Evans, R.K., Zhang, Z.Q., Simon, A.J., Trigona, W.L., Dubey, S.A, Huang L., Harris, V.A., Long, R.S., Liang, X., Handt, L., Schleif, W.A., Zhu, L., Freed, D.C., Persaud, N.V., Guan, L., Punt, K.S., Tang, Chen, A.M., Wilson, K.A., Collins, K.B., Heidecker, G.J., Fernandez, V.R., Perry, H.C., Joyce, J.G., Grimm, K.M., Cook, J.C., Keller, P.M., Kresock, D.S., Mach, H., Troutman, R.D., Isopi, L.A., Williams, D.M., Xu, Z., Bohannon, K.E., Volkin, D.B., Montefiori, D.C., Miura, A., Krivulka, G.R., Lifton, M.A., Kuroda, M.J., Schmitz, J.E., Letvin, N.L., Caulfield, M.J., Bett, A.J., Youil, R., Kaslow, D.C., and Emini, E.A. (2002). Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature* 415, 6869: 331-335.

Silvestri, G., Paiardini, M., Pandrea, I., Lederman, M.M., and Sodora, D.L. (2007). Understanding the benign nature of SIV infection in natural hosts. *Journal of Clinical Investigation* 117, 11: 3148-3154.

Sironi, M., Cagliani, R., Forni, D., and Clerici, M. (2015). Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nature Reviews Genetics*, *16*(4), 224-236.

Smith, K., and Housseau, F. (2015). An Unexpected Journey: How Cancer Immunotherapy Has Paved the Way for an HIV-1 Cure. *Discovery Medicine* 19, 104.

Smocovitis, V. B. (1992). Unifying biology: the evolutionary synthesis and evolutionary biology. *Journal of the History of Biology*, *25*(1), 1-65.

Soltis, P. S., & Soltis, D. E. (2003). Applying the bootstrap in phylogeny reconstruction. *Statistical Science*, *18*(2), 256-267.

Takahashi, R., Nishimoto, S., Muto, G., Sekiya, T., Tamiya, T., Kimura, A., Morita, R., Asakawa, M., Chinen, T., and Yoshimura, A. (2011). SOCS1 is essential for regulatory T cell functions by preventing loss of Foxp3 expression as well as IFN-γ and IL-17A production. *The Journal of experimental medicine* 208, 10: 2055-2067.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5:molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 28, 10: 2731-2739.

UniProt Consortium. (2014). UniProt: a hub for protein information. *Nucleic Acids Research*, gku989.

Velu, V., Titanji, K., Zhu, B., Husain, S., Pladevega, A., Lai, L., Vanderford, T.H., Chennareddi, L., Silvestri, G., Freeman, G.J, Ahmed, R., and Amara, R.R. (2009). Enhancing SIV-specific immunity in vivo by PD-1 blockade. *Nature* 458, 7235: 206-210.

Warren, K., Wei, T., Li, D., Qin, F., Warrilow, D., Lin, M.H., Sivakumaran, H., Apolloni, A., Abbot, C.M., Jones, A., Anderson, J.L., and Harrich, D. (2012). Eukaryotic elongation factor 1 complex subunits are critical HIV-1 reverse transcription cofactors. *Proceedings of the National Academy of Sciences* 109, 24: 9587-9592.

Xu, B., and Yang, Z. (2013). PAMLX: A graphical user interface for PAML. *Molecular biology and evolution* 30, 12: 2723-2724.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* 24, 8: 1586-1591.