Adaptation of a Diverse Simian Immunodeficiency Virus Population to a New Host Is Revealed through a Systematic Approach to Identify Amino Acid Sites under Selection

Thomas H. Vanderford,† Linda J. Demma,*†[1] Mark B. Feinberg,†‡§[2] Silvija I. Staprans,†‡§[2] and John M. Logsdon Jr*‖*

*Program in Population Biology, Evolution, and Ecology, Emory University; †Emory Vaccine Center, Emory University; ‡Department of Medicine, Emory University; §Department of Microbiology and Immunology, Emory University; and ‖Roy J. Carver Center for Comparative Genomics, Department of Biological Sciences, University of Iowa

Simian immunodeficiency viruses (SIV) have had considerable success at crossing species barriers; both human immunodeficiency virus (HIV)-1 and HIV-2 have been transmitted on multiple occasions from SIV-infected natural host species. However, the precise evolutionary and ecological mechanisms characterizing a successful cross-species transmission event remain to be elucidated. Here, in addition to expanding and clarifying our previous description of the adaptation of a diverse, naturally occurring SIVsm inoculum to a new rhesus macaque host, we present an analytical framework for understanding the selective forces driving viral adaptation to a new host. A preliminary analysis of large-scale changes in virus population structure revealed that viruses replicating in the macaques were subject to increasing levels of selection through day 70 postinfection (p.i.), whereas contemporaneous viruses in the mangabeys remained similar to the source inoculum. Three different site-by-site methods were employed to identify the amino acid sites responsible for this macaque-specific selection. Of 124 amino acid sites analyzed, 3 codons in V2, a 2–amino acid shift in an N-linked glycosylation site, and variation at 2 sites in the highly charged region were consistently evolving under either directional or diversifying selection at days 40 and 70 p.i. This strong macaque-specific selection on the V2 loop underscores the importance of this region in the adaptation of SIVsm to rhesus macaques. Due to the extreme viral diversity already extant in the naturally occurring viral inoculum, we employed a broad range of phylogenetic and numerical tools in order to distinguish the signatures of past episodes of selection in viral sequences from more recent selection pressures.

Introduction

Transmission of pathogens from animal reservoirs to humans is responsible for many of the world's most deadly and costly epidemics (Hirsch et al. 1989; Gao et al. 1999; Holmes 2005; Leroy et al. 2005; Parrish and Kawaoka 2005). In some cases the mechanisms by which animal pathogens cross the species barrier are relatively easy to identify (Qu et al. 2005; Wang et al. 2005); however, zoonotic transmissions are most often due to very complex interactions between ecological, evolutionary, biochemical, and sociological factors (Hahn et al. 2000; Palese 2004). Thus, not only is it difficult to identify the causes of past zoonotic epidemics, but these complex interactions may also hinder our ability to predict future epidemics. The development of a framework for identification of critical dynamical and biochemical factors influencing the emergence of infectious diseases is vital to our understanding of epidemics to come.

Upon transmission to a new host species, viruses must usually adapt to a new genetic and immunologic environment in order to replicate and spread to other individuals within the species (Webby et al. 2004). The high mutation and replicative rates of RNA viruses such as human immunodeficiency virus (HIV) and influenza facilitate the occurrence and fixation of such beneficial mutations (Moya et al.

2004). Viral adaptations to new hosts primarily manifest as amino acid substitutions, which can allow more efficient virus cell entry in the new host (Ito et al. 1998; Qu et al. 2005), block interactions with detrimental host proteins (Mangeat et al. 2003; Stremlau et al. 2004), and promote escape from both the new and the old host's immune responses (Smith et al. 2004; Wei et al. 2003). Should potentially adaptive mutations occur naturally before transmission (Demma et al. 2005) or be maintained in an intermediate host (Ito et al. 1998), the zoonotic virus will have a head start adapting to its new host environment. Thus, high viral genetic diversity in the reservoir host may be a dominant criterion discriminating a successful zoonosis from a dead-end exposure (Demma et al. 2006).

To understand the evolutionary processes facilitating viral zoonosis, we have been studying the adaptation of the envelope glycoprotein (*env*) of a diverse simian immunodeficiency virus (SIV) isolate from its natural sooty mangabey (SM) host after experimental inoculation of a nonnatural rhesus macaque (RM) host (Demma et al. 2005; Silvestri et al. 2005). By comparing changes in viral population structure after infection of both SMs and RMs with the identical diverse viral inoculum, we can identify the specific genetic targets of zoonotic adaptation and gain insight into what selective forces might be driving this zoonotic adaptation. Essentially, natural selection after a cross-species transmission would act like a sieve, favoring those amino acid polymorphisms critical for direct host–virus interactions in the new host, while ignoring neutral sequence variation and specific adaptations to immune responses of the previous host. Therefore, our experiment in cross-species adaptation will not only allow us to understand how a viral envelope adapts to a new host, but will also facilitate identification of particular residues in env, which are critical for its functional robustness.

Most analyses of SIV adaptation to RMs have involved inoculation of animals with clonal viruses, which had either been passaged in RMs previously or were first isolated from RMs accidentally infected with SIVsm in the 1980s. In these studies, analysis of the ratio of nonsynonymous to synonymous substitutions have revealed particular genes or amino acid sites that are under positive diversifying selection pressure due to immune responses against these sites. These studies of adaptation have been crucial to our understanding of how immunodeficiency viruses adapt to host immune responses. However, because of the preadapted condition of these virus isolates and the lack of genetic diversity upon which zoonotic selection can act, they do not sufficiently address the question of how diverse naturally occurring virus populations adapt to a new host species. Furthermore, if the spread of SIV into humans was mediated by direct contact with blood from African nonhuman primates (Hahn et al. 2000), the first infected human would have been exposed to a large, and likely very diverse, bolus of virus.

By inoculating RMs and SMs with a large bolus ($\sim 10^6$ virus copies) of an already highly diverse virus population from a naturally infected SM, we have provided the necessary sequence diversity for selection to act upon and may have more acutely approximated the circumstances of the first nonhuman to human transmission of a primate lentivirus. However, this diversity is the result of the virus having adapted to an animal whose immune responses have left their mark on the virus population's genetic structure (Demma et al. 2006). Therefore, simply calculating the ratio of nonsynonymous to synonymous substitutions at each codon in virus sequences sampled shortly after infection may not reliably distinguish between the evolutionary signatures of immune responses in the donor animal from the selective forces mediating adaptation to the recipient new host species. Thus, it is important to minimize the influence of this preexisting selective signature in zoonotic viral sequence analyses by characterizing fluctuations in the overall viral population structure as well as amino acid frequency changes at specific polymorphic sites.

Therefore, we have employed a combination of intuitive population genetic analyses and more complex models of codon substitution in order to extend our previous analyses (Demma et al. 2005) of the evolution of a natural and diverse SIVsm *env* variable loop 1 and 2 (V1V2) after transmission to a nonnatural RM host. This previous study described strong RM-specific restriction in the N-linked glycosylation (N-glyc) motif density and length of the V1 loop at days 10 and 14 postinfection (p.i.) followed by the emergence after day 100 p.i. of a diverse viral population with an allelic distribution more similar to that of the source inoculum (SI). Here, employing previously unused numerical analyses and a more comprehensive phylogenetic analysis of the same viral sequence data, we detect strong RM-specific selection in the V2 loop at days 40 and 70 p.i., which underscores this region's importance in adaptation to the RMs. Most prominently, changes in the position and frequency of an N-glyc motif in the V2 loop likely represent an adaptation either to a divergent CD4 or chemokine coreceptor or to an as-of-yet undetermined target cell population.

Methods

Experimental Inoculation of Nonhuman Primates and Collection of Clonal Sequence Data

Three SMs (FCo, FGu, and FLn) and three RMs (RHt, RZw, and RQl) were inoculated intravenously with the same diverse SIVsm population derived from a naturally infected SM (SI). Their course of viremia, immunological profiles, and viral genetic characteristics have been previously described (Demma et al. 2005; Silvestri et al. 2005). Briefly, the virus replicated well in all 3 SMs but only 2 of the 3 RMs (RHt and RZw), with peak viremia ranging from 5.0×10^7 to 1.6×10^9 viral copies per milliliter plasma and chronic phase setpoints ranging from 1×10^5 to 5×10^6 copies per milliliter. Due to its low to undetectable SIVsm viremia, RQl is not included in any subsequent analyses. Viral sequences were sampled from plasma viral RNA by reverse transcriptase–polymerase chain reaction at intervals throughout the first year of infection, and multiple clonal isolates of the virus envelope V1V2 region were amplified and sequenced. Input viral RNA copy number was not normalized prior to reverse transcription. However, viral load was not significantly different between animals at each time point, limiting potential copy number bias for within–time point comparisons. Furthermore, dilutions of a SI cDNA pool were subjected to the same PCR conditions, cloned, and then sequenced to rule out any potential input copy number bias. Finally, multiple PCR reactions were run on several samples to ensure the consistency of sequence results from each reaction. All sequences can be found in GenBank with accession numbers AY852284–AY852962.

Calculating Adaptive Events

To elucidate large-scale differences in the population structure of SIVsm infecting the SMs and the RMs, we applied to our sequence data set the analytical method developed in Williamson (2003). A majority consensus sequence was constructed from the 29 SI sequences. Alignments of sequences from each time point within each animal were compared with this consensus in SITES (Hey and Wakeley 1997) to determine the number of nonsynonymous and synonymous polymorphisms occurring at each site and at each time point. Polymorphisms were then classified as either common (>50%) or rare (<50%). Under neutrality, the ratio of common nonsynonymous to common synonymous polymorphisms should equal the same ratio of rare polymorphisms. Positive diversifying selection on amino acid sequences would manifest itself as an excess of common polymorphisms, thus

$$a = C_N - (C_S R_N / R_S)(1 + 1/R_S),$$

where a represents the excess number of common nonsynonymous polymorphisms (termed "adaptive events"), C_N and C_S are the calculated number of common nonsynonymous and synonymous polymorphisms, respectively, and R_N and R_S are the number of rare nonsynonymous and synonymous polymorphisms, respectively. Deviations of the rate of increase of adaptive events between day 14 and 70 p.i. from 0 were determined separately in the RMs and the SMs using Spearman's rank correlation.

Bayesian Phylogenetic Analysis

MrBayes (Huelsenbeck and Ronquist 2001) was used to construct phylogenetic trees (fig. 3A–E) of all isolates within each individual and the SI from our gap-stripped nucleotide alignments. Nucleotide positions within each codon were assumed to evolve at independent rates according to a gamma distribution whose rate parameter was estimated from the data. For each tree, 2 independent runs with 4 Monte Carlo Markov chains were performed on an Apple G5 8-node cluster. Each chain was run for 2,000,000 generations sampling every 1,000 generations. TRACER (http://evolve.zoo.ox.ac.uk/) was used to evaluate chain convergence. The burn-in length was determined independently for each tree, and the run with the longest burn-in (FCo: 24,000 generations; FGu: 150,000; FLn: 970,000; RHt: 87,000; RZw: 260,000) was used when building the consensus tree from both runs. The average standard deviation of the split frequencies (post–burn-in) for each tree was 0.066 (FCo), 0.010 (FGu), 0.043 (FLn), 0.045 (RHt), and 0.056 (RZw.) Bayesian posterior probabilities for each node were estimated as the proportion of trees sampled after burn-in containing each of the observed bipartitions.

Neighbor-Joining trees built using the distance method in PAUP4.10b (Swofford 2002) were consistent with our Bayesian results, although the consensus Bayesian topologies were significantly more likely (Kishino–Hasegawa test, $P < 0.001$). To ensure internal consistency between the trees of individual animals, each tree was stripped of all variants except for the 29 SI variants, using Treetool 2.0.2 (Maciukenas 1994), and the 5 resulting trees were compared with each other as well as with a maximum likelihood tree built only from the SI (data not shown). The 3 major SI clades are resolved well in all trees, although the topology of variants within these clades varies slightly. Despite difficulty in resolving SI variants 7, 8, 11, and 13 due to possible recombinant origins, these trees are largely in agreement.

Mapping Amino Acid Substitutions onto Phylogenetic Trees

Amino acid substitutions were mapped onto each individual animal Bayesian tree using the parsimony method in MacClade (Maddison WP and Maddison DR 1989). The ancestral sequence was assumed to be the node basal to the branch containing SI variants 3, 6, 14, 23, 27, and 29. Substitutions occurring on clades containing only SI variants were not counted in any analysis. After obtaining bulk site-wise counts of the number of amino acid substitutions, each site was scored for the possibility of selection. A site was considered under selection in an animal if 1) it contained multiple (>1) amino acid substitutions on terminal branches, representing parallel evolution, or 2) it contained one or more synapomorphic amino acid substitutions (Sheridan et al. 2004). The time points of variants present in clades with amino acid substitutions considered to be adaptive were recorded.

Maximum Likelihood Calculations of the Rates of Nonsynonymous and Synonymous Substitution

Due to the inability of likelihood models of codon substitution to explicitly account for differences in the time of sampling between taxa, we calculated rates of nonsynonymous (d_N) and synonymous (d_S) substitution separately at each time point within each animal. Modeltest (Posada and Crandall 1998) was run on each alignment to determine the best substitution model for building the trees. The most common substitution model was used for all trees to minimize the effects of using different nucleotide substitution models on the outcome of subsequent analyses. Maximum likelihood trees of each time point within each animal were then built in PAUP4.0b10 (Swofford 2002) using the HKY+G model of nucleotide substitution and mating base frequencies, the transition/transversion ratio, and shape of the gamma distribution separately for each alignment.

d_N and d_S were then estimated from the nucleotide alignments under a fixed effects likelihood model of codon substitution in HyPhy (Kosakovsky Pond and Frost 2005; Kosakovsky Pond et al. 2005) and scaled to the maximum likelihood trees. Unlike many other implementations of codon substitution models where d_S is estimated and fixed across the entire sequence, this particular program allows for estimation of both d_N and d_S at each individual codon. The numbers of nonsynonymous and synonymous substitutions were calculated at each site given the topology of each tree and likelihood ratio tests were performed to determine whether d_N was significantly greater than d_S. A P-value cutoff of $P < 0.25$ was chosen based on the findings of Kosakovsky Pond and Frost (2005). Sites under species-specific selection were defined as those with consistent selective patterns over time and across animals within that given species.

Results

SIVsm Undergoes Multiple Adaptive Events in RMs but Not in SMs during the Early Postacute Phase of Infection

Because previous analyses (Demma et al. 2005) of this robust data set focused primarily on viral genetic characteristics at the peak of acute infection and at late times p.i., we sought to understand the pattern of SIVsm adaptation to new host targets during the postacute phase, before the development of a strong neutralizing antibody (nAb) response. To determine whether viral sequence adaptations beyond those already described were occurring in our experimentally SIVsm-infected monkeys, we analyzed our data set using a method (Williamson 2003) designed to track changes in the frequencies of nonsynonymous and synonymous polymorphism in virus populations over time by comparing alignments of each p.i. time point with a consensus of the SI (fig. 1).

The pattern of SIVsm adaptation differs greatly between the viruses replicating in the 2 species over the first 100 days of infection (fig. 2A). The number of adaptive events occurring in SIVsm replicating in SMs remains relatively stable, fluctuating moderately around 0 through day 70 p.i. This is consistent with the notion that virus populations having replicated in SMs for centuries need not adapt to other SMs prior to the development of humoral immune responses. In contrast, SIVsm exhibits a very distinct but consistent adaptive pattern in both viremic RMs. The low, negative values at day 14 p.i. reflect the early outgrowth

of a single variant from an SI-containing multiple distinct variants (or alleles) of this region of *env*. Following this restriction, a succession of amino acid substitutions in the replicating viral populations occurs between days 14 and 70. This increase in the number of adaptive events in the RM virus populations is significant (fig. 2*B*; Spearman's rank correlation, $P < 0.05$) and, due to its occurrence before the development of strong nAb responses, is most likely the result of virus adaptation to divergent host cell receptors or target cell subsets. The subsequent decrease in adaptive events in RMs between days 70 and 100 reflects the outgrowth of viral variants more representative of the SI and presumably better able to escape emerging humoral immune responses. That virus populations in the SMs do not deviate significantly from the allelic distribution seen in the SI until after day 70 further supports the notion that humoral immune pressures first develop between day 70 and 100 in these infected animals and select for particular variants resulting in the observed increase in the number of adaptive events at day 100.

SIVsm Populations Adapting to RMs Are More Significantly Diverged from the SI than Viruses Replicating in SMs

To further elucidate temporal changes in the allelic structure of host-specific virus populations and to compare differences in viral divergence between host species, we built Bayesian phylogenetic trees of every SIVsm isolate within individual animals and the SI (fig. 3). The general placement of time points within each full Bayesian tree is highly consistent with our previous analyses (Demma et al. 2005). As before, almost all of the day 10 and day 14 RM isolates cluster with SI variants 3, 6, 14, 23, 27, and 29 (clade outlined in red on each tree), whereas the contemporaneous SM isolates are well distributed across the major clades. The relatively wider distribution of day 100 RM clones reflects the later reemergence of variants more closely related to the SI and are probably indicative of the onset of effective nAb responses. Strikingly, though, the majority of day 70 variants in both RMs fall into either 1 clade (fig. 3 RHt) or 2 clades (fig. 3, RZw). This apparent selection for distinct variants at day 70 may be at least partially responsible for the increase in RM-specific adaptive events noted earlier.

To characterize the overall divergence of the viral populations replicating in the newly infected animals from each other, patristic distance matrices were calculated from the Bayesian trees using PATRISTICv1.0 (Fourment and Gibbs 2006), and the average distance of each SI variant from its closest non-SI variant was determined within each species. Viruses establishing infection in the RMs are significantly more diverged from the SI than those replicating in the SMs (0.469 vs. 0.374; Student's *t*-test, $P < 0.005$). Additionally, the proportion of SI variants whose nearest neighbor is another SI variant is significantly greater in the RMs than in SMs (0.707 vs. 0.414; normal approximation of the binomial, $P < 0.0005$). Taken together, these data suggest that in addition to overall differences in the allelic structure of the SIVsm populations between the 2 host species, the viruses replicating in RMs are significantly more diverged from the SI than viruses replicating in the newly infected SMs.

RM-Specific Amino Acid Substitutions in V2 Mediate SIVsm Adaptation to This New Host in the Absence of nAb Responses

To identify specific amino acid sites that may be responsible for the disparate evolutionary patterns of viruses replicating in the 2 monkey species, we applied 3 site-by-site analyses to detect particular codons under selection. Due to the difficulties inherent in identifying newly selected mutations on the genetic background of a diverse virus population that already bears the mark of strong immune selective pressures (Demma et al. 2006), we applied 1) a phylogenetic based method to identify specific amino acid changes and the virus subpopulations in which they occur, 2) an analysis of the site-by-site amino acid frequencies to understand the selective changes occurring at each site, and 3) a model of codon substitution to both elucidate the genetic signature of selection in the donor and evaluate the robustness of such models in the context of such a complex selective background.

The phylogenetic method involved mapping amino acid substitutions at each individual site onto each animal's full-infection phylogenetic tree through day 100 p.i. The average number of amino acid changes per animal, cumulative across the sequence, is much greater in RMs (147) than in SMs (106), consistent with our previous findings (Demma et al. 2005) of an overall greater number of nonsynonymous substitutions in the RMs (supplementary figure 1, Supplementary Material online). We then classified sites within animals as being under selection based on criteria laid out in Sheridan et al. (2004). Despite some variation among sites and species in the temporal pattern of evolution, the sites under selection are largely the same between RMs and SMs and are primarily located in V1. Interestingly, the 3 sites (80, 81, and 105) found to be under selection exclusively in the RMs were also under selection between days 40 and 100 p.i., the same interval over which the virus populations only in the RMs underwent multiple adaptive events (fig. 4). Thus, it is likely that amino acid substitutions or changes in the relative abundance of amino acid polymorphisms at each of these sites mediate SIVsm adaptation to this new host.

The second method by which we identified amino acid sites under selection involved following changes in the frequency of the consensus amino acid at each site over time (fig. 4 and supplementary fig. 2, Supplementary Material online). This method allows the detection of fluctuations in the relative abundance of existing and de novo amino acid polymorphisms at individual codons. Sites exhibiting large fluctuations (>20% shifts) in amino acid composition were identified as being under directional selection and thus most likely represent specific adaptations to the RMs. Overall, a much larger number of amino acid positions were identified as being under RM-specific selection pressures than in the analysis of phylogenetic substitutions (fig. 4). The sites not identified in previously discussed analyses (sites 45, 53, 56, 57, and 59) are focused at the C terminus of the V1 loop and are selected for primarily at days 10 and
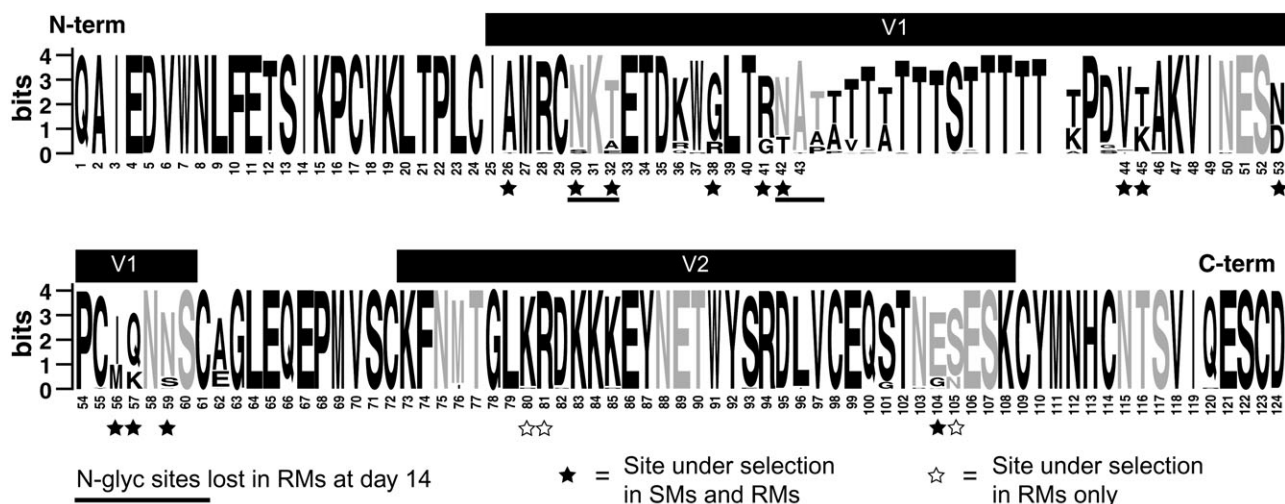
FIG. 1.—SIVsm env V1V2 is highly diverse in the plasma of the naturally infected SM used for experimental inoculation of SMs and RMs. All 29 SI sequences were aligned and summarized using WebLogo (http://weblogo.berkeley.edu). The relative height of each amino acid letter designation at any given site represents their frequency, whereas the overall height of the column indicates the amount of information contained at that site. Unnumbered amino acid sites were excluded from all analyses due to the presence of gaps. Putative N-linked glycosylation motifs (NXS/T) are represented in gray. Underlined N-linked glycosylation motifs were not present in the majority of sequences sampled from the RMs at day 14 p.i. Sites labeled with a filled star were found to be under selection in both the RMs and the SMs. Sites labeled with an open star were found to under RM-specific selection pressures.

14 p.i. That these amino acid positions were highly polymorphic in the SI (see fig. 1) and are under selection contemporaneously with the RM-specific restriction in viral diversity at the peak of acute infection suggests that they are selectively neutral and that changes in the relative abundance of these amino acid residues reflect selection at closely linked sites (i.e., N-glyc site at position 30/32).

The consensus amino acid frequencies at sites 30, 32, 80, 81, 104, and 105 (supplementary fig. 2A, B, E–H, Supplementary Material online) all illustrate various RM-specific patterns of selection, whereas sites 41 and 42 (supplementary fig. 2C–D, Supplementary Material online) demonstrate no species-specific adaptive pattern. Of the 6 RM-specific adaptive sites, 4 are part of N-glyc motifs (sites 30, 32, 104, and 105; see fig. 1). Selection at sites 30 and 32 represent the RM-specific loss of 1 N-glyc motif at day 10 and 14 p.i. previously identified (Demma et al. 2005). Although this N-glyc site remains absent or at low levels until day 100, the viral populations in the RMs continue to adapt to the new host environment through changes in consensus amino acid frequencies at 2 specific loci: the highly charged region of V2 (sites 80 and 81) and another N-glyc site (sites 104 and 105). Interestingly, the polymorphism at site 105 (Ser to Asn) causes a 2–amino acid, C-terminal shift in an N-glyc site (see fig. 1). Taken together, these data demonstrate a continued strong selection pressure during the post-acute phase of infection for specific variants, presumably better adapted to the divergent cellular and genetic environment of the RMs.

Comparing d_N and d_S at Individual Codons Does Not Discriminate Recent Episodes of Positive Selection in the Newly Infected Host from Potentially Adaptive Sites Generated by the Immune Responses of the Donor Host

Most of the polymorphisms under selection at any given time point in our experimentally infected animals were already present in the SI as demonstrated by consensus amino acid frequencies at the day 0 time point (fig. 4; see also fig. 1). This phenomenon not only underscores the importance of viral diversity within donor animals for the potential success of cross-species virus transmission but may also obscure the interpretation of more traditional site-by-site likelihood analyses of codon substitution. The large number of viruses (1×10^6) inoculated into our experimentally infected animals ensures that the signatures of selection on the virus population in the SI animal will be transmitted as well. Thus measurements of the numbers of nonsynonymous and synonymous substitutions (per nonsynonymous and synonymous site, respectively) early in the course of infection will mainly reflect selection biases predating the inoculation of our experimental animals. These sites should be easily identified as those either unaccompanied by changes in relative amino acid abundance or those which are under positive diversifying selection in both species. It is therefore necessary to compare the ability of codon substitution models to identify RM-specific sites under selection with our other sitewise analyses.

Interestingly, the distribution of sites under selection (fig. 4) is similar to our previous analyses here and elsewhere (Demma et al. 2005). Specifically, the V1 loop is under strong positive selection in both RMs and SMs, although fewer sites were evolving under positive selection in RM-specific virus populations than in the SMs. However, the SM-specific positively selected sites (sites 34, 36, and 38) were not well supported by our other analyses. In contrast, only 1 site (104) out of the 5 (sites 80, 81, 84, 102, and 104) identified to be evolving under positive selection in V2 was also detected in SMs. Two of these sites (site 84 and 102) were not identified in any other analysis, and therefore, differences between the estimated numbers of nonsynonymous and synonymous substitutions likely predated this infection experiment. Finally, the codon identified in the 2 previous site-by-site analyses, which is
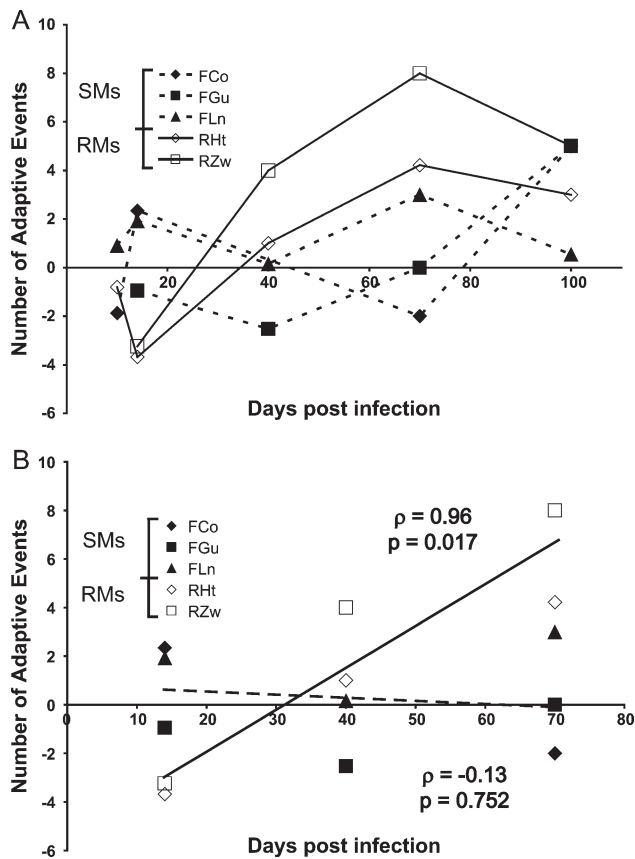
FIG. 2.—SIVsm undergoes multiple adaptive events in RMs but not SMs during the early postacute phase of infection. (*A*) The number of adaptations occurring in SIVsm populations infecting both RMs and SMs through day 100 p.i. (*B*) The rate of increase in adaptive events over time occurring in RM-specific SIVsm populations between days 14 and 70 p.i. is significantly greater than 0, whereas the slight decrease in adaptive events among SMs is indistinguishable from 0 (Spearman's rank correlation, $P < 0.05$).

responsible for a RM-specific increase in the frequency of the C-terminally shifted N-glyc motif, site 105, was not identified as positively selected in this analysis.

Discussion

Here we describe in detail the selective processes accompanying the cross-species transmission of a very diverse SIVsm virus inoculum derived from an endemically infected natural host (SM) to a nonnatural host (RM). By applying several evolutionary analyses to serially sampled virus populations for the first 100 days of infection, we have been able to identify the amino acid positions responsible for the continued adaptation of a diverse SIVsm inoculum to the new RM cellular and genetic environment. The polymorphisms mediating this adaptation were already present in the SI and primarily alter the density and position of N-glyc sites on the virus envelope. In addition, we also demonstrate the utility of a multifaceted approach to studying viral sequence evolution, which combines complex phylogenetic methods and simple intuitive analyses to identify and discriminate between sites that were selected for in the SM reservoir host from those that have only recently come under selection in the new RM host.

The *env* V1V2 sequence variants we describe here are not completely novel. Alignment of our SIVsm *env* clones from day 70 p.i. with the same region of several common SIV clones from the Los Alamos HIV sequence database (http://www.hiv.lanl.gov) reveals that the same shift in the N-glyc site at position 105 observed in our RMs is also found in most of these RM-adapted isolates (fig. 5). Because all of the sequences obtained from the database were the result of one or several accidental transmissions from SMs to RMs of a single SIVsm subtype (8) only recently described (Apetrei et al. 2005, 2006), we cannot rule out the alternative hypotheses that this site is not polymorphic in the SIVsm subtype-8 envelope or that this polymorphism was fixed in RMs due to an extreme bottleneck upon transmission. Regardless, it is probably not entirely coincidental that these subtype-8 viruses were the first to be described infecting RMs because SMs infected with other lineages of SIVsm were used in the experiments now thought to have facilitated the first SIV transmission to RMs but did not engender persistent infection of RMs (Apetrei et al. 2006).

Of all the loci identified here as sites under RM-specific selection, the C-terminal shift in an N-glyc site caused by a polymorphism at site 105 is the most intriguing. Although this shift is transient (supplementary fig. 2H, Supplementary Material online), it does occur early in infection during the brief window (through day 100) in which nAb responses, the primary immunological selection pressure on *env* (Frost et al. 2005), are thought to be low or absent (Rybarczyk et al. 2004). Thus, preferential expansion in the RMs of viruses encoding this shifted N-glyc site as well as amino acid residues at sites 80 and 81 strongly suggests that these mutations are specific adaptations to the divergent RM genetic and cellular environment, as opposed to adaptations to new host immune responses. The partial reversion of this N-glyc site to its N-terminal position beyond day 70 p.i. is due to the reemergence of the presumably more immunologically evasive variants containing both of the N-glyc sites in V1 at amino acid positions 30 and 42 (see fig. 1). The lack of linkage between the C-terminally shifted N-glyc site polymorphism in V2 and the V1 loop variants containing both N-glyc sites among all the viruses sampled in our study suggests that this adaptation in V2 may actually be disadvantageous in the presence of a highly glycosylated V1.

Even though effective nAb responses are likely low or nonexistent prior to day 100 p.i., it is possible that immune cell populations may be at least indirectly responsible for these changes in V1V2 allelic structure. It has recently become apparent that acute HIV infection of humans (Brenchley et al. 2004; Mehandru et al. 2004) and SIV infection of RMs (Li et al. 2005; Mattapallil et al. 2005) and now SMs (Silvestri G, personal communication) is associated with a massive depletion of a specific memory CD4+ T-cell subset from mucosal tissues caused by a combination of virus- and cytotoxic T lymphocyte-induced cytopathicity (Mattapallil et al. 2005; Regoes et al. 2004). Preferential depletion of this T-cell subset during the early stages of lentiviral infection identifies them as a primary target cell reservoir in both pathogenic and nonpathogenic hosts. Ultimately, this profound change in the number and type of target cells could have important consequences for SIVsm viral population structures in the early stages of infection.
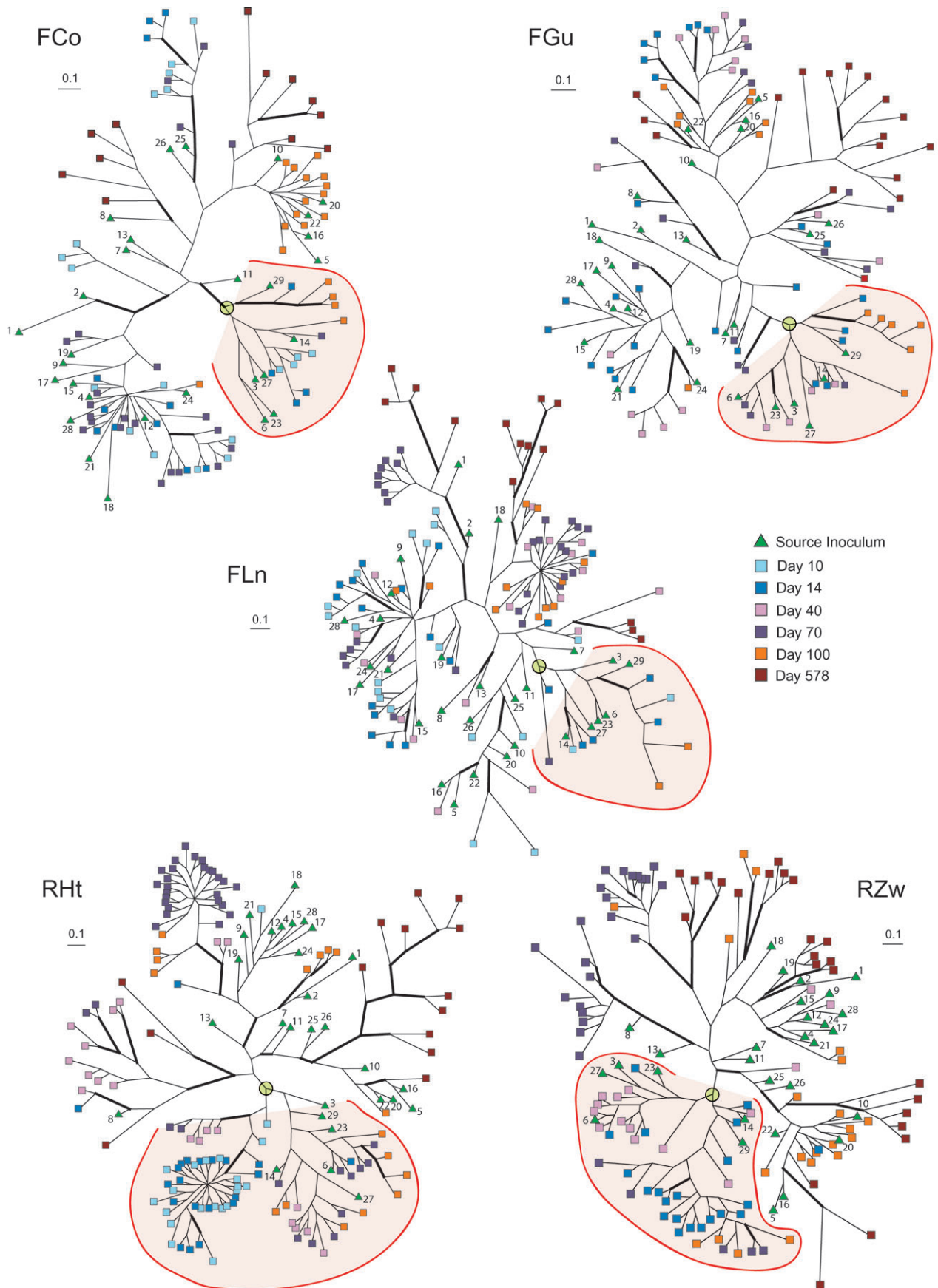
AA site	Mapped Substitutions										Frequency Patterns										Likelihood	
	Sooty Mangabey					Rhesus Macaque					Sooty Mangabey					Rhesus Macaque					SM	RM
	10	14	40	70	100	10	14	40	70	100	10	14	40	70	100	10	14	40	70	100		
2																					−	
6																					−	−
20																					−	
26			X					X						X							+	+
30				X	X	X	X	X	X	X						X	X	X	X	X	+	+
32	X	X	X	X	X														X		+	+
34																					+	
36																					+	
38			X							X		X								X	+	
41			X					X	X	X	X	X						X			+	+
42			X					X	X	X	X							X			+	+
44	X	X	X																X		+	+
45	X	X														X					+	+
46			X																		+	+
52																					−	−
53	X	X	X					X								X					+	+
56	X							X														
57	X								X										X		+	+
59		X						X										X			+	+
64																						−
78																					−	−
80								X	X									X	X			+
81									X	X								X				+
84																						+
86																					−	−
102																						+
104	X	X	X	X	X	X	X	X	X	X						X	X				+	+
105								X														+
109																					−	−
114																					−	
115																					−	−
117																					−	−

Fig. 4.—Summary of the amino acid sites found to be under selection for all 3 site-by-site analyses. Analyses are partitioned by analysis then by species and then by time point. Gray boxes indicate selection at that site and time point. "+" or "−" symbols indicate the presence of positive or purifying selection, respectively, as determined by the maximum likelihood analysis of dN and dS ($P < 0.25$, likelihood ratio test). Sites with gray forward hatching indicate the sites in V2 found to be under selection in a majority of the analyses. Sites with gray backward hatching are the sites in V2 found to be under positive selection by only the likelihood method.

More likely, however, is the possibility that the changes we observe in SIVsm viral populations are adaptations to the divergent immune cellular environment of the RMs. It has recently been reported (Pandrea et al. 2006) that a number of well-studied natural nonpathogenic SIV hosts, including SMs, exhibit far lower levels of CCR5 on these same mucosal memory CD4+ T cells than what is typically seen in nonnatural pathogenic hosts, like humans and RMs. The authors hypothesize that this could be a convergent evolutionary mechanism for ameliorating the pathogenic effects of lentiviral infection of many African nonhuman primates although not explicitly preventing host-to-host spread of the virus. However, the fact that the level of vi-remia (Broussard et al. 2001; Goldstein et al. 2005) and the rate of infected target cell turnover (Ho et al. 1995; Mohri et al. 1998) (Silvestri G, personal communication) are similar in both pathogenic and nonpathogenic hosts suggests that SIVs replicating in their natural hosts may be better able to use multiple chemokine coreceptors to facilitate entry into a broader array of short-lived target cells than their counterparts infecting pathogenic hosts. Whether this dearth of CCR5 expression in natural hosts results in lower levels of SIV infection of memory CD4+ T cells or is simply a hallmark of the lower level of immune activation seen in the natural hosts, this phenomenon underscores the dramatic difference between SMs and RMs in their immune

←

Fig. 3.—SIVsm populations adapting to RMs are distinct from those in SMs. Bayesian phylogenetic trees of all variants sampled from each individual animal and the SI. FCo, FGu, and FLn are the SMs, and RHt and RZw are the RMs. Time points sampled from individually inoculated animals are represented by colored squares. The SI is represented by green triangles, and each variant is numbered. The clade containing the 6 SI variants from which most of the day 14 RM variants are descended is outlined in red. The node on which trees were rooted to determine the ancestral sequence for inferring substitutions is indicated with a light green circle.

```
                          **                    ★
                    80         90        100        110
                 ...|....|....|....|....|....|....|..|....|....
SIVmac239        CKFNMTGLKRDKKKEYNETWYSADLVCEQGNNT--GNESRCYMNHC
SIVmac251-1A11   ...........................S.--.............
SIVsmm142B       ...........................S.--.............
SIVsmm17EC1      .....N.....................--...............
SIVsmm17EFR      .....N.....................--...............
SIVmac251        .....T.......T.........T....S.--D............
SIVmac142        ...........................S.--D............
SIVmac32H        .............T.........T....S.--D............
SIVmacH824       .............T.........T....S.--D............
SIVsmmH9         ............R.....R......-.SNGNETD.K........
SIVsmmF236       .....N.............R..I..--SA-NES..K...H....
```

Day 70 RMs
RHt
```
.....N.............R.....--S.-NES..K.......
.....K.............R.....--S.-N....K.......
.....K.............R.....--S.-N....K.......
.....K.............R....R--G.-N....K..R....
```
RZw
```
.....E.............R.....--S.-N....K.......
.....M.............R.....--S.-N....K.......
.....N.............R.....--S.-N....K.......
.....K.............R.....--S.-N....K.......
```

Day 70 SMs
FCo
```
...................R.....--S.-NES..K.......
...................K.....--S.-N.S..K.......
...................R.....--S.-N.S..K.......
```
FGu
```
...................R.....--S.-NES..K.......
...................R.....--S.-NES..K.......
...................R.....--S.-N....K.......
...................R.....--S.-N.S..K.......
```
FLn
```
.................R.S.....--S.-NES..K.......
...................R.....--S.-NES..K.......
...................R.....--S.-NES..K.......
...................R.....--S.-NES..K.......
```
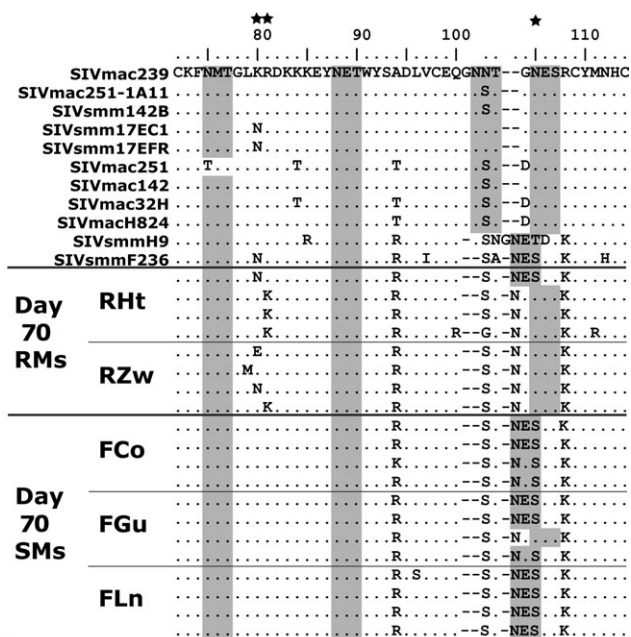
FIG. 5.—SIVsm V2 sequences at day 70 in RMs are more similar to V2 sequences from several major macaque-adapted SIV clones than to the contemporaneous SM V2 sequences. The 3 sites under significant RM-specific selection are indicated by stars, and N-glyc sites are backed in gray. The V2 sequences of RM-adapted clones were found at Los Alamos National Laboratories HIV Sequence Database (http://www.hiv.lanl.gov) and aligned by hand to maximize amino acid and codon usage similarity.

cell phenotype and thus their target cell landscape. It will be important to investigate the ability of the specific adaptations identified in this and previous studies (Demma et al. 2005) to allow the SIVsm envelope to utilize the divergent RM CD4 receptor as well as the various RM chemokine coreceptors that can mediate virus entry into target cells.

Supplementary Material

Supplementary figures 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Apetrei C, Kaur A, Lerche NW, et al. (19 co-authors). 2005. Molecular epidemiology of simian immunodeficiency virus SIVsm in U.S. primate centers unravels the origin of SIVmac and SIVstm. J Virol. 79(14):8991–9005.

Apetrei C, Lerche NW, Pandrea I, Gormus B, Silvestri G, Kaur A, Robertson DL, Hardcastle J, Lackner AA, Marx PA. 2006. Kuru experiments triggered the emergence of pathogenic SIVmac. Aids. 20(3):317–321.

Brenchley JM, Schacker TW, Ruff LE, et al. (11 co-authors). 2004. CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. J Exp Med. 200(6):749–759.

Broussard SR, Staprans SI, White R, Whitehead EM, Feinberg MB, Allan JS. 2001. Simian immunodeficiency virus replicates to high levels in naturally infected African green monkeys without inducing immunologic or neurologic disease. J Virol. 75:2262–2275.

Demma LJ, Logsdon JM Jr, Vanderford TH, Feinberg MB, Staprans SI. 2005. SIV quasispecies adaptation to a simian new host. PLoS Path. 1(1):e3.

Demma LJ, Vanderford TH, Logsdon JM Jr, Feinberg MB, Staprans SI. 2006. Evolution of the uniquely adaptable lentiviral envelope in a natural reservoir host. Retrovirology. 3:19.

Fourment M, Gibbs MJ. 2006. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evol Biol. 6:1.

Frost SDW, Wrin T, Smith DM, et al. (12 co-authors). 2005. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc Natl Acad Sci USA. 102(51):18514–18519.

Gao F, Bailes E, Robertson DL, et al. (12 co-authors). 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature. 397:436–441.

Goldstein S, Ourmanov I, Brown CR, Plishka R, Buckler-White A, Byrum R, Hirsch VM. 2005. Plateau levels of viremia correlate with the degree of CD4+-T-cell loss in simian immunodeficiency virus SIVagm-infected pigtailed macaques: variable pathogenicity of natural SIVagm isolates. J Virol. 79(8):5153–5162.

Hahn BH, Shaw GM, Cock KMD, Sharp PM. 2000. AIDS as a zoonosis: scientific and public health implications. Science. 287:607–614.

Hey J, Wakeley J. 1997. A coalescent estimator of the population recombination rate. Genetics. 145(3):833–846.

Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature. 339:389–392.

Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature. 373:123–126.

Holmes KV. 2005. Structural biology. Adaptation of SARS coronavirus to humans. Science. 309(5742):1822–1823.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Ito T, Couceiro JN, Kelm S, et al. (11 co-authors). 1998. Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. J Virol. 72(9):7367–7373.

Kosakovsky Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21(5):676–679.

Kosakovsky Pond SL, Frost SD. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 22(5):1208–1222.

Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Delicat A, Paweska JT, Gonzalez JP, Swanepoel R. 2005. Fruit bats as reservoirs of Ebola virus. Nature. 438(7068):575–576.

Li Q, Duan L, Estes JD, Ma ZM, Rourke T, Wang Y, Reilly C, Carlis J, Miller CJ, Haase AT. 2005. Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells. Nature. 434(7037):1148–1152.

Maciukenas M. 1994. Treetool 2.0.2 [Internet]. Ribosomal RNA Database Project, University of Illinois. Available from: http://iubio.bio.indiana.edu:7780/archive/00000441/.

Maddison WP, Maddison DR. 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol. 53(1–4):190–202.

Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature. 424(6944):99–103.

Mattapallil JJ, Douek DC, Hill B, Nishimura Y, Martin M, Roederer M. 2005. Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. Nature. 434(7037):1093–1097.

Mehandru S, Poles MA, Tenner-Racz K, Horowitz A, Hurley A, Hogan C, Boden D, Racz P, Markowitz M. 2004. Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. J Exp Med. 200(6):761–770.

Mohri H, Bonhoeffer S, Monard S, Perelson AS, Ho DD. 1998. Rapid turnover of T lymphocytes in SIV-infected rhesus macaques. Science. 279:1223–1227.

Moya A, Holmes EC, Gonzalez-Candelas F. 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol. 2(4):279–288.

Palese P. 2004. Influenza: old and new threats. Nat Med. 10 (Suppl 12):S82–S87.

Pandrea I, Apetrei C, Gordon S, et al. (14 co-authors). September 26, 2006. Paucity of CD4+CCR5+ T-cells is a typical feature of natural SIV hosts. Blood. 10.1182/blood-2006-05-024364.

Parrish CR, Kawaoka Y. 2005. The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. Ann Rev Microbiol. 59(1):553–586.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14(9):817–818.

Qu X-X, Hao P, Song X-J, et al. (19 co-authors). 2005. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. J Biol Chem. 280(33):29588–29595.

Regoes RR, Antia R, Garber DA, Silvestri G, Feinberg MB, Staprans SI. 2004. Roles of target cells and virus-specific cellular immunity in primary simian immunodeficiency virus infection. J Virol. 78(9):4866–4875.

Rybarczyk BJ, Montefiori D, Johnson PR, West A, Johnston RE, Swanstrom R. 2004. Correlation between env V1/V2 region diversification and neutralizing antibodies during primary infection by simian immunodeficiency virus sm in rhesus macaques. J Virol. 78(7):3561–3571.

Sheridan I, Pybus OG, Holmes EC, Klenerman P. 2004. High-resolution phylogenetic analysis of hepatitis C virus adaptation and its relationship to disease progression. J Virol. 78(7):3447–3454.

Silvestri G, Fedanov A, Germon S, Kozyr N, Kaiser W, Garber D, McClure H, Feinberg MB, Staprans SI. 2005. Divergent host responses during primary SIVsmm infection of natural mangabey and non-natural rhesus macaque hosts. J Virol. 79:4043–4054.

Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. 2004. Mapping the antigenic and genetic evolution of influenza virus. Science. 305(5682):371–376.

Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. 2004. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature. 427(6977):848–853.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sunderland (MA): Sinauer Associates.

Wang M, Yan M, Xu H, et al. (25 co-authors). 2005. SARS-CoV infection in a restaurant from palm civet. Emerg Infect Dis. 11(12):1860–1865.

Webby R, Hoffmann E, Webster R. 2004. Molecular constraints to interspecies transmission of viral pathogens. Nat Med. 10 (Suppl 12):S77–S81.

Wei X, Decker JM, Wang S, et al. (15 co-authors). 2003. Antibody neutralization and escape by HIV-1. Nature. 422:307–311.

Williamson S. 2003. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol Biol Evol. 20(8):1318–1325.