Low-Dose Rectal Inoculation of Rhesus Macaques by SIVsmE660 or SIVmac251 Recapitulates Human Mucosal Infection by HIV-1

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi: 10.1084/jem.20082831

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:4621131

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
An effective HIV-1 vaccine, microbicide, or other pre- or post-exposure prophylactic must interdict virus at or near the moment of mucosal transmission or in the early period preceding the establishment of viral latency and disseminated infection (1–4). In humans, it has been difficult to study these earliest viral host events in vivo (2, 5–13), and in tissue explant cultures or in Indian rhesus macaques the HIV-1 or simian immunodeficiency virus (SIV) inocula have typically been high to achieve uniform infection of controls or to visualize infection events in situ (14–20), thus prompting concerns about the physiological relevance of the model systems (21–24). Further complicating the analysis of early infection events in vivo is the viral “eclipse” period which obscures the events that occur shortly after infection.

We recently developed a novel strategy to identify transmitted HIV-1 genomes in acutely infected humans using single-genome amplification and a model of random virus evolution. Here, we used this approach to determine the molecular features of simian immunodeficiency virus (SIV) transmission in 18 experimentally infected Indian rhesus macaques. Animals were inoculated intrarectally (i.r.) or intravenously (i.v.) with stocks of SIVmac251 or SIVsmE660 that exhibited sequence diversity typical of early-chronic HIV-1 infection. 987 full-length SIV env sequences (median of 48 per animal) were determined from plasma virion RNA 1–5 wk after infection. i.r. inoculation was followed by productive infection by one or a few viruses (median 1; range 1–5) that diversified randomly with near starlike phylogeny and a Poisson distribution of mutations. Consensus viral sequences from ramp-up and peak viremia were identical to viruses found in the inocula or differed from them by only one or a few nucleotides, providing direct evidence that early plasma viral sequences coalesce to transmitted/founder viruses. i.v. infection was >2,000-fold more efficient than i.r. infection, and viruses transmitted by either route represented the full genetic spectra of the inocula. These findings identify key similarities in mucosal transmission and early diversification between SIV and HIV-1, and thus validate the SIV–macaque mucosal infection model for HIV-1 vaccine and microbicide research.
period during which virus replicates in mucosal and lo-
coregional lymphoreticular tissues but is not yet detectable
in the circulating plasma (25). In SIV-infected macaques, the
eclipse period is generally ~4–7 d in duration, and in
HIV-1–infected humans, it is ~7–21 d (5, 18, 25–27).

Previously, we observed that in the early stages of HIV-1
infection preceding antibody seroconversion (eclipse phase
and Fiebig stages I and II [25]), virus diversification follows
a pattern of random evolution with an almost starlike phy-
logeny and a Poisson distribution of nucleotide substitu-
tions (5). We thus hypothesized that the genetic identity of
transmitted or early founder viruses could be inferred
unambiguously by phylogenetic analysis of discrete low-
diversity viral lineages that emanate from them. This hypo-
thesis was supported by an analysis of 3,449 full-length env
genes from 102 human subjects with acute HIV-1 subtype
B infection, where we found that (a) acute viral sequences
sampled before the development of measurable adaptive
immune responses conformed to a pattern of random virus
evolution; (b) viral sequence diversity resulted in model
estimates of time to a most recent common ancestor
(MRCA) that was consistent with clinical histories and Fiebig stage classifications; and (c) in most subjects (78 of
102) there was evidence of productive infection by only a
single virus, whereas in 24 other subjects infection resulted
from transmission of at least 2 to 5 viruses, each recogniz-
able as a discrete virus lineage (5). These findings have since
been corroborated in seven additional patient cohorts in-
fected by HIV-1 subtypes A, B, C, or D (28–35). A key
innovation common to these studies was the use of singe-
genome amplification (SGA) of plasma viral RNA, fol-
lowed by direct amplicon sequencing to characterize the
virus quasispecies (5, 35–39). This method provides pro-
portional representation of plasma viral RNA (vRNA) and
precludes Taq polymerase-induced template switching (re-
combination), Taq polymerase-associated nucleotide sub-
stitutions in finished sequences, template resampling, and
cloning bias (5, 35, 37, 38, 40).

In this study, we sought to directly test our strategy for
identifying transmitted/founder viruses in the Indian rhe-
sus macaque SIV infection model where we could define
essential experimental parameters, including the route of
SIV infection, genetic composition of the inoculum, and
the duration between virus inoculation and sampling of
plasma vRNA. The primary study objectives were two-
fold: first, to determine if, as our hypothesis and model
predict, plasma SIV sequences sampled at or near peak
viremia coalesce to sequences of viruses responsible for
transmission and productive clinical infection weeks ear-
ier; and second, to determine how closely a low-dose
SIV rectal transmission model in Indian rhesus macaques
recapitulates features of human infection by HIV-1, in-
cluding the extent of the mucosal barrier to virus transmis-
sion, the number of transmitted/founder viruses leading to
productive infection, and the molecular patterns of early
virus diversification.

RESULTS
Inoculation regimen and infection kinetics
18 animals received weekly atraumatic i.r. inoculations of
cell-free SIV as part of a previous study of adaptive immune
responses after mucosal virus exposure (41). The virus inoc-
ula were uncloned SIVmac251 or SIVsmE660 strains com-
prised of virus “swarms” with env diversity comparable to
what is observed in humans 1–2 yr after infection by HIV-1
(42). 9 animals, divided into groups of 3, received log dilu-
tions of $6 \times 10^7$, $6 \times 10^6$, or $6 \times 10^5$ vRNA copies of
SIVsmE660 (Fig. 1, left). Nine other animals, divided into
groups of three, received identical dilutions of SIVmac251
(Fig. 1, right). Virus was administered weekly for 6 wk,
followed by a 3-wk observation period and a 2-mo rest period
to ensure that none of the animals experienced delayed sero-
conversion. Animals were phlebotomized weekly during the
inoculation–observation period. Animals that did not be-
come productively infected after the first inoculation–ob-
servation–rest cycle underwent a second 6-wk course of i.r.
inoculations at the same virus dose level, followed by a 3-wk
observation period and a 2-mo rest period. Animals that still
did not become infected underwent a third 6-wk course of
i.r. inoculations at the maximum virus dose level ($6 \times 10^7$
vRNA copies), followed by observation and rest periods.
One animal, AV66, received SIVsmE660 for the initial two
i.r. inoculation cycles and was then crossed over to receive
SIVmac251 for the third i.r. inoculation cycle (Fig. 1, bottom
left). 5 animals that did not become infected after the 18 i.r.
inoculations were given a single i.v. inoculation of $2 \times 10^5$
vRNA copies. All animals became infected.

The kinetics of virus infection and replication, and the
time points of plasma virus sampling for env sequence analy-
sis, are illustrated (Fig. 1). There was no significant difference
in the overall infection rate between macaques inoculated i.r.
with SIVsmE660 (6 of 8 animals) versus SIVmac251 (7 of 10
animals). Because of this similarity, and the relatively small
number of animals in the study ($n = 18$), the 6 animals in
each virus dose group were combined for purposes of statisti-
cal analysis. We found a statistically significant virus dose–
infection trend to i.r. infection, with 6 of 6 animals receiving
the highest dose of SIV becoming infected in the first 8 wk
inoculation–observation regimen, 4 of 6 animals receiving
the intermediate dose becoming infected in the same period,
and 1 of 6 animals receiving the lowest dose becoming in-
fected in the same period ($P = 0.005$ by $3 \times 2$ Fisher’s exact
test). Two animals, AV66 and AH4X, did not become in-
fected after 12 weekly i.r. inoculations at the lowest dose of
virus ($6 \times 10^5$ vRNA copies), but did become infected after
1–4 i.r. inoculations at the highest virus dose ($6 \times 10^7$ vRNA
copies). 5 animals, CP37, CR54, AV74, CG71, and CG5G,
did not become infected even after 18 weekly i.r. inocula-
tions, but each became infected after a single i.v. inoculation.
For both SIVsmE660- and SIVmac251–infected animals, the
kinetics of plasma viral load increase were quite similar,
reaching peak viremia within 1 wk of the last negative sample
in three animals and within 2 wk in 15 others. Peak viral
Figure 1. Viral kinetics after i.r. or i.v. infection. Groups of three animals received i.r. inoculations of SIVsmE660 or SIVmac251 in dosages indicated at weeks 0, 1, 2, 3, 4, and 5 (underlined) or by the i.v. route at week 0 (only) of the fourth cycle. Plasma was sampled at ramp-up (green symbols) or at peak (blue symbols) viremia for env sequence analysis. Numbers of transmitted/founder viruses found to be responsible for productive infection are indicated in parenthesis. Plasma viral load determinations were previously reported (41).
loads ranged from $1.4 \times 10^8$ to $9.4 \times 10^7$ (median $8 \times 10^8$) and were not related to the infecting virus strain or route of infection. A correlation between peak viral load and numbers of transmitted viruses was observed for SIVsmE660-infected (Spearman correlation coefficient 0.71; $P = 0.05$) and SIVmac251-infected (Spearman correlation coefficient 0.91; $P = 0.005$) macaques.

**Virus diversity in SIV inoculum stocks**

The extent of virus sequence diversity in the inoculum stocks of SIVsmE660 and SIVmac251 was determined by SGA-direct sequencing of full-length vRNA env genes, followed by pairwise sequence comparisons, neighbor-joining (NJ) phylogenetic tree construction, and Highlighter analysis (Fig. S1). Highlighter is a sequence analytical tool (www.HIV.lanl.gov) that displays the location and identity of nucleotide substitutions in a visually informative manner and allows tracing of common ancestry between sequences based on individual nucleotide polymorphisms. Maximum diversity among 42 SIVsmE660 env sequences was 1.8%, and among 61 SIVmac251 sequences, diversity was 0.8%. NJ trees of both sequence sets revealed structure in the phylogenetic relationships among sequences typical of primary virus isolates. Highlighter plots reflected this structured diversity.

**Virus diversity in SIV–infected macaques**

A total of 987 full-length SIV env genes of plasma virions from 18 animals (median of 48 per animal; range 23–137) were amplified and sequenced. 26 sequences (distributed among 12 animals) showed evidence of overt APOBEC G-to-A hypermutation (e.g., sequences RU.I3, RU.4B7, PK.I6, PK.4A29 and PK.I10 in Fig. 2). These sequences were retained in phylogenetic tree constructions but were excluded from diversity measurements and model calculations. Among the remaining 961 sequences, maximum env diversity in animals infected i.r. ranged from 0.07 to 1.43% (median = 0.33%). Maximum env diversity in animals infected intravenously ranged from 0.18 to 1.54% (median = 0.81%), which is not different from viral diversity in i.r.-infected animals ($P = 0.11$ by Wilcoxon rank sum test). i.r.-infected animals had env sequence diversities that fell into two distinctive groups, which we showed subsequently to reflect productive infection by one virus (maximum env diversity of 0.07–0.15%, median = 0.15%) or more than one virus (maximum env diversity of 0.59–1.43%, median = 0.95%; $P = 0.0003$ by Student’s $t$ test of a generalized linear model of maximum env diversity).

Sequence diversity in animal CP1W, which was infected by the i.r. route, is illustrated by a NJ tree comprised of 137 env sequences from ramp-up (green ovals) and peak (blue ovals) viremia along with 61 sequences from the SIVmac251 inoculum (black ovals; Fig. 2 A). 102 of 137 CP1W env sequences were identical to themselves and to 4 sequences (K11, K9, G2, and TB2L) in the inoculum. 22 other CP1W env sequences differed from this consensus CP1W env sequence by 1 nt, and 8 CP1W env sequences differed from the consensus by 2 nt. Most of these 30 sequences differed from the consensus by unique mutations, which is consistent with an almost starlike phylogeny (Fig. 2, A and B). Five other CP1W env sequences (RU.I3, RU.4B7, PK.I6, PK.4A29, and PK.I10) differed from the consensus by 3–7 nt, but these were notable for APOBEC-related G-to-A hypermutation (Fig. 2, A and B). Ramp-up sequences sampled 1 wk after infection, and peak viremia samples obtained 2 wk after infection, each coalesced to the same consensus sequence, which was identical to sequences in the inoculum. Thus, a single virus established productive clinical infection in CP1W, and it could be traced back from peak viremia through the eclipse phase of infection to the moment of transmission and into the inoculum.

Single, low-diversity viral lineages with starlike phylogeny and Poisson distributed substitutions were also found in i.r.-infected animals PBE (Fig. 3 A), CG87 (Fig. 4 A), AV66, CR2A, and CR53 (Table I). In each animal, sequences from ramp-up and peak viremia were not available for analysis, so we instead analyzed samples taken 2 wk after peak viremia, corresponding to 4–5 wk after infection (Fig. 1). The phylogenetic tree of CP23 sequences was distinctly different from the phylogenies of sequences from all other animals (Fig. S2). The Highlighter plot revealed an extraordinary concentration of unique and shared mutations confined to a short stretch of 19 nt in the gp41 coding region of env overlapping the nef gene (Fig. S2). Here, 32 out of 37 sequences in the env reading frame, and 34 out of 37 sequences in the nef reading frame, contained amino acid substitutions compared with the transmitted/founder sequence and the consensus of sequences present in the SIVsmE660 inoculum. The clustered mutations in the Env reading frame were contiguous with a 9-mer Env peptide (FHEAVQAVW) corresponding to a known Mamu-B*17–restricted CTL epitope (43). Animal CP23 was Mamu-B*17 positive. These findings thus indicated that monkey CP23 was infected by a single virus, but that this virus evolved extremely rapidly, most likely as a consequence of CTL selection pressure, such that by 4–5 wk after infection, >85% of viral sequences were escape variants. Similar patterns and kinetics of rapid CTL escape have been reported for SIV and HIV-1 (5, 30, 35, 44, 45).

Six other animals (CG7V, CP3C, CG7G, AK9F, CT76, and AH4X) were infected by i.r. virus inoculations but showed substantially greater maximum pairwise diversity among their sequences (0.59–1.43%) than did the animals that were productively infected by only a single virus (0.07–0.15%, excluding CP23). NJ trees and Highlighter plots of sequences from these six animals revealed two or more discrete low-diversity env lineages, each representing a distinct transmitted/founder virus (Figs. 5, 6, and S4). Overall, there was a trend observed between the virus inoculum dose that each animal received and the number of transmitted viral
Figure 2. Viral sequence analysis after intrarectal SIVmac251 infection of animal CP1W. (A) NJ tree of sequences from ramp-up (green symbols) or peak (blue symbols) viremia or from the inoculum (black symbols). (B) Highlighter alignment of all CP1W sequences compared with four identical sequences from the inoculum and to the transmitted/founder virus. Nucleotide polymorphisms are indicated by a colored tic mark (thymine in red, guanine in orange, adenine in green, and cytosine in blue). APOBEC-mediated G-to-A mutations are indicated by purple tics, and deletions are indicated by gray tics. (C) Poisson model of early viral diversity. Brown labels indicate sequences with three or more APOBEC-mediated G-to-A mutations compared with consensus.
variants leading to productive clinical infection: among the 12 animals receiving the two highest doses of virus in the initial inoculation-observation cycle, a minimum of 26 viruses were transmitted to 10 animals. Among the six animals inoculated with the lowest dose of virus, one virus was transmitted to a single animal. However, this dose–response trend was not statistically significant, as animals receiving the highest virus inoculum did not differ from animals receiving the intermediate virus inoculum in the numbers of viruses leading to productive clinical infection.

Five animals did not become infected by the i.r. route despite 18 weekly inoculations (Fig. 1). Each of these animals did, however, become productively infected after a single i.v. inoculation of virus (2 × 10^5 vRNA copies). This i.v. dose was 300-fold lower than the maximum single i.r. dose (6 × 10^7) that each of the five animals had received and ~2,000-fold lower than the cumulative i.r. dose that each animal had received (Fig. 1 and Table I). Figs. 3 B and 4 B show NJ trees and Highlighter plots of env sequences from i.v.-infected animals CG71 and CP37. Even when animals were infected with multiple viruses, the progeny of transmitted/founder viruses could be identified with certainty in cases where two or more sequences were identical (e.g., CG71 env sequences corresponding to variants 1, 2, 6, 7 and 8 in Fig. 3 B), and with high likelihood if two or more sequences differed by only one or few nucleotides (e.g., CG71 env sequences corresponding to variants 3, 4, and 5 in Fig. 3 B). Identification of transmitted/founder viruses was less certain for individual sequences that were dispersed throughout the inoculum tree because such viruses could represent either transmitted variants (that replicated less efficiently) or unique recombinant viruses between two or more transmitted lineages. Thus, we made a conservative estimate that i.v. inoculation of animal CG71 resulted in productive infection by at least eight transmitted viruses. We reached a similar conclusion for animal CP37 (Fig. 4 B), for which there was evidence of productive infection by at least nine viruses. These are minimal estimates because as the number of transmitted viruses increases, the likelihood of sampling at least one of its progeny becomes increasingly dependent on the total number of sequences analyzed. For the five animals infected by the i.v. route, the number of transmitted/founder viruses that we identified ranged from 1 to >9 with a median of 4 (Table I). Because each of these animals had not been infected by a cumulative i.r. dose of 3.7–4.3 × 10^8 vRNA equivalents, but had become infected by 2 × 10^5 vRNA equivalents given i.v., we could use these values together with the numbers of transmitted/founder viruses in each animal to estimate the relative efficiency of virus transmission by i.v. versus i.r. routes. We thus determined that i.v. transmission was 2,000–20,000-fold more efficient than i.r. transmission (Table I).

**Detection of transmitted variants with minor representation**

In some animals, the representation of viral lineages corresponding to transmitted/founder viruses was far from even. For example, animal CT76 (Fig. S4) became infected by the i.r. route by four transmitted/founder viruses, two of which were represented only transiently in the ramp-up sample (variants 3 and 4). In CT76, the data can best be explained by variants 1 and 2 outgrowing variants 3 and 4, which were lost altogether from the sampled sequences by peak viremia. Macaque CG7V (Fig. 5) is a second example of an animal infected i.r. by five viruses, two represented by predominant env sequence lineages and three by single sequences (variants 3, 4, and 5). Macaque CP3C (Fig. 6) is a third example of an animal that was infected i.r. by at least three viruses, two of which were well represented and one that was represented by only a single ramp-up sequence (variant 3). We amplified an additional 220 env genes from ramp-up plasma viruses from this animal and identified just one additional member of this transmitted/founder lineage (unpublished data). Thus, this rare variant represented <1% (2/269 sequences) of the replicating virus population. Notably, among the 269 sequences analyzed, we found no additional transmitted/founder virus lineages beyond the three depicted in Fig. 6. Animals CG7G and AK39F (Fig. S5) are two more examples of monkeys infected i.r. by multiple viruses, one or two of which were represented by lineages containing only single-sampled variants. Animals infected by the i.v. route were even more complicated because of the greater number of transmitted viruses overall (Figs. 3 B, 4 B, 7, S3, and S5). Compounding the unequal representation of sequences resulting from differences in replication efficiencies was the effect of recombination between viruses of two or more transmitted lineages. This is best seen in animal CG7V (Fig. 5), where recombinant viruses outnumbered one of the two principal transmitted lineages by nearly 3 to 1, and the minor variants by 20 to 1. Recombination was also observed in sequences from animals CG7G, CR54, CG71, and CG5G (Table I and Fig. S3).

**Phylogenetic distribution of transmitted/founder viruses in the inocula**

Sequences from all SIVmac251-infected animals and from the corresponding SIVmac251 inoculum are depicted in a single NJ tree (Fig. 7). Animals AV74, CG71, and CG5G were each inoculated by the i.v. route, and the >14 transmitted/founder sequences that productively infected these animals are distributed widely throughout the tree. Several of these sequences differ from inoculum sequences by only 1 nt. Also represented in this NJ tree are sequences from seven animals infected by the i.r. route. Sequences from each of these animals are represented by one or more discrete low-diversity lineage whose consensus either matches an inoculum sequence (CP1W) or differs by only 1 or 2 nt (CT76, CR2A, and AH4X). Again, these transmitted/founder virus lineages are distributed widely throughout the tree. A similar pattern of inoculum and transmitted/founder SIVsmE660 env sequences is shown in Fig. S5. These findings highlight the diversity of env sequences in the SIVsmE660 and SIVmac251 inocula that are capable of mediating virus transmission and productive replication in macaques by i.v. or i.r. routes.
Figure 3. Viral sequence diversity after i.r. or i.v. SIVmac251 infection. (A) NJ and Highlighter analyses of sequences from i.r.-infected animal PBE at ramp-up (green symbols) or peak (blue symbols) viremia compared with inoculum sequences (black symbols). (B) i.v.-infected animal CG71. Brown labels indicate sequences with three or more APOBEC-mediated G-to-A mutations compared with consensus.
Figure 4. Viral sequence diversity after i.r. or i.v. SIVsmE660 infection. (A) NJ and Highlighter analyses of sequences from i.r.-infected animal CG87 at ramp-up (green symbols) or peak (blue symbols) viremia compared with inoculum sequences (black symbols). (B) i.v.-infected animal CP37. Nine transmitted/founder variant lineages are indicated, although additional minor variants and recombinants are also evident (see text). Bracket indicates a single-nucleotide insertion.
Model analysis of SIV diversification
We previously described a mathematical model of HIV-1 replication and diversification in acute infection (5) using estimated parameters of HIV-1 generation time, reproductive ratio, and RT error rate and assuming that the initial virus replicates exponentially, infecting $R_0$ new cells at each generation and diversifying under a model of evolution that assumes no selection, no back mutations, and a constant mutation rate across positions and lineages. Here, we applied this model to an analysis of 987 SIV env sequences and asked if measured parameters of virus diversification were consistent with the assumptions and predictions of this model. We examined data from all 18 animals. For those animals infected by more than one transmitted/founder virus, we selected a predominant lineage for analysis and excluded minor lineages and recombinant viruses. One animal (CP37) could not be analyzed for model conformation because it was infected by so many viruses that none formed a sufficiently predominant lineage (Fig. 4 B). A second animal (CP23) was included in the analysis, but with the caveat that sequences were obtained from a time point 2 wk after peak viremia, and 4–5 wk after infection, when rapid CTL selection is known to occur in SIV-infected macaques (44, 45) and HIV-1–infected humans (5, 30, 35). For each animal, we obtained the frequency distribution of all intersequence Hamming distances (HDs; defined as the number of base positions at which two genomes differ) within a lineage and determined if it deviated from a Poisson model using a $\chi^2$ goodness–of–fit test. We then determined whether or not the observed sequences evolved under a star phylogeny model (i.e., all evolving sequences are equally likely and all coalesce at the founder) in the expected time frame based on the known or estimated date of infection. Sequences from 16 of 17 animals exhibited an almost starlike phylogeny, the one exception being animal CP23 (Table I). Sequences from only 5 of 17 animals exhibited a Poisson distribution of base substitutions. The other 12 animals had sequences that showed enrichment for G-to-A hypermutation with APOBEC signatures (Table I). In 7 of these 12 animals, the evidence for hypermutation was restricted to a single sequence, whereas in 5 others, G-to-A hypermutation was enriched across multiple sequences. When these APOBEC–associated mutations were excluded from the analysis, a good fit to the Poisson model was restored for sequences from all 12 animals (Table I). 10 animals had rare sequences

<table>
<thead>
<tr>
<th>Subject</th>
<th>Initial dose</th>
<th>Cumulative dose before first + vRNA (wk)</th>
<th>Time to infection</th>
<th>Peak viral load</th>
<th>No. of env seq. identified</th>
<th>No. of transmitted variants</th>
<th>Percent max. env diversity</th>
<th>APOBEC-mediated G-to-A mutations</th>
<th>No. of APOBEC-mutated sequences</th>
<th>Intra-host recombination</th>
<th>Starlike phylogeny</th>
<th>Poisson distribution</th>
<th>Rectal barrier^c</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIVmacE660-infected animals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG7V</td>
<td>6 $\times$ 10^9</td>
<td>1.2 $\times$ 10^8</td>
<td>2</td>
<td>i.r.</td>
<td>77,871,700</td>
<td>96</td>
<td>5</td>
<td>1.06</td>
<td>Yes</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CP3C</td>
<td>6 $\times$ 10^9</td>
<td>1.2 $\times$ 10^8</td>
<td>2</td>
<td>i.r.</td>
<td>1,112,900</td>
<td>49</td>
<td>3</td>
<td>0.84</td>
<td>Yes</td>
<td>1</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CG87</td>
<td>6 $\times$ 10^9</td>
<td>3.6 $\times$ 10^8</td>
<td>6</td>
<td>i.r.</td>
<td>479,543</td>
<td>56</td>
<td>1</td>
<td>0.15</td>
<td>No</td>
<td>–</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CG7G</td>
<td>6 $\times$ 10^9</td>
<td>6.0 $\times$ 10^8</td>
<td>1</td>
<td>i.r.</td>
<td>7,084,600</td>
<td>43</td>
<td>5</td>
<td>1.21</td>
<td>Yes</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AK9F</td>
<td>6 $\times$ 10^9</td>
<td>6.0 $\times$ 10^8</td>
<td>6</td>
<td>i.r.</td>
<td>140,795</td>
<td>35</td>
<td>4</td>
<td>1.43</td>
<td>No</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CP37</td>
<td>6 $\times$ 10^9</td>
<td>4.3 $\times$ 10^8</td>
<td>18</td>
<td>i.v.</td>
<td>94,394,600</td>
<td>62</td>
<td>&gt;9</td>
<td>1.54</td>
<td>N/A^d</td>
<td>N/A^d</td>
<td>N/A^d</td>
<td>N/A^d</td>
<td>N/A^d</td>
</tr>
<tr>
<td>CP23</td>
<td>6 $\times$ 10^9</td>
<td>1.2 $\times$ 10^9</td>
<td>2</td>
<td>i.r.</td>
<td>74,000,000</td>
<td>37</td>
<td>1</td>
<td>0.33^c</td>
<td>Yes</td>
<td>1</td>
<td>N/A</td>
<td>No^c</td>
<td>No^c</td>
</tr>
<tr>
<td>CR54</td>
<td>6 $\times$ 10^9</td>
<td>3.7 $\times$ 10^8</td>
<td>18</td>
<td>i.v.</td>
<td>1,968,800</td>
<td>29</td>
<td>4</td>
<td>1.26</td>
<td>Yes</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SIVmac251-infected animals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1W</td>
<td>6 $\times$ 10^9</td>
<td>6.0 $\times$ 10^7</td>
<td>1</td>
<td>i.r.</td>
<td>17,797,900</td>
<td>137</td>
<td>1</td>
<td>0.15</td>
<td>Yes</td>
<td>5</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CT76</td>
<td>6 $\times$ 10^9</td>
<td>6.0 $\times$ 10^7</td>
<td>1</td>
<td>i.r.</td>
<td>18,233,200</td>
<td>50</td>
<td>4</td>
<td>0.66</td>
<td>Yes</td>
<td>1</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PBE</td>
<td>6 $\times$ 10^9</td>
<td>3.0 $\times$ 10^8</td>
<td>5</td>
<td>i.r.</td>
<td>10,955,200</td>
<td>41</td>
<td>1</td>
<td>0.15</td>
<td>No</td>
<td>–</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CR2A</td>
<td>6 $\times$ 10^9</td>
<td>1.2 $\times$ 10^7</td>
<td>2</td>
<td>i.r.</td>
<td>4,130,600</td>
<td>46</td>
<td>1</td>
<td>0.07</td>
<td>No</td>
<td>–</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CR53</td>
<td>6 $\times$ 10^9</td>
<td>3.6 $\times$ 10^7</td>
<td>6</td>
<td>i.r.</td>
<td>6,341,200</td>
<td>30</td>
<td>1</td>
<td>0.15</td>
<td>Yes</td>
<td>1</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AV74</td>
<td>6 $\times$ 10^9</td>
<td>4.3 $\times$ 10^8</td>
<td>18</td>
<td>i.v.</td>
<td>18,690,000</td>
<td>71</td>
<td>1</td>
<td>0.18</td>
<td>Yes</td>
<td>3</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AH4X</td>
<td>6 $\times$ 10^9</td>
<td>2.5 $\times$ 10^9</td>
<td>16</td>
<td>i.r.</td>
<td>2,611,600</td>
<td>23</td>
<td>2</td>
<td>0.59</td>
<td>No</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CG71</td>
<td>6 $\times$ 10^9</td>
<td>3.7 $\times$ 10^8</td>
<td>18</td>
<td>i.v.</td>
<td>8,857,400</td>
<td>67</td>
<td>&gt;8</td>
<td>0.73</td>
<td>Yes</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CG5G</td>
<td>6 $\times$ 10^9</td>
<td>3.7 $\times$ 10^8</td>
<td>18</td>
<td>i.v.</td>
<td>12,938,400</td>
<td>71</td>
<td>4</td>
<td>0.81</td>
<td>Yes</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AV66</td>
<td>6 $\times$ 10^9</td>
<td>6.7 $\times$ 10^7</td>
<td>13</td>
<td>i.v.</td>
<td>4,858,300</td>
<td>44</td>
<td>1</td>
<td>0.07</td>
<td>Yes</td>
<td>1</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

^c Rectal barrier was calculated by dividing the cumulative i.r. dose by the i.v. dose multiplied by the number of transmitted viruses.

^d Not applicable; too many variants to quantify.

^e Sampled 2 wk after peak viral load.
that exhibited shared mutations. For example, CG87 (Fig. 4 A) contained 3 of 56 sequences that shared one nucleotide polymorphism, 3 other sequences that shared a different polymorphism, and 3 additional sequences that shared 1 or 2 still different polymorphisms. Animal CT76 (Fig. S4) had 3 of 50 sequences that shared a single-nucleotide polymorphism. Based on the temporal appearance and patterns of these mutations and a mathematical modeling estimate of the expected frequency of sequences having one or more shared mutations in a gene the size of env (5, 30), we could explain these rare sublineages as resulting from stochastic mutations generated shortly after transmission. In contrast, sequences from animal CP23 obtained 2 wk after peak viremia contained a far higher number of shared mutations than did sequences from any other animal. Most of these shared mutations were narrowly confined to a short stretch of 19 nt. Thus, CP23 sequences severely violated model predictions for random variation as a consequence of early selection, similar to findings we made in HIV-1–infected humans who were sampled within 2 wk after peak viremia (5, 30, 35). Sequences from animals infected by more than one virus also violated model expectations for a Poisson distribution and starlike phylogeny of mutations, but did conform to the model when env sublineages were analyzed individually (Table I).

Table II summarizes model estimates of time to a MRCA sequence in each animal. For this analysis, we used Bayesian (46, 47) and Poisson-based (5) approaches (Supplemental materials and methods). Bayesian and Poisson estimates of a MRCA of sequences in most animals were consistent with the known or estimated days since infection. This is best seen in animal CP1W, where 137 sequences (69 from ramp-up and 68 from peak viremia) were analyzed, and for which both the duration of infection before plasma sampling and the env sequence of the transmitted virus were known with certainty. Here, we observed the Poisson model estimate of a MRCA for ramp-up sequences to be 6 d (confidence interval [CI] 1–10 d) and for peak viremia sequences to be 10 d (CI 5–15 d). Bayesian calculations used sequence information from both time points to...
estimate a coalescent MRCA for the peak viremia sequences of 12 d (CI 8–20 d). Thus, Poisson and Bayesian estimates concurred closely with each other and with the known duration of infection of the animal. Model estimates of time to a MRCA were not close to the known duration of infection for animal CP23, which was sampled 2 wk after peak viremia when there was evidence of strong selection for mutations. Selected mutations are nonrandom and violate model assumptions. For CP23, the Bayesian estimate for a MRCA was 50 d (CI 28–84 d) and the Poisson estimate was 55 d (CI 47–62 d), exceeding the known duration of infection of 28–35 d. For three animals (CR2A, CG5G, and AV66), the Poisson estimates and CIs of days since a MRCA were less than the known or estimated duration of virus infection, possibly because of purifying selection, differences in virus replication parameters such as virus generation time, RT error frequency or $R_0$ compared with model assumptions, or a sampling effect.

**DISCUSSION**

Previously, we developed a conceptual framework, mathematical model, and empirical dataset of HIV-1 *env* sequences that together suggested transmitted/founder viruses responsible for productive clinical infection could be identified by phylogenetic inference in acutely infected subjects (5). A key inference from this model and the supporting HIV-1 sequence datasets (5, 28–35) is that between the time the first cell is infected in the mucosa or submucosa and the time peak viremia is reached ($\sim$3–5 wk later (eclipse phase and Fiebig stages I and II), viruses diversify essentially randomly with little or no evidence of biological selection. This is in contrast to the period immediately thereafter (Fiebig stages III-V), when evidence of strong CTL selection generally becomes apparent (5, 30, 35). Using pyrosequencing technologies that allow for a more extensive sampling of sequences, we have recently found evidence of CTL selection beginning as early as Fiebig stage II (unpublished data). A second key inference from the model and HIV-1 datasets is that by analyzing *env* sequence diversity among 30–50 plasma viruses sampled near the time of peak viremia, sufficient sequence information is available to allow for an unambiguous phylogenetic inference of transmitted/founder virus(es) weeks earlier. This conclusion is based on power calculations that provide 95% confidence limits for detecting minor sequence variants and the empirical observation.

![Figure 6. Unequal representation of transmitted viruses.](image)

Transmitted variants 1 and 2 in animal CP3C are well represented at ramp-up (green symbols) and peak (blue symbols) viremia compared with a third transmitted variant (red asterisk) that is represented by only a single sequence at ramp-up. Inoculum sequences are indicated by black symbols.
Figure 7. Composite SIVmac251 tree from i.r.- and i.v.-infected animals. The NJ tree depicts numerous transmitted/founder virus lineages distributed throughout a diverse set of inoculum sequences (black symbols). Most i.r.- or i.v.-transmitted viruses are represented by discrete, low-diversity lineages. Animal designation, route of infection, and numbers of transmitted viruses are indicated.
that between transmission and peak viremia (Fiebig stage II), it is uncommon to find sequences that exhibit shared polymorphisms resulting from immune selection (5,30); when this does occur, it is necessary to analyze sequential samples to distinguish transmitted/founder sequences from selected mutants. In this study, we tested our model assumptions and inferences directly in the SIV–macaque infection model where experimental conditions, including the composition of the virus inoculum and the interval between virus inoculation and plasma sampling for env sequence analysis, could be better defined (Fig. 1).

Fig. 2 illustrates ramp-up and peak viremia plasma vRNA sequences in animal CP1W after a typical i.r. infection by SIVmac251 and demonstrates many of the salient features of the SGA-direct sequence analysis. After elimination of five sequences that contained APOBEC-related G-to-A hypermutations, the remaining 132 env sequences conformed to a star-like phylogeny and exhibited a Poisson

### Table II. Model analysis of early SIV evolution in SIVsmE660- and SIVmac251-infected rhesus macaques

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sample time</th>
<th>Infection route</th>
<th>No. of sequences examined</th>
<th>Max. HD</th>
<th>BEAST estimated days since MRCA (95% CI)</th>
<th>Poisson estimated days since MRCA (95% CI)</th>
<th>( \lambda^a )</th>
<th>SD(^b )</th>
<th>Goodness of fit p-value(^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SIVsmE660-infected animals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG7V</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>39</td>
<td>2</td>
<td>6 (3,8)</td>
<td>0.308</td>
<td>0.082</td>
<td>0.584</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>23</td>
<td>4</td>
<td>12 (8,19)</td>
<td>13 (7,20)</td>
<td>0.762</td>
<td>0.190</td>
<td>0.840</td>
<td></td>
</tr>
<tr>
<td>CP3C</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>19</td>
<td>3</td>
<td>13 (6,20)</td>
<td>1.053</td>
<td>0.189</td>
<td>0.913</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>10</td>
<td>3</td>
<td>17 (10,29)</td>
<td>11 (3,18)</td>
<td>1.000</td>
<td>0.220</td>
<td>0.815</td>
<td></td>
</tr>
<tr>
<td>CG87</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>23</td>
<td>4</td>
<td>15 (8,22)</td>
<td>0.854</td>
<td>0.210</td>
<td>0.663</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>33</td>
<td>4</td>
<td>20 (12,34)</td>
<td>23 (18,28)</td>
<td>1.311</td>
<td>0.156</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td>CG7G</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>18</td>
<td>2</td>
<td>12 (6,17)</td>
<td>0.667</td>
<td>0.159</td>
<td>0.317</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>17</td>
<td>3</td>
<td>16 (10,27)</td>
<td>13 (6,20)</td>
<td>0.750</td>
<td>0.209</td>
<td>0.865</td>
<td></td>
</tr>
<tr>
<td>AK9F</td>
<td>Peak</td>
<td>i.r.</td>
<td>19</td>
<td>4</td>
<td>12 (2,43)</td>
<td>0.807</td>
<td>0.216</td>
<td>0.886</td>
<td></td>
</tr>
<tr>
<td>CP37(^d )</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>CP23</td>
<td>2 wk after peak</td>
<td>i.r.</td>
<td>37</td>
<td>8</td>
<td>50 (28,84)</td>
<td>55 (47,62)</td>
<td>3.129</td>
<td>0.221</td>
<td>0.038</td>
</tr>
<tr>
<td>CR54</td>
<td>Peak</td>
<td>i.v.</td>
<td>17</td>
<td>3</td>
<td>14 (2,52)</td>
<td>0.706</td>
<td>0.206</td>
<td>0.874</td>
<td></td>
</tr>
<tr>
<td><strong>SIVmac251-infected animals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1W</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>69</td>
<td>4</td>
<td>6 (1,10)</td>
<td>0.316</td>
<td>0.123</td>
<td>0.654</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>68</td>
<td>4</td>
<td>12 (8,20)</td>
<td>10 (5,15)</td>
<td>0.552</td>
<td>0.143</td>
<td>0.906</td>
<td></td>
</tr>
<tr>
<td>CT76</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>20</td>
<td>3</td>
<td>17 (10,24)</td>
<td>0.968</td>
<td>0.208</td>
<td>0.809</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>22</td>
<td>3</td>
<td>16 (7,34)</td>
<td>7 (1,12)</td>
<td>0.455</td>
<td>0.153</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>PBE</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>17</td>
<td>2</td>
<td>6 (2,11)</td>
<td>0.353</td>
<td>0.139</td>
<td>0.646</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>24</td>
<td>4</td>
<td>13 (8,21)</td>
<td>12 (7,17)</td>
<td>0.667</td>
<td>0.142</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td>CR2A</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>19</td>
<td>3</td>
<td>2 (0,5)</td>
<td>0.105</td>
<td>0.077</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>27</td>
<td>2</td>
<td>8 (7,14)</td>
<td>5 (2,9)</td>
<td>0.296</td>
<td>0.101</td>
<td>0.637</td>
<td></td>
</tr>
<tr>
<td>CR53</td>
<td>Peak</td>
<td>i.r.</td>
<td>30</td>
<td>4</td>
<td>12 (3,58)</td>
<td>0.667</td>
<td>0.166</td>
<td>0.478</td>
<td></td>
</tr>
<tr>
<td>AV74</td>
<td>Ramp-up</td>
<td>i.v.</td>
<td>35</td>
<td>3</td>
<td>5 (2,9)</td>
<td>0.284</td>
<td>0.103</td>
<td>0.703</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.v.</td>
<td>36</td>
<td>3</td>
<td>12 (8,20)</td>
<td>14 (8,20)</td>
<td>0.778</td>
<td>0.174</td>
<td>0.621</td>
<td></td>
</tr>
<tr>
<td>AH4X</td>
<td>Peak</td>
<td>i.r.</td>
<td>21</td>
<td>6</td>
<td>11 (2,80)</td>
<td>0.561</td>
<td>0.173</td>
<td>0.894</td>
<td></td>
</tr>
<tr>
<td>CG71</td>
<td>Ramp-up</td>
<td>i.v.</td>
<td>22</td>
<td>2</td>
<td>7 (2,11)</td>
<td>0.364</td>
<td>0.124</td>
<td>0.617</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.v.</td>
<td>16</td>
<td>5</td>
<td>14 (8,25)</td>
<td>17 (8,27)</td>
<td>0.983</td>
<td>0.279</td>
<td>0.630</td>
<td></td>
</tr>
<tr>
<td>CG5G</td>
<td>Ramp-up</td>
<td>i.v.</td>
<td>24</td>
<td>4</td>
<td>6 (1,11)</td>
<td>0.333</td>
<td>0.137</td>
<td>0.655</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.v.</td>
<td>17</td>
<td>4</td>
<td>9 (7,14)</td>
<td>6 (2,11)</td>
<td>0.353</td>
<td>0.136</td>
<td>0.646</td>
<td></td>
</tr>
<tr>
<td>AV66</td>
<td>Ramp-up</td>
<td>i.r.</td>
<td>20</td>
<td>2</td>
<td>5 (1,9)</td>
<td>0.300</td>
<td>0.115</td>
<td>0.657</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>i.r.</td>
<td>24</td>
<td>2</td>
<td>9 (7,28)</td>
<td>3 (0,6)</td>
<td>0.167</td>
<td>0.082</td>
<td>0.713</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Mean of best fitting Poisson distribution defined by the free parameter \( \lambda \).

\(^b\)SD of \( \lambda \) obtained by using jacknife.

\(^c\)P < 0.05 indicates that sequences deviate significantly from a Poisson distribution.

\(^d\)Not applicable; too many variants to identify major lineage.
distribution of nucleotide substitutions. At 7 d after infection, 59 of 67 (88%) sequences were identical, and at 14 d after infection, 51 of 65 (78%) sequences were identical. This degree of early SIV *env* sequence homogeneity followed by a decline in sequence identity from 88% to 78% over a 7-d period is strikingly similar to findings made in humans acutely infected by HIV-1 and conforms closely to model projections previously reported (5). Bayesian and Poisson estimations of days from a MRCA were consistent with the known duration of infection in this animal (Table II), and an unambiguous transmitted/founder SIVmac251 sequence could be inferred from the consensus of sequences at ramp-up and peak viremia time points. The validity of this conclusion was confirmed by an analysis of the SIVmac251 inoculum (Fig. 2), where four *env* sequences were found that were identical to the transmitted/founder *env* sequence in animal CP1W. This result shows that a transmitted SIV genome can be identified in an inoculum, and can then be traced across the rectal mucosa of the newly infected host and through the period of massive virus expansion in gut-associated and peripheral lymphoid tissues that leads ultimately to peak plasma viremia (2, 3, 19, 45, 48–52). Recent studies in human transmission pairs where identical HIV-1 *env* sequences were identified in donor and recipient plasma after sexual HIV-1 transmission corroborate these findings (28, 33).

Despite the identification of one or more transmitted/founder *env* genes in all 18 animals studied, we note that the detection of such discrete, low-diversity SIV lineages (representing the progeny of transmitted/founder viruses) does not mean that these represented the *only* viruses transmitted across the mucosal epithelium. It is possible that additional viruses were transmitted across the mucosa, but went undetected because of limited sampling, stochastic infection events, abortive infection, defective viruses, innate immune responses, unequal replication rates of competing virus lineages, or virus compartmentalization (26, 53–56). The same is true for HIV-1 *env* sequences identified in donor and recipient plasma after sexual HIV-1 transmission corroborate these findings (28, 33).

Three features of SIV mucosal transmission that differ from HIV-1 mucosal transmission are a shorter eclipse period for SIV, a higher frequency of G-to-A hypermutated SIV sequences, and a higher frequency of detection of minor transmitted SIV variants (identified as rare variant sequences in ramp-up or peak viremia samples). Although the design of this study, which included weekly i.r. inoculations of virus and weekly plasma sampling, was not intended to provide a precise estimate of the duration of the eclipse period, we could determine an upper boundary for the eclipse phase in animals CP1W, CT76, CG7G, and AV66. In each case, it was <7 d. This is substantially shorter than the average eclipse phase of HIV-1 in humans of ~10–14 d, and which may exceed 21 d (for review see [5]). Potential explanations for this difference include an ~10-fold lower plasma and extracellular fluid volume in macaques into which virus can distribute, together with a higher exponential growth rate and reproductive ratio for SIV compared with HIV-1 (27, 59). A second difference between SIV and HIV-1 mucosal transmission was a substantially higher frequency of detection of SIV sequences with APOBEC-related G-to-A hypermutation in the acute and early infection period. For example, SIV *env* sequences from 12 of 18 macaques initially did not conform to a Poisson model of random virus evolution because of APOBEC hypermutation, whereas HIV-1 *env* sequences from only 13 of 81 humans (5) deviated...
from the model because of APOBEC hypermutation (P < 0.0001; Fisher’s exact test). In a separate study, we also found examples of excessive G-to-A hypermutation of SIV in mucosally infected animals in the acute infection period (60). Heterogeneity of vif sequences in the SIV inoculum and in the transmitted/founder viruses that arise from them, or a greater evolutionary mismatch between SIVsm (SIV-mac) Vif proteins and the Indian rhesus macaque APOBEC3, compared with SIVcpz (HIV-1) Vif proteins and human APOBEC3 (61), may contribute to these differences. A third difference we found between mucosal SIV and HIV-1 infection was in the numbers of minor virus variants detected in ramp-up or peak viremia plasma samples. Rare virus variants were found in 5 of 13 mucosally infected macaques compared with 4 of 98 human subjects (5) (P = 0.001; Fisher’s exact test). This difference may be caused by the fact that macaques were inoculated i.r. with higher titers of virus having greater diversity in replicative potential (as a consequence of in vitro propagation) compared with viruses that is transmitted sexually in humans (4, 58, 62, 63) or to later sampling of humans (5).

In summary, this study demonstrates that i.r. infection of rhesus macaques recapitulates at a biological and molecular level many of the features of human HIV-1 mucosal transmission. When the SIV inoculum is sufficiently low, mucosal transmission is inefficient and dose related; higher SIV doses, like higher HIV-1 titers in humans (58, 62–64), are associated with higher transmission rates. The numbers of transmitted/founder viruses yielding productive clinical infection in mucosally infected humans (generally 1–5; median = 1; [5, 28, 29, 31–35]) and macaques (1–5; median = 1) were low, reflecting a substantial bottleneck in mucosal transmission in both species. In neither humans nor macaques did the numbers of transmitted viruses after mucosal exposure correspond to a Poisson distribution of independent, low-probability events (5, 31). Diversification of SIV and HIV–1 before peak viremia was essentially random and without evidence of immune selection; between peak viremia and 2 wk hence, there was evidence of strong CTL selection for escape variants in both macaques and humans (Fig. S2) (5, 30, 35, 44, 45). In acute SIV and HIV–1 infection, interlineage recombination was common (5). Further studies are needed to elucidate the precise nature of the bottleneck to mucosal transmission of SIV and HIV–1 and how it compares in rectal versus vaginal transmission. It will also be important to determine if the identification and enumeration of transmitted/founder viruses in humans or macaques can augment in a meaningful way the analysis of vaccine-mediated protection or enhancement of infection. Finally, we suggest that the analysis of full-length transmitted/founder viruses and their progeny in the SIV–macaque infection model can provide a uniquely informative view of the composite selection pressures shaping virus evolution beginning at or near the moment of transmission and extending to setpoint viremia, as shown recently for HIV–1 (30).

**MATERIALS AND METHODS**

**Study animals.** Frozen (−70°C) plasma specimens collected from 18 adult rhesus macaques (Macaca mulatta) as part of a previous study of low-dose intrarectal SIV inoculation (41) were analyzed. The SIVmac251 and SIVmac536 inoculum stocks and the inoculation procedures were previously described in detail (41). All animals were maintained in an Association for Assessment and Accreditation of Laboratory Animal Care–accredited institution and with the approval of the Animal Care and Use Committees of the National Institutes of Health and Harvard Medical School (41).

**Viral RNA extraction and cDNA synthesis.** From each plasma specimen, ~20,000 viral RNA copies were extracted using the QiAamp Viral RNA Mini kit (QIAGEN). RNA was eluted and immediately subjected to cDNA synthesis. Reverse transcription of RNA to single-stranded cDNA was performed using SuperScript III reverse transcription according to manufacturer’s recommendations (Invitrogen). In brief, each cDNA reaction included 1× RT buffer, 0.5 mM of each deoxynucleoside triphosphate, 5 mM dithiothreitol, 2 U/µl RNaseOUT (RNase inhibitor), 10 U/µl of SuperScript III reverse transcription, and 0.25 mM antisense primer SIVmac/macEnvR1 5′-TGTAAATAAATCCTGTCAGGCCCGGAGGC-3′ (nt 9454–9479 in SIVmac239). The mixture was incubated at 50°C for 60 min, followed by an increase in temperature to 55°C for an additional 60 min. The reaction was then heat-inactivated at 70°C for 15 min and treated with 2 U of RNase H at 37°C for 20 min. The newly synthesized cDNA was used immediately or frozen at −80°C.

**SGA.** cDNA was serially diluted and distributed among wells of replicate 96-well plates so as to identify a dilution where PCR-positive wells constituted <30% of the total number of reactions, as previously described (5, 35). At this dilution, most wells contain amplicons derived from a single cDNA molecule. This was confirmed in every positive well by direct sequencing of the amplicon and inspection of the sequence for mixed bases (double peaks), which would be indicative of priming from more than one original template or the introduction of PCR error in early cycles. Any sequence with mixed bases was excluded from further analysis. PCR amplification was performed in the presence of 1× High Fidelity Platinum PCR buffer, 2 mM MgSO4, 0.2 mM of each deoxynucleoside triphosphate, 0.2 µM of each primer, and 0.025 U/µl Platinum Taq High Fidelity polymerase in a 20-µl reaction (Invitrogen). First round PCR primers included sense primer SIVmacEnvF1 5′-CCTCCCCCTCCAGGACTG-3′ (nt 6127–6146 in SIVmac239) and antisense primer SIVmacEnvR1 5′-TGTAAATAAATCCTGTCAGGCCCGGAGGC-3′ (nt 9454–9479 in SIVmac239), which generated an ~3.3-kb amplicon. PCR was performed in MicroAmp 96-well reaction plates (Applied Biosystems) with the following PCR conditions: 1 cycle of 94°C for 2 min, 35 cycles of a denaturing step of 94°C for 15 s, an annealing step of 55°C for 30 s, and an extension step of 68°C for 4 min, followed by a final extension of 68°C for 10 min. Next, 2 µl from first-round PCR product was added to a second-round PCR reaction that included the sense primer SIVmacEnvF2 5′-TATGATAGACATGAGACACCTTG- AAGGAGCC-3′ (nt 6292–6323 in SIVmac239) or SIVmacEnvF2 5′-TAT- AATAGACAAGAGACACCTTGAGGGAGC-3′ (nt 6292–6323 in Mac239) and antisense primer SIVmacEnvR2 5′-ATGAGACATRTC- CATTGCAATTGTA-3′ (nt 9413–9439 in SIVmac239). The second-round PCR reaction was performed under the same conditions used for first-round PCR, but with a total of 45 cycles. Amplicons were inspected on precast 1% agarose E-gels 96 (Invitrogen). All PCR procedures were performed under PCR, clean room conditions using procedural safeguards against sample contamination, including prealiquoting of all reagents, use of dedicated equipment, and physical separation of sample processing from pre- and post-PCR amplification steps.

**DNA sequencing.** Env gene amplicons were directly sequenced by cycle-sequencing using BigDye terminator chemistry and protocols recommended by the manufacturer (Applied Biosystems). Sequencing reaction products were analyzed with an ABI 3730xl genetic analyzer (Applied Biosystems). Both DNA
strands were sequenced using partially overlapping fragments. Individual sequence fragments for each amplicon were assembled and edited using the Sequencer program 4.7 (Gene Codes). Inspection of individual chromatograms allowed for the confirmation of amplicons derived from single versus multiple templates. The absence of mixed bases at each nucleotide position throughout the entire env gene was taken as evidence of SGA from a single rRNA/cDNA template. This quality control measure allowed us to exclude from the analysis amplicons that resulted from PCR-generated in vitro recombination events or Taq polymerase errors and to obtain individual env sequences that proportionately represented those circulating in vivo in SIV virions.

Sequence alignments. All alignments were initially made with Clustal W (65). Consensus sequences were generated for the sequence set from each individual. The full alignment for SIVamE660 and SIVmac251 is available at www.hiv.lanl.gov/content/sequence/hiv/user_alignments/keeleSIV. All 1090 env sequences from ramp-up and peak viremia and from the inoculum stocks were deposited in GenBank/EMBL/DDBJ (available at accession nos. FJ578007–FJ579096).

Env diversity analysis. A total of 987 full-length SIV env genes of plasma virions from 18 animals (median of 48 per animal; range 23–137) were sequenced. Power calculations indicated that with a sample of 48 sequences, we could be 95% confident not to miss a variant transmission that comprised at least 5% of the total virus population (5). Similarly, with a sample size between 23 and 137, we could be 95% confident not to miss a variant comprising between 3 and 12% of the total virus population. Out of 987 total sequences, 26 were excluded from further analysis because of overt APOBEC-associated G-to-A hypermutation (defined as three or more G-to-A substitutions in APOBEC signature motifs). We next analyzed sequences for maximum env diversity, and then by NJ phylogenies and Highlighter plots (www.hiv.lanl.gov). For each animal having sequences that comprised more than one low-diversity lineage, we identified the lineage with the greatest sequence representation for subsequent analysis. Pairwise HD (defined as the number of base positions at which the two genotypes differ, excluding gaps) were determined for each transmitted/founder lineage. The inoculum stocks of SIVamE660 and SIVmac251 were evaluated for maximum env diversity and by NJ phylogenies and Highlighter plots.

Hypermutated samples. Enrichment for APOBEC mutations violates the assumption of constant mutation rate across positions, as the editing performed by these enzymes is base and context sensitive. Enrichment for mutations with APOBEC signatures was assessed using Hypermut 2.0 (www.hiv.lanl.gov).

Viral recombination. Recombination analysis was performed by visual inspection of Highlighter trees and confirmed by RAP (www.hiv.lanl.gov), GARD (66), or Recco (67) recombination identification tools.

Model analyses. Detailed descriptions of the mathematical models of early sequence evolution, star phylogeny determination, estimates of viral fitness, and power calculations for estimating the likelihood of detecting transmitted env variants with minor representation are provided elsewhere (5). Bayesian (46, 47) and Poisson analyses (5) are described in the Supplemental materials and methods.

Statistical analysis. We used the R software package, Version 2.8.0, to perform hypothesis tests.

Online supplemental material. Bayesian and Poisson analyses are described in the Supplemental materials and methods. Fig. S1 shows NJ trees and Highlighter plots of env sequences from the SIVamE660 and SIVmac251 inocula. Fig. S2 shows viral env sequence diversity 2 wk after peak viremia in an i.r.-infected animal CP23. Fig. S3 shows i.v. transmission of four viruses, followed by recombination in animal CG5G. Fig. S4 shows i.r. transmission of four viruses in animal CT76. Fig. S5 shows a composite NJ tree of SIVamE660 sequences from the inoculum and from i.r.- and i.v.-infected animals. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20080831/DC1.

We thank D. McPherson and B. Cochran for technical assistance; sequencing core facilities of the UAB CFAR, and J. White for manuscript preparation.

This work was supported by the Center for HIV/AIDS Vaccine Immunology and by grants from the National Institutes of Health (AI67854 and AI27767), the Bill and Melinda Gates Foundation (#37874), and the American Foundation for AIDS Research (106997–3). The authors have no conflicting financial interests.

Submitted: 18 December 2008
Accepted: 3 April 2009

REFERENCES


