LA-UR- 08-8081

Approved for public release; distribution is unlimited.

Title:	Low Dose Rectal Inoculation of Rhesus Macaques by SIVsmE660 or SIVmac251 Recapitulates
Author(s):	Brandon F. Keele,University of Alabama Hui Li, University of Alabama Gerald H. Learn, University of Alabama Peter Hraber, Los Alamos National Laboratory Elena E. Giorgi, Los Alamos National Laboratory
Intended for:	Journal of Experimental Medicine

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

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

Brandon F. Keele[1], Hui Li[1], Gerald H. Learn[1], Peter Hraber[2], Elena E. Giorgi[2,3], Truman Grayson[1], Chuanxi Sun[1], Yalu Chen[1], Wendy W. Yeh[4], Norman L. Letvin[4,5], Gary J. Nabel[5], Barton F. Haynes[6], Tanmoy Bhattacharya[2,7], Alan S. Perelson[2], Bette T. Korber[2,7], Beatrice H. Hahn[1] and George M. Shaw[1*]

[1]University of Alabama at Birmingham, Birmingham, Alabama 35223

[2]Los Alamos National Laboratory, Los Alamos NM 87545

[3]University of Massachusetts, Amherst, MA 01002

[4]Division of Viral Pathogenesis, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115

[5]Vaccine Research Center, National Institute of Allergy and Infectious Disease, National Institutes of Heath, Bethesda, MD 20892

[6]Duke University Medical Center, Durham, NC 27710

[7]Santa Fe Institute, Santa Fe, NM 87501

[*]To whom correspondence should be addressed. E-mail: gshaw@uab.edu

ABSTRACT

Recently, we developed a novel approach to identify transmitted or early founder HIV-1 genomes in acutely infected humans using single genome amplification and sequencing. Here we utilized this same approach to determine the molecular features of SIV transmission in 18 experimentally infected Indian rhesus macaques. Animals were inoculated intrarectally (IR) or intravenously (IV) 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 one to five weeks after infection. IR inoculation was followed by productive infection by one or few viruses (median 1; range 1-5) that diversified randomly with near star-like 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 few nucleotides, providing direct evidence that early plasma viral sequences coalesce to transmitted/founder virus(es). IV infection was more than 2,000-fold more efficient than IR 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.

INTRODUCTION

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), and in tissue explant cultures or in Indian rhesus macaques, the HIV-1 or simian immunodeficiency virus (SIV) inocula have typically been high in order to achieve uniform infection of controls or to visualize infection events *in situ* (6-12), thus prompting concerns about the physiological relevance of the model systems (13-16). Further complicating the analysis of early infection events *in vivo* is the viral "eclipse" period during which virus replicates in mucosal and locoregional lymphoreticular tissues but is not detectable in the circulating plasma (17). In SIV-infected macaques, the eclipse period is generally about 5-7 days in duration, and in HIV-1 infected humans, it is approximately 7-21 days (5, 10, 17-19).

Previously, we observed that in the early stages of HIV-1 infection preceding antibody seroconversion [eclipse phase and Fiebig stages I and II; (17)], virus diversification follows a pattern of random evolution with a near star-like phylogeny and a Poisson distribution of nucleotide substitutions (5). We thus hypothesized that the genetic identity of transmitted or early founder viruses could be inferred unambiguously by phylogenetic analysis of discreet low diversity viral lineages that emanate from them. This hypothesis 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 (i) acute viral sequences sampled prior to the development of measurable adaptive immune responses conformed to a pattern of random virus evolution; (ii) viral sequence diversity resulted in model estimates of time to a most recent common ancestor (MRCA)

consistent with clinical histories and Fiebig stage classifications; and (iii) 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 recognizable as a discreet virus lineage (5). These findings have since been corroborated in seven additional patient cohorts representing 238 individuals infected by HIV-1 subtypes A, B, C, or D (20-26). A key innovation common to these studies was the use of single genome amplification (SGA) followed by direct amplicon sequencing to characterize the plasma virus quasispecies. This method provides proportional representation of plasma vRNA and precludes *Taq* polymerase-induced template switching (recombination), *Taq* polymerase-associated nucleotide substitutions in finished sequences, template resampling, and cloning bias (5, 26-29).

In the present study, we sought to test directly our strategy for identifying transmitted/founder viruses in experimentally infected Indian rhesus macaques where we could define essential parameters including route of virus infection, genetic composition of the inoculum, and 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 viral sequences sampled at or near peak viremia coalesce to sequences of viruses responsible for transmission and productive clinical infection weeks earlier; 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, including the extent of the mucosal barrier to virus transmission, the number of transmitted/founder viruses leading to productive infection, and the molecular patterns of early virus diversification.

RESULTS

Inoculation regimen and infection kinetics. Eighteen animals received weekly atraumatic IR inoculations of cell-free SIV as part of a previous study of adaptive immune responses following mucosal virus exposure (30). The virus inocula were uncloned SIVmac251 or SIVsmE660 strains comprised of virus "swarms" with *env* diversity comparable to what is observed in humans one to two years after infection by HIV-1 (31). Nine animals, divided into groups of three, received log dilutions of $6x10^7$, $6x10^6$, or $6x10^5$ vRNA copies of SIVsmE660 (Fig. 1, left panels). Nine other animals, divided into groups of three, received identical dilutions of SIVmac251 (Fig. 1, right panels). Virus was administered weekly for six weeks followed by a three week observation period. Animals that did not become infected (viremic) within that time, underwent a second six week course of IR inoculations at the same virus dose level followed by a three week observation period. Animals that again did not become infected, underwent a third six week course of IR inoculations at the maximum virus dose level ($6x10^7$ vRNA copies) followed by a three week observation period. One animal, AV66, received SIVsmE660 for the initial two 8-week IR inoculation-observation cycles and was then crossed over to receive SIVmac251 for the third IR inoculation-observation cycle (Fig. 1, left bottom panel). Each 8-week IR inoculation-observation cycle was separated from the next by at least a two month rest period to ensure that none of the animals experienced delayed seroconversion. Five animals that did not become infected after the 18 weekly IR inoculations were given a single IV inoculation of $2x10^5$ vRNA copies. All animals became infected.

The kinetics of virus infection and replication, and the time points of plasma virus sampling for *env* sequence analysis, are illustrated (Fig. 1). There was no significant difference in the overall infection rate between macaques inoculated IR 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 six animals in each virus dose group were combined for purposes of statistical analysis. We found a statistically significant virus dose - infection trend to IR infection, with 6 of 6 animals receiving the highest dose of SIV becoming infected in the first 8 week 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 infected in the same period ($p = 0.005$ by 3x2 Fisher's exact test). Two animals, AV66 and AH4X, did not become infected after 12 weekly IR inoculations at the lowest dose of virus ($6x10^5$ vRNA copies) but did become infected after one to four IR inoculations at the highest virus dose ($6x10^7$ vRNA copies). Five animals, CP37, CR54, AV74, CG71 and CG5G, did not become infected even after 18 weekly IR inoculations, but each became infected after a single IV inoculation. For both SIVsmE660 and SIVmac251 infected animals, the kinetics of plasma viral load increase were quite similar, reaching peak viremia within one week of the last negative sample in three animals and within two weeks in 15 others. Peak viral loads ranged from $1.4x10^5$ to $9.4x10^7$ (median $8x10^6$) and were not related to the infecting virus strain or route of infection (Fig. 1 and Table 1).

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 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. Twenty-six sequences (distributed among 12 animals) showed evidence of overt APOBEC G-to-A hypermutation (e.g., sequences RU.I3, RU.4B7, PK.J6, 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 intrarectally 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%), not different from viral diversity in IR infected animals (p=0.11 by Wilcoxon rank sum test). IR 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 to 0.15%, median = 0.15%) or more than one virus (maximum *env* diversity of 0.59 to 1.43%, median = 0.95%; p=0.0003 by t-test of a generalized linear model of maximum *env* diversity).

Sequence diversity in animal CP1W, which was infected by the IR route, is illustrated by a NJ tree comprised of 137 *env* sequences from ramp-up and peak viremia along with 61 sequences from the SIVmac251 inoculum (Fig. 2A). 102 of 137 CP1W *env* sequences were identical to themselves and to four sequences (K11, K9, G2, and TB2L) in the inoculum. Twenty-two other CP1W *env* sequences differed from this consensus CP1W *env* sequence by

one nucleotide and eight CP1W *env* sequences differed from the consensus by two nucleotides. Most of these 30 sequences differed from the consensus by unique mutations, consistent with a near star-like phylogeny (Fig. 2A and 2B). Five other CP1W *env* sequences differed from the consensus by 3 to 7 nucleotides, but these were notable for APOBEC related G-to-A hypermutation (Fig. 2A). Ramp-up sequences sampled one week after infection, and peak viremia samples obtained two weeks 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 star-like phylogeny and Poisson distributed substitutions were also found in IR infected animals PBE (Fig. 3A), CG87 (Fig. 4A), AV66, CR2A and CR53 (Table 1). In each animal, sequences from ramp-up and peak viremia coalesced to single consensus sequences representing transmitted/founder viruses. In two of these animals (PBE and CR2A), the transmitted/founder viruses differed from sequences identified in the inoculum by only 2 nucleotides. In another IR infected animal (CP23), ramp-up and peak viremia samples were not available for analysis, so we instead analyzed samples taken 2 weeks beyond peak viremia corresponding to 4-5 weeks post-infection (Fig. 1). The phylogenetic tree of these sequences was distinctly different from the phylogenies of sequences from all other animals. The *Highlighter* plot revealed an extraordinary concentration of unique and shared polymorphisms confined to a short stretch of 19 nucleotides 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. These findings thus indicated that monkey CP23 was infected by a single virus but that this virus evolved rapidly, most likely as a consequence of CTL selection pressure, such that by 4 to 5 weeks post-infection more than 90% (34 of 37) of viral sequences were escape variants.

Six other animals (CG7V, CP3C, CG7G, AK9F, CT76, and AH4X) were infected by IR 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, S3, S4). Overall, there was a trend observed between the virus inoculum dose that each animal received and the number of transmitted viral 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 6 animals inoculated with the lowest dose of virus, 1 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 transmitted viruses each acquired.

Five animals did not become infected by the IR route despite 18 weekly inoculations (Fig. 1). Each of these animals did, however, become infected after a single IV inoculation of virus ($2x10^5$ vRNA copies). This IV dose was 300-fold lower than the maximum single IR dose ($6x10^7$) that each of the five animals had received and 2,000-fold lower than the cumulative IR dose that each animal had received (Fig. 1 and Table 1). Figures 3B and 4B show NJ trees and *Highlighter* plots of *env* sequences from IV 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. 3B), 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. 3B). Identification of transmitted/founder viruses was less certain for individual sequences that were dispersed throughout the inoculum tree, since 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 IV inoculation of animal CG71 resulted in productive infection by *at least* 8 transmitted viruses. We reached a similar conclusion for animal CP37 (Fig. 4) for which there was evidence of productive infection by *at least* 9 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 IV route, the number of transmitted/founder viruses that we identified ranged from 1 to >9 (Table 1). Because each of these animals had not been infected by a cumulative IR dose of 3.7-4.3 x 10^8 vRNA equivalents, but had become infected by $2x10^5$ vRNA equivalents given IV, 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 IV versus IR routes. We thus determined that IV transmission was 2,000 – 20,000 fold more efficient than IR transmission (Table 1).

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 IR 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 variant 1 and variant 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 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 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 example of this transmitted/founder lineage (data not shown). Thus, this rare variant represented less than 1% (2/269 sequences) of the replicating virus population. Animal CG7G (Fig. S5) is a fourth example of a monkey infected by at least five viruses, two of which were represented by only single variants. Animals infected by the IV route were even more complicated because of the greater number of transmitted viruses overall (Figs. 3B, 4B, 7, S3 and S5). Compounding the unequal representation of sequences resulting from differences in replication fitness 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 1; Fig. S3), but interestingly, not in animals CP3C or CT76 (Figs. 6 and S4).

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 IV route, and the 14-plus transmitted/founder sequences that productively infected these animals are distributed widely throughout the tree. Several of these sequences differ from inoculum sequences by only one nucleotide. Also represented in this NJ tree are sequences from seven animals infected by the IR route. Sequences from each of these animals are represented by one or more discrete low-diversity lineages whose consensus either matches an inoculum sequence (CP1W) or differs by only one or two nucleotides (CT76, CR2A, 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 Figure 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 IV or IR routes.

Model analysis of SIV diversification. We have previously described a mathematical model of HIV-1 replication and diversification in acute infection [see Keele *et al.* (5) including supplementary information available online] using estimated parameters of HIV-1 generation time, reproductive ratio, and reverse transcriptase (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 model

assumptions and predictions. 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. 4B). A second animal (CP23) was included in the analysis but with the caveat that sequences were obtained from a time point 2 weeks after peak viremia, and 4-5 weeks after infection, when CTL selection is known to occur (32). For each animal, we obtained the frequency distribution of all intersequence Hamming distances (HD, 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-squared 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 a near star-like phylogeny, the exception being animal CP23 (Table 1). 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 1). 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 1). Ten animals had rare sequences that exhibited shared mutations. For example, CG87 (Fig. 4A) contained 3 of 56 sequences that shared a single nucleotide polymorphism and three other sequences that shared one or two 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 assessment of the expected frequency of sequences having more than two shared mutations in a gene the size of *env* (5, 33), we could explain these rare sublineages as resulting from stochastic mutations generated shortly after transmission. In contrast, sequences from animal CP23 obtained 2 weeks after peak viremia contained a far higher number of shared mutations than did sequences from any other animal, most of which were confined to a short stretch of 19 nucleotides corresponding to a likely CTL epitope in Env or Nef. Thus, CP23 sequences severely violated model predictions for random variation as a consequence of early selection, similar to findings we made previously for HIV-1 (5, 26). Sequences from animals infected by more than one virus also violated model expectations for a Poisson distribution and near star phylogeny of mutations but did conform to the model when *env* sublineages were analyzed individually (Table 1).

Table 2 summarizes model estimates of time to a most recent common ancestor (MRCA) sequence in each animal. For this analysis, we employed Bayesian (BEAST) and Poisson based methods. The Poisson model evolves a viral population forward in time assuming no selection and using previously measured parameters for HIV viral replication dynamics and compares summary statistics between the model and the data. In contrast, BEAST uses the entire sequence data set for a Bayesian analysis of the coalescent starting with only the previously measured base substitution parameters as prior expectation. The use of summary statistics is expected to make the Poisson method more robust against effects like undetected recombination, whereas the Bayesian method incorporates the modeling uncertainties more completely in its estimates at the expense of statistical power. Both methods yielded comparable estimates of time to the MRCA of HIV-1 in previous studies (5, 33). Because BEAST and Poisson methods both assume that

mutations occur randomly across the genomes, and APOBEC-mediated G-to-A mutations violate this condition, we performed BEAST and Poisson calculations in two iterations. The first iteration included *env* sequences (n=461) representing the major virus lineage from each animal stripped of all APOBEC signature positions (i.e., G within the context of GRD). The second iteration did not eliminate all APOBEC signature positions but did exclude 17 overtly hypermutated sequences. Results of the first analysis are displayed in Table 2, although findings from the second analysis were not substantially different. Bayesian and Poisson estimates for 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 the duration of infection prior to plasma sampling and the *env* sequence of the transmitted virus were both known with certainty. Here, we observe the Poisson model estimate of a MRCA for ramp up sequences to be 6 days (C.I. 1-10 days) and for peak viremia sequences to be 10 days (C.I. 5-15 days). BEAST used sequence information from both time points to estimate a coalescent MRCA for the peak viremia sequences of 12 days (C.I. 8-20 days). Twelve of our animals were sampled at two time points one week apart when the viral load was increasing, and six of them were homogeneous infections that we could test using the Poisson model. The model predicts that the Poisson parameter lambda describing the sequence diversity should increase linearly with time with a coefficient of 2.38×10^{-5} per base position per day. Averaging over the 6 animals, the Poisson parameter was seen to be changing at $1.5(\pm 0.5) \times 10^{-5}$ per base per day, lower than but consistent with the model predictions. Neither model adjusts for "purifying selection" (i.e., mutations that are selected against and go unobserved because they result in less fit viruses), and as a consequence, timing estimates tend to be biased towards a low estimate. Despite this, both

models estimated a time to MRCA closely approximating the known duration of infection in animal CP1W as well as in most other study animals. This was not the case for animal CP23, which was sampled 2 weeks after peak viremia and where there was evidence of strong selection for CTL escape mutations. Selected mutations are nonrandom and violate model assumptions. For CP23, the Bayesian estimate for a MRCA was 50 days (C.I. 28-84 days) and the Poisson estimate 55 days (C.I. 47-62 days), each far exceeding the known duration of infection (28-35 days). For three animals (CR2A, CG5G and AV66), the Poisson estimates and confidence intervals of days since a MRCA were less than the known or estimated duration of virus infection, possibly due to purifying selection, differences in virus replication parameters such as virus generation time, RT error frequency or R_0 compared with model assumptions, or a stochastic sampling effect.

DISCUSSION

Previously, we developed a conceptual framework, mathematical model, and empirical data set 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). This conclusion has been supported by studies in seven other acute infection cohorts (20-26), including one study where the sequences of complete (9Kb) transmitted/founder HIV-1 genomes in twelve subjects were determined (34). A key inference from this model and the associated HIV-1 sequence data sets is that between the time the first cell is infected in the mucosa or submucosa, and peak viremia is reached approximately 3 to 5 weeks later (eclipse phase and Fiebig stages I and II), viruses diversify essentially randomly with little or no evidence

of biological selection or genetic bottlenecking. This is in contrast to the period immediately thereafter (Fiebig stages III-V) when evidence of strong CTL selection generally becomes evident (5, 26, 34). 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 data sets is that by analyzing *env* sequence diversity among 30 to 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 that between transmission and peak viremia (Fiebig stage II), it is uncommon to find sequences that exhibit shared polymorphisms resulting from immune selection; when this does occur (5, 26, 34), it is necessary to analyze sequential samples to distinguish transmitted/founder sequences from selected mutants. In the present 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).

Figure 2 illustrates ramp up and peak viremia plasma vRNA sequences in animal CP1W following a typical IR 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 distribution of nucleotide substitutions. At 7 days post-infection, 59 of 67 (88%) sequences were identical, and at 14 days post-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 day period is strikingly similar to findings made in humans acutely infected by HIV-1 and conforms closely to model projections previously reported (5). BEAST and Poisson estimations of days from a MRCA were consistent with the known duration of infection, further suggesting that a genetic bottleneck had not occurred between transmission and peak viremia (Table 2). This inference was proven by the analysis of sequence diversity in the SIVmac251 inoculum (Figure 2) where four *env* sequences were found that were *identical* to the consensus of *env* sequences sampled from animal CP1W at 7 and 14 days post-transmission. This result shows that the genome of a transmitted virus can be identified in an inoculum and then be tracked with precision across the rectal mucosa of the newly-infected host and through the period of massive virus expansion leading to peak plasma viremia (2, 3, 11, 35-40). In this animal, and in the 16 others that were studied at ramp-up and peak viremia, there was no evidence in *env* for selection of better fit viruses and no evidence of genetic bottlenecking, despite the fact that virus had undergone expansive amplification in submucosal, locoregional, gut-associated, and systemic lymphoreticular tissues. Recent studies in humans where identical and near-identical HIV-1 *env* sequences were identified in donor and recipient semen and plasma following sexual HIV-1 transmission corroborate these findings (21).

Having established that consensus *env* sequences of low diversity SIV lineages in acute infection correspond to transmitted/founder genomes, we could use this information to estimate the numbers of transmitted/founder viruses responsible for establishing productive infection in all 18 animals. The results are depicted in Figure 1 and summarized in Table 1 where the cumulative virus dose administered to each animal prior to the first positive plasma vRNA result

is also listed. The results show that rectal transmission led to productive infection by a minimum of 1-5 viruses (median = 1) whereas IV inoculation of 2000-fold less virus led to infection by at least 1 to 9 or more viruses. After accounting for differences in cumulative virus exposure, IV transmission efficiency was estimated to exceed rectal transmission efficiency by 2,000 - 20,000 fold. This finding is consistent with the demonstrably higher risk of HIV-1 transmission resulting from intravenous exposure to contaminated blood products than from sexual exposures across mucosa (4, 41).

One peculiarity in the findings illustrated in Figure 1 is the observation that while IR inoculation with the two higher viral doses resulted in greater numbers of transmitted viruses compared with the lowest dose, this relationship was not consistent or proportional across all three virus dose levels. For example, there were animals in the highest dose group (e.g., CP1W and CG87) that were infected by a single virus and other animals in the middle dose group (e.g., AK9F and CG7G) that were infected by 4 or 5 viruses. This observation in macaques has an interesting parallel in humans where two recent studies of mucosal HIV-1 infection revealed that transmission of multiple viruses did not follow an expected Poisson distribution of independent, low frequency events (5, 20). In those studies, the proportion of individuals who acquired more than one HIV-1 variant was higher than could be plausibly explained by known rates of HIV-1 sexual transmission. Moreover, the proportion of individuals who became infected by three or more viruses was disproportionately still higher. At present, there is no satisfactory biological explanation in humans or macaques for the observed non-Poisson distribution of infection events, although mucosal tears, ulceration or inflammation (23), or infection by multiply infected cells or virus aggregates, are possibilities.

Two features of SIV mucosal transmission that differ from HIV-1 mucosal transmission are a shorter eclipse period for SIV and a higher frequency of detection of minor transmitted variants of SIV (identified as single divergent sequences). Although the design of this study, which included weekly IR inoculations of virus and weekly plasma sampling, was not intended for a precise analysis 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 days. This is substantially less than the average eclipse phase of HIV-1 in humans of approximately 10-14 days and which may exceed 21 days [reviewed in Keele *et al.* (5)]. The macaque monkeys used in the present study weighed ~7 kgs (approximately 10-fold less than an average human), and accordingly, they have approximately 10-fold lower plasma and extracellular fluid volume in which virus can distribute. This difference, together with a higher exponential growth rate and reproductive ratio for SIV (19) compared with HIV-1 (42), could conceivably explain the differences observed in eclipse periods between macaques and humans.

A second difference we noted between SIV infected macaques and HIV-1 infected humans was a higher frequency of detection of multiple transmitted/founder SIVs whose progeny were represented infrequently in the plasma at ramp up or peak viremia timepoints. This was observed in 4 of 13 macaques compared with only 4 of 98 human subjects (5) [p=0.006, Fisher's exact test]. We suspect this difference may be due to the fact that the macaques were inoculated with relatively high titers of infectious virus that generally led to productive infection after only one or two IR inoculations. Such high rates of HIV-1 transmission in humans are uncommon (4, 41, 43-45). It is also possible that the shorter eclipse period and shorter time to peak viremia in macaques allows for fewer virus generations between transmission and sampling, thereby favoring the detection of viral progeny of

transmitted/founder viruses having less replicative fitness. Acutely infected humans are typically identified and studied three or more weeks after virus infection, and transmitted/founder viruses with lesser replication fitness could easily be lost by that time.

In summary, the present report demonstrates that IR 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 infrequent and dose-related; higher SIV doses, like higher HIV-1 titers in humans (43-45), are associated with higher transmission rates. The numbers of transmitted/founder viruses yielding productive clinical infection in humans and in macaques exposed to relatively small quantities of virus are relatively low [about 1 to 5 per individual (5, 20-26)], reflecting a substantial bottleneck in mucosal transmission. In neither humans nor macaques did the numbers of transmitted viruses following mucosal exposure correspond to a Poisson distribution of independent, low probability events (5, 20). Diversification of SIV and HIV-1 prior to peak viremia is essentially random and without evidence of immune selection; between peak viremia and 2 weeks hence, there is evidence of strong CTL selection for escape variants in both macaques and humans (Fig S2; see also refs. (5, 26, 32, 34, 38). In acute SIV and HIV-1 infection, APOBEC-mediated G-to-A hypermutation and inter-lineage recombination are evident (5, 46). An important difference between mucosal infection by SIV and HIV-1 is the duration of the eclipse period. Further studies are needed to define the eclipse phase in SIV infected macaques with greater precision and to estimate more accurately viral parameters of acute and early infection, including virus generation time, R_0, RT error rate and mutation frequency, all of which could be important considerations in the design and interpretation of preclinical vaccine studies. It will also be important to determine if a precise 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 the moment of transmission and extending to setpoint viremia, as shown recently for HIV-1 (34).

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 (30) were analyzed. The SIVmac251 and SIVsmE660 inoculum stocks and the inoculation procedures were described in detail (30). All animals were maintained in an Association for Assessment and Accreditation of Laboratory Animal Care-accredited institution in accordance with guidelines of the National Institutes of Health and Harvard Medical School.

Viral RNA extraction and cDNA synthesis. From each plasma specimen, approximately 20,000 viral RNA copies were extracted using the QIAamp Viral RNA Mini Kit (Qiagen, Valencia, CA). RNA was eluted and immediately subjected to cDNA synthesis. Reverse transcription of RNA to single stranded cDNA was performed using SuperScript III reverse transcriptase according to manufacturer's recommendations (Invitrogen Life Technologies, Carlsbad, CA). Briefly, each cDNA reaction included 1x RT buffer, 0.5 mM of each deoxynucleoside triphosphate, 5 mM dithiothreitol, 2 units/ml RNaseOUT (RNase inhibitor), 10 units/ml of SuperScript III reverse transcriptase, and 0.25 mM antisense primer SIVsm/macEnvR1 5'-TGTAATAAATCCCTTCCAGTCCCCCC-3 (nt 9454-9479 in Mac239).

The mixture was incubated at 50°C for 60 minutes followed by an increase in temperature to 55°C for an additional 60 minutes. The reaction was then heat-inactivated at 70°C for 15 minutes and treated with 2 units of RNase H at 37°C for 20 minutes. The newly synthesized cDNA was used immediately or frozen at –80°C.

Single genome amplification. cDNA was serially diluted and distributed among wells of replicate 96-well plates so as to identify a dilution where PCR positive wells constituted less than 30% of the total number of reactions, as previously described (5, 26). 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 carried out in the presence of 1x High Fidelity Platinum PCR buffer, 2 mM $MgSO_4$, 0.2 mM of each deoxynucleoside triphosphate, 0.2 μM of each primer, and 0.025 units/μl Platinum *Taq* High Fidelity polymerase in a 20-μl reaction (Invitrogen, Carlsbad, CA). First round PCR primers included sense primer SIVsm/macEnvF1 5′-CCTCCCCCTCCAGGACTAGC-3′ (nt 6127-6146 in SIVmac239) and antisense primer SIVsm/macEnvR1 5'-TGTAATAAATCCCTTCCAGTCCCCCC-3 (nt 9454-9479 in SIVmac239), which generated a ~3.3 kb amplicon. PCR was performed in MicroAmp 96-well reaction plates (Applied Biosystems, Foster City, CA) with the following PCR parameters: 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, 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 SIVsmEnvF2 5'-TATGATAGACATGGAGACACCCTTGAAGGAGC-3' (nt 6292-6323 in SIVmac239) or SIVmacEnvF2 5'-TATAATAGACATGGAGACACCCTTGAGGGAGC-3' (nt 6292-6323 in Mac239) and antisense primer SIVsmEnvR2 5'-ATGAGACATRTCTATTGCCAATTTGTA-3' (nt 9413-9439 in SIVmac239). The second round PCR reaction was carried out 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 Life Technologies, Carlsbad, CA). All PCR procedures were carried out under PCR clean room conditions using procedural safeguards against sample contamination, including pre-aliquoting 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; Foster City, CA). Sequencing reaction products were analyzed with an ABI 3730xl genetic analyzer (Applied Biosystems; Foster City, CA). Both DNA strands were sequenced using partially overlapping fragments. Individual sequence fragments for each amplicon were assembled and edited using the Sequencher program 4.7 (Gene Codes; Ann Arbor, MI). 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 single genome amplification from a single vRNA/cDNA template. This quality control measure enabled us to exclude from the analysis amplicons that resulted from PCR-generated in vitro recombination events or Taq

polymerase errors and to obtain multiple individual *env* sequences that proportionately represented those circulating *in vivo* in SIV virions.

Sequence alignments. All alignments were initially made with Clustal W (47). Consensus sequences were generated for the sequence set from each individual. The full alignment for SIVsmE660 and SIVmac251 is available as a supplemental data file 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 (Accession numbers XXX-YYY).

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 transmitted variant 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 neighbor-joining 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 Hamming Distances (HD, defined as the number of base positions at which the two genomes differ, excluding gaps) were determined for each transmitted/founder

lineage. The inoculum stocks of SIVsmE660 and SIVmac251 were evaluated for maximum *env* diversity and by neighbor-joining 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), as described elsewhere.

Viral recombination. Recombination analysis was performed by visual inspection of *Highlighter* trees and confirmed by RAP (www.hiv.lanl.gov), GARD (48) or Recco (49) recombination identification tools.

Model analyses. A detailed description of the mathematical model of early sequence evolution, star phylogeny determination, Bayesian and Poisson analyses, estimates of viral fitness, and power calculations for estimating the likelihood of detecting infrequent transmitted *env* variants are provided elsewhere (5). The Bayesian approach utilizes programs from the BEAST package that rely on a relaxed molecular clock (50) and a distribution of evolutionary rates based on the site replacement rate value used for the Poisson model (mean = 1.44×10^{-5}/site/day, standard deviation = 10^{-6}) determined from as many as 360,000,000 phylogenies that are sampled using a Monte Carlo-Markov Chain approach. Bayesian skyline models (51) were used that sample over these phylogenies and do not require prior specification of a demographic model or effective population size. Four independent chains of 30 to 360 million iterations with 10% burn in were analyzed with each chain showing results consistent with the others. Results from each chain were combined using LogCombiner and the combined set of trees was analyzed with TreeAnnotator. Times to MRCA were derived and visualized using Tracer version 1.4.1 (51). Because both Bayesian and Poisson analyses assume a constant mutation rate across positions,

and because we observed in this and other studies (5) an overall enhancement of base substitutions in APOBEC motifs that would be poorly accounted for in current implementations of evolutionary models, we performed model analyses on sequences that were stripped of positions that could potentially have an enhanced substitution rate due to APOBEC mutations . We then repeated the analyses on sequences that retained these sites.

Statistical analysis. We used the R software package, Version 2.8.0, to perform hypothesis tests [R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.]

ACKNOWLEDGMENTS

This work was supported by the Center for HIV/AIDS Vaccine Immunology and by grants from the NIH (AI67854, AI27767), the Bill & Melinda Gates Foundation (#37874), and The American Foundation for AIDS Research (106997-3). We thank D. McPherson and B. Cochran for technical assistance; sequencing core facilities of the UAB CFAR; and J. White for manuscript preparation.

Figure Legends

Figure 1. Viral kinetics following intrarectal or intravenous infection. Groups of three animals received IR inoculations of SIVsmE660 or SIVmac251 in dosages indicated at weeks 0, 1, 2, 3, 4 and 5 (underlined) or by the IV 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 (30).

Figure 2. Viral sequence analysis after intrarectal SIVmac251 infection of animal CP1W. (A) Neighbor-joining (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 to 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 by gray. (C) Poisson model of early viral diversity. nt, nucleotide. Brown labels indicate sequences with three or more APOBEC mediated G-to-A mutations compared with consensus.

Figure 3. Viral sequence diversity after IR or IV SIVmac251 infection. (A) NJ and *Highlighter* analyses of sequences from IR infected animal PBE at ramp-up (green symbols) or peak (blue symbols) viremia compared to inoculum sequences (black symbols). (B) IV

mutated sequence. Three sequences (RU.A1, RU.TB2 and RU.2B2) share a common polymorphism. Red asterisks indicate minor transmitted variants. Brown label indicates sequence with three or more APOBEC mediated G-to-A mutations compared with consensus.

Figure S5. Composite SIVsmE660 tree from IR and IV infected animals. NJ tree includes inoculum sequences (black symbols), which together with numerous unique IV transmitted/founder viruses, are distributed throughout the tree. Most IR or IV transmitted viruses are represented by discreet, low diversity lineages. Animal designation, route of infection, and numbers of transmitted viruses are indicated.

REFERENCES

1. Chun, T.W., D. Engel, M.M. Berrey, T. Shea, L. Corey, and A.S. Fauci. 1998. Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc Natl Acad Sci U S A* 95:8869-8873.
2. Haase, A.T. 2005. Perils at mucosal front lines for HIV and SIV and their hosts. *Nat Rev Immunol* 5:783-792.
3. Pope, M., and A.T. Haase. 2003. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nat Med* 9:847-852.
4. Shattock, R.J., and J.P. Moore. 2003. Inhibiting sexual transmission of HIV-1 infection. *Nat Rev Microbiol* 1:25-34.
5. Keele, B.F., E.E. Giorgi, J.F. Salazar-Gonzalez, J.M. Decker, K.T. Pham, M.G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J.L. Kirchherr, F. Gao, J.A. Anderson, L.H. Ping, R. Swanstrom, G.D. Tomaras, W.A. Blattner, P.A. Goepfert, J.M. Kilby, M.S. Saag, E.L. Delwart, M.P. Busch, M.S. Cohen, D.C. Montefiori, B.F. Haynes, B. Gaschen, G.S. Athreya, H.Y. Lee, N. Wood, C. Seoighe, A.S. Perelson, T. Bhattacharya, B.T. Korber, B.H. Hahn, and G.M. Shaw. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* 105:7552-7557.
6. Collins, K.B., B.K. Patterson, G.J. Naus, D.V. Landers, and P. Gupta. 2000. Development of an in vitro organ culture model to study transmission of HIV-1 in the female genital tract. *Nat Med* 6:475-479.
7. Gordon, S.N., R.M. Dunham, J.C. Engram, J. Estes, Z. Wang, N.R. Klatt, M. Paiardini, I.V. Pandrea, C. Apetrei, D.L. Sodora, H.Y. Lee, A.T. Haase, M.D. Miller, A. Kaur, S.I.

Staprans, A.S. Perelson, M.B. Feinberg, and G. Silvestri. 2008. Short-lived infected cells support virus replication in sooty mangabeys naturally infected with simian immunodeficiency virus: implications for AIDS pathogenesis. *J Virol* 82:3725-3735.

8. Hladik, F., P. Sakchalathorn, L. Ballweber, G. Lentz, M. Fialkow, D. Eschenbach, and M.J. McElrath. 2007. Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1. *Immunity* 26:257-270.
9. Hu, Q., I. Frank, V. Williams, J.J. Santos, P. Watts, G.E. Griffin, J.P. Moore, M. Pope, and R.J. Shattock. 2004. Blockade of attachment and fusion receptors inhibits HIV-1 infection of human cervical tissue. *J Exp Med* 199:1065-1075.
10. Letvin, N.L., J.R. Mascola, Y. Sun, D.A. Gorgone, A.P. Buzby, L. Xu, Z.Y. Yang, B. Chakrabarti, S.S. Rao, J.E. Schmitz, D.C. Montefiori, B.R. Barker, F.L. Bookstein, and G.J. Nabel. 2006. Preserved CD4+ central memory T cells and survival in vaccinated SIV-challenged monkeys. *Science* 312:1530-1533.
11. Miller, C.J., Q. Li, K. Abel, E.Y. Kim, Z.M. Ma, S. Wietgrefe, L. La Franco-Scheuch, L. Compton, L. Duan, M.D. Shore, M. Zupancic, M. Busch, J. Carlis, S. Wolinsky, and A.T. Haase. 2005. Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. *J Virol* 79:9217-9227.
12. Veazey, R.S., R.J. Shattock, M. Pope, J.C. Kirijan, J. Jones, Q. Hu, T. Ketas, P.A. Marx, P.J. Klasse, D.R. Burton, and J.P. Moore. 2003. Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120. *Nat Med* 9:343-346.
13. McDermott, A.B., J. Mitchen, S. Piaskowski, I. De Souza, L.J. Yant, J. Stephany, J. Furlott, and D.I. Watkins. 2004. Repeated low-dose mucosal simian immunodeficiency virus SIVmac239 challenge results in the same viral and immunological kinetics as high-dose challenge: a model for the evaluation of vaccine efficacy in nonhuman primates. *J Virol* 78:3140-3144.
14. Otten, R.A., D.R. Adams, C.N. Kim, E. Jackson, J.K. Pullium, K. Lee, L.A. Grohskopf, M. Monsour, S. Butera, and T.M. Folks. 2005. Multiple vaginal exposures to low doses of R5 simian-human immunodeficiency virus: strategy to study HIV preclinical interventions in nonhuman primates. *J Infect Dis* 191:164-173.
15. Regoes, R.R., I.M. Longini, M.B. Feinberg, and S.I. Staprans. 2005. Preclinical assessment of HIV vaccines and microbicides by repeated low-dose virus challenges. *PLoS Med* 2:e249.
16. Subbarao, S., R.A. Otten, A. Ramos, C. Kim, E. Jackson, M. Monsour, D.R. Adams, S. Bashirian, J. Johnson, V. Soriano, A. Rendon, M.G. Hudgens, S. Butera, R. Janssen, L. Paxton, A.E. Greenberg, and T.M. Folks. 2006. Chemoprophylaxis with tenofovir disoproxil fumarate provided partial protection against infection with simian human immunodeficiency virus in macaques given multiple virus challenges. *J Infect Dis* 194:904-911.
17. Fiebig, E.W., D.J. Wright, B.D. Rawal, P.E. Garrett, R.T. Schumacher, L. Peddada, C. Heldebrant, R. Smith, A. Conrad, S.H. Kleinman, and M.P. Busch. 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS* 17:1871-1879.
18. Greenier, J.L., C.J. Miller, D. Lu, P.J. Dailey, F.X. Lu, K.J. Kunstman, S.M. Wolinsky, and M.L. Marthas. 2001. Route of simian immunodeficiency virus inoculation determines

mutated sequence. Three sequences (RU.A1, RU.TB2 and RU.2B2) share a common polymorphism. Red asterisks indicate minor transmitted variants. Brown label indicates sequence with three or more APOBEC mediated G-to-A mutations compared with consensus.

Figure S5. Composite SIVsmE660 tree from IR and IV infected animals. NJ tree includes inoculum sequences (black symbols), which together with numerous unique IV transmitted/founder viruses, are distributed throughout the tree. Most IR or IV transmitted viruses are represented by discreet, low diversity lineages. Animal designation, route of infection, and numbers of transmitted viruses are indicated.

REFERENCES

1. Chun, T.W., D. Engel, M.M. Berrey, T. Shea, L. Corey, and A.S. Fauci. 1998. Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc Natl Acad Sci U S A* 95:8869-8873.
2. Haase, A.T. 2005. Perils at mucosal front lines for HIV and SIV and their hosts. *Nat Rev Immunol* 5:783-792.
3. Pope, M., and A.T. Haase. 2003. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nat Med* 9:847-852.
4. Shattock, R.J., and J.P. Moore. 2003. Inhibiting sexual transmission of HIV-1 infection. *Nat Rev Microbiol* 1:25-34.
5. Keele, B.F., E.E. Giorgi, J.F. Salazar-Gonzalez, J.M. Decker, K.T. Pham, M.G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J.L. Kirchherr, F. Gao, J.A. Anderson, L.H. Ping, R. Swanstrom, G.D. Tomaras, W.A. Blattner, P.A. Goepfert, J.M. Kilby, M.S. Saag, E.L. Delwart, M.P. Busch, M.S. Cohen, D.C. Montefiori, B.F. Haynes, B. Gaschen, G.S. Athreya, H.Y. Lee, N. Wood, C. Seoighe, A.S. Perelson, T. Bhattacharya, B.T. Korber, B.H. Hahn, and G.M. Shaw. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* 105:7552-7557.
6. Collins, K.B., B.K. Patterson, G.J. Naus, D.V. Landers, and P. Gupta. 2000. Development of an in vitro organ culture model to study transmission of HIV-1 in the female genital tract. *Nat Med* 6:475-479.
7. Gordon, S.N., R.M. Dunham, J.C. Engram, J. Estes, Z. Wang, N.R. Klatt, M. Paiardini, I.V. Pandrea, C. Apetrei, D.L. Sodora, H.Y. Lee, A.T. Haase, M.D. Miller, A. Kaur, S.I.

Staprans, A.S. Perelson, M.B. Feinberg, and G. Silvestri. 2008. Short-lived infected cells support virus replication in sooty mangabeys naturally infected with simian immunodeficiency virus: implications for AIDS pathogenesis. *J Virol* 82:3725-3735.

8. Hladik, F., P. Sakchalathorn, L. Ballweber, G. Lentz, M. Fialkow, D. Eschenbach, and M.J. McElrath. 2007. Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1. *Immunity* 26:257-270.
9. Hu, Q., I. Frank, V. Williams, J.J. Santos, P. Watts, G.E. Griffin, J.P. Moore, M. Pope, and R.J. Shattock. 2004. Blockade of attachment and fusion receptors inhibits HIV-1 infection of human cervical tissue. *J Exp Med* 199:1065-1075.
10. Letvin, N.L., J.R. Mascola, Y. Sun, D.A. Gorgone, A.P. Buzby, L. Xu, Z.Y. Yang, B. Chakrabarti, S.S. Rao, J.E. Schmitz, D.C. Montefiori, B.R. Barker, F.L. Bookstein, and G.J. Nabel. 2006. Preserved CD4+ central memory T cells and survival in vaccinated SIV-challenged monkeys. *Science* 312:1530-1533.
11. Miller, C.J., Q. Li, K. Abel, E.Y. Kim, Z.M. Ma, S. Wietgrefe, L. La Franco-Scheuch, L. Compton, L. Duan, M.D. Shore, M. Zupancic, M. Busch, J. Carlis, S. Wolinsky, and A.T. Haase. 2005. Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. *J Virol* 79:9217-9227.
12. Veazey, R.S., R.J. Shattock, M. Pope, J.C. Kirijan, J. Jones, Q. Hu, T. Ketas, P.A. Marx, P.J. Klasse, D.R. Burton, and J.P. Moore. 2003. Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120. *Nat Med* 9:343-346.
13. McDermott, A.B., J. Mitchen, S. Piaskowski, I. De Souza, L.J. Yant, J. Stephany, J. Furlott, and D.I. Watkins. 2004. Repeated low-dose mucosal simian immunodeficiency virus SIVmac239 challenge results in the same viral and immunological kinetics as high-dose challenge: a model for the evaluation of vaccine efficacy in nonhuman primates. *J Virol* 78:3140-3144.
14. Otten, R.A., D.R. Adams, C.N. Kim, E. Jackson, J.K. Pullium, K. Lee, L.A. Grohskopf, M. Monsour, S. Butera, and T.M. Folks. 2005. Multiple vaginal exposures to low doses of R5 simian-human immunodeficiency virus: strategy to study HIV preclinical interventions in nonhuman primates. *J Infect Dis* 191:164-173.
15. Regoes, R.R., I.M. Longini, M.B. Feinberg, and S.I. Staprans. 2005. Preclinical assessment of HIV vaccines and microbicides by repeated low-dose virus challenges. *PLoS Med* 2:e249.
16. Subbarao, S., R.A. Otten, A. Ramos, C. Kim, E. Jackson, M. Monsour, D.R. Adams, S. Bashirian, J. Johnson, V. Soriano, A. Rendon, M.G. Hudgens, S. Butera, R. Janssen, L. Paxton, A.E. Greenberg, and T.M. Folks. 2006. Chemoprophylaxis with tenofovir disoproxil fumarate provided partial protection against infection with simian human immunodeficiency virus in macaques given multiple virus challenges. *J Infect Dis* 194:904-911.
17. Fiebig, E.W., D.J. Wright, B.D. Rawal, P.E. Garrett, R.T. Schumacher, L. Peddada, C. Heldebrant, R. Smith, A. Conrad, S.H. Kleinman, and M.P. Busch. 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS* 17:1871-1879.
18. Greenier, J.L., C.J. Miller, D. Lu, P.J. Dailey, F.X. Lu, K.J. Kunstman, S.M. Wolinsky, and M.L. Marthas. 2001. Route of simian immunodeficiency virus inoculation determines

the complexity but not the identity of viral variant populations that infect rhesus macaques. *J Virol* 75:3753-3765.

19. Nowak, M.A., A.L. Lloyd, G.M. Vasquez, T.A. Wiltrout, L.M. Wahl, N. Bischofberger, J. Williams, A. Kinter, A.S. Fauci, V.M. Hirsch, and J.D. Lifson. 1997. Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. *J Virol* 71:7518-7525.
20. Abrahams, M.R., J.A. Anderson, C. Seoighe, E.E. Giorgi, L.-H. Ping, G.S. Athreya, F. Treurnichat, N. Wood, K. Mlisana, J. Passmore, L. Roberts, T. Bhattacharya, I. Hoffman, s. Galvin, C. Mapanje, P. Kazembre, R. Thebus, S. Fiscus, W. Hide, S. Cohen, A. Karium, B.F. Haynes, G.M. Shaw, B.T. Korber, B.H. Hahn, R. Swanstrom, and C. Williamson. 2009. Quantitating the multiplicity of infection with HIV-1subtype C. *J.Virol*. In Press.
21. Anderson, J., B.F. Keele, J.F. Salazar-Gonzalez, L.-H. Ping, L. Arney, C. Williamson, G.M. Shaw, B.H. Hahn, M.S. Cohen, R. Swanstrom, and CHAVI. 2009. Determining the genetic linkage of HIV-1 subytpe B and C transmission pairs: analyses of viral env sequences from donor and recipient. *16th Conference on Retroviruses and Opportunistic Infections* XXX:YYY.
22. Baalwa, J., S. Wang, B.F. Keele, M.G. Salazar, J.M. Decker, P. Kaleebu, J. Gilmour, E. Cormier, B.H. Hahn, and G.M. Shaw. 2009. Ugandan HIV-1 Subtype A, D and AD Recombinant Transmitted/Founder Viruses: Molecular Pathways of Evolution and Immune Escape. *16th Conference on Retroviruses and Opportunistic Infections.* XXX:YYY.
23. Haaland, R.E., P.A. Hawkins, J.F. Salazar-Gonzalez, A. Johnson, A. Tichacek, E. Karita, O. Manigart, J. Mulenga, B.F. Keele, G.M. Shaw, B.H. Hahn, S. Allen, C.A. Derdeyn, and E. Hunter. 2009. Severe genetic bottleneck in heterosexual transmission of subtype A an C HIV-1 and its mitigation by inflammatory genital infections. *PLoS Pathogen* In press.
24. Kearney, M., F. Maldarelli, W. Shao, J.B. Margolick, E.S. Daar, J.W. Mellors, V. Rao, J.M. Coffin, and S. Palmer. 2009. HIV-1 Population Genetics and Adaptation in Newly Infected Individuals. *J Virol.* In revision.
25. Li, H., B.F. Keele, E.E. Giorgi, S. Wang, G.H. Learn, J.F. Salazar-Gonzalez, B. Korber, M. Markowitz, B.H. Hahn, and G.M. Shaw. 2009. Investigation of HIV-1 Transmission and Envelope Diversity among Men Who Have Sex with Men (MSM) by Single Genome Amplification. *16th Conference on Retroviruses and Opportunistic Infections* XXX:YYY.
26. Salazar-Gonzalez, J.F., E. Bailes, K.T. Pham, M.G. Salazar, M.B. Guffey, B.F. Keele, C.A. Derdeyn, P. Farmer, E. Hunter, S. Allen, O. Manigart, J. Mulenga, J.A. Anderson, R. Swanstrom, B.F. Haynes, G.S. Athreya, B.T. Korber, P.M. Sharp, G.M. Shaw, and B.H. Hahn. 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol* 82:3952-3970.
27. Liu, S.L., A.G. Rodrigo, R. Shankarappa, G.H. Learn, L. Hsu, O. Davidov, L.P. Zhao, and J.I. Mullins. 1996. HIV quasispecies and resampling. *Science* 273:415-416.
28. Palmer, S., M. Kearney, F. Maldarelli, E.K. Halvas, C.J. Bixby, H. Bazmi, D. Rock, J. Falloon, R.T. Davey, Jr., R.L. Dewar, J.A. Metcalf, S. Hammer, J.W. Mellors, and J.M. Coffin. 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance

mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol* 43:406-413.

29. Shriner, D., A.G. Rodrigo, D.C. Nickle, and J.I. Mullins. 2004. Pervasive genomic recombination of HIV-1 in vivo. *Genetics* 167:1573-1583.
30. Letvin, N.L., S.S. Rao, V. Dang, A.P. Buzby, B. Korioth-Schmitz, D. Dombagoda, J.G. Parvani, R.H. Clarke, L. Bar, K.R. Carlson, P.A. Kozlowski, V.M. Hirsch, J.R. Mascola, and G.J. Nabel. 2007. No evidence for consistent virus-specific immunity in simian immunodeficiency virus-exposed, uninfected rhesus monkeys. *J Virol* 81:12368-12374.
31. Shankarappa, R., J.B. Margolick, S.J. Gange, A.G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C.R. Rinaldo, G.H. Learn, X. He, X.L. Huang, and J.I. Mullins. 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 73:10489-10502.
32. O'Connor, D.H., T.M. Allen, T.U. Vogel, P. Jing, I.P. DeSouza, E. Dodds, E.J. Dunphy, C. Melsaether, B. Mothe, H. Yamamoto, H. Horton, N. Wilson, A.L. Hughes, and D.I. Watkins. 2002. Acute phase cytotoxic T lymphocyte escape is a hallmark of simian immunodeficiency virus infection. *Nat Med* 8:493-499.
33. Lee, H.Y., E.E. Giorgi, B.F. Keele, B. Gaschen, G.S. Athreya, J.F. Salazar-Gonzalez, K.T. Pham, P.A. Goepfert, M. Kilby, M.S. Saag, E. Delwart, M. Busch, B.H. Hahn, G.M. Shaw, B.T. Korber, T. Bhattacharya, and A. Perelson. 2009. Modeling sequence evolution in acute HIV-1 infection. *PLoS Computational Biol.* In revision.
34. Salazar-Gonzalez, J.F., M.G. Salazar, B.F. Keele, G.H. Learn, E. Hunter, M. Markowitz, B.F. Haynes, B. Korber, B.H. Hahn, and G.M. Shaw. 2009. Complete HIV-1 sequences in acute and early infection reveal the genetic identity, biological phenotype, and precise evolutionary pathways of transmitted/founder viruses and their progeny. *16th Conference on Retroviruses and Opportunistic Infections* XXX:YYY.
35. Brenchley, J.M., T.W. Schacker, L.E. Ruff, D.A. Price, J.H. Taylor, G.J. Beilman, P.L. Nguyen, A. Khoruts, M. Larson, A.T. Haase, and D.C. Douek. 2004. CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J Exp Med* 200:749-759.
36. Li, Q., L. Duan, J.D. Estes, Z.M. Ma, T. Rourke, Y. Wang, C. Reilly, J. Carlis, C.J. Miller, and A.T. Haase. 2005. Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells. *Nature* 434:1148-1152.
37. Mehandru, S., M.A. Poles, K. Tenner-Racz, A. Horowitz, A. Hurley, C. Hogan, D. Boden, P. Racz, and M. Markowitz. 2004. Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. *J Exp Med* 200:761-770.
38. Reynolds, M.R., E. Rakasz, P.J. Skinner, C. White, K. Abel, Z.M. Ma, L. Compton, G. Napoe, N. Wilson, C.J. Miller, A. Haase, and D.I. Watkins. 2005. CD8+ T-lymphocyte response to major immunodominant epitopes after vaginal exposure to simian immunodeficiency virus: too late and too little. *J Virol* 79:9228-9235.
39. Veazey, R.S., M. DeMaria, L.V. Chalifoux, D.E. Shvetz, D.R. Pauley, H.L. Knight, M. Rosenzweig, R.P. Johnson, R.C. Desrosiers, and A.A. Lackner. 1998. Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. *Science* 280:427-431.
40. Veazey, R.S., and A.A. Lackner. 2004. Getting to the guts of HIV pathogenesis. *J Exp Med* 200:697-700.

41. Powers, K.A., C. Poole, A.E. Pettifor, and M.S. Cohen. 2008. Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. *Lancet Infect Dis* 8:553-563.
42. Stafford, M.A., L. Corey, Y. Cao, E.S. Daar, D.D. Ho, and A.S. Perelson. 2000. Modeling plasma virus concentration during primary HIV infection. *J Theor Biol* 203:285-301.
43. Pilcher, C.D., G. Joaki, I.F. Hoffman, F.E. Martinson, C. Mapanje, P.W. Stewart, K.A. Powers, S. Galvin, D. Chilongozi, S. Gama, M.A. Price, S.A. Fiscus, and M.S. Cohen. 2007. Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection. *AIDS* 21:1723-1730.
44. Quinn, T.C., M.J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M.O. Meehan, T. Lutalo, and R.H. Gray. 2000. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med* 342:921-929.
45. Wawer, M.J., R.H. Gray, N.K. Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo, F. Nalugoda, F. Wabwire-Mangen, M.P. Meehan, and T.C. Quinn. 2005. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* 191:1403-1409.
46. Simon, V., V. Zennou, D. Murray, Y. Huang, D.D. Ho, and P.D. Bieniasz. 2005. Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* 1:e6.
47. Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
48. Kosakovsky Pond, S.L., D. Posada, M.B. Gravenor, C.H. Woelk, and S.D. Frost. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23:1891-1901.
49. Maydt, J., and T. Lengauer. 2006. Recco: recombination analysis using cost optimization. *Bioinformatics* 22:1064-1071.
50. Drummond, A.J., S.Y. Ho, M.J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88.
51. Drummond, A.J., A. Rambaut, B. Shapiro, and O.G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185-1192.

Table I. Characterization of SIV transmission by intrarectal or intravenous inoculation

Subject	Initial dose	Cumulative dose prior to first positive vRNA	Time to first + vRNA (wks)	Infection route	Peak viral load	No. of *env* seq. per animal	No. of identified transmitted variants	Percent max. *env* diversity	APOBEC mediated G-to-A mutations	No. of APOBEC mutated sequences	Intra-host recom-bination	Star-like phylogeny[a]	Poisson distribution[a,b]	Rectal barrier[c]
SIVsmE660 infected animals														
CG7V	$6x10^7$	$1.2x10^8$	2	IR	77,871,700	96	5	1.06	Yes	2	Yes	Yes	Yes	
CP3C	$6x10^7$	$1.2x10^8$	2	IR	1,112,900	49	3	0.84	Yes	1	No	Yes	Yes	
CG87	$6x10^7$	$3.6x10^8$	6	IR	479,543	56	1	0.15	No	–	N/A	Yes	Yes	
CG7G	$6x10^6$	$6.0x10^6$	1	IR	7,084,600	43	5	1.21	Yes	1	Yes	Yes	Yes	
AK9F	$6x10^6$	$3.6x10^7$	6	IR	140,795	35	4	1.43	No	–	No	Yes	Yes	
CP37	$6x10^6$	$4.3x10^8$	18	IV	94,394,600	62	>9	1.54	N/A[d]	N/A[d]	N/A[d]	N/A[d]	N/A[d]	>19,350
CP23[e]	$6x10^5$	$1.2x10^6$	2	IR	74,000,000	37	1	0.33[e]	Yes	1	N/A	No[e]	No[e]	
CR54	$6x10^5$	$3.7x10^8$	18	IV	1,968,800	29	4	1.26	Yes	1	Yes	Yes	Yes	7,400
SIVmac251 infected animals														
CP1W	$6x10^7$	$6.0x10^7$	1	IR	17,797,900	137	1	0.15	Yes	5	N/A	Yes	Yes	
CT76	$6x10^7$	$6.0x10^7$	1	IR	18,233,200	50	4	0.66	Yes	1	No	Yes	Yes	
PBE	$6x10^7$	$3.0x10^8$	5	IR	10,955,200	41	1	0.15	No	–	N/A	Yes	Yes	
CR2A	$6x10^6$	$1.2x10^7$	2	IR	4,130,600	46	1	0.07	No	–	N/A	Yes	Yes	
CR53	$6x10^6$	$3.6x10^7$	6	IR	6,341,200	30	1	0.15	Yes	1	N/A	Yes	Yes	
AV74	$6x10^6$	$4.3x10^8$	18	IV	18,690,000	71	1	0.18	Yes	3	N/A	Yes	Yes	2,150
AH4X	$6x10^5$	$2.5x10^8$	16	IR	2,611,600	23	2	0.59	No	–	No	Yes	Yes	
CG71	$6x10^5$	$3.7x10^8$	18	IV	8,857,400	67	>8	0.73	Yes	3	Yes	Yes	Yes	>14,800
CG5G	$6x10^5$	$3.7x10^8$	18	IV	12,938,400	71	4	0.81	Yes	2	Yes	Yes	Yes	7,400
AV66	$6x10^5$	$6.7x10^7$	13	IR	4,858,300	44	1	0.07	Yes	1	N/A	Yes	Yes	

[a] If infected with more than one variant, the major lineages were examined for a star-like phylogeny and Poisson distribution of substitutions excluding APOBEC mediated changes.

[b] APOBEC mutated sequences excluded.

[c] Rectal barrier was calculated by dividing the cumulative IR dose by the IV dose multiplied by the number of transmitted viruses.

[d] Not applicable; too many variants to quantify.

[e] Sampled 2 weeks post peak VL.

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

Subject	Sample time	Infection route	No. of sequences examined	Max. Hamming Distance (HD)	BEAST estimated days since MRCA (95% CI)	Poisson estimated days since MRCA (95% CI)	Lambda[a]	SD[b]	Goodness of fit P value[c]
SIVsmE660 infected animals									
CG7V	Ramp-up	IR	39	2		6 (3,8)	0.308	0.082	0.584
	Peak	IR	23	4	12 (8,19)	13 (7,20)	0.762	0.190	0.840
CP3C	Ramp-up	IR	19	3		13 (6,20)	1.053	0.189	0.913
	Peak	IR	10	3	17 (10,29)	11 (3,18)	1.000	0.220	0.815
CG87	Ramp-up	IR	23	4		15 (8,22)	0.854	0.210	0.663
	Peak	IR	33	4	20 (12,34)	23 (18,28)	1.311	0.156	0.220
CG7G	Ramp-up	IR	18	2		12 (6,17)	0.667	0.159	0.317
	Peak	IR	17	3	16 (10,27)	13 (6,20)	0.750	0.209	0.865
AK9F	Peak	IR	19	4	12 (2,43)	14 (7,22)	0.807	0.216	0.886
CP37[d]	–	–	–	–	–	–	–	–	–
CP23	2wk post	IR	37	8	50 (28,84)	55 (47, 62)	3.129	0.221	0.038
CR54	Peak	IV	17	3	14 (2,52)	12 (5,20)	0.706	0.206	0.874
SIVmac251 infected animals									
CP1W	Ramp-up	IR	69	4		6 (1,10)	0.316	0.123	0.654
	Peak	IR	68	4	12 (8,20)	10 (5,15)	0.552	0.143	0.906
CT76	Ramp-up	IR	20	3		17 (10, 24)	0.968	0.208	0.809
	Peak	IR	22	3	16 (7,34)	7 (1,12)	0.455	0.153	0.701
PBE	Ramp-up	IR	17	2		6 (2,11)	0.353	0.139	0.646
	Peak	IR	24	4	13 (8,21)	12 (7,17)	0.667	0.142	0.211
CR2A	Ramp-up	IR	19	3		2 (0,5)	0.105	0.077	0.750
	Peak	IR	27	2	8 (7,14)	5 (2,9)	0.296	0.101	0.637
CR53	Peak	IR	30	4	12 (3,57)	12 (6,18)	0.667	0.166	0.478
AV74	Ramp-up	IV	35	3		5 (2,9)	0.284	0.103	0.703
	Peak	IV	36	3	12 (8,20)	14 (8,20)	0.778	0.174	0.621
AH4X	Peak	IR	21	6	11 (2,80)	10 (4,16)	0.561	0.173	0.894
CG71	Ramp-up	IV	22	2		7 (2,11)	0.364	0.124	0.617
	Peak	IV	16	5	14 (8,25)	17 (8,27)	0.983	0.279	0.630
CG5G	Ramp-up	IV	24	4		6 (1,11)	0.333	0.137	0.655
	Peak	IV	17	4	9 (7,14)	6 (2,11)	0.353	0.136	0.646
AV66	Ramp-up	IR	20	2		5 (1,9)	0.300	0.115	0.657
	Peak	IR	24	2	9 (7,28)	3 (0,6)	0.167	0.082	0.713

[a] Mean of best fitting Poisson distribution defined by the free parameter lambda.

[b] Standard deviation 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.

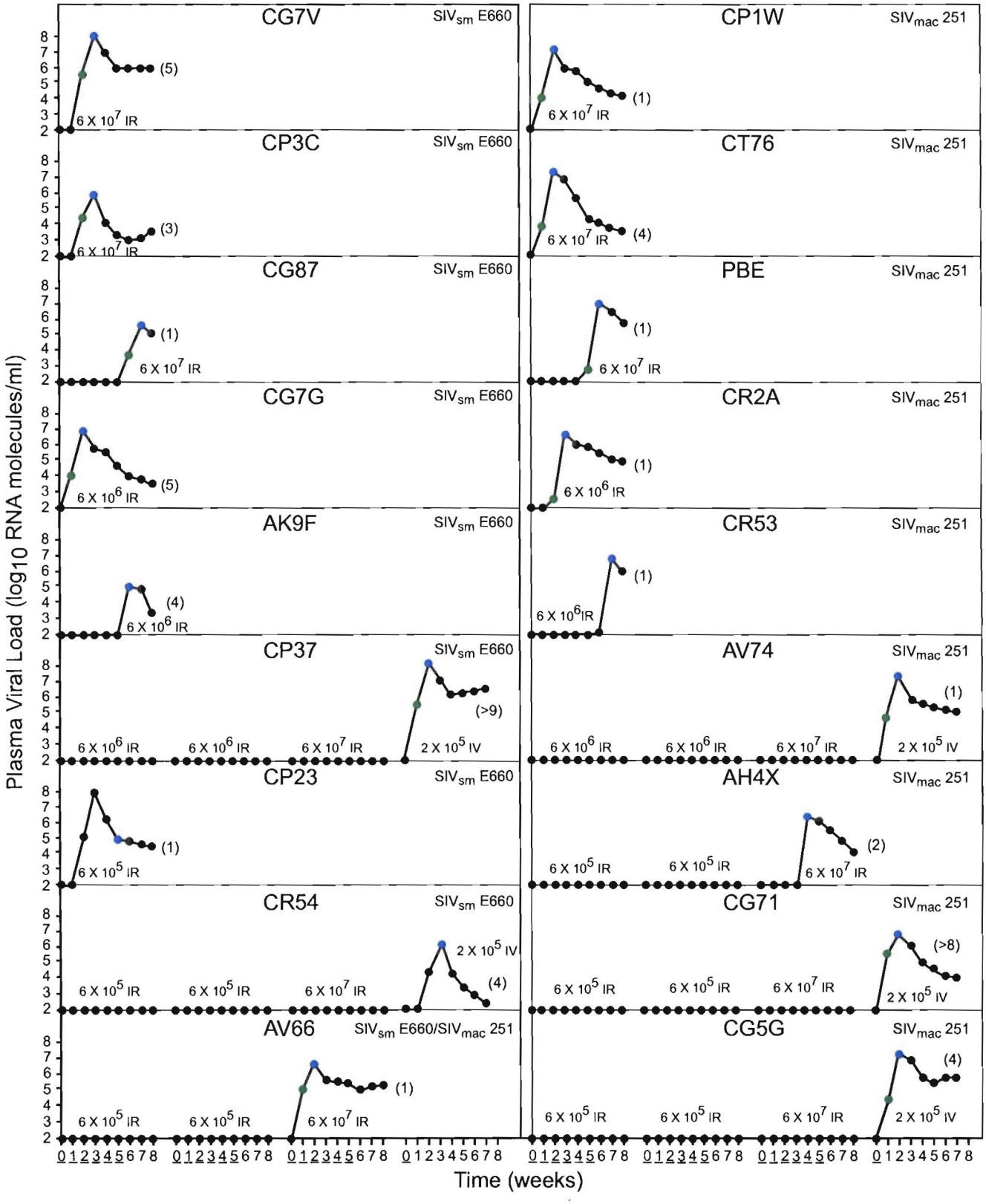
CG7V
SIVsm E660
(5)
6 X 10^7 IR
CP1W
SIVmac 251
(1)
6 X 10^7 IR
CP3C
SIVsm E660
(3)
6 X 10^7 IR
CT76
SIVmac 251
(4)
6 X 10^7 IR
CG87
SIVsm E660
(1)
6 X 10^7 IR
PBE
SIVmac 251
(1)
6 X 10^7 IR
CG7G
SIVsm E660
(5)
6 X 10^6 IR
CR2A
SIVmac 251
(1)
6 X 10^6 IR
AK9F
SIVsm E660
(4)
6 X 10^6 IR
CR53
SIVmac 251
(1)
6 X 10^6IR
CP37
SIVsm E660
(>9)
6 X 10^6 IR
6 X 10^6 IR
6 X 10^7 IR
2 X 10^5 IV
AV74
SIVmac 251
(1)
6 X 10^6 IR
6 X 10^6 IR
6 X 10^7 IR
2 X 10^5 IV
CP23
SIVsm E660
(1)
6 X 10^5 IR
AH4X
SIVmac 251
(2)
6 X 10^5 IR
6 X 10^5 IR
6 X 10^7 IR
CR54
SIVsm E660
2 X 10^5 IV
(4)
6 X 10^5 IR
6 X 10^5 IR
6 X 10^7 IR
CG71
SIVmac 251
(>8)
2 X 10^5 IV
6 X 10^5 IR
6 X 10^5 IR
6 X 10^7 IR
AV66
SIVsm E660/SIVmac 251
(1)
6 X 10^5 IR
6 X 10^5 IR
6 X 10^7 IR
CG5G
SIVmac 251
(4)
2 X 10^5 IV
6 X 10^5 IR
6 X 10^5 IR
6 X 10^7 IR
Plasma Viral Load (log10 RNA molecules/ml)
Time (weeks)

Figure 1

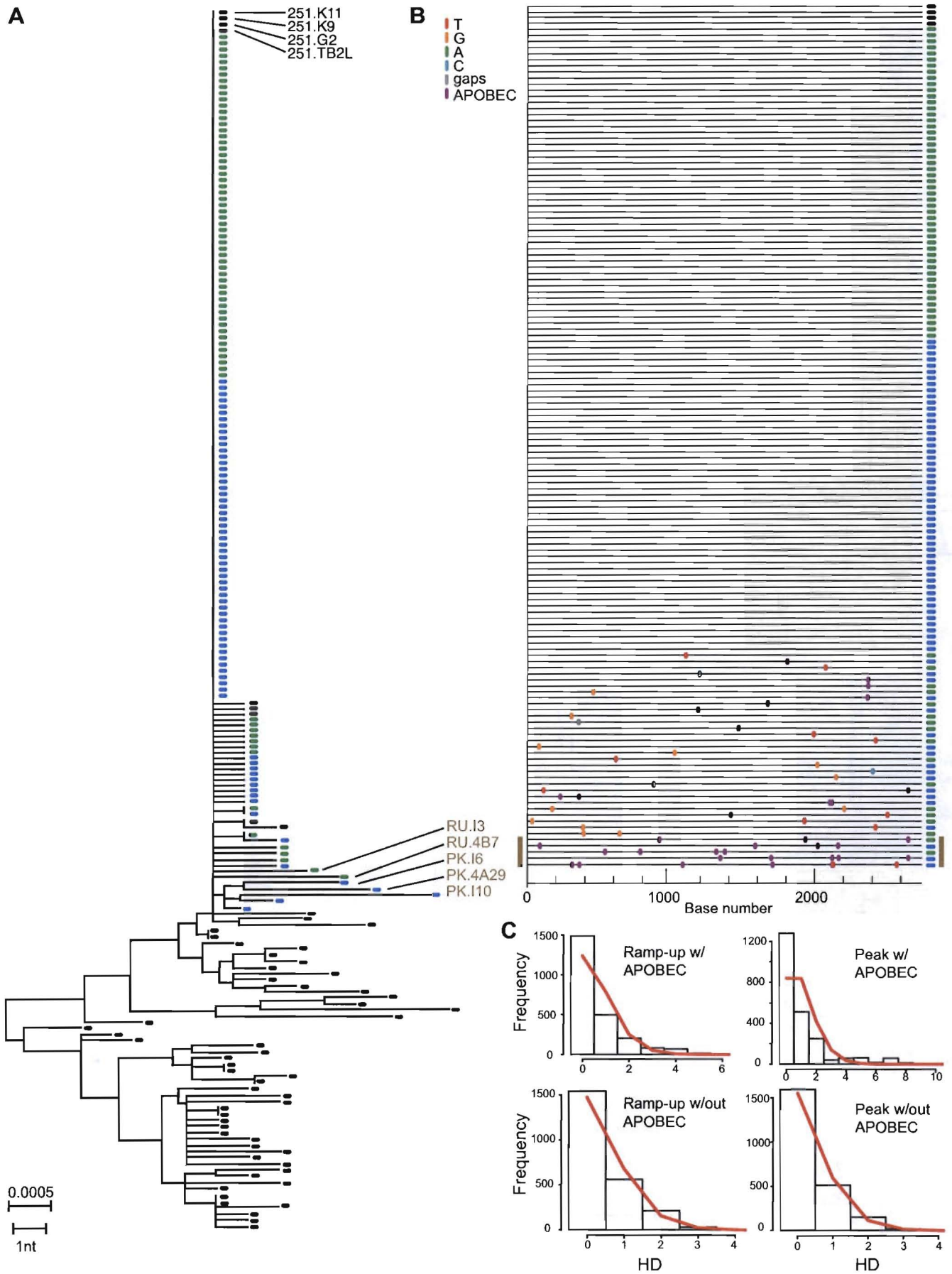
A
251.K11
251.K9
251.G2
251.TB2L
RU.I3
RU.4B7
PK.I6
PK.4A29
PK.I10
0.0005
1nt
B
T
G
A
C
gaps
APOBEC
0
1000
2000
Base number
C
Frequency
Ramp-up w/ APOBEC
Peak w/ APOBEC
Ramp-up w/out APOBEC
Peak w/out APOBEC
HD
HD

Figure 2

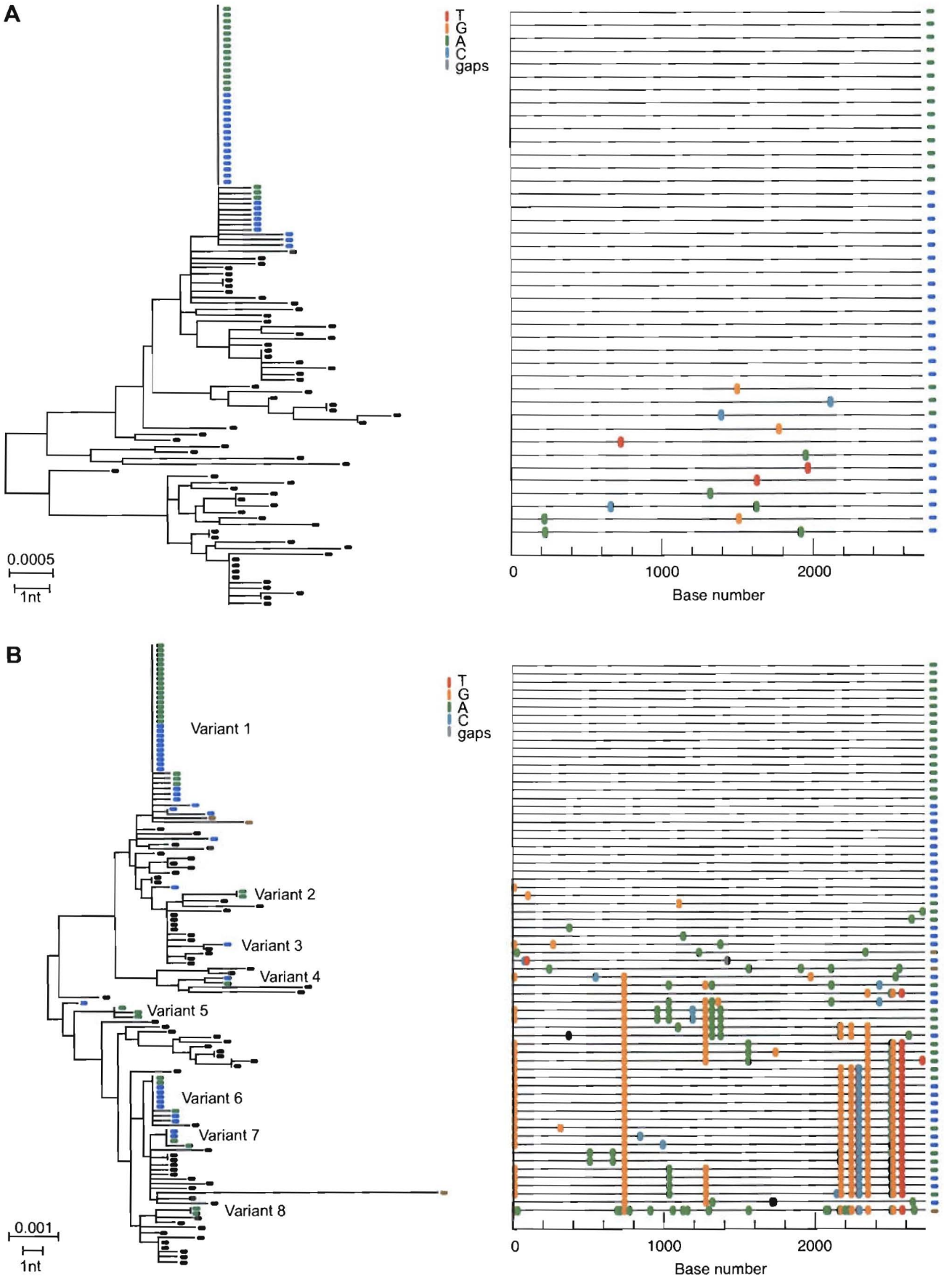
A
T
G
A
C
gaps
0.0005
1nt
0
1000
2000
Base number
B
Variant 1
Variant 2
Variant 3
Variant 4
Variant 5
Variant 6
Variant 7
Variant 8
0.001
1nt
T
G
A
C
gaps
0
1000
2000
Base number

Figure 3

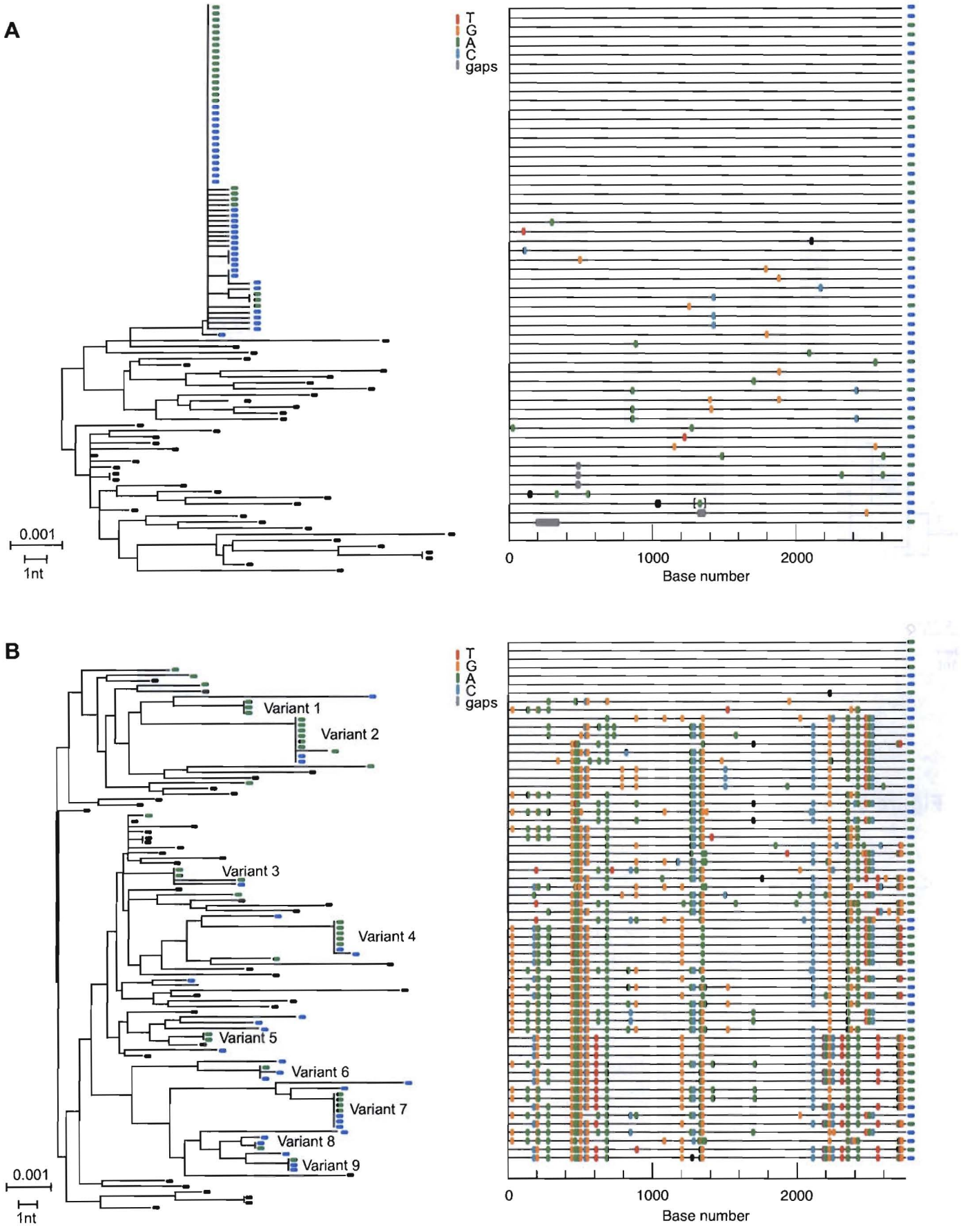
A
T
G
A
C
gaps
0.001
1nt
0
1000
2000
Base number
B
Variant 1
Variant 2
Variant 3
Variant 4
Variant 5
Variant 6
Variant 7
Variant 8
Variant 9
0.001
1nt
T
G
A
C
gaps
0
1000
2000
Base number

Figure 4

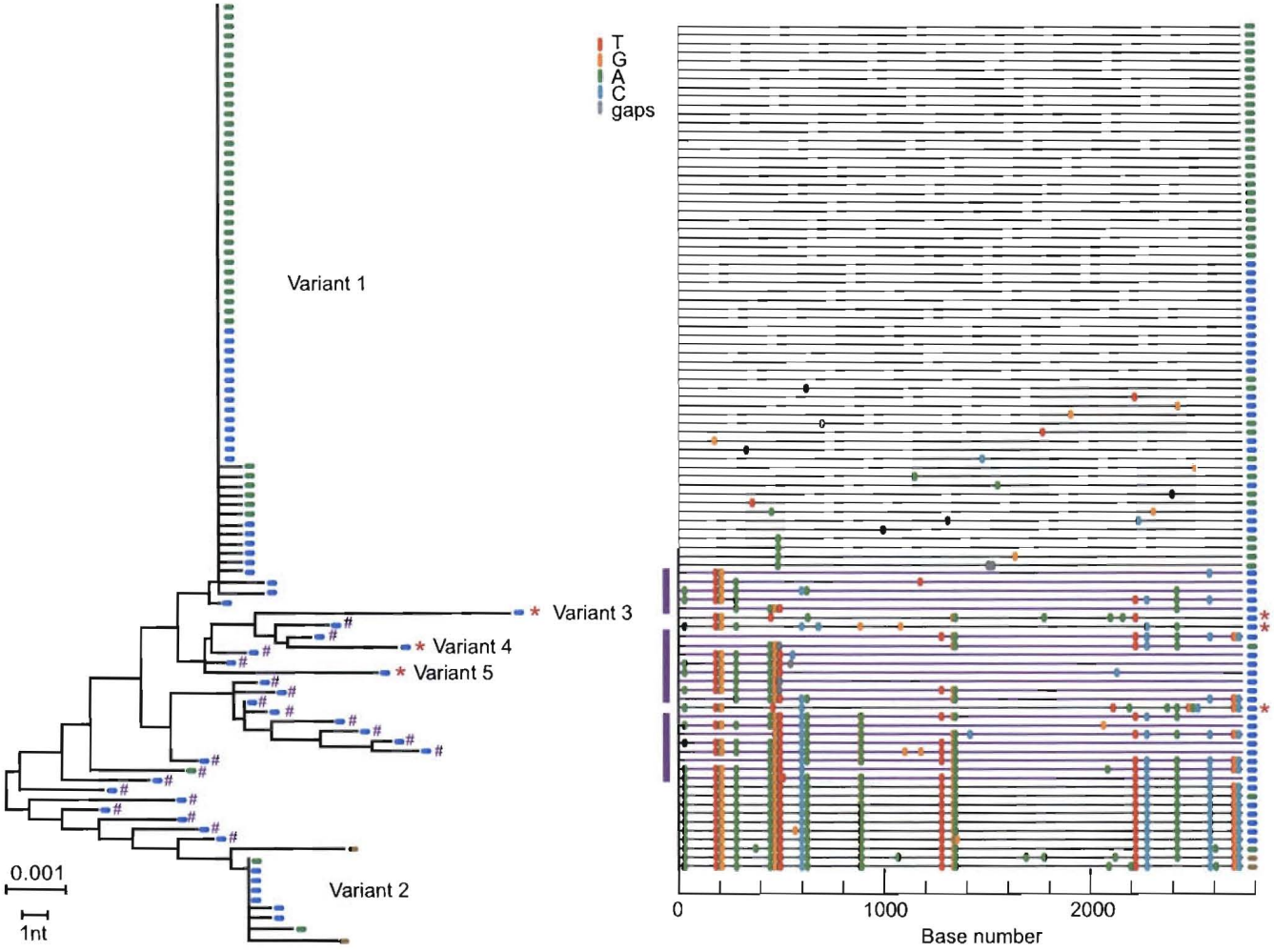
Variant 1
Variant 3
Variant 4
Variant 5
Variant 2
0.001
1nt
T
G
A
C
gaps
0
1000
2000
Base number

Figure 5

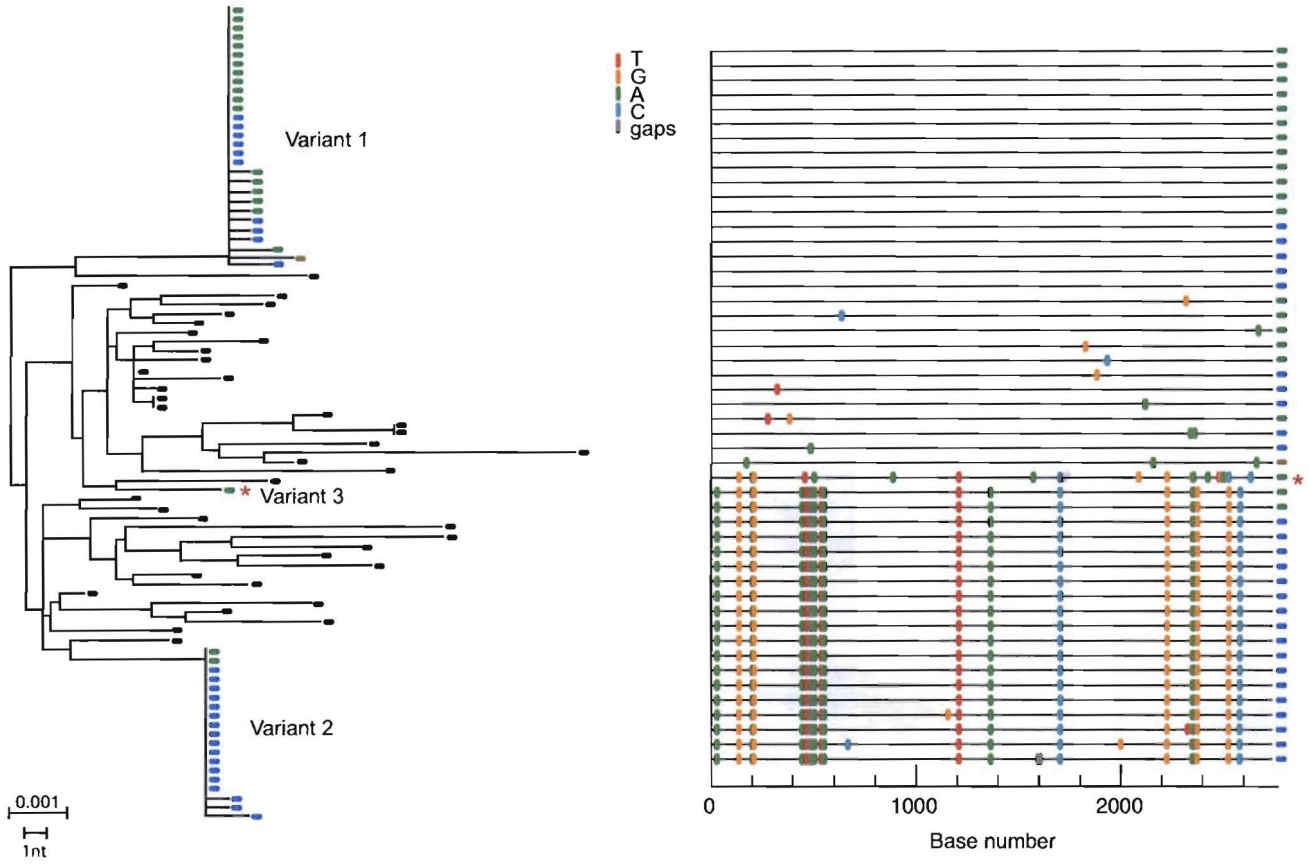
Variant 1
Variant 3
Variant 2
0.001
1nt
T
G
A
C
gaps
0
1000
2000
Base number

Figure 6

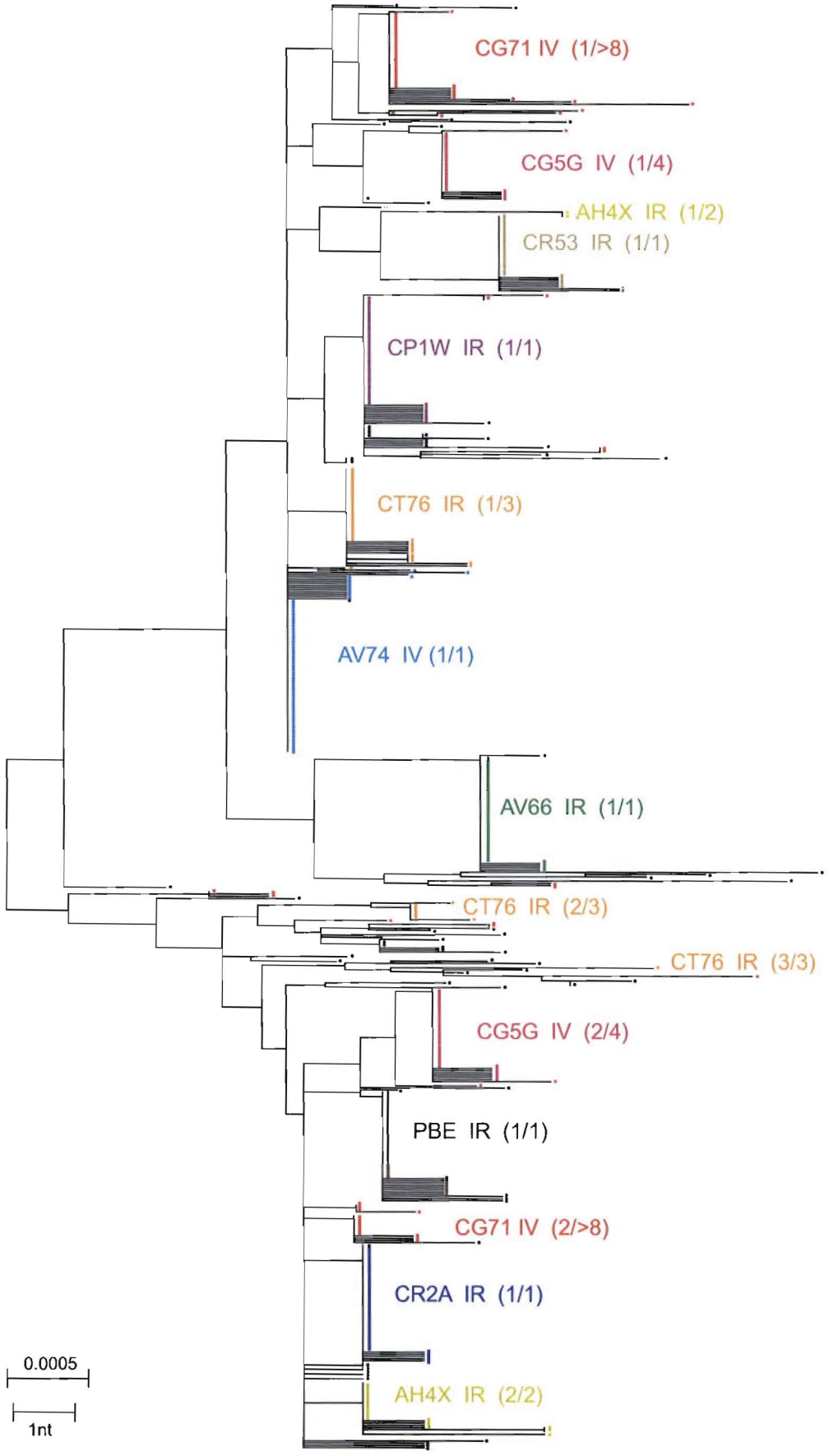
CG71 IV (1/>8)
CG5G IV (1/4)
AH4X IR (1/2)
CR53 IR (1/1)
CP1W IR (1/1)
CT76 IR (1/3)
AV74 IV (1/1)
AV66 IR (1/1)
CT76 IR (2/3)
CT76 IR (3/3)
CG5G IV (2/4)
PBE IR (1/1)
CG71 IV (2/>8)
CR2A IR (1/1)
AH4X IR (2/2)
0.0005
1nt

Figure 7

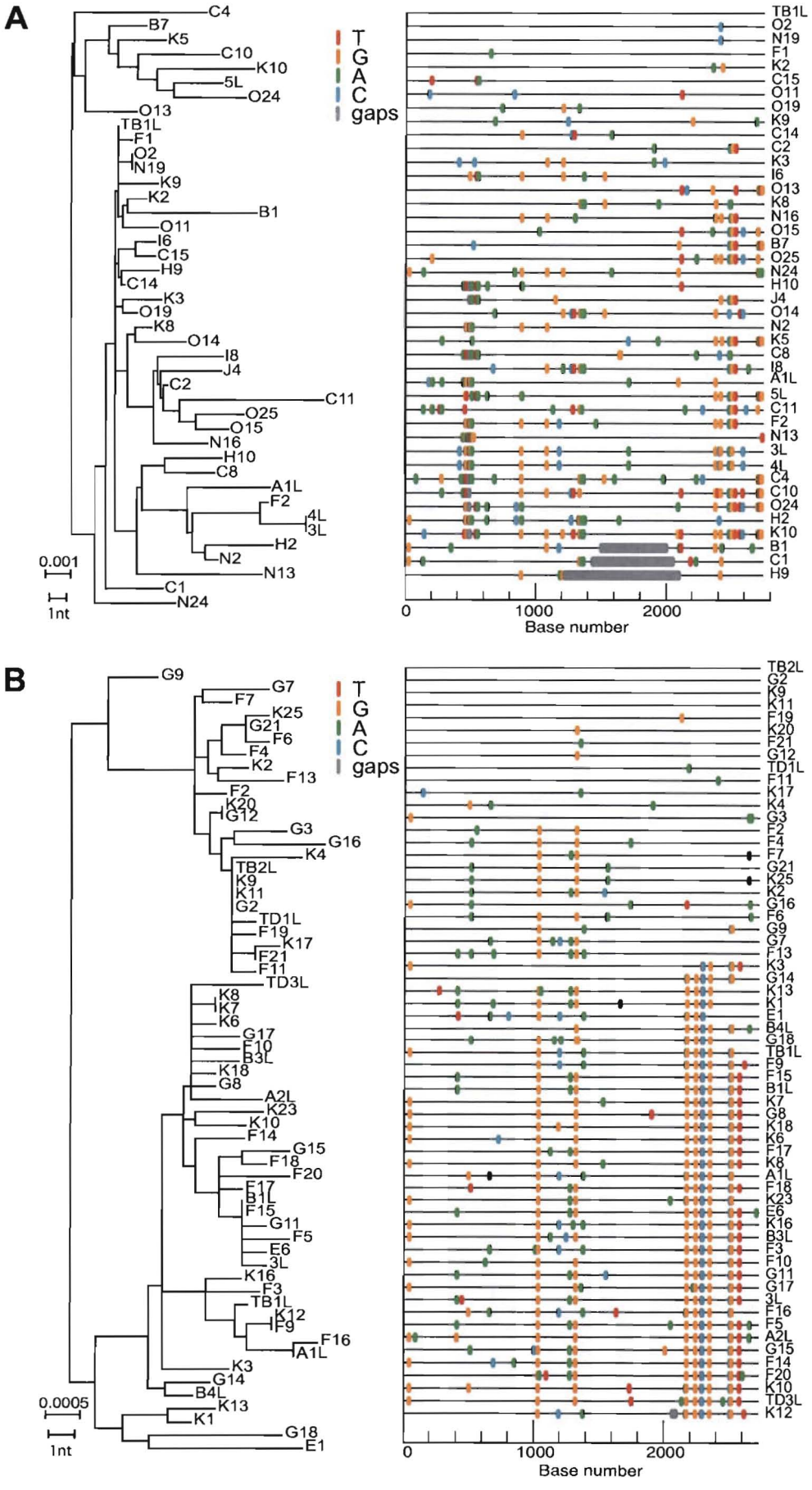
A
C4
B7
K5
C10
K10
5L
O24
O13
TB1L
F1
O2
N19
K9
K2
B1
O11
I6
C15
H9
C14
K3
O19
K8
O14
I8
J4
C2
C11
O25
O15
N16
H10
C8
A1L
F2
4L
3L
H2
N2
N13
C1
N24
0.001
1nt
T
G
A
C
gaps
TB1L
O2
N19
F1
K2
C15
O11
O19
K9
C14
C2
K3
I6
O13
K8
N16
O15
B7
O25
N24
H10
J4
O14
N2
K5
C8
I8
A1L
5L
C11
F2
N13
3L
4L
C4
C10
O24
H2
K10
B1
C1
H9
0
1000
2000
Base number
B
G9
G7
F7
K25
G21
F6
F4
K2
F13
F2
K20
G12
G3
G16
K4
TB2L
K9
K11
G2
TD1L
F19
K17
F21
F11
TD3L
K8
K7
K6
G17
F10
B3L
K18
G8
A2L
K23
K10
F14
G15
F18
F20
F17
B1L
F15
G11
F5
E6
3L
K16
F3
TB1L
K12
F9
F16
A1L
K3
G14
B4L
K13
K1
G18
E1
0.0005
1nt
T
G
A
C
gaps
TB2L
G2
K9
K11
F19
K20
F21
G12
TD1L
F11
K17
K4
G3
F2
F4
F7
G21
K25
K2
G16
F6
G9
G7
F13
K3
G14
K13
K1
E1
B4L
G18
TB1L
F9
F15
B1L
K7
G8
K18
K6
F17
K8
A1L
F18
K23
E6
K16
B3L
F3
F10
G11
G17
3L
F16
F5
A2L
G15
F14
F20
K10
TD3L
K12
0
1000
2000
Base number

Figure S1

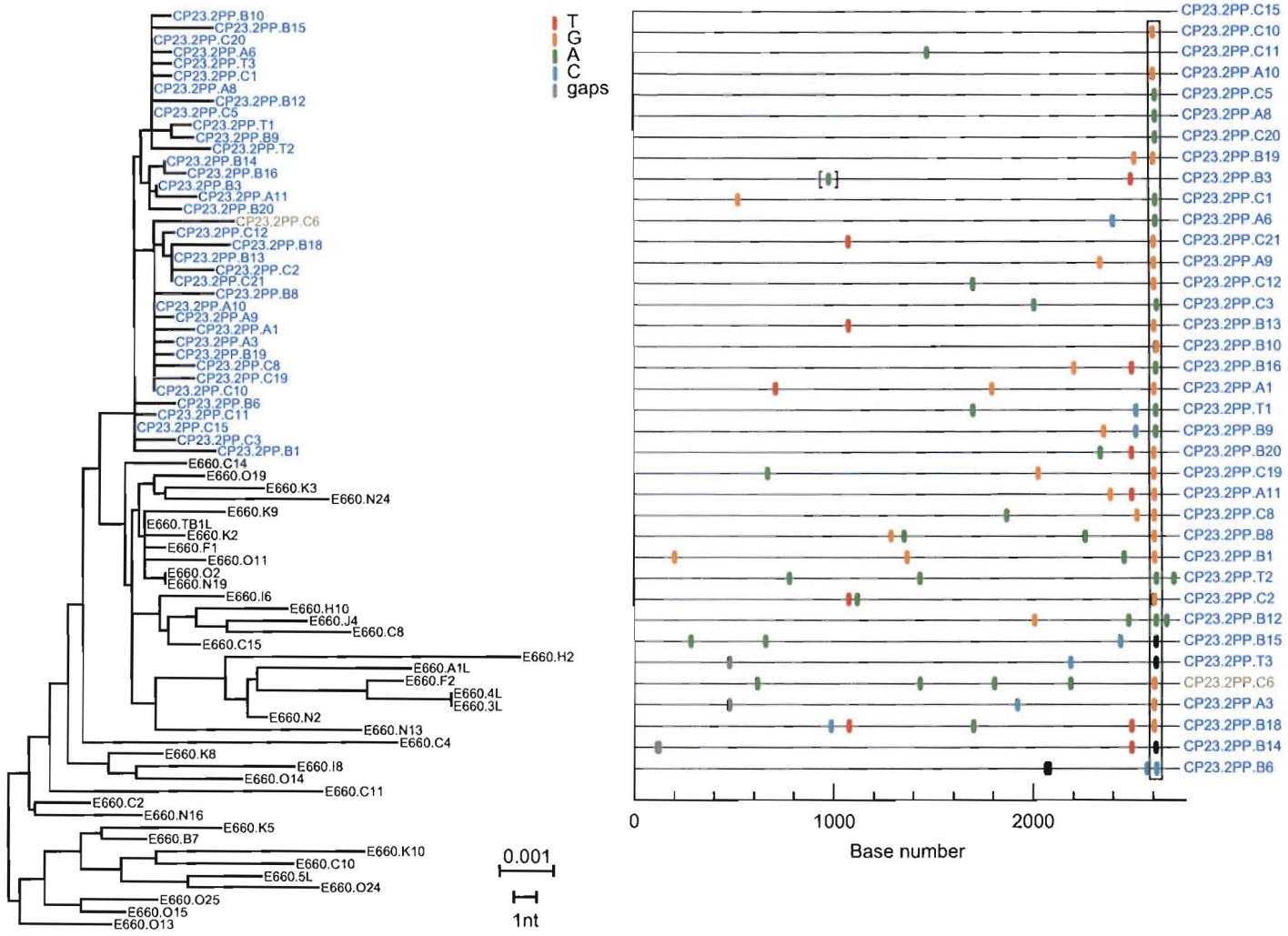
CP23.2PP.B10
CP23.2PP.B15
CP23.2PP.C20
CP23.2PP.A6
CP23.2PP.T3
CP23.2PP.C1
CP23.2PP.A8
CP23.2PP.B12
CP23.2PP.C5
CP23.2PP.T1
CP23.2PP.B9
CP23.2PP.T2
CP23.2PP.B14
CP23.2PP.B16
CP23.2PP.B3
CP23.2PP.A11
CP23.2PP.B20
CP23.2PP.C6
CP23.2PP.C12
CP23.2PP.B18
CP23.2PP.B13
CP23.2PP.C2
CP23.2PP.C21
CP23.2PP.B8
CP23.2PP.A10
CP23.2PP.A9
CP23.2PP.A1
CP23.2PP.A3
CP23.2PP.B19
CP23.2PP.C8
CP23.2PP.C19
CP23.2PP.C10
CP23.2PP.B6
CP23.2PP.C11
CP23.2PP.C15
CP23.2PP.C3
CP23.2PP.B1
E660.C14
E660.O19
E660.K3
E660.N24
E660.K9
E660.TB1L
E660.K2
E660.F1
E660.O11
E660.O2
E660.N19
E660.I6
E660.H10
E660.J4
E660.C8
E660.C15
E660.H2
E660.A1L
E660.F2
E660.4L
E660.3L
E660.N2
E660.N13
E660.C4
E660.K8
E660.I8
E660.O14
E660.C11
E660.C2
E660.N16
E660.K5
E660.B7
E660.K10
E660.C10
E660.5L
E660.O24
E660.O25
E660.O15
E660.O13
0.001
1nt
T
G
A
C
gaps
CP23.2PP.C15
CP23.2PP.C10
CP23.2PP.C11
CP23.2PP.A10
CP23.2PP.C5
CP23.2PP.A8
CP23.2PP.C20
CP23.2PP.B19
CP23.2PP.B3
CP23.2PP.C1
CP23.2PP.A6
CP23.2PP.C21
CP23.2PP.A9
CP23.2PP.C12
CP23.2PP.C3
CP23.2PP.B13
CP23.2PP.B10
CP23.2PP.B16
CP23.2PP.A1
CP23.2PP.T1
CP23.2PP.B9
CP23.2PP.B20
CP23.2PP.C19
CP23.2PP.A11
CP23.2PP.C8
CP23.2PP.B8
CP23.2PP.B1
CP23.2PP.T2
CP23.2PP.C2
CP23.2PP.B12
CP23.2PP.B15
CP23.2PP.T3
CP23.2PP.C6
CP23.2PP.A3
CP23.2PP.B18
CP23.2PP.B14
CP23.2PP.B6
0
1000
2000
Base number

Figure S2

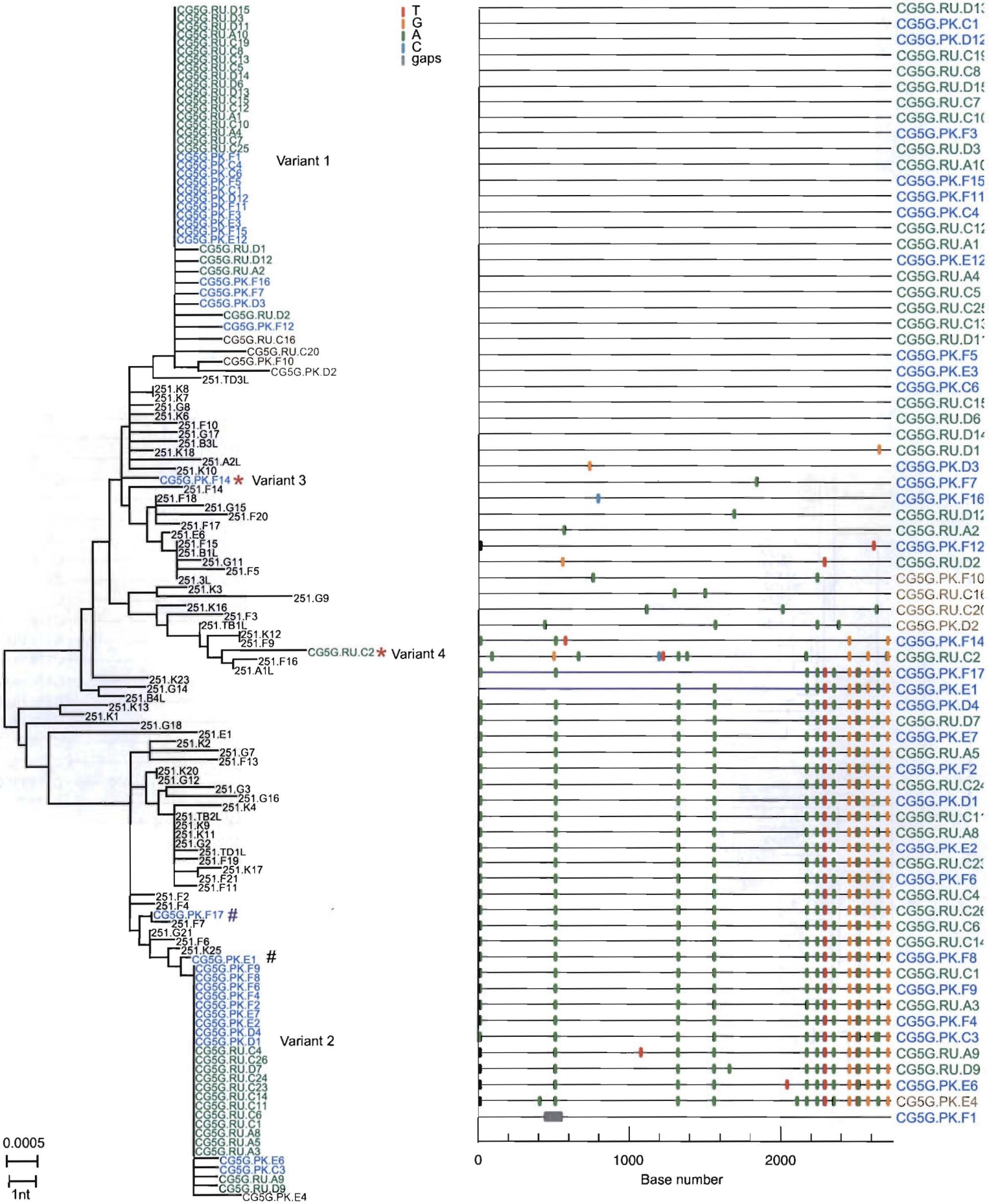
T
G
A
C
gaps
Variant 1
Variant 3
Variant 4
Variant 2
0.0005
1nt
0
1000
2000
Base number

Figure S3

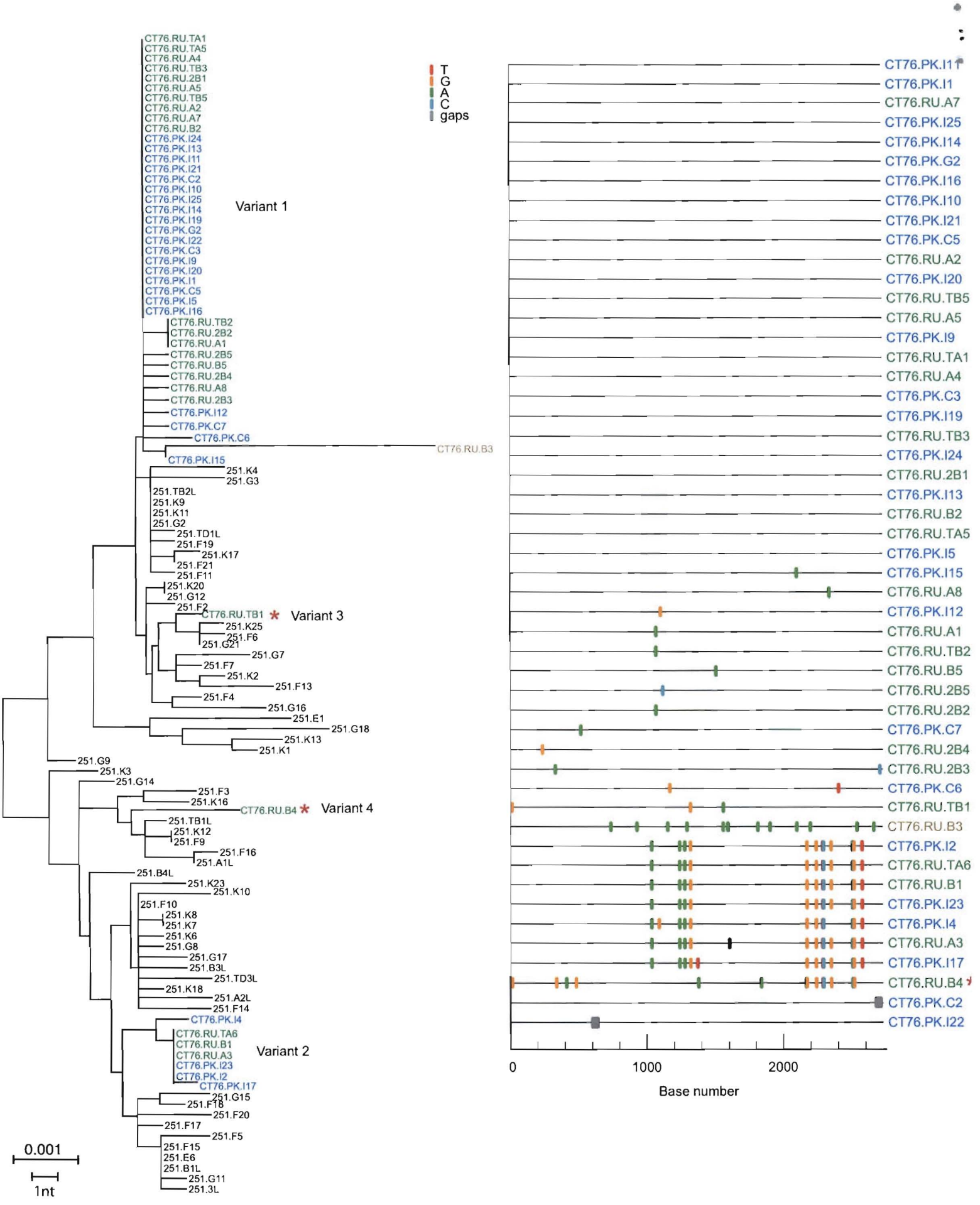
T
G
A
C
gaps
Variant 1
Variant 3
Variant 4
Variant 2
CT76.RU.TA1
CT76.RU.TA5
CT76.RU.A4
CT76.RU.TB3
CT76.RU.2B1
CT76.RU.A5
CT76.RU.TB5
CT76.RU.A2
CT76.RU.A7
CT76.RU.B2
CT76.PK.I24
CT76.PK.I13
CT76.PK.I11
CT76.PK.I21
CT76.PK.C2
CT76.PK.I10
CT76.PK.I25
CT76.PK.I14
CT76.PK.I19
CT76.PK.G2
CT76.PK.I22
CT76.PK.C3
CT76.PK.I9
CT76.PK.I20
CT76.PK.I1
CT76.PK.C5
CT76.PK.I5
CT76.PK.I16
CT76.RU.TB2
CT76.RU.2B2
CT76.RU.A1
CT76.RU.2B5
CT76.RU.B5
CT76.RU.2B4
CT76.RU.A8
CT76.RU.2B3
CT76.PK.I12
CT76.PK.C7
CT76.PK.C6
CT76.RU.B3
CT76.PK.I15
251.K4
251.G3
251.TB2L
251.K9
251.K11
251.G2
251.TD1L
251.F19
251.K17
251.F21
251.F11
251.K20
251.G12
251.F2
CT76.RU.TB1
251.K25
251.F6
251.G21
251.G7
251.F7
251.K2
251.F13
251.F4
251.G16
251.E1
251.G18
251.K13
251.K1
251.G9
251.K3
251.G14
251.F3
251.K16
CT76.RU.B4
251.TB1L
251.K12
251.F9
251.F16
251.A1L
251.B4L
251.K23
251.K10
251.F10
251.K8
251.K7
251.K6
251.G8
251.G17
251.B3L
251.TD3L
251.K18
251.A2L
251.F14
CT76.PK.I4
CT76.RU.TA6
CT76.RU.B1
CT76.RU.A3
CT76.PK.I23
CT76.PK.I2
CT76.PK.I17
251.G15
251.F18
251.F20
251.F17
251.F5
251.F15
251.E6
251.B1L
251.G11
251.3L
0.001
1nt
CT76.PK.I11
CT76.PK.I1
CT76.RU.A7
CT76.PK.I25
CT76.PK.I14
CT76.PK.G2
CT76.PK.I16
CT76.PK.I10
CT76.PK.I21
CT76.PK.C5
CT76.RU.A2
CT76.PK.I20
CT76.RU.TB5
CT76.RU.A5
CT76.PK.I9
CT76.RU.TA1
CT76.RU.A4
CT76.PK.C3
CT76.PK.I19
CT76.RU.TB3
CT76.PK.I24
CT76.RU.2B1
CT76.PK.I13
CT76.RU.B2
CT76.RU.TA5
CT76.PK.I5
CT76.PK.I15
CT76.RU.A8
CT76.PK.I12
CT76.RU.A1
CT76.RU.TB2
CT76.RU.B5
CT76.RU.2B5
CT76.RU.2B2
CT76.PK.C7
CT76.RU.2B4
CT76.RU.2B3
CT76.PK.C6
CT76.RU.TB1
CT76.RU.B3
CT76.PK.I2
CT76.RU.TA6
CT76.RU.B1
CT76.PK.I23
CT76.PK.I4
CT76.RU.A3
CT76.PK.I17
CT76.RU.B4
CT76.PK.C2
CT76.PK.I22
0
1000
2000
Base number

Figure S4

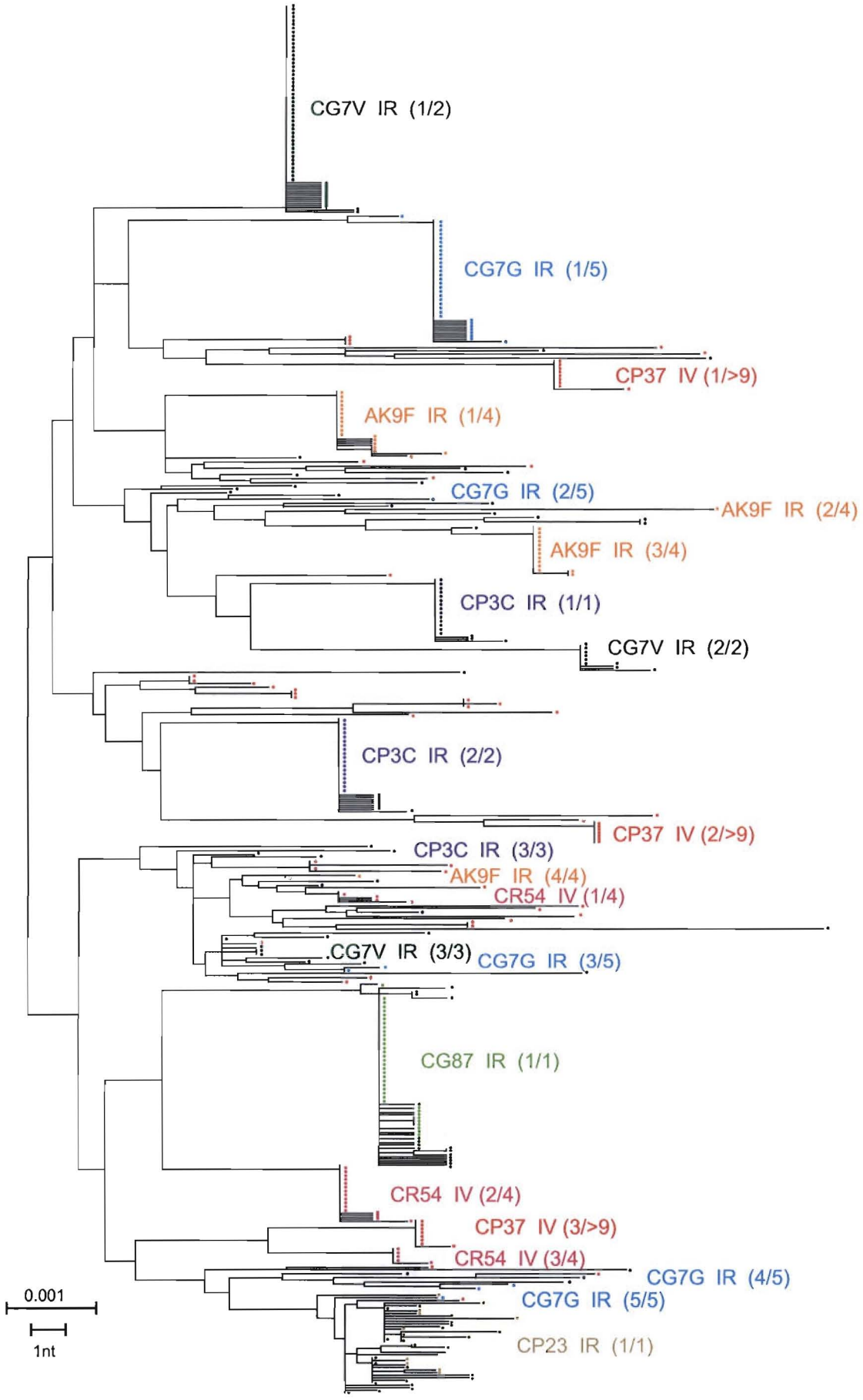
CG7V IR (1/2)
CG7G IR (1/5)
CP37 IV (1/>9)
AK9F IR (1/4)
CG7G IR (2/5)
AK9F IR (2/4)
AK9F IR (3/4)
CP3C IR (1/1)
CG7V IR (2/2)
CP3C IR (2/2)
CP37 IV (2/>9)
CP3C IR (3/3)
AK9F IR (4/4)
CR54 IV (1/4)
CG7V IR (3/3)
CG7G IR (3/5)
CG87 IR (1/1)
CR54 IV (2/4)
CP37 IV (3/>9)
CR54 IV (3/4)
CG7G IR (4/5)
CG7G IR (5/5)
CP23 IR (1/1)
0.001
1nt

Figure S5