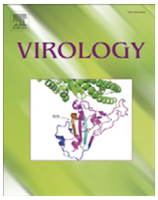
Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro



Estimating the date of origin of an HIV-1 circulating recombinant form

Kok Keng Tee [a,b,1], Oliver G. Pybus [c,1], Joe Parker [c], Kee Peng Ng [d], Adeeba Kamarulzaman [b], Yutaka Takebe [a,*]

[a] *Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan*
[b] *Department of Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[c] *Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*
[d] *Department of Medical Microbiology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia*

ARTICLE INFO

Article history:
Received 20 January 2009
Returned to author for revision
4 February 2009
Accepted 11 February 2009
Available online 9 March 2009

Keywords:
HIV
Recombination
Coalescence
Reticulate evolution
Genesis

ABSTRACT

HIV is capable of frequent genetic exchange through recombination. Despite the pandemic spread of HIV-1 recombinants, their times of origin are not well understood. We investigate the epidemic history of a HIV-1 circulating recombinant form (CRF) by estimating the time of the recombination event that lead to the emergence of CRF33_01B, a recently described recombinant descended from CRF01_AE and subtype B. The *gag*, *pol* and *env* genes were analyzed using a combined coalescent and relaxed molecular clock model, implemented in a Bayesian Markov chain Monte Carlo framework. Using linked genealogical trees we calculated the time interval between the common ancestor of CRF33_01B and the ancestors it shares with closely related parental lineages. The recombination event that generated CRF33_01B (t_{rec}) occurred sometime between 1991 and 1993, suggesting that recombination is common in the early evolutionary history of HIV-1. The proof-of-concept approach provides a new tool for the investigation of HIV molecular epidemiology and evolution.

Introduction

RNA viruses are characterized by their capacity to generate and accumulate large numbers of genomic mutations (Holland et al., 1982). Many RNA viruses are also capable of exchanging genetic material with one another by homologous recombination (Lai, 1992) — contributing to viral evolution and survival by generating genetic variation and creating new viruses (Hahn et al., 1988).

Human immunodeficiency virus (HIV) is among the most genetically variable human viruses and is characterized by high rates of mutation, viral replication and recombination (Ho et al., 1995; Perelson et al., 1996; Robertson et al., 1995). Separate cross-species transmission of simian immunodeficiency virus in chimpanzees (SIVcpz) to humans has resulted in the diversification of HIV type 1 (HIV-1) into three groups, denoted M, N and O (Keele et al., 2006; Van Heuverswyn et al., 2006), with the major pandemic group M viruses being further divided into subtypes and subsubtypes (A to D, F to H, J and K). In addition to pure subtypes, HIV-1 circulating recombinant forms (CRFs) are generated by recombination between distinct subtypes and/or other CRFs and are spreading at epidemic rates in various parts of the world (Robertson et al., 2000). HIV-1 evolves particularly quickly, resulting in the generation and accumulation of significant numbers of mutations over short timescales. Many studies have exploited this rapid evolution and used statistical methods to reconstruct the origins and evolutionary dynamics of HIV from viral sequence data (Korber et al., 2000; Lemey et al., 2003; Salemi et al., 2001; Tee et al., 2008; Worobey et al., 2008; Zhu et al., 1998). However, the molecular evolution and epidemic history of HIV-1 recombinants have not been fully explored – despite the dramatic expansion of CRFs worldwide – possibly due to a lack of statistical methods that can reliably reconstruct the evolutionary dynamics of sequences with recombination. HIV-1 CRF33_01B, for example, is a recently emerged recombinant in Southeast Asia descended from CRF01_AE and subtype B (Tee et al., 2006, 2005a,b). Thought to be originated in Malaysia, CRF33_01B has been expanding and found commonly among HIV-infected individuals from various risk groups in the region (Lau et al., 2008; Tee et al., 2006, in press; Wang et al., 2007). In this study, we attempt to assess the age of the putative recombination event that gave rise to CRF33_01B by incorporating a relaxed clock Bayesian phylogenetic model into a classical framework of reticulate evolution. This straightforward, proof-of-concept approach provides a practical and robust evolutionary biology method for estimating the so-called “date of birth” of recombinant lineages for which closely related non-recombinant lineages have been sampled.

Results

Estimation of recombination history

The shared ancestry of a population that is derived from a recombination event or events (e.g. a recombinant lineage of viruses) can be represented in the form of an ancestral recombination graph, as

[*] Corresponding author. Fax: +81 3 5285 1258.
E-mail address: takebe@nih.go.jp (Y. Takebe).
[1] K.K.T. and O.G.P. contributed equally to this work.

0042-6822/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2009.02.020

illustrated in Fig. 1. This graph is drawn on a real calendar timescale, such that the vertical branch lengths represent time in months or years. The graph results from the combination of two phylogenies, representing the two parental clades, x and y, and includes the putative xy recombinant generated by recombination between a member of clade x and a member of clade y. In the graph we can discern the time of the most recent common ancestor (tMRCA) of the xy recombinant (denoted t_{xy}) and the divergence times of the most closely-related x and y parental lineages with the recombinant, denoted t_x and t_y, respectively. Assuming there is a single recombination event, the time of origin of the xy recombinant (denoted t_{rec}) can be estimated as the interval between t_{xy} and the most recent of either t_x or t_y. For example, in Fig. 1, the divergence time of clade x and the recombinant is more recent than that of clade y, hence in this instance t_{rec} is defined as the interval between t_x and t_{xy}.

To investigate the origin and recombination history of HIV-1 CRF33_01B (a recombinant derived from CRF01_AE and subtype B) partial CRF33_01B genome sequences were obtained and formed the basis of three alignments with different parental subtypes: CRF01_AE$_{\mathrm{gag}}$, CRF01_AE$_{\mathrm{gag-env}}$ (a concatenated alignment of CRF01_AE$_{\mathrm{gag}}$ and the *env* genes), and subtype B$_{\mathrm{gag-pol}}$ (Fig. 2A). Phylogenetic reconstruction of the three non-overlapping regions showed that contemporary CRF33_01B isolates grouped as a distinct monophyletic cluster within each tree, demonstrating its descent from the parental subtypes of Southeast Asian origin (Kalish et al., 1995; Ou et al., 1993) (Supplementary Fig. 1). Further evolutionary analysis was performed on CRF01_AE and subtype B reference sequences with known sampling dates, spanning 17 and 23 years, respectively. Table 1 summarizes the evolutionary parameters estimated from these data sets using a Bayesian MCMC relaxed clock approach. To check that our results are robust to model specification, we obtained posterior distributions under different nucleotide substitution and demographic models. The estimated evolutionary rates under the GTR substitution and constant size population model (expressed in 10^{-3} substitutions/site/year) for each genome region were as follows: 2.6 (95% credible region [CR]: 2.4–2.7) for CRF01_AE$_{\mathrm{gag}}$; 3.4 (95% CR: 3.1–3.6) for CRF01_AE$_{\mathrm{gag-env}}$; 3.7 (95% CR: 3.5–4.0) for subtype B$_{\mathrm{gag-pol}}$. The estimated values of the coefficient of variation parameter were well above zero, in line with previous estimates for HIV-1 (de Oliveira et al., 2006), indicating that the data exhibit significant variation in evolutionary rate among lineages. The relaxed clock analysis was used to estimate the date of the common ancestor of CRF33_01B and its parental lineages. As shown in Table 1 and Fig. 2B, the mean estimated dates for the MRCA of CRF33_01B were similar among the three alignments (CRF01_AE$_{\mathrm{gag}}$, CRF01_AE$_{\mathrm{gag-env}}$ and subtype B$_{\mathrm{gag-pol}}$), ranging from 1992.5 to 1993.8. This concordance among estimates obtained from different genome regions indicates that the sampled CRF33_01B isolates did share a single common ancestor. To obtain a consensus MRCA date for CRF33_01B, we used the average of the three MRCA estimates, which was 1993.1 (denoted tMRCA$_{\mathrm{CRF33}}$). We also estimated the time at which CRF33_01B diverged from the most closely-related parental lineages sampled, which was 1985.5 and 1986.6 for the CRF01_AE parental clade, and 1991.1 for the subtype B parental clade. Therefore, given the reference strains available, we were able to estimate a more recent divergence for CRF33_01B and subtype B than for CRF01_AE and CRF33_01B. Following the model described above (Fig. 1), we conclude that t_{rec} – the "date of origin" – of CRF33_01B occurred sometime between 1991.1 and 1993.1. The most conservative estimate possible for the date of CRF33_01B origin can be obtained by measuring between the lowest lower limit and the highest upper limit of the 95% credible regions of the relevant divergence dates, which for our data is 1987.3 and 1997.1. Additional analyses performed using the HKY substitution model were parallel to those obtained under the GTR model (shown in Supplementary Table 2). Of note, the exponential demographic model did not fit these data sets well, as indicated by unreasonable effective sample size (ESS) scores (data not shown), suggesting that there could be insufficient information in the data to accommodate a complex tree model.

Discussion

We have demonstrated that the time of an HIV recombination event, representing the date of genesis of a recombinant lineage, can be estimated by using relaxed molecular clock models to enforce a timescale onto linked phylogenies. As an example, we have attempted to estimate the origin time of a recently identified HIV-1 recombinant, CRF33_01B, which is derived from two well-characterized HIV-1 clades, namely CRF01_AE and subtype B. Thought to be originated and endemic in Malaysia, CRF33_01B has been spreading among various risk populations (Tee et al., 2006). Using a Bayesian MCMC phylogenetic framework in the context of a simple model of reticulate evolution, we investigated the evolutionary history of CRF33_01B (Fig. 1). Given HIV-1 sequence data with known dates of

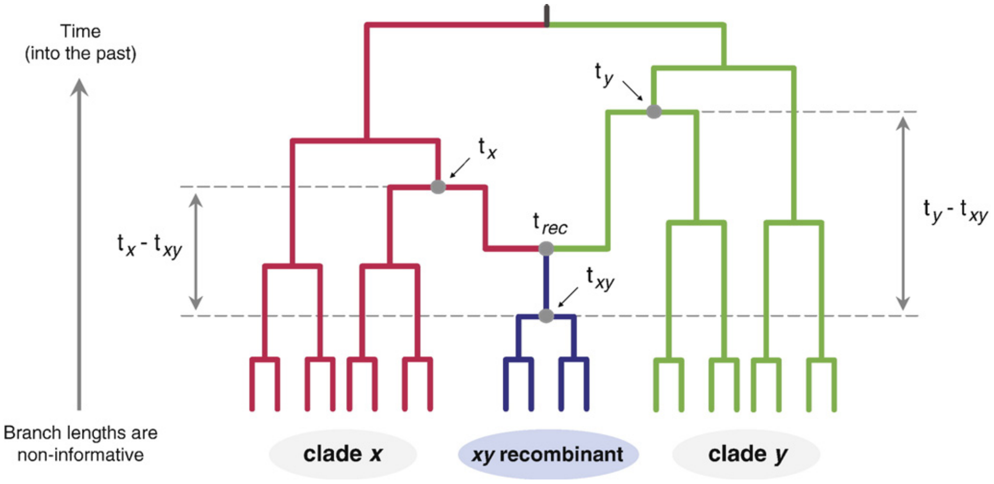


Fig. 1. An ancestral recombination graph representing the classical model of reticulate evolution. In a recombinogenic virus population, intertypic recombination between clades x and y generates the recombinant xy, which shares the evolutionary histories of both parent lineages, thus generating a closed loop (or 'reticulation') in the graph. If each clade-specific subgraph (i.e. phylogeny) is estimated using a molecular clock model, then the coalescence time of the xy recombinant clade (t_{xy}) can be estimated. In addition, the time that xy diverged from its most closely-related ancestral lineages can be estimated (denoted t_x and t_y). Taken together, these times can be used to define an interval during which the xy recombination event must have occurred. See text for further descriptions.

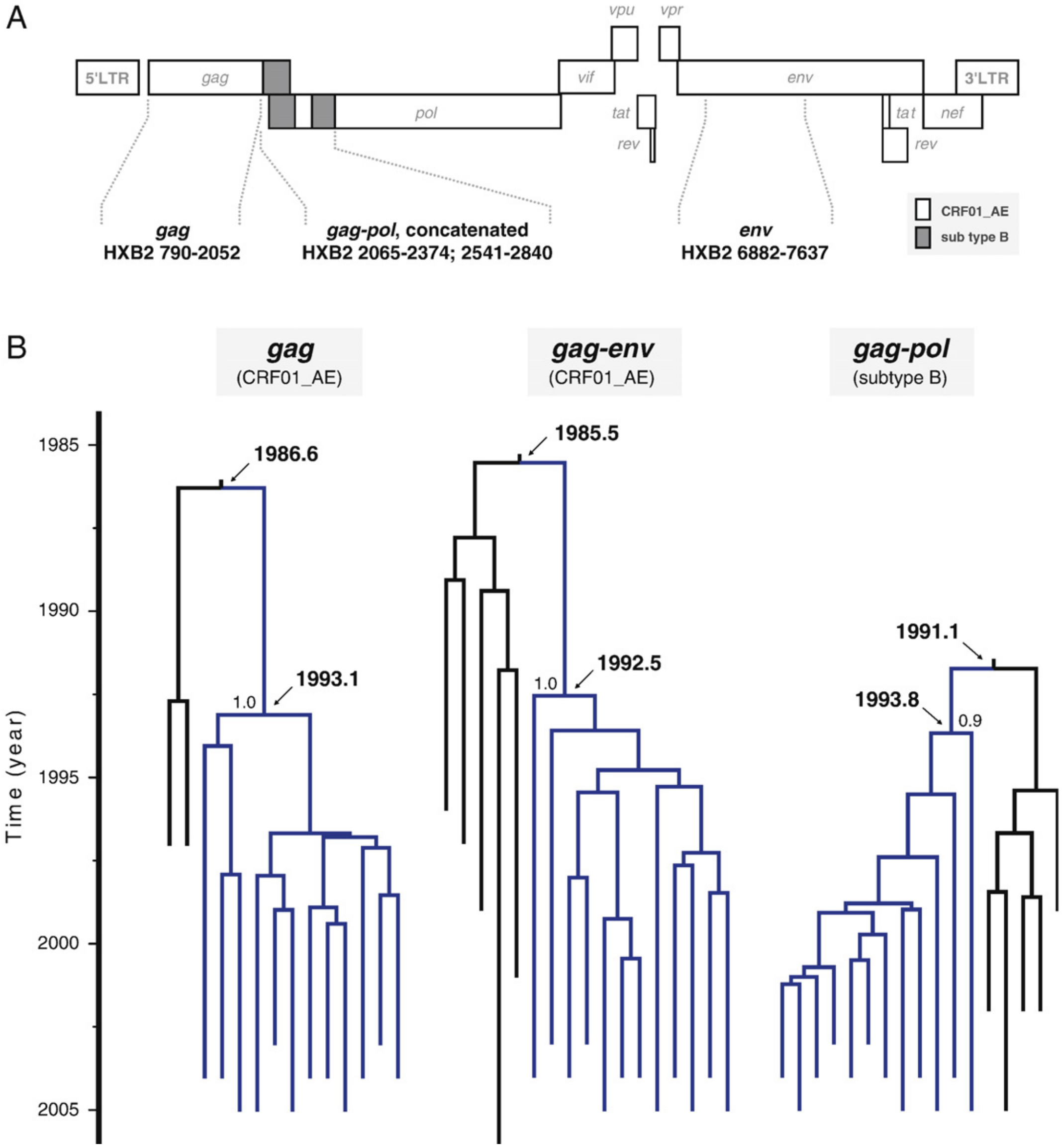


Fig. 2. (A) Genome structure of HIV-1 CRF33_01B. In this recombinant lineage, subtype B fragments are found within the *gag–pol* gene region, in a genomic background that matches CRF01_AE. Mosaic genome structure was determined and confirmed by bootscanning and by informative-sites plus sub-region tree analyses, as described elsewhere (Tee et al., 2006). We created three non-overlapping alignments from CRF33_01B: CRF01_AE$_{gag}$, subtype B$_{gag-pol}$ (concatenated) and CRF01_AE$_{env}$. To achieve a sufficient level of phylogenetic signal, the CRF01_AE$_{gag}$ and CRF01_AE$_{env}$ data sets were concatenated (CRF01_AE$_{gag-env}$) for subsequent analyses. Nucleotide positions corresponding to HXB2 prototype strain (accession number K03455) are indicated. (B) Phylogenetic reconstructions of HIV-1 CRF33_01B and the closely related parental lineages belonging to clades CRF01_AE and subtype B. Maximum clade credibility phylogenies were obtained using BEAST v1.4 (Drummond and Rambaut, 2007) from alignments that represent the regions CRF01_AE$_{gag}$ (HXB2 790–2052), CRF01_AE$_{gag-env}$ (HXB2 790–2052 and 6882–7637) and subtype B$_{gag-pol}$ (HXB2 2065–2840). Reference sequences belonging to CRF01_AE (*gag* and *env*) and subtype B (*gag–pol*) were obtained from the HIV sequence database (www.hiv.lanl.gov) (only closely-related parental lineages are shown in this figure — see Supplementary Fig. 1 for larger phylogenies). Monophyletic clusters ($P \geq 0.9$) of CRF33_01B sequences are colored. Tree branches are scaled in units of time. Dates of the most recent common ancestor (MRCA) of CRF33_01B and the divergence times of closely-related parental lineages (as estimated using BEAST) are indicated on the respective nodes. Estimated evolutionary parameters obtained for these data sets are provided in Table 1.

sampling, we estimated that the recombination event that generated CRF33_01B occurred between year 1991.1 and 1993.1. (Fig. 2B, Table 1), with a conservative interval between 1987.3 and 1997.1. This period coincided with the early phase of the HIV epidemic in Malaysia, during which annual incidence was increasing rapidly (Fig. 3A). Although the simple and robust method is practical for estimating the date of birth of

Table 1
Bayesian Markov Chain Monte Carlo (MCMC) evolutionary parameters for HIV-1 CRF33_01B

Data sets	Rate of evolution	Coefficient of variation	Date of MRCA of CRF33_01B (year)	Date of divergence of CRF33_01B closely-related parents (year)
gag (CRF01_AE)	2.6 (2.4, 2.7)	0.4 (0.2, 0.6)	1993.1 (1990.6, 1995.5)	1986.6 (1983.0, 1989.3)
gag and *env* (CRF01_AE, concatenated)	3.4 (3.1, 3.6)	0.4 (0.3, 0.6)	1992.5 (1989.8, 1995.1)	1985.5 (1983.8, 1987.0)
gag and *pol* (subtype B, concatenated)	3.7 (3.5, 4.0)	0.5 (0.3, 0.6)	1993.8 (1990.6, 1997.1)	1991.1 (1987.3, 1995.0)

Posterior distributions are estimated using a relaxed molecular clock model (Drummond et al., 2006). The general time-reversible substitution model was used with gamma-distributed among-site rate variation (GTR + γ_4) and a constant-sized demographic model. Nucleotide substitution rates are expressed in 10^{-3} substitutions per site per year. 95% highest posterior densities (HPDs) are shown in parentheses.

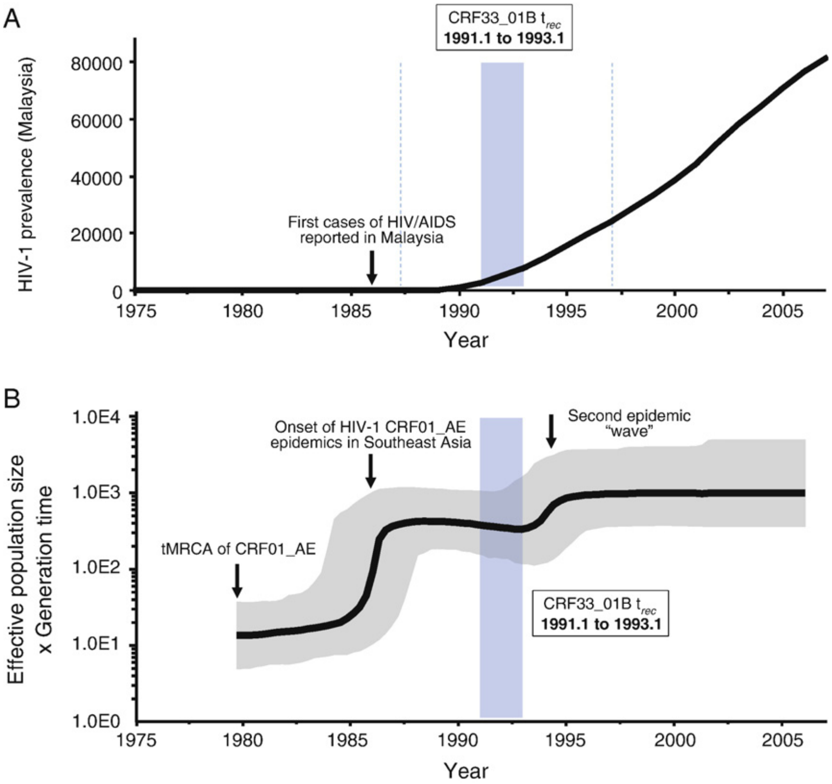


Fig. 3. (A) Epidemiological dynamics of HIV-1 prevalence in Malaysia from the 1980s to 2007 (source: AIDS/STI Unit, Ministry of Health Malaysia). The time of recombination event between HIV-1 CRF01_AE and subtype B that gave rise to CRF33_01B (t_{rec}) is depicted as a vertical bar. This 2-year period coincides with the early stages of the HIV epidemic in Malaysia. Dotted vertical lines indicate the most conservative possible date range for CRF33_01B origin (1987.3 to 1997.1, see text for details). (B) Bayesian skyline plot (Drummond et al., 2005) for the CRF01_AE clade circulated throughout Southeast Asia. Population dynamics were estimated using the GTR + γ_4 parameter in BEAST 1.4. See Materials and methods for full analysis details. Note that the *y*-axis is proportional to effective population size, which reflects prevalence. The black line represents the median estimate and the shaded region represents the 95% highest posterior density (HPD) credible region. The plot indicates that Asian CRF01_AE first emerged in 1979.6 (95% CR: 1976.4–1982.7) and later established in Southeast Asia around mid-1980s, followed by a second but less significant "wave" a decade later. The time of recombination event that generated CRF33_01B (t_{rec}) is also depicted as a vertical bar.

recombinant viruses, the analysis could be affected by several factors: (i) poor sampling, (ii) complex mosaic genomes structures of CRFs, (iii) low diversity genome regions (e.g. the polymerase gene), and (iv) difficulty in estimating genetic distance at saturated sites in which too many mutations occur at the same site.

We further investigated the epidemic history of CRF01_AE in Southeast Asia by estimating a Bayesian skyline plot for the CRF01_AE$_{\mathrm{gag-env}}$ data set, which depicts the change in effective population size through time (Drummond et al., 2005) (Fig. 3B). The plots show that the CRF01_AE epidemic in Southeast Asia began with a rapid exponential increase in effective population size. The plot also suggests a more modest second epidemic "wave" a decade later; however this second increase is not statistically significant given the confidence limits obtained. The CRF01_AE epidemic has since matured and, to date, accounts for more than 80% of all infections in South and Southeast Asia (excluding India) (Hemelaar et al., 2006). Moreover, the population genetic data correlate with the epidemic timeline of CRF01_AE across the region (Weniger et al., 1994); for example, the first cases of HIV/AIDS in Malaysia were reported around 1986 (Goh et al., 1987). To our knowledge, this is the first genetic evidence that specifically defines the emergence and dynamics of CRF01_AE in Southeast Asia.

During the early phase of the HIV epidemic in Southeast Asia, CRF01_AE and subtype B were circulating relatively independently among patients who acquired infections through heterosexual contact and injecting drug use, respectively (Weniger et al., 1994). However, dual infection as a result of co-infection or super-infection by multiple HIV-1 strains within a single host was not uncommon among high-risk groups (Xin et al., 1995; Zhu et al., 1995). This presents an ideal environment for the *in vivo* "mixing" of distinct viral types, leading to cross-clade recombination (Ramos et al., 2002; Takebe et al., 2003). The recombinant virus could have had disseminated widely beyond a single host when conditions that favor viral spread arose within the drug injection or sexual transmission networks. Accordingly, our analyses indicate that recombination may be common during the establishment phase of an HIV-1 outbreak, perhaps a few years after the first founding virus appears (Fig. 3A), and, as in the case of CRF33_01B, was associated with complex networks of persons at risk in Southeast Asia in the early 1990s (Weniger et al., 1991).

In summary, we have shown that date of a CRF-generating HIV-1 recombination event can be estimated using a simple phylogenetic model. This approach could be applied to investigate the spatiotemporal history of other HIV-1 circulating recombinant forms. Furthermore, it could be extended to other recombinant virus species that contain highly prevalent recombinant lineages, including members of the retroviruses, flaviviruses, picornaviruses and alphaviruses.

Materials and methods

Study subjects, PCR and sequencing

One hundred eighty four HIV-1 infected patients who acquired infection through various risk practices (injecting drug user, hetero/bisexual, men who have sex with men) were recruited between July 2003 and August 2005 in Kuala Lumpur, Malaysia, where CRF33_01B is endemic. HIV-1 molecular subtypes/CRFs were screened using the *gag*, polymerase (*pol*) and envelope (*env*) genes from plasma or peripheral blood mononuclear cells (PBMCs). Briefly, HIV-1 RNA was extracted from plasma with High Pure Viral RNA Kit (Roche Diagnostic GmbH, Mannheim, Germany) according to the manufacturer's instructions. To amplify the *gag*, *pol* and *env* genes, reverse transcription PCR was performed using TaKaRa One Step RNA PCR Kit (AMV) (TaKaRa, Shiga, Japan) followed by nested PCR by *Premix Taq* (*Ex Taq* Version) (TaKaRa). Three overlapped regions were amplified encompassing the *gag* and *pol* genes and primers were designed to obtain sufficient number of overlapping nucleotides to ensure that recombinants are not generated by assembling sequence fragments derived from different HIV-1 subtypes amplified from an individual. PCR primers used in primary and secondary (nested) amplifications are listed in Supplementary Table 1. Primary thermocycling conditions were 1 cycle at 50 °C for 30 min (reverse transcription); 1 cycle at 94 °C for 2 min; 30 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 3 min; and 1 cycle at 72 °C for 7 min. Nested PCR in a 50 µl reaction was performed using 5 µl of amplicon generated from the first PCR. The PCR cycling conditions were similar to the primary amplification except for the single denaturation cycle, which was set at 94 °C for 1 min. Amplicons were purified and directly sequenced by an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA). For proviral DNA isolation, PBMCs from HIV-1 patients were separated by Ficoll–Hypaque density gradient centrifugation (Amersham Biosciences AB, Uppsala, Sweden). PBMCs were co-cultured with phytohemagglutinin-stimulated (1 µg/ml) CD8+ T cell-depleted PBMCs (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) from HIV-negative donors in RPMI 1640 supplemented with 10% fetal calf serum and interleukin-2 (20 U/ml). Virus production was measured by a virion-associated RT assay as described elsewhere (Kato et al., 1999). HIV-infected PBMCs were then harvested and proviral DNA was extracted with guanidine detergent for near full-length viral DNA amplification and sequencing, as described previously (Tee et al., 2006). Recombination structures of HIV-1 CRF33_01B were determined by bootscanning and confirmed by informative-sites and subregion tree analyses described elsewhere (Tee et al., 2006, 2005a). In this novel recombinant, two fragments of subtype B, each about 300 bp long, can be found in the *gag–pol* gene region. The remainder of the genome is most closely related to CRF01_AE.

Phylogenetic reconstruction and Bayesian coalescent inference

Twelve CRF33_01B sequences with known dates of sampling were selected for further phylogenetic and coalescent analysis. The CRF33_01B genome was separated into three alignments, corresponding to genome regions with different parent clades: (i) CRF01_AE$_{gag}$ (HXB2 790–2052), (ii) subtype B$_{gag–pol}$ (HXB2 2065–2374 and 2541–2840), and (iii) CRF01_AE$_{env}$ (HXB2 688−−7637). The two short subtype B fragments in the *gag–pol* region were concatenated to form a longer alignment in order to maximize phylogenetic signal. Likewise, the *gag* and *env* genes of CRF01_AE origin were also concatenated to achieve higher phylogenetic resolution (CRF01_AE$_{gag–env}$). For each sub-genomic or concatenated alignment, the CRF33_01B sequences were combined with the most closely related subtype B or CRF01_AE reference sequences (including all sequences of Asian origin) for which the date of sampling was known. Phylogenetic trees were estimated for each data set using the maximum-likelihood approach implemented in PAUP* v4.0 beta (Swofford, 2003). Reference sequences were obtained from the HIV sequence database (www.hiv.lanl.gov).

Evolutionary rates could not be estimated directly from the CRF33_01B sequences because they were not sampled over a sufficiently wide range of dates. Therefore specific rates of evolution for the different genome regions under study were estimated from independent sets of ‘serially-sampled’ subtype B and CRF01_AE sequences with known dates of sampling (retrieved from the HIV sequence database). The *gag* and *env* CRF01_AE reference sequences were sampled between 1990 and 2006 ($n = 37$). The corresponding *gag–pol* subtype B reference sequences were sampled between 1983 and 2005 ($n = 55$). Evolutionary rates were obtained using the Bayesian MCMC approach implemented in BEAST v1.4 (Drummond and Rambaut, 2007). An uncorrelated lognormal relaxed molecular clock was chosen, which assumes no *a priori* correlation between a lineage's rate of evolution and that of its ancestor (Drummond et al., 2006). During analysis, evolutionary rates and tree topologies were analyzed using the general time-reversible (GTR) (Rodriguez et al., 1990) and Hasegawa–Kishino–Yano (HKY) (Hasegawa et al., 1985) substitution models with gamma distributed among-site rate variation with four rate categories (γ_4) (Yang, 1994). Constant-sized and exponentially growing coalescent models were used in each case (Drummond et al., 2002) and each MCMC chain was run for 20–30 million states, sampled at every 10,000 states. Posterior probability densities were calculated and chains were checked for convergence in Tracer v1.4 (available from http://tree.bio.ed.ac.uk) with 10% of each chain discarded as burn-in. Estimated posterior distributions of evolutionary rates obtained from the serially-sampled data sets were incorporated as prior probability distributions in subsequent analyses in order to infer the timescale of CRF33_01B evolution (Pybus et al., 2003). The sequences reported in this paper have been deposited in the GenBank database (accession numbers EU785944-EU785961 and DQ366659-DQ366661).

Acknowledgments

The authors gratefully acknowledge Timothy D. Mastro for critical reading of the manuscript and Naoki Yamamoto for support.

The study was supported in part by the Ministry of Health, Labour and Welfare, the Ministry of Education, Science and Technology, the Japanese Foundation for AIDS Prevention (JFAP), and a Royal Society International Project Fund to O.G.P. and Y.T. K.K.T. is a recipient of the JFAP fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2009.02.020.

References

de Oliveira, T., Pybus, O.G., Rambaut, A., Salemi, M., Cassol, S., Ciccozzi, M., Rezza, G., Gattinara, G.C., D'Arrigo, R., Amicosante, M., Perrin, L., Colizzi, V., Perno, C.F., 2006. Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature 444, 836–837.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307–1320.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192.

Goh, K.L., Chua, C.T., Chiew, I.S., Soo-Hoo, T.S., 1987. The acquired immune deficiency syndrome: a report of the first case in Malaysia. Med. J. Malaysia 42, 58–60.

Hahn, C.S., Lustig, S., Strauss, E.G., Strauss, J.H., 1988. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. U. S. A. 85, 5997–6001.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.

Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20, W13–W23.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373, 123–126.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., VandePol, S., 1982. Rapid evolution of RNA genomes. Science 215, 1577–1585.

Kalish, M.L., Baldwin, A., Raktham, S., Wasi, C., Luo, C.C., Schochetman, G., Mastro, T.D., Young, N., Vanichseni, S., Rubsamen-Waigmann, H., vonBriesen, H., Mullins, J.I., Delwart, E., Herring, B., Esparza, J., Heyward, W.L., Osmanov, S., 1995. The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. AIDS 9, 851–857.

Kato, K., Sato, H., Takebe, Y., 1999. Role of naturally occurring basic amino acid substitutions in the human immunodeficiency virus type 1 subtype E envelope V3 loop on viral coreceptor usage and cell tropism. J. Virol. 73, 5520–5526.

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F., Sharp, P.M., Shaw, G.M., Peeters, M., Hahn, B.H., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313, 523–526.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S., Bhattacharya, T., 2000. Timing the ancestor of the HIV-1 pandemic strains. Science 288, 1789–1796.

Lai, M.M., 1992. RNA recombination in animal and plant viruses. Microbiol. Rev. 56, 61–79.

Lau, K.A., Wang, B., Kamarulzaman, A., Ngb, K.P., Saksena, N.K., 2008. Continuous crossover(s) events of HIV-1 CRF01_AE and B subtype strains in Malaysia: evidence of rapid and extensive HIV-1 evolution in the region. Curr. HIV Res. 6, 108–116.

Lemey, P., Pybus, O.G., Wang, B., Saksena, N.K., Salemi, M., Vandamme, A.M., 2003. Tracing the origin and history of the HIV-2 epidemic. Proc. Natl. Acad. Sci. U. S. A. 100, 6588–6592.

Ou, C.Y., Takebe, Y., Weniger, B.G., Luo, C.C., Kalish, M.L., Auwanit, W., Yamazaki, S., Gayle, H.D., Young, N.L., Schochetman, G., 1993. Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. Lancet 341, 1171–1174.

Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., Ho, D.D., 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271, 1582–1586.

Pybus, O.G., Drummond, A.J., Nakano, T., Robertson, B.H., Rambaut, A., 2003. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol. Biol. Evol. 20, 381–387.

Ramos, A., Hu, D.J., Nguyen, L., Phan, K.O., Vanichseni, S., Promadej, N., Choopanya, K., Callahan, M., Young, N.L., McNicholl, J., Mastro, T.D., Folks, T.M., Subbarao, S., 2002. Intersubtype human immunodeficiency virus type 1 superinfection following seroconversion to primary infection in two injection drug users. J. Virol. 76, 7444–7452.

Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S., Korber, B., 2000. HIV-1 nomenclature proposal. Science 288, 55–56.

Robertson, D.L., Sharp, P.M., McCutchan, F.E., Hahn, B.H., 1995. Recombination in HIV-1. Nature 374, 124–126.

Rodriguez, F., Oliver, J.L., Marin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Salemi, M., Strimmer, K., Hall, W.W., Duffy, M., Delaporte, E., Mboup, S., Peeters, M., Vandamme, A.M., 2001. Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. FASEB J. 15, 276–278.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) 4.0 beta: Sinauer Associates, Sunderland, Massachusetts.

Takebe, Y., Motomura, K., Tatsumi, M., Lwin, H.H., Zaw, M., Kusagawa, S., 2003. High prevalence of diverse forms of HIV-1 intersubtype recombinants in Central Myanmar: geographical hot spot of extensive recombination. AIDS 17, 2077–2087.

Tee, K.K., Li, X.J., Nohtomi, K., Ng, K.P., Kamarulzaman, A., Takebe, Y., 2006. Identification of a novel circulating recombinant form (CRF33_01B) disseminating widely among various risk populations in Kuala Lumpur, Malaysia. J. Acquir. Immune Defic. Syndr. 43, 523–529.

Tee, K.K., Pon, C.K., Kamarulzaman, A., Ng, K.P., 2005a. Emergence of HIV-1 CRF01_AE/B unique recombinant forms in Kuala Lumpur, Malaysia. AIDS 19, 119–126.

Tee, K.K., Pybus, O.G., Li, X.J., Han, X., Shang, H., Kamarulzaman, A., Takebe, Y., 2008. Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08_BC and 07_BC in Asia. J. Virol. 82, 9206–9215.

Tee, K.K., Saw, T.L., Pon, C.K., Kamarulzaman, A., Ng, K.P., 2005b. The evolving molecular epidemiology of HIV type 1 among injecting drug users (IDUs) in Malaysia. AIDS Res. Hum. Retroviruses 21, 1046–1050.

Tee, K.K., Takebe, Y., Kamarulzaman, A., in press. Emerging and re-emerging viruses in Malaysia, 1997–2007. Int. J. Infect. Dis.

Van Heuverswyn, F., Li, Y., Neel, C., Bailes, E., Keele, B.F., Liu, W., Loul, S., Butel, C., Liegeois, F., Bienvenue, Y., Ngolle, E.M., Sharp, P.M., Shaw, G.M., Delaporte, E., Hahn, B.H., Peeters, M., 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature 444, 164.

Wang, B., Lau, K.A., Ong, L.Y., Shah, M., Steain, M.C., Foley, B., Dwyer, D.E., Chew, C.B., Kamarulzaman, A., Ng, K.P., Saksena, N.K., 2007. Complex patterns of the HIV-1 epidemic in Kuala Lumpur, Malaysia: evidence for expansion of circulating recombinant form CRF33_01B and detection of multiple other recombinants. Virology 367, 288–297.

Weniger, B.G., Limpakarnjanarat, K., Ungchusak, K., Thanprasertsuk, S., Choopanya, K., Vanichseni, S., Uneklabh, T., Thongcharoen, P., Wasi, C., 1991. The epidemiology of HIV infection and AIDS in Thailand. AIDS 5 (Suppl. 2), S71–S85.

Weniger, B.G., Takebe, Y., Ou, C.Y., Yamazaki, S., 1994. The molecular epidemiology of HIV in Asia. AIDS 8 (Suppl. 2), S13–S28.

Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J.J., Kabongo, J.M., Kalengayi, R.M., Van Marck, E., Gilbert, M.T., Wolinsky, S.M., 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455, 661–664.

Xin, K.Q., Ma, X.H., Crandall, K.A., Bukawa, H., Ishigatsubo, Y., Kawamoto, S., Okuda, K., 1995. Dual infection with HIV-1 Thai subtype B and E. Lancet 346, 1372–1373.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39, 306–314.

Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M., Ho, D.D., 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391, 594–597.

Zhu, T., Wang, N., Carr, A., Wolinsky, S., Ho, D.D., 1995. Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconvertor. J. Virol. 69, 1324–1327.