

Molecular Epidemiology of Human Immunodeficiency Virus

Bum Sik Chin

Center for Infectious Diseases, National Medical Center, Seoul, Korea

During the evolution of human immunodeficiency virus (HIV), transmissions between humans and primates resulted in multiple HIV lineages in humans. This evolution has been rapid, giving rise to a complex classification and allowing for worldwide spread and intermixing of subtypes, which has consequently led to dozens of circulating recombinant forms. In the Republic of Korea, 12,522 cases of HIV infection have been reported between 1985, when AIDS was first identified, and 2015. This review focuses on the evolution of HIV infection worldwide and the molecular epidemiologic characteristics of HIV in Korea.

Key Words: Human immunodeficiency virus; Molecular epidemiology; Phylogeny

Introduction

Human immunodeficiency virus (HIV) is highly heterogeneous even within infected individuals, owing to rapid turnover rates, high viral load, and an error-prone reverse transcriptase enzyme, and recombination between viral genomes [1-5]. The rapid evolution and exceptional diversity of HIV facilitate the application of a variety of phylogenetic approaches to study its various aspects such as its geographical distribution and dispersion patterns. The number of newly identified Korean HIV-1-infected individuals has increased in recent years [6] and an understanding of HIV epidemiology would contribute to better management of the recent HIV epidemic in Korea.

Origin of the HIV epidemic

HIV originated from multiple zoonotic transmissions of simian immunodeficiency virus (SIV) from nonhuman primates to humans in West and Central Africa. More than 40 different nonhuman primate species harbor SIV strains, each of which exhibits species-specificity [7]. Independent zoonotic transmission events from nonhuman primates to humans have generated several HIV lineages: HIV type 1 (HIV-1) groups M, N, O, and P; and HIV type 2 (HIV-2) groups A-H. While HIV-1 group M (Main) is responsible for the global HIV pandemic [8], not all HIV lineages have been identified and new cross-species transmissions may take place in the future. HIV-1 groups M and N originate directly, but independently, from SIV found in the chimpanzee *Pan troglodytes* (SIVcpz) in

Received: February 6, 2017 **Published online:** March 22, 2017

Corresponding Author : Bum Sik Chin, MD, PhD

Center for Infectious Diseases, National Medical Center, Euljiro 6th St. 245 Jung-gu, Seoul 04564, Korea

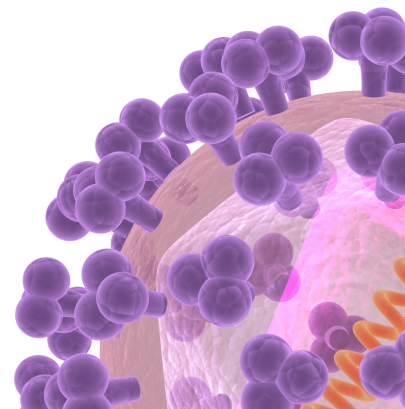
Tel: +82-2-2262-4748, Fax: +82-2-2268-0803

E-mail: moberrer@nmc.or.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyrights © 2017 by The Korean Society of Infectious Diseases | Korean Society for Chemotherapy

www.icjournal.org



West-Central Africa [9, 10]. Group N has remained confined to a small number of individuals in Cameroon, but group M spread south via the Congo River to what is now Kinshasa in the Democratic Republic of the Congo (DRC; formerly Leopoldville, Zaire), where the global HIV-1 epidemic probably began [10]. The earliest direct evidence of HIV infection in humans was found retrospectively in a serum sample and a lymph node biopsy specimen stored in 1959 and 1960, respectively, in Kinshasa, DRC [11, 12]. Molecular clock computer programs used to estimate the time since the most recent common ancestors (tMRCAs) and evolutionary rates of various HIV lineages revealed that groups M and SIVcpz shared a common ancestor in 1853 (1799–1904), and cross-species transfer is therefore inferred to have taken place around this time [13].

Global epidemiology of HIV

Group M causes the majority of HIV-1 infections, owing to its high number of subtypes and circulating recombinant forms (CRFs). Variation within group M is greatest in the Congo River basin, while is the probable site of the initial zoonotic jumps and regional diversification. Initial subtype distribution indicated dominance of subtype B in the western world and of subtype A in sub-Saharan Africa. In the past 15 years, however, the rapid emergence of new subtypes, and intermixing of strains, has altered the geographical distribution of subtypes [14, 15]. Sub-Saharan Africa bears the highest burden of HIV-1 in terms of prevalence and diversity [16]. Most, if not all, subtypes, sub-subtypes, and CRFs have been reported in the DRC and Cameroon [17, 18]. Although HIV-1 genetic diversity is high in West and Central Africa, the highest prevalence shifted in the late 1990s from East Africa (Uganda, Kenya, and Tanzania) to the southern African region. On average, close to 20% of the human population of South Africa, Lesotho, Botswana, and Zimbabwe is thought to be infected with HIV-1 [19].

This shift provides strong evidence for the founder-effect theory (a single introduction followed by a rapid spread), since the southern African epidemic is due almost entirely to the spread of HIV-1 subtype C [20]. Independently, a rapid spread of subtype C in east Asia has also contributed to this subtype being responsible for more than 50% of all HIV-1 infections worldwide [21]. The HIV epidemic in Asia is dynamic and all subtypes that circulate are due to multiple founder events. Subtype B was the first to be introduced to Asia, in the mid-1980s, and it was seen mainly in China, India, and Thai-

land. This strain has been named B' (also known as Thai B) because of its divergence from the subtype B that occurs in the Americas. Deng and colleagues [22] reported that this divergence occurred about 15 years after the B subtype began to spread, which coincides roughly with the introduction of CRF01_AE into Thailand. CRF01_AE has now gained dominance in Thailand; likewise, subtype C is currently dominant in most East Asian countries [23]. Subtype C was reportedly introduced into India from South Africa, and then into China from India [24]. In China, subtypes B and C have recombined to form CRF07_B'C and CRF08_B'C [25]. In eastern Europe, the breakdown of the former Soviet Union has coincided with an increase in the number of HIV-1 infections, and most notably, the spread of sub-subtype A1, which has been linked to intravenous drug use (IDU), and the spread of subtype B (and to a lesser extent CRF03_AB), mainly through sexual transmission. In western Europe, as in north America and Australia, subtype B predominates. However, the prevalence of non-B strains has increased owing to the influx of immigrants from Africa and Asia [26, 27]. In south America, the prevalence and diversity of HIV-1 are highest in Brazil and Argentina, with substantial circulation of subtypes B, C, and F, and BC and BF recombinants. The subtype C epidemic is now the fastest-emerging epidemic in south America [28, 29].

Circulating recombinant forms and unique recombinant forms

Analyses of multiple genome regions, and in particular full-length genome sequencing, have revealed that recombination between strains is a frequent occurrence; this has also been demonstrated between strains of different groups of HIV-1 (groups M and O), as well as between and within group M subtypes [30, 31]. If recombinants of different HIV-1 group M subtypes have been fully sequenced and are found in three or more epidemiologically unlinked individuals, they are defined as circulating CRFs. Alternatively, they may be defined as unique recombinant forms (URFs) if they do not meet these criteria. Newly discovered CRFs are named in the order in which they are discovered and described; thus, their numbering does not reflect their evolutionary history or the chronological order in which they were formed. CRF01_AE is the result of a recombination between subtypes A and E, and CRF02_AG is a recombinant of subtypes A and G. While no 'pure' subtype E lineage has been found, CRF01_AE is probably the result of a recombination event early on in the epidem-

ic after which the putative parental 'pure' subtype E became extinct. The extension 'cpx', representing complex, is given if the CRF consists of contributions from three or more different subtypes (e.g., CRF04_cpx is composed of A, G, H, K, and U). As of January 2017, 88 CRFs have been registered by the Los Alamos National Laboratory [32].

HIV genetic diversity and dual infection

Recombination is one of the reasons of high variability of HIV, which involves shuttling of mutations between viral genomes and leads to major antigenic shifts or alterations in virulence [33-35]. One of the most important factors in the worldwide spread of HIV is this enormous genetic variability and rapid evolution. In particular, viral sequences within a single individual can differ genetically by up to 10% [36]; however, productive infections are initiated by a single infectious unit (a single virus or infected cell), and the remainder are the result of multiple variants. The mucosal barrier in inflammatory genital infections is thought to be important for this genetic bottleneck [37]. This reduction in viral diversity from a diverse swarm in the donor to a very limited number of variants in the recipient is dependent on the transmission route. For example, approximately 80% of productive infections are caused by a single virus in heterosexual transmission, but only 60% in men who have sex with men (MSM) and 40% in IDU [38, 39].

Within a subtype of the HIV-1 group M (subtypes A-D, F-H, J, and K), variation at the amino acid level is in the order of 8-17%, but can be as high as 30%. Variation between subtypes is usually between 17% and 35%, but can be up to 42% depending on the subtypes and genome regions examined [36]. Further, dual infection can increase the viral genetic diversity within an infected person [40]. Dual infection can be due to simultaneous infection with two heterologous strains (co-infection) or to sequential infection, in which a second infection with a heterologous strain occurs after seroconversion to the initial infection (superinfection). Dual infection is detected in 10-20% of HIV-infected individuals in regions of Africa where different variants co-circulate. Moreover, dual infection may be associated with a higher viral load setpoint and more rapid disease progression [41, 42].

Phylogenetic analysis of HIV

As mentioned previously, phylogenetic analyses can bring

to light dimensions of HIV evolution, such as the "where," "when," and "how" of its spread across the globe, which are impossible to assess using other approaches [13, 43, 44]. In addition, phylogenetic analyses of HIV sequences have been used to investigate a variety of problems, such as potential transmission of the HIV virus among individuals [45], cross-species transmissions [9], origins [10, 13], subtyping [46], and drug resistance [47]. The geographical distribution and dispersion patterns of HIV have been inferred using phylogenetic analysis [48, 49]. However, if the dataset is enlarged, the number of cases increases exponentially. For example, the number of possible trees that can be generated using 100 sequences is 1.700459×10^{82} [50]. Therefore, it is necessary to generate a representative set of trees rather than finding a single best answer. If 95% of the trees contain a particular feature, the confidence of this feature is 95%. Resampling techniques, such as bootstrapping [51], are used to estimate the variance of a statistic for which the underlying distribution is either unknown or difficult to derive analytically. Each new sample obtained by resampling is called a "pseudosample." From each pseudosample a new tree is estimated, and the number of times that a specific internal branch appears in the whole set of trees is recorded as the bootstrap proportion of that branch [52, 53].

Transmission network analysis of HIV

The origin and geographic expansion of HIV-1 have been well characterized using phylogenetic approaches [43], but these methods are suboptimal for describing recent HIV transmission. Phylogenies are well suited for differentiating distinct viral lineages, but not for identifying transmission partners. It is generally accepted that phylogenetic analysis is the most powerful tool for excluding potential transmission partners, rather than establishing linkages [54]. Conversely, transmission networks focus on similarity and directly link genetically similar viruses.

Regarding HIV transmission clusters, investigation typically begins by inferring a phylogeny and then identifying those clades that have the appropriate statistical support. The branching structure within the phylogeny is irrelevant after these clades have been identified. A conceptual problem arises as these clades are not easily resolved into transmission pairs and clusters, and within a transmission cluster all individuals are treated as equally related/connected; the inferred branching structure within these clusters is again discarded,

wasting the considerable computational effort expended to infer the complete phylogeny [55] (Fig. 1). Furthermore, high statistical support (e.g., bootstrap) for any specific clade does not indicate that the members of the clade itself are necessarily closely related to each other [52].

The transmission network is derived from a matrix of pairwise distances between each pair of nodes or sequences in a dataset, and a network can be created much more rapidly than a tree for large numbers of sequences because no attempt is made to take into account the patterns of shared ancestry of the individual sequences. The distribution of cluster sizes in an HIV-1 population sequence is usually found to be heterogeneous in network analysis, consistent with different rates of transmission for different subpopulations or individuals [56, 57]. This heterogeneity suggests that interventions to prevent transmission should be targeted towards those with the highest risk of transmitting the virus to others [58, 59]. Not surprisingly, studies in which sequence data are combined with behavioral risk factors have shown that sequences from individuals with the same risk factors are likely to cluster with one another [56].

Phylogenetic studies among Korean HIV-infected individuals

The first report regarding HIV phylogenetics in Korea was published in 1998 [60], 13 years after the first HIV case was identified. This study analyzed the *nef* gene in 46 Korean HIV-1-infected individuals; 35 of the 41 Korean subtype B isolates formed a distinct monophyletic clade that was not related to any of the sequences reported from other countries available in the Los Alamos Database or GenBank. Subsequently, several studies consistently reported this unique Korean clade B (KCB), analyzing various genes such as *nef* [60, 61], *env* [62, 63], *vif* [61], and *pol* [64, 65]. In 2012, Kim et al. phylogenetically analyzed 560 Korean and 2,347 global subtype B *env* sequences; 491 (88%) of the 560 Korean subtype B *env* sequences belonged to the distinct KCB (75% bootstrap value), which was differentiated from those of other countries [66] (Fig. 2). In this report, the mean tMRCA was estimated to be around 1961 and 1967 for subtype B and KCB, respectively. A Bayesian skyline plot estimated that the effective number of infections increased rapidly until the early 1980s, followed by a slower increase until the mid-1990s, thereafter reaching a steady-state. The mean growth rates were estimated to be 0.15 and 0.17 per year for subtypes B and KCB, respectively, which were approx-

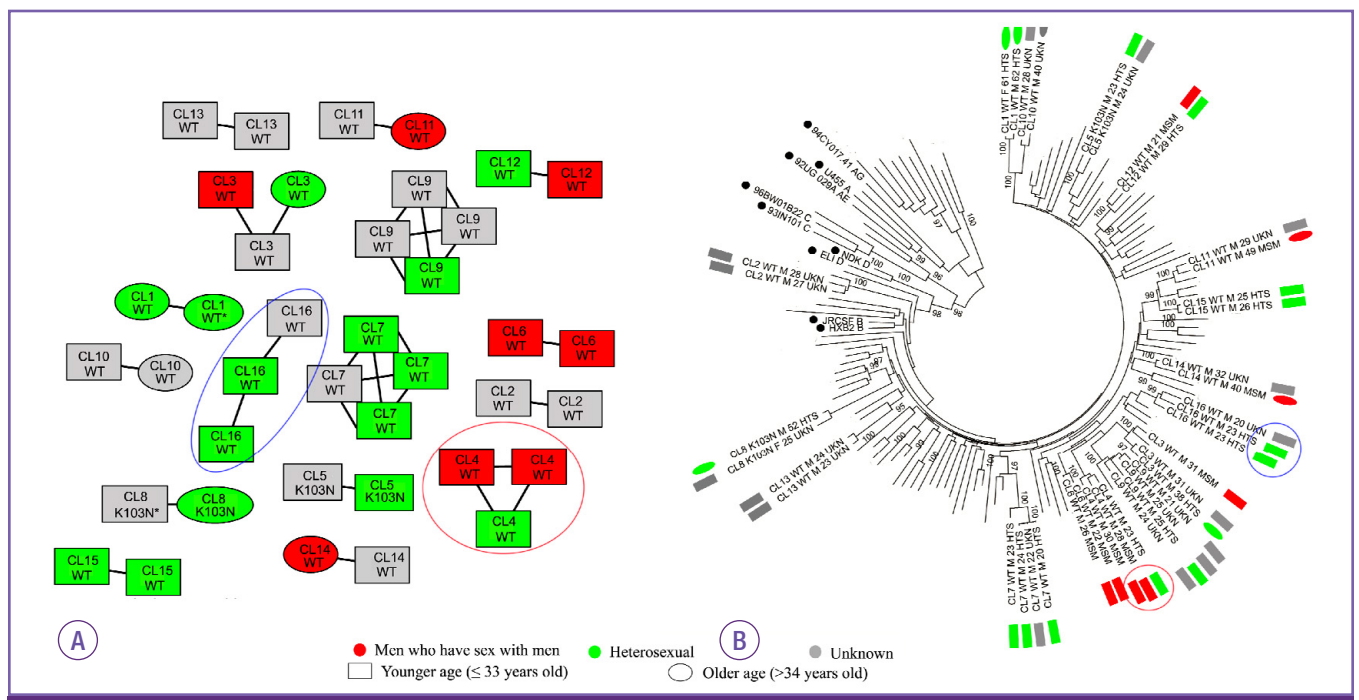


Figure 1. Comparison of network analysis (A) and phylogenetic tree (B). While the level of interconnection in the cluster is indistinguishable in the phylogenetic tree, network analysis reveals the level of connections. In cluster 4 (red circle), every node is connected with the other nodes while only one node is interconnected with the other two nodes in cluster 16 (blue circle) [74].

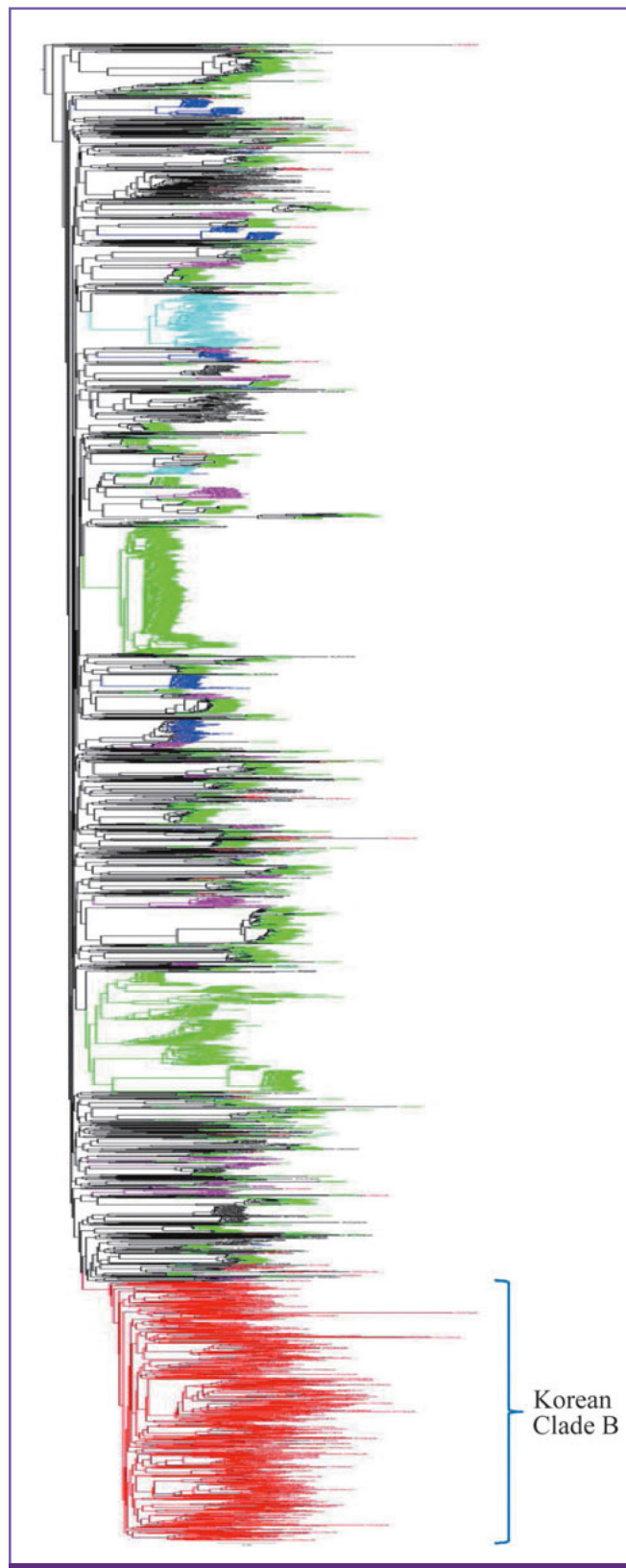


Figure 2. Neighbor-joining phylogenetic relationship generated from 2,907 HIV-1 subtype B gp120 *env* sequences. Korean clade B (in red) formed a distinct cluster from worldwide HIV-1 subtype B sequences [66].

imately three- to five-fold lower than those reported for other countries [49, 67]. This relatively slow growth rate could be explained by the demographic factors of HIV acquisition in Korea, in which heterosexual contact was the most common cause of transmission (51.9%), followed by homosexual contact (39.2%), while few patients described themselves as IDUs so far [6]. However, many HIV-infected men may not have disclosed their true sexual risk considering that the male-to-female ratio among newly reported HIV infected individuals is greater than 20:1, and a substantial proportion (36.0%) refused to refer to their risk factor for HIV acquisition [6].

Phylogenetic analyses were also utilized for epidemiologic evaluation of HIV-infected Korean hemophilia patients [68, 69], and investigation of HIV-1-transmitted drug resistance (TDR) in Korea [64, 68, 70-72]. Studies on TDR revealed that the drug resistance mutation was not common among treatment-naïve HIV-infected individuals in Korea, and there was little evidence to suggest that TDR in Korea is related to the clonal spread of drug-resistant strains. By employing the transmission network analysis, limited genetic variation in the *env* gene was observed within subtype B strains from Korean MSMs [73], and heterosexual exposure as well as younger age at HIV-1 diagnosis were associated with cluster formation among recently diagnosed Korean HIV-infected individuals [74].

Summary

The origin of HIV and its epidemics are beginning to be understood by molecular epidemiology utilizing the characteristics of HIV, which include rapid turnover rates, high viral load, and error-prone reverse transcriptase enzymes. While most subtypes and CRFs have been reported in sub-Saharan Africa (such as in the DRC and Cameroon, where HIV was first introduced to humans), the heterogeneity of its worldwide distribution can be explained by the founder effect of a single introduction followed by rapid spread. In Korea, the unique monophyletic subtype B has predominated throughout the HIV epidemic. While low growth rates of the HIV-infected population were suggested to explain KCB formation, a recent increase in the number of newly diagnosed HIV-infected individuals suggests the necessity of molecular investigation of the HIV epidemic in this area.

Acknowledgements

I thank Professor Davey M. Smith of the University of California, San Diego for his comprehensive advice during the manuscript preparation. This study was supported by the intramural fund of the Department of Education and Research, National Medical Center, Seoul, Republic of Korea.

Conflicts of Interest

No conflicts of interest.

ORCID

Bum Sik Chin <http://orcid.org/0000-0003-3021-1434>

References

- Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 1995;373:123-6.
- Preston B, Poiesz B, Loeb L. Fidelity of HIV-1 reverse transcriptase. *Science* 1988;242:1168-71.
- Preston BD, Dougherty JP. Mechanisms of retroviral mutation. *Trends Microbiol* 1996;4:16-21.
- Roberts J, Bebenek K, Kunkel T. The accuracy of reverse transcriptase from HIV-1. *Science* 1988;242:1171-3.
- Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH, Saag MS, Shaw GM. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 1995;373:117-22.
- Korea Centers for Disease Control and Prevention. Annual report on the notified HIV/AIDS in Korea. 2015. Available at: <http://cdc.go.kr/CDC/info/CdcKrInfo0128.jsp?menuIds=HOME001-MNU1130-MNU1156-MNU1426-MNU1448>. Accessed 10 January 2017.
- Aghokeng AF, Ayoub A, Mpoudi-Ngole E, Loul S, Liegeois F, Delaporte E, Peeters M. Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provide insights into risks for potential new cross-species transmissions. *Infect Genet Evol* 2010;10:386-96.
- Los Alamos National Laboratory. Distribution of all HIV-1 sequences: world. Available at: <https://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp>. Accessed 1 February 2017.
- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* 1999;397:436-41.
- Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen Y, Wain LV, Liegeois F, Loul S, Ngole EM, Bienvenue Y, Delaporte E, Brookfield JF, Sharp PM, Shaw GM, Peeters M, Hahn BH. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 2006;313:523-6.
- Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 1998;391:594-7.
- Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, Muyembe JJ, Kabongo JM, Kalengayi RM, Van Marck E, Gilbert MT, Wolinsky SM. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008;455:661-4.
- Wertheim JO, Worobey M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Comput Biol* 2009;5:e1000377.
- Robertson DL, Hahn BH, Sharp PM. Recombination in AIDS viruses. *J Mol Evol* 1995;40:249-59.
- Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C, Learn GH, Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, Korber B. HIV-1 nomenclature proposal. *Science* 2000;288:55-6.
- UNAIDS. Global AIDS update 2016. Available at: <http://www.unaids.org/en/resources/documents/2016/Global-AIDS-update-2016>. Accessed 10 February 2017.
- Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W, Sema H, Tshimanga K, Bongo B, Delaporte E. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol* 2000;74:10498-507.
- Triques K, Bourgeois A, Saragosti S, Vidal N, Mpoudi-Ngole E, Nzilambi N, Apetrei C, Ekwalanga M, Delaporte E, Peeters M. High diversity of HIV-1 subtype F strains in Central Africa. *Virology* 1999;259:99-109.
- UNAIDS. AIDSinfo HIV Prevalence. Available at: <http://aidsinfo.unaids.org>. Accessed 21 February 2017.
- Rambaut A, Robertson DL, Pybus OG, Peeters M, Holmes EC. Human immunodeficiency virus: Phylogeny and the origin of HIV-1. *Nature* 2001;410:1047-8.
- Ghys PD, Sadel T, Vu HT, Savtchenko I, Erasilova I, Masho-

- logu YS, Indongo R, Sikhosana N, Walker N. Growing in silence: selected regions and countries with expanding HIV/AIDS epidemics. *AIDS* 2003;17 (Suppl 4):S45-50.
22. Deng X, Liu H, Shao Y, Rayner S, Yang R. The epidemic origin and molecular properties of B': a founder strain of the HIV-1 transmission in Asia. *AIDS* 2008;22:1851-8.
23. de Silva UC, Warachit J, Sattagowit N, Jirapongwattana C, Panthong S, Utachee P, Yasunaga T, Ikuta K, Kameoka M, Boonsathorn N. Genotypic characterization of HIV type 1 env gp160 sequences from three regions in Thailand. *AIDS Res Hum Retroviruses* 2010;26:223-7.
24. Rodenburg CM, Li Y, Trask SA, Chen Y, Decker J, Robertson DL, Kalish ML, Shaw GM, Allen S, Hahn BH, Gao F; UN-AIDS and NIAID Networks for HIV Isolation and Characterization. Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. *AIDS Res Hum Retroviruses* 2001;17:161-8.
25. Piyasirisilp S, McCutchan FE, Carr JK, Sanders-Buell E, Liu W, Chen J, Wagner R, Wolf H, Shao Y, Lai S, Beyrer C, Yu XF. A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. *J Virol* 2000;74:11286-95.
26. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 2006;20:W13-23.
27. Pérez-Alvarez L, Muñoz M, Delgado E, Miralles C, Ocampo A, García V, Thomson M, Contreras G, Nájera R; Spanish Group for Antiretroviral Resistance Studies in Galicia. Isolation and biological characterization of HIV-1 BG inter-subtype recombinants and other genetic forms circulating in Galicia, Spain. *J Med Virol* 2006;78:1520-8.
28. Soares EA, Martínez AM, Souza TM, Santos AF, Da Hora V, Silveira J, Bastos FI, Tanuri A, Soares MA. HIV-1 subtype C dissemination in southern Brazil. *AIDS* 2005;19 (Suppl 4):S81-6.
29. Rodrigues R, Manenti S, Romão PR, de Paula Ferreira JL, Batista JP, Siqueira AF, de Macedo Brigido LF. Young pregnant women living with HIV/AIDS in Criciúma, Southern Brazil, are infected almost exclusively with HIV type 1 clade C. *AIDS Res Hum Retroviruses* 2010;26:351-7.
30. Peeters M, Liegeois E, Torimiro N, Bourgeois A, Mpoudi E, Vergne L, Saman E, Delaporte E, Saragosti S. Characterization of a highly replicative intergroup M/O human immunodeficiency virus Type 1 recombinant isolated from a Cameroonian patient. *J Virol* 1999;73:7368-75.
31. Rousseau CM, Learn GH, Bhattacharya T, Nickle DC, Heckerman D, Chetty S, Brander C, Goulder PJ, Walker BD, Kiepiela P, Korber BT, Mullins JL. Extensive intrasubtype recombination in South African human immunodeficiency virus type 1 subtype C infections. *J Virol* 2007;81:4492-500.
32. Los Alamos National Laboratory. HIV circulating recombinant forms (CRFs). Available at: <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>. Accessed 10 February 2017.
33. Hu WS, Temin HM. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci USA* 1990;87:1556-60.
34. Temin HM. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci USA* 1993;90:6900-3.
35. Burke DS. Recombination in HIV: an important viral evolutionary strategy. *Emerg Infect Dis* 1997;3:253-9.
36. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull* 2001;58:19-42.
37. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, Manigart O, Mulenga J, Keele BF, Shaw GM, Hahn BH, Allen SA, Derdeyn CA, Hunter E. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathogens* 2009;5:e1000274.
38. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, Salazar-Gonzalez JF, Salazar MG, Learn GH, Morgan CJ, Schumacher JE, Hraber P, Giorgi EE, Bhattacharya T, Korber BT, Perelson AS, Eron JJ, Cohen MS, Hicks CB, Haynes BF, Markowitz M, Keele BF, Hahn BH, Shaw GM. High multiplicity infection by HIV-1 in men who have sex with men. *PLoS Pathogens* 2010;6:e1000890.
39. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, Sun C, Wang S, Learn GH, Morgan CJ, Schumacher JE, Haynes BF, Keele BF, Hahn BH, Shaw GM. Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J Virol* 2010;84:6241-7.
40. Powell RL, Urbanski MM, Burda S, Kinge T, Nyambi PN. High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West Central Africa. *J Acquir Immune Defic Syndr* 2009;50:84-92.
41. Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Grobler J, Li F, Liu SL, Rademeyer C, Learn GH, Karim SS, Williamson C, Corey L, Margolick JB, Mullins JL. Dual HIV-1 infection

- associated with rapid disease progression. *Lancet* 2004;363:619-22.
42. Grobler J, Gray CM, Rademeyer C, Seoighe C, Ramjee G, Karim SA, Morris L, Williamson C. Incidence of HIV-1 dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C-infected female sex workers. *J Infect Dis* 2004;190:1355-9.
 43. Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* 2011;11:45-56.
 44. Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* 2012;18:182-92.
 45. Krushkal J, Li WH. Use of phylogenetic inference to test an HIV transmission hypothesis, In: Crandall KA, ed. *The Evolution of HIV*. Baltimore, MD: The Johns Hopkins University Press; 1999.
 46. Simon F, Maucière P, Roques P, Loussert-Ajaka I, Müller-Trutwin MC, Saragosti S, Georges-Courbot MC, Barré-Sinoussi F, Brun-Vézinet F. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* 1998;4:1032-7.
 47. Hattori J, Shiino T, Gatanaga H, Mori H, Minami R, Uchida K, Sadamasu K, Kondo M, Sugiura W; Japanese Drug Resistance HIV-1 Surveillance Network. Characteristics of transmitted drug-resistant HIV-1 in recently infected treatment-naive patients in Japan. *J Acquir Immune Defic Syndr* 2016;71:367-73.
 48. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM. Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci USA* 2003;100:6588-92.
 49. Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, Brown TM, Salemi M, Vandamme AM, Kalish ML. U.S. human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J Virol* 2003;77:6359-66.
 50. Posada D, Crandall KA, Hillis DM. Phylogenetics of HIV. In: Rodrigo AG, Learn Jr GH., eds. *Computational Analysis of HIV Molecular Sequences*. Dordrecht, Netherlands: Kluwer Academic Publishers; 2001;121-60.
 51. Efron B, Tibshirani RJ. *An Introduction to the bootstrap*. New York: Chapman & Hall; 1993.
 52. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39:783-91.
 53. Felsenstein J. Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* 1988;22:521-65.
 54. Pillay D, Rambaut A, Geretti AM, Brown AJ. HIV phylogenetics. *BMJ* 2007;335:460-1.
 55. Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, Kosakovsky Pond SL. The global transmission network of HIV-1. *J Infect Dis* 2014;209:304-13.
 56. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 2008;5:e50.
 57. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT; UK HIV Drug Resistance Collaboration. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. *J Infect Dis* 2011;204:1463-9.
 58. Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, May S, Smith DM. Using HIV networks to inform real time prevention interventions. *PLoS One* 2014;9:e98443.
 59. Wang X, Wu Y, Mao L, Xia W, Zhang W, Dai L, Mehta SR, Wertheim JO, Dong X, Zhang T, Wu H, Smith DM. Targeting HIV prevention based on molecular epidemiology among deeply sampled subnetworks of men who have sex with men. *Clin Infect Dis* 2015;61:1462-8.
 60. Kang MR, Cho YK, Chun J, Kim YB, Lee I, Lee HJ, Kim SH, Kim YK, Yoon K, Yang JM, Kim JM, Shin YO, Kang C, Lee JS, Choi KW, Kim DG, Fitch WM, Kim S. Phylogenetic analysis of the nef gene reveals a distinctive monophyletic clade in Korean HIV-1 cases. *J Acquir Immune Defic Syndr Hum Retrovirol* 1998;17:58-68.
 61. Park CS, Kim MS, Lee SD, Kim SS, Lee KM, Lee CH. Molecular phylogenetic analysis of HIV-1 vif gene from Korean isolates. *J Microbiol* 2006;44:655-9.
 62. Kim YB, Cho YK, Lee HJ, Kim CK, Kim YK, Yang JM. Molecular phylogenetic analysis of human immunodeficiency virus type 1 strains obtained from Korean patients: env gene sequences. *AIDS Res Hum Retroviruses* 1999;15:303-7.
 63. Daniels RS, Kang C, Patel D, Xiang Z, Douglas NW, Zheng NN, Cho HW, Lee JS. An HIV type 1 subtype B founder effect in Korea: gp160 signature patterns infer circulation of CTL-escape strains at the population level. *AIDS Res Hum Retroviruses* 2003;19:631-41.
 64. Sung H, Foley BT, Bae IG, Chi HS, Cho YK. Phylogenetic analysis of reverse transcriptase in antiretroviral drug-naive Korean HIV type 1 patients. *AIDS Res Hum Retroviruses* 2001;17:1549-54.
 65. Sung H, Foley BT, Ahn SH, Kim YB, Chae JD, Shin YO, Kang HI, Cho YK. Natural polymorphisms of protease in protease inhibitor-naive HIV-1 infected patients in Korea: a novel L63M in subtype B. *AIDS Res Hum Retroviruses* 2003;19:525-30.

66. Kim GJ, Yun MR, Koo MJ, Shin BG, Lee JS, Kim SS. Estimating the origin and evolution characteristics for Korean HIV type 1 subtype B using Bayesian phylogenetic analysis. *AIDS Res Hum Retroviruses* 2012;28:880-4.
67. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL, Teixeira SL, Morgado MG. Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol* 2007;7:263-70.
68. Cho YK, Jung YS, Foley BT. Phylogenetic analysis of full-length pol gene from Korean hemophiliacs and plasma donors infected with Korean subclade B of HIV type 1. *AIDS Res Hum Retroviruses* 2011;27:613-21.
69. Cho YK, Kim JE, Foley BT. Phylogenetic analysis of the earliest nef gene from hemophiliacs and local controls in Korea. *Biores Open Access* 2012;1:41-9.
70. Choi JY, Kim EJ, Park YK, Lee JS, Kim SS. National survey for drug-resistant variants in newly diagnosed antiretroviral drug-naive patients with HIV/AIDS in South Korea: 1999-2005. *J Acquir Immune Defic Syndr* 2008;49:237-42.
71. Chin BS, Choi JY, Han Y, Kuang J, Li Y, Han SH, Choi H, Chae YT, Jin SJ, Baek JH, Lim YS, Kim CO, Song YG, Yong D, Li T, Kim JM. Comparison of genotypic resistance mutations in treatment-naive HIV type 1-infected patients in Korea and China. *AIDS Res Hum Retroviruses* 2010;26:217-21.
72. Chin BS, Shin HS, Kim G, Wagner GA, Gianella S, Smith DM. Short communication: increase of HIV-1 K103N transmitted drug resistance and its association with efavirenz use in South Korea. *AIDS Res Hum Retroviruses* 2015;31:603-7.
73. Kim GJ, Nam JG, Shin BG, Kee MK, Kim EJ, Lee JS, Kim SS. National survey of prevalent HIV strains: limited genetic variation of Korean HIV-1 clade B within the population of Korean men who have sex with men. *J Acquir Immune Defic Syndr* 2008;48:127-32.
74. Chin BS, Chaillon A, Mehta SR, Wertheim JO, Kim G, Shin HS, Smith DM. Molecular epidemiology identifies HIV transmission networks associated with younger age and heterosexual exposure among Korean individuals. *J Med Virol* 2016;88:1832-5.